Electrostatics and Dynamics in Molecular Functions:
A Spectroscopic Approach

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

By

Pranab Deb
Registration Number: 10CC14J26007

Under the guidance of
Dr. Sayan Bagchi

Physical and Materials Chemistry Division
CSIR- National Chemical Laboratory
Pune-411008, India

Certificate

This is to certify that the work incorporated in this Ph.D. thesis entitled "**Electrostatics and dynamics in molecular functions: a spectroscopic approach**" submitted by Mr. **Pranab Deb** to Academy of Scientific and Innovative Research (AcSIR) in fulfillment of the requirements for the award of the Degree of **Doctor of Philosophy**, embodies original research work under my/our supervision/guidance. I/We 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 material obtained from other sources has been duly acknowledged in the thesis. Any text, illustration, table etc., used in the thesis from other sources, have been duly cited and acknowledged.

It is also certified that this work done by the student, under my supervision, is plagiarism free.

Pranab Deb
(Student)

Dr. Sayan Bagchi
(Supervisor)

Date: 10th July, 2019.

Place: Physical & Materials Chemistry Division,
CSIR – National Chemical Laboratory (NCL),
Pune – 411008, India.

Declaration

I hereby declare that the thesis entitled **"Electrostatics and dynamics in molecular functions: a spectroscopic approach"** submitted for the award of the Degree of Doctor of Philosophy in Chemical Sciences to the Academy of Scientific and Innovative Research (AcSIR), has been carried out by me under the supervision of Dr. Sayan Bagchi at the Physical and Materials Chemistry Division of CSIR - National Chemical Laboratory, Pune. 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 material obtained from other sources has been duly acknowledged in the thesis.

Pranab Deb
Pranab Deb
(Senior research fellow)

Date: 10th July, 2019.

Place: Physical & Materials Chemistry Division,
        CSIR – National Chemical Laboratory (NCL),
        Pune – 411008, India.

To my Guru,
Swami Sarvananda

[Shri Gurudham, Yogoda Satsanga Ashram, Swami Sarvananda Sevashram, Vadia-365480, Dist-Amreli, Gujarat]

Acknowledgements

It is a great pleasure for me to present my research work in the form of this Ph. D. thesis. I would like to take this privilege to express my gratitude and appreciation to all who helped me to accomplish my destination.

Foremost, I express my sincere gratitude to my guide Dr. Sayan Bagchi for introducing me to this exciting area of research. I thank him for his dedicated help, advice, inspiration, enthusiasm, encouragement, patience, continuous support and constructive criticisms during the entire tenure of my Ph.D. work. He had given me a platform where I could work with absolute freedom. His wisdom, unique approach towards solving scientific problems, optimistic outlook, confidence, and strong determination have made a great impression on me. Here, I must admit that I find it very difficult to encapsulate in words how his guidance has influenced me and what he has done for me. I would finally say that working with him was a wonderful opportunity in my life.

I would like to sincerely thank my doctoral advisory committee members, Dr. Neelanjana Sengupta (now at IISER Kolkata), Dr. E. Balaraman (now at IISER Tirupati), Dr. B L V Prasad, Dr.Sarika Maitra Bhattacharyya and Dr. Santosh Kumar Jha for their constant support and suggestions throughout my Ph.D.

I express my special appreciation to Dr. Sayam Sen Gupta (now at IISER Kolkata), Dr. Mahesh J Kulkarni and Dr. Arnab Mukherjee (IISER Pune) for their generous support. My thesis work would have been incomplete without the help of our collaborators. I sincerely thank to all our collaborators, Dr. Suman Chakrabarty (SNBNCBS, Kolkata), Dr. Anirban Hazra (IISER Pune), Dr. Aloke Das (IISER Pune), Dr. Debashree Ghosh (IACS, Kolkata) and Yung Sam Kim (UNIST, Korea).

I extend my sincere thanks to Prof. Ashwini Kumar Nangia (the present Director of CSIR-NCL), former directors of CSIR-NCL, Dr. P. A. Joy (-present chair of the Physical and Materials Chemistry division) and former chairs of the division for allowing me to avail all the instrument facilities. I express my thanks to the eminent scientists of NCL for enriching me with knowledge and valuable information. I am thankful to the student academic office of NCL for all the

<div align="right">

Pranab Deb
10th July, 2019

</div>

Contents

Abstract

Understanding the physical basis of biomolecular processes is a continued pursuit in molecular biology. Non-covalent interactions (hydrogen bonding, hydrophobic interactions, electrostatic interactions, van der Waals interactions, etc.), fundamental to several biological processes, are exploited to regulate biomolecular functions involving molecular-recognition, protein folding, enzyme-catalysis and protein-protein interactions. In addition to these interactions, biological macromolecules have dynamic structures and the fluctuations of these structures are intimately related to their biological functions. Even at normal temperatures under thermal equilibrium conditions, the proteins are never at rest. Almost all biological processes involve structural and conformational changes in proteins. Proteins, in course of their transient interactions and conformational changes, behave in astounding ways, leading to enzyme catalysis, protein folding and misfolding, signaling and transport, charge transfer and energy relaxation and transduction. Proteins are the functional entities of cell and malfunction of proteins leads to several diseases. Therefore a cognitive understanding of interactions and dynamics, which play the central role in protein functions, is also crucial in the realm of drug discovery.

Apart from understanding biomolecular functions, a feasible and economically viable synthetic strategy is required to obtain desired organics motifs of clinical importance. To this end, modern organic chemistry is longing for developing suitable catalysts to introduce chemical changes in reactions with a high level of precision. The core of catalyst design in homogeneous medium requires the knowledge of different kinds of interactions involved in catalysis along with the knowledge of the nature of fluctuating solvent medium. Various interactions, responsible in catalyzing reactions, are often non-covalent in nature. However, due to the lack of a general approach in describing/quantifying non-covalent interactions and dearth of our knowledge on the role of solvent dynamics in reaction mechanism, a unique catalyst design principle is still not known to date.

Molecular functions are largely governed by intermolecular interactions. Usual theoretical approaches use some geometry and distance criteria to describe different intermolecular interactions. However often such descriptions lead to rise of confusion. Moreover, describing

intermolecular interactions through geometric and distance constraints conceals the fact that such interactions can be explained and understood well by electrostatics. In this thesis, using electrostatics and dynamics, we have addressed few important challenges in biology and synthetic chemistry, the two principal contributing domains toward developing molecules relevant to the pharmaceutical industry. We have described how we have been able to surmount the challenges those came along the way.

List of publications

1. Somnath M. Kashid, **Pranab Deb**, Tapas Haldar, Partha Pratim Roy and Sayan Bagchi. Measuring equilibrium ultrafast structural and conformational dynamics in bio-molecules: A two dimensional infrared (2D-IR) spectroscopic approach. *Sci. Lett. J.,* **2015**, 4, 173.

2. **Pranab Deb**, Tapas Haldar, Somnath M Kashid, Subhrashis Banerjee, Suman Chakrabarty and and Sayan Bagchi. Correlating Nitrile IR Frequencies to Local Electrostatics Quantifies Noncovalent Interactions of Peptides and Proteins. *J. Phys. Chem. B,* **2016**, 120, 4034–4046.

3. Tapas Haldar, Somnath M Kashid, **Pranab Deb**, Sandeep Kesh and Sayan Bagchi. Pick and Choose the Spectroscopic Method to Calibrate the Local Electric Field inside Proteins. *J. Phys. Chem. Lett.,* **2016**, 7, 2456–2460.

4. **Pranab Deb**, Geun Young Jin, Santosh K. Singh, Juran Moon, Hyejin Kwon, Aloke Das, Sayan Bagchi and Yung Sam Kim. Interconverting Hydrogen-Bonding and Weak n → π* Interactions in Aqueous Solution: A Direct Spectroscopic Evidence. *J. Phys. Chem. Lett*., **2018**, 9, 5425–5429.

5. **Pranab Deb** and Sayan Bagchi. Infrared Spectroscopy Probes Electrostatic Catalysis of Diels-Alder Reaction in Solution. *Manuscript communicated*, (**2019**).

N.B.: The content of the thesis has been majorly drawn from publication 2, 4 and 5 as mentioned under the list of publications.

Chapter 1

Introduction

1.1 Electrostatics and molecular function

Electrostatic interaction has been widely studied in understanding the enzyme catalysis. Arieh Warshel and Michael Levitt are the pioneers of this field. Like any other molecules, proteins also contain charged and polarizable atoms. The electrostatic energy of a group of charges in the protein active site is considered to be governed by two different contributions,[1,2]

(i) Charge-charge interaction:

$$V_{QQ} + V_{Qq} = \sum_{i>j} \frac{Q_i Q_j}{r_{ij}} + \sum_{ik} \frac{Q_i q_k}{r_{ik}}$$ [1.1]

Where, V is the interaction energy, Q and q are the charges of inside and outside of the enzyme active site region respectively, and r is the distance between charges. Subscript, i, j and k represents different interacting charges.

(ii) Interactions between charges and the induced dipoles on the protein atoms:

$$V_{ind} = -\frac{1}{2}[\sum_{i,k} \frac{Q_i \mu_k \vec{r}_{ik}}{r^3_{ik}} + \sum_{j,k} \frac{q_j \mu_k \vec{r}_{jk}}{r^3_{jk}}]$$ [1.2]

Where V_{ind} is the electrostatic interaction energy due to the charge-induced dipole interactions and μ_k represents the induced dipoles of the system.

Enzymes perform a significant role in the functioning of cellular machinery. In almost all life processes - energy transduction, transcription and translation of genetic information - enzymes plays the key role. The fascinating thing about enzymes is their unique capacity to catalyze reactions at a very high rate compare with the uncatalyzed reactions. Without such catalytic proficiency of enzymes, it would have been impossible for cells to function. Among different hypotheses[3] in understanding remarkable catalytic power of enzymes, contribution of electrostatics has been illustrated at a great extent by both theory and experiment. We will focus on one example that will help us to understand the importance of electrostatics on molecular function, which is an integral part of the thesis.

Keto steroid isomerize (KSI), a member of isomerases family, catalyzes enolization process of β, γ-unsaturated 3-oxosteroids (Figure 1.1).[4,5] KSI is known to have one of the highest unimolecular rate constants in biology.[6] The enolization process of KSI involves an anionic intermediate. In a computational simulation study, Warshel et al. investigated the origin of large

2

catalytic proficiency of KSI.[4] Empirical valence bond (EVB) was used to calculate the different contributing factors to the energetics of the catalytic reaction of KSI (Figure 1.1). The rate

| Reactants | Intermediate | Products |
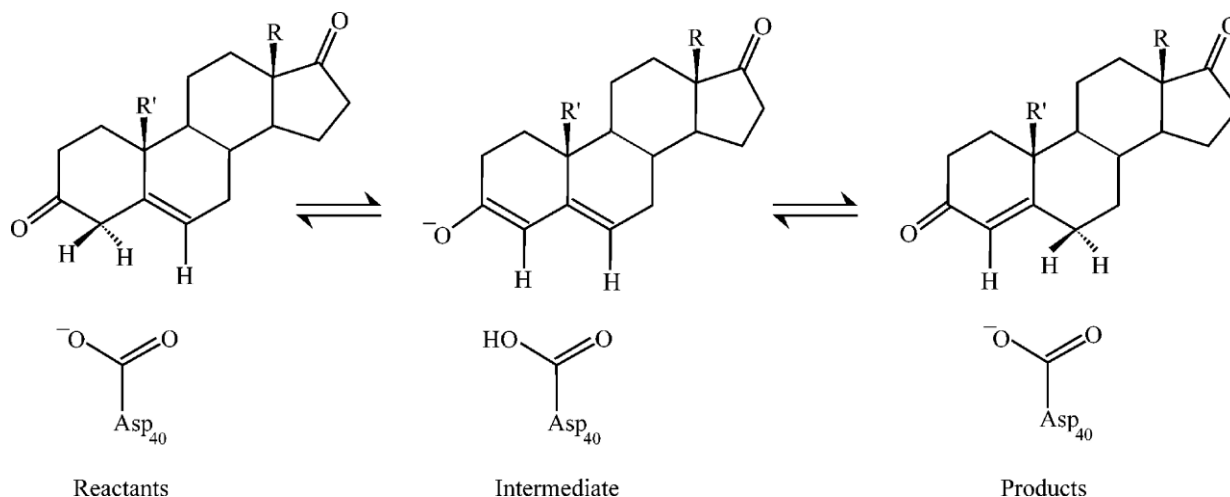
Figure 1.1: *Schematic description of the catalytic reaction of KSI. This figure describes only the chemical parts of the system so that it can be compared conveniently to the corresponding reference reaction in water. The reaction is described in terms of the actual states rather than by arrows that correspond to electronic rearrangements.[4] This figure has been taken from reference 4.*

determining step is the first step where charge transfer happens from ASP-40 residue to the enolate oxygen of steroid. Free energy of activation for this reaction in aqueous solution (~21.9 kcal/mol) is much higher compared to the free energy of activation inside enzyme active site pocket (10.3 kcal/ mol). Calculations revealed that the electrostatic (described in equation 1.1 and 1.2) stabilization of the transition state by a preorganized enzyme active site contribute largely (~10.3 kcal/mol) in reduction of the activation free energy of the reaction. To get the confirmation about electrostatic effect on the catalysis, simulations were done removing the electrostatic effect (i.e. by keeping residual charges as zero for all the atoms which are part of the reaction). Interestingly, it was observed that the catalytic effect of KSI disappeared in the absence of electrostatic effect (Figure 1.2). It is also important to know here that the calculations
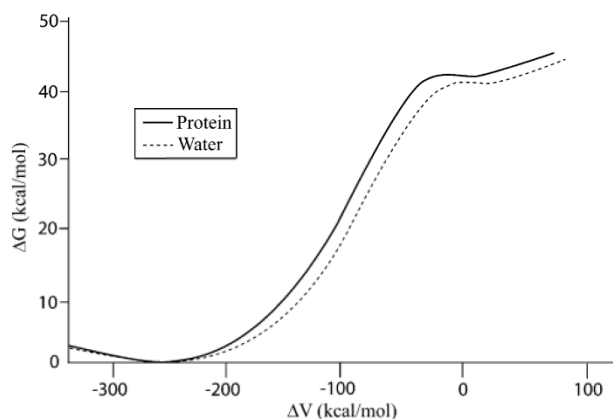
Figure 1.2*: Free energy surfaces for the reaction of KSI (s) and the corresponding reference reaction in water (- - -) for the case where the residual charges of the reacting atoms are set to zero. This figure shows that the reduction of the chemical barrier is due to electrostatic effects. This figure has been taken from reference 4.*

used in this study were successful in predicting the binding energy of the transition state analogues of KSI, an experimental quantity obtained from NMR spectroscopy[7]. Thus, in this work, Warshel illustrated the significant contributions of electrostatics toward catalytic proficiency of KSI.

The electrostatic origin of enzyme catalysis in KSI got complemented in a ground breaking experimental work by Boxer and coworkers.[8] Vibrational Stark effect (VSE) spectroscopy was used in this study. Linear VSE quantitatively reports the frequency change ($\Delta \bar{v}$ /cm[-1]) of a vibrational probe (vibrationally active functional groups) in response to the change in electric field ($\Delta \vec{F}$ /MVcm[-1])) acting on to the probe:[9]

$$\Delta \bar{v} = \Delta \vec{\mu} . \Delta \vec{F}$$
[1.3]

where, $\Delta \vec{\mu}$ is termed as Stark tuning rate which is a constant. $\Delta \vec{\mu}$ describes the sensitivity of a probe to electric field (for details please see chapter 2). Electric field mentioned in equation 1.3 either can be applied externally through an electrical circuit[10] or can be endogenously generated by charged or polar groups around a probe[9]. Equation 1.3 holds only when the interactions between the probe and the electric field are electrostatic (dipolar) is nature. This relation was

used by Boxer and coworkers in studying KSI.[8] Carbonyl (C=O) vibrational frequency of an inhibitor, which directly binds to the active site of KSI, was studied to extract the electrostatic field of the active site (Figure 1.3A). Catalytic activities of different mutants of KSI were investigated. Interestingly, the electric fields at the active sites of different mutants showed a strong linear correlation with the free energy barrier of the isomerization reaction (Figure 1.3B). Boxer and coworkers have also demonstrated in an another work that the electrostatic stabilization of transition state is not only in KSI, but is also crucial for catalytic processes in various other enzymes.[11] These results showed that electrostatics plays the pivotal role in molecular functions like enzyme catalysis.
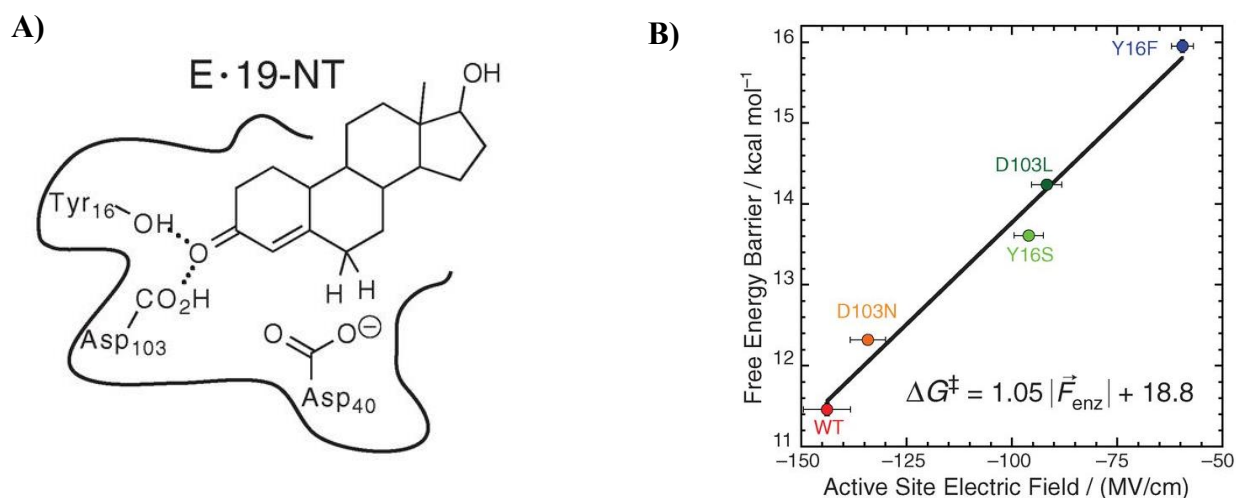
Figure 1.3: *(A) Complex between KSI and the product-like inhibitor, 19-NT, used in this study to probe electric fields in the KSI active site. (B) Plot of enzymatic unimolecular free energy barrier against the electric field, 19-NT's C=O group experiences in each of the active sites of KSI variants.[8] This figure has been taken from the reference 8.*

1.2 Dynamics and molecular function

The understanding of protein functions got revolutionized from immense contribution of X-ray crystallography. This may be the primary cause of looking at proteins as static biological entities. However, from a physicist's point of view, proteins are bio-analogues of polymeric molecules which samples large number of conformations centering the average structure. Apart from concept of electrostatics in understanding molecular functions, a second proposal, well known in the scientific community involves the dynamical effect. The pioneers of illustrating dynamical

effects in controlling biomolecular functions are G. Careri, P. Fasella, E. Gratton and Martin Karplus.[12,13] Here dynamics is defined as the time dependent change of atomic coordinates of a molecule.[14] Due to the contribution from thermal energy, molecules are inevitably never at rest even at normal temperature at thermal equilibrium condition. The energy landscape theory,[15,16] which is used regularly in studying folding-unfolding processes, demonstrates that proteins consists of large ensemble of conformational states in the multidimensional energy landscape. Probabilities of such conformational states are governed by the equilibrium thermodynamics, whereas energy required in crossing the barrier separating the states is related to dynamics. Apart from protein's own dynamics, the fluctuations of solvents and ligands also have an immense impact on the time evolution process of protein structures. The first evidence of dynamic structural fluctuation in proteins was instanced by Lakowicz and Weber in 1973 by fluorescence quenching measurements on globular proteins.[17] In 1977, an elegant work by McCammon, Gelin and Karplus presented how protein dynamics can be studied using molecular dynamics simulations.[18] This study opened a new dimension to get the atomistic picture of not only proteins but also of any molecular system. The dynamic nature of proteins is further supported by several other facts like: penetration of oxygen molecules into the binding site of heme proteins[19] and conformational adjustments of binding sites during complex formation of proteins with ligands including substrate, drug molecules, and hormones.[13] Wide varieties of dynamic motion shown by proteins are classified on the basis of time scale (10^{-15} to 10^{3}s), amplitude (0.1 to 100 Å), energy (0.1 to 100 kcal) and the structural type (local, rigid body, large scale and collective). Based on the time scale of dynamics various spectroscopic techniques such as: NMR, fluorescence, IR and Raman spectroscopy, with different ranges of time resolution, can be employed to study the dynamic processes. We will now see an example illustrating the importance of dynamics in biomolecular function.

Dihydrofolate reductase (DHFR) catalyzes the stereospecific hydride transfer (Figure 1.4) from NADPH to dihydroflorate (DHF) to form tetrahydroflorate (THF) which plays the key role in cell growth and replication[16]. During the biosynthesis of DNA, THF gets oxidized to DHF in an irreversible path. DHFR regenerates THF from DHF. In absence of THF the DNA production in cell gets inhibited and this leads to cell death. Escherichia coli DHFR (ecDHFR) has been a subject of immense interest in cancer research. Cancer cells replicates very fast. To engineer a drug molecule which will stop the replication process, understanding DHFR mechanism
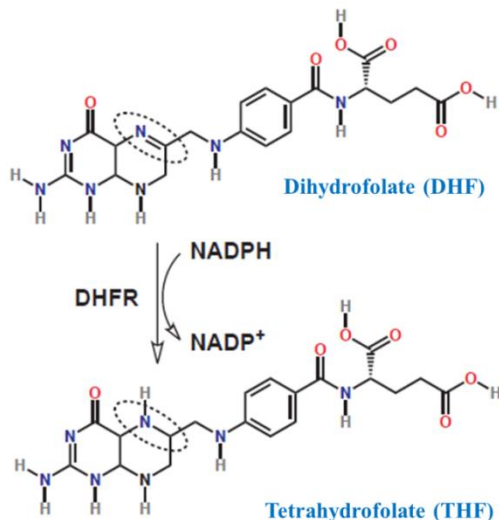
Figure 1.4: *Activity of DHFR.*

is crucial. DHFR contains a Met-20 loop near the active site (Figure 1.5A).[20] Met-20 loop takes closed and occluded conformation during the catalytic circle (Figure 1.5B). For the hydride transfer to happen, the active site loops must cycle between closed and occluded states.
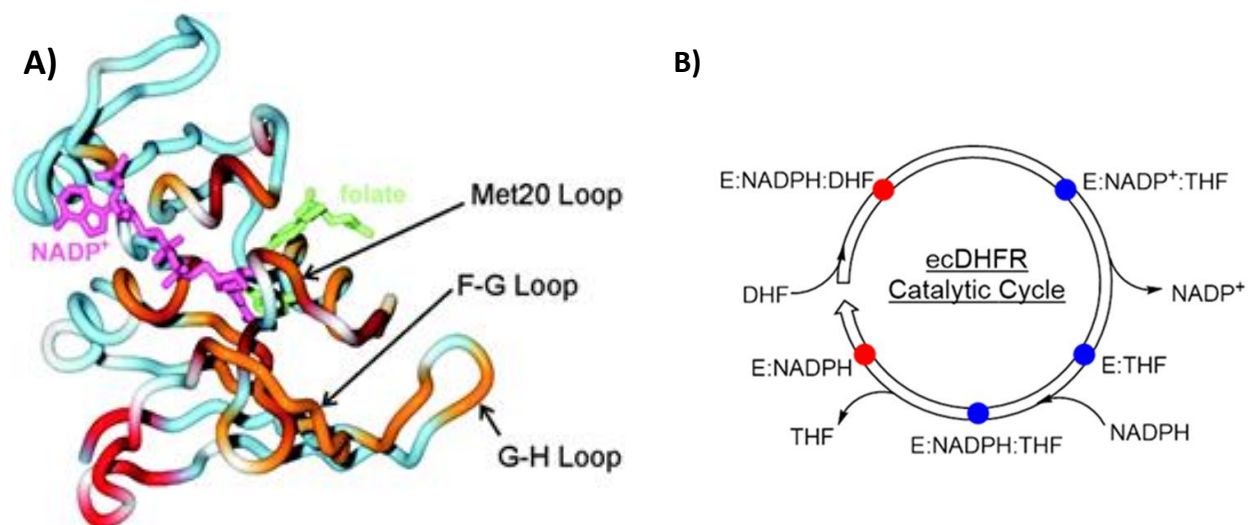
Figure 1.5: *(A) Structure of the ternary complex of ecDHFR with bound folate andNADP[+].[21] This figure has been taken from reference 21. (B) Catalytic cycle of ecDHFR.[22] Red and blue solid circles represent the closed and occluded conformational states respectively. This figure has been taken from reference 22.*

In a seminal work, Bhabha et al. designed a mutant of ecDHFR to study the dynamical effect.[23] The N23PP/S148A (PP: polyproline) mutant of ecDHFR was synthesized. X-ray and NMR analysis in conjunction with the pKa measurements (Table 1.1) suggested the similar

electrostatic environment for both the mutated and wild type enzyme active sites. However the hydride transfer rate sharply decreased in the N23PP/S148A mutant (Table 1.1). NMR relaxation measurements revealed that the enzyme active site undergoes structural fluctuations at millisecond time scale. Whereas in the active site of N23PP/S148A mutant, the millisecond time scale fluctuations abrogated. The absence of the active site fluctuations was supported by the fact that N23PP/S148A mutant was unable to adopt the occluded conformation as observed in NMR. Interestingly, the effect in dynamics due to mutation was seen only in the active site of the enzyme. These observations lead to conclude that the conformational dynamics plays the key role in catalytic activity of ecDHFR instead of electrostatic pre-organization.

Table 1.1: *Kinetic parameters of mutated and wild type ecDHFR[23]*

	Wild type ecDHFR	**N23PP-S148A ecDHFR**
pKa	6.5 ± 0.1	6.7 ± 0.1
k_{hyd} (s^{-1})	220 ± 10	13.9 ± 0.6

1.3 Tussle between electrostatics and dynamics

The conclusion of the study by Bhabha et al.[23] on ecDHFR directly contradicted Warshel's electrostatic pre-organization concept which is the main responsible factor leading to enzyme catalysis. Confronting this dynamical concept, Warshel and coworkers published an article in the same year with extensive calculations on the wild type and mutated ecDHFR enzymes.[24] EVB (mentioned in section 1.1) calculations were performed to obtain the activation energy barriers of the ecDHFR variants. The calculated activation energy values showed the similar trend as obtained from experimental work of Bhabha et al (Table 1.2) [23] and thereby establishing the fact that the effect of mutation is most likely due to change in free energy and not due to the dynamical effect. The EVB calculated activation energies showed linear correlation between electrostatic preorganization energy and the free energy of activation. This indicated the main role of electrostatics on catalytic function instead of dynamics. The calculations also showed the presence of different electrostatic environment in enzyme variants. This was completely opposite to what observed from NMR, X-ray and pKa measurements by Bhabha et al. Warshel

8

Table 1.2: *Calculated and experimentally observed activation energies of ecDHFR variants.*[24]

System	$\Delta G^{\neq}_{\text{calculated}}$ in kcal/mol	$\Delta G^{\neq}_{\text{experimentally observed}}$ in kcal/mol
Wild type ecDHFR	14.3	14.4
N23PP-S148A ecDHFR	15.1	16.1
S148A ecDHFR	14.6	14.6

mentioned that the information on electrostatic preorganization obtained from NMR or X-ray lacked the microscopic insight. In addition to this, it is important to mention here that the only interpretation given by Bhabha et al. on the cause of the abrogation of active site fluctuations, as seen in N23PP-S148A mutant, was the inability of the mutant to form the occluded conformational state. There was no direct evidence showing the relation between millisecond timescale fluctuation and it's impact on enzyme catalysis.

To resolve the issue raised in case of DHFR, Liu et al. performed a study on ecDHFR mutants by inserting a vibrational probe at a site specific location in the active site. SCN vibrational probe was incorporated in the 54th (L54C-CN) residue.[25] Vibrational spectra of the SCN mutated enzymes were recorded for all the five enzyme–complexes which constitute the catalytic cycle. It was observed that the nitrile frequency changed over a frequency span of 10 cm^{-1} throughout the catalytic cycle. This indicated the change in electrostatic environment of the enzyme active site. Moreover, spectral line width analysis showed that the full width at half maxima (FWHM) of the SCN probe is around two times smaller in either DHF or THF bound state of enzyme as compared to the unbound states. Line width of a spectrum indicates the variation of the chemical environment around a probe. As small line width corresponds to a less fluctuating environment, the observations by Liu et al. suggested a less fluctuating and well-structured electrostatic environment in the active site of ecDHFR. Moreover, in the study of KSI by Boxer and coworkers (described in section 1.1), it was noted that the spectral line width of C=O vibrational probe of the inhibitor was significantly lower (~4 times) when bound to the enzyme active site as compared with the water.

From the discussion, it can be concluded that a very high catalytic rate shown by the enzymes is due to the dynamics assisted electrostatic stabilization of the transition state. The electrostatic contribution to enzyme catalysis is direct and can be estimated quantitatively. At present the

electrostatic concept is established on a solid platform supported by both theory and experiments. Whereas, the contribution of the dynamical effect on enzyme activity is yet be illustrated to access it's impact on catalysis in a quantitative manner. However, it's effect cannot be neglected. Thus, to understand the molecular function, both the electrostatic and dynamic perspective must be taken into account.

1.4 Outline of the thesis

In this thesis, using steady state infrared (IR) spectroscopy and two dimensional infrared (2DIR) spectroscopy, we have gained the information of electrostatics and dynamics respectively in condensed phase molecular systems. Moreover, in order to get some atomistic insight, computational calculations employing both classical and quantum mechanics have been performed to complement the spectroscopic studies. The aim of the thesis is twofold. On one hand, we would like to address some important challenges in studying proteins or enzymes which are responsible for different diseases. On the other hand we would like to address existing challenges in the field of chemical catalysis, which breathe life into the drug development process. The outline of the thesis is following:

In the present chapter (**Chapter 1**), the aim of the thesis, along with the approach adopted to achieve the aim, have been stated. A background study has also been given to get a better understanding of the theme of the thesis.

Chapter 2 describes experimental and computational methods used in this thesis to understand and measure electrostatics and dynamics.

Chapter 3 addresses a long lasting issue in obtaining quantitative information of electrostatics in molecular environments using a vibrational probe, nitrile (C≡N) probe, with the aid of vibrational Stark spectroscopy. C≡N group absorbs the infrared light frequency in a spectral region that is transparent in terms of background absorption of biomolecules (proteins, enzymes, peptides etc.). This imparts an outstanding opportunity in obtaining quantitative estimation of electric field utilizing nitrile vibrational probe. However, in the presence of hydrogen bonding, nitrile stretching frequency shows a blue shift which is not expected as per electrostatics (or Stark effect) is concerned. This weird behaviour of nitrile kept it out of the play. It is worth to

mention here that C≡N group is present in many lifesaving drug molecules and even take part in the active site binding of targeted proteins responsible for diseases. Therefore the study of nitrile is extremely important to get microscopic understanding of protein- ligand/ drug molecule binding. IR spectroscopy of nitrile containing small organic molecules was performed along with classical molecular dynamics simulations to obtain a solution for the nitrile probes in hydrogen bonding environments. Interestingly, we found that solvent-induced frequency shifts in C≡N vibrations (of hydrogen bonding solvation environments) correlate quantitatively with the ensemble-average electric field experienced by the C≡N group in the various solvation environments. The result was rigorously tested for nitrile containing molecules including peptides and proteins. It has been seen that irrespective of the size of molecule and complexity of the molecular environments, our observation was successful to accomplish in obtaining electrostatic information in both hydrogen and non-hydrogen bonding environment for varieties of nitriles. The outcome of the chapter 3 opens up a wide access to study nitriles in understanding various important biological processes.

Chapter 4 deals with dynamics and shows the role of intrinsic molecular fluctuations in regulating non-covalent interactions, which are significant to the structure and function of biomolecules, small drug molecules, and materials. The basic feature of non-covalent interactions is governed by the delocalization of electron density. Hydrogen bonding and n→π*, the two conceptually homologues non-covalent interactions, are ubiquitously present in biology. Although existence and nature of hydrogen bonding are well versed in literature, a little knowledge is available regarding the n→π* interaction and this is due to the much smaller energy contribution of n→π* interaction (0.3−1.5 kcal/mol) compared with the hydrogen bonding interactions ((4−10 kcal/ mol). Knowledge on n→π* interaction is majorly limited to the computational studies. However, statistical analysis of protein structure obtained from high resolution X-ray crystallographic technique suggests that n→π* interaction prolifically present in proteins (~34%). Notwithstanding the lower energy contribution of individual n→π* interaction, their high prevalence illustrates a large contribution to the overall energy related to protein function. An open question that still remains unanswered is: how does a weak interaction like n→π* interaction exists in competition with strong interaction like hydrogen bonding? Using two-dimensional infrared spectroscopy, we present direct experimental signature of the

11

coexistence of n→π* and hydrogen-bonding interactions, interconverting in picosecond timescale. The dynamic equilibrium between the two interactions provides a unique perspective toward the existence of n→π* interactions in aqueous solution.

Chapter 5 presents an electrostatic approach towards understanding catalytic process of organic reactions. Catalysis has become the pinnacle principle in pharmaceutical industries to develop atom-economic and eco-benign synthetic methods of new lifesaving molecules.[26] In this regard, organic chemistry is heading towards formulating a design principle that will be able to predict the catalysts to carry out a reaction with absolute precision in product selectivity. However, even after striking advancements in theory, computational technique, spectroscopy and synthetic methodology, the arduous trial and error synthetic process concludes the fate of catalytic reactions. In search of a way to construct reaction model to design de novo catalysts, it is extremely important to get mechanistic details and ability to describe the catalysis process from a very general viewpoint. To achieve the goal, it is desirable to look how enzymes catalyzes reactions. It has now been well established that electrostatics play a great role behind catalytic proficiency of enzymes. However, the role of electrostatics has never been interrogated in catalysis of organic reactions happening in the condense phase. In this chapter we have studied Diels-Alder reactions. Diels Alder reaction has brought a revolution in making carbon-carbon bond formation. Moreover, it has made the synthesis of numerous bioactive compounds easier. Through a combined approach of density functional theory (DFT), reaction rate measurements, molecular dynamics (MD) simulation and IR spectroscopy, we provide experimental evidence of electrostatic catalysis as a general catalytic principle for Diels-Alder reactions by exploiting the internal electric field exerted by homogeneous solution environment on the reactants.

Chapter 6 provides future aspects for all three working chapters (Chapter 3, Chapter 4 and Chapter 5) and the summary of the thesis.

1.5 References

(1)   Warshel, A.; Levitt, M. Theoretical Studies of Enzymic reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227-249.

(2)   Warshel, A. Energetics of enzyme catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **1978**, *75*, 5250-5254.

(3)   Schramm, V. L. Introduction: Principles of Enzymatic Catalysis *Chem. Rev.* **2006**, *106*, 3029-3030.

(4)   Warshel, A.; Sharma, P. K.; Chu, Z. T.; Åqvist, J. Electrostatic Contributions to Binding of Transition State Analogues Can Be Very Different from the Corresponding Contributions to Catalysis: Phenolates Binding to the Oxyanion Hole of Ketosteroid Isomerase. *Biochemistry* **2007**, *46*, 1466-1476.

(5)   Pollack, R. M.; Thornburg, L. D.; Wu, Z. R.; Summers, M. F. Mechanistic Insights from the Three-Dimensional Structure of 3-Oxo-Δ5-steroid Isomerase. *Arch. Biochem. Biophys.* **1999**, *370*, 9-15.

(6)   Pollack, R. M. Enzymatic Mechanisms for Catalysis of Enolization: Ketosteroid Isomerase *Bioorg. Chem.* **2004**, *32*, 341-353.

(7)   Kraut, D. A.; Sigala, P. A.; Pybus, B.; Liu, C. W.; Ringe, D.; Petsko, G. A.; Herschlag, D. Testing Electrostatic Complementarity in Enzyme Catalysis: Hydrogen Bonding in the Ketosteroid Isomerase Oxyanion Hole. *PLOS Biology* **2006**, *4*, e99.

(8)   Fried, S. D.; Bagchi, S.; Boxer, S. G. Extreme Electric Fields Power Catalysis in the Active Site of Ketosteroid Isomerase. *Science* **2014**, *346*, 1510-1514.

(9)   Fafarman, A. T.; Boxer, S. G. Nitrile Bonds as Infrared Probes of Electrostatics in Ribonuclease S. *J. Phys. Chem. B* **2010**, *114*, 13536-13544.

(10)   Boxer, S. G. Stark Realities. *J. Phys. Chem. B* **2009**, *113*, 2972-2983.

(11)   Schneider, S. H.; Boxer, S. G. Vibrational Stark Effects of Carbonyl Probes Applied to Reinterpret IR and Raman Data for Enzyme Inhibitors in Terms of Electric Fields at the Active Site. *J. Phys. Chem. B* **2016**, *120*, 9672-9684.

(12)   G Careri; P Fasella, a.; Gratton, E. Enzyme Dynamics: The Statistical Physics Approach. *Annu. Rev. Biophys. Bioeng* **1979**, *8*, 69-97.

(13)   Kurplus, M.; McCammon, J. A. Dynamics of Proteins: Elements and Function *Annu. Rev. Biochem* **1983**, *52*, 263-300.

(14)   Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature* **2007**, *450*, 964.

(15)   Wolynes, P. G. Recent successes of the energy landscape theory of protein folding and function *Q. Rev. Biophys* **2005**, *38*, 405-410.

(16)   Austin, R. H.; Beeson, K. W.; Eisenstein, L.; Frauenfelder, H.; Gunsalus, I. C. Dynamics of ligand binding to myoglobin. *Biochemistry* **1975**, *14*, 5355-5373.

(17)   Lakowicz, J. R.; Weber, G. Quenching of Protein Fluorescence by Oxygen. Detection of Structural Fluctuations in Proteins on the Nanosecond Time Scale. *Biochemistry* **1973**, *12*, 4171-4179.

(18)   McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977**, *267*, 585-590.

(19)   Perutz, M. F.; Mathews, F. S. An X-ray Study of Azide Methaemoglobin. *J. Mol. Biol.* **1966**, *21*, 199-202.

(20)   Sawaya, M. R.; Kraut, J. Loop and Subdomain Movements in the Mechanism of Escherichia coli Dihydrofolate Reductase: Crystallographic Evidence. *Biochemistry* **1997**, *36*, 586-603.

(21)   McElheny, D.; Schnell, J. R.; Lansing, J. C.; Dyson, H. J.; Wright, P. E. Defining the Role of Active-site Loop Fluctuations in Dihydrofolate Reductase Catalysis. *Proc. Natl. Acad. Sci. U. S.A.* **2005**, *102*, 5032.

(22)   Hanoian, P.; Liu, C. T.; Hammes-Schiffer, S.; Benkovic, S. Perspectives on Electrostatics and Conformational Motions in Enzyme Catalysis. *Acc. Chem. Res.* **2015**, *48*, 482-489.

(23)   Bhabha, G.; Lee, J.; Ekiert, D. C.; Gam, J.; Wilson, I. A.; Dyson, H. J.; Benkovic, S. J.; Wright, P. E. A Dynamic Knockout Reveals That Conformational Fluctuations Influence the Chemical Step of Enzyme Catalysis. *Science* **2011**, *332*, 234.

(24)   Adamczyk, A. J.; Cao, J.; Kamerlin, S. C. L.; Warshel, A. Catalysis by Dihydrofolate Reductase and Other Enzymes Arises From Electrostatic Preorganization, Not Conformational Motions. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 14115.

(25) Liu, C. T.; Layfield, J. P.; Stewart, R. J.; French, J. B.; Hanoian, P.; Asbury, J. B.; Hammes-Schiffer, S.; Benkovic, S. J. Probing the Electrostatics of Active Site Microenvironments along the Catalytic Cycle for Escherichia coli Dihydrofolate Reductase J. Am. Chem. Soc. **2014**, 136, 10349-10360.

(26)   Busacca, C. A.; Fandrick, D. R.; Song, J. J.; Senanayake, C. H. The Growing Impact of Catalysis in the Pharmaceutical Industry. *Adv. Synth. Catal.* **2011**, *353*, 1825-1864.

Chapter 2

Experimental and Computational Methods

2.1 Vibrational Stark effect

Vibrational Stark effect has been successfully utilized in literature to experimentally measure electrostatic interaction by quantitative estimation of electric field of a chemical environment which is under investigation.[1] The electric field ($\vec{F}$) at a distance r from a point charge (q) is defined by the Coulomb's law:[2]

$$\vec{F} = \frac{q}{4\pi\varepsilon_0 r^2} \vec{r}$$

[2.1]

Where ε_0 is the dielectric constant of the medium.

Due to the anharmonic nature of the molecular potential energy curve, the equilibrium bond distance of a diatomic oscillator at higher vibrational energy states are larger than the vibrational ground state (Figure 2.1A).[1] If a diatom possesses charge separation, a difference dipole moment
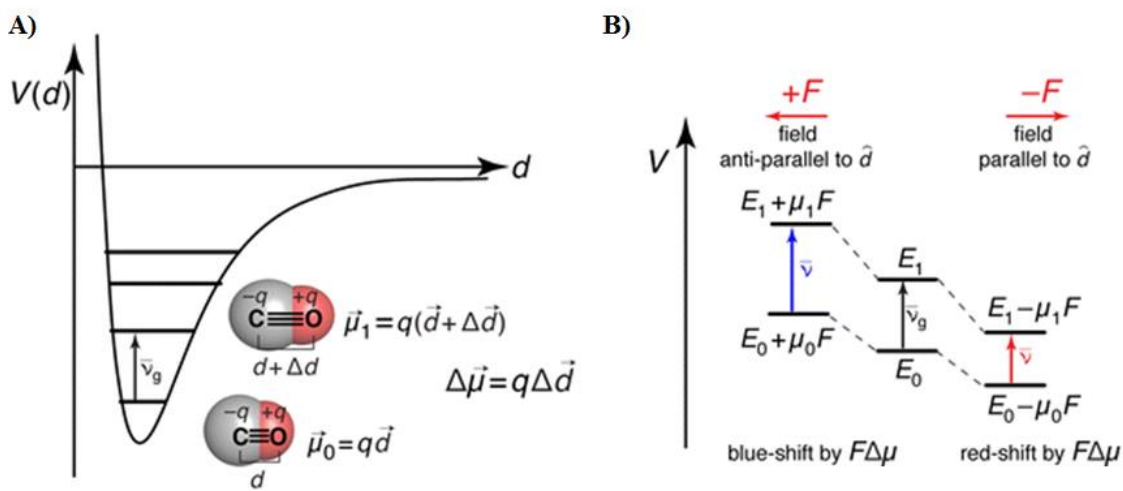
Figure 2.1*: Vibrational Stark effect can be explained in terms of anharmonicity. (A) The anharmonic form of molecular potential energy surfaces implies that bonds will be slightly longer (and so possess slightly larger dipole moments) in their vibrational excited states; ($\vec{d}$) denotes the unit vector aligned with the diatom (CO) bond axis. (B) Therefore, ground and excited vibrational states are stabilized differently by an electric field, resulting in a shift in the transitionenergy (vibrational frequency). The left side energy gap (blue arrow) corresponds to anti parallel arrangement between $\vec{F}$ and $\vec{d}$, the middle energy gap (black arrow) corresponds to the perpendicular arrangement between $\vec{F}$ and $\vec{d}$, and the right side energy gap (red arrow) corresponds to the parallel arrangement between $\vec{F}$ and $\vec{d}$. This figure is taken from the reference 1.*

($\Delta\vec{\mu}$) between ground and first excited vibrational state is established. In presence of an electric

field ($\vec{F}$), the energy gap between these two states is modulated by the scalar product of $\Delta\vec{\mu}$ and $\vec{F}$ vectors i.e. $|\vec{F}||\Delta\vec{\mu}|cos\theta$.[3] Where θ is the angle between $\vec{F}$ and the $\Delta\vec{\mu}$ which lies along the unit vector ($\vec{d}$) of the diatom bond axis (Figure 2.1A). Three different extreme orientations are possible between the $\vec{d}$ and the $\vec{F}$: i) parallel ($\theta = 0°$), ii) anti-parallel ($\theta = 180°$), and iii) perpendicular ($\theta = 90°$). These three different types of orientations lead to the generation of the three different energy gaps between ground and the first excited states of the diatomic oscillator (Figure 2.1B). Such changes in the vibrational energy gap in presence of electric field is referred to as the vibrational Stark effect in this thesis. Since IR spectroscopy probes the energy gaps between the first two vibrational energy levels, the changes in energy gap is seen as a change in the vibrational frequency in field dependent IR study. Within the limit of electrostatic interaction, the frequency change ($\Delta\bar{\nu}$) is linearly related to the field change ($\Delta\vec{F}$) by the linear Stark equation:[3]

$$\Delta\bar{\nu} = \Delta\vec{\mu}.\Delta\vec{F} \qquad\qquad\qquad 2.2$$

$\Delta\vec{\mu}$ is also known as the Stark tunning rate. As the $\Delta\vec{\mu}$ is directly related to the sensitivity of the vibrational probe to an external electric field, it is referred to as the Stark tuning rate. The Stark tuning rate of several vibrational probes has been reported. The unit of Stark tuning rate, cm^{-1}/ (MV/cm), corresponds to the frequency shift (in cm^{-1}) that is accompanied by the application of a unit electric field (1 MV/cm= 0.01 V/Å) projected along the axis of vibration. In a vibrational Stark experiment, the molecule containing the vibration of interest is dissolved in a glass forming solvent and rapidly cooled to a frozen state using liquid nitrogen.[4] Absorption spectra are collected both in the presence and absence of an external electric field. By analyzing small differences between the absorption spectra, Stark tuning rate is estimated. Once $\Delta\vec{\mu}$ is determined, direct experimental prediction of electric fields from the surroundings can be made from simple IR absorption experiments using site-specific vibrational probes.

A simpler experiment to estimate the probe's sensitivity to the surroundings consists of vibrational solvatochromism where the molecule is dissolved in solvents of varying polarity. The normal mode frequencies and transition dipoles of vibrational probes are strongly affected by local electric field (or solvent reaction field) induced by surrounding solvent molecules.[5,6] Several empirical models exist to characterize the solvent electric field.[7] However, molecular dynamics (MD) simulations captures the field information in a microscopic manner. Calculating electric fields from MD

simulations and correlating the solvent-by-solvent frequency variation for an individual solute against the average solvent field calculated for each solvent have demonstrated a linear frequency-field correlation which is conceptually similar to that predicted by vibrational Stark effect.[8] This correlation can also be used successfully to estimate electric field by measuring the probe's response in IR. In this thesis we have estimated the electrostatic interactions using the solvatochromic approach. The details of the field calculation is described in the computational section.

2.2 2D IR spectroscopy

Due to the contribution from thermal energy, molecules are inevitably never at rest even at normal temperature at thermal equilibrium condition. The probabilities of different conformational states of a molecule are governed by the equilibrium thermodynamics. Whereas, energy required to cross the barrier between states is related to dynamics. Dynamics can be quantified by measuring the time scale. A slow timescale corresponds to a high energy barrier and a fast timescale correspond to a low energy barrier. Time resolved nuclear magnetic resonance (NMR) can access timescale which falls in the range of ms to µs. 2D IR is analogous to 2D-NMR and 2D IR provides important information on the structural fluctuations which happens at ultrafast timescales (fs to ps).

2D IR is a third order non-linear spectroscopic technique, where three successive femtosecond IR pulses are focused at the sample to induce the subsequent emission of the non-linear signal known as the vibrational photon echo, which is detected with frequency and phase resolution by an IR array detector.[9] Pulse sequence of a 2D IR experiment is shown in figure 2.2A. The time interval between pulses 1 and 2, is known as the coherence period (τ). Where as the time interval between pulse 2 and 3 is known as the population period/ waiting time (T_W). The time delay between the third laser pulse and the detector signal is termed as detection period (t). The action of the pulses are shown in Figure 2.2B. The first pulse generates a coherence in the 0-1 states. The second pulse stops the coherence and fixes the molecules either 0 or 1 state. Finally third pulse generates coherences in 0-1 and 1-2 states. A 2D IR spectrum is a correlation map between the pump frequencies (at the coherence period) and the probe frequencies (at the detection period). For any oscillator, a typical 2D IR spectrum consists of a pair of peaks along the diagonal (Figure 2.2C). The peak which appears on the diagonal is due to the 0→1 transition. Whereas, the off diagonal peak appears due to the 1→2 transition, separated by the vibrational anharmonicity from the diagonal peak. The phases of the diagonal and the off-diagonal peaks are opposite to each other (as shown in Figure 2.2C with

different colors). The presence of cross peaks denotes either dipolar coupling (if present at $T_W = 0$) or energy relaxation/ chemical exchange (if they evolve over increasing waiting time). Waiting time dependent 2D-IR spectra provide estimates about ultrafast dynamics timescales. By varying the time interval between pulses 2 and 3, contributions of dynamics to the spectral lineshape arising due to spectral diffusion, ultrafast fluctuations and chemical exchange can be obtained. At $T_W = 0$, the system is examined just after the excitation and hence a strong correlation between the pump and probe frequency gives the diagonally elongated spectrum (Figure 2.2D, peaks arising due to only 0→1 transition are shown for simplicity). At non-zero waiting time, the system relaxes and the correlation is gradually lost as a function of T_W. The shape of the 2D IR spectrum becomes circular at long enough waiting times. By analyzing the waiting time dependent change in the 2D IR spectral lineshape, the fluctuation timescales of ultrafast molecular motions can be estimated. Similarly in presence of an equilibrium exchange dynamics, the time evolution of the cross peaks can be analyzed to obtain the chemical exchange timescale. In Figure 2.2D, the appearance of cross peaks are shown for a finite value of the population period. The cross peaks also appear for the off-diagonal 1→2 transitions.
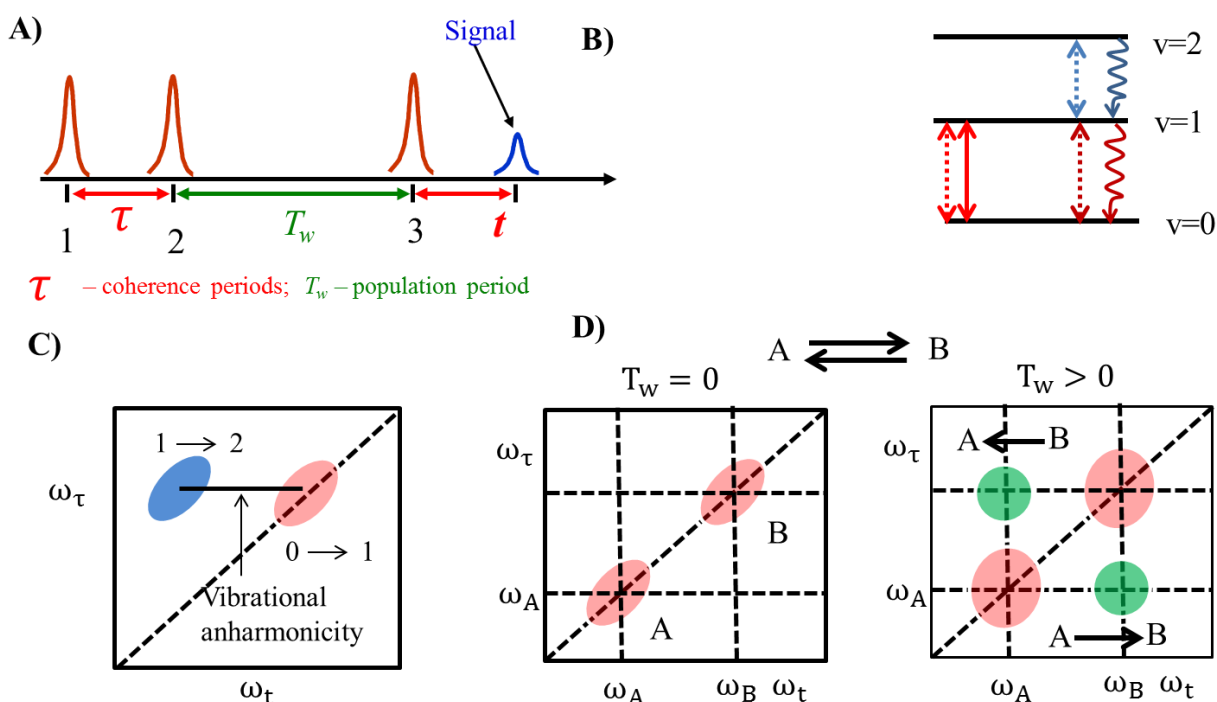
Figure 2.2: *Schematic diagram of the 2D IR spectroscopy. (A) Pulse sequence used in a 2D IR spectroscopy. (B) Vibrational energy levels. (C) A 2D IR spectra. The pink peak along the*

diagonal is arisen due to the 0 →1 transition and the blue, off-diagonal, peak is seen due to 1→2 transition. (D) 2D IR with exchange. Exchange between only 0 →1 transition (green cross peaks) is shown for simplicity. However, the cross peaks due to exchange also appear for the 1→2 transition.

The 2D IR data used for the completion of the thesis have been recorded by prof. Yung-Sam Kim of the Dept of Chemistry, UNIST, Korea. The 2D IR technique of our lab is presently at development stage.

2.3 Molecular dynamics simulations

All atom molecular dynamics (MD) simulations were carried out using the GROMACS[10] (version: 4.6.x) software. The details of the force field parameters will be given in the respective working chapters. Here we will discuss briefly the general methods involved in the MD simulations and how we have calculated the electric field. The standard steps involved in the simulations are following: (i) A cubic simulation box of required dimension was prepared. (ii) The probe molecule was inserted into the centre of the box. (iii) Solvent molecules were put into the box. (iv) The solvated system was sequentially processed through a) energy minimization using steepest descent algorithm, b) equilibration in NVT ensemble at 300K temperature for 100ps using the velocity rescale thermostat,[11] and c) equilibration in NPT ensemble at 300K temperature and 1 bar pressure (using Parrinello-Rahman barostat) for 1ns.[12] The positions of the solute molecules were restrained during NVT and NPT equilibrium step. (v) Finally, the production MD run was given with positions and forces on each atom being saved at every 0.2 ps. The electric field (F_{A-B}) exerted onto the A-B bond by the environment was calculated by projecting the total electrostatic field due to the rest of the system onto the A-B bond vector following the same protocol as used by Boxer and co-workers.[8] The equations used to define (F_{A-B}) are the following,

$$\vec{F}_i = \vec{f}_{i,\ electro}/q_i \qquad\qquad [2.3]$$

$$F_{A-B} = \frac{1}{2} < \vec{F}_A \cdot \vec{\mu}_{A-B} + \vec{F}_B \cdot \vec{\mu}_{A-B} > \qquad\qquad [2.4]$$

where, $\vec{F}_i$ denotes the electric field and $\vec{f}_{i,\ electro}$ denotes the total electrostatic (Coulomb) force exerted on the i-th reference atom (A and B atoms of the A-B bond), and q_i denotes the partial charge obtained from the force field. $\vec{\mu}_{A-B}$ is the unit vector along the A-B bond. Thus, the reported F_{A-B} values were averaged over both the A and B atoms as well as over the

configurational ensemble along the MD trajectory until convergence was achieved. For complete generality and consistent treatment of the electric field, the total exerted electric field was calculated from the full solvation system.

2.4 Quantum calculations

All the calculations in this thesis have been performed with density functional theory (DFT), with the aid of the Gaussian09 suite of programs.[13]

2.4.1 Geometry optimization

Geometry optimizations were performed with "tight" convergence criteria. Initial structures were generated based on the chemical intuition. Harmonic frequency calculations were performed for all points to confirm them as either local minima or transition state structures.

2.4.2 NBO calculations

Natural bond orbital (NBO) analysis were performed using NBO 6.0 program developed by Weinhold and co-workers.[14,15] The NBO method enables to understand the fundamental concept of the localized Lewis-like chemical bonds. The delocalized molecular orbitals of the molecule or complex are converted into localized Lewis type and non-Lewis type natural bond orbitals (NBOs). A Lewis-type NBO, also called as donor orbital, has filled electrons and lower in energy while the non-Lewis type (anti-bonding NBO), also called as acceptor orbital, has low occupancy (vacant) and higher in energy. Non-covalent interactions like $n \rightarrow \pi^*$ and hydrogen bonding, involves delocalization of electron density from the filled bonding NBO to the empty anti-bonding NBO which results in the stabilization of the system. This delocalization energy is determined by the second order perturbative energy, $E^{(2)}_{i \rightarrow j*}$, where, i and j* stand for the donor and acceptor orbitals. It measures the extent of the overlap of donor orbital (i) with the acceptor (j*) anti-bonding orbital and hence the strength of the interaction. NBO analysis has been used to determine the strength of the $n \rightarrow \pi^*$ and the hydrogen bonding interactions present in phenyl formate and phenyl formate…water complex mentioned in chapter 4.

The second order perturbative energy is given by the equation:

$$E(2) = \Delta E_{ij} = q_i \frac{F(i,j)^2}{\varepsilon_j - \varepsilon_i} \qquad [2.5]$$

where q_i is the donor orbital occupancy (2 for closed-shell, 1 for open-shell), ε_i, ε_j are diagonal elements (orbital energies), and F(i,j) is the off-diagonal NBO Fock matrix element.

The higher is the second order perturbative energy value for a given donor-acceptor pair, the greater is the strength of the interaction. NBO analysis has been used extensively to visualize the n→π* interaction in the molecules, to determine its strength and its effect on nearby orbital interactions.

2.4.3 Geometry optimization in presence of electric field

We ventured to see how free energy of activation of a reaction gets affected in presence of electric field. In this regard, we optimized the transition state structure and the structures of the reactants. We then found out the unit vector along the dipole moment of either transition state or reactants. Again we re-optimized the transition state geometry in presence of an electric field which was applied in the opposite direction of the dipole moment unit vector using "Field=Read" keyword. The "Nosymm" keyword was used to avoid the rotation of the Cartesian axes during the process of optimization. Geometry optimizations were done at electric field strengths between 0.0V/Å to -0.70 V/Å at -0.10 V/Å (-0.10 V/Å= 0.1950×10^{-2} au) intervals.

2.5 References

(1)   Fried, S. D.; Boxer, S. G. Measuring Electric Fields and Noncovalent Interactions Using the Vibrational Stark Effect. *Acc. Chem. Res.* **2015**, *48*, 998-1006.

(2)   Walker, J.; Resnick, R.; & Halliday, D.; Hoboken, N. J. W. S., Inc.. Halliday & Resnick fundamentals of physics (Tenth edition.) *Hoboken, NJ: John Wiley & Sons, Inc..* **(2014)**.

(3)   Boxer, S. G. Stark Realities *J. Phys. Chem. B* **2009**, *113*, 2972-2983.

(4)   Suydam, I. T.; Boxer, S. G. Vibrational Stark Effects Calibrate the Sensitivity of Vibrational Probes for Electric Fields in Proteins. *Biochemistry* **2003**, *42*, 12050-12055.

(5)   Choi, J.-H.; Cho, M. Vibrational Solvatochromism and Electrochromism of Infrared Probe Molecules Containing C≡O, C≡N, C=O, or C−F Vibrational Chromophore. *J. Chem. Phys.* **2011**, *134*, 154513.

(6)   Liptay, W. Electrochromism and Solvatochromism *Angew. Chem., Int. Ed.* **1969**, *8*, 177-188.

(7)   Fried, S. D.; Wang, L.-P.; Boxer, S. G.; Ren, P.; Pande, V. S. Calculations of the Electric Fields in Liquid Solutions. *J. Phys. Chem. B* **2013**, *117*, 16236-16248.

(8)   Fried, S. D.; Bagchi, S.; Boxer, S. G. Measuring Electrostatic Fields in Both Hydrogen-Bonding and Non-Hydrogen-Bonding Environments Using Carbonyl Vibrational Probes. *J. Am. Chem. Soc.* **2013**, *135*, 11181-11192.

(9)   Kashid, S. M.; Deb, P.; Haldar, T.; Roy, P. P.; Bagchi, S. Measuring Equilibrium Ultrafast Structural and Conformational Dynamics in Bio-molecules: A Two Dimensional Infrared (2D-IR) Spectroscopic Approach. *ScienceJet .,* **2015**, *4*, 173.

(10)   Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435-447.

(11)   Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling *J. Chem. Phys.* **2007**, *126*, 014101.

(12)   Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182-7190.

(13)   Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J., A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. B., J. C.; Iyengar, S. S. T., J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A. S., P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Wallingford, CT, 2009.

(14)   Weinhold, F.; Landis, C. R. Valency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective. *Cambridge university press, Cambridge, UK.* **2005**.

(15)   Reed, A. E.; Curtiss, L. A.; Weinhold, F. Intermolecular Interactions From a Natural Bond Orbital, Donor-Acceptor Viewpoint. *Chem. Rev.* **1988**, *88*, 899-926.

Chapter 3

Unraveling Non-Covalent Interactions in Biomolecules Using Nitrile Probe

3.1 Introduction

Non-covalent interactions play a crucial role towards bio-molecular functions.[1-4] A key challenge in the fields of biochemistry and biophysics is to understand the role of these interactions towards molecular recognition, protein folding and enzyme catalysis within the electrostatic environments of the proteins and enzymes. Some of the commonly used descriptors of non-covalent interactions include dissociation constant, hydrogen-bonds (H-bonds), and interactions involving π-electron cloud of aromatic residues.[1] These concepts are either difficult to relate to microscopic parameters or are defined by arbitrary geometrical criteria.[1] Recently a wide attention has been drawn into vibrational spectroscopy, as infrared (IR) probes are sensitive reporters of the local electrostatic environments.[5-15] Vibrational Stark effect (VSE) predicts the linear dependence of the vibrational frequencies on the electric field due to the environment $(\Delta \bar{v} = \Delta \vec{\mu} . \Delta \vec{F})$.[16] It has been suggested that electric field can serve as a quantitative and microscopic parameter, independent of the specificity of the intermolecular interactions.[1] The inherent advantage of IR based methodology is that VSE allows us to predict the electric fields using vibrational probes incorporated into specific protein residues or ligands, providing a quantitative understanding of the non-covalent interactions.[16-18] One of the numerous IR probes in focus to estimate the electric field inside proteins is the nitrile (C≡N) chemical group.[6,18-25] C≡N stretch is a local mode and absorbs in the mid-IR spectral region devoid of any background absorption from the protein background and side-chains. Interestingly, C≡N is also a very common substituent in small molecules used as inhibitors, including a remarkable number of drugs currently in use or in clinical trials.[26] Several reports are now available for introducing C≡N probes into proteins.[6,18,19,22-24] These advantages suit C≡N as an ideal IR probe to investigate protein structure, interactions, and dynamics.[27-30]

One unusual characteristic of the C≡N group, as compared with other commonly used IR probes (e.g. C=O, O-H, N-H), is the shift of the nitrile stretch to higher frequencies (blue-shift) in the IR spectrum upon H-bonding.[31-35] However, unlike carbonyls, it has been suggested that the H-bonding interactions in nitriles are not purely electrostatic [30,33,35-37] [31,34,36-39] and the non-electrostatic contribution to nitrile stretch can be attributed to polarization and other quantum mechanical aspects of H-bonding interactions.[30,33,35,36] This makes the estimation of electric field elusive as VSE cannot describe nitriles in the H-bonding environments.[33,35,36] To date, this has

severely limited the utility of C≡N IR probes because H-bonds are often critical components of the environment one seeks to characterize.

In this work, we have used a combination of IR spectroscopy and atomistic classical molecular dynamics (MD) simulations to obtain the nitrile stretching frequencies and to estimate the average electric fields $(F_{C≡N} = \vec{F}.\vec{u}_{C≡N})$[38] on the C≡N bond of an aromatic (benzonitrile, PhCN) and an aliphatic (methyl thiocyanate, MeSCN) nitrile exerted by different solvation environments (both H-bonded and non-H-bonded) of varying polarity. Surprisingly, our results demonstrate that a linear correlation between electric field and IR frequency still exists, albeit not described by VSE, for H-bonded nitriles. As the linear field-frequency correlation for the blue-shifting H-bonded nitriles has never been reported till date, we have performed various stringent tests on molecules of varying sizes, structural compactness and solvation environments to validate the robustness of our results. We have shown the validity of the field-frequency correlation on small molecule nitriles in presence of high concentration of a denaturant (urea) in aqueous solution. It has been reported earlier that urea, a H-bond donor, can induce spectral shifts in H-bonding nitrile stretching frequencies.[39] We have tested the extendibility of the field-frequency correlation for H-bonded nitriles to a field value larger than that obtained in case of the model nitrile by using an aromatic nitrile modified cysteine in water. In case of the nitrile modified cysteine, the nitrile peak frequency is blue shifted to that of PhCN in aqueous solution. Using the above mentioned correlation data as a calibration curve, and the estimated electric fields from classical MD simulations, we have predicted the nitrile peak frequencies of multiple mutants of two nitrile modified peptides and compared them with the experimental observations. The two peptides are chosen such that an amino acid has been modified to an aromatic nitrile in one and to an aliphatic nitrile in the other. The nitrile group of these peptides were previously reported to be exposed to the aqueous solvent environment.[31,40] Moreover, the predicted electric fields of the nitrile group from field-frequency correlations in case of a nitrile conjugated to an amino acid of a protein, have been compared with that estimated from classical MD simulation to obtain a quantitative estimate of non-covalent interactions in H-bonded nitriles from experimental IR frequencies. Overall, we have used two different nitriles (aromatic and aliphatic) in a wide range of solvents/solvent mixtures of varying polarity, and have successfully tested the predictive power of our field-frequency correlation curve on nitrile modified amino acid, disordered peptides and a structured protein system. Our results illustrate the utility of nitrile

chemical group as an effective reporter of the electric field irrespective of the specificity of the non-covalent interactions.

3.2 Results and discussions

3.2.1 Vibrational solvatochromism

Nitrile probes are generally introduced into proteins and peptides as aliphatic thiocyanates (SCN) via cysteine modification (Cys-SCN) or as aromatic nitriles via cyano phenylalanine (p-CN-Phe). It has been reported earlier that benzonitrile (PhCN) and methyl thiocyanate (MeSCN) can serve as model nitriles for p-CN-Phe and Cys-SCN, respectively.[38,41] We have performed solvatochromic IR experiments PhCN and MeSCN, where different organic solvents of varying polarity and neat water serve as the benchmark solvents to elucidate the nature of nitrile (C≡N) group in different electrostatic environments. Further, the nitriles are dissolved in aqueous binary mixtures of dimethyl sulphoxide (DMSO) and dimethylformamide (DMF) with varying water content (v/v) to illustrate the sensitivity of the nitrile stretching frequencies ($\bar{v}_{C \equiv N}$) in a wide range of H-bonding environments. The solvatochromic linear IR absorption experiments are performed with a 10 mM concentration of the nitrile in neat solvents or in binary solvent mixtures. The technical details of vibrational spectroscopic study is given in section 3.4.

The solvation environments of the nitriles were scanned through a wide range of polarity by dissolving in various solvents having different dielectric constants. The effect of solvent polarity on the nitrile group is reflected in the linear absorption spectra of the C≡N stretching mode (Figure 3.1). In the most nonpolar solvent, hexane, the nitriles PhCN and MeSCN show peak maxima at 2233.4 cm^{-1} and 2164.3 cm^{-1} respectively. The C≡N peak frequencies, $\bar{v}_{C \equiv N}$, gradually decrease (red-shift) in aprotic solvents with increasing solvent polarity. In the most polar aprotic solvent, DMSO, the nitrile peak maxima for PhCN and MeSCN are at 2227.5 cm^{-1} and 2154.2 cm^{-1}, respectively. A 5.9 cm^{-1} red shift in frequency is observed for PhCN as the solvation environment is changed from hexane to DMSO; MeSCN shows a red-shift of 10.1 cm^{-1} for the same. The greater red-shift in the case of MeSCN indicates its greater sensitivity towards solvent polarity as compared with PhCN. This result reinforces the earlier results from VSE spectroscopy experiments, where the aliphatic nitriles are found to have a larger Stark tuning rate than the aromatic nitriles.[24,25] However the overall red-shifting trend from hexane to DMSO is

the same for both the nitriles and matches with the earlier reported results.[1,24,25] A similar trend in aprotic solvents has been reported for carbonyl (C=O) vibrational probes where a red-shift in
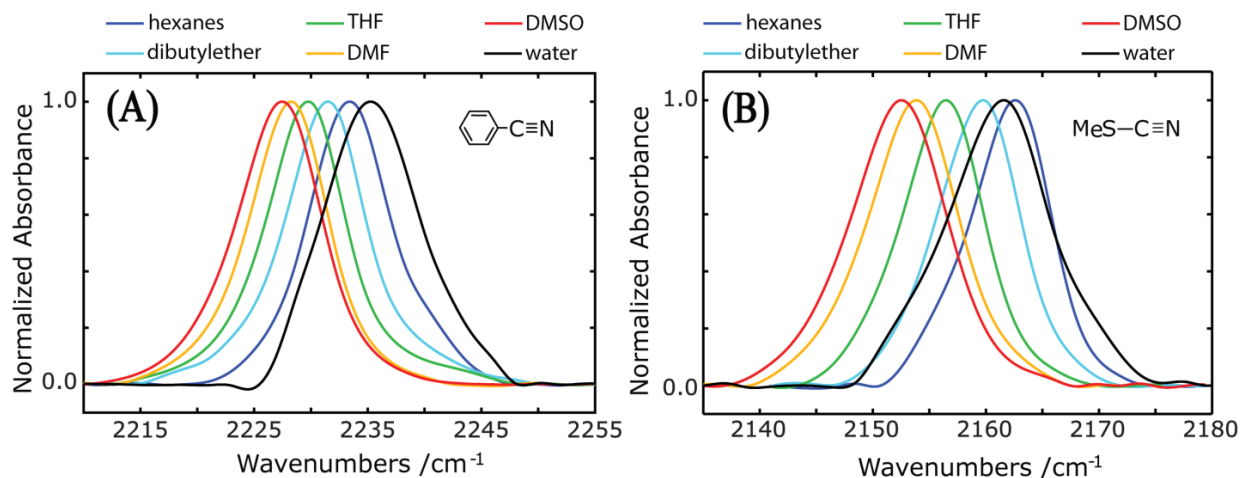
Figure 3.1*: FTIR absorption spectra of the nitrile for (A) PhCN and (B) MeSCN in different solvents at 10 mM concentration. With increasing solvent polarity from hexane to DMSO, the nitrile stretching frequency gradually decreases (red shift).THF stands for tetrahydrofuran. In the protic solvent, water, the nitrile stretching frequency shows a blue shift as compared to DMSO.*

the C=O stretching frequency is also observed. In addition, C=O shows an even larger red-shift when dissolved in a polar protic solvent like water.[1,38,42-44] In contrary, when PhCN or MeSCN is dissolved in water, the C≡N stretching frequency shows a blue shift as compared with the peak frequency in DMSO (Figure 3.1). It has been well established that nitriles accept a H-bond when exposed to water and the blue shift in H-bonded nitrile stretching frequency contradicts the interaction to be a strictly dipole-dipole interaction based on solvatochromic surveys and ab-initio quantum mechanical calculations.[31,33,36,45-49] It has been suggested that the H-bonding interactions in nitriles consists of non-electrostatic contributions, thereby the frequency shifts in H-bonded nitriles are not described by VSE.[24,33,48]

The blue-shifted C≡N stretching frequencies in water as compared with the red-shift expected from a purely dipole-dipole interaction is quite intriguing. To check the response of nitriles in a wide range of H-bonding solvation environments of intermediate polarities, linear IR absorption experiments were performed in aqueous binary solvent mixtures of DMSO and DMF. Upon gradually increasing the water content with 10% (v/v) increments, the C≡N stretching

frequencies of both PhCN and MeSCN show an increasing blue-shift from that in DMSO (Figure 3.2). Similar behaviour has been seen in DMF-water mixtures also.
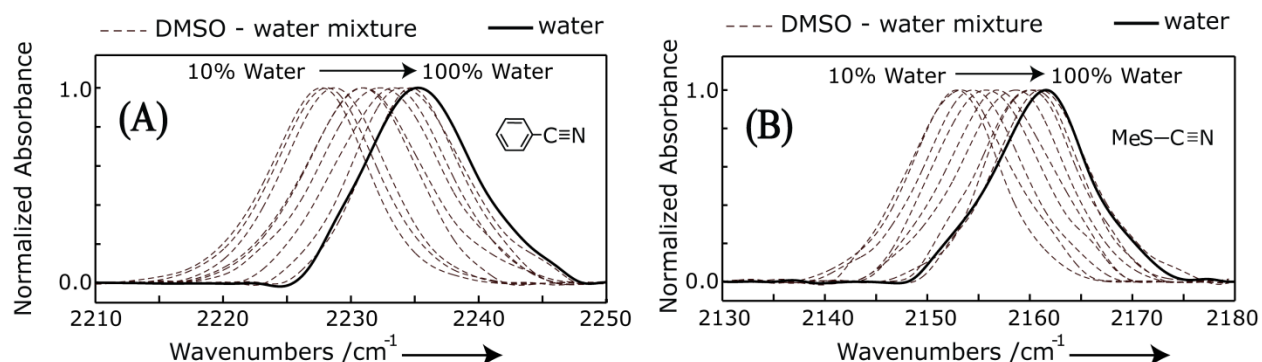
Figure 3.2: *FTIR spectra of C≡N stretching band of (A) PhCN and (B) MeSCN in water and DMSO-water binary solvent mixtures at 10 mM concentrations. Nitrile stretching frequency gradually increases upon increasing the water content in DMSO/water solutions by 10% (v/v). Whereas, a similar set of linear IR experiments on C=O vibrational probes in DMSO/water solution reported a monotonic red-shift in C=O frequencies from that in DMSO to water.[43] However, the monotonically increasing blue-shift in C≡N frequency with increasing water content of illustrates that the C≡N groups of PhCN and MeSCN are H-bonded in all the DMSO/water solutions of varying water content.*

3.2.2 Electric field - IR frequency correlation

Average electric field exerted on to the C≡N group, $F_{C≡N}$ by different solvents are calculated using atomistic MD simulation (the details of the simulation protocols is given in section 3.5). Electric fields in aprotic solvents show an excellent linear correlation with the experimentally recorded frequencies of both PhCN (Figure 3.3A) and MeSCN (Figure 3.3B). The slopes of the best fit lines are 0.22 and 0.37 for Figure 3.3A and 3.3B, respectively. These slopes are theoretically equivalent to the Stark tuning rates, which have been independently measured for the nitrile group in both aromatic and aliphatic nitriles.[24,25] The reported values of the Stark tuning rate for aromatic aliphatic nitriles are 0.61 cm^{-1}/(MV/cm) and 0.72 cm^{-1}/(MV/cm), respectively.[24,25] It is noteworthy that the solvatochromic slopes are smaller by factors of 2.77 and 1.95 from the corresponding experimentally measured Stark tuning rates. This difference arises from the local field effect and is consistent with the recently reported value of the local field correction factor, f ~ 2 .[1] Henceforth in this manuscript, the estimated electric fields from MD simulations are reported after applying a correction factor of 2.77 and 1.95 for
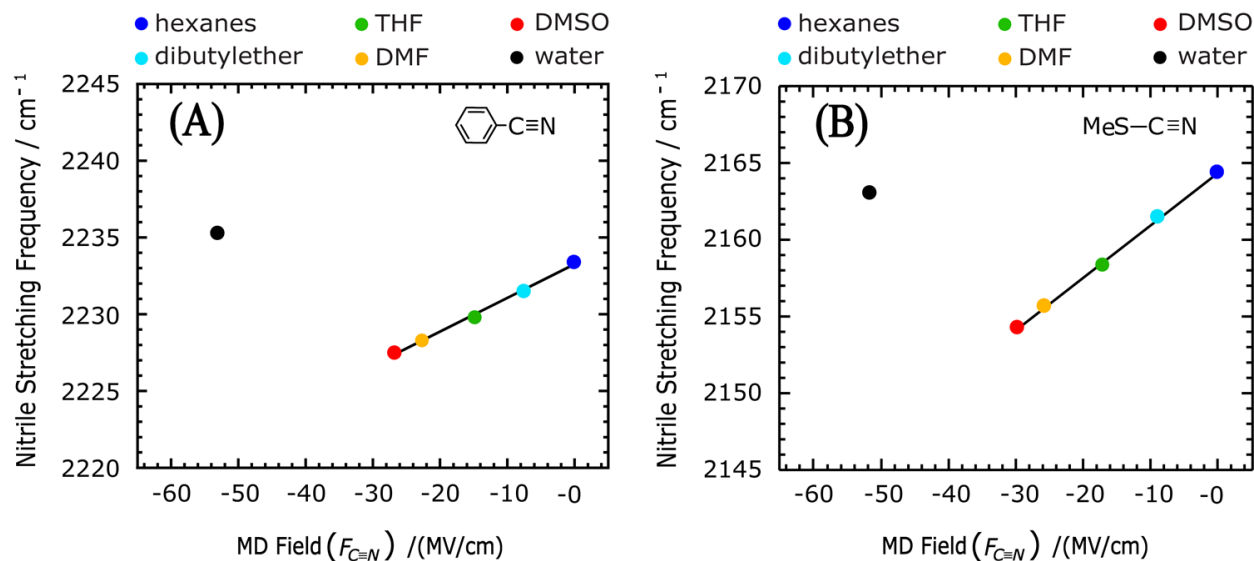
Figure 3.3*: Field-frequency calibration for (A) PhCN and (B) MeSCN in non-H-bonding solvents. In purely electrostatic environment $\bar{v}_{C\equiv N}$ varies linearly with the electric field for both aromatic and aliphatic nitriles. The equations of the best-fit lines for (A) PhCN and (B) MeSCN are $\bar{v}_{C\equiv N} = 0.22 \ (F_{C\equiv N}) + 2233.3 \ (R^2 = 0.99)$ and $\bar{v}_{C\equiv N} = 0.35 \ (F_{C\equiv N}) + 2164.3 \ (R^2 = 0.99)$ respectively. The field-frequency paired point (black circle) for the nitriles dissolved in water (H-bonded nitrile) falls off the best-fit lines. After correcting the electric field values to make the slopes of the best-fit lines the same as corresponding Stark tuning rates, the corrected equations for (A) PhCN and (B) MeSCN are $\bar{v}_{C\equiv N} = 0.61 \ (F_{C\equiv N}) + 2233.3 \ (R^2 = 0.99)$ and $\bar{v}_{C\equiv N} = 0.72 \ (F_{C\equiv N}) + 2164.3 \ (R^2 = 0.99)$ respectively.*

PhCN and MeSCN respectively to make the slopes equal to the respective Stark tuning rates. The electric fields calculated from MD simulations for neat solvents, along with the corresponding corrected field values are tabulated in Table 3.1 and equations of the best fit lines of the two plots using the corrected fields are given in the captions of Figures 3.3A and 3.3B.

Although nitriles show excellent field-frequency correlations in aprotic solvents, the correlations are violated when the nitriles are dissolved in a protic solvent (e.g. water). In the field-frequency plots (Figure 3.3A and 3.3B), the $(F_{C\equiv N}, \bar{v}_{C\equiv N})$ ordered point arising from neat water deviates from the best fit lines for both the nitriles. Earlier reports where Onsager reaction field model was used to estimate the electric fields, along with a recent report using MD simulations for the same, also predict similar results.[1,24,25] The electrostatic interactions of the C≡N group with the aprotic solvent molecules leads to the linear dependence of the electric fields to C≡N frequencies through VSE. However, due to the non-electrostatic contribution to H-bonded C≡N frequencies,

the frequency shifts upon changing the polarity of the H-bonding environment cannot be described by VSE. [24,33,36]

Table 3.1:MD estimated and corrected electric fields experienced by the nitriles in neat solvents.

solvents	MD estimated electric field / (MV/cm)		corrected electric field /(MV/cm)[a]	
	PhCN	MeSCN	PhCN	MeSCN
Hexanes	-0.08	-0.12	-0.03	-0.06
Dibutylether	-7.57	-8.16	-2.73	-4.18
Tetrahydrofuran (THF)	-14.85	-15.47	-5.36	-7.93
Dimethyl formamide (DMF)	-22.69	-23.55	-8.19	-12.08
Dimethyl sulphoxide (DMSO)	-26.81	-27.89	-9.68	-14.30
Water	-53.11	-49.59	-19.17	-25.43

[a]:*MD estimated fields are divided by 2.77 and 1.95 for PhCN and MeSCN respectively to get the corresponding corrected electric fields.*

To systematically understand the local electrostatics in H-bonding environments, we have estimated the electric fields experienced by C≡N vibrational probe in aqueous binary mixtures of DMSO and DMF using MD simulations. $F_{C\equiv N}$ for PhCN and MeSCN in different binary mixtures when plotted against the corresponding experimental frequencies (as shown in Figure 3.4) show a linear correlation with R^2 values of 0.99 and 0.98 respectively. The slopes of the best fit lines, 0.88 for both PhCN and MeSCN, are opposite in sign to that obtained in neat solvents. This illustrates that, unlike non-H-bonded nitriles, an increase in $|F_{C\equiv N}|$ leads to an increase in the IR frequencies for the H-bonded nitriles. The absolute values of the slopes being larger for the H-bonded nitriles indicate a larger blue shift in $\overline{\nu}_{C\equiv N}$ as compared with the red-shift

observed for the non-H-bonded nitrile for the same change in $|F_{C\equiv N}|$. Figure 3.4A and 3.4B show the field-frequency correlation for H-bonded nitriles, for PhCN and MeSCN, respectively. The MD estimated and the corrected fields exerted on the C≡N group by various binary mixtures of are given in Table 3.2 and Table 3.3.
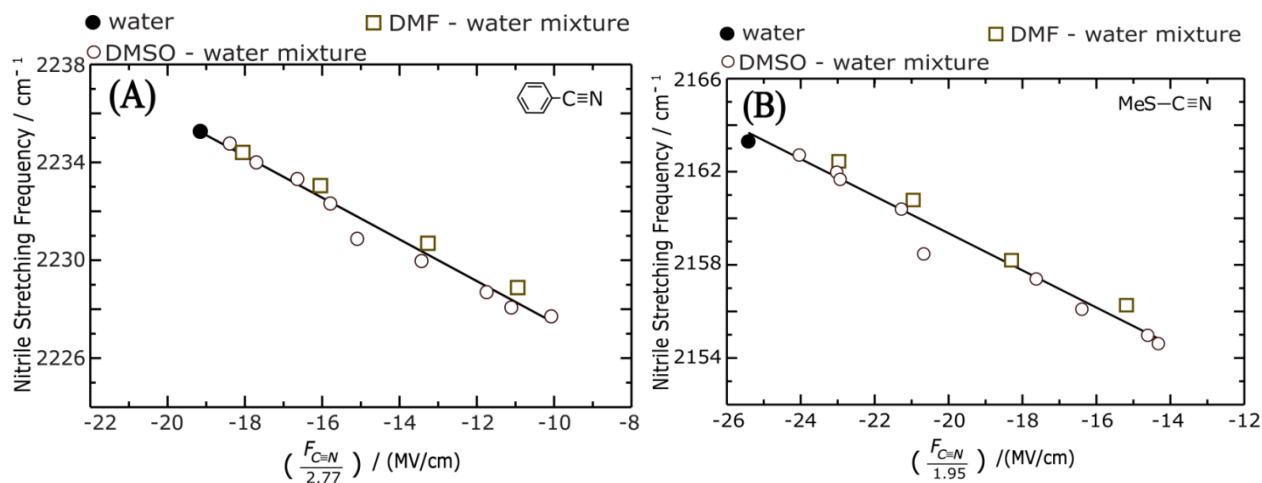
Figure 3.4: *Field-frequency calibration for (A) PhCN and (B) MeSCN in of DMSO/water and DMF/water binary mixtures (H-bonding environments). A linear correlation between nitrile frequency and the electric field exerted by the H-bonding environments exists for both the aromatic and sulfur based nitriles. The equations of the best fit lines for (A) PhCN and (B) MeSCN are $\bar{v}_{C\equiv N}$ = -0.88 ($F_{C\equiv N}$/2.77) + 2218.4, (R^2 = 0.987) and $\bar{v}_{C\equiv N}$ = -0.80 ($F_{C\equiv N}$/1.95) + 2143.4, (R^2 = 0.97), respectively.*

Table 3.2: *MD estimated and corrected electric fields experienced by the nitriles in different DMSO-water mixtures.*

% of water in DMSO-water mixtures.	MD estimated electric field / (MV/cm)		corrected electric field / (MV/cm) [a]	
	PhCN	MeSCN	PhCN	MeSCN
10	-27.98	-27.99	-10.10	-14.36
20	-30.77	-28.42	-11.11	-14.82
30	-32.59	-31.94	-11.77	-16.16
40	-37.20	-34.4	-13.43	-17.38

50	-41.84	-40.42	-15.10	-20.02
60	-43.77	-41.48	-15.80	-21.46
70	-46.16	-44.71	-16.66	-22.04
80	-48.97	-44.91	-17.68	-22.41
90	-50.92	-46.94	-18.38	-23.40

[a] *MD estimated fields are divided by 2.77 and 1.95 for PhCN and MeSCN respectively to get the corresponding corrected electric fields.*

Table 3.3: *MD estimated and corrected electric fields experienced by the nitriles in different DMF-water mixtures.*

% of water in DMF-water mixtures.	MD estimated electric field / (MV/cm)		corrected electric field / (MV/cm) [a]	
	PhCN	MeSCN	PhCN	MeSCN
20	-30.4	-29.66	-10.97	-15.21
40	-36.87	-35.71	-13.31	-18.31
60	-44.45	-40.87	-16.05	-20.96
80	-50.02	-44.91	-18.06	-23.03

[a] *:MD estimated fields are divided by 2.77 and 1.95 for PhCN and MeSCN respectively to get the corresponding corrected electric fields.*

3.2.3 Significance of the field-frequency correlation in H-bonded nitriles

The electric fields, as shown in cases of PhCN and MeSCN, can be calculated directly from MD simulations. However, apart from the approximations involved in the force field, the determination of the electric field for large biological macromolecules like peptides and proteins is itself computationally expensive. To date, the linear correlation between $\overline{\nu}_{C \equiv N}$ and $F_{C \equiv N}$ for non-specific interactions of the nitrile group has been known from VSE.[24,25] This permitted the

quantitative estimation of the non-covalent interactions in nitriles and nitrile modified biomolecules directly from IR absorption experiments, provided the nitrile group is not H-

bonded. The non-electrostatic nature associated with H-bonded nitrile could not be described by VSE; this severely limited the use of a nitrile to probe specific non-covalent interactions. As specific interactions like H-bonding with side chains, bound ligands and neighbouring water molecules are prevalent in biomolecules, the nitrile chemical group, in spite of having several advantages as a biomolecular site-specific spectroscopic probe, could not be extensively used to quantify non-covalent interactions where the environment was not known a priori. The linear correlation reported here for H-bonded nitriles, coupled with the ideal nature of the nitrile IR probe, opens up a new and direct experimental avenue to report on the specific non-covalent interactions in biomolecules.

3.2.4 Validation of the field-frequency correlation in H-bonded nitriles

To test the generality of the linear dependence of the nitrile IR frequencies to the electric fields in H-bonded environments, we have investigated its robustness for H-bonded aromatic and aliphatic nitriles in a wide range of nitrile containing molecules of varying size, compactness and solvation environments using the calibration lines of Figure 3.4.

In aqueous binary mixtures, nitriles can only form H-bonds with the water molecules present. On the contrary, in presence of a denaturant like urea in aqueous environment, C≡N can form H-bonds either with water or urea molecules. It has been suggested that studying the effect of denaturants on the $\overline{v}_{C \equiv N}$ in aqueous solutions of model nitriles can provide insights about the mechanism of protein denaturation.[39] We have performed IR absorption experiments to obtain nitrile IR frequencies for PhCN and MeSCN in aqueous urea (5M) solution. The peak frequencies of the C≡N group of both the aliphatic and aromatic nitriles in 5M urea shifts to a lower value compared to that in aqueous solution; a similar red-shift of the C≡N stretching frequency in the presence of urea has been previously reported by Gai and co-workers for acetonitrile.[39] From the respective calibration curves (Figures 3.4 A and B), we have predicted the electric fields, which agree with the independently estimated fields from MD simulations. This agreement validates the field-frequency calibration of H-bonded nitriles for small molecules in a complex solvation environment. The predicted and estimated electric fields are listed in Table 3.4.

To investigate whether H-bonded $\overline{\nu}_{C\equiv N}$ can quantitatively predict the electric fields in biomolecules, we have tested our calibration curve on an amino acid, since they serve as the building block of any peptide or protein. We have synthesized a modified cysteine (Compound **3** of the section 3.6 of this chapter) by incorporating a PhCN unit into the amino acid. The details of the synthesis of **3** are given in the section 3.6. The synthesis of **3**, using a slightly different protocol, has been reported earlier to demonstrate that the cysteine alkylation can serve as an effective alternative to introduce aromatic nitriles into peptides and proteins.[50] The experimental $\overline{\nu}_{C\equiv N}$ (2236.7 cm^{-1}) of the modified amino acid corresponds to a field value of 20.80 MV/cm as predicted by extrapolating the best fit line of Figure 3.4A. The MD estimated field for **3** shows

Table 3.4: *Comparison of the predicted electric fields with the MD estimated values in model nitriles in complex solvation environment, in a nitrile modified amino acid and in a nitrile modified protein.*

	nitrile stretching frequency /(cm^{-1})	figure used [a]	corrected electric field on nitrile /(MV/cm)		prediction error /(MV/cm) [b]
			predicted	estimated from MD	
PhCN in aqueous urea (5M) solution	2234.4	Figure 3.4 A	-18.18	-19.35	1.17
MeSCN in aqueous urea (5M) solution	2162.0	Figure 3.4 B	-23.38	-26.33	3.04
Nitrile modified amino acid (Compound 3)	2236.7	Figure 3.4 A	-20.80	-21.07	0.27
Nitrile Modified Protein (Compound 7)	2234.2	Figure 4	-17.96	-19.08	1.12
[p-CN-Phe]RNase S	2231	3.3A	-3.77	-2.82	0.95

[a] :indicates the best-fit line of the figure used to predict the electric field.
[b] :Prediction error is the absolute value of difference between the Predicted and the Estimated electric fields.

an excellent agreement and is within 0.27 MV/cm of the predicted field (Table 3.4). Here $\overline{\nu}_{C\equiv N}$ is 1.4 cm^{-1} higher than that of PhCN in neat water and thereby the value of the electric field is larger than the field experienced by the model nitrile (PhCN) in aqueous environments. This

result underpins the extendibility of the field-frequency correlation to higher absolute field values than that obtained for H-bonded model nitrile in water. The predicted and estimated electric fields on the nitrile group of the modified cysteine are listed in Table 3.4.

Further to validate the robustness of the calibration curves in peptides, multiple mutants of two peptides, one with an aromatic nitrile and the other with an aliphatic nitrile are surveyed. For each peptide we have performed MD simulations to estimate $F_{C\equiv N}$. The corresponding nitrile IR frequencies predicted using the calibration lines are compared with the earlier reported experimental results.[31,40] The C≡N group in the respective peptides has been known to be solvent (water) exposed in each of the mutants, thereby accepting a H-bond from the solvent molecule.[31,40] We have estimated the electric fields experienced by the C≡N probe along the

Table 3.5: *Comparison of the experimental and predicted IR frequencies in nitrile modified peptides.*

peptides	mutant position	corrected electric field on nitrile /(MV/cm)	figure used [a]	nitrile stretching frequency /(cm[-1])		prediction error [b] /(cm[-1])
				Predicted	Experimental	
Mastoparan X	6	-18.48	Figure 3.4A	2234.7	2234.9	0.2
	9	-19.32		2235.4	2235.1	0.3
	10	-19.35		2235.4	2235.2	0.2
CM15	4	-23.19	Figure 3.4B	2161.9	2161.6	0.3
	10	-22.91		2161.6	2161.2	0.4

[a]: indicates the best-fit line of the figure used to predict the IR frequency.[b] :Prediction error is the absolute value of difference between the Predicted and the Estimated frequencies.

nitrile bond for nitrile-modified mutants of Mastoparan X (Mp-X) and the antimicrobial peptide CM15. For the aromatic nitrile modified Mp-X, MD field calculations have been done for three mutants in water, where p-CN-Phe is introduced at positions 6, 9, and 10 of the peptide sequence in each mutant respectively. The predicted nitrile stretching frequencies along with the experimental values are listed in Table 3.5. In an earlier report, Gai and co-workers reported the

nitrile stretching frequencies for these three mutants and mentioned that the p-CN-Phe residues to be fully hydrated (i.e. C≡N is H-bonded) when dissolved in water.[40] Our predicted frequencies for the MP-X mutants match with the experimental frequencies within 0.3 cm^{-1} (Table 3.5). Further, we have performed MD simulations to find out $F_{C≡N}$ for two of the mutants of CM15 peptide (Cys-SCN introduced at either positions 4 and 10) using the same protocol used for Mp-X mutants. Our predicted frequencies are in excellent agreement with the experimental frequencies as recorded by Londergan and coworkers.[31] These results confirm that the field-frequency correlation is not only applicable to small nitrile containing molecules, but also for biologically relevant peptides. These peptides are inherently structurally disordered, and can exist in a wide distribution of conformations in aqueous solutions. The excellent agreement of the predicted and experimental $\overline{v}_{C≡N}$ suggests that site-specific non-covalent interactions can be quantitatively estimated in nitrile modified peptides in an inhomogeneously broadened aqueous environment.

In biological macromolecules like proteins and enzymes, theoretical determination of the electric field is computationally expensive. It would be extremely helpful to get the quantitative information of local electric field exerted on the nitrile modified proteins in H-bonding environments from simple IR experiments using the field-frequency calibration. Linear IR absorption experiments are performed to obtain $\overline{v}_{C≡N}$ of nitrile modified protein (compound **7** of
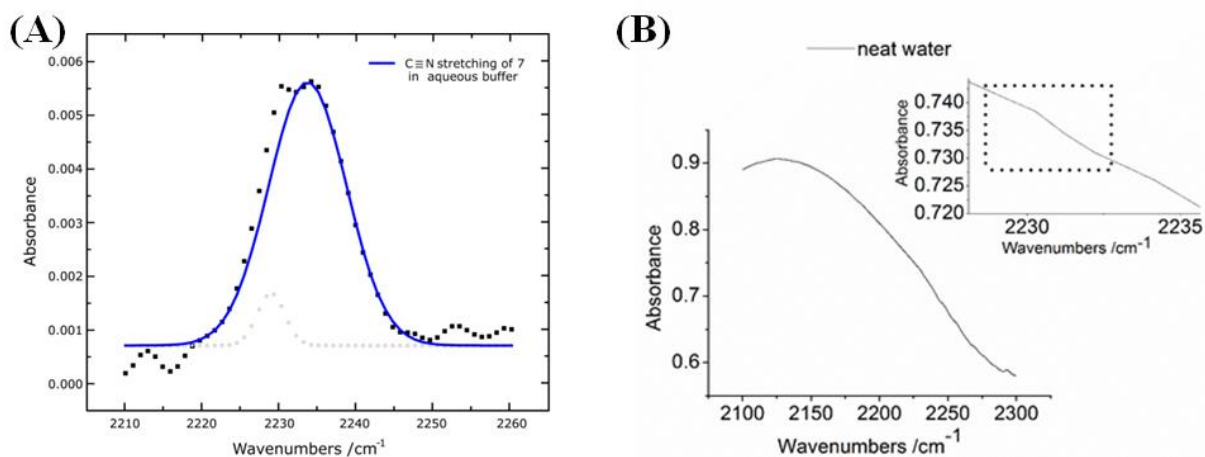
Figure 3.5: *(A) FTIR absorption spectra of the C≡N stretching band (at a resolution of 1 cm^{-1}) of **7** (~2.5mM) in aqueous buffer (40 mM NaH$_2$PO$_4$, pH=7.6) is represented by the black squares. The spectum has been fitted by using two Gaussian function. The high frequency peak (blue curve) at 2234.2 cm^{-1} is the nitrile stretching frequency of **7**. The low frequency peak (gray circles) at 2229.4 cm^{-1} arises from a very weak water band. The existence of this water band has*

*been confirmed by analysing the FTIR spectrum of neat water (B). A zoomed version of neat water spectra is shown in the inset of figure B. The weak peak contributes to the lineshape of the nitrile stretch when the nitrile peak intensity is low. The maximum possible concentration attained for the nitrile modified protein (**7**) is ~2.5 mM as compared to the 10 mM concentration used for the model nitriles. The peak arsing from water has also been verified multiple times with nitriles of known frequencies.*
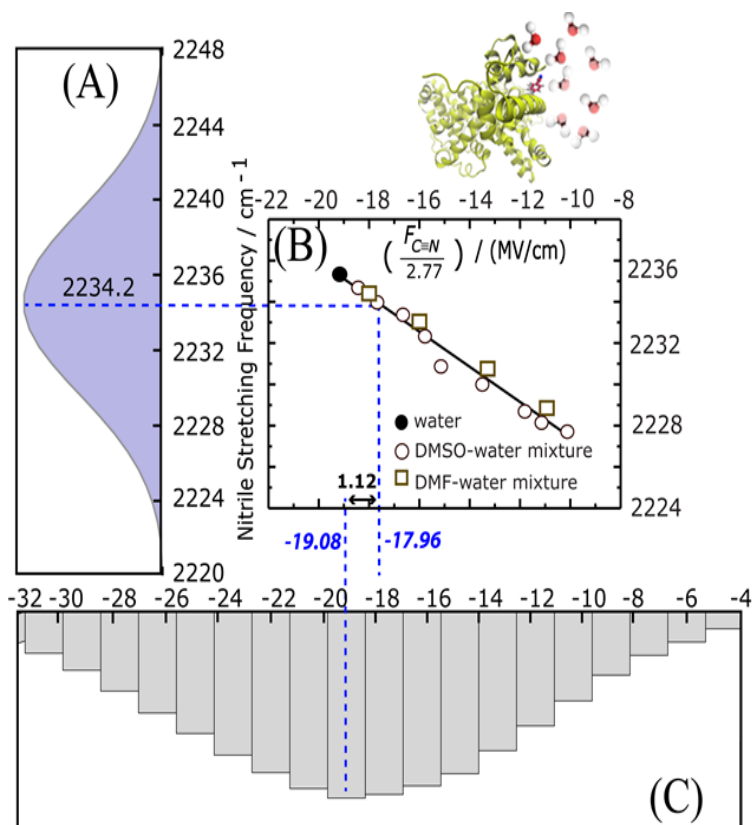
Figure 3.6*: Field-frequency correlation acts a calibration curve to predict the electric field experienced by the nitrile modified bovine serum albumin (compound 7). (A) The nitrile stretch region of the FTIR spectrum of compound 7, fitted to a Gaussian profile. (B) The field-frequency correlation plot from figure 3.4(A) to predict $F_{C\equiv N}$ of the nitrile modified protein from the C≡N peak frequency of compound 7. (C) Histograms of the estimated electric fields from MD simulations (divided by the 2.77, the scaling factor for PhCN) of compound 7. The predicted field value from the calibration curve, -17.96 MV/cm, differs from the estimated field value, -19.08 MV/cm, by 1.12 MV/cm.*

Section 3.6) where PhCN has been attached with the Cys-34 unit of the Bovine Serum Albumin (BSA) through a linker (compound **6** of the section 3.6). The C≡N stretching frequencies of the linker in different solvents (THF and DMSO) are found to be within 0.1 cm^{-1} for the linker and PhCN. This result suggests that PhCN can be used as a model nitrile for the linker and the linker

conjugated protein. Synthesis of **6** and **7** have been discussed in the section 3.6. The bio-conjugation of the aromatic nitrile group with Cys-34 (only free cysteine in BSA) is achieved under non-reductive conditions to keep the disulfide bonds between the other Cys residues unperturbed. From tryptic digestion of **7** it has been confirmed that modification occurred only at cys-34 residue of BSA. The nitrile stretching frequency of **7** is recorded in aqueous buffer (Figure 3.5) and the electric field experienced by the nitrile group is predicted to be -17.96 MV/cm from the field-frequency correlation plot which is in good agreement with the electric field value of -19.08 MV/cm, determined independently using MD simulations (Figure 3.6 and Table 3.4). This observation firmly suggests that our model is appropriate for predicting electric fields experienced by H-bonded nitrile groups in biological macromolecules. The protein is structurally ordered as compared with the earlier mentioned disordered peptides. The ability to estimate specific non-covalent interactions in both disordered and ordered biomolecules illustrates the generality of the field-frequency correlation for H-bonded nitriles.

3.2.5 Prediction of electric fields with a priori knowledge about H-bonding

Combining Figure 3.4 with the field-frequency correlation for nitrile in non-H-bonded environments (Figure 3.3) provides an interesting V-shaped plot (Figure 3.7) in which line L_1 represents the behavior of nitriles in non-H-bonded environments and line L_2 represents the same in H-bonded environments. As can be seen from Figure 3.7, the ensemble average value of $\overline{\nu}_{C\equiv N}$ obtained from IR experiments can correspond to two values of the electric field, depending on the specificity of the non-covalent interaction. To predict the site-specific electric field experienced by a nitrile modified biomolecule directly from IR absorption experiments, the knowledge of the H-bonding status of the nitrile is required, which can be confirmed either from IR-NMR correlations or from the structural information of the nitrile containing compounds.[24,25] We have already validated a direct correlation between nitrile stretching frequency and the electric field experienced by the nitrile group in H-bonding environments (Figure 3.4). We have also checked the validity of the L1 line (Figure 3.7 A) using a nitrile mutated protein, [p-CN-Phe]RNase S where nitrile group is buried in a non-hydrogen bonding environment. From the nitrile IR absorption of the protein we have predicted the field and compare to the reported field.[25] Our prediction closely matches with the reported field (Table 3.4). Thus we have validated both L1 and L2. From our observations it is clear that once the H-bonding status is

39

known, the corresponding best-fit line (either L_1 or L_2) can be utilized to quantify the non-covalent interactions. This way, electric field approach can be generalized to quantify any non-covalent interaction of nitrile containing biomolecules or ligands, specific or non-specific.
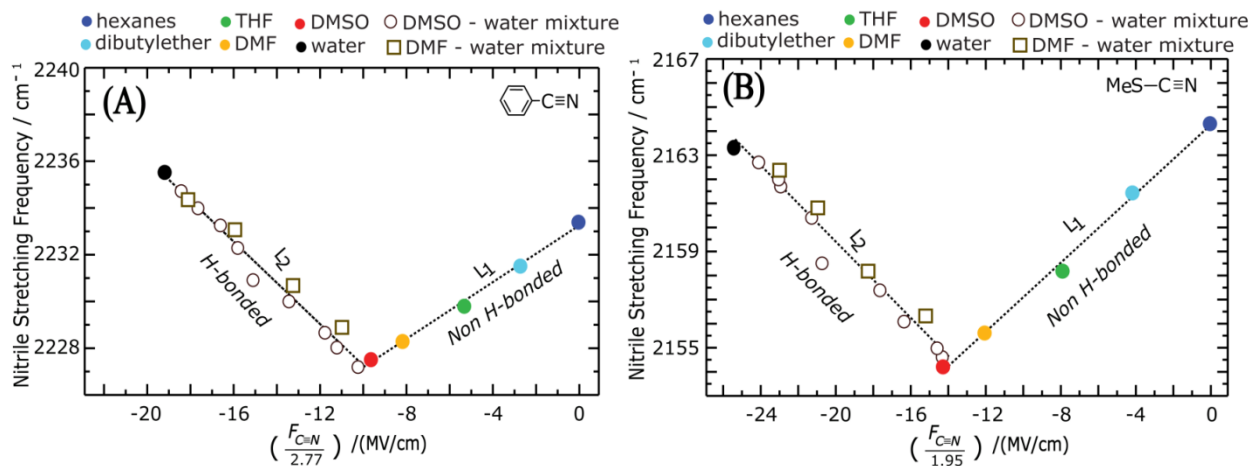

Figure 3.7*: "V-shaped" combined field-frequency correlation can act as a calibration curve for (A) PhCN and (B) MeSCN in H-bonding and non-H-bonding environments. Depending upon the specificity of the non-covalent interactions, L_1 or L_2 can be chosen as the calibration curve to predict electric field from experimental IR frequency.*

3.3 Concluding Remarks

We have shown and validated a linear field-frequency correlation for specific chemical interactions in nitriles in a range of molecules of different size and compactness. Such a correlation has never been reported to date and our current finding will be extremely useful to quantitatively estimate electric fields experienced by the H-bonded nitriles. Further, this result will allow us to describe the H-bonded and non-H-bonded environments of a nitrile vibrational probe in a consistent manner through average electric fields arising from the environment, which has been suggested as one of the important descriptors of non-covalent interactions. The field-frequency correlation obtained for both the aromatic and aliphatic nitriles illustrates that nitrile vibrational probes, like carbonyls, can also be benchmarked against reference data, providing a consistent way out to interpret nitrile frequencies in complex molecular systems, in both H-bonding and non-H-bonding environments. This finding will also allow the prediction of experimental nitrile IR frequencies without performing computationally expensive quantum mechanical calculations. Our results will allow researchers to obtain a quantitative understanding

of the non-covalent interactions and their roles towards molecular recognition and molecular self-assembly in either nitrile containing drugs or nitrile modified proteins in any chemical environment.

3.4 FTIR spectroscopy

IR absorption spectra were recorded on a FTIR-8300 (Shimadzu) spectrometer with 2 cm^{-1} resolution at room temperature. For each sample, ~60 microliters of the sample solution was loaded into a demountable cell consisting of two windows (CaF$_2$, 3 mm thickness, Shenzen Laser), separated by a mylar spacer of 56 micrometers thickness. Both the nitriles (PhCN & MeSCN) were dissolved independently in, aprotic solvents, water, water/DMSO, water/DMF, and water/THF binary mixtures, and aqueous urea solutions (5M) so that the final concentration of liquid sample is 10 mM for the IR studies. For recording the IR spectra of nitriles in aprotic solvents, 16 scans were collected whereas 100 scans were recorded for nitriles in water, water/DMSO binary mixtures and aqueous urea solutions. For nitrile mutated amino acid (compound **3**) in water and nitrile mutated BSA (compound **7**) in aqueous buffer 100 scans were collected. A polynomial fit was used to baseline each FTIR absorbance spectrum. The processed absorption spectra were fit to a Gaussian to obtain the reported peak frequencies.

3.5 Computational methods

Molecular dynamics (MD) simulations were carried out using the GROMACS[51] (version: 4.6.5) software. General AMBER force field (GAFF)[52] was used to model benzonitrile, methyl thiocyanate and all other non-aqueous solvent molecules. The force field parameters for the solvent molecules were obtained from virtualchemistry.org,[53] a database of parameters and physical properties of a wide range of organic liquids, which have been thoroughly tested and benchmarked for reliability.[54] CHARMM force field parameters[55] were used to model the THF and THF/water binary mixtures, since this combination has been shown to reproduce the temperature dependent miscibility in THF/water binary mixtures.[56] For the simulations involving binary mixtures (water with DMSO /DMF /THF), the non-bonded interaction parameters between water and the other solvent molecules were generated using the standard Lorentz-Berthelot mixing rule given by $\sigma_{ij} = \frac{1}{2}\left(\sigma_i + \sigma_j\right)$ and $\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$, where σ_i and ε_i are the

Lennard-Jones parameters for the individual solvent molecules. Amber ff99SB-ILDN force field was used to model the standard amino acid residues in peptides and proteins[57]. The modified residue p-CN-Phe in the peptide, Mastoparan X, was created by combining the parameters of benzonitrile (GAFF) and PHE residue (Amber ff99SB-ILDN). Similarly the cysteine modified residues in the peptide CM15 and protein BSA were created by combining the parameters of the Cystine residue (Amber ff99SB-ILDN) with those of methyl thiocyanate (GAFF) and modified amino acid (GAFF), respectively. The structure of the BSA protein was obtained from the PDB data bank (ID: 4F5S). The simulations of the Mastoparan X peptide were started from multiple initial conformations including a fully extended conformation in order to rigorously check the equivalence of the calculated average electric field from independent runs. TIP3P water model[58] was used in the simulations involving water molecules.

A single probe molecule (benzonitrile or methyl thiocyanate) was solvated in a cubic box of 5nm dimension of respective solvent molecules. The solvated system was sequentially processed through (i) energy minimization using steepest descent algorithm, (ii) equilibration in NVT ensemble at 300K temperature for 100ps using the velocity rescale thermostat,[59] and (iii) equilibration in NPT ensemble at 300K temperature and 1 bar pressure (using Parrinello-Rahman barostat[60]) for 1ns. For the systems involving benzonitrile and methyl thiocyanate in various solvents, the production run was continued for 10 ns, with positions and forces on each atom being saved at every 0.2 ps. For the Mastoparan X and CM15 peptides much longer production runs of 50ns were used, since the electrostatic field needed to be averaged over various conformations of the peptide. For the protein (BSA) production runs of 50 ns were used starting with different conformations and the average electric field has been calculated from multiple trajectories.

3.6 Details of synthesis

A) Synthesis of 3

Scheme1.

1

2

3

.
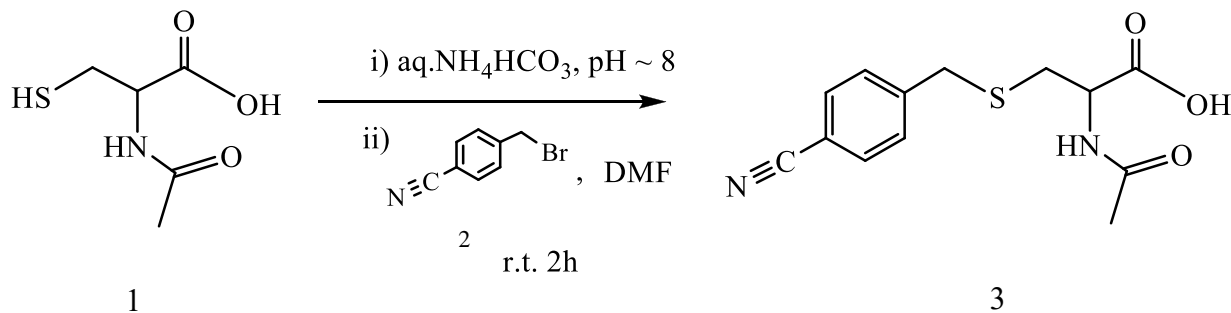
N-acetyl cysteine (1, 250mg, 99 mmol) was dissolved in minimum volume of degassed aq. NH_4HCO_3 solution (pH ≈ 8) in a small round bottom flask (rb). In a separate rb, 4-cyanobenzyl bromide (2, 274 mg, 118.8 mmol) is dissolved in DMF (5.5 mL) and degassed with nitrogen. The solution of 2 in DMF was added dropwise into the aq. NH_4HCO_3 solution of 1. The reaction mixture was stirred for 2 h at room temperature under nitrogen atmosphere. Product was extracted in ethyl acetate, washed with brine to remove DMF, dried over Na_2SO_4, filtered and concentrated in vacuum. The residual solid was purified by column chromatography using ethyl acetate/hexane (60:20) mixture to yield 3 (white solid).

Compound 3: ^{1}H NMR (400 MHZ, DMSO - d_6) δ 8.17 (br, 1H), 7.49-7.77 (m, 4H), 4.37 (m, 1H), 3.81 (s, 2H), 2.60-2.77 (m, 2H), 1.84 (s, 3H); ^{13}C NMR (500 MHZ, DMSO-d_6) δ 169.67, 144.75, 132.50, 130.09, 119.04, 109.75, 52.03, 35.06, 32.84, 22.57. HRMS (ESI) calculated for $C_{13}H_{15}N_2O_3S$ [M+H[+]] is 279.3348, Found 279.0793.

B) Synthesis of 6

Step1 :
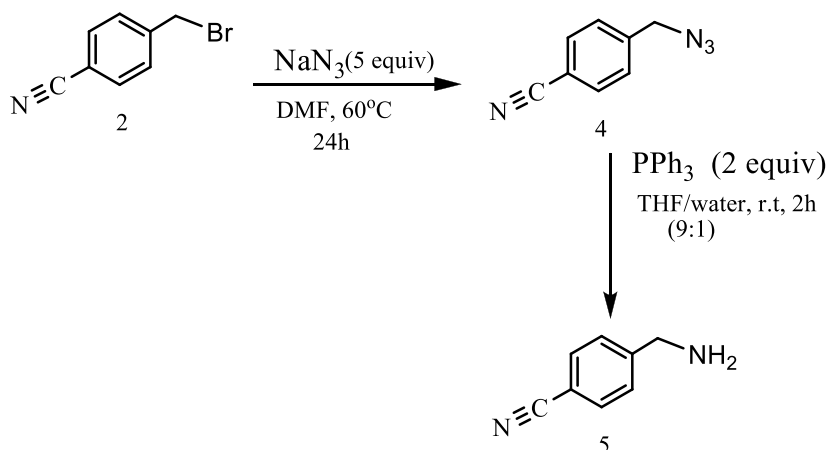
Step2 :
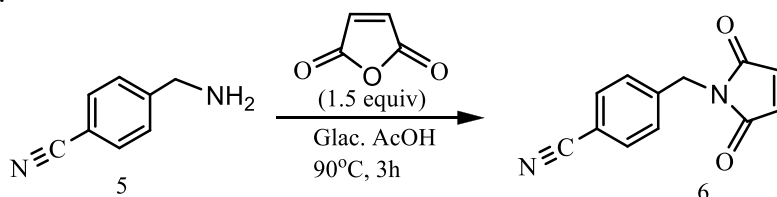
Preparation of 5 (Step1): To a stirring solution of 2 (500mg, 2.55 mmol) in DMF (10 mL), 5

equivalent of NaN$_3$ (explosive) added. A condenser was fitted with the rb with continuous supply of cold water. After stirring at 60° C for 24 h, the reaction mixture was cooled to room temperature and the product was extracted with diethyl ether and washed with brine. Formation of the azide 4 was confirmed by TLC using PPh$_3$ (as charring agent) in hexane. Since benzyl azides are explosive, 4 was subjected to immediate reduction without further analysis.

4 was dissolved in 6 mL solvent mixture of THF/water (9:1). PPh$_3$ (2 equiv.) added into it and the reaction mixture was stirred for 4 h. Reaction mixture was dried over Na$_2$SO$_4$. Crude product was purified by silica-gel column chromatography using EtOAc/hexane (95:5) to yield pure 5 (yellowish liquid)

Compound 5: ^{1}H NMR (200 MHZ, DMSO-d$_6$) δ 6.46-6.72 (m, 4H), 2.74 (s, 1H); ^{13}C NMR (200 MHZ, DMSO-d$_6$) δ 149.47, 132.25, 128.29, 119.31, 109.19, 45.12.; HRMS (ESI) HRMS (ESI) calculated for C$_8$H$_9$N$_2$ [M+H$^+$] is 133.1705, Found 133.0760.

Preparation of 6 (Step2): Maleic anhydride (300mg, 3.06 mmol) was added to the rb containing amine 5 (269 mg, 2.04 mmol). A water condenser was fitted with the rb. Glacial acetic acid (3.5 ml) added into the rb and the reaction mixture was refluxed at 90 ^{o}C for 3 h at stirring condition. Product was extracted with ethyl acetate followed by washing with saturated aqueous solution of NaHCO$_3$ to remove excess acetic acid. The ethyl acetate layer was dried over Na$_2$SO$_4$ and concentrated using a rotary evaporator. Finally column chromatography was performed using EtOAc/hexane (40:60) to yield 6 (yellow solid) which was characterized using NMR and mass spectrometry.

Compound 6: ^{1}H NMR (200 MHZ, CDCl$_3$) δ 7.42-7.64 (m, 4H), 6.76 (S, 2H), 4.72 (s, 2H); ^{13}C NMR (200 MHZ, CDCl3) δ 170.04, 141.10, 134.34, 132.54, 128.98, 118.43, 111.90, 40.93; HRMS (ESI) calculated for C$_{12}$H$_9$N$_2$O$_2$ [M+H$^+$] 213.2121, Found 213.0654.

C) Synthesis of 7

Scheme3:
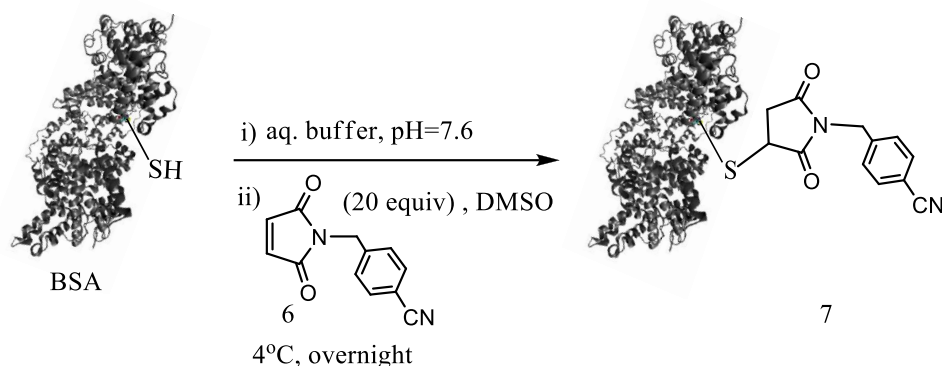
Preparation of 7 : A 50 mM stock solution of 6 was prepared by dissolving 6 in DMSO followed by degassing with nitrogen. 6 µL of the DMSO solution was added dropwise to a degassed aqueous buffer (40 mM sodium phosphate and 2mM EDTA, pH=7.6) containing 150 µM BSA (Bovine Serum Albumin). The final reaction volume was 100 µL and the concentration of reagent 6 was 20 times higher than the protein concentration. The mixture was incubated for 12 h at 4°C. Completion of the reaction was confirmed from Elmann's assay.[61] The conjugate 7 was purified using gell filtration chromatography. HRMS (ESI) analysis of the tripsin digested peptides of the conjugate revealed that only Cys-34 of BSA was modified. We have followed the tripsin digestion procedure as reported by Sen Gupta and co-workers.[62] The masses of one of the peptide fragments containing Cys-34 derived from the unmodified and modified BSA were 974.4420 and 1186.5020 respectively. The difference in the mass, 212.06, corresponds to the exact mass of the modified unit (6).

3.7 References

(1)   Fried, S. D.; Boxer, S. G. Measuring Electric Fields and Noncovalent Interactions Using the Vibrational Stark Effect. *Acc. Chem. Res.* **2015**, *48*, 998-1006.

(2)   Lehn, J. M. Supramolecular Chemistry - Scope and Perspectives Molecules, Supermolecules, and Molecular Devices. *Angew. Chem., Int. Ed.* **1988**, *27*, 89-112.

(3)   Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. Electrostatic Basis for Enzyme Catalysis. *Chem. Rev.* **2006**, *106*, 3210-3235.

(4)   Warshel, A.; Aqvist, J. Electrostatic Energy and Macromolecular Function. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 267-298.

(5)   Park, E. S.; Andrews, S. S.; Hu, R. B.; Boxer, S. G. Vibrational Stark Spectroscopy in Proteins: A Probe and Calibration for Electrostatic Fields. *J. Phys. Chem. B* **1999**, *103*, 9813-9817.

(6)   Suydam, I. T.; Snow, C. D.; Pande, V. S.; Boxer, S. G. Electric Fields at the Active Site of an Enzyme: Direct Comparison of Experiment With Theory. *Science* **2006**, *313*, 200-204.

(7)   Thielges, M. C.; Case, D. A.; Romesberg, F. E. Carbon-deuterium Bonds as Probes of Dihydrofolate Reductase. *J. Am. Chem. Soc.* **2008**, *130*, 6597-6603.

(8)   Ye, S.; Huber, T.; Vogel, R.; Sakmar, T. P. FTIR Analysis of GPCR Activation Using Azido Probes. *Nat.Chem.Biol.* **2009**, *5*, 397-399.

(9)   Li, T. S.; Quillin, M. L.; Phillips, G. N.; Olson, J. S. Structural Determinants of The Stretching Frequency of CO Bound to Myoglobin. *Biochemistry* **1994**, *33*, 1433-1446.

(10)   Taskent-Sezgin, H.; Chung, J.; Banerjee, P. S.; Nagarajan, S.; Dyer, R. B.; Carrico, I.; Raleigh, D. P. Azidohomoalanine: A Conformationally Sensitive IR Probe of Protein Folding, Protein Structure, and Electrostatics. *Angew. Chem. Int. Ed.* **2010**, *49*, 7473-7475.

(11)   Hu, W.; Webb, L. J. Direct Measurement of the Membrane Dipole Field in Bicelles Using Vibrational Stark Effect Spectroscopy. *J. Phys. Chem. Lett.* **2011**, *2*, 1925-1930.

(12)   Shrestha, R.; Cardenas, A. E.; Elber, R.; Webb, L. J. Measurement of the Membrane Dipole Electric Field in DMPC Vesicles Using Vibrational Shifts of p-Cyanophenylalanine and Molecular Dynamics Simulations. *J. Phys. Chem. B* **2015**, *119*, 2869-2876.

(13)   Fafarman, A. T.; Boxer, S. G. Nitrile Bonds as Infrared Probes of Electrostatics in Ribonuclease. S *J. Phys. Chem. B* **2010**, *114*, 13536-13544.

(14)   Pollack, R. M.; Thornburg, L. D.; Wu, Z. R.; Summers, M. F. Mechanistic Insights from the Three-Dimensional Structure of 3-Oxo-Δ5-steroid Isomerase. *Arch. Biochem. Biophys.* **1999**, *370*, 9-15.

(15)   Ma, J.; Pazos, I. M.; Zhang, W.; Culik, R. M.; Gai, F. Site-Specific Infrared Probes of Proteins. *Annu. Rev. Phys. Chem.* **2015**, *66*, 357-377.

(16)   Suydam, I. T.; Boxer, S. G. Vibrational Stark Effects Calibrate the Sensitivity of Vibrational Probes for Electric Fields in Proteins. *Biochemistry* **2003**, *42*, 12050-12055.

(17)   Park, E. S.; Thomas, M. R.; Boxer, S. G. Vibrational Stark spectroscopy of NO Bound to Heme: Effects of Protein Electrostatic Fields on the NO Stretch Frequency. *J. Am. Chem. Soc.* **2000**, *122*, 12297-12303.

(18)   Fafarman, A. T.; Sigala, P. A.; Schwans, J. P.; Fenn, T. D.; Herschlag, D.; Boxer, S. G. Quantitative, Directional Measurement of Electric Field Heterogeneity in the Active Site of Ketosteroid Isomerase. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, E299-E308.

(19)   Webb, L. J.; Boxer, S. G. Electrostatic Fields Near the Active Site of Human Aldose Reductase: 1. New Inhibitors and Vibrational Stark Effect Measurements. *Biochemistry* **2008**, *47*, 1588-1598.

(20)   Perutz, M. F.; Mathews, F. S. An X-Ray Study of Azide Methaemoglobin. *J. Mol. Biol.* **1966**, *21*, 199-202.

(21)   Lindquist, B. A.; Furse, K. E.; Corcelli, S. A. Nitrile Groups as Vibrational Probes of Biomolecular Structure and Dynamics: an Overview. *Phys. Chem. Chem. Phys.* **2009**, *11*, 8119-8132.

(22)   Warshel, A.; Bora, R. P. Perspective: Defining and Quantifying the Role of Dynamics in Enzyme Catalysis. *J. Chem. Phys.* **2016**, *144*, 180901-180901.

(23)   Getahun, Z.; Huang, C.-Y.; Wang, T.; De León, B.; DeGrado, W. F.; Gai, F. Using Nitrile-Derivatized Amino Acids as Infrared Probes of Local Environment. *J. Am. Chem. Soc.* **2002**, *125*, 405-411.

(24)   Wolynes, P. G. Recent Successes of the Energy Landscape Theory of Protein Folding and Function. *Q. Rev. Biophys* **2005**, *38*, 405-410.

(25)   Warshel, A. Energetics of enzyme catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **1978**, *75*, 5250-5254.

(26)   McElheny, D.; Schnell, J. R.; Lansing, J. C.; Dyson, H. J.; Wright, P. E. Defining the Role of Active-site Loop Fluctuations in Dihydrofolate Reductase Catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 5032.

(27)   Bagchi, S.; Boxer, S. G.; Fayer, M. D. Ribonuclease S Dynamics Measured Using a Nitrile Label with 2D IR Vibrational Echo Spectroscopy. *J. Phys. Chem. B* **2012**, *116*, 4034-4042.

(28)   Urbanek, D. C.; Vorobyev, D. Y.; Serrano, A. L.; Gai, F.; Hochstrasser, R. M. The Two-Dimensional Vibrational Echo of a Nitrile Probe of the Villin HP35 Protein. *J. Phys. Chem. Lett* **2010**, *1*, 3311-3315.

(29)   Levinson, N. M.; Boxer, S. G. A Conserved Water-Mediated Hydrogen Bond Network Defines Bosutinib's Kinase Selectivity. *Nat.Chem.Biol.* **2014**, *10*, 127-132.

(30)   Liu, C. T.; Layfield, J. P.; Stewart, R. J.; French, J. B.; Hanoian, P.; Asbury, J. B.; Hammes-Schiffer, S.; Benkovic, S. J. Probing the Electrostatics of Active Site Microenvironments Along the Catalytic Cycle for Escherichia coli Dihydrofolate Reductase. *J. Am. Chem. Soc.* **2014**, *136*, 10349-10360.

(31)   Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, Electrostatic and Steric stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227-249.

(32)   Maienschein-Cline, M. G.; Londergan, C. H. The CN Stretching Band of Aliphatic Thiocyanate is Sensitive to Solvent Dynamics and Specific Solvation. *J. Phys. Chem. A* **2007**, *111*, 10020-10025.

(33)   G Careri; P Fasella, a.; Gratton, E. Enzyme Dynamics: The Statistical Physics Approach. *Ann. Rev. Biophys. Bioeng.* **1979**, *8*, 69-97.

(34)   Boxer, S. G. Stark Realities. *J. Phys. Chem. B* **2009**, *113*, 2972-2983.

(35)   Lindquist, B. A.; Corcelli, S. A. Nitrile Groups as Vibrational Probes: Calculations of the C≡N Infrared Absorption Line Shape of Acetonitrile in Water and Tetrahydrofuran. *J. Phys. Chem. B* **2008**, *112*, 6301-6303.

(36)   Purcell, K. F.; Drago, R. S. Studies of the Bonding in Acetonitrile Adducts1. *J. Am. Chem. Soc.* **1966**, *88*, 919-924.

(37)   Layfield, J. P.; Hammes-Schiffer, S. Calculation of Vibrational Shifts of Nitrile Probes in the Active Site of Ketosteroid Isomerase upon Ligand Binding. *J. Am. Chem. Soc.* **2012**, *135*, 717-725.

(38)   Hammes-Schiffer, S.; Benkovic, S. J. Relating Protein Motion to Catalysis. *Annu. Rev. Biochem* **2006**, *75*, 519-541.

(39)   Pazos, I. M.; Gai, F. Solute's Perspective on How Trimethylamine Oxide, Urea, and Guanidine Hydrochloride Affect Water's Hydrogen Bonding Ability. *J. Phys. Chem. B* **2012**, *116*, 12473-12478.

(40)   Tucker, M. J.; Getahun, Z.; Nanda, V.; DeGrado, W. F.; Gai, F. A New Method for Determining the Local Environment and Orientation of Individual Side Chains of Membrane-Binding Peptides. *J. Am. Chem. Soc.* **2004**, *126*, 5078-5079.

(41)   Warshel, A.; Sharma, P. K.; Chu, Z. T.; Åqvist, J. Electrostatic Contributions to Binding of Transition State Analogues Can Be Very Different from the Corresponding Contributions to Catalysis: Phenolates Binding to the Oxyanion Hole of Ketosteroid. Isomerase *Biochemistry* **2007**, *46*, 1466-1476.

(42)   Pazos, I. M.; Ghosh, A.; Tucker, M. J.; Gai, F. Ester Carbonyl Vibration as a Sensitive Probe of Protein Local Electric Field. *Angew. Chem., Int. Ed.* **2014**, *53*, 6080-6084.

(43)   Kashid, S. M.; Bagchi, S. Experimental Determination of the Electrostatic Nature of Carbonyl Hydrogen-Bonding Interactions Using IR-NMR Correlations. *J. Phys. Chem. Lett* **2014**, *5*, 3211-3215.

(44)   McElheny, D.; Schnell, J. R.; Lansing, J. C.; Dyson, H. J.; Wright, P. E. Defining the Role of Active-site Loop Fluctuations in Dihydrofolate Reductase Catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 5032-5037.

(45)   Kurplus, M.; McCammon, J. A. Dynamics of Proteins: Elements and Function. *Annu. Rev. Biochem* **1983**, *52*, 263-300.

(46)   McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977**, *267*, 585-590.

(47)   Malaspina, T.; Fileti, E. E.; Riveros, J. M.; Canuto, S. Ab initio study of the isomeric equilibrium of the HCN center dot center dot center dot H2O and H2O center dot center dot center dot HCN hydrogen-bonded clusters. *J. Phys. Chem. A* **2006**, *110*, 10303-10308.

(48)   Reimers, J. R.; Hall, L. E. The Solvation of Acetonitrile. *J. Am. Chem. Soc.* **1999**, *121*, 3730-3744.

(49)   Reimers, J. R.; Zeng, J.; Hush, N. S. Vibrational stark spectroscopy .2. Application to the CN stretch in HCN and acetonitrile. *J. Phys. Chem.* **1996**, *100*, 1498-1504.

(50)   Jo, H.; Culik, R. M.; Korendovych, I. V.; DeGrado, W. F.; Gai, F. Selective Incorporation of Nitrile-Based Infrared Probes into Proteins via Cysteine Alkylation. *Biochemistry* **2010**, *49*, 10354-10356.

(51)   Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435-447.

(52)   Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

(53)   van der Spoel, D.; van Maaren, P. J.; Caleman, C. GROMACS Molecule & Liquid Database. *Bioinformatics* **2012**, *28*, 752-753.

(54)   Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.* **2012**, *8*, 61-74.

(55)   Vorobyov, I.; Anisimov, V. M.; Greene, S.; Venable, R. M.; Moser, A.; Pastor, R. W.; MacKerell, A. D., Jr. Additive and Classical Drude Polarizable Force Fields for Linear and cyclic ethers. *J. Chem. Theory Comput.* **2007**, *3*, 1120-1133.

(56)   Smith, M. D.; Mostofian, B.; Petridis, L.; Cheng, X.; Smith, J. C. Molecular Driving Forces Behind the Tetrahydrofuran-Water Miscibility. Gap *J. Phys. Chem. B* **2016**, *120*, 740-747.

(57)   Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950-1958.

(58)   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926-935.

(59)   Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, -.

(60)   Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182-7190.

(61)   Ellman, G. L. Tissue sulfhydryl groups. *Arch. Biochem. Biophys.* **1959**, *82*, 70-77.

(62)   Kumari, S.; Panda, C.; Mazumdar, S.; Sen Gupta, S. A Molecular Fe-complex as a Catalyst Probe for In-gel Visual Detection of Proteins Via Signal Amplification. *Chem. Commun.* **2015**.

Chapter 4

Solvent Dynamics in the Interplay Between n→ π* Interaction and Hydrogen Bonding

4.1 Introduction

Non-covalent interactions regulate molecular conformation, function, and reactivity in chemical and biological systems.[1,2] The n→π* interaction, an electronic delocalization effect analogous to hydrogen bonding (H bonding), has emerged as a key non-covalent interaction to stabilize the structures of nucleic acids, proteins, neurotransmitters, amino acids and small drug molecules.[3-18] The electrons in the lone-pair (n) orbital of the donor are delocalized into the σ* orbital of the acceptor in H bonds, whereas n→π* interaction involves electron delocalization into the π* orbital of a carbonyl (C=O) or an aryl acceptor.[15] The counterintuitive nature, considering the expected electron repulsion between an electronegative atom and the electron π-cloud of the arene, has led to extensive debates regarding the existence of n→π* interactions.[19] Computations estimated the n→π* interactions to be energetically weaker (0.2−1 kcal/mol) by at least one order of magnitude than the H bonds (5−10 kcal/mol)[3]. Thus to decipher the weak n→π* interactions in solution, within the milieu of H bonds and Coulombic interactions in complex biomolecules, is a formidable challenge. The existence of n→π* interactions, either between two backbone carbonyls or between a backbone C=O and a side-chain aryl group, has mostly been demonstrated from the analysis of X-ray crystal structures in the Protein Data Bank (PDB) and Cambridge Structural Database (CSD).[5,9,14,16] On an average, 34% of the protein residues are predicted to engage in n→π* interactions, providing a substantial energetic contribution toward protein stability even though the individual energetic contribution is very small.[5] Installing isosteric substituents in small molecules, Raines and coworkers reported n→π* interactions of esters, amides, and thioamides using X-ray diffraction analysis, NMR spectroscopy, and *ab initio* calculations.[3,4,9] Implications of n→π* interactions on the structure and reactivity of prebiotic molecules, on the functional role of potassium ion (K⁺) channels, and on the prevention of structural distortion of green fluorescent protein (GFP) have also been predicted using crystal structure analyses and computations.[5,6,13] Recently, Das and coworkers founded direct spectroscopic evidence of n→π* interaction in the gas phase for phenyl formate by monitoring the C=O stretching frequency using isolated gas-phase infrared (IR) spectroscopy.[12] In phenyl formate, the lone-pair electrons on the C=O are reported to be delocalized into the π* orbital of the phenyl ring (figure 4.1).
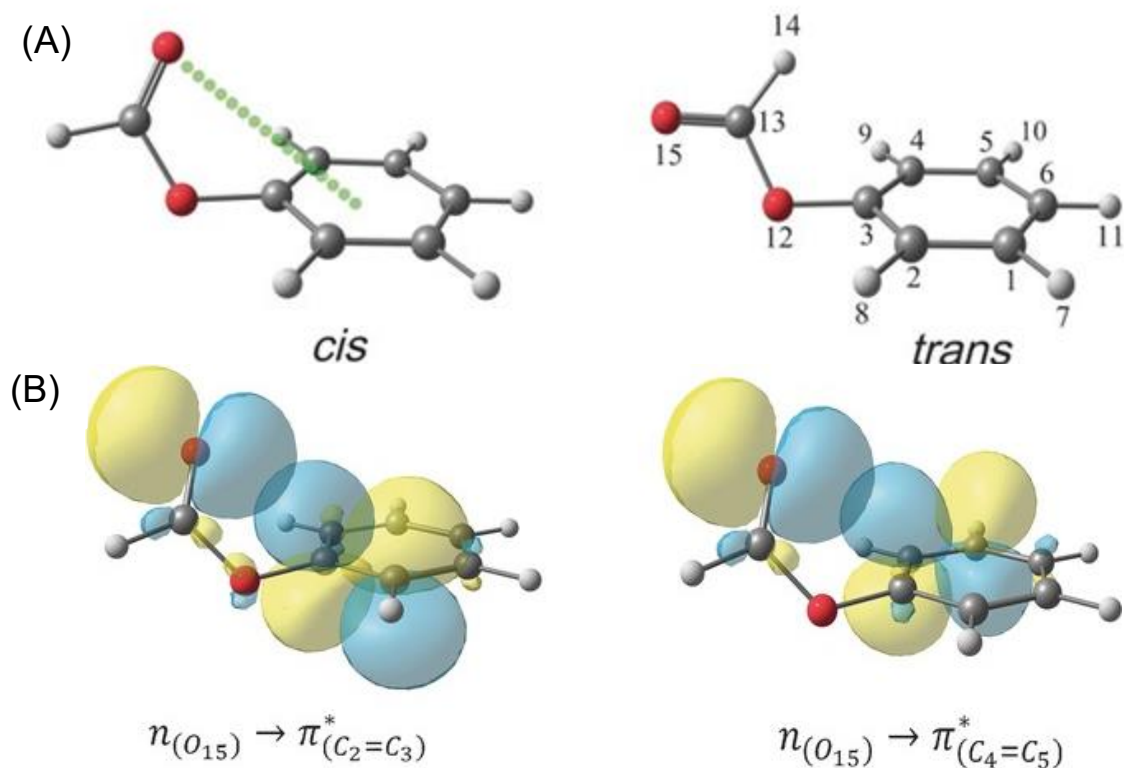
$$n_{(O_{15})} \rightarrow \pi^*_{(C_2=C_3)} \qquad\qquad n_{(O_{15})} \rightarrow \pi^*_{(C_4=C_5)}$$

Figure 4.1: *(A) Phenyl formate exists in cis and trans conformation. (B) Cis conformations shows n→π* interaction between n-orbitals on the carbonyl oxygen and the π* orbitals of the aryl ring. The figure is taken form ref 12 of this chapter.[12]*

Knowledge of these weak but abundant n→π* interactions is the key to our understanding of the biomolecular systems. A detailed description of these unconventional interactions can provide accurate force field parameters for computational investigation of proteins.[9] Since both H bonding and n→π* interactions involve C=O lone-pair electrons, any cooperativity or interplay between these two interactions is likely to have an immense impact on the structural and dynamical aspects of chemical biology. It was speculated from statistical analyses of protein crystal structures that compromised H-bond geometry in constrained protein environments might be interrelated to n→π* interactions.[9]

In this chapter, using linear infrared (IR) and two-dimensional infrared (2D IR) spectroscopy[20-23] in solution, we show a direct evidence of the coexistence of n→π* interaction and H bonds involving C=O group at room temperature. Our results indicate that H bonds, though energetically favorable, are in dynamic equilibrium with the n→π* interactions in an aqueous environment and interconvert in picosecond timescale.

4.2 Results and discussions

4.2.1 Infrared signature of C=O n→π* interaction in solution

The model compound chosen for this work is phenyl formate which has already been reported to show IR spectroscopic evidence of n→π* interaction in the gas phase.[12] In a non-polar aprotic solvent like *n*-hexanes (hexane), where the nature of interaction should not greatly differ from that in the gas phase, two C=O transitions are observed at 1781 and 1755 cm^{-1} (Figure 4.2A). Both the peaks are red-shifted as compared to those reported in the gas-phase IR experiments (1797 and
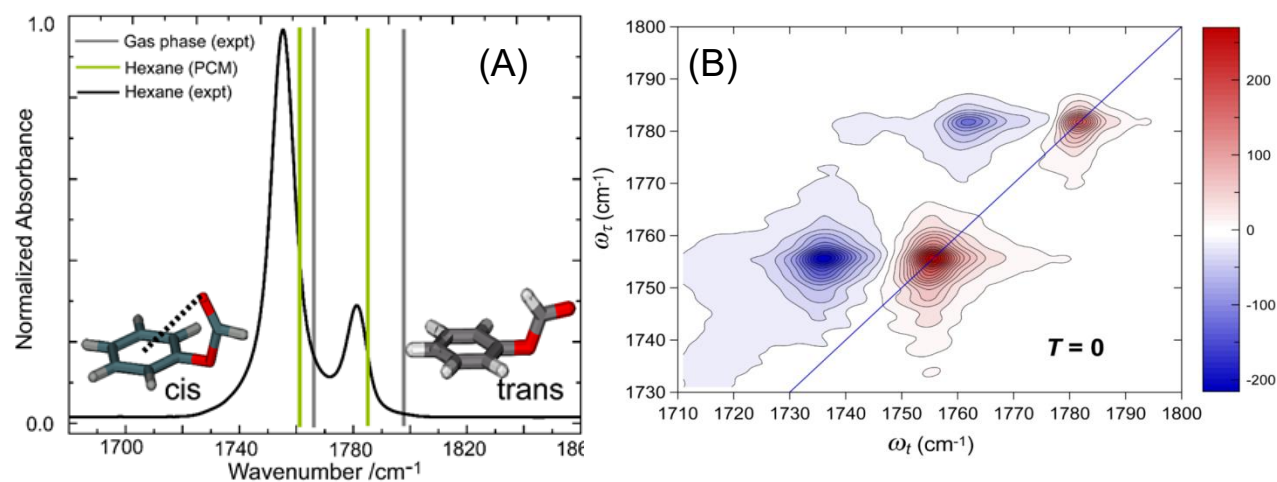

Figure 4.2*: Assignment of IR spectral signature of n→π* interaction in solution. (A) Linear IR spectrum of phenyl formate in hexane shows two transitions at 1755 and 1781 cm^{-1}. The gray lines represent corresponding peak positions in gas-phase IR experiments[12]. The green lines represent the peak positions estimated from DFT calculations in hexane . (B) 2D IR spectrum of phenyl formate in hexane at T = 0. The diagonal peak pairs (red and blue) correspond to 0→1 and 1→2 transitions, respectively. Absence of cross peaks in the 2DIR spectrum implies no Fermi resonance and confirms that the two transitions are from cis and trans conformations of phenyl formate.*

1766 cm^{-1}) or obtained from DFT calculations (1800 and 1770 cm^{-1});[12] the lowering of the C=O stretching frequencies is a common phenomenon going from the gas phase to the solution phase.[24] The separation of 26 cm^{-1} between the two C=O peaks in hexane is similar to that observed in the gas phase (30 cm^{-1}). Previous gas-phase study has attributed the peaks to the trans and cis conformations of phenyl formate respectively (Figure 4.2A), where the C=O of the cis conformer participates in n→π* interaction with the phenyl ring and is more stable than the

trans conformer by 1.32 kcal/mol.[12] When the effect of solvent dielectric for hexane ($\varepsilon = 1.9$) is considered using a polarizable continuum model (PCM), the energy difference (ΔE_{rel}) between the two conformers decreases to 0.99 kcal/mol. The peak positions and the relative areas of the two peaks ($K_{cis/trans} = 5.0$) in hexane are in excellent agreement with the PCM calculations (Table 4.1). Although the spectral signatures in hexane are similar to the gas-phase results and the population ratio of the peaks conforms to the theoretical predictions, further independent experiments have been performed to confirm the origin of the observed IR transitions in solution. Another plausible reason for the existence of the two IR peaks in solution can be Fermi resonance between the C=O and the overtone/combination band of a low-frequency mode, as

Table 4.1: *Tunability of n→π* interaction by local polarity around the C=O group.*

Solvent	ε^a	ΔE_{rel} (kcal/mol)	$K_{cis/trans}$ (theo)	$K_{cis/trans}$ (expt)	$\nu_{C=O,\,cis}{}^c$ (theo)	$\nu_{C=O,\,cis}{}^c$ (expt)	$\nu_{C=O,\,trans}{}^c$ (theo)	$\nu_{C=O,\,trans}{}^c$ (expt)
Gas Phase	0	1.32	9.4	-	1765	1766[b]	1795	1797[b]
Hexane	1.9	0.99	5.4	5.0	1761	1755	1785	1781
THF	7.4	0.70	3.3	2.1	1751	1744	1774	1766

[a]Dielectric constant of the solvents used in the calculation.
[b]Gas phase values are taken from reference 12.
[c] All values are in cm^{-1}.

previously reported in esters.[25] Although the theoretical descriptions of Fermi resonance are well established,[26] linear IR alone does not characterize these couplings unequivocally. For phenyl formate in hexane, 2D IR spectroscopy (discussed later in details), sensitive to anharmonic coupling[27-31] between vibrational modes, confirms the absence of Fermi resonance (Figure 4.2 B) and indicates that the two C=O transitions in hexane arise from two conformations of phenyl formate, as predicted from theoretical and gas-phase results.[12] We have also measured the IR spectra in tetrahydrofuran (THF) which is more polar than n-hexane. In THF, also two peaks are observed with the peak frequencies red shifted from n-hexane peaks (Table 4.1). The reason of using THF can be better understood in the next section. From Table 4.1, it is evident that DFT analysis goes in hand in hand with the experimental observations. To find out the origin of the two non-covalent interactions, Natural bond orbital (NBO) calculation on the DFT optimized cis geometry were performed. NBO analysis shows the presence of n→π* interaction in hexane as well as in THF. The details of NBO analysis are mentioned in the chapter 2. From the structure

of phenyl formate, another possibility is the existence of $CH\cdots O$ hydrogen bonding. However the structural analysis shows that the distance and angle criteria are far away from the required distance and angle range to form the $CH\cdots O$ type interaction (Figure 4.3 and Table 4.2).
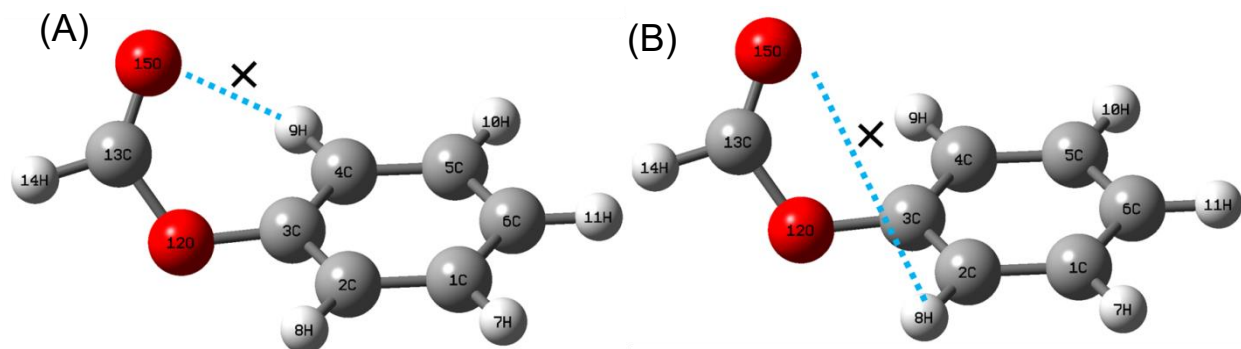

Figure 4.3: *Schematic representation of the absence $CH\cdots O$ type interaction in cis phenyl formate.*

Table 4.2 *Few geometrical parameters of the cis conformer of phenyl formate in gas phase, n-hexanes and THF calculated at M05-2X/cc- pVDZ level of theory.*

	Gas phase	Hexane	THF
∠H9-C4-C3-C2	179.8°	179.9°	179.8°
∠H8-C2-C3-C4	179.3°	179.5°	179.7°
∠H9-C4-C3-O12	4.1°	3.6°	3.0°
∠H8-C2-C3-O12	3.1°	3.0°	2.9°
∠C4-H9…O15	93.0°	86.8°	80.9°
∠C2-H8…O15	62.6°	64.9°	67.9°
H9…O15	2.7 Å	2.9 Å	3.1 Å
H8…O15	4.2 Å	4.0 Å	3.8 Å

4.2.2 Coexistence of n→π* and H-bonding interactions in solution

Linear IR spectra in protic solvation environment show signatures of C=O H-bonding interaction. When dissolved in water (D_2O), the overall spectrum shifts to lower frequencies (Figure 4.4). The spectrum can apparently be fitted to four peaks using Voigt lineshape, consisting of two additional C=O peaks at 1704 (11%) and 1724 cm^{-1} (44%) along with the peaks arising from non-H-bonded trans (11%) and cis (34%) conformation (Table 4.2).
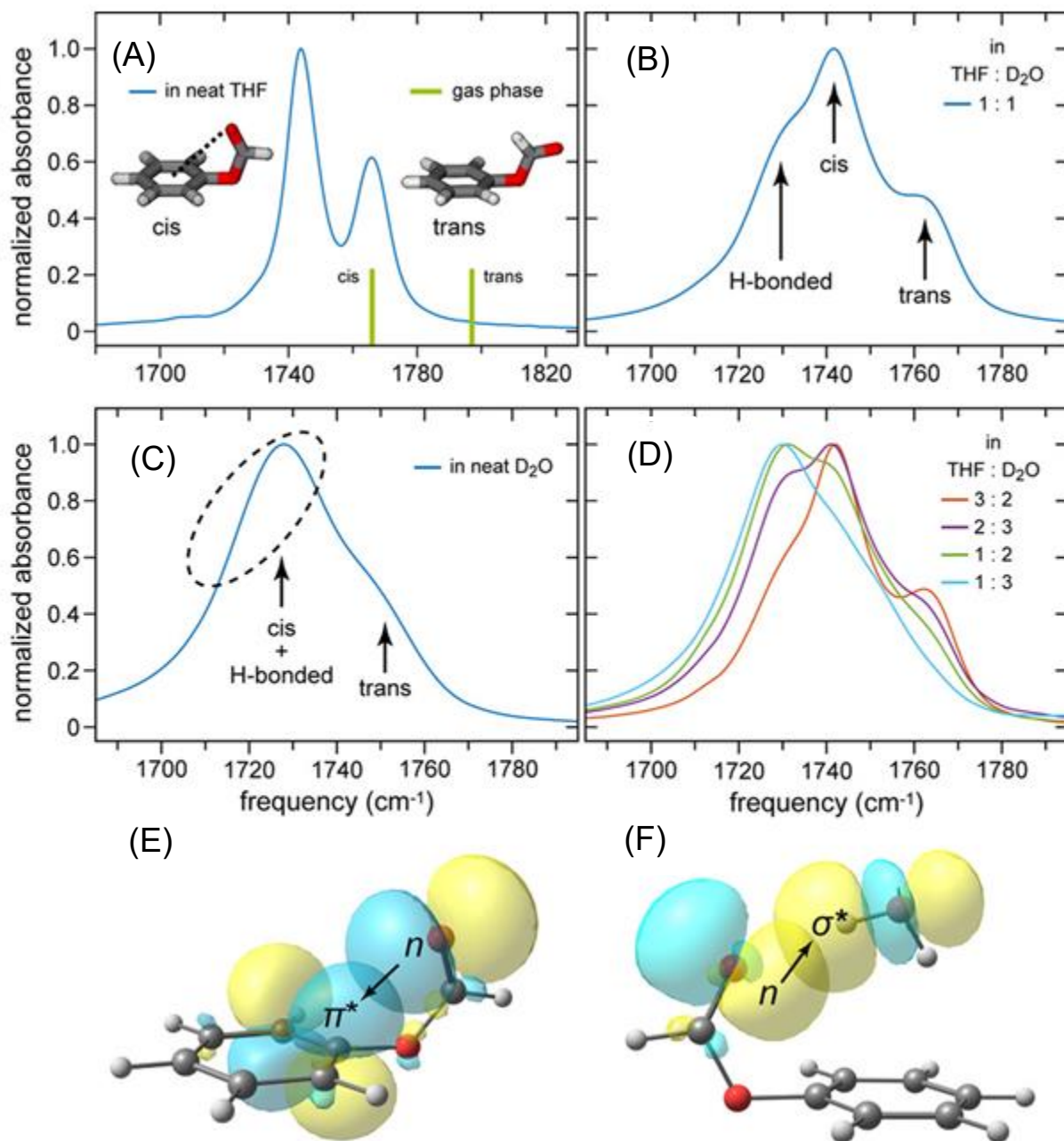
Figure 4.4: *Coexistence of Coulombic interaction, n→π* interaction and H-bonding interactions. Linear IR spectra of phenyl formate in solvents of (A) neat THF, (B) 1:1 THF/D_2O (v/v), (C) neat D_2O and (D) different aqueous THF solutions. Orbitals corresponding to (E) $n_{CO}→\pi^*_{ar}$ and (F) $n_{CO}→\sigma^*$(H-bond) NBO interactions for the cis and H-bonded conformers.*

The additional peak positions agree with the earlier reported H-bonded ester C=O transitions in water due to the formation of one and two H bonds between C=O and water.[32,33] To further investigate on the coexistence of n→π* and H-bond interactions, we have performed IR

experiments in aqueous binary mixtures of THF varying the volume percentage of water. Usage of aqueous binary mixtures allows us to systematically change the H-bonding environment around a probe. Fitting each linear IR spectrum with multiple Voigt lineshapes indicates that an increase in water content causes a decrease in the populations of both the non-H-bonded trans and cis conformers along with a concomitant increase in single H-bonded C=O population (Figure 4.4 and Table 4.3). At higher water concentrations, evidence of the double H-bonded population can be observed (Table 4.3), – a similar trend has been previously

Table 4.3: *Peak positions and population percentages of each phenyl formate conformer in aqueous solutions of THF.*

Water fraction (v/v)	Peak positions/cm^{-1}				Percentage contribution			
	cis	non-H-bonded trans	single H-bonded	double H-bonded	cis	non-H-bonded trans	single H-bonded	double H-bonded
0	1744	1766	-	-	66	34	-	-
0.2	1743	1763	1731	-	55	30	15	-
0.4	1743	1763	1730	1713	46	23	28	3
0.5	1743	1762	1729	1713	44	20	30	6
0.6	1743	1761	1729	1713	43	16	33	7
0.67	1742	1760	1728	1713	42	12	38	8
0.75	1739	1754	1727	1707	37	11	42	10
1.0	1735	1751	1724	1704	34	11	44	11

reported.[34] This experimental observation is in accordance with the theoretical prediction of H-bond to be energetically favorable. Thus the C=O group forms H bond in the expense of weaker non-covalent interaction.[9] Raines and coworkers suggested attenuation of the prevalence of n→π* interaction in the presence of H-bonds based on gas-phase theoretical calculations using NBO analysis on asparagine residues.[9] However, they surmised the H bond geometry would be compromised within the local structural constraints in proteins and thereby lead to a smaller attenuation of n→π* interaction as compared to the calculated gas-phase results. It is interesting to note that, though H bonded C=O is energetically favorable, 34% of the conformers of phenyl formate participate in n→π* interaction in neat water, elucidating the coexistence of H bond and n→π* interaction even in an unconstrained environment.

Atomistic molecular dynamics (MD) simulations, performed on phenyl formate in neat water to obtain a better representative picture of the experiments (Details of the simulation method is given in section 4.6.2) . MD shows the phenyl formate C=O (cis) is predominantly single H-

bonded to water (~56%), along with a 12% double H bonded and a 32% non-H-bonded population. Similar to the cis conformer, C=O forming one H bond with water is the predominant H bonded species for the trans conformer. Cis-trans interconversion was not observed during the MD run, irrespective of the starting configuration of phenyl formate. The extra non-H-bonded population in phenyl formate possibly arises due to the n→π* interaction. MD shows that the persistent solvent fluctuation dynamics result in instantaneous making and breaking of H-bonds, thereby a dynamical equilibrium exists between the H bonded and the non-H bonded forms.

4.2.3 Interconverting n→π* and H-bonding interaction in solution

The time-averaged linear IR results cannot capture the intrinsic dynamics of the water molecules around the C=O group. Waiting-time-dependent (T-dependent) 2D IR spectroscopy of C=O vibrations can provide the signature of conformational interconversion in the picosecond time regime, limited by the lifetime of the vibrational probe.[34-38] At values of T greater than the exchange timescale, interchanging conformational populations in equilibrium are manifested in the 2D IR spectra by the evolution of T-dependent cross peaks. We have observed the T-dependent evolution of cross peaks between the cis and the H bonded conformers in the 2D IR spectra of phenyl formate in a series of THF/water solutions of varying water content (Figure 4.5). Our results indicate a tug of war between the n→π* and the H bonding interactions in the picosecond timescale that can be directly visualized using 2D IR spectroscopy. The kinetic rates of population exchange for each THF/water composition have been estimated from the least-squares fitting of the corresponding numerically simulated 2D IR spectrum.[34,38] The interconversion rate increases with increasing water content in the aqueous binary mixtures (Figure 4.6).
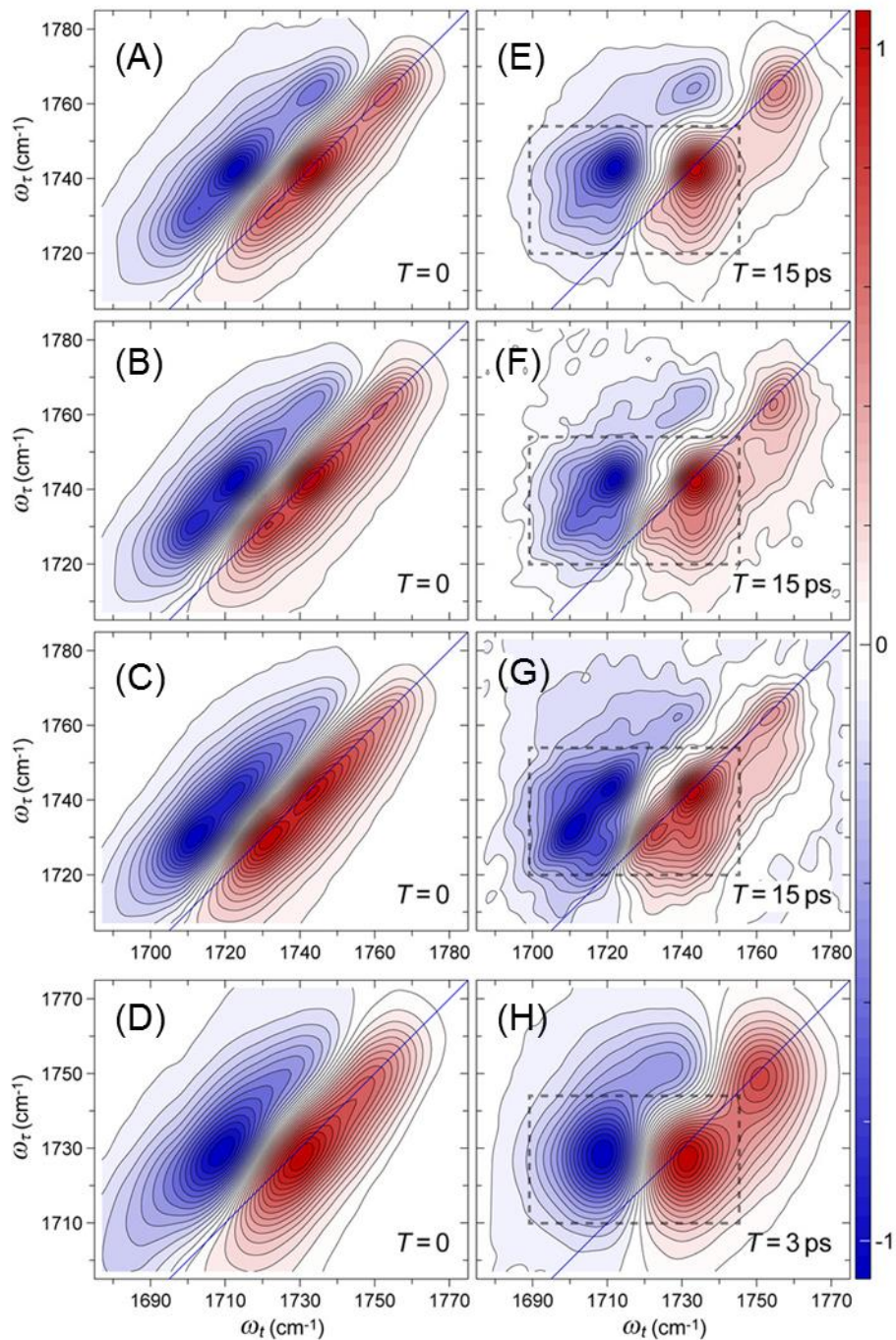
Figure 4.5 2D IR spectra with varying water content at different T values. 2D IR spectra of phenyl formate at T = 0 in (A) 1:1 THF/water, (B) 2:3 THF/water, (C) 1:2 THF/water, (D) neat water. (E−G) 2D IR spectra at T = 15 ps for corresponding THF/water solutions and (H) at T = 3 ps in neat water. The diagonal peak pairs show three C=O populations arising from trans, cis (n → π*), and H bonded conformers of phenyl formate in the order of decreasing frequencies. The cross peaks (within dashed boxes) between the cis and the H bonded conformations evolve as T increases.

Figure 4.6 *: Interconversion rate between n→π* interaction and H-bond.*

Earlier interpretation interrelating the two interactions utilized extensive statistical surveys of static protein crystal structures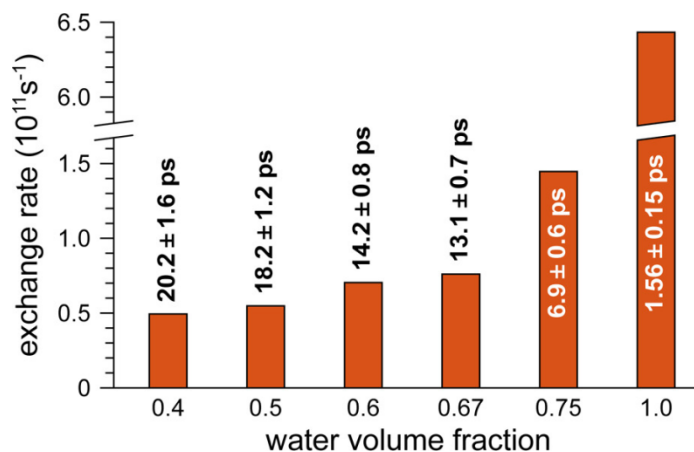[9] and NBO calculations,[39] however, direct experimental evidence of a dynamic equilibrium between the two competing n→π* and H bond interactions has been elusive to date. The trans conformer, where the C=O is rotated away from the phenyl ring, is also expected to undergo C=O H bond making and breaking.[34,38] Signature of such events in the 2D IR spectra is not as prominent as the interconversion between n→π* and H-bond interactions. We surmise that the larger population of the cis conformer in the corresponding THF/water mixtures (Table 4.3) as compared to the trans, enables us to clearly observe the dynamic equilibrium involving the cis form. The 2D IR results provide direct evidence of the interplay between two analogous interactions where the inherent fluctuations of the water molecules around the C=O allows the coexistence of n→π* and H bond interactions. We anticipate that the interchanging dynamics might have an important role in molecular stability.

4.3 Biological perspective

The linear IR and 2D IR results, though obtained for a small molecule, bear an important perspective for the n→π* interaction in biomolecules. Statistical analyses of protein crystal structures have previously been performed to identify regions of protein conformational space that could contain significant n→π* interactions.[5] We suggest a similar analysis based on water accessibility of the backbone carbonyls can provide valuable information on the role of n→π* interactions in stabilizing protein structure. The picosecond interchange illustrated in this work,

however, cannot be obtained from the static snapshots of the protein crystal structure, yet is crucial for our knowledge on n→π* interactions in biomolecules.

n→π* interactions are present in drug molecule like aspirin (acetylsalicylic acid, Figure 4.7). Aspirin is used as a medication to treat pain or inflammation. The pain relieving ability of aspirin stems from its ability to inhibit an enzyme, cyclooxygenases.[40] Aspirin contains an acetyl group which it transfers to a serine residue in the active site of cyclooxygenases. Once the serine -OH group got acetylated the activity of cyclooxygenases gets hindered. However, long term use of aspirin reduces thrombotic events, and leads to colon cancer and Alzheimer's disease[40]. Aspirin has mixed anyhydric character.[41] To reach in the active site, aspirin has
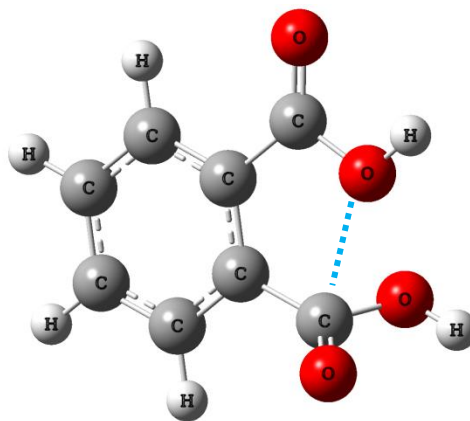
Fig 4.7: n→π* interaction in aspirin

to go through hydrophobic paraffin like pocket. So it was an open question how aspirin enter into a hydrophobic channel from an hydrophilic environment.[40] In a computational study Raines and coworkers first found that aspirin consists of a n→π* interaction. The electron density from the non-bonding orbital on hydroxyl oxygen are shared with the π* orbital of the ester C=O group (Figure 4.7).[42] They postulated that by engaging in n→π* interactions hydroxyl oxygen can prevent from the H bonding interactions. However the mechanism was not clear. But from our study we see that water dynamics plays a crucial role in the existence of n→π* interactions. Even though we have not studied with aspirin, our observations give a mechanistic insight in a qualitative manner. Form our study we can predict that dynamical surrounding of aspirin helps it exist in conformations having n→π* interactions. Thus it can pass through a hydrophobic channel.

4.4 Conclusion

The work provides a spectroscopic signature of n→π* interaction in solution by comparing IR spectra and theoretical calculations on phenyl formate. When exposed to water, n→π* and H bonding interactions are found to remain in equilibrium, interconverting between one another in picosecond timescale. The interchanging n→π* and H bonded populations, directly observed

from 2D IR experiments, provide important inferences from the biological perspective, which cannot be comprehended from steady-state experiments or protein crystal structure analysis.

4.5 Linear IR and 2D IR spectroscopy

Linear IR and 2D IR spectroscopy were performed on solutions such that the solute concentration is 50 mM. Linear IR spectra at different solute concentrations up to 50 mM were collected to check for aggregation. No aggregation was found. All spectra showed multiple populations. The solvent subtracted spectra were fitted to Voigt profiles to obtain the peak positions and the peaks areas. phenyl formate showed two transitions in the 1700-1800 cm-1 range. In aqueous binary mixtures, the spectra were fitted to either three or four Voigt profiles. FTIR spectrum for each sample was recorded before the 2D IR experiments. All linear IR spectra were measured in Bruker Vertex-70 spectrometer. The 2D IR spectra were recorded by prof. Yung-Sam Kim of the Dept of Chemistry, UNIST, Korea.

4.6 Computational study

4.6.1 Quantum calculations

Geometry optimizations and harmonic vibrational frequency calculations of different conformers of phenyl formate in gas phase and solution were performed at the M05-2X/aug- cc-pVDZ level of theory. For the solution phase calculation in different solvents (hexane and THF), we used polarizable continuum model (PCM) by incorporating the dielectric constant of individual solvent in the calculation. The DFT frequency is scaled frequencies as reported in ref 12. The details of NBO calculations has been given in chapter 2.

4.6.2 MD simulations

Molecular dynamics (MD) simulations were carried out on phenyl formate (starting from the cis and the trans configurations) using the GROMACS[43] (version: 4.6.7) software. General AMBER force field (GAFF)[44] was used to model cis and trans isomers of phenyl formate following the work of Caleman et al.[45] TIP3P water model was used in the simulations involving water molecules.[46] A single solute molecule was solvated in a cubic box of 5 nm dimension of water molecules. The solvated system was sequentially processed through (i) energy minimization using steepest descent algorithm, (ii) equilibration in NVT ensemble at 300 K temperature for 100 ps using the velocity rescale thermostat,[47] and (iii) equilibration in NPT ensemble at 300 K

temperature and 1 bar pressure (using Parrinello−Rahman barostat [48]) for 1 ns. Final MD production runs were continued for 10 ns, with positions and forces on each atom being saved at every 0.2 ps. A 10 ns long MD trajectory was used for analysis for each solute molecule. We adopted a geometric definition for defining a hydrogen bond between the carbonyl oxygen and a water molecule. The water molecule was considered to be hydrogen bonded with the solute C=O if the O (water) – O (C=O) distance was below 3.2Å and the H−O (donor) – O (acceptor) angle was below 20°.

4.7 References

(1)   Anfinsen, C. B. Principles that Govern the Folding of Protein. Chains *Science* **1973**, *181*, 223-230.

(2)   Dill, K. A. Dominant Forces in Protein Folding. *Biochemistry* **1990**, *29*, 7133-7155.

(3)   Hinderaker, M. P.; Raines, R. T. An Electronic Effect on Protein. Structure. *Protein Sci.* **2003**, *12*, 1188-1194.

(4)   Choudhary, A.; Gandla, D.; Krow, G. R.; Raines, R. T. Nature of Amide Carbonyl−Carbonyl Interactions in Proteins. *J. Am. Chem. Soc.* **2009**, *131*, 7244-7246.

(5)   Bartlett, G. J.; Choudhary, A.; Raines, R. T.; Woolfson, D. N. n→π* Interactions in Proteins. *Nat. Chem. Biol.* **2010**, *6*, 615-620.

(6)   Choudhary, A.; Kamer, K. J.; Raines, R. T. A Conserved Interaction with the Chromophore of Fluorescent Proteins. *Protein Sci.* **2012**, *21*, 171-177.

(7)   DeRider, M. L.; Wilkens, S. J.; Waddell, M. J.; Bretscher, L. E.; Weinhold, F.; Raines, R. T.; Markley, J. L. Collagen Stability: Insights from NMR Spectroscopic and Hybrid Density Functional Computational Investigations of the Effect of Electronegative Substituents on Prolyl Ring Conformations. *J. Am. Chem. Soc.* **2002**, *124*, 2497-2505.

(8)   Kuemin, M.; Nagel, Y. A.; Schweizer, S.; Monnard, F. W.; Ochsenfeld, C.; Wennemers, H. Tuning the cis/trans Conformer Ratio of Xaa–Pro Amide Bonds by Intramolecular Hydrogen Bonds: The Effect on PPII Helix Stability. *Angew. Chem. Int. Ed.* **2010**, *49*, 6324-6327.

(9)   Bartlett, G. J.; Newberry, R. W.; VanVeller, B.; Raines, R. T.; Woolfson, D. N. Interplay of Hydrogen Bonds and n→π* Interactions in Proteins. *J. Am. Chem. Soc.* **2013**, *135*, 18682-18688.

(10)   Singh, S. K.; Kumar, S.; Das, A. Competition Between n→π* and Conventional Hydrogen bonding (N-H···N) Interactions: an Ab Initio Study of the Complexes of 7-azaindole and Fluorosubstituted Pyridines. *Phys. Chem. Chem. Phys.* **2014**, *16*, 8819-8827.

(11)   Singh, S. K.; Das, A. The n→π* interaction: a Rapidly Emerging Non-covalent Interaction. *Phys. Chem. Chem. Phys.* **2015**, *17*, 9596-9612.

(12)   Singh, S. K.; Mishra, K. K.; Sharma, N.; Das, A. Direct Spectroscopic Evidence for an n→π* Interaction *Angew. Chem. Int. Ed.* **2016**, *55*, 7801-7805.

(13)   Choudhary, A.; Kamer, K. J.; Raines, R. T. An n→π* Interaction in Aspirin: Implications for Structure and Reactivity. *J. Org. Chem.* **2011**, *76*, 7933-7937.

(14)   Egli, M.; Sarkhel, S. Lone Pair−Aromatic Interactions: To Stabilize or Not to Stabilize. *Acc. Chem. Res.* **2007**, *40*, 197-205.

(15)   Gorske, B. C.; Bastian, B. L.; Geske, G. D.; Blackwell, H. E. Local and Tunable n→π* Interactions Regulate Amide Isomerism in the Peptoid Backbone. *J. Am. Chem. Soc.* **2007**, *129*, 8928-8929.

(16)   Mooibroek, T. J.; Gamez, P.; Reedijk, J. n→π* Interactions: a New Supramolecular Bond? *CrystEngComm* **2008**, *10*, 1501-1515.

(17)   Rahim, A.; Saha, P.; Jha, K. K.; Sukumar, N.; Sarma, B. K. Reciprocal Carbonyl–Carbonyl Interactions in Small Molecules and Proteins. *Nat. Commun* **2017**, *8*, 78.

(18)   Newberry, R. W.; Raines, R. T. The n→π* Interaction *Acc. Chem. Res.* **2017**.

(19)   Neel, A. J.; Hilton, M. J.; Sigman, M. S.; Toste, F. D. Exploiting non-covalent π Interactions for Catalyst Design. *Nature* **2017**, *543*, 637-646.

(20)   Park, S.; Kwak, K.; Fayer, M. D. Ultrafast 2D-IR Vibrational Echo Spectroscopy: a Probe of Molecular Dynamics. *Laser Phys Lett* **2007**, *4*, 704-718.

(21)   Hochstrasser, R. M. Two-dimensional spectroscopy at infrared and optical frequencies. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 14190-14196.

(22)   Ghosh, A.; Ostrander, J. S.; Zanni, M. T. Watching Proteins Wiggle: Mapping Structures with Two-Dimensional Infrared Spectroscopy. *Chem. Rev.* **2017**.

(23)   Fayer, M. D. Dynamics of Liquids, Molecules, and Proteins Measured with Ultrafast 2D IR Vibrational Echo Chemical Exchange Spectroscopy. *Annu. Rev. Phys. Chem.* **2009**, *60*, 21-38.

(24)   Edington, S. C.; Flanagan, J. C.; Baiz, C. R. An Empirical IR Frequency Map for Ester C═O Stretching Vibrations *J. Phys. Chem. A* **2016**, *120*, 3888-3896.

(25)   Sibert, E. L.; Tabor, D. P.; Kidwell, N. M.; Dean, J. C.; Zwier, T. S. Fermi Resonance Effects in the Vibrational Spectroscopy of Methyl and Methoxy Groups. *J. Phys. Chem. A* **2014**, *118*, 11272-11281.

(26)   Kondratyuk, P. Analytical formulas for Fermi Resonance Interactions in Continuous Distributions of States. *Spectrochim. Acta Mol. Biomol.* **2005**, *61*, 589-593.

(27)   Demirdöven, N.; Cheatum, C. M.; Chung, H. S.; Khalil, M.; Knoester, J.; Tokmakoff, A. Two-Dimensional Infrared Spectroscopy of Antiparallel β-Sheet Secondary Structure. *J. Am. Chem. Soc.* **2004**, *126*, 7981-7990.

(28)   Krummel, A. T.; Zanni, M. T. DNA Vibrational Coupling Revealed with Two-Dimensional Infrared Spectroscopy: Insight into Why Vibrational Spectroscopy Is Sensitive to DNA Structure. *J. Phys. Chem. B* **2006**, *110*, 13991-14000.

(29)   Gnanakaran, S.; Hochstrasser, R. M. Conformational Preferences and Vibrational Frequency Distributions of Short Peptides in Relation to Multidimensional Infrared Spectroscopy. *J. Am. Chem. Soc.* **2001**, *123*, 12886-12898.

(30)   Scheurer, C.; Piryatinski, A.; Mukamel, S. Signatures of β-Peptide Unfolding in Two-Dimensional Vibrational Echo Spectroscopy: A Simulation Study *J. Am. Chem. Soc.* **2001**, *123*, 3114-3124.

(31)   Edler, J.; Hamm, P. Two-dimensional Vibrational Spectroscopy of the Amide I Band of Crystalline Acetanilide: Fermi resonance, Conformational Substates, or Vibrational Self-Trapping? *J. Chem. Phys.* **2003**, *119*, 2709-2715.

(32)   Chuntonov, L.; Pazos, I. M.; Ma, J.; Gai, F. Kinetics of Exchange between Zero-, One-, and Two-Hydrogen-Bonded States of Methyl and Ethyl Acetate in Methanol. *J. Phys. Chem. B* **2015**, *119*, 4512-4520.

(33)   Pazos, I. M.; Ghosh, A.; Tucker, M. J.; Gai, F. Ester Carbonyl Vibration as a Sensitive Probe of Protein Local Electric Field. *Angew. Chem. Int. Ed.* **2014**, *53*, 6080-6084.

(34)   Kashid, S. M.; Jin, G. Y.; Chakrabarty, S.; Kim, Y. S.; Bagchi, S. Two-Dimensional Infrared Spectroscopy Reveals Cosolvent-Composition-Dependent Crossover in Intermolecular Hydrogen-Bond Dynamics. *J. Phys. Chem. Lett* **2017**, *8*, 1604-1609.

(35)   Kim, Y. S.; Hochstrasser, R. M. Chemical exchange 2D IR of hydrogen-bond making and breaking. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 11185-11190.

(36)   Kim, Y. S.; Hochstrasser, R. M. Comparison of Linear and 2D IR Spectra in the Presence of Fast Exchange. *J. Phys. Chem. B* **2006**, *110*, 8531-8534.

(37)   Zheng, J.; Kwak, K.; Asbury, J.; Chen, X.; Piletic, I. R.; Fayer, M. D. Ultrafast Dynamics of Solute-Solvent Complexation Observed at Thermal Equilibrium in Real Time. *Science* **2005**, *309*, 1338-1343.

(38)   Kashid, S. M.; Jin, G. Y.; Bagchi, S.; Kim, Y. S. Cosolvent Effects on Solute–Solvent Hydrogen-Bond Dynamics: Ultrafast 2D IR Investigations. *J. Phys. Chem. B* **2015**, *119*, 15334-15343.

(39)   Newberry, R. W.; Orke, S. J.; Raines, R. T. n→π* Interactions Are Competitive with Hydrogen Bonds. *Org. Lett.* **2016**, *18*, 3614-3617.

(40)   William L. Smith; David L. DeWitt, a.; Garavito, R. M. Cyclooxygenases: Structural, Cellular, and Molecular Biology. *Annu. Rev. Biochem* **2000**, *69*, 145-182.

(41)   Davidson, D.; Auerbach, L. The Acid Anhydride Character of Aspirin. *J. Am. Chem. Soc.* **1953**, *75*, 5984-5986.

(42)   Choudhary, A.; Kamer, K. J.; Raines, R. T. An n→π* Interaction in Aspirin: Implications for Structure and Reactivity. *J. Org. Chem.* **2011**, *76*, 7933-7937.

(43)   Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435-447.

(44)   Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

(45)   Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.* **2012**, *8*, 61-74.

(46)   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.

(47)   Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, -.

(48)   Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182-7190.

Chapter 5

Electric Field Catalysis of Diels-Alder Reactions

5.1 Introduction

Modern organic chemistry is driven by the objective of controlling reaction selectivity to accomplish targeted outcomes with desired functional groups. To this end, identification and optimization of effective catalysts have remained a constant quest in the realm of reaction discovery. All possible interactions, covalent and non-covalent (H-bonds), alter the transition state (TS) energies and lead to divergent reaction outcomes[1]. Due to the ambiguities in unraveling the contributions of the diverse non-covalent interactions, an explicit design-principle of controlling reaction pathway has remained challenging[2-6]. Electric field, a fundamental concept from physics, can provide a unified theory for comparing the relative importance of specific (H-bond) and non-specific (electrostatic) non-covalent interactions in a reaction[7,8]. Similar to the role of temperature and pressure, electric field of the order of V/Å (1 V/Å = 100 MV/cm) can distinctly alter the activation barrier of a chemical reaction, a process known as electrostatic catalysis[9,10]. Hence, discerning the role of electric field in an organic reaction is relevant to get a better control on the catalyst design process.

The concept of electric field induced chemical reaction originated from the electrostatic catalysis of bond heterolysis reaction in concentrated ionic solution by Pocker et al[11]. Since then, catalysis by external electric field (EEF) has extensively been utilized in several theoretical studies. Shaik and coworkers demonstrated that EEF can act as an effector for Diels-Alder (DA) reaction (Figure 5.1)[12]. Very recently, this theoretical prediction has been verified in an elegant experiment by Coote and coworkers for a DA reaction between a surface tethered furan and a non-polar dienophile[13], where electric field was created by applying a voltage difference between an STM tip holding the diene and the surface (Figure 5.2). The observed TS stabilization in presence of the applied electric field has opened up a new avenue of manipulating chemical reactions by means of electrostatics in heterogeneous catalysis.
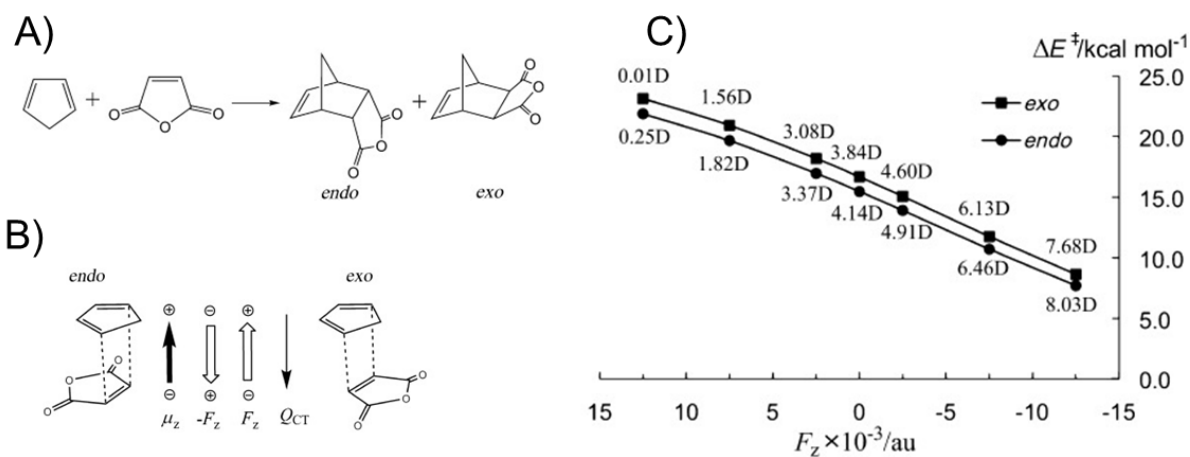
Figure 5.1: *This figure has been taken form the work of Sason Shaik.[12] A) Scheme of a Diels-Alder reaction between cyclopentadiene and maleic anhydride. B) Optimized transition states and the directions of transition state dipolemoment (z-component, μ_z), EEF ($\pm F_z$) and charge transfer (Q_{CT}). (C) Change in the energy barriers in presence of EEF. It is clear that the energy barrier decreases with increasing stabilizing electric field ($-F_z$).*
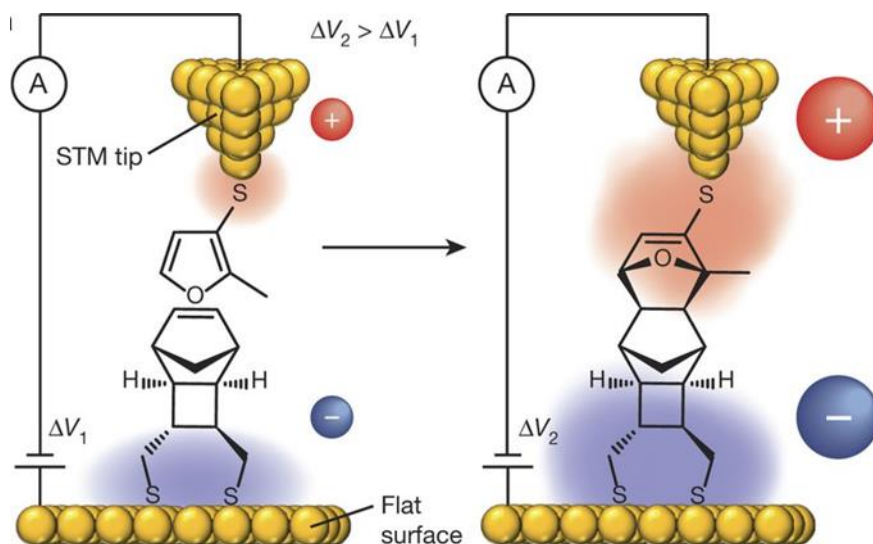
Figure 5.2: *This Figure has been taken from the work of Michelle L. Coote[13]. Diene is attached with STM tip and dienophile is is attached to a gold surface. In presence of a certain voltage difference (ΔV) the reaction happens.*

However, applying external electric field is a daunting task for the routinely performed organic reactions in homogeneous solutions. As electrostatic effects are strongly directional in nature, persistent solvent fluctuations in solution make it difficult to orient an external electric field along a fixed molecular axis. This makes the experimental realization of electrostatic catalysis of

organic reactions in solution unexplored to date. However, a typical polar environment (solvent, catalyst, or enzyme) in a homogeneous solution exerts an internal electric field (IEF) on the dissolved reactants. A large internal electric field arising from the charged/polar/polarizable groups at the enzyme active site of Ketosteroid isomerase (KSI) was shown by Boxer and coworkers to play a major role toward the TS stabilization and rate acceleration of an isomerization reaction (discussed in chapter 1)[14,15]. Interestingly, the internal fields for different KSI mutants showed a linear correlation with the free energy barrier of the reaction. Similar examples of catalytic activities of other enzymes have been reported where the electrostatic stabilization of TS is the crucial part of the catalytic process[16]. Although all previous examples span enzyme catalysis, these findings illustrate that IEF, conceptually similar to EEF, might be explored to decipher the fundamental understanding of the factors guiding the outcomes of organic reactions in solution.

Vibrational frequencies report on local electric field experienced by the vibration and are sensitive to the changes in the field through vibrational Stark effect (VSE). Thus, the IEF can be directly measured from infrared (IR) absorption spectroscopy. In other words, the IR peak positions can act as an experimental proxy of internal electric field to decipher electrostatic catalysis of organic reactions. To understand an organic reaction in the language of electric field, we have studied the DA reaction between cyclopentadiene (CP, diene) and methyl acrylate (MA, dienophile) in solution. Exploiting the concept of electric field induced changes in the IR peak position, we show that the internal solvent electric field catalyzes the DA reaction by stabilizing the polar TS of the reaction. Through a combined approach of IR spectroscopy, molecular dynamics (MD) simulation and reaction rate measurements, we show a strong linear correlation between the ensemble-averaged electric field induced by solvent and the free energy of activation of the DA reaction. We also employ density functional theory (DFT) to provide a deeper insight into the electrostatic catalysis of the organic reaction.

5.2 Results and discussion

5.2.1 Effect of EEF on the DA reaction

DA reaction, one of the most powerful carbon-carbon bond formation reactions, has afforded unparalleled solutions to numerous challenges in organic synthesis[17]. The reaction (Figure 5.3A) between cyclopentadiene (CP, diene) and methyl acrylate (MA, dienophile) yields two isomeric

products, exo (X) and endo (E), where the endo form is the kinetically controlled product. To ascertain whether the DA reaction between MA and CP is sensitive to electric field, we performed quantum chemical calculations of the reaction to get the free energy barriers ($\Delta G^{\#}$) at different EEF. The range of the applied EEF correspond to the reported range of IEF exerted on a small molecule solute by the different commonly used non-polar, polar-aprotic, and polar protic solvents used for DA reactions[15,18]. In Figure 5.3B we show the optimized geometry, direction of dipole moment and EEF direction for reactants and endo-TS. In presence of a large range of EEF, we observe that the $\Delta G^{\#}$ decreases as the stabilizing field increases more and more (Figure 5.3C). Figure 5.3C indicates that if one could determine $\Delta G^{\#}$ by measuring the rates of the DA reaction between CP and MA in different solvents and estimate the IEF exerted by the corresponding solvents along the TS dipole, one could determine whether the language of electric field can be extended to homogeneously catalyzed DA reactions.
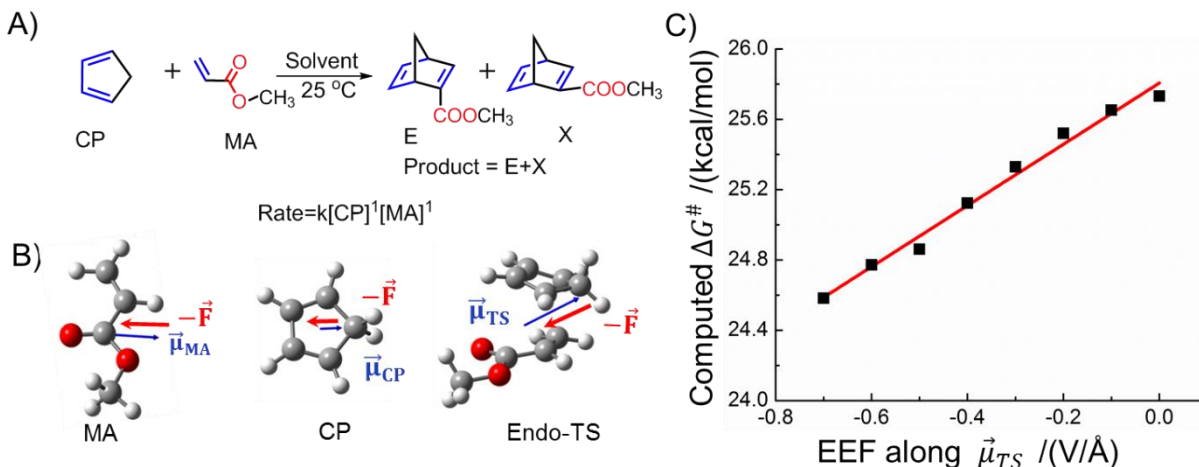

Figure 5.3: *Effect of the external electric-field (EEF) on the free energy barrier of the Diels-Alder reaction. A) The Diels-Alder (DA) reaction between CP (diene) and MA (dienophile) generates two diasteriomeric products, endo (E) and exo (X). Rate of the reaction is first order with respect to both the diene and dienophile and the E isomer is kinetically favored over the X isomer. B) The geometry optimization and frequency calculations were performed at M06-2X /6-311+G(d,p) level on the endo TS and the reactants in the presence of EEF (see section 5.4). The optimized structures of the reactants and the TS, along with the EEF ($\vec{F}$; the direction of positive field vector in Gaussian09 is defined from negative charge to the positive charge) and dipole moments ($\vec{\mu}$) directions are shown. Experimentally it is observed that the E isomer is the major product under kinetically controlled reaction condition and from computational study we saw that exo isomer does not show a clear field sensitivity. Therefore all quantum calculations were done for endo TS. C) Computed free energy barrier ($\Delta G^{\#}$) shows a linear dependence on the EEF applied along the transition state (TS) dipole. The least squares regression line (shown in*

73

red) is y = 0.57x – 14.62 with R² = 0.99. The robust linear correlation between computed $\Delta G^{\#}$ and EEF ascertains that the DA reaction between CP and MA is sensitive to electric-field.

5.2.2 Solvent effect on DA reaction kinetics

We performed reaction kinetics studies in different neat solvents of varying polarity and proticity (Figure 5.4 A and section 5.5). The solvent's role in DA reactions is commonly explained using frontier molecular orbital (FMO) theory[19], however, we observe no trend between the reaction rates and the corresponding quantum chemically estimated HOMO-LUMO energy gaps (Figure 5.4 B,C &D). This indicates we have to look beyond FMO to explain the solvent dependence for such reactions. Although measuring reaction rates is routine in synthetic chemistry, experimental estimation of IEF along the TS dipole is tricky and requires an explicit strategy.
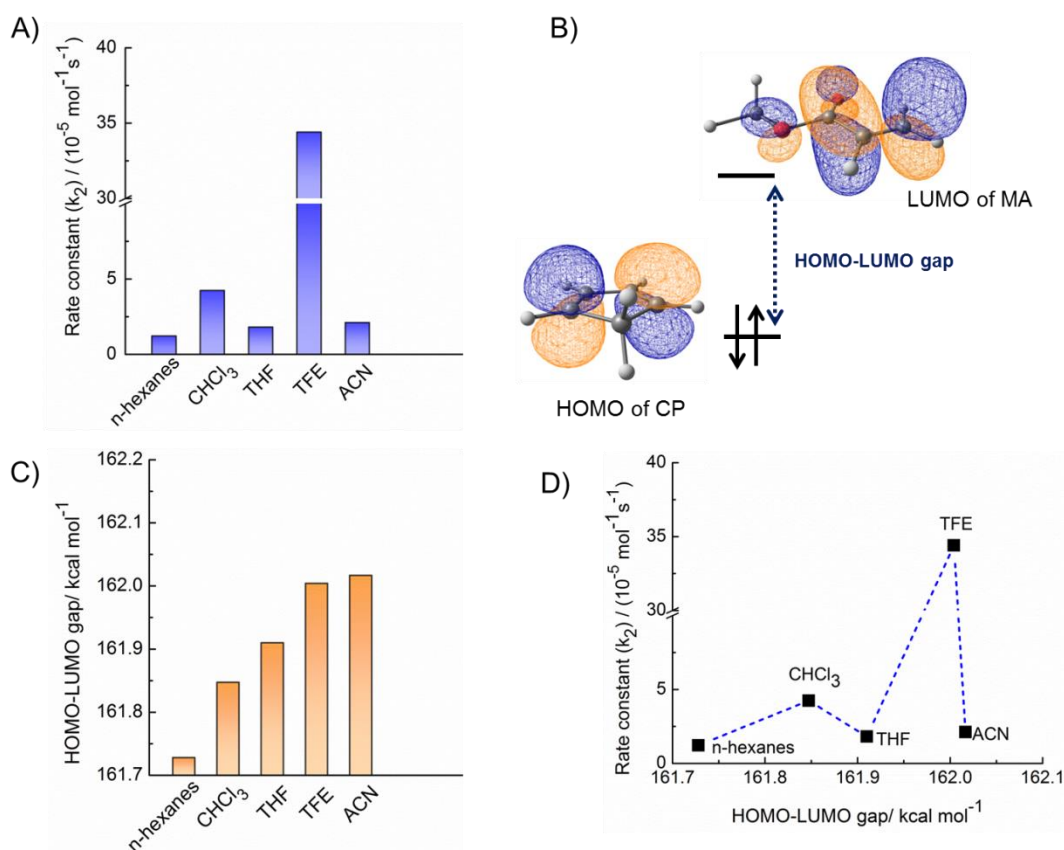
Figure 5.4*: Reaction kinetics and frontier molecular orbital (FMO) theory. (A) Second order rate constants for the DA reaction between CP and MA, measured in solvents of varying polarity and proticity, are shown. Solvent abbreviations are, CHCl₃: chloroform, THF: tetrahydrofuran, TFE: 2,2,2-trifluoro ethanol, ACN: acetonitrile. (B & C) Quantum calculations were performed employing polarizable continuum model (please see section 5.4) to estimate the HOMO - LUMO energy gaps in different solvents. (D) The HOMO – LUMO gaps show no correlation with the measured rate constants.*

5.2.3 Strategy to illustrate the role of IEF on reaction

If the angle between a certain bond dipole (IR probe) and the molecular dipole remains constant in the TS for different solvents, the solvent exerted IEF along the bond can serve as a surrogate for that along the TS dipole. Further, it is known that the solvent IEF induces a linear change in the frequencies of certain vibrations of the solute through linear vibrational Stark effect (VSE).[20,21] Through this effect, infrared (IR) frequencies of these vibrations can be used as probes of IEF along the bond. In particular, the carbonyl (C=O) bond has been shown to be sensitive to electric field in earlier reported solvatochromic and enzyme catalysis studies. [15,16,18] However, to use IR frequency of the C=O as a proxy of the electric field along the TS dipole, the angle between C=O and the TS dipole (θ, Figure 5.5A) has to remain constant throughout the solvents. One can evaluate θ between the C=O bond dipole and TS dipole in different solvents by combined quantum mechanics/molecular mechanics (QM/MM) calculations [22]. Another way of evaluation is to apply different EEF along the TS dipole and monitor the variation in θ because it has been recently reported that a uniform EEF can efficiently replicate the solvent effect on a reaction [23,24]. We optimized the TS structures by applying a range of EEF, comparable to those exerted by solvents of varying polarity and proticity, along the TS dipole. We observe only a small variation of θ ($< 5^{\circ}$, Figure 5.5 B) for the entire range of electric fields which indicates that if one could measure the C=O frequency ($\overline{\nu}_{C=O}$) of the TS in different solvents, one could effectively estimate the corresponding IEF along the TS dipole.
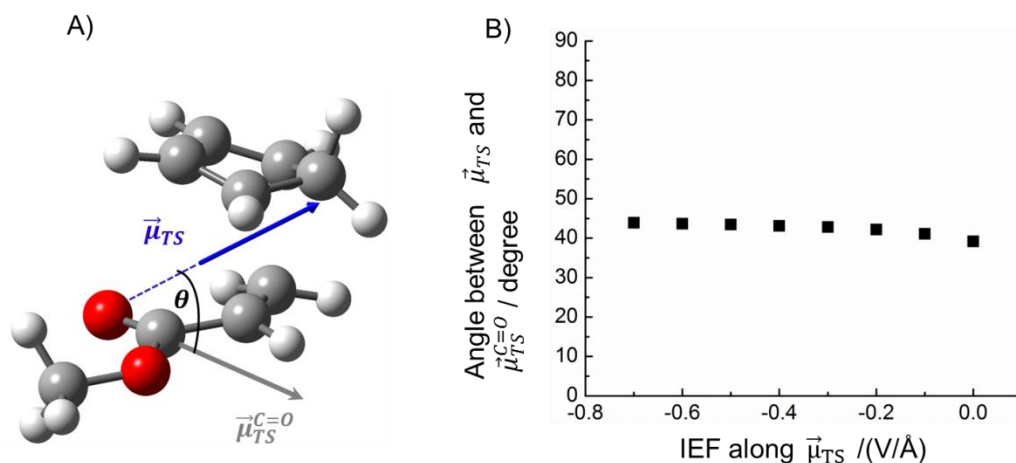

Figure 5.5: A)'θ' is the angle between $\vec{\mu}_{TS}$ and $\vec{\mu}_{TS}^{C=O}$ in Endo-TS. B) Variation of 'θ' in presence of EEF along $\vec{\mu}_{TS}$ of Endo-TS.

However, the TS cannot be isolated during the reaction. The energy profile obtained from the DFT calculations (Figure 5.6A) shows that the DA reaction has an early TS and the TS is structurally similar to the reactants. Our observation is in accordance with the Hammond's postulate.[25] We then estimated the electric field along the C=O bond for both the reactants (MA and CP) and the TS from atomistic molecular dynamics (MD) simulations in different solvents (Figure 5.6B). The estimated fields show a linear correlation between the IEF exerted on the C=O's of MA and TS (slope ~ 1, intercept ~ 0), which indicates that both the reactant and the TS experience a similar electric field along the C=O bond. Combining this result with the linear VSE of C=O implies that $\overline{v}_{C=O}$ of MA, which can be directly measured using IR spectroscopy, can act as the experimental proxy for the IEF along the TS dipole. Thus, in order to investigate whether electrostatic catalysis can be extended to DA reactions in solutions, we have employed the experimental strategy of measuring the $\overline{v}_{C=O}$ of MA in different solvents (Table 5.1).
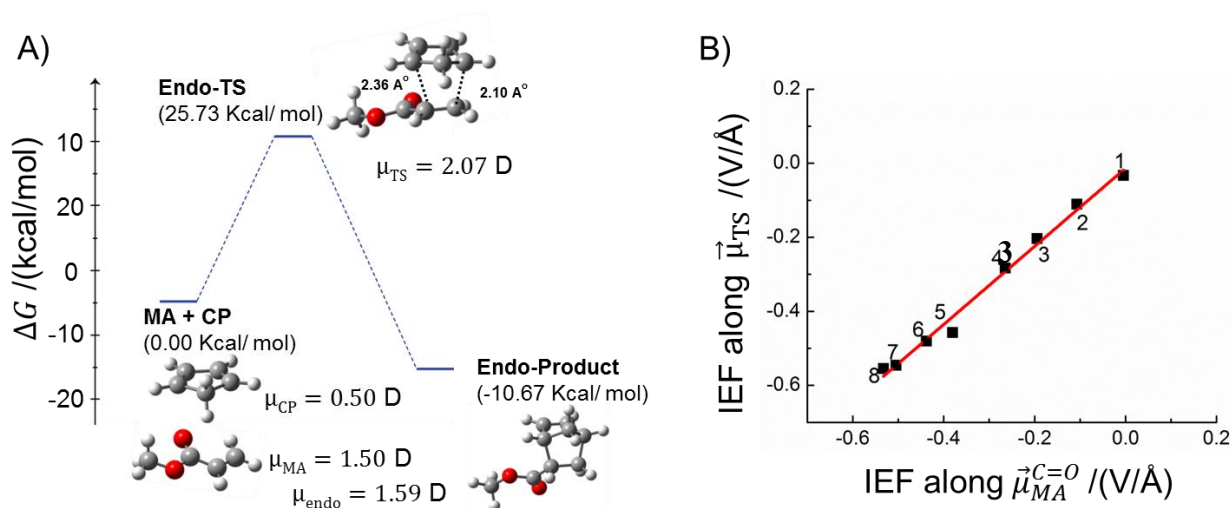
Figure 5.6: *(A) Gas phase energy profile shows that the endo TS of the DA reaction between CP and MA is an early TS. (B) Molecular dynamics (MD) simulations show a linear correlation between the computed IEF along the C=O bonds of MA and TS (see section 5.4 for details) for all the solvents (**1**- n-hexanes, **2**- THF, **3**- ACN, **4**- CHCl$_3$, **5**- 60ACN:40TFE(v/v), **6**- 40ACN:60TFE(v/v), **7**- 20ACN:80TFE(v/v), **8**- TFE). The equation of the least squares regression line (red) y = 1.06x - 0.01 (R² = 0.99) implies that the IEF along the TS C=O is similar to that along the reactant C=O.*

Figure 5.7 *(A) IR spectra of MA show increasing redshift in C=O stretching frequencies with increase in polarity and proti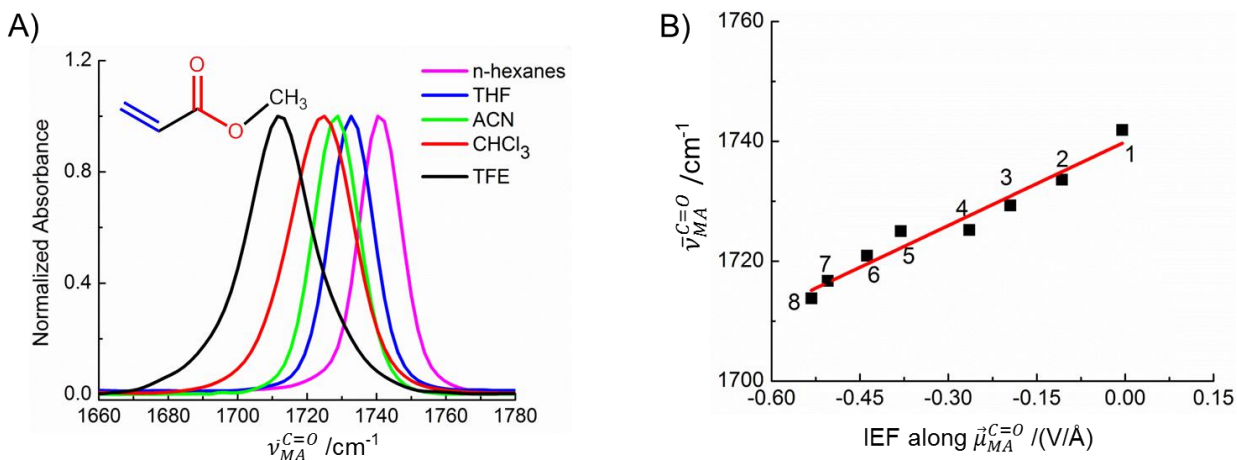city of the solvent. (B) The linear variation of the C=O frequency with the computed IEF in each solvent validates vibrational stark effect (VSE). The least squares regression line (red) is y = 46.50x + 1739.90 (R² = 0.95).*

Table 5.1 $\overline{v}_{C=O}$ of MA in solvent and solvent mixtures.

Solvents or solvent mixtures	C=O IR stretch (cm⁻¹) of MA
n-hexanes	1741.9
THF	1733.6
ACN	1729.3
CHCl₃	1725.2
TFE	1713.8
TFE80:ACN20	1716.7
TFE60:ACN40	1720.9
TFE40:ACN60	1725

IR measurements (Figure 5.7A and Table 5.1) show that $\overline{v}_{C=O}$ shifts consistently to the red with increasing solvent polarity, from 1741.9 cm⁻¹ in nonpolar hexane to 1713.8 cm⁻¹ in the polar protic solvent, 2,2,2-trifluoroethanol (TFE). Similar solvatochromic IR shifts in $\overline{v}_{C=O}$ of small

solute molecules have been reported earlier and the large redshift in a protic solvent has been explained by the large IEF created by the solvent's hydrogen bond (H-bond) [15]. The solvent exerted IEF along the MA C=O bond, computed from MD simulations, are linearly correlated to $\bar{\nu}_{C=O}$. The regression line implies that the change in $\bar{\nu}_{C=O}$ in different solvents can be explained as a field effect (Figure 5.7 B) and validates that $\bar{\nu}_{C=O}$ can report on the solvent IEF experienced by the vibration.

The estimated $\Delta G^{\#}$ values for the DA reaction, has been obtained using the Eyring equation:

$$k_2 = \frac{k_B T}{h} \exp\left(-\frac{\Delta G^{\#}}{RT}\right) \qquad\qquad 5.1$$

Where k_2 is second order rate constant obtained from kinetic measurements in solvent, k_B is the Boltzman constant, T is absolute temperature at which reactions were performed, h is Planck's constant, R is the gas constant and $\Delta G^{\#}$ is free energy of activation.[26] $\Delta G^{\#}$ show a robust linear correlation either with $\bar{\nu}_{C=O}$ or with the MD computed IEF for different solvents (Figure 5.8). The reduction of $\Delta G^{\#}$ with increase in IEF of the solvent is akin to previously reported electric field catalysis where an EEF lowers the reaction free energy barrier, as compared to an
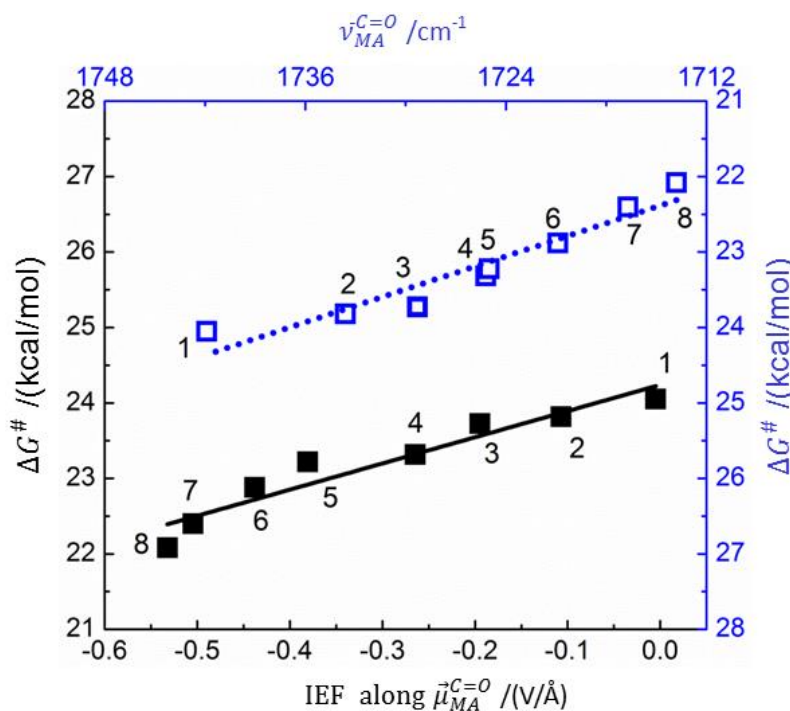

Figure 5.8: *Electrostatic catalysis by internal electric-field (IEF). A linear correlation is observed between $\Delta G^{\#}$ for the DA reaction and the IEF acting along the C=O bond of MA*

uncatalyzed reaction.[13] This demonstrates that the DA reactions routinely performed in the laboratory in solutions are governed by the same principles of electrostatic catalysis, as was earlier shown for heterogeneous catalysis of DA reaction. DA reactions are also performed in protic-aprotic solvent mixtures. To capture the effect of binary solvent mixtures on DA reactions, we performed both reaction rate and IR measurements in different acetonitrile/TFE mixtures where the volume ratios of the protic to aprotic solvent were varied. Figure 5.8 validates electrostatic catalysis of DA reactions in both neat solvents and binary solvent mixtures. This can be explained by considering that IEF depends on the partial atomic charges and their respective distances from the point where the IEF is estimated, irrespective of the molecular identity of the solvent. Apparently, the variation in reaction rate in protic solvents can be explained with the extent of H- bonding (as seen in different acetonitrile/TFE mixtures). However, a solvent capable of H-bonding to the C=O group exerts a larger IEF. This can be rationalized by considering that an H- bond donor positions a large dipole very close to the C=O group due to the small van der Waals radius of hydrogen. This suggests that an apparent rate increase with the strength of H-bond can be explained in terms of electrostatic catalysis.

It is evident that the solvent IEF reduces the activation barrier of the DA reaction. In solution, a dipolar solute orients the dipole of solvent molecules to minimize system's potential energy resulting in a net stabilizing IEF along the solute's dipole.[27] In addition to stabilizing the reactants, the solvent molecules can reorient during the reaction course to stabilize the TS. Quantum calculations show that both the reactants and the TS have nonzero dipole moments. However, as the dipole moment of the TS is larger than that of the reactants (Figure 5.6 A), the solvent IEF stabilizes the TS more than the reactant, resulting in the electrostatic catalysis of the DA reaction. A larger electric field in polar protic solvents, resulting in a larger stabilization of the TS as compared to the reactants, forms the basis of the experimentally observed reaction rate enhancement.

5.2.4 Generality of the electrostatics catalysis in DA reaction

To seek for the generality of electrostatic catalysis in DA reactions we have looked into a hetero Diels-alder reaction, between an electron rich diene and p-methoxy anisaldehyde (Figure 5.9 A), where the reported reaction rates were measured by NMR spectroscopy.[28] Using IR spectroscopy, we measured $\overline{v}_{C=O}$ of the dienophile in the solvents where reaction rates have been previously reported (Figure 5.9B). The IR peak appears as a doublet, where the higher frequency peak is from C=O absorption and the lower frequency peak arises from Fermi resonance (please see the section 5.6).[29] Similar to our findings for the reaction between MA and CP, $\overline{v}_{C=O}$ (higher frequency peak,) gradually redshifts (Figure 5.9B) with increasing solvent polarity and shows linear correlation with $\Delta G^{\#}$ estimated from the reported rate constants using Eyring equation (Figure 5.9 C). This illustrates that the electrostatic catalysis is not limited to the DA reaction between CP and MA, but is applicable to other DA reactions. We believe that electrostatic catalysis is a general phenomenon for DA reactions in solution where the larger electric field from polar protic solvents stabilizes the polar TS more than the reactants and thereby results in reaction rate acceleration.
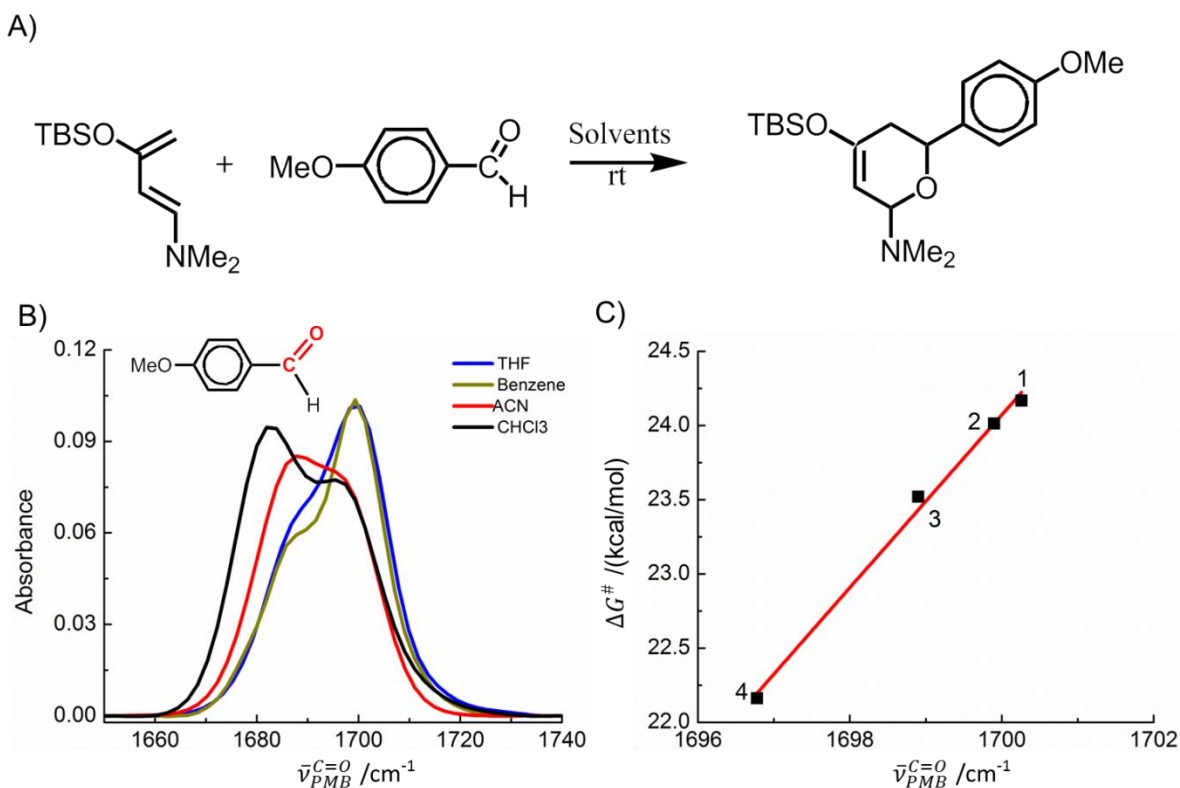

Figure 5.9: *Electrostatics catalysis in hetero DA reaction. A hetero DA reaction (A) between p-methoxy benzaldehyde (PMB) and an electron rich diene was studied. (B) IR solvatochromic*

*study of PMB were performed in different solvents. The higher frequency peaks arise from C=O absorption whereas the lower frequency peaks of PMB are from the Fermi resonance between some low frequency mode (section 5.6). (C) $\Delta G^{\#}$ of the reaction in solvents (**1-** THF, 2: Benzene, **3-** ACN, **4-** CHCl₃) were calculated from the second order rate constants using Eyring equation (equation 5.1). An excellent correlation between $\Delta G^{\#}$ and C=O stretching frequency (higher frequencies) of PMB enforces the role of electrostatics in organic reactions which involve polar transition states. The least squares regression line (red) is y = 0.58x - 965.66 with R² = 0.99.*

5.3 Conclusion

This study provides the first evidence (to our knowledge) of the electrostatic catalysis of DA reactions in solutions. We show that IEF catalyzes DA reaction between cyclopentadiene (CP) and methyl acrylate (MA) in solution by stabilizing the polar TS of the reaction. We have also extended our study to another hetero DA reaction involving different reactants. Moreover, we report a robust linear correlation between the IEF induced by solvent and $\Delta G^{\#}$ of these DA reactions. Our results demonstrate that IEF exerted by an organized environment can catalyze solution phase DA reactions because the TS is more polar as compared to the reactants.

5.4 Computational details

All the calculations in this study have been performed with density functional theory (DFT), with the aid of the Gaussian09 suite of programs,[30] using the M06-2X functional.[31,32] The 6-311+G(d,p) basis set has been employed. We have not included empirical dispersion correction separately as M06-2X functional already taken care of dispersion correction. The values reported are ΔG values, with zero point energy corrections, internal energy and entropic contributions included through frequency calculations on the optimized minima with the temperature taken to be 298.15 K. Harmonic frequency calculations were performed for all stationary points to confirm them as a local minima or transition state structures.

The optimized Cartesian coordinates and the energy values are given below:

## **CP**
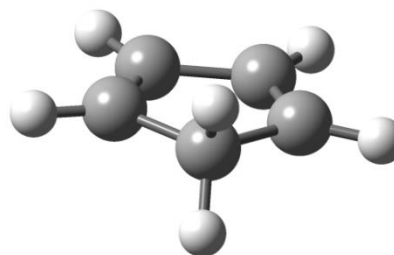
0 1

C   -1.17614600 -0.28069600   0.00006700
C   -0.73498400   0.98776600 -0.00011500
C   0.73496800   0.98777900   0.00011700
C   1.17615100 -0.28067400 -0.00007900
H   -2.20672400 -0.60715700   0.00012200
H   -1.34994300   1.87796900 -0.00016700
H   1.34990200   1.87800000   0.00018300
H   2.20673200 -0.60712500 -0.00012000
C   0.00000900 -1.21453600   0.00000400
H   -0.00001600 -1.86977200 -0.87944300
H   0.00005900 -1.86975100   0.87946800

Zero-point correction=
0.093201 (Hartree/Particle)
Thermal correction to Energy=
0.097331
Thermal correction to Enthalpy=
0.098275
Thermal correction to Gibbs Free Energy=
0.066612
Sum of electronic and zero-point Energies=
-193.965826
Sum of electronic and thermal Energies=
-193.961696
Sum of electronic and thermal Enthalpies=
-193.960752
Sum of electronic and thermal Free
Energies=   -193.992414
Number of Imaginary Frequency=0

MA
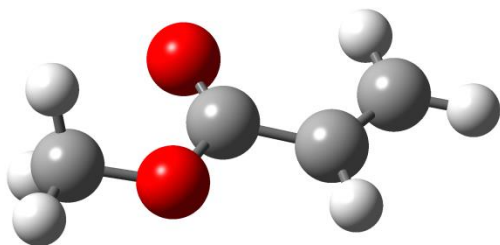
0 1

C   2.47623000 -0.00207400 -0.00000700
H   3.42090000 -0.53137500 -0.00006700

H   2.49139800   1.08243900   0.00011800
C   1.31791700 -0.64987000 -0.00008600
H   1.24539200 -1.73046000 -0.00021000
C   -2.29189200 -0.08256900   0.00002300
H   -2.40330300   0.54075100 -0.88774500
H   -3.02255900 -0.88685900 -0.00001000
H   -2.40324400   0.54062800   0.88788400
O   -1.01352300 -0.71890900 -0.00006300
C   0.04090500   0.11117600   0.00000500
O   -0.05992000   1.30952200   0.00011600

Zero-point   correction=
0.095881 (Hartree/Particle)
Thermal   correction   to   Energy=
0.102560
Thermal   correction   to   Enthalpy=
0.103504
Thermal correction to Gibbs Free Energy=
0.065151
Sum of electronic and zero-point Energies=
-306.331306
Sum of electronic and thermal Energies=
-306.324627
Sum of electronic and thermal Enthalpies=
-306.323683
Sum of electronic and thermal Free
Energies=   -306.362036
Number of Imaginary Frequency=0
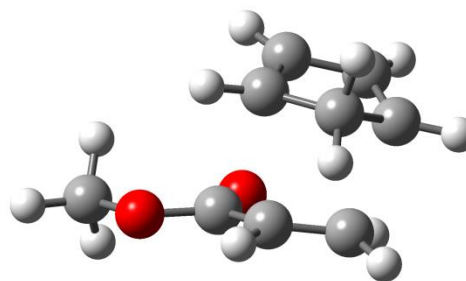

Endo-TS

0 1

C   0.87638100   1.68801400 -0.04756700
H   1.52212800   2.34486100 -0.61811900
H   0.68745900   1.97964700   0.97744600
C -0.10340600   0.96289400 -0.70800600
H -0.19429300   0.95639000 -1.78573300

C -3.35881100 -0.54090600 -0.04737500
H -3.83995100   0.25173100   0.52697200
H -4.02847300 -0.91391600 -0.81849300
H -3.07443400 -1.34505600   0.63329700
O -2.21532300 -0.03934000 -0.73311300
C -1.25821400   0.49196200   0.05922700
O -1.36147700   0.53561000   1.26305000
C 2.40293900   0.29644500   0.31362200
C 1.78318900 -0.53669200   1.24825000
C 0.90227700 -1.40058100   0.58123400
C 0.96518900 -1.14181900 -0.77861100
H 3.24643400   0.93409600   0.54843100
H 1.86235500 -0.43757600   2.32250900
H 0.18986700 -2.05474900   1.06501700
H 0.41968800 -1.67830200 -1.54410800
C 2.22091400 -0.35947900 -1.03435200
H 2.19799700   0.30981500 -1.89226300
H 3.04286300 -1.07614200 -1.17298000

Zero-point   correction=
0.192393(Hartree/Particle)
Thermal   correction   to   Energy=
0.202557
Thermal   correction   to   Enthalpy=
0.203501
Thermal correction to Gibbs Free Energy=
0.156644
Sum of electronic and zero-point Energies=
-500.277694
Sum of electronic and thermal Energies=
-500.267530
Sum of electronic and thermal Enthalpies=
-500.266586
Sum of electronic and thermal Free
Energies=   -500.313444
Number of Imaginary Frequency=1, Value
of Imaginary Frequency= - 464.12

Exo-TS
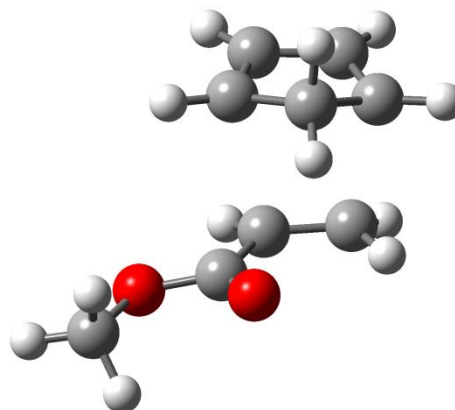
0 1

C 0.75683200 -1.38471300 -0.90655600
H 0.50207900 -2.23808500 -0.28715800
H 1.39532500 -1.58900500 -1.75426500
C -0.14583900 -0.33453100 -0.96524300
H -0.14419000   0.36660100 -1.78709000
C -3.49095200   0.58740200   0.15328000
H -3.31748900   0.78527300   1.21206200
H -4.07824900   1.38591900 -0.29335800
H -4.00500500 -0.36919300   0.05364100
O -2.25724100   0.56779700 -0.56023200
C -1.38178600 -0.38450100 -0.17405100
O -1.62298400 -1.15697100   0.72518800
C 2.29101300 -0.68907800   0.39333400
C 2.81050700   0.35322500 -0.37996400
C 1.94545100   1.45348700 -0.30275000
C 0.88834300   1.13421100   0.53990200
H 2.83284900 -1.59799900   0.62332100
H 3.66619100   0.27590400 -1.03802000
H 2.02262100   2.35561300 -0.89480000
H 0.09684000   1.81378200   0.83020500
C 1.28672400 -0.07459400   1.33181600
H 0.47357600 -0.71891600   1.65960000
H 1.83549600   0.27804300   2.21761300
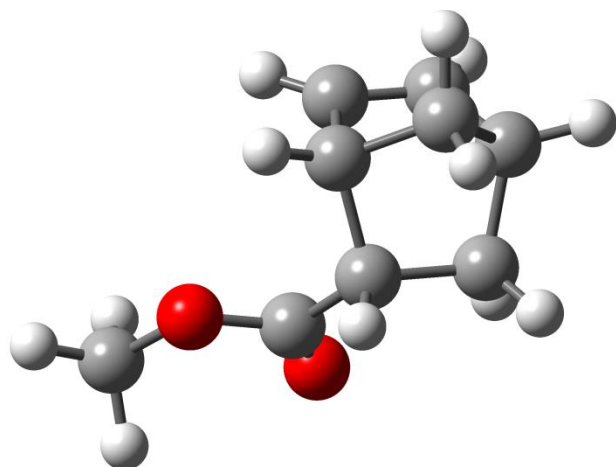

Zero-point   correction=
0.192379 (Hartree/Particle)
Thermal   correction   to   Energy=
0.202553
Thermal   correction   to   Enthalpy=
0.203497
Thermal correction to Gibbs Free Energy=
0.156644

Sum of electronic and zero-point Energies= -500.276959
Sum of electronic and thermal Energies= -500.266785
Sum of electronic and thermal Enthalpies= -500.265841
Sum of electronic and thermal Free Energies=   -500.312693
Number of Imaginary Frequency=1, Value of Imaginary Frequency= - 461.51

Endo-product (E)

0 1

```
C   1.14063700   1.38956400  -0.33315200
H   1.48820800   2.01284800  -1.15904100
H   0.73858500   2.03876300   0.44366300
C   0.08860100   0.36388700  -0.81908300
H  -0.09110500   0.44911500  -1.89475800
C  -3.35023000  -0.52716300   0.12620800
H  -3.89611400   0.36592400  -0.17957400
H  -3.85034300  -1.42343500  -0.23179100
H  -3.27171000  -0.53913100   1.21386000
O  -2.05081600  -0.54415900  -0.46513200
C  -1.25436100   0.49189200  -0.14218900
```

O -1.60495100   1.37840100   0.58817000
C   2.29535700   0.47394100   0.18286200
C   1.74675700 -0.24252600   1.40360000
C   0.84669400 -1.13362300   0.98180700
C   0.77237000 -1.02076400 -0.53086300
H   3.24793100   0.98405300   0.31087100
H   1.96976400   0.02268100   2.42915900
H   0.18863600 -1.74018900   1.59075700
H   0.31588700 -1.85249500 -1.06139300
C   2.23284000 -0.66198300 -0.85683000
H   2.37325700 -0.31260900 -1.88333800
H   2.92114000 -1.47879500 -0.63687900

Zero-point   correction= 0.197286 (Hartree/Particle)
Thermal   correction   to   Energy= 0.206768
Thermal   correction   to   Enthalpy= 0.207713
Thermal correction to Gibbs Free Energy= 0.161911
Sum of electronic and zero-point Energies= -500.336075
Sum of electronic and thermal Energies= -500.326593
Sum of electronic and thermal Enthalpies= -500.325648
Sum of electronic and thermal Free Energies=   -500.371450
Number of Imaginary Frequency=0

Exo-product (X)

C 2.81919900   0.41911500   -0.34151100
C 2.01092400   1.43499100   -0.03135200
C 0.75562500   0.84407800   0.58303800
H 2.74458400   -1.74646800   0.16988600
H 3.75418300   0.47248700   -0.88455700
H 2.14652000   2.48164400   -0.27091300
H 0.12711100   1.52216100   1.15780000
C 1.34003400   -0.38763500   1.29920800
H 0.57976800   -1.09865700   1.62894100
H 1.99423400   -0.11078500   2.12642300

Zero-point correction= 0.197286 (Hartree/Particle)

Thermal correction to Energy= 0.206768

Thermal correction to Enthalpy= 0.207713

Thermal correction to Gibbs Free Energy= 0.161911

Sum of electronic and zero-point Energies= -500.336075

Sum of electronic and thermal Energies= -500.326593

Sum of electronic and thermal Enthalpies= -500.325648

Sum of electronic and thermal Free Energies= -500.371450
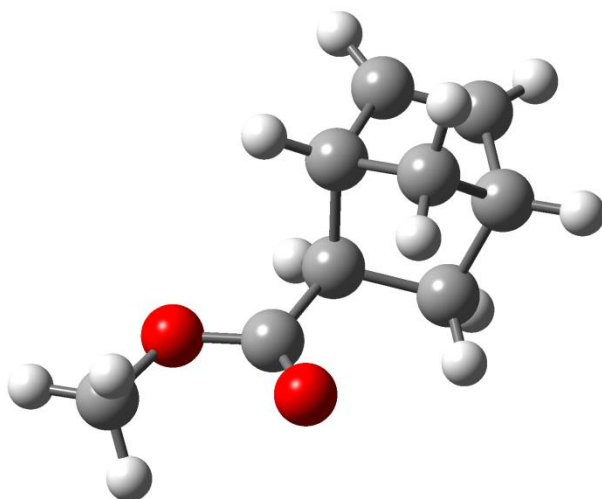
Number of Imaginary Frequency=0

0 1

C 0.93670100   -1.04842200   -0.94899800
H 0.42113800   -1.99077300   -0.75814800
H 1.27328800   -1.03243000   -1.98538500
C 0.00647300   0.14480400   -0.62030200
H -0.10840400   0.84843400   -1.44377600
C -3.58237900   0.45134300   0.11591600
H -3.57968900   0.26300000   1.19003700
H -4.16273600   1.34044100   -0.11642500
H -3.99220300   -0.42004000   -0.39550600
O -2.25793100   0.72056800   -0.34523400
C -1.37044400   -0.27328100   -0.16798700
O -1.66592300   -1.32638300   0.33356500
C 2.11604000   -0.86540900   0.05781700

We optimized the transition state structure at M06-2X /6-311+G(d,p) lebel of theory and found out the unit vector along the dipole moment of the transition state. Again we reoptimized the transition state geometry in presence of an electric field which was applied in the opposite direction of the dipole moment unit vector using "Field=Read" keyword. The "Nosymm" keyword was used to avoid the rotation of the Cartesian axes during the process of optimization. Geometry optimizations were done at electric field strengths between 0.0V/Åto -0.70 V/Å at -0.10 V/Å (-0.10 V/Å= 0.1950 x 10^{-2} au) intervals.

Energy differences between HOMO of CP and LUMO of MA were calculated following the method stated by Doren and coworkers.[33] Geometry optimization was performed for both MA and CP at M06-2X /6-311+G(d,p) level of theory. Vibrational frequency calculations of the

optimized structures show the absence of negative frequency and therefore confirm the stable optimized geometry. Effect of solvents on the HOMO-LUMO gap was studied using polarization continuum model (PCM).[34-36] The HOMO-LUMO gaps are given in Table 5.2. We have also performed the HOMO-LUMO calculations at HF level which show that the same trend as observed at M06-2X /6-311+G(d,p) level of theory.

*Table 5.2: Calculated HOMO - LUMO gap in solvents with different dielectric constants*

System	Dielectric constant (ε)	HOMO-LUMO energy gap at M06-2X/6-311+G(d,p) (kcal/mol)	HOMO-LUMO energy gap at HF (kcal/mol)
Gas phase	0.00	161.7	262.5
n-hexane	1.88	161.7	262.6
THF	7.43	161.9	262.9
ACN	35.69	162.0	263.0

Molecular dynamics (MD)simulations were carried out using the GROMACS[37] (version: 4.6.7) software. General AMBER force field (GAFF)[38] was used to model MA. Topology file of MA was generated using the procedure as described by Spool and coworkers.[39] Geometry optimization of MA was performed at HF/6-311G** level of theory using Gaussian 09 and the partial charges were obtaine using MerzSinghKollman (MK) scheme[40] in Gaussian. The Gaussian output (with .gout extension) file was used to create mol2 file using Antichamber.[38,41] The charges in mol2 file are RESP charges created by Antichamber. Finally GROMACS toplology file was generated using acpype script.[42] OPLS-AA force field parameters with 1.14*CM1A partial atomic charges were used to model the solvent molecules with the help of a well-known web-based service, LigParGen, developed by the Jorgenson group.[43-45] A single MA molecule was solvated in a cubic box of 5 nm dimension of respective solvent molecules. The solvated system was sequentially processed through energy minimization using steepest descent algorithm, equilibration in NVT ensemble at 300 K temperature for 100 ps using the velocity rescale thermostat,[46] and equilibration in NPT ensemble at 300 K temperature and 1 bar pressure

(using Parrinello−Rahman barostat)[47] for 1ns. Finally the production run was continued for 5 ns, with positions and forces on each atom being saved at every 0.2 ps. The electric field exerted onto the C=O bond by the solvation environment was calculated by projecting the total electrostatic field due to the rest of the system onto the C=O bond vector.[18,48] The solvent electric field on the C=O of endo/exo TS was calculated in a similar manner putting positioned restrained on the coordinates on the coordinates of the TS. Finally average electric field on the transition state were calculated using the experimentally obtained product ratio of the endo and exo product.

5.5 Kinetic study – experimental details

The reaction between methyl acrylate (MA) and cyclopentadiene (CP) in all solvents were carried out under second order condition (equal concentration of MA and CP). CP was cracked from dicyclopentadiene prior to the kinetics measurement (scheme5.2). 1.34 µL of MA (1M) and 0.96 µL of CP (1M) were added to 9.4 µL of solvent or solvent mixtures in a 25 mL round bottomed flask jacketed with a glass beaker whose inlet and outlet were connected to the chiller which was used to maintain the reaction temperature at $25 \pm 0.1\ ^0C$.

Dicyclopentadiene   ~190⁰C →   2 Cyclopentadiene (CP)
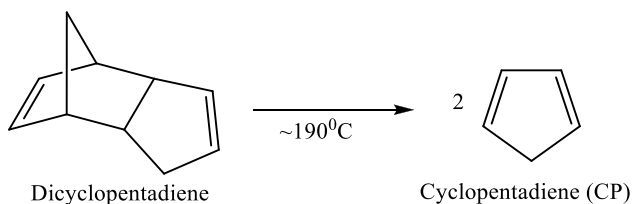
Scheme5.2: *Cracking of dicyclopentadiene to CP*

The solution was magnetically stirred at nitrogen atmosphere. After every 15 min interval of time aliquots were taken out from the reaction mixture, diluted by solvents and dipped into a cold bath made up of liquid nitrogen and methanol (~ -100^0C) to quench the reaction. Finally an internal standard (chclorobenzene) of fixed concentration was added to the solution at room temperature and rate constants (Table 5.3) were measured by monitoring the product concentration over time intervals in Gas chromatography (GC: Agilent 7890A).

Table 5.3*: Experimentally measured Second order rate constants of the Diels Alder reactions between CP and MA*

Solvents or solvents mixtures	Second order rate constants (k_2) in $M^{-1}sec^{-1}$
n-hexanes	1.22×10^{-5}
THF	1.81×10^{-5}
ACN	2.11×10^{-5}
CHCl$_3$	4.24×10^{-5}
ACN60-TFE40	4.98×10^{-5}
ACN40-TFE60	8.87×10^{-5}
ACN20-TFE80	20.00×10^{-5}
TFE	34.40×10^{-5}

5.6 Vibrational spectroscopy

For each solution, 60 µLof the sample solution (50 mM) was loaded into a demountable cell consisting of two CaF2 windows (3 mm thickness) separated by a mylar spacer of 56 µm thickness. All the IR spectra were recorded in Bruker Vertex 70.

p-anisaldehyde shows two distinct peaks in the C=O region. Nyquist mentioned that these two peaks are coming due to the Fermi resonance between the fundamental C=O stretch and overtone of the out of plane bending of aldehydic CH bending mode.[29] The Fermi resonance arise in substituted benzaldehydes. We took the simple benzaldehyde and measured FTIR spectra in acetonitrile and chloroform (Figure 5.10). The spectra show single C=O stretch in the higher frequency region of p-anisaldehyde (Figure 5.10 & Table 5.4). So we assign the higher frequency peaks as fundamental C=O IR stretching frequencies of p-anisaldehyde in different solvents.
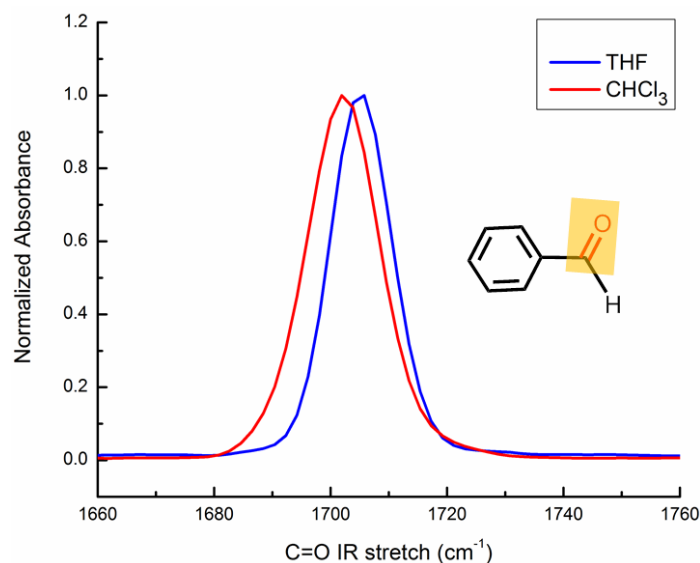
Figure 5.10 *Experimentally measured IR spectra of benzaldehyde.*

Table 5.4: *Peak positions of benzaldehyde in solvents.*

Solvents or solvent mixtures	C=O IR stretch (cm^{-1}) of benzaldehyde
THF	1706
CHCl3	1703.3

5.7 References

(1)   Davis, H. J.; Phipps, R. J. Harnessing Non-covalent Interactions to Exert Control Over Regioselectivity and Site-selectivity in Catalytic Reactions. *Chem. Sci.* **2017**, *8*, 864-877.

(2)   Toste, F. D.; Sigman, M. S.; Miller, S. J. Pursuit of Noncovalent Interactions for Strategic Site-Selective Catalysis. *Acc. Chem. Res.* **2017**, *50*, 609-615.

(3)   Huynh, P. N. H.; Walvoord, R. R.; Kozlowski, M. C. Rapid Quantification of the Activating Effects of Hydrogen-Bonding Catalysts with a Colorimetric Sensor. *J. Am. Chem. Soc.* **2012**, *134*, 15621-15623.

(4)   Grabowski, S. J. What Is the Covalency of Hydrogen Bonding? *Chem. Rev.* **2011**, *111*, 2597-2625.

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

(6)   Contreras-Garcia, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J. P.; Beratan, D. N.; Yang, W. T. NCIPLOT: A Program for Plotting Noncovalent Interaction Regions. *J. Chem. Theory Comput.* **2011**, *7*, 625-632.

(7)   Fried, S. D.; Boxer, S. G. Measuring Electric Fields and Noncovalent Interactions Using the Vibrational Stark Effect. *Acc. Chem. Res.* **2015**, *48*, 998-1006.

(8)   Fried, S. D.; Boxer, S. G. In *Annu. Rev. Biochem.* Kornberg, R. D., Ed. 2017; Vol. 86, p 387-415.

(9)   Kreuzer, H. J. Physics and Chemistry in High Electric Fields. *Surf. Interface Anal.* **2004**, *36*, 372-379.

(10)   Che, F.; Gray, J. T.; Ha, S.; Kruse, N.; Scott, S. L.; McEwen, J.-S. Elucidating the Roles of the Electric Fields in Catalysis: A Perspective. *ACS Catal.* **2018**.

(11)   Pocker, Y.; Buchholz, R. F. Electrostatic catalysis by ionic aggregates. II. Reversible Elimination of Hydrogen Chloride from Tert-butyl Chloride and the Rearrangement of 1-Phenylallyl Chloride in Lithium Perchlorate-Diethyl Ether Solutions. *J. Am. Chem. Soc.* **1970**, *92*, 4033-4038.

(12)   Meir, R.; Chen, H.; Lai, W. Z.; Shaik, S. Oriented Electric Fields Accelerate Diels-Alder Reactions and Control the endo/exo Selectivity. *Chemphyschem* **2010**, *11*, 301-310.

(13)   Aragones, A. C.; Haworth, N. L.; Darwish, N.; Ciampi, S.; Bloomfield, N. J.; Wallace, G. G.; Diez-Perez, I.; Coote, M. L. Electrostatic Catalysis of a Diels-Alder Reaction. *Nature* **2016**, *531*, 88-91.

(14)   Knowles, J. R. Enzyme catalysis: Not Different, Just Better. *Nature* **1991**, *350*, 121.

(15)   Fried, S. D.; Bagchi, S.; Boxer, S. G. Extreme Electric Fields Power Catalysis in the Active Site of Ketosteroid Isomerase. *Science* **2014**, *346*, 1510-1514.

(16)   Schneider, S. H.; Boxer, S. G. Vibrational Stark Effects of Carbonyl Probes Applied to Reinterpret IR and Raman Data for Enzyme Inhibitors in Terms of Electric Fields at the Active Site. *J. Phys. Chem. B* **2016**, *120*, 9672-9684.

(17)   Nicolaou, K. C.; Snyder, S. A.; Montagnon, T.; Vassilikogiannakis, G. The Diels–Alder Reaction in Total Synthesis. *Angew. Chem. Int. Ed.* **2002**, *41*, 1668-1698.

(18)   Fried, S. D.; Bagchi, S.; Boxer, S. G. Measuring Electrostatic Fields in Both Hydrogen-Bonding and Non-Hydrogen-Bonding Environments Using Carbonyl Vibrational Probes. *J. Am. Chem. Soc.* **2013**, *135*, 11181-11192.

(19)   Fukui, K. Role of Frontier Orbitals in Chemical Reactions *Science* **1982**, *218*, 747-754.

(20)   Andrews, S. S.; Boxer, S. G. Vibrational Stark Effects of Nitriles I. Methods and Experimental Results. *J. Phys. Chem. A* **2000**, *104*, 11853-11863.

(21)   Boxer, S. G. Stark Realities *J. Phys. Chem. B* **2009**, *113*, 2972-2983.

(22)   Yang, Z.; Doubleday, C.; Houk, K. N. QM/MM Protocol for Direct Molecular Dynamics of Chemical Reactions in Solution: The Water-Accelerated Diels–Alder Reaction. *J. Chem. Theory Comput.* **2015**, *11*, 5606-5612.

(23)   Ramanan, R.; Danovich, D.; Mandal, D.; Shaik, S. Catalysis of Methyl Transfer Reactions by Oriented External Electric Fields: Are Gold–Thiolate Linkers Innocent? *J. Am. Chem. Soc.* **2018**, *140*, 4354-4362.

(24)   Sola, M.; Lledos, A.; Duran, M.; Bertran, J.; Abboud, J. L. M. Analysis of Solvent Effects on the Menshutkin Reaction. *J. Am. Chem. Soc.* **1991**, *113*, 2873-2879.

(25)   Hammond, G. S. A Correlation of Reaction Rates *J. Am. Chem. Soc.* **1955**, *77*, 334-338.

(26)   Eyring, H. The Activated Complex in Chemical Reactions. *J. Chem. Phys.* **1935**, *3*, 107-115.

(27)   Liptay, W. Electrochromism and Solvatochromism. *Angew. Chem. Int. Ed.* **1969**, *8*, 177-188.

(28)   Huang, Y.; Rawal, V. H. Hydrogen-Bond-Promoted Hetero-Diels−Alder Reactions of Unactivated Ketones. *J. Am. Chem. Soc.* **2002**, *124*, 9662-9663.

(29)   Nyquist, R. A. Infrared Study of 4-Substituted Benzaldehydes in Dilute Solution in Various Solvents: The Carbonyl Stretching Mode. *Appl. Spectrosc.* **1992**, *46*, 306-316.

(30)   Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J., A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. B., J. C.; Iyengar, S. S. T., J.; Cossi, M.; Rega, N.; Millam, J. M.;

Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A. S., P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Wallingford, CT, 2009.

(31)   Zhao, Y.; Truhlar, D. G. Density Functionals with Broad Applicability in Chemistry. *Acc. Chem. Res.* **2008**, *41*, 157-167.

(32)   Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition elements: Two New Functionals and Systematic Testing of Four M06-class Functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215-241.

(33)   Salavati-fard, T.; Caratzoulas, S.; Doren, D. J. DFT Study of Solvent Effects in Acid-Catalyzed Diels–Alder Cycloadditions of 2,5-Dimethylfuran and Maleic Anhydride. *J. Phys. Chem. A* **2015**, *119*, 9834-9843.

(34)   Pascual-ahuir, J. L.; Silla, E.; Tuñon, I. GEPOL: An Improved Description of Molecular surfaces. III. A New Algorithm for the Computation of a Solvent-excluding Surface. *J. Comput. Chem.* **1994**, *15*, 1127-1138.

(35)   Miertuš̃, S.; Tomasi, J. Approximate Evaluations of the Electrostatic Free Energy and Internal Energy Changes in Solution Processes. *Chem. Phys.* **1982**, *65*, 239-245.

(36)   Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic Interaction of a Solute with a Continuum. A Direct Utilizaion of AB Initio Molecular Potentials for the Prevision of Solvent Effects. *Chem. Phys.* **1981**, *55*, 117-129.

(37)   Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435-447.

(38)   Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

(39)   Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.* **2012**, *8*, 61-74.

(40)   Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. . *J. Comput. Chem.* **1984**, *5*, 129-145.

(41)   Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247-260.

(42)   Sousa da Silva, A. W.; Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interfacE *BMC Research Notes* **2012**, *5*, 367.

(43)   Jorgensen, W. L.; Tirado-Rives, J. Potential Energy Functions for Atomic-level Simulations of Water and Organic and Biomolecular SystemsProc. Natl. Acad. Sci. U. S. A. **2005**, *102*, 6665-6670.

(44)   Dodda, L. S.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. 1.14*CM1A-LBCC: Localized Bond-Charge Corrected CM1A Charges for Condensed-Phase Simulations. *J. Phys. Chem. B* **2017**, *121*, 3864-3870.

(45)   Cabeza de Vaca, I.; Tirado-Rives, J.; Dodda, L. S.; Jorgensen, W. L. LigParGen Web Server: an Automatic OPLS-AA Parameter Generator for Organic Ligands. *Nucleic Acids Res.* **2017**, *45*, W331-W336.

(46)   Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(47)   Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182-7190.

(48)   Deb, P.; Haldar, T.; Kashid, S. M.; Banerjee, S.; Chakrabarty, S.; Bagchi, S. Correlating Nitrile IR Frequencies to Local Electrostatics Quantifies Noncovalent Interactions of Peptides and Proteins. *J. Phys. Chem. B* **2016**, *120*, 4034-4046.

Chapter 6

Summary and Future Scope of the Thesis

6.1 Summary

Electrostatics and dynamics are the two fundamental concepts in achieving a rational mechanistic insight about processes involving small molecules as well as complex macromolecules. In this thesis, we have looked into different problems from the perspective of electrostatics and dynamics. Steady state IR spectroscopy and time resolved two-dimensional IR spectroscopy were used to interrogate electrostatics and dynamics respectively in molecular functions. The spectroscopic studies were complemented by classical and quantum mechanical computational studies.

In chapter 3 and 4 we have addressed some important issues for non-covalent interactions which are ubiquitously present in biology and play key role in the function of biomolecules. Chapter 3 demonstrates how precisely electrostatics information can be obtained for biologically important environments using a very useful vibrational probe - nitrile - which has been sidelined to date because of its certain limitations. The reason behind investigating on the nitrile probe was its prevalence in many potent drug molecules. After resolving the electrostatics issues for nitrile in specific non-covalent interactions (hydrogen bonding), we looked into another non-covalent interaction, $n \rightarrow \pi^*$ whose origin of existence has been a subject of debate in the scientific community. $n \rightarrow \pi^*$ Interaction is largely present in biomolecules. The work of chapter 4 has revealed that studying dynamics is very crucial besides studying electrostatics.

In the chapter 5, we have looked into a problem which belongs to a different dimension. We extended the approach of the thesis in understanding catalysis of organic reactions. We have constructed a reaction model based on electrostatics. Our model successfully assesses the catalysis behaviour from a very general point of view. We have seen that like proteins/enzymes in biology, electrostatics is also harnessed in the catalytic process of organic reactions.

In the thesis, we have studied two different fields which are seemingly uncorrelated. However, if development of new molecules, which will be used as remedy for life threatening diseases, is considered, it is important to surmount the challenges existing in biology as well as in chemistry. In this regard, the work of the thesis is relevant and an indication of the importance of both electrostatics and dynamics in studying molecular functions.

6.2 Future scope of the thesis

6.2.1 Information of electrostatics in hydrogen bonded nitrile in potent inhibitors

Protein kinase plays the pivotal role in cell growth. In cancer therapy the study of protein kinase is crucial. The kinase inhibitors which are used as potent drug for cancer therapy have very limited target specificity.[1] Therefore understanding the binding mechanism of kinase inhibitors is important in the realm of new inhibitor development. Bosutinib (Figure 6.1A) is one of the kinase inhibitors.[2] Bosutinib shows its inhibitory activity towards src-kinases.[3] Boxer and coworkers have reported that bosutinib recognizes it's target by engaging hydrogen bonding between structured water in the active site of Src and the nitrile group bosutinib (Figure 6.1B).[1] In this study the effect of different amino acid residues in Src-bosutinib binding was qualitatively illustrated by using mutants of Src. However the role of electrostatic environment in the active site is still unexplored as hydrogen bonded nitrile does not follow the linear Stark effect.[4] In
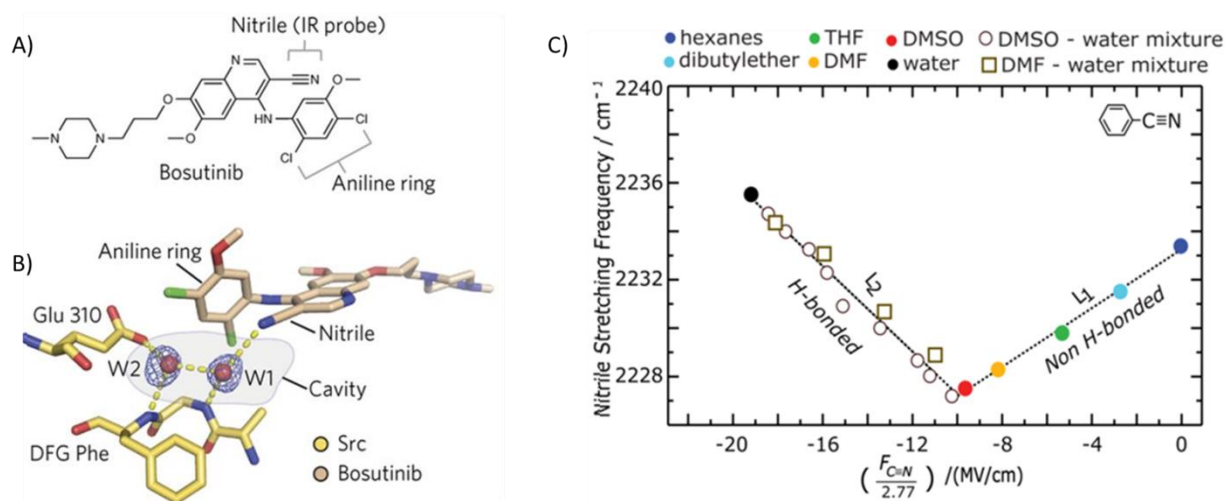

Figure 6.1: *A) structure of bosutinib.[1] B) Crystal structure of bosutinib bound to src[1]. From the structure it is evident that nitrile group of bosutinib is hydrogen bonded to a structured water in the active site. C) Field frequency calibration for model compound, benzonitrile. For more details about this figure, please see chapter 3.*

chapter 3 we have illustrated that even in presence of hydrogen bonding, the information about electrostatic can be extracted precisely by using a suitable model nitrile for bosutinib unit and constructing a field frequency calibration lines similar to Figure 6.1C. The electric field would be extremely useful to engineer inhibitors with better selectivity by providing insight about the electrostatics of the active site environments of kinases.

6.2.2 Role of polarity on the weak non-covalent interaction like $n \rightarrow \pi^*$ interaction

In chapter 4, we have shown that the intrinsic solvent fluctuation dynamics helps in the coexistence of two electronically analogous but energetically disparate interactions, $n \rightarrow \pi^*$ and hydrogen bonding, of a C=O group. Even though the energetic contribution $n \rightarrow \pi^*$ interaction is low (0.2−1 kcal/mol),[5] their abundance in biomolecules has made them an important energy contributor. In proteins the charged/polar residues are ubiquitously present. Therefore, understanding the effect of polarity on the $n \rightarrow \pi^*$ interaction is crucial.

To understand the effect of polarity on the n→π* interaction, we have performed linear IR experiments in aprotic solvents of varying dielectric constants. These solvents, not capable of forming hydrogen bonds with the C=O group, provide a direct way to assess the sensitivity of n→π* interaction toward the polarity of the environment. Both the C=O transitions shift to lower frequencies with increase in the solvent polarity (Figure 6.2 A). Variations in solvent polarity also
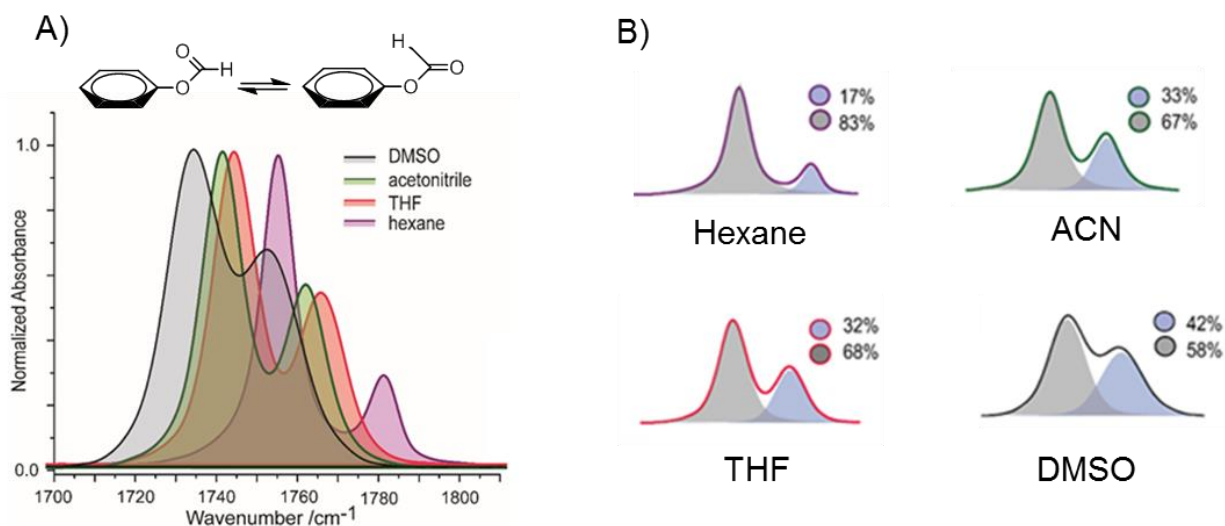

Figure 6.2: *A) Solvatochromism of phenyl formate. For any solvent, the lower frequency peak corresponds to the cis isomer where n→π* interaction is present and the higher frequency peak corresponds to trans isomer where n→π* interaction is absent. From the solvatochromism we can see that both the peaks shift to the lower frequency values with increasing solvent polarity from n-hexanes to DMSO. B) Percentage of the cis and trans population in different solvents. With increasing solvent polarity the population of trans isomer increases.*

demonstrate changes in the population distribution of the two conformers. A systematic increase in the solvent polarity shows that a gradual surge in the trans population is accompanied by a consequent decline in the cis counterpart (Figure 6.2 B). A plausible reason for such observation is that the increasing dielectric screening in polar solvents might destabilize the n→π* interaction. These results open up the possibility of studying the effect of polarity on the n→π* interaction using a combination of spectroscopic (IR, NMR, etc.) and theoretical investigations.

6.2.3. Developing a reaction model for de novo catalyst design

In chapter 5 we have shown that the solvent electric field catalyzes the Diels-Alder reaction. However, we have seen a marginal enhancement in rate constant ($\sim 10^{1.5}$) for a large amount of electric field (~0.55 V/Å). The reason of such small enhancement is not known to us at present. A detailed investigation considering the dynamical effect of solvents needs to be carried out.

So far we have only shown DA reactions with solvents. However, for routinely performed organic reactions, various catalysts are screened to obtain the best catalyst. By studying varieties of reactions with our multipronged approach we aim at developing a robust and general reaction model to predict future catalysts successfully.

6.3 References

(1) Levinson, N. M.; Boxer, S. G. A Conserved Water-mediated Hydrogen Bond Network Defines Bosutinib's Kinase Selectivity. *Nat. Chem. Biol.* **2013**, *10*, 127.

(2) Cortes, J. E.; Kantarjian, H.; Brümmendorf, T.; Khoury, H. J.; Kim, D.; Turkina, A.; Volkert, A.; Wang, J.; Arkin, S.; Gambacorti-Passerini, C. Safety and Efficacy of Bosutinib (SKI-606) in Patients (pts) With Chronic Phase (CP) Chronic Myeloid Leukemia (CML) Following Resistance or Intolerance to Iimatinib (IM). *J. Clin. Oncol.* **2010**, *28*, 6502-6502.

(3) Kennedy, L. C.; Gadi, V. Dasatinib in Breast Cancer: Src-ing for Response in All the Wrong Kinases *Ann. Transl. Med.* **2018**, *6*.

(4) Fafarman, A. T.; Sigala, P. A.; Herschlag, D.; Boxer, S. G. Decomposition of Vibrational Shifts of Nitriles into Electrostatic and Hydrogen-Bonding Effects. *J. Am. Chem. Soc.* **2010**, *132*, 12811-12813.

(5) Hinderaker, M. P.; Raines, R. T. An Electronic Effect on Protein Structure. *Protein Sci.* **2003**, *12*, 1188-1194.

The thesis ends here