Studies Of Some Biophysical Processes Using Ultrafast Spectroscopic Techniques

A thesis submitted for the award of the degree of

Doctor of Philosophy (Science)

in

Physics (Experimental)

By

Partha Pyne

Department of Physics

University of Calcutta

May 2022

Dedicated to

My Parents

&

My supervisor

Declaration

I hereby declare that the research works manifested in the thesis *"Studies Of Some Biophysical Processes Using Ultrafast Spectroscopic Techniques"* are original. All the experiments were carried out by me with the help of my lab-mates, collaborators and my supervisor, Prof. Rajib Kumar Mitra. I further declare that these results have not been used elsewhere for the award of any degree/ diploma from any university or institute.

Sign:_ _ _ _ _ _ _ _ _ _ _ _ _ _ Date: 25th May, 2022

Partha Pyne

(email: parthapyne100@gmail.com)

S N Bose National Centre for Basic Sciences

Block JD, Sector-III, Salt Lake, Kolkata-700106, India

The page is left blank intentionally

Abstract

To understand the mechanisms that mediate the activity of biological systems and/or biophysical processes at molecular and sub-cellular levels, it is necessary to recognize the associated physical processes; a potential way to study about those processes is to understand the structure and dynamics of those biological systems. Structural information enables the relative arrangement of atoms and/or molecules and links their functionality to their respective equilibrium location, which is time independent. On the other hand, dynamics traces the response of the biophysical processes at different time under the influence of external stimulation (like change in solvent polarity, pH, temperature, viscosity, stress, chemical reaction with other molecules etc.) while evolving from one configuration to another to perform its biological function along the way. Many biological processes occur on long time scale (from milli second to second or even hours) while some others evolve within nanosecond or even faster. Ultrafast spectroscopy is a very consistent tool to study such ultrafast biophysical process. This thesis contains the investigation of four different biophysical processes using different ultrafast spectroscopic techniques, namely, (i) Polyethylene glycol affects electron transfer rate in phenosafranin-DNA complex, (ii) Correlating solvation with conformational pathways of proteins in alcohol–water mixtures: a THz spectroscopic insight, (iii) Alteration of water network during fibrillation pathway and (iv) Excipients do regulate phase separation in Lysozyme and so also its hydration. We first investigate electron transfer (ET) process between drug and DNA using time resolved fluorescence spectroscopy and observe that two types of electron transfer process do coexist: one is through space ET which depends on the solvent polarity, and this type of ET gets prohibited in presence of polyethylene glycol (PEGs), used to mimic the cellular environment. Another one is guanine mediated ET which does not depend on the externally applied PEGs. Hydration (solvation) behaviour of some model proteins (BSA, βlg, Lysozyme, human Insulin) of different conformations (protein folding-unfolding, fibril and liquid-liquid phase separation (LLPS)) has been investigated in detail using THz (terahertz) spectroscopy. We find that alcohol mediated protein hydration (solvation) shows non-linear oscillatory behaviour. We then correlate the structural makeover of proteins during (un)folding and fibrillation with its associated hydration (solvation). We also observe that externally added crowders regulate the fibrillation and LLPS process of model proteins and the associated hydration do also follow the crowders mediated regulation of both LLPS and fibrillation process.

List of Publications

Inside thesis

1. Polyethylene glycols affect electron transfer rate in phenosafranin-DNA complex; P Pyne, N Samanta, A Patra, A Das, P Sen, R K Mitra; ***Spectrochimica Acta A (2020), 225, 117464***

2. Alteration of water absorption in THz region traces the onset of fibrillation in proteins; P Pyne, N Samanta, H Gohil, SS Prabhu, R K Mitra; ***Chem. Commun. (2021), 57, 998-1001***

3. Correlating solvation with conformational pathways of proteins in alcohol-water mixtures: A THz spectroscopic insight; P Pyne, DD Mahanta, H Gohil, SS Prabhu, RK Mitra; ***Phys. Chem. Chem. Phys. (2021), 23, 17536-17544***

4. Excipients do regulate phase separation in Lysozyme and thus also its hydration; P Pyne, R K Mitra; ***J. Phys. Chem. Lett. (2022), 13, 931-938***

5. Water release from protein interior dictates Insulin fibrillation; P Pyne, S Pyne, RK Mitra ***(Under Preparation)***

Outside thesis

1. Thermal stability modulation of the native and chemically-unfolded state of bovine serum albumin by amino acids; S Pal, P Pyne, N Samanta, S Ebbinghaus, RK Mitra; ***Phys. Chem. Chem. Phys. (2020), 22, 179-188***

2. Effect of Surfactant Tail Length on the Hydroxypropyl Cellulose-Mediated Premicellar Aggregation of Sodium *n*-Alkyl Sulfate Surfactants; S Mondal, P Pyne, A Patra, RK Mitra, S Ghosh; ***Langmuir (2021), 37, 6168–6177***

3. Molecular Insight into Dye–Surfactant Interaction at premicellar Concentrations: A combined Two-Photon Absorption and Molecular Dynamics Simulation Study; SKI Islam, P Pyne, DK Das, S Mukherjee, S Chakrabarty, RK Mitra; ***Langmuir (2022), 38, 3105-3112***

4. Probing the inner hydration in surfactant/cholesterol vesicles using THz spectroscopy measurements; S Pyne, P Pyne, RK Mitra ***(In Communication)***

5. A subtle interplay between hydrophilic and hydrophobic hydration governs butanol (de)mixing in water: a combined spectroscopic and simulation investigation, S Chakraborty, P Pyne, DD Mahanta, RK Mitra ***(In Communication)***

6. Addition of cholesterol alters the hydration at the surface of model lipids: a spectroscopic investigation; S Pyne, P Pyne, RK Mitra ***(In Communication)***

Acknowledgement

The thesis work will remain incomplete till acknowledging some persons who are involved directly or indirectly to support and encourage me to complete my PhD work.

First of all, I want to bow my parents. I am here only for their ceaseless support, unconditional love, affection, ideology, and inspiration. I have no word to acknowledge them.

I wish to convey my deepest gratitude to my supervisor, Prof. Rajib Kumar Mitra for his continued support, guidance, fruitful discussions, and motivation. He taught me a lot of things related to research and also outside of research, necessary in daily life. I learn from him how to deal rationally of any problem (both research problem and problems outside the research). I believe that I am really lucky to have such an opportunity to work under his supervision.

Beside my supervisor I would like to convey my gratitude to Prof. S S Prabhu (TIFR Mumbai) and Prof. Pratik Sen (IIT Kanpur) for giving me the opportunity to access their laboratories and research facilities and also for their suggestions. Prof. Prabhu also helped me a lot in fixing the Menlo THz set up (Tera-K8) in our laboratory. I acknowledge specially to Prof. Dipak Palit (CEBS Mumbai) for helping to design optical pump probe spectroscopy. Both of them (Prof. Palit and Prof. Prabhu) visited S.N. Bose centre and specially our laboratory to help in building up optical pump probe spectroscopy set up and fixing time domain THz spectrometer respectively.

I would like to thank specially two of my lab seniors: Dr. Animesh Patra and Dr. Nirnay Samanta for their constant supports. Dr. Animesh Patra (Animesh ka) encouraged me to join this group. Nirnay da taught me a lot of things starting from very basic things like how to prepare samples, how to use pipette, how to handle with data, how to plot in plotting software. Animesh ka always discussed with me about the basic things of self-assemble systems to build up my interest in such an interdisciplinary field. Being a physics background student I did not have any basic idea about this field. Without the help of Nirnay da and Animesh ka it was impossible for me to survive in this field, and I am very much grateful to both of them.

I express my appreciation to all-other lab seniors Debasish da, Dipak da, Debanjan da, Arindam da, Chaitrali di and Subhadip da for helping and suggesting me in a various way. I have discussed a lot of scientific problems with Debasish da. I learnt basic things to deal with amplified laser and different optical set ups from Dipak da. Debanjan da helped me in understating the basic physics and the set-up of THz time domain spectroscopy. Thanks goes to present lab members: Imadul, Saikat, Didhiti, Sumana, Ria, Sudip and Subhajit for some fruitful research discussion and for spending a great time in lab even outside the research work.

Special thanks to Sumana, who is not only my lab junior but also junior member of my family (my sister), as she always discusses various kinds of research problems even outside the lab. As Both of us work at same place we also spend all the time with each other. I feel that I am really lucky to have one of my family members always with me even in the lab also.

We, the lab members spent some delightful evenings with sir and madam (Dr. Sukanya Chakrabarty) at sir's place. I am thankful to them for arranging such enjoyable moments with special dinners.

Beside the research work and lab environment I would like to acknowledge some nice persons whom I met during my PhD time and spent some enjoyable times with full of happiness and a lot of fun. Out of them Nirnay da, Semanti di, Animesh ka, Sonali di, Nitish da, Ishita di are the leading names with whom I spent some joyful times. I would like to thank some of the SNBosians (people carrying research work from S.N. Bose centre are kidding by SNBosians): Rakesh da, Rajkumar da, Supriyo da, Sumanta da and present members with whom I used to discuss different types of topics excluding research. I also spent some memorable time with SNB cricket (though it is small time).

Here I am going to start taking names of my close friends: Somnath, Souvik, Gouranga, Sumanta who also use to encourage and support me whenever I needed some mental supports. We have spent good times whenever I went home from S.N. Bose centre. But I don't like to thank them as I believe there is no existence of the word "thanks" between the close friends.

I am indebted to my first teacher, my choto-masi (Mrs Chandana Goswami), with whom I started my first journey of education. I am also obligated to my father from whom I learnt the mathematics and started loving mathematical problem solving during my school and college time. I am grateful to all my teachers who taught and inspired me at different phases during my education time. I also acknowledge Sanjay Sengupta, my english teacher for his careful teaching and all-time inspiration.

I convey my gratitude to my family members (grandfather (recently passed away) and grandmother, my parents, Sumana, uncle, aunty and Ritwik (my brother)). I also thank Paulomi to be with me and supporting me in my tough time.

I am thankful to all the reviewers of different journals for their judicious evaluation and thanks to the thesis committee. Thanks to all the academic and non-academic members of S.N. Bose centres for their cooperation and also SNB canteen and SNB mess. Finally I would like to acknowledge S.N. Bose centre and DST for the financial supports and providing the opportunity to carry out my research work.

Acronyms

AFM	Atomic Force Microscopy
Arg	Arginine
BSA	Bovine Serum Albumin
CD	Circular Dichroism
CTDNA	Calf Thymus Deoxyribonucleic Acid
DIC	Differential Interference Contrast
DLS	Dynamic Light Scattering
ET	Electron Transfer
FTIR	Fourier Transform Infrared
HB	Hydrogen Bond
LLPS	Liquid-Liquid Phase Separation
Lys	Lysozyme
PBS	Phosphate Buffer Saline
PC	Principal Component
PCA	Principal Component Analysis
PEG	Polyethylene Glycol
PSF	Phenosafranine
SVD	Singular Value Decomposition
TCSPC	Time Correlated Single Photon Counting
TFE	2, 2, 2-Trifluoro Ethanol
ThT	Thioflavin-T
THz	Terahertz
Ubi	Ubiquitin
βlg	Beta-lactoglobulin

Table of Contents

Chapter 1

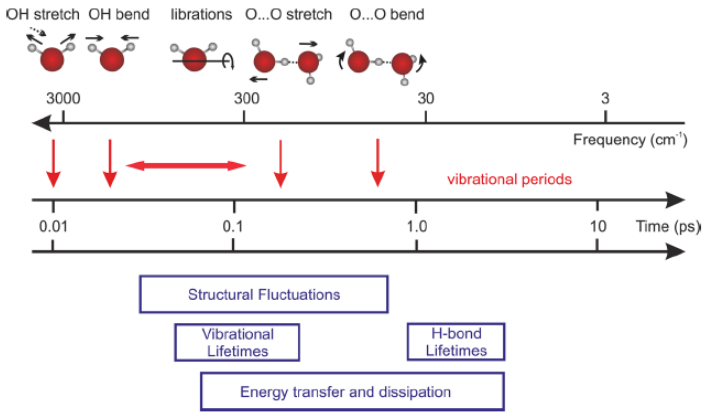
1.I. Introduction:

Introduction

Ever since human being learned to think a question haunted them: *"What is life and how does it evolve?".* In spite of an enormous effort in searching the answer, the result is still in vague. The term *"life"* could not be defined in a sentence as no one appears absolute and perfect. The latest definition of life has come from an astrobiology program organised by NASA stating, *"Life is a self-sustaining chemical system capable of Darwin evolution".*[1] However, scientists are still in search of how it (life) evolves. It is evident that a good number of physical laws and chemical reactions are associated with life, however, the specificity of the laws and the chemical processes are the wonder of the nature. The term *biological process*[2] finds its significance as it deals with the different processes (including chemical reactions) associated with the life starting from a single cellular body bacteria to a trillion-celled human body. To understand the mechanisms that mediate the activity of biological systems at molecular and sub-cellular levels, it is necessary to recognize the associated physical processes and the way to study about those processes is to understand the structure and dynamics of such biological systems.[3, 4] Structural information enables the relative arrangement of atoms and/or molecules and links their functions to their special location, which is time independent. On the other hand, dynamics traces the respond of the biophysical processes at different time under external stimulation (like polarity of the solvent, pH, temperature, stress, chemical reaction with other molecules etc.) while evolving from one configuration to another in order to perform its biological function along the way. Many biological processes occur on long time scale (from milli second to second or even hours).[5] Sensory and motor processing including speech recognition and motor coordination in human being takes place on millisecond to second time scale.[5, 6] Human's sleep-wake cycles accomplish in the time scale of hours.[5] On the other hand, some other elementary processes (like charge transfer, translation and/or rotation of a molecule) take place at noticeably shorter times (10^{-12} s or even faster).[7] These faster time scale characterizes the motion of nuclei over interatomic distances. Examples of biophysical processes occurring within such fast time scale include energy transfer between chromophores involved in the photosynthesis process,[7, 8] isomerization of rhodopsin (a primary steps towards vision)[9, 10] etc. The photochemical reaction in rhodopsin completes within 200 fs[6] with the efficiency of ~65%. In photosynthesis process antenna system (e.g. carotenoid) first absorbs

photon from (sun) light and transfers it to the reaction centre (e.g. chlorophyll) with almost ~100% efficiency and this process occurs within sub-100 femtoseconds.[11]

Water plays the most crucial role in the evolution of life as it is the major prerequisite of the living cells to be alive and functional.[12, 13] Water is present both inside and the outside the cell. About 70% of the total weight of mammalian body accounts only water;[14, 15] out of which 46% is inside the cells and the rest 24% occupies outside of the cells. It is widely involved in transporting materials and molecular machinery and facilitates most of the chemical reactions undergoing in the cells.[15] It is thus utmost important to consider the dynamics of water molecules while studying the dynamics of biomolecules involved in the biophysical processes. Water can significantly affect both the structure and dynamics of biomolecules within it.[13, 15] It mostly interacts with biomolecules through hydrogen bond (H-bond),[16] long range Coulombic forces and hydrophobic forces.[17] Both the structural and vibrational dynamics of bulk water have been well investigated by using experimental as well as theoretical studies.[18, 19] Each water molecule forms more than three but less than four H-bonds with the nearest water molecules; two via hydrogen donating OH group and two via hydrogen accepting O atom.[20] This configuration is dynamic and the breaking and making of these hydrogen bonds take place within picosecond (ps) time scale (1 ps= 10^{-12} s).[21] Intramolecular OH stretching and bending of bulk water molecules accomplish in 10 and 20 femtoseconds (1 fs=10^{-15} s), respectively.[22] Librational motion of water molecules occur within tens to hundreds of femtosecond time-scale.[22] Intermolecular hydrogen-bond (H-bond) stretching and bending vibrations have time period of 200 fs and 600-800 fs, respectively.[23, 24]



Scheme 1.1: Different mode of water molecule and their relaxation time. Red arrows represent the periods of vibrational and librational degrees of freedom as illustrated schematically at the top of the figure. The time range covered by particular processes of bulk water dynamics has been shown by the horizontal boxes below the time axis (Image Source: *Chem. Rev. 2017, **117**, 10694−10725).*

In spite of having detailed understanding about the structure and the dynamics of bulk water, there has been a lack of studies about the behaviour of water, buried within the biomolecules or the water at the surface of the biomolecules i.e. the water within its hydration shell. Recent studies reveal that the spatial arrangement of water molecules in the first (few) hydration shell(s) (i.e. the water layer just around biomolecules) are significantly different than that in the bulk water.[25-27] The polar and/ or charge moieties of the biomolecules interact with the first (few) layer(s) of water molecules via electrostatic interaction and H-bond causing a structural heterogeneity in the hydration layer.[28] This alteration causes the rotational and vibrational degrees of freedom of the bound water called as *"biological water"* to be slowed down with a concomitant increase in the H bond lifetime.[29-31] A bulk like behaviour is observed a few layers away from the surface of the biomolecules i.e. beyond a few hydration shells.

The dynamics of water molecules around biomolecules (hydration water) and/or different biophysical processes (as mentioned earlier) occur within nanosecond to picosecond time scale or even faster.[22, 25, 29] In this context ultrafast spectroscopy plays a pioneering role to determining such dynamics experimentally.[32] The term ultrafast signifies dealing with the time of sub nanosecond or even shorter than that.[32] LASER has been developed with ultrashort light pulse (with femto-second or attosecond duration) with frequency tuneability and reliability to investigate the photoinduced dynamical process of such elementary biological processes having dynamics of sub nanosecond or even faster.[33] Different experimental techniques e.g. neutron diffraction,[34] femtosecond resolved fluorescence up-conversion spectroscopy,[29, 35] NMR spectroscopy,[36] pump probe spectroscopy,[37] terahertz (THz) spectroscopy,[38, 39] have been developed with the unique features of ultrafast time scales which have been utilised to study the structure and dynamics of hydration water around biomolecules. Each of the techniques has their own unique features. Small angle X-ray and neutron scattering (SAXS and SANS) can determine the radial distribution of counterion around biomolecules and can directly probe the hydration shell structure.[40, 41] Nuclear magnetic resonance (NMR), neutron scattering, and infrared ultrafast spectroscopy can probe the dynamics of individual water molecules; however, the acquired signal is the average of all molecules taken together. NMR spectroscopy coupled with nuclear overhauser effect (NOE) have also been deployed to study the hydration around biomolecules[42, 43]which can delicately probe the long range dipolar coupling with the water molecules.[44] Displacement of individual water hydrogen atom can be monitored using quasi elastic neutron scattering (QENS) technique. Thus QENS can precisely provide the information of the water hydrogen atom in spite of the presence of hydrogen atoms in the biomolecules.[45] Femtosecond pump probe spectroscopy has also become a useful

technique to probe the reorientation dynamics of water molecules around biomolecules.[37] Collective hydrogen bond network of water molecule can be explored using THz spectroscopy. Any change in the THz spectra owning to the biomolecules directly infers about the hydration behaviour of biomolecules.[38, 39] Another advantage of THz spectroscopy is that it is a label free tool and THz radiation (1 THz corresponds to ~4 meV) is feeble to harm the biomolecules. Beside these above-mentioned tools, which directly probe the water molecules, another very widely explored tool is time resolved fluorescence spectroscopy (both time correlated single photon counting and fluorescence up-conversion spectroscopy)[29, 35] by which the dynamics of water molecules at the local environment can be investigated indirectly. In this method, either externally employed fluorophore (an aromatic molecule which emits when excited with proper wavelength of light) or the fluorophore present in the biomolecules (e.g. Trp (Tryptophan), Phe (Phenylalanine), Tyr (Tyrosine) residues present in the protein)[29] are tagged to a particular environment and subsequently the information of that local environment can be analysed.

The primary objective of this thesis is to investigate four different biophysical processes using ultrafast spectroscopic techniques. Two kinds of biomolecules: DNA (deoxyribonucleic acid) and proteins have been chosen to investigate some biophysical processes as described later. DNA is the carrier of genetic information in living body (eukaryotic cell).[14, 46] Each DNA molecule consists of two antiparallel long polynucleotide chains with four types of nucleotides namely *adenine, guanine, thymine, cytosine*. These nucleotides link together via hydrogen bonds between *guanine-cytosine* and *adenine-thymine* base pairs. The genetic information stored in the nucleic acid is used to synthesis protein which is the in-charge of many functionalities of the cell including the rebuilding of nucleic acids. Protein synthesis in the cell cytoplasm is initiated with genetic information through two major steps: "*transcription*" (through which the stored information from DNA is transferred to ribonucleic acid, RNA to carry out the information from cell nucleus to cytoplasm) and "*translation*" (through which protein synthesise is initiated in the cell cytoplasm).[14, 46] Proteins are large molecules consisting of small sub-units, called *amino acids* which are connected through covalent peptide bonds.[14, 46] There are twenty different amino acids with different chemical properties that the cell can synthesize upon receiving the information from DNA. Each type of protein molecule consists of unique sequences of these amino acids. Protein has different structures: primary (amino acids link one after another by peptide bonds), secondary (folded into α-helix, β-sheet, turns and coils by H-bond), tertiary (further folded into compact structure via H-bonds, van der Waal interaction and electrostatic interaction) and quaternary.[46, 47] Protein is functional only when it is properly folded (where the free energy of the protein is minimized[45]). Addition of physical

stress (e.g. temperature, pH, pressure etc) or chemical agents (e.g. urea, guanidinium chloride etc) disrupts its folded form causing a denaturation (unfolding) of protein.[48, 49] Again, during the energy minimization process, proteins sometimes undergo misfolded pathways and becomes aggregated, which is responsible for various types of neurodegenerative diseases.[50-52] This thesis discusses the following biophysical processes occurring in DNA and proteins: electron transfer in DNA under crowded environment; conformational modification, and their hydration (solvation) behaviour of different proteins suffering denaturation; aggregation (amyloid fibril) and liquid-liquid phase separation in proteins (chapters 3, 4, 5 and 6).

The four biophysical processes, investigated in this thesis, are (i) Polyethylene glycol affects electron transfer rate in phenosafranin-DNA complex, (ii) Correlating solvation with conformational pathways of proteins in alcohol–water mixtures: a THz spectroscopic insight, (iii) Alteration of water network during fibrillation pathway and (iv) The role of excipients to regulate phase separation in Lysozyme and so also its hydration. The structural conformation of both DNA and proteins have been studied using circular dichroism (CD) spectroscopy. Melting temperature of DNA has been investigated using temperature dependent absorption spectroscopy. Local information around the protein and DNA have been obtained from fluorescence spectroscopy measurements. Time resolved fluorescence spectroscopy (both time correlated single photon counting (TCSPC) and fluorescence up-conversion spectroscopy) shades light on the different types of electron transfer processes associated with drug-DNA interactions. Dynamic light scattering (DLS) has been deployed to estimate the size of protein oligomers. Morphology of oligomers and fibrils have been investigated using atomic force microscopy (AFM). Optical differential interference contrast (DIC) microscopy provides information on the appearance of droplets during the phase separation process of proteins. Finally terahertz spectroscopy (both frequency domain THz spectroscopy and far infrared Fourier transform spectroscopy in the THz region) have been utilised to investigate the hydration behaviour of proteins of different state (native, unfolded, aggregated and phase separated).

This thesis consists of following chapters as described briefly:

Chapter 1 deals with general introduction about the biophysical processes, ultrafast spectroscopies, and the necessity of using ultrafast spectroscopy in biophysical processes to investigate the different phenomena (basically hydration) associated with the biophysical processes.

Chapter 2 consists of some basic theories, approaches with mathematical description towards understanding the biophysical processes studied and also about the instruments which have been used in measuring or extracting different parameters in the thesis.

Chapter 3 contains the detailed investigation on the electron transfer (ET) from DNA to drug and how molecular crowders influence the ET process. For this, we chose calf thymus DNA and phenosafranine (PSF) as the drug molecule. ET takes place when PSF intercalates with DNA (PSF:DNA=1:20). We use monomeric ethylene glycol (EG) and polyethylene glycol of different molecular weights (200, 400, 1000) as the molecular crowders. Circular dichroism and temperature induced absorption spectroscopy were used to monitor the structural perturbation of DNA in presence of PEGs. Time resolved fluorescence spectroscopy (both time correlated single photon counting and fluorescence up-conversion spectroscopy) was used to study the ET process. We observed that "through space" ET solely depends on the PEGs whereas "guanine mediated" ET is independent of molecular crowders.

Chapter 4 reveals the investigation of the effect of alcohols on the structure and the hydration of two diverse proteins: Lysozyme and β-lactoglobulin. Secondary structural analysis (from CD spectroscopy) of these proteins in presence of alcohols infers that protein alcohol interactions are protein specific. The hydration study from the terahertz (THz) spectroscopy explores that protein hydration in presence of alcohols is oscillatory in nature. We finally construct phase diagram to correlate the structural makeover with the associated hydration.

Chapter 5 describes the formation of amyloid fibril and investigate the alteration of water network during the fibrillation pathway. Amyloid fibril is important in pathology as it is associated with many neurodegenerative diseases. We first prepare the fibril of a model protein, bovine serum albumin (BSA). We observe that the hydrodynamic diameter of spherically shaped oligomers does not change appreciably with incubation time. Atomic force measurement reveals that fibrils are non-spherical, elongated structures. Finally THz spectroscopy provides the idea that the water network gets altered during the fibrillation pathway. As an extension of the fibril study, we then choose a more realistic protein: human insulin (associated with diabetes II) and study its fibrillation process in presence of crowders, which delay amyloid formation and observe, using FAR IR (THz) spectroscopy, that water gets released from the protein interior depending on the crowder(s). We also investigate the physical behaviour of trapped water inside the protein and find that the water network differs from bulk water in each phase of fibrillation process.

Chapter 6 deals with the investigation of liquid-liquid phase separation (LLPS) of Lysozyme protein and the effect of excipients (L-Arg, sucrose, BSA, ubiquitin) on the stability of the LLPS process. We measure the turbidity of the protein containing solutions as turbidity ensures the onset of LLPS process. We observe that sucrose stabilizes the LLPS wile BSA prohibits the same. The effect of L-Arg is nominal and its stability strongly depends on the ubiquitin concentration. We finally measure the hydration of LLPS under different condition and also in presence of excipients using FIR FTIR spectroscopy

Chapter 7 summarizes the key findings of chapters 3, 4, 5 & 6. It also consists of some ideas for future research in this direction.

1.II References:

1. Benner, S. A., Defining Life. *Astrobiology* **2010,** *10*, 1021-1030.
2. Mossio, M.; Montévil, M.; Long, G., Prog. Biophys. Mol. Biol. *Theoretical principles for biology: Organization* **2016,** *122*, 24-35.
3. Kubelka, J., Time-resolved methods in biophysics. 9. Laser temperature-jump methods for investigating biomolecular dynamics. *Photochem. Photobiol. sci* **2009,** *8*, 499-512.
4. Cerullo, G.; Manzoni, C.; Lüera, L.; Pollia, D., Time-resolved methods in biophysics. 4. Broadband pump–probe spectroscopy system with sub-20 fs temporal resolution for the study of energy transfer processes in photosynthesis. *Photochem. Photobiol. sci* **2007,** *6*, 135-144.
5. Buonomano, D. V., The biology of time across different scales. *Nat. Chem. Biol.* **2007,** *3*, 594-597.
6. Mauk, M. D.; Buonomano, D. V., The neural basis of temporal processing. *Annu. Rev. Neurosci.* **2004,** *27*, 304–340.
7. Brixner, T.; Stenger, J.; Vaswani, H. M.; Cho, M.; Blankenship, R. E.; Fleming, G. R., Two-dimensional spectroscopy of electronic couplings in photosynthesis. *Nature* **2005,** *434*, 625–628.
8. Cerullo, G.; Polli, D.; Lanzani, G.; De Silvestri, S.; Hashimoto, H.; Cogdell, R. J., Photosynthetic Light Harvesting by Carotenoids: Detection of an Intermediate Excited State. *Science* **2002,** *298*, 2395-2398.
9. Wang, Q.; Schoenlein, R. W.; Peteanu, L. A.; Mathies, R. A.; Shank, C. V., Vibrationally coherent photochemistry in the femtosecond primary event of vision. *Science* **1994,** *266*, 422-424.
10. Kukura, P.; McCamant, D. W.; Yoon, S.; Wandschneider, D. B.; Mathies, R. A., Structural Observation of the Primary Isomerization in Vision with Femtosecond-Stimulated Raman. *Science* **2005,** *310*, 1006-1009.
11. Berera, R.; Grondelle, R. V.; Kennis, J. T. M., Ultrafast transient absorption spectroscopy: principles and application to photosynthetic systems. *Photosynth Res* **2009,** *101*, 105–118.
12. Rothschild, L. J.; Mancinelli, R. L., Life in extreme environments. *Nature* **2001,** *409*, 1092–1101.
13. Ball, P., Water as an Active Constituent in Cell Biology. *Chem. Rev.* **2008,** *108*, 74–108.
14. Cooper, G. M., *The Cell: A molecular approach*. 8 ed.; Oxford University press, U. S. A.: New York, 2001.
15. Levy, Y.; Onuchic, J. N., Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.* **2006,** *35*, 389−415.
16. Arunan, E.; Desiraju, G. R.; Klein, R. A.; Sadlej, J.; Scheiner, S.; Alkorta, I.; Clary, D. C.; Crabtree, R. H.; Dannenberg, J. J.; Hobza, P.; Kjaergaard, H. G.; Legon, A. C.; Mennucci, B.; Nesbitt, D. J., Defining the hydrogen bond: An account. *Pure Appl. Chem.* **2011,** *83*, 1619–1636.

17. Chandler, D., Interfaces and the driving force of hydrophobic assembly. *Nature* **2005,** *437*, 640–647.
18. Nibbering, E. T. J.; Elsaesser, T., Ultrafast Vibrational Dynamics of Hydrogen Bonds in the Condensed Phase. *Chem. Rev.* **2004,** *104*, 1887–1914.
19. Bakker, H. J.; Skinner, J. L., Vibrational Spectroscopy as a Probe of Structure and Dynamics in Liquid Water. *Chem. Rev.* **2010,** *110*, 1498–1517.
20. Laage, D.; Hynes, J. T., A Molecular Jump Mechanism of Water Reorientation. *Science* **2006,** *311*, 832-835.
21. Steinel, T.; Asbury, J. B.; Zheng, J.; Fayer, M. D., Watching Hydrogen Bonds Break: A Transient Absorption Study of Water. *J Phys Chem A.* **2004,** *108*, 10957–10964.
22. Laage, D.; Elsaesser, T.; Hynes, J. T., Water Dynamics in the Hydration Shells of Biomolecules. *Chem. Rev.* **2017,** *117*, 10694–10725.
23. Laage, D.; Elsaesser, T.; T., H. J., Water Dynamics in the Hydration Shells of Biomolecules. *Chem. Rev.* **2017,** *117*, 10694−10725.
24. Fecko, C. J.; Eaves, J. D.; Loparo, J. J.; Tokmakoff, A.; Geissler, P. L., Ultrafast Hydrogen-Bond Dynamics in the Infrared Spectroscopy of Water. *Science* **2003,** *301*, 1698-1702.
25. Bagchi, B., Water Dynamics in the Hydration Layer around Proteins and Micelles. *Chem. Rev.* **2005,** *105*, 3197−3219.
26. Drew, H. R.; Dickerson, R. E., Structure of a B-DNA dodecamer. III. Geometry of hydration. *J. Mol. Biol.* **1981,** *151*, 535−556.
27. Fumagalli, L.; Esfandiar, A.; Fabregas, R.; Hu, S.; Ares, P.; Janardanan, A.; Yang, Q.; Radha, B.; Taniguchi, T.; Watanabe, K.; Gomila, G.; Novoselov, K. S.; Geim, A. K., Anomalously low dielectric constant of confined water. *Science* **2018,** *360*, 1339–1342.
28. Schneider, B.; Patel, K.; Berman, H. M., Hydration of the Phosphate Group in Double-Helical DNA. *Biophys. J.* **1998,** *75*, 2422−2434.
29. Pal, S. K.; Zewail, A., Dynamics of Water in Biological Recognition. *Chem. Rev.* **2004,** *104*, 2099−2124.
30. Furse, K. E.; Corcelli, S. A., The Dynamics of Water at DNA Interfaces: Computational Studies of Hoechst 33258 Bound to DNA. *J. Am. Chem. Soc.* **2008,** *130*, 13103−13109.
31. Pal, S. K.; Peon, J.; Bagchi, B.; Zewail, A., Biological Water: Femtosecond Dynamics of Macromolecular Hydration. *J. Phys. Chem. B* **2002,** *106*, 12376–12395.
32. Holzwarth, A. R., Applications of ultrafast laser spectroscopy for the study of biological systems. *Quarterly Reviews of Biophysics* **1989,** *22*, 239-326.
33. Maiuri, M.; Garavelli, M.; Cerullo, G., Ultrafast Spectroscopy: State of the Art and Open Challenges. *J. Am. Chem. Soc.* **2020,** *142*, 3-15.
34. Savage, H.; Wlodawer, E., Determination of water-structure around bimolecules using X-ray and neutron-diffraction methods. *Methods Enzymol.* **1986,** *127*, 162–83.
35. Pal, S. K.; Peon, J.; Zewail, A. H., Biological water at the protein surface: dynamical solvation probed directly with femtosecond resolution. *Proc. Natl. Acad. Sci. USA* **2002,** *99*, 1763–68.
36. Denisov, V. P.; Jonsson, B. H.; Halle, B., Hydration of denatured and molten globule proteins. *Nat. Struct. Biol.* **1999,** *6*, 253–60.
37. Brotzakis, Z. F.; Groot, C. C. M.; Brandeburgo, W. H.; Bakker, H. J.; Bolhuis, P. G., Dynamics of Hydration Water Around Native and Misfolded α-Lactalbumin. *J. Phys. Chem. B* **2016,** *120*, 4756−4766.
38. Ebbinghaus, S.; Kim, S. J.; Heyden, M.; Yu, X.; Heugen, U.; Gruebele, M.; Leitner, D. M.; Havenith, M., An Extended Dynamical Hydration Shell Around Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104*, 20749−20752.
39. Conti Nibali, V.; Havenith, M., New Insights into the Role of Water in Biological Function: Terahertz Absorption Spectroscopy and Molecular Dynamics Simulations Studies of the Solvation Dynamics of Biomolecules. *J. Am. Chem. Soc.* **2014,** *136*, 12800−12807.
40. Pabit, S. A.; Meisburger, S. P.; Li, L.; Blose, J. M.; Jones, C. D.; Pollack, L., Counting Ions Around DNA with Anomalous Small-Angle X-Ray Scattering. *J. Am. Chem. Soc.* **2010,** *132*, 16334−16336.

41. Svergun, D. I. S.; Richard, S.; Koch, M. H. J.; Sayers, Z.; Kuprin, S.; Zaccai, G., Protein Hydration in Solution: Experimental Observation by X-Ray and Neutron Scattering. . *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95*, 2267−2272.
42. Otting, G.; Liepinsh, E.; Wüthrich, K., Protein Hydration in Aqueous Solution. Science. **1991,** *254*, 974−980.
43. Liepinsh, E.; Otting, G.; Wüthrich, K., NMR Observation of Individual Molecules of Hydration Water Bound to DNA Duplexes: Direct Evidence for a Spine of Hydration Water Present in Aqueous Solution. *Nucleic Acids Res.* **1992,** *20*, 6549−6553.
44. Halle, B., Cross-Relaxation Between Macromolecular and Solvent Spins: The Role of Long-Range Dipole Couplings. *J. Chem. Phys.* **2003,** *119*, 12372−12385.
45. Bellissent-Funel, M. C. Z., J. M.; Chen, S. H. , Slow Dynamics of Water Molecules on the Surface of a Globular Protein. *Faraday Discuss.* **1996,** *103*, 281−294.
46. Alberts, B.; Johnson, A.; Lewis, J.; Morgan, D.; Raff, M.; Roberts, K.; Walter, P., *Molecular Biology of teh Cell*. Garland Science, Taylor & Francis Group: New York, 2015.
47. Kessel, A.; Ben-Tal, N., *Introduction to Proteins: Structure, Function and Motion*. CRC Press, Taylor & Francis Group: NewYork, 2018.
48. de Oliveira, G. A. P.; Silva, J. L., A hypothesis to reconcile the physical and chemical unfolding of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112*, E2775-E2784.
49. Shirley, B. A., *Protein Stability and Folding: Theory and Practice*. Humana Press: 1995.
50. Ferreira, S. T.; Vieira, M. N. N.; De Felice, F. G., Soluble protein oligomers as emerging toxins in alzheimer's and other amyloid diseases. *IUBMB Life* **2007,** *59*, 332-345.
51. Stefani, M.; Dobson, C. M., Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J. Mol. Med.* **2003,** *81*, 678-699.
52. Dobson, C. M., Protein misfolding, evolution and disease. *Trends Biochem. Sci* **1999,** *24*, 329-332.

Chapter 2

2. Theories, Material & methods, and Instruments

2.I. Basic theories:

2.I.A Physics behind the liquid-liquid phase separation (LLPS) of proteins.

Most of organelles, present inside the cell to perform the physicochemical activity smoothly are fenced by semi-permeable lipid membrane to isolate from cell cytoplasm.[1] However, there are some other organelles having no membrane[2-4] (membranelles organelles) which maintain their structure by the liquid-liquid phase separation (LLPS) process.[5-7] Among the biomolecules proteins mostly undergo to the LLPS process. To understand LLPS process occurring inside the cellular bodies it is important to consider the free energy associated with it.[8] For that we can approach in the following way.[9, 10]

A protein can be considered as a polymer chain having x number of identical spherical amino acids. Let us consider again that N_p number of such protein molecules are dissolved in water having N_s number of water molecules in the solution. So the total number of molecules present in the volume is $N = N_w + xN_p$. Now we need to find the number of ways by which all the N molecules can be configured in all different possibilities.

1st amino acid of 1st protein: can be placed in any of the N_w+xN_p ways.

2nd amino acid of 1st protein: has the z_c number of alternatives (z_c is the coordination number of protein molecule).

3rd amino acid of 1st protein has the ($z_c - 1$) number of alternatives.

……

1st amino acid of 2nd protein has the alternatives of $N_w + xN_p - x$

Thus the total number of different configurations for all the molecules by which they can be placed is

$$\Omega = \frac{\left(\frac{N_w}{x}+N_p\right)!\left(N_w+xN_p\right)^{N_w\frac{x-1}{x}}(xz_c)^{N_p}(z_C-1)^{(x-2)N_p}}{\left(\frac{N_w}{x}\right)!N_p!N_w^{\frac{x-1}{x}N_w}e^{(x-1)N_p}\sigma^{N_p}} \qquad (2.I.A1)$$

Where σ is the symmetry factor.

So the entropy associated with the system is $S = k_B \ln \Omega$; where k_B is the Boltzmann constant.

Putting the value of Ω in the equation of S and simplifying we get

$$S = k_B[(N_w + N_p)\ln(N_w + xN_p) - N_w\, lnN_w - N_p\, lnN_p + N_p \ln\left(\frac{z_c}{\sigma}\right) + x(x - 2)\, N_p \ln(z_c - 1) - (x-1)N_p] \quad \text{(2.I.A2)}$$

Now the entropy of pure water is $S_w = 0$ (by putting N_p=0 in the equation 2.I.A2)

Entropy of pure protein, $S_p = k_B[N_p \ln(\frac{z_c x}{\sigma}) + (x-2)\, N_p \ln(z_c - 1) - (x-1)N_p]$ (by putting N_w=0 in the equation 2.I.A2)

Hence, the mixing of entropy, $\Delta S_{mix} = S - S_w - S_p$

$$\Rightarrow \Delta S_{mix} = k_B\left[N_w \ln\left(\frac{N_w + xN_p}{N_w}\right) + N_p \ln\left(\frac{N_w + xN_p}{xN_p}\right)\right] = -k_B[N_s \ln \phi_s + N_p \ln \phi_p] \quad \text{(2.I.A3)}$$

Where, ϕ_w $(= \frac{N_w}{N})$ and $\phi_p (= \frac{N_w}{N})$ are the volume fraction of water and protein respectively.

Addition to entropic effect, the change in enthalpy also contributes to determining the free energy. In this context three different type of molecular interaction can be occurred: amino acid-amino acid interaction, u_{pp}; water-water interaction, u_{ww}; and amino acid-water interaction, u_{pw}. So, the energy increase upon interaction of each amino acid with water is

$$\Delta u = u_{pw} - \frac{1}{2}(u_{pp} + u_{ww}) \quad \text{(2.I.A4)}$$

As the total number of amino acids in the solution is xN_p and z_c is the coordination number of proteins, the total number of nearest neighbour site to all amino acid is $xN_p z_c$. Now multiplying with ϕ_w that any such site is occupied by water molecule we can obtain the total number of protein-water molecular interaction.

Thus, the enthalpy change is equal to the energy change per unit amino acid-water interaction multiplied by the number of interactions i.e.

$$\Delta H = N_w \phi_p z_c \Delta u = xN_p \phi_w z_c \Delta u \text{ (as, } N_w \phi_p = xN_p \phi_w) \quad \text{(2.I.A5)}$$

Here Flory exponent can be defined as $\chi = \frac{z_c \Delta u}{k_B T}$.

Thus ΔH can be re-written as $\Delta H = k_B T N_w \phi_p \chi$ (2.I.A6)

Combining equation (2.I.A3) and (2.I.A6) we get,

$$\Delta G = k_B T\left[N_w \ln \phi_w + N_p \ln \phi_p\right] + k_B T N_w \phi_p \chi$$

So, the free energy per lattice site is,

$$F = \frac{\Delta G}{N} = k_B T[\phi_w \ln \phi_w + (\frac{\phi_p}{x}) \ln \phi_p + \phi_w \phi_p \chi\ ] \quad (2.I.A7)$$

Here the magnitude of χ quantifies the energetic cost. The higher the value the larger is the cost. Above a critical value of χ the energetic term will overcome the entropy such that the free energy of mixing has a negative curvature. As a result an instability of the mixing of protein in water is occurred which is conquered by the phase separation process; one phase becomes protein deficit, and another phase enriches with protein.

2.I.B Different relaxation mode of molecules on illuminating with THz radiation:

As the nucleus of an atom is massive compared to electron, the molecule can be considered as a spring connected with a heavy mass and the system can be considered as a harmonic oscillator. The natural frequency of oscillation can be derived from the Hook's law and Newton's 2nd law. We can add a damping force, related to the friction of molecules having proportional to the velocity of the molecule. Now when THz radiation is applied on the sample a driving force (originated from electric field of the THz radiation) will come to act. So the one-dimensional equation of motion of molecule with driving force is[11]

$$\frac{d^2x}{dt^2} + 2\gamma \frac{dx}{dt} + \omega_0^2 x = f cos\omega t \quad (2.I.B1)$$

$$\text{With } \gamma = \frac{b}{2m}; \omega_0 = \sqrt{\frac{k}{m}}; f = -\frac{eE}{m}$$

The solution of equation 2.I.B1 yields:

$$x(t) = \frac{f cos(\omega t - \psi)}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\gamma^2\omega^2}} \text{ with } \psi = \tan^{-1}(\frac{2\gamma\omega}{\omega_0^2 - \omega^2}) \quad (2.I.B2)$$

Now the absorption of radiation by the ions per unit time is $P = -N < \vec{F} \cdot \vec{v} >$

where $\vec{F} = -e\vec{E}\cos(\omega t)$ and $\vec{v} = \frac{d\vec{x}}{dt}$

Putting the value of $x(t)$ (as in equation 2.I.B2) and calculating the power we get,

$$P(\omega) = \frac{N(eE)^2\omega^2\gamma}{m\left[(\omega_0^2 - \omega^2)^2 + 4\gamma^2\omega^2\right]} \quad (2.I.B3)$$

Again the absorption coefficient, $\alpha(\omega)$ is related with the dissipation as

$$P(\omega) = \alpha(\omega)\frac{c}{n} \cdot \frac{1}{2}\,\epsilon E^2;$$

Where "n" is the refractive index of the medium, "c" is the velocity of the light; "ϵ" is the dielectric constant of the medium.

Putting the power as the dissipation and simplifying we get,

$$\alpha(\omega) = A\frac{\omega^2\gamma}{\left[\left(\omega_0^2-\omega^2\right)^2+4\gamma^2\omega^2\right]} \qquad \text{(2.I.B4)}$$

$$\text{Where } A = 2\frac{n}{c}\frac{Ne^2}{m\epsilon}$$

Now if the molecule undergoes more than one underdamping motion then the absorption coefficient can be written as the sum of the damping motion as[12-14]

$$\alpha(\omega) = \sum_i A_i\frac{\omega^2\gamma_i}{\left[\left(\omega_{0,i}^2-\omega^2\right)^2+4\gamma_i^2\omega^2\right]} \qquad \text{(2.I.B5)}$$

2.I.C Collective dipole moment estimation from Isotropic absorption coefficient as measured from FAR IR (THz) spectroscopy: a quantum mechanical approach:

Terahertz (Far IR) absorption spectroscopy has been emerged as powerful tool to explore the hydration of solutes and the dynamical coupling of the collective mode of water network as it directly probes the reorientation of permanent and induced dipole moment.[15, 16] THz radiation (external electric field) used to probe the sample couples itself with the sample. Here we use a two-state model to estimate how the absorption coefficient of molecule depends on the dipole moment of the molecules using quantum mechanics.[17]

Let H_0 be the Hamiltonian of the unperturbed time independent system having in the initial state ψ_i (with energy E_i). Now the molecules can transit from ψ_i to state ψ_f (with energyE_f) upon interacting weekly with an external electric field of frequency ω ($\approx \frac{E_f-E_i}{\hbar}$). For a monochromatic light, the electric filed can be described as $\vec{E}(t) = \epsilon E_0 cos\omega t$. Now if the wavelength of the light is much larger than molecular dimension, the interaction between molecule and the light can be written as $\hat{V}(t) = -\hat{\mu}\cdot\vec{E}(t)$ (2.I.C1)

Where $\hat{\mu}$ is the dipole moment operator (collective dipole moment of the system).

So, the total Hamiltonian of the system can be written as $H = H_0 + \hat{V}(t)$.

According to the time dependent perturbation theory the transition probability from ψ_i to ψ_f is

$$P_{if}(\omega) = \frac{1}{\hbar^2}\left|\int_0^\infty < \psi_f|\, \hat{V}(t)|\psi_f > e^{i\omega_{fi}t} dt\right|^2 \qquad (2.I.C2)$$

Now as IR light excites the vibrational states and vibrational states are almost continuum so we can apply the Fermi golden rule:

$$P_{if}(\omega) = \frac{\pi E_0^2}{2\hbar^2}\left|< \psi_f|\, \epsilon\hat{\mu}\, |\psi_f >\right|^2 \left[\delta(\omega_{fi} + \omega) + \delta(\omega_{fi} - \omega)\right] \qquad (2.I.C3)$$

AS δ term corresponds to the energy conservation, so to avoid the loss of energy, $\delta(\omega_{fi} + \omega)$ will be invalid and the equation (2.I.C3) leads to

$$P_{if}(\omega) = \frac{\pi E_0^2}{2\hbar^2}\left|< \psi_f|\, \epsilon\hat{\mu}\, |\psi_f >\right|^2 \left[\delta(\omega_{fi} - \omega)\right] \qquad (2.I.C4)$$

The transition rate for all the molecules is given by the average of $P_{if}(\omega)$ over the initial state. Thus for canonical ensemble the average transition rate is

$$P(\omega) = \frac{\pi E_0^2}{2\hbar^2}\sum_i \rho_i \left|< \psi_f|\, \epsilon\hat{\mu}\, |\psi_f >\right|^2 \left[\delta(\omega_{fi} - \omega)\right] \qquad (2.I.C5)$$

Where ρ_i is the density matrix eigenvalue $\left(\rho_i = \frac{e^{-\beta E_i}}{Tr(e^{-\beta H_0})}\right)$

Now the Schrodinger picture can be converted to interaction picture as

$$P(\omega) = \frac{E_0^2}{4\hbar^2}\int dt \sum_i \rho_i < \psi_i|\, \epsilon\hat{\mu}\, |\psi_f >< \psi_f|\, \epsilon\hat{\mu}\, |\psi_i > e^{i(\omega_{fi} - \omega)t}$$

$$\Rightarrow P(\omega) = \frac{E_0^2}{4\hbar^2}\int dt e^{-i\omega t} \sum_i \rho_i < \psi_i|\, \epsilon\hat{\mu}\, |\psi_f >< \psi_f|\, \epsilon\hat{\mu}\,(t)|\psi_i > \qquad (2.I.C6)$$

Now as ψ_f is the eigenstate of unperturbed Hamiltonian, they form a complete basis of Hilbert space, so $|\psi_f >< \psi_f| = \mathbb{I}$ ($\mathbb{I}$ is the identity operator) holds good and the equation (2.I.C6) reduces to,

$$P(\omega) = \frac{E_0^2}{4\hbar^2}\int dt e^{-i\omega t} \sum_i \rho_i < \psi_i|\, \epsilon\hat{\mu}(0)\, \epsilon\hat{\mu}\,(t)|\psi_i > \qquad (2.I.C7)$$

Finally putting the value of ρ_i and summing over we get,

$$P(\omega) = \frac{1}{3}\frac{E_0^2}{4\hbar^2}\int dt e^{-i\omega t} < \hat{\mu}(0)\hat{\mu}(t) > \qquad (2.I.C8)$$

"1/3" term appears due to the consideration of the isotropic system for which the response is independent of the direction of the radiation.

Now another case may appear for $\omega < 0$; in that case $\delta(\omega_{fi} - \omega)$will be invalid. Physically this case will imply the emission process and the transition rate for such process can be recalculated by replacing ψ_i to ψ_f and interchanging i to f in equation (2.I.C5) as follows:

$$P(-\omega) = \frac{\pi E_0^2}{2\hbar^2} e^{-\beta\hbar\omega} \sum_{i,f} \rho_i \left| < \psi_i | \epsilon\hat{\mu} | \psi_f > \right|^2 \left[\delta\left(\omega_{fi} - \omega\right)\right] = e^{-\beta\hbar\omega} P(\omega) \qquad (2.I.C9)$$

$$\left[\text{As } \rho_f = \frac{e^{-\beta E_f}}{Tr(e^{-\beta H_0})} = \frac{e^{-\beta(E_i+\hbar\omega)}}{Tr(e^{-\beta H_0})} = \rho_i e^{-\beta\hbar\omega}\right]$$

So, net energy absorbed per unit time, $\Pi(\omega) = \hbar\omega[P(\omega) - P(-\omega)]$

$$\Rightarrow \Pi(\omega) = \frac{E_0^2}{12\hbar}\int dt e^{-i\omega t} < \hat{\mu}(0)\hat{\mu}(t) > (1 - e^{-\beta\hbar\omega}) \qquad (2.I.C10)$$

Now the absorption coefficient can be estimated as $\alpha(\omega) = \frac{\Pi(\omega)}{V|S|}$

Putting the value of Poynting vector, |S| and assuming $(1 - e^{-\beta\hbar\omega}) \approx \beta\hbar\omega$ we finally get,

$$\alpha(\omega) = \frac{1}{4\pi\varepsilon_o}\frac{2\beta\omega^2\pi}{3Vcn(\omega)}\int dt e^{-i\omega t} < \hat{\mu}(0)\hat{\mu}(t) > \qquad (2.I.C11)$$

Thus the equation (2.I.C11) enables a direct correlation between absorption coefficient (as directly measured from THz spectroscopy) and the total (collective) dipole moment of sample.

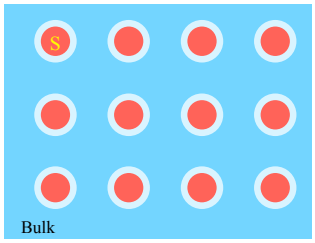
2.I.D Why THz spectroscopy emerges as a potential tool to probe the solute hydration?

Hydration water (water in front of the molecule) around biomolecules plays a pivotal role in maintaining the structure, stability, and the dynamics of the biomolecules.[18-20] Hydration water around solute (biomolecules) are quite distinct in nature compared to the bulk water. This hydration water can directly be probed using THz spectroscopy.[21, 22] As discussed in section 2.I.C, THz spectroscopy can estimate the total absorption coefficient of the system which solely depends on the dipole moment fluctuation. Thus, it is a direct measure of the collective dynamics of water molecules on spatial scale. That is why THz absorption can probe even a small perturbation of hydrogen bond network owning to the presence of solutes enabling to

investigate the dynamics of hydration water. For a homogeneous solution, the absorption coefficient, α(ν) as estimated from THz spectroscopy can be described using a three-component model consisting of solute, hydration water, and bulk water (scheme 2.I):[21, 22]

$$\alpha(\nu) = \alpha_s(\nu)\frac{V_s}{V_T} + \alpha_{sh}(\nu)\frac{V_{sh}}{V_T} + \alpha_b(\nu)\frac{V_b}{V_T} \qquad (2.I.D1)$$

Where $\alpha_s(\nu)$, $\alpha_{sh}(\nu)$, $\alpha_b(\nu)$ are the absorption coefficient of bulk water, hydration shell and solute (biomolecules) respectively and V_T, V_s, V_{sh}, V_b are respectively the total volume, volume occupied by the solute, volume occupied by the dynamical hydration shell and volume of bulk water. Water in the dynamical hydration shell offers a THz absorption coefficient different from bulk water. THz spectroscopy thus manifests a direct investigation of the dynamical hydration shell.



Scheme 2.I: Schematic diagram with solute molecules (red) along with hydration shell of the solute (white).

2.I.E Principal component Analysis:

Principal component analysis (PCA) is widely used to reduce the component of the dataset. In this method dataset is first decomposed. Singular value decomposition (SVD) method, as described earlier[13, 23] is one of the most popular techniques in decomposing the dataset. A data matrix, A is constructed for each measurement as row vectors and the variables as the column vector. Singular value decomposition of the matrix yields three matrices, namely score matrix, U, singular value matrix S and loading matrix V which are related as

$$A = USV^T \qquad (2.I.E1)$$

Where V^T is the transpose matrix of V.

Principal components (PC_i) can be achieved from the following formulae:

$$PC_i = s_i v_i \qquad (2.I.E2)$$

Where s_i (known as the singular values) are the diagonal elements of S matrix and v_i are the row vector of V. The entries of score matrix U quantifies the contribution of the variables of corresponding PC. Total number of principal components is equal to the number of variables. PCA analysis was done in Mathematica 9 without any pre-processing of the data. The fist principal component signifies the averaged spectrum, and second component indicates the change over the averaged spectrum.

2.II. Experimental techniques

2.II.A.1. DNA structure analysis:

Absorption spectroscopy is a pivotal tool to verify the purity of DNA. If the ratio of A_{260}/A_{280} occupies in between 1.8 to 2.0 then there is no protein content in the DNA and satisfactorily good for the experiment.[24] The structure of native DNA (B- DNA) can also be characterised by CD spectroscopy as DNA shows positive long wavelength band(s) at about 260-280 nm and a negative band around 245 nm.[25] However, the position and amplitude of this band varies with sequence and conformational properties of DNA.

2.II.A.2. Melting temperature of DNA:

Unfolding of DNA and measurement of melting temperature of DNA can be studied using temperature dependent absorption spectroscopy. Melting temperature of DNA, (T_m) is defined as the temperature at which double stranded DNA starts to be single stranded. The change in the temperature dependent absorption curve, monitored at 260 nm is fitted using following sigmoidal equation:

$$A_{260}(T) = \frac{a}{1+e^{-k(T-T_m)}} \qquad (2.II.1)$$

Where a and k is the parameters and T_m is the melting temperature.

2.II.B.1. Protein secondary structure analysis:

Secondary structure of protein can be analysed by using CD spectroscopy in Far UV region (190-250 nm). Secondary structural parameters of protein e.g. α-helix, β-sheet (parallel, anti-parallel, strand) and random coil can be analysed by deconvoluting far UV CD spectrum in different softwares available[26] (e.g. CDNN, dichroweb, K2D2, SELCON, CONTIN, BESTSEL etc). The CD spectra of α-helix shows negative bands at 222 nm (due to $n \rightarrow$

π^* transition) and 208 nm (due to exciton splitting of $\pi \rightarrow \pi^*$ transition)[26] and a positive band at 190 nm; β-sheet illustrates negative band at 218 nm and positive band at 196 nm and random coil displays two bands at 212 nm (positive band) and 195 nm (negative band).[26] The spectrum appears at near UV region due to the presence of aromatic residues (e.g. phenylalanine, tyrosine, tryptophan) and disulphide bonds.

2.II.B.2. Phase diagram analysis:

Protein during folding-unfolding pathways (in presence of external agents e.g. cosolvents, osmolytes, molecular crowders etc or environmental stress e.g. temperature, pressure, pH etc) drives through different intermediate pathways. Such intermediates can be found from phase diagram plot.[27, 28] Extensive parameters (parameters whose value is causally related to the amount of analysed matter of a system) (e.g. CD spectroscopy, fluorescence spectroscopy) are only considered for such analysis. If the analysed matters obtained from same spectroscopic method or from two different spectroscopic techniques are $I(\lambda_1)$, $I(\lambda_2)$, they can be related as $I(\lambda_1) = f[I(\lambda_2)]$. The linear dependency of this $I(\lambda_1) = f[I(\lambda_2)]$ indicates all or none transition between two conformers of the protein. On the contrary the nonlinear dependency reveals about the sequential transition between two states. The number of segments (both linear and nonlinear) present in the phase diagram from native to final state of proteins reflects at least that number of intermediates involved in the protein folding, un(mis)folding pathways.

2.II.C. Fluorescence measurements of proteins:

Fluorescence spectroscopy is a well-known method to investigate the immediate local environment. As, DNA does not contain any fluorophore, external fluorophores must be attached to get the information of the DNA interior. On the other hand, some of the aromatic amino acids present in the protein (Tryptophan (Trp), Phenyl alanine (Phe), Tyrosine (Tyr)) can fluoresce. So, the information of local environment of protein can be obtained from these intrinsic fluorophores[29]. Out of these amino acid residues, Trp absorbs at longest wavelength and has largest molar extinction coefficient ($\varepsilon_{280} = 5579\ M^{-1}cm^{-1}$). So, we excite protein at 295 nm to avoid the excitation of Tyr and Phe. Another advantage of choosing Trp is that Trp is extremely sensitive to local environment and presence of less amount in protein enables precise interpretation. Beyond this, when protein aggregates a new fluorescence, known as "*blue fluorescence"* [30-32] appears with emission maxima at ~ 445 nm region with the excitation at 375 nm, which is completely inattentive in the case of low concentrated folded protein.

2.II.C.I Steady state fluorescence anisotropy measurement:

Anisotropy is the measure of rigidity of a molecule in various solvent environment. It is based on the principle of the photo-selective excitation of fluorophores by polarized light. Most of the fluorophores absorbs only that component of light which is parallel to the transition moment of them. For anisotropy measurements, emission polarisation is set to be parallel or perpendicular with respect to the excitation and obtained anisotropy is expressed as

$$r = \frac{(I_{\parallel} - I_{\perp})}{(I_{\parallel} + 2I_{\perp})} \qquad (2.II.2)$$

where $I_{\parallel}$ and $I_{\perp}$ are intensities measured with a linear emission polarizer placing parallel and perpendicular with respect to the excitation polarisation.

2.II.C.II Time resolved fluorescence measurements:

Time resolved fluorescence spectroscopy (TRFS) (both time correlated single photon counting, TCSPC and fluorescence up-conversion spectroscopy) enables to measure the time taken by a fluorophore to relax from excited state through a radiative pathway i.e. lifetime of the fluorophore. TRFS has advantages over steady state measurement as it is independent of the concentration of fluorophores used. Fluorophores are excited by laser with noticeably short pulse width and the repetition rate is set to be at least four time larger than the lifetime of the fluorophores. Fluorophores relax from excited state exponentially. The decay transients are globally fitted after convoluting with laser pulse. The decay can be represented mathematically as

$$I(t) = \sum_i a_i e^{-\frac{t}{\tau_i}} \text{ with } \sum_i a_i = 1 \qquad (2.II.3)$$

and the average lifetime of the fluorophore can be expressed as $\tau = \sum_i a_i \tau_i$; where τ_i is the lifetime of ith species of fluorophores with contribution a_i.

2.II.D. THz measurement:

Terahertz (THz) (1THz =10^{12} Hz) frequency region, a bridge between FAR IR and microwave has recently been grown up as a potential label free tool to probe the intermolecular collective dynamics of water network[21, 33]. Absorption coefficient, α (ν) is manifested with the measure of hydration dynamics as it is directly correlated with permanent and induced dipole of the molecule[34] (also see section 2.I.C for detailed derivation). We use both THz frequency domain spectroscopy and Fourier transformed infrared spectroscopy (in FAR IR range) for THz measurement. From the FIR FTIR measurement we get the idea of intermolecular H bond stretching and librational motion (hindered rotation) of water molecule.[35]

2.II.D.a Data analysis as measured from THz frequency domain spectroscopy:

THz data as obtained from frequency domain THz spectroscopy is analysed to extract the frequency dependent absorption coefficient, $\alpha(\nu)$ as follows[36]:

$$\alpha(\nu) = \frac{1}{d}\ln\left(\frac{I_{ref}(\nu)}{I_{sample}(\nu)}\right) \quad (2.II.4)$$

Where $I_{ref}(\nu)$ and $I_{sample}(\nu)$ are the intensity of light of reference and sample, respectively. "d" is the thickness of the sample. We used air medium as the reference. A heavy dry nitrogen is purged in the sample chamber during the experiment to remove the water vapour present in the air. Each THz spectrum presented here is the average of four individual measurements. All the data are validated with repeated experiments. The Typical THz power used in the experiment varies from about 65 microwatts to few microwatts at the spectral range end (~1.2 THz).

2.II.D.b Data analysis as measured from Fourier transformed infrared spectroscopy:

All the spectra were collected in attenuated total reflection (ATR) mode and have been converted to the absorbance using the following relation:

$$Abs(\nu) = \frac{intensity\ in\ ATR\ unit(\nu)\times 1000}{Wave\ number} \quad (2.II.5)$$

The absorption coefficient, α was then estimated as

$$\alpha(\nu) = \frac{Abs_{sample}(\nu) - Abs_{air}(\nu)}{d_p} \quad (2.II.6)$$

Where d_p is penetration depth[37], defined as $d_p = \frac{\lambda}{2\pi\sqrt{(n_d Sin\theta)^2 - n_s^2}}$ (2.II.7)

where λ in the wavelength of the light; n_d and n_s are the refractive indices of the diamond crystal and the sample respectively. θ is the incident angle (Here $\theta = \frac{\pi}{4}$).

$\Delta\alpha$ is fitted using a sum of damped harmonic oscillator functions as derived in equation (2.I.B5):

$$\Delta\alpha(\nu) = \sum_{i=1}^{3} \frac{a_i \omega_i \nu^2}{\nu^2\omega_i^2 + \pi^2\left(\nu_i^2 + \frac{\omega_i^2}{4\pi^2} - \nu^2\right)^2} \quad (2.II.8)$$

Where a_i, ω_i, ν_i are the amplitude, width and the centre frequency of i[th] resonance. The unperturbed centre frequency can be obtained as $\nu_{0,i} = \sqrt{\nu_i^2 + \frac{\omega_i^2}{4\pi^2}}$. (2.II.9)

2.II.E. Image analysis as obtained from atomic force microscope and optical microscope:

Freshly cleaved mica was used to drop-cast the samples to observe the morphology of samples using atomic force microscopy. The drop-casted mica plate was kept for 10-15 mins to adsorb the sample and then rinsed with milli-Q water. The mica plates were then dehydrated using dry nitrogen before measurements. Collected images of oligomers and matured fibrils are analysed in *Image J* and *Gwyddion* software to extract the diameter of the particles (protein oligomers during aggregation). Image of liquid droplets (phase separation process) are recorded from optical differential interference contrast (DIC) microscope and the images are analysed using Mathematica 9 (custom written code) to obtain their sizes.

2.II.F. Glass slide preparation for optical microscope:

To prepare silanized glass slides,[38] slides and cover slips were first immersed in a 1:1 solution of methanol and HCl (35%) for 45 min, then thoroughly washed with distilled water and dried in warm air. After that, the slides were again immersed in sulphuric acid (98%) for 1 hr and washed thoroughly with distilled water and dried in warm air. The silanized slides were then coated with Pluronic F127 solution for surface treatment of the glass slides.[38] Pluronic F127 solution was prepared by dissolving 10 wt% Pluronic in 10 mM PBS buffer (pH 7.4) and the glass slides were immerged in the solution for 1 hr at room temperature, then washed with water and dried in hot air.

2.III. Chemicals used:

2.III.A DNA:

Deoxyribonucleic acid (DNA) sodium salt from calf thymus (CT) is a double stranded long thin and locally cylindrical polyelectrolyte chain with molar extinction coefficient of 13600 M^{-1} cm^{-1} at 258 nm.[39] Each base pair of DNA consists of two elementary negative charges. The estimated number of base pairs in CT DNA is 640.[40]

2.III.B proteins:

a. *Bovine Serum Albumin* (BSA) is a globular helical protein (α-helix ~64% in native form[41]), extracted from blood plasma of cow. It contains 583 number of amino acids and 17 disulphide bridges with average molecular weight of ~66.4 kDa. BSA has two tryptophan moieties (Trp

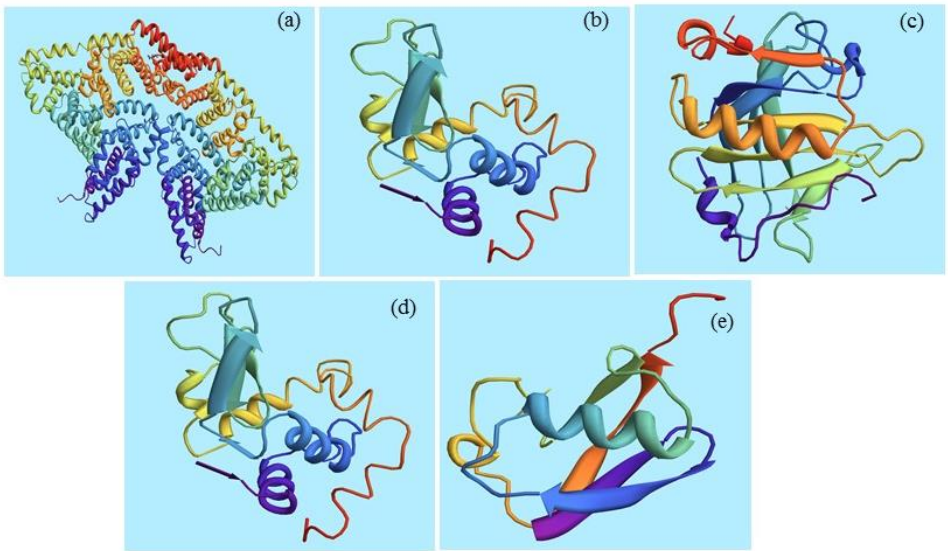
134 and Trp 213) which emits fluorescence.[42] Isoelectric point (pI) of BSA is 5.8.[43] It is often used as a blocking agent in immunohistochemistry.[44]

b. *Hen Egg White Lysozyme* (Lys) is also a globular single chain polypeptide with 129 number of amino acids having molecular weight of ~14.3 kDa. Isoelectric point of Lys is 11.2.[45] It basically acts as an enzyme which is able to lyse bacterial cell membranes to kill.[46]

c. *β-lactoglobulin* (β-lg) is a predominant sheet protein having molecular weight of ~18.6 kDa with 162 number of amino acid residues. This protein retains its dimeric form in neutral environment (pH~ 7) whereas it is in monomeric from in acidic solution.[47]

d. *Human recombinant insulin* (Ins) is a small protein hormone (molecular weight ~5.8 kDa and 51 amino acid residues. It consists of two polypeptide chains: A-chain (21 residues) and B-chain (30 residues) linked via two disulphide bonds.[48] It is secreted from the β-cell of pancreas and plays a crucial role in controlling blood sugar level and hence widely used in diabetes II treatment.[49]

e. *Ubiquitin* (Ubi) is a small protein present in eukaryotic cell having 76 number of amino acid (molecular weight ~8.6 kDa) which is widely used in eukaryotic cell to modifying protein and modulating their functions.[50]

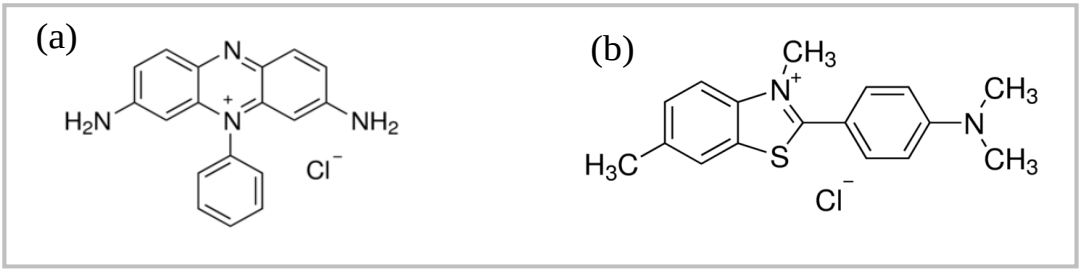


Scheme 2.III.B.1 Crystallin structure of (a) BSA (4f5s), (b) Lysozyme (1lyz), (c) β-lg (3npo) (d) human insulin (3i40) and (e) Ubiquitin from bovine erythrocyte (1ubq) as obtained from protein data bank (PDB). [pdb files are viewed and modified in *Mathematica 9*]

2.III.C Dye molecules:

Phenosafranine, PSF (3,7-diamino-5-phenylphenazinium chloride) is a cationic phenazinium dye with absorption and fluorescence emission maxima in water at ~521 nm and ~580 nm respectively. Molar extinction of PSF in water at 520 nm is 35600 $M^{-1}cm^{-1}$.[51] It is a well-known intercalator while binding with DNA.

Thioflavin-T, ThT (2-[4-(dimethylamino)phenyl]-3,6-dimethyl-1,3-benzothiazol-3-ium chloride) is a benzothiazole salt, commonly known as amyloid fibril marker. It emits weakly at ~427 nm in water (pH 7.4) but when it binds with amyloid aggregates (β sheet) the fluorescence intensity enhances sharply[52-54] with emission maxima at ~482 nm when excited at 440 nm.



Scheme 2.III.B.2 Molecular structure of (a) PSF, (b) ThT (structure has been collected from *www.sigmaalddrich.com*)

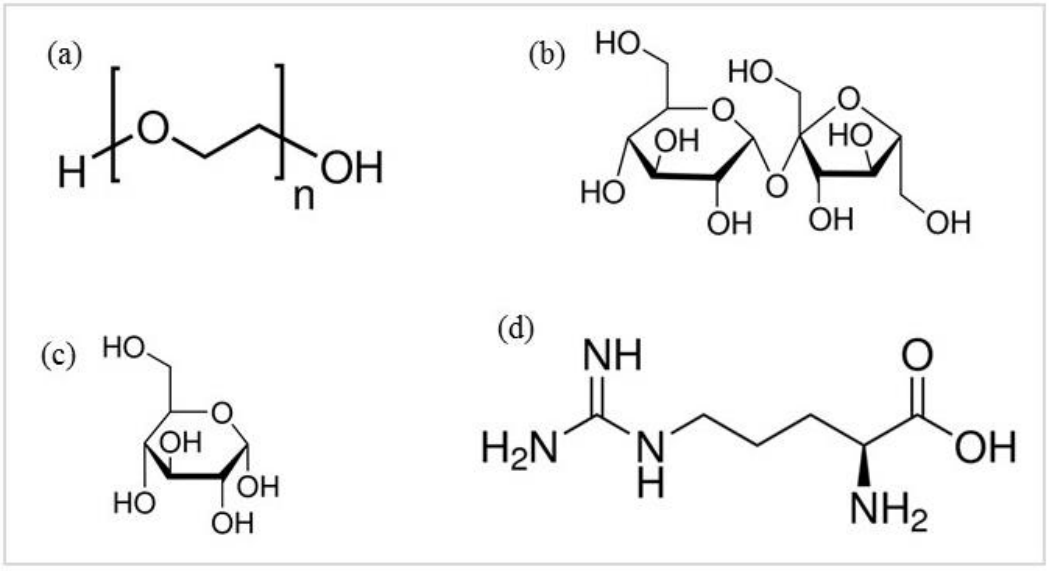
2.III.D Molecular crowders/ Excipients:

a. Polyethylene glycol (PEG) of different chain length has been used to mimic the cellular environments and also to increase the viscosity of the solution. The general molecular formula of PEG is $C_{2n}H_{4n+2}O_{n+1}$. Depending on the value of "n" we can get different molecular weight PEG e.g. the value of "n" for PEG200 with average molecular weight of 200 (gm/mol) is 4-5.

b. Sucrose (molecular weight 342.3 gm/mol) is a polar molecule along with some hydrophobic moiety. It consists of one glucose and one fructose molecule connected with each other. It is widely used as a stabilizing agent of protein during unfolding process.[55]

c. Glucose (molecular weight 180.15 gm/mol) is a monomer of sucrose. It is also used as molecular crowders.

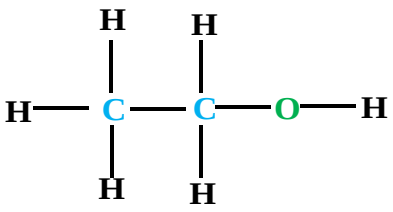
d. L-Arginine (molecular weight 174.2 gm/mol) is a basic amino acid and also used as molecular crowder.



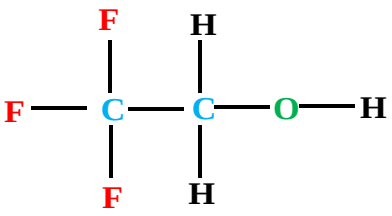
Scheme 2.III.B.3 Molecular structure of (a) PEG, (b) sucrose, (c) glucose and (d) L-Arg (structure has been collected from *www.sigmaalddrich.com*).

2.III.E Cosolvents:

Ethanol and its fluorine derivative *2,2,2-trifluoroethanol* (TFE) are very well known cosolvents used to investigate the protein folding-unfolding in presence of them. Both these alcohols offer comparable dielectric behaviour ($\epsilon_{ethanol}$ ~24 and ϵ_{TFE} ~27 at room temperature);[56] however they show distinct nature on protein stability. [57]



(a) Ethanol



(b) 2, 2, 2-trifluoroethanol

Scheme 2.III.B.4 Molecular structure of (a) ethanol, (b) TFE.

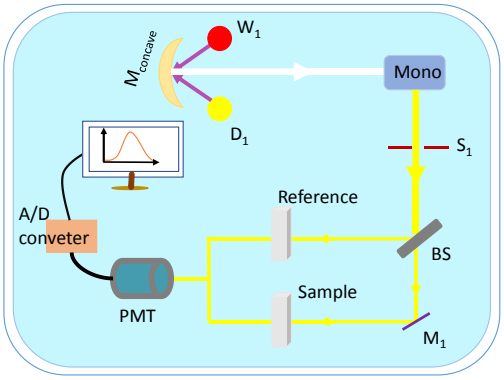
2.IV Instruments:

2.IV.A Ultraviolet-Visible Spectrophotometer:

UV-vis spectrophotometer (*Shimadzu UV 2600*) (scheme 2.IV.A) has been used to record the absorption spectra. This spectroscopy infers an idea about the transition of a molecule from ground state to higher electronic state (e.g. S1, S2, etc.). A tungsten lamp is used as visible source whereas deuterium lamp as UV source with lamp interchange wavelength at 282-293 nm. A monochromator with highly blazed holographic grating is used to disperse white light to a single tuneable light. Upon irradiating continuously on sample some fraction of the light is absorbed by the sample and rest is transmitted and the transmitted light is collected via a photo-multiplier tube (PMT) attached with the spectrophotometer. Mathematically the absorption of a molecule can be expressed as $A = \log_{10}(\frac{I_0}{I})$ where I_0 is the incident intensity of light falling on sample and I is the intensity leaving the sample. According to Beer-Lambert the absorption can be expressed as $A_\lambda = \varepsilon_\lambda cl$ where ε_λ is the molar extinction coefficient of the sample at wavelength λ, c is the concentration of sample in solution and l is the path length of the sample.

So, $A_\lambda = \log_{10}(\frac{I_0}{I}) = \varepsilon_\lambda cl$ (2.IV.A1)

A temperature controller with circulated water bath was used during temperature dependent absorption measurements. Solution turbidity has been estimated by monitoring the absorbance at 400 nm.[58]



Scheme 2.IV.A Schematic diagram of Absorption spectrophotometer. W_1 and D_1 is the tungsten and deuterium lamp. Mono is the monochromator.M_1 is the plane mirror and S_1 is the slit. BS is the beam splitter. $M_{concave}$ is a concave mirror.

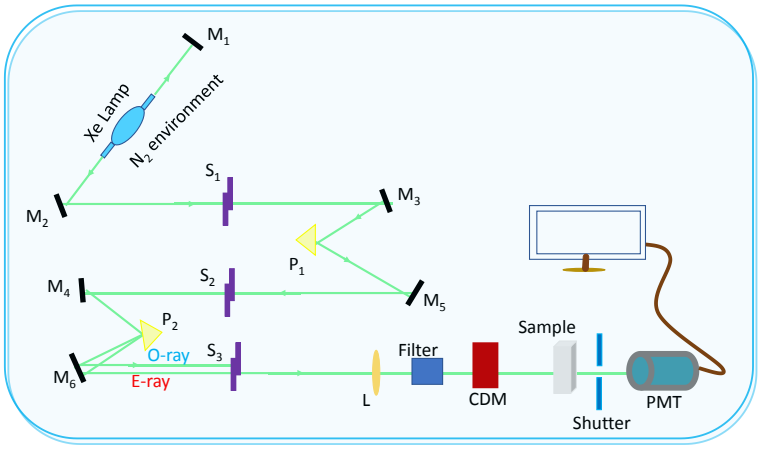
2.IV.B Circular dichroism (CD) Spectrometer:

The structure of DNA and proteins are analysed with the help of CD spectroscopy (model number: *Jasco-J 815*). A schematic diagram of CD spectroscopy has been demonstrated in scheme 2.IV.B. A Xe lamp is used as the light source and nitrogen gas is purged to remove the ozone gas. When left and right circularly polarised light passes through a chiral molecule, the sample absorbs differently.[59] CD signal, θ_λ is expressed as the difference between the absorption due to left circularly polarised (LCP) and right circularly polarised (RCP) light i.e.[26]

$$\theta_\lambda = A_\lambda\,(LCP) - A_\lambda\,(RCP) \qquad \text{(2.IV.B1)}$$

Molar ellipticity, ε is defined as[26] $\varepsilon\ (\mathrm{deg\,cm^2 mol^{-1}}) = \frac{\theta_\lambda\,(mdeg)}{C_M \times l \times 10}$ (2.IV.B2)

Where $\theta_\lambda\,(mdeg)$is the CD signal at wavelength λ in the unit of mdeg in C_M is the molar concentration of sample and *l* is path length of the sample in cm.



Scheme 2.IV.B Schematic diagram of circular dichroism spectroscopy. M_i, S_i, P_i, E-ray, O-ray, L, CDM, PMT represent mirror, slit, reflecting prism, extra-ordinary ray, ordinary ray, lens, circular dichroism modulator, photomultiplier tube respectively.

2.IV.C Dynamic light scattering:

Dynamic light scattering (DLS) measurement has been carried out in *Nano S Malvern* instrument (schematic 2.IV.C), engaging a 4 mW laser (λ =632.8 nm) equipped with a thermostatic sample chamber, to measure the hydrodynamic diameter of samples. Light (photon) upon colliding with the particles' scatters in all direction, and the photons only scattered at 173º are collected. Detail instrumentation may be found elsewhere.[60] DLS

measures correlation curve from where diffusion coefficient is calculated directly by fitting the curve using the equation as

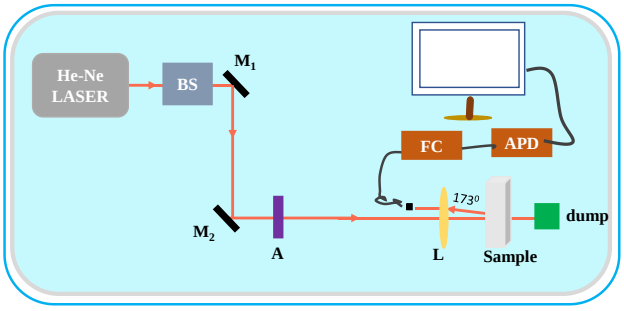
$$G(t) = a[1 + b\, e^{-2\Gamma t}] \qquad \text{(2.IV.C1)}$$

Where "a" is the baseline of the correlation function, b is the intercept, $\Gamma = D\left(\frac{4\pi n}{\lambda_0}\right)^2 Sin^2(\frac{\theta}{2})$; D is the diffusion coefficient, n is the refractive index of the dispersant, λ_0 is the wavelength of the laser and θ is the scattering angle.

Finally, hydrodynamic diameter of the particle is determined using the following formula:

$$d_H = \frac{k_B T}{3\pi\eta D} \qquad \text{(2.IV.C2)}$$

Where k_B is the Boltzman constant, η is the viscosity of solvent, T is the temperature.



Scheme 2.IV.C Schematics of dynamic light scattering technique. BS, FC, APD and dump stands for the beam sampler, fibre coupler, avalanche photodiode, beam dump, respectively. M_i is the plane mirror and L is the lens and A represents attenuator.

2.IV.D Fluorescence Spectroscopy:

2.IV.D.a Steady state fluorescence:

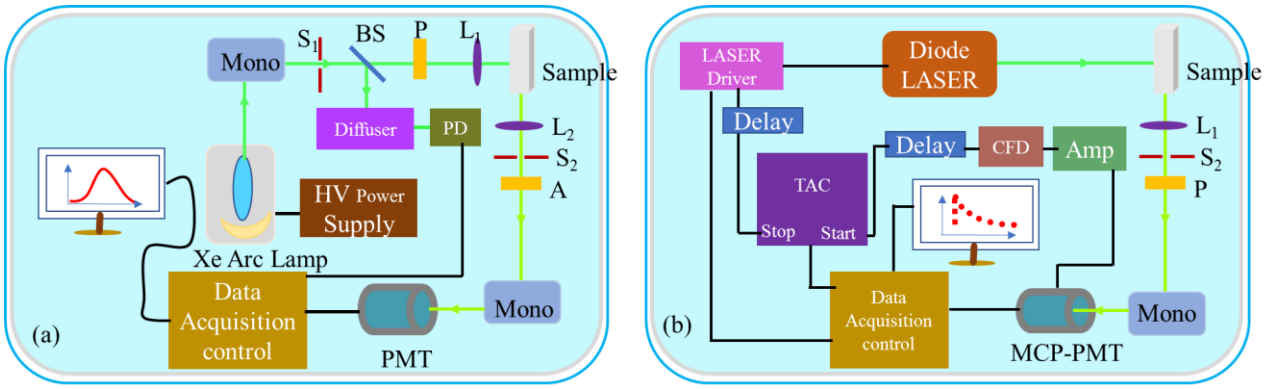
We have used Fluorolog (Horiba Jobin Yvon) for steady state fluorescence measurements. Detail instrumentation and working principle can be found elsewhere.[61] A schematic diagram of steady state fluorescence spectroscopy is shown in 2.IV.Da. Ozone free Xe-Arc lamp is used here as continuous light source. Light coming out from the source is passed through a monochromator (excitation monochromator) to get tuneable monochromic light to excite the sample. According to Kasha, sample fluoresces from lower most excited electronic state (S_1) (exception: malachite green, which fluoresces from S_2 state). Sample emits continuum

fluorescence in all direction, but it should be collected perpendicularly to minimize the scattering. A second monochromator (emission monochromator) is used to collect the fluorescence intensity at a particular wavelength. Finally, the signal is stored in PMT and data is collected through computer.

2.IV.D.b Time Resolved fluorescence:

2.IV.D.b.1 Time correlated single photon counting:

We record pico-second (ps) resolved fluorescence transients using time correlated single photon counting (TCSPC) technique (Life Spec II Edinburgh Instruments, UK) (schematic 2.IV.Db). Instrument response function (IRF) of the instrument varies from 70-250 ps depending upon the choice of laser. Polarisers in the emission side is kept at 54.7^o (magic angle) with respect to the polariser at excitation side to minimize the anisotropic effect. Measured transients are fitted globally using exponential decay equation in F900 software.

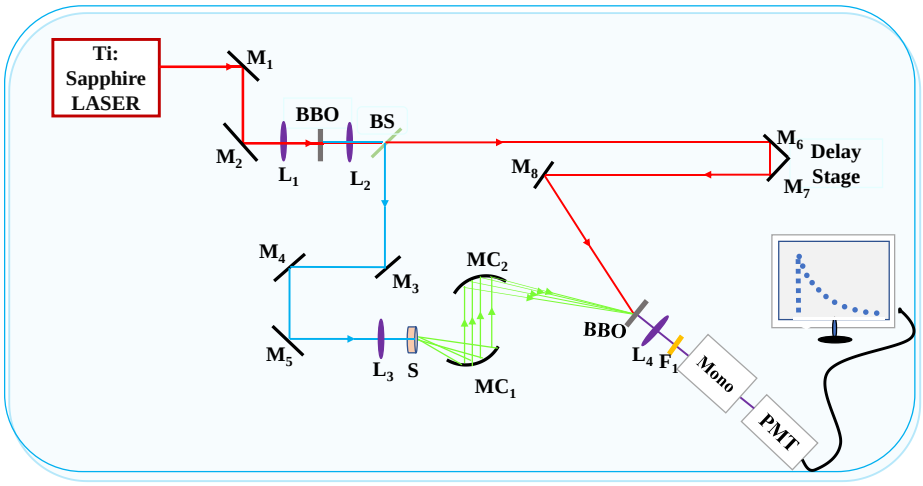


Scheme 2.IV.D Schematics of (a) steady state fluorescence and (b) time resolved fluorescence spectroscopy (TCSPC). S_i, BS, P, L_i stands for the slit, beam splitter, polariser, lens respectively. TAC and CFD in (B) are the abbreviation of time to amplitude converter and constant fraction discriminator.

2.IV.D.b.2 Fluorescence up-conversion spectroscopy:

Sub ps lifetime component of the fluorescence transient measurements was executed in a fluorescence up-conversion set up (FOG-100, CDP Corp., Russia) (Scheme 2.IV.E). The details of fluorescence up-conversion spectroscopy can be found elsewhere[62]. A mode-locked femtosecond Ti: Sapphire oscillator (MaiTai HP, Spectra Physics, USA) laser with pulse width of 80 fs is used as the fundamental light source and second harmonics is generated by focusing fundamental light onto a 0.2 mm β-barium borate (BBO) crystal to use as excitation source. The excitation light at magic angle polarization is focused onto a rotating cuvette and the

emitted fluorescence from the sample and the gate pulse is mixed on another 0.2 mm BBO crystal to generate the sum frequency light, which is dispersed in a monochromator and detected using a PMT coupled with a photon counter. The power of the excitation light is generally kept ~2-3 mW and the overall instrument response function is to be ~250 fs.



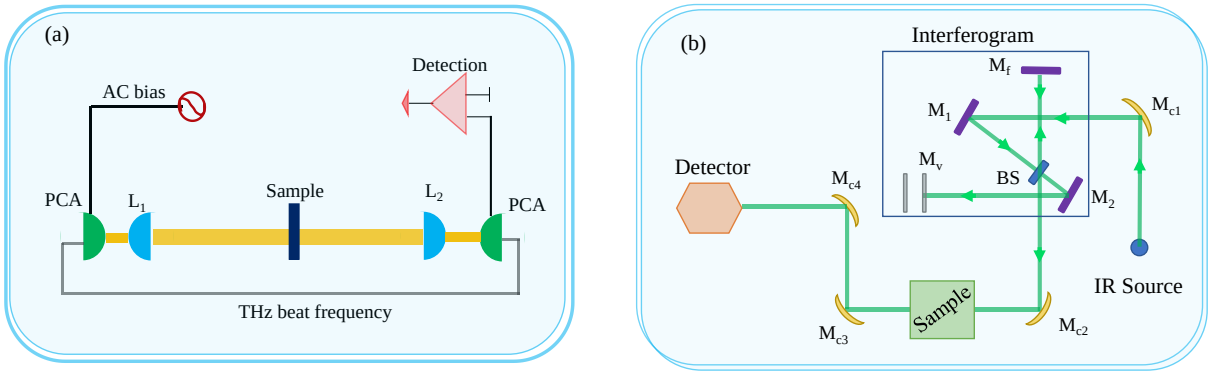
Scheme 2.IV.E: Schematic diagram of fluorescence up-conversion spectroscopy. M_i, L_i, MC_i are flat mirror, convex lenses, and concave mirrors, respectively. S stands for rotating sample chamber. BS is the dichroic beam splitter. MONO and PMT are the abbreviation of monochromator and photo multiplier tube, respectively. F_1 is the 400 nm high pass filter.

2.IV.E Frequency Domain THz spectroscopy:

Terahertz (THz) frequency domain spectroscopy (TFDS) measurements were performed on Toptica (Tera-scan 1550).[63, 64] A schematic diagram has been illustrated in Scheme 2.IV.Fa. A continuous wave (CW) THz radiation is obtained by optical heterodyning technique. Lights form two CW lasers (one has the fixed wavelength of 1550 nm and the other one has the variable wavelength ranging from 1550-1570 nm) are mixed in the photo-mixer in order to generate THz frequency exactly at the difference frequency of the lasers. A bias voltage is applied to the electrodes present in the photo-mixer to generate photocurrents which oscillates at the beat frequency. An antenna near photo-mixer emits electromagnetic wave in the THz frequency (0.1-1.2 THz). The typical power emitted is about 1 microwatt and beyond 1.2 THz the Signal to Noise is quite low due to the frequency fall-off response of the THz detector. During the experiment a heavy nitrogen is purged at the sample chamber to remove the unwanted water vapour.

2.IV.F Far IR FTIR spectroscopy:

Far infrared (FIR) spectra in the range of 50-700 cm^{-1} were collected from Vertex 70V (Bruker, Germany) Fourier transform infrared (FTIR) spectrometer equipped with a complex material (DLaTGS) detector (schematic diagram has been shown in Scheme 2.IV.Fb). Fourier spectroscopy in the FIR range is widely used to investigate the hydration behaviour of solute molecules and biomolecules.[13, 21, 23, 65] Measurements were carried out in ATR (attenuated total reflection) mode (a diamond crystal is used in the ATR attachment). The sample compartment was evacuated (pressure < 1hPa) using a vacuum- pump every time prior to the experiment. A polyvinyl chloride (PVC) disk was used to prevent the sample from being washed out from the diamond crystal after evacuating the instrument. The working principle of FTIR follows the Michelson interferometer principle.[66]



Scheme 2.IV.F: Schematic diagram of (a) THz frequency domain and (b) FTIR spectroscopy. L_i, PCA in scheme (a) stands for plano-convex lens and photo-conductive antenna respectively. M_i, M_f, M_v, M_{ci} in figure (b) represent plane mirror, plane fixed mirror, movable mirror (for delay stage) and concave mirror respectively. BS is the beam-splitter.

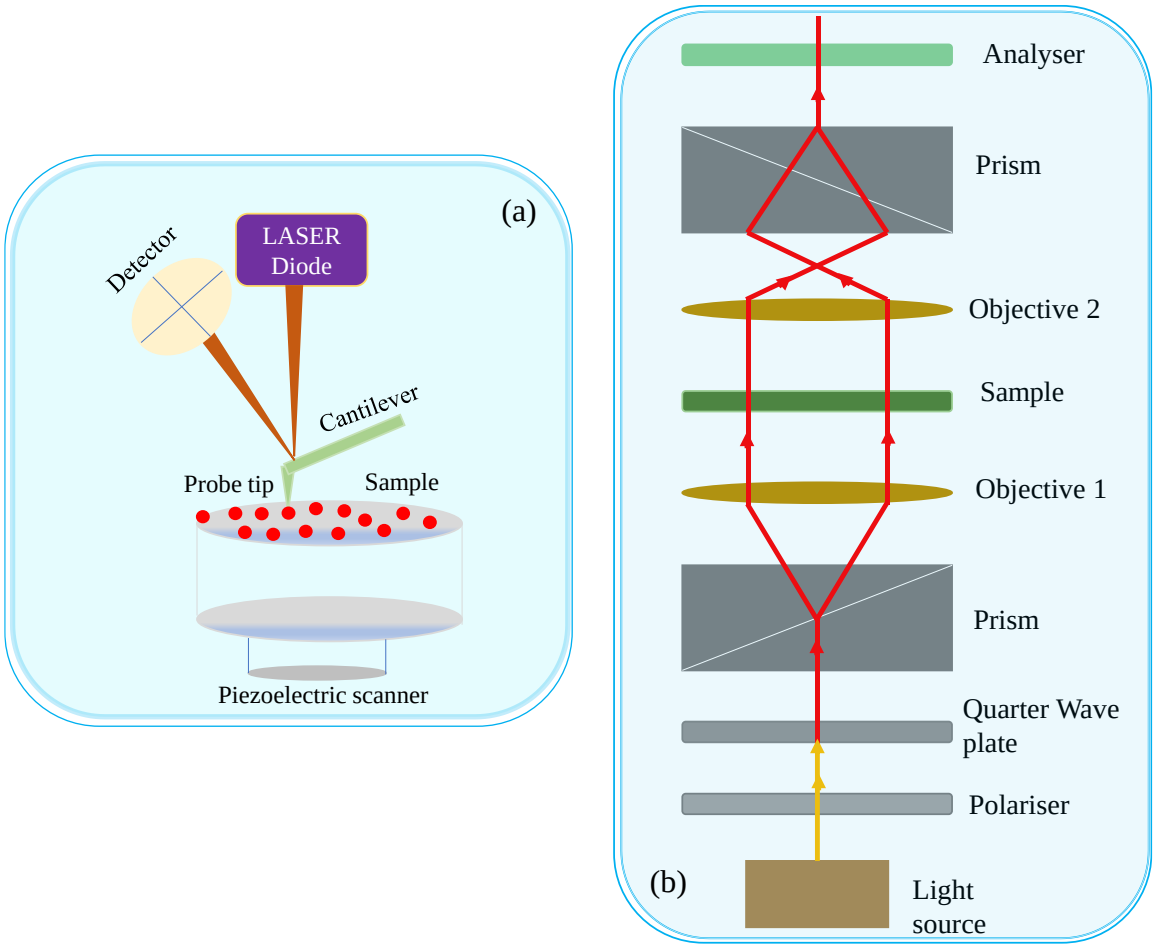
2.IV.G Atomic force microscopy:

Morphology of oligomers and fibrils were being observed with the help of atomic force microscopy (AFM; model number di INNOVA). AFM is one type of scanning probe microscope by which the topographical image of a sample surface is accomplished.[67, 68] A typical AFM consists of a probe (tip) made of silicon or silicon nitride at the free end of the cantilever (see 2.IV.Ga for the schematic diagram of the AFM). The basic principle of AFM is the interaction between sample surface and the tip, present in the microscope.[69, 70] When the probe approaches towards very close to the sample surface (i.e. in a range of interatomic

distance) attractive or repulsive forces appears due to the interaction between probe and sample. These attractive or repulsive force causes the cantilever to suffer a negative or positive bending respectively. This bending is noticed through the laser beam. In general the probe is kept fixed, and sample is moved in the x, y, z direction using piezoelectric material. Piezoelectric materials can enlarge or shrink depending on the applied voltage. Thus a precise movement in all directions can be achievable. Laser beam is focused onto the back of the cantilever and the reflected light is focused on the four-quadrant photo-detector. The bending of the cantilever can be accurately measured using the position sensitive photodiode. The cantilever deflects according to the atomic force variation between tip and the sample and thereby the detector measures the deflection. The created topographical image of the sample surface is recorded through the computer. Depending on the sample type, generally contact mode (for hard sample) or tapping mode (for soft bio-samples) are utilised in accumulating the topography image from AFM.

2.IV.H Optical DIC microscopy:

All the images of liquid-liquid phase separation were observed in Leica DM6M differential interference contrast (DIC) microscope. In DIC microscope[71] a linearly polarised light is used to illuminate the sample. The polarised light is first dispersed into two distinct rays having polarisation perpendicular to each other. These two rays travel extremely near to each other, but they do not interfere due to their polarisation orthogonal to each other. Upon passing through the specimen (sample) an alteration of the effective pathlength (effective pathlength means the path travel by light multiplied by the refractive index of the medium) of both rays takes place due to the different thickness and refractive index of the sample occurring a phase shift (as phase shift $\frac{2\pi}{\lambda}$ path difference λ being the wavelength). Then they enter through objective where they are focused above their rear focal plane. Now upon recombining they will interfere with each other, and the resultant light becomes elliptically polarized. This polarization can be converted into an amplitude shift via an analyser. The light finally travels through objective or camera where the image is constructed with difference in intensity and colour. A schematic diagram of this DIC microscope is illustrated scheme 2.IV.Gb.



Scheme 2.IV.G: Schematic diagram behind the operating principle of (a) atomic force microscopy and (b) Optical DIC microscopy.

2.V References:

1. A. A. Spector and M. A. Yorek, *J. Lipid Res.*, 1985, **26**, 1015-1035.
2. M. Feric, N. Vaidya, T. S. Harmon, D. M. Mitrea, L. Zhu, T. M. Richardson, R. W. Kriwacki, R. V. Pappu and C. P. Brangwynne, *Cell*, 2016, **165**, 1686-1697.
3. T. E. Kaiser, R. V. Intine and M. Dundr, *Science*, 2008, **322**, 1713-1717.
4. A. Molliex, J. Temirov, J. Lee, M. Coughlin, A. P. Kanagaraj, H. J. Kim, T. Mittag and J. P. Taylor, *Cell*, 2015, **163**, 123-133.
5. S. F. Banani, H. O. Lee, A. A. Hyman and M. K. Rosen, *Nat. Rev. Mol. Cell Biol.*, 2017, **18**, 285–298.
6. S. Boeynaems, S. Alberti, N. L. Fawzi, T. Mittag, M. Polymenidou, F. Rousseau, J. Schymkowitz, J. Shorter, B. Wolozin, L. V. D. Bosch, P. Tompa and M. Fuxreiter, *Trends Cell Biol.*, 2018, **28**, 420-435.
7. Y. Shin and C. P. Brangwynne, *Science*, 2017, **357**, eaaf4382.
8. C. P. Brangwynne, P. Tompa and R. V. Pappu, *Nat. Phys.*, 2015, **11**, 899–904.
9. M. L. Huggins, *J. Phys. Chem.*, 1942, **46**, 151-158.
10. P. J. Flory, *J. Chem. Phys.*, 1942, **10**, 51-61.
11. T. Dodo, M. Sugawa, E. Nonaka, H. Honda and S. Ikawa, *J. Chem. Phys.*, 1995, **102**, 6208.

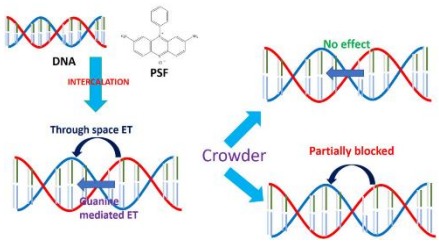
12. D. A. Schmidt, O. Birer, S. Funkner, B. P. Born, R. Gnanasekaran, G. W. Schwaab, D. M. Leitner and M. Havenith, *J. Am. Chem. Soc.*, 2009, **131**, 18512–18517.
13. F. Sebastiani, C. Y. Ma, S. Funke, A. Baumer, D. Decka, C. Hoberg, A. Esser, H. Forbert, G. Schwaab, D. Marx and M. Havenith, *Angew. Chem. Int. Ed.*, 2021, **60**, 3768–3772.
14. P. Schienbein, G. Schwaab, H. Forbert, M. Havenith and D. Marx, *J. Phys. Chem. Lett.*, 2017, **8**, 2373−2380.
15. H. Vondracek, J. Dielmann-Gessner, W. Lubitz, M. Knipp and M. Havenith, *J. Chem. Phys.*, 2014, **141**, 22D534.
16. N. Samanta, D. D. Mahanta and R. K. Mitra, *Phys. Chem. Chem. Phys.*, 2014, **16**, 23308-23315.
17. V. Vitale, J. Dziedzic, M. M. Dubois, H. Fangohr and C. K. Skylaris, *J. Chem. Theory Comput.*, 2015, **11**, 3321−3332.
18. B. Bagchi, *Chem. Rev.*, 2005, **105**, 3197−3219.
19. H. R. Drew and R. E. Dickerson, *J. Mol. Biol.*, 1981, **151**, 535−556.
20. L. Fumagalli, A. Esfandiar, R. Fabregas, S. Hu, P. Ares, A. Janardanan, Q. Yang, B. Radha, T. Taniguchi, K. Watanabe, G. Gomila, K. S. Novoselov and A. K. Geim, *Science*, 2018, **360**, 1339–1342.
21. V. Conti Nibali and M. Havenith, *J. Am. Chem. Soc.*, 2014, **136**, 12800−12807.
22. S. Ebbinghaus, S. J. Kim, M. Heyden, X. Yu, U. Heugen, M. Grubele, D. M. Leitner and M. Havenith, *Proc. Natl. Acad. Sci. U.S.A*, 2007, **104**, 20749-20752.
23. G. Schwaab, F. Sebastiani and M. Havenith, *Angew. Chem. Intl. Ed.*, 2019, **58**, 3000–3013.
24. D. Sarkar, P. Das, S. Basak and N. Chattopadhyay, *J. Phys. Chem. B*, 2008, **112**, 9243–9249.
25. J. Kypr, I. Kejnovska´, D. Renc˘iuk and M. Vorlıc˘kova, *Nucleic Acids Res.*, 2009, **37**, 1713-1725.
26. N. J. Greenfield, *Nature Protocols*, 2006, **1**, 2876–2890.
27. I. M. Kuznetsova, K. K. Turoverov and V. N. Uversky, *J. Proteome Res.*, 2004, **3**, 485-494.
28. D. C. Jenkins, I. D. Sylvester and J. T. Pinheiro, *FEBS J.*, 2008, **275**, 1323–1335.
29. S. K. Pal and A. Zewail, *Chem. Rev.*, 2004, **104**, 2099−2124.
30. L. L. del Mercato, P. P. Pompa, G. Maruccio, A. D. Torre, S. Sabella, A. M. Tamburro, R. Cingolani and R. Rinaldi, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 18019-18024.
31. N. Amdursky, M. Molotskii, D. Aronov, L. Adler-Abramovich, E. Gazit and G. Rosenman, *Nano Lett.* , 2009, **9**, 3111–3115.
32. P. Hanczyc, M. Samoc and B. Norden, *Nat. Photon.*, 2013, **7**, 969–972.
33. S. Ebbinghaus, S. J. Kim, M. Heyden, X. Yu, U. Heugen, M. Gruebele, D. M. Leitner and M. Havenith, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 20749−20752.
34. V. Conti Nibali and M. Havenith, *J. Am. Chem. Soc.*, 2014, **136**, 12800−12807.
35. M. Heyden, J. Sun, S. Funkner, G. Mathias, H. Forbert, M. Havenith and D. Marx, *Proc. Natl. Acad. Sci. U.S.A.*, 2010, **107**, 12068-12073.
36. S. Ebbinghaus, S. J. Kim, M. Heyden, X. Yu, U. Heugen, M. Gruebele, D. M. Leitner and M. Havenith, *Proc. Nat. Acad. Sci. USA*, 2007, **104**, 20749–20752.
37. M. Milosevic, *Appl. Spectrosc. Rev*, 2004, **39**, 365-384.
38. A. M. Forson, H. C. van der Mei and J. Sjollema, *Sci. Rep*, 2020, **10**, 12093.
39. N. K. Modukuru, K. J. Snow, B. S. Perrin and C. V. Kumar, *J. Phys. Chem. B*, 2005, **109**, 11810-11818.
40. C. Niedergang, H. Okazaki and P. Mandel, *Eur. J. Biochem.*, 1979, **102**, 43-57.
41. R. G. Reed, R. C. Feldhoff, O. L. Clute and T. Peters, *Biochemistry*, 1975, **14**, 4578-4583.
42. Y. Moriyama, D. Ohta, K. Hachiya, Y. Mitsui and K. Takeda, *J. Protein Chem.*, 1996, **15**, 265-272.
43. A. Salis, M. Bostro¨m, L. Medda, F. Cugia, B. Barse, D. F. Parsons, B. W. Ninham and M. Monduzzi, *Langmuir*, , , 2011, **27**, 11597–11604.
44. R. Ahirwar, S. Bariar, A. Balakrishnan and P. Nahar, *RSC Adv.*, 2015, **5**, 100077.
45. M. Bostro¨m, F. W. Tavares , S. Finet, F. Skouri-Panet, A. Tardieu and B. W. Ninham, *Biophys. Chem.*, 2005, **117** 217 – 224.
46. T. Ganz, *Encyclopedia of Respiratory Medicine (Article: Lysozyme, page: 649-653)*, Academic Press, 2006.

47. J. R. Albani, J. Vogelaer, L. Bretesche and D. Kmiecik, *J. Pharm. Biomed. Anal.*, 2014, **91**, 144– 150.
48. D. P. Hong and A. L. Fink, *Biochemistry*, 2005, **44**, 16701-16709.
49. P. V. Röder, B. Wu, Y. Liu and W. Han, *Exp. Mol. Med.*, 2016, **48**, e219.
50. D. Mukhopadhyay and H. Riezman, *Science*, 2007, **315**, 201-205.
51. M. Taniguchi and J. S. Lindsey, *Photochem. Photobiol.*, 2017, **94**, 290-327.
52. B. Urbanc, L. Cruz, R. Le, J. Sanders, K. H. Ashe, K. Duff, H. E. Stanley, M. C. Irizarry and B. T. Hyman, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 13990–13995.
53. P. Friedhoff, A. Schneider, E. M. Mandelkow and E. Mandelkow, *Biochemistry*, 1998, **37**, 10223–10230.
54. T. Kampers, P. Friedhoff, J. Biernat, E. M. Mandelkow and E. Mandelkow, *FEBS Lett*, 1996, **399**, 344–349
55. J. C. Lee and S. N. Timasheff, *J. Biol. Chem.*, 1981, **256**, 7193-7201.
56. C. Wohlfarth, *Static Dielectric Constants of Pure Liquids and Binary Liquid Mixtures*, Springer2015.
57. S. Maity, S. Sardar, Pal, S., H. Parvej, J. Chakraborty and U. C. Halder, *RSC Adv.*, 2016, **6**, 74409-74417.
58. H. Cinar and R. Winter, *Sci. Rep*, 2020, **10**, 17245.
59. G. D. Fasman, *Circular Dichroism and the Conformational Analysis of Biomolecules*, Plenum Press, NewYork, 1996.
60. R. K. Mitra, S. S. Sinha and S. K. Pal, *Langmuir*, 2007, **23**, 10224-10229.
61. J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*2006.
62. P. Mukherjee, A. Das and P. Sen, *Chem. Phys.*, 2018, **513**, 141-148.
63. R. Sarkar, D. Ghindani, K. M. Devi, S. S. Prabhu, A. Ahmad and G. Kumar, *Sci. Rep.*, 2019, **9**, 18068.
64. S. Han, L. Cong, Y. K. Srivastava, B. Qiang, M. V. Rybin, A. Kumar, R. Jain, W. X. Lim, V. G. Achanta, S. S. Prabhu, Q. J. Wang, Y. S. Kivshar and R. Singh, *Adv. Mater*, 2019, **31**, 1901921.
65. J. Ahlers, E. M. Adams, V. Bader, S. Pezzotti, K. F. Winklhofer, J. Tatzelt and M. Havenith, *Biophysical J.*, 2021, **120**, 1266–1275.
66. V. Saptari, *Fourier-Transform Spectroscopy Instrumentation Engineering*, SPIE Optical Engineering Press.
67. G. Lamour, C. K. Yip, H. Li and J. Gsponer, *ACS Nano*, 2014, **8**, 3851–3861.
68. Y. F. Dufrêne, T. Ando, R. Garcia, D. Alsteens, D. Martinez-Martin, A. Engel, C. Gerber and D. J. Müller, *Nat. Nanotechnol*, 2017, **12**, 295–307.
69. N. Gavara, *Microsc. Res. Tech.*, 2016, **80**, 75-84.
70. R. W. Carpick and M. Salmeron, *Chem. Rev.*, 1997, **97**, 1163−1194.
71. D. B. Murphy, *Fundamentals of Light Microscopy and Electronic Imaging*, A John Willey & Sons, NewYork, 2001.

Chapter 3

3. Polyethylene glycols affect electron transfer rate in Phenosafranin-DNA complex

Long distance electron transfer (ET) between small ligands and DNA is a much-studied phenomenon and is principally believed to occur through electron (or hole) hopping. Several studies have been carried out in aqueous environments while in real biological milieu the DNA molecules experience a denser and more heterogeneous environment containing otherwise indifferent molecular crowders. It is therefore expected that the ET could get modified in the presence of crowding agent and to investigate that we have made elaborate studies on steady state and time-resolved (picosecond and femtosecond-resolved) emission properties of a phenosafranine (PSF) intercalated to calf thymus (CT) DNA in the presence of ethylene glycol (EG) and polyethylene glycols (PEG) of different chain lengths (PEG 200, 400 and 1000). The emission of PSF gets considerably quenched when intercalated to DNA; the quenching is released when PEGs are added into it. The structural integrity of the CT DNA has been established using circular dichroism spectroscopy. CD measurements have evidenced only marginal changes in the DNA structure upon the addition of PEG. Picosecond (ps)-resolved fluorescence measurements show significant decrease in the contribution of the DNA induced quenched time-constant of PSF upon the addition of PEGs, however, femtosecond (fs)-resolved measurements show less noticeable changes in the time constants. Our study shows that the electron hopping rate through the guanine base in DNA core remains unaffected whereas the 'through space' electron transfer process does get affected in the presence of molecular crowders.



3.I. Introduction

Interaction of small molecules (ligands) with bio-macromolecules like protein, DNA, etc. has appeared as one of the most enthralling fields of research in the recent past.[1] Small molecules bind to DNA in three different manners: (i) through intercalation (mostly seen in planar molecules which inserts between the base pairs of DNA), (ii) through minor groove binding (applicable for non-planer curved molecules, which fit inside the minor grove of DNA) and (iii) through electrostatic interaction (mostly applicable for positively charged ligands), which often has been evidenced to initiate the other two binding modes.[2] Intercalative binding is often recognized as the most important one medicinally.[3] Binding of small molecules has been noticed to be mostly entropy driven.[4] Drug molecules are found to get de-intercalated from the DNA interior in presence of organized assemblies which has been explained in term of the drug-DNA dissociation equilibrium shifting.[5-7] Such process has been found to be primarily entropy driven irrespective of the binding mode of the drug molecules.[8] Interestingly most of these studies have been performed in aqueous milieu whereas the real biological environment is highly crowded.[9] Such crowded environment could be achieved by adding molecular crowders in the aqueous solutions.[10] In spite of enormous effort in studying the effect

of such crowders on proteins, comparatively less attention has been paid on their effects on drug-DNA interactions.[11, 12] Here we have investigated the effect of poly-ethylene glycol (PEG) of different chain sizes on the intercalation of phenosafranine (PSF) in calf-thymus DNA (CT-DNA).

Molecular crowders can modify the rates as well as can shift the equilibrium of biomolecular reactions.[10, 13] They can also alter the water structure at the DNA surface,[14] compaction of the DNA structure,[15, 16] oxidative damage and electron transfer (ET) in oligonucleotides,[12] DNA melting behaviour[17] etc. The most commonly used molecular crowder is synthetic PEG based polymers owing to their high solubility in water.[18-20] However, their use as a pure 'molecular crowder' has recently been critically discussed.[19, 21] PEG molecules have been found to be either interactive or indifferent towards DNA depending upon their chain length and concentration.[22-24] Addition of PEGs could also induce drastic conformational transition in DNA between elongated coil to compact globule.[25, 26] Oh et al has reported that the binding mode of intercalative drugs does not get modified in the presence of molecular crowders.[27] The effect of PEG induced crowding on DNA (specifically G-quadruplexes) stability and functionality has also been investigated.[14, 28, 29] These studies point out that DNA-mediated ET between ligand and DNA base-pairs could also get affected in presence of molecular crowders as they can modify the diffusion rate, solvent polarity and water activity.

Phenosafranin (PSF), a cationic dye with potential application in photochemical and photophysical processes of biological importance, is known to bind to DNA.[30, 31] Our group previously showed that binding of PSF with DNA is electrostatic in nature and addition of sodium dodecyl sulphate (SDS) micelles de-intercalate PSF from DNA to relocate it at the micellar interface, the effect being predominantly entropy driven.[6] The role of SDS micelles as being molecular crowder or direct DNA binder remains elusive. In the present investigation we add PEG into intercalated PSF-DNA complex and observe a relative decrease in the emission intensity of the DNA intercalated PSF. We have employed picosecond (ps) and femtosecond (fs) time-resolved fluorescence spectroscopy (TRFS) coupled with circular dichroism (CD) measurements to understand the underlying mechanism. We observed that upon intercalation, the emission intensity is quenched heavily and the reason behind this is electron transfer (ET). We assigned two different types of ET during drug-DNA interaction: guanine mediated hopping ET which is independent on environment and through space ET solely dependent of environment.

3.II. Materials and Methods:

Calf thymus (CT) DNA, 3,7-diamino-5-phenylphenazinium chloride (phenosafranine, PSF), ethylene glycol (EG), polyethylene glycol of different chain lengths (PEG200, PEG400, PEG1000) were procured from Sigma-Aldrich of highest purity (~99%) and used without further purification. Details of the chemicals are described in section 2.III. Ultra-pure (Mili Q) water was used as solvents for all the measurements. CT-DNA solution was prepared in sodium phosphate buffer (10 mM) of pH ~7 by stirring the solution gently with magnetic stirrer. All the experiments were carried out at 22 °C. Used instruments are described in section 2.IV

3.III. Results and Discussions:

CD measurements: We investigate the structural stability of the DNA in presence of PEGs using far-UV CD measurements. A representative CD profile of the DNA in the presence of PEG200 has been depicted in the inset of figure 3.III.1a. It is observed that the absorption profile does not change appreciably, however, moderate changes in the peak intensity at 275 and 245 nm are identified. Such marginal change in the CD profile was also reported earlier.[32] For a comparative understanding, we plot the difference in the CD signal (as defined by $\Delta\varepsilon = \varepsilon^{DNA}_{crowder} - \varepsilon^{DNA}_{water}$) in the entire wavelength range in the presence of 20% PEG (figure 3.III.1a). It is observed that only marginal changes occur for EG, PEG200 and PEG400, while an appreciable change has been noticed for PEG1000, specially around the 245 nm peak. We also measure the thermal stability of the DNA in the presence of these molecular crowders. We determine the melting temperature (T_m, the temperature at which the ds-DNA transforms into a single strand) from UV absorption measurements monitoring the change in absorbance at 260 nm[14] (a representative plot is provided in figure 3.III.1b and 3.III.1c). The T_m values are plotted as a function of PEG concentration in figure 3.III.1d. We observe a near-linear decrease in T_m with increasing PEG concentration except for PEG 400, where the changes are drastic, especially in its low concentration range. An earlier study by Nakano et al.[33] reported a decrease in T_m of DNA oligomer duplexes in the presence of EG, PEG200 and PEG1000, with the effect being the most prominent in PEG200. The same group has also reported decrease in T_m of DNA quadruplexes in the presence of PEG200[14]. The CD and T_m measurements thus confirm subtle, yet definite changes in the native structure of CT-DNA in the presence of the PEGs. We then investigate whether intercalation of PSF induces changes in the DNA structure. Figure 3.III.2a depicts the CD profile of CT DNA in the presence of PSF. We observe noticeable changes in the CD profile, especially in the 275 nm peak. Intercalative binding of ligand

generally causes elongation in the DNA structure,[34, 35] as manifested in the changes in the CD profile. Addition of PEG into the DNA-PSF complex induces further changes in the CD profile (figure 3.III.2a). The extent of the change could be apprehended from the difference plot (figure 3.III.2b). The significant change in the 275 nm region, as induced by PSF, diminishes upon the addition of PEG. Alongside a noticeable change in the 245 nm region is observed, which was not evident either for PSF or for the PEGs individually. Thus PEG, grossly acting as a molecular crowder, induces further non-native structural perturbation in PSF-DNA complex. Here we note that PSF itself does not show any CD signal in the far and near UV region.[6] Intercalation of this drug into the DNA interior, however, induces chirality into it as the intercalated complex produces induced CD signal in the 500-580 nm region.[6, 8] Induced CD absorption spectra of DNA bound PSF in absence and in the presence of PEG200 are shown in figure 3.III.2c. The PSF-DNA complex (1:20 molar ratio) shows an intense peak at ~538 nm. The peak intensity gradually decreases as PEG200 is added (figure 3.III.2c). Interestingly, even at 30% PEG there remains a finite induced CD signal, which strongly concludes PSF molecules to still remain bound to the DNA even at this high PEG concentration.

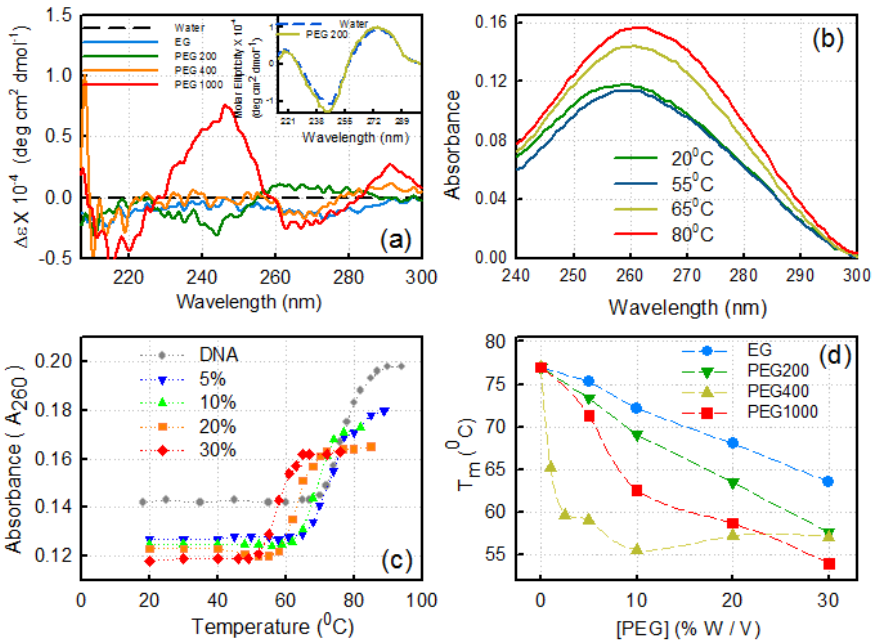


Figure 3.III.1 (a) Difference of ellipticity ($\Delta\varepsilon = \varepsilon^{DNA}_{crowder} - \varepsilon^{DNA}_{water}$) in the presence of 20% PEGs. The inset shows the representative molar ellipticity of CT DNA in water and 20% (w/v) of PEG200. (b) Representative profile of absorbance spectra of 20% PEG 200 as a function of temperature. (c) Representative change of absorbance at 260 nm for different concentration of PEG 200. (d) Melting temperature (T_m) of CT-DNA as a function of PEG concentration.

We plot the relative change in the induced CD peak as a function of PEG concentration (figure 3.III.2d). We observe moderate decrease in the induced peak intensity up to 10% PEG concentration beyond which it reduces gradually following the order EG<PEG200<PEG400. The observation in PEG1000 is contrasting as the induced CD intensity grows initially (up to 10% PEG) and then decreases to saturate at 25% PEG, which eventually is identical to the value that of the PSF-DNA complex. The induced CD measurement thus unambiguously concludes that even at 30% PEG concentration, a definite fraction of the ligand does remain intercalated to the DNA.

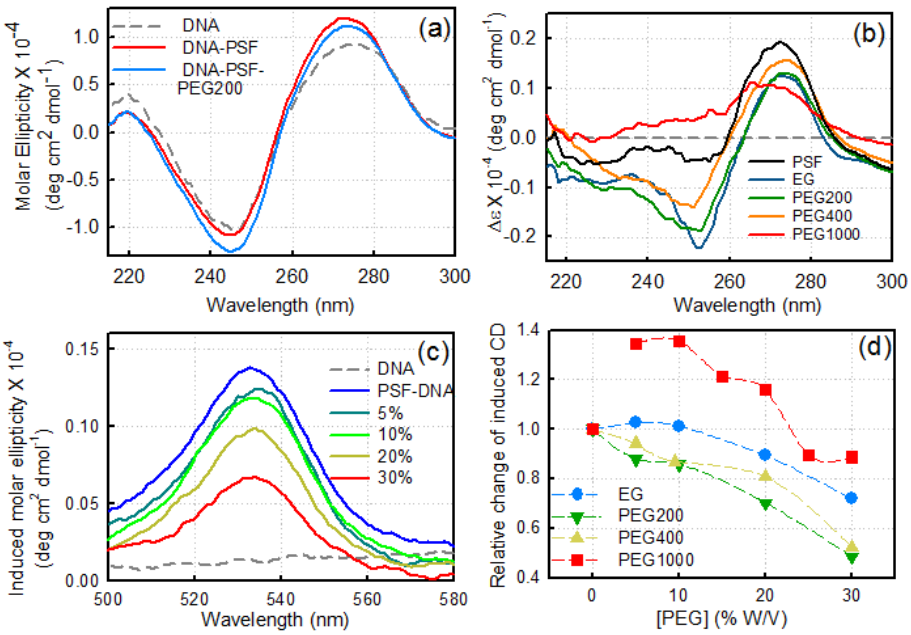


Figure 3.III.2. (a) Representative molar ellipticity of CT DNA in water and PSF bound DNA (PSF:DNA = 1:20) in the presence of 20% (w/v) of PEG200. (b) Difference of ellipticity (Δε) of different PEGs. (c) Induced CD spectra of PSF-DNA complex with increasing concentration of PEG200. (d) Induced CD signal at 532 nm for PSF-DNA complex in the presence of different concentration of PEGs.

Steady state fluorescence measurements: PSF in water shows emission maximum at 580 nm (figure 3.III.3a). Its intensity increases as PEG is added with a slight blue shift of ~5 nm (figure 3.III.3a). An earlier result by Bhattacharya et al.[36] also corroborates this finding, where they observed an increase in safranine T fluorescence intensity in the presence of PEG200. The observed shift may be due to an increase in the viscosity of the solution upon the addition of PEG. The increase in the emission intensity coupled with the blue shift clearly indicates either

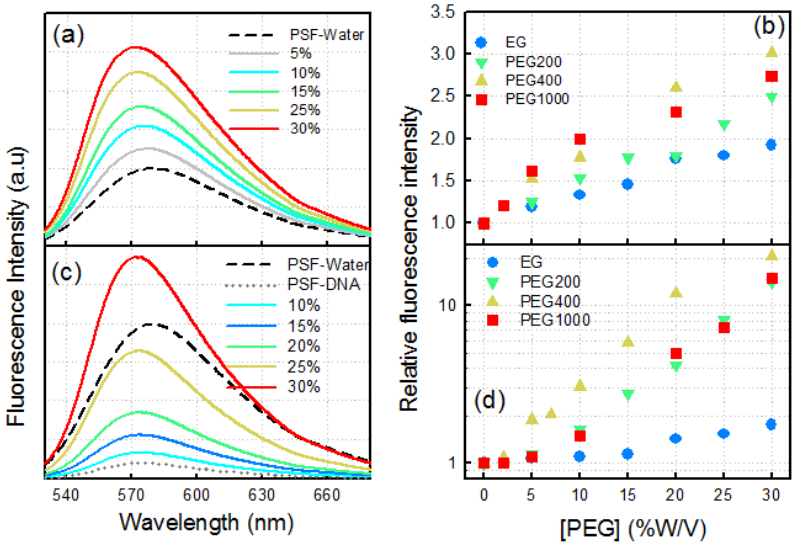


Figure 3.III.3. (a) Representative profile of fluorescence intensity of PSF in presence of PEG200 in absence of DNA. (b) Relative change in fluorescence intensity with crowding agents (c) Steady state fluorescence spectra of PSF in water, DNA and PSF-DNA complex in the presence of PEG200 at different concentrations. (d) Relative intensity (as expressed by the ratio of the fluorescence intensities measured at 580 nm of PSF-DNA and PSF-DNA-PEG systems) as a function of PEG concentrations.

---

relocation of the dye in a relatively non-polar PEG-environment or some specific interaction of the dye with the polymer or both.[37] The relative change in the intensity is found to follow a near-linear increase as a function of PEG concentration, with slopes being more or less comparable for all the PEGs under study (figure 3.III.3b). Addition of DNA severely quenches PSF emission coupled with a ~5 nm blue shift (figure 3.III.3c).[6, 38] Such quenching manifests excited state electron transfer (ET) through the G-C base pair stacking in the DNA.[39] It could be noted here that interaction of PSF with DNA is initially electrostatic in nature due to their opposite charge type, however, at the experimental concentration ratio (PSF:DNA=1:20) the binding is predominantly intercalative in nature,[6] which unambiguously concludes ET to be the acceptable mechanism responsible for the quenching. Addition of PEG partially recovers the quenching with no apparent change in the peak position. A representative emission profile in the presence of PEG200 is shown in figure 3.III.3c. The relative fluorescence intensity ($= \frac{I_{DNA-PSF-PEG}}{I_{DNA-PSF}}$) of DNA-PSF-PEG systems as a function of PEG concentration is plotted in figure 3.III.3d, and we found it to increase exponentially. The change is only moderate in the presence of EG, however, relatively steep rise is observed for the other polymers, and the effect is most pronounced in PEG400. Addition of PEG could either partially extract the ligand from

the DNA interior and/or prohibits the ET process, resulting in the fluorescence recovery of PSF.

Steady state anisotropy measurements: To understand the rotational restriction of the probe in the DNA interior we measure the steady state anisotropy of the PSF-DNA complex in the presence of PEGs. We measure first the anisotropy of PSF in water and in PEG solutions (figure 3.III.4a). In water it is rather low (~0.03)[38] and increases linearly with increasing PEG content in the mixture, the slope being the minimum for EG, and those in the other PEGs are comparable. Addition of PEG reduces the restriction almost linearly with concentration (figure 3.III.4b). Anisotropy is relatively high for the PSF-DNA complex presumably due to the intercalation of the probe. Rotational anisotropy of PSF in 30% EG aqueous solution is noticeably low (0.06) as compared to PSF-DNA complex (0.21) in the presence of 30% EG. The induced CD measurements have also suggested a relatively high fraction of PSF to remain bound to the DNA in the presence of 30% EG. Although the DNA native structure gets noticeably altered in the presence of EG (figure 3.III.2b), the drug-DNA binding does not reciprocally weaken. The change is rather prominent in case of the PEGs, e.g., at 30% PEG concentration the values are comparable to those of the corresponding PEG-water solutions. It can be inferred that rotational restriction of PSF is released either by DNA deformation or by PEG induced partial extraction of PSF or both.

Time resolved fluorescence measurements: Picosecond resolved fluorescence transient of PSF in water (at 580 nm) follows a single exponential decay with a time constant of 0.77 ns (figure 3.III.5a, table 3.T1), which is in comparable agreement with previously published results by Afzal et al.[38] The decay profile remains single-exponential in the presence of the PEGs, however, the time constant increases linearly. The slopes are comparable for the PEGs and are individually higher than that of EG. The increase in the viscosity of the PEG-water solution, which also changes more or less linearly in the studied concentration region, perhaps causes the observed stabilization of the excited state. The emission is severely quenched in PSF-DNA complex and the emission transient could only be fitted bi-exponentially with time constants of 0.28 ns (72%) and 1.67 ns (28%) (table 3.T1). The appearance of more than one time constant for the PSF-DNA complex is not surprising as it has been reported in the past that PSF bound to macromolecules usually shows bi-exponential fluorescence decay.[40] The ps-resolved fluorescence transients of PSF-DNA-PEG systems could only be fitted using a tri-exponential function. For example, in the presence of 10% PEG200 the time constants are τ_1~0.28 ns (57%), τ_2~1.02 ns (28%) and τ_3~1.88 ns (15%).

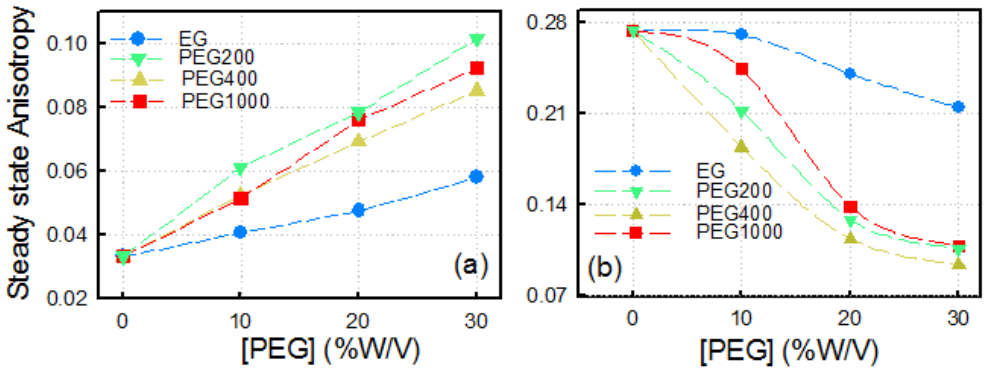


Figure 3.III.4 Steady State Anisotropy of PSF in different concentration of different crowding agents (a) in absence and (b) in presence of DNA.

A point to be noted here is that for better comparison we purposefully fix two components τ_1 and τ_2 as that of PSF in the corresponding water-PEG mixtures. The fitting after fixing these two parameters are reasonably good and does not deviate much when they are floated free. The contribution of the 0.28 ns component, which is the quenched emission originating from the DNA bound PSF, decreases with increasing PEG concentration (figure 3.III.5b). This time component could be argued to signify a rather slow ET process between PSF and DNA which perhaps operates through the medium.[12] Also, the contribution of the DNA-bound unquenched PSF decreases from 28% to 15%. Emergence of the extra component of 1.02 ns, which reciprocates the timescale of the PSF-PEG complex, clearly indicates a finite population of PSF unbound to DNA. This phenomenon is evident in all the polymers, however, with varying extent. We plot the relative weightage of τ_1 ($= \frac{a_{1,crowder}}{a_{1,water}}$) as a function of PEG concentration in figure 3.III.5b. A similar plot for τ_2 has been shown in figure 3.III.5c. For EG the change for τ_1 contribution is only modest, for PEG200 and PEG1000 it is almost linear, and for PEG400 it decreases more sharply (figure 3.III.5b). In a similar manner, the change in the τ_2 contribution is linearly increasing for EG, PEG200 and PEG1000, whereas, for PEG400 it does not change appreciably (figure 3.III.5c).

Fluorescence Up-conversion Study: To investigate the early time fluorescence response of PSF intercalated within CT-DNA, we have performed the femtosecond fluorescence up-conversion study. The early time fluorescence transients with the femtosecond time resolution are shown in figure 3.III.6. The fluorescence transient for free PSF in buffer was fitted with a sum of two exponential functions (table 3.T5). The two observed time constants are 2.6 and 760 ps.

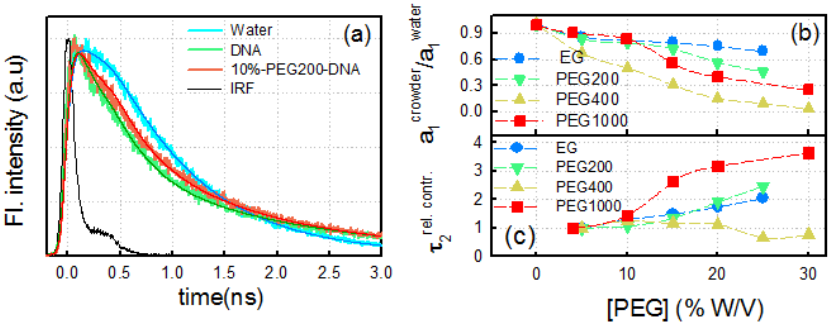


Figure 3.III.5. (a) Decay transients of PSF in different environments measured at 580 nm. (b) The relative weightage in the contribution of the faster component (fixed at 100 ps), as expressed by $a_{1,crowder}/a_{1,water}$ as a function of PEG concentrations. (c) Relative contribution of τ_2 as a function of crowder concentration.

The 2.6 ps component is unrecoverable in the TCSPC setup and is inherently related to some internal photo physics of PSF. The long-time component is in comparable agreement with the time constant obtained from TCSPC measurements (table 3.T1). However, for the fluorescent transients of PSF intercalated to DNA, we needed an extra exponential component in our fitting equation to fit the data satisfactorily. The time constants obtained are 0.8 ps (30%), 5.4 ps (50%) and 290 ps (20%). The 290 ps time constant matches well with the faster time constant obtained from the TCSPC measurements (table 3.T1). As we have already assigned the time constants obtained from the TCSPC measurement in the previous section of our discussion, here we have focused our discussion on the faster time constants only. The photo physics of PSF in sub-picosecond time domain has not been reported previously and is surely beyond the scope of this article. With that in mind we are leaving that task as a promising future work. We rather focus on apprehending the plausible excited state processes and assign the time constants obtained from our experiments accordingly. A careful analysis of the observed time constants reveals that the 2.6 ps component of PSF in buffer increases to 5.4 ps in DNA and an additional faster time constant of 0.8 ps appears. The 2.6 ps time component could be assigned to an excited state process, presumably the internal charge transfer (ICT) process along some structural deformation coordinate of PSF. The retardation of this ICT process with a time constant of 5.4 ps in case of PSF-DNA complex is probably because of the hindered structural deformation of PSF when intercalated to DNA. In this context we would like to introduce the control data of PSF emission in buffer in the presence of PEG400 as crowder. As we increase the crowder concentration up to 30% the 2.6 ps time constant gradually increases up to 6.0 ps.

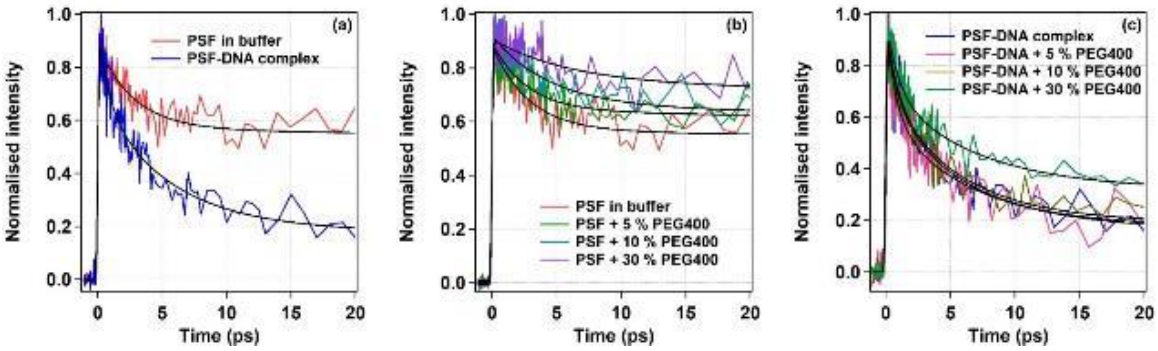


Figure 3.III.6. Femtosecond fluorescence transient of PSF (a) in buffer and intercalated in DNA, (b) in buffer in the presence of PEG400 of various concentration, and (c) within PSF-DNA complex in the presence of varying concentration of PEG400. The solid black lines are the fittings after deconvolution with a Gaussian shaped instrument response function of 200 fs.

With increase in the crowder concentration the viscosity around the free PSF molecule increases and hinders the structural deformation of the molecule. According to this argument, addition of crowder in PSF intercalated in DNA is not expected to affect the 5.4 ps time constant much, and we observe that the time constant lies within the range of 5.5-6.5 ps for 0-30% PEG400 concentration (table 3.T5), which unambiguously demonstrates the origin of this time constant. The most important part of this time-resolved experiment in fs resolution is the observation of the sub-ps time constant of 0.8 ps in PSF-DNA complex, which is completely absent in free PSF. We assign it to the ET process between PSF and DNA bases which occurs within the DNA interior through the guanine base. As the PSF molecule is bound to the DNA there is no role of its diffusion in the ET process and the observed time component is probably the pure ET rate constant. This process is due to guanine mediated hopping ET and is, therefore, independent of the local environment. So, the addition of crowder is found not to affect the ET process as evident from the unaltered 0.8–0.9 ps time component of PSF-DNA complex in the presence of 0-30% PEG400 (table 3.T5). Hence, we can conclude that the intercalation of PSF into CT-DNA alters the excited state process of PSF significantly, whereas addition of crowder in the PSF-DNA complex does not change the dynamics.

3.IV Overall comprehension

The most fascinating finding of this investigation is the PEG induced release in the quenching, which in turn is a consequence of ET, of PSF intercalated in DNA. The extent of such release is found to be dependent on the type as well as on the concentration of the polymer used. Interestingly the most efficient release is observed for an intermediate chain length (PEG400).

There could be several factors responsible for this observation. The most obvious factor could be the exclusion of PSF from the DNA interior influenced by the PEGs resulting in an overall decrease in ET. A clear analytical rationale in its support could be found from the TCSPC time resolved data (tables 3.T1-4). The emission decay of all the PSF-DNA-PEG systems could only be fitted using a triple exponential decay function, and one of the timescales matches with that in the PSF-PEG environment (τ_2). The relative weightage of this timescale increases near-linearly (except for PEG400, figure 3.III.5c) with PEG content, which unambiguously indicates the exclusion of DNA bound PSF in water-PEG environment, which in turn increases with increasing PEG content and PEG chain length. For example, in the presence of 20% EG, ~35% of PSF de-intercalates while in the presence of 20% PEG1000, the exclusion increases to as high as ~80%. It is interesting to note that the increase in a_2 occurs at the expense of both a_1 and a_3, indicating that the PEG induced release of PSF from DNA is less selective on the base-pair sequence.

The exact mechanism behind the PEG induced exclusion of PSF from DNA is not very evident at present. A possible rationale could be the PEG mediated change in the DNA structure. CD measurements have evidenced only marginal changes in the DNA structure upon the addition of PEG (except for PEG1000, figure 3.III.1a). The polymers, however, induce noticeable thermal instability in the DNA (figure 3.III.1b) coupled with significant hydration change at the DNA surface.[14, 41] On the other hand, intercalation of PSF into DNA produces modest changes in its native structure. Addition of PEG into the PSF-DNA intercalated complex perturbs the structure further (figure 3.III.2b), however, a recent study has shown that the intercalative binding mode of the ligand remains intact under such situations.[27] Exclusion of PSF from the DNA interior in the presence of PEGs could be a possible reason behind the structural alteration. The induced CD measurements also shine evidence on this de-intercalation process, however, the exclusion is not as efficient as in micelles, in which almost the entire fraction of the ligand gets excluded from the DNA.[6] It also is interesting to note that the CD signal of the DNA gets comparably altered in the presence of EG and PEG200 and 400, however, PSF bound to DNA experiences more rotational restriction in the presence of EG compared to those of PEG200 and 400.

The unusual behaviour of PEG400 over the other polymers is also intriguing. The release in quenching is most efficient in the presence of PEG400, however, the perturbation of DNA structure induced by PEG400 is comparable to those of EG and PEG200 and are less adverse than in PEG1000 (figure 3.III.1a). It is interesting also to notice the sharp change in a_1

(TCSPC measurements) with increasing PEG400, at the expense of an increase in a_3, the a_2 remains more or less constant (table 3.T3). For the other polymers, the otherwise reduction in a_1 has been compensated by an overwhelming increase in a_2 considering a steady decrease in a_3. The unusual behavior of PEG400 as an osmolyte has previously been reported for nucleic acids[42] and proteins[43, 44]; the effect had been explained on the basis of transition of PEG-biomolecular interaction from being excluded volume to direct interaction.

Quenching of ligand emission in DNA has mostly been explained in terms of electron (or hole) hopping through guanine residues as an electron-hole is localized on the ligand upon excitation and a hole is injected into the DNA base[45] or electron transfer mediated 'through space' at the DNA surface, which eventually is external environment controlled.[12] In our present investigation, the ultrafast quenched PSF emission timescale of 0.8 ps is attributed to the former process while the relatively slower component of 0.28 ns manifests the ET process mediated 'through space'. The transfer is governed by both the rate of electron injection (and so on the coupling between ligand and base) and on the transport through the base sequences. It has been concluded that the rate of electron (hole) transfer changes with distance[46] and decreases exponentially.[39] The change in the DNA structure could also bring about change in the ET process. Also increase in viscosity might restrict the motion of the base-pairs, which in turn could affect the ET.[47] Presence of crowders as well can decrease the oxygen diffusion rate and thereby could reduce the ET rate.[48] Our investigation shows that the ultrafast ET rate does not suffer considerable change in the presence of the crowders (table 3.T5) whereas the 0.28 ns component shows systematic change with the crowder concentration (figure 3.III.5b). Thus, it could be concluded that the guanine mediated ET process of PSF in CT DNA is more or less independent of the molecular crowders, but the process occurring 'through space' indeed suffers perturbation when PEGs of different chain lengths are added. It should be mentioned here that the effect could well be ligand-DNA pair specific and a generalized overview needs to be approached through further systematic investigations.

Table 3.T1. Multi-exponential fitting parameters for fluorescence decay of PSF in buffer and in DNA in presence of ethylene glycol (EG) of different concentrations.

System	a_1	τ_1 (ns)	a_2	τ_2 (ns)	a_3	τ_3 (ns)
In buffer						
0%	-	0.77	-	-	-	-
5%	-	0.85	-	-	-	-
10%	-	0.90	-	-	-	-
15%	-	0.96	-	-	-	-
20%	-	1.03	-	-	-	-
25%	-	1.10	-	-	-	-
In presence of DNA						
0%	0.72	0.28	0.28	1.67	-	-
5%	0.62	0.28	0.20	0.87	0.18	1.98
10%	0.59	0.28	0.26	0.97	0.15	2.08
15%	0.57	0.28	0.30	0.99	0.13	2.10
20%	0.54	0.28	0.35	1.06	0.11	2.21
25.00%	0.50	0.28	0.41	1.13	0.09	2.31

Table 3.T2. Multi-exponential fitting parameters for fluorescence decay of PSF in buffer and in DNA in presence of PEG200 of different concentrations.

[EG] w/v	a_1	τ_1 (ns)	a_2	τ_2 (ns)	a_3	τ_3 (ns)
In buffer						
5%	-	0.93	-	-	-	-
10%	-	1.03	-	-	-	-
15%	-	1.13	-	-	-	-
20%	-	1.22	-	-	-	-
25%	-	1.31	-	-	-	-
In presence of DNA						
5%	0.60	0.28	0.26	0.92	0.15	2.04
10%	0.57	0.28	0.28	1.02	0.15	1.88
15%	0.52	0.28	0.35	1.13	0.13	1.74
20%	0.41	0.28	0.50	1.22	0.09	1.74
25%	0.33	0.28	0.64	1.31	0.02	1.88

Table 3.T3. Multi-exponential fitting parameters for fluorescence decay of PSF in buffer and in DNA in presence of PEG400 of different concentrations.

[PEG200] w/v	**a_1**	**τ_1 (ns)**	**a_2**	**τ_2 (ns)**	**a_3**	**τ_3 (ns)**
In buffer						
5%	--	0.96	-	-	-	-
10%	--	1.06	-	-	-	-
20%	--	1.21	-	-	-	-
30%	--	1.32	-	-	-	-
In presence of DNA						
5%	0.48	0.28	0.30	0.74	0.23	1.85
10%	0.36	0.28	0.38	0.82	0.26	1.79
15%	0.22	0.28	0.35	0.87	0.43	1.58
20%	0.11	0.28	0.34	0.94	0.55	1.58
25%	0.07	0.28	0.20	0.97	0.73	1.54
30%	0.02	0.28	0.23	1.02	0.74	1.61

Table 3.T4. Multi-exponential fitting parameters for fluorescence decay of PSF in buffer and in DNA in presence of PEG1000 of different concentrations.

[PEG200] w/v	**a_1**	**τ_1 (ns)**	**a_2**	**τ_2 (ns)**	**a_3**	**τ_3 (ns)**
In buffer						
4%	-	0.98	-	-	-	-
10%	-	1.11	-	-	-	-
15%	-	1.18	-	-	-	-
20%	-	1.25	-	-	-	-
30%	-	1.37	-	-	-	-
In presence of DNA						
4%	0.65	0.28	0.22	0.81	0.13	1.83
10%	0.60	0.28	0.32	0.91	0.08	2.08
15%	0.40	0.28	0.58	1.03	0.02	2.95
20%	0.29	0.28	0.70	1.07	0.02	2.89
30%	0.18	0.28	0.80	1.13	0.02	3.62

Table 3.T5. Multi-exponential fitting parameters for ultrafast fluorescence decay of PSF in buffer and in DNA in presence of PEG 400 at different weight percentage.

[PEG]	a_1	τ_1 (ps)	a_2	τ_2 (ps)	a_3	τ_3 (ps)
In buffer						
0%	-	-	0.32	2.8	0.57	770
5%	-	-	0.26	2.7	0.63	810
10%	-	-	0.27	4.5	0.65	730
30%	-	-	0.17	6.0	0.74	885
In presence of DNA						
0%	0.30	0.80	0.50	5.4	0.20	294
5%	0.38	0.80	0.42	6.5	0.20	220
10%	0.27	0.85	0.52	5.5	0.21	370
30%	0.29	0.91	0.38	6.6	0.33	680

3.V References:

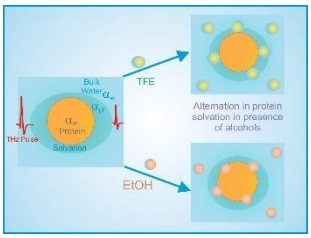
1. J. B. Chaires, *Curr Opin. Struct. Biol.*, 1998, **8**, 314-320.
2. S. Satyanarayana, J. C. Dabrowiak and J. B. Chaires, *Biochemistry*, 1993, **32**, 2573-2584.
3. L. H. Hurley and F. L. Boyd, *Trends in Pharmacological Sciences*, **9**, 402-407.
4. J. B. Chaires, *Arch. Biochem. Biophys.*, 2006, **453**, 26-31.
5. F. Westerlund, L. M. Wilhelmsson, B. Norden and P. Lincoln, *J. Am. Chem. Soc.*, 2003, **125**, 3773-3779.
6. A. Patra, S. Hazra, G. S. Kumar and R. K. Mitra, *J. Phys. Chem. B*, 2014, **118**, 901-908.
7. S. Banerjee, D. Bhowmik, P. K. Verma, R. K. Mitra, A. Sidhhanta, G. Basu and S. K. Pal, *J. Phys. Chem. B*, 2011, **115**, 14776-14783.
8. A. Patra, S. Hazra, N. Samanta, G. S. Kumar and R. K. Mitra, *Int. J. Biol. Macromol.*, 2016, **82**, 418-424.
9. R. J. Ellis and A. P. Minton, *Nature*, 2003, **425**, 27-28.
10. A. P. Minton, *J. Biol. Chem.*, 2001, **276**, 10577-10580.
11. P. Hanczyc, P. Lincoln and B. Norden, *J. Phys. Chem. B*, 2013, **117**, 2947-2954.
12. M. Tanaka, H. Iida and T. Matsumoto, *Chem. Lett.*, 2018, **47**, 62-64.
13. H. X. Zhou, G. Rivas and A. P. Minton, *Annu. Rev. Biophys.*, 2008, **37**, 375-397.
14. D. Miyoshi, H. Karimata and N. Sugimoto, *J. Am. Chem. Soc.*, 2006, **128**, 7957-7963.
15. A. Zinchenko, N. V. Berezhnoy, Q. Chen and L. Nordenskiold, *Biophys. J.*, 2018, **114**, 2326-2335.
16. A. N. Gupta and J. R. C. van der Maarel, *Macromolecules*, 2017, **50**, 1666-1671.
17. A. Singh and N. Singh, *Phys. Chem. Chem. Phys.*, 2017, **19**, 19452-19460.
18. M. Bello-Roufaï, O. Lambert and B. Pitard, *Nucleic Acids Res.*, 2007 **35**, 728-739.
19. M. C. Miller, R. Buscaglia, J. B. Chaires, A. N. Lane and J. O. Trent, *J. Am. Chem. Soc.*, 2010, **132**, 17105-17107.
20. R. Hänsel, F. Löhr, S. Foldynová-Trantírková, E. Bamberg, L. Trantírek and V. Dötsch, *Nucleic Acids Res.* , 2011, **39**, 5768-5775.
21. L. Petraccone, B. Pagano and C. Giancola, *Methods*, 2012, **57**, 76-83.
22. J. T. Lis and R. Schleif, *Nucleic Acids Res.* , 1975, **2**, 383-389.
23. C. P. Kimpton, G. Corbitt and D. J. Morris, *J. Virol. Methods*, 1990, **28**, 141-145.
24. T. Adali, A. Bentaleb, N. Elmarzugi and A. M. Hamza, *Int. J. Biol. Macromol.*, 2013 **61**, 373-378.

25. V. V. Vasilevskay, *J. Chem. Phys.*, 1995, **102**, 6595.
26. N. Biswas, M. Ichikawa, A. Datta, Y. T. Sato, M. Yanagisawa and K. Yoshikawa, *Chem. Phys. Lett.*, 2012 **539-540**, 157-162.
27. Y. S. Oh, J. H. Park, S. W. Han, S. K. Kim and Y.-A. Lee, *J. Biomol. Struct. Dyn.*, 2018, **36**, 3035-3046.
28. H. Yu, X. Gu, S.-i. Nakano, D. Miyoshi and N. Sugimoto, *J. Am. Chem. Soc.* , 2012, **134**, 20060−20069.
29. A. Arora and S. Maiti, *J. Phys. Chem. B*, 2009, **113**, 8784-8792.
30. S. Das and G. S. Kumar, *J. Mol. Struct.*, 2008, **872**, 56-63.
31. P. Das, A. Chakrabarty, A. Mallick and N. Chattopadhyay, *J. Phys. Chem. B*, 2007, **111**, 11169-11176.
32. S.-i. Nakano, L. Wu, H. Oka, H. T. Karimata, T. Kirihata, Y. Sato, S. Fujii, H. Sakai, M. Kuwahara, H. Sawai and N. Sugimoto, *Mol. Biosys.*, 2008, **4**, 579-588.
33. S.-i. Nakano, H. Karimata, T. Ohmichi, J. Kawakami and N. Sugimoto, *J. Am. Chem. Soc.*, 2004, **126**, 14330-14331.
34. S. S. Jain, M. Polak and N. V. Hud, *Nucleic Acids Res.* , 2003, **31**, 4608-4615.
35. D. Sarkar, P. Das, S. Basak and N. Chattopadhyay, *J. Phys. Chem. B*, 2008, **112**, 9243–9249.
36. S. C. Bhattacharya, P. Ray and S. P. Moulik, *J. Photochem. Photobiol. A: Chem.*, 1995, **88**, 139-145.
37. M. F. Broglia, S. G. Bertolotti, C. M. Previtali and H. A. Montejano, *J. Photochem. Photobiol. A: Chem.*, 2006, **180**, 143-149.
38. M. Afzal, S. Ghosh, S. Das and N. Chattopadhyay, *J. Phys. Chem. B*, 2016 **120**, 11492−11501.
39. B. Giese, *Annu. Rev. Biochem.*, 2002, **71**, 51-70.
40. K. Viswanathan and P. Natarajan, *J. Photochem. Photobiol. A: Chem.*, 1996, **95**, 245-253.
41. S.-i. Nakano, D. Miyoshi and N. Sugimoto, *Chem. Rev.*, 2014, **114**, 2733-2758.
42. D. B. Knowlesa, A. S. LaCroixa, N. F. Deinesa, I. Shkela and J. Record, M.T., *Proc. Natl. Acad. Sci. USA*, 2011, **108**, 12699-12704.
43. N. Samanta, D. Das Mahanta, S. Hazra, K. G.S. and R. K. Mitra, *Biochimie*, 2014, **104**, 81-89.
44. N. Samanta, T. Q. Luong, D. Das Mahanta, R. K. Mitra and M. Havenith, *Langmuir*, 2016, **32**, 831-837.
45. C. Wan, T. Fiebig, O. Schiemann, J. K. Barton and A. H. Zewail, *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 14052-14055.
46. S. O. Kelley and J. K. Barton, *Science*, 1999, **283**, 375-381.
47. M. A. O'Neil and J. K. Barton, *J. Am. Chem. Soc.*, 2004, **126**, 11471-11483.
48. W. B. Davis, C. C. Bjorklund and P. S. Cho, *J. Phys. Chem. C*, 2010, **114**, 20821-20833.

Chapter 4

4. Correlating solvation with conformational pathways of proteins in alcohol-water mixtures: A THz spectroscopic insight

Water being an active participant in most of the biophysical process, it is important to trace how protein solvation changes as its conformations evolves in presence of solutes or co-solvents. In this present study, we investigate how the secondary structures of two diverse proteins: lysozyme and β-lactoglobulin changes in the aqueous mixtures of two alcohols: ethanol and 2,2,2-trifluoroethanol (TFE) using circular dichroism measurements. We observe that these alcohols do change the secondary structures of these proteins and the changes are protein specific. Subsequently we measure the collective solvation dynamics of these two proteins in absence and in presence of the alcohols by measuring the frequency dependent absorption coefficient ($\alpha(\nu)$) in the THz (0.1-1.2 THz) frequency domain. The alcohol-water mixtures exhibit a non-ideal behaviour with a highest absorption difference ($\Delta\alpha$) obtained at $X_{alcohol}$=0.2. The protein solvation in presence of alcohols shows an oscillating behaviour in which $\Delta\alpha_{protein}$ changes with $X_{alcohol}$. Such non-monotonic solvation behaviour of proteins results from a delicate interplay between protein-water, protein-alcohol and water-alcohol associations. We attempt to correlate the various structural conformations of the proteins with the associated solvation.



4.I. Introduction

Protein (mis)folding problem has been a prime focus of research during a last few years. Protein folding/unfolding mechanisms have been investigated in great details in presence of various types of macromolecules, osmolytes, cosolutes and cosolvents.[1-5] Among the cosolvents alcohols are one of the most ubiquitous choices to study the folding-unfolding mechanism[6-10] since the effects of alcohols having different chain lengths and/or substituted atomic groups (e.g. halogens, benzene etc.) on proteins are not straightforward, and are often specific and depend on the individuality of the protein and alcohol molecules involved.[7, 9, 11-14] Alcohols have both hydrophobic moiety and hydrophilic OH-group which create a spatially inhomogeneous hydrogen bond (H-bond) environment (structure and strength) around themselves.[15] Moreover, their propensity of self-aggregation due to the hydrophobic parts creates (micro)heterogeneous sub-domain in aqueous environments.[16] Such local heterogeneity can in turn bother the solvation of proteins, which could eventually modify protein stability. Such effects have so far been described in terms of both indirect and direct protein-alcohol interactions.[9] Certain alcohols (e.g. methanol, ethanol, butanol etc.) collapse the compactness of proteins causing their unfolding.[17, 18] The possible reason behind this is a direct mechanism through which the alcohol molecules preferentially bind to the non-polar moieties of the proteins. The lower polarity of alcohol as compared to that of water tends to weaken the

intramolecular interaction inside the proteins and therefore forces them to unfold.[9] On the contrary, some other alcohols (e.g. 2,2,2-trifluoroethanol (TFE), hexafluoro isopropanol (HFIP) etc.) have a tendency to stabilize the secondary structure, especially the helical content in proteins.[7, 10, 19] This can be explained in term of the indirect mechanism where alcohol molecules disrupt the water network around proteins, or they replace the solvation water and gather around the protein surface. This causes to the strengthening of the intramolecular electrostatic and van der Waals interactions between neighbouring residues inside proteins resulting to an increase in their folded structures.[20]

In spite of enormous efforts, alcohol specific interactions and the exact mechanisms by which alcohol-water mixtures alter the conformational stability of the proteins still remains debatable. The effect of different alcohols (methanol, ethanol, TFE and HFIP) on β-lactoglobulin (βlg) and melittin proteins have been studied experimentally and observed that clustering of alcohol molecules decides the fate of protein stability.[10] Uversky et al.[11] have studied the effect of methanol, ethanol, isopropanol on βlg and found a strong correlation between the solvent dielectric constants (ϵ) and protein stability. A contrasting nature of the preferential solvation on the surface of tetrapeptide (NAc-Ser-Phe-Val-Gly-OMe) by TFE and ethanol has been concluded by Fiorini et al. using a combined experimental (diffusion NMR and intermolecular nuclear overhauser effect (NOE)) and molecular dynamics (MD) simulation study.[21] In a different study TFE induced protein aggregation and the structural intermediates were investigated using light scattering and circular dichroism (CD) spectroscopy.[22] Later the folding and self-assemble propensity of unfolded βlg have been explored experimentally in presence of different alcohols e.g. methanol, 2-propanol, tert-butanol and TFE.[23] Recently, Mohanta et al.[24] using MD simulation study observed the unfolding phenomenon of chymotrypsin inhibitor 2 (CI2) in presence of methanol, ethanol and TFE and concluded that water-alcohol (for methanol) and alcohol-alcohol (for TFE) H-bonded clusters determine the nature of perturbation in CI2.

It could be emphasized here that the nature of solvation around biomolecules, slaved by the fluctuations of transient H-bonded network,[25] plays a crucial role in their stability and activity. It is therefore of utmost importance to observe the correlation between the structural perturbation and its consequence on the solvation in proteins induced by alcohol molecules. Water H-bonded networks in presence of alcohol and proteins are largely influenced by the local electric fields, hydrophilic/hydrophobic interfaces, and the ability/disability of the molecular groups to form H-bonds. In this context Terahertz (THz) spectroscopy has emerged

as a potential label free tool to probe solute solvation and dynamical coupling of the low-frequency collective modes of water networks.[26, 27] Such modes are responsible for the relaxation in the intermolecular H-bond network, which originates from the rotational and translational diffusion of water molecules.[28] This spectral feature is also coupled with the protein intramolecular vibrations including the skeleton motions exhibiting their signature in 1-10 THz frequency region.[29] Our group has previously studied the solvation properties of bovine serum albumin (BSA) protein in presence of alcohols of varying chain lengths (ethanol, 2-propanol and tert-butyl alcohol) using THz time domain spectroscopy (TTDS).[5] We found that alcohol/water mixtures are not homogeneous over the entire concentration range. We also observed a non-monotonic oscillating behaviour of the collective solvation behaviour of BSA as a function of alcohol concentrations, which, we have argued to originate from the delicate balance between alcohol-water, alcohol-alcohol, and alcohol-protein preferential interactions. The question that remained unaddressed is whether this behaviour is protein specific or alcohol specific or both. To address this issue and for a deeper insight into the correlations of protein solvation and structural makeover, we compare the effect of two different alcohols, ethanol and its fluorinated derivative trifluoroethanol (TFE) (scheme 2.II.B4) on two proteins, namely lysozyme (Lys) and β-lactoglobulin (βlg) (scheme 2.II.B1 b & c). Lys from chicken egg white is a globular protein containing 129 amino acids while bovine βlg is a 162 amino acid protein found in animal milk. Both these alcohols offer comparable dielectric behaviour ($\epsilon_{\text{ethanol}}$ ~24 and ϵ_{TFE} ~27 at room temperature),[30] however, TFE can form extensive intermolecular H-bonds due to the presence of fluorine atoms. Moreover, the OH group of TFE has a lower electron donating ability than ethanol. This makes its affinity toward water higher than ethanol.[31] These alcohols are known to produce contrasting impacts on protein stability.[10, 14, 32] The probable reason for this dissimilarity is rooted into their specific interactions with the protein and perhaps with water also. It is therefore intriguing to investigate how these alcohols perturb water structure and whether such effects have direct consequences on the protein stability. As THz probes the fluctuations in the collective dipole moment of water, it serves as a unique and explicit platform to obtain such information.[33] Our group has put forward a substantial effort in order to understand the change in collective water dynamics in mixed solvents and in presence of biologically important molecules using THz spectroscopic technique.[4, 5, 27, 34] Here we study the intermediate conformational states of these proteins in presence of these two alcohols using CD spectroscopy. We correlate these conformational changes with their surrounding solvation as alcohol concentration (defined as the mole fraction of alcohol in the

solution, $X_{alcohol}$) changes. We found that the alcohol-induced protein folding/unfolding pathway is not necessarily a simple two state process, rather there exists certain intermediate conformational steps associated with a non-monotonic fluctuation in the solvation behaviour which results from the convolution of a subtle interplay between the various intramolecular interactions.

4. II. Materials and Methods

Lysozyme (Lys) and bovine β-lactoglobulin (βlg) were procured from Sigma Aldrich and the alcohols (ethanol and TFE) were purchased from Merck with their highest available grade and were used without any further purification. Freshly prepared 10 mM phosphate buffer (pH ~ 7) was used to prepare the protein solutions. We recorded the CD spectra with Jasco J-815 spectropolarimeter in the far-UV range (190-260 nm) under constant nitrogen flow using 1 mm path length quartz cuvette. Each spectrum was an average of three scans with a scan speed of 50 nm/min and a response time of 2 sec. Protein concentrations were kept fixed at 10 μM for the CD experiments. Throughout our discussion we represent the alcohol concentration as the mole fraction of alcohol in the solutions defined as

$$X_{alcohol} = \frac{n_{alcohol}}{n_{alcohol}+n_{water}} \quad (4.1)$$

where $n_{alcohol}$ and n_{water} are the number of moles of alcohol and water molecules respectively in the solutions. THz frequency domain spectroscopy (TFDS) measurements were carried out using a commercial set up from Toptica (Tera-scan 1550). Detail of the spectrometer is described in section 2.IV. Frequency dependent absorption coefficient was calculated as described in *2.II.D.a.* To investigate the alcohol induced changes in the protein solvation we kept both the protein concentrations fixed at 200 μM to avoid any possibility of protein aggregations in the solutions.

4. III. Results and Discussions

Secondary structure of the proteins: We measure the far-UV (190 nm-260 nm) CD signal of Lys and βlg in buffer (grey broken lines) and in presence of different concentrations of ethanol and TFE (figure 4.III.1a and b, respectively). We also estimate the relative abundance of the secondary structures and found that the α-helical content in the native state is 27% in Lys while it is ~21% in βlg. With the progressive increase in ethanol content in the mixture we observe the CD signal of Lys to first decrease up-to $X_{alcohol}$ ~ 0.3, beyond which it increases (upper panel of figure 4.III.1a).

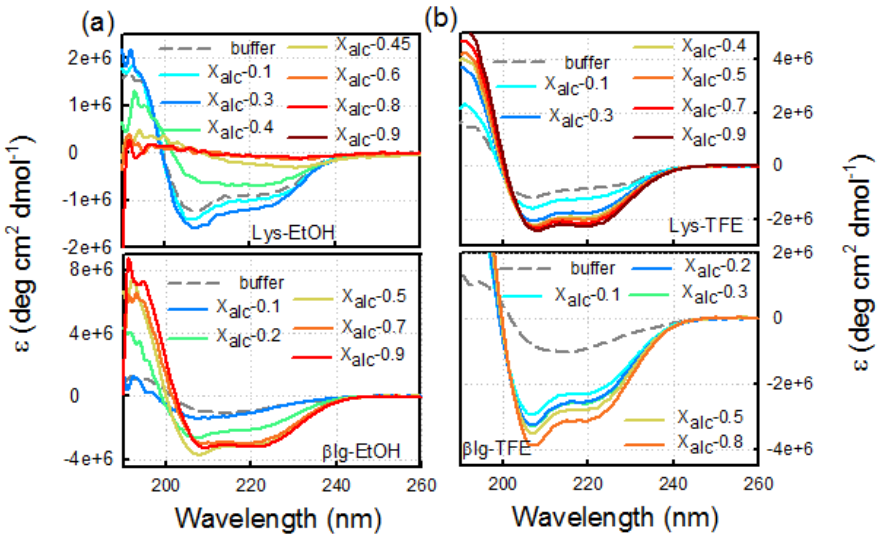


Figure 4.III.1. CD signals (in term of molar ellipticity, ε) of Lys (upper Panel) and β-lg (lower panel) in presence of (a) ethanol and (b) TFE at different alcohol mole fraction ($X_{alcohol}$). The grey broken lines are the CD signals obtained for the native proteins (in 10 mM phosphate buffer). Lys and βlg concentrations are kept fixed at 10 μM throughout the experiments.

It has been reported earlier that ethanol, at low concentrations, stabilizes the native form of Lys.[35] However, in presence of TFE the negative CD signal monotonically enhances indicating stabilization of Lys (upper panel of figure 4.III.1b). To check whether this contrasting effect is protein specific, we measure the CD signals of βlg (lower panels of figures 4.III.1a and b). We observe that both ethanol and TFE further stabilize the secondary structures of βlg. This confirms that the structural modifications induced by ethanol and TFE are protein specific, at least for βlg and Lys. It is evident from upper panel of figure 4.III.2a that the α-helix content of Lys increases mildly (27% to 33%) up-to $X_{ethanol}$ =0.3; beyond which it decreases substantially (to ~12% at $X_{ethanol}$ =0.8) at the expense of a concomitant increase in the β-structure (from ~42% to ~55%), while random coil abundance does not change much. On the other hand, addition of TFE promotes the α-helical content in Lys (up to 60% at X_{TFE}=0.7) by reducing both β-structure (up to 23% from 38%) and random coil (30% to 17%) conformations. In βlg addition of both ethanol and TFE drastically promotes the α-helical content (from ~24% to ~75%) by diminishing the other native structures; the effect being more prominent in TFE (at X_{TFE}=0.1, it increases to 52%).

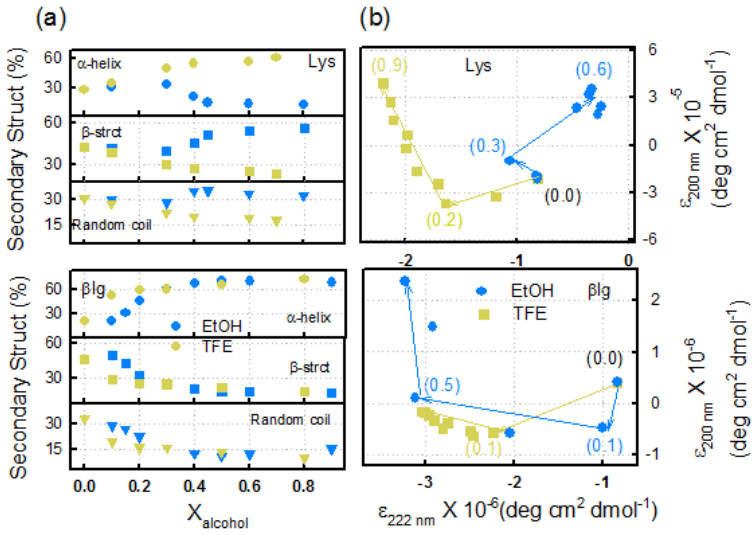


Figure 4.III.2. (a) Abundance (in percentage) of secondary structures in Lys (upper panel) and βlg (lower panel) as a function of ethanol (blue symbols) and TFE (yellow symbols) mole fraction. (b) Plot of molar ellipticity (ε) measured at 222 nm vs. 200 nm of Lys (upper panel) and βlg (lower panel) at different mole fractions of ethanol (blue symbols) and TFE (yellow symbols). Some representative mole fractions are shown within parenthesis. The arrows indicate increasing alcohol mole fraction.

The observed difference in the effect of these alcohols with proteins lies on fact that the nature of hydrophobicity and consequently their specific interaction with the proteins are distinct. Earlier studies have concluded that the presence of electronegative F atoms makes the TFE molecules an efficient proton donor, but a weak acceptor,[31] and hence TFE tends to reside at the protein surface. It therefore induces water exclusion from the protein surface and eventually increases non-covalent intra-molecular interactions between the amino acid residues.[7, 21] As a consequence, the pre-existing helix-coil equilibrium shifts towards a more structured conformation in Lys (figure 4.III.2a; upper panel). On the contrary, being hydrophobic in nature, ethanol preferentially binds to the hydrophobic pockets at the core of Lys and forces it to unfold.[35, 36] The effect is however found to be protein specific as both these alcohols induce enhanced structuring in βlg (figure 4.III.2a; Lower panel).

It is important to note here that for many proteins structural transitions mediated by co-solvents are not a merely a two state process, rather it might include certain intermediate transition steps, too.[37] Possibility of such a multistep transition can be established by constructing protein spectral phase diagrams as described in section 2.II.B.2.[38, 39] We construct such a spectral phase diagram correlating the CD signals at two different wavelengths (ε_{222nm} vs ε_{200nm}) for both of the proteins in presence of ethanol and TFE (figure 4.III.2b). The choice of the wavelengths

lies on the fact that the amplitude of the negative signal at 222 nm mainly characterises the helices whereas the positive band appearing at 195 nm manifests the β-structure content of proteins.[40] A linear segment in the phase diagram concludes a single transition pathway, at least for the secondary structural contents. On the other hand, if a nonlinear phase path is observed, it can be concluded that the protein conformational changes occur through sequential transitions.[39] For Lys, we found two individual linear segments for TFE (yellow lines) and ethanol (blue lines) (upper panel of figure 4.III.2b). It, therefore, signifies the existence of at least one intermediate transition state at $X_{TFE} \sim 0.2$, and $X_{ethanol} \sim 0.3$. These two threshold points correspond to the secondary structural content transition points in figure 4.III.2b (upper panel). For βlg (lower panel of figure 4.III.2b) we found one (and two) such intermediate states for TFE ($X_{TFE} = 0.1$) (and ethanol ($X_{ethanol} = 0.1$ and $X_{ethanol} = 0.5$). In the subsequent sections we address the question whether solvation behaviour does follow the observed multi-fold structural makeover.

Solvation of alcohols: To investigate the possible correlation between hydrophobic association (in terms of alcohol-alcohol aggregation) and non-covalent intermolecular H-bonded (water-water and alcohol-water) interactions we measure the absorption coefficient [$\alpha_{THz}(\nu)$] of water-alcohol mixtures in 0.1-1.2 THz frequency range (figure 4.III.3a and 4.III.3b). The parameter $\alpha_{THz}(\nu)$ is of importance since it manifests the water collective dynamics, or in other words manifests the number density of water exhibiting collective intermolecular solvation dynamics around the solute molecules with a frequency ν_{THz}.[29] Pure water shows high absorption (figure 4.III.3c; blue broken line) in this frequency region.[5, 41-44] We observe a "THz defect"[45, 46] i.e., $\alpha(\nu)$ decreases noticeably even upon the addition of small amount of either of the alcohols ($X_{alcohol} = 0.1$) (figure 4.III.3c). Similar THz defect has previously been observed for small co-solutes and co-solvents.[4, 5, 34, 47] Absorption of TFE/water is higher than that of ethanol/water mixture ($\alpha_{\mathrm{TFE/water}} > \alpha_{\mathrm{ethanol/water}}$) at any $X_{alcohol}$ in this frequency window. For a quantitative analysis, we plot the difference of absorption coefficient, $\Delta\alpha$ ($=\alpha_{\mathrm{alcohol-water}} - \alpha_{\mathrm{water}}$) at 1 THz (averaged from 0.95 THz to 1.05 THz) (figure 4.III.3d) as a function of $X_{alcohol}$. We found that $\Delta\alpha$ decreases non-linearly with increasing $X_{alcohol}$ in the mixtures. The extracted $\Delta\alpha$ for ethanol at 1 THz is consistent with our earlier report.[5] We found that the decrease is sharp up to $X_{alcohol}$=0.2, beyond which the effect is rather mild. Such non-linearity cannot be explained using a two-component model considering water is replaced with alcohol. Instead one needs to take into consideration the combined effects of the defect in the H-bonding network of water around the hydrophobic interfaces of alcohols, alcohol-alcohol

self-association and alcohol associated dynamic solvation shell. The reason behind the marginally higher absorption of water-TFE mixture compared to water-ethanol mixtures is due to the difference in the H-bonding pattern of the two alcohols with water.[31] In spite of the fact that TFE offers a larger affinity compared to ethanol to interact with water, we found that both these alcohols decrease α_{THz}, inferring their gross role in the perturbation of water structure. To quantify the extent of inhomogeneity in the mixtures we calculate $\Delta\alpha'$, which is a measure of the non-ideality of the solutions, defined as,[5]

$$\Delta\alpha' = \alpha_{\text{ideal}} - \alpha_{\text{expt}} \tag{4.2}$$

where α_{ideal}, the algebraic sum of the contributions originates from each individual component of the mixtures (water and alcohol) assuming they do not interact, is defined as

$$\alpha_{\text{ideal}} = \frac{\rho}{\sum_i X_i \rho_i} \sum_i X_i \alpha_i \tag{4.3}$$

ρ is the measured density of the alcohol/water binary mixtures; X_i is the mole fraction, ρ_i is the density and α_i is the measured absorption coefficient at 1 THz for the i^{th} species (pure water or pure alcohol) in the solution. We plot $\Delta\alpha'$ against $X_{alcohol}$ for both ethanol and TFE (figure 4.III.3e) and observe a nonlinear bell shaped curve.[5, 42] The maximum deviation is found at $X_{alcohol} = 0.2$ for both the alcohols. In this concentration region, the alcohol molecules are fully solvated by water clusters,[48] which might increase the number of dangling water around the alkyl (CH_3/CF_3) groups[49] and thereby induces the maximum heterogeneity and so also the observed maximum increase in $\Delta\alpha$. Previously we have observed heterogeneity in water-DME solutions.[34] Existence of such inhomogeneity clearly suggests that the notion of non-interacting individual contribution concept does not suffice to explain the $\alpha(\nu_{THz})$ behaviour, rather one needs to consider an additional intermolecular interaction (solvation),

$$\alpha_{\text{alc/water}} = \phi_w \alpha_w + \phi_{alc} \alpha_{alc} + \phi_{ah} \alpha_{ah} \tag{4.4}$$

where α_w, α_{alc}, and α_{ah} are the absorption co-efficients of pure water, pure alcohol and water in the solvation shell of alcohol respectively. ϕ_i are the corresponding volume fractions with $\phi_w + \phi_{alc} + \phi_{ah} = 1$.

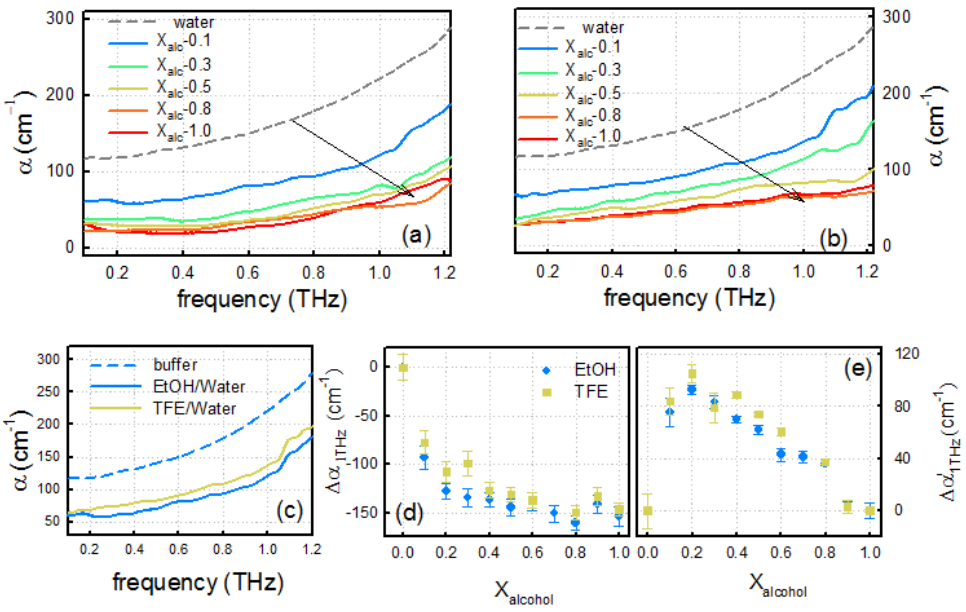


Figure 4.III.3 Absorption co-efficient of water/alcohol binary mixture as a function of frequency for (a) ethanol (b) TFE. (c) Representative profiles of absorption co-efficient, α as a function of THz frequency for buffer (broken line) and alcohol/water mixtures with $X_{alcohol}$ =0.1. (d) $\Delta\alpha$ ($= \alpha_{alcohol-water} - \alpha_{water}$) and (e) $\Delta\alpha'(= \alpha_{ideal} - \alpha_{expt})$ (see equation 4.3 in the text) measured at 1 THz as a function $X_{alcohol}$. We observe a sharp increase in the non-ideality ($\Delta\alpha'$) at $X_{alcohol}$=0.2 for both the alcohols.

Solvation of proteins: Next, we measure the solvation $\alpha(\nu_{THz})$ of proteins in presence of alcohol/water mixtures. Both the proteins show lower absorption than that of pure water (figure 4.III.4a) as highly absorbing water molecules are replaced by merely absorbing protein molecules.[5, 43] Yamamoto et al.[50] have previously carried out a detailed broadband dielectric spectroscopy (in sub-GHz to THz region) of lysozyme at different solvation level while Vondracek et al.[26] have established the collective solvation behaviour of βlg using intense THz radiation in the 2.4-2.7 THz region. The latter study has concluded that one needs to consider a three-component model to justify the change in $\alpha(\nu_{THz})$ profile for protein solutions as one needs to include an additional protein solvation term into concern. We herein adopt a simple model in which the absorption of protein aqueous solutions can be written in the form,

$$\alpha\text{protein/water} = (1 - \phi_p - \phi_{ph})\alpha_w + \phi_p\alpha_p + \phi_{ph}\alpha_{ph} \qquad (4.5)$$

Pure protein has a very low THz absorption as compared to that of water ($\alpha_p \sim 0$), it can therefore be considered as a spherical cavity in a dielectric continuum.

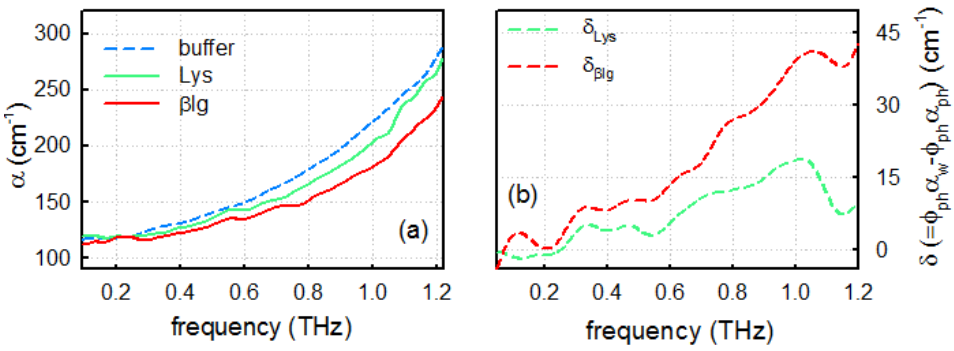


Figure 4.III.4 (a) Absorption coefficient (α) of buffer, Lys, and βlg as a function of THz frequency. (b) Change in the THz absorption, $\delta\ (= \phi_{ph}\alpha_w - \phi_{ph}\alpha_{ph})$ of two proteins due to the conversion of bulk water to protein solvation water.

Equation 4.5 thus can be rewritten as

$$\alpha\text{protein/water} = \alpha_w - \phi_p\alpha_w + \phi_{ph}\alpha_{ph} - \phi_{ph}\alpha_w \qquad (4.6)$$

The change in the THz absorption, $\delta\ (= \phi_{ph}\alpha_w - \phi_{ph}\alpha_{ph})$, due to the conversion of bulk water to protein solvation water can then be expressed as,

$$\delta = (\phi_{ph}\alpha_w - \phi_{ph}\alpha_{ph}) = (1 - \phi_p)\alpha_w - \alpha\text{protein/water} \qquad (4.7)$$

We calculate the total volume of the proteins in the solutions using $V_p = n_p \frac{4}{3}\pi r_p^3$, where r_p and n_p are the radius and the number of protein molecules in the solution, respectively, and consequently the volume fraction ϕ_p. It should be noted here that the model is an approximate one and includes the gross assumptions that the shape of the protein molecules are spherical, protein molecules are fully hydrated by water and the solvation shells of the proteins do not overlap (the concentration of protein has been kept low to avoid this). Following equation 4.7, we calculate δ for both the proteins as a function of frequency (figure 4.III. 4b). We found a definite change in THz absorption (δ) due to the conversion of bulk water to protein solvation water. However, it is not the absolute change in the α value, rather it is the difference in the product $\alpha\phi$, which is not trivial to deconvolute, and therefore the change could be due to either in the change in α or in ϕ or both. A previous simulation study by Heyden et al.[51] using biologically relevant solutes have shown that α changes with solvation layer (and thereby with ϕ also) and also depends upon the solute molecule involved. We found that the change in solvation is larger for βlg than that for Lys in their native forms (figure 4.III.4b). A plausible

rationale for this could be a larger size of the former and therefore a higher ϕ_{ph}. Moreover, a difference in the secondary structure content and the higher SASA value of βlg could also account for the observed difference.

Addition of alcohols in the protein solutions decreases the $\alpha(\nu)$ profile further. Some representative examples are depicted in the figure 4.III.5. For such a tri-component system we first try to identify whether the presence of proteins affects the independent solvation of the alcohols and to quantify this we calculate the explicit change in alcohol solvation: $\Delta\alpha_{alc}$ $(= \alpha_{(alcohol/protein/water)} - \alpha_{(protein/water)})$ at 1 THz (figure 4.III.6a and b). We compare these changes with those in absence of the proteins: $\Delta\alpha_{alc}(= \alpha_{(alcohol/water)} - \alpha_{water})$. We found that $\Delta\alpha_{alc}$ values are comparable in absence and in presence of protein. This observation leads us to a-priori conclude that the interaction of alcohols with protein involves mostly their hydrophobic milieus and the solvation of alcohols, which is preferentially through the OH bonds, remains almost intact in presence of the proteins. It could also be argued here that the alcohol molecules are only feebly hydrated and during their interaction with the protein surface that number does not change appreciably. In an earlier study we found a similar protein independent solvation of monovalent cations.[52]

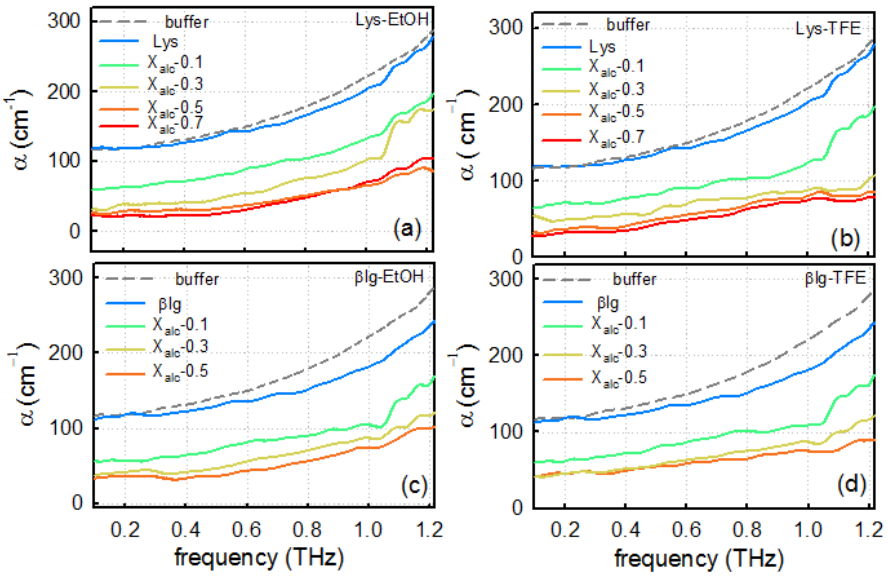


Figure 4.III.5 Some representative profile of the absorption coefficient as a function of frequency for Lysozyme in (a) ethanol/ water mixture, (b) TFE/ water mixture and βlg in (c) ethanol/water mixture, (d) TFE/ water mixture.

Solvation behaviour of proteins in presence of alcohols: We now focus on the protein solvation and see whether it is altered in presence of the alcohols. We calculate $\Delta\alpha_p$ defined as, $\Delta\alpha_p = \alpha(\text{protein/alcohol/water}) - \alpha_{(\text{alcohol/water})}$ and plot it against $X_{alcohol}$ (figure 4.III.6c and d). As evident from the definition, $\Delta\alpha_p$ explicitly identifies the alcohol induced changes in the solvation dynamics of the protein molecules. We observe an interesting nonlinear, more specifically, an oscillatory pattern of $\Delta\alpha_p$ (as it changes with $X_{alcohol}$) for both the proteins. It shows a sharp peak at $X_{alcohol} = 0.2$ and at a higher alcohol concentration a rather feeble periodic increase and decrease is observed. Previously, we have observed such patterned oscillating solvation in BSA protein-alcohol solutions.[5] Here we further observe that the change in solvation is prominent in ethanol in comparison to that in TFE for both the proteins. The observed marked change in the solvation at $X_{ethanol} = 0.2$ in Lys (figure 4.III.6c) can be correlated with a pronounced decrease in the α-helical content with a concomitant increase in

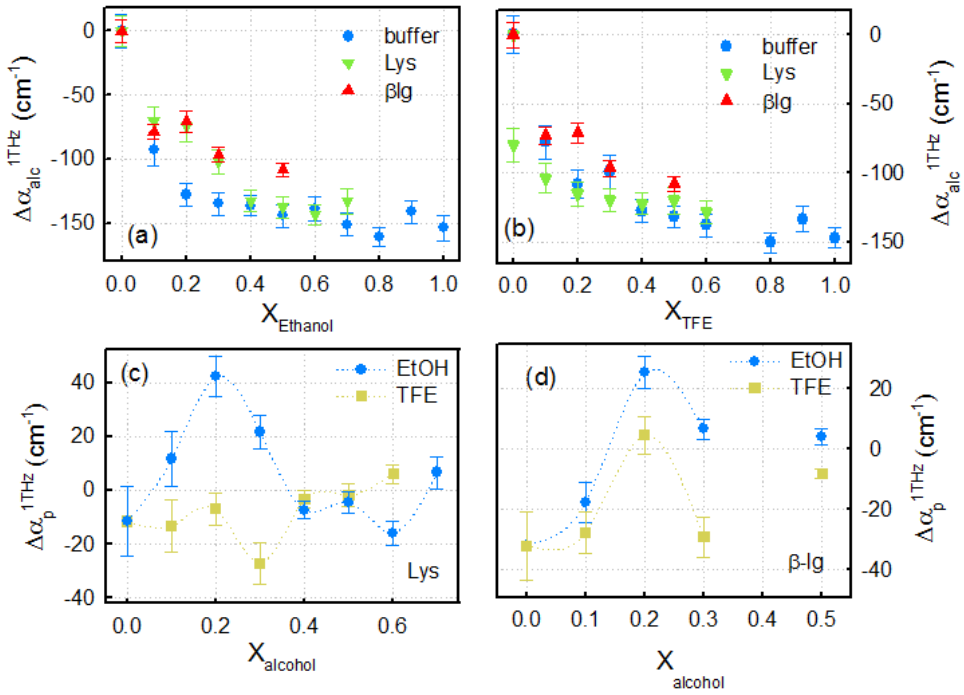


Figure 4.III.6 (a and b) Change in solvation of alcohol in presence of proteins defined as: $\Delta\alpha_{alc} = \alpha_{(protein-alchol-water)} - \alpha_{(protein-water)}$ calculated at 1 THz for ethanol and TFE as a function of $X_{alcohol}$. (c and d) Change in protein solvation in presence of alcohols defined as $\Delta\alpha_p$ ($= \alpha_{protein-water-alcohol} - \alpha_{water-alcohol}$) measured at 1 THz plotted as a function of $X_{alcohol}$. Concentration of Lys and βlg were kept fixed at 200 μM. The dotted lines are used as guide for the visuals. $\Delta\alpha_p$ shows a fluctuating behaviour with $X_{alcohol}$.

the β-structures and random coil conformations, which in turn increases the nonpolar solvent accessible surface area (SASA). It, therefore, could be concluded at this point that the observed oscillations in protein dynamics is a direct consequence of protein structural changes induced by alcohols.

Correlation of protein structure and solvation: In order to underline whether protein solvation change (as evidenced in figures 4.III.6c and 4.III.6d) and protein structures are correlated we construct the phase diagrams as mentioned earlier. The CD measurements have already established multi-step native to unfolding transitions for both the proteins in presence of the alcohols (figure 4.III.2b). To correlate structural changes with solvation we plot the changes in molar ellipticity at 222 nm ($\Delta\varepsilon_{222\mathrm{nm}} = \varepsilon_{222\mathrm{nm}}^{\mathrm{protein-alc-buf}} - \varepsilon_{222\mathrm{nm}}^{\mathrm{protein-buf}}$, as obtained from CD measurements) with the difference in the absorption coefficient at 1 THz ($\Delta\alpha_{1\mathrm{THz}} = \alpha_{1\mathrm{THz}}^{\mathrm{protein-alc-buf}} - \alpha_{1\mathrm{THz}}^{\mathrm{alc-buf}}$, as obtained from THz measurements) (figure 4.III.7). Each point in this diagram therefore represents a particular phase of the protein in terms of structure and solvation. The arrows indicate the traces as alcohol mole fraction is increased (denoted in the respective parenthesis). The phase diagrams appear to be complex and cannot be expressed with single lines. For both the protein-alcohol pairs, a sharp transition occurs at $X_{alcohol}$ = 0.2, which, as evident from the CD measurements, is a critical concentration beyond which the helix-coil equilibrium gets perturbed (figure 4.III.2b). At this threshold alcohol concentration, the alcohol molecules either start interacting directly with the hydrophobic core of the proteins (for ethanol) or accumulate around the protein surface (for TFE). Correspondingly the protein solvation (α) is lifted (see figure 4.III.6). Interestingly, in the TFE systems, we find a dip (at $X_{alcohol}$=0.3) followed by a rise in the phase profile, while such changes are not reflected in the CD phase diagrams (figure 4.III.2b) as well as in the secondary structural content (figure 4.III.2a). The mild oscillations beyond the sharp transition point also does not correlate any gross secondary structural fluctuations. While the sharp change at $X_{alcohol}$ could unambiguously be correlated with the protein structural perturbation, the oscillations are mostly due to the subtle change in the (de)solvation of the protein surface without any gross change in the protein structures. Since we explicitly measure the protein solvation here, any change in $\Delta\alpha$ due to the alcohol solvation is excluded. This leads to infer that the protein solvation behaviour at higher $X_{alcohol}$ is mostly due to the inclusion-exclusion of water in the protein surface as the concentration of alcohol in the bulk changes, which turns out to be applicable in both the proteins, however, the extent of such process is protein specific.

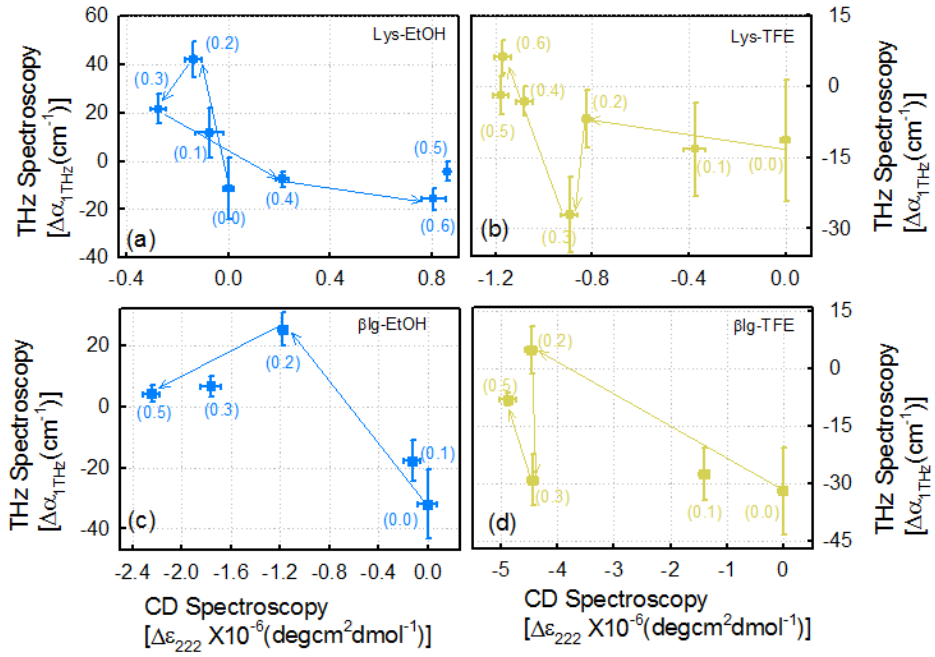


Figure 4.III.7 Protein solvation (assigned to be the difference in absorption coefficient at 1 THz, $\Delta\alpha_{1THz} = (\alpha_{1THz}^{protein-alc-buf} - \alpha_{1THz}^{alc-buf})$ is plotted as a function of protein structure (defined as the difference in molar ellipticity at 222 nm, $\Delta\varepsilon_{222nm=}$ $(\varepsilon_{222nm}^{protein-alc-buf} - \varepsilon_{222nm}^{protein-buf})$ obtained from CD measurements for different protein-alcohol pairs. The values given in the parenthesis in the figures correspond to a particular alcohol mole fraction ($X_{alcohol}$). The arrows trace the path of increasing $X_{alcohol}$. It is evident that the transitions do follow a non-linear trend.

4.IV Conclusions

In this study we intended to correlate the structural makeover of proteins in presence of alcohols with the associated changes in its solvation dynamics. CD measurement reveals that the effects of the alcohols on protein conformational stability are protein specific in nature. These effects could be ascribed in terms of direct and indirect interactions of the alcohols with the protein surfaces. We found that multiple steps are involved in these structural makeovers. We explore the correlation of the protein's structural stability associated with its solvation by probing the non-covalent intermolecular modes of water network using THz spectroscopy. We observe that the protein solvation changes in a non-monotonic fashion and bares a one-to-one correlation with the structural makeover of the protein in the low $X_{alcohol}$ region as evidenced from CD measurements. In the higher alcohol content the solvation change does occurs without substantial protein structural modifications. The overall change is governed by a delicate interplay between the different interactions present in the tri-component system.

4.V References:

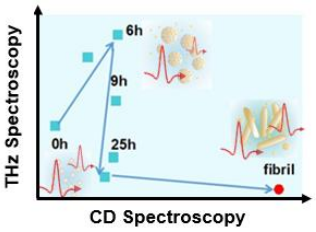
1. D. R. Canchi and A. E. García, *Annu. Rev. Phys. Chem.*, 2013, **64**, 273-293.
2. S. N. Timasheff, *Annu. Rev. Biophys. Biomol. Struct.*, 1993, **22**, 67–97.
3. A. Wang and D. W. Bolen, *Biochemistry*, 1997, **36**, 9101–9108.
4. D. D. Mahanta, N. Samanta and R. K. Mitra, *J. Phys. Chem. B*, 2017, **121**, 7777-7785.
5. D. K. Das, D. D. Mahanta and R. K. Mitra, *Chem. Phys. Chem.*, 2017, **18**, 1-7.
6. K. Shiraki, K. Nishikawa and Y. Goto, *J. Mol. Biol.*, 1995, **245**, 180–194.
7. D. Roccatano, G. Colombo, M. Fioroni and A. E. Mark, *Proc. Nat. Acad. Sci.*, 2002, **99**, 12179-12184.
8. T. Nakano and A. L. Fink, *J. Biol. Chem.*, 1990, **265**, 12356-12362.
9. M. Buck, *Quart. Rev. Biophys.*, 1998, **31**, 297-355.
10. D. P. Hong, M. Hoshino, R. Kuboi and Y. Goto, *J. Am. Chem. Soc.*, 1999, **121**, 8427-8433.
11. V. N. Uversky, N. V. Narizhneva, S. O. Kirschstein, S. Winter and G. Löber, *Folding & Design*, 1997, **2**, 163-172.
12. S. M. Singh, J. Cabello-Villegas, R. L. Hutchings and K. M. G. Mallela, *Proteins*, 2010, **78**, 2625-2637.
13. Q. Shao, Y. Fan, L. Yang and Y. Q. Gao, *J. Chem. Phys.*, 2012, **136**, 115101.
14. M. Perham, J. Liao and P. Wittung-Stafshede, *Biochemistry*, 2006, **45**, 7740-7749.
15. A. Ghoufi, F. Artzner and P. Malfreyt, *J. Phys. Chem. B*, 2016, **120**, 793–802.
16. L. Weirich and C. Merten, *Phys. Chem. Chem. Phys.*, 2019, **21**, 13494-13503.
17. Y. Yang, S. Barker, M. J. Chen and K. H. Mayo, *J. Biol. Chem.*, 1993, **268**, 9223-9229.
18. M. Buck, S. E. Radford and C. M. Dobson, *Biochemistry*, 1993, **32**, 669-678
19. R. M. Culik, R. M. Abaskharon, I. M. Pazos and F. Gai, *J. Phys. Chem. B*, 2014, **118**, 11455−11461.
20. P. D. Thomas and K. A. Dill, *Protein Sci.*, 1993, **2**, 2050-2065.
21. M. Fioroni, M. D. Diaz, K. Burger and S. Berger, *J. Am. Chem. Soc.*, 2002, **124**, 7737-7744.
22. V. L. Anderson and W. W. Webb, *Biophysical J.*, 2012, **102**, 897–906.
23. S. Maity, S. Sardar, Pal, S., H. Parvej, J. Chakraborty and U. C. Halder, *RSC Adv.*, 2016, **6**, 74409-74417.
24. D. Mohanta and M. Jana, *Phys. Chem. Chem. Phys.*, 2018, **20**, 9886-9896.
25. D. Laage, T. Elsaesser and H. J. T., *Chem. Rev.*, 2017, **117**, 10694−10725.
26. H. Vondracek, J. Dielmann-Gessner, W. Lubitz, M. Knipp and M. Havenith, *J. Chem. Phys.*, 2014, **141**, 22D534.
27. N. Samanta, D. D. Mahanta and R. K. Mitra, *Phys. Chem. Chem. Phys.*, 2014, **16**, 23308-23315.
28. H. Elgabarty, T. Kampfrath, D. J. Bonthuis, V. Balos, N. K. Kaliannan, P. Loche, R. R. Netz, M. Wolf, T. D. Kühne and M. Sajadi, *Science Adv.*, 2020, **6**, eaay 7074.
29. Y. Xu and M. Havenith, *J. Chem. Phys.*, 2015, **143**, 70901.
30. C. Wohlfarth, *Static Dielectric Constants of Pure Liquids and Binary Liquid Mixtures*, Springer2015.
31. M. Matsugamia, R. Yamamotob, T. Kumaib, M. Tanakab, T. Umeckyb and T. Takamukub, *J. Mol. Liq.*, 2016, **217**, 3-11.
32. B. Lai, A. Cao and L. Lai, *Biochimica et Biophysica Acta*, 2000, **1543**, 115-122.
33. S. Ebbinghaus, S. J. Kim, M. Heyden, X. Yu, U. Heugen, M. Gruebele, D. M. Leitner and M. Havenith, *Proc. Nat. Acad. Sci. USA*, 2007, **104**, 20749–20752.
34. D. D. Mahanta, A. Patra, N. Samanta, T. Q. Luong, B. Mukherjee and R. K. Mitra, *J. Chem. Phys.*, 2016, **145**, 164501.
35. S. Chattoraj, A. K. Mandal and K. Bhattacharyya, *J. Chem. Phys.*, 2014, **140**, 115105.
36. R. Ghosh, S. Roy and B. Bagchi, *J. Phys. Chem. B*, 2013, **117**, 15625−15638.
37. U. M. Yasin, P. Sashi and A. K. Bhuyan, *J. Phys. Chem. B*, 2014, **118**, 6662−6669.
38. I. M. Kuznetsova, K. K. Turoverov and V. N. Uversky, *J. Proteome Res.*, 2004, **3**, 485-494.
39. D. C. Jenkins, I. D. Sylvester and J. T. Pinheiro, *FEBS J.*, 2008, **275**, 1323–1335.
40. N. Greenfield and G. D. Fasman, *Biochemistry*, 1969, **8**, 4108-4116.
41. J. Xu, K. W. Plaxco and S. J. Allen, *J. Phys. Chem. B*, 2006, **110**, 24255-24259.

42. R. Li, C. D'Agostino, J. McGregor, M. D. Mantle, J. A. Zeitler and L. F. Gladden, *J. Phys. Chem. B*, 2014, **118**, 10156−10166.
43. U. Heugen, G. Schwaab, E. Bründermann, M. Heyden, X. Yu, D. M. Leitner and M. Havenith, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 12301-12306.
44. M. C. Beard, G. M. Turner and C. A. Schmuttenmaer, *J. Phys. Chem. B*, 2002, **106**, 7146-7159.
45. D. M. Leitner, M. Gruebele and M. Havenith, *HFSP J.*, 2008, **2**, 314–323.
46. B. Born, S. J. Kim, S. Ebbinghaus, M. Gruebelebc and M. Havenith, *Faraday Discuss.*, 2009, **141**, 161–173.
47. D. D. Mahanta, S. I. Islam, S. Choudhury, D. K. Das, R. K. Mitra and A. Barman, *J. Mol. Liq.*, 2019, **290**, 111194.
48. S. Y. Noskov, G. Lamoureux and B. Roux, *J. Phys. Chem. B*, 2005, **109**, 6705-6713.
49. J. R. Robalo, L. M. Streacker, D. M. Oliveira, P. Imhof, D. Ben-Amotz and A. V. Verde, *J. Am. Chem. Soc.*, 2019, **141**, 15856–15868.
50. N. Yamamoto, K. Ohta, A. Tamura and K. Tominaga, *J. Phys. Chem. B*, 2016, **120**, 4743–4755.
51. M. Heyden, E. Bründermann, U. Heugen, G. Niehues, D. M. Leitner and M. Havenith, *J. Am. Chem. Soc.*, 2008, **130**, 5773–5779.
52. D. D. Mahanta, N. Samanta and R. K. Mitra, *J. Mol. Liq.*, 2016, **215**, 197-203.

Chapter 5

5. Alteration of water network during fibrillation pathway

Amyloid fibril, occurring owning to protein aggregation in the living body causes many neurodegenerative diseases. The reason behind protein aggregation is still not very obvious. It is assumed that protein-protein interaction is a key factor in the aggregation process; however it is not explored much whether water plays any role in the fibrillation process. Here, we, using terahertz (THz) spectroscopy first establish the alteration of the collective hydration of water during the fibrillation process (native →intermediate →fibril) of a model protein, bovine serum albumin and establish the correlation between structural modification and associated hydration. As an extension of the above findings, we use a more realistic protein human insulin and observe that water releases from the insulin interior during protein aggregation. We also use two crowders, sucrose and its monomer, glucose to monitor the effect of crowders on the fibrillation process and how water release amount depends on the crowders.



5.I Introduction

Molecular self-aggregation is an important biological phenomenon which induces the formation of complex biological structures.[1] One of the most important examples of self-aggregation in living body is protein aggregation, specifically the formation of amyloid fibrils[2] as it could lead to neurodegenerative diseases like Parkinson, Alzheimer, amyotrophic lateral sclerosis (ALS).[2-4] Fibrils are enriched with β sheets, and are stabilized by dense network of hydrogen bonds.[5] They are thermodynamically the most stable complex compared to any other form of proteins.[6] There exist several hypotheses explaining the fibril formation mechanism. One of these proposes that fibril grows up inside multivesicular bodies due to the formation of prefibrillar oligomers and intracellular amyloids releases to extracellular space.[7] Upon uncontrollable growth of these pre-fibrils, living cells suffer cell death and intracellular amyloids are released to extracellular environment. The process essentially involves three steps: lag phase (at which amyloid materials do not increase detectably), elongation phase (where non amyloid like oligomers convert into amyloid ones) and asymptotic phase (where no further change in amyloidosis is observed).[8] It is important to detect the lag phase as the process can be inverted from this phase.[9]

In spite of enormous efforts, a molecular level understanding on the pathway, by which native protein converts to amyloid fibrillar form, and also what factors govern these transitions still remain an open concern. Various studies suggest that formation of amyloid fibrils from native protein takes place via conformational change of proteins[10] and protein-protein interaction plays a key role in this regard.[11] It is still less explored, at least experimentally, whether protein

hydration plays a pivotal role on protein aggregation or to put it in another words, whether protein hydration follows the traces of protein aggregation. An answer to this could lead to an avenue for early identification of protein aggregation process. The role of water in protein aggregation has theoretically been investigated by correlating hydration free energy of a protein with the protein aggregation propensity in aqueous solution and it was inferred that it is the water around the protein rather than the protein itself which controls the extent to which the protein is aggregation prone.[12] It has also been theoretically revealed that water releases during the protein aggregation from protein interior.[13, 14] However, the experimental verification of these claims has been sparsely attempted. Fichou et al, using neutron scattering techniques, investigated the dynamic behaviour of hydration water around tau protein[15] and it was found to be more mobile in the fibrillar state compared to that in the monomeric form. In a subsequent work, dynamical heterogeneity of water around an intrinsically disordered protein, α-synuclein was investigated using femtosecond fluorescence up-conversion and molecular dynamics simulation studies.[16]

In most of these earlier studies on protein aggregation the focus was made from the protein's perspective only while less attention was paid on the concomitant effect on the associated hydration. Protein hydration can unambiguously be traced using terahertz (THz) spectroscopy, which has emerged as a powerful tool to probe the dynamical coupling of the collective modes of water network and provides explicit information on hydration dynamics as it directly probes the water dipole orientations.[17, 18] In a previous study we established that reversible and irreversible thermal denaturation of serum albumin could successfully be trailed by monitoring the alterations in collective hydration network using THz spectroscopy[19]. Here, we use THz spectroscopy to address: (i) whether the collective hydration of water is altered during the fibrillation process (native →intermediate →fibril), (ii) the release of water molecules from protein interior during amyloid formation and its modification in presence of crowders which modulates protein aggregation, (iii) the physical behaviour of remaining trapped water or hydration layer of protein during aggregation in absence and in presence of crowders. While a model protein, bovine serum albumin (BSA) is used to investigate the former concern, a more realistic protein, human insulin (Ins) is preferred to address the latter two problems. Since Ins plays a key role in controlling the blood sugar level and hence widely used in diabetes II treatment.[20] Thus understanding the Ins fibrillation process could also assist preventing the aggregation for long term storage.

5.II Material & Method:

All the chemicals: BSA, Ins, glucose, sucrose, Thioflavin-T (ThT) and Potassium Chloride (KCl) were procured from Sigma Aldrich with their highest purity (> 99%) and used without further purification. Hydroxyl chloride (HCl) (strength 33%), di-Sodium hydrogen phosphate (Na_2HPO_4) and Sodium phosphate monobasic (NaH_2PO_4) were the product of Merck. Na_2HPO_4 and NaH_2PO_4 were used to prepare phosphate buffer (PBS) of pH 7.4.

BSA oligomers and fibril formation: BSA of 200 µM was denatured by 60% (V/V) EtOH and then incubated (Innova 42, Brunswick Scientific) at 62 °C for 6 hours. It was then kept at 20 °C for 24 hours to form matured fibrils.[21-23] Prior to every measurement the fibril was centrifuged for 10 minutes and then the supernatant liquid was used for the measurements. For oligomers, 400 µM BSA was incubated at 62 °C for different times. Freshly prepared 10 mM phosphate buffer (PBS) (using Na_2HPO_4 and NaH_2PO_4) with pH~7.4 was used as the solvent throughout the experiment.

Ins fibril formation: 150 µM Ins was dissolved in a 200 mM buffer (KCl and HCl) of pH ~1.5 to unfold the protein first. The protein solution was kept at 4 °C for 12 hours to stabilize and ensure unfolding. The protein solution was then incubated (Innova 42, Brunswick Scientific) at 70 °C for different time to prepare Ins oligomers, proto-fibrils and matured fibrils. To prepare the Ins fibril in presence of crowders, we dissolve the crowders in the protein solutions and started incubation after 30 mins.

Dynamic light scattering (DLS) measurements were carried out in Nano S Malvern instrument with 632.8 nm laser and the photons are collected at 173° scattering angle. Circular dichroism (CD) measurements were performed in Jasco 815 spectrophotometer. The secondary structural analysis was performed in CDNN software in the wavelength range between 190-260 nm. Fluorolog3 (Horiba, Jobin, Yvon, USA) fluorimeter was used to collect the steady state fluorescence spectra. We used time correlated single photon counting (TCSPC) form Edinberg Instrument (Life Spec-II, U.K.) to measure fluorescence lifetime (375 nm excitation with ~80 ps instrument response function, IRF). Morphologies of different BSA structures were studied using atomic force microscopy (AFM) in di INNOVA microscope in tapping mode. Prepared samples were drop casted either on thoroughly cleaned silicon wafer substrate or freshly cleaved mica which were then kept inside desiccator to facilitate vacuum environment for 12 hours for vaporisation of water before the measurement of AFM images. The size of the various structures, as evident in AFM measurements, are estimated using Gwydion and Mathematica

9 software. Terahertz (THz) frequency domain spectroscopy (TFDS) measurements were performed in Toptica (Tera-scan 1550). Details about this spectroscopy can be found in section 2.IV. Far infrared (FIR) spectra in the range of 50-570 cm^{-1} (1.5-17 THz) were collected in a Vertex 70V (Bruker, Germany) Fourier transform infrared (FTIR) spectrometer equipped with a DLaTGS detector as described in detail in section 2.IV. Frequency dependent absorption coefficient, α(ν) is calculated using the method described in section 2.II.D.

5.III Results and discussions:

5.III.A Alteration of water absorption in THz region traces the onset of fibrillation in proteins

Dynamic light scattering measurement: Hydrodynamic diameter (measured from dynamic light scattering (DLS)) of native BSA (5μM) (figure 5.III.A.1a, grey dotted line) is found to be ~ 6 nm. Increase in the BSA concentration to 400μM does not change the diameter much with only subtle increase in the FWHM. Previous studies have confirmed that BSA, in the concentration range of 30-500 μM, does not form any aggregation.[24] Incubation of 400 μM BSA at 62 °C for different times produces solutions containing larger structures, with average diameter of ~ 35 nm. These structures could be identified as soluble oligomers assembled via hydrophobic interaction.[25] The dimension of these oligomers does not change appreciably with increasing incubation time (figure 5.III.A.1). It is to note here that in addition to these ~35 nm globules, we also observe the signature of the protein monomers (peak ~6 nm), the intensity of which decreases with increasing incubation time. It can be inferred that initially BSA monomers do coexist, however, after 25 hrs of incubation almost all the protein molecules form the ~ 35 nm oligomers. The red broken line (figure 5.III.A.1a) represents the scattering profile of the fibril. We observe an intense and sharp peak at ~ 500 nm.

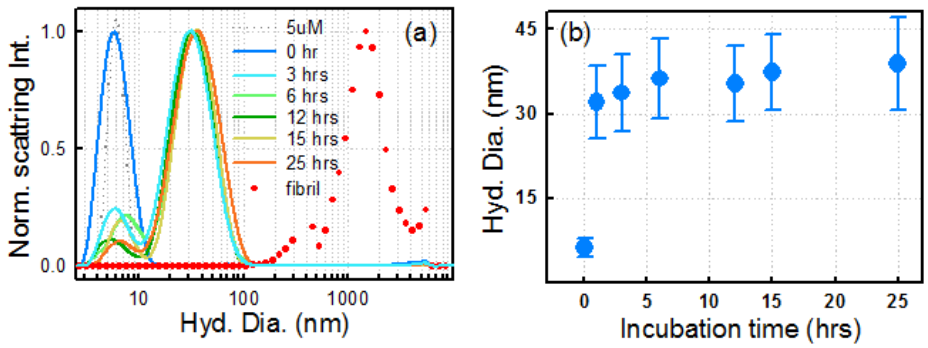


Figure 5.III.A.1 Normalised scattering profile of BSA of different condition. Red dotted symbol indicates the scattering profile of fibril. Hydrodynamic diameter after different incubation.

AFM measurement: To make further insight into the structures we measure the AFM images of the solutions. We do not identify any large structure at 100μM BSA. AFM images of 6 hrs incubated 400 μM BSA (figure 5.III.A.2; left panel) show the presence of small-sized more or less spherical oligomers with diameter 45±6 nm (figure 5.III.A.2; left panel). AFM images further infer that the micro-globules (oligomers)[26] do exist even after 25 hrs of incubation. We present the AFM images of mature fibrils (figure 5.III.A.2; right panel),[27] which unambiguously confirms the presence of large non-spherical elongated chain like structures along with a few oligomers. The DLS and AFM measurements thus affirm the formation of intermediate globules like oligomers which can be considered as an onset of the formation of fibrils.

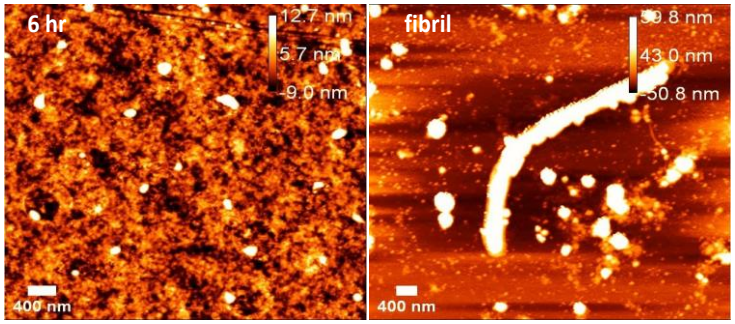


Figure 5.III.A.2 AFM images of oligomers (after six hours of incubation) in the left panel and matured fibril in the right panel.

Circular dichroism measurement for secondary structure and phase diagram: The secondary structural modification of BSA during the fibrillation process is studied using circular dichroism (CD) spectroscopy in far UV range (190 -260 nm) (figure 5.III.A.3a). It is observed that upon incubation of BSA the CD signal decreases, exhibiting a clear signature of partial unfolding of the protein, which serves as the essential first step towards oligomerisation. CD signal quenches subsequently in case of mature fibril (red line and the inset of figure 5.III.A.3a). Secondary structure analysis (figure 5.III.A.3b) reveals that BSA in its native form contains ~ 63% α-helix and ~20% β-structure (including sheet and turn).[28] We observe that after 6 hrs of incubation, the α-helical content decreases from ~63% to ~45% with a concomitant increase in the β-structure content (from 20% to ~28%). Further incubation does not change the secondary structure much (figure 5.III.A.3b, table 5.T1).

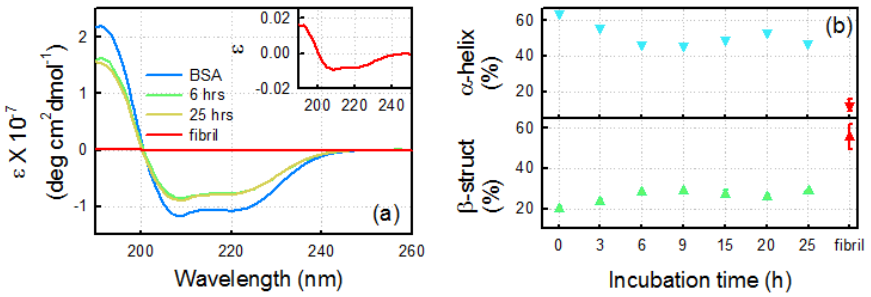


Figure 5.III.A.3 Some of the representative CD signals with different conditions. ε stands for molar ellipticity. Inset is the zoomed version of fibril. (b) percentage of α helix (upper panel) and β-structure (lower panel) profile of differently incubated BSA including fibril.

The α-helical content of mature fibril reduces to ~ 12.5% while that of the β-structure increases to 55.7%; such a sharp change in the secondary content of BSA upon fibrillization is consistent with previous literature.[29] Construction of phase diagram is a very efficient way to qualitatively identify the existence of intermediate conformations during protein structural transitions.[30] We construct phase diagrams for both far UV-far UV ($\varepsilon_{200\,nm}$ vs $\varepsilon_{218\,nm}$) and far UV-near UV ($\varepsilon_{260\,nm}$ vs $\varepsilon_{218\,nm}$) CD values as a function of incubation time[30] (figure 5.III.A.4) to get a precise conclusion about the intermediate state(s) as these can be identified or missed depending on the parameter chosen to construct phase diagram.[30] We observe at least three linear segments corresponding to four different stages (native state, oligomers at 6 hours, at 25 hours of incubation and fibril) to appear during the pathway. Secondary structure determination (figure 5.III.A.3b) and the correlating phase diagram (figure 5.III.A.4) reveal that although oligomer formation initiates after one hour of incubation, a sharp change in the conformation is evident after six hours of incubation.[31]

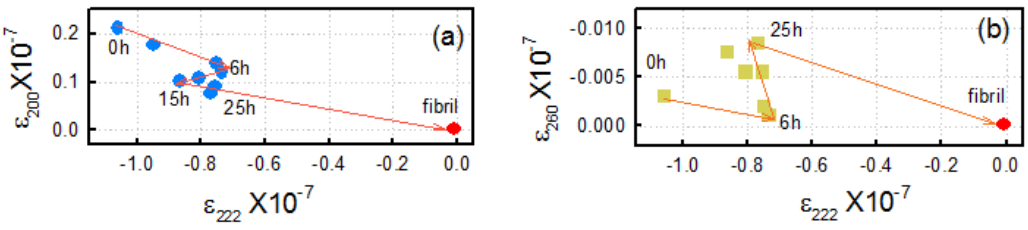


Figure 5.III.A.4 Phase diagram during fibrillation process to find out the intermediates by CD spectroscopy. Molar ellipticity of (a) FAR UV vs FAR UV and (b) FAR UUV vs Near UV.

Fluorescence measurement: Tryptophans (Trp134 and Trp213) present in BSA emit fluorescence when excited at 295 nm. This fluorescence is quenched when incubated (figure 5.III.A.5a, relative intensity is depicted in the inset). BSA, during the fibrillation pathway, produces an additional fluorescence in the visible region when excited at 375 nm (figure 5.III.A.5b). In contrast to the intrinsic fluorescence due to Trp the intensity of this emission, having maximum at 445 nm, increases regularly without any spectral shift with the incubation time (figure 5.III.A.5b; relative intensity is depicted in the inset). A similar behaviour has earlier been reported by Mercato et al in Poly (ValGlyGlyLeuGly) amyloid like fibrils.[32] The reason behind this inherent fluorescence is still debatable and it is argued that dipolar coupling among the excited state of aromatic amino acids residue present inside the protein,[33, 34] formation of intra and/intermolecular H bonds with nearby water molecules and delocalisation of electrons through this formed H-bonded networks,[32] π stacking interaction[35] and also carbonyl double bonds present in the amino acids and proteins[36] could be the possible reasons.

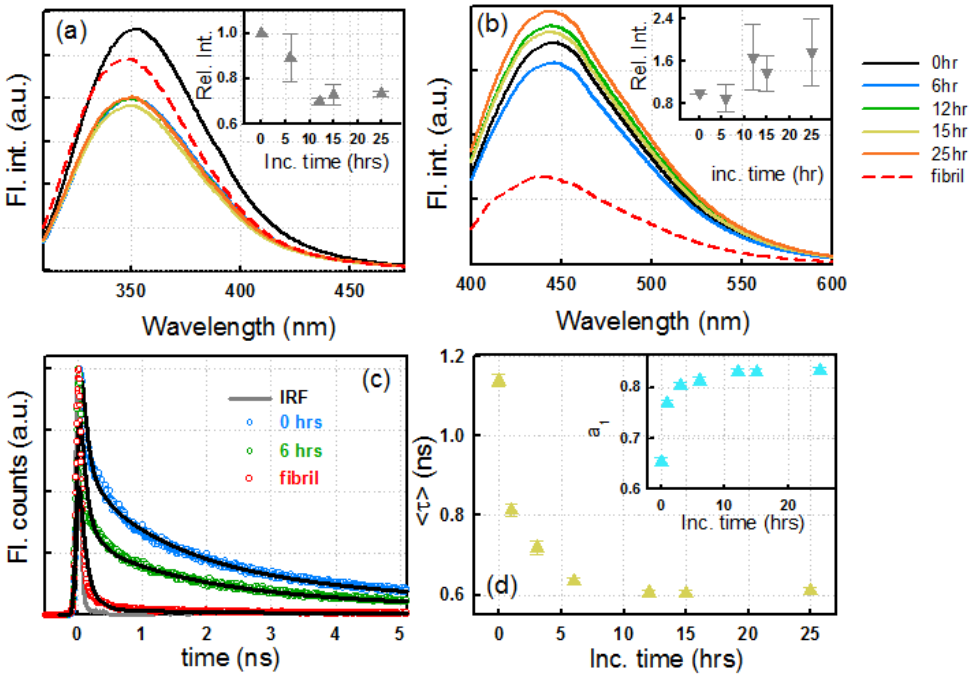


Figure 5.III.A.5 Steady state fluorescence of BSA of different state with excitation at (a) 295 nm and (b) 375 nm. Inset plot represents the respective relative fluorescence intensity. (c) Some representative profiles of time resolved fluorescence decay transients of different condition of BSA, excited at 375 nm. (d) Average lifetime of BSA globules after incubation of different time, monitored at 448 nm. Faster component contribution is shown in the inset of the figure.

We, additionally, notice that lifetime of the oligomers decreases regularly (from ~1.14 ns to ~0.62 ns) up-to six hrs of incubation and beyond that it does not change appreciably (figure 5.III.A.5c, 5.III.A.5d and table 5.T2). For mature fibril, the lifetime becomes noticeably short (~100 ps). The fluorescence characteristics of oligomers formed before six hrs incubation is therefore found to be significantly different than those after six hrs of incubation and fibrillization (table 5.T2). It can then a-priori be claimed that this fluorescence characteristic is an efficient intrinsic indicator of protein aggregation.

Hydration study using THz spectroscopy: Water strongly absorbs in the low frequency THz region as manifested by its high frequency dependent absorption coefficient, $\alpha(\nu)$, due to its resonant intermolecular H-bond network vibrating modes[17]. Since THz frequency probes the collective water dipole fluctuations it provides with a direct tool, in the form of $\Delta\alpha\ (\nu)$, to infer on the alteration of the hydration dynamics (see section 2.I.D for details). Earlier studies reveal that $\alpha(\nu_{THz})$ of water corresponds to the H-bonds which are dynamic in sub-pico second range (~8 ps for cooperative reordering of the H-bonds, ~200 fs for the rotation of water dipole).[18] In a more recent study[37] based on MD simulation, it is evidenced that THz frequency can differentiate short range hydration termed as the "HB wrap" in the closer vicinity of the solute and long range "HB hydration2bulk" water. Pure water (buffer) offers a high $\alpha(\nu_{THz})$ value (figure 5.III.A.6a; cyan solid line).[18] We observe that addition of 400 μM BSA in water decreases the $\alpha(\nu)$ (figure 5.III.A.6a) as protein merely absorbs in the THz radiation. The incubated BSA solutions also offer smaller $\alpha(\nu)$ values than that in buffer. Mature fibrils absorb even less in this frequency window (red line in figure 5.III.A.6a). Since mature fibrils are formed in alcohol solutions only, we, as a control, measure the absorption profile of ethanol/water mixture (60% V/V) as has been depicted in figure 5.III.A.6a (grey line). The water-ethanol mixture also shows lower absorption than pure buffer. To establish quantitatively the change in hydration as BSA traverse the fibrillation pathway (native →oligomers→fibril) we monitor the change in α as defined by $\Delta\alpha\ \ (= \alpha_{protein} - \alpha_{buffer}$; for fibril we replace $\alpha_{buffer}\ with\ \alpha_{ethanol\left(60\%\frac{V}{V}\right)}$) at 1 THz (we chose this frequency as apart from the water dynamics intramolecular vibrations of proteins including the skeleton motions exhibit their signature in 1-10 THz frequency region[38]) and plot it against incubation time. We observe that the $\Delta\alpha$ value is positive and increases up to 6 hrs of incubation beyond which it decreases and acquires a negative value (figure 5.III.A.6a; inset). The mature fibril has the lowest $\Delta\alpha$. It is interesting to note here that the trend in $\Delta\alpha$ is not straightforward as one can observe intersection between the $\alpha(\nu_{THz})$ profiles of buffer and the incubated protein solutions. This

signifies that beyond a certain frequency, $\alpha_{protein}(\nu)$ of the incubated protein solution absorbs more than water. We previously observed similar crossover in urea (>2M) (and its hydrophobic derivatives)-water mixtures.[18]

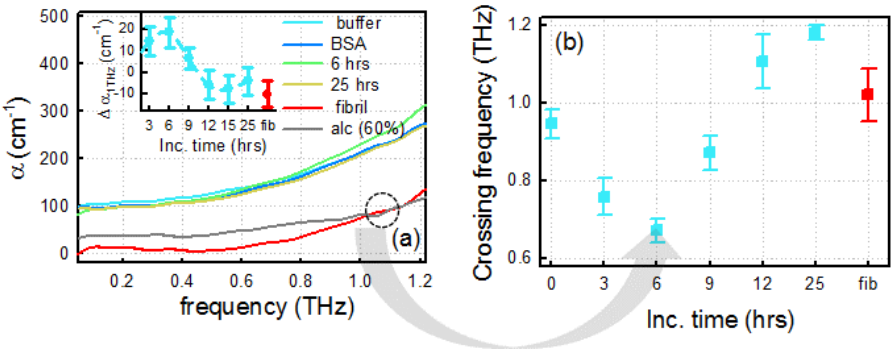


Figure 5.III.A.6 (a) Representative absorption coefficient (α) as a function of frequency. Hollow circle stands for the representative intersection. Difference in absorption coefficient, Δα (=α_{sample}-α_{buffer}) at 1THz is depicted in the inset of the figure. (b) Crossing frequency of different condition of BSA.

For a quantitative apprehension we plot the $\nu_{crossing}$ as a function of incubation time for all the protein solutions (figure 5.III.A.6b). We observe that up-to 6 hrs of incubation, $\nu_{crossing}$ experiences a red shift (lower frequency) than water, while beyond 6 hrs incubation it suffers a blue shift. A previous study with protein hydration has inferred that water, in the vicinity of hydrophobic and hydrophilic moieties (amino acid residues) in proteins, offers contrasting frequency depending absorbance and these two curves indeed intercept each other; the crossing frequency shifts towards higher when water interacts with a hydrophilic surface and vice-versa.[39] The initial lowering of $\nu_{crossover}$ thus infers that water experiences hydrophobic environment when proteins start to form oligomers (as has also been concluded from the CD measurements that the protein is unfolded enough to expose the hydrophobic residues to the aqueous environment). In the pre-fibrillation stage water molecules are released (entropy driven) from the hydrophobic surface of the partially denatured proteins and eventually intermolecular hydrophobic-hydrophobic interactions enhances.[13] As a consequence, water network interacts more to the hydrophilic surfaces. Beyond 6 hrs incubation the observed increase in the $\nu_{crossover}$ suggests that it is the hydrophilic residues of the oligomers that interact more with surrounding water molecules. Interestingly, for mature fibrils, $\nu_{crossover}$ shifts toward a lower frequency compared to that of 25 hrs of incubation. This is due to the contribution from the water molecules present in the fibril interior,

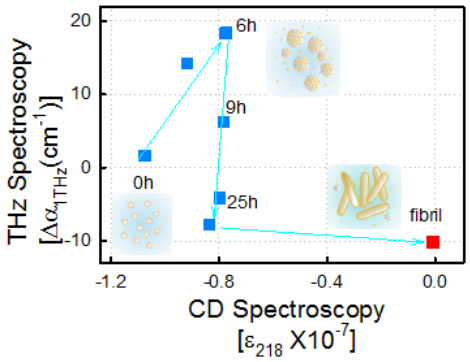


Figure 5.III.A.7 Phase diagram during fibrillation process to correlate structural modification (molar ellipticity at 218 nm from CD spectroscopy) and extended collective hydration (Δα at 1THz from THz spectroscopy).

---

otherwise absent in the oligomers, which abundantly interacts with the hydrophobic moieties in the fibrils.

As we found that the hydration behaviour of BSA shows marked non-linearity as fibrillation proceeds, we now try to correlate whether the structural makeover during this process is correlated with the associated hydration, and to do so we construct the phase diagram of $\Delta\nu_{1THz}$ against the CD signal obtained at 218 nm (figure 5.III.A.7). This study unambiguously concludes a strong correlation between the change in hydration (as obtained from THz spectroscopy) and the structural perturbation (as evidenced from CD) as two clear transitions are observed in the fibrillation pathways. The CD phase diagram has clearly established (figure 5.III.A.4) that the protein secondary as well as tertiary structure changes noticeably (i) at 6 hrs incubation and (ii) upon fibrillation. It has been expected that such changes would show definite impact on the collective hydration dynamics, and we indeed found that hydration does trace back the conformational make overs.

5.III.B Water release amount modifies during fibrillation process depending on the crowders

ThT fluorescence measurements: In this phase, we first monitor the fluorescence of thioflavin T (ThT) upon binding with the Ins aliquots after different time of incubation on the basis of the fact that the fluorescence intensity of ThT increases significantly while binding with the β-sheet of the fibril.[40-42] We notice that the kinetics of ThT on binding with Ins follows sigmoidal like appearance with a lag phase (where the fluorescence intensity does not change

appreciably), elongated phase (where the fluorescence intensity enhances rapidly) and a saturated phase (where no further fluorescence intensity enhances) suggesting the fibril formation (figure 5.III.B.1a and 1b). We fit the normalised relative fluorescence intensity using the following sigmoidal equation to estimate the lag time, τ_{Lag} and the time where the half of the monomers converts to amyloid form, $\tau_{1/2}$:

$$y(t) = \frac{a}{1+e^{-k(t-\tau_{1/2})}} \qquad (5.1)$$

With k being the nucleation growth rate, $\tau_{1/2}$ is the time where the half of the monomers converts to amyloid form. The lag time, τ_{Lag} is defined as $\tau_{1/2} - \frac{2}{k}$.[43]

We find that τ_{Lag} and $\tau_{1/2}$ of Ins are ~52 min and 63 min respectively. These two times, τ_{Lag} and $\tau_{1/2}$ are the measure of the onset of fibrillation process. Now with the addition of sucrose we observe an increase in both τ_{Lag} and $\tau_{1/2}$ (figure 5.III.B.1b, inset and table 5.T3) manifesting the delay in the onset of amyloid formation. Previously Marasini et al reported the delay of Ins fibrillation in presence of sucrose owning to the modification of conformers of Ins in presence of sucrose.[44] We then use glucose, a monomer of sucrose to verify whether it has also any impact on the fibrillation process of Ins. We also notice that both τ_{Lag} and $\tau_{1/2}$ increases regularly with the increase of glucose concentration (inset of figure 5.III.B.1b and table 5.T3) manifesting glucose also delays Ins fibrillation. However the effect of glucose is lesser than sucrose as both τ_{Lag} and $\tau_{1/2}$ increases less amount in glucose than sucrose (table 5.T3). The reason might be the crowding effect; sucrose being larger molecule than glucose has more significant effect on delaying amyloid formation process.

CD measurements: We now monitor the *β*-structure (parallel and anti-parallel *β*-sheet) of Ins in absence and in presence of both glucose and sucrose (figure 5.III.B.1c and 1d) to observe the secondary structural modification of Ins protein during fibrillation pathway. *β*-structure of Ins also follows a sigmoidal trend in consistent with the ThT kinetics (figure 5.III.B.1a and 1b) as the amyloid is formed (grey symbol of figure 5.III.B.1c and 1d); the *β*-structure value being almost constant in the lag phase, increases rapidly in the elongated phase and becomes constant with a high value (from 28% to 50%) in the saturated phase (when the matured fibril develops). Now with the addition of glucose and sucrose, *β*-structure starts to increase at a delayed condition than only Ins solution (figure 5.III.B.1c and 1d) which also corroborates the delaying of Ins fibril formation (in consistent with ThT kinetics). Thus both ThT kinetics and secondary structure of Ins unambiguously confirm the amyloid formation.

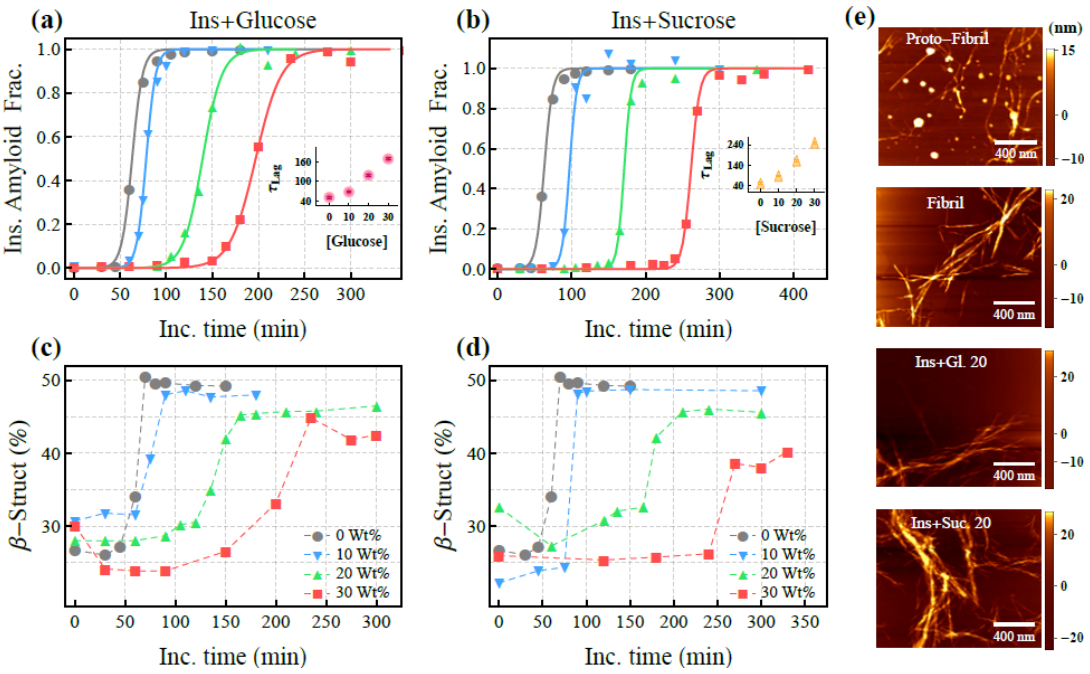


Figure 5.III.B.1. Normalised relative fluorescence profile of ThT, bind with Ins solution in presence of (a) glucose and (b) sucrose to represent the fraction of amyloid growth. Inset of both figures represents the lag time (in min) during the fibrillation process of Ins in presence of crowders. Concentration of both ThT and Ins was kept fixed at 5 μM. β-structure of Ins in presence of (c) glucose and (d) sucrose during the fibril growth. (e) AFM images of different phases of Ins. Protofibril (at elongated phase) and fibril (saturated phase) are shown in the upper two panels of Ins. Lower two panels represent images of Ins fibril in presence of crowders. Gl. 20 and Suc. 20 represents glucose and sucrose of 20 Wt% respectively.

AFM measurements: On verifying the amyloid formation using CD and fluorescence spectroscopy, we image the Ins solutions in atomic force microscopy (AFM) to get the morphological structure of fibrils and also in presence of crowders (figure 5.III.B.1e). We find that Ins during amyloid growth (in the elongated phase) first produces spherical oligomers with average diameter of 240±60 nm along with non-spherical chain like protofibril (upper panel of figure 5.III.B.1e). When matured fibril develops (after 120 mins of incubation), we do not observe any spherical oligomers, rather only rod like structure appears (figure 5.III.B.1e). We also find the appearance of rod like fibres in presence of 20 Wt% glucose and sucrose (figure 5.III.B.1e, lower two panels). Now to verify whether the structural modification of fibres occurs in presence of crowders depending on crowders molecular weight, we analyse the AFM images to get the statistical distribution of length and height of Ins fibril in absence and in presence of crowders (figure 5.III.B.2). We fit the statistical data using normal distribution and

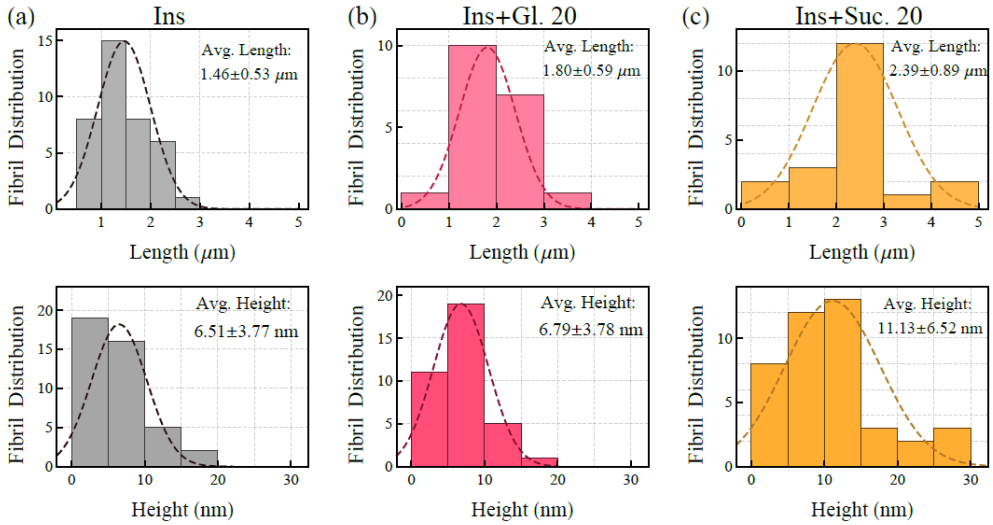


Figure 5.III.B.2. Length and height distribution of matured fibril of Ins (a) in absence of any crowders and in presence of 20 Wt% (b) glucose and (c) sucrose.

observe that the average length and height of Ins fibril is 1.46±0.53 μm and 6.51±3.77 nm respectively (figure 5.III.B.2a). Both the length and height of matured fibrils increases in presence of crowders; the effect is more significant in sucrose than glucose (figure 5.III.B.2). AFM images thus suggest that morphological structure of fibril gets modified in presence of crowders, corroborated with both CD and fluorescence spectroscopy.

Hydration study using FIR-THz spectroscopy: Upon establishing the amyloid formation and the structure of fibrils in presence of crowders, we now investigate the associated protein hydration using far IR FTIR spectroscopy in THz fingerprint region (50-570 cm^{-1}; 1.5-17 THz). In this frequency window, both water and protein offers different inter/intra molecular vibration (specifically skeleton motions of proteins[45]). It is therefore expected that the water vibrational modes, coupled with the proteins, would exhibit specific signature in the overall hydration

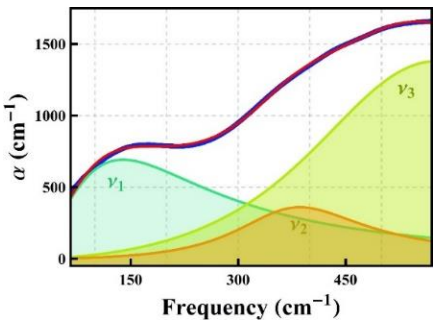


Figure 5.III.B.3. Fitting profile of water using three component damped harmonic oscillator model. Peak frequency with ν_1, ν_2 and ν_3 respectively represents HB stretch, in-plane and out of plane librational mode.

dynamics during fibril growth. We measure the frequency dependent absorption coefficient, $\alpha(\nu)$ (see section 2.II.D for details) of water and protein solution under different conformations. Water offers two strong bands with peak ~133 cm^{-1} and ~550 cm^{-1} (figure 5.III.B.3) corresponding to the intermolecular HB stretch and librational motion respectively.[46] However when we further dissect the librational band we find two peaks exhibiting at ~ 359 cm^{-1} and ~565 cm^{-1} which can be attributed as "*in plane*" and "*out of plane*" O-H librational mode respectively (figure 5.III.B.3).[45, 47, 48] Addition of Ins in the buffer solution decreases $\alpha(\nu)$ due to the replacement of high absorbing water molecules by low absorbing protein molecules.[49, 50] In order to explicitly extract the change in protein hydration during fibrillation process, we calculate the difference in absorption coefficient, $\Delta\alpha(\nu)$ $(=\alpha_{sample}(\nu)-\alpha_{buffer}(\nu))$ at different incubation time and observe a distinct change in the $\Delta\alpha(\nu)$ (figure 5.III.B.4a) manifesting the change in protein hydration during amyloid formation. As the fibrillation process consists of three phases (lag, elongated and saturated phase), a lot of data set (following ThT kinetics as shown in figure 5.III.B.1a and 1b) is required to completely express all three phases, and it is not very evident form figure 5.III.B.4a how the hydration changes in each phases during amyloid growth. To overcome this challenge,

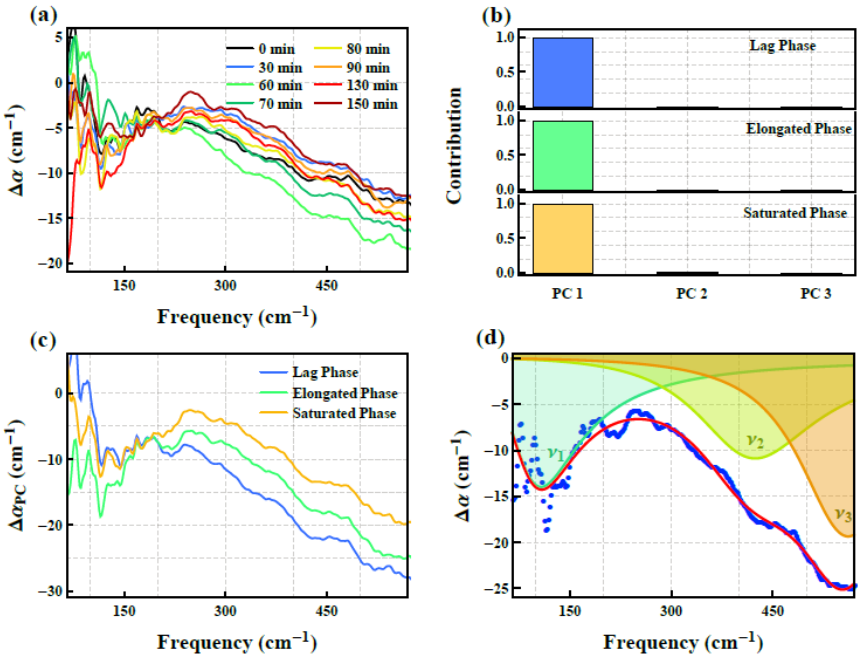


Figure 5.III.B.4. (a) Representative Δα of Ins protein solution at different incubation time during amyloid growth. (b) Contribution of first three principal components of three different phases of Ins during its fibrillation pathway. (c) Representative 1st principal component of Δα of Ins solution of three different phases: lag phase, elongated phase and saturated phase. (d) Representative fitting of Ins solution at its elongated phase. The fitting has been done using three component damped harmonic oscillator model (equation 5.2).

we perform principal component analysis (see section 2.I.E for details) on $\Delta\alpha(\nu)$ of each phase. Representative contribution of first three principal components (PCs) of all three phases (lag phase, elongated phase and saturated phase) while converting from Ins monomer to aggregated amyloid fibril has been depicted in figure 5.III.B.4b. As the first PC of each phase has the extremely high contribution (>98%) we consider only the first PC to conclude our results. First PC of each phase of Ins protein during amyloid growth has been shown in figure 5.III.B.4c. $\Delta\alpha_{PC}(\nu)$ value above 200 cm^{-1} decreases regularly in the following order: lag phase, elongated phase and saturated phase; however, the data, being noisy is not much reliable below 200 cm^{-1}. Decrease of $\Delta\alpha_{PC}(\nu)$ in the successive phases can be ascribed by water release (protein de-wetting) mechanism. When the monomers start to aggregate, protein de-wetting occurs exhibiting a decrease in $\Delta\alpha_{PC}(\nu)$ in elongated phase than lag phase, and in saturated phase where a complete aggregation takes place, more water molecules release suggesting further decrease in $\Delta\alpha_{PC}(\nu)$. Water release (de-wetting) mechanism during amyloid formation has been hypothesized earlier.[13, 51] Although water releases from the protein interior and/or protein hydration layer during the aggregation process, the fibril core environment does not become completely dry rather some water molecules persist there under trapped condition as evidenced from the negative $\Delta\alpha_{PC}(\nu)$ value of saturated phase (figure 5.III.B.4c). Existence of such water pool, trapped in Aβ fibril has been reported earlier.[52]

For a more quantitate view on the decrease of $\Delta\alpha_{PC}(\nu)$ or in other words to get an idea about the release of water molecules from the hydration layer/protein interior of Ins during the fibrillation process and also how the release of water molecules modifies in presence of crowders during the amyloid formation we calculate $\Delta\alpha_{PC}$ at 359 and 565 cm^{-1} (figure 5.III.B.5). We choose these two particular frequencies on the basis of the fact that these frequencies are the characteristic peaks of librational mode of water (figure 5.III.B.3). $\Delta\alpha_{PC}$ of Ins at 359 and 565 cm^{-1} decreases gradually in three successive phases (grey symbol in figure 5.III.B.5a and b) confirming water release during amyloid growth. We now perform PCA on $\Delta\alpha(\nu)$ ($=\alpha_{protein\text{-}crowder\text{-}buffer}-\alpha_{crowder\text{-}buffer}$) in presence of each crowders in a way similar to only Ins solution. We collect the 1st PC of each sample (every concentration of each crowder) for further analysis since the contribution of first PC is also overwhelmingly high (more than 95%). We notice that $\Delta\alpha_{PC}$ at both 359 and 565 cm^{-1} also decreases in each phase as the aggregation precedes in presence of glucose (figure 5.III.B.5a); however the change is glucose concentration dependent. In a similar way, we observe that in presence of sucrose, the change in $\Delta\alpha_{PC}$ depends on sucrose concentration (figure 5.III.B.5b). Before proceeding further, we measure the

difference in absorption coefficient, $\Delta\alpha_c$ ($\Delta\alpha_c=\alpha_{crowder\text{-}buffer}-\alpha_{buffer}-\phi_{crowder}\,\alpha_{dry\ crowder}$) in presence of only crowders to verify any change in hydration of crowders. Since, either glucose or sucrose itself offers a change in $\Delta\alpha_c$ (figure 5.III.B.6b and S4c) we can unambiguously infer that the hydration of crowders also impacts on the protein hydration, causing a change in the $\Delta\alpha_{PC}$ value of lag phase in presence of crowders compared to native protein (figure 5.III.B.5).

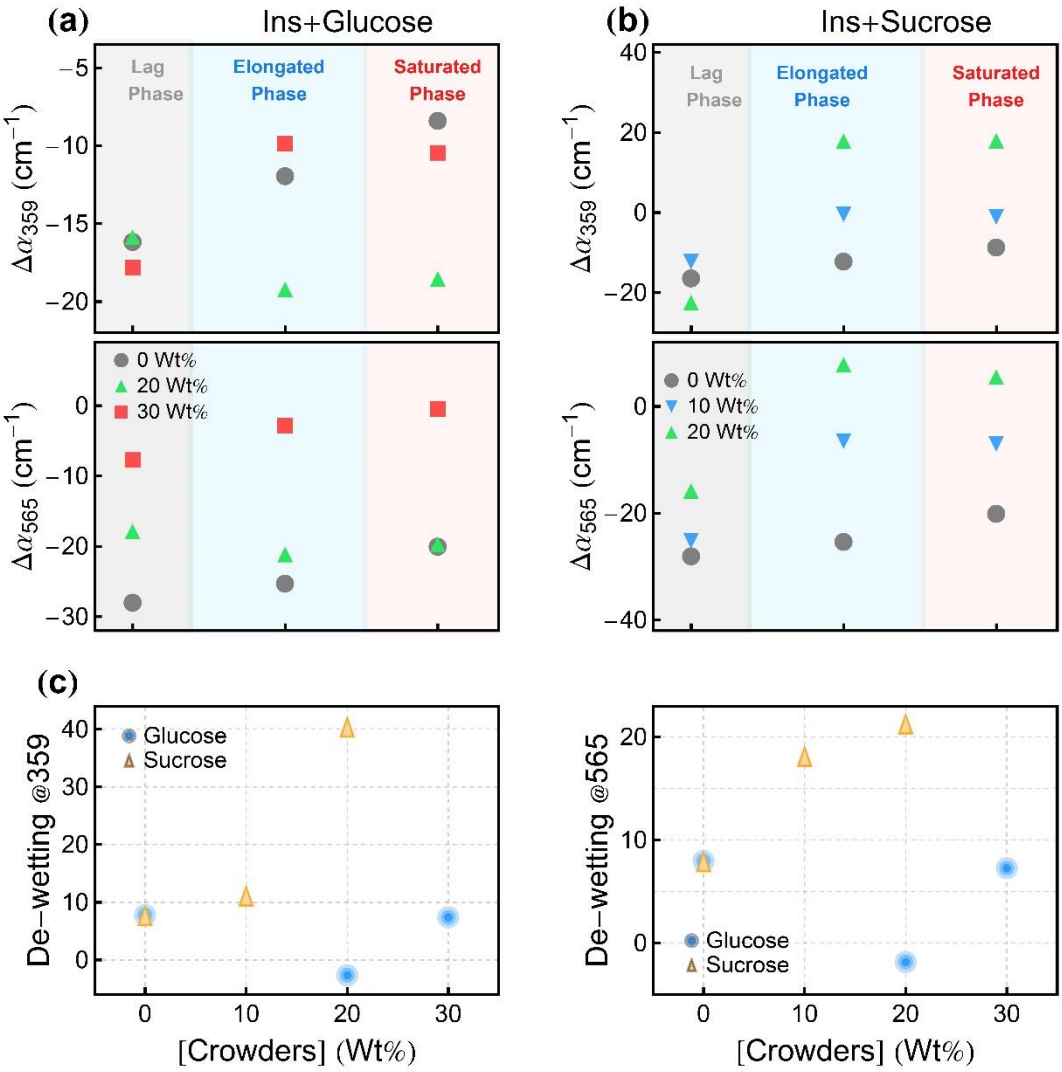


Figure 5.III.B.5. $\Delta\alpha_{PC}$ of Ins in presence of (a) glucose and (b) sucrose for three different phases at 359 and 565 cm^{-1}. The reason behind the selection of these two frequencies is that water shows two librational modes at these frequencies. The $\Delta\alpha$ at 133 cm^{-1} (owning to the HB stretch of water) is not considered due to the noisy data of $\Delta\alpha$ (ν) below 200 cm^{-1}. (c) water release during amyloid growth has been estimated in presence of glucose and sucrose at 359 and 565 cm^{-1} respectively. Water release process (de-wetting) during amyloid growth has been estimated in term of the difference between $\Delta\alpha$ in the saturated phase and lag phase (i.e. $\Delta\alpha_{saturated\ phase} - \Delta\alpha_{lag\ phase}$).

The change in $\Delta\alpha$ in both elongated phase and saturated phase in presence of crowders compared to the only Ins solution might be either due to the crowder hydration or the crowder mediated conformational modification of protein or both.

We now focus to estimate the release of water during amyloid formation. Since $\Delta\alpha$ is a measure of hydration, so any decrease in $\Delta\alpha$ manifests the decrease of volume fraction of hydration, ϕ_h; hence decrease (release) of water molecule in (from) the hydration layer/protein interior. Thus the difference of $\Delta\alpha$ between saturated and lag phase (i.e. $\Delta\alpha_{\text{saturated phase}} - \Delta\alpha_{\text{lag phase}}$) would be an estimation of water release during the amyloid formation process. We differentiate the $\Delta\alpha$ at both 359 and 565 cm^{-1} and plot in figure 5.III.B.5c and observe that water release increases gradually with the increase of either glucose or sucrose concentration. Addition to this we also observe that the more amount of water releases in sucrose than glucose at a particular concentration (figure 5.III.B.5c). AFM image analysis of Ins fibres manifests that length and height is greater in the presence of sucrose than glucose (figure 5.III.B.2). The larger is the fibre length and height the more is the water release which corroborates with water release analysis (figure 5.III.B.5c). Again, ThT kinetics and β structure analysis confirms a delayed fibril growth in presence of crowders (figure 5.III.B.1a-d). This result thus a priori concludes that the more is the delay on the onset of amyloid growth the more is the release of water molecules from protein interior.

On establishing the release of water molecules from the protein interior during amyloid formation and also in presence of crowders, it is affirmed that some water molecules remain trapped in the core of the proto-fibril and matured fibril. We now explore, in a quantitative way, the physical nature of remaining trapped water and whether it is distinct from tetrahedral network of bulk water. We fit the first PC of $\Delta\alpha(\nu)$ of each phase of Ins using a sum of three component damped harmonic oscillation model as follows:[53]

$$\Delta\alpha_{PC}(\nu) = \sum_{i=1}^{3} \frac{a_i \omega_i \nu^2}{\nu^2\omega_i^2 + \pi^2\left(\nu_i^2 + \frac{\omega_i^2}{4\pi^2} - \nu^2\right)^2} \tag{5.2}$$

where a_i, ω_i, ν_i are the amplitude, spectral line-width, and the centre frequency of the i^{th} resonance. The unperturbed centre frequency can be defined as,

$$\nu_{0,i} = \sqrt{\nu_i^2 + \frac{\omega_i^2}{4\pi^2}}. \tag{5.3}$$

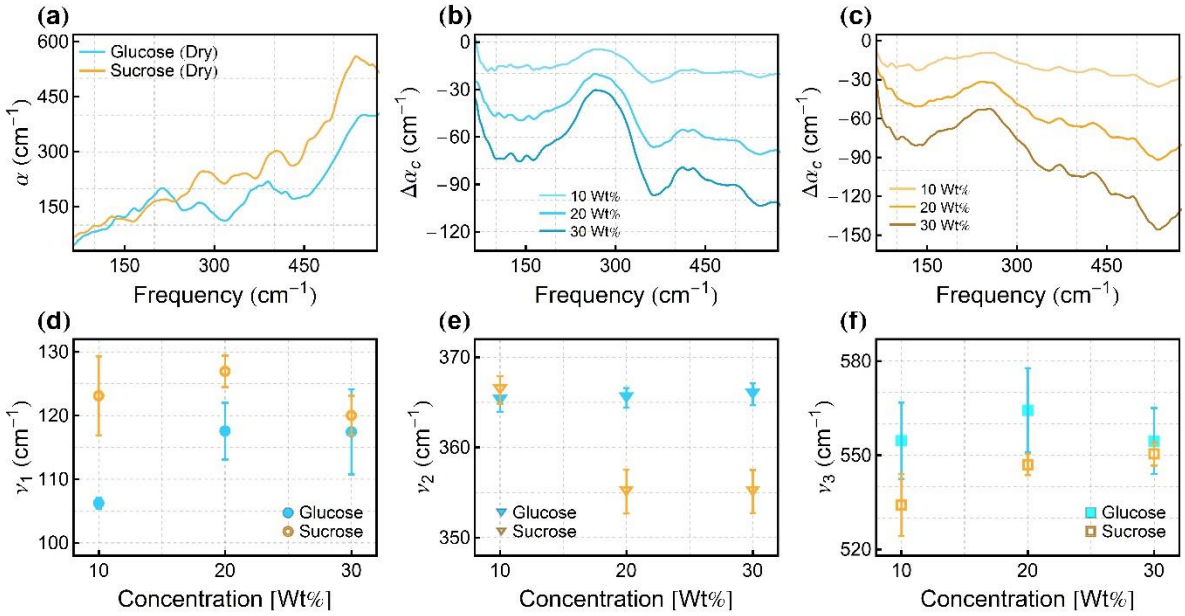


Figure 5.III.B.6. (a) Absorption coefficient of pure glucose and sucrose (dry). $\Delta\alpha_c(\nu)$ ($=\alpha_{\text{crowder-buffer}}-\alpha_{\text{buffer}}-\phi_{\text{crowder}}\,\alpha_{\text{dry crowder}}$) profile with increasing concentration of (b) glucose and (c) sucrose solution. ϕ_{crowders} is the corresponding volume fraction of each crowder. (d), (e) and (f) represent peak frequency of ν_1, ν_2 and ν_3 with the increase of crowder concentration.

A representative fitting of elongated phase of Ins protein is shown in figure 5.III.B.4d. We obtain three curves with peak frequencies at ν_1 (peak 1, corresponding to HB stretch) ~138 cm^{-1}, ν_2 (peak 2, *in plane* librational motion) ~445 cm^{-1} and ν_3 (peak 3, *out of plane* librational motion) ~542 cm^{-1} in the lag phase of Ins solution. Any blue (red) shift in HB stretch and librational motion signifies strengthening (weakening) of hydrogen bond coupled with more (less) hindered rotation respectively.[54, 55] Ins hydration in the lag phase suffers stronger hydrogen bonds along with more hindered rotation of "in plane" librational motion whereas "out of plane" librational motion experiences less hindered rotation. A red shift occurs in all three bands (ν_1, ν_2, and ν_3) in both elongated and saturated phase compared to the lag phase (figure 5.III.B.7, grey symbol) exhibiting that trapped water in the proto-fibril (elongated phase) and fibril (saturated phase) suffers weaker H bonds with less hindered rotation than the water in lag phase. The result thus establishes that tetrahedral network of trapped water becomes more fragile while developing proto-fibrils and fibrils. Now to investigate the behaviour of trapped water around proto-fibril and fibril in presence of crowders we fit the $\Delta\alpha_{PC}(\nu)$ using the equation 5.2 and extract all three peaks, ν_1, ν_2 and ν_3 for all three phases (figure 5.III.B.7 and table 5.T4). But before proceeding further to discuss the change in the

water network during fibril growth in presence of crowers it is important to verify whether the crowders themselves modify the water network. We fit each $\Delta\alpha_c$ using equation 5.2 and peak frequencies: ν_1, ν_2 and ν_3 are plotted in figure 5.III.B.6d-f. We observe an almost monotonic change in all three frequencies with the increase of both glucose and sucrose concentration. The results thus unambiguously conclude that crowders themselves do modify the water network.

We observe that all peak frequencies get shifted (either blue or red end) compared to only Ins solution depending on the concentration, crowders and phase (see figure 5.III.B.7 and table 5.T4). The observed peak shifts suggest that water network at the protein interior/hydration layer gets modified in each phase during amyloid growth and water network in each phase is distinct from the other phases which also depends on the crowders and also its

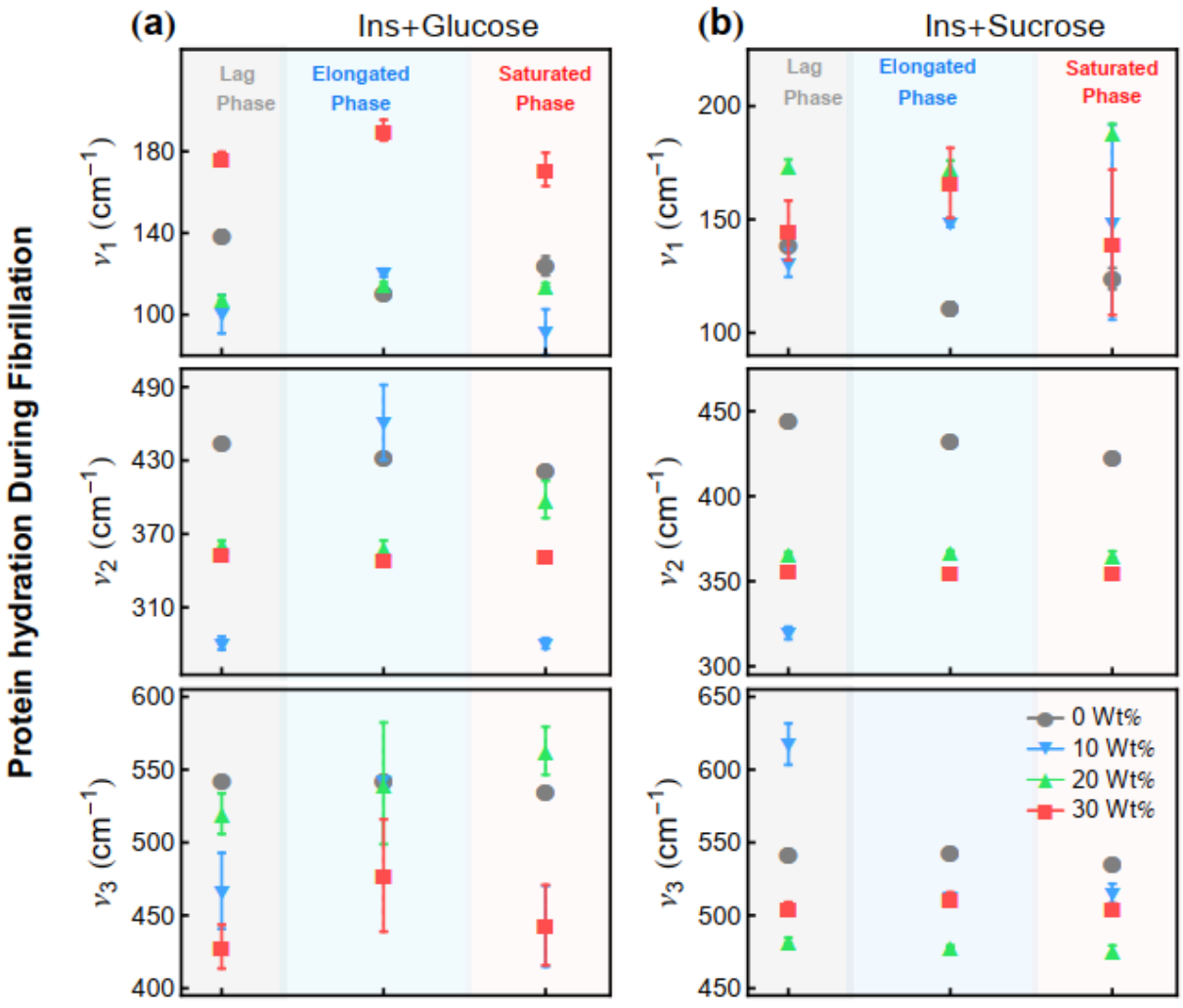


Figure 5.III.B.7. Peak frequency of Ins in presence of (a) glucose and (b) sucrose for three different phases. Fitting is performed on the 1st PC of each phases.

concentration. Such non-monotonous change in the water network perhaps appears due to the various factors like the perturbation of water network by the crowders themselves, conformational modification of proteins owning to aggregation, however the exact origin of such change in the water network is not very evident, rather a detailed simulation study might assist towards understanding this

Our study certainly addresses the water release process during the amyloid growth and the amount of released water strongly depends on the crowders; the crowders which delays more in amyloid formation, assists to release more water molecules from protein interior indicating water release process and the hydration dictates the fibril inhibition process. We also find that the behaviour of trapped water (water remaining even after release) is distinct in each phase depending on crowders concentration and differ quietly from bulk water.

5. IV Conclusion:

The above two in vitro studies are a primary step towards experimentally identifying the onset of fibrillation process using THz spectroscopy in term of alteration of water network structure. We first demonstrate the alteration of water network during the fibrillation of BSA and correlate the structural makeover of BSA with the associated hydration. We then investigate the water release process and the physical behaviour of trapped water in the protein during insulin amyloid formation. We also observe the crowders induced delay of Ins fibrillation and the modification of both water release amount and water network in presence of crowders. Our study unambiguously establishes that the amount of released water modifies depending on the crowders which inhibits the fibrillation process of insulin protein. Moreover, the delay in Ins fibrillation could also assist the long term storage.

Table 5.T1: Secondary structure information of BSA in different condition:

System	α-Helix (%)	β-Antiparallel (%)	β-Parallel (%)	β-Turn (%)	Random coil (%)
Buffer	62.83±0.55	2.99±0.42	4.05±0.18	13.30±0.37	16.83±0.36
3 hr	55.94±0.57	4.05±0.43	5.04±0.16	14.57±0.38	20.44±0.35
6 hr	45.23±0.58	5.95±0.36	6.69±0.09	16.00±0.27	26.10±0.42
9 hr	45.18±0.54	6.13±0.30	6.66±0.11	16.07±0.33	25.96±0.39
15 hr	48.87±0.54	5.40±0.33	6.08±0.12	15.65±0.39	24.00±0.38
20 hr	52.28±0.61	4.77±0.37	5.58±0.16	15.18±0.42	22.16±0.37
25 hr	46.22±0.55	5.94±0.30	6.49±0.12	16.00±0.36	25.31±0.42
fibril	12.52±1.37	30.85±3.38	9.57±1.74	15.32±1.03	31.74±4.06

Table 5.T2: Time resolved fluorescence data of BSA after different time of incubation.

time (hrs)	a_1	t_1 (ns)	a_2	t_2 (ns)	a_3	t_3 (ns)	t_{avg} (ns)
0	0.65	0.10	0.23	1.39	0.12	6.53	1.14
1	0.77	0.10	0.16	1.51	0.07	6.82	0.81
3	0.80	0.10	0.14	1.58	0.06	6.98	0.72
6	0.81	0.10	0.14	1.58	0.05	7.07	0.64
12	0.83	0.10	0.12	1.60	0.05	7.05	0.60
15	0.83	0.10	0.12	1.59	0.05	7.05	0.60
25	0.83	0.10	0.12	1.63	0.05	7.24	0.61
fibril	0.98	0.10	0.02	2.65	--	--	0.15

Table 5.T3: τ_{Lag} and $\tau_{1/2}$ of Ins fibrillation process and also in presence of crowders.

Glucose (Wt%)	τ_{Lag} (min)	$\tau_{1/2}$ (min)		Sucrose (Wt%)	τ_{Lag} (min)	$\tau_{1/2}$ (min)
0	52±3	63±2		0	52±3	63±2
10	68±1	78±1		10	87±4	97±2
20	119±3	139±1		20	161±4	171±2
30	169±3	197±2		30	250±6	261±4

Table T4. Peak frequencies: ν_1, ν_2 and ν_3 of three different phases during amyloid formation in presence of crowders

System (Wt%)	Lag Phase			Elongated Phase			Saturated Phase		
	ν_1 (cm^{-1})	ν_2 (cm^{-1})	ν_3 (cm^{-1})	ν_1 (cm^{-1})	ν_2 (cm^{-1})	ν_3 (cm^{-1})	ν_1 (cm^{-1})	ν_2 (cm^{-1})	ν_3 (cm^{-1})
Glucose									
0	138±3	445±3	542±2	110±3	432±2	543±2	124±5	422±2	535±2
10	100±10	281±5	467±26	120±2	461±30	542±3	92±11	281±5	442±27
20	108±1	361±4	520±14	115±1	359±6	541±41	115±1	395±15	563±16
30	176±3	354±2	429±15	190±5	349±3	477±38	172±8	352±2	443±27
Sucrose									
10	131±6	320±4	618±14	149±2	355±1	513±1	148±43	354±1	515±7
20	174±3	366±1	483±2	173±3	367±1	478±1	189±3	365±2	477±3
30	145±13	356±2	505±4	166±15	356±1	512±4	140±32	355±1	505±4

5.V References:

1. G. M. Whitesides, J. P. Mathias and C. T. Seto, *Science*, 1991, **254**, 1312-1319.
2. C. M. Dobson, *Trends Biochem. Sci*, 1999, **24**, 329-332.
3. S. T. Ferreira, M. N. N. Vieira and F. G. De Felice, *IUBMB Life*, 2007, **59**, 332-345.
4. M. Stefani and C. M. Dobson, *J. Mol. Med.*, 2003, **81**, 678-699.
5. R. Nelson, M. R. Sawaya, M. Balbirnie, A. Ø. Madsen, C. Riekel, R. Grothe and D. Eisenberg, *Nature*, 2005, **435**, 773-778.
6. A. J. Baldwin, T. P. J. Knowles, G. G. Tartaglia, A. W. Fitzpatrick, G. L. Devlin, S. L. Shammas, C. A. Waudby, M. F. Mossuto, S. Meehan, S. L. Gras, J. Christodoulou, S. J.

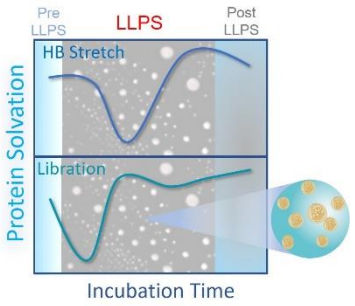
Anthony-Cahill, P. D. Barker, M. Vendruscolo and C. M. Dobson, *J. Am. Chem. Soc.*, 2011, **133**, 14160-14163.
7. R. P. Friedrich, K. Tepper, R. Rönicke, M. Soom, M. Westermann, K. Reymann, C. Kaether and M. Fändrich, *Proc. Natl. Acad. Sci.*, 2010, **107**, 1942-1947.
8. F. Bemporad and F. Chiti, in *Amyloid Fibrils and Prefibrillar Aggregates*, Wiley Online Library2013, ch. 8.
9. P. Arosio, T. P. J. Knowlesa and S. Linse, *Phys. Chem. Chem. Phys.*, 2015, **17**, 7606-7618.
10. V. N. Uversky, J. Li and A. L. Fink, *J. Biol. Chem.*, 2001, **276**, 10737–10744.
11. F. Chiti, M. Stefani, N. Taddei, G. Ramponi and C. M. Dobson, *Nature*, 2003, **424**, 805 – 808.
12. S. H. Chong and S. Ham, *Angew. Chem. Int. Ed.*, 2014, **53**, 3961 –3964.
13. D. Thirumalai, G. Reddy and J. E. Straub, *Acc. Chem. Res.*, 2012, **45**, 83-92.
14. M. Perutz, J. Finch, J. Berriman and A. Lesk, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 5591–5595.
15. Y. Fichou, G. Schirò, F.-X. Gallat, C. Laguri, M. Moulin, J. Combet, M. Zamponi, M. Härtlein, C. Picart, E. Mossou, H. Lortat-Jacob, J.-P. Colletier, D. J. Tobias and M. Weik, *Proc. Natl. Acad. Sci. U.S.A.*, 2015, **112**, 6365–6370.
16. S. Arya, A. K. Singh, K. Bhasne, P. Dogra, A. Datta, P. Das and S. Mukhopadhyay, *Biophysical J.*, 2018, **114**, 2540–2551.
17. S. Ebbinghaus, S. J. Kim, M. Heyden, X. Yu, U. Heugen, M. Gruebele, D. M. Leitner and M. Havenith, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 20749-20752.
18. N. Samanta, D. D. Mahanta and R. K. Mitra, *Chem. Asian J.*, 2014, **9**, 1 – 8.
19. T. Q. Luong, P. K. Verma, R. K. Mitra and M. Havenith, *Biophysical J.* , 2011, **101**, 925–933.
20. P. V. Röder, B. Wu, Y. Liu and W. Han, *Exp. Mol. Med.*, 2016, **48**, e219.
21. B. Farruggia, F. Rodriguez, R. Rigatuso, G. Fidelio and G. Picó, *J. Protein Chem.*, 2001, **20**, 81-89.
22. A. Michnik, *J.Therm. Anal. Cal.*, 2003, **71**, 509-519.
23. Z. Yaseen, S. U. Rehman, M. Tabish, A. H. Shallaa and K. U. Dinc, *RSC Advances*, 2015, **5**, 58616.
24. N. R. Rovnyagina, N. N. Sluchanko, T. N. Tikhonova, V. V. Fadeev, A. Y. Litskevich, A. A. Maskevich and E. A. Shirshin, *Int. J. Biol. Macromol.*, 2018, **108**, 284–290.
25. M. Bhattacharya, N. Jain and S. Mukhopadhyay, *J. Phys. Chem. B*, 2011, **115**, 4195–4205.
26. V. Mittal, G. Devitt, M. Nedeljkovic, L. G. Carpenter, H. M. H. Chong, J. S. Wilkinson, S. Mahajan and G. Z. Mashanovich, *Biomed. Opt. Express.*, 2020, **11**, 4714-4722.
27. R. Khurana, C. Ionescu-Zanetti, M. Pope, J. Li, L. Nielson, M. Ramírez-Alvarado, L. Regan, A. L. Fink and S. A. Carter, *Biophysical J.*, 2003, **85**, 1135–1144.
28. R. G. Reed, R. C. Feldhoff, O. L. Clute and T. Peters, *Biochemistry*, 1975, **14**, 4578-4583.
29. Z. Yaseen, U. R. Sayeed, M. Tabish, A. H. Shalla and K. U. Din, *RSC Advances*, 2015, **5**, 58616-58624.
30. I. M. Kuznetsova, K. K. Turoverov and V. N. Uversky, *J. Proteome Res.*, 2004. , **3**, 485–494.
31. I. Usov, J. Adamcik and R. Mezzenga, *Faraday Discuss.* , 2013, **166**, 151.
32. L. L. del Mercato, P. P. Pompa, G. Maruccio, A. D. Torre, S. Sabella, A. M. Tamburro, R. Cingolani and R. Rinaldi, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 18019-18024.
33. N. Amdursky, M. Molotskii, D. Aronov, L. Adler-Abramovich, E. Gazit and G. Rosenman, *Nano Lett.* , 2009, **9**, 3111–3115.
34. P. Hanczyc, M. Samoc and B. Norden, *Nat. Photon.*, 2013, **7**, 969–972.
35. M. J. Krysmann, V. Castelletto, A. Kelarakis, I. W. Hamley, R. A. Hule and D. J. Pochan, *Biochemistry* 2008, **47**, 4597–4605.
36. C. Niyangoda, T. Miti, l. Breydo, V. N. Uversky and M. Muschol, *PLoS One*, 2017, **12**, e0176983.
37. V. C. Nibali, S. Pezzotti, F. Sebastiani, D. R. Galimberti, G. Schwaab, M. Heyden, M. P. Gaigeot and M. Havenith, *J. Phys. Chem. Lett.* , 2020, **11**, 4809−4816.
38. Y. Xu and M. Havenith, *J. Chem. Phys.*, 2015, **143**, 70901.
39. M. Heyden and M. Havenith, *Methods*, 2010, **52**, 74-83.
40. B. Urbanc, L. Cruz, R. Le, J. Sanders, K. H. Ashe, K. Duff, H. E. Stanley, M. C. Irizarry and B. T. Hyman, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 13990–13995.

41. P. Friedhoff, A. Schneider, E. M. Mandelkow and E. Mandelkow, *Biochemistry*, 1998, **37**, 10223–10230.
42. T. Kampers, P. Friedhoff, J. Biernat, E. M. Mandelkow and E. Mandelkow, *FEBS Lett*, 1996, **399**, 344–349
43. L. Nielsen, R. Khurana, A. Coats, S. Frokjaer, J. Brange, S. Vyas, V. N. Uversky and A. L. Fink, *Biochemistry*, 2001, **40**, 6036-6046.
44. C. Marasini, V. Foder`a and B. Vestergaard, *RSC Adv.*, 2017, **7**, 10487.
45. Y. Xu and M. Havenith, *J. Chem. Phys.*, 2015, **143**, 70901.
46. M. Heyden, J. Sun, S. Funkner, G. Mathias, H. Forbert, M. Havenith and D. Marx, *Proc. Natl. Acad. Sci. U.S.A.*, 2010, **107**, 12068-12073.
47. H. R. Zelsmann, *J. Mol. Struct.*, 1995, **350**, 95−114.
48. J. Ceponkus and B. Nelander, *J. Phys. Chem. A*, 2004, **108**, 6499–6502.
49. S. Ebbinghaus, S. J. Kim, M. Heyden, X. Yu, U. Heugen, M. Grubele, D. M. Leitner and M. Havenith, *Proc. Natl. Acad. Sci. U.S.A*, 2007, **104**, 20749-20752.
50. T. Q. Luong, P. K. Verma, R. K. Mitra and M. Havenith, *Biophys J.*, 2011, **101**, 925-933.
51. P. H. Nguyen, M. S. Li, G. Stock, J. E. Straub and D. Thirumalai, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 111–116.
52. T. Wang, H. Jo, W. F. DeGrado and M. Hong, *J. Am. Chem. Soc.* , 2017, **139**, 6242−6252.
53. T. Dodo, M. Sugawa, E. Nonaka, H. Honda and S. Ikawa, *J. Chem. Phys.*, 1995, **102**, 6208.
54. V. Conti Nibali, S. Pezzotti, F. Sebastiani, D. R. Galimberti, G. Schwaab, M. Heyden, M. P. Gaigeot and M. Havenith, *J. Phys. Chem. Lett.* , 2020, **11**, 4809–4816.
55. J. Ahlers, E. M. Adams, V. Bader, S. Pezzotti, K. F. Winklhofer, J. Tatzelt and M. Havenith, *Biophysical J.*, 2021, **120**, 1266–1275.

Chapter 6

6. Excipients do regulate phase separation in Lysozyme and thus also its hydration

While liquid-liquid phase separation (LLPS) process in proteins has been studied in great details, it has not been widely explored how the associated protein solvation changes during the process and how important is its role in the process itself. In this contribution we experimentally explore the alteration of lysozyme solvation during its LLPS process using attenuated total reflection (ATR)-FTIR spectroscopy in the THz frequency range (1.5-21 THz). Additionally, we explore the role of excipients (L-arginine, sucrose, bovine albumin (BSA) and ubiquitin (Ubi)) in regulating the process and found that while sucrose stabilizes the LLPS, BSA inhibits it. The effect of Arg in the LLPS stands subtle and that of Ubi is concentration dependent. We made a detailed analysis of the solvation profile of Lys in presence of these excipients and observe that solvation changes make a definite signature assisting or resisting the process.



6.I Introduction:

Compartmentalization of eukaryotic cells is essential to maintain their functionality and activities. Organelles inside cells are mostly fenced with semi-permeable bilayer lipid membrane segregating them from the cell cytoplasm so that their physicochemical processes remain unperturbed.[1] On the contrary, some other organelles[2-4] maintain their structures and perform physiochemical activities inside them in spite of having no physical barrier. Recent reports suggest that assembling and preservation of such membranelles organelles (MLO) proceeds through a biophysical process commonly termed as liquid-liquid phase separation (LLPS),[5-7] which is associated with several functionality in the living cells. Several proteins (e.g. fused in sarcoma, heterogeneous nuclear ribonucleoprotein A2/B1, TAR DNA binding protein 43 etc) associated with MLO have been reported to form liquid like droplets even in-vitro condition depending upon the protein concentration and/or external stress (like temperature, pH); study of such phenomena assists to envisage the mechanism of LLPS in-vivo.[8] During liquid droplet formation, weakly interacting proteins (and also oligonucleotides[9]) minimize their free energy by separating into two different phases: protein depleted dilute phase and protein enriched dense phase.[10, 11] LLPS could also induce proteins to self-aggregate over time to form amyloid fibrils,[12-14] an early signature of many neurodegenerative diseases.[15-17] Therefore, understanding the process of phase separation at a molecular level has become an emergent area of research in the recent past. It has now been accepted that mostly electrostatic as well as hydrophobic, including some weak, homotypic (protein-protein) and heterotypic

(protein-nucleic acid), interactions are responsible for the formation and the intriguing dynamic nature of LLPS.[10, 18] When protein-solvent interaction exceeds a certain value of the enthalpy of the combined effect of protein-protein and solvent-solvent interactions, the free energy becomes positive, causing an instability of the solution, which is then resolved by a phase separation process.[19]

It is often desired to either inhibit or facilitate the LLPS process for certain biological (dis)advantages. Excipients have been known to promote or inhibit LLPS formation and thus also the consequent amyloid fibril formation by modifying the intra and/or inter-molecular interactions. Different kinds of excipients,[12] like macromolecules, salts, proteins, oligonucleotide etc. have been employed to observe their effects on protein LLPS process.[20-24] Depending upon their specific natures, some of the excipients (e.g. polyethylene glycol, ficoll, dextran etc.) promote protein phase separation[22, 23, 25] while there are reports of excipients to inhibiting the process (e.g. urea for γD-crystallin,[22] human Lamin B receptor peptide for nPhos-heterochromatin protein 1,[26] phase separation of SNAP protein is inhibited by BSA, lysozyme and RNase A[27]). Even some excipients (e.g. bovine serum albumin) could exhibit a dual nature by promoting the phase separation of a driver protein and on the other hand suppressing that for others.[27]

It has now been established that alteration of protein-protein interaction and/or protein conformational stability are the crucial factors influencing the phase separation process. While LLPS process has thoroughly been investigated from the protein's perspective (in terms of altering its native configuration and/or its immediate environment), less attention has been paid on the possible role of solvation mediated protein-protein interactions, in spite of the fact that solvation plays a key role in all biophysical processes in terms of their stability and activity.[28] Or putting the statement inversely, how the solvation of proteins alter as the LLPS sets in, has seldomly been considered; recognition of which could serve as an early detection of the process. Recently, our group, using terahertz (THz) spectroscopy, have explored the alteration of protein solvation during fibrillation pathway of a model protein, bovine serum albumin, which could act as a marker for the onset of the protein fibrilization process.[29] This observation assimilates a pertaining question about the possible alteration in protein solvation during the LLPS process also, which has very sparsely been addressed so far. In a very recent study, the change in the H-bonding network of water during the phase separation of *fused in sarcoma* (FUS) protein has been reported using attenuated total reflection (ATR) spectroscopy in the

THz frequency range[30]. Such studies are even more limited in presence of excipients with fundamental questions remaining unaddressed: that excipients themselves often alter the water network dynamics and whether such a perturbation could in turn assist/resist the LLPS stability; or inversely, whether excipients directly interact with proteins, resulting in the onset of LLPS, which in turn produces an overall alteration in its solvation. The present investigation is an attempt to experimentally address this issue.

We experimentally explore the competition/connection between the presence of excipients and protein's solvation dynamics to regulating the phase separation process of a model protein lysozyme (Lys). We have used four different types of excipients: sucrose, L-arginine (Arg), bovine serum albumin (BSA) and ubiquitin (Ubi). Sucrose is a well-known molecular crowder and protein stabilizer.[31] BSA is a globular protein consisting of 583 amino acids, often used as a blocking agent in immunohistochemistry.[32] Ubi is a smaller protein (76 amino acids) which is widely used in eukaryotic cells to modify proteins and modulating their functions.[33] We observe that addition of sucrose facilitates LLPS process of Lys while BSA inhibits it. The effect of Ubi on LLPS stability has been found to be the dependent on its concentration. As both BSA and Ubi offer significant imprints on the LLPS process, we also use Arg, a basic amino acid, as an independent excipient and we found that low concentration of Arg destabilizes the LLPS process while beyond a certain concentration the effect is nominal. We use differential interference contrast (DIC) microscopy to establish the formation of LLPS droplets in Lys in presence of the excipients. Fluorescence measurements using Thioflavin-T as the fluorophore confirm that no aggregated species (e.g. fibrils) or oligomers was produced throughout the process. We finally explore the change in the protein solvation during the LLPS process using Fourier transform infrared (FTIR) spectroscopy in the THz frequency region (50-700 cm^{-1}; 1.5-21 THz). THz spectroscopy has now been established as a powerful label free tool to probe the dynamical coupling of intermolecular collective hydrogen bond network of water; more specifically, intermolecular translational motion of H-bonds and librational motion of water in its three dimensional network leave their decisive imprints in this elusive frequency window.[34, 35] The uniqueness of these measurements lies in the estimation of dynamic fluctuation in the collective water dipole moment which enables to detect the solvation dynamics without the aid of any external probe.[36] A Fourier transform in the correlation of fluctuations in the collective dipole moment results in the frequency dependent absorption coefficient ($\alpha(\nu)$), which can experimentally be estimated. Thus, any fluctuation of dipole moment in protein associated water molecules induces a change in absorption coefficient

allowing to detect the alteration in solvation. We observe that both the intermolecular translational motion of H-bond (HB stretch) and their librational motion do suffer changes during the LLPS process establishing a definite alteration of protein solvation. Addition of excipients changes the spectral characteristics of both the bands reflecting further alteration in water network. Our study exhibits a strong correlation between the stability of LLPS in presence of excipients and the associated solvation.

6.II Experimental Section

Lys was dissolved in a 200 mM buffer (KCl and NaOH) of pH 12.6; the isoelectric point of Lys is 11.2.[37] We observe that Lys forms LLPS only above the isoelectric point and no phase separation is observed at representative pH values of 2 or 7.4. It is to note here that Lys undergoes phase separation only in the presence of 200 mM KCl. In absence of KCl i.e. 200 mM NaOH solution (pH 12.6) did not produce any turbid solution. Temperature dependent turbidity measurements were carried out in a Shimadzu U2600 UV-visible spectrophotometer by monitoring the absorbance at 400 nm.[22] Leica (DM6 M) microscope was used to obtain the images. All the images were then processed in Mathematica 9 software. We recorded the fluorescence spectra of ThT in a Fluorolog (Jovin Yvon) fluorimeter. We kept both the ThT and protein concentrations in the solutions to be 10 μM for the fluorescence measurements and excited the ThT at 440 nm and collect the emission in 455 -550 nm window. We measured the secondary structures of Lys under different conditions by using circular dichroism (CD) spectroscopy (model number: Jasco J815). CD signal was collected in 190-260 nm wavelength region with a scanning speed of 50 nm/min. Far infrared (FIR) spectra in the range of 50-700 cm^{-1} were collected in a Vertex 70V (Bruker, Germany) fourier transform infrared (FTIR) spectrometer equipped with a DLaTGS detector. Details of the FTIR spectroscopy, data acquisition and analysis technique has been described in chapter 2.

6.III Results and discussions

Turbidity measurement: To determine the physiological condition for Lys phase separation we vary both the protein concentration and the temperature keeping the pH fixed at 12.6 (the isoelectric point of Lys is 11.2.[37] We observe that Lys forms LLPS only above the isoelectric point and no phase separation is observed at representative pH values of 2 or 7.4). We construct a temperature dependent phase diagram (figure 6.III.1) monitoring the solution turbidity on the basis of the fact that phase separation brings on turbidity in the solutions. In the phase diagram,

the blue shaded regions indicate a concentration-temperature phase space where LLPS sets in. We fix the Lys concentration at 600 μM for the rest of the studies as it undergoes phase separation over a wide range of temperature (figure 6.III.1a). The Lys solution is then incubated at different temperatures and turbidity of each solution is recorded as a function of incubation time. The Lys solution starts getting turbid after ~2 min of incubation (incubation temperature >35 °C) and remains turbid up to a certain time (depending on the physical parameters like concentration, temperature etc.) beyond which it becomes clear again. As the solution gets clear, it remains in that state even after a long time of incubation (more than 100 hrs). A representative plot of the turbidity of Lys in presence of 100 mM Arg as a function of time is provided in figure 6.III.1b. For a quantitative understanding we define *turbid time*, τ_{Turbid} (difference between the times when the turbidity of the solution becomes half of its maximum value: $\tau_{Turbid} = |t_2 - t_1|$, t_1 and t_2 are the respective time when turbidity is half of its maximum value, see figure 6.III.1b). τ_{Turbid} provides with an estimation of the time duration during which the protein remains phase-separated; the larger is the value of τ_{Turbid} the more favoured is the LLPS process while a smaller value of τ_{Turbid} suggests suppression of LLPS. To identify temperature dependency on LLPS stability of Lys we vary the incubation temperature and observe that increasing the incubation temperature from 37 °C to 70 °C causes a drastic decrease in τ_{Turbid} from ~30 h to ~85 min; such a decrease a priori suggests that the stability of Lys phase separation is dependent on the incubation temperature.

We next check the influence of the excipients on the Lys phase separation behaviour. We fix the incubation temperature at 70 °C and measure τ_{Turbid} in presence of the excipients. The value of τ_{Turbid} for Lys is ~85 min; in presence of 30 mM Arg it decreases to ~66 min; however, then it increases to ~115 min as the Arg concentration increases and does not change appreciably up to 300 mM Arg (figure 6.III.1c). Amino acids like Pro and Arg have been reported to suppress the phase separation of monoclonal antibodies (mAbs) and the phenomenon has been attributed to the shielding of attractive intermolecular electrostatic (direct) interaction in proteins.[38] Arg, in the present case, also reduces the attractive protein-protein interaction of Lys causing the observed reduction of τ_{Turbid} at lower concentrations. However, at higher Arg concentration (>100 mM) it starts acting more as a conventional molecular crowder and due to the volume exclusion effect, it starts to stabilizing the LLPS. Earlier studies have also reported stabilization of γD crystallin LLPS owning to excluded volume effect offered by polyethylene glycol.[39] At even higher concentrations, perhaps these two interactions balance to produce no net effect on the LLPS formation.

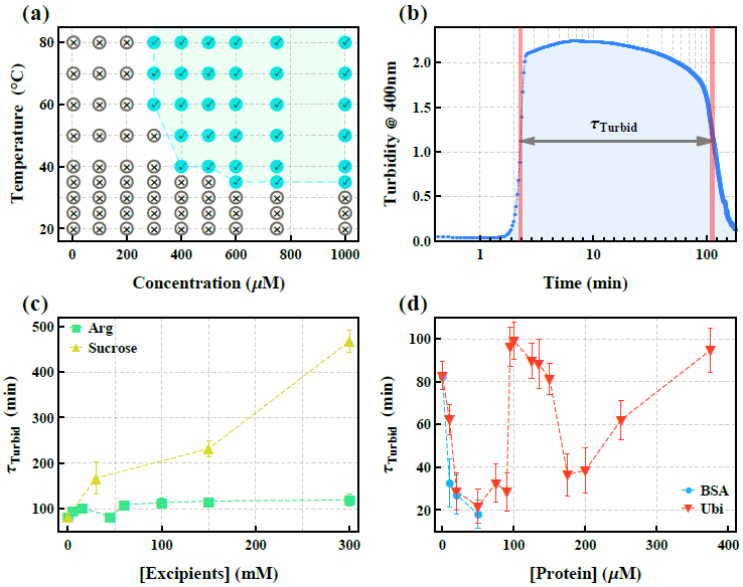


Figure 6.III.1. (a) Temperature induced phase plot of Lys in 200 mM buffer of pH 12.6. The blue circles (right marked) indicate a concentration-temperature phase point at which the protein gets phase separated and the grey circles (cross marked) indicate a single phase. (b) Turbidity of Lys (measured from the absorbance at 400 nm) as a function of time in presence of 100 mM Arg at 70^0 C. τ_{Turbid} is defined as the time difference when the solution turbidity becomes half of its maximum value. (c) and (d) show τ_{Turbid} as a function of excipient (Arginine, Sucrose, BSA, Ubi) concentration (concentration and temperature of the Lys solution is kept fixed at 600 μM and 70^0 C, respectively).

We, on the other hand, observe a rather drastic enhancement in the τ_{Turbid} value in presence of sucrose and this, in fact, increases monotonically with the sucrose concentration (figure 6.III.1c) indicating continued stabilization of LLPS. It perhaps is the site-specific weak binding of the sugar molecules with the aromatic amino acids present in the Lys that overwhelms the steric exclusion effect of sucrose causing the eventual stabilization of the LLPS process of Lys.[40]

We now investigate the effect of proteins (BSA and Ubi) added as the external crowders. We observe that in the low BSA concentration regime (up to 50 μM), τ_{Turbid} decreases sharply to ~27 mins (figure 6.III.1d) suggesting a suppression of the LLPS process. Interestingly, beyond 50 μM BSA, no turbidity in the solution is observed, which implies a complete inhibition of Lys phase separation as induced by concentrated BSA. Ubi, on the other hand, produces an unique and non-monotonous concentration dependency on the LLPS behaviour (figure 6.III.1d). τ_{Turbid} decreases sharply (up-to 20 μM Ubi), in a manner similar to that obtained in case of BSA, and then only mild changes are registered up to 90 μM. At an Ubi concentration beyond this, τ_{Turbid} rises sharply to reach a value of ~100 min, which is even higher than that of pure Lys. It decreases mildly up to 150 μM Ubi, and again drops

sharply up to 200 μM Ubi; at an even higher concentration it increases monotonically. It is intriguing to note that neither pure BSA nor pure Ubi produces any phase separation under the same pH condition and at the same concentration range studied here. Thus, any observed change in the turbidity emanates solely from the Lys phase separation. Protter et al. have previously observed the inhibition of LLPS of hnRNPA1$_{IDR}$ protein in presence of 500 μM BSA and the observation has been attributed to the specific as well as promiscuous non-specific interactions involving weak binding, long range electrostatic, cation-π and π-π interactions between BSA and IDRs.[27] Ubi has previously also been applied to gradually reduce phase separation of ubiquilin2 (UBQLN2) (complete inhibition takes place at 1:1 ratio); the effect being attributed to the specific non-covalent interaction between Ubi and the Ubi associating domain (UBA) of UBQLN2, which disrupts the multivalent interaction of UBQNL2 involving the UBA domain.[41] However, the non-monotonous behaviour as observed in the present investigation, has not been reported earlier.

Imaging of LLPS: At this point it appears important to investigate whether the increase in solution turbidity arises due to the formation of LLPS or any other type of aggregated species (like amyloid fibril etc.) do also appear. To establish this, we monitor the solutions using differential interference contrast (DIC) microscopy (figure 6.III.2). We observe that 600 μM Lys does not form any droplet at room temperature (figure 6.III.2a). After incubation at 50 °C we notice the appearance of spherical droplets of various diameters (sub-μm to μm) (figure 6.III.2b), which unambiguously establishes the formation of LLPS only.[14] For a quantitative view we plot the droplet distribution with different droplet diameters (figure 6.III.2g). Formation of higher number of droplets with larger diameters compared to pure Lys signifies the stabilization of LLPS process while an opposite observation stands for the inhibition process. Addition of 100 mM Arg in the Lys solution does not modify the droplets much (figure 6.III.2c). More droplets are found in sucrose (10 mM) containing Lys solution after incubation suggesting the stabilization of LLPS (figure 6.III.2d). Addition of 50 μM BSA, on the other hand, reduces the droplet size and their numbers also (figure 6.III.2e), reflecting suppression of LLPS. The effect of Ubi in the LLPS process is concentration dependent. Addition of 50 μM Ubi induces less droplet formation compared to that at 100 μM. The microscopy study thus infers that it is the phase separation rather than protein aggregation that is responsible for the observed alteration in the turbidity in the solutions (figure 6.III.1c and d). This study also verifies the impacts of excipients on LLPS process as obtained from the τ_{Turbid} measurements.

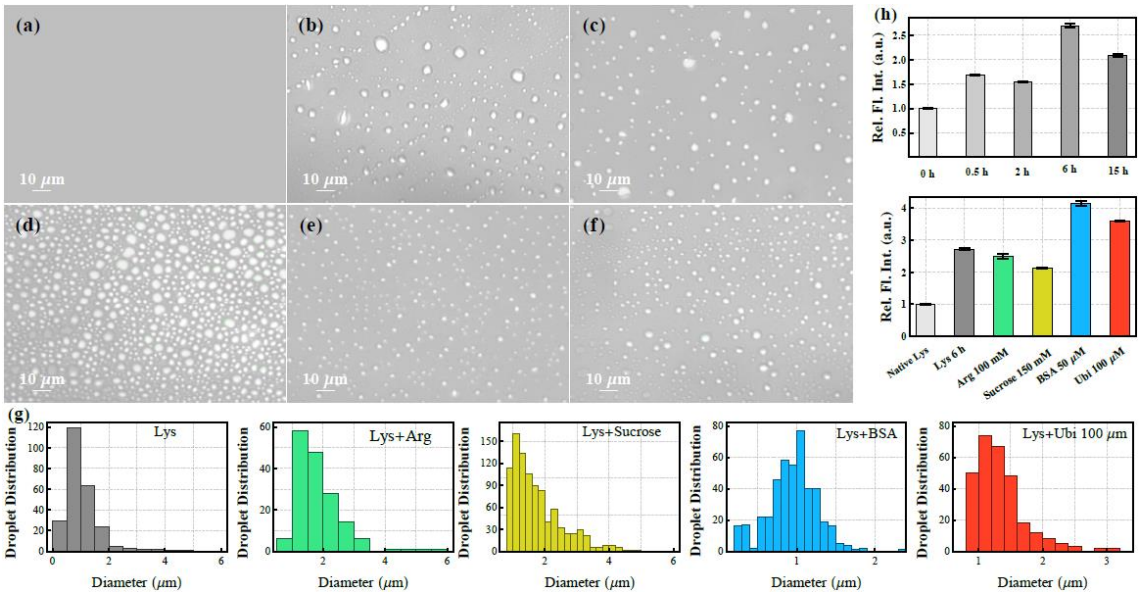


Figure 6.III.2. (a)-(f) Microscopic image of Lys under different conditions: (a) No phase separation (b) Only Lys (600 µM), (c) Lys-100 mM Arg solution, (d) Lys- 10 mM sucrose solution, (e) Lys-50 µM BSA solution and (f) Lys-100 µM Ubi solution. All the measurements were carried out after 6 h incubation at 50^0 C. Concentration of Lys was kept fixed at 600 µM and the pH of the solution was 12.6. (g) Droplet distribution with different diameters of the images shown in (b)-(f). (h) Relative fluorescence intensity of Thioflavin-T (ThT) with respect to native Lys (without incubation) under different conditions: Lys in buffer under different time of incubation (upper panel), in presence of excipients after 6 h of incubation (lower panel).

ThT Fluorescence measurements: To further strengthening the observation that no oligomers or rigid fibrils have been formed, we perform Thioflavin-T (ThT) fluorescence measurements based on the fact that fluorescence intensity of ThT enhances significantly when bind to oligomers and/or fibrils.[42] We observe that upon incubation of Lys at 50 °C for different time the relative fluorescence intensity of ThT (defined as the fluorescence intensity with respect to that obtained in the native protein) does not enhance appreciably (figure 6.III.2h, upper panel) obscuring the possibility of any oligomers or fibrils formation. In the presence of excipients also (after incubation for 6 hrs.), we observe that the relative fluorescence intensity does not change appreciably (figure 6.III.2h, lower panel) which unambiguously establishes the fact that no aggregated structure is formed in Lys even in the presence of excipients.

Hydration (solvation) during LLPS: Upon establishing the fact that excipients do modulate the LLPS process in Lys, we now focus on to the associated change in protein hydration using THz spectroscopy. Water shows a strong absorption feature in this frequency window.

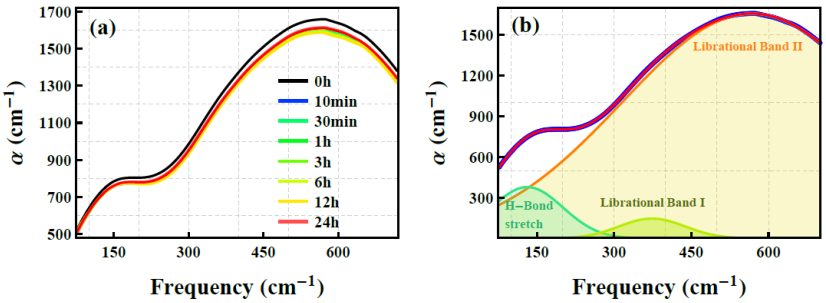


Figure 6.III.3. (a) Representative absorption coefficient of Lys solution undergoing LLPS after different time of incubation (Lys concentration was 600 μM and the pH of the solution was 12.6). (b) Deconvolution of absorption of water into three sub gaussian bands.

Also various intramolecular vibrations of proteins including the skeleton motions exhibit their signature in this frequency region.[43] It is therefore expected that the water vibrational modes would be coupled with the proteins and would offer specific signature in the overall solvation dynamics during the phase separation. We measure the frequency dependent absorbance of these systems in the THz fingerprint region and estimate the frequency dependent absorption coefficients, α(ν) (see section 2.II.D.b for details). In this frequency window bulk water offers two intense bands peaking at ~ 130 cm^{-1} associated with the intermolecular H-bond stretching and at ~570 cm^{-1} associated with the intermolecular librational (hindered rotation) motion of water molecules (figure 6.III.3).[36] The librational band can further be deconvoluted into two sub-gaussian bands having peak frequencies at ~370 and ~565 cm^{-1} (figure 6.III.3b) and these peaks can be attributed to the intermolecular in-plane and out of plane O-H librational bands, respectively.[43-45] Lys solution in buffer has an altogether lower value of α(ν) compared to that of bulk water itself (figure 6.III.3a) due to the replacement of water molecules with low absorbing protein molecules. Such a decrease in the THz absorption profile of water in presence of small molecules and protein have previously been reported.[46, 47] In order to identify the explicit change in the protein solvation during the LLPS process we measure the difference in absorbance, $\Delta\alpha(\nu)$ $(= \alpha_{sample}(\nu) - \alpha_{buffer}(\nu))$ and observe a significant change in the $\Delta\alpha(\nu)$ profile upon varying the time of incubation (figure 6.III.4a), which unambiguously affirms that protein solvation does get altered during the LLPS process. For a quantitative insight onto the collective hydrogen bond network of protein solvation and also on the modification of its tetrahedral structure (as obtained from the alteration in the librational band[36]) during LLPS, we deconvolute the $\Delta\alpha(\nu)$ profile using a sum of a three component damped harmonic oscillation model (see section 2.I.B for details) as follows:[48]

$$\Delta\alpha(\nu) = \sum_{i=1}^{3} \frac{a_i \omega_i \nu^2}{\nu^2 \omega_i^2 + \pi^2 \left(\nu_i^2 + \frac{\omega_i^2}{4\pi^2} - \nu^2 \right)^2} \quad (6.1)$$

where a_i, ω_i, ν_i are the amplitude, width, and the centre frequency of the i^{th} resonance. The unperturbed centre frequency can be defined as,

$$\nu_{0,i} = \sqrt{\nu_i^2 + \frac{\omega_i^2}{4\pi^2}}. \quad (6.2)$$

A representative deconvolution analysis of native Lys solution is shown in figure 6.III.4b; we obtain three curves having peak frequencies at ν_1~ 103 cm^{-1} (peak 1), ν_2~ 355 cm^{-1} (peak 2) and ν_3~510 cm^{-1} (peak 3) (figure 6.III.7; hollow symbols). Addition of Lys shifts all the bands towards low frequency manifesting more hydrogen bond formation[49] along with less hindered rotation. Upon incubation vis-à-vis phase separation, a red shift (~ 26 cm^{-1}) in ν_1 (HB stretching) is observed up-to 1 hr of incubation (figure 6.III.7; upper panel, hollow symbol), then it is blue shifted (by 50 cm^{-1} compared to 1 hr incubation) (figure 6.III.7, upper panel). This red or blue shift in ν_1 indicates respectively the formation and dissociation of HB in the protein solvation layer.[49] We observe a similar red followed by a blue shift in both the other two peaks (ν_2 and ν_3) (figure 6.III.7; middle and lower panel). The observed peak shift (either blue or red) of the librational bands represents a definite alteration in the hindered rotation of water molecule.[30] Recently Ahlers et al. have established the contribution of three different types of water population in protein: bound (interacting directly with the polar amino acids), wrap (water population around hydrophobic amino acids of protein) and bulk (which do not interact with the protein and far away from the protein) during the phase separation process of FUS protein and concluded that phase separation removes the water molecules from the protein surface rupturing the hydration shell which enhances the FUS concentration inside the droplet.[30] The observed changes in the solvation dynamics in the present study is also in line of the previous observation. It is interesting to mention here that beyond the completion of phase separation, i.e. 12 hrs after the incubation, when the solution becomes clear again, the peak maxima do not necessarily return to the values obtained before the phase separation condition, which unambiguously mark a permanent perturbation of the protein structure during the LLPS process.

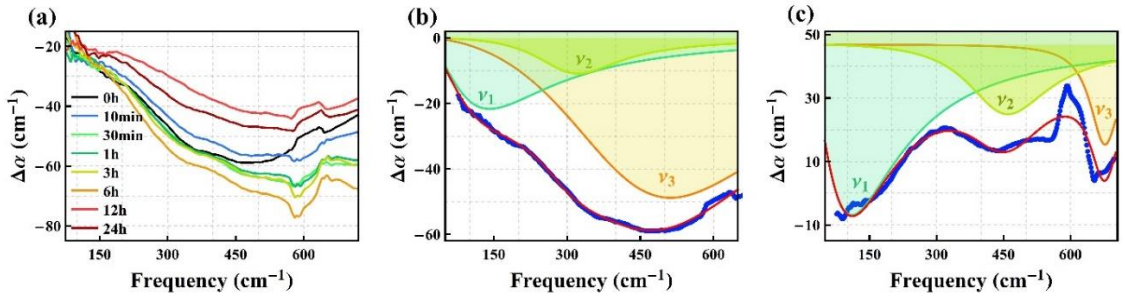


Figure 6.III.4. (a) Change in absorption co-efficient ($\Delta\alpha$) as a function of frequency of Lys in buffer in its native state and under incubation (at 50^0 C) for different time. A representative deconvolution of the $\Delta\alpha$ profile into three different vibration bands (HB stretching (ν_1) and librational modes (twisting, ν_2 and rocking, ν_3))) for (b) native Lys (c) and Lys-sucrose (150 mM) solution after incubation for 24 h (see equation S4 for details). The blue line is the experimental curve and the red line is the overall fitted curve.

To understand whether this irreversibility is related to protein conformational change during phase separation, we study the protein's secondary structure using CD spectroscopy (figure 6.III.5) before and after the phase separation process and observe that upon incubation (10 mins of incubation) the CD signal quenches substantially, indicating unfolding of the protein. After 1 hrs of incubation, the CD signal starts to increase again and after complete phase separation (after 12 hrs of incubation) the CD signal does not change further and does not return to its native state (before incubation). These studies suggest that alteration of protein conformation and the associated change in its solvation during the phase separation process are correlated. Our group has recently established similar correlation between structural makeover and change in protein solvation during protein fibrillation pathway.[29]

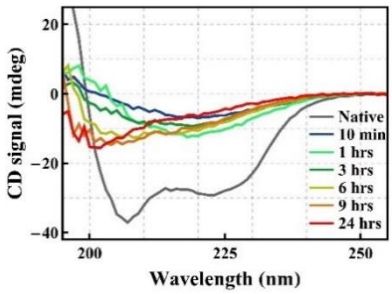


Figure 6.III.5. FAR UV CD signal (mdeg) of Lys after different time of incubation. The concentration of each solution was diluted to be 20 µM before measurement as the Lys of 600 µM solution causes a signal saturation in CD spectrum.

The above results unambiguously affirm that solvation around Lys during LLPS formation do get altered as excipients are added. Before we progress to any further analysis it is important to check how the excipients themselves alter the protein solvation. We measure the change $\Delta\alpha(\nu)$ in excipients water solutions (figure 6.III.6). We observe a reduction in the HB stretching peak frequency from 159 to 99 cm^{-1} with the Arg concentration while ν_2 moves from 433 to 398 cm^{-1} and ν_3 from 661 to 600 cm^{-1} (figure 6.III.6a) implying a modification of the water network in presence of Arg molecules. As we fit the $\Delta\alpha(\nu)$ of sucrose solution we observe a blue shift in ν_1 (frequency increases from ~95 cm^{-1} to 105 cm^{-1}) with sucrose concentration from 10 mM to 150 mM; however, beyond that sucrose concentration the change is subtle (figure 6.III.6b). A gradual decrease of ν_2 is also observed whereas ν_3 increases regularly. The changes are also monotonic in Ubi-water solution; we observe a linear decrease of ν_1 (from ~140 cm^{-1} to ~60 cm^{-1}) with Ubi concentration while ν_2 and ν_3 increases monotonically with the increase of Ubi concentration (figure 6.III.6c). Addition of BSA in water also changes water dynamics as reflected in the change in $\Delta\alpha(\nu)$, but due to very small BSA concentration we could not fit the data. These observations affirm that excipients themselves do modify the water dynamics, however, the changes are more or less monotonically dependent on the excipient's concentration.

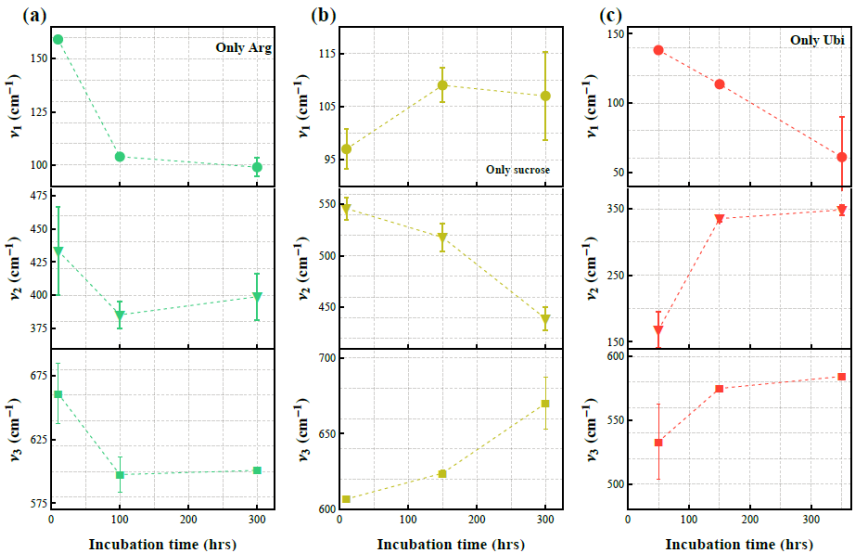


Figure 6.III.6. Fitting parameters, ν_1, ν_2 and ν_3 as a function of concentration of (a) Arg, (b) sucrose, (c) and (c) Ubi.

With the background of phase separation details of Lys we now investigate how the water network modifies in presence of the excipients during the LLPS process. Arg of different

concentrations (10, 100 and 300 mM) is added to Lys solution and each sample is then incubated for different periods of times. After incubation we measure the THz absorption profile of each sample and fit the $\Delta\alpha(\nu)$ curves in a manner similar to that in case of Lys solution. We find that all the peaks are shifted towards higher frequency compared to that in the Lys solution (figure 6.III.7), the extent of the shift is found to be Arg concentration dependent. We do observe comparable peak shifts in ν_1 (~15 cm^{-1} towards higher frequency) in 10 and 100 mM Arg whereas addition of 300 mM Arg shows 25 cm^{-1} blue shift (figure 6.III.7a; upper panel). The ν_2 peak frequency is found to be higher (~100 cm^{-1}) compared to the bare Lys solution (figure 6.III.7a; middle panel). A significant blue shift (~75 cm^{-1}) in ν_3 is also observed at all Arg concentrations (figure 6.III.7a; lower panel). Addition of sucrose also produces changes in the three peak frequencies (figure 6.III.7b, a representative deconvolution is shown in figure 6.III.4c). A red shift in ν_1 (~35 cm^{-1}) is observed upon the addition of 150 mM sucrose (figure 6.III.7b; upper panel); it then increases mildly beyond three hrs. of incubation; however, the $\nu_1(t)$ profile always remain red shifted compared to that of Lys solution. In presence of 300 mM sucrose an initial blue shift is observed, however, after 1 hr of incubation it gets red shifted compared to the Lys solution, which indicates a stronger H-bonded environment causing Lys LLPS to stabilize more, as has also been evidenced from the turbidity measurements (figure 6.III.1d). A blue shift is observed in ν_2 and ν_3 compared to Lys upon the addition of sucrose (figure 6.III.7b; middle and lower panel). We notice an initial change in ν_2 and ν_3 between 150 and 300 mM sucrose. However, after 10 hrs. of incubation no noticeable change is detected. This suggests that addition of sucrose induces restricted hindered motion of water molecule around Lys. It is interesting to note here that sucrose induces a blue shift in both $\nu_2(t)$ and $\nu_3(t)$ after phase separation of Lys compared to the native Lys.

We then explore the effects of the proteins on LLPS process. Addition of 20 µM BSA produces an initial blue shift in ν_1 (figure 6.III.7c; upper panel). After 1 hr of incubation, we interestingly observe a crossing in the $\nu_1(t)$ profile between BSA/Lys and pure Lys solutions and an overall red shift of 40 cm^{-1} is also noticed. With 50 µM BSA, however, we observe a monotonic blue shift in $\nu_1(t)$. A nominal change in ν_2 is observed upon the addition of 20 µM BSA (figure 6.III.7c; middle panel). At a BSA concentration of 50 µM, $\nu_2(t)$ changes first modestly, and then a sharp enhancement is observed after 3 hrs. of incubation, which eventually decreases beyond 6 hrs. of incubation to reach a value similar to that of Lys. Such a sharp enhancement is associated with more restricted motion of water molecules. We notice a blue shift followed by a red shift in ν_3 upon addition of 20 µM BSA and a reverse trend with further

addition of BSA (50 µM) (figure 6.III.7c; lower panel). In case of Ubi also, we observe comparable ν(t) profiles for all the three peak frequencies. We notice a nominal change in ν_1 upon the addition of 50 µM and 150 µM Ubi (figure 6.III.7d; upper panel); however, the increase is significant (~ 170 cm^{-1}) as the concentration is increased to 350 µM which diminishes beyond 1 hr. of incubation and beyond 12 hrs. of incubation the ν_1 is red shifted. ν_2 increases regularly with Ubi concentration and we observe a blue shift in all Ubi mediated Lys solutions compared to bare Lys solution after any incubation time (figure 6.III.7d; middle panel). Addition of 50 µM Ubi causes a red shift in ν_3, however, further addition of Ubi does not alter the ν_3 much in comparison to only Lys solution (figure 6.III.7d; lower panel).

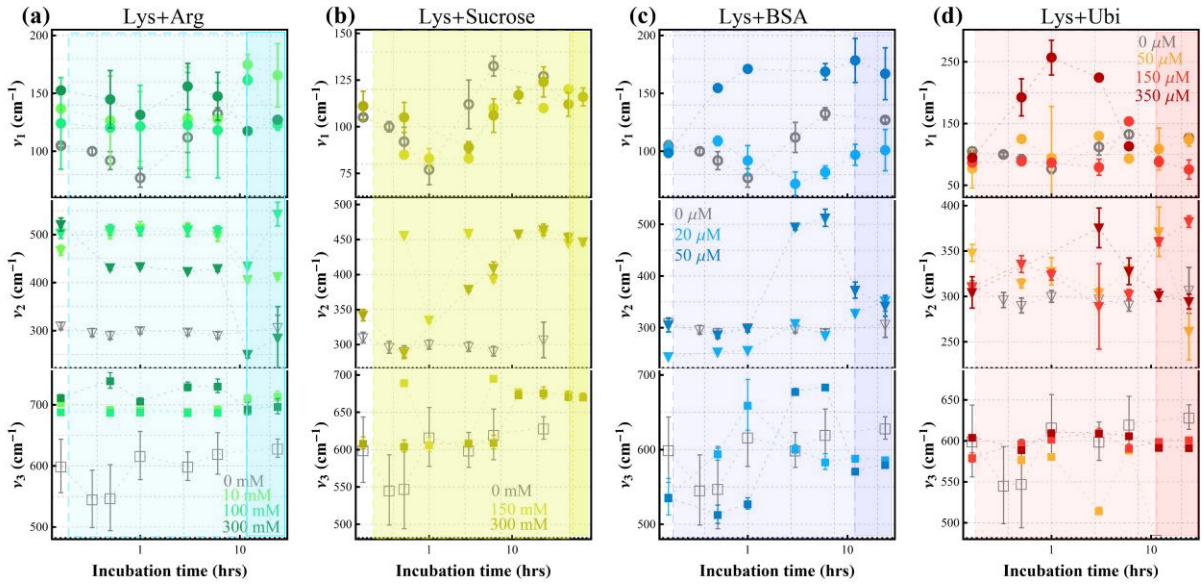


Figure 6.III.7. Fitting parameters, ν_1, ν_2 and ν_3 for the various vibrational modes of water around Lys (see equation S4 for details) as a function of incubation time in presence of (a) Arg (10, 100 and 300 mM), (b) sucrose (150, 300 mM), (c) BSA (20, 50 μM) and (d) Ubi (50, 150 and 350 μM). The hollow grey symbols represent Lys in absence of any excipient.

The above analysis suggests that the effect of excipients on protein solvation is not straight forward, rather it could vary non-linearly depending mostly upon two external parameters: excipient concentration and incubation time (at a constant temperature). Since our analysis also takes care of three different vibrational modes of water the complexity in the analysis even enhances. To quantitatively summarise the global effect of excipients on the protein solvation during the LLPS process we perform principal component analysis (PCA) (see section 2.I.E for details) on Δα for each excipient. Representative profiles of the first two

principal components (PCs) along with the scores and the contributions of each PC of Lys solution during the phase separation process is shown in figure 6.III.8. The first PC of Lys during the LLPS formation offers a contribution of >99% and the score changes nonlinearly with the incubation time. The 2nd PC increases to ~5% upon the addition of the excipients (a representative plot is shown in figure 6.III.8c in presence of 10 mM Arg). Since the first PC has overwhelmingly high contribution (typically >95%) for each sample (pure lysozyme and Lys containing the excipients) we fit the first PC by the damping equation (equation 6.1). The results of the fitting are summarized in table S1. We notice that protein solvation undergoes an overall perturbation in terms of alteration of both *in plane* and *out of plane* librational modes as well as the HB stretching modes during the LLPS process in presence of the excipients (as also evidenced from figure 6.III.7).

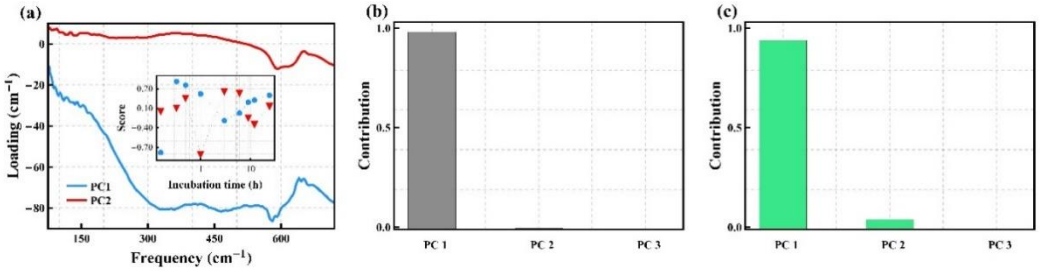


Figure 6.III.8. (a) Representative principal components (PC) of Lys solution during the phase separation process. Score of corresponding PC as a function of incubation time are shown in the inset. Contribution of the first three PC (b) only in Lys and (c) in presence of 10 mM Arg.

For a deeper insight we make a relative comparison ($\nu_{Lys\text{-}ex}/\nu_{Lys}$; where $\nu_{Lys\text{-}ex}$ and ν_{Lys} indicate the fitted peak frequency of a particular band of phase separated Lys in presence and in absence of excipients, respectively) for both HB stretch and librational modes of solvated water during the LLPS process as a function of the excipients (figure 6.III.9). We observe a relative blue shift in the peak frequency for both the librational modes and the HB stretch bands during the LLPS process in presence of all the excipients, however, interestingly barring BSA, where we observe a completely reverse trend. This analysis thus points out that protein solvation during the LLPS process in the presence of BSA behaves in a manner completely opposite that of the other excipients (Arg, Sucrose, Ubi).

Our present study has been focused to address two major aspects: whether external agents could modify LLPS process and if so, how does that connects with the protein hydration? We have

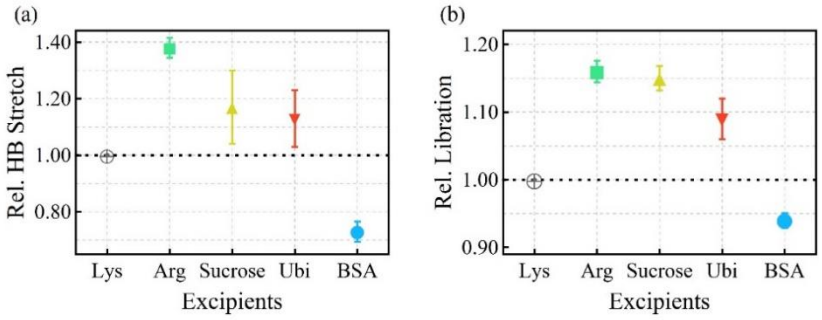


Figure 6.III.9. Relative change of (a) HB stretch and (b) Librational motion in the overall hydration during the LLPS formation of Lys in presence of excipients (considering LLPS in Lys as the unity for both HB stretching and librational mode)

---

deployed two rather conventional crowders, sucrose and arginine, and two otherwise indifferent macromolecules, ubiquitin and BSA. We found that the answer of the first question is straightforward, and we conclude that macromolecular crowders can definitely affect LLPS. Both sucrose and Arg facilitates LLPS formation, the effect being much pronounced in sucrose. Earlier studies from our group have established that both arginine and sucrose act as water structure makers as they can form slow moving solvation layer around themselves.[50, 51] Thus, they induce a volume exclusion effect on the protein solvation, thereby assisting the LLPS formation. BSA, on the other hand, surprisingly opposes the process, in spite of the fact that it possesses a slow solvation layer around itself, as of the other conventional crowders. Ubi shows some unique non-monotonous behaviour, at low concentrations, it behaves like BSA, however, at higher concentrations it either facilitates or represses the LLPS process. This unique behaviour leads us to conclude that conventional crowders and macromolecules behave differently with proteins during the LLPS process. One way to understand such behaviour is to look into their solvation nature.

It is known that water molecules at the protein solvation layers offer different behaviour than bulk. We also have noticed that the vibrational peaks suffer considerable shifts in protein solutions. It is to note here that since we are fitting the $\Delta\alpha(\nu)$ curves, it subtracts the contribution of the bulk water (also to note here that proteins share very small contribution to the overall absorbance) and provides the explicit information of protein solvation. Any non-zero value of $\Delta\alpha(\nu)$ indicates a definite modification in the bulk water structure. In all the cases, we observe a negative $\Delta\alpha(\nu)$, which manifests a "*THz defect"* as solvation shell absorbs less than the bulk water.[51, 52] Further analysis using the damping vibrational model provides with

the additional information on the specific protein-water interaction modes. Since we have carried out a vast number of analyses a global picture on the solvation is achieved employing a PCA, which makes an overall comprehension easier looking into figure 6.III.9. If the ν_1 peak shifts towards a higher frequency it signifies a hindered rotation[30] and/or feebler H-bonded network[49] compared to that in bulk, which we have indeed observed in case of the excipients. On the other hand, presence of BSA shifts the relative peak frequency to a lower value signifying a higher number of H-bond formation[49] and less restricted motion. This behaviour is also consistent with the observed stabilization of the LLPS by excipients (as evidenced from τ_{Turbid} observed from figure 6.III.1c and 6.III.1d). It is to note here that LLPS is a thermodynamically controlled process and in the present systems it can be approximately estimated in terms of an enthalpic term, Flory parameter (as described in detail in section 2.I.A). For the LLPS process in presence of excipients this Flory parameter gets modified. Depending on whether LLPS gets stabilized or inhibited in presence of excipients would be decided on the sign of $\Delta F = F_p^e - F_p$ (where F_p and F_p^e are respectively the free energy of protein undergoing LLPS in absence and in presence of excipients). If $\Delta F > 0$ the excipient will stabilize the LLPS (like the sucrose in the present case) whereas for BSA, $\Delta F < 0$ as it inhibits the LLPS process. Thus, for the excipients which stabilize LLPS ($\Delta F > 0$) must satisfy $\chi_p^e > \chi_p$ (χ_p^e is the modified Flory parameter in presence of excipients), which manifests that protein-solvent interaction is higher in presence of excipients compared to that in pure Lys. With the same argument for BSA, the protein-solvent interaction enthalpy should be less than that of pure Lys solution. Protein-solvent interaction involves mostly van der Waals forces, polar charged, ionic and hydrogen bonds;[53] thus any change in the free energy of LLPS formation would be reflected by the modification in term of strengthening/weaking of the hydrogen bonds (which are otherwise manifested in the peak shift of intermolecular HB stretch and librational motion bands) of the solvation layer around proteins (figure 6.III.9) as we probe by measuring $\Delta\alpha(\nu)$. The exact mechanism of the interaction (direct or indirect) of the excipients with the proteins, which leads to the strengthening of the H-bonds is still not very evident, perhaps a detailed simulation would lead to a step forward, however, the experimental results clearly identify that both protein and excipient solvation play decisive roles in regulating the LLPS process.

6. IV Conclusions:

In summary, our investigation establishes a possible route towards regulation (either stabilization or prohibition) of Lys LLPS formation by excipients (Arg, sucrose, BSA, Ubi), in terms of excipient concentration and incubation time. We also conclude that solvation around

Lys gets altered during the LLPS process depending upon the specific nature of the excipients. Monitoring a change in solvation, therefore, could act as a potential marker for an early and easy detection of LLPS onset. We anticipate that this in-vitro study would be found useful in understanding how LLPS occurs in vivo and with proteins with intrinsically disordered region (IDR).

6.V References:

1. A. A. Spector and M. A. Yorek, *J. Lipid Res.*, 1985, **26**, 1015-1035.
2. M. Feric, N. Vaidya, T. S. Harmon, D. M. Mitrea, L. Zhu, T. M. Richardson, R. W. Kriwacki, R. V. Pappu and C. P. Brangwynne, *Cell*, 2016, **165**, 1686-1697.
3. T. E. Kaiser, R. V. Intine and M. Dundr, *Science*, 2008, **322**, 1713-1717.
4. A. Molliex, J. Temirov, J. Lee, M. Coughlin, A. P. Kanagaraj, H. J. Kim, T. Mittag and J. P. Taylor, *Cell*, 2015, **163**, 123-133.
5. S. F. Banani, H. O. Lee, A. A. Hyman and M. K. Rosen, *Nat. Rev. Mol. Cell Biol.*, 2017, **18**, 285–298.
6. S. Boeynaems, S. Alberti, N. L. Fawzi, T. Mittag, M. Polymenidou, F. Rousseau, J. Schymkowitz, J. Shorter, B. Wolozin, L. V. D. Bosch, P. Tompa and M. Fuxreiter, *Trends Cell Biol.*, 2018, **28**, 420-435.
7. Y. Shin and C. P. Brangwynne, *Science*, 2017, **357**, eaaf4382.
8. T. M. Franzmann, M. Jahnel, A. Pozniakovsky, J. Mahamid, A. S. Holehouse, E. Nüske, D. Richter, W. Baumeister, S. W. Grill, R. V. Pappu, A. A. Hyman and S. Alberti, *Science*, 2018, **359**, eaao5654.
9. N. Martin, L. Tian, D. Spencer, A. Coutable-Pennarun, J. L. R. Anderson and S. Mann, *Angew.Chem.Int. Ed.*, 2019, **58**, 14594 –14598.
10. G. L. Dignon, R. B. Best and J. Mittal, *Annu. Rev. Phys. Chem.*, 2020, **71**, 53–75.
11. E. P. Bentley, B. B. Frey and A. A. Deniz, *Chem. Eur. J.*, 2019, **25**, 5600 – 5610.
12. S. Wegmann, B. Eftekharzadeh, K. Tepper, K. M. Zoltowska, R. E. Bennett, S. Dujardin, P. R. Laskowski, D. MacKenzie, T. Kamath, C. Commins, V. Vanderburg, A. D. Roe, Z. Fan, A. M. Molliex, A. H. Vega, D. Muller, A. A. Hyman, E. Mandelkow, J. P. Taylor and B. T. Hyman, *EMBO J*, 2018, **37**, e98049.
13. A. Patel, H. O. Lee, L. Jawerth, S. Maharana, M. Jahnel, M. Y. Hein, S. Stoynov, J. Mahamid, S. Saha, T. M. Franzmann, A. Pozniakovski, I. Poser, N. Maghelli, L. A. Royer, M. Weigert, E. W. Myers, S. Grill, D. Drechsel, A. A. Hyman and S. Alberti, *Cell*, 2015, **162**, 1066–1077.
14. S. Ambadipudi, J. Biernat, D. Riedel, E. Mandelkow and M. Zweckstetter, *Nat. Commun.*, 2017, **8**, 275
15. F. Chiti and C. M. Dobson, *Annu. Rev. Biochemi.*, 2006, **75**, 333-366.
16. M. Stefani and C. M. Dobson, *J. Mol. Med.*, 2003, **81**, 678-699.
17. S. T. Ferreira, M. N. N. Vieira and F. G. De Felice, *IUBMB Life*, 2007, **59**, 332-345.
18. A. C. Murthy, G. L. Dignon, Y. kan, G. H. Zerze, S. H. Parekh, J. Mittal and N. L. Fawzi, *Nat. Struct. Mol. Biol.*, 2019, **26**, 637–648.
19. C. P. Brangwynne, P. Tompa and R. V. Pappu, *Nat. Phys.*, 2015, **11**, 899–904.
20. A. Ghosh, K. Mazarakos and H. X. Zhou, *Proc. Natl. Acad. Sci. U.S.A*, 2019, **116**, 19474-19483.
21. W. M. Babinchak, B. K. Dumm, S. Venus, S. Boyko, A. A. Putnam, E. Jankowsky and W. K. Surewicz, *Nat. Commun.*, 2020, **11**, 5574.
22. H. Cinar and R. Winter, *Sci. Rep*, 2020, **10**, 17245.
23. Y. Lin, D. S. Protter, M. K. Rosen and R. Parker, *Mol. Cell*, 2015, **60**, 208–219
24. S. Maharana, J. Wang, D. K. Papadopoulos, D. Richter, A. Pozniakovsky, I. Poser, M. Bickle, S. Rizk, J. Guillén-Boixet, T. M. Franzmann, M. Jahnel, L. Marrone, Y. T. Chang, J. Sterneckert, P. Tomancak, A. A. Hyman and S. Alberti, *Science*, 2018, **360**, 918–921.
25. K. Julius, J. Weine, M. Gao, J. Latarius, M. Elbers, M. Paulus, M. Tolan and R. Winter, *Macromolecules*, 2019, **52**, 1772-1784.

26. A. G. Larson, D. Elnatan, M. M. Keenen, M. J. Trnka, J. B. Johnston, A. L. Burlingame, D. A. Agard, S. Redding and G. J. Narlikar, *Nature*, 2017, **547**, 236-240.
27. D. S. W. Protter, B. S. Rao, B. V. Treeck, Y. Lin, L. Mizoue, M. K. Rosen and R. Parker, *Cell Rep.*, 2018, **22**, 1401-1412.
28. D. Laage, T. Elsaesser and J. T. Hynes, *Chem. Rev.*, 2017, **117**, 10694−10725.
29. P. Pyne, N. Samanta, H. Gohil, S. S. Prabhu and R. K. Mitra, *Chem. Commun.*, 2021, **57**, 998-1001.
30. J. Ahlers, E. M. Adams, V. Bader, S. Pezzotti, K. F. Winklhofer, J. Tatzelt and M. Havenith, *Biophysical J.*, 2021, **120**, 1266–1275.
31. J. C. Lee and S. N. Timasheff, *J. Biol. Chem.*, 1981, **256**, 7193-7201.
32. R. Ahirwar, S. Bariar, A. Balakrishnan and P. Nahar, *RSC Adv.*, 2015, **5**, 100077.
33. D. Mukhopadhyay and H. Riezman, *Science*, 2007, **315**, 201-205.
34. J. B. Brubach, A. Mermet, A. Filabozzi, A. Gerschel and P. Roy, *J. Chem. Phys.*, 2005, **122**, 184509.
35. G. Schwaab, F. Sebastiani and M. Havenith, *Angew. Chem. Int. Ed. Engl.*, 2019, **58**, 3000–3013.
36. M. Heyden, J. Sun, S. Funkner, G. Mathias, H. Forbert, M. Havenith and D. Marx, *Proc. Natl. Acad. Sci. U.S.A.*, 2010, **107**, 12068-12073.
37. M. Bostro¨m, F. W. Tavares , S. Finet, F. Skouri-Panet, A. Tardieu and B. W. Ninham, *Biophys. Chem.*, 2005, **117** 217 – 224.
38. D. D. Banks and J. F. Cordia, *Mol. Pharmaceutics*, 2021, **18**, 1285−1292.
39. O. Annunziata, N. Asherie, A. Lomakin, J. Pande, O. Ogun and G. B. Benedek, *Proc. Natl. Acad. Sci. U.S.A*, 2002, **99**, 14165–14170.
40. A. S. Raut and D. S. Kalonia, *Mol. Pharmaceutics*, 2016, **13**, 774−783.
41. T. P. Dao, R. M. Kolaitis, H. J. Kim, K. O'Donovan, B. Martyniak, E. Colicino, H. Hehnly, J. P. Taylor and C. A. Castan˜ eda, *Mol. Cell*, 2018, **69**, 965-978.e966.
42. P. Friedhoff, A. Schneider, E. M. Mandelkow and E. Mandelkow, *Biochemistry*, 1998, **37**, 10223–10230.
43. Y. Xu and M. Havenith, *J. Chem. Phys.*, 2015, **143**, 70901.
44. H. R. Zelsmann, *J. Mol. Struct.*, 1995, **350**, 95−114.
45. J. Ceponkus and B. Nelander, *J. Phys. Chem. A*, 2004, **108**, 6499–6502.
46. S. Ebbinghaus, S. J. Kim, M. Heyden, X. Yu, U. Heugen, M. Grubele, D. M. Leitner and M. Havenith, *Proc. Natl. Acad. Sci. U.S.A*, 2007, **104**, 20749-20752.
47. T. Q. Luong, P. K. Verma, R. K. Mitra and M. Havenith, *Biophys J.*, 2011, **101**, 925-933.
48. T. Dodo, M. Sugawa, E. Nonaka, H. Honda and S. Ikawa, *J. Chem. Phys.*, 1995, **102**, 6208.
49. X. Li, L. Liu and H. B. Schlegel, *J. Am. Chem. Soc.*, 2002, **124**, 9639-9647.
50. N. Samanta, D. Das Mahanta and R. K. Mitra, *Chem Assian J.*, 2014, **9**, 3457-3463.
51. N. Samanta, D. Das Mahanta, S. Choudhury, A. Barman and R. K. Mitra, *J. Chem. Phys.*, 2017, **146**, 125101.
52. B. Born, S. J. Kim, S. Ebbinghaus, M. Gruebelebc and M. Havenith, *Faraday Discuss.*, 2009, **141**, 161–173.
53. N. Prabhu and K. Sharp, *Chem. Rev.*, 2006, **106**, 1616−1623.

Chapter 7

7. Summery and Future Perspective

7.I Summery

The present thesis deals with the study of four different biophysical processes using ultrafast spectroscopic techniques and for that we choose two different type biomacromolecules: DNA and protein. First, we have studied the electron transfer (ET) in between a model drug, PSF and CTDNA and observed how ET process gets modulated in presence of PEGs (act as molecular crowders). We, using CD spectroscopy and temperature dependent absorption spectroscopy, observed the structural perturbation of DNA in presence of PEGs. A significant quenching of PSF emission has been noticed while intercalating with DNA and the reason was attributed as ET process. Further ET process has been thoroughly investigated using time resolved fluorescence spectroscopy. We observed two different kind of ET process associated with the PSF-DNA complex: one is "*guanine mediated hopping ET*" (time scale of ~1 ps), another one is "*through space ET*" (time scale ~280 ps). We further noticed that the former one is independent of PEGs while the later one is solely affected by the PEGs.

Our next studies aim to observe the solvation of protein of different conformation in presence of additives. We, using CD spectroscopy observed that EtOH unfolds the secondary structure of Lys while TFE stabilizes the same. On the contrary both alcohols stabilize the secondary structure of βlg implying that alcohol protein interaction is protein specific. We then measure the protein solvation in presence of both alcohols using THz spectroscopy and noticed an oscillatory change in the hydration behaviour of both proteins with increasing alcohol concentration, however at higher alcohol concentration the effect becomes nominal. We finally correlate between change in hydration (as obtained from THz spectroscopy) and conformation of proteins (as obtained from CD spectroscopy) in presence of alcohols by constructing phase diagram. We then investigate how the protein hydration follows the trace of fibrillation pathway (native→ intermediate → fibril) of BSA. Here we also observed the change in the hydration as the fibrillation proceeds and finally correlate the hydration with the conformational change during the fibrillation pathway. Additionally we noticed a new kind of fluorescence phenomenon appearing only in the oligomers and fibrils which could be the marker of the fibril formation. As an extension of the work, we systematically study the fibril growth of human insulin protein and its (fibril growth) delay in presence of two crowders, sucrose and glucose. We notice that water releases from the protein interior/hydration layer

during the amyloid formation and the amount depends on the crowders. We finally monitor the alteration of protein hydration during the phase separation (LLPS) process of Lys as LLPS is important in maintaining the structure and activity of membranelles organelles inside living cell. We observed that during the LLPS process, both intermolecular hydrogen bond (HB) stretching and the librational motion of the water network gets altered. Addition of excipients alter the stability of the LLPS process: sucrose stabilizes the LLPS while BSA inhibits the same; L-Arg has the nominal effect and ubiquitin effect is concentration dependent. The hydration behaviour of Lys during LLPS gets further altered in the presence of excipients. Principal component analysis yields that addition of BSA alters both HB stretching and librational motion during the LLPS in a way opposite to the other excipients.

In brief the key findings of this thesis are:

- Modulation of electron transfer (ET) process between PSF and CTDNA in presence of PEGs. The through space ET depends on PEGs while guanine mediated ET is independent of PEGs.
- Alcohols modifies the secondary structure of both Lys and βlg and it is protein specific. The protein hydration shows oscillatory nature in presence of both ethanol and TFE and the structural makeover of protein and the associated solvation are correlated.
- Protein hydration follows the traces of fibrillation pathway (native→ oligomers→ fibril) of BSA. Water releases from protein interior during the fibrillation process of human insulin and it depends on the crowders used to delay the fibrillation process.
- Excipients (sucrose, BSA, Ubi, Arg) influence the LLPS of Lys and hydration is associated with the LLPS process and it coupled with excipients plays a role in influencing the LLSP process.

7.II Future perspective:

The studies carried out in this thesis could lead some new kind of research on the amyloid fibril and LLPS process. As the studies included in this thesis confirm that protein solvation follows the trace of both amyloid fibril formation and LLPS process of model proteins, one could attempt to study the hydration behaviour of some intrinsically disordered proteins (e.g. α-synuclein, tau protein, prion, amyloid β etc) in-vitro and also in-vivo, directly related with many neurodegenerative diseases,[1-3] as this could assist early detection of fibrils. Additionally, studies can be carried out to observe the effect of different type of molecular crowders, cosolvents to arrest further fibril growth. Some other research can also be accomplished on the

LLPS process. Some proteins (e.g. tau[4], α-synuclein[5]) can form LLPS under a certain condition and converts to its amyloid form over time. So, hydration study may be performed starting from native form to fibril through LLPS of such proteins to get an idea about the hydration of each state and how it is modified as the process proceeds. In this context, fluorescence spectroscopy (binding with ThT and/or Congo red), CD spectroscopy, atomic force microscopy and transverse electron microscopy will act as leading role in detecting the amyloid fibrils and THz spectroscopy can be utilised to study the alteration of protein solvation. DIC and confocal microscope will be consistent tool to observe the droplets appearing during the LLPS process.

7.III Reference:

1. S. T. Ferreira, M. N. N. Vieira and F. G. De Felice, *IUBMB Life*, 2007, **59**, 332-345.
2. M. Stefani and C. M. Dobson, *J. Mol. Med.*, 2003, **81**, 678-699.
3. C. M. Dobson, *Trends Biochem. Sci*, 1999, **24**, 329-332.
4. A. Siegert, M. Rankovic, F. Favretto1, T. Ukmar-Godec, T. Strohäker, S. Becker and M. Zweckstetter, *Protein Sci.*, 2021, **30**, 1326-1336.
5. S. Ray, N. Singh, R. Kumar, K. Patel, S. Pandey, D. Datta, J. Mahato, R. Panigrahi, A. Navalkar, S. Mehra, L. Gandhe, D. Chatterjee, A. Singh Sawner, S. Maity, S. Bhatia, J. A. Gerez, A. Chowdhury, A. Kumar, R. Padinhateeri, R. Rieck, G. Krishnamoorthy and S. K. Maji, *Nat. Chem.*, 2020, **12**, 705-716.

Index