

Computational Studies of Noncovalent Interactions in Understanding and Designing New Systems of Biological and Chemical Significance

Thesis Submitted to AcSIR For the Award of
the Degree of
DOCTOR OF PHILOSOPHY
In Chemical Sciences



By
Mrityunjay Kumar Tiwari
(10CC11A26035)

Under the guidance of
Dr. Kumar Vanka

Physical and Material Chemistry Division,
CSIR-National Chemical Laboratory, Pune (India)
Pune - 411008, India

May 2018



CERTIFICATE

This is to certify that the work incorporated in this Ph.D. thesis entitled “**Computational Studies of Noncovalent Interactions in Understanding and Designing New Systems of Biological and Chemical Significance**” submitted by **Mrityunjay Kumar Tiwari** to **Academy of Scientific and Innovative Research (AcSIR)** in fulfillment of the requirements for the award of the Degree of **Doctor of Philosophy in Chemical Sciences**, embodies original research work under my supervision. I further certify that this work has not been submitted to any other University or Institution in part or full for the award of any degree or diploma. Research materials obtained from other sources have been duly acknowledged in the thesis. Any text, illustration, table etc., used in the thesis from other sources, have been duly cited and acknowledged.

Mrityunjay Kumar Tiwari

Mrityunjay Kumar Tiwari
(Research Student)
(Reg. No. **10CC11A26035**)

Kumar Vanka

Dr. Kumar Vanka
(Research Supervisor)

Date: 11/5/2018

Place: CSIR-NCL, Pune

DECLARATION BY THE CANDIDATE

I hereby declare that this thesis entitled “**Computational Studies of Noncovalent Interactions in Understanding and Designing New Systems of Biological and Chemical Significance**” submitted for the Degree of **Doctor of Philosophy in Chemical Sciences** to the **Academy of Scientific and Innovative Research (AcSIR)** is the outcome of theoretical and computational investigations carried out by me under the supervision of **Dr. Kumar Vanka**, Scientist, Physical and Material Chemistry Division, CSIR-National Chemical Laboratory, Pune. I further affirm that the work incorporated is original and has not been submitted to any other academy, university or institute for the award of any other degree or diploma.



Date: 11/05/2018

Mrityunjay Kumar Tiwari

Place: CSIR-NCL, Pune-411008.

Dedicated to

*To My Eldest Sister “Srimati Manorama
Kumari”*

Acknowledgement

No one knows what destiny a drop of water will fall to while it leaves the cloud. Whether it will get evaporated even before reaching the ground or will meet an ember or a dust mound to lose its existence. Whether it will fall on a petal to shine like a pearl or will fall on the rosy cheeks of a beautiful maiden to ire her Romeo. Whether it will fall on a seed to help that start a new life or will get into the mouth of a thirsty bird to quench his thirst. It is the speed and direction of the air and the temperature in the region, which determine the fate of the drop. Similarly, I knew nothing about what will I meet in my academic career when I left my home for the first time after matriculation for higher studies. I had never thought that I would be writing my Ph.D. thesis at some point of time. Since this thing has come to a reality, I must acknowledge the people who have been instrumental (directly or indirectly) in bringing me to this stage.

First and foremost, I would like to express my gratitude to my Ph.D. supervisor, Dr. Kumar Vanka (CSIR-NCL) for his support, without which it would not have been possible for me to write this thesis. Especially, his teaching skills and widespread of knowledge had been inspirational to me during my entire stay with his research group. His comprehension and explanations to the subject had been so clear to me that I never needed to refer any references for further clarification for whatever he had taught to me.

I express my sincere thanks to my present and past Doctoral Advisory Committee members: Dr. Nayana Vaval, Dr. Durba Sengupta, Dr. Mugdha Gadgil, Dr. Neelanjana Sengupta, Dr. Sudip Roy and Dr. Rajnish Kumar. I am also grateful to Dr. Ashwini Nangia, the Director, CSIR-NCL, and Dr. Sourav Pal and Dr. Vijayamohan K. Pillai, former directors of CSIR-NCL. I would also like to thank the present and the past heads of the Physical and Materials Chemistry Division for their support during my Ph.D.

Knowledge is indeed the most precious asset for a human life. Therefore, I thank all my teachers who had taught me during my Ph.D. studies at CSIR-NCL: Dr. Kumar Vanka, Dr. Sourav Pal, Dr. Darbha Srinivas, Dr. Sayam Sengupta, Dr. Nayana Vaval, Dr. Durba Sengupta, Dr. Mugdha Gadgil and Dr. Kavita Joshi. I had been thinking this for a long period that quantum chemistry could not be taught because no one has understood this subject properly. However, I have been forced to change my opinion after Dr. Pal's quantum chemistry classes at CSIR-NCL. I am thankful to him for introducing me to this subject. Further, I am highly obliged to my college and school teachers for their contributions in making me eligible to carry forward scientific research. I wish to thank the scientific fraternity and ancient Indian sages whose creations and contributions have percolated down to my brain cells. No words can describe the admiration and the respect I bear for my elder brother, Dr. Dhananjay Kumar Tiwari. He had been a true Guru to me. His teachings have been precious for my entire existence on this planet. I also thank Prof.

Sitaramam for whatever the small things I have learned from him although they have come at a much greater cost.

I would like to acknowledge my research collaborators, Dr. Sayam Sengupta, Kundan, Dr. Neelanjana Sengupta and Asis for their trust and belief in me. I would also like to thank Dr. Pradip Kumar Tripathi and Dr. Asish K. Bhattacharya for their moral support during my Ph.D.

I would like to extend my sincere thanks to all the support staffs of CSIR-NCL, AcSIR, and CSIR Delhi office for every small pain they have taken in my favor. I feel highly indebted to all the mess and canteen staffs who had fed me during my Ph.D.

It is very difficult to find friends of your own mindset, particularly when you got the honesty, daring, and truthfulness to the level of sickness. I was fortunate enough to find as many as three people at CSIR-NCL with same ailments that had nailed me down. And, developed a bond with them that has the strength to endure till the end of my life: Dr. Mudit Dixit (the discoverer of the T-virus), Dr. Kundan K. Singh (the Pahalwan), Dr. Brijesh M. Sharma (the Sankatmochan). The company of Dr. Sunil Rai has been just amazing near the end of my Ph.D.

My stay at CSIR-NCL campus has been made pleasant by the research scholars from many other research groups. A few people who had been always kind to me are: Pushkar, Sumantra da, Krunal, Prakash, Kshirodra, Mishraji, Guptaaji, Jha Ji, Arun, Ketan, Chand Bhai, Deepak Kumar, Deepak Chand, Akshay, Pragati, Deepak Kumar, Yashpal, Manoj Sharma, Amarnath, Achintya, Turbashu, Ravi, Anagha, Rima and Susanta. I thank all of them from the bottom of my heart, in particular to Deepak Kumar and Pushkar.

I express my heartfelt thanks to my labmates: Tamal, Vipin, Jugal, Subhrashish, Siddharth, Shailaja, Ruchi, Shantanu, Yuvraj and Anagh. I give my special thanks to those who extended their helping hands to me when it was needed.

At last but the most, I thank my family members: my parents, my brothers and sisters and their kids and spouses. I feel extremely sorry that I could not spend sufficient time with them since past many years. I am highly obliged to my wife, Anuradha, for her unconditional supports during my Ph.D. She had suffered every good and bad that had happened to me since we got married. In all, no words are enough to describe the affection and assistance I have received from my family during my Ph.D., especially from my eldest sister, Jijaji, Kanchan and Akash. I am grateful to the God for giving me such a lovable, supportive and understanding family.

Above all, I extend my gratitude to the Almighty for blessing me with this wonderful life with its unique challenges and protecting my all the loving once.

— **Mrityunjay Kumar Tiwari**

न त्वहं कामये राज्यं न स्वर्गं ना पुनर्भवं।
कामये दुःखतप्तानां प्राणिनामार्तिनाशनं॥

(श्रीमदभागवत पुराण)

न तो मुझे राज्य कामना है, ना ही स्वर्ग की। और ना ही मैं जन्म के बंधनो से मुक्ति ही चाहता हूँ। दुःख-संतप्त प्राणियों की पीड़ा हरने में सहायक बन सकूँ, यही मेरी कामना है।

(An English translation by Dr. Shanti Swaroop Bhatnagar)

I do not for a royal realm aspire,
For release or for paradise,
To serve those bent with grief I desire,
And calm their sorrows and help them rise.

Contents

Abbreviations	iv
Physical Constants	v
Synopsis	vi
1 A Brief Introduction of Noncovalent Chemistry	1
Abstract	1
1.1 Introduction	1
1.2 A link to the past	3
1.3 Types of noncovalent interactions	4
1.4 Cooperativity in noncovalent interactions	5
1.5 Schematic representation of noncovalent interactions	7
1.6 Some general nomenclatures	7
1.7 Quantum Chemical Methods for Noncovalent Interactions	8
1.8 Electrostatic Dominated Noncovalent Interactions	12
1.8.1 Hydrogen bonding	12
1.8.2 Directional noncovalent interactions	13
1.8.3 Literature Precedence	15
1.9 Objective of the Thesis	21
1.10 Scope of the thesis	21
1.11 References	23
2 The Fundamentals of Density Functional Theory	30
Abstract	30
2.1 Introduction	30
2.2 Elementary Quantum Mechanics	31
2.2.1 The Schrödinger Equation	31
2.2.2 The Born-Oppenheimer approximation	32
2.2.3 Wave Function, Probability and Normalization	33
2.2.4 The Variational Principle	34
2.2.5 Functional	36
2.3 The Fundamentals of Density Functional Theory	37

2.3.1	The Electron Density	38
2.3.2	The Pair Density	39
2.3.3	The Thomas-Fermi Model	40
2.3.4	Hohenberg-Kohn (HK) Theorems	41
2.3.5	The Kohn-Sham (KS) approach	48
2.4	References	53
3	The Directional Nature of Long Range Electrostatic Interactions	55
	Abstract	55
3.1	Introduction	55
3.2	Computational Details and Background Theory	59
3.3	Results and Discussion	62
3.4	Scope of the Work	72
3.5	Conclusions	72
3.6	References	73
4	Exploiting Long Range Secondary Electrostatic Forces for Regulating Electrostatics Dominated Noncovalent Interactions	77
	Abstract	77
4.1	Introduction	77
4.2	Computational Details	79
4.3	Results and Discussion	82
4.4	Conclusions	101
4.5	References	102
5	Influence of the Explicit Presence of Solvent Molecules on Hydrogen Bond Strength	106
	Abstract	106
5.1	Introduction	106
5.2	Computational Details	109
5.3	Results and Discussion	114
5.4	Association constant of hydrogen-bonded complexes	132
5.5	Conclusions	134

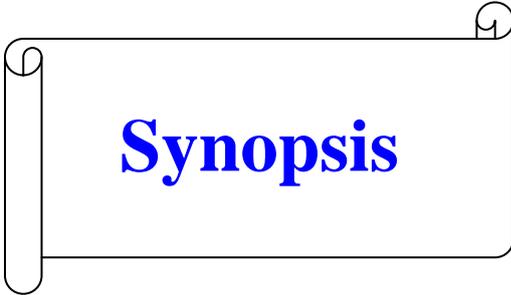
5.6	References	135
6	Does Electrostatic Field Created by Solvent Molecules Influence Reaction Barriers?	140
	Abstract	140
6.1	Introduction	140
6.2	Computational Details	145
6.3	Theoretical Background	147
	6.3.1 Field-equivalent-solvent	147
	6.3.2 Barrier in field-equivalent-solvent	152
6.4	Results and Discussion	155
6.5	Conclusions	162
6.6	References	163
7	Future aspects of the Work	166
	Abstract	166
7.1	Introduction	166
7.2	Computational Details	167
7.3	Results and Discussion	168
7.4	Conclusions	181
7.5	References	181

Abbreviations

DFT	density functional theory
MP2	second-order Møller–Plesset perturbation theory
RI	resolution of identity
MARIJ	multipole accelerated resolution of identity
CI	configuration interaction
CC	coupled cluster
KS	Kohn-Sham
B3LYP	Becke, 3-parameter, Lee-Yang-Parr
M06	Minnesota 06
PBE	Perdew, Burke and Ernzerhof
TZVP	triple zeta valence plus polarization
TS	transition state
NBO	natural bond orbital
ESP	electrostatic potential
COSMO	conductor-like screening model
CPCM	conductor like polarization continuum model
QM	quantum mechanical
EF	electrostatic force
SEI	secondary electrostatic interactions
H-bond	hydrogen bond

Physical Constants

Speed of Light	$c = 2.99792458 \times 10^8 \text{ ms}^{-1}$
Elementary charge	$e = 1.602176565 \times 10^{-19} \text{ C}$
Planck's constant	$h = 6.62606957 \times 10^{-34} \text{ Js}$
Boltzmann's constant	$k = 1.3806488 \times 10^{-23} \text{ JK}^{-1}$
Avogadro's constant	$N_A = 6.02214129 \times 10^{23} \text{ mol}^{-1}$
Gas constant	$R = 8.3144621 \text{ JK}^{-1}\text{mol}^{-1}$
Faraday's constant	$F = 9.64853365 \times 10^4 \text{ Cmol}^{-1}$
Mass of electron	$m_e = 9.10938291 \times 10^{-31} \text{ kg}$
Mass of proton	$m_p = 1.672621777 \times 10^{-27} \text{ kg}$
Mass of neutron	$m_n = 1.674927351 \times 10^{-27} \text{ kg}$
Atomic mass unit	$u = 1.660538921 \times 10^{-27} \text{ kg}$
Bohr magneton	$\mu_B = 9.27400968 \times 10^{-24} \text{ JT}^{-1}$
Bohr radius	$a_0 = 5.291772109 \times 10^{-11} \text{ m}$
Rydberg constant	$R = 1.097373157 \times 10^5 \text{ cm}^{-1}$
Standard acceleration of free fall	$g = 9.80665 \text{ ms}^{-2}$
Gravitational constant	$G = 6.67384 \times 10^{-11} \text{ Nm}^2 \text{ kg}^{-2}$



Synopsis

	Synopsis of the Thesis to be submitted to the Academy of Scientific and Innovative Research for Award of the Degree of Doctor of Philosophy in Chemistry
Name of the Candidate	Mrityunjay Kumar Tiwari
AcSIR Enrolment No. & Date	Ph. D in Chemical Sciences (10CC11A26035); Aug. 2011
Title of the Thesis	Computational Studies of Noncovalent Interactions in Understanding and Designing New Systems of Biological and Chemical Significance
Research Supervisor	Dr. Kumar Vanka

Keywords: *density functional theory, hydrogen bond, noncovalent bond, Long Range Interactions, Electrostatics Force.*

Noncovalent interactions are at the center of many biological and chemical processes. There has been a conscious effort in recent times to exploit these interactions in order to achieve specific targets.¹ However, to fully unravel and exploit the potential of these weak interactions, it is necessary to understand their effects and efficacy in greater detail. Noncovalent interactions that have garnered maximum attention in terms of the ubiquity, bond-strength, and practical applications are those that are dominated by electrostatic contributions.¹⁻²

Noncovalent interactions are defined to be the combination of some physically well defined contributors: electrostatics, dispersion, donor-acceptor charge transfer, polarization and Pauli repulsion. Recent literature reveals that the polarization and donor-acceptor charge transfer are electrostatic in nature.³ According to the Feynman interpretation, even the dispersion interaction is electrostatic in origin.⁴ Therefore, proper treatment of electrostatic interactions is adequate to describe all kinds of noncovalent interactions fully. This thesis is dedicated to the studies of those noncovalent interactions that are dominated by electrostatic contributions.

References

- (1) E. A. C. Davie, S. M. Mennen, Y. Xu and S. J. Miller, *Chem. Rev.*, 2007, **107**, 5759.
- (2) A. S. Mahadevi, G. N. Sastry, *Chem. Rev.*, 2016, **116**, 2775.
- (3) T. Clark, P. Politzer, J. S. Murray, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*,

2015, **5**, 169.

(4) R. P. Feynman, *Phys. Rev.*, 1939, **56**, 340.

Statement of the Objective: Existing literature signifies that the long range electrostatic interactions between remotely placed atoms on two partners are crucial in determining the association constant of hydrogen bonded complexes.¹ Based on this finding the other pertinent questions that could be asked here and are: (i) Does the long range electrostatic influence arising due to explicit solvent molecules affect the hydrogen bond strength? (ii) What role do the long range electrostatic interactions play in determining reaction barriers? (iii) Do long range electrostatic interactions lead to some amount of chemical selectivity? (iv) Do the chemical phenomena that happen involving charge systems get influenced by long range electrostatic interactions? Also, the multi-point long range electrostatic interactions have inbuilt directional properties. However, the directional aspects of these interactions have been totally overlooked so far in the literature. In this thesis, we have aimed to answer, to some extent, the above-specified questions. Firstly, we have proposed a method that suggests estimating the impact of long range electrostatic interactions in terms of the electrostatic force that can take care of the directionality. Later on, we have investigated certain representative examples in order to answer the above questions that are of fundamental interest for a broad area of science. The thesis is divided into seven chapters. A brief introduction of each chapter is provided below.

References

(1) P. L. Popelier, L. Joubert, *J. Am. Chem. Soc.*, 2002, **124**, 8725.

Proposed Chapter 1: A Brief Introduction of Noncovalent Chemistry

In this chapter, we have provided a brief introduction to noncovalent bonding and its impact on chemistry. We have also discussed a brief history of the presiding literature to describe how the awareness of the significance of long range electrostatic interactions has evolved.

Proposed Chapter 2: The Fundamentals of Density Functional Theory and Theoretical Background

We have exploited density functional theory (DFT) as a tool for investigating problems of interest in this thesis. In this chapter, the development and fundamental aspects of DFT have been described in brief.

Proposed Chapter 3: The Directional Nature of Long Range Electrostatic Interactions

It has been well established that long range multipoint electrostatic interactions have a significant effect on the stability of hydrogen bonded complexes.¹ Interestingly, multipoint electrostatic interactions are directional in nature, ignoring which may lead to a loss in certain important information about the bonding. This point can be understood with the help of following hypothetical model (Fig. 1). If all of the individual distances between the charges on the smaller and the larger subunits in Fig. 1 are equal in the two complexes, the intermolecular electrostatic potential of both the systems would be the same. However, the electrostatic force of bonding will vary: the structure on the right will be more tightly held. One can also see from Fig. 1 that the line of approach of the two interacting species is significant, further underlining the importance of directionality in electrostatic dominated interactions. In this chapter, we have described in detail the directional aspect of these interactions, as well as their

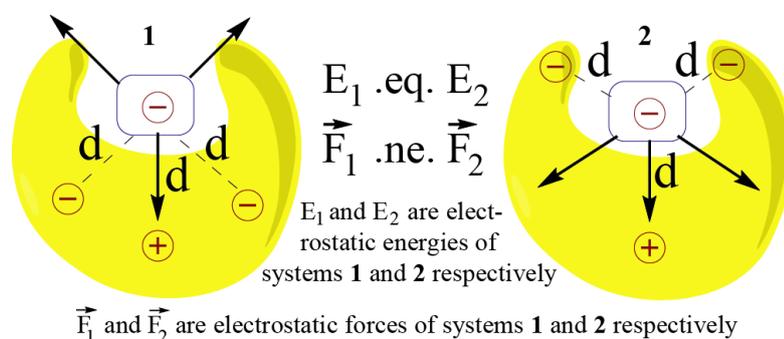


Figure 1. The representation of electrostatic forces between two molecular partners bonded by electrostatic dominated noncovalent interactions and having different charge distributions in three-dimensional space; d is the distance between the two charges.

consequences. We have proposed here that the consideration of electrostatic force can provide better information about the bonded systems when bonding is predominantly electrostatics. We have further proposed that the electrostatic force may be calculated considering atoms in molecules as point charges and employing Coulomb's electrostatic equation for force. We have benchmarked this method against the forces that have been obtained employing the finite difference method with the help of energy decomposition analysis.

References

- (1) P. L. Popelier, L. Joubert, *J. Am. Chem. Soc.*, 2002, **124**, 8725.

Proposed Chapter 4: Exploiting Long Range Secondary Electrostatic Forces for Regulating Electrostatics Dominated Noncovalent Interactions

In this chapter, we have shown that the electrostatic force of binding correlates linearly with binding energy within a family of molecular complexes when the interactions between corresponding partners are dominated by electrostatics and the major varying factor in different complexes is atomic charges and not the distances between the atoms. This indicates that any virtual change in the electrostatic force will lead to change in binding energy in a linear fashion. We have exploited this idea in designing new, superior systems of diverse binding by designing new molecules having association constants up to fifteen orders of magnitude higher than the best-reported association constants for hydrogen bonded complexes,¹ in which the goal of researchers was to obtain as strong a binding as possible; and by further designing an anion that has six orders of magnitude lower association tendency with the cation than the best reported example in the field of contact-ion pairs² where researchers aim to design anions with as weak a binding as possible. We have further shown that the consideration of electrostatic force leads to two distinct regions around the electrostatic dominated noncovalent bonds: binding and anti-binding regions, similar to what has been reported by Berlin³ and Bader⁴ on covalent compounds considering the actual electron density.

References

- (1) B. A. Blight, C. A. Hunter, D. A. Leigh, H. McNab, P. I. Thomson, *Nat. Chem.*, 2011, **3**, 246.
- (2) Hannant *et al.*, *Dalton Trans.*, 2006, 2415.
- (3) T. Berlin, *J. Chem. Phys.*, 1951, **19**, 208.
- (4) R. F. W. Bader, *J. Am. Chem. Soc.*, 1964, **86**, 5070.

Proposed Chapter 5: Influence of Explicit Presence of Solvent Molecules on Hydrogen Bond Strength

The hydrogen bonded systems that exist in the solvent environment are prone to get affected by the long range influence of explicit solvent molecules. Despite the ubiquitous presence of solvents molecules around most of the noncovalently bonded systems in biology and chemistry, this incidence has not earned proper attention in the literature. Moreover, estimating the long range influence of solvent molecules on such interactions employing conventional theoretical methods is not possible. The method that we have proposed recently that describes calculating the electrostatic force of binding to estimate the strength of electrostatic dominated interactions, can effectively be employed to obtain a qualitative estimation of the long range influence of explicit solvent molecules. In this chapter, we have reported an extensive theoretical and computational study that was intended to explore this phenomenon on three distinctly different types of hydrogen bonded model systems. We have obtained in our studies that the explicit solvent molecules indeed influence the strength of hydrogen bonds and that this influence can be expressed in two ways: (i) through charge transfer or induction effect involving solvent molecules and (ii) through the long electrostatic range influence of the solvent molecules.

Proposed Chapter 6: Does Electrostatic Field Created due to Solvent Molecules Influence Reaction Barriers?

Solvent's influence on reaction barriers has been understood in four ways, through (i) effect of polarity¹ (macrosolvation), (ii) noncovalent bonding effect (microsolvation),² (iii) proton relay effect³ and (iv) coordinating role of solvent⁴. A reaction barrier gets altered due to differential stabilization of transition state and reactant structures by (i)

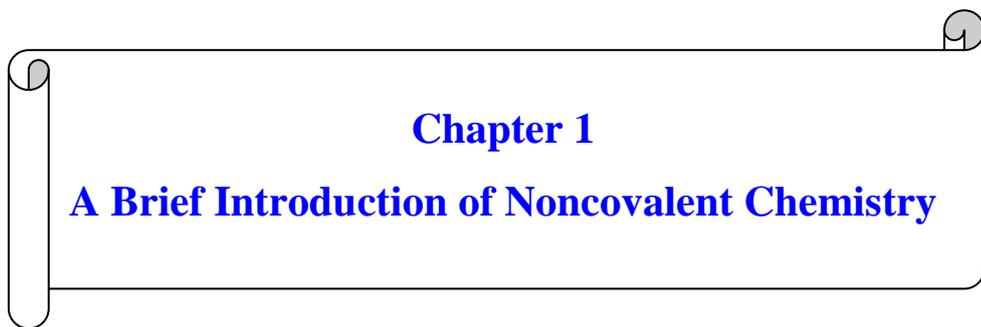
and (ii) and by the direct participation of solvent molecules that reduce strain in transition state structures through (iii) and (iv). The first two conditions are almost common to all kind of reactions in a homogeneous environment. However, the last two effects are restricted to only a few specific cases. The reason argued behind differential stabilization of reactants and transition state structures is the differential charge distribution in reactants and transitions state structures (TSs).¹ However, what must be noted here is that if the charges on the atoms in reactants and TSs varies, there is a response in the local electric field that is created by explicit presence of solvent molecules near the reaction sites that will also vary and may lead to further discrimination in stabilization of reactants and TS structures depending upon the orientation of the field. It has been seen in the past that the external electric field indeed affects the reaction barriers of all kinds of chemical transformations.⁵ However, a general rule to quantify electric field effect of solvents is still lacking in the literature. In this chapter, we have proposed a general scheme to enumerate the effect of the electric field created by the solvent molecules. We have obtained a quantitative trend in barrier heights in the presence of the actual solvent molecules and in the field-equivalent solvent.

References

- (1) D. E. Hughes, C. K. Ingold, *J. Chem. Soc.* 1935, 244.
- (2) Singh *et al.*, *Inorg. Chem.*, 2015, **54**, 1535.
- (3) M. P. Patil, R. B. Sunoj, *Chem. Eur. J.*, 2008, **14**, 10472.
- (4) S. Jain, K. Vanka, *Chem. Eur. J.*, 2017, **23**, 13957.
- (5) Aragonès *et al.*, *Nature*, 2016, **531**, 88.

Proposed Chapter 7: Future aspects of the Work

In this chapter, we have shown other miscellaneous examples where consideration of directional nature of long range electrostatic interactions might be helpful to obtain deeper insights into such important systems.



Chapter 1
A Brief Introduction of Noncovalent Chemistry

Chapter 1

A Brief Introduction of Noncovalent Chemistry

Abstract: Noncovalent interactions are at the center of biological processes. Right from molecular recognition to molecular assembly to enzyme catalysis to the dynamic behavior of biomolecules, everything is assisted by noncovalent interactions. In recent times, applications of noncovalent interactions in problems related to diverse areas of chemistry, biology and physics have become inevitably profound and are at the verge of increasing even further. Noncovalent interactions that have garnered maximum attention and coverage in terms of their ubiquity and bond-strength are those that are dominated by electrostatic contributions. Electrostatic interactions are long range and directional in nature. However, the directional aspects of electrostatic dominated noncovalent interactions that arise due to the intrinsic long range influential nature of electrostatic contributions have been totally overlooked so far into the literature. This thesis essentially showcases the significance and directional nature of long range electrostatic interactions in order to fully elucidate the electrostatic contributions. In this chapter, we have provided a brief introduction to noncovalent bonds and their impact on chemistry. We have also discussed a brief history of the presiding literature in order to describe how the awareness of the significance of long range electrostatic interactions has evolved over time.

1.1 Introduction

The interactions between atoms under certain conditions form molecules. These pairs of atoms are said to form a chemical bond, which could be of covalent or ionic or polar character. These bonds are strong, with usually maximum attractions between bonded pairs of atoms. However, there also exist some bonds between atoms or molecules or ions that are relatively much weaker but are still strong enough to determine physical and/or chemical properties of a plethora of chemical substances when many of such bonds work in tandem. One classic example of such properties is the condensed phase of matter, which is determined by the extent and recurrence of weak bonds between molecules of the given material at a given condition. These weak

bonds are commonly called noncovalent bonds. Noncovalent bonds literally refer to the types of bonds that are not covalent in origin, *i.e.*, the bonds that do not occur as a consequence of the actual sharing of electrons between the interacting atoms are noncovalent bonds.¹⁻² However, the interactions between ions in ionic crystal lattices or atoms in metallic lattices are also not covalent in origin, but they are generally not considered as noncovalent bonds. Traditionally, only the weak interactions that are of non-covalent nature are referred to as noncovalent bonds.¹⁻² For this reason, noncovalent bonds are also sometimes termed simply as weak bonds. These bonds are so weak in comparison to the normal chemical (electrovalent or covalent) bonds that they have been plainly neglected for practical applications for a significant period of time after their discovery. It has been realized much later that noncovalent interactions are significant enough to induce selective effects or functionality into the systems. Recently, techniques have been evolved to exploit these interactions in order to achieve specifically designed goals in diverse disciplines of science.³⁻⁴ This field is growing continuously and has attained a multidisciplinary and multifaceted recognition with a vast number of researchers across the globe from every discipline of chemistry, biology and material science. The large number of papers that have appeared containing the term “noncovalent interaction” in the recent past are a manifestation of the popularity that this area has gained in recent years.³⁻⁵

It is to be noted that the energy of formation of noncovalent bonds is typically in the range of 1-5 kcal/mol/bond.⁶ Since, the average kinetic energy of molecules at room temperature is about 0.6 kcal/mol, many molecules acquire enough energy to break these bonds, and hence these bonds are considered to be dynamic and are often referred to simply as interactions. However, multiple noncovalent bonds get to stabilize a specific shape and conformation of a large molecule, and specific associations between two or more large molecules.^{1,5-7} For example, several types of noncovalent bonds are critical in maintaining the three-dimensional structures of proteins, nucleic acids and supramolecular structures.^{1,5-7} Additionally, noncovalent bonds also enable one large molecule to bind transiently in different accessible conformations, making them the basis of many dynamic biological processes.⁸ The molecular recognition that is very vital for selectivity and the reactivity of biomolecules is assisted by noncovalent interactions.⁸ The dynamic nature of enzymes allows many types of molecules to penetrate within the active site. However, only

certain patterns in noncovalent interactions specific to certain kinds of substrates lead to the triggering of catalysis.⁹ Certain cell surface receptors recognize a specific pattern of a noncovalent bond with the ligand in order to differentiate between the native and foreign molecules that has to pass through a membrane.¹⁰ The solubility of solutes in a solvent, the viscosity of a liquid, the melting and the boiling point of chemical substances, and the association tendencies of small peptides and molecular clusters are other key areas where noncovalent interactions play a vital and decisive role.¹⁻⁶ It is important to note here that some specific instances of strong noncovalent interactions have also been reported (*e.g.*, hydrogen bond involving ions¹¹ and cation- π interactions¹² between small cations and easily polarizable π -systems), which occurs due to strong covalent contributions in noncovalent bonding. However, weak noncovalent interactions are omnipresent and of greater significance due to their decisive role in governing molecular behavior. In this thesis, we will focus exclusively on those noncovalent interactions that are weak, and henceforth weak noncovalent interactions will be referred to simply as noncovalent interactions.

When two noncovalent bonded partners come in the vicinity of each other, the partner that has a higher electric potential near the site of interaction polarizes the electron cloud of the other partner in such a manner that some of the electron density of the latter enters into the molecular boundary of the first partner, which is accounted for as the donor-acceptor term or the covalent contribution to the noncovalent bond.¹³ Thus, the formation of noncovalent bonds often accompanies the transfer of charge (quantitatively very small) between bonded partners. The effect of polarization of electron clouds of bonded partners also results in the perturbation of atomic charges on the individual atoms in comparison to their infinitely separated state, and hence, the strength of these bonds also gets affected, which is considered the induction or polarization contribution to the noncovalent bonds.¹³

1.2 A link to the past

The first ever implication of noncovalent interactions has been discovered and described by van der Waals in the 1870s. He explained that the discrepancies in the state functions of ideal and real gases is due to the attractive interactions between the atoms (of noble gases) or the molecules (of polyatomic gases).¹⁴ These interactions have been referred to as van der Waals interactions. In 1930, Fritz London described

the fundamental nature of van der Waals interactions employing the recently born quantum mechanics,¹⁵ which was further extended by Hans Hellmann.¹⁶ Later, many other kinds of interactions that prevail in nature were recognized and extensively explored.^{1,2} Their respective roles in chemistry was also understood to a large extent. On the basis of the structure of the interacting moieties they can be separated into the following different categories: hydrogen bonds (classical and non-classical), σ -hole bonds, cation- π interactions, anion- π interactions, π - π stacking, hydrophobic interactions, lone pair interactions and lone pair-cation- π interactions. The strength and properties of these bonds differ considerably. However, their fundamental origin may be considered as the same. Fundamentally, they occur due to the attraction between a positively charged (which may be permanent or transient) center/centers and a negatively charged center that may have a permanent negative charge or a pi-electron cloud or a lone pair of electrons or an easily polarizable electron cloud.¹³

1.3 Types of noncovalent interactions

There are various physically well defined contributors to the noncovalent interactions: electrostatic, dispersion (London dispersion), polarization (induction), exchange-repulsion (Fermi repulsion) and donor acceptor charge transfer interactions (covalent contribution). Even though these contributors are very well defined, none of them are physical observables. Depending on the fundamental nature of these contributors, their influence could be of longer or shorter range. The origin of short range interactions is in the overlap of molecular orbitals (or the transfer of charge from one partner to another), and hence they do not act at long distances. Generally, in molecular dynamics simulations, the forces that decrease with distance quicker than r^{-d} (where d is the dimensionality of the system, usually 3) are considered to be short range forces. Evidently, the electrostatic or electrostatic dominated noncovalent forces are long range forces as they decrease with distance as r^{-2} . The effects of electrostatic interactions have been seen to be non-negligible even beyond a distance of 15 Å.¹⁷ Polarization and dispersion interactions are also considered to be long range interactions. However, they are of relatively shorter range than the electrostatic. Their effects are considered to be significant only up to a distance of 5 Å.¹⁸ However, it has been reported that the influence of London dispersion interactions can be seen up to a distance of 8 Å.¹⁹ The exchange-repulsion and donor acceptor charge transfer

interactions are considered to be short range interactions.¹⁹ Based on the dominance of the attractive individual energy contributors, noncovalent interactions can be classified into two distinct major classes: (1) the electrostatic dominated interactions (*e.g.* classical H-bond, dipole-dipole interactions, the σ -hole bonding, the interaction between the ion-pairs in a solution, and so on) and (2) the dispersion dominated interactions (π - π interactions, anion- π interactions, CH- π interactions *etc.*). This thesis is mostly dedicated to the electrostatic dominated noncovalent interactions.

Since the long range influence of electrostatic contributions are significant and since they have largely been overlooked in the literature, we will focus mainly on the long range effect of electrostatic contributions on various aspects of chemistry, ranging from the bond energy of noncovalent bonds to reaction barriers. As electrostatic interactions are directional in nature, and the same is thus true for long range electrostatic contributions, we will look into the long range influence of electrostatic interactions as a perspective of electrostatic force rather than the electrostatic energy. It is necessary to note here that the directional aspect of long range electrostatic interactions has been completely ignored in the literature.

1.4 Cooperativity in noncovalent interactions

It has been seen that the strength of a noncovalent interaction depends highly on the local arrangement of atoms in bonded partners^{17,20} and also on the local chemical environment²¹. Further, the mutual interactions between different units of noncovalent interactions have been found to be nonadditive.¹⁹ For example, the average bond strength of hydrogen bonds in water clusters increases with the simultaneous decrease in bond length and it becomes double in a decamer.²² This nonadditivity in noncovalent interactions are considered to be a consequence of cooperativity, noncooperativity and anticooperativity in noncovalent interactions.^{19,23} When a pair of noncovalent interactions strengthen each other, they are said to be cooperative, whereas when they weaken each other they are said to be anticooperative. Cooperativity implies that the net interaction energy of two interactions together is larger than the simple sum of the individual interaction energies. A general trend in the cooperativity in noncovalent interactions has been shown in Figure 1.1.

The origin of nonadditivity is attributed to the long range interactions between the atoms of bonded partners and between the atoms of bonded partners and the

surrounding molecules.^{19,23} There are basically three kinds of long-range effects that can be considered in this context: (i) the electrostatic effect (ii) dispersion effect and (iii) the polarization effect. The individual long range electrostatic interactions between a pair of atoms on bonded partners or between surrounding atoms and the bonded partner, which may be attractive or repulsive, are one of the major factors behind the nonadditivity of noncovalent interactions, as the long range electrostatic interaction brings directionality into play.²⁰ Dispersion refers to the attractive term of the vdW equation. It arises due to the electron correlation effect.²⁴ The contribution of dispersion to the many body interaction is considered to be very small.¹⁹ Polarization effects are also attractive effects and arise due to the polarization of electron density of the bonded partners in the electric field of its neighboring molecules including the solvent, as discussed in Section 1.1. As the fields of several neighboring molecules may add or cancel out depending on the direction, the polarization effects are also nonadditive.²⁵ It has also been seen in certain systems that the formation of the first noncovalent interaction facilitates the formation of the second, the third and further interactions.²⁶ This is particularly due to the fact that the entropic cost of making

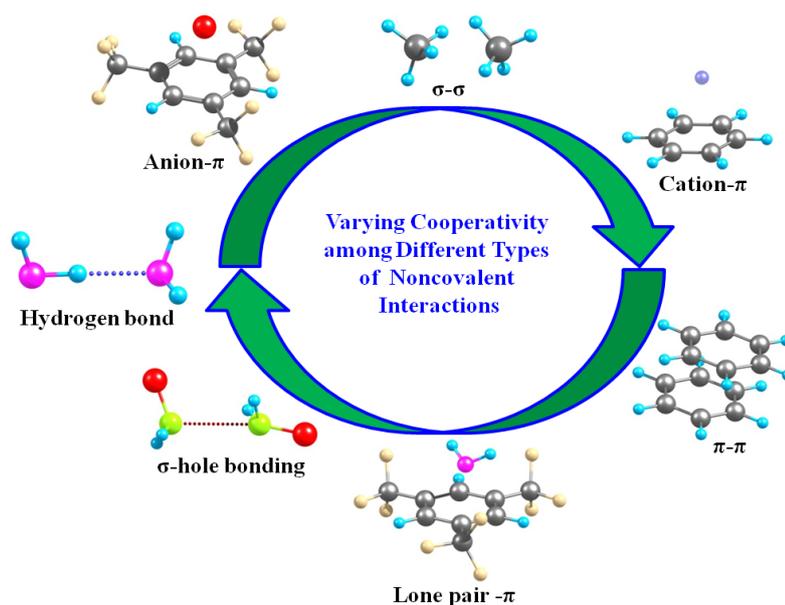


Figure 1.1. Representation of different types of noncovalent interactions and a general trend in cooperativity among them. Color representation: black – carbon, cyan – hydrogen, blue – lithium cation, red – chloride anion, yellow green – phosphorous, pink – oxygen, cream – fluorine.

subsequent noncovalent interactions once the first one is already formed becomes smaller.^{19,27} This effect is commonly called the chelate effect, and is suggested to be the reason why helix propagation is seen to be cooperative.²⁶

1.5 Schematic representation of noncovalent interactions

Schematically, noncovalent interactions are represented as three dots (...), instead of a full line, which represents a covalent bond. For example, the covalent bonds between hydrogen and oxygen atoms in water molecules will be represented as H-O, whereas a hydrogen bond between two water molecules will be denoted as H...O.

1.6 Some general nomenclatures

The formation of any kind of bond is associated with the release of energy, which accounts for the stabilization of the systems by the amount of bond energy or bond dissociation energy. The strength or stability of covalent bonds is generally reported in terms of bond dissociation energy. However, there is no standard scale to indicate the strength of noncovalent bonds. Generally, in experimental techniques, the strength of noncovalent bond/bonds is reported in terms of the association constant, which accounts for the degree of formation of the noncovalent bonded complex from the corresponding separated partners. A representative example of how the association constant can be obtained from experiments is illustrated below. Let us assume that two species A and B interact with each other through a noncovalent bond to form a molecular complex A...B as shown below



Since noncovalent bonds are weak, the complex formation can be considered to be a reversible process under ambient conditions. The association constant for the complex formation can be calculated from the following equation once the reaction has reached equilibrium,

$$K_a = [A...B]/([A].[B]) \quad (1.1)$$

Where K_a is association constant for the formation of A...B. The unit of association constant is mole^{-1} . Theoretically, association constants can be computed from following equation:

$$K_a = \exp(-\Delta G/RT) \quad (1.2)$$

Where ΔG is the free energy of formation of A...B and can be calculated from the difference in free energies of complex A...B and the sum of the free energies of the infinitely separated A and B. The other common energy terms that have been used frequently in theoretical calculations to denote the strength of noncovalent bonds are the binding energy and the interaction energy. The binding energies are calculated considering the electronic energies of the systems at 0 K (sometimes zero point energies are also considered). As per the abovementioned scheme, this is the difference between the electronic energy of A...B and the sum of electronic energies of the infinitely separated A and B. On the other hand, the interaction energy is calculated as the binding energy when the two partners are separated from an optimized complex to obtain the single point energy of each structure in the same geometry that they possessed when they were in the complex. It is therefore clear that the interaction energy also includes the deformation energy term, which appears due to the geometrical perturbation in the structure of the partners in a complex as compared to their structures in isolation.

It is worth noting here that there is no experimental or theoretical technique available so far, to the best of our knowledge, that can pinpoint the strength of individual noncovalent bonds in a multi-noncovalent bonded system. These bonds are so weak that their selective dissociation is impossible to obtain, at least to date. The work reported in this thesis where we have shown that the strength of electrostatic dominant noncovalent interactions can be stated in terms of the electrostatic force thus opens an avenue for estimating the relative strength of the individual noncovalent bonds in terms of the binding force when the bonding is dominated by electrostatic contributions. The electrostatic force of binding for each bond in such cases can be calculated along the line joining the donor and acceptor atoms as per Coulomb's equation for the electrostatic force, where the atoms in molecules are considered to be point charges.

1.7 Quantum Chemical Methods for Noncovalent Interactions

The efficient and accurate description of noncovalent interactions is very important to understand several properties of molecular systems. Similar to covalent interactions,

noncovalent interactions can also be studied by standard methods of quantum chemistry. There are two basic quantum chemical approaches that have been employed extensively in investigating noncovalent interactions: (i) the *ab initio* wave function approach and (ii) density functional theory (DFT). The semiempirical molecular orbital (MO)-based methods, and classical methods (atomistic molecular dynamics, coarse grained force fields and continuum mechanics) are other alternatives. Classical methods are based on classical mechanics which does not treat electrons explicitly, and are therefore less sophisticated approaches in comparison to quantum chemical methods. Unfortunately, the efficacy and accuracy of these methods does not work in synergy. One has to pick and choose among these choices depending on the system size and the available computational resources. High level, electron correlation wave function theory such as CCSD(T) accurately describes all energy contributors of noncovalent interactions²⁸ and they are thus considered to be among the best methods for dealing with noncovalent interactions. However, these methods are computationally very expensive and cannot be employed to even moderate size molecular systems. Recent reports suggest that these methods can be applied only to systems having up to a hundred atoms.²⁹ Therefore, these methods are mostly used in the benchmarking of DFT based methods. The perturbation theory based approaches such as MP2 (second order Møller-Plesset perturbation theory) are the other recommended quantum chemical methods that have been suggested to capture noncovalent interactions with high accuracy.³⁰ However, these methods are also expensive and cannot be employed to investigate large size molecular systems. With recent development using RI (resolution of identity) approximations, they can effectively be applied for systems having up to a thousand atoms.³¹ However, the full geometry optimization of moderate size molecular systems is still tricky with these methods.

DFT, on the other hand, represents a less expensive method and can be employed smoothly to moderate size molecular systems of up to a thousand atoms.²⁸ Even full geometry optimizations with DFT are within reach because of the speed and efficiency of contemporary supercomputers. However, the omnipresent electron correlation, in DFT is typically expressed in a local framework, and hence these methods are ineffective in capturing the dispersion interactions (which is one of the important contributors in noncovalent interactions) accurately because of nonlocal

(long-range electron correlation) nature of dispersion effects.³² Similar difficulties have been seen with the Hartree-Fock (HF) approach as well, which lacks the electron-electron correlation effects. The failure of standard Kohn–Sham DFT in capturing London dispersion is due to the true wave function based origin of the dispersion energy. Various approaches have been made in recent years in order to introduce London dispersion effects in DFT. As a result, many approximate methods are available nowadays that provide very good accuracy in DFT calculations and have now become the standard in the field.³³ These methods are called dispersion corrected density functional theory, DFT-D approaches.³³ A brief introduction to DFT has been provided in Chapter 2 of this thesis.

The other problem with the DFT based methods in dealing with the noncovalent bonded molecular complexes is that they are basis set inconsistent. This inconsistency in DFT approach is associated with the employment of the variational principle in the Kohn-Sham formalism. In variational calculations, supersystems are described as the sum of the basis sets of all subsystems, which lead to a more complete basis set description of subsystems in the complex in comparison to the isolated cases, and subsequently causes the artificial lowering of energy of the supersystems.² In more descriptive terms, this error originates due to incomplete treatment of subsystems, which try to improve the treatment by borrowing orbitals from neighboring subsystems. Therefore, this error is purely mathematical in origin and has no physical meaning as such. Boys and Bernardi came with the solution of this problem and termed it as a counterpoise correction method and this problem as a basis set superposition error, BSSE.³⁴ This error can be eradicated either by describing the basis sets of subsystems in terms of the basis sets of the complete supersystem or by using an infinite AO basis set. In DFT methods, the counterpoise correction is performed by introducing the concept of ghost-orbitals. Ghost subsystems have basis set functions at the place of atomic positions but without the nuclei and electrons. This simply means that like the interaction energy, the BSSE is geometry dependent and we will obtain different values of BSSE for different orientations of atoms (ghost systems) even for the same complex. It has been seen that the BSSE-corrected properties converge much faster to the limit values than the BSSE-uncorrected ones. The second approach to eliminate BSSE is available through an extrapolation procedure to the complete basis set (CBS) limit, for which BSSE vanishes

automatically by definition. To test whether a particular variational procedure is free of BSSE or not, one can compare the stabilization energy obtained by this method with the stabilization energy obtained by employing perturbation methods, which are, by definition, free of this issue. It is to be noted that BSSE is not limited only to molecular clusters. The same error can be seen in an isolated system when one part of the system improves its basis set description by borrowing functions from the other part.³⁵ Many methods have been suggested in order to solve this error.³⁵ The solution to intramolecular-BSSE is suggested by Asturial *et al.* who suggested correcting this error by considering one molecule as a different fragments, with potentially as many fragments as the number of atoms. The most famous among them is the one suggested by Jensen.^{35,36}

Other low cost methods for electronic structure calculations for even larger systems (up to approximately $10^3 - 10^4$ atoms) are semiempirical MO methods.³⁷ Two general classes of semiempirical approaches that exist in the literature are - approximations to either DFT or to the Hartree-Fock method.²⁸ In both of these cases, minimal AO basis sets are used to expand the MOs. The semiempirical integral and basis set approximations lead to errors in electrostatic, polarization and exchange-repulsion terms in addition to the dispersion interaction. The repulsion term is somewhat recovered through empirical element-specific pair potentials. However, the poor representation of electrostatic and induction interactions lead to a bad description of electrostatic dominated interactions.³⁸ In order to minimize these errors, geometry-dependent noncovalent bond corrections have been proposed.³⁹ Both of the abovementioned semiempirical approaches had been recently augmented with the D3 dispersion correction and had been reported to give better performance for various kinds of systems.⁴⁰ The errors that arise due to the negligence of the three- and four centered two-electron integrals in the Hartree-Fock formalism have been repaired through one-electron contributions and element-specific pair potentials.⁴¹ Despite several improved treatments in recent times, semiempirical methods still lack the reasonable accuracy for a variety of systems, and hence are considered to be relatively approximate and crude methods for application to moderate size molecular systems.

1.8 Electrostatic Dominated Noncovalent Interactions

As described in Section 1.3, some noncovalent interactions are dominated by electrostatic contributions. This means that the major part of their binding energy comes from the electrostatic term. The best example of this kind is the classical hydrogen bond. The non-classical hydrogen bonds of type X-H (highly polar)... π also fall into this category. The other examples are σ -hole bonds. The interactions between ion-pairs in solutions can also be considered as the electrostatic dominated noncovalent interactions. The electrostatic dominated interactions have been established as long range interactions. It has also been shown that the electrostatic contributions in hydrogen bonded systems have a long range influence on binding. The following subsections will deal with various aspects of electrostatic dominant noncovalent interactions.

1.8.1 Hydrogen bonding

Hydrogen bonding is the most widely prevalent form of noncovalent interaction. Hydrogen bonds are generally represented as X-H...Y, where X is an atom having stronger electronegativity than the H atom and Y is an electron rich species: any kind of Lewis base. Here, Y is called the hydrogen bond acceptor as it accepts H in a hydrogen bond and X-H is referred to as the hydrogen bond donor. It is to be noted that the hydrogen bond does not merely mean the H...Y interactions. Instead, the entire X-H...Y assembly is collectively referred to as a hydrogen bond. Hydrogen bonds are considered to be of two basic types: (i) the classical hydrogen bond (when Y is an electronegative atom with a lone pair of electrons on it, *e.g.*, the hydrogen bond between water molecules) and (ii) the non-classical hydrogen bond (when Y is a π -electron rich moiety, *e.g.*, in the case of XH... π interaction, X = O, N, C, halogens). Classical hydrogen bonds are ubiquitous in nature and they are believed to be primarily electrostatic in type.¹³ A representative example of a classical hydrogen bond has been shown in Figure 1.2. The electrostatic interactions involving the hydrogen bond acceptors and donors that are directly involved in hydrogen bonding with each other are referred to as primary interactions. The electrostatic interactions between any pair of atoms between two partners that are not directly involved in hydrogen bonding have been defined as secondary electrostatic interactions. All kinds

of electrostatic interactions (including primary and secondary interactions) have been referred to collectively as long range electrostatic interactions.

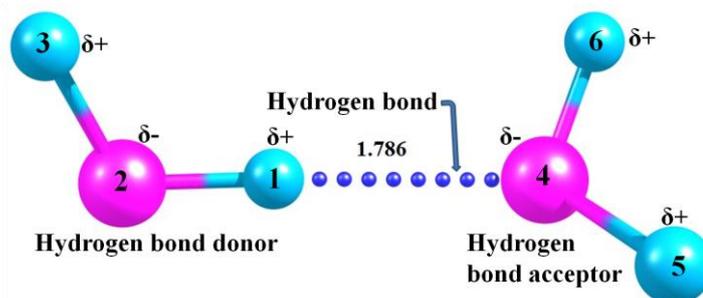


Figure 1.2. An optimized geometry of the water dimer at the COSMO(water)/PBE-D3/TZVP level of theory using Turbomole 6.4. The water molecules are bonded here through a classical hydrogen bond. Color coding: pink – oxygen, cyan – hydrogen, dotted blue line – hydrogen bond. O4 is a hydrogen bond acceptor and H1 is a donor. The interaction between H1 and O4 denote primary interactions and the remaining combinations of atom-pairs interactions between two partners (*e.g.*, between H3 and O4) are referred to secondary electrostatic interactions. All the interactions between each atom of the first water molecule with every atom of the second water molecule are commonly defined as long range electrostatic interactions.

1.8.2 Directional noncovalent interactions

It has been believed for long that the electrostatic (Coulomb) interactions in a noncovalent bond exhibit little directionality, *i.e.*, their energy barely depends on the bond angle. This incorrect assumption could be due to two fundamental misconceptions: (i) the atoms that are involved in these interactions are charged in an isotropic fashion and (ii) long range secondary electrostatic interactions are negligible (or their spatial arrangements cancel out the effect of directionality in these interactions). The small directionality in noncovalent interactions has been mostly attributed to the covalent contributions. However, the molecular electrostatic potential (MEP) distribution suggests that the potential around atoms in a molecule is not symmetric, because the distribution of electron density around an atom in a molecule is anisotropic due to the electronegativity or polarizing power (which decreases with distance) of covalently connected atoms. When this differential distribution of the electron cloud around an atom is large and distinct, it brings strong directionality into

the noncovalent interactions, which is of Coulombic origin. The MEP defines the electrostatic interaction of an isolated molecule with its surroundings. It represents the sum of contributions from nuclei and the electron density and is commonly visualized by a color coded isodensity surface.⁴² MEP provides a very useful view of the molecular electrostatic potential at the surface of molecules. A useful consequence of this MEP is that the area of the molecular surface for which the electron density is thin will exhibit a positive potential.⁴² In case of heavier atoms of groups IV-VII, a positive patch of MEP appears as the extension of the covalent bond with stronger electronegative groups.⁴³⁻⁴⁴ This positive area is called the σ -hole (see Figure 1.3 below), which is physically a consequence of the anisotropic distribution of electron density around the heavy atoms of groups IV-VII.⁴³⁻⁴⁴ It has also been seen that σ -holes often exist together with regions of negative potential. If we look at the other side of the heavy atoms in Figure 1.3, a net negative potential can be seen to have accumulated. These σ -holes selectively pull atoms or groups from their negative potential surface side to make a noncovalent bond. These bonds are called as σ -hole bonds. Therefore, σ -hole bonds are noncovalent interactions between covalently bonded heavy atom of Groups IV-VII and a negative site, which could be a lone pair of a Lewis base or an anion. At the same time, the negative potential side of the same heavy atom in the same or different arrangement may attract positive groups or σ -holes, which results in multiple noncovalent bonds exhibited by heavy atoms of groups IV-VII.⁴³⁻⁴⁴ The angular positions of each of these bonds (σ -hole bonds) are found to be almost fixed and thus they are directional in nature. Examples of σ -hole bonds are halogen bonds, chalcogen bonds and pnictogen bonds.

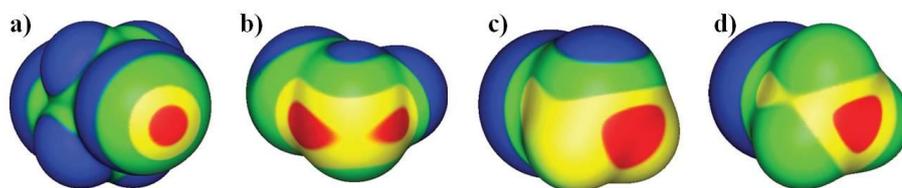


Figure 1.3. Molecular surface electrostatic potential computed on the 0.001 au contour of the electronic density at the M06-2X/6-311G* level of theory: a) ICF₂CF₂I, b) SeFCl, c) PH₂Cl, and d) GeH₃Br. Color schemes: red – strong positive potential; yellow - weakly positive potential; green - weakly negative potential; blue – negative potential. This figure has been taken from the reference 43.

The most common example of electrostatic dominated interactions is the classical hydrogen bond¹³ (henceforth will be referred to only as the hydrogen bond). Even though these bonds are recognized to be predominately electrostatic, the observed preference in the linear X-H...Y angle is credited to the covalent contributions in these bonds, where the donation of the electron has been argued to occur from the lone pair of Y to the antibonding X-H orbital (σ_{XH}^*). Since the overlap between orbitals is largest in the collinear case, X-H...Y is seen to prefer a linear arrangement.^{13b} However, the factor that is often neglected when the directionality in hydrogen bonding is discussed is that the strength of a covalent interaction depends on two parameters – the symmetry or extent of overlap between interacting orbitals and the energy difference between them.⁴⁵ Usually, the X-H bonds are so strong that their bonding σ_{XH} lies very low and the corresponding σ_{XH}^* is very high.^{13b} Since, high-lying antibonding orbitals are poor acceptors of electrons, the covalent interaction in classical hydrogen bonds are very unlikely whenever the X-H bond is strong. The covalent contributions in the noncovalent bonds fundamentally arise due to the shift in electron density of Y to X-H bond, which is caused by the polarization of Y by the electric field of the X-H bond, and not due to mixing of lone pair of Y with σ_{XH}^* , as discussed in Section 1.0.⁴⁶

Thus, the directionality in the noncovalent interactions is basically of electrostatic origin, which can be attributed to two factors: (i) the anisotropic electron density distribution around the nuclei and (ii) the long range electrostatic interactions. The anisotropic distribution of electron density prefers directional interactions, as discussed above, and affects all of the interaction terms that contribute to the overall interaction energy, as they all depend on the electron density and thus are anisotropic in nature.⁴⁷ The nature of electrostatic interactions is directional itself. When the multi-point long range electrostatic interactions are brought into consideration, the directionality arises automatically in the net noncovalent interaction by virtue of the nature of electrostatic interactions.

1.8.3 Literature Precedence

The Watson and Crick model of DNA (1953) describes DNA as a double helical structure in which nitrogen bases of one strand of DNA form multiple hydrogen bonds with nitrogen bases of the complementary strand.⁴⁸ There are in total four

nitrogen bases that are found in DNA: adenine (A), guanine (G), cytosine (C) and thymine (T). A and G are purine ring based, while C and T are pyrimidine ring based (Figure 1.4 below). Out of four of these bases, we normally have two kinds of base pairs that are found predominantly in DNAs^{*1}: the C...G (triple hydrogen bond, C-G) and A...T (double hydrogen bond, A-T) (Figure 1.5). They are the Watson-Crick base pairs. However, if we look at the overall possibility of hydrogen bonding (including both homo- and hetero- patterns), consisting of these four bases, there can be as many as 29 possible base pairs.^{*2} 28 of them were described by Donohue (1960)⁴⁹ and the last one was recognized by Poltev and Shulyupina (1986).⁵⁰ The question that has been of fundamental interest since the discovery of the DNA structure is this - are the two base pairs occurring in DNA the most stable ones? A plethora of studies has been devoted to describing the structure and stabilization energies of different base pairs, employing semiempirical and empirical potentials in accordance with the computational resources available at that time.⁵⁰⁻⁵¹ Experimentally, different analogues of base pairs had been synthesized in order to understand the association tendency of nitrogen base analogues in comparison to the Watson-Crick base pairs.⁵²⁻⁵⁵ These studies opened the gate and later enforced the belief in the scientific community that long range electrostatic interactions in hydrogen bonded complexes are unquestionably significant. A brief history of how the general understanding about the significance of long range secondary electrostatic interactions has evolved is described below. Hitherto, it had been believed exclusively that only primary electrostatic interactions between the hydrogen bond acceptor and donor accounted for the strength of hydrogen bonds.

In Scheme 1.1, the association constant (K_a) of complex **1**, a closer analogue of the G-C base pair, in chloroform was found to be *ca.* 10^4 - 10^5 M^{-1} .⁵² The association constant of a similar complex **2** in chloroform was found to be $1.7 * 10^5$ M^{-1} .⁵³ A careful analysis of the structures of these two complexes suggests that there are two primary O...H interactions and one primary N...H interaction in complex **1**, whereas

^{*1} 42 other unusual base pairs, which are rarely observed, have also been recognized to be found in the DNA structure. The most famous examples among them are the Hoogsteen base pairs.⁵⁶

^{*2}The N9-H of purines and N1-H of pyrimidines (see Figure 1.4 for numbering in nitrogen bases) had been omitted for consideration since N9 of purines and N1 of pyrimidines form covalent bonds with the sugar moiety in the DNA structure and are not available anymore for hydrogen bonding.

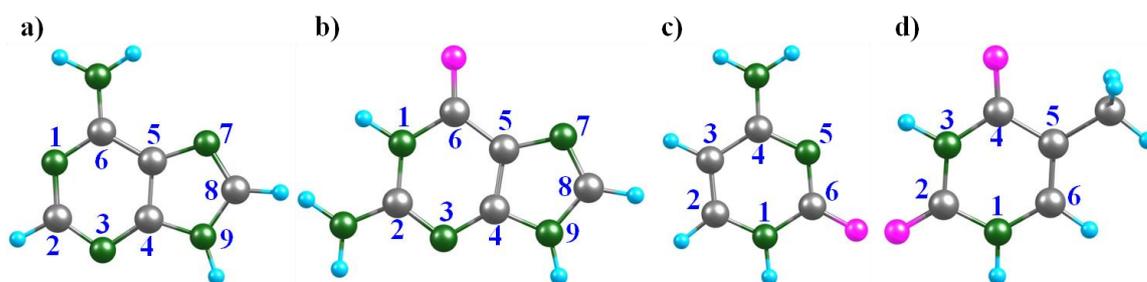


Figure 1.4. Different types of nitrogen bases found in DNA: a) adenine, b) guanine, c) cytosine and d) thymine. Color representation: black – carbon, green – nitrogen, cyan – hydrogen, pink – oxygen. Figure also illustrates the numbering of atoms in the heterocyclic ring.

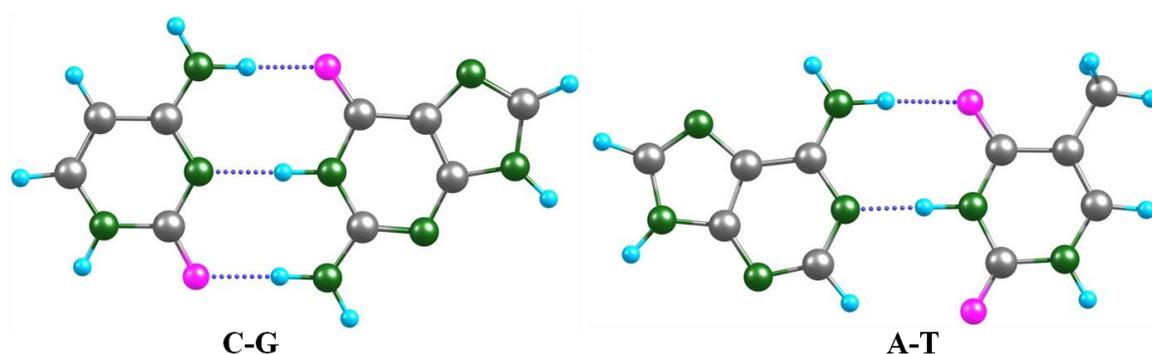
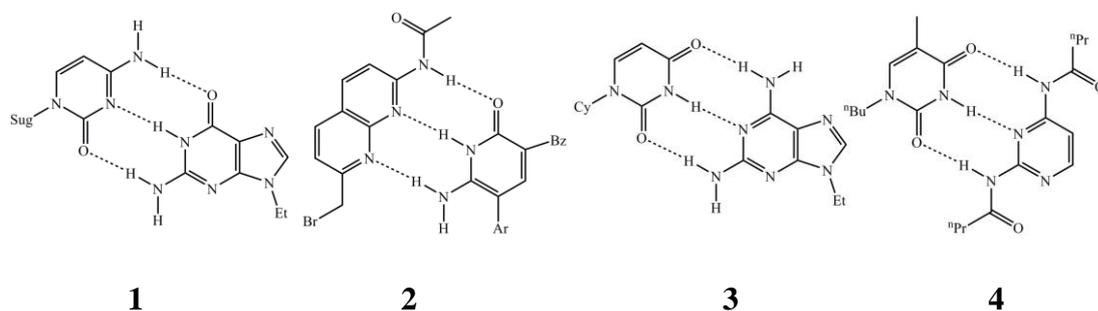


Figure 1.5. Usual hydrogen bonding pattern between nitrogen bases of DNA. Color representation: black – carbon, green – nitrogen, cyan – hydrogen, pink – oxygen, dotted blue line – hydrogen bonding.

there is one O...H interaction and two N...H primary interactions in complex **2**. Since the primary interaction for the O...H pair is stronger than the N...H pair (O acquires higher partial charge than N as its electronegativity is higher than N), the association constant of complex **1** should, in principle, be higher than that of complex **2**. However, the association constants of these complexes are found to be of the same order with **1** having a slightly lower value. A comparison of the primary interactions between complex **2** and complex **3** suggests that the association constant of **3** should be higher than that of **2**. However, the association constant of complex **3** in chloroform was found to be 170 M^{-1} ,⁵⁴ *i.e.*, three orders of magnitude smaller than for complex **2**. Furthermore, comparing the nature of primary interactions in complexes **3** and **4**, one would expect similar association constants for the two, which was not the

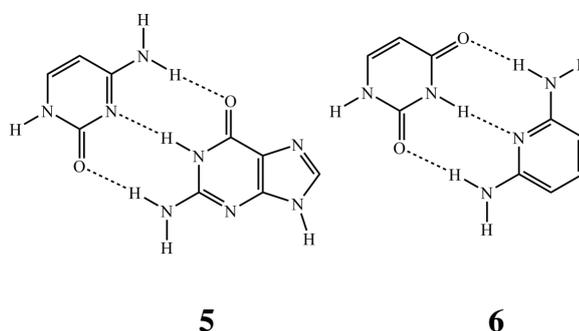
case. The experimentally observed association constant of **4** in chloroform is merely 90 M^{-1} ,⁵⁵ considerably smaller than the K_a of **3**.



Scheme 1.1. Schematic representation of triple hydrogen bonded analogues of DNA base pairs

A computational study in 1990 by Jorgensen and Pranata reports that the free energy of formation for the G-C complex **5** (Scheme 1.2) is -21.1 kcal/mol , and the free energy of formation of Uridine-2,6-diaminopyridine **6** is -11.4 kcal/mol .⁵⁷ These values were found to be in close agreement with experiments. However, the nature of primary interactions in complexes **5** and **6** are exactly the same. Jorgen and Pranata further came up with a very simple and brilliant explanation for the observed deviations in the binding behavior of hydrogen bonded complexes. They suggested that the consideration of only primary electrostatic interactions is not enough to explain the differential association constants of nitrogen base analogues. Perhaps the secondary electrostatic interactions between immediate nonhydrogen bonded acceptors and donor are significant enough to alter the association constant of hydrogen bonded complexes, as they are not much farther and have acquired certain amount of charge.⁵⁷ For instance, if we consider a triply hydrogen bonded complex, there will be three acceptor and three donor atoms in each complex. These acceptor and donor atoms on different partners can be arranged in three different ways: (i) all acceptors on one partner and all donors on the other, (ii) two consecutive acceptors and a donor on one partner and the complementary arrangement on the other partner and (iii) donors and acceptors following an alternative arrangement on each partner. Three representative cases of triply hydrogen bonded complexes have been shown below in Figure 1.6. Let us assume that all the donor and acceptor atoms in these three cases are the same and that they carry the same partial charge as well. Looking at all

possible direct interactions between the donor and acceptor atoms (only frontier hydrogen bonded atoms) of hydrogen bond constituting partners separately - there exist three primary and four secondary electrostatic interactions in all of the three cases. The primary interactions in all of the three cases are attractive. However, the nature of secondary electrostatic interactions varies in each case. All the secondary interactions in case 1 are attractive. In case 2, two of the secondary interactions are attractive and two are repulsive, while all of the secondary interactions in case 3 are repulsive. Based on this analysis, they predicted that the hydrogen bonded complexes of type 1 will have the highest association constant followed by those of type 2 and then those of type 3.⁵⁷ In a further study, they have shown that the relative association constant of some representative nitrogen base analogues follow this order.⁵⁸



Scheme 1.2. A schematic representation of the G-C base pair and the uridine-2,6-diaminopyridine base pair.

The breakthrough idea suggested by Jorgensen and coworkers triggered experimentalists to design hydrogen bonded complexes with all donors on one and acceptors on another partner in order to get the highest association constant. In many of the cases, they have obtained agreeable results.⁵⁹⁻⁶⁶ However, some countering examples had also been observed in the meantime in some of the studies, where Jorgensen's hypothesis failed completely.^{59-60,67} In 2001, Popelier *et al.* in their comprehensive QTAIM (Quantum Theory of Atoms-in-Molecules) study on 28 nitrogen base pairs complexes showed that the electrostatic energy of interaction between many remote atom pairs across a hydrogen bond is also influential to the binding, and hence the consideration of the electrostatic interactions of all the atoms of one partner with all the atoms of the other may be necessary in order to get the proper picture of the long range electrostatic influence on the binding.¹⁷ In Figure 1.7,

the electrostatic interaction energy between different atom pairs of the C-G complex has been shown, as described in reference 17. It is clear from Figure 1.7 that the interaction energies for certain atom pairs (*e.g.*, CN pair) are non-negligible even at a distance of 17 Å. Thus, they have suggested that considering only secondary electrostatic interactions between frontier hydrogen bonded atoms, for the sake of simplicity, is incorrect and may lead to erroneous results, and that the efficacy of frontier atom secondary electrostatic interactions are only limited to specific cases.¹⁷

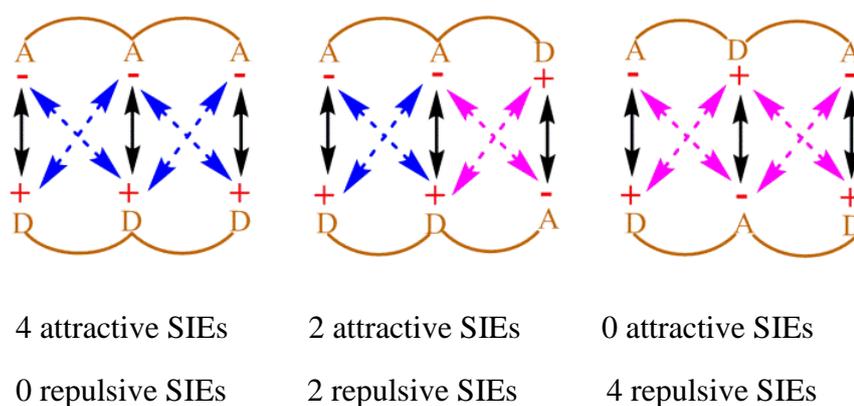


Figure 1.6. Electrostatic interactions between H-bond acceptors and donors in triply H-bonded complexes; A and D denote hydrogen bond acceptor and donor atoms respectively; black, blue and pink arrows represent attractive primary, attractive secondary and repulsive secondary interactions, respectively; the association constant decreases with an increasing number of repulsive secondary interactions.

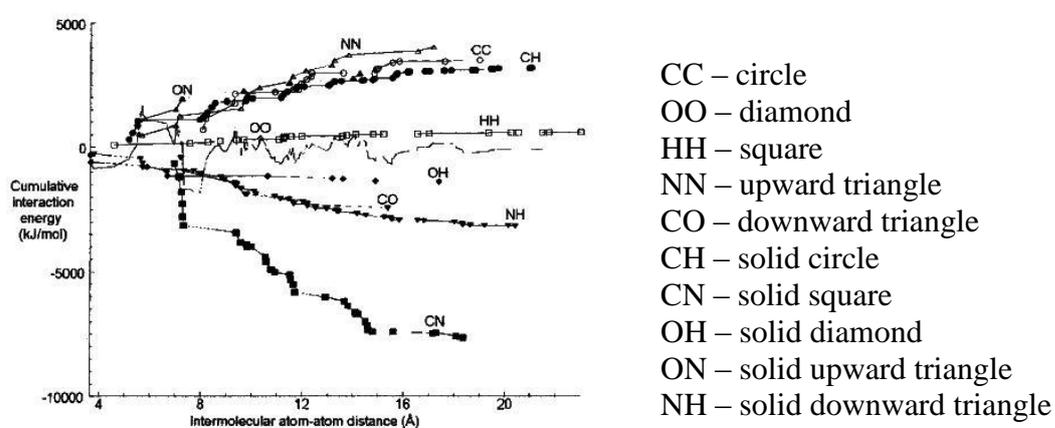


Figure 1.7. The energy partitioning of atom types in the G-C base pair as reported in reference 17. The cumulative interaction energy is the energy between atom types of two different bases in the G-C complex.

These studies established the decisive contributions of long range electrostatic interactions on the overall binding energy of electrostatic dominated systems. However, merely the consideration of energy to fully understand the strength of electrostatic dominated noncovalent interactions is not sufficient; as the energy is a scalar quantity and considering it alone will lead to loss in a certain amount of information regarding the binding. What is referred to specifically here is the directionality. Chapter 3 of this thesis will describe in detail the directional aspect of these interactions, as well as their consequences.

1.9 Objective of the Thesis

Considering the above benchmarking studies, we have asked three fundamental questions that are the basis of this thesis:

1. Since long range electrostatic interactions are significant, should they not be directional?
2. If long range electrostatic interactions are indeed directional, what could be the possible implications?
3. How would the directional long range electrostatic interactions be exploited in understanding and designing new chemical systems of prime importance?

1.10 Scope of the thesis

Long range secondary electrostatic interactions have been shown to have a distinct and deterministic influence on the association constant of multi-hydrogen bonded systems, signifying that the structure, orientation, and position of even distant substituents on the hydrogen bonding partners will have a decisive impact on the hydrogen bond strength.¹⁷ Intriguingly, in *in vivo* or homogeneous *in vitro* environments, hydrogen bonded systems largely exist surrounded by a quantum of solvent molecules. The long range secondary electrostatic influence exerted by the explicit presence of solvent molecules similarly has the potential to maneuver hydrogen bond strength, depending upon the orientation, position and structure of the solvent molecules. The current literature does not classify any study, to the best of our knowledge, that precisely pinpoints the effect of explicit solvent molecules on the

stability or strength of hydrogen bonds that are less exposed to the solvent or are present at the distal or interior sites in the chemical architecture of solutes (*e.g.*, proteins and carbohydrates.). Also, all the *in vivo* and homogeneous *in vitro* reactions happen in the solvent environment where the solvent plays a critical role in defining the barrier of key rate determining steps and hence the rate of the reactions. Since covalent bonds are believed to be comprised of many charge-separated resonance contributors,⁶⁸ does not the local electric field created by explicit solvent molecules affect the reaction barriers? The long range electrostatic influence on reaction barriers can also be correlated with the reactions that are being facilitated by the external architecture of molecular cages.⁶⁹ Furthermore, secondary coordination spheres are found to be extremely important in enzyme catalyzed reactions. These coordination spheres in enzymes are provided by amino acid residues that surround the active sites of the enzyme.⁷⁰ It has been seen in the past that a mimic of only active sites of enzymes leads to a weaker reactivity of coordination complexes.⁷¹ Therefore, the question that has to be asked here is - what role does the secondary coordination sphere play in enzyme catalysis? α -helices and β -pleated sheets, the two most common secondary structural motifs that are present in a typical protein have completely distinct hydrogen bonding patterns.⁷² These differential patterns in the hydrogen bonding bring disparity in stabilization of protein structural motifs.⁷³ However, deeper insights into why the differences in hydrogen bonding patterns affects the hydrogen bond strength of protein secondary structural motifs are still lacking. The other question of real significance is whether long range electrostatic interactions lead to chemical selectivity.

In this thesis, we have aimed to answer some of these problems to a certain extent and touched upon each of the aforementioned areas. First of all, we have proposed a method that suggests estimating the impact of long range secondary electrostatic interactions in terms of the electrostatic force, which will take care of the directional nature of these interactions. Later on, we have investigated certain representative examples that fall into the previously mentioned areas to answer the questions that are of fundamental interest to chemists, biologists and material science researchers.

The thesis as such is divided into 7 chapters. The first chapter, the current one, is dedicated to providing an overall view explaining the importance of the

electrostatic dominated noncovalent interactions and their long range influence. The second chapter explains the fundamentals of density functional theory, the method that has been employed extensively to carry out the work that has been reported in this thesis. The subsequent five chapters will deal with specific examples to showcase the directional influence of long range electrostatic interactions.

1.11 References

1. E. V. Anslyn, D. A. Dougherty *Modern Physical Organic Chemistry*, University Science Books, **2006**.
2. P. Hobza; K. Müller-Dethlefs, *Non-covalent Interactions. Theory and Experiment*, RSC Publishing, Cambridge, 2010.
3. E. A. C. Davie, S. M. Mennen, Y. Xu, S. J. Miller, *Chemical reviews*, 2007, **107**, 5759-5812; J.-M. Lehn, *Chemical Society Reviews*, 2007, **36**, 151-160; L. Brunsveld, B. Folmer, E. Meijer, R. Sijbesma, *Chemical Reviews*, 2011, **101**, 4071-4098; C. B. Aakeröy, K. R. Seddon, *Chemical Society Reviews*, 1993, **22**, 397-407; F. Zhao, M. L. Ma, B. Xu, *Chemical Society Reviews*, 2009, **38**, 883-891; A. G. Doyle, E. N. Jacobsen, *Chemical reviews*, 2007, **107**, 5713-5743.
4. T. Šmejkal, B. Breit, *Angewandte Chemie*, 2008, **120**, 4010-4013; R. R. Knowles, E. N. Jacobsen, *Proceedings of the National Academy of Sciences*, 2010, **107**, 20678-20685; K. Biradha, *CrystEngComm*, 2003, **5**, 374-384; A. Bauzá, T. J. Mooibroek, A. Frontera, *CrystEngComm*, 2016, **18**, 10-23; R. G. Gonnade, M. S. Shashidhar, M. M. Bhadbhade, *Journal of the Indian Institute of Science*, 2012, **87**, 149; B. Sellergren, M. Lepistö, K. Mosbach, *J. Am. Chem. Soc.*, 1988, **110**, 5853-5860; T. Rossow, S. Seiffert, *Supramolecular Polymer Networks and Gels, Vol. 268* (Eds.: S. Seiffert), *Springer*, 2015, pp. 1-46; D. J. Hill, M. J. Mio, R. B. Prince, T. S. Hughes, J. S. Moore, *Chemical reviews*, 2001, **101**, 3893-4012; D.-W. Zhang, X. Zhao, J.-L. Hou, Z.-T. Li, *Chemical Reviews*, 2012, **112**, 5271-5316; D. Wang, G. Tong, R. Dong, Y. Zhou, J. Shen, X. Zhu, *Chemical Communications*, 2014, **50**, 11994-12017; H. Yang, B. Yuan, X. Zhang, O. A. Scherman, *Accounts of chemical research*, 2014,

- 47, 2106-2115; B. R. Beno, K. S. Yeung, M. D. Bartberger, L. D. Pennington, N. A. Meanwell, *Journal of medicinal chemistry*, 2015, **58**, 4383-4438; T. F. A. De Greef, M. M. J. Smulders, M. Wolffs, A. P. H. J. Schenning, R. P. Sijbesma, E. W. Meijer, *Chem. Rev.*, 2009, **109**, 5687–5754.
5. H. Lodish, A. Berk, P. Matsudaira, C. A. Kaiser, M. Krieger, M. P. Scott, L. Zipursky, J. Darnell, *Noncovalent bonds – Molecular Cell Biology*, 5th edition, 2000.
6. B. Alberts *et al.*, *Molecular Biology of the Cell* (Garland, New York, ed. 3, 1994); H. Lodish *et al.*, *Molecular Cell Biology* (Scientific American Books, New York, 5th edition, 2004).
7. C. A. Schalley, *Noncovalent Bonding in Supramolecular Chemistry*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2007; pp 1–4.
8. K. Müller-Dethlefs, P. Hobza, *Chem. Rev.*, 2000, **100**, 143–167.
9. S. Hammes-Schiffer, S. J. Benkovic, *Annu Rev Biochem*, 2006, **75**, 519-541; J. P. Changeux, S. J. Edelstein, *Science*, 2005, **308**, 1424-1428; I. Bahar, A. J. Rader, *Curr Opin Struct Biol*, 2005, **15**, 586-592; F. Tama, C. L. Brooks, *Annu Rev Biophys Biomol Struct*, 2006, **35**, 115-133; Ivet Bahar, Chakra Chennubhotla and Dror Tobi, *Current Opinion in Structural Biology*, 2007, **17**, 633–640.
10. R. K. Bledsoe, V. G. Montana, T. B. Stanley, C. J. Delves, C. J. Apolito, D. D. McKee, T. G. Consler, D. J. Parks, E. L. Stewart, T. M. Willson, M. H. Lambert, J. T. Moore, K. H. Pearce, H. E. Xu, *Cell*, 2002, **110**, 93–105; G. G. J. M. Kuiper, B. Carlsson, K. Grandien, E. Enmark, J. H. Ggblad, S. Nilsson, and J.-Å. Gustafsson, *Endo*, 1997, **138**, 863-870.
11. P. Gilli, V. Bertolasi, V. Ferretti, G. Gilli, *J. Am. Chem. Soc.*, 1994, **116**, 909-915, 1994; J. W. Larson, T. B. McMahon, *J. Am. Chem. Soc.*, 1983, **105**, 2944-2950.
12. J. R. Premkumar, D. Vijay, G. N. Sastry, *Dalton Trans.*, 2012, **41**, 4965–4975; D. Umadevi, G.N. Sastry, *J. Phys. Chem. C*, 2011, **115**, 9656–9667; A. S. Mahadevi, G.N. Sastry, *Chem. Rev.*, 2013, **113**, 2100–2138.

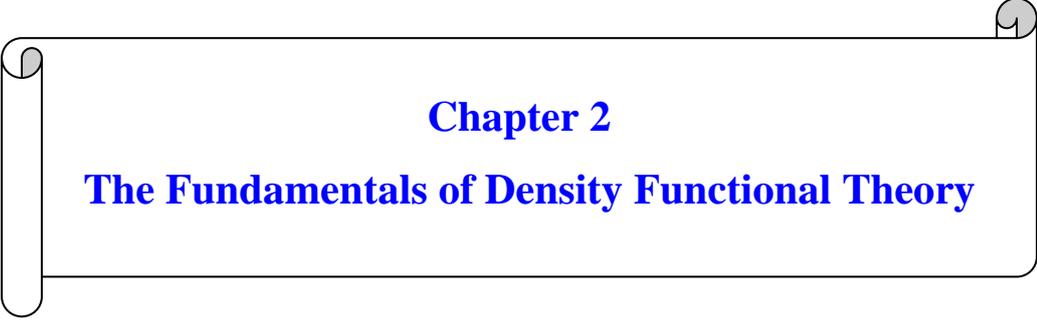
13. a) T. Clark, P. Politzer, J. S. Murray, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2015, **5**, 169–177; b) T. Clark, in *The Chemical Bond* (Eds.: G. Franking, Sason Shaik), Wiley, Chapter 18, pp 523-536.
14. van der Waals, J. D. *Over de Continuïteit van den Gas- en Vloeistofoestand* (On the continuity of the gaseous and liquid state), Ph.D. Dissertation, University of Leiden, 1873.
15. F. London, *Zeitschrift für Physik*, 1930, **63**, 245.
16. H. Hellmann (Moscow – 05.04.1935). On the quantum mechanical calculation of polarizabilities and dispersion forces (in German). *Acta Physicochim. U.R.S.S.*, 1935, **2**, 273-290.
17. P. L. A. Popelier, L. Joubert, *J. Am. Chem. Soc.*, 2002, **124**, 8725–8729.
18. M. J. Plevin, D. L. Bryce, J. Boisbouvier, *Nat. Chem.*, 2010, **2**, 466–471.
19. A. Subha Mahadevi, G. Narahari Sastry, *Chem. Rev.* 2016, **116**, 2775–2825.
20. M. K. Tiwari, K. Vanka, *Chemical Science*, 2017, **8**, 1378 – 1390.
21. P. Ball, *Chem. Rev.*, 2008, **108**, 74–108; M. K. Tiwari, A. K. Jana, K. Vanka, N. Sengupta, *Phys. Chem. Chem. Phys.*, 2016, **18**, 5910-5924; C. Beeson, N. Pham, G. Jr. Shipps, T. A. Dix, *J. Am. Chem. Soc.*, 1993, **115**, 6803–6812; M. G. Sarwar, B. Dragisic, L. J. Salsberg, C. Gouliaras, M. S. Taylor, *J. Am. Chem. Soc.*, 2010, **132**, 1646–1653; L. R. Olano, S. W. Rick, *J. Am. Chem. Soc.*, 2004, **126**, 7991; A. C. English, S. H. Done, L. S. Caves, C. R. Groom, R. E. Hubbard, *Proteins*, 1999, **37**, 628-640; K. Kalyanasundaram, J. K. Thomas, *J. Am. Chem. Soc.*, 1977, **99**, 2039-2044.
22. L. J. Prins, D. N. Reinhoudt, P. Timmerman, Noncovalent Synthesis Using Hydrogen Bonding. *Angew. Chem. Int. Ed.*, 2001, **40**, 2382–2426.
23. C. A. Hunter, H. L. Anderson, *Angew. Chem. Int. Ed.* 2009, **48**, 7488–7499; D. H. Williams, M. S. Westwell, *Chem. Soc. Rev.*, 1998, **27**, 57–63.
24. J. P. Wagner, P. R. Schreiner, *Angew. Chem. Int. Ed.*, 2015, **54**, 12274–12296.
25. A. J. Stone, *The Theory of Intermolecular Forces*, 2 ed.; Oxford University Press, 2013.
26. M. Siedlecka, G. Goch, A. Ejchart, H. Sticht, A. Bierzyski, *Proc. Natl. Acad. Sci. U.S.A.* 1999, **96**, 903-908; J. X. Yang, K. Zhao, Y. X. Gong, Y. Vologodskii, N. R. Kallenbach, *J. Am. Chem. Soc.*, 1998, **120**, 10646-10652.

-
27. M. S. Searle, D. H. Williams, *J. Am. Chem. Soc.*, 1992, **114**, 10690–10697.
28. J. G. Brandenburg, M. Hochheim, T. Bredow, S. Grimme, *J. Phys. Chem. Lett.*, 2014, **5**, 4275–4284.
29. C. Riplinger, B. Sandhoefer, A. Hansen, F. Neese, *J. Chem. Phys.*, 2013, **139**, 134101.
30. B. J. ki , R. M. ki , K. Szalewicz, *Chem. Rev.*, 1994, **94**, 1887-1930.
31. D. Cremer, *John Wiley & Sons, Ltd. WIREs Comput Mol Sci*, 2011, **1**, 509–530.
32. S. Kristyán, P. Pulay, *Chem. Phys. Lett.*, 1994, **229**, 175–180; J. M. Pérez-Jordá, A. Becke, *Chem. Phys. Lett.*, 1995, **233**, 134–137.; P. Hobza, J. Šponer, T. Reschel, *J. Comput. Chem.*, 1995, **16**, 1315–1325.
33. K. E. Riley, M. Pitoňák, P. Jurečka, P. Hobza, *Chem. Rev.*, 2010, **110**, 5023–5063; S. Grimme, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 211–228; L. A. Burns, A. Vazquez-Mayagoitia, B. G. Sumpter, C. D. Sherrill, *J. Chem. Phys.*, 2011, **134**, 084107.
34. S. F. Boys, F. Bernardi, *Mol. Phys.*, 1970, **19**, 553.
35. M. L. Senent, S. Wilson, *International Journal of Quantum Chemistry*, 2001, **82**, 282–292.
36. F. Jensen, *Chem Phys Lett*, 1996, **261**, 633.
37. T. Bredow, K. Jug, *Theor. Chem. Acc.*, 2005, **113**, 1–14; M. Hennemann, T. Clark, *J. Mol. Model.*, 2014, **20**, 2331.
38. J. Dannenberg, *J. Mol. Struct.: Theochem*, 1997, **401**, 279–286; G. I. Csonka, J. G. Ángyán, *J. Mol. Struct.: Theochem*, 1997, **393**, 31–38; P. Winget, C. Selçuki, A. H. C. Horn, B. Martin, T. Clark, *Theor. Chem. Acc.*, 2003, **110**, 254–266.
39. M. Korth, *J. Chem. Theory Comput.*, 2010, **6**, 3808–3816; J. Řezáč, K. E. Riley, P. Hobza, *J. Chem. Theory Comput.*, 2012, **8**, 141–151.
40. J. G. Brandenburg, S. Grimme, *J. Phys. Chem. Lett.*, 2014, **5**, 1785–1789.
41. J. J. P. Stewart, *J. Comput. Chem.*, 1989, **10**, 209–220; B. Ahlswede, K. Jug, *J. Comput. Chem.*, 1999, **20**, 563–572; W. Weber, W. Thiel, *Theor. Chem. Acc.*, 2000, **103**, 495–506.
42. R. W. F. Bader, M. T. Carroll, J. R. Cheeseman, C. Chang, 1987, **109**, 7968-7979.
-

-
43. P. Politzer, J. S. Murray, T. Clark, *Phys. Chem. Chem. Phys.*, 2013, **15**, 11178—11189.
44. P. Politzer, J. S. Murray, T. Clark, G. Resnati, *Phys. Chem. Chem. Phys.*, 2017, **19**, 32166—32178.
45. K. Fukui, T. Yonezawa, C. Nagata, *J. Chem. Phys.*, 1957, **26**, 831-841.
46. L. Fielder, J. Gao, D. G. Truhlar, *J. Chem. Theory Comput.*, 2011, **7**, 852-856.
47. A. J. Stone, C.-S. Tong, *J. Comput. Chem.*, 1994, **15**, 1377-1392; A. J. Stone, *The Theory of Intermolecular Forces*, 1996, Clarendon, Oxford; G. Schürer, P. Gedeck, M. Gottschalk, T. Clark, *Int. J. Quantum Chem.*, 1999, **75**,17; B. Martin, P. Gedeck, T. Clark, *Int. J. Quantum Chem.*, 2000, **77**, 473-497; G. J. Williams, A. J. Stone, *J. Chem. Phys.*, 2003, **119**, 4620-4628.
48. J. D. Watson, F. H. C. Crick, *Nature*, 1953, **171**, 737–738.
49. J. Donohue, *Proc. Natl. Acad. Sci. U.S.A.*, 1956, **42**, 60. J. Donohue, K. N. Trueblood, *J. Mol. Biol.*, 1960, **2**, 363.
50. V. Poltev, N. V. J. Shulyupina, *Biomol. Struct. Dyn.*, 1986, **3**, 739.
51. H. De Voe, I. Tinoco, Jr. *J. Mol. Biol.*, 1962, **4**, 500; B. Pullman, P. Claverie, J. Caillet, *J. Proc. Natl. Acad. Sci. U.S.A.*, 1966, **55**, 904; M. N. Stamatiadou, T. J. Swisler, J. R. Rabinowitz, R. Rein, *Biopolymers*, 1972, **11**, 1217; Z. G. Kudritskaya, V. J. Danilov, *Theor. Biol.*, 1976, **59**, 303; R. G. A. R. Maclagan, *Aust. J. Chem.* 1979, **32**, 1635; R. Rein, In *Intermolecular Interactions: From Diatomics to Biopolymers*, B. Pullman, Ed., Wiley-Interscience: New York, 1978; pp 308-363; J. Langlet, P. Claverie, F. Caron, J. C. Boeue, *Int. J. Quantum. Chem.*, 1981, **20**, 299; P. Hobza, C. Sandorfy, *J. Am. Chem. Soc.*, 1987, **109**, 1302-1307.
52. Kyogoku *et al.*, *Biochim. Biophys. Acta*, 1969, **179**, 10.
53. Kelly *et al.*, *J. Am. Chem. Soc.*, 1989, **111**, 3744.
54. Kyogoku *et al.*, *Proc. Nutl. Acad. Sci. U.S.A.*, 1967, **57**, 250.
55. Hamilton *et al.*, *J. Am. Chem. Soc.*, 1987, **109**, 5035.
56. Base Pairing in DNA: Unusual Patterns, Wiley, W. H Gmeiner, B. J Walberer, Published Online: 2001.
57. W. L. Jorgensen, J. Pranata, *J. Am. Chem. Soc.*, 1990, **112**, 2008-2010.
58. J. Pranata, S. G. Wierschke, W. L. Jorgensen, *J. Am. Chem. Soc.*, 1991, **113**, 2810-2819.
-

59. B. A. Blight, C. A. Hunter, D. A. Leigh, D. A., H. McNab, P. I. Thomson, *Nature Chemistry*, 2011, **3**, 244-248.
60. T. J. Murray, S. C. Zimmerman, *J. Am. Chem. Soc.*, 1992, **114**, 4010-4011.
61. B. A. Blight, C.-C. Amaya, S. Djurdjevic, M. Kaller, D. A. Leigh, F. M. McMillan, H. McNab, A. M. Slawin. *J. Am. Chem. Soc.*, 2009, **131**, 14116-14122.
62. S. Djurdjevic, D. A. Leigh, H. McNab, S. Parsons, G. Teobaldi, F. Zerbetto, *J. Am. Chem. Soc.*, 2007, **129**, 476-477.
63. D. A. Leigh, C. C. Robertson, A. M. Slawin, P. I. Thomson, *J. Am. Chem. Soc.*, 2013, **135**, 9939-9943.
64. J. Sartorius, H. J. Schneider, *Chemistry-A European Journal*, 1996, **2**, 1446-1452.
65. F. H. Beijer, H. Kooijman, A. L. Spek, R. P. Sijbesma, E. Meijer, *Angew. Chem. Int. Ed.*, 1998, **37**, 75-78.
66. J. R. Quinn, S. C. Zimmerman, J. E. D. Bene, I. Shavitt, *J. Am. Chem. Soc.*, 2007, **129**, 934-941.
67. H. Zeng, R. S. Miller, R. A. Flowers, B. Gong, *J. Am. Chem. Soc.*, 2000, **122**, 2635-2644; R. R. Gardner, S. H. Gellman, *Tetrahedron*, 1997, **53**, 9881-9890; J. Yang, S. H. Gellman, *J. Am. Chem. Soc.*, 1998, **120**, 9090-9091; K. S. Jeong, T. Tjivikua, A. Muehldorf, G. Deslongchamp, M. Famulok, J. Jr. Rebek, *J. Am. Chem. Soc.*, 1991, **113**, 201-209.
68. G. Sini, P. Maitre, P. C. Hiberty, S. S. Shaik, *J. Mol. Struct. Theochem*, 1991, **229**, 163-188 ; S. Shaik, D. Danovich, W. Wu, P. C. Hiberty, *Nature Chem.*, 2009, **1**, 443-449.
69. J. Kang, J. R. Jr, *Nature*, 1997, **385**, 50-52.
70. M. S. Rogers, E. M. Tyler, N. Akyumani, C. R. Kurtis, R. K. Spooner, S. E. Deacon, S. Tamber, S. J. Firbank, K. Mahmoud, P. F. Knowles, S. E. V. Phillips, M. J. McPherson, D. M. Dooley, *Biochemistry*, 2007, **46**, 4606-4618; S. J. Benkovic, S. Hammes-Schiffer, *Science*, 2003, **301**, 1196-1202; B. A. Springer *et al.*, *Chem. Rev.* 1994, **94**, 699; Y. Lu, J. S. Valentine, *Curr. Opin. Struct. Biol.*, 1997, **7**, 495.
71. D. M. Kurtz, *Chem. Rev.*, 1990, **90**, 585, and references therein; M. Mukai *et al.*, *J. Am. Chem. Soc.*, 1997, **119**, 1758. Y.-D. Wu *et al.*, *Inorg. Chem.*, 1992,

- 31, 718; C. E. MacBeth, A. P. Golombek, V. G. Young, C. Yang, K. Kuczera, M. P. Hendrich, A. S. Borovik, *Science*, 2000, **289**, 938.
72. A. Fersht, *Structure and Mechanism in Protein Science*; Freeman: New York, 1999.
73. S. Vijayakumar, S. Vishveshwara, G. Ravishanker, D. L. Beveridge, *Biophysical Journal*, 1993, **65**, 2304-2312; R. Tomar, V. K. Dubey, M. V. Jagannadham, *Biochimie*, 2009, **91**, 951-960.



Chapter 2
The Fundamentals of Density Functional Theory

Chapter 2

The Fundamentals of Density Functional Theory

Abstract: We have exploited density functional theory (DFT) as a tool for investigating problems of interest in this thesis. In this chapter, the development and fundamental aspects of DFT have been described in brief. The DFT methods offer a way to compute energy and other properties of atomic and molecular systems as a functional of electron density, which is otherwise a physical observable and has direct consequences on the chemical and physical behavior of chemical systems. These methods work excellently under the Born-Oppenheimer approximation for stationary point calculations of ground state geometries and provide an accurate and computationally less expensive alternative to wave function based methods, which are unrealistic for computing properties of real (rather than model) molecular systems, due to their high scalability.

2.1 Introduction

The laws of thermodynamics and kinetics have paved a reliable and precise means of understanding the feasibility of chemical transformations. Specifically, they have provided an understanding of the change of the energy (the enthalpy, the free energy, the activation energy, the association energy, the binding energy or the interaction energy) during chemical changes, for a given system. Notably, the differential stabilities of reaction intermediates and transitions states essentially provide complete information about the feasibility and the pathway of a chosen chemical reaction and/or interaction. In course of time, several methods, including experimental techniques, have evolved to provide numbers corresponding to energy changes, in order to yield a simple, precise and accurate understanding of chemical changes. The laws and postulates of quantum mechanics (QM) provide one of the routes to obtain these energy terms based on pure mathematical formulations. However, it has been recognized that the *ab initio* quantum chemical methods (coupled cluster¹ and configuration interaction,² to name two such approaches) that solve the non-relativistic Schrodinger equations with high accuracy are highly expensive, and are

thus impractical to be employed for a real systems. The best theoretical method available to date to handle larger chemical systems is density functional theory (DFT).³ With recent advancements, DFT has become able to provide a range of computational tools that are necessary to compute the complete set of properties for any given chemical system with a sufficiently reliable accuracy.³ Thus, in recent years, DFT has emerged as a pivotal computational tool for investigating various aspects of chemical reactions and/or chemical interactions and has emerged as the popular option both among computational and experimental chemists. Considering the aforementioned advantages, we have chosen DFT as the prime theoretical tool for investigating problems of our interest, a description of which have been provided in the succeeding chapters (Chapter 3 to Chapter 7) of this thesis. At present, the fundamentals of DFT have been described in subsequent sections of this chapter. In order to lay the foundations of DFT, the elementary concepts of quantum chemistry have also been reviewed in short in this chapter.

2.2 Elementary Quantum Mechanics

2.2.1 The Schrödinger Equation

The central goal of modern quantum mechanics is to solve the time dependent non-relativistic Schrödinger equation that was proposed by Erwin Rudolf Josef Alexander Schrödinger in 1926. However, the time dependent interactions are often not significant in most of the problems relating to chemistry, if one considers the potential energy of the system to be independent of time. Therefore, the time independent form of the Schrödinger equation (TISE) is generally considered as the chief equation in modern quantum chemistry. Henceforth the TISE will be referred to as the Schrödinger equation (SE). The many body time-independent Schrödinger equation for the system with n number of electrons and m number of nuclei is given by

$$\hat{H}\Psi(\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n, \vec{r}_1, \vec{r}_2, \vec{r}_3, \dots, \vec{r}_m) = E\Psi(\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n, \vec{r}_1, \vec{r}_2, \vec{r}_3, \dots, \vec{r}_m) \quad (2.1)$$

where \hat{H} denotes the molecular Hamiltonian operator, \vec{x} and \vec{r} are the coordinates of respective electrons and nuclei, $\Psi(\vec{x}_n, \vec{r}_m)$ is a many particle wave function for a given

system and is a function of the $3n$ space coordinates and n spin coordinates of electrons and $3m$ space coordinates of nuclei, and E is the total energy of the system.

Equation (2.1) is also called the *eigenvalue equation*, where E is the eigenvalue, $\Psi(\vec{x}_n, \vec{r}_m)$ is an eigenfunction of operator \hat{H} , which is a *Hermitian operator* and returns only real numbers as the eigenvalues. The wave function $\Psi(\vec{x}_n, \vec{r}_m)$ has to fulfill certain requirements, which defines $\Psi(\vec{x}_n, \vec{r}_m)$ to be well behaved, in order to be allowed for quantum chemical consideration.

The Hamiltonian operator in atomic unit is represented as,

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^n \nabla_i^2 - \frac{1}{2} \sum_{a=1}^m \frac{\nabla_a^2}{m_a} - \sum_{a=1}^m \sum_{i=1}^n \frac{Z_a}{r_{ia}} + \sum_a \sum_{b>a}^m \frac{Z_a Z_b}{r_{ab}} + \sum_{i=1}^n \sum_{j>i}^n \frac{1}{r_{ij}} \quad (2.2)$$

The first two terms in the Hamiltonian represent the kinetic energy of electrons and the nuclei respectively. The last three terms are the nuclear-electron attraction, nuclear-nuclear repulsion and electron-electron repulsion respectively. Indices i and j indicate a total number of n electrons, whereas indices a and b denote a total of the m nuclei of the system. Other associated terms in the equation have their usual meaning.

To apply QM theory to address real life chemical problems, we need to solve the SE for multi-electron multi-nuclear molecular systems. However, the exact solution to the SE is limited to only few simplistic ideal cases, such as a particle in a box, the harmonic oscillator, the rigid rotor and the hydrogen atom. Hence, to make this theory applicable to larger systems, approximate methods have been proposed over the years. The first approximation that comes into the consideration is the Born-Oppenheimer approximation, which is believed to be a good approximation for the stationary point calculations.

2.2.2 The Born-Oppenheimer approximation

It is evident that nuclei are much heavier than electrons. Hence, nuclei move much slower relative to electrons. Therefore, we can assume that all of the electrons exist in the field of fixed nuclei. The nuclear kinetic energy can therefore be considered to be zero and the nuclear potential energy to merely be a constant. This is known as the

Born-Oppenheimer approximation. The *Born-Oppenheimer approximation* allows the separation of the SE into the electronic and the nuclear parts. Ultimately, the Hamiltonian shown in Equation (2.2) can be simplified to only the electronic part \hat{H}_{el} as,

$$\hat{H}_{el} = -\frac{1}{2} \sum_{i=1}^n \nabla_i^2 - \sum_{a=1}^m \sum_{i=1}^n \frac{Z_a}{r_{ia}} + \sum_{i=1}^n \sum_{j>i}^n \frac{1}{r_{ij}} = \hat{T} + \hat{V}_{ne} + \hat{V}_{ee} \quad (2.3)$$

The solution of the Schrödinger Equation with the electronic Hamiltonian \hat{H}_{el} yields the electronic energy E_{el} when \hat{H}_{el} is operated on the electronic wave function Ψ_{el} ,

$$\hat{H}_{el} \Psi_{el} = E_{el} \Psi_{el} \quad (2.4)$$

The total energy can be obtained as the sum of the electronic energy and the nuclear repulsion energy E_{nn} as follows:

$$E_{tot} = E_{el} + E_{nn} \quad (2.5)$$

Where E_{nn} is given by,

$$E_{nn} = \sum_a^m \sum_{b>a}^m \frac{Z_a Z_b}{r_{ab}} \quad (2.6)$$

2.2.3 Wave Function, Probability and Normalization

We have seen, in the previous section, how the energy E can be obtained from operating Hamiltonian \hat{H} on $\Psi(x_n, r_m)$. In this section, we will give a brief description regarding the wave function $\Psi(x_n, r_m)$. The definition of the wave function has come from the Copenhagen interpretation. According to this, a wave function is a function that contains all information of a given system for the state that has been represented by this wave function and all the information of that system for this state can be extracted from the wave function on operating it with suitable operators. However, the wave function itself has no physical significance as it is not a physical observable. It returns the value of an observable only when is treated with an operator, the origins of

which is in classical mechanics. The physical interpretation of a wave function comes instead with the square of the $\Psi(x_n, r_m)$, which gives the probability density

$$\text{probability density} = |\Psi(x_1, x_2, \dots, x_n)|^2 \quad (2.7)$$

The wave function $\Psi(x_n, r_m)$ can take on complex or negative values but not the probability density. The probability of finding electrons in a given volume element can be obtained from the following expression:

$$|\Psi(x_1, x_2, \dots, x_n)|^2 dx_1 dx_2 \dots dx_n \quad (2.8)$$

Equation (2.8) represents the probability that electrons 1, 2, ..., n are found simultaneously in volume $dx_1 dx_2 \dots dx_n$. Since the electrons are indistinguishable, the probability density remains unchanged if two electrons interchange their positions,

$$|\Psi(x_1, x_2, \dots, x_i, x_j, \dots, x_n)|^2 = |\Psi(x_1, x_2, \dots, x_j, x_i, \dots, x_n)|^2 \quad (2.9)$$

However, electrons are fermions with spin $s=1/2$. Therefore, Ψ must be antisymmetric with respect to interchange of the spatial and spin coordinates of any two electrons according to the quantum mechanical generalization of *Pauli's exclusion principle* ('no two electrons can occupy the same state').

$$\Psi(x_1, x_2, \dots, x_i, x_j, \dots, x_n) = -\Psi(x_1, x_2, \dots, x_j, x_i, \dots, x_n) \quad (2.10)$$

Furthermore, the probability of finding n electrons over all space should be unity. Hence, the integral of Equation (2.8) over the entire space should be equal to 1.

$$\int \dots \int |\Psi(x_1, x_2, \dots, x_n)|^2 dx_1 dx_2 \dots dx_n = 1 \quad (2.11)$$

A wave function that satisfies this condition is called a *normalized* wave function.

2.2.4 The Variational Principle

In Equation (2.1) there are two unknown quantities $\Psi(x_n, r_m)$ and E . In order to solve the SE for any chosen molecule, the prerequisite is to construct the Hamiltonian operator corresponding to the system and then to find the eigenfunction followed by

solving the SE to obtain the eigenvalue E for the system. However, there is no definite way to find the exact eigenfunction for a given system. The variational principle helps to offer a method to resolve this problem. If a system is in the state Ψ , the expectation value of energy for that particular system can be obtained by

$$E[\Psi] = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \quad (2.12)$$

Where $\langle \Psi | \hat{H} | \Psi \rangle = \int \Psi^* \hat{H} \Psi d\tau$.

Now, according to the variational principle ‘*The energy calculated using a guess wave function Ψ is always an upper bound to the original ground state energy (E_0) of the system of interest*’

$$E[\Psi] = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \geq E_0 = \frac{\langle \Psi_0 | \hat{H} | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle} \quad (2.13)$$

Here, Ψ_0 is true ground state wave function. A full minimization of the given functional $E[\Psi]$ with respect to all of the allowed n-electronic wave functions, under the Born-Oppenheimer approximation, will provide the true ground state Ψ_0 and the corresponding energy $E[\Psi] = E_0$. What “allowed” means in the present context is that the trial wave functions must follow certain criteria that ensure that these functions make physical sense. For instance, to be qualified as a wave function, Ψ must be continuous everywhere and be square integrable. Otherwise, the normalization of Equation (2.11) would not be possible. This can be expressed as,

$$E_0 = \min_{\Psi \rightarrow n} E[\Psi] = \min_{\Psi \rightarrow n} \langle \Psi | \hat{T} + \hat{V}_{ne} + \hat{V}_{ee} | \Psi \rangle \quad (2.14)$$

where $\Psi \rightarrow n$ indicates that Ψ is an acceptable n-electron wave function. However, one must note here that a search over *all acceptable* functions is practically impossible. However, the variational principle can also be applied to subsets of all allowed functions. One possible way to surpass this problem is to choose these subsets such that the minimization in Equation (2.14) can be done in some algebraic scheme, which will lead to the best approximation to the exact wave function that can

Like a function, a functional can also have derivatives. The formulation is similar to the derivatives of functions. The differentiation of a functional $F[g]$ is defined as,

$$\delta F[g] = F[g + \delta x] - F[g] = \int \frac{\delta F}{\delta g(x)} \delta g(x) dx \quad (2.17)$$

The rules of differentiation are also similar to the functions,

$$\frac{\delta}{\delta f(x)} (c_1 F_1 + c_2 F_2) = c_1 \frac{\delta F_1}{\delta f(x)} + c_2 \frac{\delta F_2}{\delta f(x)} \quad (2.18)$$

$$\frac{\delta}{\delta f(x)} (F_1 F_2) = \frac{\delta F_1}{\delta f(x)} F_2 + \frac{\delta F_2}{\delta f(x)} F_1 \quad (2.19)$$

2.3 The Fundamentals of Density Functional Theory

The conventional quantum mechanical treatments, which offer solutions to the SE, are wave function based, as they take wave function Ψ as input to compute the energy and other properties of the system. However, these methods were found to be highly expensive and this made theoretical chemists look out for other methods that are not only accurate but also computationally economical. The method that has naturally emerged as the best alternative of wave function dependent methods in terms of accuracy and efficacy is density functional theory (DFT). DFT provides an alternative route for computing the properties of a system by employing electron density as the fundamental parameter. In other words, the DFT utilizes the electron density as such a basic quantity that can be interpreted to predict chemical and physical properties of any system. Intriguingly, using electron density as the fundamental parameter to compute properties of quantum mechanical systems has always been a tempting area to theoretical chemists because of the following advantages it possesses over the wave function: (i) unlike the wave function, the electron density of a quantum mechanical system is an experimentally measurable quantity and (ii) the electron density depends only on three spatial coordinates, whereas the wave function of a system composed of n particles has a dependency on $4n$ variables: $3n$ for coordinates and n for spin. Thus, density based methods are expected to be more economical than the wave function based ones if one considers computational time and resources, which is what has been achieved with recent developments with the help of smart and robust algorithms that

have been implemented in many quantum chemical packages. Consequently, DFT is available now for computing all kinds of properties of many different types of systems. In the following subsections, we will give a brief introduction to density functional theory.

2.3.1 The Electron Density

In the section 2.3 we have learnt that the DFT postulates employing the electron density as a fundamental parameter to obtain all the essential properties of a quantum chemical system. In this section, we have described the theoretical formulation of the electron density, an ingredient on which the entire recipe of DFT is based. The electron density is defined as the probability to find one electron of arbitrary spin within a particular volume element while the rest of the electrons may be anywhere else in the space. The electron density is a physically observable quantity and can be measured employing experimental techniques such as X-ray diffraction, transmission electron microscopy (TEM), scanning tunneling microscopy (STM), and atomic force microscopy (AFM). Theoretically, the physical interpretation of Equation (2.8) directly leads to the electron density $\rho(\vec{r})$. Mathematical expression of the total electron density of a n-electronic system can be obtained as the following multiple integral of the modulus square of the wave function over all the spin coordinates (s) of all the electrons and over all but one of the spatial coordinates (\vec{r}),

$$\rho(\vec{r}) = n \int \dots \int |\Psi(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n)|^2 ds_1 d\vec{x}_2 \dots d\vec{x}_n \quad (2.20)$$

Here $\vec{x} = \vec{r}.s$, and $\rho(\vec{r})$ is the probability of finding any of the n electrons in the volume element $d\vec{r}_1$ but with arbitrary spin. The rest of the n-1 electrons will have arbitrary positions and spins in the state represented by Ψ . Strictly speaking, $\rho(\vec{r})$ represents the probability density, but calling it the electron density is a common practice. The multiple integral represents the probability of finding one particular electron in the volume element $d\vec{r}_1$ (here, $\vec{x}_1 = \vec{r}_1.s_1$). However, since electrons are indistinguishable, the probability of finding any electron out of total n at this position is just n times the probability of finding one particular electron. $\rho(\vec{r})$ follow some specific properties, which have been described below,

1. $\rho(\vec{r})$ is a non-negative function of the three spatial variables, which integrates to the total number of electrons and vanishes at infinity,

$$\rho(\vec{r}) \geq 0 \quad (2.21)$$

$$\int \rho(\vec{r}) d\vec{r}_1 = n \quad (2.22)$$

$$\rho(\vec{r} \rightarrow \infty) = 0 \quad (2.23)$$

2. $\rho(\vec{r})$ exhibits a maximum possessing a finite value at a given position of the atom due to the attractive force exerted by the positive nuclei. At this position, the gradient of density exhibits a discontinuity resulting in a *cusp* because of the singularity in the $-\frac{Z_a}{r_{ia}}$ part in the Hamiltonian (Equation (2.2)), as $r_{ia} \rightarrow 0$ at this position. The properties of the cusp are intimately related to the nuclear charge of the atom,

$$\lim_{r_{ia} \rightarrow 0} \left[\frac{\partial}{\partial r} + 2Z_a \right] \bar{\rho}(\vec{r}) = 0 \quad (2.24)$$

where Z_a is the nuclear charge and $\bar{\rho}(\vec{r})$ is the spherical average of $\rho(\vec{r})$.

3. $\rho(\vec{r})$ shows an asymptotic exponential decay for large distances from the nuclei for any system,

$$\rho(\vec{r}) \propto \exp[-2\sqrt{2I} |\vec{r}|] \quad (2.25)$$

where I is the exact first ionization energy of the system.

2.3.2 The Pair Density

The pair density is the probability of finding a pair of electrons with spins σ_1 and σ_2 simultaneously within two different volume elements $d\vec{r}_1$ and $d\vec{r}_2$ while the other $n-2$ electrons have arbitrary positions and spins. It is given as,

$$\rho(\vec{x}_1, \vec{x}_2) = n(n-1) \int \dots \int |\Psi(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n)|^2 d\vec{x}_3 \dots d\vec{x}_n \quad (2.26)$$

Like the electron density, the pair density is also a positive quantity and is normalized to the total number of non-distinct pairs of electrons, *i.e.*, $n(n-1)^*$. It is symmetric in the coordinates. The pair density is of great importance since it contains all information about electron correlation.

2.3.3 The Thomas-Fermi Model

Although the journey towards modern DFT began just a few decades ago, the first attempt to use electron density to obtain information about atomic and molecular systems has been dated back to the early days of quantum chemistry, just shortly after the introduction of the Schrödinger equation (1926). The first approximation of such a kind was proposed by Llewellyn Thomas (1927) and Enrico Fermi (1927). They introduced electron density in place of the wave function as a means to understanding the electronic structure of many-body systems. This model is known as the Thomas–Fermi (TF) model.⁴⁻⁶ In this quantum statistical model, Thomas and Fermi used the concept of the uniform electron gas with a constant electron density. They further derived the expression of the kinetic energy of a quantum mechanical system. The expression for the kinetic energy in the TF model can be obtained by following functional:

$$T_{TF}[\rho(\vec{r})] = \frac{3}{10} (3\pi^2)^{2/3} \int \rho^{5/3}(\vec{r}) d\vec{r} \quad (2.27)$$

All other electronic contributions due to nuclear-electron attraction and electron-electron repulsion, according to this model, are treated in a completely classical manner. The total energy can thus be written as:

$$E_{TF}[\rho(\vec{r})] = \frac{3}{10} (3\pi^2)^{2/3} \int \rho^{5/3}(\vec{r}) d\vec{r} - Z \int \frac{\rho(\vec{r})}{r} d\vec{r} + \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 \quad (2.28)$$

This model possesses the following limitations. It gives a very rough estimate of the actual kinetic energy of the system, as the electrons in this model have been considered to be part of a gas of a constant electron density. Furthermore, the exchange and correlation effects are totally ignored. Thus, upon implementing the TF

*It is also common to use a different normalization coefficient $n(n-1)/2$, which corresponds to the distinct number of pairs of electrons.

model for molecular systems, it has been observed that the model is unable to describe the existence of a chemical bond. Therefore, the TF model shows extremely poor performance in describing real molecular systems. However, the real importance of the TF model is not in the accuracy of the method but due to the fact that it was the first prescription for representing energy using only the electron density without any additional information. In particular, it should be noted that no recourse to the wave function was taken in this model.

Now that we have a functional expressing the energy in terms of the electron density alone, the next important step is to find and identify the correct density. In order to find out the correct density for use in Equation (2.28), Thomas and Fermi employed the variational principle. They assumed that the ground state of the system is related to the density, for which the expression of the total energy is minimized under the constraint $\int \rho(\vec{r}) d\vec{r}_1 = n$. However, at that point of time, it was still unknown whether expressing the total energy of a system in terms of density is physically justified and whether a method employing the variational principle on the density is conceptually acceptable in this context. Thus, this model has mostly come out of intuition than solid physical ground. Therefore, this model is considered as a limited model and in the present context, the TF approach only has historical value rather than any practical significance.

2.3.4 Hohenberg-Kohn (HK) Theorems

The density functional theory we are acquainted with and exercise today was born in 1964 when a landmark paper was published by Hohenberg and Kohn in *Physical Review*.⁷ They proposed two fundamental theorems regarding the electron density of the system with simple proofs. These two theorems not only solve the queries regarding the justification of the approximations used in the Thomas Fermi model but also provide the sound theoretical foundation needed to construct DFT rigorously in the ground state. In fact, these theorems represent the key theoretical pillars on which all modern day density functional theories are constructed. The two theorems can be stated as: (1) every observable of a stationary quantum mechanical system, including energy, can be calculated, in principle exactly, from the ground state density alone, and (2) the ground state density can be calculated, in principle exactly, using the

variational method involving only the density.⁴⁻⁵ It is to be noted that the original HK theorems were intended for the time independent stationary ground states, but were later also been extended to excited states and time dependent systems as well.⁸⁻⁹ The details of the two HK theorems and their proof have been provided below.

Theorem 1 *‘The external potential $\hat{V}_{ext}(\vec{r})$ is (to within a constant) a unique functional of $\rho(\vec{r})$; since, in turn $\hat{V}_{ext}(\vec{r})$ fixes \hat{H} we see that the full many particle ground state is a unique functional of $\rho(\vec{r})$.*

In order to prove validity of the first theorem, Hohenberg and Kohn contrive some legitimate assumptions based on the fundamental principles of quantum chemistry, which are as follows: under the Born-Oppenheimer approximation, the ground state of an electronic system is a direct consequence of the potential exerted by the nuclei. They called the potential exerted by nuclei as the *external potential*, \hat{V}_{ext} . This assumption can further be clarified by analyzing the expression of the electronic Hamiltonian (\hat{H}_e) of Equation (2.3), which has been derived under the Born-Oppenheimer approximation. In Equation (2.3), the kinetic energy of electrons (\hat{T}_e) and the electron-electron repulsion term (\hat{V}_{ee}) simply adjust themselves to the external potential (which is represented as \hat{V}_{ne} in Equation (2.3)) so that the net energy of the system is minimized. Thus, for a specific value of \hat{V}_{ext} , every other variable of the system, including the electron density, accommodate themselves to provide the lowest possible ground state energy of the electronic system. Thus, one can consider \hat{V}_{ext} as the only variable term in the electronic Hamiltonian - the other parameters ultimately rely on it.

Based on the aforesaid assumption, Hohenberg and Kohn asked some very fundamental questions: “could the parameter \hat{V}_{ext} be uniquely determined just from the knowledge of electron density $\rho(\vec{r})$ alone? Is it possible (at least in principle) to get the information about the position and type of the nuclei if we accurately know the density $\rho(\vec{r})$ of the ground state? Does there exist a precise path for mapping from the

density ($\rho(\vec{r})$) to the external potential (\hat{V}_{ext})?" The answers to all of the abovementioned questions were found to be positive. Essentially, the mapping of \hat{V}_{ext} from $\rho(\vec{r})$ has been observed to be accurate within the limit of a constant, which is not a big concern as it is well known that the SE with two different Hamiltonian of \hat{H}_e and $\hat{H}_e + const$ provides precisely the same eigenstates. The only thing that will vary in this case is the value of the energy, which will be shifted by the value of this constant. Thus, if this is true, the knowledge of only the density is enough to get complete information about the system. As $\rho(\vec{r})$ yields the total number of electrons, $n = \int \rho(\vec{r}) d\vec{r}_1$, as well determining the \hat{V}_{ext} , it is logical to conclude that the knowledge of $\rho(\vec{r})$ is an adequate substitute to the knowledge of Ψ , the wave function of the system.

The proof of Theorem 1 that has been provided by Hohenberg and Kohn is disarmingly simple and is based on the *reductio ad absurdum* principle. The proof runs as follows:

Let us consider that $\rho(\vec{r})$ represents the exact ground state density of a non-degenerate system* of n electrons and that Ψ is the wave function of the ground state. Let us further assume that for this specific density $\rho(\vec{r})$, there exist two possible external potentials \hat{V}_{ext} and \hat{V}'_{ext} which differ with each other by more than a constant, since the wavefunction, and hence the electron density, remains unaffected if only a constant is added to the potential. Then these two different external potentials will certainly correspond to two separate electronic Hamiltonian operators (\hat{H} and \hat{H}') and the two Hamiltonians \hat{H} and \hat{H}' belong to two different ground state wave functions, Ψ and Ψ' respectively. The ground state energies corresponding to the two wave functions are E_0 and E'_0 respectively, where $E_0 = \langle \Psi | \hat{H} | \Psi \rangle$, $E'_0 = \langle \Psi' | \hat{H}' | \Psi' \rangle$ and $E_0 \neq E'_0$.

$$\hat{H} = \hat{T} + \hat{V}_{ee} + \hat{V}_{ext} \tag{2.29}$$

*It is worth noting that later investigations have proven that HK theorems can easily be extended for degenerate ground states as well.¹⁰

$$\hat{H}' = \hat{T} + \hat{V}_{ee} + \hat{V}'_{ext} \quad (2.30)$$

Since both wave functions give rise to the same electron density (this is possible considering how the charge density that is constructed from a wave function by quadrature, using Equation 2.20, is not unique), the overall scenario can be represented schematically as

$$V_{ext} \Rightarrow \hat{H} \Rightarrow \Psi \Rightarrow \rho(\vec{r}) \Leftarrow \Psi' \Leftarrow \hat{H}' \Leftarrow V'_{ext}$$

If we apply the variational principle on the expectation value of energy for the Ψ' with the Hamiltonian \hat{H} ,

$$E_0 < \langle \Psi' | \hat{H} | \Psi' \rangle = \overbrace{\langle \Psi' | \hat{H}' | \Psi' \rangle}^{E'_0} + \langle \Psi' | \hat{H} - \hat{H}' | \Psi' \rangle \quad (2.31)$$

Putting the values of \hat{H} and \hat{H}' from Equations (2.29) and (2.30) into the Equation (2.31),

$$E_0 < E'_0 + \langle \Psi' | \hat{T} + \hat{V}_{ee} + \hat{V}_{ext} - \hat{T} - \hat{V}_{ee} - \hat{V}'_{ext} | \Psi' \rangle = E'_0 + \int \rho(r) \{ \hat{V}_{ext} - \hat{V}'_{ext} \} dr \quad (2.32)$$

Similarly, applying the variational principle on the expectation value of energy for the Ψ with the Hamiltonian \hat{H}' yields,

$$E'_0 < \langle \Psi | \hat{H}' | \Psi \rangle = \overbrace{\langle \Psi | \hat{H} | \Psi \rangle}^{E_0} + \langle \Psi | \hat{H}' - \hat{H} | \Psi \rangle = E_0 - \int \rho(r) \{ \hat{V}_{ext} - \hat{V}'_{ext} \} dr \quad (2.33)$$

Adding Equations (2.32) and (2.33) leads us to a contradictory result,

$$E_0 + E'_0 < E'_0 + E_0 \quad \text{or} \quad 0 < 0 \quad (2.34)$$

It is therefore proved that there cannot be two different \hat{V}'_{ext} that give rise to the same ground state density. Thus, the ground state density $\rho(\vec{r})$ uniquely determines the external potential \hat{V}'_{ext} . Using again the concept of the variational principle, we can further state that all the properties of the ground state, including the kinetic energy of

electrons ($T[\rho]$) and the energy of electron interactions ($E_{ee}[\rho]$) can be uniquely determined by the ground state density. Thus, the total ground state energy of the system can be represented as simply a functional of density alone,

$$E[\rho] = \underbrace{E_{ne}[\rho]}_{\text{system dependent}} + \underbrace{T[\rho] + E_{ee}[\rho]}_{\text{universally valid}} = \int \rho(\vec{r}) \hat{V}_{ne}(\vec{r}) d\vec{r} + T[\rho] + E_{ee}[\rho] \quad (2.35)$$

Here, $E_{ne}[\rho] = \int \rho(\vec{r}) \hat{V}_{ne}(\vec{r}) d\vec{r}$ (the potential energy due to nuclei-electron attraction) is the system dependent term whereas $T[\rho] + E_{ee}[\rho]$ depends only on the electron density and is universal, *i.e.*, its mathematical form does not depend on the type of system (*i.e.*, on the n , R and Z of the system). Here, we have retained the subscript 'ne' to highlight the fact that the type of external potential in this case is the electron and nuclear attraction. Now, we can further group together the functionals that are just responses (the system independent part) and can be considered secondary as compared to the $E_{ne}[\rho]$,

$$F_{HK}[\rho] = T[\rho] + E_{ee}[\rho] \quad (2.36)$$

The system independent part $F_{HK}[\rho]$ is called the *Hohenberg-Kohn functional*. Now the total energy functional can be written as

$$E[\rho] = \int \rho(\vec{r}) \hat{V}_{ne}(\vec{r}) d\vec{r} + F_{HK}[\rho] \quad (2.37)$$

This, simple-looking (at first glance) $F_{HK}[\rho]$ functional is the most challenging term used in DFT. If the explicit form of both the terms that are involved in it were known, we would have solved the SE exactly. And, since it is completely system independent, it applies equally well to all kind of systems, from the hydrogen atom to gigantic molecules. However, the exact forms of both the terms are still unknown. Digging further, we can see that the second term of the $F_{HK}[\rho]$, the electron-electron interaction term $E_{ee}[\rho]$ contains exchange, Coulomb and the self interaction correction terms. Thus, it can be separated into two parts: the classical Coulomb part ($J[\rho]$, the explicit expression for which is known) and the non-classical contribution (

$E_{nc}[\rho]$, containing the Coulomb correlation, exchange and self interaction correction terms). Thus,

$$E_{ee}[\rho] = J(\rho) + E_{nc}(\rho) = \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 + E_{nc}(\rho) \quad (2.38)$$

The total energy functional now will be

$$E[\rho] = \int \rho(\vec{r}) \hat{V}_{ne}(\vec{r}) d\vec{r} + \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 + E_{nc}(\rho) + T[\rho] \quad (2.39)$$

There are two terms here whose mathematical forms are unknown: $E_{nc}(\rho)$ and $T[\rho]$.

In the next section, we will show how the major part of the $T[\rho]$ can be obtained by the Kohn-Sham (KS) approach.

We have seen until now that the proof of the first HK theorem showcases the ground state density alone is in principle sufficient enough to estimate all properties of interest. However, it does not guide us as to how to be sure that the density of our choice is the correct ground density for the system. A formal prescription for this has been provided in the second HK theorem in terms of the variational principle introduced earlier in Section 2.2.4 of this chapter.

Theorem 2 $F_{HK}[\rho]$, the functional that delivers the ground state energy of the system delivers the lowest energy if and only if the input density is the true ground state density, ρ .

This theorem can be expressed as,

$$E_0 \leq E[\tilde{\rho}] = T[\tilde{\rho}] + E_{ne}[\tilde{\rho}] + E_{ee}[\tilde{\rho}] \quad (2.40)$$

Thus, for any chosen trial density $\tilde{\rho}(\vec{r})$ - which satisfies the necessary boundary conditions like, $\tilde{\rho}(\vec{r}) \geq 0$ and $\int \rho(\vec{r}) d\vec{r} = n$, and which corresponds to some external potential \hat{V}_{ext} - the obtained energy $E[\tilde{\rho}]$ (from functional given in Equation (2.35))

will be an upper bound to the exact ground state energy $E_0[\rho]$. $E[\tilde{\rho}]$ will only be equal to E_0 if and only if the true ground state density is inserted into Equation (2.35).

The proof of the inequality (2.40) is simple. Since it is known that any trial density $\tilde{\rho}(\vec{r})$ defines its own Hamiltonian \tilde{H} and hence its own wave function $\tilde{\Psi}$, this wave function can be taken as the trial wave function for the Hamiltonian originated from the true external potential \hat{V}_{ext} . Therefore,

$$\langle \tilde{\Psi} | \hat{H} | \tilde{\Psi} \rangle = T[\tilde{\rho}] + E_{ee}[\tilde{\rho}] + \int \tilde{\rho}(\vec{r}) V_{ext} d\vec{r} = E[\tilde{\rho}] \geq E_0[\rho] = \langle \Psi | \hat{H} | \Psi \rangle \quad (2.41)$$

which is what was desired.

Before proceeding further, let us scratch out a few more formal theoretical problems that should be essentially addressed in the current context. As we have mentioned in the previous paragraph, in order for the second HK theorem to be valid, the trial density $\tilde{\rho}(\vec{r})$ must satisfy some predetermined requirements. These conditions are abbreviated as the representability of density. The first one is called the N-representability problem, which states that the $\tilde{\rho}(\vec{r})$ must sum up to the total number of electrons ‘n’.^{4-5,11-13} This condition is easy to achieve and automatically ensured when $\tilde{\rho}(\vec{r})$ stems from an antisymmetric wave function. Since nearly all practical applications in some way are technically related to the wave function, all densities that occur in these applications virtually satisfy this condition. The second condition is known as the \hat{V}_{ext} -representability (or simply as ν -representability) problem, which states that the trial density must be associated with some external potential \hat{V}_{ext} .^{4-5,14} This condition is not as trivial as that of the N-representability one, because many among the available trial densities are not suitable in accordance with the HK theorem. Only those densities are eligible that are mapped with an antisymmetrized wave function and a Hamilton operator with some type of \hat{V}_{ext} . The question that arises here is how such densities can be recognized. This is yet an open problem in DFT since so far it is not known which conditions a trial density must satisfy in order to be ν -representable. By taking some reasonable trial densities, Levy (1982) and Lieb

(1983) have demonstrated that they cannot be mapped to any \hat{V}_{ext} .¹⁵⁻¹⁶ Thus, if one chooses any of those densities it would be unfeasible to converge to any physically relevant ground state *via* variational optimization. Thus, it must be noted that the second HK theorem will only be valid if we restrict ourselves to only N- as well as v -representable trial densities. However, while this is an important problem in theoretical DFT, it has only minor relevance from an application point of view. Moreover, these rules can be reduced to a much weaker requirement that the density must originate from an antisymmetrized wave function without having explicit connection to an external potential (the N-representable condition).

2.3.5 The Kohn-Sham (KS) approach

As discussed in the previous section, the Hohenberg-Kohn theorems allow us to construct a rigorous many body theory using the electron density as the fundamental quantity. Using this framework the ground state energy of the atomic or molecular system can be written as⁷

$$E_0[\rho] = \min_{\rho \rightarrow n} \left(\int \rho(\vec{r}) \hat{V}_{ne}(\vec{r}) d\vec{r} + F_{HK}[\rho] \right) \quad (2.42)$$

where $F_{HK}[\rho]$ is the universal functional consisting of the kinetic energy, the classical Coulomb and the non-classical contribution terms as shown below,

$$F_{HK}[\rho] = T[\rho] + J(\rho) + E_{nc}(\rho) \quad (2.43)$$

Among these, the exact expression of only $J(\rho)$ is known to us. The true explicit form of the other two terms remains a mystery. Unfortunately, the expression of the kinetic energy ($T[\rho]$), which is the major contributor to total energy, is also not known with adequate accuracy. Even the final expression with the extended Thomas-Fermi model is underdeveloped and does not work for molecular systems. Then, the question that is of greatest significance here is how to make the Hohenberg-Kohn theorems practically applicable for real molecular systems. An alternative to bypass this limitation was proposed by Kohn and Sham in their second major paper of modern DFT, which appeared in 1965.¹⁷ Their central idea was to focus mostly on how exactly the kinetic energy can be determined. In order to ease the problem and

realizing that the kinetic energy of a system can easily be calculated from a known wave function (as in the case of the Hartree Fock method), Kohn and Sham introduced the concept of the non-interacting reference system (whose electron density is the same as the real interacting system of interest) built from a set of orbitals (one electron functions) in the Hohenberg-Kohn formalism such that the major portion of the kinetic energy can be obtained exactly. A non-interacting system is a fictitious system where the electrons behave simply as uncharged fermions and therefore do not interact with each other *via* Coulomb repulsion. The remainder of the kinetic energy is merged with the non-classical contributions, which are also unknown, but are usually fairly small. By this approach, as much information as possible is computed exactly, leaving only a small portion of the total energy to be determined by an approximate functional. Thus,

$$T_S = -\frac{1}{2} \sum_i^n \langle \psi_i | \nabla^2 | \psi_i \rangle \quad (2.44)$$

and

$$\rho_s(\vec{r}) = \sum_i^n \sum_s |\psi_i(\vec{r}, s)|^2 = \rho(\vec{r}) \quad (2.45)$$

Where ψ_i and T_S are the wave function and the kinetic energy respectively of the reference system. It is evident that $T_S \neq T$ even if both the interacting and non-interacting systems have the same electron density. However, it is expected that a major part of $T[\rho]$ is recovered *via* T_S . In order to correct this error, Kohn and Sham suggested the following partitioning of the universal functional $F_{HK}[\rho]$,

$$F_{HK}[\rho] = T_S[\rho] + J[\rho] + E_{XC}[\rho] \quad (2.46)$$

Where $E_{XC}[\rho]$ is defined as the *exchange-correlation energy* and is given by,

$$E_{XC}[\rho] = (T[\rho] - T_S[\rho]) + (E_{ee}[\rho] - J[\rho]) \quad (2.47)$$

This functional $E_{XC}[\rho]$ contains all the components of energy that are unknown and not accounted for: the contribution of electron exchange, the contribution of electron

correlation (which is the crucial part of the energy for systems containing interacting electrons), the residual portion of the kinetic energy (which is not included in the term T_s), the correction for the self-interaction (which originated from the classical Coulomb potential). In summary, all the contributors of the total energy whose explicit forms are not known and are difficult to obtain *via* theoretical techniques are put together within this functional. Today, many superior and accurate approximations are available for this functional.

The question that is crucial to be asked at this venture is how we would find the potential ($V_s(\vec{r})$) for the non-interacting reference system so that it leads to an antisymmetrized wave function (the Slater determinant) that is associated with the exact same density as our real interacting system. To answer this, we will rewrite the energy of our system of interest in terms of the newly separated $F_{HK}[\rho]$,

$$E[\rho] = T_s[\rho] + J[\rho] + E_{EX}[\rho] + E_{ne}[\rho] \tag{2.48}$$

The further expansion of these terms will lead to

$$\begin{aligned} E[\rho] &= T_s[\rho] + \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 + E_{EX}[\rho] + \int V_{ne}\rho(\vec{r})d\vec{r} \tag{2.49} \\ &= -\frac{1}{2} \sum_i^n \langle \psi_i | \nabla^2 | \psi_i \rangle + \frac{1}{2} \sum_i^n \sum_j^n \iint |\psi_i(\vec{r}_1)|^2 \frac{1}{r_{12}} |\psi_j(\vec{r}_2)|^2 d\vec{r}_1 d\vec{r}_2 + E_{EX}[\rho] - \sum_i^n \int \sum_a^m \frac{Z_a}{r_{1a}} |\psi_i(\vec{r}_1)|^2 d\vec{r}_1 \\ &\dots\dots\dots \tag{2.50} \end{aligned}$$

The only unknown variable in Equation (2.50) is $E_{EX}[\rho]$. Now, if we apply the variational principle on Equation (2.50) to minimize the energy under the constraint:

$\langle \psi_i | \psi_j \rangle = \delta_{ij}$, we will obtain the *Kohn-Sham equation*,

$$\begin{aligned} &\left(-\frac{1}{2} \nabla^2 + \left[\int \frac{\rho(\vec{r}_2)}{r_{12}} d\vec{r}_2 + V_{XC}(\vec{r}_1) - \sum_a^m \frac{Z_a}{r_{1a}} \right] \right) \psi_i \\ &= \left(-\frac{1}{2} \nabla^2 + V_{eff}(\vec{r}_1) \right) \psi_i = \epsilon_i \psi_i \tag{2.51} \end{aligned}$$

The resulting equation is very similar to the eigenvalue-equation of the Hartree-Fock (HF) method. However, in the HF method, the Fock operator contains the non-local potential that differs for every electron whereas in the Kohn-Sham operator the V_{eff} potential depends only on \vec{r} (and is called as local potential) and not on the index of the electrons, and hence it is equal for all electrons. On comparing this equation with the one-particle equations from the non-interacting reference system we will get the expression for V_s , which is nothing but V_{eff} .

$$V_s(\vec{r}) = V_{eff}(\vec{r}) = \int \frac{\rho(\vec{r}_2)}{r_{12}} d\vec{r}_2 + V_{xc}(\vec{r}_1) - \sum_a^m \frac{Z_a}{r_{1a}} \quad (2.52)$$

ψ_i in Equation (2.51) is called the *Kohn-Sham orbital* (or briefly *KS orbital*), which can easily be derived from Equation (2.51) and can be used to compute the density (using Equation (2.45)) which can further be utilized for generating a new improved V_{eff} , which will eventually lead to a new self-consistent cycle as shown in the flow chart in Figure 2.1. That is, once we know the various contributions in Equation (2.52), we can get the potential V_{eff} which we need to insert into the one-particle equations, which in turn determines the orbitals and consequently the ground state density and the ground state energy by employing the energy expression (2.50). As V_{eff} already depends on the density (and correspondingly on the orbitals) through the classical Coulomb term as shown in Equation (2.50), the Kohn-Sham one-electron Equation (2.51) also have to be solved iteratively until self-consistency is achieved, just like the Hartree-Fock equations.

V_{xc} in Equation (2.51) is simply defined as the functional derivative of E_{EX} with respect to ρ as its explicit form is also unknown, similar to the E_{xc} ,

$$V_{xc} = \frac{\partial E_{xc}}{\partial \rho} \quad (2.53)$$

It is necessary to realize here that the ψ_i s are not equivalent to the real orbitals of the system and hence they do not connect to any real physically meaningful system. Their exclusive purpose is to provide a theoretical mapping between the kinetic energy and

the density. It is also useful to note that the KS wave function is a single determinant approach and it fails where multiple determinants are needed in order to describe the system; for example, the dissociation of a molecule. A general scheme for a KS iteration cycle (during a single point or a geometry optimization process) is illustrated in Figure 2.1 below. Notably, if the exact forms of E_{EX} and V_{XC} were known, the Kohn-Sham strategy would give the exact value of the energy, *i.e.*, the correct eigenvalue of the Hamilton operator. The formalism that has been illustrated so far in this section does not contain any approximation. Thus, unlike the Hartree-Fock model, where the approximation is introduced right from the beginning, the *Kohn-Sham approach is in principle exact!* The approximation enters only when we need to decide on an explicit form of the unknown functional (by any technique) for the exchange-correlation energy E_{EX} and the corresponding potential V_{XC} . The major goal in development of modern DFT is therefore to find better and better approximations to these two quantities. This has led to a large number of functionals that are available to us for the computation of the energy for a variety of systems. These functionals are found to be highly system specific and require some sort of theoretical understanding, instead of being used blindly.

In course of the past few decades, DFT has become the pivotal computational tool that has been employed by a vast range of researchers across the globe. They include both theoretical as well as experimental chemists/physicists. However, many physicists and chemists regard the journey of DFT from just a theoretical concept to the present state of glory as merely a fortuitous venture. This might have happened due to the easy accomplishment of accuracy, which is often regarded as a challenging task in *ab initio* wave function based methods. A number of theoretical chemists also consider the fundamental theories somewhat dubious and the simplistic single deterministic approach provokes as a cause for further skepticism. However, considering the accuracy in addition to the reduction in computational cost, the widespread preference for DFT seems like a rational choice. Due to the widespread use of modern DFT, performing fruitful calculations for a large number of systems of real size has become possible within a limited span of time. Supporting experimental results or predicting new chemistry by theoretical calculations has now been reduced to merely a few days job with the aid of DFT. On the other hand, the wave function

based methods are still impractical to be applied to the actual systems related to chemistry, biology and material sciences. This implies that the extensive availability of DFT software packages and the ever growing DFT-community must be viewed as a natural progress rather than a miracle.

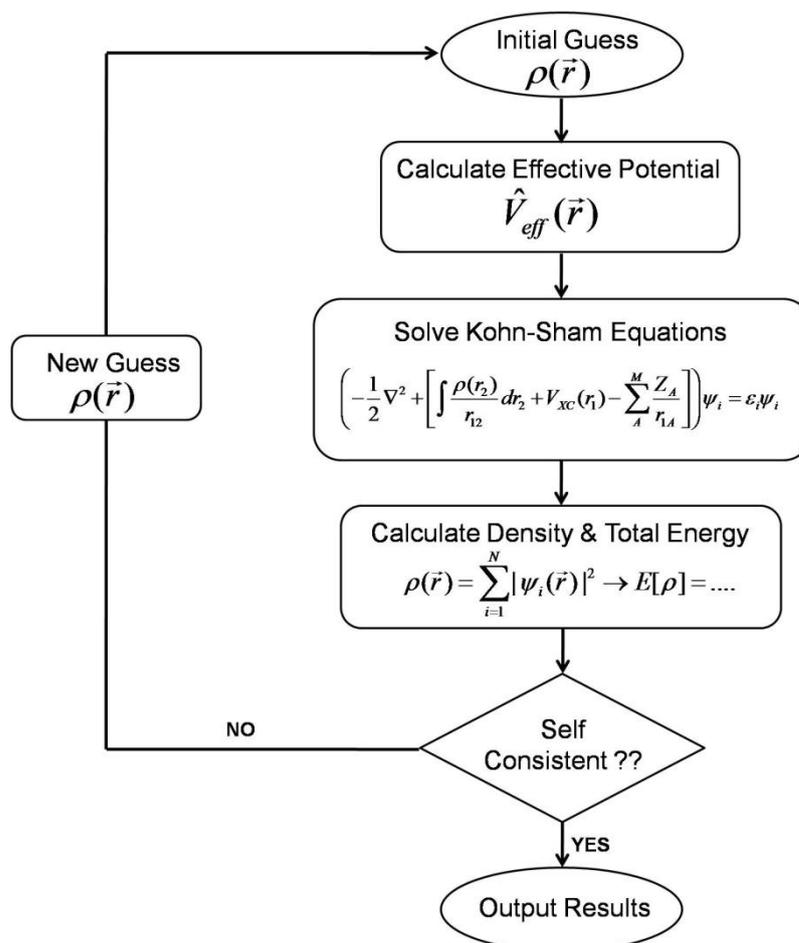
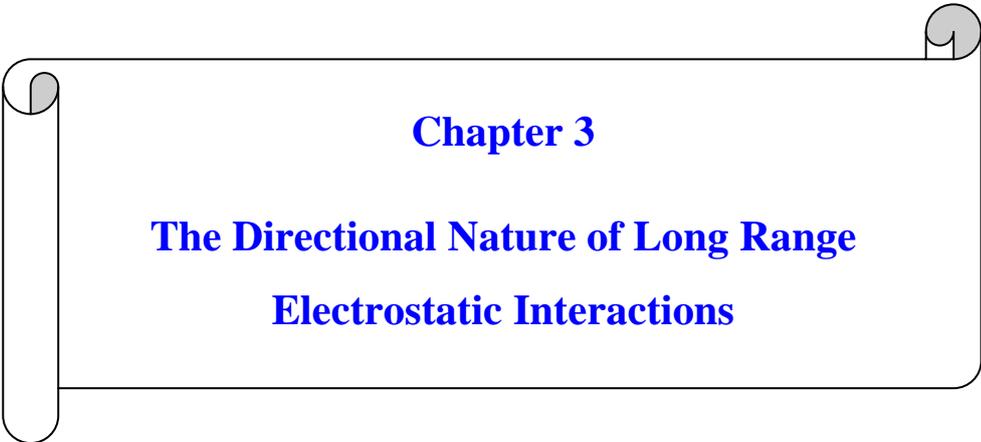


Figure 2.1 The flow chart for Kohn Sham iterations for the single point or the single step during optimization.

2.4 References

1. H. G. Kümmel, *Int. J. Mod. Phys. B*, 2003, **17**, 5311; R. F. In Bishop, T. Brandes, K. A. Gernoth, N. R. Walet, Y. Xian, H. Kümmel, in *Proceedings of the 11th International Conference on Recent Progress in many-body theories, Singapore: World Scientific Publishing*, 2002, 334–348; I. Shavitt, R. J. Bartlett, *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory*, Cambridge University Press. 2009.

2. C. J. Cramer, *Chichester: John Wiley & Sons, Ltd.* 2002, 191–232; C. D. Sherrill, H. F. Schaefer III, *The Configuration Interaction Method: Advances in Highly Correlated Approaches*, *San Diego: Academic Press*, 1999, 143–269.
3. M. H. N. Assadi, D. A. H. Hanao *J. Appl. Phys.*, 2013, **113**, 233913; M. D. Segall, P. J. D. Lindan, M. J. Probert, C. J. Pickard, P. J. Hasnip, S. J. Clark, M. C. Payne, *J. Phys: Condens. Matter*, 2002, **14**, 2717-2743; F. R. Somayeh, H. Soleymanabadi, *J. Mol. Model.*, 2014, **20**, 2439-2445; F. R. Somayeh, H. Soleymanabadi, *J Mol Model* , 2013, **19**, 3733-3740; D. Music, R. W. Geyer, J. M. Schneider, *Surface & Coatings Technology*, 2016, **286**, 178.
4. R. G. Parr, W. Yang, *Density functional theory of atoms and molecules*, *Oxford University Press, New York*: 1989, pp 1–352.
5. W. Koch, M. C. Holthausen, *A chemist's guide to density functional theory*; *John Wiley & Sons*: 2015.
6. L. H. Thomas, *Mathematical Proc. Cambridge Phil. Soc.*, 1927, **23**, 542-548; E. Fermi, *Rend. Accad. Naz. Lincei.* 1927, **6**, 602–607.
7. P. Hohenberg, W. Kohn, *Physical review* 1964, **136**, B864.
8. E. Gross, S. Kurth, *Density Functional Theory: The Modern Treatment of Electron Correlations*; Springer US: 1994, pp 367–409.
9. E. Gross, W. Kohn, *Advances in quantum chemistry* 1990, **21**, 255–291.
10. R. M. Dreizler, E. Gross, *Density Functional Theory: An Approach to the Quantum Many-Body Problem* Springer; Berlin: 1990.
11. T. Gilbert, *Physical Review B* 1975, **12**, 2111.
12. E. A. H. Lieb, A. Shimony, H. Feshbach (*MIT, Cambridge, MA*) 1982, pp. 111-149.
13. J. E. Harriman, *Physical Review A*, 1981, **24**, 680.
14. W. Kohn, *Physical review letters*, 1983, **51**, 1596.
15. M. Levy, *Physical Review A*, 1982, **26**, 1200.
16. E. H. Lieb, *International Journal of Quantum Chemistry*, 1983, **24**, 243–277.
17. W. Kohn, L. J. Sham, *Physical review*, 1965, **140**, A1133.



Chapter 3

The Directional Nature of Long Range Electrostatic Interactions

Chapter 3

The Directional Nature of Long Range Electrostatic Interactions

Abstract: It has been well established that long range multipoint electrostatic interactions have a significant effect on the stability of hydrogen bonded complexes. Interestingly, multi-point electrostatic interactions are directional in nature. However, the directional aspects of long range electrostatic interactions have been totally overlooked so far in the literature. In this chapter, we have proposed a method that estimates the impact of long range electrostatic interactions in terms of the electrostatic force. The method can take care of the directionality as well as the magnitude of the interactions. We have suggested here that the consideration of the electrostatic force can provide better information about the binding strength of systems when the bonding is predominantly electrostatic. We have further proposed that the electrostatic force may simply be calculated considering the atoms in molecules as point charges and employing Coulomb's electrostatic equation for force. The benchmarking of the proposed method has been done against the forces that have been obtained employing the finite difference method with the help of energy decomposition analysis for 16 planar hydrogen bonded complexes.

3.1 Introduction

Noncovalent interactions are of great significance in several varied and important areas of chemistry and biology. Given their significance, there has been a conscious effort in recent times to exploit such interactions in order to achieve specifically designed goals, in areas as diverse as those of asymmetric catalysis,¹⁻⁴ supramolecular chemistry,^{1,5-6} crystal engineering,⁷⁻¹⁰ polymer chemistry,^{6,11-12} peptido-mimetic chemistry,^{4,13-14} and molecular medicine.¹⁵⁻¹⁸ However, in order to fully unlock and exploit the potential of such noncovalent interactions, it is necessary to properly understand the factors that determine their strength. Some of these interactions are *de facto* dominated by

electrostatics (*e.g.* XH-Y H-bonds),¹⁹ whereas some are dominated by dispersion and other factors (*e.g.* CH- π ,²⁰ π - π ²¹). In this computational and theoretical study, our focus has been on the development of understanding of systems where electrostatic noncovalent interactions are the principal influencing factor.

In such systems, the fundamental question is that of understanding the extent of long-range electrostatic interactions in determining important properties of the system, such as the binding behaviour of two partners into a single complex. A typical family of complexes that has generated great interest in this regard is that of the planar hydrogen bonded complexes. Jorgensen *et al.* in their seminal theoretical study on triply hydrogen bonded nitrogenous bases explained that it is inadequate to consider only primary electrostatic interactions in determining the association constant for a system with two partners held together by hydrogen bonding.²² They concluded that the electrostatic interaction between the immediate non-hydrogen bonded donors and acceptor, which they defined as secondary electrostatic interactions (SEI), also contribute significantly to the binding (Figure 3.1a). This has been supported by a plethora of experimental studies.²³⁻³⁰ However, the results of some subsequent studies^{23-24,31} have brought this hypothesis into question. Popelier *et al.* in their comprehensive QTAIM (Quantum Theory of Atoms-in-Molecules) study on 28 base pairs complexes have shown that the electrostatic energy of interaction between many remote atom pairs across a hydrogen bond is also influential to the binding, and hence the consideration of the electrostatic interaction of all the atoms of one partner with all the atoms of the other may be necessary in order to get the proper picture of the long range electrostatic influence on the binding.³²

The question that has sparked the current investigation is this: since long range SEIs have been demonstrated to be significant, should not the electrostatic *force*, rather than the electrostatic energy of interaction, be the more important property that needs to be evaluated, in order to get proper understanding and insight into such systems? For example, in structures such as those of proteins and DNA, hydrogen bonds are surrounded by atoms from every side in a three dimensional framework. In such a situation, the

electrostatic energy cannot define the strength of the interaction unambiguously. Electrostatic force, which has directionality, thus becomes a more significant factor. This point is illustrated by a simple model shown in Figure 3.1b. If all of the individual distances between the charges on the smaller subunit and the charges on the larger subunit in Figure 3.1b are equal in the two complexes, the intermolecular electrostatic potential of both the systems would be the same. However, the electrostatic force of interaction will vary: the structure on the right will be more tightly held. One can also see from Figure 3.1b that the line of approach of the two species is of significance. Therefore, it is more important to consider the electrostatic force of interaction rather than the electrostatic energy.

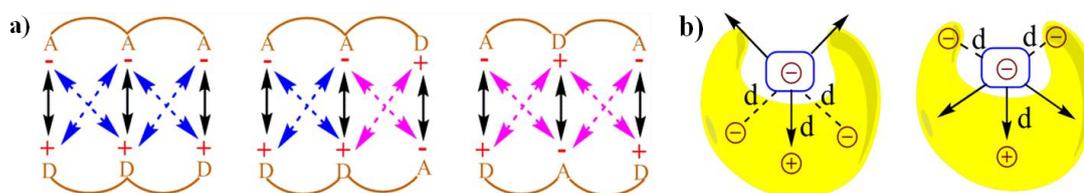


Figure 3.1 a) Electrostatic interactions between H-bond acceptor and donors in triply H-bonded complexes as proposed by Jorgensen *et al.*; A and D symbolize hydrogen bond acceptor and donor atoms respectively; black, blue and pink arrows represent attractive primary, attractive secondary and repulsive secondary interactions respectively; the association constant decreases with the increasing number of repulsive secondary interactions. b) The representation of electrostatic forces between two molecular segments bonded by noncovalent interactions and having different charge distributions in the three dimensional space; d is the distance between the two charges. This model considers both primary and secondary interactions.

Several methods such as the EDA (energy decomposition analysis), NCE (natural Coulomb electrostatics) and QTAIM have been developed and are being practiced regularly for segregating and quantifying the electrostatic contribution in the interaction between two partners or fragments in a system. However, all of these methods rely on the computation of the electrostatic energy rather than the force. Credit for the pioneering contributions to the electrostatic force analysis, however, should be given to Berlin, who partitioned

diatomic molecules (for example, a covalently linked H₂ molecule) into binding and nonbinding regions on the basis of a binding force function $f(r)$ obtained by the actual computed electron density under the Born-Oppenheimer approximation.³³ Many attempts were made to extend Berlin's approach to analyze covalent bonds in polyatomic molecules by Bader *et al.*,³⁴ Johnsen,³⁵ Koga *et al.*,³⁶ and other authors.³⁷ However, this approach had also been criticized by Epstein,³⁸ Koga *et al.*³⁶ and Silberbach³⁹ for different reasons. Overall, the concepts of a binding and an antibinding region become unrealistic when applied to real life chemistry, specifically to polyatomic molecules. The noted alternative method to obtain directionality in noncovalent interactions is provided by the Buckingham-Fowler model, according to which directionality in noncovalent bonds can be achieved as a function of the relative orientation of interacting partners by putting them in van der Waals contact of each other and then by allowing one of them to roll over the other in search of the minimum electrostatic energy (till the global minimum is achieved).⁴⁰ However, their method does not describe the consideration of the electrostatic force of binding to obtain the directionality in noncovalent interactions.

In the current work, we propose the determination and understanding of the electrostatic forces (EFs) as a viable alternative to account for the strength and nature of electrostatic interactions. We have further shown that the forces may be calculated simply by employing Coulomb's law, with the atoms being considered as point charges. The charges on the atoms have been determined from quantum chemical calculations, and the net electrostatic force of interaction between two partners has been determined in a particular direction assigned after careful analysis of the molecular structures (see Figure 3.6, Results and Discussion section). To showcase the reliability of the method, we have benchmarked it against the forces that have been obtained employing the finite difference method with the help of the energy decomposition analysis for the sixteen near-planar hydrogen bonded complexes where the electrostatic interaction has been known to be the significant contributing factor. These complexes have been studied extensively by Leigh and co-workers, as well as others, and are proposed to be the standard models for important biological systems (Figure 3.4).^{23,25-27} Furthermore, we have also discussed the scope of the work, thereby underlining its significance in several different areas of chemistry, biology and material sciences.

3.2 Computational Details and Background Theory

All the DFT calculations, until unless mentioned specifically, were carried out using the Turbomole 6.4 suite of quantum-chemical programs.⁴¹ Geometry optimizations were performed using the PBE⁴² functional in the solvent phase with the Conductor like Screening Model (COSMO)⁴³ employing chloroform (epsilon = 4.81) as the solvent. The electronic configuration of the atoms was described by a triple-zeta basis set augmented by a polarization function (TURBOMOLE basis set TZVP). The resolution of identity (RI),⁴⁴ along with the multipole accelerated resolution of identity (marij)⁴⁵ approximations were employed for an accurate and efficient treatment of the electronic Coulomb term in the density functional calculations. The option “disp” provided in Turbomole package (DFT-D2, a general, empirical dispersion correction proposed by Stefan Grimme for density functional calculations) was used for dispersion corrected DFT calculations for all the calculations with Turbomole.⁴⁶

The Energy Decomposition analysis (EDA) was carried out using Turbomole 7.0 at the same level of theory and under the same conditions of solvent and dispersion that have been used for geometry optimizations using the Turbomole version 6.4 (the EDA implementation is not available with Turbomole 6.4).

The Mulliken⁴⁷, NBO⁴⁸ and Löwdin⁴⁹ charges have been used to calculate electrostatic forces on each fragment of every complex along a particular direction that described the intermolecular interactions most suitably (*i.e.* along the line of direction, see the figures 7a and 7b). To compute the magnitude of the net force of binding, *i.e.*, the binding force, the vector sum of electrostatic forces experienced by each fragment along the above said direction has been considered.

It is to be noted here that interactions involving the hydrogen bond acceptors and donors that are directly involved in hydrogen bonding with each other are referred to as primary interactions. The electrostatic interactions between any pair of atoms between two partners that are not directly involved in hydrogen bonding have been defined as long range secondary electrostatic

interactions. These interactions also include the Jorgensen's type of secondary electrostatic interactions (SEIs).

A recent review by Clark, Politzer, Murray and others states that even the polarization/induction and covalent (or donor-acceptor charge transfer) factors in the noncovalent bonds are essentially electrostatic in nature.¹⁹ According to the Feynman interpretation, even the dispersion interaction is electrostatic in nature.⁵⁰ This indicates that the proper treatment of electrostatic interactions is adequate enough to fully describe all kinds of noncovalent interactions. The Born-Oppenheimer approximation reveals molecules to be a collection of point charge nuclei and a cloud of indistinguishable electrons described by electron density. To distinguish a molecule in the form of atoms and bonds, in a point charge analysis, the electron density is divided on the criteria of basis sets, *i.e.*, the atomic orbitals. When a noncovalent bonding partner is kept in the electric field of another partner, the induction causes a change in the local electron density of the first, and *vice versa*. This is reflected in the point charge calculation as the modification in the charges of the atoms when the charges are calculated for partners in the complex in comparison to the charges when calculated with the two partners infinitely separated. The point charges that are calculated for the atoms in a complex will, therefore, also include the polarization and donor-acceptor charge transfer in an approximate way. This kind of analyses has been done by other groups as well to account for the effect of polarization and donor-acceptor contributions.⁵¹ Recently, noncovalent interactions between covalently bonded atoms of the Groups IV-VII and a negative site (e.g. a Lewis base) have been discovered, and are commonly referred to as σ -hole bonding, because of the presence of a positive cap on the electrostatic potential surface on the opposite side of one of the covalent bonds of the atom, labeled as a σ -hole.⁵²⁻⁵⁴ One of the most common known cases of σ -hole bonding is halogen bonding. The σ -hole arises due to anisotropy of the atomic charge distribution, and can be visualized through an anisotropic electrostatic potential around the atom. This results in unusual behavior of atoms that have σ -holes. Such atoms can have regions of both positive and negative electrostatic potential on their surfaces, and they can thus interact attractively with both negative and positive sites respectively, in

different directions.^{52,54} Assigning a single atomic charge to such atoms in molecular complexes will fail to describe σ -hole bonding. More recently, methods have been developed in molecular dynamics (MD) simulations to address such behavior in a more accurate way. Force fields have been developed where the positive region of halogen atoms are represented as an extra point of charge in a way so that the net formal charge and the electrostatic potential assigned to the atom remain the same.^{55,56} The results of this method have also been corroborated by high level quantum chemical calculations.⁵⁵ Therefore, assigning one extra-point charge to every positive hole in a molecule/complex was found to give a good reasonable approximation for modulating other electrostatic properties. However, σ -hole bonding is not found to be universally manifested by all the elements of Groups IV-VII in every chemical composition. In general, σ -hole bonding is not exhibited by fluorines and is insignificant in other elements of the same period, particularly when they are not covalently linked to more electronegative atoms.^{55,57} This can be justified by looking at the molecular electrostatic potential surface map of the corresponding compounds. Molecules that do not possess σ -holes may not require additional treatment (of assigning extra point charges for representing σ -holes) to accurately address the electrostatic properties. To further confirm whether the individual noncovalent hydrogen bonded partners, considered in this study, possess σ -holes on their electrostatic potential surfaces, we have constructed the molecular surface electrostatic potential map of a set of representative noncovalent partners containing nitrogen and fluorine atoms at their interactive sites. The obtained electrostatic potential surfaces reveal that none of the moieties that have been considered in this study possess σ -holes on their surface (Figure 3.2 below). It is, therefore, correct to conclude that the point charge calculations employed in the current work provide a good approach to estimating the net electrostatic interactions, especially when determined from charge calculations at a high level of theory. However, additional caution should be taken for molecules containing σ -holes in them.

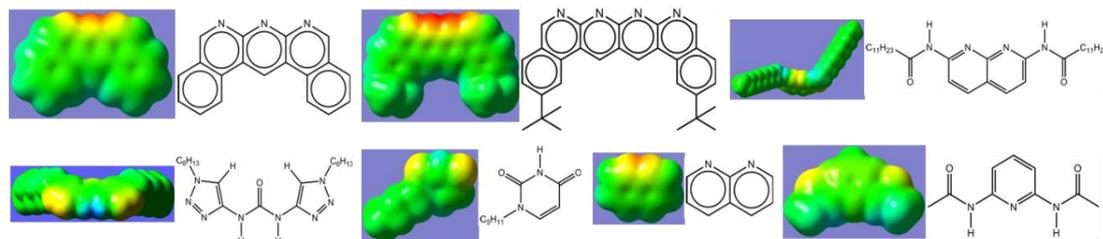


Figure 3.2 The molecular surface electrostatic potential in Hartrees, computed on the 0.0004 au contour of the electron density, using the GaussView software at the CPCM(CHCl₃)/M062X/6-31G** level of theory with the geometries optimized using G09. The blue colour indicates positive electrostatic potential and the red color indicates negative electrostatic potential, whereas the intermediate colours indicate intermediate electrostatic potential. The colour scale was kept uniform in all the cases.

3.3 Results and Discussion

In order to determine the electrostatic force of binding between two noncovalent bonded partners, the approach that has been proposed here has been to determine the Coulombic force between each pair of atoms, with the atoms being chosen from the different partners, and then taking a vector sum up all these individual contributions to obtain the net force of binding. The point charge approximation has been proposed as a practical and more general choice for this purpose, and each atom has been considered to have a charge, determined by the QM calculations, as described in the previous section. Since force has direction, the Coulombic interaction has been proposed to be considered along a certain common line to obtain the force of binding for two bonded partners: the “line of direction” for the given molecular complex (see Figure 3.6). Therefore, only the component of the individual pair-wise interactions between two partners is proposed to be considered along that given line (Figure 3.6c). The algorithm for the code that incorporates this approach has been provided in Figure 3.3 below. This method has been explained below in details considering sixteen planar hydrogen bonded complexes.

The family of the planar hydrogen bonded complexes involves two partners interacting through X-H---Y hydrogen bonded interactions, with the number of such hydrogen bonds varying from two to four. All the hydrogen

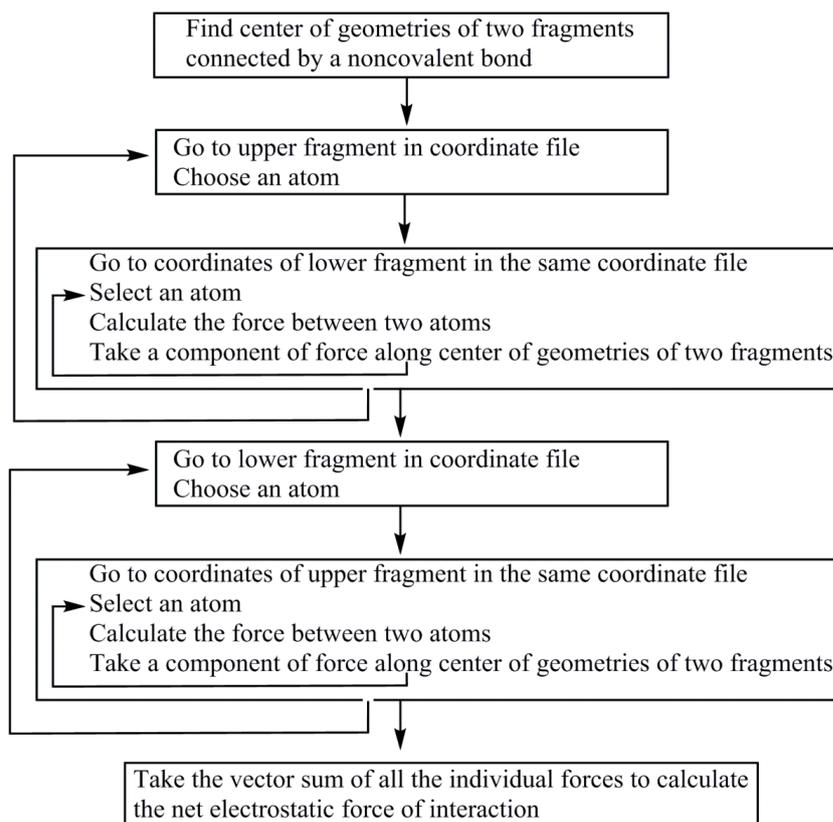


Figure 3.3 A flow chart of the Fortran 90 code used for calculating the net force between two partners in hydrogen bonded complexes.

bonds in these complexes exist in one plane. These complexes have been proposed to serve as useful models for investigating and understanding multipoint hydrogen bonding, which is of great importance to biological systems as well as to multifunctional materials and supramolecular polymers.^{23,25-27,58} This has led to publication of a large number of reports in reputed international journals in recent times that have discussed different planar hydrogen bonded structures belonging to this family. This family of planar hydrogen bonded complexes has been chosen specifically to test our approach because it has been shown that SEIs are very important for determining the stability of these complexes,²³⁻³⁰ which has led chemists to the design principle of having all the hydrogen bond acceptors on one partner and all the hydrogen bond donors on the other.²² We have therefore taken a sample set of sixteen different representative complexes to show the viability of our approach. The optimized structures of these complexes at the

COSMO(CHCl₃)/PBE/TZVP level of theory have been shown below in Figure 3.4.

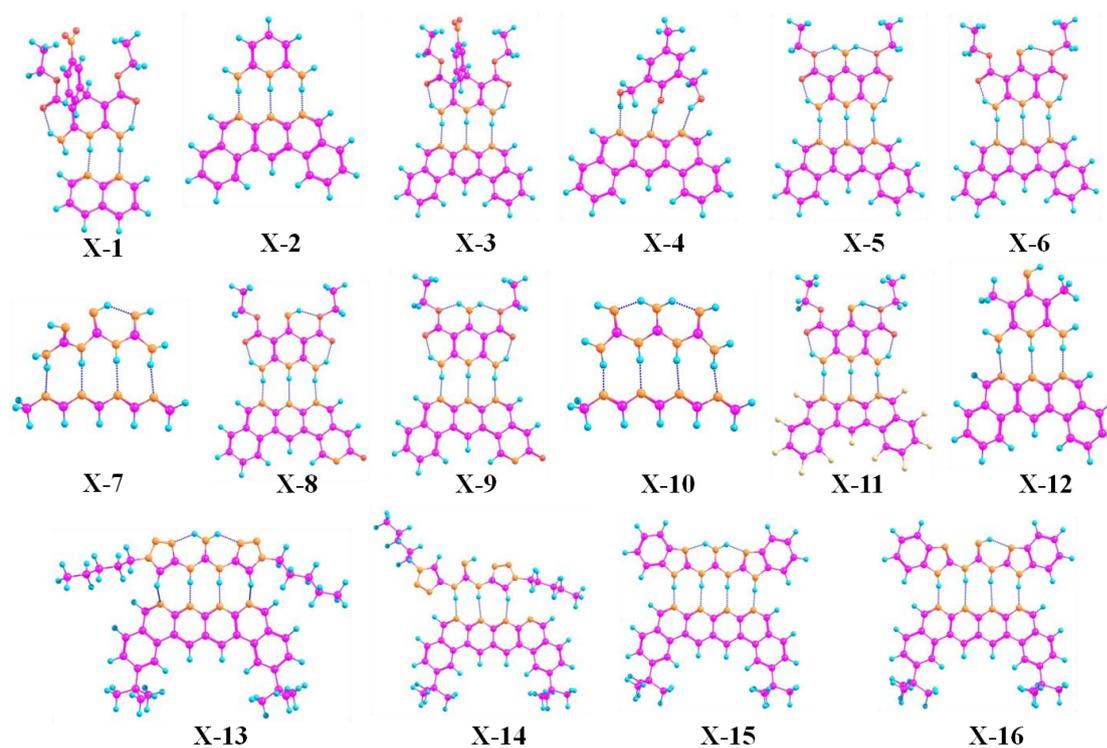


Figure 3.4 The optimized geometries of planar hydrogen bonded complexes at the COSMO(CHCl₃)/PBE/TZVP level of theory. Pink, cyan, brown and white colors represent carbon, hydrogen, nitrogen and fluorine atoms respectively, whereas, dotted blue lines represent hydrogen bonds.

It has been reported that long range electrostatic interaction energies, which vary with r as r^{-1} , are significant in determining the net binding energy of hydrogen bonded complexes.³² The magnitudes of electrostatic forces, however, vary with r as r^{-2} . The question that is of significant relevance in this context is this: do the electrostatic interaction forces in hydrogen bonded partners also have long range influence? To examine this, we have calculated all the pair-wise electrostatic forces between atoms on two partners in **X-15** (see Figure 3.4 for the structure). The pair-wise forces in this case have been computed along the line joining the center of geometries (geometrical centers) of the frontier hydrogen bonded atoms in two partners in **X-15**, as this line seems to represent binding interactions between the two partners most

appropriately. A graph with the magnitude of pair-wise electrostatic force (considering Mulliken and NBO charges) on the Y axis and the corresponding atom-atom distance on the X axis is shown in Figure 3.5 below. It can be deduced from Figure 3.5 that the electrostatic forces are significant even beyond 12 Å of distance between two atoms at two bonded partners. Thereby, it is concluded that the long range electrostatic forces are also significant in the hydrogen bonded systems.

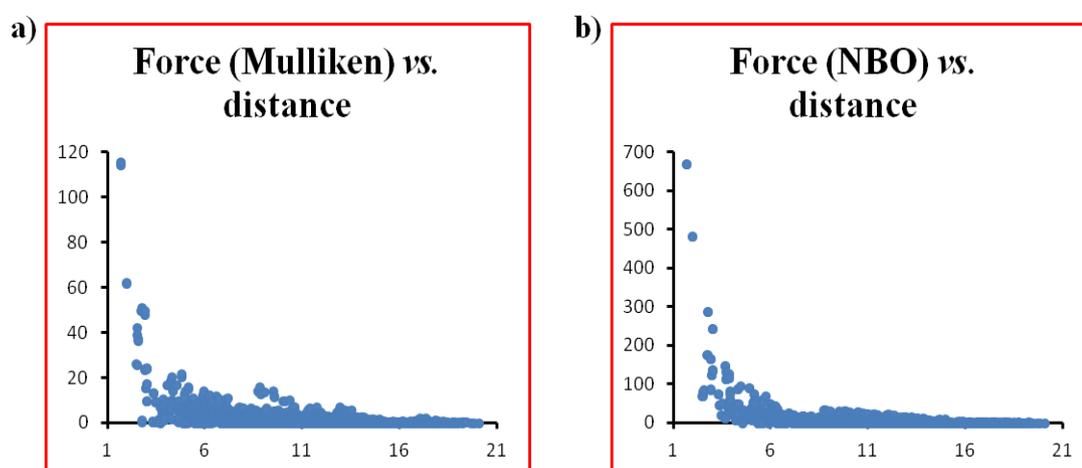


Figure 3.5 The pair-wise atomic electrostatic force *vs.* the corresponding atom-atom distance: a) with Mulliken charges, and b) with NBO charges. Only the magnitude of the pair-wise forces has been shown here.

However, a careful inspection of molecular structure is necessary in order to determine the direction to compute the net intermolecular electrostatic force of interaction to represent the net binding force. A thorough analysis reveals that the line joining the center of geometry of the frontier atoms in the planar hydrogen bonded complexes describes the interaction forces most appropriately, because this is the line along which the two partners that will hydrogen bond would approach and bind with each other, as revealed by the molecular surface electrostatic potential map in Figure 3.6a. The frontier atoms of a hydrogen bonding partner are atoms that are nearest to the complementary partner and are directly involved in the hydrogen bonding (Figure 3.6b). The center of geometry of the frontier atoms of a partner is the geometrical center of the frontier atoms. The X, Y and Z coordinates of the geometrical center of a particular partner were calculated by taking the average of the corresponding

coordinates of all the frontier atoms. It has been assumed here that the two partners will approach each other along the line connecting the center of geometries of the frontier atoms. This hypothesis has been tested by a thorough analysis where we have calculated all the pair-wise binding forces between atoms at two partners along lines (in the same plane, as hydrogen bonds in these cases are almost planar) making angles of 0° , 30° , 60° , 90° , and so on up to 360° for a sample set of 16 representative complexes of the planar hydrogen bond family using Mulliken charges (Table 3.1 below). The net forces of binding were also calculated along these lines. The magnitude of the total binding force was found to be most favorable (negative) along the originally chosen reference line of direction (i.e., at 0° and 360°) for 10 out of the total 16 structures. A further analysis using NBO charges gives 15 structures having the most favorable binding along this line, and analysis using Löwdin charges gives 13 structures with maximum binding force along this line of reference (Table 3.1). Therefore, the line joining the center of geometries of the two hydrogen bonding partners was found to be the most appropriate, common line for the line of direction calculations for the family of planar hydrogen bonded complexes (Figure 3.6b). All the forces on one partner due to the presence of the other were calculated along this “line of direction” but in the direction of mutual approach as illustrated in Figure 3.6c below by a simple two dimensional (2D) model complex. It should be noted here that Figure 3.6c is a 2D model chosen for the purpose of simplicity and clarity. The complexes considered in this study are three dimensional (3D). Suitable measures have been taken in order to calculate the component of the forces along the line of direction by employing vector algebra.

The idea behind choosing a particular line as a line of direction has been derived from the following logic: when the electrostatic interaction is the dominant factor in the binding, the two binding partners, in general, would approach each other along a direction that is most favorable electrostatically, i.e., along the direction that maximizes the favorable electrostatic force of interaction. This implies that the magnitude of the electrostatic force of interaction must be the highest (with negative sign) along this direction, which we define here as the “line of direction”. This point can be understood by

looking at the potential energy surface of the individual partners (please see Figure 3.6a). Having hypothesized that there exists a direction (or a line) along which the favorable electrostatic force of interaction between two partners in a hydrogen bonded complex is the greatest, we have tried to deduce this direction by looking at the molecular geometries of the individual partners in the complex, and by employing the force analysis. With a simple approximation that all the hydrogen bonds in the complex are similar in strength, we can reach the conclusion that the line joining the center of geometries of the individual partners would be the best line to generalize as a line of direction in planar hydrogen bonded complexes, as also suggested by our force analysis results (Table 3.1). It can also be observed from Table 3.1 that the binding forces obtained at the chosen angles are positive (unfavorable), indicating that bonding between the two partners in a given complex along these lines is not possible. Thus, the results of our study indicate that there exist two distinct binding and anti-binding regions around the hydrogen bonded systems (or any electrostatic dominated noncovalent bonded system) as well, similar to what had been observed and described by Berlin for covalent compounds using actual electron density in his seminal 1951 paper.

After ensuring a particular line as the line of direction, we have corroborated our electrostatic force analysis method with the electrostatic forces obtained by the help of the EDA analysis method, which we have termed “EF (EDA)”. EF (EDA) was calculated as a finite difference force obtained by taking the spatial derivative of the electrostatic energies (EEs), *i.e.*, it was calculated by taking the negative gradient of electrostatic energies between the two points: $-\{(EE2 - EE1)/(0.1)\}$. One of the two points considered is the optimized geometry, with the corresponding EDA obtained electrostatic energy: EE1, and the other is the geometry optimized after translating one partner by 0.1 Å away from the other partner along the line of direction, with the corresponding EDA obtained electrostatic energy: EE2. It should be noted that these newly obtained geometries for each complex case were optimized *via* a constrained geometry optimization, where the frontier atoms of each complex were frozen, in order to preserve the center of geometry, and by maintaining the additional distance of 0.1 Å between the centers of geometry in the two

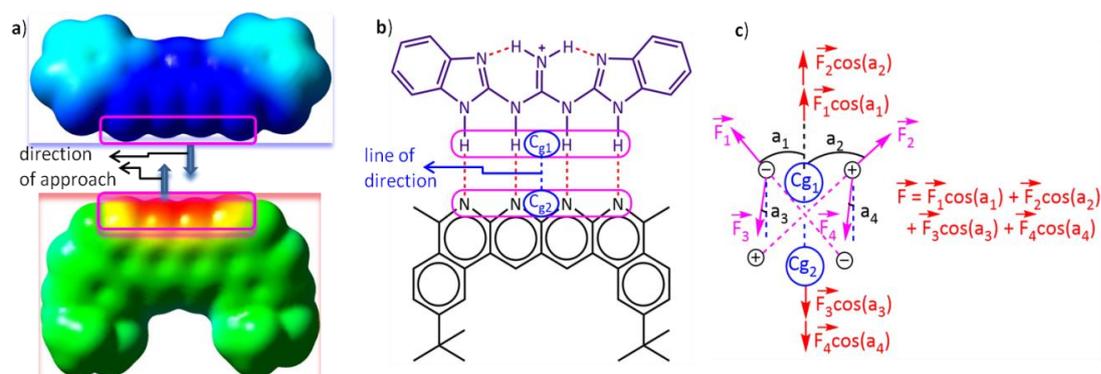


Figure 3.6 A schematic representation of a) the direction of approach of two partners represented through the molecular surface electrostatic potential, b) the center of geometry of frontier atoms on two partners describing the line of direction and c) the electrostatic force vector acting on one partner due to the charges on the other in a two dimensional planar model complex. The Molecular Surface electrostatic potential was computed in Hartrees on the 0.0004 au contour of the electron density using the GaussView software at the CPCM(CHCl₃)/M062X/6-31G** level of theory. The blue colour on the potential surface indicates a positive electrostatic potential and the red colour indicates a negative electrostatic potential, whereas the intermediate colours indicate intermediate electrostatic potentials. The colour scale was kept uniform for both the partners while constructing the potential surface. C_{g1} and C_{g2} in b) and c) are the geometric centers of frontier atoms of partner 1 (upper) and 2 (lower) respectively. The line joining C_{g1} and C_{g2} is defined to be the line of direction, which is the line along which the two partners will approach and interact with each other, as described by the nature of the electrostatic potential surfaces of the two partners in a). F₁, F₂, F₃ and F₄ in c) are the electrostatic forces acting on the atoms in one partner due to the charges on the atoms of the complementary partner. a₁, a₂, a₃, and a₄ are angles subtended by the forces F₁, F₂, F₃ and F₄ respectively from the line of direction. Therefore, F₁cos(a₁), F₂cos(a₂), F₃cos(a₃), and F₄cos(a₄) are the components of the forces F₁, F₂, F₃ and F₄ respectively along the line of direction. F is the net electrostatic force acting along the line of direction.

Table 3.1 Electrostatic Forces along the lines making angles 0°, 30°, 60°, 90°, up to 360° from the line connecting the centre of geometries of frontier atoms of hydrogen bonding partners for a sample set of 16 representative complexes of the planar hydrogen bond family

		0°	30°	60°	90°	120°	150°	180°	210°	240°	270°	300°	330°
X-1	a	-163.4	-122.2	-48.3	38.6	115.1	160.8	163.4	122.2	48.3	-38.6	-115.1	-160.8
	b	-451.8	-354.8	-162.8	72.9	289.0	427.7	451.8	354.8	162.8	-72.9	-289.0	-427.7
	c	-101.3	-77.8	-33.5	19.8	67.8	97.6	101.3	77.8	33.5	-19.8	-67.8	-97.6
X-2	a	-280.2	-207.8	-79.8	69.6	200.4	277.5	280.2	207.8	79.8	-69.6	-200.4	-277.5
	b	-634.9	-524.8	-274.2	49.9	360.7	574.8	634.9	524.8	274.2	-49.9	-360.7	-574.8
	c	-152.7	-120.1	-55.3	24.3	97.4	144.4	152.7	120.1	55.3	-24.3	-97.4	-144.4
X-3	a	-253.6	-190.0	-75.6	59.1	178.0	249.1	253.6	190.0	75.6	-59.1	-178.0	-249.1
	b	-541.1	-436.2	-241.5	17.9	272.5	454.1	514.1	436.2	241.5	-17.9	-272.5	-454.1
	c	-133.8	-106.6	-51.1	18.2	82.7	125.0	133.8	106.6	51.1	-18.2	-82.7	-125.0
X-4	a	-135.1	-96.6	-32.3	40.7	102.8	137.4	135.1	96.6	32.3	-40.7	-102.8	-137.4
	b	-609.9	-447.7	-165.6	160.9	444.3	608.7	609.9	447.7	165.6	-160.9	-444.3	-608.7
	c	-81.8	-59.7	-21.6	22.3	60.2	82.0	81.8	59.7	21.6	-22.3	-60.2	-82.0
X-5	a	-326.4	-242.7	-93.9	80.0	232.5	322.7	326.4	242.7	93.9	-80.0	-232.5	-322.7
	b	-667.5	-562.9	-307.6	30.2	359.9	593.1	667.5	562.9	307.6	-30.2	-359.9	-593.1
	c	-161.8	-130.8	-64.8	18.6	97.1	149.5	161.8	130.8	64.8	-18.6	-97.1	-149.5
X-6	a	-202.8	-148.4	-54.3	54.4	148.5	202.8	202.8	148.4	54.3	-54.4	-148.5	-202.8
	b	-487.3	-409.9	-222.7	24.3	264.7	434.2	487.3	409.9	222.7	-24.3	-264.7	-434.2
	c	-117.0	-93.1	-44.2	16.5	72.8	109.6	117.0	93.1	44.2	-16.5	-72.8	-109.6
X-7	a	-105.0	-74.5	-24.0	32.9	81.0	107.4	105.0	74.5	24.0	-32.9	-81.0	-107.4
	b	-621.9	-459.1	-173.4	158.9	448.5	618.0	621.9	459.1	173.4	-158.9	-448.5	-618.0
	c	-108.0	-78.5	-28.1	29.9	79.9	108.5	108.0	78.5	28.1	-29.9	-79.9	-108.5
X-8	a	-267.0	-195.1	-71.0	72.2	196.0	267.3	267.0	195.1	71.0	-72.2	-196.0	-267.3
	b	-568.4	-484.9	-271.5	14.7	296.9	499.6	568.4	484.9	271.5	-14.7	-296.9	-499.6
	c	-151.3	-120.1	-56.7	21.9	94.7	142.0	151.3	120.1	56.7	-21.9	-94.7	-142.0
X-9	a	-450.8	-333.7	-127.2	113.4	323.6	447.1	450.8	333.7	127.2	-113.4	-323.6	-447.1
	b	-812.1	-680.5	-366.6	45.6	445.5	726.1	812.1	680.5	366.6	-45.6	-445.5	-726.1
	c	-245.9	-193.6	-89.3	38.8	156.6	232.4	245.9	193.6	89.3	-38.8	-156.6	-232.4
X-10	a	-243.8	-185.2	-77.0	51.8	166.8	237.0	243.8	185.2	77.0	-51.8	-166.8	-237.0
	b	-846.0	-670.8	-315.8	123.8	530.3	794.6	846.0	670.8	315.8	-123.8	-530.3	-794.6
	c	-166.2	-126.8	-53.4	34.3	112.8	161.1	166.2	126.8	53.4	-34.3	-112.8	-161.1
X-11	a	-75.8	-51.6	-13.5	28.2	62.3	79.7	75.8	51.6	13.5	-28.2	-62.3	-79.7
	b	-321.0	-256.2	-122.7	43.7	198.3	299.8	321.0	256.2	122.7	-43.7	-198.3	-299.8
	c	-66.1	-51.1	-22.4	12.3	43.7	63.4	66.1	51.1	22.4	-12.3	-43.7	-63.4
X-12	a	-95.5	-58.4	-5.7	48.5	89.8	107.0	95.5	58.4	5.7	-48.5	-89.8	-107.0
	b	-307.7	-262.0	-146.1	9.0	161.7	271.0	307.7	262.0	146.1	-9.0	-161.7	-271.0
	c	-89.3	-70.7	-33.1	13.3	56.2	84.0	89.3	70.7	33.1	-13.3	-56.2	-84.0
X-13	a	-301.8	-208.6	-59.5	105.5	242.2	314.1	301.8	208.6	59.5	-105.5	-242.2	-314.1
	b	-674.3	-455.8	-115.1	256.4	559.2	712.2	674.3	455.8	115.1	-256.4	-559.2	-712.2
	c	-160.0	-107.8	-26.8	61.5	133.2	169.3	160.0	107.8	26.8	-61.5	-133.2	-169.3
X-14	a	-167.9	-135.3	-66.4	20.3	101.6	155.6	167.9	135.3	66.4	-20.3	-101.6	-155.6
	b	-680.4	-547.3	-267.6	83.9	412.8	631.2	680.4	547.3	267.6	-83.9	-412.8	-631.2
	c	-164.7	-139.6	-77.0	6.1	87.7	145.7	164.7	139.6	77.0	-6.1	-87.7	-145.7
X-15	a	-521.1	-399.3	-170.6	103.9	350.5	503.2	521.1	399.3	170.6	-103.9	-350.5	-503.2
	b	-	-931.5	-440.1	169.2	733.2	1100.7	1173.3	931.5	440.1	-169.2	-733.2	-
	c	1173.3	-	-	-	-	-	-	-	-	-	-	1100.7
X-16	a	-241.0	-190.4	-88.7	36.7	152.3	227.1	241.0	190.4	88.7	-36.7	-152.3	-227.1
	b	-273.3	-212.8	-96.2	46.1	176.0	258.8	273.3	212.8	96.2	-46.1	-176.0	-258.8
	c	-869.5	-684.0	-315.2	138.1	554.3	822.0	869.5	684.0	315.2	-138.1	-554.3	-822.0
	c	-163.2	-129.9	-61.8	22.9	101.4	152.8	163.2	129.9	61.8	-22.9	-101.4	-152.8

a = computed electrostatic force using Mulliken Charges; b = computed electrostatic force using NBO charges; c = computed electrostatic force using Löwdin Charges; all forces are in pN; the values at 360° are exactly same to the values at 0°, and hence have not been shown here.

Please see Figure 3.4 for the optimized geometries X-1 to X-16.

partners. The EF (EDA) thus obtained was compared with the EF obtained by our charge analysis approach. Linear plots with correlation coefficients of 0.88, 0.90 and 0.84 were obtained for the Mulliken (Figure 7a), NBO (Figure 7b) and Löwdin (Figure 7c) charge analyses methods. Furthermore, since a change of 0.1 Å in the distance between the two centers of geometries should not lead to any appreciable change in the geometries, it is expected that single point calculations with the translated geometries, followed by the EDA analysis, will also provide the same results. This was indeed found to be the case: plots of EF vs the EF(EDA) obtained by our approach showed similar correlation of 0.87, 0.92 and 0.85 for the Mulliken (Figure 7d), NBO (Figure 7e) and Löwdin (Figure 7f) charge analyses methods. These results suggest that the EFs obtained by employing our point charge analysis correlates excellently and linearly with the EFs obtained from the EDA analysis, and thereby validates our approach. Hence, this suggests that this method (employing point charges) can be employed for computing the electrostatic force of binding for any noncovalent bonded systems when the bonding is predominantly dominated by electrostatic contributions.

Herein, we have discussed two methods to obtain the electrostatic force of binding in hydrogen bonded complexes: (i) the one that employs atoms as point charges and (ii) the EF (EDA) method, that employs the electrostatic energy of binding obtained from EDA analysis. There are several advantages of (i) over (ii). It is a simple and general method and can be exploited effectively employing any quantum chemical (QM) package. However, the EDA analysis is not implemented in all QM packages and hence its application is limited. The exact contributions from every individual atom, atom-pairs or molecular fragments towards the overall stabilization of the system can be using (i). Thus, (i) can also be exploited for the rational designing of new superior noncovalent bonded systems or in the fine-tuning of the binding strength of the existing hydrogen bonded systems by just allowing minimal changes in the molecular structure as per the requirement. It is to be noted here that there is no method reported so far that can be employed effectively to pinpoint the binding strength of individual hydrogen bonds in multi-hydrogen bonded systems. However, the binding strength of each and every hydrogen bond in multi-hydrogen bonded systems may be obtained in terms of the binding force using (i), when the pair-wise atomic forces between different partners are computed along the line joining

acceptor and donor atoms of a given hydrogen bond. Considering the multi-hydrogen bonded complexes that are prevalent in biological systems such as proteins, DNAs, RNAs, carbohydrates and in supramolecular complexes and synthetic polymers, method (i) is of immense significance. On the other hand, method (ii) lags behind on all of these assessing parameters.

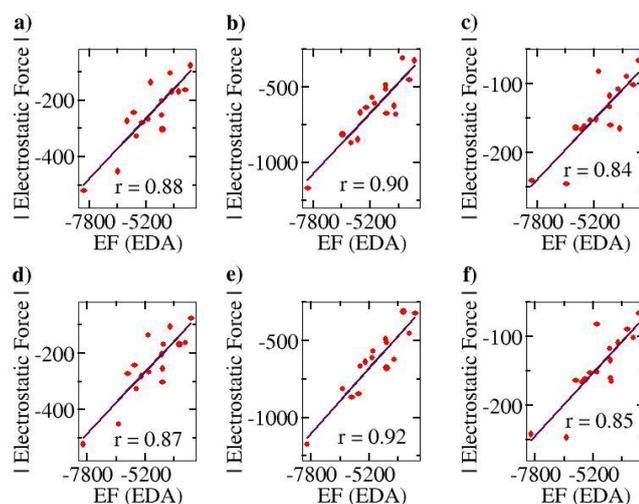


Figure 3.7 The Pearson correlation graph for planar hydrogen bonded molecules. EF (EDA) represents the EF (electrostatic force) obtained by calculating the negative of the gradient of the electrostatic energies of interaction between two points, calculated by the energy decomposition analysis (EDA) method. One of the two points considered for each complex is the optimized geometry, and the other point is the geometry obtained after translating one partner of each complex by 0.1 \AA away from the corresponding partner along the line of direction. a), b) and c) represent EF vs. EF (EDA) plots where the EF was calculated using Mulliken, NBO and Löwdin population analyses respectively and the EF (EDA) was calculated by constrained geometry optimization (keeping frontier atoms fixed) for the complexes obtained after translation. c), d) and e) are EF vs. EF (EDA) plots obtained when EF was calculated using Mulliken, NBO and Löwdin population analyses respectively, and the EF (EDA) was calculated by single point energy calculations of the complexes obtained after translation.

3.4 Scope of the Work

The sections above have demonstrated how the current approach can effectively be employed in determining the binding strength of electrostatic dominated noncovalent bonded systems, and contributions from different molecular parts or atom pairs. There are many areas in chemistry and biology where just a fine-tuning of hydrogen bond strength (or distance between binding partners) triggers the selectivity and the yield of the reaction.⁵⁹ It is likely, therefore, that the approach can be useful in providing deeper insights into many other areas of chemistry as well. Biomolecules, for instance, are primarily governed by hydrogen bonds that are dominated by long range electrostatic interactions. Furthermore, the unfolding behavior of proteins in acidic and basic media, as well as in certain salt solutions, can be better understood with the current approach. The mobility of ions in a liquid is significantly governed by long range solute solvent electrostatic interactions. The current approach would also help in understanding the behavior of hydrogen bonded solvents - this is because long range SEI have been found²²⁻³² to be impactful in determining the H-bond strength. The importance of hydrogen bonding and electrostatic interactions has also been seen in foldamer chemistry.^{4,13,14} It is also to be noted that the exciting and rapidly developing field of ionic liquids would be highly benefited by an understanding of long range electrostatic interactions, as the behavior and solvation properties of solutes in ionic liquids will depend on the SEI. Furthermore, noncovalent interactions have recently been exploited in stereochemical induction, where the approach of prochiral substrates to the chiral catalysts was allowed only from a specific direction.^{1-4,59} Our EF analysis approach would provide meaningful insights into the mechanistic understanding of such reactions, and that would help in tuning such systems for superior catalytic performance. This model could also be useful in understanding the behavior of ionic crystals.

3.5 Conclusions

The current work showcases a simple method based on evaluating the electrostatic force (EF) of interaction between two partners in molecular complexes where noncovalent electrostatic interaction is the dominant factor. More importantly, the current work emphasizes the significance of long range secondary interactions

between all atoms of one binding partner with all the atoms of the other. The work also shows that the consideration of directionality in defining such interactions is important and demonstrates how it can be implemented to obtain the binding strength of hydrogen bonded complexes. It has been shown that in multi-hydrogen bonded planar complexes, the line joining the center of geometries of the frontier atoms of two partners defines the molecular interactions most suitably. Forces obtained considering atoms in molecules as point charges correlates linearly with the forces obtained with the help of EDA analysis using the finite difference scheme. Given the diverse areas of chemistry, material sciences and biology where long range electrostatic interactions play a significant role in determining the strength of the overall interaction, the current approach is of significant relevance.

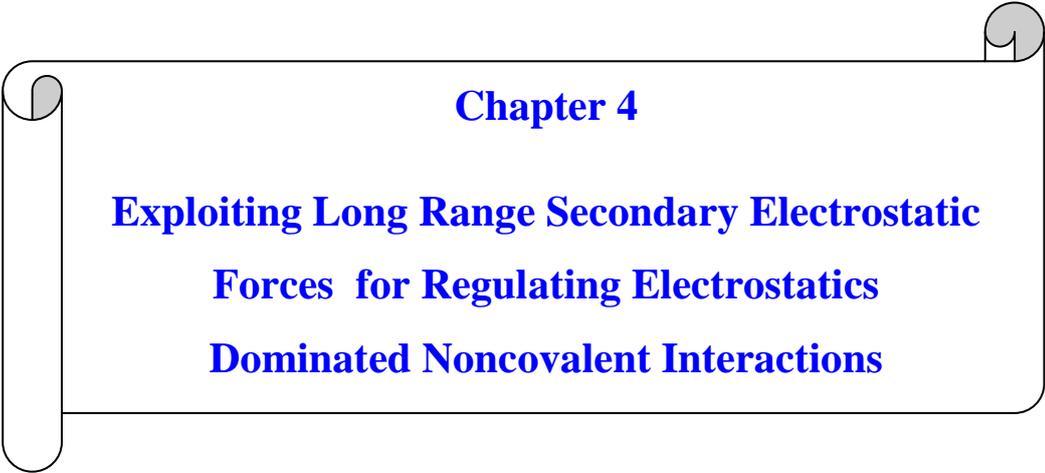
3.6 References

1. T. Šmejkal, B. Breit, *Angewandte Chemie*, 2008, **120**, 4010-4013.
2. A. G. Doyle, E. N. Jacobsen, *Chemical reviews*, 2007, **107**, 5713-5743.
3. R. R. Knowles, E. N. Jacobsen, *Proceedings of the National Academy of Sciences*, 2010, **107**, 20678-20685.
4. E. A. C. Davie, S. M. Mennen, Y. Xu, S. J. Miller, *Chemical reviews*, 2007, **107**, 5759-5812.
5. J.-M. Lehn, *Chemical Society Reviews*, 2007, **36**, 151-160.
6. L. Brunsveld, B. Folmer, E. Meijer, R. Sijbesma, *Chemical Reviews*, 2011, **101**, 4071-4098.
7. K. Biradha, *CrystEngComm*, 2003, **5**, 374-384.
8. C. B. Aakeröy, K. R. Seddon, *Chemical Society Review*,s 1993, **22**, 397-407.
9. A. Bauzá, T. J. Mooibroek, A. Frontera, *CrystEngComm*, 2016, **18**, 10-23.
10. R. G. Gonnade, M. S. Shashidhar, M. M. Bhadbhade, *Journal of the Indian Institute of Science*, 2012, **87**, 149.
11. B. Sellergren, M. Lepistöe, K. Mosbach, *J. Am. Chem. Soc.*, 1988, **110**, 5853-5860.
12. T. Rossow, S. Seiffert, *Supramolecular Polymer Networks and Gels, Vol. 268* (Eds.: S. Seiffert), Springer, 2015, pp. 1-46.

13. D. J. Hill, M. J. Mio, R. B. Prince, T. S. Hughes, J. S. Moore, *Chemical reviews*, 2001, **101**, 3893-4012.
14. D.-W. Zhang, X. Zhao, J.-L. Hou, Z.-T. Li, *Chemical Reviews*, 2012, **112**, 5271-5316.
15. D. Wang, G. Tong, R. Dong, Y. Zhou, J. Shen, X. Zhu, *Chemical Communications*, 2014, **50**, 11994-12017.
16. H. Yang, B. Yuan, X. Zhang, O. A. Scherman, *Accounts of chemical research*, 2014, **47**, 2106-2115.
17. F. Zhao, M. L. Ma, B. Xu, *Chemical Society Reviews*, 2009, **38**, 883-891.
18. B. R. Beno, K. S. Yeung, M. D. Bartberger, L. D. Pennington, N. A. Meanwell, *Journal of medicinal chemistry*, 2015, **58**, 4383-4438.
19. T. Clark, P. Politzer, J. S. Murray, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2015, **5**, 169-177.
20. M.J. Plevin, D. L. Bryce, J. Boisbouvier, *Nat. Chem.*, 2010, **2**, 466-471.
21. Kawase, T. & Kurata, H. *Chemical reviews* **106**, 5250-5273 (2006).
22. W. L. Jorgensen, J. Pranata, *J. Am. Chem. Soc.*, 1990, **112**, 2008-2010; J. Pranata, S. G. Wierschke, W. L. Jorgensen, *J. Am. Chem. Soc.*, 1991, **113**, 2810-2819.
23. B. A. Blight, C. A. Hunter, D. A. Leigh, D. A., H. McNab, P. I. Thomson, *Nature Chemistry*, 2011, **3**, 244-248.
24. T. J. Murray, S. C. Zimmerman, *J. Am. Chem. Soc.*, 1992, **114**, 4010-4011.
25. B. A. Blight, C.-C. Amaya, S. Djurdjevic, M. Kaller, D. A. Leigh, F. M. McMillan, H. McNab, A. M. Slawin. *J. Am. Chem. Soc.*, 2009, **131**, 14116-14122.
26. S. Djurdjevic, D. A. Leigh, H. McNab, S. Parsons, G. Teobaldi, F. Zerbetto, *J. Am. Chem. Soc.*, 2007, **129**, 476-477.
27. D. A. Leigh, C. C. Robertson, A. M. Slawin, P. I. Thomson, *J. Am. Chem. Soc.*, 2013, **135**, 9939-9943.
28. J. Sartorius, H. J. Schneider, *Chemistry-A European Journal*, 1996, **2**, 1446-1452.
29. F. H. Beijer, H. Kooijman, A. L. Spek, R. P. Sijbesma, E. Meijer, *Angew. Chem. Int. Ed.*, 1998, **37**, 75-78.

-
30. J. R. Quinn, S. C. Zimmerman, J. E. D. Bene, I. Shavitt, *J. Am. Chem. Soc.*, 2007, **129**, 934-941.
31. H. Zeng, R. S. Miller, R. A. Flowers, B. Gong, *J. Am. Chem. Soc.*, 2000, **122**, 2635-2644; R. R. Gardner, S. H. Gellman, *Tetrahedron*, 1997, **53**, 9881-9890; J. Yang, S. H. Gellman, *J. Am. Chem. Soc.*, 1998, **120**, 9090-9091; K. S. Jeong, T. Tjivikua, A. Muehldorf, G. Deslongchamp, M. Famulok, J. Jr. Rebek, *J. Am. Chem. Soc.*, 1991, **113**, 201-209.
32. P. L. Popelier, L. Joubert, *J. Am. Chem. Soc.*, 2002, **124**, 8725-8729.
33. T. Berlin, *J. Chem Phys.*, 1951, **19**, 208-213.
34. R.F. W. Bader, W.H. Henneker, P. E. Cade, *J. Chem Phys.*, 1967, **46**, 3341-3363; R. F. W. Bader, *J. Am. Chem. Soc.*, 1964, **86**, 5070-5075; R.F.W. Bader, H.J.T. Preston, *Can. J. Chem.*, 1966, **44**, 1131-1145.
35. O. Johnson, *Chem Scr.*, 1974, **6**, 202-297.
36. T. Koga, H. Nakatsuji, T. Yonezawa, *J. Am. Chem. Soc.*, 1978, **100**, 7522-7527.
37. X. Wang, Z. Peng, *International Journal of Quantum Chemistry*, 1993, **47**, 393-404.
38. S. T. Epstein, *The Variational Method in Quantum Chemistry Academic*, New York, 1974, 250.
39. H. Silberbach, *J. Chem. Phys.*, 1991, **94**, 2977-2985.
40. A. D. Buckingham, P. W. Fowler, *The Journal of chemical physics*, 1983, **79**, 6426-6428; A. D. Buckingham, P. W. Fowler, *Can. J. Chem.*, 1985, **63**, 2018-2025.
41. R. Ahlrichs, *Chem. Phys. Lett.*, 1989, **162**, 165-169.
42. J. P. Perdew, *Physical Review Letters*, 1996, **77**, 3865.
43. A. Klamt, G. Schüürmann *J. Chem. Soc. Perkin Trans. 2*, 1993, **5**, 799–805.
44. K. Eichkorn, *Chemical Physics Letters*, 1995, **240**, 283-289.
45. M. Sierka, *J. Chem. Phys.*, 2003, **118**, 9136-9148.
46. J. Hepburn, G. Scoles, R. Penco. *Chem. Phys. Lett.*, 1975, **36**, 451–456; R. Ahlrichs, R. Penco, G. Scoles. *Chem. Phys.*, 1977, **19**, 119–130; S. Grimme, *J. Comput. Chem.*, 2004, **25**, 1463–1473; S. Grimme, *J. Comput.*

- Chem.*, 2006, **27**, 1787–1799; S. Grimme; J. Antony; S. Ehrlich; H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
47. R. S. Mulliken, *J. Chem. Phys.*, 1955, **23**, 1833-1840.
48. A. E. Reed, R. B. Weinstock, F. Weinhold, *J. Chem. Phys.*, 1985, **83**, 735-46.
49. P. O. Löwdin, *J. Chem. Phys.*, 1955, **21**, 374-375.
50. R. P. Feynmann, *Phys. Rev.*, 1939, **56**, 340-343.
51. P. Politzer, K. E. Riley, F. A. Bulat, J. S. Murray, *Comput. Theor. Chem.*, 2012, **998**, 2-8; M. Hennemann, J. S. Murray, P. Politzer, K. E. Riley, T. Clark, *J. Mol. Model.*, 2012, **18**, 2461-2469; K. Dyduch, M. P. Mitoraj, A. Michalak, *J. Mol. Model.*, 2013, **19**, 2747-2758.
52. P. Politzer, J. S. Murray, and T. Clark, *Phys. Chem. Chem. Phys.*, 2013, **15**, 11178-11189.
53. M. R. Scholfield, C. M. V. Zanden, M. Carter, P. S. Ho, *Protein Science*, 2013, **22**, 139-152.
54. P. Politzer, J. S. Murray, and T. Clark, σ -Hole Bonding: A Physical Interpretation. *In halogen Bonding I, Vol. 358*, Springer International Publishing, 2014, PP. 19-42.
55. M. A. A. Ibrahim, *Journal of computational chemistry*, 2011, **32**, 2564-2574.
56. W. L. Jorgensen, P. Schyman, *Journal of chemical theory and computation*, 2012, **8**, 3895-3901; M. Kolář, P. Hobza, *Journal of chemical theory and computation*, 2012, **8**, 1325-1333.
57. P. Politzer, J. S. Murray, T. Clark, *Physical Chemistry Chemical Physics*, 2010, **12**, 7748-7757; P. Politzer, J. S. Murray, *ChemPhysChem* 2013, **14**, 278-294; P. Politzer, J. S. Murray, G. V. Janjić, S. D. Zarić, *Crystals*, 2014, **4**, 12-31.
58. Y. Hisamatsu, N. Shirai, S. I. Ikeda, K. Odashima, *Organic letters*, 2010, **12**, 1776-1779.
59. Yoichiro K., Haruka I., Mitsumi N. & Motomu K. *Nature Chemistry*, 2015, **7**, 712-716.



Chapter 4

**Exploiting Long Range Secondary Electrostatic
Forces for Regulating Electrostatics
Dominated Noncovalent Interactions**

Chapter 4

Exploiting Long Range Secondary Electrostatic Forces for Regulating Electrostatics Dominated Noncovalent Interactions

Abstract: It has been well established that long range secondary electrostatic interactions (SEIs) have a significant effect on the stability of supramolecular complexes. However, general rules to exploit SEIs for the rational design of diverse supramolecular complexes have been difficult to obtain. In this work, we outline a quantum chemical approach that can be exploited for this purpose. This approach is seen to provide excellent correlation between the electrostatic force of binding and the binding energy of two partners in hydrogen bonded complexes, as well as of two ions in ion-pair complexes. Furthermore, we illustrate how the comprehensive analysis of the electrostatic force of binding allows for the rational design of new complexes where the association constant between the two partners can be increased or decreased as desired, by several orders of magnitude. Hence, the current work showcases a general, simple and powerful method of understanding and exploiting long range secondary electrostatic interactions. We have further shown that the consideration of electrostatic force leads to two distinct regions around the electrostatic dominated noncovalent bonds: binding and anti-binding regions, similar to what has been reported by Berlin, Bader and others on covalent compounds, by considering the actual electron density.

4.1 Introduction

Noncovalent interactions play crucial roles in defining specific three dimensional structures of biomolecules such as DNAs, RNAs and proteins. These discrete and definitive structures of biomolecules are responsible for diverse biomolecular functions. Noncovalent interactions also govern certain physical and chemical properties of almost all chemical substances including water and other key solvents. Drawing inspiration from nature, noncovalent interactions have been exploited in

recent times to accomplish certain targets in multiple and diverse areas of chemistry and biology such as stereoselective catalysis,¹⁻⁴ supramolecular chemistry,^{1,5-6} crystal engineering,⁷⁻¹⁰ polymer chemistry,^{6,11-12} peptido-mimetic chemistry,^{4,13-14} and molecular medicines¹⁵⁻¹⁸ and so on. The interactions that have been exploited maximally in such cases are those that are dominated by electrostatic contributions such as hydrogen bonding, ion-pair interactions, σ -hole interactions, CH(polar)- π interactions and so on.

It has been shown that long range secondary electrostatic interactions (SEIs) play a significant role in determining the stability of hydrogen bonded complexes,¹⁹⁻²⁰ which is dominated by electrostatic contributions.²¹ This had led experimentalists to a design principle of putting all the donors groups on one partner and acceptors on the other (also referred to as Jorgensen's hypothesis¹⁹) in order to obtain the strongest binding in the planar hydrogen bonded complexes having a hydrogen bonded pattern similar to the nitrogen bases.²²⁻²⁹ However, the results of some subsequent studies have also questioned the reliability of this design principle.^{20,22-23,30} This is particularly because the interactions between remotely placed atoms present at two partners have also been found to be significant.²⁰ Moreover, a general rule to exploit SEIs for the rational design of diverse supramolecular complexes has proved difficult to obtain. In addition to this, a simple and general method to pinpoint the contributions of every part or region of hydrogen bonded partners towards the overall stability of an electrostatic dominated noncovalent interaction is lacking. It must also be noted that all the hydrogen bonded complexes that exist in nature or those that are synthesized chemically are not planar in structure. Also, it is not always desirable to have very strong noncovalent bonding between noncovalent bonded partners in each and every system. For example, in the case of homogeneous olefin polymerization catalysts, it is desirable to have as weak binding as possible between the two counter ions in order to produce a vacancy at the active site of the cationic metallocene center so as to allow new olefins to come, bind and polymerize.³¹⁻³⁴ There are also some cases where just a little tuning in the hydrogen bond strength is preferable in order to achieve improved stereoselectivity in the product.³⁵

In the current work, we propose the determination and understanding of the electrostatic forces (EFs) of binding between two partners as a viable method for the rational designing of new superior systems of desirable binding

strength. The details of the method have been provided in Chapter 3. The forces, in this study, have been calculated by employing Coulomb's law, with the atoms being considered as point charges. The charges on the atoms have been obtained from quantum chemical calculations. All the pair-wise interactions between atoms at the two bonded partners have been considered in order to account for the long range electrostatic influence. The net electrostatic force of interaction between two partners has been determined in a particular direction assigned after careful analysis of the molecular structures.

The current investigation focuses on two completely different families of complexes where the electrostatic interaction has been known to be the significant contributing factor: (i) the near-planar hydrogen bonded molecular complexes that have been studied extensively by Leigh and co-workers, as well as others, and that are models for biological systems,^{22,24-26} and (ii) contact ion-pairs that are very significant in homogenous olefin polymerization,³¹⁻³⁴ which have been modeled by binding the cationic zirconocene, Cp_2ZrMe^+ with several different counterions. These two illustrative examples have been specifically chosen in order to highlight the efficacy of the current approach, because they represent two completely different challenges in their structure and function: while the goal of researchers working with the class of molecules in (i) has been to obtain as strong a binding as possible between the two partners, the objective in the field of homogenous olefin polymerization (case (ii)) is to make the interaction between the cation and the counterion as weak as possible. The current approach allows each of these objectives to be realized, showing its general versatility and usefulness, and allowing for the rational design of new systems that are significantly better than the state-of-the art in the different fields.

4.2 Computational Details

All the DFT calculations, until unless mentioned specifically, were carried out using the Turbomole 6.4 suite of quantum-chemical programs.³⁶ Geometry optimizations were performed using the PBE³⁷ functional in the solvent phase using the Conductor like Screening Model (COSMO)³⁸ employing chloroform (epsilon=4.81) as the solvent. The electronic configuration of the atoms was described by a triple-zeta basis set augmented by a polarization function

(TURBOMOLE basis set TZVP). The resolution of identity (RI)³⁹, along with the multipole accelerated resolution of identity (mari-j)⁴⁰ approximations were employed for an accurate and efficient treatment of the electronic Coulomb term in the density functional calculations. The option “disp” provided in Turbomole package (DFT-D2, a general, empirical dispersion correction proposed by Stefan Grimme for density functional calculations) was used for dispersion corrected DFT calculations for all the calculations with Turbomole.⁴¹ Only the electronic energies were considered for calculating the binding and interaction energies. The free energies of binding were calculated to determine the association constant of specific systems wherever mentioned in the manuscript. The binding energy (E_b) between two noncovalently bonded fragments was calculated using the following formula

$$E_b = E_{\text{comp}} - (E_{\text{o,frag1}} + E_{\text{o,frag2}}) \quad (4.1)$$

Where, E_{comp} is the energy of noncovalently bonded complex, and $E_{\text{o,frag1}}$ and $E_{\text{o,frag2}}$ are energies of two independently optimized fragments involved in the weak interactions. The interaction energy (E_i) between two noncovalently bonded fragments was calculated as

$$E_i = E_{\text{comp}} - (E_{\text{frag1}} + E_{\text{frag2}}) \quad (4.2)$$

Where, E_{comp} is the energy of noncovalently bonded complex, and E_{frag1} and E_{frag2} are single point energies of two fragments being separated from an optimized complex.

The Energy Decomposition analysis (EDA) was carried out using Turbomole 7.0 at the same level of theory and under the same conditions of solvent and dispersion that have been used for geometry optimizations using the Turbomole version 6.4 (the EDA implementation is not available with Turbomole 6.4).

The Gaussian geometries were optimized at the M06-2X/6-31G** level of theory⁴² using the Gaussian 09 suite of quantum-chemical programs.⁴³ The solvent effect was added through the Conductor like Polarization Continuum Model (CPCM) using chloroform as a common solvent for all the geometries considered.⁴⁴ To ensure that none of the atoms in any chosen system posses the σ -hole (as described in Computational Details and Back ground theory section of Chapter 3), the Molecular Surface electrostatic potential have been computed

on the 0.0004 au contour of the electron density for some representative partners, using the GaussView software at the CPCM(CHCl₃)/M062X/6-31G**of theory (Figure 4.1 below).

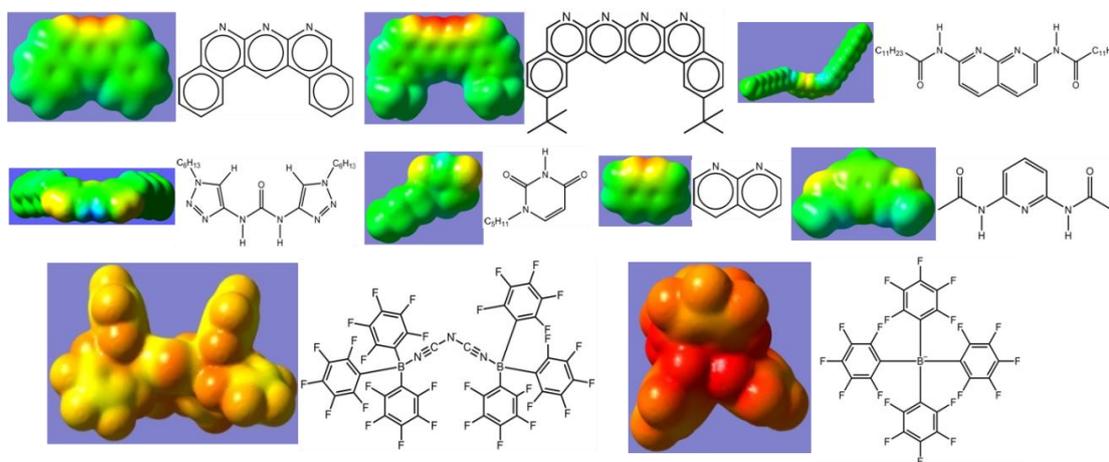


Figure 4.1 The molecular surface electrostatic potential in Hartrees, computed on the 0.0004 au contour of the electron density, using the GaussView software at the CPCM(CHCl₃)/M062X/6-31G**of theory with the geometries optimized using g09. The blue colour indicates positive electrostatic potential and the red color indicates negative electrostatic potential, whereas, the intermediate colours indicate intermediate electrostatic potential. The colour scale was kept uniform in all the cases.

The Mulliken⁴⁵ and NBO⁴⁶ charges have been used to calculate electrostatic forces on each fragment of every complex along a particular direction that described the intermolecular interactions most suitably (*i.e.* along the line of direction, see the Figures 4.3a and 4.3b). To compute the magnitude of the net force of binding, *i.e.*, the binding force, the vector sum of electrostatic forces experienced by each fragment along the above said direction has been considered. A flow chart revealing the algorithm of the code that was employed for computing forces is provided in Figure 3.3 of Chapter 3 and Figure 4.13 of this chapter. Löwdin⁴⁷ and CHelpG⁴⁸ charge analyses have also been employed for calculating the forces and for providing further validation of our method for a representative set of 16 planar hydrogen bonded structures.

A careful analysis reveals that the line joining the center of geometries of the frontier atoms in the planar hydrogen bonded complexes describes the interaction forces most appropriately. We have defined this line as line of direction. The detailed analysis of how a particular line has been chosen as the line of direction has been provided into Chapter 3. The frontier atoms of a hydrogen bonding partner are atoms that are nearest to the complementary partner and are directly involved in the hydrogen bonding. It has been assumed here that the two partners will approach each other along the line connecting the center of geometries of the frontier atoms. The pair-wise forces between atoms at two partners were, therefore, calculated along these lines to obtain the net force of binding for the planar hydrogen bonded complexes. However, in the case of the olefin polymerization catalysts, the line joining the central (metal) atom of the cation and the central atom of the anion was found to be most appropriate line of direction for portraying the intermolecular electrostatic force of interaction, as the magnitude of electrostatic force of binding was found to be more (with Mulliken and NBO charges) along this line in comparison to the line parallel to the Zr-F bond. The forces on one partner due to the presence of the other were calculated along this “line of direction” but in the direction of mutual approach as illustrated in Chapter 3.

4.3 Results and Discussion

In order to determine the EF existing between two partners, the approach that has been adopted has been to determine the Coulombic force between each pair of atoms, with the atoms being chosen from the different fragments, and then summing up the forces. This provides the net EF of interaction between the two fragments. As mentioned in the Introduction, the point charge approximation has been employed for this purpose, and each atom has been considered to have a charge, determined by the NBO and/or the Mulliken charge analysis. Since force has direction, the Coulombic interaction has been considered along a certain common line: the “line of direction” for the given molecule (see Chapter 3 for details of the method). Therefore, only the component of the force along that given line has been considered.

(i) **The Planar Hydrogen Bonded Structures Case.** This family of complexes involves two planar fragments interacting through X-H...Y hydrogen bonded interactions, with the number of such hydrogen bonds varying from two to four. The significance of such complexes lies in the fact that they can serve as model structures for investigating and understanding multipoint hydrogen bonding, which is highly relevant to biological systems as well as to multifunctional materials and supramolecular polymers.^{22,24-26,49} Hence, there have been a large number of reports in recent times that have discussed different planar structures belonging to this family. A typical optimized structure is shown in Figure 4.2a below: with four N-H...N hydrogen bonds connecting two fragments.

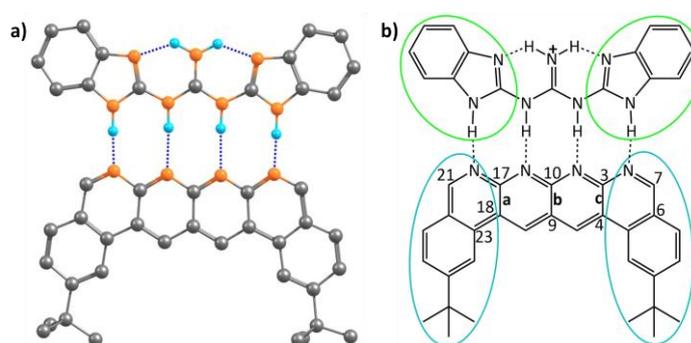


Figure 4.2 a) The optimized geometry of the cationic complex reported by Leigh *et al.*; black, cyan and orange colours represent carbon, hydrogen and nitrogen atoms respectively; dotted lines represent the hydrogen bonding interactions. Hydrogen atoms other than those involved in hydrogen bonding interactions have been deleted for clarity. b) A schematic picture of the cationic complex showing the three regions into which each of partners was divided for the electrostatic force analysis.

This family of complexes is an ideal choice for testing our approach because it has been shown that SEIs are very important for determining the stability of these complexes,²²⁻²⁹ which has led experimentalists to the design principle of having all the hydrogen bond acceptor groups in one fragment and all the hydrogen bond donor groups in the other¹⁹. We have therefore taken a sample set of sixteen different representative complexes from this family, optimized the structures (with Turbomole 6.4 and the COSMO(CHCl₃)/PBE/TZVP level of theory, see Figure 3.4 in Chapter 3) and obtained the total force of electrostatic interaction between the two fragments for

each case. The line of direction along which the force had been considered is the “center of geometry” of the frontier atoms of the two fragments, because this is the line along which the two partners that will hydrogen bond would approach and bind with each other. The flow chart of the code being implemented to calculate the EFs in this case has been provided in Figure 3.3 of Chapter 3. Furthermore, we have obtained the energy of binding of the two fragments for each case. A graph with the EF (with the charges obtained from a Mulliken charge analysis) on the Y Axis and the corresponding binding energies on the X Axis is shown in Figure 4.3a below. Gratifyingly, we find a near linear correlation ($r = 0.92$) between the two quantities for the sixteen cases considered. In order to show that the result is not an artifact of the method of charge analysis, we have repeated the EF calculations by taking the charges from the NBO analysis and have obtained a graph of comparable linear correlation ($r = 0.82$), as shown in Figure 4.3b below. It is believed that systems dominated by polarization are better represented by NBO charges in comparison to the Mulliken.⁴⁷ Since the long range electrostatic interactions between atoms in the two hydrogen bonding partners are unlikely to be influenced by polarization (as the atoms are from the first and second periods of the Periodic Table), this helps explain why Mulliken charges give better results than NBO, especially since the calculations have been done with good basis sets. In addition to this, in order to show that similar results would be obtained by other population calculation approaches, we have done the analysis with another (conceptually different) charge analysis method: Löwdin, and obtained a linear correlation ($r = 0.93$, see Figure 4.3c) for this as well. Furthermore, in order to show that the results are not dependent on the choice of basis set and functional, we have repeated the optimization calculations at the CPCM(CHCl₃)/M06-2X/6-31G** level of theory with Gaussian 09, with a slightly more diverse group of sixteen structures, and have found a similar correlation between the total EF of interaction and binding energies for both the Mulliken ($r = 0.89$, Figure 4.3d) and NBO charge ($r = 0.81$, Figure 4.3e) analysis cases. A further analysis with the charges obtained from the CHelpG method (Charges from Electrostatic Potentials using a Grid based method) for the same set of 16 molecules also gave a satisfactory correlation constant of $r = 0.78$ (Figure 4.3f). It is to be noted that charges assigned to the atoms by the CHelpG method account for the electrostatic potential around each atom, and therefore, employing such charges for the force

calculations in our method can be considered a means of accounting for the electron density around each atom.⁴⁸

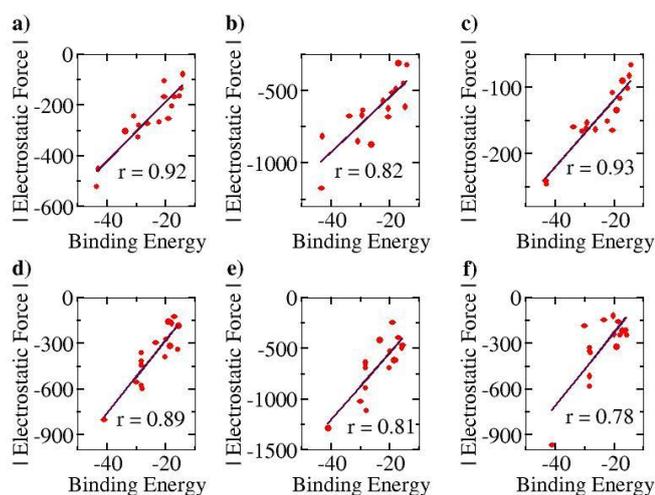


Figure 4.3 The Pearson correlation graph for planar hydrogen bonded molecules a) EF vs. E_b for Mulliken charges at the COSMO/PBE/TZVP Turbomole 6.4 geometries b) EF vs. E_b for NBO charges at the COSMO(CHCl₃)/PBE/TZVP Turbomole 6.4 geometries c) EF vs. E_b for Löwdin charges at the COSMO/PBE/TZVP Turbomole 6.4 geometries d) EF vs. E_b for Mulliken charges at the CPCM/M06-2X/6-31G** Gaussian 09 geometries e) EF vs. E_b for NBO charges at the CPCM(CHCl₃)/M06-2X/6-31G** Gaussian 09 geometries f) EF vs. E_b for ChelpG (Charges from Electrostatic Potentials using a Grid based method) charges at the CPCM(CHCl₃)/M06-2X/6-31G** Gaussian 09 level of theory.

To further corroborate our results with the electrostatic component of binding energy, we did an energy decomposition analysis (EDA) using Turbomole 7.0. A graph with the EF on the Y Axis and the corresponding electrostatic component of binding energies obtained from EDA method, “EE (EDA)”, on the X Axis, is shown in Figure 4.4 below. We found an improved correlation between the two quantities for the sixteen cases considered, for the Mulliken ($r = 0.94$, Figure 4.4a) and NBO ($r = 0.86$, Figure 4.4b) cases, as compared to the EF Vs. E_b plots (Figure 4.3a and 4.3b) shown earlier. This further shows that our approach correctly captures the electrostatic interaction between the two partners. Likewise, a good correlation was also obtained for the Löwdin population analysis ($r = 0.90$, Figure 4.4c) case.

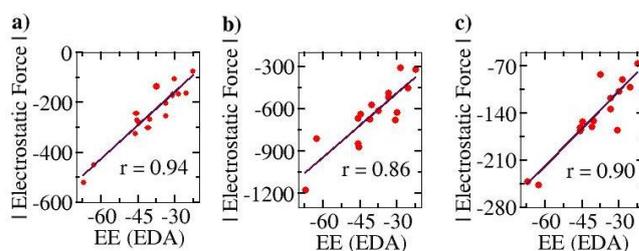


Figure 4.4 The Pearson correlation graph for planar hydrogen bonded molecules a) EF vs. EE (EDA) for Mulliken charges at the COSMO/PBE/TZVP Turbomole 6.4 geometries, b) EF vs. EE (EDA) for NBO charges at the COSMO(CHCl₃)/PBE/TZVP Turbomole 6.4 level of theory and c) EF vs. EE (EDA) for Löwdin charges at the COSMO(CHCl₃)/PBE/TZVP Turbomole 6.4 level of theory. EE (EDA) represents the Electrostatic component of the binding energy obtained from the Energy Decomposition Analysis (EDA) method implemented in Turbomole 7.0.

In order to further understand why the electrostatic force correlates so well with the binding energy, additional calculations have been done with a simple model system that is shown in Figure 4.5 below. The figure shows two hydrogen bonding partners that each possess two charged centres that are equidistant from their opposite charged counterparts. When the charges are all negative or all positive, the net force is repulsive (as shown in the figure), but when the charges in one partner are positive and the charges in the other partner are negative, then the net force would be attractive (not shown in the figure). The net electrostatic force (along with the direction obtained) was calculated for each charge combination. Furthermore, for each combination, the net electrostatic energy was also calculated. This was done for different sets of values for the charges. The set of thirty electrostatic force values thus obtained was correlated with the thirty corresponding electrostatic energy values, and an exact correlation was obtained (see Figure 4.5c). When an angle 180° different from the resultant line of force was taken, an exact negative correlation was obtained (see Figure 4.5d). This indicates that at a given distance and for a particular arrangement of atoms, the electrostatic force of binding (when considered along the direction of approach and along the line of direction) correlates perfectly with the electrostatic energy, i.e., the greater the magnitude of the force, the greater is the electrostatic binding energy. In other words, the electrostatic force and the electrostatic energy change by the same proportion when charges on atoms are

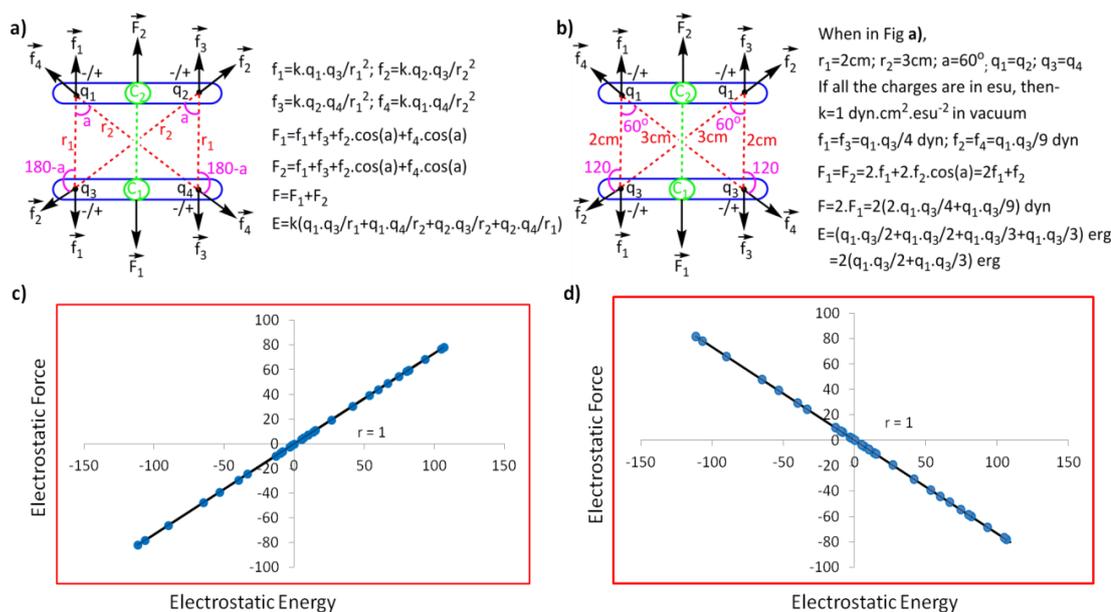


Figure 4.5 a) A simple model representing the electrostatic interaction between the two partners, each made up of two point charges of the same nature, in a two dimensional plane. f_1, f_2, f_3, f_4 represent the magnitude of forces experienced by the corresponding charged particles due to the charges on the other partner. C_1 and C_2 are the center of geometries of the two particles on the respective partners. The line joining C_1 and C_2 is the line of direction. F_1 and F_2 are the magnitude of forces experienced by the respective partners. F is the magnitude of the force of interaction between two partners. E is the electrostatic energy of interaction. b) A further simplified model when the magnitude and the nature of both the charges on each fragment are the same. c) The Pearson correlation graph between the electrostatic force of interaction along the direction of approach of the two partners and the electrostatic energy of interaction under the conditions described in b) for a set of 30 different values of q_1 and q_2 . The forces are in dyn and the energies are in erg. d) The Pearson correlation graph between the electrostatic force of interaction along the direction opposite to the direction of approach of the two partners and the electrostatic energy of interaction under the conditions described in b) for the same set of 30 different values of q_1 and q_2 .

changed in a complex while the partners remain at a fixed separation. This explains why good linear correlation has been obtained between the EF and binding energy within the chosen family of complexes in this study: the major influencing factor

within each family (when we move from one complex to the other in the family) is the charges on the atoms, and not the distances between the atoms, as the relative atomic arrangement of the atoms within each family of complexes is nearly the same. Furthermore, the reason why the total electrostatic force of interaction is seen to correlate so well with the binding energy is because of the fact that greater electrostatic interaction between the two partners allows them to overcome the Pauli repulsion force to a greater extent, thereby allowing them to bind more strongly in their equilibrium structures. Therefore, determining the electrostatic force of interaction along the line of direction, which is the line along which the two partners approach and bind, and the line along which the interaction between the partners is the greatest, is shown to be the correct approach for understanding the binding between the two partners.

Therefore, the results showcase the validity of our approach, and also illustrate the importance of taking the SEIs due to *all* the atoms in each fragment into consideration, rather than the SEIs for only the frontier atoms, as has been the traditional view.^{19,22-29} Indeed, a plot of the EF obtained by considering only the frontier atoms, versus the binding energy shows a poorer correlation ($r = 0.75$, for the Mulliken charge analysis case), as opposed to $r = 0.92$ when all atoms are taken into account. It is to be noted here that the interactions of the frontier atoms in quadruple hydrogen bonded complexes also include the secondary interactions between diagonal atoms, which were not taken into account in Jorgesen's SEIs hypothesis¹⁹. Furthermore, when the EF was calculated by calculating the forces (for all the atoms) along a line of direction perpendicular to the line employed in the calculations (the line connecting the center of geometries of the frontier atoms, as stated earlier), we observed a *negative* correlation of the EF with the binding energy, with $r = 0.85$ (the Mulliken charge analysis case), with the net EFs for each of the sixteen structures now found to be positive (see Figure 4.6). This further shows the significance of taking the direction of the electrostatic interaction into account. It is to be noted that Berlin⁵⁰ in the 1950s and Bader⁵¹ and coworkers in the 1960s had attempted to understand covalent bonding in compounds by dividing a molecule into "binding" and "anti-binding" regions based on electron density calculations. In effect, our current work shows that the line of direction is analogous to the "binding" and "anti-binding" regions that had been discussed by Berlin and Bader earlier, with a further important

distinction that the line of direction is now applied to non-covalently bonded supramolecular complexes.

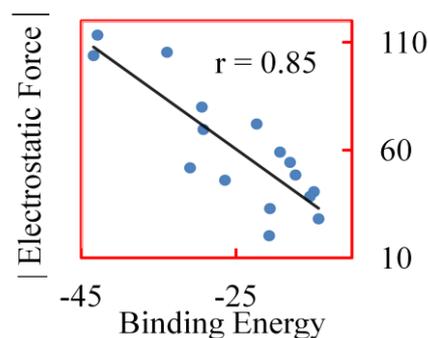


Figure 4.6 The EF vs. E_b Pearson Correlation graph for planar hydrogen bonded molecules for forces calculated along a line perpendicular to the line of direction of hydrogen bonds, by employing Mulliken charges for the geometries obtained at the COSMO(CHCl_3)/PBE/TZVP level of theory using Turbomole 6.4 package.

We further note here that Popelier *et al.* had also suggested that including only the frontier atom SEIs leads to erroneous results for different hydrogen bonded cases.²⁰ They pointed out that taking frontier atom SEIs as an indicator and a design principle was flawed, because the efficacy of frontier atom SEIs was only limited to specific cases. We have employed our approach (structures optimized at the COSMO(CHCl_3)/PBE/TZVP level of theory, Figure 4.7) to determine the correlation between the total EF (including all the atoms, between the two base pair fragments) and the binding energy (between the two base pair fragments), for the 28 base pair cases (including the uracil-diaminopyridine: U-DAP interaction, see Figure 4.7) that Popelier *et al.* had studied. The line of direction was taken, as before, to be the center of geometry of the frontier hydrogen bonding atoms. The results indicate significant correlation: $r = 0.74$ with the Mulliken charge analysis (Figure 4.8a), and $r = 0.79$ with the NBO charge analysis (Figure 4.8b). It is to be noted that the EF was calculated between the hydrogen bonding partners obtained after geometry optimization of the complexes, and the binding energy (E_b) of the complexes was obtained with respect to the infinitely separated partners. In order to investigate the effect of geometrical variance in the presence and absence of the hydrogen bonded partner, we have also done a Pearson Correlation analysis for the EF vs. the

interaction energy (E_i) plot (please see the Computational Details section for the description of how the E_b and the E_i have been calculated). A marginally improved correlation coefficient was obtained for both Mulliken ($r = 0.79$, Figure 4.8c) and NBO ($r = 0.81$, Figure 4.8d) charge analyses, indicating only a minor change in the geometry of molecular fragments while optimized independently, mainly due to their rigid aromatic framework. Hence, the calculations of the EF with our approach for the 28 base pair cases also shows good correlation with the binding energy, indicating that our method works for different families of hydrogen bonded complexes.

However, the advantage of employing the Jorgensen approach – that of looking at only the frontier atoms to understand the strength of interaction between the two interacting partners - was that it provided a simple means of designing systems that would bind more strongly and effectively. While the current work shows the limitations of that approach, does it then also provide a superior means of design that would lead to more strongly binding systems? It does indeed do so: since the code determines the specific force of interaction between each atom in a given partner and all the atoms of the other partner, one can write down the value for each such interaction in an output file, and then look at the output to determine which specific

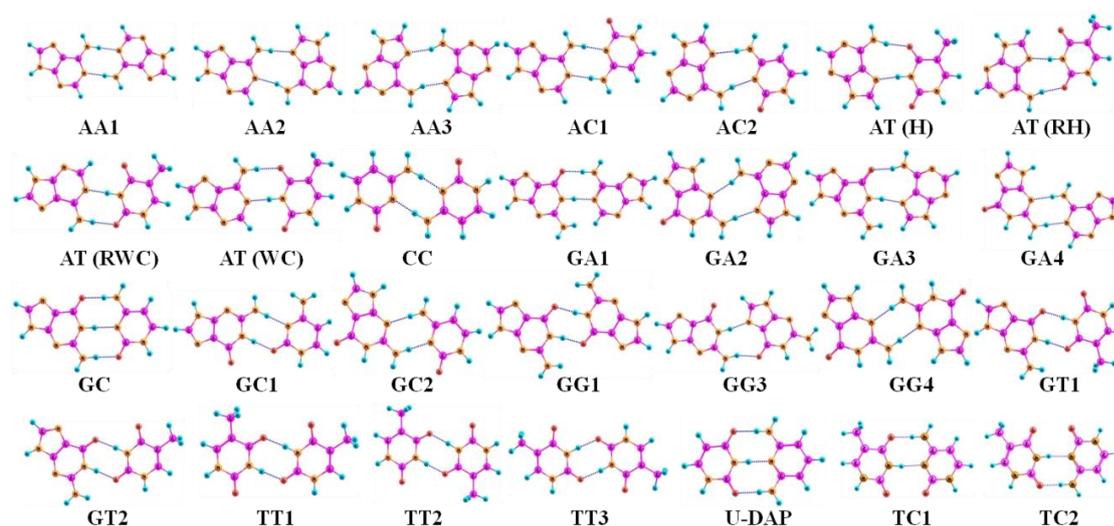


Figure 7. The optimized geometries of 28 base pairs considered by Popelier *et al.* in their QTAIM studies. The same convention of nomenclature of base pairs is followed. Colour representation: pink – carbon, cyan – hydrogen, brown – nitrogen, strawberry – oxygen, dotted blue lines - hydrogen bonds.

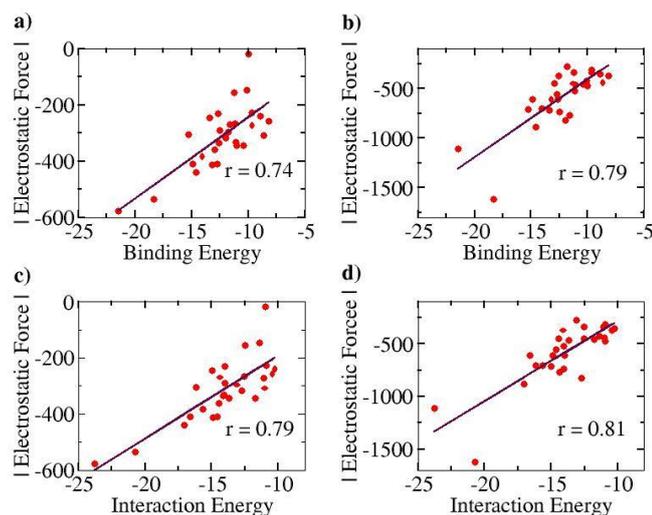


Figure 4.8 The Pearson correlation graph for nitrogen base pairs a) EF vs. E_b for Mulliken charges b) EF vs. E_b for NBO charges c) EF vs. E_i for Mulliken charges d) EF vs. E_i for NBO charges.

interactions are the strongest. As an example, we consider the structure shown in Figure 4.2, which has been shown to have the strongest binding,²² among all the planar hydrogen bonded structures that have been studied to date. Indeed, we obtained the binding energy for the two partners in this case to be -43.4 kcal/mol, which is the highest among all the cases considered in this study. We aimed to improve the binding energy of the system by changing the non-frontier atoms (which were not considered in the Jorgensens theory) in order to ascertain the impact of long range secondary interactions. We first calculated the net interaction forces experienced by every non-frontier atom on the acceptor partner due to all the atoms on the donor partner and *vice versa*. A perusal of the output files shows that both the strongest attractive and repulsive interactions are experienced by atoms from the middle region (see Figure 4.2b) of the acceptor partner. The atoms 3 and 17 face most attractive interactions followed by atoms 10, 7 and 21 respectively, whereas the most repulsive interactions experienced by atoms other than H are 9 and 4 followed by 23, 18 and 6 respectively (see Figure 4.2b). An examination of individual atom-atom interactions between the rear atoms of one partner and each atom of the other partner as well suggest the same trend in the electrostatic forces, as atoms 3 and 17 face the most attractive and 9 and 4 face the most repulsive interactions respectively, all belonging to the middle region of the acceptor segments. Interestingly, as revealed from the

output files, the magnitude of the attractive interactions are greater than the magnitude of the repulsive interactions. Based on this result, in a further analysis, we divided each donor and acceptor molecule into three regions (left, middle and right), as shown in Figure 4.2b, and calculated the net force experienced by each region due to all the atoms on the complementary partner. The output further suggests that the middle region of the acceptor contributes the most to the net attractive interaction between the two partners.

Based on this analysis, in order to enhance the net attractive EF between the two partners, we first replaced bond **a** with the isoelectronic B-N bond to obtain a new structure. Another structure was obtained by doing the same with bond **b**. Then, a new structure was obtained by replacing bonds **a** and **c** together. Finally, all of the bonds **a**, **b** and **c** were substituted at once with the isoelectronic B-N bonds, as shown in Figure 4.9 below. The newly designed AAAA-DDDD cationic systems were obtained with improved binding energies of -49.6, -49.2, -55.7, -60.2 kcal/mol respectively (see Figure 4.9 a-d). This translates to association constants that are orders of magnitude greater than those obtained for the best known case to date, for which the binding energy is -43.4 kcal/mol. An electrostatic force analysis reveals that the increase in binding affinity occurs mainly due to more favorable interactions between the frontier atoms of the newly designed complex, caused by the altered electronics of the frontier atoms due to modification at the remote sites. An electrostatic potential surface map also suggests accumulation of more negative electrostatic potential near the frontier atoms in the newly designed acceptor partner with respect to the originally synthesized molecule (see Figure 4.9e). It is to be noted here that the frontier atom interactions in quadruply hydrogen bonded complexes also include the secondary interactions between diagonal atoms, which have been overlooked in Jorgensen's analysis¹⁹ (Chapter 3, Figure 3.1a). However, the contributions from the remote atoms cannot be neglected.

This modification has been extended to other similar molecules that have been considered in this study (Scheme 4.1 below). In all the cases, the binding energy of the newly obtained complexes was found to be increased. Recent reports suggest that replacements of C-C bonds of aromatic systems with isoelectronic B-N moieties are possible with specified strategies;⁵² hence the synthesis of such proposed compounds would be quite feasible.

Furthermore, it has been shown in the literature that electrostatic interactions are weak in CH- π interactions, where the attractive dispersion interaction is the more dominant.⁵³ Since our current work shows the significance of interactions in atoms far away from the frontier hydrogen bonding region, another simple recipe for improving the bonding would be to add phenyl rings in one of the two partners, as shown in Figure 4.10 below. The primary electrostatic impact would be marginal, and all the EFs between the phenyl ring atoms and the atoms of the partner would have a very small component along the line of direction. Therefore, the electrostatic effect of adding the phenyl rings would be small, while the system would benefit from attractive dispersion interactions, thereby improving the binding strength. This is indeed seen to be true: as shown in Figure 4.10a and 10b, which are modifications of the structure in Figure 4.2b, with phenyl rings at the end of the frontier line, have binding energies of -50.2 and -50.8 kcal/mol respectively, i.e., about 7.0 kcal/mol

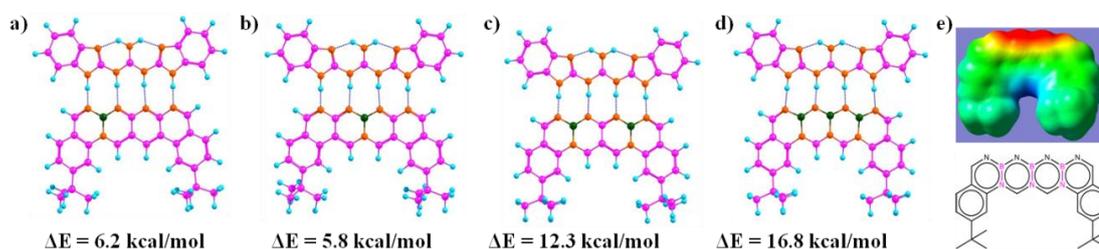
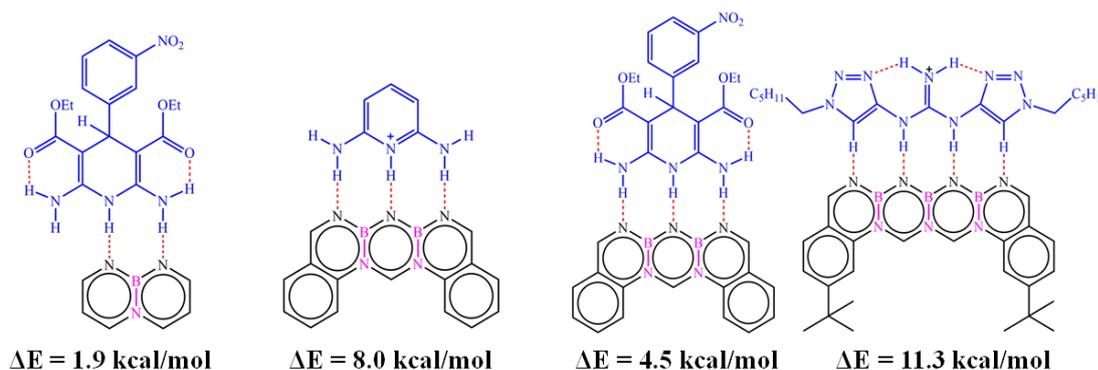


Figure 4.9 The optimized geometries of newly designed cationic AAAA-DDDD hydrogen bonded complexes, where the C-C bond on the middle region of the acceptor partner is replaced with the isoelectronic B-N bonds (a to d). Pink, cyan, brown and green colors represent carbon, hydrogen, nitrogen and boron atoms respectively, whereas, dotted blue lines represent hydrogen bonds. ΔE is the increase in binding energy over their parental complexes ($E_b = -43.4$ kcal/mol) from where they are derived. e) The molecular surface electrostatic potential of acceptor partner in d) in Hartrees, computed on the 0.0004 au contour of the electron density, using the GaussView software at the CPCM(CHCl₃)/M062X/6-31G** of theory with the geometries optimized using g09. The blue colour indicates positive electrostatic potential and the red color indicates negative electrostatic potential, whereas, the intermediate colours indicate intermediate electrostatic potential. The colour scale was kept uniform for all the cases.

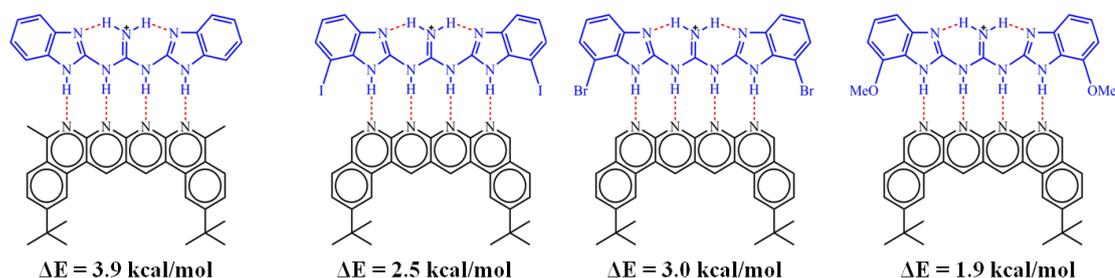


Scheme 4.1 A schematic picture of some newly designed acceptor-donor planar hydrogen bonded complexes based on our electrostatic force analysis, where central C-C bonds on acceptor moieties are replaced by B-N bonds; ΔE = the improved binding energy over their corresponding parental complexes, from where they are derived.

stronger than the binding energy obtained for the Figure 4.2b structure. To further examine the effect of non-directional attractive dispersion forces on the binding strength of complex 2a, we next substituted frontier non-hydrogen bonded hydrogen atoms on acceptor and donor partners with sterically less demanding substituents - CH_3 , I, Br and $-\text{OCH}_3$ respectively, as shown in Scheme 4.2. Improved binding energies were obtained with respect to the parent AAAA-DDDD complex, which suggests that the binding gets benefited by dispersion interactions in these designed complexes. However, dispersion can be closely counteracted by sterics, which works towards destabilizing the complex by weakening the existing hydrogen bonds. Hence, the unhelpful steric effect of adding new groups should also be kept in mind when designing such systems.

Now, taking a hypothetical “best case” design improvement of the Figure 4.2b structure (Figure 4.9d), one can put three B-N pairs in the place of the C-C bonds, and put *ortho* and *para* methyl substituted (so as to increase dispersion interactions) phenyl rings at each end, in order to get a new structure (see Figure 4.10c). Such a structure would be expected to have the best binding energy. This is indeed seen to be the case: we find that the binding energy of this complex is as high as -69.6 kcal/mol , i.e. 60.4 % greater binding than for the strongest binding structure known to date! Such a complex would have an association constant that would be as much as $1.9 \times$

10^{15} times greater than the association constant of the “best” known complex of this family, which shows the great power of understanding and exploiting SEIs by the current approach.



Scheme 4.2 A schematic picture of some newly designed acceptor-donor planar hydrogen bonded complexes, where the attractive non-directional dispersive force was exploited for improved binding; ΔE = the improved binding energy over their parental complex ($E_b = -43.4$ kcal/mol), from where they are derived.

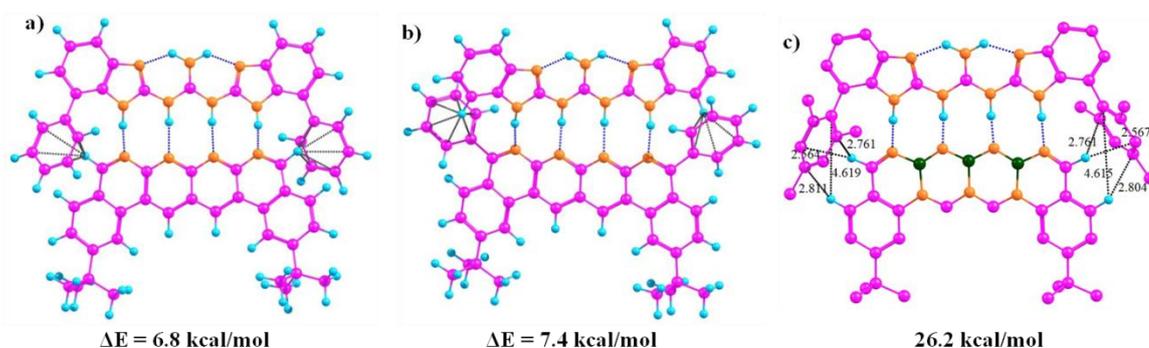


Figure 4.10 The optimized geometries of newly designed cationic AAAA-DDDD hydrogen bonded complexes, where hydrogen atoms in frontier lines are replaced with the phenyl groups (a and b), and the optimized geometry of hypothetical best case designed cationic AAAA-DDDD hydrogen bonded complex (9d), where hydrogen atoms in frontier lines are replaced with the 2,4,6-trimethylphenyl groups and three C-C bonds in the middle region of the acceptor partner is replaced with the isoelectronic B-N bonds. Pink, cyan, brown and green colors represent carbon, hydrogen, nitrogen and boron atoms respectively, whereas dotted blue lines represent hydrogen bonds and dotted black lines represent CH- π interactions. ΔE is the increase in binding energy over the parental AAAA-DDDD cationic complexes ($E_b = -43.4$ kcal/mol), from where they are derived.

We note that there might be individual cases in the considered families where the electrostatic force would be seen to not correlate with the binding energy. A reason for such a deviation is the selection of the common “line of direction” (line joining the center of geometries of the frontier atoms in each complex) for all the complexes of a family. The best line of direction may vary slightly for specific complexes in the family, thus leading to the possibility of having two complexes with similar binding energies but slightly different electrostatic forces of binding. However, for a given family of complexes of sufficient sample size, the electrostatic force of binding will fairly represent the binding strength of the electrostatic dominated noncovalent bonds (as suggested by the linear nature of the correlation plot between EFs and E_b), and thus, it can be exploited for the design of new, superior systems by tuning the strength of the noncovalent bonds.

(ii) **The Contact Ion-Pair Case.** The second family of structures that we have considered is that of the contact ion-pair catalysts employed in homogeneous olefin polymerization. This is an important area of research, beginning from the pioneering work done by Kaminsky *et al.* in the 1980s.⁵⁴ The active catalyst in these systems is the cationic species. One of the major foci of investigation in these systems has been to make the counterion as weakly coordinating as possible (several hundreds of papers have been produced in this pursuit). This is because the reduced interaction with the cation would allow the counterion to be displaced easily when the olefin substrate approaches the cationic metal center. A large variety of counterions have been employed over the years, with one of the most successful being the “BARF anion”: $B(C_6F_5)_4^-$.³¹ More recent counterions that achieve weaker interaction with the cation even in comparison to $B(C_6F_5)_4^-$ are the dinuclear counterions that have been proposed by Bochmann and coworkers.³⁴ They will henceforth be referred to as the “Bochmann anions”. We decided to investigate the contact ion-pair catalyst systems by focusing on how one could rationally design better counterions, i.e., ones that would interact more weakly with the counterions even in comparison to the Bochmann anions. The zirconocene cation Cp_2ZrMe^+ , was employed as the model cation in these studies, and the total electrostatic interaction of this cation has been considered for different contact ion-pair cases, with a range of

different counterions considered. Figure 4.11 below shows the optimized ion-pair complexes of the zirconocene cation with eight different counteranions. The zirconocene complex with the best of the three Bochmann anions has been shown in Figure 4.12.

Since the two binding partners of the ion-pair complex are charged, the dominant noncovalent interactions would be likely to be electrostatic in nature, which makes this family of complexes appropriate for investigation with our approach. It is also to be noted that the sum total of the long range secondary EF in the complexes studied was found to be approximately half that of the primary EF, which is the same ratio that had been found for the planar hydrogen bonded structures studied in (i).

For the purpose of analyzing the interaction by our approach, the line of direction chosen was the line connecting the zirconium atom in the cation and the central atom in the anion. The code being implemented to calculate EFs in this case is provided in Figure 4.13 below. The results of our calculations (with Turbomole 6.4, at the COSMO(CHCl₃)PBE/TZVP level of theory) are shown in Figure 4.14 below. Like for the planar hydrogen bonded complexes, we observed a high amount of correlation between the attractive EF between the two ions, and the binding energy. This was true for both the Mulliken ($r = 0.88$, Fig 14a) and the NBO ($r = 0.80$, Fig 14b) charge analyses. This result further showcases the viability of our approach. Also, as for case (i), we were interested in exploiting our approach for designing new anions that would serve as better counterions than the state of the art, thereby improving the efficiency of the homogenous olefin polymerization systems. Now, however, unlike in case (i), the focus was upon *reducing* the binding between the two interacting partners. For doing so, we analyzed the nature of the interaction between the two ions for the case where the binding energy had been seen to be the weakest: for the [Cp₂ZrMe-N{CNB(C₆F₅)₃}₂] case – the structure shown in Figure 4.12b. For this case, the analysis revealed that the -C₆F₅ ring that is in direct contact with the cation had the greatest contribution to the EF (-90.5 pN), followed by the central N of the anion (-56.4 pN) (see Table 4.1). Based on this information, it became clear that increasing the distance of the zirconium center from the

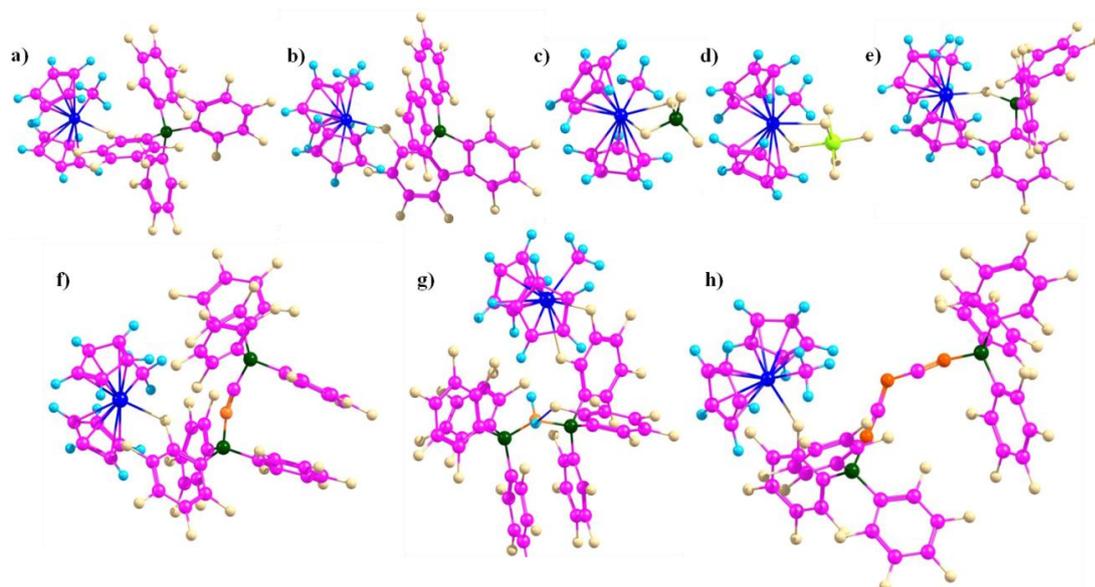


Figure 4.11 The optimized geometries of ion-pair complexes at the COSMO(CHCl_3)/PBE/TZVP level of theory using Turbomole 6.4. Pink, cyan, brown, green, blue, lime and white colors represent carbon, hydrogen, nitrogen, boron, zirconium, phosphorous and fluorine atoms respectively, whereas dotted blue lines represent hydrogen bonds.

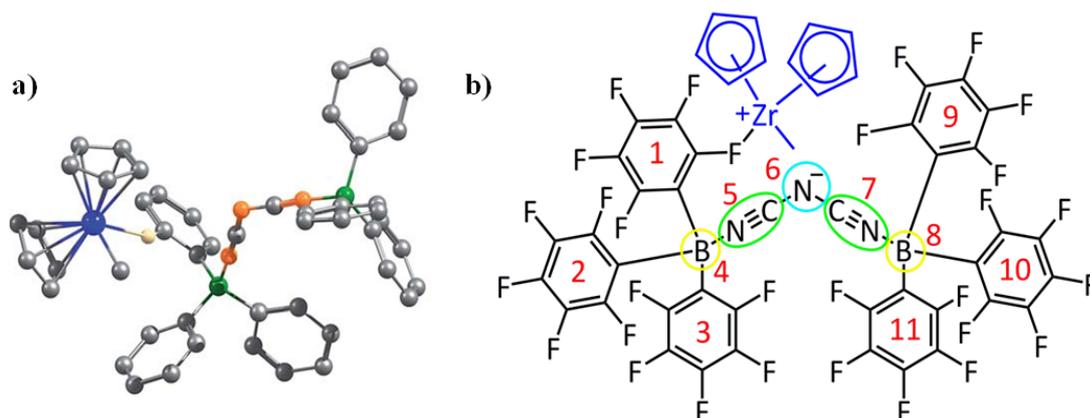


Figure 4.12 a) The optimized geometry of a non-coordinating dinuclear anion and the cationic zirconocene complex; black, orange, green, blue, and white represent carbon, nitrogen, boron, zirconium and fluorine atoms respectively. All the hydrogens of the zirconocene and the fluorides of the anion other than the one involved in coordination with the cation have been deleted for clarity. b) A schematic picture of the non-coordinating dinuclear anion and the cationic zirconocene complex illustrating different regions of the anion considered in the force analysis.

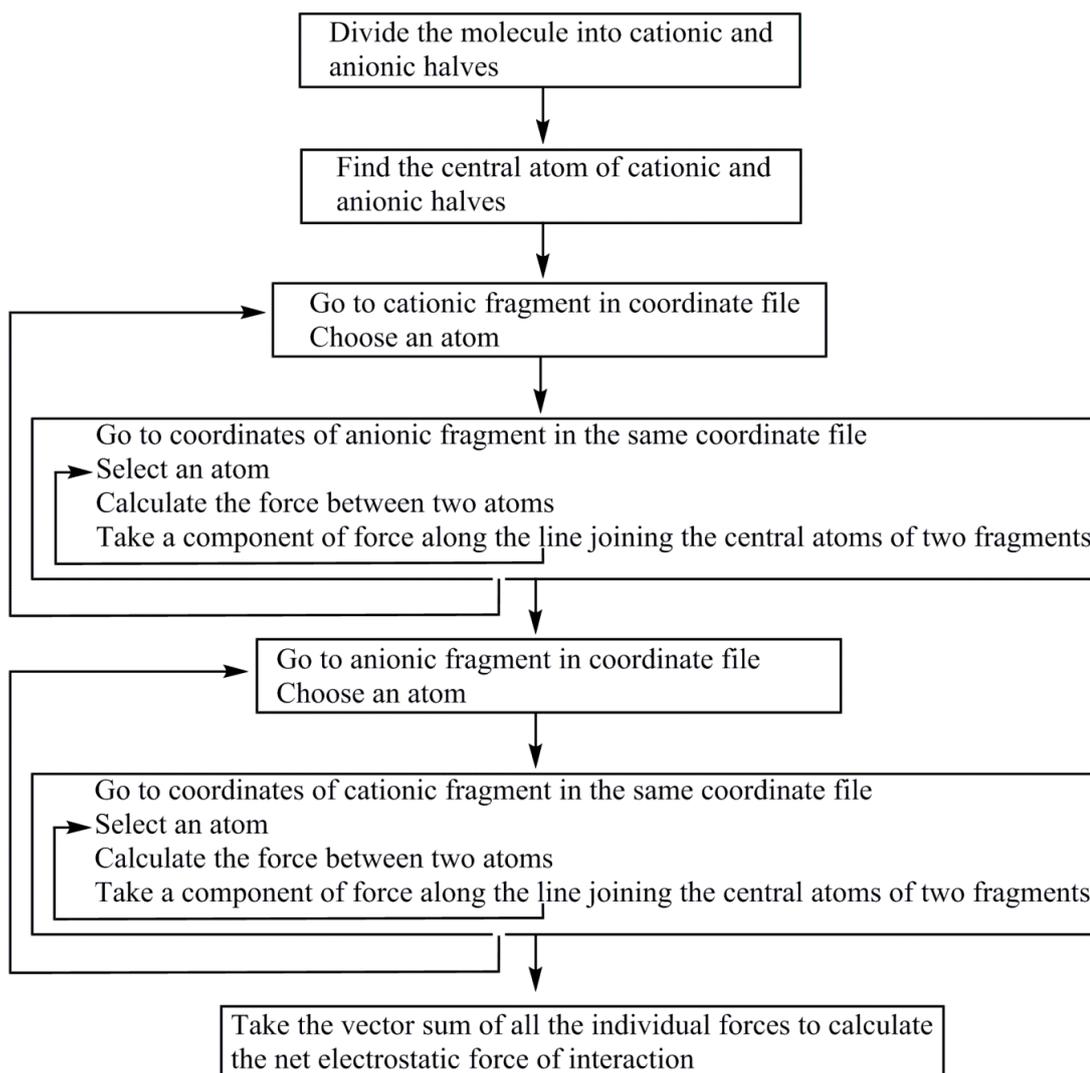


Figure 4.13 A flow chart of the Fortran 90 code used for calculating the net force between two partners in the contact ion-pairs case.

central nitrogen atom in the Figure 4.12b structure and/or decreasing the electronic density from the ring directly associated to the cation would lead to a decrease in the EF between the two species. We therefore propose anion **1** shown in Scheme 4.3, where the fluorides at the *ortho* and *para* positions of the phenyl ring of the counterion have been replaced by CF₃ groups. The distance between the zirconium and nitrogen atoms in the new complex is seen to have increased: from 5.008 Å in the Figure 4.9b structure to 6.538 Å. The EF was seen to be decreased (from -94.1 pN to -88.5 pN for the Mulliken charge analysis case, and from -185.7 pN to -137.1 pN with the NBO charge analysis) and the binding energy was seen to have reduced from -30.7 kcal/mol to -28.7 kcal/mol, thus indicating that the newly proposed anion would

be better than the best counterion for homogeneous olefin polymerization systems. In a further attempt, in order to reduce the attractive interaction of the phenyl ring of the anion that is connected to the cation, we replaced all C_6F_5 groups of Bochmann's best anion with the $C_4(CF_3)_4N^-$ group (Structure 2 in Scheme 4.3), which led to even weaker binding: of -23.8 kcal/mol, which is as much as 6.9 kcal weaker than the best Bochmann anion considered in this study. The computed association constant of this ion-pair system would be 2.2×10^6 times less than the state-of-the-art for this family of complexes, again illustrating the power of the present method to significantly improve upon existing systems. An electrostatic potential map of this anion shows weaker negative potential on its exposed outer surface than Bochmann's best anion, which further corroborates our results (Figure 4.15).

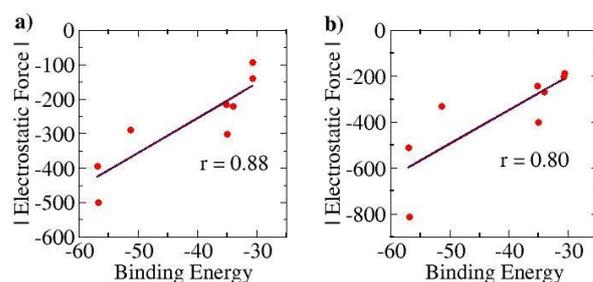
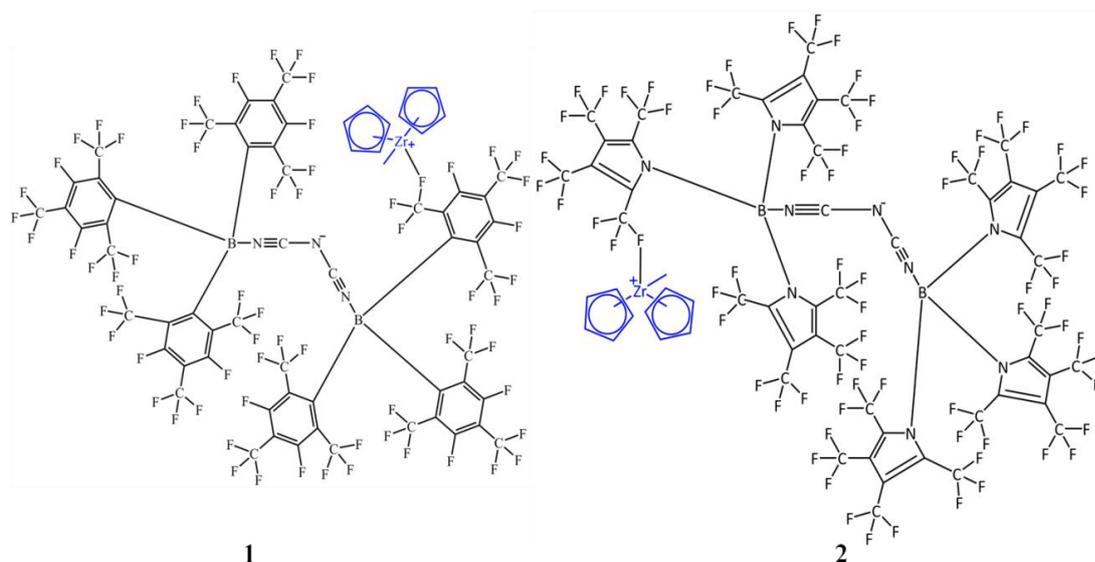


Figure 4.14 The Pearson correlation graph for the olefin polymerization catalyst: a) EF vs. E_b for Mulliken charges and b) EF vs. E_b for NBO charges.



Scheme 4.3 The schematic picture of the zirconocene complexes of the newly designed weakly coordinating anions.

Table 4.1 Electrostatic Forces analysis for the zirconocene complex with the best Bochmann's anion based on the Mulliken charge analysis for the geometries obtained at the COSMO/PBE/TZVP level of theory using Turbomole 6.4. All the Forces are in pN.

Region	1	2	3	4	5	6	7	8	9	10	11
Force	-90.5	-2.1	-9.6	22.6	38.4	-56.4	31.5	6.0	-6.9	-14.9	-12.2

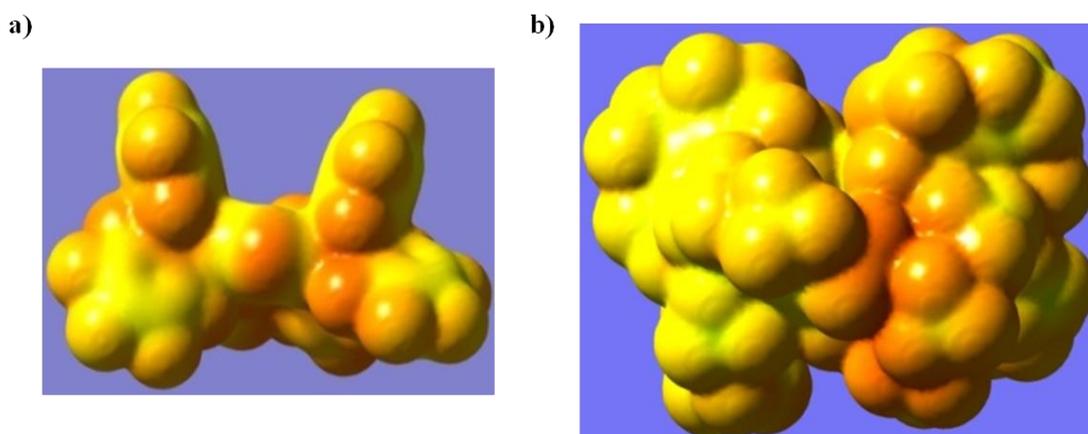


Figure 4.15 The molecular surface electrostatic potential in Hartrees, computed on the 0.0004 au contour of the electron density, using the GaussView software at the CPCM(CHCl₃)/M062X/6-31G** of theory with the geometries optimized using G09: a) the best of Bochmann's anions and b) the newly designed anion **2** in Scheme 4.3. In this scale, the blue colour indicates positive electrostatic potential and the red color indicates negative electrostatic potential, whereas the intermediate colours indicate intermediate electrostatic potential values. The colour scale was kept uniform in all the cases.

4.4 Conclusions

The current work employs a simple method based on evaluating the electrostatic force (EF) of interaction between all atoms of one binding partner with all the atoms of the other in molecular complexes where the noncovalent electrostatic interaction is the dominant factor. Excellent correlation is seen between the electrostatic force of binding and the binding energy of two partners within a family of molecular complexes when the interactions between

corresponding partners are dominated by electrostatics and the major varying factor in different complexes is atomic charges and not the distances between the atoms. The significance of the work lies in the fact that such an approach provides insight into the nature of bonding in the different systems studied, and can be exploited to design new systems with significantly increased or decreased binding, as desired, in comparison to the state-of-the-art. Multifold increase and decrease in binding energies has been obtained by altering distant atoms in the noncovalently bonded partners for two distinctly different families of complexes.

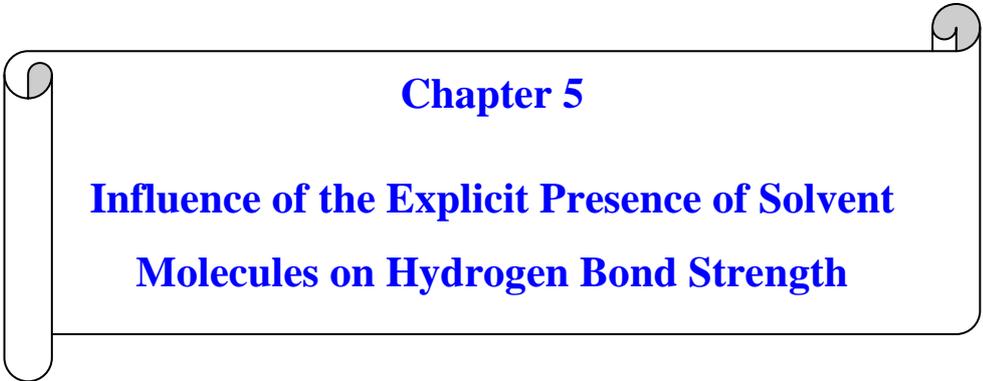
4.5 References

1. T. Šmejkal, B. Breit, *Angewandte Chemie*, 2008, **120**, 4010-4013.
2. A. G. Doyle, E. N. Jacobsen, *Chemical reviews*, 2007, **107**, 5713-5743.
3. R. R. Knowles, E. N. Jacobsen, *Proceedings of the National Academy of Sciences*, 2010, **107**, 20678-20685.
4. E. A. C. Davie, S. M. Mennen, Y. Xu, S. J. Miller, *Chemical reviews*, 2007, **107**, 5759-5812.
5. J.-M. Lehn, *Chemical Society Reviews*, 2007, **36**, 151-160; S. K. Yang, S. C. Zimmerman, *Isr. J. Chem.* 2013, **53**, 511 – 520.
6. L. Brunsveld, B. Folmer, E. Meijer, R. Sijbesma, *Chemical Reviews*, 2011, **101**, 4071-4098.
7. K. Biradha, *CrystEngComm*, 2003, **5**, 374-384.
8. C. B. Aakeröy, K. R. Seddon, *Chemical Society Review,s* 1993, **22**, 397-407.
9. A. Bauzá, T. J. Mooibroek, A. Frontera, *CrystEngComm*, 2016, **18**, 10-23.
10. R. G. Gonnade, M. S. Shashidhar, M. M. Bhadbhade, *Journal of the Indian Institute of Science*, 2012, **87**, 149.
11. B. Sellergren, M. Lepistöe, K. Mosbach, *J. Am. Chem. Soc.*, 1988, **110**, 5853-5860.
12. T. Rossow, S. Seiffert, *Supramolecular Polymer Networks and Gels, Vol. 268* (Eds.: S. Seiffert), Springer, 2015, pp. 1-46.
13. D. J. Hill, M. J. Mio, R. B. Prince, T. S. Hughes, J. S. Moore, *Chemical reviews*, 2001, **101**, 3893-4012.

14. D.-W. Zhang, X. Zhao, J.-L. Hou, Z.-T. Li, *Chemical Reviews*, 2012, **112**, 5271-5316.
15. D. Wang, G. Tong, R. Dong, Y. Zhou, J. Shen, X. Zhu, *Chemical Communications*, 2014, **50**, 11994-12017.
16. H. Yang, B. Yuan, X. Zhang, O. A. Scherman, *Accounts of chemical research*, 2014, **47**, 2106-2115.
17. F. Zhao, M. L. Ma, B. Xu, *Chemical Society Reviews*, 2009, **38**, 883-891.
18. B. R. Beno, K. S. Yeung, M. D. Bartberger, L. D. Pennington, N. A. Meanwell, *Journal of medicinal chemistry*, 2015, **58**, 4383-4438.
19. W. L. Jorgensen, J. Pranata, *J. Am. Chem. Soc.*, 1990, **112**, 2008-2010; J. Pranata, S. G. Wierschke, W. L. Jorgensen, *J. Am. Chem. Soc.*, 1991, **113**, 2810-2819.
20. P. L. Popelier, L. Joubert, *J. Am. Chem. Soc.*, 2002, **124**, 8725-8729.
21. T. Clark, P. Politzer, J. S. Murray, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2015, **5**, 169-177.
22. B. A. Blight, C. A. Hunter, D. A. Leigh, D. A., H. McNab, P. I. Thomson, *Nature Chemistry*, 2011, **3**, 244-248.
23. T. J. Murray, S. C. Zimmerman, *J. Am. Chem. Soc.*, 1992, **114**, 4010-4011.
24. B. A. Blight, C.-C. Amaya, S. Djurdjevic, M. Kaller, D. A. Leigh, F. M. McMillan, H. McNab, A. M. Slawin. *J. Am. Chem. Soc.*, 2009, **131**, 14116-14122.
25. S. Djurdjevic, D. A. Leigh, H. McNab, S. Parsons, G. Teobaldi, F. Zerbetto, *J. Am. Chem. Soc.*, 2007, **129**, 476-477.
26. D. A. Leigh, C. C. Robertson, A. M. Slawin, P. I. Thomson, *J. Am. Chem. Soc.*, 2013, **135**, 9939-9943.
27. J. Sartorius, H. J. Schneider, *Chemistry-A European Journal*, 1996, **2**, 1446-1452.
28. F. H. Beijer, H. Kooijman, A. L. Spek, R. P. Sijbesma, E. Meijer, *Angew. Chem. Int. Ed.*, 1998, **37**, 75-78.
29. J. R. Quinn, S. C. Zimmerman, J. E. D. Bene, I. Shavitt, *J. Am. Chem. Soc.*, 2007, **129**, 934-941.

30. H. Zeng, R. S. Miller, R. A. Flowers, B. Gong, *J. Am. Chem. Soc.*, 2000, **122**, 2635-2644; R. R. Gardner, S. H. Gellman, *Tetrahedron*, 1997, **53**, 9881-9890; J. Yang, S. H. Gellman, *J. Am. Chem. Soc.*, 1998, **120**, 9090-9091; K. S. Jeong, T. Tjivikua, A. Muehldorf, G. Deslongchamp, M. Famulok, J. Jr. Rebek, *J. Am. Chem. Soc.*, 1991, **113**, 201-209.
31. E. Y.-X. Chen, T. J. Marks, *Chemical reviews*, 2000, **100**, 1391-1434.
32. A. Macchioni, *Chemical reviews*, 2005, **105**, 2039-2074.
33. M. Delferro, T. J. Marks, *Chemical reviews*, 2011, **111**, 2450-2485.
34. M. H. Hannant, J. A. Wright, S. J. Lancaster, D. L. Hughes, P. N. Horton, M. Bochmann, *Dalton Transactions*, 2006, **20**, 2415-2426.
35. Yoichiro K., Haruka I., Mitsumi N. & Motomu K. *Nature Chemistry*, 2015, **7**, 712-716.
36. R. Ahlrichs, *Chem. Phys. Lett.*, 1989, **162**, 165-169.
37. J. P. Perdew, *Physical Review Letters*, 1996, **77**, 3865.
38. A. Klamt, G. Schüürmann *J. Chem. Soc. Perkin Trans. 2*, 1993, **5**, 799–805.
39. K. Eichkorn, *Chemical Physics Letters*, 1995, **240**, 283-289.
40. M. Sierka, *J. Chem. Phys.*, 2003, **118**, 9136-9148.
41. J. Hepburn, G. Scoles, R. Penco. *Chem. Phys. Lett.*, 1975, **36**, 451–456; R. Ahlrichs, R. Penco, G. Scoles. *Chem. Phys.*, 1977, **19**, 119–130; S. Grimme, *J. Comput. Chem.*, 2004, **25**, 1463–1473; S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787–1799; S. Grimme; J. Antony; S. Ehrlich; H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
42. Zhao, Y.; Truhlar, D.G. *Theor. Chem. Acc.*, 2008, **120**, 215-241; Zhao, Y.; Truhlar, D.G. *J. Phys. Chem. A*, 2006, **110**, 13126-13130; Zhao, Y.; Truhlar, D.G. *J. Chem. Phys.*, 2006, **125**, 194101.
43. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J.

- Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, Gaussian09 Revision A.01, Gaussian Inc., Wallingford, CT, 2009.
44. V. Barone, M. Cossi, *J. Phys. Chem. A.*, 1998, **102**, 1995-2001; b) M. Cossi, N. Rega, G. Scalmani, V. Barone, *J. Comput. Chem.*, 2003, **24**, 669-681.
45. R. S. Mulliken, *J. Chem. Phys.*, 1955, **23**, 1833-1840.
46. A. E. Reed, R. B. Weinstock, F. Weinhold, *J. Chem. Phys.*, 1985, **83**, 735-46.
47. P. O. Löwdin, *J. Chem. Phys.*, 1955, **21**, 374-375.
48. C. M. Breneman, K. B. Wiberg, *J. Comp. Chem.*, 1990, **11**, 361-73.
49. Y. Hisamatsu, N. Shirai, S. I. Ikeda, K. Odashima, *Organic letters*, 2010, **12**, 1776-1779.
50. T. Berlin, *J. Chem Phys.*, 1951, **19**, 208-213.
51. R.F. W. Bader, W.H. Henneker, P. E. Cade, *J. Chem Phys.*, 1967, **46**, 3341-3363; R. F. W. Bader, *J. Am. Chem. Soc.*, 1964, **86**, 5070-5075; R.F.W. Bader, H.J.T. Preston, *Can. J. Chem.*, 1966, **44**, 1131-1145.
52. P. G. Campbell, A. J. V. Marwitz, S.-Y. Liu, *Angew. Chem. Int. Ed.*, 2012, **51**, 6074-6092; A. M. Pejlovas, A. M. Daly, A. J. Ashe, S. G. Kukulich, *J. Chem. Phys.*, 2016, **144**, 114303; A. Mazzanti, E. Mercanti, M. Mancinelli. "Axial Chirality about Boron–Carbon Bond: Atropisomeric Azaborines." *Organic Letters*, 2016; X. Wang, F. Zhang, J. Gao, Y. Fu, W. Zhao, R. Tang, W. Zhang, X. Zhuang, X. Feng, *The Journal of organic chemistry*, 2015, **80**, 10127-10133.
53. M.J. Plevin, D. L. Bryce, J. Boisbouvier, *Nat. Chem.*, 2010, **2**, 466-471.
54. H. Sinn, W. Kaminsky, H.-J. Vollmer, D.-C. R. Woldt, *Angew. Chem. Int. Ed. in English*, 1980, **19**, 390-392.



Chapter 5

Influence of the Explicit Presence of Solvent Molecules on Hydrogen Bond Strength

Chapter 5

Influence of the Explicit Presence of Solvent Molecules on Hydrogen Bond Strength

Abstract: Long range electrostatic interactions between atoms of hydrogen bonded partners have been found to be significant in determining the strength of electrostatic dominated noncovalent interactions. Similarly, the hydrogen bonded systems that exist in the solvent environment are prone to get affected by the long range influence of explicit solvent molecules. Despite the ubiquitous presence of solvent molecules around most of the noncovalently bonded systems in biology and chemistry, this fact has not earned adequate attention in the literature. Moreover, estimating the long range influence of solvent molecules on such interactions employing conventional theoretical methods are not possible. However, the method that we have recently proposed, which describes calculating the electrostatic force of binding to estimate the strength of electrostatic dominated interactions, can effectively be employed to obtain a qualitative picture of the long range influence of explicit solvent molecules. In this chapter, we have reported an extensive theoretical and computational study that was intended to explore this phenomenon on three distinctly different types of hydrogen bonded model systems. We have seen in our study that the explicit solvent molecules indeed influence the strength of hydrogen bonds, and that this influence is expressed primarily in two ways: (a) through charge transfer and induction effects involving solvent molecules and (b) through the long electrostatic range influence of solvent molecules.

5.1 Introduction

Noncovalent interactions are at the center of biological processes. Many chemical and physical phenomena of high relevance in chemistry and material science are also governed by weak interactions.¹ Recently, in target-oriented synthesis and catalysis, noncovalent interactions have been exploited as tools to accomplish specific targets.^{1,2} However, the influence of these interactions and their dependence on external factors has not been understood completely. To fully unravel and exploit the potential of

these weak interactions, it is necessary to understand their effects and efficacy in greater detail. Noncovalent interactions that have attracted maximum attention so far are those that are dominated by electrostatic contributions, for example, for the case of hydrogen bonding.^{1,2} The fact that has earned significant attention regarding electrostatic dominated noncovalent interactions in recent times is this - their influence is long range, and their strengths are influenced considerably by long range electrostatic interactions between remotely placed atoms.³ For instance, the long range secondary electrostatic interactions between the remote atoms associated with hydrogen bonded partners have been shown to have a distinct and deterministic influence on the association constant of multi-hydrogen bonded systems. This signifies that the structure, orientation, and position of even distant atoms or substituents on the hydrogen bonding partners will have a decisive impact on hydrogen bond strength.^{3,4} Intriguingly, hydrogen bonded systems largely exist surrounded by a quantum of solvent molecules inside the synthetic or cellular environment. The explicit presence of solvent molecules, therefore, may potentially alter the strength of hydrogen bonds through the long range electrostatic interactions. Orientation, position, structure, and the number of solvent molecules that surround the hydrogen bonded system within a solvation shell or even beyond, all, have the potential to influence the strength of any interaction where electrostatic contributions are predominant. Since electrostatic interactions are directional, the long range electrostatic influence due to solvent molecules may also alter the geometry of such systems. Current literature reports show that the solvent indeed affects the geometries and the stability of noncovalent bonded systems.⁵⁻⁹ However, it does not classify any study, to the best of our knowledge, which precisely pinpoints the effect of explicit solvent molecules on the stability, strength or geometry of noncovalent bonds. Moreover, existing literature elucidates the following properties of solvents that influence the H-bond strength: (a) the dielectric constant of the solvent or the medium,⁶ (b) the competitive role of hydrogen bonding solvents,^{5,8} which competes for hydrogen bonding with the solvent-exposed hydrogen bonding units of solutes, and (c) the local arrangement of solvent dipoles near the hydrogen bonded units of solutes.⁹ The effect of solvent molecules on the geometry of noncovalent bonded solutes such as proteins has also been understood through hydrophobic or solute-solvent interactions.⁵ However, the hydrogen bonds that are less exposed to the

solvent or that are present at the interior sites in the chemical architecture of solutes (particularly in large molecules such as proteins, DNAs, and polysaccharides) feel the weaker dielectric effect. The dipolar arrangement near hydrogen bonds and the contest between solvent molecules and solute for making hydrogen bonds at such sites is also unlikely. On the other hand, atoms within solvent molecules acquire certain net partial charges, which can interact electrostatically with the charged atoms of hydrogen bonded partners even from longer distances in the given region of influence, and hence can alter the strength of a given hydrogen bond. Therefore, consideration of only the above effects is not adequate to quantify the complete role of solvent on the strength and geometry of hydrogen bonds.

Considering the ubiquity and wide applicability of electrostatic dominated noncovalent interactions and their practical dependence on the long range electrostatic effect of solvent molecules, it is of great practical significance to fundamentally understand the influence of explicit solvent molecules on the strength and geometry of noncovalent bonded solutes. However, conventional theoretical methods are not eligible to be applied directly to quantify such effects. The method that we have recently proposed that is based on electrostatic force calculations to represent the binding strength may be employed in this case to understand the long range effect of solvent molecules on the strength of electrostatic dominated interactions.⁴ We have employed this method to study the influence and impact of solvent molecules on the hydrogen bond strength of three inherently different kinds of hydrogen bonded systems in this study. They are: (i) a small model peptide synthesized by Paramjit Arora and coworkers,¹⁰ (ii) a 50mer water cluster studied extensively by Dieter Cremer and coworkers¹¹ and (iii) the cation-water shell in water.¹² These examples from three different areas have been chosen specifically to illustrate the generality of the result. The electrostatic force of binding for a chosen hydrogen bond of all systems has been calculated along the line joining the hydrogen bond acceptor and donor atoms (hydrogen bonding line) employing Coulomb's equation for force, with the atoms being considered as point charges. The charges on the atoms have been determined from quantum chemical calculations. Two slightly modified approaches (with two slightly different algorithm and codes) has been employed to compute the binding forces in case (i), and for cases (ii) and (iii). These approaches have been

discussed further in the Computational Details section below. A flow chart of Fortran codes that have been implemented is provided in Figures 5.1 and 5.2.

5.2 Computational Details

All the DFT calculations were carried out using the Turbomole 6.4 suite of quantum-chemical programs in the solvent phase employing the Conductor-like Screening Model (COSMO)¹³ using water (epsilon = 78.5) as the solvent.¹⁴ Geometry optimizations were performed employing the PBE functional.¹⁵ The electronic configuration of the atoms was described by a triple-zeta basis set augmented by a polarization function (TURBOMOLE basis set TZVP). The resolution of identity (RI),¹⁶ and the multipole accelerated resolution of identity (MARI-J)¹⁷ approximations were employed for an accurate and efficient treatment of the electronic Coulomb term in the density functional calculations. All the calculations were performed with DFT-D2, a general empirical dispersion correction proposed by Stefan Grimme for density functional calculations.¹⁸ Only electronic energies were considered for energy analysis of all the systems.

The Mulliken,¹⁹ NBO²⁰ and Löwdin²¹ charges have been used to calculate the electrostatic forces of binding for each hydrogen bond of every system. The forces were calculated along the line joining the donor and acceptor atoms of a particular hydrogen bond. To compute the magnitude of the net force of binding, i.e., the binding force, the vector sum of the electrostatic forces experienced by each fragment along the aforementioned line has been considered. A flow chart revealing the algorithm of the code that was employed for computing the forces have been shown in Figures 5.1 and 5.2. Two slightly modified versions of the original code⁴ have been used to compute the binding forces. For the peptidic systems, case (i), where there is no clear-cut demarcation of hydrogen bonding partners, the entire molecular space corresponding to each hydrogen bond has been divided into three regions to obtain the force of binding: acceptor, donor and the middle region. Either side of two parallel planes (donor and acceptor planes) that are drawn perpendicular to the hydrogen bonding line and pass through the donor and acceptor atoms respectively constitute these regions (see Figure 5.3). The region that is between the donor and acceptor planes is defined as a middle region. The region that is on the other side of the acceptor plane (that is away from the donor plane) is termed as the acceptor region.

Similarly, the region that is on the other side of the donor plane, and that is away from the acceptor plane is considered as the donor region (please see Figure 5.3). All the pair-wise interactions between atoms of donor and acceptor regions and the atoms of these two regions with the atoms of the middle region have been considered in this case in order to compute the binding force for a chosen hydrogen bond. The same procedure had been repeated to compute the bond strength of each of the hydrogen bonds for (i) (see Figure 5.1). For the cases (ii) and (iii) where there is the clear-cut demarcation of the hydrogen bonding partners, all the pair-wise atomic electrostatic interactions between two partners have been considered to obtain the electrostatic force of binding for any chosen hydrogen bond (see Figure 5.2). The effect of explicit solvent molecules on hydrogen bond strength has been computed as the vector sum of the component of individual pair-wise forces obtained along the hydrogen bonding line when all the pair-wise atomic interactions between all atoms of solvent molecules and solute have been considered (Figures 5.1 and 5.2).

Recent literature reveals that the polarization and the donor-acceptor charge transfer terms that contribute significantly to noncovalent interactions are electrostatic in nature.²² The Feynman interpretation describes even dispersion interaction to be of electrostatic origin.²³ This suggests that the proper treatment of electrostatic interactions is adequate for a complete description of all kinds of noncovalent interactions. In fact, these effects are being captured approximately in population analysis schemes. The Born-Oppenheimer approximation reveals molecules as a collection of point charge nuclei and electron density distributed over the molecular space. To distinguish atoms and bonds in a molecule in point charge analyses under the Born Oppenheimer approximation, the electron density is divided based on atomic orbitals. When a noncovalent bonding unit or a partner is kept in the vicinity of another molecule or partner, the induction causes a change in the local electron density of all the atoms close by. This gets reflected in the point charges of atoms as a modification in the charges when the charges calculated for molecules or partner in the vicinity of another molecule or partner and the charges calculated keeping these molecules or partners at infinite separation are compared. This kind of analysis has also been done earlier by other research groups to demonstrate the effect of polarization and donor-acceptor contributions.²⁴ These partitioning schemes of molecules into atoms (different population analyses schemes) have also been able to

provide a qualitative estimation of the amount of charge that is being transferred from one partner to another during the formation of a noncovalent bond.²⁴ The amount of the charge getting transferred may differ depending on the scale of the particular charge analysis method that is being chosen at a time. Similarly, a comparison of atomic charges for a molecule in absence and presence of explicit solvent molecules will provide a qualitative estimate of the charge getting transferred from solvent molecules to the solute or *vice versa*. The magnitude or direction of transferred charges may depend on the partitioning scheme in this case as well. It is worth to note here that the charge transfer effect and induction effect involving solvent molecules will eventually get reflected as a difference in the magnitude of the electrostatic force of binding for a hydrogen bonding unit of a solute, when forces are calculated from the charges derived in presence and absence of explicit solvent molecules. The caution that must be taken in employing the point charge approach that we have described here, particularly in σ -hole bonded complexes,²⁵ is that the atoms or molecules that are under investigation must not possess σ -holes in them. Such atoms can have regions of both positive and negative electrostatic potential on their surfaces, and assigning a single atomic charge to such atoms may lead to erroneous results. Recently, strategies have been devised to deal with such systems more accurately. Such atoms are proposed to be given one extra point charge to the positive region (σ -hole) so that the net formal charge and the electrostatic potential assigned to the atom remain the same.^{26,27} However, σ -holes are found to be manifested by only heavier atoms of groups IV-VII when these atoms are linked covalently with a highly electronegative group.^{26,28} The systems that we have studied here do not possess such elements or groups bonded to them, and hence no extra effort has been required to treat σ -holes for these cases.

It is to be noted that interactions involving the hydrogen bond acceptors and donors that are directly involved in hydrogen bonding with each other are called primary interactions. The electrostatic interactions between any pairs of atoms between two hydrogen bonded partners have been commonly referred to as long range electrostatic interactions. Therefore, the long range interactions involve all primary and secondary electrostatic interactions together.

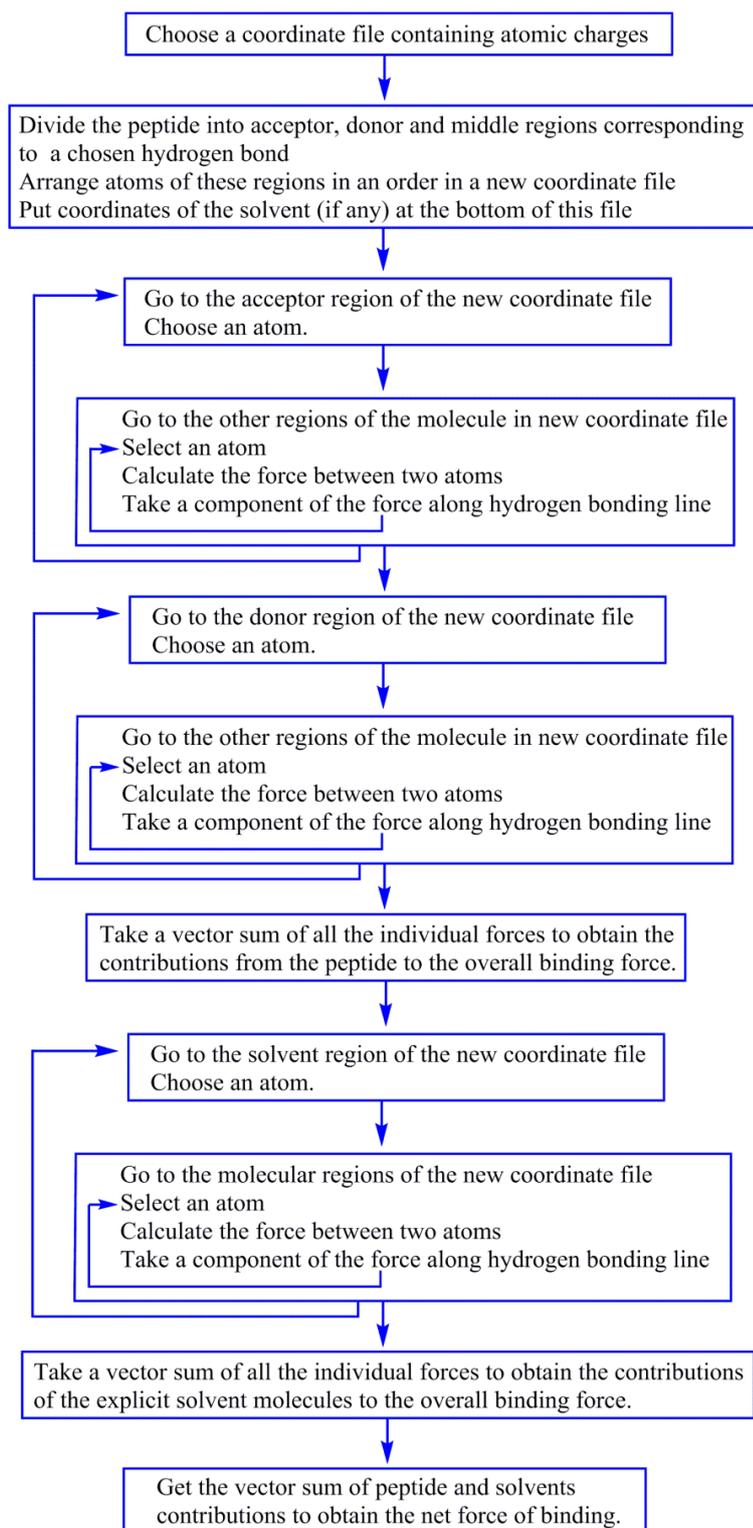


Figure 5.1 The flowchart of the Fortran 90 code used for calculating the electrostatic force of binding for a given hydrogen bond in a peptide in the presence of explicit solvent molecules.

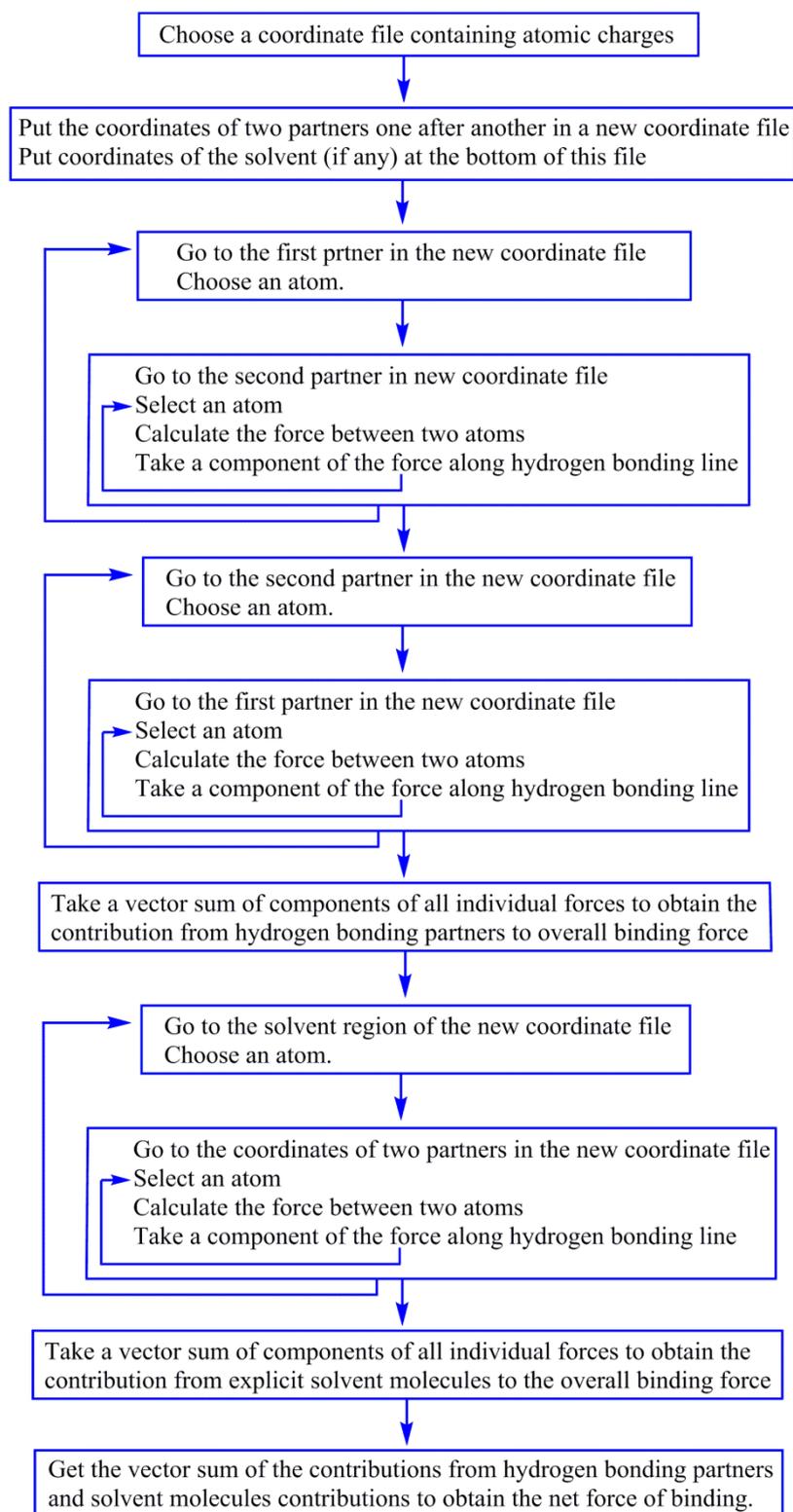


Figure 5.2 The flowchart of the Fortran 90 code used for calculating the electrostatic force of binding for a given H-bond between two partners in the presence of explicit solvent molecules.

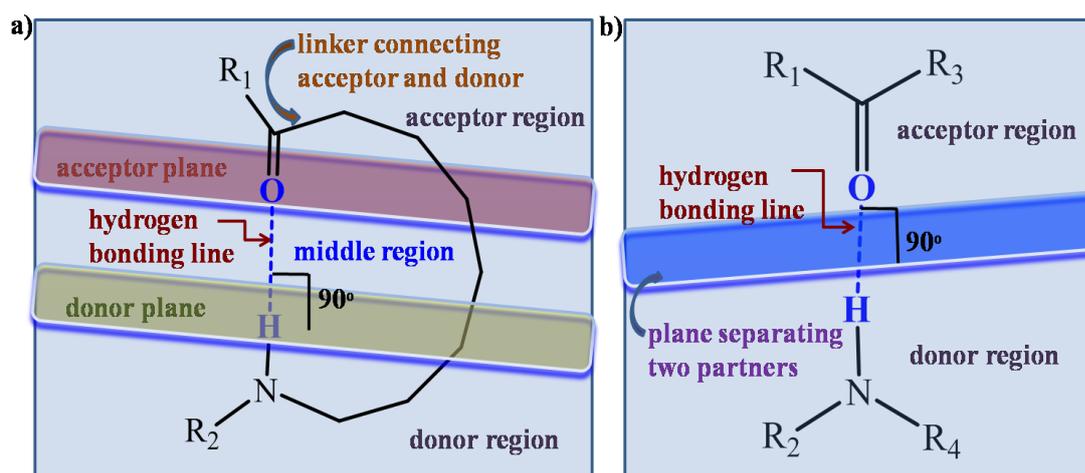


Figure 5.3 Schematic representation of different regions of hydrogen bonds: (a) in cases where a clear-cut demarcation of hydrogen bond acceptors and donors partners are not possible, and (b) in the cases where the demarcation of acceptor and donor partners is clear. The entire space in (a) is divided into three regions by two parallel planes that are drawn perpendicular to the hydrogen bonding line and passes through acceptor and donor atoms. The hydrogen bonding space in (b) is divided into two regions by a plane that passes *via* any point between donor and acceptor atoms and is perpendicular to the hydrogen bonding line.

5.3 Results and Discussion

In this theoretical and computational study, we intended to investigate the influence of long range secondary electrostatic influence due to explicit solvent molecules on the stability of hydrogen bonds. In this study, these effects have been examined in terms of the electrostatic force exerted by solvent molecules on the hydrogen bonded unit of solutes. In order to determine the electrostatic force (EF) of binding of a given hydrogen bond, the approach that has been adopted has been to determine the Coulombic force between each pair of atoms, with the atoms being chosen from the different regions or partners (as described in Computational details section), and then summing them up the force values vectorially. Since force has direction, the electrostatic interaction has been considered along the line that describes a hydrogen bond most appropriately: the line joining acceptor and donor atoms of a chosen hydrogen bond (hydrogen bonding line) for any given system (see Fig. 3). Therefore, only the component of these pair-wise forces along that given line has been

considered to obtain the net force of binding. The influence of solvent molecules on hydrogen bond strength has been determined as the net electrostatic force exerted together by all atoms comprising solvent molecules on every atom of the hydrogen bonded system in a given geometry, while the component of pair-wise forces have been taken along the aforementioned line to get the net force. As mentioned in the previous sections, the point charge approximation has been employed for this purpose, and each atom has been considered to have a charge, determined by the NBO, the Mulliken and the Löwdin population analyses.

(i) The Small Peptide Case: protein secondary structure motifs are basically composed of α -helices and β -sheets. They are structural blocks of three dimensional tertiary structures of functional proteins and the basic structure of many structural proteins. These protein motifs are distinguished based on differential hydrogen bonding patterns involving the peptide backbone, which is specific to each of these types. In a cellular environment, these structures (in soluble proteins) are surrounded by different shells of water molecules. Thus, understanding the effect of solvent molecules (water in this case) on the stability of protein secondary structure motifs is of fundamental and practical significance. To investigate this problem, we have chosen a representative hydrogen bond surrogate derived α -helix that has been synthesized by Arora and coworkers.¹⁰ The crystal structure of this dodecapeptide has been reported in the dimeric form with 15 water molecules intact. The reason behind choosing this specific example is the self-contained helical nature of this peptide. Usually, α -helices composed of 10 or less amino acid residues are not stable due to a low nucleation probability according to the helix-coil transition theory.²⁹ To overcome the nucleation barrier, Arora and coworkers have devised a strategy of using a preorganized α -turn at the N-terminus of the peptide by replacing the main chain hydrogen bond between the C=O of the i^{th} amino acid residue and the N-H of the $i+4^{\text{th}}$ amino acid with the C-C covalent bond.³⁰ This covalent linker is called as the hydrogen bond surrogate, which has been able to induce stability in helical conformations of small peptides. Therefore, to investigate the influence of solvent molecules on the hydrogen bond strength of proteins or peptidic structures, a self-contained helix was preferred over a small peptide model in our study. A geometry optimization of this dimeric helix in the presence (dimer-water complex) and absence

(bare dimer) of 15 water molecules (that have been reported with the crystal structure) have been carried out. The optimized geometry of the dimer in the presence and absence of 15 explicit water molecules has been shown below in Figure 5.4. A single point energy calculation of the dimer isolated from the optimized dimer-water complex reveals that the peptide dimer in the dimer-water complex is unstable as compared to the optimized bare dimer (*i.e.*, dimer in absence of 15 water molecules) by 31.6 kcal/mol. This could be because of the weakening of the main chain or interhelix hydrogen bonds, or geometrical or conformational changes that have occurred in the peptide during geometry optimization in the presence of explicit solvent molecules, which has brought the peptide into a different local minima. There could be two possible reasons for this: (i) the long range influences of solvent molecules (that might have altered the bond parameters in the peptide) and (ii) the mutual interaction between solvent molecules that have brought the peptide into the other conformation, which is inevitable in the presence of explicit solvent molecules, and can be observed easily by comparing the two structures. Comparisons of corresponding dihedral angles that belong to heavy atoms (except hydrogen atoms) in the bare dimer and isolated dimer from the dimer-water complex support this assertion in this case.

It is well understood that even a slight perturbation in bonding parameters (bond lengths or bond angles) due to external factors (*e.g.*, electric field, that may arise due to the long range electrostatic interactions of solvent) brings the molecule into the other local minimum, specific to this external effect. The single point energy of this geometry in the absence of the external factor will always be higher than the energy of the geometry that is obtained after optimization in the absence of the external factor, because a minimum in one condition does not generally belong to the minimum in another condition (see Table 5.1, for example). This effect can be observed in geometry optimizations from the exact crystal structure geometries. The geometry optimization from the crystal structure of any molecule at any level of theory in a neat condition (in vacuum or in implicit solvent) does not complete in a single step (the optimization has often been seen to occur with the fall in energy of the molecule on the potential energy surface), indicating that the experimental environment has not been exactly mimicked in the theoretical calculations. The changes in the bond parameters of bonds that do not comprise hydrogen atoms (since

hydrogen atoms are generally not traced in X-ray diffraction) may also be observed in optimization. All of this simply means that the two structures that are obtained from experiments and from geometry optimizations belong to two different minima under two different conditions. The conformational changes that arise from changes in the dihedral angles in a molecule, on the other hand, may or may not lead to a difference in the energy of the system, as two different conformational minima may or may not have the same energy. However, for peptides, it is difficult to distinguish and identify the influence of these two effects on the energy of the system.

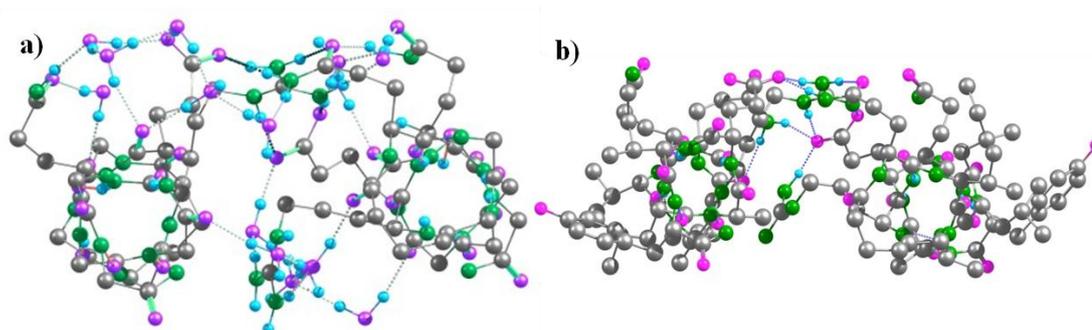


Figure 5.4 The optimized geometry of (a) dimer-water complex and (b) the bare peptide dimer at the COSMO(water)/PBE-D2/TZVP level of theory. Hydrogen atoms of the dimer that are not involved in hydrogen bonding have been deleted for clarity. Color coding: cyan - hydrogen, light black – carbon, pink – oxygen, green – nitrogen. Bonds with broken lines represent hydrogen bonds.

However, these effects can be studied separately in structures where only one of these effects is prevalent, like the water clusters. To elaborate this point we have explored the cluster of 15 water molecules that has been isolated from the optimized dimer-water complex. Figure 5.5 below shows the cluster of 15 water molecules isolated from the optimized dimer-water complex. Comparative data for structural changes that have occurred due to the influence of other neighboring molecules and their impact on the energy of each water molecule in the dimer-water complex has been provided in Table 5.1. The sum of single point energies of 15 water molecules isolated from the optimized dimer-water complex was found to be 6.0 kcal/mol higher than the sum of energies of the infinitely separated 15 water molecules (*i.e.*, 15 times the energy of an optimized water molecule). One can see from the structure of a water molecule that the conformational changes in the water molecules due to external

perturbation (for example, interactions with peptide dimer or neighboring water molecules in this case) is not possible, as water does not contain any dihedral angle. This indicates that the long range influence of peptide and the neighboring solvent molecules (if applicable) have distorted the geometry of each water molecule in such a way that the overall effect is their destabilization. The energy of destabilization of water molecules ranges between 0.0 to 0.7 kcal/mol, and sums up to a total of 6.0 kcal/mol (Table 5.1). The long range influence of one molecule on the bonding parameters of another can be understood in two ways: (i) through induction and (ii) through the charge transfer effect. Both of these effects cause restructuring of the molecular electron density, and hence affect the bonding and geometry of the molecule. The induction effect can further be understood as the influence of the electric field created by the former on the electronic distribution of the latter. In cases where the charge transfer effect is negligible, the long range influence of one molecule on the geometry of another can be approximated as the effect of the electric field (which would mostly be anisotropic because of the shape of molecules as well as the differential distribution of electron density over the molecular space) that is created by the former in which the latter is kept and *vice versa*. Since the magnitude of an anisotropic electric field in a given space depends on the distance as well as on the direction, different bonds in a molecule in the presence of another may locally feel different values of the electric field, which will depend upon the orientation and distance of the chosen bond from the neighboring molecule. When there are many such molecules (say, solvent molecules) present in the vicinity of one molecule (say solute), a bond (in the solute) may be assumed to be kept in the resultant field of electric fields created by all other molecules, including distant atoms of solute, which may take any value, including zero. Since covalent bonds are believed to be comprised of many charge-separated resonance contributors, a covalent species can be stabilized or destabilized *via* minor charge-separated resonance contributors in the presence of the external electric field.³¹ Therefore, depending upon the magnitude and direction, the resultant electric field created by neighboring molecules may increase or decrease the bond dissociation energy of a given bond. Different bonds in the molecule will experience different effects (Table 5.1, columns 2-4), which will eventually lead to additional stability or instability in the molecule. Since the covalent

bonds are stronger and rigid, structural perturbations in such bonds are very limited in a weak electric field.

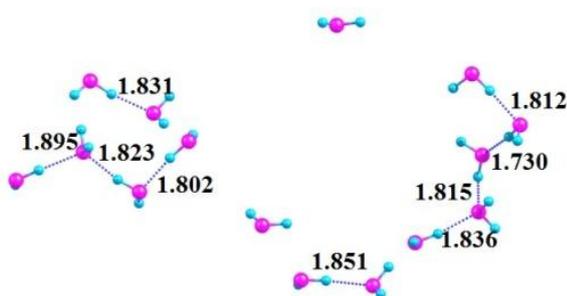


Figure 5.5 The geometry of an isolated cluster of 15 water molecules from the optimized dimer-water complex at the COSMO(water)/PBE-D2/TZVP level of theory. Color scheme: cyan - hydrogen, pink – oxygen. Bonds with broken lines represent hydrogen bonds.

However, in case of noncovalent bonds, even a small value of the electric field may alter the geometry drastically. Owing to the long range effect of electrostatic interactions, an electrostatic dominated noncovalent bond, X-H...Y, can also be thought to be composed of an assembly of many charged contributors, as shown in Figure 5.6. The multi-point electrostatic interactions between two partners can be uniquely represented as a two-point electrostatic interaction between a positive and a negative equivalent charge when these charges are kept in place of donor and the acceptor partners respectively at a separation so that the magnitude and direction of the electrostatic interaction remain unchanged. To keep these charges at equilibrium, the Pauli repulsion between two charges should also be the same as the actual interacting species. Thus a classical hydrogen bond of the form X-H...Y can be argued to be the result of the interaction between two representative charged species, $[X-H...Y \leftrightarrow H^{+q}p...Y^{-q}n]$ (see Figure 5.6). An electric field acting along this line ($H^{+q}p...Y^{-q}n$ line, the line joining donor and acceptor atoms) will affect the strength of this bond significantly. Fields that are applied perpendicular to this line are expected to affect the geometry of the system.

To evaluate the effect of the solvent, it can be assumed that this species inside the solvent is surrounded by a shell of solvent molecules. Since every atom of the solvent acquires a certain formal charge (that may be represented as a point charge) or

a set of charges (in cases where σ - and π -holes may exist) due to the differential electronegativity of covalently linked atoms that compose the solvent molecules, the collective effect of these charges may produce a local electrostatic field. The direction and magnitude of this field may vary in space depending upon the local chemical environment, and the structure and local arrangement of the solvent molecules. Consequently, the field created by the explicit presence of solvent molecules may alter the strength of the hydrogen bond and many other properties of the solutes such as the geometry, conformations, and hence the overall stabilization of the solute. It is to be noted that a solvent molecule within a solvent or in a solution can be also be considered as the solute to see the effect of other neighboring molecules on its geometry and energy. What must also be noted here is that the consideration of the electric field created by solvent molecules will not account for the entire effect of long range electrostatic interactions due to solvent molecules on the energy or stability of the hydrogen bonded solutes. The electric field mainly covers the induction effect of long range electrostatic interactions. However, the energy of a charged system also depends on the interactions with other charged centers, which are not covered in the consideration of the equivalent electric field for a given hydrogen bonding unit. However, the method that we have proposed based on electrostatic force calculations has an upper hand in covering all these effects in a qualitative manner to account for the influence of solvent molecules on the strength of a hydrogen bond in the solute.

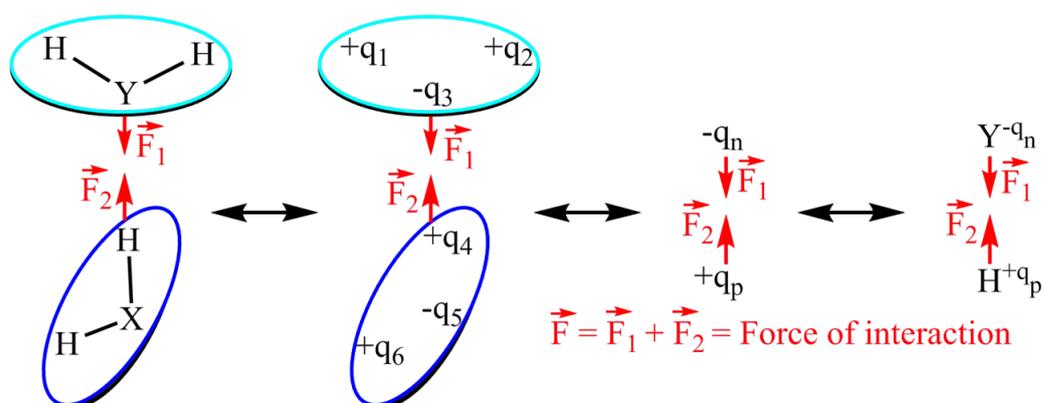


Figure 5.6 A schematic representation of the electrostatic dominated noncovalent interaction between two species H_2X and H_2Y , which is represented as the multipoint electrostatic interaction between atoms at complementary partners; each of them acquires a certain formal charge.

Table 5.1 Geometrical parameters of an isolated cluster of 15 water molecules from optimized dimer-water complex as well as the optimized water molecule at the COSMO(water)/PBE-D2/TZVP level of theory.

	S. N. o.	O-H ₁ bond length	O-H ₂ bond length	H-O-H bond angle	ΔE (kcal/mol)	amount of charge transferred w.r.t. single isolated water		
						NBO	Mulliken	Löwdin
Cluster of 15 water molecules isolated from optimized dimer-water complex	1	0.984	0.984	101.9	0.1	0.05518	0.09871	0.14071
	2	0.974	0.975	104.6	0.0	0.01577	0.03105	0.03785
	3	0.989	0.989	107.2	0.5	-0.01188	0.00063	-0.03859
	4	0.994	1.001	103.0	0.6	-0.03973	-0.05019	-0.09725
	5	0.981	0.990	105.9	0.3	0.00415	0.02242	0.03194
	6	1.009	0.973	105.4	0.7	-0.00247	0.02334	-0.00664
	7	0.981	0.986	105.5	0.2	-0.02594	-0.00715	-0.05012
	8	0.985	0.982	109.3	0.6	-0.04127	-0.0243	-0.09681
	9	0.998	0.991	103.6	0.5	-0.03291	-0.01161	-0.06368
	10	0.973	0.993	105.5	0.3	0.01303	0.00723	0.01004
	11	0.981	0.992	104.1	0.2	-0.0207	-0.02651	-0.05247
	12	0.974	1.000	105.6	0.5	-0.03624	-0.00125	-0.04692
	13	0.992	0.993	106.0	0.6	0.00578	0.02755	0.03667
	14	1.003	0.981	103.6	0.5	-0.01026	0.01979	-0.01581
	15	0.993	0.992	103.8	0.4	0.02559	0.05561	0.08215
Optimized water molecule	1	0.973	0.973	103.6	0.0	0.00	0.00	0.00

ΔE = energy of an water molecule in the dimer-water complex with respect to the energy of an optimized water molecule.

Though this method explains the long range influence of solvent molecules on the hydrogen bond strength, its applicability is general and can be employed to enumerate the effect of long range electrostatic interactions due to any kind of system; and on the interactions that are dominated by electrostatic contributions in general.

Thus, the geometrical changes due to the long range influence in the individual molecules in the dimer-water complex lead to an overall destabilization of the system by 37.5 kcal/mol. However, this disadvantage is easily compensated for by

the formation of the hydrogen bonds (in total = 33) between water molecules (a total of 9, see Figure 5.5) and between water molecules and peptide dimer (a total of 24 hydrogen bonds have been found between them). The interaction energy (see Section 2.4 of Chapter 1 for definition) of 9 hydrogen bonds formed within solvent molecules has been found to be -41.4 kcal/mol. The interaction energy of the 24 peptide-water hydrogen bonds has been obtained to be -133.1 kcal/mol. Therefore, a total stabilization of -174.5 kcal/mol has been obtained due to the formation of a total of 33 intermolecular hydrogen bonds. It must also be noted that several XH (X=N,C)-O interactions between peptide and water molecules has also been observed, along with the classical hydrogen bonds. The overall effect is the stabilization of the dimer-water complex by 136.9 kcal/mol (binding energy = -136.9 kcal/mol, see Section 2.4 of Chapter 1 for definition) with respect to the infinitely separated dimer and water molecules. It is to be noted here that the optimized geometries have been obtained in the dielectric of water to exclusively pinpoint the effect of the explicit solvent molecules.

To obtain the effect of solvent molecules on the hydrogen bond strength, the electrostatic force of binding for all hydrogen bonds of the peptide backbone in the dimer-water complex has been calculated (Table 5.2). The net contribution from all the 15 water molecules towards the electrostatic force of binding for each hydrogen bond has been found to be positive, *i.e.*, unfavorable, irrespective of the charge analysis schemes (columns 3, 6, and 9 of Table 5.1). This result suggests that explicit water molecules weaken the main chain hydrogen bonds of the peptide dimer. To further elaborate upon the role of explicit water molecules, we have calculated the electrostatic force of binding for each main chain hydrogen bond in the bare dimer (Table 5.3). Table 5.3 suggests that the net force of binding for 8 hydrogen bonds out of a total 12 are larger for the dimer-water complex than for the bare dimer (with the exceptions being the 4th, the 6th, the 8th and the 10th hydrogen bonds), and this result is consistent with the types of charges being employed. This implies that though the explicit contribution from the 15 molecules of water in the dimer-water complex is unfavorable, the net effect is still favorable, and it strengthens most of the hydrogen bonds. This could be because of the induction caused due to the electric field of 15 water molecules or the charge transfer effect. To see the amount of charge being transferred from the water molecules to peptide or *vice versa*, see Table 5.1. These

effects affect the magnitude of atomic charges, which lead to the electrostatic force of binding.

Table 5.2 The calculated electrostatic force of binding in pN for the main chain hydrogen bonds in the optimized geometry of the dimer-water complex when the optimization is performed at the COSMO(water)/PBE-D2/TZVP level of theory from the crystal structure.

H-bond distance (Å) ‡	NBO charges			Mulliken charges			Löwdin charges		
	force helix	force water	net force	force helix	force water	net force	force helix	force water	net force
1.797	-876.6	42.7	-833.9	-217.3	18.2	-199.1	-137.0	10.2	-126.5
1.993	-760.8	53.3	-707.4	-184.5	25.7	-158.8	-119.5	15.0	-104.5
2.03	-712.9	46.2	-666.7	-157.3	23.7	-133.6	130.3	13.9	-116.4
1.812	-491.0	16.1	-475.0	-125.0	8.7	-116.6	-96.1	7.2	-88.9
2.36	-685.0	23.3	-661.8	-186.3	12.8	-173.5	-130.3	8.1	-122.2
1.834	-319.8	23.1	-296.7	-104.7	10.7	-94.1	-75.4	8.0	-67.4
1.921	-621.7	17.8	-603.9	-153.7	9.0	-144.8	-98.1	6.2	-91.9
2.035	-815.8	42.8	-773.0	-231.7	18.0	-213.7	-157.9	11.3	-146.7
1.889	-493.2	18.0	-475.1	-141.6	9.9	-131.7	-82.3	7.6	-74.7
1.85	-636.6	10.3	-626.3	-138.4	4.3	-134.0	-82.4	3.1	-79.3
2.116	-718.7	8.7	-710.0	-181.1	8.4	-172.7	-138.5	4.2	-134.3
1.846	-404.0	9.7	-394.4	-94.5	6.8	-87.8	-64.2	3.1	-61.1

helix = contributions from the peptide dimer

water = contributions from 15 explicit water molecules

net = the vector sum of the forces obtained from peptide and solvent contributions

‡ for the water-dimer complex obtained after optimization of the crystal structure geometry

To further ensure the generality of this result, we performed *ab initio* MD simulations on the dimer-water complex at the COSMO(water)/HF/3-21G** level of theory with a 0.5 fs time step and at 298.15 K for 3.2 fs. Geometries were analyzed at 2.55, 2.85 and 3.15 ps for equilibration. The geometry obtained at 3.15 ps was optimized at the COSMO(water)/PBE-D2/TZVP level of theory. This new geometry

Table 5.3 The calculated electrostatic force of binding for the main chain hydrogen bonds of the peptide dimer when optimization is performed at the COSMO(water)/PBE-D2/TZVP level of the theory under three different conditions.

H-bond length (Å) ‡	net force (pN)			net force (pN)			net force (pN)		
	(optimized dimer-water complex form crystal geometry)			(optimized crystal geometry in the absence of 15 water molecules)			(Optimized dimer-water complex after 3.15 ps <i>ab initio</i> MD simulation)		
	a	b	c	a	b	c	a	b	c
1.797	-833.9	-199.1	-126.5	-539.3	-125.2	-88.3	-577.1	-172.5	-103.1
1.993	-707.4	-158.8	-104.5	-654.0	-158.4	-120.6	-764.5	-192.5	-128.1
2.03	-666.7	-133.6	-116.4	-604.4	-162.9	-122.9	-756.5	-186.7	-137.1
1.812	-475.0	-116.6	-88.9	-477.1	-121.6	-93.3	-682.0	-175.3	-123.7
2.36	-661.8	-173.5	-122.2	-542.1	-153.7	-98.6	-557.3	-123.6	-76.9
1.834	-296.7	-94.1	-67.4	-444.3	-140.7	-106.9	-401.1	-114.7	-72.2
1.921	-603.9	-144.8	-91.9	-522.7	-120.1	-63.8	-640.9	-157.6	-116.3
2.035	-773.0	-213.7	-146.7	-815.4	-219.9	-167.5	-632.7	-166.0	-110.7
1.889	-475.1	-131.7	-74.7	-400.5	-118.5	-74.1	-520.8	-139.0	-83.3
1.85	-626.3	-134.0	-79.3	-702.2	-176.8	-135.4	-516.0	-112.6	-76.2
2.116	-710.0	-172.7	-134.3	-647.5	-153.2	-104.5	-801.6	-198.7	-142.6
1.846	-394.4	-87.8	-61.1	-294.6	-83.6	-50.4	-581.8	-176.6	-110.5

a = the net force obtained from NBO charges

b = the net force obtained from Mulliken charges

c = the net force obtained from Löwdin charges

‡ for the dimer-water complex obtained after optimization of the crystal structure geometry

was found to be 4.0 kcal/mol lower in energy than the previous one. Force analysis of all backbone hydrogen bonds of the peptide dimer in this geometry of dimer-water complex further suggest the positive (unfavorable) contributions from water molecules for each hydrogen bond irrespective of the type of charges being employed. However, the net attractive force of binding due to the contributions of both, the peptide and solvent molecules, has been found to be larger for most of the hydrogen bonds consistently with all the three types of charges (9 with NBO and 8 with both Mulliken and Löwdin charges), than the bare dimer (see Table 5.3 above). The reason

for this stabilization can further be explained in terms of induction and charge transfer effects caused due to the electric field of 15 water molecules, and not due to the effect of explicit charges on atoms of the solvent in the dimer-water complex. Thus, it can be concluded statistically that the explicit solvent molecules enhance the strength of main chain hydrogen bonds, in general, in the dimer-water complex that we have considered.

(ii) The Water Cluster Case. Water is basic to life on the earth. Life originated in water. Also, all the vital processes for life happen in the aqueous environment. It has been known for decades that liquid water is a network of intermolecular hydrogen bonds, which is eventually responsible for the unique properties of water. Several quantum chemical studies have been carried out in the past to understand the complex structure of liquid water.^{11,32-37} In all of these studies, water has been modeled as a cluster of water molecules. The methods that have been used mostly in these studies have been Hartree-Fock, density functional theory and perturbation theory. These studies have been mostly targeted to obtain the binding energy of different types and sizes of water clusters.^{32,33} The effect of shape and size of water clusters on the dipole moment,³³ polarizability,³⁴ vibrational spectra,^{35,36} intermolecular charge flow,³⁴ electron density at hydrogen bond critical points³⁷ and the cooperativity effect of hydrogen bonding³⁸ have all been studied in detail in water clusters. Despite the extensive list of literature dedicated to understanding the structure of water, no explicit conclusions has yet been obtained about the structure of liquid water. This is because of the fluidity and dynamics in the hydrogen bonded array formed by the water molecules. However, it has been noticed clearly that the arrangement of different water molecules around the hydrogen bond acceptor and donor vary locally in hydrogen bonded clusters, which lead to the differential strength of hydrogen bonds formed between them. Based on the coordination number of hydrogen bond donor and acceptor molecules, a total of 36 kinds of hydrogen standard bonds are predicted to be present in water when penta-coordination of oxygen and di-coordination of hydrogen atoms are excluded.¹¹ Lenz and co-workers have studied water clusters up to the size of 30 water molecules and categorized some of these hydrogen bonds based on the O-H vibrational stretch of the donors.³⁶ Tao *et al.* have studied four 50mer water clusters and recognized 16 out of 36 hydrogen bonds in them.¹¹ They

have also organized these hydrogen bonds in the order of their increasing strength using the local H-bond stretching force constant.¹¹ The coordination pattern of the strongest and the weakest hydrogen bonds has been shown below in Figure 5.7(a). However, the long range influence of surrounding water molecules on the H-bond strength of these hydrogen bonds have not been evaluated in their study. By calculating the dipole moment and intra- and intermolecular bond order in water clusters Bako and co-workers have shown that the electrostatic effects on H-bonding caused by the non-immediate environment in liquid water is important.³⁸ However, they have not exactly pinpointed the contributions of neighboring molecules on the hydrogen bond strength. To understand the actual contributions and long range influence of neighboring water molecules on the differential stability of hydrogen bonds, we have done a force analysis of the strongest and weakest type of hydrogen bonds that have been described by Tao *et al.* We have optimized all of the four (A-D) 50mer water clusters reported in their studies at the COSMO(water)/PBE-D2/TZVP level of theory. The cluster (cluster D) with lowest energy has been chosen for further force analysis. The optimized geometry of this cluster is shown in Figure 5.7(b). The strongest and weakest hydrogen bonds have been recognized based on following criteria, as described in reference 11 by Tao *et al.* The strongest hydrogen bond (20-02): the hydrogen bond donor partner is associated with two more water molecules through hydrogen bonding involving the oxygen atom and the acceptor molecule is involved in two extra hydrogen bonds through hydrogen atoms. Both hydrogen bonded partners in this case are tri-coordinated through hydrogen bonding. One hydrogen atom in the donor partner is free from hydrogen bonding, and an unstaured hydrogen bonding space is available at the oxygen atom in the acceptor partner (Figure 5.7b). The weakest hydrogen bond (11-11): both the partners in this case are involved in two additional hydrogen bonds involving one each with hydrogen and oxygen atoms. One free vacancy for hydrogen bonding is available on both the partners: at the oxygen atom on the donor and at one of the hydrogens on the acceptor (Figure 5.7b). Three hydrogen bonds of type 20-02 and seven hydrogen bonds of type 11-11 have been recognized in this cluster. In the force analysis, the hydrogen bonded molecules have been considered as a solute while the rest of the surrounding molecules have been considered as the solvent for each hydrogen bond type. The results of force analysis are summarized in Table 5.4.

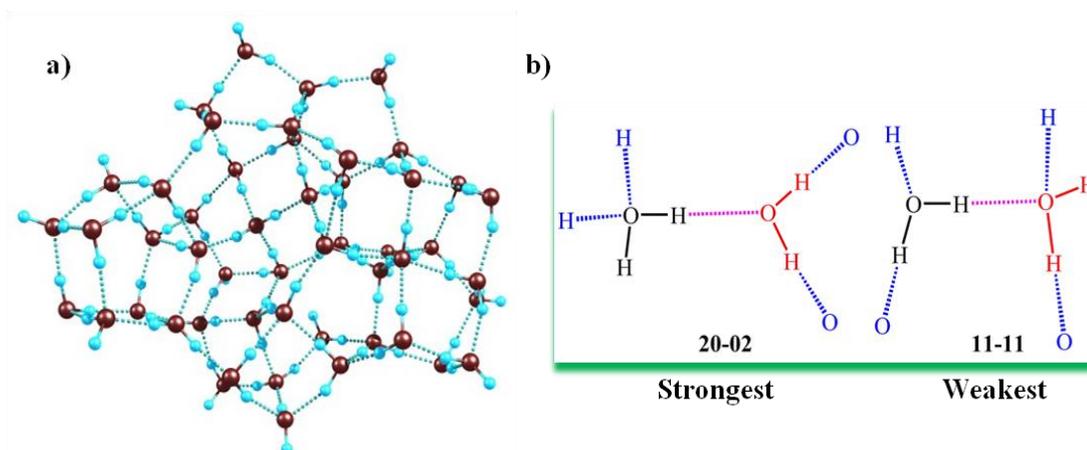


Figure 5.7 a) The optimized geometry of the 50mer water cluster at the COSMO(water)/PBE-D2/TZVP level of theory. b) Schematic representation of strongest and weakest hydrogen bonds in 50mer water cluster. Dotted lines in pink represent hydrogen bonds of interest; whereas dotted lines in blue color represent hydrogen bonding with the immediate environment for the given hydrogen bonded water dimer. The acceptor and donor molecules in two cases have been shown in red and black colors respectively.

Contributions from surrounding water molecules for all the hydrogen bonds of type 20-02 (strongest) have been found to be negative, i.e., favorable, whereas the contributions from the surrounding for the hydrogen bonds of type 11-11 (weakest) have been found to be positive (unfavorable) for 4 out of the total 7 cases. This result has been consistent with different types of charges being employed. The average values of different contributing forces in both of the types have been provided in Table 5.4, with the standard deviation inside parenthesis. The contributions from water molecules that are directly involved in hydrogen bonding in the 20-02 case have been larger (larger negative values) than for the 11-11 case, for all of the hydrogen bonds, indicating the role of the charge transfer effect. The charge transfer effect is nothing but the covalent contributions. The difference in the values of the forces for hydrogen bonded water molecules in two cases can be attributed to the covalent contributions to the hydrogen bonding. Our results suggest that the covalent contributions in 20-02 should be larger than for the 11-11 case (Table 5.4, column “contributions from water complex”, last three rows), which is in agreement with the supposition made by Tao *et al.* They have suggested that the 20-02 is largely covalent in nature, while 11-11 is electrostatic. It is worth noting here that the dipole moment

Table 5.4 The calculated average electrostatic force of binding in pN for 20-02 and 11-11 types of hydrogen bonds present in the 50mer water cluster revealing contributions from different parts of the cluster, when the geometry was optimized at the COSMO(water)/PBE-D2/TZVP level of theory.

H-bond type	H-bond distance (Å)	primary interactions force	net Force from water dimer	Force from surrounding molecules	net force of binding	
20-02 (strongest)	1.639 (0.01)	a	-107.6 (1.52)	-37.1 (2.55)	-15.8 (4.57)	-52.9 (2.39)
		b	-38.3 (0.51)	-13.5 (0.93)	-5.4 (1.32)	-18.9 (0.58)
		c	-19.9 (0.29)	-7.7 (0.64)	-2.9 (0.64)	-10.6 (0.41)
11-11 (weakest)	1.794 (0.02)	a	-90.0 (2.19)	-9.9 (8.67)	5.0 (13.64)	-4.9 (14.92)
		b	-32.1 (1.20)	-2.5 (3.81)	2.1 (5.09)	-0.3 (6.04)
		c	-15.5 (0.30)	-0.3 (1.95)	1.1 (2.51)	0.8 (3.14)
Strongest	-0.155	a	-17.6	-27.2	-20.8	-48.0
-weakest		b	-6.2	-11	-7.5	-18.6
		c	-4.4	-7.4	-4.0	-11.4

a = the net force obtained from NBO charges

b = the net force obtained from Mulliken charges

c = the net force obtained from Löwdin charges

Values outside parentheses are average values and inside parenthesis are standard deviations for a given type of hydrogen bond

of a water molecule in a cluster changes even with the addition of water molecules in the non-immediate environment from where direct charge transfer is negligible.³⁸ This suggests that the charges (redistributed in the cluster with respect to the isolated case) on the atoms in the water molecules present in the non-immediate environment also play a definitive role in determining the properties of water molecules. These long range effects cannot be captured by the traditional methods that have been employed in understanding hydrogen bonding strength in water clusters. The long range contributions of the surrounding water molecules (in the given electronic distribution in a cluster) on hydrogen bonds in water clusters can, however, be obtained from our method. As described earlier, these forces have been found to be attractive and aligned along the hydrogen bonding line for all hydrogen bonds in the 20-02 case.

The average value of this force for all hydrogen bonds has been provided in the column “contributions from solvent molecules” in Table 5.4.

Thus, the net force of binding for all hydrogen bonds in the 20-02 case has been found to be more negative than the hydrogen bonds of 11-11 type, indicating 20-02 to be stronger than 11-11, as reported by Tao *et al.*¹¹

(iii) The Cation-shell Case. Solvated ions are present everywhere: in fresh drinking water to the oceans and in biological systems.³⁹ Many chemical systems and solutions also comprise of solvated ions.⁴⁰ Thus, understanding the solvation of ions and the factors that determines their solvation is very important in chemistry and for practical applications. The purification of water,⁴¹ electrochemical applications,⁴² and the reactivity or conductivity of ions in a solution^{42,43} all depend on the solvation of ions in a given solvent. Differential solvation of ions in different solutions have also been exploited in solvent separation techniques.⁴⁴ Different concentric shells of solvent molecules are often formed around ions in solution. The energy released in the formation of these shells is one of the key factors responsible for the solvation of ions. The arrangement of solvent molecules in solvation shells depend on the nature, charge and size of the ion as well as on the nature of the solvent. Solvation (hydration) of ions inside water is accompanied by the formation of hydrogen bonds between water molecules that form different layers of solvent shells. It is thus a question of fundamental interest whether solvent molecules from other shells or different water molecules from the same shell affect the strength of hydrogen bond that is formed between water molecules from two shells. In order to answer this question, we have chosen two specific examples for a comparative study. It is reported that the length and strength of hydrogen bonds formed between the first and the second shells of Mg^{2+} - and Al^{3+} - cations vary significantly.¹² To find the contributions of other water molecules from the second shell in order to understand this problem, we optimized the Mg^{2+} - and the Al^{3+} -water shells considering water molecules only up to the second shell, and did a comprehensive force analysis for the hydrogen bonds in two geometries belonging to two different cations. The starting structure for geometry optimization of Al^{3+} -shells has been taken in reference with the MD simulation results reported by Farao *et al.*⁴⁵ A similar geometry has also been taken as a starting point for the optimization of the Mg^{2+} -shell. The optimized geometries of these cations

shells have been shown in Figure 5.8 below. It has been seen that the water molecules in the first shell are coordinated with the cations through covalent bonding in an octahedral geometry. However, water molecules of the first shell interact with that of the second shell through hydrogen bonding. The hydrogen bond lengths for all the 12 hydrogen bonds formed between two shells have been listed in Table 5.5. It can be seen from Table 5.5 (columns 1 and 6) that the hydrogen bonds formed between the first and the second shells of Mg^{2+} cation are significantly larger in comparison to the shells for the Al^{3+} cation.

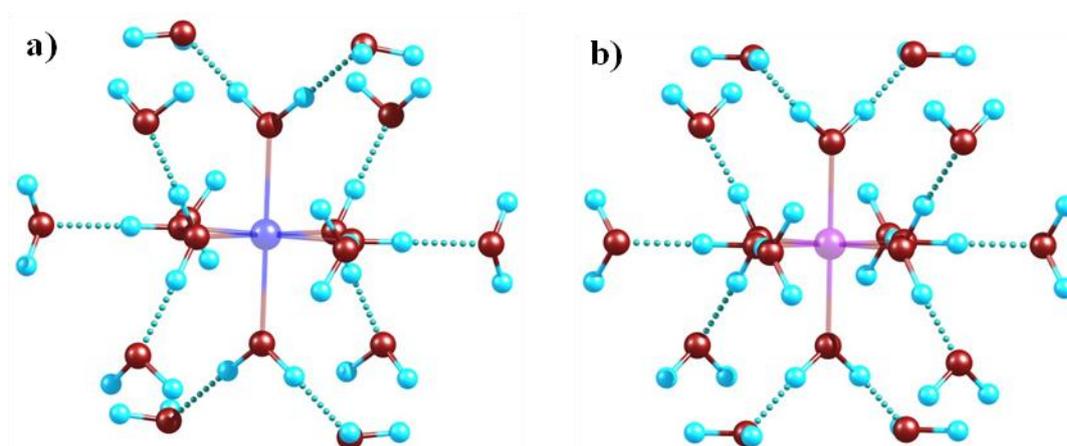


Figure 5.8 The optimized geometry at the COSMO(water)/PBE-D2/TZVP level of theory of (a) the Mg^{2+} -water shell and (b) the Al^{3+} -water shell. Color symbols: cyan – hydrogen, blue – Mg, maroon – oxygen, magenta – Al.

The water molecules from the second shell other than the one that is involved in hydrogen bonding have been considered as explicit solvent molecules and their influence on hydrogen bonds formed between the two shells has been obtained by the force analysis. The results from the force analysis (with charges obtained from NPA) of all 12 hydrogen bonds in these cation-shells has been provided in Table 5.5. It can be seen from Table 5.5 that the contributions from the explicit solvent molecule in both of the cases for all hydrogen bonds are repulsive, which indicates that the explicit presence of these molecules weakens the hydrogen bond strength. To further corroborate this result, we optimized these cation shells removing 11 explicit water molecules from the second shell in each of the two cases. The optimized geometries for the two cases are shown below in Figure 5.9. It can be observed from Figure 5.9 that the hydrogen bond distances between the aforementioned shells in two cases

decreases significantly, which is in accordance with our results. The hydrogen bond in Al^{3+} -water shell is closer to a weak covalent bond than a noncovalent bond with slight covalent contributions. We have further separated out the contributions from the first shell and the cations (Table 5.5, columns 4 and 9) to the overall bonding. Contributions from the first shell and the cation in both the cases are significantly negative, making the overall force attractive, thereby make these hydrogen bonds feasible. It can further be noticed from Table 5.5 that the contributions from the first shell of water and the trivalent aluminium cation in the Al^{3+} -water shell is much larger than the Mg^{2+} -water shell. This provides the reason why hydrogen bonds between the first and second shells of the Al^{3+} -water shell is smaller (and hence stronger) than the Mg^{2+} -water shell.

Table 5.5 The calculated electrostatic force of binding for different hydrogen bonds revealing contributions from different parts of the Mg^{2+} -water and Al^{3+} -water shells at the COSMO(water)/PBE-D2/TZVP level optimized geometries.

Solvated Mg cation (forces in pN, NBO)					Solvated Al cation (forces in pN, NBO)				
H-bond length (Å)	water dimer	water from 2 nd shell	1 st shell and metal	Net force of binding	H-bond length (Å)	water dimer	water from 2 nd shell	1 st shell and metal	net force of binding
1.747	11.9	11.7	-57.6	-34.0	1.608	-4.1	30.1	-98.6	-72.6
1.747	11.6	12.0	-57.1	-33.5	1.609	-4.1	30.2	-98.7	-72.5
1.748	12.1	12.0	-58.4	-34.4	1.608	-4.1	30.2	-98.7	-72.6
1.748	11.9	12.4	-58.2	-33.9	1.609	-4.1	30.1	-98.6	-72.6
1.748	12.1	12.4	-58.8	-34.4	1.609	-4.1	30.2	-98.7	-72.6
1.747	11.5	12.6	-57.1	-33.0	1.609	-4.1	30.3	-98.7	-72.5
1.748	12.2	12.2	-59.5	-35.0	1.608	-4.1	30.0	-98.5	-72.6
1.748	11.7	12.8	-58.7	-34.2	1.609	-4.1	30.0	-98.4	-72.5
1.748	12.1	12.5	-59.2	-34.7	1.609	-4.1	30.0	-98.5	-72.5
1.747	11.6	12.1	-57.1	-33.5	1.608	-4.0	29.8	-98.3	-72.5
1.748	12.0	12.4	-58.7	-34.3	1.608	-4.0	30.0	-98.5	-72.5
1.747	11.5	12.7	-57.5	-33.3	1.608	-4.0	30.0	-98.5	-72.5

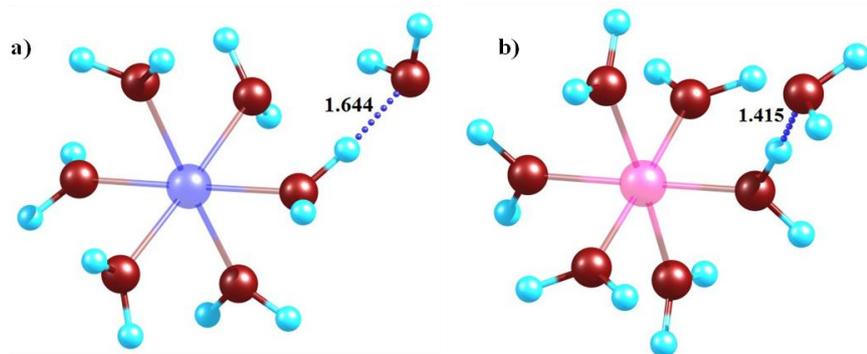


Figure 5.9 The optimized geometry at the COSMO(water)/PBE-D2/TZVP level of theory of (a) the Mg^{2+} -water shell and (b) the Al^{3+} -water shell. Color symbols: cyan – hydrogen, blue – Mg, maroon – oxygen, magenta – Al.

5.4 Association constant of hydrogen-bonded complexes

We have shown that the long range secondary electrostatic interactions bestow directionality into the electrostatic dominated noncovalent interactions. This effect can be estimated in terms of the electrostatic force considering the atoms as point charges.⁴ Contributions from the explicit solvent molecules in terms of the net electrostatic force exerted by them on hydrogen bonding solutes can also be obtained. The derivative of the electrostatic force with respect to a positive test charge is the electrostatic field. The field thus obtained can be applied on the molecule to obtain the influence of solvent molecules on the hydrogen bonding of solutes. The component of the field along the hydrogen bond will exclusively influence the strength of the hydrogen bond. Furthermore, the components that are perpendicular to the hydrogen bonding plane will affect the geometry of the hydrogen bonded systems. However, these geometrical changes may lead to a strain in the molecule and thus affect the net stabilization indirectly. However, the electric field calculated thus may not provide the exact value of the force for the water molecules when applied to a hydrogen bonded solute, as the field is calculated with respect to a test charge whose magnitude is different than the actual charge that is present in the system. Theoretical methods that are available today provide the electric field employing a reference test charge, which may lead to erroneous results. The calculated field-equivalent involving actual charges of the system is expected to provide a better force for replacing the solvent

molecules in the calculations, provided the charge analysis being employed to obtain atomic charges is reliable.

Once the exact-field equivalent for the solvent is known, the association constant of two hydrogen bonding partners can be estimated in the presence of explicit solvent molecules, employing the field equivalent of the solvent through the following equation. It should be noted here that there is no theoretical method available to date that can give an estimate of the effect of explicit solvent molecules on the association constant of hydrogen bonded complexes or that can provide association constant of partners in the presence of explicit solvent molecules.

$$K_s = \exp(-\Delta G_s/RT) \quad (5.1)$$

Where K_s is association constant of a hydrogen bonded complex in the presence of the explicit solvent molecules and ΔG_s is the free energy of association in the presence of explicit solvent molecules, which can be calculated from the following equation:

$$\Delta G_s = \Delta G_x + \Delta G_y + \Delta G_z - 2.\Delta G \quad (5.2)$$

Where ΔG_x , ΔG_y and ΔG_z are the free energies of association when the field-equivalent for the solvent is applied along X, Y and Z axes respectively, and ΔG is the free energy of association in the absence of explicit solvent molecules. The derivation of Equation 1.1 is as follows .

When the free energy of association of hydrogen bonded complexes in absence of explicit solvent molecules is ΔG , the association constant K for a hydrogen bonded complex in the absence of any external factor is written as,

$$K = \exp(-\Delta G/RT) \quad (5.3)$$

When an external electric field or any other perturbation is applied along a hydrogen bond, say along the X-axis

$$K_x = \exp(-\Delta G_x/RT) \quad (5.4)$$

Now, ΔG_x can be written as the sum of ΔG and the perturbation energy due to external factor, say λ_x . Therefore, Equation (5.4) can be written as,

$$K_x = \exp(-(\Delta G + \lambda_x)/RT)$$

$$\Rightarrow K_x = \exp(-\Delta G/RT) \cdot \exp(-\lambda_x/RT)$$

$$\Rightarrow K_x = K \cdot K' \quad (5.5)$$

$$\text{where } K' = \exp(-\lambda_x/RT)$$

Similarly, when the external electric field is applied along the Y and the Z axes, the respective association constants would be –

$$K_y = K \cdot K'' \quad (5.6)$$

and

$$K_z = K \cdot K''' \quad (5.7)$$

respectively, where $K'' = \exp(-\lambda_y/RT)$ and $K''' = \exp(-\lambda_z/RT)$.

Therefore, when the electric fields are applied along all the axes, the association constant can be written as,

$$K = K \cdot K' \cdot K'' \cdot K''' = (K \cdot K') \cdot (K \cdot K'') \cdot (K \cdot K''') / K^2 = K_x \cdot K_y \cdot K_z / K^2$$

$$\Rightarrow \Delta G_s = \Delta G_x + \Delta G_y + \Delta G_z - 2 \cdot \Delta G,$$

which is the desired equation.

5.5 Conclusions

We have shown that the strength of an individual H-bond in multi-hydrogen bonded systems can be estimated by computing the electrostatic force of binding along the chosen H-bond when the forces are derived from atomic charges. Thus, the influence of solvent molecules on hydrogen bonds can also be estimated as the difference in the electrostatic force of binding in the presence and absence of solvent molecules. The results of the force analysis of hydrogen bonds in three different kinds of systems suggest that the mere explicit presence of solvent molecules may affect the H-bond strength, which may depend upon the abundance and orientation of the solvent molecules. These effects can be seen to be operated by two means: (i) the charge transfer and the induction effect of solvent molecules and (ii) the long range electrostatic interactions involving solvent molecules.

5.6 References

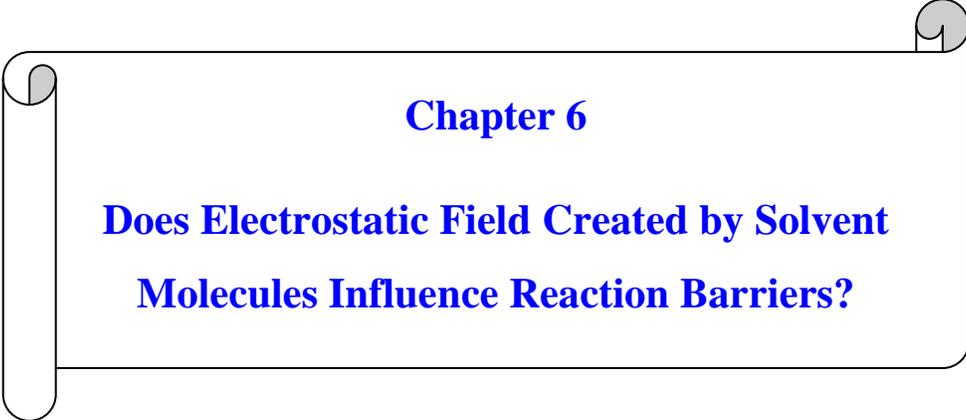
1. E. A. C. Davie, S. M. Mennen, Y. Xu, S. J. Miller, *Chemical reviews*, 2007, **107**, 5759-5812; J.-M. Lehn, *Chemical Society Reviews*, 2007, **36**, 151-160; L. Brunsveld, B. Folmer, E. Meijer, R. Sijbesma, *Chemical Reviews*, 2011, **101**, 4071-4098; C. B. Aakeröy, K. R. Seddon, *Chemical Society Review*,s 1993, **22**, 397-407; F. Zhao, M. L. Ma, B. Xu, *Chemical Society Reviews*, 2009, **38**, 883-891; A. G. Doyle, E. N. Jacobsen, *Chemical reviews*, 2007, **107**, 5713-5743.
2. T. Šmejkal, B. Breit, *Angewandte Chemie*, 2008, **120**, 4010-4013; R. R. Knowles, E. N. Jacobsen, *Proceedings of the National Academy of Sciences*, 2010, **107**, 20678-20685; K. Biradha, *CrystEngComm*, 2003, **5**, 374-384; A. Bauzá, T. J. Mooibroek, A. Frontera, *CrystEngComm*, 2016, **18**, 10-23; R. G. Gonnade, M. S. Shashidhar, M. M. Bhadbhade, *Journal of the Indian Institute of Science*, 2012, **87**, 149; B. Sellergren, M. Lepistoe, K. Mosbach, *J. Am. Chem. Soc.*, 1988, **110**, 5853-5860; T. Rossow, S. Seiffert, *Supramolecular Polymer Networks and Gels*, Vol. 268 (Eds.: S. Seiffert), Springer, 2015, pp. 1-46; D. J. Hill, M. J. Mio, R. B. Prince, T. S. Hughes, J. S. Moore, *Chemical reviews*, 2001, **101**, 3893-4012; D.-W. Zhang, X. Zhao, J.-L. Hou, Z.-T. Li, *Chemical Reviews*, 2012, **112**, 5271-5316; D. Wang, G. Tong, R. Dong, Y. Zhou, J. Shen, X. Zhu, *Chemical Communications*, 2014, **50**, 11994-12017; H. Yang, B. Yuan, X. Zhang, O. A. Scherman, *Accounts of chemical research*, 2014, **47**, 2106-2115; B. R. Beno, K. S. Yeung, M. D. Bartberger, L. D. Pennington, N. A. Meanwell, *Journal of medicinal chemistry*, 2015, **58**, 4383-4438.
3. P. L. Popelier, L. Joubert, *J. Am. Chem. Soc.*, 2002, **124**, 8725-8729; J. Pranata, S. G. Wierschke, W. L. Jorgensen, *J. Am. Chem. Soc.*, 1991, **113**, 2810-2819.
4. M. K. Tiwari, K. Vanka, *Chemical Science*, 2017, **8**, 1378 – 1390.
5. P. Ball, *Chem. Rev.* 2008, **108**, 74–108.
6. M. K. Tiwari, A. K. Jana, K. Vanka, N. Sengupta, *Phys. Chem. Chem. Phys.*, 2016, **18**, 5910-5924; C. Beeson, N. Pham, G. Jr. Shipps, T. A. Dix, *J. Am.*

- Chem. Soc.* 1993, **115**, 6803–6812; M. G. Sarwar, B. Dragisic, L. J. Salsberg, C. Gouliaras, and M. S. Taylor, *J. Am. Chem. Soc.* 2010, **132**, 1646–1653; L. R. Olano, S. W. Rick, *J. Am. Chem. Soc.* 2004, **126**, 7991.
7. J. L. Cook, C. A. Hunter, C. M. R. Low, A. Perez-Velasco, J. G. Vinter, *Angew. Chem., Int. Ed.* 2007, **46**, 3706–3709.
 8. A. C. English, S. H. Done, L.S. Caves, C. R. Groom, R. E. Hubbard, *Proteins*, 1999, **37**, 628-640.
 9. K. Kalyanasundaram, J. K. Thomas, *J. Am. Chem. Soc.*, 1977, **99**, 2039-2044.
 10. J. Liu, D. Wang, Q. Zheng, M. Lu, P. S. Arora, *J. Am. Chem. Soc.* 2008, **130**, 4334-4337.
 11. Y. Tao, W. Zou, J. Jia, W. Li, D. Cremer, *J. Chem. Theory Comput.* 2017, **13**, 55–76.
 12. D. Spångberg, K. Hermansson, *J. Chem. Phys.*, 2004, **120**, 4829-4842.
 13. A. Klamt, G. Schüürmann *J. Chem. Soc. Perkin Trans. 2*, 1993, **5**, 799–805.
 14. R. Ahlrichs, *Chem. Phys. Lett.*, 1989, **162**, 165-169.
 15. J. P. Perdew, *Physical Review Letters*, 1996, **77**, 3865.
 16. K. Eichkorn, *Chemical Physics Letters*, 1995, **240**, 283-289.
 17. M. Sierka, *J. Chem. Phys.*, 2003, **118**, 9136-9148.
 18. J. Hepburn, G. Scoles, R. Penco. *Chem. Phys. Lett.*, 1975, **36**, 451–456; R. Ahlrichs, R. Penco, G. Scoles. *Chem. Phys.*, 1977, **19**, 119–130; S. Grimme, *J. Comput. Chem.*, 2004, **25**, 1463–1473; S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787–1799; S. Grimme; J. Antony; S. Ehrlich; H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
 19. R. S. Mulliken, *J. Chem. Phys.*, 1955, **23**, 1833-1840.
 20. A. E. Reed, R. B. Weinstock, F. Weinhold, *J. Chem. Phys.*, 1985, **83**, 735-46.
 21. P. O. Löwdin, *J. Chem. Phys.*, 1955, **21**, 374-375.
 22. T. Clark, P. Politzer, J. S. Murray, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2015, **5**, 169-177.
 23. R. P. Feynmann, *Phys. Rev.*, 1939, **56**, 340-343.

24. P. Politzer, K. E. Riley, F. A. Bulat, J. S. Murray, *Comput. Theor. Chem.*, 2012, **998**, 2-8; M. Hennemann, J. S. Murray, P. Politzer, K. E. Riley, T. Clark, *J. Mol. Model.*, 2012, **18**, 2461-2469; K. Dyduch, M. P. Mitoraj, A. Michalak, *J. Mol. Model.*, 2013, **19**, 2747-2758.
25. P. Politzer, J. S. Murray, and T. Clark, *Phys. Chem. Chem. Phys.*, 2013, **15**, 11178-11189; M. R. Scholfield, C. M. V. Zanden, M. Carter, P. S. Ho, *Protein Science*, 2013, **22**, 139-152; P. Politzer, J. S. Murray, and T. Clark, σ -Hole Bonding: A Physical Interpretation. *In halogen Bonding I, Vol. 358*, Springer International Publishing, 2014, PP. 19-42.
26. M. A. A. Ibrahim, *Journal of computational chemistry*, 2011, **32**, 2564-2574.
27. W. L. Jorgensen, P. Schyman, *J. Chem. Theory Comput.*, 2012, **8**, 3895-3901; M. Kolář, P. Hobza, *J. Chem. Theory Comput.*, 2012, **8**, 1325-1333.
28. P. Politzer, J. S. Murray, T. Clark, *Phys. Chem. Chem. Phys.*, 2010, **12**, 7748-7757; P. Politzer, J. S. Murray, *ChemPhysChem* 2013, **14**, 278-294; P. Politzer, J. S. Murray, G. V. Janjić, S. D. Zarić, *Crystals*, 2014, **4**, 12-31.
29. M. Siedlecka, G. Goch, A. Ejchart, H. Sticht, A. Bierzyski, *Proc. Natl. Acad. Sci. U.S.A.* 1999, **96**, 903-908; J. X. Yang, K. Zhao, Y. X. Gong, Y. Vologodskii, N. R. Kallenbach, *J. Am. Chem. Soc.* 1998, **120**, 10646-10652.
30. R. N. Chapman, P. S. Arora, *Org. Lett.* 2006, **8**, 5825-5828; G. Dimartino, D. Wang, R. N. Chapman, P. S. Arora, *Org. Lett.* 2005, **7**, 2389-2392.
31. G. Sini, P. Maitre, P. C. Hiberty, S. S. Shaik, *J. Mol. Struct. Theochem*, 1991, **229**, 163-188 ; S. Shaik, D. Danovich, W. Wu, P. C. Hiberty, *Nature Chem.*, 2009, **1**, 443-449.
32. R. M. Shields, B. Temelso, K. A. Archer, T. E. Morrell, G. C. Shields, *J. Phys. Chem. A*, 2010, **114**, 11725-11737; F. Ramírez, C. Hadad, D. Guerra, J. David, B. Temelso, K. Archer, G. Shields, *J. Phys. Chem. A*, 2011, **115**, 12034-12046; B. Temelso, G. C. Shields, *J. Chem. Theory Comput.*, 2011, **7**, 2804-2817; J. C. Howard, G. S. Tschumper, *WIREs Comput. Mol. Sci.*, 2014, **4**, 199-224; V. Bryantsev, M. Diallo, A. van Duin, W. Goddard, *J. Chem.*

- Theory Comput.*, 2009, **5**, 1016–1026; S. Iwata, *Phys. Chem. Chem. Phys.*, 2014, **16**, 11310; M. Elango, V. Subramanian, N. Sathyamurthy, *J. Chem. Sci.*, 2009, **121**, 839–848; N. Sahu, S. Khire, S. Gadre, *Mol. Phys.*, 2015, **113**, 2970–2979; P. Qian, W. Song, L. Lu, Z. Yang, *Int. J. Quantum Chem.*, 2009, **110**, 1924–1937; J. Liu, L. Wang, J. Zhao, *J. Comput. Theor. Nanosci.*, 2009, **6**, 454–458; R. Ludwig, A. Appelhagen, *Angew. Chem. Int. Ed.* 2005, **44**, 811–815; O. Loboda, V. Goncharuk, *Chem. Phys. Lett.*, 2010, 484, 144–147; L. Turi, *J. Chem. Phys.*, 2014, **140**, 204317; Y. Wang, V. Babin, J. Bowman, F. Paesani, *J. Am. Chem. Soc.*, 2012, **134**, 11116–11119; W. Zou, R. Kalescky, E. Kraka, D. Cremer, *J. Chem. Phys.*, 2012, **137**, 084114; M. Freindorf, E. Kraka, D. Cremer, *Int. J. Quantum Chem.*, 2012, **112**, 3174–3187. M. Ahtee, *Phys. Educ.*, 1969, **4**, 379–380.
33. R. Parthasarathi, M. Elango, V. Subramanian, N. Sathyamurthy, *J. Phys. Chem. A*, 2009, **113**, 3744–3749.
34. F. Yang, X. Wang, M. Yang, A. Krishtal, C. van Alsenoy, P. Delarue, P. Senet, *Phys. Chem. Chem. Phys.* 2010, **12**, 9239.
35. G. Fanourgakis, E. Apra, W. de Jong, S. Xantheas, *J. Chem. Phys.*, 2005, 122, 134304; L. Huang, S. Lambrakos, A. Shabaev, N. Bernstein, L. Massa, *C. R. Chim.*, 2015, **18**, 516–524.
36. A. Lenz, L. Ojamae, *J. Phys. Chem. A*, 2006, **110**, 13388–13393.
37. Y. Neela, A. Mahadevi, G. Sastry, *J. Phys. Chem. B*, 2010, **114**, 17162–17171.
38. I. Bako, I. Mayer, *J. Phys. Chem., A*, 2016, **120**, 631–638.
39. W. Kaim, B. Schwederski, *Bioinorganic Chemistry: Inorganic Elements in the Chemistry of Life : An Introduction and Guide*; Chichester : Wiley: 1994; D. A. Dougherty, D. A. Stauffer, *Science*, 1990, **250**, 1558; Dennis A. Dougherty, *Science*, 1996, **271**, 163.
40. D. T. Richens, *The Chemistry of Aqua Ions*, Chichester, Wiley, 1997.
41. J. K. Edzwald, *Water Quality and Treatment*, 6th Edition, New York:McGraw-Hill, 2011; A. A. Zagorodni, *Ion exchange materials: properties and applications*. Elsevier, 2007.
42. Hamann, C. H.; Hamnett, A.; Vielstich, W. *Electrochemistry*; Weinheim: VCH: 1998.

43. Burgess, J. *Metal Ions in Solution*; Chichester, Ellis Horwood: 1978.
44. Marcus, Y. *Ion solvation*; Chichester:Wiley: 1985 and references therein.
45. T. M. C. Faro, G. P. Thim, M. S. Skaf, *J. Chem. Phys.* 2010, **132**, 114509.



Chapter 6

**Does Electrostatic Field Created by Solvent
Molecules Influence Reaction Barriers?**

Chapter 6

Does Electrostatic Field Created by Solvent Molecules Influence Reaction Barriers?

Abstract: Solvent's influence on reaction barriers has been understood in four basic ways, through (i) macrosolvation, (ii) microsolvation, (iii) the proton relay effect and (iv) the coordinating effect. A reaction barrier gets altered due to differential stabilization of transition state and reactant structures in almost all chemical transformations by (i) and (ii), and by direct participation of solvent molecules in a few that reduces strain in transition state structures through (iii) and (iv). The major reason argued behind differential stabilization of reactants and transition state structures (TSs) is the differential charge distribution in them. The effect of differential charge distribution has also been seen in the differential stabilization of the two structures in an externally applied electric field. Importantly, explicit solvent molecules that are present near reaction sites may also create their own local electric field, which may then affect the reaction barriers. In this chapter, we have proposed a general scheme to enumerate the effect of the electric field created by solvent molecules on reaction barriers. We have further shown that a field-equivalent-solvent approach may be employed for representing the effect of actual solvent molecules in the reaction profile, in order to obtain the activation barriers in the condensed phase. The barriers obtained in this field-equivalent of solvent approach by employing our method correlate linearly with the barriers obtained in the actual presence of solvent molecules for eight sample cases. Although this method has been described by employing the field equivalent of solvent molecules, it has general significance and can be exploited to compute the effect of any molecule or molecular fragment on the barrier heights in terms of their corresponding field-equivalent in QM calculations.

6. 1 Introduction

All homogeneous *in vivo* or *in vitro* reactions happen in a solvent environment. The presence of a specific solvent environment has been found to be critical for the

occurrence of a number of chemical transformations.¹⁻¹⁰ It has been shown that the solvent does not merely act as a spectator or just a medium, but also takes part actively in many chemical transformations.³⁻⁹ It has also been reported that the presence of a specific solvent and its coordination in specific arrangements induces stereoselectivity.¹² Moreover, the existing literature elucidates four basic features of the solvent that can affect the reaction barrier: (i) macrosolvation or the bulk solvation or the polarity effect,¹ (ii) microsolvation or the noncovalent bonding effect,^{11,13} (iii) the proton relay effect of polar protic solvents^{12,14} and (iv) the coordinating effect of organic or basic solvents¹⁵⁻¹⁶ (see Figure 6.1). A reaction barrier gets affected due to differential stabilization of the transition state and reactant structures by (i) and (ii), and by direct participation of solvent molecules that reduces strain in transition state structures (TSs) through (iii) and (iv). Conditions (i) and (ii) are common to almost all kinds of reactions happening in homogeneous environments. However, the last two effects are restricted to only a few specific cases. The major reason argued behind differential stabilization of reactants and TSs *via* (i) and (ii) is the differential charge distribution of atoms that are part of the reaction coordinates in the given reaction. Effects (i) and (ii) have been suggested to be captured in quantum chemical calculations by the cluster continuum model that employs geometry optimizations in solvent continuum with explicitly added solvent molecules.¹⁷ However, this model is lengthy and has many flaws in its implementation, in comparison to actual solvent systems in full QM calculations. For example, consideration of a rigid shell of solvent molecules (which is often formed due to covalent coordination of the solvent molecules with cations) covering the reactants or TSs may not be true in cases of uncharged or diffused charged systems. Thus, determining the appropriate number of solvent molecules to be considered in implementing this model for a given reaction is a tedious task. Even in the cases where a rigid cluster of solvent molecules surrounding the reacting species may be obtained, the long range electrostatic influence of weakly coordinated solvent molecules (with reacting species) on the reaction barriers may not be captured employing this model. More importantly, a complete shell of solvent molecules cannot be employed to a large or moderate size molecular system in QM calculations. The intermolecular interactions and the large conformational space of larger solvent clusters add further complications in implementing this model to the real chemical systems. Furthermore, the unresolved

structure of the solvent in a given condition of reaction put in additional practical difficulties for the successful exploitation of this model.

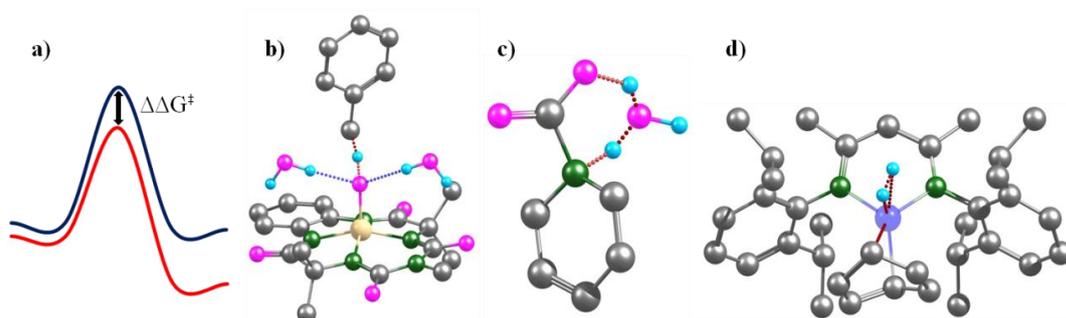


Figure 6.1 Different roles of solvent in determining the reaction barrier: a) effect of solvent polarity, b) effect of microsolvation through noncovalent bonding, c) the proton relay effect involving the polar protic solvent and d) the coordinating effect of the aromatic solvent. Colour representation: black – carbon, pink – oxygen, green – nitrogen, cyan – hydrogen, Corn silk – Iron, Blue – aluminium. The dotted blue line represents the hydrogen bond and the dotted red line represents bonds that are part of the reaction coordinate. Hydrogen atoms that are not the part of the transition state and the hydrogen bonding solvent are omitted for the sake of clarity.

The effect of long range electrostatic influence on reaction barriers can also be correlated to the reactions that are performed exclusively inside molecular cages.¹⁸ Furthermore, the secondary coordination spheres of enzyme active sites that are liable to exert long range electrostatic influence have been found to be extremely important in enzyme catalysis.¹⁹ It has been seen in bio-mimetic reactions, in particular in metalloenzymes, that a mimic of only active sites of enzymes leads to poor reactivity and selectivity.²⁰ Different approximate approaches such as quantum mechanics/molecular mechanics (QM/MM),²¹ quantum chemical cluster approach,²² empirical valence bond (EVB) approach²³ have been developed and employed for the quantum chemical modelling of enzymatic reactions in order to capture the long range electrostatic influence and steric effects. Another important method that has been reported that deals with the enhanced efficiency of QM calculations of larger molecular systems is the molecular tailoring approach.²⁴ However, all of these approaches are approximate and apply full QM calculations on only small model

systems, with the additional disadvantage of their having no accurate partitioning or truncations schemes. The two basic advantages these methods offer in taking a whole molecular system or a bigger model system over small models are the improved treatment of steric and the electronic effects. It is to be noted that the electronic effect in non-conjugated systems is nothing but the electrostatic effects. Considering the popularity and importance of the aforementioned methods, it is apparent that the long range electrostatic influence of distant substituents on reacting species and/or catalyst or the explicit solvent molecules on rate determining barriers (RDBs) are of general and practical significance. Herein, we have proposed an alternative approach to account for the long range influence of distal substituents or the explicit solvents molecules on the barrier heights of a given chemical reaction.

Since the charge distribution of atoms (particularly those atoms that are part of reaction coordinates) in reactants and TSs vary, their response in the externally applied electric field will also differ, which may lead to discrimination in the stabilization of the two structures depending upon the orientation and the magnitude of the applied field. Recently, electrostatic catalysis had attracted significant attention.²⁵ For both chemical and biological systems and for a wide range of chemical transformations, it has been shown that the applied external electric field indeed affects the reaction barriers.²⁵ This point has been illustrated employing both experimental as well as the computational approaches.²⁵⁻²⁶ This fact has led to the central aspect of our approach: *since atoms in solvent molecules acquire certain electric charges, they may create their own local electric field whose magnitude and orientation may vary point to point in the reaction site.* Thus, different bonds in reactants and TSs may feel different values of the electrostatic field, and hence will be stabilized differently, which will eventually be reflected in the barrier heights obtained in the presence of solvent molecules. The same can also be obtained by determining the electrostatic field that is applied by the solvent molecules (field-equivalent-solvent) on the bonds that undergo charge redistribution during the reaction. The electrostatic field effect created by solvent molecules may be understood in terms of the long range electrostatic interactions, as the multipoint electrostatic interactions show directional behaviour.

Also, as described earlier, including actual solvent molecules brings multiple additional effects and hence complexity into the calculations that need to be taken

care of, addressing which may make the QM calculations unreasonably slow and expensive. In this theoretical and computational study, we have proposed a method to bypass the multivariant effect of the actual solvent molecules with only one variable, *i.e.*, replacing actual solvent molecules with their field-equivalent in the quantum chemical calculations. Geometry optimizations applying field equivalent of solvents and the implicit solvent model together are proposed to provide a better picture of the solvent effect as compared to the existing solvent models. The field-equivalent-solvent in this study has been obtained along all the bonds that are part of the reaction coordinates, as all bonds may feel their own unique field equivalent in the given solvent. The method that has been employed to obtain the field-equivalent-solvent along reaction coordinates has been described in the “Field-equivalent-solvent” subsection of “Theoretical Background” section. Thus obtained field-equivalent-solvents along different reactions coordinates may be applied along corresponding bonds to obtain the corresponding barriers employing full QM calculations. Then, the net barrier in the field-equivalent-solvent may be obtained by the following Equation (a detailed description and the proof of the Equation has been provided in the “Barrier in field-equivalent” subsection of the “Theoretical Background” section):

$$\Delta G_s^\ddagger = \Delta G_1^\ddagger + \Delta G_2^\ddagger + \dots + \Delta G_n^\ddagger - (n-1) \cdot \Delta G^\ddagger \quad (6.1)$$

Where ΔG_s^\ddagger is the net barrier in field-equivalent of solvent, which corresponds to the barrier in presence of the actual solvent molecules; ΔG_1^\ddagger , ΔG_2^\ddagger , ..., ΔG_n^\ddagger are barriers when corresponding field-equivalent-solvents have been applied along different bonds that are part of the reaction coordinates; ΔG^\ddagger is the barrier in the absence of explicit solvent molecules and their field-equivalent, and n is the number of bonds that are part of the reaction coordinates for a given reaction. A linear correlation between the net barriers obtained in the field-equivalent-solvent and the barriers obtained in the presence of the actual solvent molecules have been obtained for eight sample cases. This indicates that the quantitative trend has been maintained in the barriers heights and thus the method can be synergised with the implicit solvent model to account for the solvent effect in a more accurate way in the actual practical applications employing QM calculations. Although this method has been described estimating electric field effects of explicit solvent molecules, it has a general

significance and can be exploited to compute the effect of any molecule or molecular fragment on the barrier heights in terms of their corresponding field-equivalent employing full QM methods. For example, in cases such as enzyme catalysis where it has been proposed that the electric field created by the protein environment affect the rate and the selectivity of the reaction, this method would be very relevant and significant. Therefore, this method can be employed in conjunction with the Quantum Chemical Cluster approach for an improved treatment of electronic and steric effects of enzymatic reactions.

This long range electrostatic effect of explicit solvent molecules has not been effectively included in the available implicit models of the solvent.²⁷⁻²⁹ The solvent models such as COSMO-RS (conductor like screening model for real system)²⁸ and mean field theory of solvent²⁹ addresses this issue only to a certain extent. Therefore, the current approach represents a significant advance in the modelling of distal electrostatic effects.

6.2 Computational Details

All the DFT calculations, until unless mentioned specifically, were carried out at the M06-2X/6-31G** level of theory³⁰ using the Gaussian 09 suite of quantum-chemical programs.³¹ All the QM calculations, for all the geometries of reactants and transition states (in both presence and absence of explicit solvent molecules), have been performed in the solvent phase employing Conductor like Polarization Continuum Model (CPCM).³² The solvent systems that have been considered in this study are water ($\epsilon = 78.3553$), DMSO ($\epsilon = 46.826$), methanol ($\epsilon = 32.613$) and toluene ($\epsilon = 2.3741$). The optimized reactants geometries for all the reactions, in the presence or in the absence of explicit solvent molecules, have been obtained by keeping all reactant molecules in one structure. Only the rate and selectivity determining steps for all the chosen reactions have been considered to illustrate our approach, which accounts for the long range influence of solvent molecules on reaction barriers. It has been assumed that the bonds that undergo electronic redistribution in the transition state will only be influenced by the charges on the atoms of the explicit solvent molecules. Therefore, all the bonds that undergo physical changes during the course of reaction due to redistribution of electron density or electric charges (*i.e.*, the bonds that are part

of the reaction coordinate) have been considered in the force analysis. However, it is worth noting that all the bonds in the reactants and transition states are liable to get affected by the long range electrostatic influence of solvent molecules, but the bonds that do not undergo any changes during the course of the reaction will get affected equally (or nearly equally) in the reactant and the transition states (provided that the relative position of the solvent molecules in the two structures are the same), and hence will not alter the reaction barrier. Therefore, the bonds that are not part of the reaction coordinates have not been considered for force analysis investigating the long range electrostatic influence of the solvent molecules.

In order to estimate long range electrostatic influence of solvent molecules on the reaction barriers, all of the selected bonds have been considered to be made up of two subsystems. The NBO³³ charges have been used to calculate the net electrostatic force experienced by each subsystem related to a chosen bond in a molecule (or a line along which a new bond is going to be formed between two reactant molecules in the transition state). The electrostatic force on each fragment (or subsystem) of a chosen bond due to the other fragment of the same bond or the other molecules in the vicinity has been calculated along that chosen bond. The magnitudes of the forces for any fragment or subsystem belonging to any chosen bond has been calculated employing the dielectric constant of the same solvent (as per Coulomb's Law) that has been employed in the QM calculations of that system. The procedures described in Chapter 5 to obtain the electrostatic force of binding for hydrogen bonded complexes have been applied here as well in order to obtain the net electrostatic forces corresponding to every subsystem that represent the aforementioned bonds most appropriately. Two different variants of the code have been employed for the force calculations for bonds in acyclic and cyclic systems. This point has been explained in detail in the next section. The forces experienced by different fragments of all the chosen bonds have been employed for the calculation of the corresponding field-equivalent-solvent, which have further been employed to compute the corresponding barriers. These barriers have then been utilized to obtain the net barrier (according to Equation (6.1)) in the field-equivalent-solvent, which corresponds to the barrier in presence of actual solvent molecules. This approach has been described in detail in the following section.

6.3 Theoretical Background

In order to estimate the long range electrostatic influence of explicit solvent molecules, we have introduced the concept of the field-equivalent-solvent. The electrostatic forces experienced by each subsystem belonging to a selected bond, in the presence and absence of explicit solvent molecules, for a given set of reactants in a given reaction, have been used for the computation of the field-equivalent-solvents, and this procedure has been applied for all the bonds that belong to reaction coordinates in a chosen chemical reaction. In principle, the field-equivalent-solvent is the electric field, which is when applied in absence of solvent molecules, exerts the same force (of same magnitude and direction) on the given subsystem that had been exerted by the explicit solvent molecules. Thus, the field-equivalent-solvent has been proposed to be used as a proxy to obtain the long range influence of solvent molecules on reaction barriers without considering the solvent molecules explicitly in the calculations. A detailed description of the field-equivalent-solvent and how it may be calculated and implemented in the calculations for different bonds in different conditions of bonding for any given chemical transformation has been provided in subsequent subsections.

6.3.1 Field-equivalent-solvent

The field-equivalent-solvent has been defined here as the electric field, which when applied will create the same effect on the barrier that is created by the actual solvent molecules. It is very important to note here that the field-equivalent-solvent is different from the actual electric field created by the solvent molecules along a chosen bond. An electric field due to a charged particle at any point in space is conceptually defined as

$$\vec{E} = \lim_{q_0 \rightarrow 0} \left(\frac{\vec{F}}{q_0} \right) \quad (6.2)$$

where \vec{F} is the electrostatic force experienced by a reference positive point test charge q_0 under the condition $q_0 \rightarrow 0$. In the current scenario, the electrostatic force \vec{F} exerted by the solvent molecule on the reference positive point test charge q_0 at the

bonding points (positions of the atoms that make the bond), would be considered for the calculation of the electric field \vec{E} . When this \vec{E} will be applied along a chemical bond, the subsystems belonging to that bond, say each having equivalent charge q_e , will experience the force $q_e \cdot (\vec{F}/q_0)$, which is different from the actual force (which is \vec{F}) applied by the solvent molecules on the subsystems, and hence will lead to a different amount of stabilization to that bond. Thus, \vec{E} cannot be applied as a proxy for the actual solvent molecules. It must also be noted here that the available quantum chemical softwares provide the values of \vec{E} at different points in a given geometry. The field-equivalent-solvent for a chosen subsystem belonging to a bond in a given system is, however, defined in this study as,

$$\vec{E}_e = \vec{F}/q_e \quad (6.3)$$

so that when \vec{E}_e is applied on the subsystems, it will exert the force $q_e \cdot \vec{E}_e = \vec{F}$ on them, which is the same force that had been applied by the actual solvent molecules. However, there is no exact way to calculate the field-equivalent-solvent. Herein, we have proposed an approximate approach to calculate the field-equivalent-solvent for every chosen bond that belongs to the reaction coordinate for a given chemical reaction.

In order to calculate the field-equivalent-solvent specific to a bond, each bond has been considered to be made up of two subsystems, which have been joined together by this bond. It has been considered that any alteration in the electrostatic force experienced by these subsystems due to the presence of explicit solvent molecules or any external factor (say, the electric field) will lead to a change in the bond length, and hence the relative stabilization of that bond. Based on the structure of the reacting molecules, there could be two possible situations in which any chosen bond (whether it is already existing in any of the given reactants or going to be formed for the first time in the transition state structure during the course of the product formation) may be present: (i) the bond is part of a cyclic ring or (ii) the bond is not a part of any cyclic ring. This point has been illustrated in Figure 6.2 below. The reaction between a proline enamine (an intermediate in the proline catalyzed

asymmetric Aldol reaction) and benzaldehyde has been considered here for this purpose. Figure 6.2a showcases the direction in which the electron density gets polarized during the course of the reaction, which elucidates the fact that the bonds are part of the reaction coordinate in the enamine intermediate. Figure 6.2b and 6.2c showcase the specific examples of bond type (ii) and (i) respectively in the enamine. As illustrated in Figure 6.2b, identifying subsystems in situation (ii) is easy and they can be obtained by recognizing two groups (present at either side of the bond) that make this bond. This situation resembles the type of the hydrogen bonded systems that have been shown in Figure 5.3b of Chapter 5. However, identification of two subsystems in situation (i) is not trivial. In order to obtain the two subsystems for a bond that is a part of a cyclic ring in a reacting species, the strategy that has been applied here is to divide the entire molecular space of that reactant into three regions by drawing two parallel planes that are perpendicular to this bond and that pass through two atoms that make this bond, as shown in Figure 6.2c. This situation resembles the type of hydrogen bonded system that has been shown in Figure 5.3a of Chapter 5. The two outer regions have been considered as the two subsystems in this case (Figure 6.2c). This is because any perturbation in the electrostatic force experienced by atoms of these two regions due to external effects will directly lead to a change in the binding parameters of the related bond, which is represented by these subsystems. It has been assumed that the perturbation in the electrostatic force experienced by atoms in the middle region due to external factors will not directly lead to any change in the binding parameters of the bond type (i), until unless the external parameters do not lead to a change in the value of atomic charges. The electrostatic effect of all the atoms that fall in the middle region on the bonding parameter of (i) has been, therefore, included in calculation of the net force experienced by the two subsystems while calculating the field-equivalent-solvent for this bond, as an atom of this region will also be interacting with atoms of the two outer regions in a given geometry. It has been considered that there will be mutual electrostatic interaction between every atom of subsystem 1 with every atom of subsystem 2, in both, in the presence as well as in the absence of the explicit solvent molecules. Notably, mutual electrostatic interactions within subsystems for any bond type have not been considered for the calculation of the field-equivalent-solvent, as these interactions may not affect the stabilization of the chosen bond. The rest of the

neighboring molecules (or the atoms in the middle region of bond type (i)) may interact simultaneously with the atoms of both the subsystems, and, hence, may affect the electrostatic stabilization of the chosen bond. Therefore, electrostatic interactions due to atoms of neighboring systems, whenever present, have been accordingly taken into account while calculating the net binding force experienced by subsystems along the chosen bond in a given scenario. A schematic explanation of the bonds that undergo changes and hence constitute the reaction coordinate has been provided in Figure 6.2d with the example of the self-dehydration of the carbonic acid. Bonds 1, 2, 3 and 4 in Figure 6.2d are the part of the reaction coordinate.

Once the subsystems related to every bond belonging to the reaction coordinate have been identified for both the reactants as well as the transition state structures, the field-equivalent-solvent for every subsystem and subsequently for every bond can be calculated as per the following procedure. The field-equivalent-solvent of a bond is nothing but the net field-equivalent-solvent along that bond, which may be calculated by taking the average value of the field-equivalent-solvent of the two subsystems that represent this chosen bond.

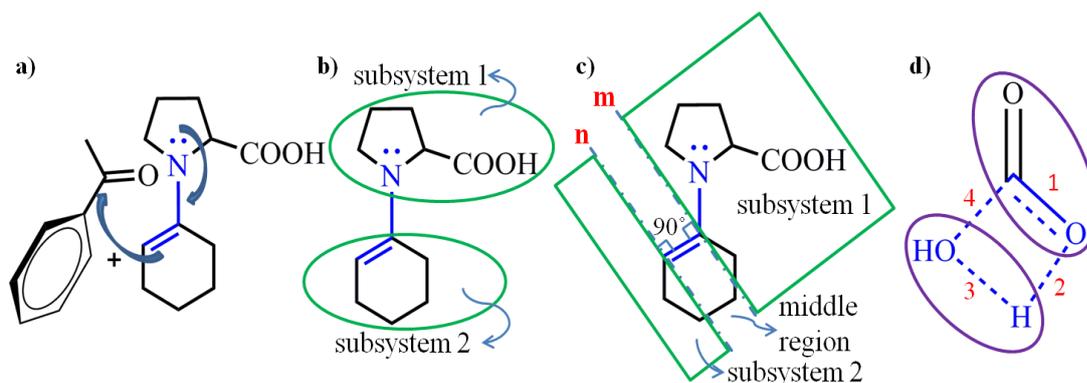


Figure 6.2 Schematic illustration of: a) the direction of electron density flow in enamine during the reaction with benzaldehyde, revealing the bonds that are part of the reaction coordinate in enamine, b) two subsystems related to a bond in a reactant species that is not the part of the cyclic ring, c) two subsystems related to a bond in a reactant species that is a part of the cyclic ring; m and n represent two perpendicular planes passing through two extreme points of this bond, and d) bonds that are part of the reaction coordinate in the self-dehydration reaction of carbonic acid.

Let us consider that the electrostatic force experienced by subsystems 1 and 2 along a chosen bond in the absence of explicit solvent molecules are \vec{F}_1 and \vec{F}_2 respectively, and the forces experienced by these subsystems in the presence of explicit solvent molecules along the same bond are \vec{F}_1' and \vec{F}_2' respectively. It is now clear that the difference of these forces in the presence and in the absence of solvent molecules on each subsystem along the chosen bond will be the contributing force from the solvent molecules in the given orientation of the solvent. It is to be noted that the values of \vec{F}_1' and \vec{F}_2' may change in a different orientation of the solvent molecules, since electrostatic force is a directional quantity. Thus, $\vec{F}_1 - \vec{F}_1'$ will be the net force applied by the solvent molecules on the subsystem 1 along the chosen bond in a chosen orientation of the solvent molecules. Similarly, $\vec{F}_2 - \vec{F}_2'$ will be the net force applied by the solvent molecules on the subsystem 2 along the chosen bond in that same orientation of the solvent molecules. Now, if we know the equivalent charges of the subsystems, we will easily get the field-equivalent-solvent for them by the help of Equation (6.2).

The problem that arises here is how the equivalent charge that will represent the electrostatic interactions of the entire subsystem will be obtained. It is reasonable to assume here that the charges on the atoms that do not take part in the reaction remain unchanged or change by a very nominal amount in the transition state structure. If we keep the orientation of the solvent molecules in the reactant and transition state structures fixed, the interactions of these charges with the solvent molecules will not differ significantly in the reactant and the transition state structures. Even the induction or charge transfer effects of the solvent molecules will perturb these charges by almost the same amount in the two structures and their mutual interactions with the solvent molecules almost be the same in the two structures, provided that the distance of these atoms from the solvent molecules remain unchanged, which can be ensured by keeping the solvent's orientation fixed in the two structures. Thus, the interactions of these charges with the solvent molecules will, in principle, not lead to differential stabilization of the two structures. However, the atoms that undergo charge rearrangement in the transition state structure will experience differential electrostatic influence of solvent molecules in the reactant and the transition state structures. These atoms are the ones that are part of the reaction

coordinate. Therefore, atoms that constitute the selected bond for which field-equivalent-solvent has to be calculated are the most significant. The charges on these atoms have been considered in this study as the equivalent charges of the respective subsystems, which belong to a chosen bond. Let us consider now that the charge on these atoms in the subsystem 1 and 2 in the absence of solvent molecules are $q_{e,1}$ and $q_{e,2}$ respectively. Since we want to estimate the electric field effect of solvent molecules as an external perturbation, the charges on these atoms in the absence of water molecules have only been considered for both reactants and the TSs. Now, the field-equivalent-solvent $\vec{E}_{e,1}$ and $\vec{E}_{e,2}$ for subsystem 1 and 2, according to the Equation (6.2), will be $\vec{F}_1 - \vec{F}_1' / q_{e,1}$ and $\vec{F}_2 - \vec{F}_2' / q_{e,2}$ respectively. The field-equivalent-solvent of the chosen bond then can be obtained as $(\vec{E}_{e,1} + \vec{E}_{e,2}) / 2$.

The electrostatic force of binding for the subsystems joined together by a bond type (ii), as described earlier in Figure 6.2b, have been calculated using the procedure that has been applied to calculate the binding force for the hydrogen bonded systems where the hydrogen bonding partners are easily recognizable, like for the case of the water dimer (please see Figure 5.2 of Chapter 5 for the algorithm of the code). While the electrostatic force of binding for the subsystems belonging to bond type (i) (see Figure 6.2c), have been calculated employing the procedure that has been applied for calculating the binding forces for hydrogen bonds in the peptidic systems (please see Figure 5.1 of Chapter 5 for the algorithm of the code). To compute the magnitude of the net force on each subsystem, the vector sum of electrostatic forces experienced by each atom of this subsystem due to all the other atoms of the entire system along the bond joining the other subsystem, has been considered. This procedure has been followed uniformly for all geometries of reactants and transition state that have been obtained in the presence or in the absence of the explicit solvent molecules.

6.3.2 Barrier in field-equivalent-solvent

Having obtained the field-equivalent-solvent corresponding to each bond that belongs to the reaction coordinate of a given reaction, the question that comes next is how to apply this concept to obtain the net barrier in the field-equivalent-solvent, which will correspond to the barrier in the actual presence of solvent molecules. It is important to

note here that the field-equivalent-solvent is, in reality, an electric field, whose dimension and other properties are exactly same to that of the electric field. The only difference between the electric field and the field-equivalent-solvent is in the way they are calculated, which leads to a difference in their magnitude for a given bond. Therefore, the field-equivalent-solvent is the electric field, whose magnitude is different in comparison to the actual electric field created by the solvent molecules along a chosen bond. Thus, in the subsequent studies, we have applied the electric field having magnitude equal to the field-equivalent-solvent along the bond for which it has been calculated. This field has been applied to both reactant as well as in transition state geometries for the bare (in absence of solvent) case, in fresh QM calculations, in order to obtain the barrier in the field-equivalent-solvent. This procedure has been repeated for all the bonds that belong to the reaction coordinate in a given reaction in order to obtain the corresponding barriers in the field-equivalent-solvent of each bond. Now, these barriers in the field-equivalent-solvent corresponding to every bond in the reaction coordinate have been applied in Equation (6.1) to obtain the net barrier in the field-equivalent-solvent.

The proof of Equation (6.1) is really simple, which is as follows -

According to the Eyring-Polanyi Equation, the rate constant k of a chosen reaction in bare conditions can be written as,

$$k = C.exp(-\Delta G^\ddagger / RT) \quad (6.4)$$

where ΔG^\ddagger is the free energy barrier in the bare condition.

Let us consider that the barrier becomes ΔG_1^\ddagger when an external perturbation (say an electric field) is applied exclusively along one of the bonds that is part of the reaction coordinate. Then the rate constant k_1 for this case can now be written as,

$$k_1 = C.exp(-\Delta G_1^\ddagger / RT) \quad (6.5)$$

Now, ΔG_1^\ddagger can be written as the sum of ΔG^\ddagger and a perturbation energy, say λ_1 , for this case. Therefore, k_1 can now be rewritten as

$$k_1 = C.exp(-(\Delta G^\ddagger + \lambda_1) / RT) \quad (6.6)$$

which on further simplification becomes

$$\begin{aligned} k_1 &= C.exp(-\Delta G^\ddagger / RT).exp(-\lambda_1 / RT) \\ \Rightarrow k_1 &= k.k' \end{aligned} \quad (6.7)$$

where $k' = exp(-\lambda_1 / RT)$

Similarly, when an external perturbation is applied exclusively along another bond of the reaction coordinate, the obtained rate constant k_2 can be written as,

$$k_2 = k.k'' \quad (6.8)$$

where $k'' = exp(-\lambda_2 / RT)$

When a perturbation will be applied exclusively along n^{th} bond belonging to the reaction coordinate, the rate constant k_n be

$$k_n = k.k^{n'} \quad (6.9)$$

where $k^{n'} = exp(-\lambda_n / RT)$

Similarly, when n number of external perturbations are applied simultaneously along n number of bonds that are part of the reaction coordinate, the net rate constant k_s , which is the net rate constant in field-equivalent-solvent in the current scenario, can be written as,

$$k_s = k.k'.k'' \dots k^{n'} \quad (6.10)$$

Putting values of rate constants in Equation (6.10) from previous equations

$$k_s = (k.k').(k.k'') \dots (k.k^{n'}) / k^{n-1} \quad (6.11)$$

$$\Rightarrow k_s = (k_1.k_2 \dots k_n) / k^{(n-1)} \quad (6.12)$$

$$\Rightarrow C.exp(-\Delta G_s^\ddagger / RT) = [C.exp(-\Delta G_1^\ddagger / RT). C.exp(-\Delta G_2^\ddagger / RT)...C.exp(-\Delta G_n^\ddagger / RT)] [C^{n-1}.exp(n-1)(-\Delta G^\ddagger / RT)$$

$$\Rightarrow \Delta G_s^\ddagger = \Delta G_1^\ddagger + \Delta G_2^\ddagger + \dots + \Delta G_n^\ddagger - (n-1).\Delta G^\ddagger \quad (6.1)$$

which is the desired Equation.

In the derivation of Equation (6.1), the external electric field (field-equivalent-solvent) specific to a chosen bond have been considered to be applied exclusively along that bond. However, this condition cannot be ensured in a chemical system, as all the bonds that are part of the reaction coordinate are not perpendicular to each other. Therefore, when an electric field is applied along a bond, its component along other bonds will come into the effect as well, which will lead to an error in the calculation of net barrier in field-equivalent-solvent. In order to minimize this error, the strategy that has been employed in this study has been to find the principle reaction coordinate. The principle reaction coordinate is identified as the bond that undergoes the maximum amount of stabilization (relative stabilization of the bond in the transition state with respect to the reactants) or destabilization per unit of the electric field, which has been calculated as difference in barriers (in field-equivalent-solvent of a chosen bond and bare case) divided by the field applied along that bond. A component of electric fields applied along the rest of the bonds of the reaction coordinate have been calculated along this bond and have further been applied along this to get the related barriers, which have further been applied together in equation (6.1) to get the net barrier in the field-equivalent-solvent.

6.4 Results and Discussion

In this theoretical and computational study, we have aimed to estimate and quantify the impact of the long range electrostatic influence of explicit solvent molecules on the rate determining barriers (RDBs). As long range electrostatic interactions have directional property, their effect on the reaction barriers have been considered in the form of field-equivalent-solvent that is created by the solvent molecules on all the bonds that belong to the reaction coordinate for a given reaction. Depending upon the orientation of the solvent molecules, different values of field-equivalent-solvent may

be experienced by different bonds in a given reaction. Hence they will get differentially stabilized in the presence of solvent molecules. The bonds that do not undergo charge reorganization will attain the same extent of stabilization in the transition state as well as in the reactant structures, provided that the orientation of the solvent molecules do not get changed in the reactant and the transition state structures. All the bonds that undergo charge reorganization in the transition state will attain different amount of stabilization for the transition state and reactant structures, which will lead to differential stabilization of the two structures, and hence will alter the reaction barrier. The net barrier obtained in the presence of the explicit solvent molecules includes this directional (long range electrostatic) influence of the solvent molecules, whose magnitude and direction may vary from point to point in the molecular space of the reacting species. If the above hypothesis is true, the barriers obtained in the explicit solvent molecules should have some directional dependence, which can be obtained by varying the orientation of solvent molecules, while keeping it fixed for the corresponding reactant and transition state structures. In order to examine this hypothesis, we have chosen six different kinds of reactions, where different solvent systems have been employed. Reactions with different solvent systems have been chosen specifically to verify the generality of the directional influence of the solvent. The reactions that have been chosen belong to two different classes of catalysis: (i) organocatalysts and (ii) organometallic catalysts. Overall, four different kinds of reactions in four different solvent systems have been considered: (i) the proline catalysed Aldol reaction in DMSO, studied by Hauk and co-workers,³⁴ (ii) the histidine catalysed Aldol reaction in water, studied by Hauk and co-workers,³⁵ (iii) the proline catalysed Michael addition reaction in methanol, studied by Patil *et al.*¹² and (iv) the hydrogen insertion step in the hydroformylation reaction in toluene, studied by Dangat *et al.*³⁶ The optimized transition state geometries pertaining to all six cases in the absence of solvent molecules have been shown in Figure 6.3 below. For all of these cases (except one for which a smaller sample space has been searched due to a very small reaction barrier in the bare condition, 3.0 kcal/mol) we have obtained more than 20 structures of transition states and reactants by varying the orientation and the number of the solvent molecules, while keeping these variants fixed for every pair of transition state and reactant structures. Only a slight distortion in the position and orientation of the solvent molecules in the transition state and the

corresponding reactant structure have been observed for all the chosen cases, which is inevitable because of the difference in the compactness of the two structures. The transition state structures are generally more compact for all the intramolecular reactions. All the calculations for every case have been carried out employing the CPCM implicit solvent model under the influence of the dielectric constant of the corresponding solvent system for every reaction. It is very important to note that only the rate and selectivity determining barriers for these cases, which have been already reported employing full QM calculations, have been considered. A range of different values of barriers have been obtained for each of the selected cases, which has been summarized briefly in Table 6.1. The entire landscape of reaction barriers for the case shown in Figure 6.3d have been provided in Table 6.2 on two different levels of theory. The difference in the highest and the lowest values of the barrier for all of these cases have been found to be more than 4 kcal/mol (last column of Table 6.1),

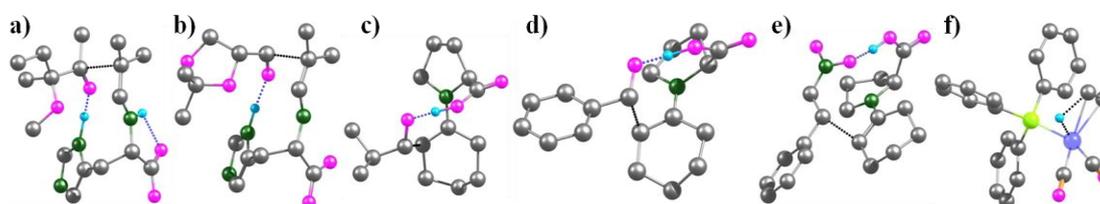


Figure 6.3. The rate and selectivity determining transition state geometries for six different reactions that have been considered in this study, to look into the effect of the orientation and the number of the solvent molecules in the first coordination shell on the barrier height. The geometries for (a) to (e) have been obtained at the CPCM(solvent)/M06-2X/6-31G** level of theory and for (f) at the CPCM(solvent)/M06/6-31G** level of theory, since (f) contains the rhodium atom in its structure. The solvents for (a) and (b), (c) and (d), (e) and (f) are water, DMSO, methanol and toluene respectively. The rate and selectivity determining transition state structures have been taken from previous QM studies by different groups. Colour representation: black – carbon, cyan – hydrogen, pink – oxygen, green – nitrogen, yellow green – phosphorus, blue – rhodium, dotted blue lines – hydrogen bonds, and dotted black lines – bonds representing the transition state. Hydrogen atoms that are not involved in hydrogen bonding or the transition state have been removed for the purposes of clarity.

Table 6.1 A brief summary of DFT calculations for the rate and selectivity determining step of six different chemical reactions that have been chosen for the evaluation of the effect of orientation and the number of the solvent molecules in the first coordination shell on the rate and selectivity determining barriers.

Reaction specification	bare	lowest	highest	highest-lowest
Histidine catalyzed Aldol reaction of 3-methoxy-3-methyl-2-Pentanone in water	13.3	11.9	16.6	4.7
Histidine catalyzed Aldol reaction of Isopropylidene-glyceraldehyde (R) in water	3.0	0.4	5.8	5.4
Proline catalyzed Aldol reaction of 2-methylpropanone in DMSO	6.2	3.0	8.1	5.1
Proline catalyzed Aldol reaction of benzaldehyde in DMSO,	7.7 (6.6)	2.6 (1.4)	8.4 (7.7)	5.8 (6.3)
Hydrogen insertion step of Hydroformylation reaction in toluene	10.1	5.3	11.5	6.2
Proline catalyzed Michael addition reaction of Nitrostyrene in methanol	13.6	10.8	14.9	4.1

bare = barrier in absence of solvent molecules.

highest = highest barrier in the presence of solvent molecules.

lowest = lowest barrier in the presence of solvent molecules.

highest - lowest = the difference in the highest and the lowest barrier values.

The values inside parentheses for one given case are at the CPCM (DMSO)/B3LYP6-D2/6-31g** level of theory. The rest of the values are at the CPCM (DMSO)/M062X/6-31g** level of theory.

which is very high and accounts for more than 99% of enantioselectivity when the energy of the corresponding diastereomeric transition state structures differ by this amount.

Having established that the different orientations of explicit solvent molecules affect the reaction barriers differentially for all kinds of reactions irrespective of the solvent system being employed, we chose eight different sample cases out of the above specified structures in the presence of solvent molecules for all cases, for further studies. The cases having the reaction barriers at the interval of ~ 2.0 kcal/mol only have been chosen for further studies (Table 6.3 below), considering the accuracy of the DFT calculations. This has been done specifically keeping in mind that multiple DFT calculations may be needed to be performed in-field-equivalent-solvent for any chosen reaction along all the bonds belonging to reaction coordinates. All the covalent bonds near the reaction sites, with bond lengths differing by more than 0.002 \AA in the transition state structure (with respect to the corresponding bond in the reactant structure) have been considered to be part of the reaction coordinate. A total of six to eleven such bonds have been recognized in different cases considered here. The optimized transition state geometries for these eight cases have been shown in Figure 6.4 below. Then, the field-equivalent-solvents for all of these bonds in all geometries have been calculated according to the procedure explained in Section 6.3.1. Further, the barriers in the field-equivalent-solvent of each bond that is part of the reaction coordinate in each system (the transition state and reactant structures in corresponding field-equivalent-solvent could not be located along N-H line for 4d), and then the net barrier in field-equivalent-solvent for each system has been calculated according to the procedure described in Section 6.3.2. However, only the electronic energies (although the derivation of the Equation (6.1) has been shown employing free energy barriers) at 0 K have been considered for further analysis in all these cases. This is because the barriers in all these cases have been calculated with respect to the reactant structure in which two reactants have been put together in one geometry. This procedure reduces the effect of entropy and zero point energies in the calculation of free energy barriers. Therefore, the values of energy that have been obtained by direct quantum chemical treatment of the above system only have been considered to evaluate the efficacy of our approach.

A graph with the net barrier in the field-equivalent-solvent on the Y axis and the corresponding barrier in presence of actual solvent molecules on the X axis is shown in Figure 6.5. Gratifyingly, we obtained an excellent linear correlation between

the barrier heights obtained employing two methods, which suggests that the quantitative trend in barrier heights obtained employing our approach is maintained.

Table 6.2 A summary of DFT calculations for the rate and selectivity determining barriers of 3d under different conditions of explicit solvent molecules.

	1		2		3		4		5		6	
	ΔE^\ddagger	ΔG^\ddagger										
1DMSO	7.8 (5.2)	10.4 (5.7)	6.6 (5.8)	8.0 (5.4)	6.9 (8.5)	8.2 (8.8)	6.6 (5.6)	6.0 (5.8)	8.5 (7.7)	10.7 (7.2)	6.6 (5.3)	7.7 (5.1)
2DMSO	5.8 (4.7)	7.7 (3.9)	7.1 (7.1)	8.0 (6.8)	8.1 (6.9)	7.6 (5.9)	6.4 (5.2)	7.7 (4.8)	7.2 (5.9)	6.1 (6.0)	7.5 (6.5)	7.7 (5.4)
3DMSO	6.5 ()	5.5 (.)	7.9 (7.6)	8.1 (8.1)	7.7 (6.8)	9.6 (5.6)	5.3 (4.3)	6.6 (3.7)	5.5 (12.7)	5.6 (9.5)	6.3 (4.3)	6.8 (6.7)
4DMSO	6.6 (5.2)	7.2 (3.7)	5.8 (4.4)	5.9 (4.5)	7.6 (5.9)	6.7 (5.8)	6.5 (5.4)	8.3 (5.4)	7.3 (3.3)	10.1 (2.8)	4.2 (4.1)	5.6 (3.3)
5 DMSO	4.4 (3.3)	7.6 (5.2)										
6 DMSO	4.1 (3.0)	6.3 (2.4)										
8 DMSO	2.6 (1.4)	3.3 (0.6)										

Values outside and inside the parentheses are at the CPCM(DMSO)/M06-2X/6-31G** and CPCM(DMSO)/B3LYP-D2/6-31G** levels of theory respectively.

1, 2, 3, 4, 5 and 6 represent different orientations of solvent molecules around the reactant and transition state structures.

ΔE^\ddagger and ΔG^\ddagger are electronic and free energy barriers respectively.

This shows the efficacy of our approach, which further suggests that our approach could be applied to get the corresponding barrier in the presence of explicit solvent

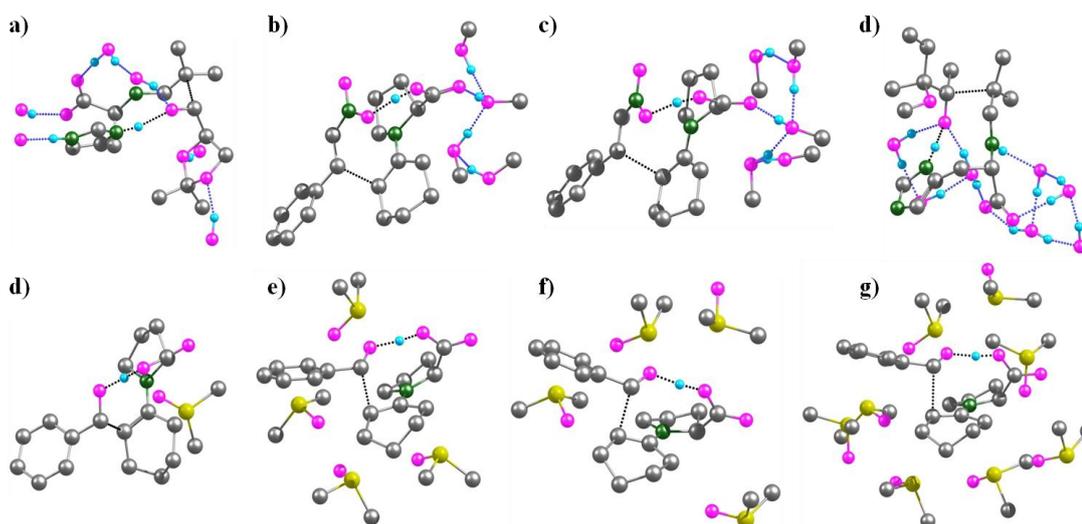


Figure 6.4. The transition state geometries of eight chosen cases in the presence of different solvent molecules specific to each system for which the field-equivalent-solvent and the corresponding barriers have been considered for evaluating the efficacy of our approach. Colour representation: black – carbon, cyan –hydrogen, pink – oxygen, green – nitrogen, yellow – sulphur, dotted blue lines – hydrogen bonds, and dotted black lines – a near single bond that is going to be broken or formed in the product. Hydrogen atoms that are not involved in hydrogen bonding or in the transition state have been removed for the purpose of clarity.

Table 6.3 The net barriers in the field-equivalent-solvent and the corresponding barriers in the presence of the actual solvent molecules for eight different chosen cases, as shown in Figure 6.4, at the CPCM(solvent)/M06-2X/6-31G** level of theory.

	4a	4b	4c	4d	4e	4f	4g	4h
ΔE^\ddagger	0.4	2.6	4.2	6.5	8.5	10.9	11.9	14.9
ΔE_s^\ddagger	2.3	7.2	8.3	8.3	8.5	16.8	13.0	18.1

ΔE^\ddagger = the barrier in the presence of the actual solvent molecules.

ΔE_s^\ddagger = net barrier in the field-equivalent-solvent.

molecules even without using them explicitly into the QM calculations, as adding actual solvent molecules in full QM calculations add many other disadvantages.

One of the reasons why exact correlation between the two barriers has not been obtained is that the NBO charges do not represent the exact atomic charges in a given system. Further studies that showcase the direct applications of this approach in order to obtain deeper insights into the reaction mechanism are in progress.

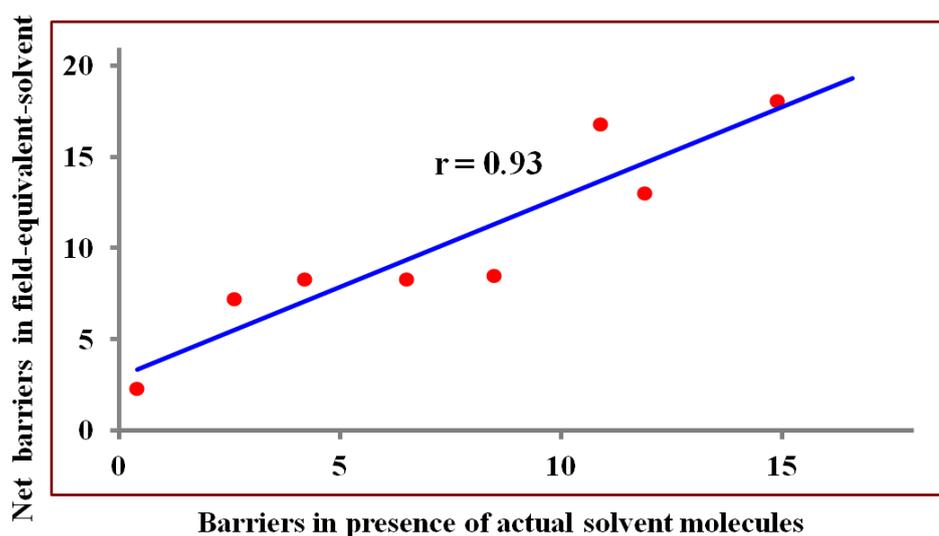


Figure 6.5 A Pearson correlation graph between the net barriers obtained in the field-equivalent-solvents and the corresponding barriers in the presence of the actual solvent molecules for the eight chosen sample cases. The field-equivalent-solvents for the chosen cases have been obtained by employing NBO charges. The obtained linear correlation between the barriers employing two methods suggests that the quantitative trend between the net barriers obtained in the field-equivalent-solvent and the barriers obtained in the presence of the actual solvent molecules is maintained for these cases.

6.5 Conclusions

An approach to compute the reaction barriers in the presence of solvent molecules without using them explicitly has been proposed herein. The approach employs consideration of the field-equivalent-solvent experienced by every bond that belongs to the reaction coordinate of a given reaction, which can be calculated by considering the atoms in molecules as point charges, the simplest method that could be employed

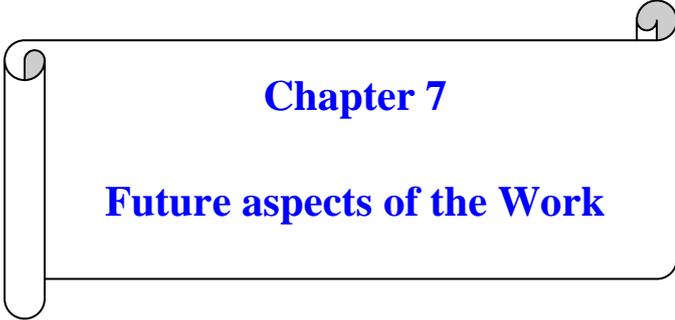
for the calculation of the electrostatic force of binding for a given bond. An excellent linear correlation between the net barriers obtained in the field-equivalent-solvent and the corresponding barriers in the presence of actual solvent molecules have been obtained for eight sample cases, which shows the efficacy of our approach. Although this method has been described by employing the field-equivalent of solvent molecules, it has general significance and can be exploited to compute the effect of any molecule or molecular fragment on the barrier heights in terms of their corresponding field-equivalent in QM calculations.

6.6 References

1. D. E. Hughes, C. K. Ingold, *J. Chem. Soc.*, 1935, 244.
2. A. J. Parker, *Chemical Reviews*, 1969, **69**, 1-32.
3. C. Reichardt, *Chem. Rev.*, 1994, **94**, 2319; C. Reichardt, VCH: Weinheim, 1998.
4. H. B. Kagan, O. Riant, *Chem. Rev.*, 1992, **92**, 1007-1019.
5. B. Subramaniam, M. A. McHugh, *Ind. Eng. Chem. Process Des. Dev.* 1986, **25**, 1-12.
6. D. L. Hughes, *Organic Preparations and Procedures International*, 1996, **28**, 127-164.
7. H. Yorimitsu, T. Nakamura, H. Shinokubo, K. Oshima, K. Omoto, H. Fujimoto, *J. Am. Chem. Soc.*, 2000, **122**, 11041-11047.
8. F. Proutiere, F. Schoenebeck, *Angew. Chem. Int. Ed.*, 2011, **50**, 8192–8195.
9. Y. Inoue, *Chem. Rev.*, 1992, **92**, 741-770.
10. Ž. Knez, M. Habulin, *The Journal of Supercritical Fluids*, 2002, **23**, 29-42.
11. Singh *et al.*, *Inorg. Chem.*, 2015, **54**, 1535.
12. M. P. Patil, R. B. Sunoj, *Chem. Eur. J.* 2008, **14**, 10472.
13. J. Chandrasekhar, S. Shariffskul, W. L. Jorgensen, *J. Phys. Chem. B*, 2002, **106**, 8078-8085.
14. C.-H. Lim, A. M. Holder, C. B. Musgrave, *J. Am. Chem. Soc.*, 2013, **135**, 142–154.
15. S. Jain, K. Vanka, *Chem. Eur. J.*, 2017, **23**, 13957 – 13963.
16. E. Y.-X. Chen, T. J. Marks, *Chemical reviews*, 2000, **100**, 1391-1434.
17. J. R. P. Jr, J. M. Riveros, *J. Phys. Chem. A*, 2001, **105**, 7241-7247.

18. J. Kang, J. Rebek Jr, *Nature*, 1997, **385**, 50-52.
19. M. S. Rogers, E. M. Tyler, N. Akyumani, C. R. Kurtis, R. K. Spooner, S. E. Deacon, S. Tamber, S. J. Firbank, K. Mahmoud, P. F. Knowles, S. E. V. Phillips, M. J. McPherson, D. M. Dooley, *Biochemistry*, 2007, **46**, 4606-4618; S. J. Benkovic, S. Hammes-Schiffer, *Science*, 2003, **301**, 1196-1202; B. A. Springer *et al.*, *Chem. Rev.* 1994, **94**, 699; Y. Lu, J. S. Valentine, *Curr. Opin. Struct. Biol.*, 1997, **7**, 495.
20. D. M. Kurtz, *Chem. Rev.*, 1990, **90**, 585, and references therein; M. Mukai *et al.*, *J. Am. Chem. Soc.*, 1997, **119**, 1758; Y.-D. Wu *et al.*, *Inorg. Chem.*, 1992, **31**, 718; C. E. MacBeth, A. P. Golombek, V. G. Young, C. Yang, K. Kuczera, M. P. Hendrich, A. S. Borovik, *Science*, 2000, **289**, 938.
21. S. F. Sousa, A. J. M. Ribeiro, R. P. P. Neves, N. F. Brás, N. M. F. S. A. Cerqueira, P. A. Fernandes, M. J. Ramos, *WIREs Comput. Mol. Sci.*, 2017, **7**, e1281; M. G. Quesne, T. Borowski, S. P. de Visser, *Chem. - Eur. J.*, 2016, **22**, 2562–2581; K. Swiderek, I. Tunon, V. Moliner, *WIREs Comput. Mol. Sci.*, 2014, **4**, 407–421; M. W. van der Kamp, A. J. Mulholland, *Biochemistry*, 2013, **52**, 2708–2728; C. Rovira, *WIREs Comput. Mol. Sci.*, 2013, **3**, 393–407.
22. F. Himo, *J. Am. Chem. Soc.*, 2017, **139**, 6780–6786.
23. S. C. L. Kamerlin, A. Warshel, *WIREs Comput. Mol. Sci.*, 2011, **1**, 30–45.
24. S. R. Gadre, R. N. Shirsat, A. C. Limaye, *J. Phys. Chem.*, 1994, **98**, 9165-9169.
25. S. Shaik, D. Mandal, R. Ramanan, *Nature Chemistry*, 2016, **8**, 1091-1098, and references therein.
26. Aragonès *et al.*, *Nature*, 2016, **531**, 88; Fried *et al.*, *Science*, 2014, **346**, 1510-1513; Shaik *et al.*, *J. Am. Chem. Soc.*, 2004, **126**, 11746-11749.
27. M. Orozco, F. J. Luque, *Chem. Rev.*, 2000, **100**, 4187–4225; R. Cammi, R. Fukuda, M. Ehara, H. Nakatsuji, *The Journal of Chemical Physics*, 2010, **133**, 024104; F. Hirata, Ed. *Molecular theory of solvation*; Kluwer Academic Publishers: Dordrecht, Netherlands, 2003; M. V. Fedorov, A. A. Kornyshev, *Molecular Physics*, 2007, 105, 1–16; T. L. Beck, M. E. Paulaitis, L. R. Pratt, *The potential distribution theorem and models of molecular solutions*; Cambridge University Press: New York, 2006.
28. F. Eckert, A. Klamt, *AIChE Journal*, 2002, **48**, 369-385.

29. N. Matubayasi, *The Journal of Chemical Physics*, 1996, **104**, 5265; G. Kotliar, *Reviews of Modern Physics*, 2006, **78**, 865-951; M. D. Whitmore, J. D. Vavasour, *Macromolecules*, 1992, **25**, 2041-2045.
30. Zhao, Y.; Truhlar, D.G. *Theor. Chem. Acc.*, 2008, **120**, 215-241; Zhao, Y.; Truhlar, D.G. *J. Phys. Chem. A*, 2006, **110**, 13126-13130; Zhao, Y.; Truhlar, D.G. *J. Chem. Phys.*, 2006, **125**, 194101.
31. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, Gaussian09 Revision A.01, Gaussian Inc., Wallingford, CT, 2009.
32. V. Barone, M. Cossi, *J. Phys. Chem. A.*, 1998, **102**, 1995-2001; b) M. Cossi, N. Rega, G. Scalmani, V. Barone, *J. Comput. Chem.*, 2003, **24**, 669-681.
33. A. E. Reed, R. B. Weinstock, F. Weinhold, *J. Chem. Phys.*, 1985, **83**, 735-46.
34. S. Bahmanyar, K. N. Houk, H. J. Martin, B. List, *J. Am. Chem. Soc.*, 2003, **125**, 2475-2479.
35. Y.-h. Lam, K. N. Houk, U. Scheffler, R. Mahrwald, *J. Am. Chem. Soc.*, 2012, **134**, 6286-6295.
36. Y. Dangat, M. A. Rizvi, P. Pandey, K. Vanka, *Journal of Organometallic Chemistry*, 2016, **801**, 30-41.



Chapter 7
Future aspects of the Work

Chapter 7

Future aspects of the Work

Abstract: In this chapter, we have shown other miscellaneous examples where the consideration of the electrostatic force of binding in determining the strength of noncovalent interactions has useful implications in obtaining deeper insights into the binding behavior of important chemical and biological systems. Three different examples from three diverse areas have been chosen in order to further showcase the generality and wide applicability of the electrostatic force analysis approach that we have proposed earlier in this thesis work.

7.1 Introduction

It has long been established that the long range electrostatic interactions are significant in determining properties of important chemical systems.¹ Recently, we have proposed a method that describes the directional nature of these interactions.² The method determines the electrostatic force of binding employing atoms in molecular systems as point charges, which may be determined by quantum chemical (QM) calculations (see Chapter 3 for details). The efficacy of this approach has been in its simple applications and wider relevance in diverse areas of chemistry and biology and material sciences. The robustness of the method is in its application employing any kind of charge analysis method and with all kinds of QM packages. In Chapters 4-6, we have shown different useful applications of this method. To further explain the viability and generality of this approach, in this chapter, we have investigated three different problems corresponding to three different areas of chemistry and biology where it can be employed successfully to obtain deeper insights into deducing reaction mechanisms of important chemical systems and/or in the comprehensive understanding of the binding behavior of biological systems. The questions that had been attempted to be answered to some extent in this chapter are as follows:

(i) Do long range electrostatic interactions induce some sort of selectivity in chemical transformations?

- (ii) How do long range electrostatic interactions influence the mechanism of a chemical transformation?
- (iii) What causes the differential stability of protein secondary structures such as α -helices and β -sheets?

The importance of the directional nature of long range electrostatic interactions in inducing selectivity in a chemical transformation has been explained by considering two examples of S_N2 reactions. The influence of long range electrostatic interactions on the reaction pathway of a physicochemical phenomenon that had been reported in 2009 by Wiemann *et al.*³ has been explained by employing simple model systems containing 18-crown-6, dimethylether, 2,5,8-trioxynonane and monoprotated ethylenediamine. And, the differential stability of protein secondary structural motifs has been explained by employing small model or chopped peptidic systems.

7.2 Computational Details

All the DFT calculations were carried out using the Turbomole 6.4 suite of quantum-chemical programs in the solvent phase employing the Conductor-like Screening Model (COSMO)⁴ using acetone (20.493), methanol ($\epsilon = 32.613$) and water ($\epsilon = 78.5$) as solvents.⁵ Geometry optimizations were performed employing the PBE functional.⁶ The electronic configuration of the atoms was described by a triple-zeta basis set augmented by a polarization function (TURBOMOLE basis set TZVP).⁷ The resolution of identity (RI),⁸ and the multipole accelerated resolution of identity (MARI-J)⁹ approximations were employed for an accurate and efficient treatment of the electronic Coulomb term in the density functional calculations. All the calculations were performed with DFT-D2, a general empirical dispersion correction proposed by Stefan Grimme for density functional calculations.¹⁰ Only electronic energies were considered for energy analysis of all the systems considered in this chapter. The Mulliken,¹¹ NBO¹² and ESP (electrostatic potential) fit¹³ charges have been used to calculate the electrostatic force of binding for every system. The forces of binding for every chosen hydrogen bond in hydrogen bonded systems were calculated along the line joining the donor and acceptor atoms of that particular hydrogen bond. The force of binding between two species in S_N2 reactions in a given orientation have been obtained along the line joining nucleophile (iodide) and

nucleophilic carbon atoms of the substrates. To compute the magnitude of the net force of binding, i.e., the binding force, the vector sum of the electrostatic forces experienced by each atom of every fragment due to the charges on each atom of the complementary fragment along the aforementioned line has been considered. Please refer to Chapter 3 for the detailed description of the force analysis method.

7.3 Results and Discussion

(i) **S_N2 reactions:** it is well known that in S_N2 reactions, the attack of (a generally strong) nucleophile (in a polar aprotic solvent) occurs from the opposite side of the leaving group. The formation of the carbon bond between the nucleophile and the substrate due to the creation of an σ -hole at the nucleophilic carbon atom opposite to the leaving group and at the extension of C-X (where X is a leaving group) covalent bond has been recently explained to be the reason behind this preference.¹⁴ However, the formation of the σ -hole opposite to the weakly or moderately electronegative leaving group (for example, the -SH group) may not always be possible in all RX substrates, as the carbon atom is weakly polarizable due to its smaller size. Another possible reason given behind this selectivity is the unfavorable lone pair – lone pair repulsive interactions between the leaving group and the nucleophile when the nucleophilic attack occurs from the front. However, lone pair - lone pair repulsive interactions are effective only at small separations. As the electrons in the molecular systems are not localized, the attractive dispersion interactions, are expected to dominate at the larger separations between two groups containing lone pairs of electrons. Such effects have been seen to be deterministic in the systems involving lone pair – π , and anion - π interactions.¹⁵ Furthermore, the dispersion interactions are stronger for larger size, easily polarizable strong nucleophiles, such as the iodide and the -SH group, which also favors the S_N2 mechanism over S_N1. Also, at smaller separations where lone pair – lone pair repulsions become effective, some orbital overlap between the nucleophile and nucleophilic carbon is also possible. Therefore, even the lone pair – lone pair repulsion explanation is dubious and may not hold true for every type of S_N2 system. However, preferential back-side attack is favored by almost all combinations of nucleophiles and leaving groups related to any alkylic systems.

To understand the preferential back-side nucleophilic attack in the S_N2 reactions with the viewpoint of long range electrostatic interactions, we have done a force analysis on two representative model systems containing two different substrates (CH_3F and CH_3SH) with iodide as the common nucleophile. Iodide has been especially chosen as the nucleophile for the conceptual representation of electrostatic dominated interactions between the nucleophile and the nucleophilic substrates, even at the distances smaller than the sum of van der Waal radii of carbon and iodide, as the orbital mismatch between the two is very high for the formation of a strong covalent bond. Also, the iodide is a very strong nucleophile, which is expected to favor the S_N2 pathway. The substrates in both the cases were first optimized and then the iodide ion was kept manually at around 3 Å and 3.5 Å distance from the nucleophilic carbon atoms at angles (I-C-X angle, where X = F and S) of around 60° and the 180° with both the substrates. The two different I-C-X angles have been specifically chosen to model the probable front side and the back side attack of the nucleophile respectively. The two different I-C distances (both below the sum of van der Waal radii of the I and the C atoms, ~4.0 Å, but above the C-I covalent bond length, ~2.14 Å in CH_3I) have been chosen in order to evaluate the electrostatic nature of the interactions between the nucleophile and the substrate, as it has been reported that the covalent character in the noncovalent interactions increases with the decrease in the noncovalent bond length due to stronger induction and charge transfer effects at smaller separations.¹⁶ Also, at smaller separations, the lone pair – lone pair repulsion will be more effective in these systems. All the four combinations of angles and distances have been considered to obtain the corresponding atomic charges, with all geometries belonging to both the substrates. Subsequently, the electrostatic forces of binding for both the substrates (please see Figure 7.1) in each system have been calculated. The basic assumption that has been made in this case is to that during the directional attack of the nucleophile, the structures of the substrates do not change due to the influence of the nucleophile. It is to be noted that acetone ($\epsilon = 20.493$) has been considered as the common polar aprotic solvent in all the calculations that have been performed here for this case.

The results of the force analysis have been summarized in Table 7.1 below. Table 7.1 reveals that the magnitude of the electrostatic force of binding increases significantly with decrease in the I-C distance for both the substrates and at all angles,

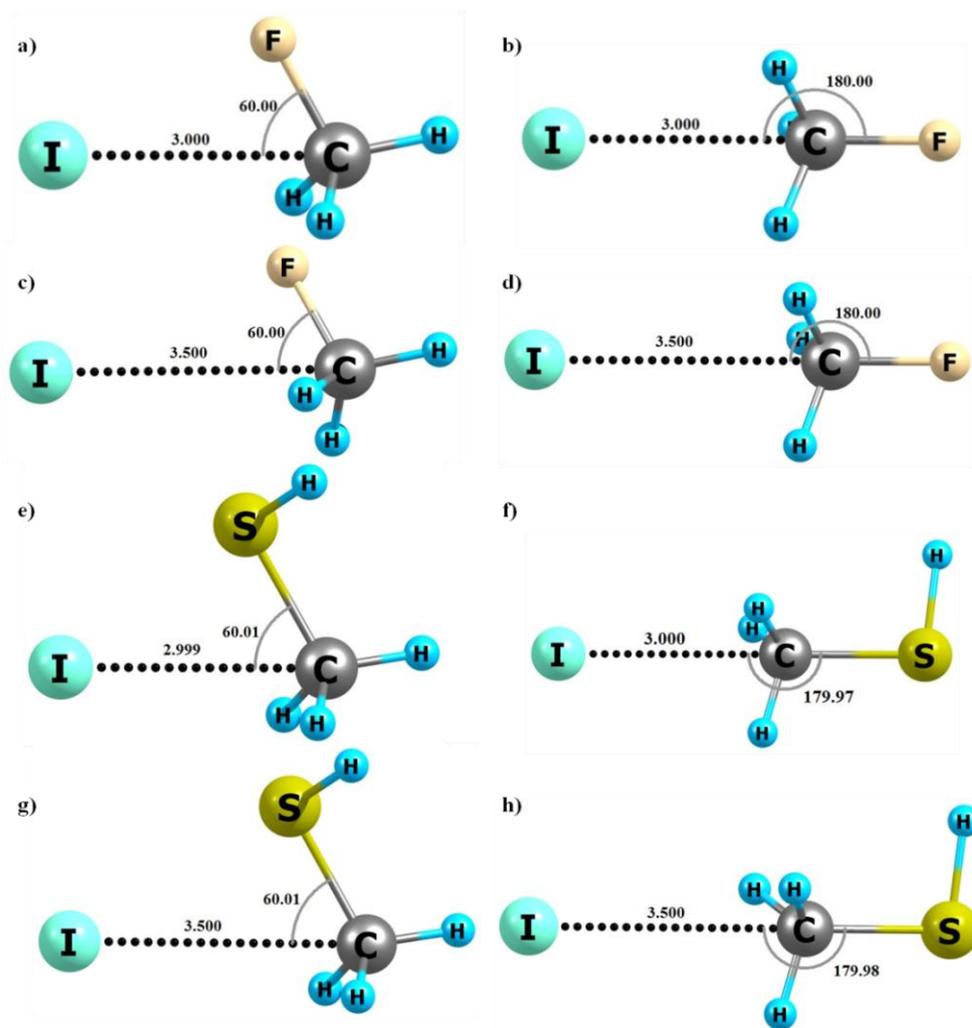


Figure 7.1 Different considered arrangements of the nucleophile with respect to two different substrates that have been considered in this study to model S_N2 reactions.

which suggests that the initial interactions between the nucleophile and the substrate is dominated by the electrostatic contributions. Further, the electrostatic forces of binding for both the substrates for all I-C distances have been found to be more favorable when the nucleophile makes an I-C-X angle of 180° . These results are consistent with different kinds of charge analyses methods. This indicates that the attack of nucleophile is guided by the long range electrostatic interactions. Thus, it can be concluded that the long range electrostatic interactions induce chemical selectivity into the chemical reactions by virtue of the directional nature of these interactions, which come into effect automatically when the multipoint electrostatic interactions between the two reacting species are taken into account.

Table 7.1 The electrostatic force of binding for the two representative S_N2 systems at the COSMO(acetone)/PBE-D2/TZVP level of theory, employing three different charge analyses methods. Mulliken, NBO and ESP fit represent electrostatic forces obtained by employing Mulliken, NBO and ESP fit charges respectively.

type of interaction		I...CH ₃ F interactions				I...CH ₃ SH interactions			
		3 Å		3.5 Å		3 Å		3.5 Å	
I-C-X (X = F, S) angle		60°	180°	60°	180°	60°	180°	60°	180°
electrostatic binding force (pN)	Mulliken	19.8	-34.3	17.6	-28.3	27.3	-19.2	22.6	-15.5
	NBO	44.7	-41.5	28.1	-36.3	57.5	5.9	31.7	-2.3
	ESP fit	-75.0	-92.0	-23.1	-42.5	-87.1	-100.1	-35.3	-45.0

(ii) The crown ether ammonium complexes: the ammonium cations are known to form hydrogen bonded complexes with different kinds of crown and *aza* crown ethers.¹⁷ Recently, Weimann *et al.* have reported the highly dynamic motion of crown ethers along the oligolysine peptide chain when the medium of the solution was kept slightly acidic.³ They have proved that the direct hopping of the crown ether from ammonium to ammonium or to the amine group with intermediate dissociation of the ammonium ether complex does not happen.³ They have further shown that the crown ether moves from ammonium to the amine group with a proton *via* hydrogen bonding intermediates, which proceeds *via* breaking of the first two hydrogen bonds between the ammonium and the crown ether and then the cleavage of the N-H covalent bond of the ammonium group takes place.³ The question that is pertinent here is: why is the cleavage of a N-H covalent bond preferable over the breaking of the last N-H...O hydrogen bond, when the breaking of the first two hydrogen bonds has been facile under the same experimental conditions? In order to investigate the mechanism of the crown ether transfer on the oligolysine chain, we have performed DFT calculations on monoprotonated ethylenediamine (EDA ammonium) complexes of dimethyl ether, 2,5,8-trioxynonane and 18-crown-6 ether. The optimized geometries of these hydrogen bonded complexes have been shown in Figure 7.2 below. All the calculations for every complex in this case, including the force analyses (whose results have been explained below), have been performed considering methanol as the common solvent ($\epsilon = 32.613$) for every system considered in this case. The sum of the energies of ethylenediamine (EDA) and the monoprotonated crown ether has been

found to be 29.4 kcal/mol higher, as compared to the sum of the energies of the EDA ammonium and the crown ether, which suggests that the equilibrium will be highly shifted towards the protonated amine moiety. Further, the binding energy of the ammonium-crown complex (Figure 7.2c) has been obtained to be favourable by 28.9 kcal/mol (see Section 1.6 of Chapter 1 for the definition of the binding energy). Thus, the sum of energies of the protonated crown and the EDA is 58.3 kcal/mol higher than the energy of the ammonium crown complex. This explains that the direct hopping of the crown *via* intermediate breaking of the complex **c** (in Figure 7.2) through all possible routes is thermodynamically prohibited at the experimental temperature, 40°C, which is in accordance with the experimental findings of Wiemann *et al.*

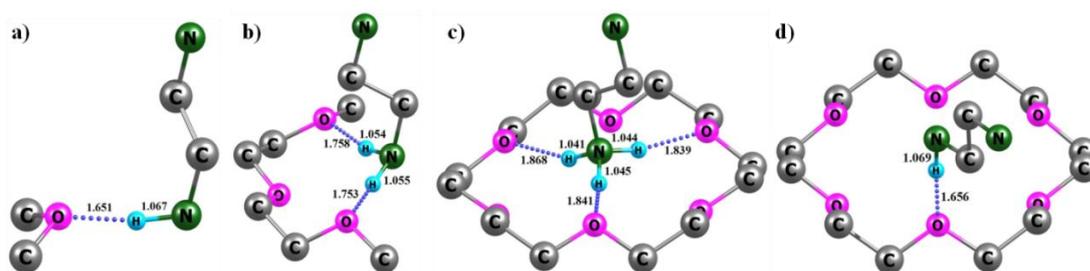


Figure 7.2 The optimized geometries at the COSMO(methanol)/PBE-D2/TZVP level of theory revealing hydrogen bonding parameters of different hydrogen bonded complexes of ammonium with: a) dimethyl ether, b) 2,5,8-trioxynonane, c) 18-crown-6 ether in tricoordination and d) 18-crown-6 ether in monocoordination. The hydrogen atoms which are not the part of hydrogen bonding are omitted for the sake of clarity. The average covalent bond length of nonhydrogen bonded N-H groups, which have not been shown in this figure, in (a), (b) and (d) are 1.027, 1.025 and 1.029 Å respectively.

A comparison of the hydrogen bond lengths in complexes **a** to **c** in Figure 7.2 (henceforth, they will be referred to as **2a**, **2b** and **2c** respectively) suggest that the hydrogen bond length increases significantly with the simultaneous decrease in the corresponding N-H covalent bond length (Figure 7.2) as the molecular size of the corresponding ether partner increases, on going from **2a** to **2c**. The average hydrogen bond length in **2a**, **2b** and **2c** has been found to be 1.651, 1.756, 1.849 Å respectively (see column 3 in Table 7.2). This suggests that the hydrogen bond strength decreases continuously on going from complex **2a** to complex **2c**. However, the binding

energies of the corresponding complexes exhibit the reverse trend. The binding energies of complexes **2a**, **2b** and **2c** have been found to be -7.1, -16.3 and -28.9 kcal/mol respectively, which suggests that complex **2c** is the most stable, followed by complex **2b** and then complex **2a**. To further understand the binding behavior of these complexes, we have performed a comprehensive electrostatic force analyses on complexes **2a-2c**. The summary of the force analysis have been provided in Table 7.2 below.

The force analysis reveals that the electrostatic force of binding per hydrogen bond decreases continuously on going from complex **2a** to complex **2c**. This indicates that the force constant of the hydrogen bond in **2a** is significantly larger than the force constants of the hydrogen bonds in **2c** (the last three columns of Table 7.2), which is in accordance with the corresponding hydrogen bond length. Therefore, further stretch or cleavage of any hydrogen bond in **2c** is much easier and more facile than that in **2a**, despite having significantly greater electrostatic stabilization of the complex **2c**. In order to make a direct comparison of a singly hydrogen bonded complex with the triply hydrogen bonded complex **2c**, we have translated the EDA ammonium group in complex **2c** to the periphery of the crown ether so that it can form only one hydrogen bond with the crown ring, analogous to the complex **2a**. The optimized geometry of the newly obtained complex is shown in Figure 7.2d above (henceforth, this will be referred to as complex **2d**). The binding parameters for this complex have been tabulated in the last row of Table 7.2. The hydrogen bonding parameters (including bond lengths and the electrostatic force of binding) in **2d**, an analogous structure of **2a**, have been obtained to be nearest to the complex **2a**, and fit in the same orders that have been observed for **2a**, **2b** and **2c** previously. The binding energy of this new complex has been found to be 1.7 kcal/mol lower than the binding of energy of the complex **2a**. However, the hydrogen bond and the corresponding N-H bond lengths in **2d** have been found to be just slightly larger than **2a** (see Figure 7.2). The reason that this causes difference in the structural parameters and hence the binding forces of hydrogen bonds in **2c** in comparison to the binding parameters in **2a** or **2d** is the electrostatic pull created by charges on the atoms of the opposite side of the crown ring, which pulls the ammonium ion away from the acceptor oxygen atom in order to bring it to the center of the ring (possessing the center of symmetry) where the electrostatic potential energy is the lowest, which weakens the electrostatic force of

binding per hydrogen bond. To further prove this point, we have performed a force analysis on different fragments of the complex **2c** (considering different fragments of 18-crown-6 only) that resembles more closely the structures of the complexes **2a** and **2b**. The structures of these fragments of the complex **2c** can be obtained by removing one hydrogen atom from both termini of the ether partners in the complexes **2a** and **2b** (say **2a'** and **2b'**). The electrostatic force of binding of these fragments (**2a'**, **2b'** and **2c**) follows the same trend as followed by complexes **2a**, **2b** and **2c**. The electrostatic forces of binding of the fragment that resembles complex **2a** have been found to be -28.2, -57.4 and -8.3 pN for Mulliken, NBO and ESP fit charges respectively, whereas the average binding forces for two hydrogen bonds of the fragment that resembles the complex **2b** have been found to be -18.8, -29.2, and -4.6 pN respectively. As reported in Table 7.2, the average binding forces for hydrogen bonds in complex **2c** employing the above specified charges have been obtained to be -8.1, 8.3 and 6.6 pN respectively. This clearly indicates that the consideration of larger fragments of the 18-crown-6 causes a decrease in the electrostatic binding force for each hydrogen bond in **2c**, which is because of the electrostatic pull created by the larger envelope of the 18-crown-6. This phenomenon resembles the levitation effect in the molecular systems and suggests that the breaking of the first or second hydrogen bonds in **2c** is much easier in comparison to the third hydrogen bond, which is what have been observed experimentally.

The N-H covalent bond length corresponding to the hydrogen bond donor group in the complex **2d** is 0.025 Å larger than the average N-H bond length in the complex **2c** and the hydrogen bond length in the complex **2d** is 0.193 Å smaller than the average hydrogen bond length in the complex **2c**. The change in the covalent bond length of the non-donor N-H group in **2a**, **2b** and **2d** is very marginal (below 0.002 Å). This suggests that the cleavage of hydrogen bonds in **2c** (and **2b** as well) is easier in comparison to the hydrogen bond in **2d** or **2a**, which is what has also been suggested by our force analysis, as the electrostatic force of binding per hydrogen bond has been found to be significantly larger for the complexes **2a** and **2d** in comparison to the complexes **2b** and **2c**. Further, the donor N-H bonds in **2a** and **2d** are more stretched and hence more activated in comparison to the corresponding N-H bonds in **2b** and **2c**, suggesting that the cleavage of the donor N-H bond in **2a** and **2d** is more facile than that in **2b** and **2c**. This corroborates the experimental findings of

the Weimann *et al.*, and showcases the power of the electrostatic force analysis method that we have proposed in finding the reaction mechanism of an important physicochemical phenomenon.

Table 7.2 Hydrogen bonding parameters and the electrostatic force of binding for different hydrogen bonds in hydrogen bonded complexes of EDA ammonium with different ether moieties as shown in the Figure 7.2 at the COSMO(methanol)/PBE-D2/TZVP level of theory employing three different charge analyses methods.

System	E_b (kcal/mol)	avg. H- bond length (Å)	avg. N-H bond length (Å)	avg. binding force (pN)		
				Mulliken	NBO	ESP_fit
2a	-7.1	1.651	1.067	-45.4	-89.3	-5.4
2b	-16.3	1.756	1.055	-13.5	-19.4	-5.9
2c	-28.9	1.849	1.044	-8.1	8.3	6.6
2d	-5.4	1.656	1.069	-64.4	-138.0	-5.3

2a, 2b, 2c and **2d** are the hydrogen bonded complexes of ammonium cation with different ethers as shown in Figure 7.2.

avg. = average

N-H bond length represents the length of N-H bonds that act as the hydrogen bond donor in the given complexes

Mulliken, NBO, ESP fit represent electrostatic forces obtained by employing Mulliken, NBO and ESP fit charges respectively.

(iii) The secondary structure motif case of the proteins: the secondary structure motifs of protein are basically composed of α -helices and β -sheets. They are structural (sometimes functional as well) blocks of three dimensional tertiary structures of functional proteins and the basic structural unit of numerous structural proteins. These protein motifs are distinguished based on differential hydrogen bonding patterns involving the peptide backbone, which is specific to each of these types. The β -sheets have been considered to be structurally more stable than the α -helices and can withstand a higher temperature before denaturation.¹⁸⁻¹⁹ In order to understand the reasons for the differential stabilization of the α -helices and β -sheets, we have

performed a frequency distribution analysis of the hydrogen bond distances in the proteins containing both α -helices and β -sheets. The O...N distances corresponding to N-H...O hydrogen bonds in the crystal structures of proteins, which are reported with resolution 2.25 Å or lower numeric values only, have been considered in this study as the standard measure of the hydrogen bond length. This is because the X-ray diffraction does not precisely locate the hydrogen atoms positions, and hence coordinates of hydrogen atoms are usually not provided in to the PDB files of protein crystals. The external addition of hydrogen atoms with the aid of softwares has been avoided to obtain the H...O distances (the H...O distances represent actual hydrogen bond lengths in an N-H...O hydrogen bond) in order to avoid any discrepancies that may be caused due to the inefficiency of available approximate methods in locating the exact position of hydrogen atoms in relation to the rest of the molecule. The O...N distances ranging between 2.5 to 3.5 Å have been considered as the measure of the hydrogen bond, assuming that the magnitude of the component of the corresponding N-H bond vector along the O...N line is close to 1.0 Å. The residues following α -helical and β -sheet conformations were first isolated from the PDB files and saved in other PDB files keeping all the information of the parent PDB file intact. The O...N distances between the atoms of two residues having a difference of 2 or more in their positional residue number (the serial number of the chosen residue, which starts with the N-terminus residue and ends at the C-terminus one) provided in the PDB file have been considered for the measure of the hydrogen bond length in both of the protein motifs for any given protein. Therefore, some O...N distances might represent hydrogen bonds formed by the side chain of some hydrogen bonding residues with the atoms from the main chain. Some of the hydrogen bonds may even correspond to the 3_{10} or π -helical conformations that are mixed with the α -helices in the given PDB structure, particularly those O...N distances that lie at higher values. Figure 7.3 below shows the frequency distribution of the above specified O...N distances in the α -helices and the β -sheets conformations of two proteins, Flavodoxin and Nuclease, which have been studied experimentally at cryogenic (100K) as well as at above freezing temperature (above 273 K). The plot illustrates the frequency distribution of the two proteins at both the temperatures. The plot suggests that the hydrogen bond lengths in the β -sheets are generally smaller than the hydrogen bond lengths in α -helices for the chosen proteins. Eight other samples of proteins have been analyzed in

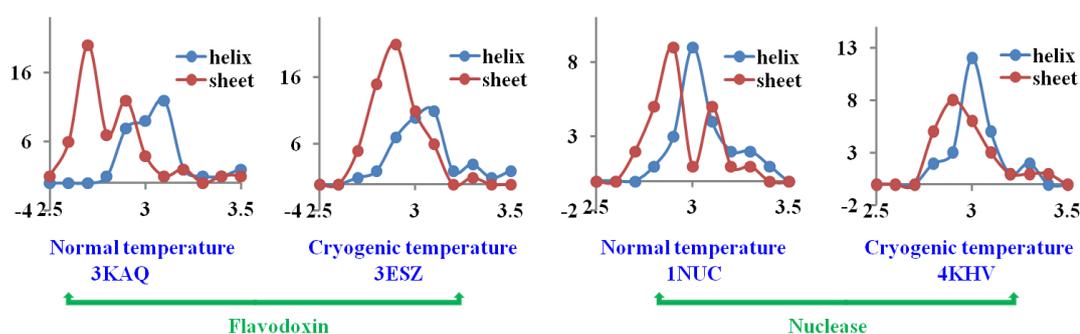


Figure 7.3 The frequency distribution of O...N hydrogen bond distances corresponding to α -helices and β -sheets in Flavodoxin and Nuclease at cryogenic and normal temperatures. The values on abscissa represent O...N distances and on the ordinate represent frequency. 3KAQ, 3ESZ, 1NUC and 4KHV are the PDB IDs of two proteins at aforementioned temperatures for which the frequency distribution of O...N distances corresponding to hydrogen bonds in helices and β -sheets have been plotted. The normal temperature denotes temperature above 273 K, *i.e.*, the melting point of water.

order to further validate this assertion. All of them were seen to follow the same trend in hydrogen bonding as described above, which indicates that the hydrogen bonds in β -sheets are generally stronger than the hydrogen bonds in α -helices. However, hydrogen bonds are weak noncovalent interactions. A little difference in hydrogen bond length from the equilibrium position may not alter the hydrogen bond strength significantly, as the potential energy surface of weak interactions are shallow. Also, the hydrogen bond length (*i.e.*, the distance between acceptor and donor atoms) is a weak descriptor of electrostatic interactions, as there can be situations in bulky molecules like proteins where the two groups that are making hydrogen bonds are held apart at larger distances due to steric or net electronic effects of other molecular parts but have very high long range electrostatic attractions between them, which is in fact the major determinant of the actual strength of a hydrogen bond in a given situation. Therefore, the measure of just the hydrogen bond length of hydrogen bonds may not always lead to a proper description of the hydrogen bond strength, as the hydrogen bonds have been reported to be predominantly electrostatic interactions. It is important to note here that there is no method available so far that can directly

measure the hydrogen bond strength of every individual bond in multi-hydrogen bonded systems, such as proteins or DNAs.

In order to further investigate this problem, we have performed force analysis on some chopped and small model peptides. Three sets of small peptides, each containing α -helical and β -sheet structures have been chosen for this analysis. These three examples have been chosen as the representative cases of the large sample space acquired by α -helices and β -sheets in proteins crystal structures that are available today. It has been assumed that the basic structure and hydrogen bonding pattern in these model peptides will remain intact during the optimization, which has indeed found to be true. All the geometries of the selected peptides from each type (α -helical and β -sheet) have been optimized at the COMSO(water)/PBE-D2/TZVP level of theory. The effect of explicit water molecules has not been included in this analysis, as it has been reported that β -sheets are more stable than α -helices in even the solid state structures of proteins, where there exist presumably very few or no water molecules trapped inside the peptide backbone. However, the electrostatic effect of the solvent has been included by employing the implicit solvent model considering water ($\epsilon = 78.5$) as the common solvent for all the peptidic structures. Two of the selected peptides of each type have been the chopped peptides, whose basic structures have been isolated by chopping off the secondary structural parts (specific to each type by a random selection) from the actual protein's PDB structures. The third helix and sheet have been selected from a class of the self contained small synthetic peptides whose crystal structures have already been reported. The assumption that has been made here is that the basic nature and pattern of noncovalent binding in these small self-contained peptides will be similar to the proteins, while the effect of long range electrostatic interaction due the entire molecules would be captured. Further, they have been specifically chosen in order to demonstrate that the differential stability of the β -sheets and the α -helices are intrinsic to their structure type and does not depend on the long range electrostatic effect exerted by one part of the protein over the other, which may happen in the case of the actual protein containing both of the structures together (with many other structural parts of proteins), as has been seen above in the cases of Flavooxin and Nuclease.

The two chopped helices containing eight and seven main chain hydrogen bonds have been obtained from the PDB structures of proteins having the PDB IDs 1IYT and 1N5U respectively. The hydrogen bond surrogate helix, which has been synthesized by P. Arora and coworkers, has been chosen as the third helix, which is self-contained.²⁰ The crystal structure of this peptide has been reported in the dimeric form. Therefore, the entire helix dimer, whose optimized geometry shows a total of 12 main chain hydrogen bonds, has been considered in this study. The structure of the two β -sheets containing eight and seven main chain hydrogen bonds have been obtained from the PDB structures of proteins having PDB IDs 2BEG and 1LSH respectively. A third example of the β -sheet has been chosen as the synthetic analogue of the Amyloid β -protein (PDB ID: 3T4G), which is a self contained peptide containing a total of eight hydrogen bonds in it.²¹ The crystal structure of this peptide has been taken as the initial input for geometry optimization. All the β -sheets that have been considered here are anti parallel β -sheets. It is to be noted here that all the helices and the sheets that have been chosen here are taken from the different proteins structures. This has been specifically done in order to see the generality of the result, whether the differential bond length and stabilization in protein secondary structure motifs are specific to each protein structure or is general and independent from the protein types and local environment of the protein. The frequency distribution of all the 23 main chain hydrogen bonds present in the three β -sheets and 27 hydrogen bonds present in the three helices has been obtained and shown in Figure 7.4a. The H...O distances corresponding to each hydrogen bond has been considered (here) in the optimized geometries of the peptide structures. The frequency distribution of the H...O distances in these optimized structures also suggests that the hydrogen bond length in the β -sheets are generally smaller than the α -helices (Figure 7.4a).

Following this, the electrostatic force of binding, along the line joining donor H and acceptor O atoms, corresponding to each hydrogen bond has been calculated. The method that has been employed for calculating the electrostatic forces of binding for the peptidic system in Chapter 5 has been employed in this case as well for calculating the binding forces for individual hydrogen bonds in all the six model peptides chosen for force analysis. The obtained electrostatic binding forces (the statistical distribution) in β -sheets have been found to be generally stronger (with

more negative values of the binding force) than the α -helices. The frequency distribution of the electrostatic forces of binding for all the 23 hydrogen bonds in sheets and 27 in helices (from three geometries in each case) employing Mulliken and NBO charges have been shown in Figure 7.4b and 7.4c respectively, which further suggests that the hydrogen bonds (the electrostatic force of binding) in β -sheets are generally stronger than in α -helices, as the frequency graph for β -sheets has been shifted to the left (to larger values of the magnitude of the force) in comparison to the α -helices. Therefore, the cleavage of the hydrogen bonds (during denaturation of proteins) in β -sheets requires a higher amount of energy or force, which is what has been observed experimentally (see the first paragraph of this subsection for references). Thereby, our method provides a physical interpretation of understanding the relative stability of hydrogen bonds in helices and sheet structures of proteins, which is otherwise very difficult to estimate employing existing methods. Furthermore, it shows that the method can be employed directly in estimating the relative strength of individual hydrogen bonds in multi-hydrogen bonded systems. Considering the importance of proteins and the number of papers appearing in the literature, our method holds high significance for biologists.

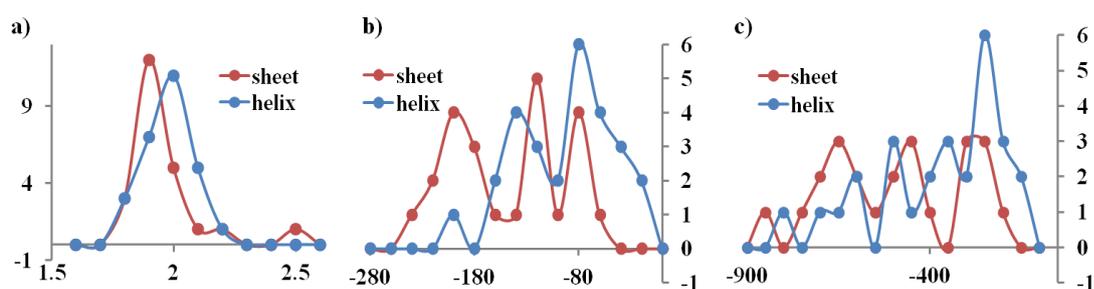


Figure 7.4 a) Frequency distributions of (N)H...O hydrogen bond distances corresponding to main chain hydrogen bonds in the optimized geometries of three representative small model peptides at the COSMO(water)/PBE-D2/TZVP level of theory. Three peptides from each of α -helices and β -sheets have been chosen together for the frequency distribution calculations. b) and c) illustrate frequency distributions of the electrostatic forces of binding for the above-mentioned hydrogen bonds employing Mulliken and NBO charges respectively. Negative values of force in (b) and (c) indicate attractive electrostatic interactions.

7.4 Conclusions

We have shown the general relevance of our method in obtaining the physical interpretation of many physical and chemical phenomena, which have largely been overlooked so far. This is because of the consideration of directionality, taken into account and implemented in our method, which is extremely important in the systems where electrostatic interactions hold significance. This point has been explained taking three examples from three different areas of chemistry and biology, which explain the multidisciplinary significance of our approach. The importance of long range electrostatic interactions in determining the chemical selectivity has been explained by the example of two S_N2 reactions. The molecular levitation phenomenon has been explained taking examples of ammonium and 18-crown-6 interactions. The differential binding in protein secondary structure motifs has been explained by choosing small model and self-contained peptides. Although the sample space in our studies, particularly the peptides one, has been very small, the preliminary results are promising and suggest that the method can be employed in screening the entire sample space of proteins as well. These results also suggest that it may be necessary to look into molecular interactions from a new perspective: by including directionality in deducing reaction mechanisms and the nature and strength of binding in chemical systems.

7.5 References

1. P. L. Popelier, L. Joubert, *J. Am. Chem. Soc.*, 2002, **124**, 8725-8729; J. Pranata, S. G. Wierschke, W. L. Jorgensen, *J. Am. Chem. Soc.*, 1991, **113**, 2810-2819.
2. M. K. Tiwari, K. Vanka, *Chemical Science*, 2017, **8**, 1378 – 1390.
3. D. P. Weimann, H. D. F. Winkler, J. A. Falenski, B. Kokschi, C. A. Schalley, *Nature Chemistry*, 2009, **1**, 573-577.
4. A. Klamt, G. Schüürmann *J. Chem. Soc. Perkin Trans. 2*, 1993, **5**, 799–805.
5. R. Ahlrichs, *Chem. Phys. Lett.*, 1989, **162**, 165-169.
6. J. P. Perdew, *Physical Review Letters*, 1996, **77**, 3865.
7. S. Ansgar, H. Christian, A. Reinhart, *J. Chem. Phys.* 1994, **100**, 5829.

8. K. Eichkorn, *Chemical Physics Letters*, 1995, **240**, 283-289.
9. M. Sierka, *J. Chem. Phys.*, 2003, **118**, 9136-9148.
10. J. Hepburn, G. Scoles, R. Penco. *Chem. Phys. Lett.*, 1975, **36**, 451-456; R. Ahlrichs, R. Penco, G. Scoles. *Chem. Phys.*, 1977, **19**, 119-130; S. Grimme, *J. Comput. Chem.*, 2004, **25**, 1463-1473; S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787-1799; S. Grimme; J. Antony; S. Ehrlich; H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
11. R. S. Mulliken, *J. Chem. Phys.*, 1955, **23**, 1833-1840.
12. A. E. Reed, R. B. Weinstock, F. Weinhold, *J. Chem. Phys.*, 1985, **83**, 735-46.
13. U. C. Singh, P.A. Kollman, *J. Comput. Chem.*, 1984, **5**, 129-145.
14. D. Mani, E. Arunan, *J. Phys. Chem. A*, 2014, **118**, 10081-10089; D. Mani, E. Arunan, *Phys. Chem. Chem. Phys.*, 2013, **15**, 14377.
15. A. K. Jana, M. K. Tiwari, K. Vanka, N. Sengupta, *Phys. Chem. Chem. Phys.*, 2016, **18**, 5910-5924; J. P. Gallivan, D. A. Dougherty, *Org. Lett.*, 1999, **1**, 103-105.
16. T. Clark, in *The Chemical Bond* (Eds.: G. Franking, Sason Shaik), Wiley, Chapter 18, pp 523-536.
17. V. Rüdiger, H.-J. Schneider, V. P. Solov'ev, V. P. Kazachenko, O. A. Raevsky, *Eur. J. Org. Chem.* 1999, 1847-1856.
18. S. Vijayakumar, S. Vishveshwara, G. Ravishanker, D. L. Beveridge, *Biophysical Journal*, 1993, **65**, 2304-2312; S. G. Anema, A. B. McKenna, *J. Agric. Food Chem.*, 1996, **44**, 422-428; A. Kato, T. Takagi, *J. Agric. Food Chem.*, 1988, **36**, 1156-1159.
19. K. A. H. Wildman, D.-K. Lee, A. Ramamoorthy, *Biopolymers*, 2002, **64**, 246-254.
20. J. Liu, D. Wang, Q. Zheng, M. Lu, P. S. Arora, *J. Am. Chem. Soc.* 2008, **130**, 4334-4337.
21. P.-N. Cheng¹, C. Liu, M. Zhao, D. Eisenberg, J. S. Nowick, *Nature Chemistry*, 2012, **4**, 927-933.

Publications

1. M. K. Tiwari, K. Vanka, *Chemical Science*, Exploiting Directional Long Range Secondary Forces for Regulating Electrostatics-Dominated Noncovalent Interactions, 2017, **8**, 1378-1390.
2. M. K. Tiwari,[†] A. K. Jana,[†] K. Vanka, N. Sengupta, Unraveling origins of the heterogeneous curvature dependence of polypeptide interactions with carbon nanostructures, *Phys. Chem. Chem. Phys.*, 2016, **18**, 5910-5924; ([†] = equal authors), *Inner front cover*.
3. K. K. Singh, M. K. Tiwari, B. B. Dhar, K. Vanka, S. S. Gupta, Mechanism of Oxygen Atom Transfer from Fe^V(O) to Olefins at Room Temperature, *Inorg. Chem.*, 2015, **54**, 6112-6121.
4. K. K. Singh, M. K. Tiwari, M. Ghosh, C. Panda, A. Weitz, M. P. Hendrich, B. B. Dhar, K. Vanka, S. S. Gupta, Tuning the Reactivity of Fe^V(O) toward C–H Bonds at Room Temperature: Effect of Water, *Inorg. Chem.*, 2015, **54**, 1535-1542.
5. Dissection of the Catalytic Site of Crucial Gut Microbiome Enzyme: Bile Salt Hydrolase; Y. Yadav, M. K. Tiwari, D. Chand, A. Pundle, S. Ramasamy; *manuscript communicated*.
6. M. K. Tiwari, K. Vanka, A New Approach to Determine the Influence of Explicit Solvent Molecules on Hydrogen bond Strength, *manuscript prepared*.
7. H. Chand, M. K. Tiwari, A. Bhattacharya, Glycol Mediated Synthesis of Fagomine, 4-epifagomine, 2-Deoxynojirimycin and an Advance Intermediate Iminoglycol: Added Insights from DFT Studies, *manuscript prepared*.
8. M. K. Tiwari, K. Vanka, A New Insight into Differential Stability of Proteins Secondary Structure Motifs from Electrostatic Force Analysis, *manuscript communicated*.
9. M. K. Tiwari, K. Vanka, A General Method to Evaluate the Long Range Electrostatic Influence of Solvent Molecules on Reaction Barriers, *manuscript under preparation*.
10. M. K. Tiwari, K. Vanka, Directional Influence of Long Range Electrostatic Interactions on Reactivity and Selectivity, *manuscript under preparation*.

अमंत्रमक्षरं नास्ति नास्ति मूलमनौषधम् ।
अयोग्यः पुरुषो नास्ति योजकस्तत्र दुर्लभः॥

(ऋग्वेद)

कोई अक्षर ऐसा नहीं जिससे मन्त्रोत्पत्ति न की जा सके, कोई वनस्पति ऐसी नहीं जिसमे कोई औषधीय गुण न हो। कोई मनुष्य ऐसा नहीं जो अयोग्य हो, परन्तु ऐसे योजक जो अपनी रचनाशीलता से इन्हें उपयोगी व सार्थक बना सकने में सक्षम हों, अवश्य दुर्लभ हैं॥

Mrityunjay Kumar Tiwari

Physical and Materials Chemistry Division
National Chemical Laboratory,
Dr. Homi Bhabha Road, Pune, (India)- 411 008.
Mob. +91 8605556270
Email ID: tiwari084@gmail.com



Mrityunjay Kumar Tiwari has done his B.Sc. in Chemistry (Honors) from Banaras Hindu University (BHU) and M.Sc in Chemistry from Indian Institute of Technology Bombay (IITB) in the years 2007 and 2009 respectively. He joined Dr. Vanka's research group at the CSIR-NCL in July 2011 for his doctoral work.

His research has mostly been focused on the noncovalent bonded systems where electrostatic interactions are dominating. More precisely, he has worked on understanding the directional influence of long-range electrostatic interactions on a diverse range of chemical and biological systems and their chemistry. He has also worked on organometallic and main group chemistry as well for certain period with Dr. Vanka's research group.

He has received following awards and prizes during his Ph.D. studies at CSIR-NCL:

1. Best Research Scholar Award, CSIR-NCL, 2016.
2. Best Poster Award, Science Day Celebration 2017, CSIR-NCL.
3. Junior Research Fellowship
4. Senior Research Fellowship
5. First Prize in the Essay Competition organized by the CSIR-NCL on the Swachhh Bharat Mission 2017.

Extracurricular Activities:

1. He has been editor-in-chief of NCL's first student Magazine "Dhwani".
2. Meanwhile he has published many unique pieces of poetries, riddles and stories (in Hindi) in some reputed literary magazines.
3. He has been auditor of the Golden Jubilee hostel for the year 2014.