Summer 2020

Studies on the stability of plasma proteins and their fragments

Paul Praveen Nakka

Department: Chemical and Biochemical Engineering

STUDIES ON THE STABILITY OF PLASMA PROTEINS AND THEIR FRAGMENTS

by

PAUL PRAVEEN NAKKA

A DISSERTATION

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

CHEMICAL ENGINEERING

2020

Approved by:

Dr. Daniel Forciniti, Advisor
Dr. Douglas Ludlow
Dr. Dipak Barua
Dr. Xinhua Liang
Dr. David Westenberg

PUBLICATION DISSERTATION OPTION

This dissertation consists of the following three articles, formatted in the style used by the Missouri University of Science and Technology:

Paper I, found on pages 9–43, is intended for submission to Biomacromolecules *Journal*.

Paper II, found on pages 44–90, has been published in the ACS Omega *Journal*.

Paper III, found on pages 91–129, is intended for submission to Biochemistry *Journal*.

ABSTRACT

The stability of a protein-based biological drug is an important concern during processing, storage or patient administration. There are various forces which contribute to the stabilizing of proteins. Both hydrogen bonds and hydrophobic interactions play an important role in protein stability. Improper folding or misfolding of proteins leads to aggregation, which is a major problem found in the bioprocessing industry. The objectives of this research are: 1) to determine the effect of environmental factors such as pH, temperature, additives or salts, and cosolvents on the stability human serum albumin (HAS), which is the most abundant protein in plasma, 2) to determine the effect of environmental factors on the propensity to aggregate of insulin fragments, and 3) to determine the effect of post-translational modifications such as glycosylation on the propensity to aggregate of insulin fragments. Small-Angle Neutron Scattering (SANS) was used to explore the effect of environmental factors on the intra and inter molecular interactions of HSA. The effect of cosolvents, pH and temperature on the protein conformation and on protein-protein interactions was explored. The propensity to aggregate, particularly the propensity to form amyloid fibrils of the region of the insulin A-chain (where amino acid sequence for bovine and human insulins is different) was then studied. Aggregation was monitored by fluorescence and infrared spectroscopy and optical microscopy. Both fragments formed fibrils at pH 1.6 and a temperature of 60°C. These peptides were glycosylated with D-glucose by the Maillard reaction to increase their stability. Glycosylated peptides were highly stable and they did not show signs of fibril formation during the course of the experiment.

ACKNOWLEDGMENTS

First, I would like to express my sincere gratitude to my advisor, Dr. Daniel Forciniti for his valuable guidance and encouragement through deepest waters when most doors seemed closed. Thanks to him, I have grown both professionally and personally learning what it takes to balance life.

I would also like to express my deepest gratitude to Dr. Douglas Ludlow, Dr. Dipak Barua, Dr. Xinhua Liang and Dr. David Westenberg for their time and endeavor in serving as committee members and reviewing this dissertation. A special thanks to all the staff in the Department of Chemical and Biochemical Engineering for their help and support during my research.

Thanks to my parents, my in-laws and my brother, Paul Preetham Nakka for their prayers and believing and supporting me through out the journey. Special thanks to my friends for their wonderful support. Finally, I would like to acknowledge my best half, Mounica Priyadarshini Tumati who stood by me through thick and thin especially during this COVID-19 pandemic and helped me reach the other side.

I would like to dedicate this dissertation to my parents Sam Sundar Rao Nakka and Ratna Kumari Jyothula for their unconditional love and support throughout the journey.

TABLE OF CONTENTS

Page

LIST OF ILLUSTRATIONS

PAPER I Page

PAPER II

PAPER III

LIST OF TABLES

NOMENCLATURE

Symbol	Description
pH	power of Hydrogen
UV	Ultra Violet
NaCl	Sodium Chloride
BIF	Bovine Insulin Fragment
HIF	Human Insulin Fragment
FTIR	Fourier Transform Infra Red
Ala	Alanine
Thr	Threonine
Val	Valine
Ile	Iso Leucine
ThT	Thioflavin T
ATR	Attenuated Total Reflection
HBTU	O-(Benzotriazol-1-yl)-N,N,N′,N′-tetramethyluronium hexafluorophosphate
DMF	Dimethyl Formamide
TFA	Tri Fluoro Acetic Acid
DIPEA	N,N-Diisopropylethylamine
Fmoc	Fluorenyl methoxy carbonyl
DCM	Dichloro Methane
NMR	Nuclear Magnetic Resonance

HPLC	High Pressure Liquid Chromatography
MS	Mass Spectrometry
QTOF	Quadrupole Time of Flight
DTT	Dithiothreitol
pH	power of Hydrogen

1. INTRODUCTION

Proteins and peptides are important for life as we know it. They serve as catalysts for the biochemical reactions that take place inside cells, they transport materials from place to place, they play key functions in our immune system, they serve as signal transmitters, etc. Amino acids are the building blocks of proteins and peptides. Long chains of amino acids make proteins whereas shorter ones make peptides. Proteins have four levels of organization: primary, secondary, tertiary and quaternary structure. While primary structure is simply the sequence of amino acids in a polypeptide chain, secondary structure refers to the local folded structures that form within a protein. The three-dimensional shape of the protein is its tertiary structure and the number and arrangement of multiple folded subunit complexes is the quaternary structure of a protein.

Biopharmaceutical companies invest a large portion of their resources to generate libraries of protein therapeutics. The stability of a protein is an important concern during expression, purification, formulation and storage of these proteins [1]. The consequences of lack of stability (on the shelf) of a biologic can be devastating for the patient whereas the lack of stability during processing results in financial loses. Stability manifests in two different dimensions: a) within each molecule (intramolecular) it is associated with the ability to maintain the three-dimensional shape that the protein needs to perform its biological functions and b) in between molecules (intermolecular) it is associated with the lack of formation of aggregates.

The intramolecular protein stability is the result of the net balance of forces, which determines whether a protein will be in its native folded conformation or in a denatured

state [2]. There are various forces that contribute to stabilize proteins [3]. Both hydrogen bonds and hydrophobic interactions play an important role in protein stability. Each hydrogen bond upon folding contributes 1.1±0.8 Kcal/mol to protein stability. An average of 85% of the nonpolar side chains bury to form 1.1 hydrogen bonds per residue, which contributes to proper protein folding [4,5]. Also, buried polar groups that are not hydrogen bonded can contribute to protein stability. Other forces which contribute to protein stability are disulfide bonds. For example, addition of 3 disulfide bonds increases the melting temperature of T4 lysozyme by 23.4°C [6].

Protein folding results in the release of the water that is hydrogen-bonded to the polypeptide chain. That release increases the configurational entropy of water and thus stabilizes the folded structure. Under physiological conditions in an aqueous solution, a protein folds into α-helices, turns or β-sheets depending upon its primary structure [3]. A typical change in the free energy of 5-15 Kcal/mol between the folded and unfolded structures is observed in most proteins.

The intermolecular stability is controlled by electrostatic, hydration and hydrophobic forces. Electrostatic forces are long-range forces that originate on direct Coulombic as well as dipole-dipole and induced dipole-induced dipole interactions. They weaken as the ionic strength of the solution or the dielectric constant of the solvent increases. Coulombic forces maybe attractive or repulsive whereas dipole-dipole and induced dipole forces are always attractive. Hydration forces are strong and short-range repulsive forces between charged or polar surfaced molecules separated by a thin water layer. Hydrophobic forces are attractive forces between two non-polar surfaces.

Proteins in solution may be de-stabilized to different extents. The repulsive forces that keep the proteins in solution may be weakened by external factors causing the proteins to aggregate and precipitate. The de-stabilizing agent may be a change in pH, a change in the polar characteristic of the solvent, or the addition of salts or polymers [7]. A change in pH, temperature or solvent polarity may also induce a disruption in tertiary or secondary structure of the protein [8]. These changes in the conformation of the protein may lead to precipitation but also the denatured protein may remain in solution [9]. These two different levels of perturbations to a protein solution may lead to the formation of aggregates that will eventually form a precipitate. The protein in these aggregates may be native or may be completely or partially denatured. Our attention has been focused on the effect of the polarity of the solvent, pH, temperature, addition of denaturants, and stirring on protein stability [10]. The effects of these parameters are explained in the following paragraphs.

The effect of the polarity of a solvent or its capability to participate in hydrogen bonds on protein stability is significant. The folded structure of a protein is marginally stable and, therefore, the folding/unfolding equilibrium of proteins in aqueous solution can shift either way upon addition of cosolvents. While methanol cannot substitute the essential water from the protein hydration layer, glycerol can induce protein preferential hydration particularly at hydrophobic patches [11]. Studies show that the interaction of water with cosolvents like trimethylamine N-oxide (TMAO) plays an important role in determining the preferential exclusion from protein surfaces [12]. Ionic liquids are also important in the preservation of protein [13]. Ionic liquids are salts consisting of poorly coordinated ions, which remain in liquid state at temperatures < 100°C. Because of their unusual properties like low melting point, low vapor pressure, low toxicity as well as high polarity and thermal

stability they are considered as ‘ green solvents’ [14]. For example, lysozyme is stable in a solution of ionic liquids containing $[emim]^+$ cations and other anions [15].

Temperature affects protein stability. Incubating protein solutions at high temperatures can cause changes in the tertiary structure, which can lead to irreversible aggregation [16]. It is well known that the aggregation happens well below the equilibrium melting temperature of a protein which infers that it is not the fully unfolded proteins which aggregate but rather partially unfolded molecules [17]. Of course, denaturants like urea and guanidine chloride will affect the stability of the protein. Interfaces like solid/liquid or gas/liquid can also induce partial unfolding and thus protein aggregation. Stirring incorporates air into the solution creating a gas/liquid interface, which can affect stability.

Ionic strength and pH affect Coulombic interactions between the side chains of amino acids and that affects protein stability. They also affect intra and inter dipole-dipole and induced dipole-induced dipole interactions. Moreover, charge-charge interactions and salt bridges contribute to protein stability. Salt bridges are oppositely charged groups within 5Å distance in the proteins. A salt bridge on the surface contributes to < 1 Kcal/mol to the stability while a buried salt bridge can contribute to more than 4 Kcal/mol to the free energy of folding [18,19]. Finally, the degree of ionization of different groups on the protein’s surface depends on the pH and on ionic strength. The degree of ionization of the surface group determines the protein net charge, which defines the range and strength of Coulombic repulsions. That electrostatic repulsive force between protein molecules is the prevailing force preventing protein aggregation.

As we explained above, the de-stabilization of a protein in solution may lead to the formation of aggregates. Aggregates may be disordered or ordered; amyloid fibrils are one

type of ordered aggregates. They are β-sheet rich linear aggregates formed from circular oligomers. X-Ray diffraction of amyloid fibrils shows a cross-β diffraction pattern which is a trademark of intermolecular β-sheet structures [20-23]. This intermolecular β-sheet structure is easily identified also by FTIR. It has been recently proven that oligomers are the elemental blocks to build a fibril [24,25]. The propensity to form amyloid fibrils seems to be correlated with the hydrophobicity of proteins [26]. Amyloid fibrils are a unique type of ordered aggregates formed by different proteins and peptides in vivo as well as in vitro. Formation of amyloid fibrils is associated with various neurodegenerative diseases like Alzheimer's, Parkinson's and Huntington's disease [27]. Although amyloid fibrils are mostly related to disease causing proteins, it was also reported that if appropriate incubation conditions were provided even a non-pathogenic protein can form a fibrillar structure [28]. Amyloid aggregates can also be used as scaffolds, as peptide hormones storage and natural adhesives [29]. Fibrils can also be synthesized in vitro by subjecting the protein to favorable destabilizing conditions [30,31]. For example, hen lysozyme incubated at pH 2.0 and 37°C forms fibrils [32].

Post-translational modification of a protein affects both its biological activity and stability. There are different types of post translational modifications that occur at a cellular level. Most common among them are phosphorylation, glycosylation, lipodation, methylation and acetylation. Changes in the glycosylation pattern is one approach to change the stability of a protein therapeutic. Glycosylation consists of the addition of various sugars to the polypeptide chain [33,34]. There is no consensus about the role played by glycosylation on protein stability. It has been argued that protein glycosylation protects against proteolytic degradation, oxidation, chemical cross linking, precipitation,

aggregation, kinetic deactivation, and denaturation induced by changes in pH, addition of denaturants, and high and low temperatures. A better understanding of the role of glycosylation on protein stability will facilitate the engineering of proteins/ peptides with a glycosylation pattern that increases their shelf-life, minimizes losses during processing but does not alter the potency of the protein therapeutics nor changes its allergenic properties.

Glycosylation happens in vitro via two linkages i.e., via N-linkage known as N-glycosylation and via O-linkage known as O-glycosylation. Amino acids like serine, glutamine, lysine can undergo glycosylation. Protein/peptide glycosylation can be N-type or O-type depending upon the type of amino acid and its availability whether it is on the surface or buried inside the polypeptide chain. Glycopeptides can be synthesized either by direct incorporation of a glyco amino acid during peptide synthesis or by heavily glycosylating the peptide with a known sugar following the Maillard reaction [35]. Direct incorporation of sugars during the peptide synthesis can lead to a controlled reaction where the number of carbohydrate moieties being attached can be monitored. The main challenge lies in synthesizing the glycoamino acids to be used for peptide synthesis. Synthesizing glyco amino acids is quite expensive as opposed to the direct glycosylation of a peptide. Synthesis of glyco amino acids typically takes place in organic solvents like DCM (Dichloro Methane). The challenge is to remove the organic solvent from the synthesized glyco amino acid [36]. A rotary evaporator which is typically used to remove solvents is not effective in this case (unpublished data).

The main objective of this dissertation was to study the stability of plasma proteins and their fragments. This dissertation is a humble effort to augment our knowledge of

protein stability. My studies ranged from experiments probing molecular level phenomena to experiments at the macroscopic scale. This dissertation is divided into three parts that are briefly described below.

In Paper I, the effect of cosolvents, pH and temperature on the structure and stabilizing forces for human serum albumin (HSA) using small angle neutron scattering (SANS) was explored. The experimental data for this project was provided by Dr. Daniel Forciniti who conducted the SANS experiments at Oak Ridge National laboratory, TN. We explored intra and intermolecular interactions between protein molecules. The results of this work help to better understand the aggregation process. The interaction potential that can be extracted from SANS data offers an insight into the type and strength of the intermolecular interactions in solution. In the absence of intermolecular interactions, the shape of HSA changed from a prolate to an oblate ellipsoid as the concentration increased. Methanol increases the disorder of a solution of HSA whereas glycerol makes the solution more structured.

In Paper II, we studied the amyloidogenic regions of insulin from different species at various conditions using a screening experimental design in order to determine the favorable conditions for fibrillation. The studies were focused on short peptides having well-defined secondary structures, which mimic protein domains. The study of individual domains seems to be a logical approach towards the understanding of proteins responses to stresses and how that response may be affected by the presence of the stabilization groups. Peptides were synthesized using an F-moc chemistry in a solid phase peptide synthesizer. Mass and NMR spectroscopies were used to identify the synthesized products and to quantify them. A fluorometric assay was used to monitor the progress of fibrillation.

Fourier Transform Infra-Red (FTIR) spectroscopy was used to monitor the secondary structural changes in the peptides. A digital microscope is used to obtain visual evidence of the fibrils. The effects of temperature, urea, ionic strength, pH and stirring on the fibrillation process were explored. It was found that the bovine insulin fragments aggregate faster than the human fragments. Both the fragments showed an increase in the β-sheets content with time.

In Paper III, the peptides used in Paper II were glycosylated to study the effect of that post-translation modification on their ability to form amyloid fibrils. Insulin A-chain fragments from bovine and human species were glycosylated using α-D-Glucose. Experiments were performed at pH 1.6 and 60°C under stirring conditions in the presence of 1M urea and 20 mM NaCl. Aggregation was monitored using a fluorescent assay as well as by turbidity. The aggregates were characterized using FTIR and Congo red staining. The glycosylated peptides did not show any signs of fibril formation. The turbidity assay revealed the underlying oligomer formation mechanism, which is not seen using a fluorescent assay.

PAPER

I. EFFECT OF METHANOL AND GLYCEROL ON THE STRUCTURE OF HUMAN SERUM ALBUMIN SOLUTIONS

ABSTRACT

Small angle scattering (SAS) is an important tool to study the structural properties and interactions of biological macromolecules. Numerous studies have been published about the effect of organic solvents on intramolecular protein interactions. However, little attention has been paid to the effect of organic solvents on both intra and inter molecular interactions in protein solutions. The solution behavior of human serum albumin (HSA) in the presence of methanol and glycerol was studied by small angle neutron scattering (SANS). The experiments covered a wide range of protein, salt, and cosolvent concentrations at 5, 25, and 45 °C. UV and infrared spectra of selected samples were also obtained to monitor protein denaturation upon addition of the cosolvents. An ellipsoid form factor in combination with Hayter-Penfold rescaled mean spherical approximation were used to fit the data at high protein concentrations. The overall dimensions of HSA obtained in this work are consistent with previously reported data. Pairwise distribution function (P(r)) profiles were used to deduce the particle shape using GNOM. With an increase in the concentration of protein, HSA changes from a prolate to an oblate ellipsoid form. It was observed that methanol makes the interaction potential more attractive whereas glycerol makes it more repulsive.

1. INTRODUCTION

The distribution of proteins in a solvent depends not only on direct interactions between proteins but also on solvent-mediated forces. The former depends on the chemical structure of the protein and on the dielectric constant of the solvent (ε), whereas the latter depends on the chemical nature of both the protein and the solvent. Two solvent-related forces that affect protein structure and protein-protein interactions are hydrophobic attraction and hydrophobic hydration (repulsive). As two nonpolar molecules (or two hydrophobic portions of the same molecule) approach one another they push ordered water molecules into the bulk causing an increase in entropy. Conversely, we can argue that increase in entropy causes an attraction between hydrophobic surfaces (hydrophobic attraction). These forces affect the folding of the molecule and may induce protein aggregation [1]. Protein molecules in solution are surrounded by a layer of water molecules attached to the protein surface mainly by hydrogen bonds. The addition of an organic solvent distorts this layer of water and as a result the protein may denature [2]. On the contrary, solvents that promote the stabilization of the hydrogen-bonded network of liquid water will stabilize the water layer increasing the repulsion between proteins (hydrophobic hydration) [2]. Other additives, like salts or nonionic polymers, also affect intermolecular forces [3,4]. To understand the effect of water on the 3D shape of proteins in solution, researchers have looked up the effect that partial replacement of water by other solvents has on a protein 3D organization. The addition of a cosolvent stabilizes or destabilizes a protein structure depending on the difference between the preferential interaction of the mixed solvent with the protein in the unfolded and native states [5,6].

Gupta *et al.*, [7] examined the effect that various organic solvents have on several proteins and concluded that methanol cannot substitute "essential" water from the protein hydration layer whereas glycerol can. Thus, these two solvents not only alter electrostatic interactions by decreasing the dielectric constant of the solvent ($\varepsilon_{methanol}$= 32.7 and $\varepsilon_{glycerol}$= 42.5, ε_{water}= 78), but also change the solution behavior of the protein by changing its interactions with the solvent. Studies of protein solutions in which part of the water is replaced by methanol or glycerol provide an insight into the relative importance of solvent-mediated forces and therefore, they constitute an astringent test for mean field approximations. Glycerol induces protein preferential hydration particularly at hydrophobic patches, increases the solubility of proteins, and reduces the protein's surface area since the protein adopts a more compact conformation to compensate for the increase in chemical potential [7-9]. A mono alcohol (such as methanol) can participate in limited hydrogen bonding. Therefore, a mixture of water and alcohol will exhibit a hydrogen bonding network that will be different from that of pure water. Methanol is used to precipitate proteins and it is a well-known denaturant at considerable high concentrations. The addition of methanol, a less polar (dipolar moment = 1.70 Debye, ε = 32.7) solvent than water (dipolar moment = 1.85 Debye, ε = 78.54) most likely displaces water from hydrophobic patches disrupting the clathrate structure and decreasing the hydration repulsion and perhaps increasing the hydrophobic attraction. [10].

The effect of cosolvents on intra and intermolecular structure was studied using SANS. The general approach to interpret SANS data is as follows. The experimental data obtained under conditions where intermolecular interactions can be neglected (low protein concentrations, high ionic strength), is fitted using a model for the intramolecular structure

factor. Then, the intermolecular structure factor is calculated using, for example, an appropriate closure for the Ornstein Zernike equation like the Percus-Yevick or mean spherical approximations with a intermolecular potential that captures the main intermolecular forces [11]. The intra and intermolecular structure factors are combined and used to fit SANS data at all protein concentrations and solvent conditions.

Human Serum Albumin (HSA) was chosen in this work because it is the most abundant and better-known plasma protein. Its molecular weight is 66,248 Da and its isoelectric point is 4.7. A detailed crystallographic structure of HSA has been determined [12,13]. The molecule has a heart shape that can be approximated by a solid equilateral triangle with sides of 80 Å and depth of 30 Å. HSA consists of three well-defined domains. Two of them (I and II) form the wider portion of the heart whereas the third one (III) forms its tip. Domain III is positively charged at physiological pHs. HSA is a very stable protein. Changes in the conformation of this protein upon changes in pH have been observed in proteins from several species including bovine and human. However, changes in the overall size have not been detected using low resolution techniques such as photon correlation spectroscopy in the pH range from 5 to 10 [14].

The dependence of protein structure and protein-protein interactions on pH, salt and protein concentrations has been studied [15,16]. Unfortunately, those studies offer limited information since by manipulating those variables researchers have failed to explore the effect of the chemical nature of the solvent. Therefore, solvent-mediated phenomena may have been overlooked and their effects inadvertently included into the model parameters. New information can be obtained by studying the behavior of proteins in mixtures of methanol, glycerol and water at different temperatures. These two solvents have different

effects on the structure and solution behavior of proteins. The effect of temperature, although it may be small, can be used to discriminate between different hypothesis.

Here, first we present and analyze experimental data obtained under conditions where intermolecular interactions are minimized. This is done using Guinier analysis, P(r) analysis and fitting the data using an ellipsoid model in Igor PRO. Pairwise distribution function (P(r)) profiles were used to deduce the particle shape using GNOM [17]. After that, we present and discuss experimental results where intermolecular interactions dominate the structure of the solution. Intermolecular interactions can be captured by combining the form factor with a suitable interparticle structure factor.

2. EXPERIMENTAL SECTION

2.1. MATERIALS AND METHODS

Essentially fatty acid free HSA was bought from Sigma and used as received. Fully deuterated solvents were from Sigma (St. Louis, MO). HSA solutions at various concentrations were made ready by dissolving the protein in binary mixtures of fully deuterated methanol or glycerol with deuterium oxide in 20 mM phosphate buffer. The pH was adjusted to 6. Insoluble matter was removed by centrifuging and the concentration of protein in the sample was measured spectrophotometrically at 280 nm *(ε = 0.531 L/g cm)*. A wavelength scan was done for each sample from 350 nm to 270 nm. Negligible adsorption at 350 nm was taken as an indication of completely transparent samples. In the few cases where the absorption at 350 nm was not negligible the absorbance at 350nm was extrapolated to 280nm and the value subtracted from the sample. This correction was

always smaller than 5%. The samples were run at three different temperatures, 5, 25 and 45 ^{0}C.

Ultraviolet and infrared spectroscopy were used to determine the stability of HSA in mixtures of methanol and water. UV spectra were measured in a Shimadzu spectrophotometer using a protein concentration of 3 mg/ml. Appropriate blanks were discounted before calculating the second derivative spectra. Fourier Transform infrared spectroscopy was done using a Nicolet 6700 spectrophotometer. The protein solution (approximately 10 mg/ml) was placed in a $CaCl_2$ cell (0.1 mm path length). After 15 minutes stabilization, the samples were scanned 400 times. Background spectra of the solvents blank (either water or methanol/water mixtures) were subtracted from each sample spectra. Amide I band of each spectrum was Fourier self-deconvoluted to resolve peaks that provide information about secondary structure.

Neutron scattering experiments were conducted at the 30m SANS instrument at the High Flux Isotope Reactor (HFIR) within the national center for small angle scattering research at Oak Ridge National Laboratory. The protein solutions were placed in flat cylindrical quartz cells having 1 mm thick circular windows and a path length of 1- or 2-mm. Multiple samples were mounted on an automatic rotating sample exchanger maintained at constant temperature by circulating water. To prevent the deuterated water from becoming contaminated with atmospheric moisture, all SANS samples remained sealed throughout the experiments.

The experiments were performed on the W. C. Koehler 30m SANS facility at the Oak Ridge National Laboratory. The neutron wavelength was 0.475 nm (Dl/l= ±5%) and the source (3.5 cm in diameter) and sample (1.0 cm in diameter) slits (irises) were separated

by 7.5 m. The sample-detector distance was either 13.5 or 2.49 m and the data were corrected for instrumental backgrounds and detector efficiency on a cell-by-cell basis before radial averaging to give a momentum transfer (Q) range of $0.05 < Q < 2.5$ nm^{-1}. The coherent intensities of the sample were obtained by subtracting the intensities of the corresponding blanks (either heavy water or the appropriate heavy water/organic solvent mixture), which formed only a minor correction (<10%) to the sample data. The net intensities were converted to an absolute (±5%) differential cross section per unit sample volume (in units of cm^{-1}) by comparison with pre-calibrated secondary standards, based on the measurement of beam flux, vanadium incoherent cross section, the scattering from water and other reference materials. The efficient calibration was based on the scattering from light water and this led to angle-independent scattering for vanadium, H-polymer blanks and water samples of different thickness in the range 1-10 mm.

The transmission of the sample was measured in a separate experiment by collimating the beam with slits (irises) approximately 1 cm in diameter, separated by 7.5 m. A strongly scattering sample, porous carbon, was placed at the sample position to spread the beam over the whole detector, placed at a sample-detector distance of 10 m. Without the carbon in position, the beam would either be blocked by the beam stop or concentrated in a few detector cells, with the possibility of saturating or damaging the detector. The total count summed over the whole detector (>10^5) was recorded for ~1 minute and the sample being measured was placed over the slit, thus attenuating the beam. The count was repeated over the same time interval and the transmission is given by the ratio of the two counts after minor corrections (< 0.1%) for the beam-blocked background due to electronic noise, cosmic rays, etc. In this geometry only scattering from the sample at Q-values < 10^{-2} nm^{-1}

can enter the second iris and be scattered by the porous carbon and hence be counted by the detector.

2.2. DATA ANALYSIS

The analysis of SANS data is well documented in the literature [18]. The intensity (I(Q)) of the scattered neutrons in terms of the form factor (F(Q)) and intermolecular (S(Q)) structure factor is given by:

$$I(Q) = \frac{cNa}{M}\left(\frac{vM}{Na}\right)^2 < F(Q)^2 > \Delta\rho^2 \quad \left\{1 + \left(\frac{< F(Q) >^2}{< F(Q)^2 >}\right) \quad [S(Q) - 1]\right\} \qquad (1)$$

where c is the concentration of the protein in g/cm^3, Na is the Avogadro's number, M is the protein's molecular weight, v is protein's specific volume in cm^3/g, and $\Delta\rho$ is the difference between the scattering density of the scatters, ρ_p and the scattering density of the solvent, ρ_s, in cm^{-2}. The brackets indicate angular averages.

The form factor can be calculated from the crystal structure of the protein, by modeling the proteins as an assembly of spheres or by assuming a shape for the protein molecule [19]. Calculations based on the crystal structure are formally exact but highly complex. The use of an assembly of spheres model is attractive but also quite complex and the structure obtained may not be unique. Finally, calculations based on the assumption of a simple form for the protein molecule (ellipsoids, spheres, or cylinders) are computationally simpler but less precise. This simple approach yields excellent fittings and it is computationally cheap. The interparticle structure factor, S(Q), was calculated using a

simple approach by assuming the particles to be charged and spherical in a dielectric medium using Hayter- Penfold rescaled MSA [20-22].

To interpret SANS data correctly changes in intramolecular structure must be identified and quantified before intermolecular interactions can be studied. To be able to extract information about the effect of the added cosolvents on the structure of protein solutions, one needs to rule out or account for major changes in protein conformation or aggregation induced by the addition of the solvents. One way of obtaining information about size changes and to detect the presence of aggregates is to perform a Guinier analysis of the data.

A plot of ln(I(Q)) vs. Q^2 yields a straight line whose slope is proportional to the radius of gyration of the scatters and whose intercept is proportional to their molecular weight,

$$I(0) = \frac{cNa}{M}\left(\frac{vM}{Na}\right)^2 \Delta\rho^2 \tag{2}$$

The approximations that lead to Guinier analysis are valid in the low Q region when using dilute protein solutions in the presence of high concentrations of salt. The addition of NaCl or other salts is necessary since otherwise the analysis cannot be performed because of the presence of a first correlation peak in the spectrum (a result of interference between waves scattered by more than one protein molecule). The data, however, must be analyzed with caution since extrapolation at other conditions is not always valid. For example, aggregation may not occur under conditions in which Guinier analysis is valid (absence of intermolecular interaction, low Q values) but occurs in more concentrated

solutions, even in the absence of NaCl. Similarly, aggregation may occur in the presence of NaCl but not in its absence.

Data analysis was performed using Igor Pro (v 8.02; waveMetrics, Lake Oswego, OR, USA) using the algorithms based on NCNR analysis package [23]. The particle distance distribution function P(r) determination was done using a program suite ATSAS 2.8.4 [24]. The inverse transformation of I(Q) allows for the calculation of P(r) given by

$$P(r) = \frac{r}{2\pi^2}\int_0^\infty QI(Q)\sin(Qr)\,dQ \tag{3}$$

The dielectric constant for the binary solvent mixture with respect to solvent composition (Ø) and temperature (T) is calculated using the following equation.

$$\ln \varepsilon_m = \phi_1 \ln \varepsilon_1 + \phi_2 \ln \varepsilon_2 + \phi_1\phi_2 \sum_{j=0}^{2}\left[\frac{A_j(\phi_1 - \phi_2)}{T}\right] \tag{4}$$

where ε_m, ε_1, ε_2 corresponds to the dielectric constants of mixture and solvents 1 and 2 respectively. A_j corresponds to the constants for the specific binary system obtained from literature [25-27]. The values of the dielectric constants of the three solvents at 5, 25, and 40 oC are summarized in Table 1.

Table 1. Dielectric constants of the solvents

Solvent	Temperature (°C)	ε_m
D_2O	5	85.7
	25	78
	45	71.5
Methanol	5	36.4
	25	33.1
	45	28.4
Glycerol	5	44.2
	25	42.5
	45	37.3

3. RESULTS AND DISCUSSION

It can be argued that upon addition of methanol the conformation of the protein changes and that those changes are not captured by SANS [28]. To evaluate these possible changes, we conducted infrared and UV spectroscopy experiments of selected samples. Figures 1 and 2 show the UV second derivative spectra for buffer and 30% methanol at 45°C and the Fourier self-deconvoluted spectra (amide I band in the wave number range 1700- 1600 cm^{-1}) of buffer and HSA solutions at 30% methanol concentration and 40% methanol concentration respectively.

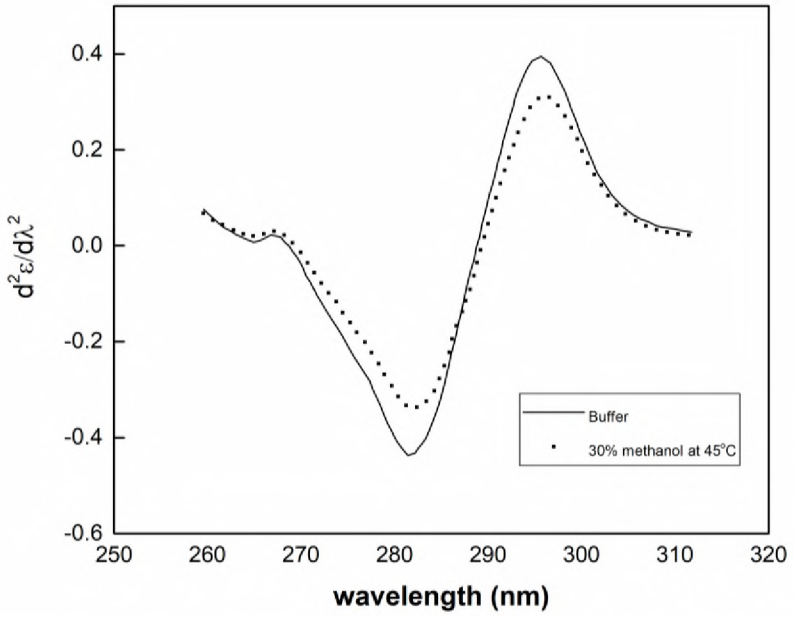

Figure 1. UV second derivative spectra for HSA solutions

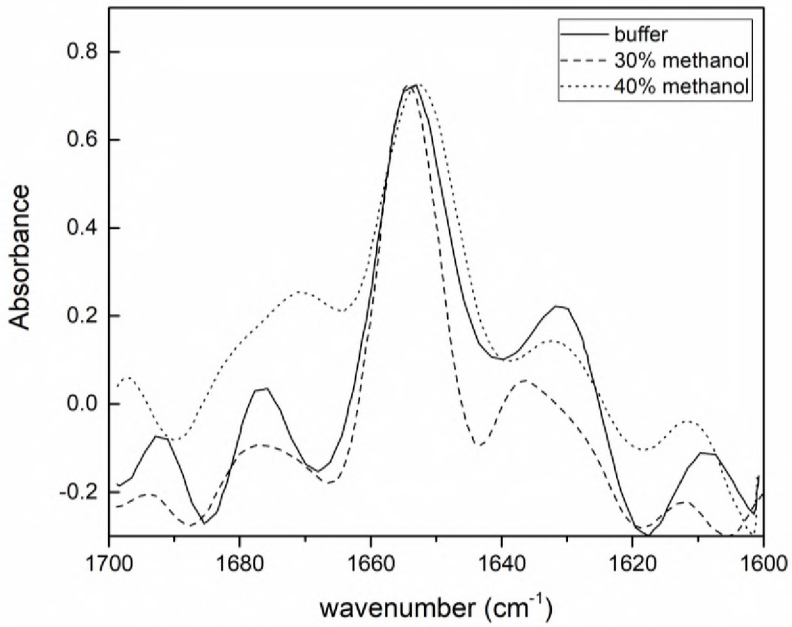

Figure 2. Fourier self deconvoluted amide I spectra for HSA solutions

Up to 30% methanol at room temperature, the IR spectra do not show any significant change. However, at a concentration of methanol of 40% there is a decrease in the peak area at 1677 cm^{-1} relative to the area of the peak at 1654 cm^{-1} (α-helix). This indicates a change in the conformation of the protein at these methanol concentrations [29]. In addition to this, there is a shift in the amide II peak to lower frequencies, which is associated with the degree of hydrogen-deuterium exchange. The presence of methanol in the medium may facilitate this exchange by "opening" the secondary structure of the protein. At 40% methanol, we also observed a peak at 1900 cm^{-1} that is not present at the other methanol concentrations. This peak may be due to proline chains that become exposed to the solvent. The addition of up to 30% methanol at room temperature to an HSA solution does not produce any change in its second derivative spectra (this is not shown because overlaps perfectly with the spectrum in pure buffer). However, after incubation of the solution for 30 minutes at 45^0 C we observe the formation of large aggregates. This solution was filtered through a 0.2 μm filter and the UV spectra were recorded. The difference in absorbance before the aggregates were formed and after filtration indicates that approximately 10% (w/w) of the protein is present as large aggregates. The second derivative spectrum also shows some differences. After heating, there is a decrease in the intensity in the tryptophan region (291-295 nm) that is due to partial exposure of these residues to the solvent [30]. Therefore, except at 40% methanol concentrations the protein is stable at the conditions at which SANS data was obtained. This evidence shows that at low protein concentrations, HSA remains native when moderate amounts of methanol are added to a salt-free protein solution. On the contrary, the simultaneous addition of salt and methanol induce aggregation of the protein at methanol concentrations higher than 15%.

No apparent change in the shape of the protein upon addition of the cosolvents is observed at low to ambient temperature. At high temperatures and high methanol concentrations the protein denatures and form aggregates.

A first set of SANS experiments were performed using low protein concentrations and relatively high electrolyte (NaCl) concentrations to determine the effect of the cosolvents on the shape of the protein. Figure 3 shows a representative Guinier plots of HSA dissolved in D_2O and D_2O/ methanol at low protein concentrations. The values of the radius of gyration, I(0) and I (0)/c$\Delta\rho^2$ for various samples are reported in Table 2 together with 95% confidence intervals.

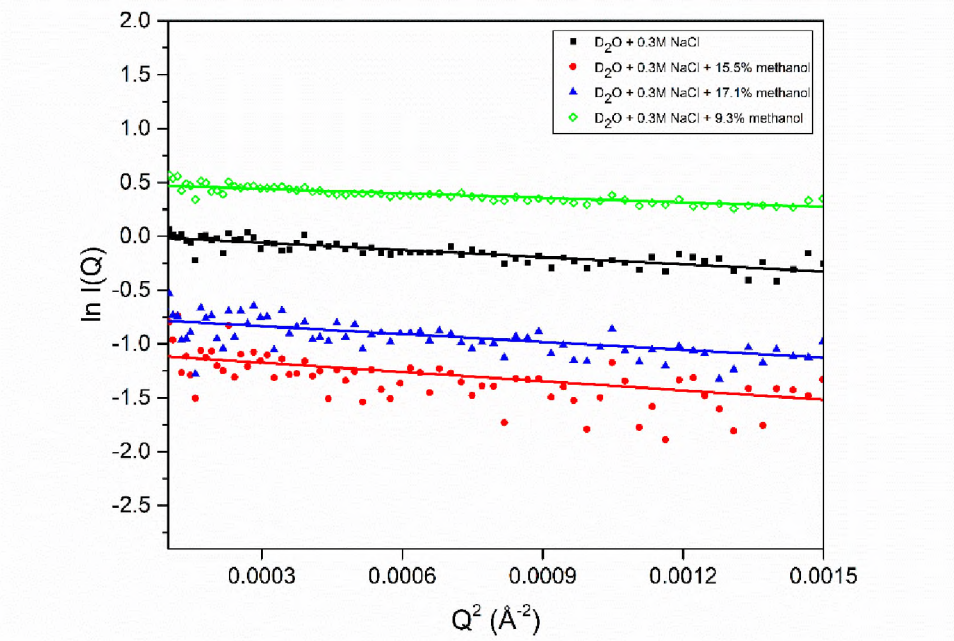

Figure 3. Guinier plots for HSA at low concentrations

We report I(0) and I (0)/c$\Delta\rho^2$ because I (0)/c$\Delta\rho^2$ should be a constant for a given molecular weight and in the absence of intermolecular interactions. Guinier analysis yields a radius of gyration of 2.6 ± 0.1 nm and an I (0) of 1.01± 0.04 cm^{-1} for HSA in pure D_2O. The value for the radius of gyration is smaller than the value reported by the SAXS study of Olivieri and Craivich who reported values for R_g of 3.1 nm at pH 5 and 3.3 nm at pH 7 [31]. Discrepancies between the values of R_g from SANS and SAXS have been explained by Svergun *et al* in terms of a hydration shell [32]. The value of the intensity at zero momentum transfer compares very well with the expected value of 0.96 cm^{-1} calculated from Eq. 2. This is a definite evidence that HSA remains monomeric in D_2O. The increase in I(0) in presence of moderate methanol concentrations suggests the formation of aggregates.

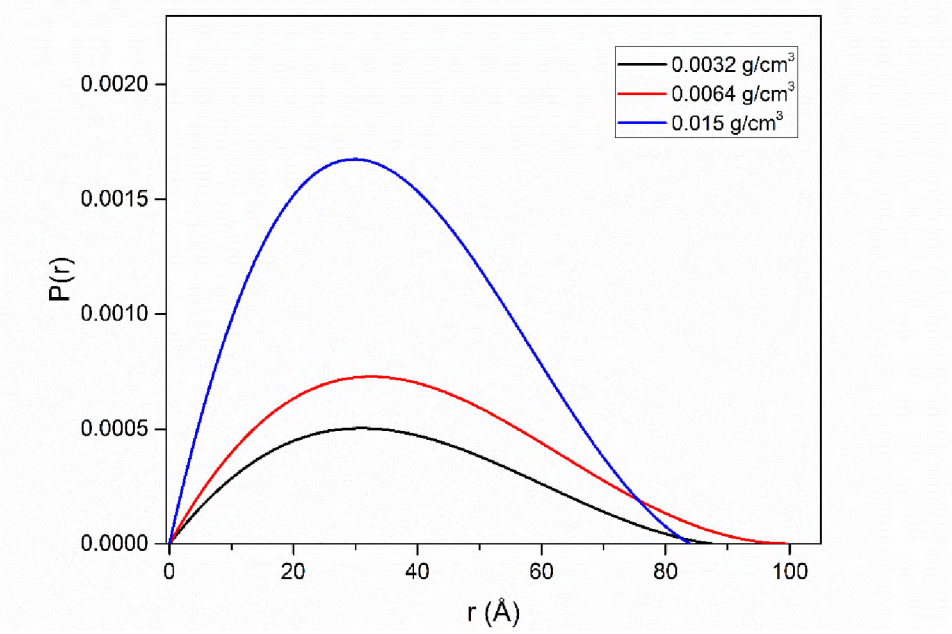

Figure 4. Pair distance distribution functions for HSA at low concentrations

Table 2. Guinier analysis summary

Solvent	Protein concentration (g/cm^3)	Radius of Gyration (nm)	R_g from P(r) curves	$I(0)/c\Delta\rho^2$ ($cm^6\ g^{-1}\ 10^{-20}$)*	$I(0)^{calc}$	$I(0)^{exp}$
D_2O	0.015	2.6±0.1	2.7±0.1	6.3±0.2	0.96± 0.05	1.01±0.04
D_2O+9.3% Methanol	0.032	2.3±0.1	2.5±0.1	6.8±0.1	1.40±0.07	1.62±0.01
D_2O+15.5 % Methanol	0.0038	2.9±0.2	3.6±0.5	12.2±0.4	0.16±0.01	0.34±0.01
D_2O+17.1 % Methanol	0.0060	2.7±0.1	2.8±0.2	10.7±0.2	0.26±0.01	0.47±0.01
* Values for each solvent are calculated according to Eqn 2						

This is easily confirmed comparing the calculated values of I(0) with the experimental ones particularly at methanol concentrations higher that 9.3%. On the other hand, the variations in Rg under different solvent conditions are not statistically significant

at 95% confidence. To better understand the effect of the cosolvents and of protein concentration on the intramolecular structure of HSA, pair distance distribution functions (P(r)) were calculated at the three protein concentrations used for Guinier analysis. Figure 4 shows the P(r) curves for HSA with increasing concentration that were calculated using the GNOM program [33]. The maximum diameter of the macromolecule (D_{max}) can be obtained from the P(r) curves (P(r) tends to zero for $r>D_{max}$). The maxima in the peak corresponds to the most frequently occurring interatomic distances within the structure in real space.

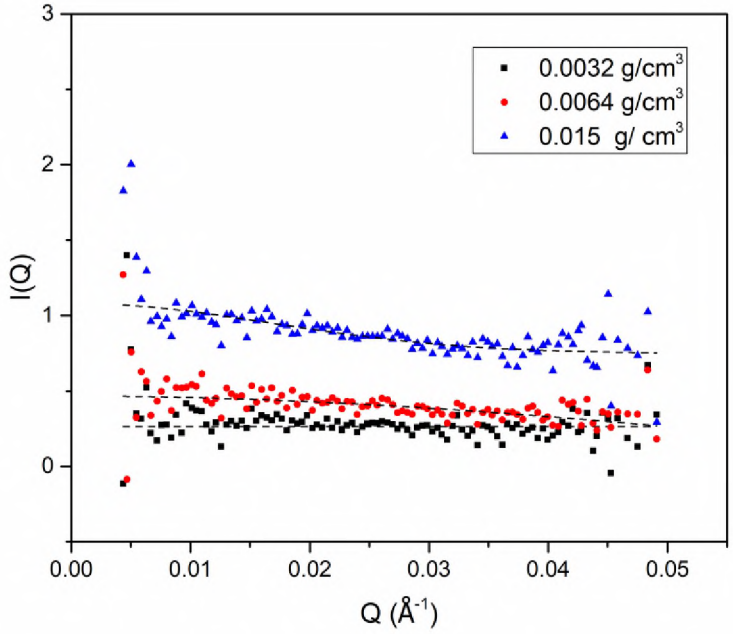

Figure 5. A plot of I(Q) vs Q for HSA in D_2O at 0.3 M NaCl and at three different concentrations

HSA at 0.0032 g/cm^3 and 0.0064 g/cm^3 is a prolate ellipsoid whereas at 0.015 g/cm^3 becomes an oblate ellipsoid.[34] Proteins tend to evolve into a shape which avoids random collisions from their sites of interaction [34]. The values of R_g obtained using P(r) (also reported in Table 2) are also in good agreement with the values obtained from Guinier analysis but the observed differences under different solvent conditions are not statistically significant.

Figure 5 shows a plot of I(Q) vs Q for HSA in D_2O at 0.3 M NaCl and at three different concentrations (0.0032 g/cm^3, 0.0064 g/cm^3 and 0.015 g/ cm^3) of the protein along with its fits. The values of SLD for protein and D_2O were taken as 3.053 x 10^{-6} Å^{-2} and 6.335 x 10^{-6} Å^{-2}. The fits were obtained by assuming the protein to be a prolate ellipsoid at the two lower concentrations and an oblate ellipsoid at a concentration of 0.015 g/cm^3. The minor and major axis are R_a and R_b respectively. This is somehow justified by recent findings showing that proteins larger than 270 amino acids (HSA has 585 amino acids) trend to have an ellipsoidal form [26].

Table 3. Using ellipsoids to fit low concentration data

Solvent	Protein Conc (g/cc)	R_a (nm)	R_b (nm)	R
D_2O + 0.3M NaCl	0.0032	13.6±0.2	1.0±0.1	2.4±0.1
	0.0064	14.0±0.2	0.9±0.1	2.2±0.1
	0.015	1.0±0.2	4.6±0.2	1.6±0.2

The decision to use different shapes was based on the analysis of P(r). Moreover, Liu et al. measured the distance distribution function for human hemoglobin at different concentrations which shows that the shape changes with change in protein concentration [15]. The fitting parameters R_a and R_b are reported in Table 3. The protein becomes more compact as its concentration increases. The effect of the cosolvents on HSA may be different in the absence of the salt. To elucidate this point further, we performed a few experiments using low protein concentrations without added salts. Under these conditions, the Guinier approximation is not valid but the data can still be analyzed by fitting the scattering curve in a wider range of momentum transfers using a simple model for the form factor and the Hayter- Penfold rescaled MSA for the intermolecular structure factor.

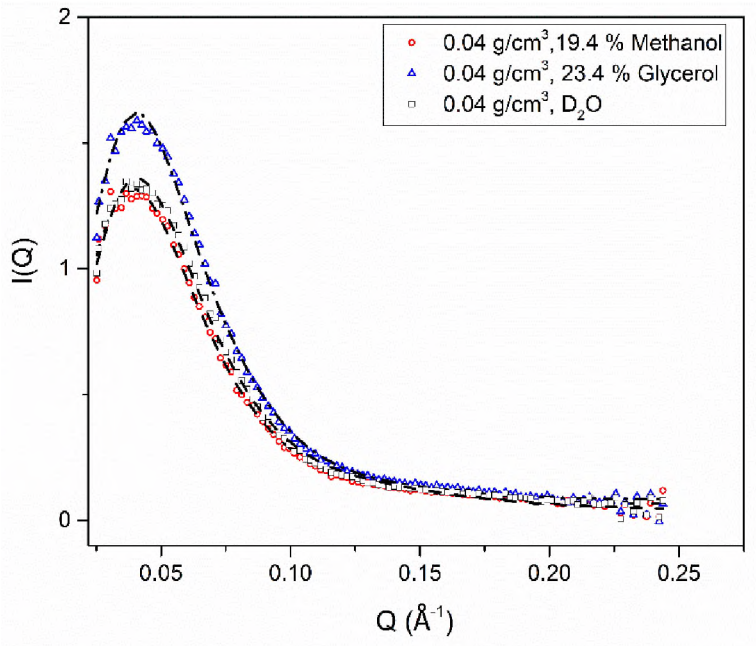

Figure 6. Effect of cosolvents on HSA at 0.04 g/cm^3 and 25°C

Table 4. Fitting parameters obtained using HPMSA model

Description	R_a (nm)	R_b (nm)	Equivalent Radius (nm)	Charge
D_2O; 0.04 g/cm^3; 25°C	1.0±0.1	4.0±0.2	1.6±0.1	7.2±0.1
19.3% methanol; 0.04 g/cm^3; 25°C	1.4±0.2	4.0±0.2	2.0±0.3	6.6±0.1
23.4% glycerol; 0.04 g/cm^3; 25°C	1.6±0.1	3.9±0.1	2.1±0.1	7.1±0.1
D2O; 0.14 g/cm^3; 25°C	0.7±0.1	4.5±0.1	1.3±0.1	7.8±0.5
19.9% methanol; 0.18 g/cm^3; 25°C	1.4±0.1	3.8±0.3	1.9±0.1	6.0±0.5
23.4% glycerol; 0.17 g/cm^3; 25°C	1.6±0.1	3.5±0.1	2.1±0.1	8.6±0.5
D_2O; 0.30 g/cm^3; 5°C	0.6±0.1	4.0±0.2	1.1±0.1	10.0±0.1
D_2O; 0.30 g/cm^3; 25°C	0.6±0.1	4.0±0.2	1.1±0.1	9.5±0.1
D_2O; 0.30 g/cm^3; 45°C	0.6±0.2	4.0±0.2	1.1±0.1	9.0±0.1
6.5% methanol; 0.19 g/cm^3; 5°C	1.3±0.2	3.8±0.2	1.8±0.3	7.4±0.5
12.8% methanol; 0.18 g/cm^3; 5°C	1.3±0.1	3.8±0.1	1.8±0.1	6.7±0.5
19.9% methanol; 0.18 g/cm^3; 5°C	1.4±0.1	3.9±0.1	1.9±0.1	5.6±0.5
28.4% methanol; 0.16 g/cm^3; 5°C	1.4±0.1	4.0±0.4	2.0±0.2	4.7±0.5
6.5% methanol; 0.19 g/cm^3; 5°C	1.3±0.2	3.8±0.2	1.8±0.3	7.4±0.5
6.5% methanol; 0.19 g/cm^3; 25°C	1.3±0.1	3.7±0.1	1.8±0.1	7.6±0.3
6.5% methanol; 0.19 g/cm^3; 45°C	1.3±0.1	3.8±0.1	1.8±0.1	7.5±0.3

The protein was modelled as an oblate ellipsoid and screened coulombic potential was used as the intermolecular potential. The parameters volume fraction, temperature, dielectric constant, ionic strength, SLD of protein and solvent were all fixed, and the parameters R_a, R_b, charge and incoherent background were allowed to vary. Plots of ln I(Q) vs Q for HSA solutions (0.04 g/cm^3) in D_2O, 19.9% methanol and 23.4% glycerol in the absence of NaCl and at 25°C are shown in Figure 6 and the fitted parameters presented in Table 4. The symbols are the experimental data and the dashed lines the results of the fitting. The first correlation peak becomes prominent in the absence of NaCl. The ellipsoid used in this work to fit the data is similar to the one used by Sjoberg and Mortensen [35] who modeled HSA at a concentration of 0.26 g/cm3 as an oblate ellipsoid of semi axes Ra =1.9 nm and Rb = 6.8 nm.

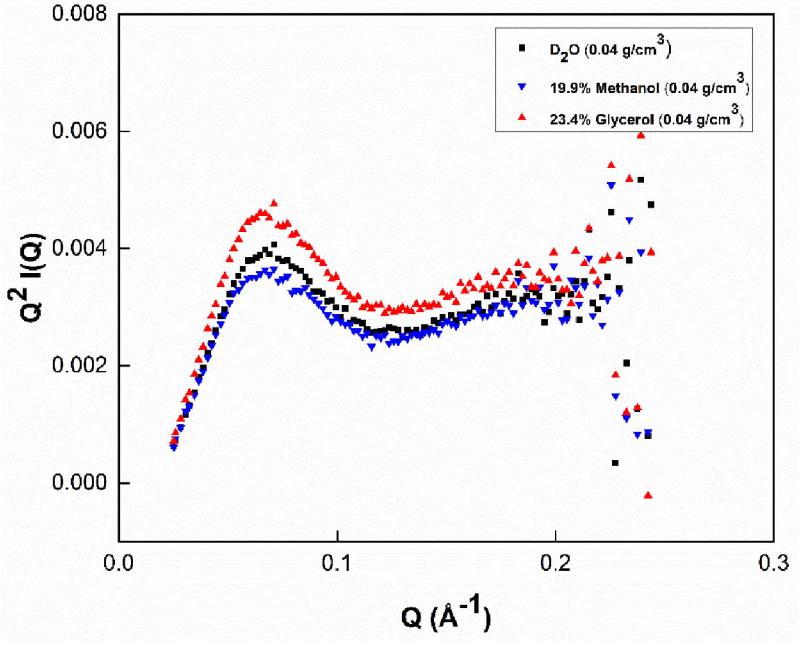

Figure 7. Kratky plots showing the ffect of cosolvents on HSA at 0.04 g/cm^3

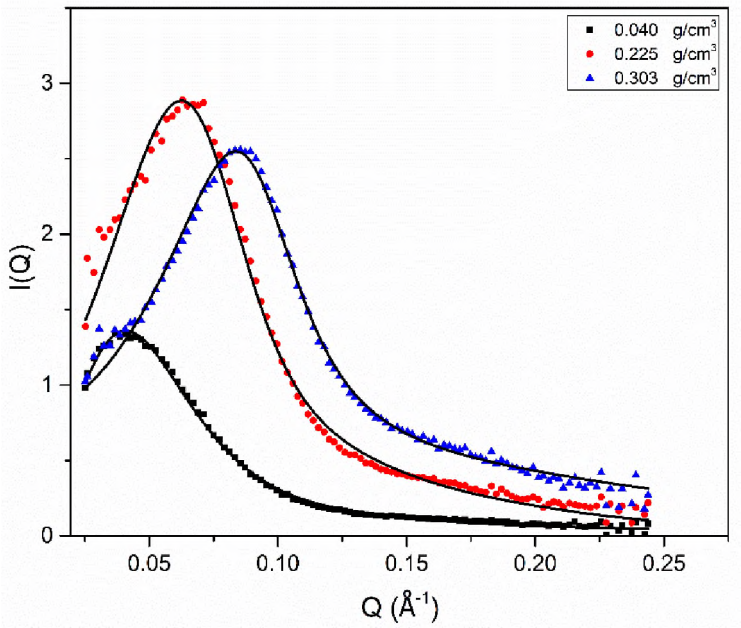

Figure 8. Experimental data for HSA in D_2O at high concentrations with its corresponding fits

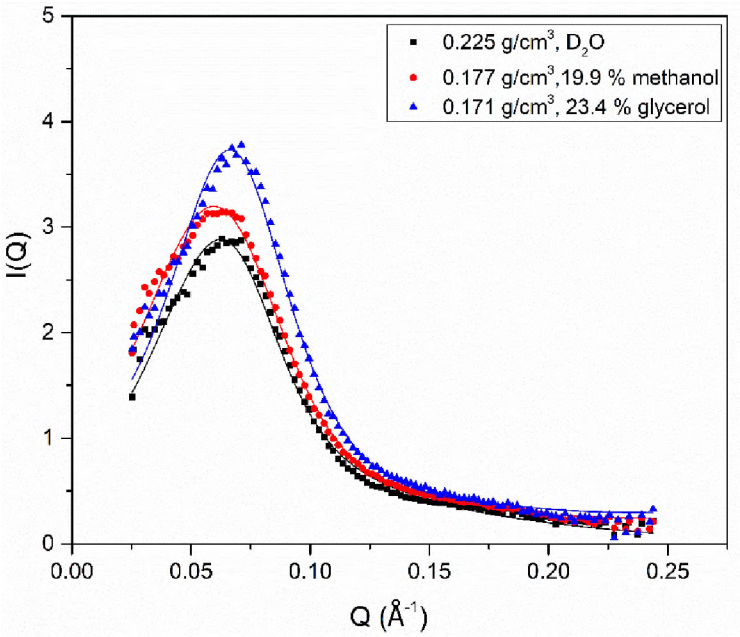

Figure 9. Effect of cosolvents on HSA at high concentrations

But it is quite different than the data reported by Carter et al. who modeled HSA at a concentration of 0.12 to 0.18 g/cm3 as an oblate ellipsoid of semi axes 14 and 4 nm [12]. The variations in Ra and Rb are not statistically significant but the increase in charge is. Notice that the size and the shape of the scatters in the presence and absence of NaCl are the same within the statistical noise.

The degree of compactness of the protein can be determined using a Kratky plot. A plot of Q^2. I(Q) vs Q for the two cosolvents in comparison with D_2O is shown in Figure 7. The three curves shown in Figure 7 for HSA at 0.04 g/cm^3 have a bell-shaped curve, which is a characteristic of globular macromolecules. The maxima in Figure 7 are located at 0.656 (glycerol), 0.684 (D2O) and 0.704 (Methanol). Therefore, the protein takes a more extended configuration as the solvent is changed from glycerol to water to methanol.

Plots of ln I(Q) vs Q with its corresponding fits for HSA in D_2O at various concentrations and in the presence of the two cosolvents are shown in Figures 8 and 9. As the concentration increases from 0.04 g/cm^3 to 0.3 g/cm^3 there is a clear shift of the peak towards higher Q values (Figure 8); so, the scatters are closer to each other. A comparison of the effect of methanol and glycerol at higher protein concentrations is shown in Figure 9. The correlation peak in the presence of glycerol moves towards higher Q values whereas methanol has the opposite effect. There is an increase in the apparent charge as we move from methanol, to water and to glycerol. This suggests an increase in repulsions; the solution becomes more liquid like.

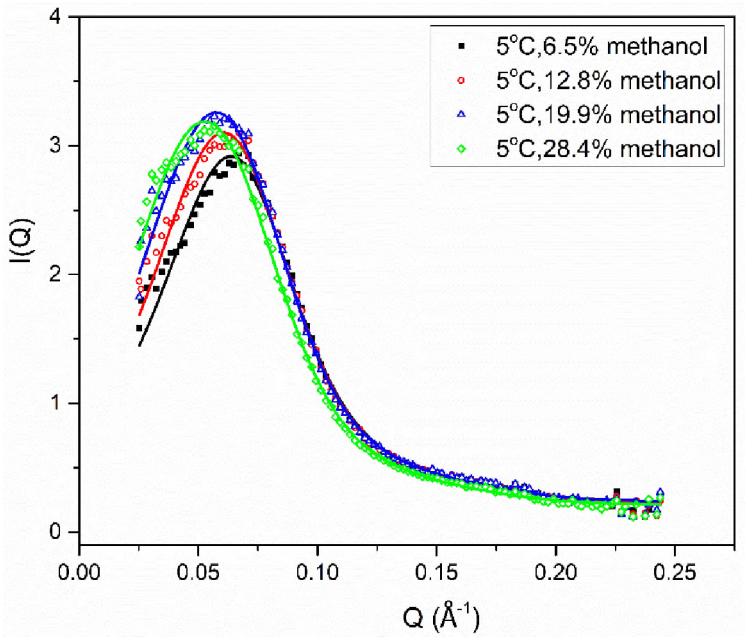

Figure 10. Effect of methanol on HSA in D_2O at 5°C

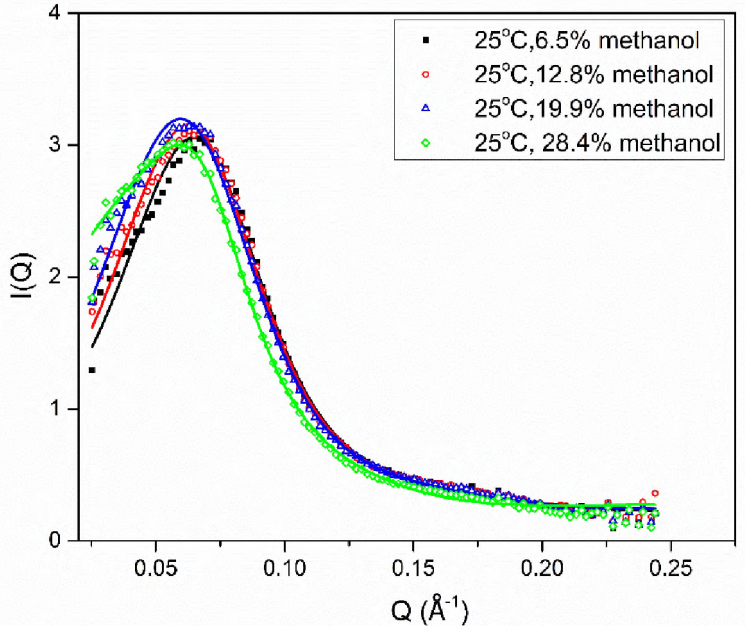

Figure 11. Effect of methanol on HSA in D_2O at 25°C

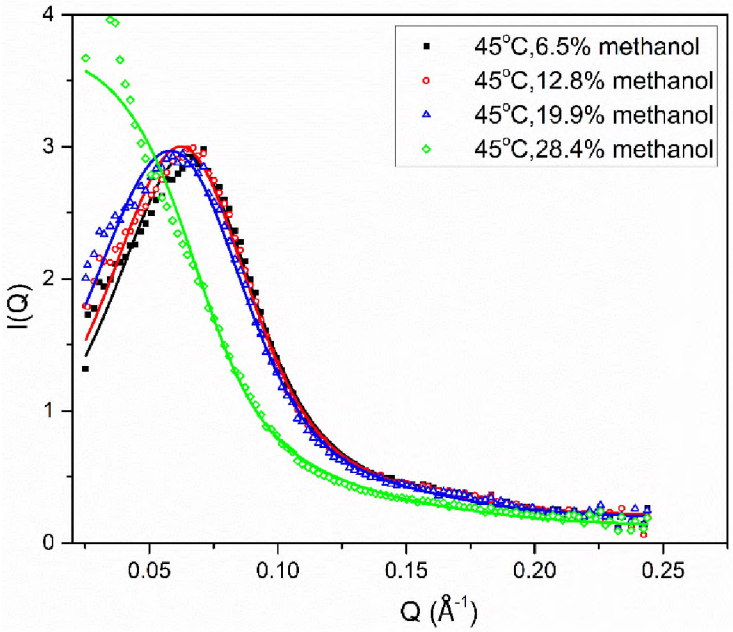

Figure 12. Effect of methanol on HSA in D_2O at 45°C

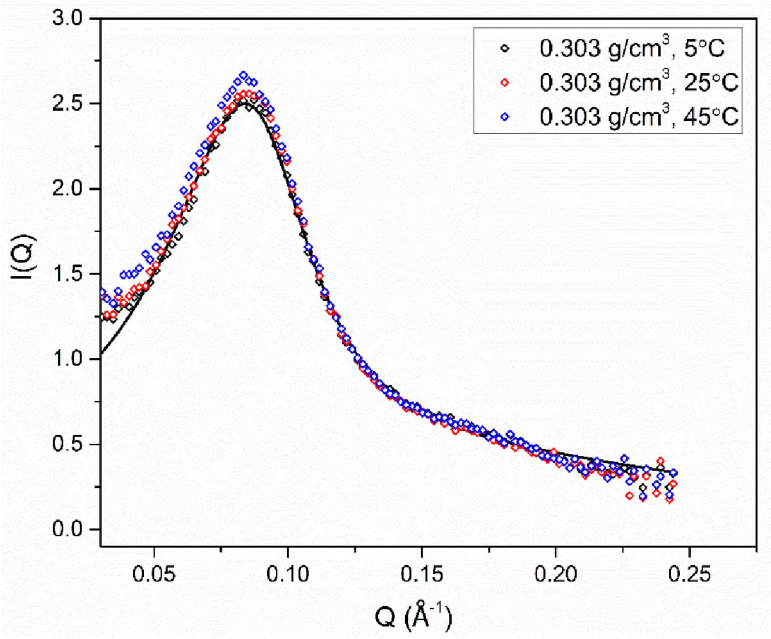

Figure 13. Effect of temperature for HSA in D_2O at 0.303 g/cm^3

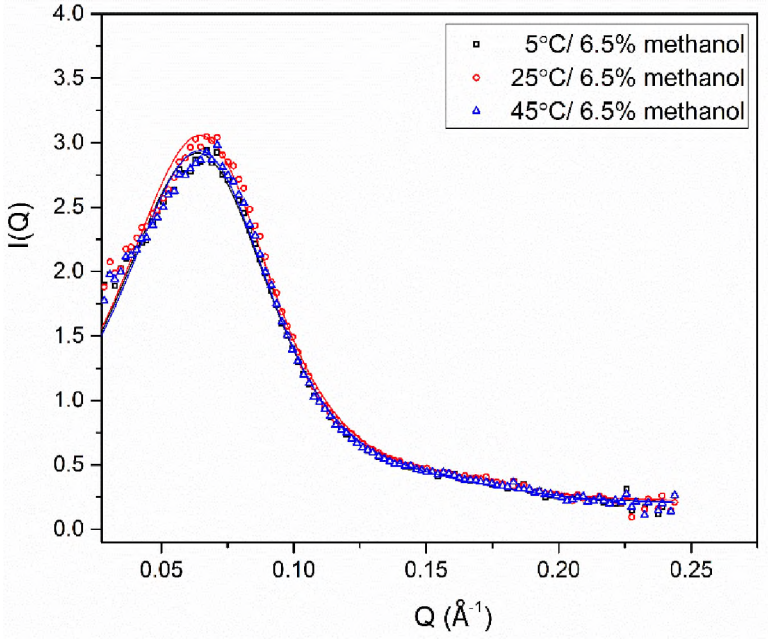

Figure 14. Effect of temperature on HSA in D_2O at 6.5% methanol and 0.19 g/cm^3

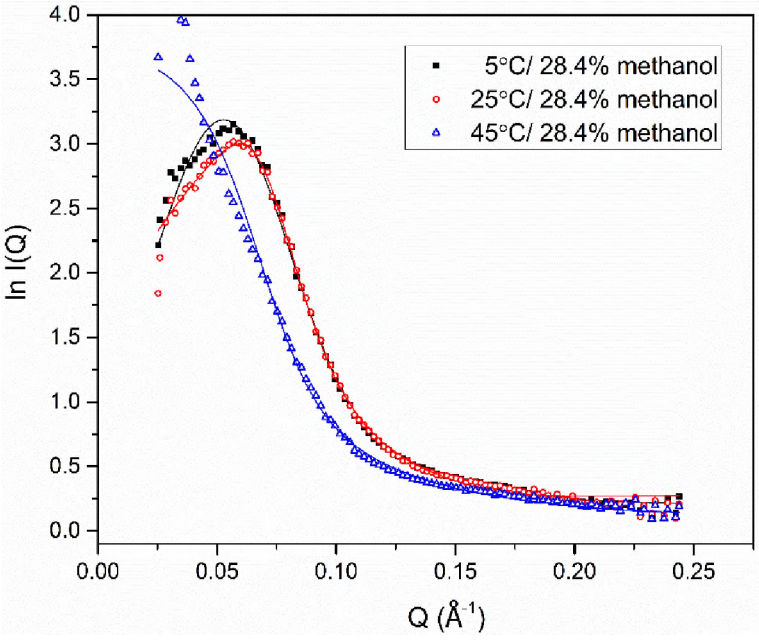

Figure 15. Effect of temperature on HSA in D_2O at 28.4% methanol and 0.16 g/cm^3

The effect on the structure of the HSA solution of increasing amounts of methanol at 5, 25, and 45°C is shown in Figures 10, 11 and 12 respectively. The symbols are the experimental points and the lines the best fits. The figures show that the intensity of the first correlation peak increases and the peak moves towards lower Q values as the methanol concentration increases (Figures 10-12). Although it is possible that this pattern is cause by the aggregation of the protein, the typical bending upwards of the scattering curve that reflects the presence of large aggregates is absent.

The FT-IR experiments indicate that no changes in the ternary or secondary structure of the protein that will expose more hydrophobic groups to the solvent are taking place. The addition of methanol is effectively decreasing the range of the force and this is captured by the screened potential as a decrease in the net charge.

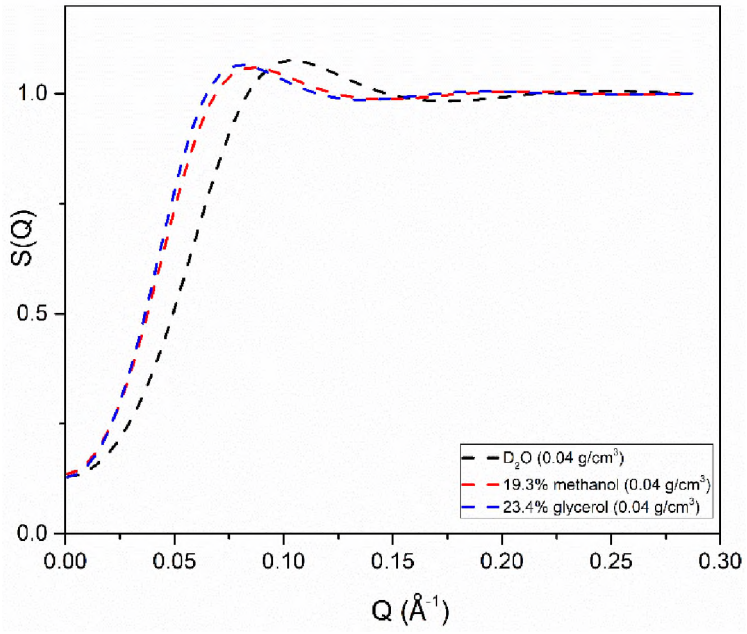

Figure 16. Interparticle structure factor calculations for HSA at 0.04 g/cm^3 in the presence of different solvents

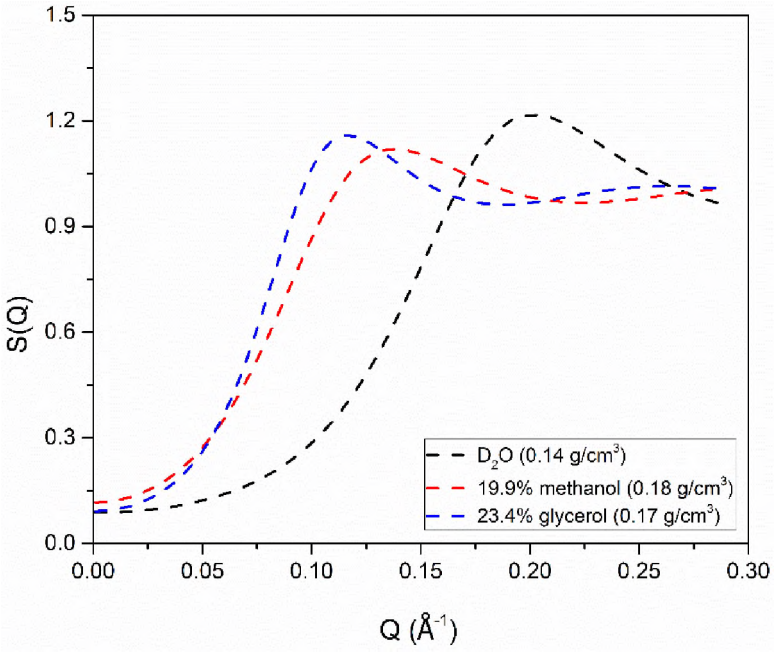

Figure 17. Interparticle structure factor calculations for HSA at higher concentration in the presence of different solvents

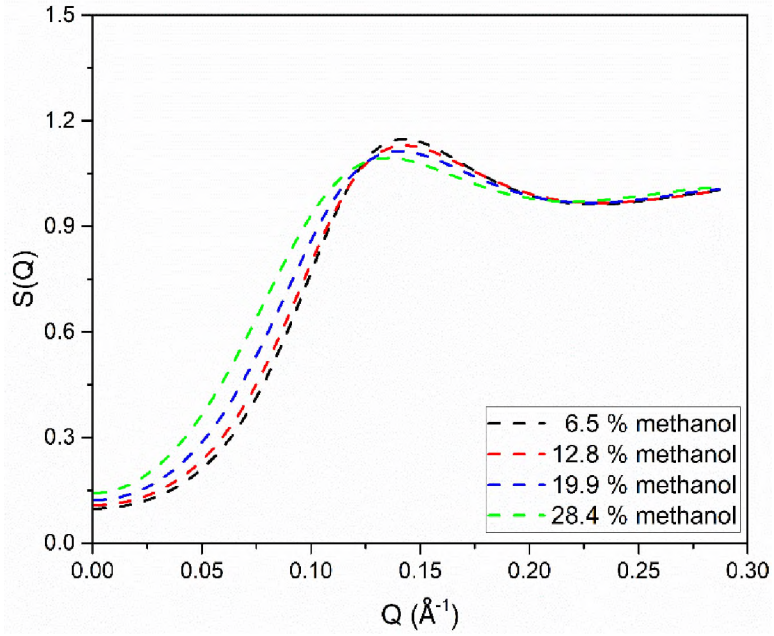

Figure 18. Interparticle structure factor calculations for HSA at 5°C and different methanol concentrations

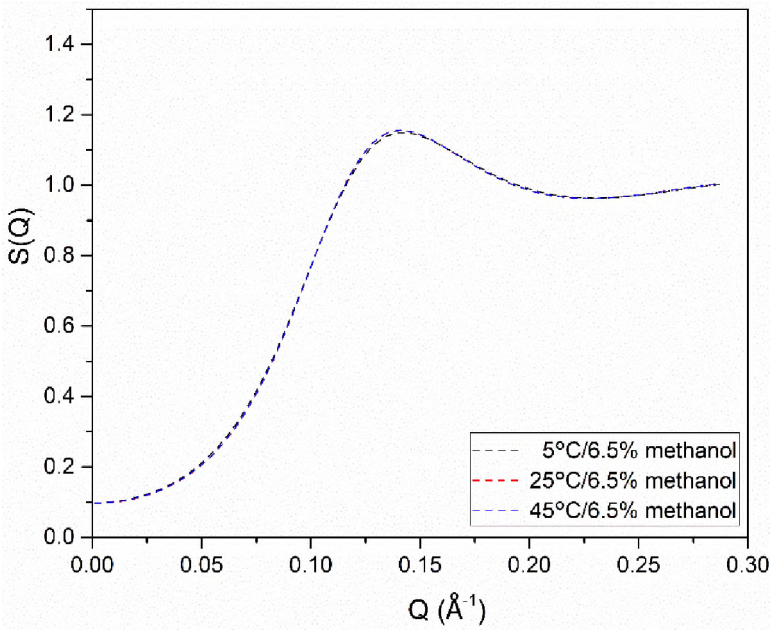

Figure 19. Interparticle structure factor calculations for HSA at 6.5% methanol and different temperatures

The effect of temperature on the scattered intensity in D_2O, 6.5% and 28.4% methanol concentrations is shown in Figure 13-15 respectively. The lines represent the best fit using the same approach described above. The fitted parameters are shown in Table 4. The curves in D_2O and low methanol concentrations are almost identical. Aggregation has likely occurred at 45°C and 28.4% methanol.

The interparticle structure factor (S(Q)) for HSA at various conditions was calculated using Hayter-Penfold rescaled mean spherical approximation (HPMSA) for a system of charged particles in a dielectric medium [23,36,37]. The fitted parameters (equivalent radius, charge and incoherent background) obtained in the fitting of I(Q) were used. The results are summarized in Figures 16-19. Figures 16 and 17 show a plot of S(Q) vs Q at 25°C for D_2O, 19.3% methanol and 23.4% glycerol for HSA at 0.04 g/cm^3 and 0.22

g/cm^3 respectively. The cosolvents show an almost identical on the structure of the solution. The effective radius increased from 1.6 nm for D_2O to 2.0 nm for both methanol and glycerol at low protein concentration which shifted the peak towards lower Q-values for both the solvents. Similarly, when the concentration of the protein is increased (Figure 17), the effective radius increased from 1.3 nm for D_2O to 2.0 nm for both methanol and glycerol respectively. The dielectric constants of both methanol (70) and glycerol (70.2) are smaller than the one for water (78). Therefore, the screened coulombic potential should become stronger in the presence of the solvents. The charges used to calculate S(Q) decrease in the order methanol<glycerol<D_2O at low protein concentration while at higher concentration it is methanol<D_2O <glycerol. Therefore, the potential should be stronger in the presence of glycerol and weaker in the presence of methanol.

Figure 18 shows calculated S(Q) for various concentrations of methanol at a protein concentration equal to 0.16 g/cm^3 and 5°C. As the cosolvent concentration increases, the peak in S(Q) moves towards lower Q values. Both the dielectric constant and the effective charge decrease as the concentration of methanol increases. The net result is a decrease in the intermolecular potential as the concentration of the alcohol increases. A similar trend is observed at 25°C. A negligible effect of temperature on S(Q) has been found (Figure 19). It has been shown[30] that the addition large amounts of methanol (50%) causes proteins to "swell" exposing more of the hydrophobic core but simultaneously to increase the amount of secondary structure (alpha helixes). Those effects are caused by direct binding of methanol to the polypeptide change via single hydrogen bonds and by indirectly increasing intramolecular polar interactions. We do not observed that here since the size of the protein remains nearly constant from 6.5% to almost 30% methanol.

4. CONCLUSIONS

HSA shape changed from a prolate to an oblate ellipsoid in the absence of intermolecular forces as the protein concentration was increased. At higher concentrations and in the presence of intermolecular forces the size and shape of the protein remains invariant within the statistical noise. By fitting the experimental data with a simple model, we demonstrated that the protein solution becomes more liquid like (more disordered) in the presence of methanol and more crystal like (more ordered) in the presence of glycerol.

ACKNOWLEDGEMENT

This work was supported by a grant from the Whitaker Foundation. The research at Oak Ridge was supported by the Division of Materials Sciences, U.S. Department of Energy under contract No. DE-AC05-96OR22464 with Lockheed Martin Energy Research Corp. We would like to thank Prof. R. Triolo for helping me in obtaining the SANS data and Ms. M. Roberts for obtaining the FT-IR spectra.

REFERENCES

(1) Krüger, P.; Schnell, S. K.; Bedeaux, D.; Kjelstrup, S.; Vlugt, T. J. H.; Simon, J.-M. Kirkwood–Buff Integrals for Finite Volumes. J. Phys. Chem. Lett. 2013, 4 (2), 235–238.

(2) Khmelnitsky, Y. L.; Belova, A. B.; Levashov, A. V.; Mozhaev, V. V. Relationship between Surface Hydrophilicity of a Protein and Its Stability against Denaturation by Organic Solvents. FEBS Lett. 1991, 284 (2), 267–269.

(3) Zhang, F.; Roosen-runge, F.; Skoda, M. W. A.; Jacobs, R. M. J.; Wolf, M.; Callow, P.; Frielinghaus, H. Hydration and Interactions in Protein Solutions Containing Concentrated Electrolytes Studied by Small-Angle Scattering. 2012, 2483–2493.

(4) Ianeselli, L.; Zhang, F.; Skoda, M. W. A.; Jacobs, R. M. J.; Martin, R. A.; Callow, S.; Pre, S.; Schreiber, F. Protein - Protein Interactions in Ovalbumin Solutions Studied by Small-Angle Scattering : Effect of Ionic Strength and the Chemical Nature of Cations. 2010, 3776–3783.

(5) Kita, Y.; Arakawa, T.; Lin, T. Y.; Timasheff, S. N. Contribution of the Surface Free-Energy Perturbation to Protein Solvent Interactions. Biochemistry 1994, 33 (50), 15178–15189.

(6) Chin, J. T.; Wheeler, S. L.; Klibanov, A. M. On Protein Solubility in Organic Solvent. Biotechnol. Bioeng. 1994, 44 (1), 140–145.

(7) Gonnelli, M.; Strambini, G. B. Glycerol Effects on Protein Flexibility: A Tryptophan Phosphorescence Study. Biophys. J. 1993, 65 (1), 131–137.

(8) Almagor, A.; Priev, A.; Barshtein, G.; Gavish, B.; Yedgar, S. Reduction of Protein Volume and Compressibility by Macromolecular Cosolvents: Dependence on the Cosolvent Molecular Weight. Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol. 1998, 1382 (1), 151–156.

(9) Kim, D.; Lee, Y. J. Effect of Glycerol on Protein Aggregation: Quantitation of Thermal Aggregation of Proteins from CHO Cells and Analysis of Aggregated Proteins. J. Therm. Biol. 1993, 18 (1), 41–48.

(10) Char, K.; Frank, C. W.; Gast, A. P.; Tang, W. T. Hydrophobic Attraction of Pyrene-End-Labeled Poly(Ethylene Glycol) in Water and Water-Methanol Mixtures. Macromolecules 1987, 20 (8), 1833–1838.

(11) Haro, M. L. de; Robles, M. The Structure Factor and Equation of State of Hard-Sphere Fluids. J. Phys. Condens. Matter 2004, 16 (22), S2089–S2096.

(12) Carter, D. C.; He, X. M.; Munson, S. H.; Twigg, P. D.; Gernert, K. M.; Broom, M. B.; Miller, T. Y. Three-Dimensional Structure of Human Serum Albumin. Science 1989, 244 (4909), 1195–1198.

(13) He, X. M.; Carter, D. C. Atomic Structure and Chemistry of Human Serum Albumin. Nature 1992, 358 (6383), 209–215.

(14) Sontum, P. C.; Christiansen, C. Photon Correlation Spectroscopy Applied to Characterisation of Denaturation and Thermal Stability of Human Albumin. J. Pharm. Biomed. Anal. 1997, 16 (2), 295–302.

(15) Zhang, F.; Skoda, M. W. A.; Jacobs, R. M. J.; Martin, R. A.; Martin, C. M.; Schreiber, F. Protein Interactions Studied by SAXS: Effect of Ionic Strength and Protein Concentration for BSA in Aqueous Solutions. J. Phys. Chem. B 2007, 111 (1), 251–259.

(16) O'Brien, E. P.; Brooks, B. R.; Thirumalai, D. Effects of PH on Proteins: Predictions for Ensemble and Single-Molecule Pulling Experiments. J. Am. Chem. Soc. 2012, 134 (2), 979–987.

(17) Liu, H.; Zwart, P. H. Determining Pair Distance Distribution Function from SAXS Data Using Parametric Functionals. J. Struct. Biol. 2012, 180 (1), 226–234.

(18) Rice, S. A. Small Angle Scattering of X-Rays. A. Guinier and G. Fournet. Translated by C. B. Wilson and with a Bibliographical Appendix by K. L. Yudowitch. Wiley, New York, 1955. 268 Pp. $7.50. J. Polym. Sci. 1956, 19 (93), 594–594.

(19) Bendedouch, D.; Chen, S.; Koehler, W. C.; Lin, J. S. A Method for Determination of Intra- and Interparticle Structure Factors of Macroions in Solution from Small Angle Neutron Scattering. J. Chem. Phys. 1982, 76 (10), 5022–5026.

(20) Frenkel, D.; Vos, R. J.; de Kruif, C. G.; Vrij, A. Structure Factors of Polydisperse Systems of Hard Spheres: A Comparison of Monte Carlo Simulations and Percus–Yevick Theory. J. Chem. Phys. 1986, 84 (8), 4625–4630.

(21) Sarangapani, P. S.; Hudson, S. D.; Jones, R. L.; Douglas, J. F.; Pathak, J. A. Critical Examination of the Colloidal Particle Model of Globular Proteins. Biophys. J. 2015, 108 (3), 724–737.

(22) Kaieda, S.; Lund, M.; Plivelic, T. S.; Halle, B. Weak Self-Interactions of Globular Proteins Studied by Small-Angle X-Ray Scattering and Structure-Based Modeling. J. Phys. Chem. B 2014, 118 (34), 10111–10119.

(23) Kline, S. R. Reduction and Analysis of SANS and USANS Data Using IGOR Pro. J. Appl. Crystallogr. 2006, 39 (6), 895–900.

(24) Franke, D.; Petoukhov, M. V.; Konarev, P. V.; Panjkovich, A.; Tuukkanen, A.; Mertens, H. D. T.; Kikhney, A. G.; Hajizadeh, N. R.; Franklin, J. M.; Jeffries, C. M.; Svergun, D. I. ATSAS 2.8: A Comprehensive Data Analysis Suite for Small-Angle Scattering from Macromolecular Solutions. J. Appl. Crystallogr. 2017, 50, 1212–1225.

(25) Jouyban, A.; Soltanpour, S.; Chan, H. K. A Simple Relationship between Dielectric Constant of Mixed Solvents with Solvent Composition and Temperature. Int. J. Pharm. 2004, 269 (2), 353–360.

(26) Jouyban, A.; Soltanpour, S. Prediction of Dielectric Constants of Binary Solvents at Various Temperatures. J. Chem. Eng. Data 2010, 55 (9), 2951–2963.

(27) Åkerlöf, G. Dielectric Constants of Some Organic Solvent-Water Mixtures at Various Temperatures. J. Am. Chem. Soc. 1932, 54 (11), 4125–4139.

(28) Uversky, V. N.; Narizhneva, N. V.; Kirschstein, S. O.; Winter, S.; Löber, G. Conformational Transitions Provoked by Organic Solvents in β-Lactoglobulin: Can a Molten Globule like Intermediate Be Induced by the Decrease in Dielectric Constant? Fold. Des. 1997, 2 (3), 163–172.

(29) Nakka, P. P.; Li, K.; Forciniti, D. Effect of Differences in the Primary Structure of the A-Chain on the Aggregation of Insulin Fragments. ACS Omega 2018.

(30) Havel, H. A.; Kauffman, E. W.; Plaisted, S. M.; Brems, D. N. Reversible Self-Association of Bovine Growth Hormone during Equilibrium Unfolding. Biochemistry 1986, 25 (21), 6533–6538.

(31) Olivieril, J. R.; Craievich, A. F. The Subdomain Structure of Human Serum Albumin in Solution under Different PH Conditions Studied by Small Angle X-Ray Scattering. Eur. Biophys. J. 1995, 24 (2), 77–84.

(32) Svergun, D. I.; Richard, S.; Koch, M. H.; 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 (5), 2267–2272.

(33) Svergun, D. I.; IUCr. Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. J. Appl. Crystallogr. 1992, 25 (4), 495–503.

(34) Shannon, G.; Marples, C. R.; Toofanny, R. D.; Williams, P. M. Evolutionary Drivers of Protein Shape. Sci. Rep. 2019, 9 (1), 1–15.

(35) Sjöberg, B.; Mortensen, K. Interparticle Interactions and Structure in Nonideal Solutions of Human Serum Albumin Studied by Small-Angle Neutron Scattering and Monte Carlo Simulation. Biophys. Chem. 1994, 52 (2), 131–138.

(36) Antonio, M. R.; Bera, M. K. PH-Dependent Interactions between Keggin Heteropolyanions in Dilute Solutions. Eur. J. Inorg. Chem. 2019, 2019 (3–4), 367–373.

(37) Artykulnyi, O. P.; Shibaev, A. V.; Avdeev, M. M.; Ivankov, O. I.; Bulavin, L. A.; Petrenko, V. I.; Philippova, O. E. Structural Investigations of Poly(Ethylene Glycol)-Dodecylbenzenesulfonic Acid Complexes in Aqueous Solutions. J. Mol. Liq. 2020, 308, 113045.

II. EFFECT OF DIFFERENCES IN THE PRIMARY STRUCTURE OF THE A-CHAIN ON THE AGGREGATION OF INSULIN FRAGMENTS

ABSTRACT

Bovine and human insulin have similar primary structures. In this article, the region of the insulin A-chain of bovine and human insulin where the amino acid composition is different was studied. Bovine (BIF) and human (HIF) fragments were synthesized in solid phase peptide synthesis. The effect of pH, temperature, urea, ionic strength and stirring on the formation of fibrils was studied using a fractional factorial resolution III experimental design. Fibrillation was monitored by fluorescence and infrared spectroscopy and optical microscopy. Both fragments formed fibrils at pH 1.6 and a temperature of 60°C. The lag time and apparent aggregation growth rate constant were determined using a two-parameter kinetic model. It was found that the bovine insulin fragment has a shorter lag time than the human insulin one while the exponential phase rate was faster for HIF than for BIF. An increase in β-sheets content with time was observed in both fragments. The increase in β-sheets was preceded by an initial decrease in α-helices followed by an intermediate increase during the transition from the lag phase to elongation phase. Temperature and ionic strength are among the most important experimental factors during the lag phase whereas ionic strength is replaced by pH during the elongation phase for both the fragments. Congo red binding confirmed the presence of ring-like oligomer structures rich in anti-parallel β-sheets which tend to form fibrils rich in parallel β-sheets.

1. INTRODUCTION

Protein aggregation is a major problem found in the bioprocessing industry. Aggregation can occur at any step, from production to purification and packing [1]. Protein therapeutics might lose their biological activity and cause problems during treatment if they aggregate. Proteins can form aggregates both in their native and denatured state. They can aggregate either during the lyophilization step or during their long-term storage. Aggregation also happens during the unfolding/refolding reaction in the production of recombinant proteins.

Protein aggregation in vivo is associated with disease. Formation of amyloid fibrils is the manifestation of various amyloidogenic diseases like Alzheimer's, neurofibrillary tangle, neurodegeneration, Parkinson's, etc. A total of 36 proteins/ peptides have been identified to form amyloid fibrils in vivo till 2016 [2]. Amyloid fibrils are β-sheet rich linear aggregates formed from circular oligomers. X-Ray diffraction of amyloid fibrils shows a cross-β diffraction pattern which is a trademark of intermolecular β-sheet structures [3-6]. This intermolecular β-sheet structure is easily identified by FTIR. It has been recently proved that oligomers are the elemental blocks to build a fibril [7,8]. The propensity to form amyloid fibrils seems to be correlated with the hydrophobicity of proteins [9]. In spite of its association with disease, amyloid aggregates can also be used as scaffolds, as peptide hormones storage and natural adhesives [10].

Although amyloid fibrils are mostly related to disease causing proteins, it was also reported that if appropriate incubation conditions were provided even a non-pathogenic protein can form a fibrillar structure [11]. Fibrils can also be synthesized in vitro by

subjecting the protein to favorable destabilizing conditions [12,13]. For example, hen lysozyme incubated at pH 2.0 and 37°C forms fibrils [14].

Insulin aggregates at acidic pH and high temperature [15]. Insulin is a highly-conserved hormone exhibiting minor differences in the amino acid sequence from species to species. For instance, Ala replaces Thr at position 8, Val replaces Ile at position 10 and Ala replaces Thr at position 30 in bovine and human insulin respectively [16]. Our previous studies showed that the aggregation kinetics (both lag times and rate constants) is very different for bovine and human insulin [17]. Moreover, dynamic light scattering and FTIR experiments showed that both insulins seem to follow different aggregation pathways as shown by different populations of aggregate sizes and different secondary structure changes. Still, mature fibrils were practically identical. Previously, it was demonstrated that the B chain fragment LVEALYL plays a key role in aggregation [18]. A recent review shows that different insulin analogues show different aggregation pathways [19]. In an attempt to identify the causes for those differences in the aggregation kinetics, we decided to study fragments of the insulin A-chain containing the sequence where bovine and human insulin are different [16]. Insulin A-chain is a 21-residue peptide in which only the amino acids at positions 8 and 10 differ for bovine and human insulin [20]. Bovine insulin has alanine and valine at positions 8 and 10 whereas human insulin has threonine and isoleucine. We selected the sequences **ASV**CSLYQLENK$_3$ (BIF) and **TSI**CSLYQLENK$_3$ (HIF) for the bovine and human fragments respectively. Bold letters identify the amino acids that are different in both fragments. A lysine tail was added to each fragment to increase its solubility [21].

To study the effects of temperature, urea, ionic strength, pH and stirring on the fibrillation process, a screening experimental design was chosen. The factors were chosen based on the following criteria. Temperature affects protein folding and therefore the formation of the nuclei. Urea, a known chaotropic, was chosen to study the effect of water structure on fibril formation.

Table 1. Experimental design FF0508

Run	pH	Temp. [°C]	Urea [M]	Ionic [M]	Stirring
1	-	-	-	-	+
2	-	-	+	+	-
3	-	+	-	+	-
4	-	+	+	-	+
5	+	-	-	+	+
6	+	-	+	-	-
7	+	+	-	-	-
8	+	+	+	+	+

Ionic strength and pH affects coulombic interactions among the side chains of amino acids which affects fibrillation. Interfaces (solid/liquid or gas/liquid) induce protein aggregation. Air was incorporated into samples by stirring to study the effect of gas/liquid interfaces on fibrillation. The evolution of the aggregation process with time was monitored using a thioflavin T (ThT) fluorescence assay. Secondary structure changes were

monitored using attenuated total reflection (ATR)-FTIR spectroscopy. The presence of Congo red stained fibrils [22] was confirmed by light microscopy.

pH, temperature, urea, ionic strength and stirring were chosen as factors for the experimental design. A two-level fractional factorial design FF0508 of resolution III as shown in Table 1 consisting of five factors at two levels each with a total of 8 runs was chosen.

Table 2. Experimental factors with lower and higher levels

Factor	-	+
pH	1.6	5
Temperature	25°C	60°C
Urea (M)	0	1
Ionic Strength (M)	0.02	1
Stirring	Off	On

Each main effect in this design was confounded with one or more two-factor interactions. The lower and higher levels chosen for the factors are shown in Table 2.

2. EXPERIMENTAL SECTION

2.1. MATERIALS

All Fmoc amino acids, Wang resin and HBTU were purchased from CS BIO Co. (USA). ThT was purchased from Acros Organics (USA). Piperidine, DMF, TFA, DIPEA and deuterium oxide; 99.9 atom % D were purchased from Sigma-Aldrich (USA). All other solvents were purchased from Fischer Scientific (USA).

2.2. PEPTIDE SYNTHESIS

The peptides BIF and HIF were synthesized in a CS Bio 336 peptide synthesizer using Fmoc technique in solid phase peptide synthesis [61]. Wang resin (0.5 mmol/g) was used as a solid support matrix for this purpose and the synthesis was carried out from the C- to the N- terminus [62]. One mmol per synthesis of each Fmoc amino acid was used. Deprotection of Fmoc group was done using 20% piperidine in DMF. HBTU combined with DIPEA was used as an activator to facilitate the coupling of amino acids. The complete synthesis took approximately 18 hours. The product from the reaction vessel was suspended in DCM and the peptide was cleaved from the swollen resin using 95% TFA. The cleaved peptide was then re-suspended in cold diethyl ether and centrifuged twice using RC-3B Refrigerated centrifuge, (Sorvall Instruments) at 4650 x g, -5°C for 30 min. The obtained pellet was dissolved in 10 mM acetic acid and lyophilized to obtain the final peptide fragment.

2.3. NMR SPECTROSCOPY

The amino acid composition of the peptides was confirmed using liquid state one-dimensional (1D) ^{1}H NMR spectroscopy. Samples for the analysis were prepared by dissolving 10 mg of the peptide in 750 µl D_2O (99.9 % Deuterium). Spectra were obtained using a Bruker 400-MHz spectrometer equipped with a BBO probe. ^{1}H NMR data were obtained with a sweep width of 3997.8 MHz, 0.256 s acquisition time, 16 scans and a 2s recycle delay.

2.4. MASS SPECTROMETRY

To determine the purity and the mass of the peptides LC-MS data were acquired on an Agilent 6520 Q-TOF LCMS system. Lyophilized peptides were suspended in 0.1% formic acid in water to obtain a final concentration of 1mg/ml. 5 µL of the sample was loaded on an Agilent Technologies 43mm C18 Chip column. Agilent Technologies 1200 Series Quaternary HPLC was used. LC Gradient: initial conditions 3% B (A: 0.1% formic acid in water; B: 99.9% acetonitrile, 0.1% formic acid), gradient to 80% B over 9 min, hold at 80% B for 3 min, ramp back to (0.5 min) and hold at (2.5min) initial conditions. An Agilent Technologies 6520A Accurate Mass QTOF MS with integrated Chip Cube source was used to collect data. Data across a total of 15 minutes of elution were collected. MS data were collected between 295 and 2500 m/z and MS/MS data were collected between 70-2500 m/z at 2 spectra/sec. Peptides with counts >2000 were picked for MS/MS.

LC-MS for the fibrils was obtained using an LTQ Orbitrap XL mass spectrometer. Samples were precipitated using ice-cold acetone. The pellet was recovered by centrifugation (16,000 x g) and washed twice with 80% acetone (in water). The pellet was

dried and resuspended in 100 μL of urea buffer (6M urea, 2M thiourea, 0.1M ammonium bicarbonate). No particulates were observed, suggesting fibrils were completely dissolved in the buffer.

2.5. IN VITRO FIBRILLATION

Samples were prepared with a peptide concentration of 10 mg/ml. A total of 16 samples, 8 for each fragment varying the experimental conditions were made. All the samples were prepared in 0.22 μm filtered phosphate buffer at their respective pH (1.6 and 5). Urea and sodium chloride with two different concentrations were added. Samples were incubated with ThT at two different temperatures (25°C and 60°C) under stirring and no stirring conditions. The evolution of the fibrillation process with time was followed by fluorescence assay and the changes in secondary structure were studied using FT-IR. The water used for the preparation of buffers was purified to a resistivity of 17.5 MΩ.cm.

2.6. THT FLUORESCENCE ASSAY

ThT assay was done using a NanoDrop 3300 Fluorospectrometer (Thermo Fisher Scientific) using white light as the excitation source (460-650nm). The emission wavelength was set to 482 nm. Free dye has an excitation wavelength of 385 nm and an emission wavelength of 445 nm. A stock solution for the ThT assay was prepared by adding 0.8 mg of ThT dye to 50 ml of 0.22 μm filtered phosphate buffer (pH 7) and stored at 4 °C in a dark place to prevent quenching. 0.05 mM of ThT solution was added to each sample at the beginning of incubation. A ThT solution with only the buffer was used as a blank for the assay. Fluorescence for each sample was recorded every three days. Increase in

viscosity or the appearance of gel/ precipitate with time upon visual inspection is a characteristic of the ThT assay [22]. All the reported fluorescence readings were an average of 5 replicates.

2.7. FTIR ANALYSIS

Analysis was carried out with a Nicolet 6700 FT-IR spectrometer from Thermoscientific equipped with a KBr beam-splitter and a DTGC detector cooled by liquid nitrogen. A smart performer Attenuated Total Reflectance sampling accessory with a Germanium (Ge) crystal was used. Interferograms were obtained for each sample at a resolution of 16 cm^{-1} and 256 scans were collected for each run to obtain a good signal-to-noise ratio.

2.8. CONGO RED STAINING

Congo red dye was used to stain and obtain the micrographs of the fibrils. A stock solution was prepared by adding 7 mg of Congo red to 1 ml of 0.22 μm filtered water. Samples containing fibrils were first centrifuged using a micro centrifuge at 9450 x g and then washed thoroughly with water. The pellet was then resuspended in 1 ml of water and 5 μL of Congo red solution was added and incubated at 25°C for one hour. After obtaining a red precipitate, the sample was centrifuged again. The pellet was washed twice with water to remove the unattached dye. The fibrils were then resuspended in a small amount of water. 30 μL of Congo red stained fibrils were placed on a microscopic slide and air dried. Micrographs were obtained at 1000x using Hirox KH-8700 digital microscope.

3. RESULTS AND DISCUSSION

The amino acid composition of both BIF and HIF was determined using ^{1}H NMR spectroscopy. ^{1}H resonance assignments for each amino acid were done based on its corresponding proton chemical shifts. All the data related to proton chemical shifts was obtained from the Biological Magnetic Resonance Data Bank [23]. Figure 1 and 2 shows the full ^{1}H NMR spectra for BIF and HIF respectively with their corresponding amino acid resonance assignments for various protons.

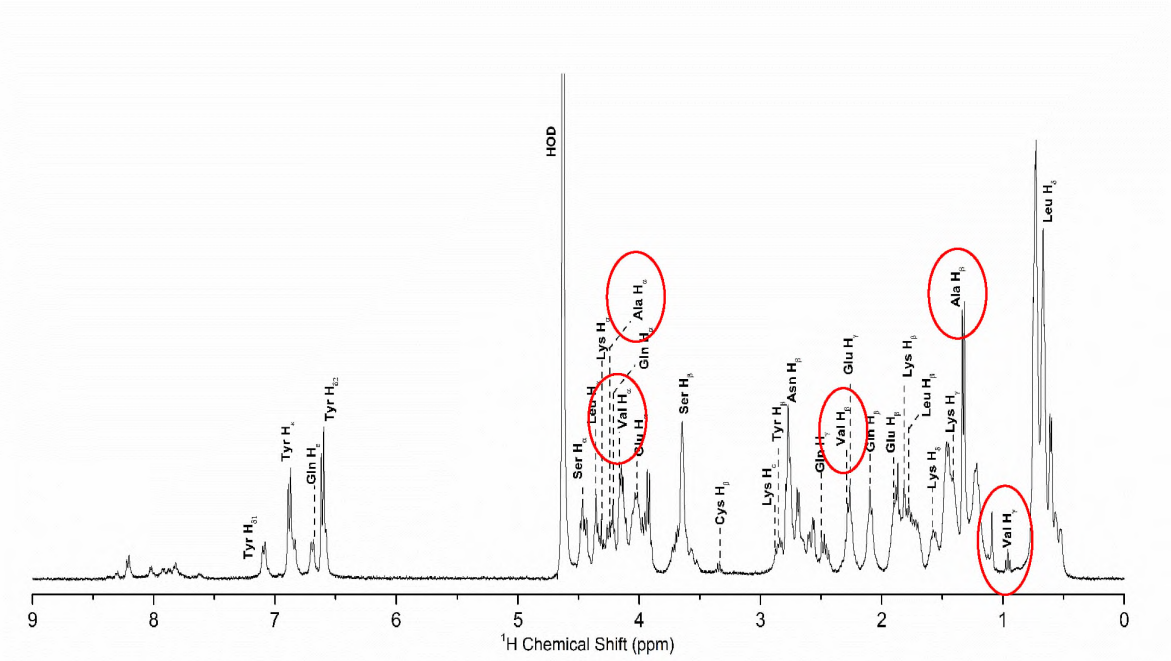

Figure 1. ^{1}H NMR spectrum of BIF highlighting the chemical shifts corresponding to Alanine and Valine at 25°C

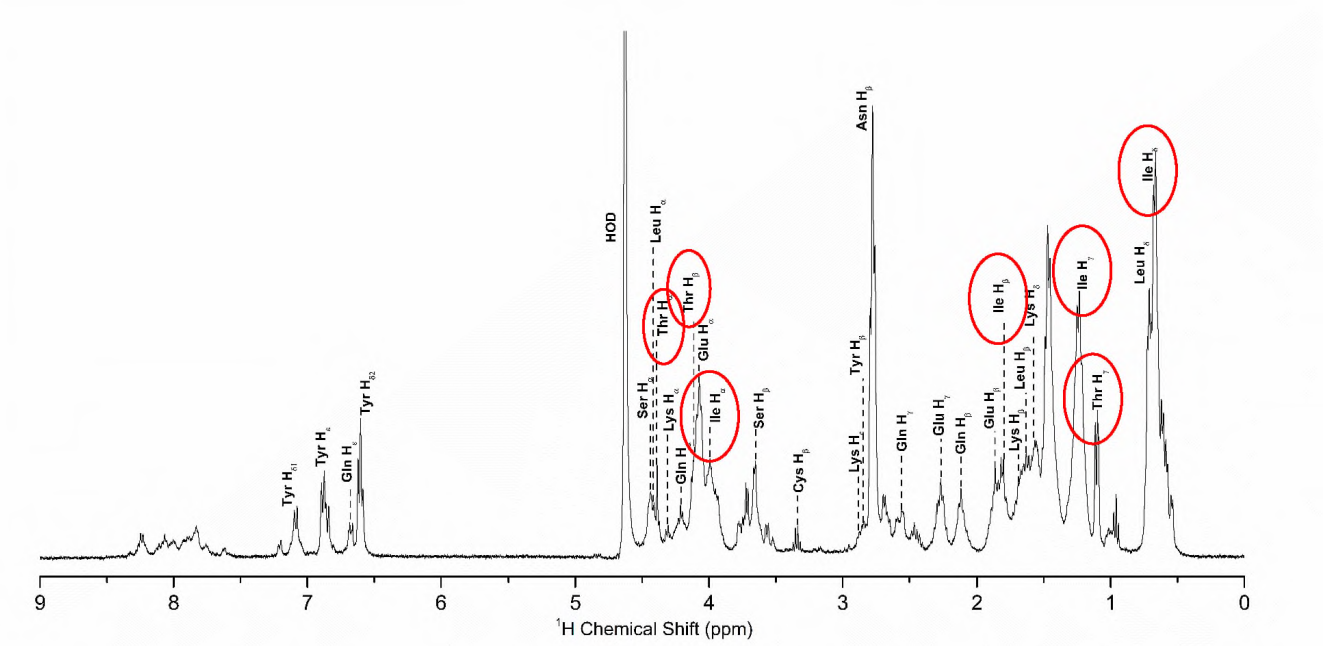

Figure 2. ^{1}H NMR spectrum of HIF highlighting the chemical shifts corresponding to Threonine and Iso leucine at 25°C

Both spectra look similar with a few minor differences in the peaks at the backbone and the sidechain region. Figure 1 has peaks at 4.25 ppm, 1.35 ppm, 4.17 ppm and 2.28 ppm which correspond to Ala H_α, Ala H_β, Val H_α and Val H_β respectively and seen only in BIF. Similarly Figure 2 has peaks at 4.38 ppm, 4.11 ppm, 3.99 ppm and 1.78 ppm which correspond to Thr H_α, Thr H_β, Ile H_α and Ile H_β respectively and seen in only HIF. Peaks assigned to rest of the amino acids are same in both the fragments (Table 3 and 4).

The purity and mass of both peptide fragments were calculated from the LC-MS data obtained using a Q-TOF LC/MS system. Figures 3 and 4 show the relative abundance and mass for native BIF and HIF respectively. The purity of the samples was calculated based on the relative abundance of each fraction. BIF full intact peptide (1-3 Lysines at the tail) has a purity of 72%. The major impurity found in both peptides is a Fmoc group (mass

of 164 Da) due to an inefficient final deprotection step. Fmoc attached peptide (impurity) is ~ 28% in BIF while it is ~ 66 % in HIF (53% Fmoc full peptide plus 13 % Fmoc with no Thr). Since the peptides were used without further purification, it is important to determine the composition of the fibrils obtained in the experiment using LC-MS to estimate the effect of the impurity on fibrillation.

Table 3. ^{1}H NMR chemical shifts for BIF

	Proton chemical shift (ppm)				
Amino Acid	H_α	H_β	H_γ	H_δ	H_ε
Ala	4.25	1.35	-	-	-
Ser	4.46	3.65	-	-	-
Val	4.17	2.28	0.96	-	-
Cys	4.62	3.32	-	-	-
Leu	4.35	1.77	-	0.66	-
Tyr	-	2.84	-	7.07,6.59	-
Gln	4.21	2.09	2.5	-	6.67
Glu	4.01	1.9	2.24	-	-
Asn	-	2.76	-	-	-
Lys	4.29	1.81	1.39	1.58	2.88

Table 4. ^{1}H NMR chemical shifts for HIF

Proton chemical shift (ppm)					
Amino Acid	H_α	H_β	H_γ	H_δ	H_ε
Thr	4.38	4.11	1.09	-	-
Ser	4.46	3.65	-	-	-
Ile	3.99	1.78	1.21	0.67	-
Cys	4.62	3.32	-	-	-
Leu	4.35	1.77	-	0.66	-
Tyr	-	2.84	-	7.07,6.59	-
Gln	4.21	2.09	2.5	-	6.67
Glu	4.01	1.9	2.24	-	-
Asn	-	2.76	-	-	-
Lys	4.29	1.81	1.39	1.58	2.88

Four samples, one with high β-sheet content (pH 1.6, 60°C, 1M urea, 0.02M NaCl) and the other with low β-sheet content (pH 5, 25°C, 1M urea, 0.02M NaCl) for both BIF and HIF were chosen for this analysis. Aggregates formed at pH 5 and 25°C in both the fragments consist of 93% Fmoc (BIF) and 100% HIF attached peptide in abundance as shown in Figures 5 and 6.

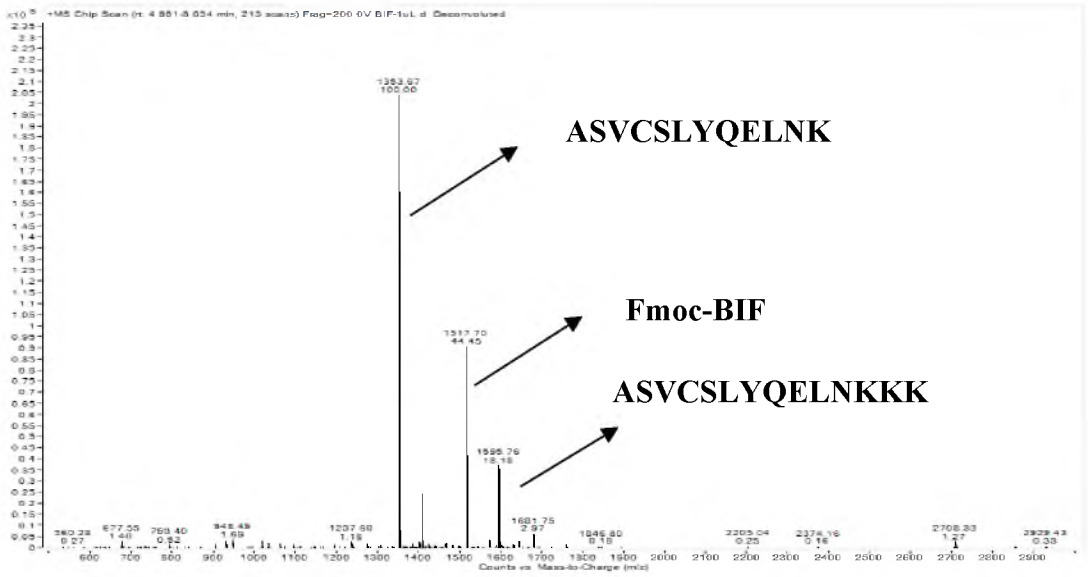

Figure 3. LC-MS of native BIF showing the native BIF (with 1 and 3 Lys tail) and impurity (Fmoc-BIF)

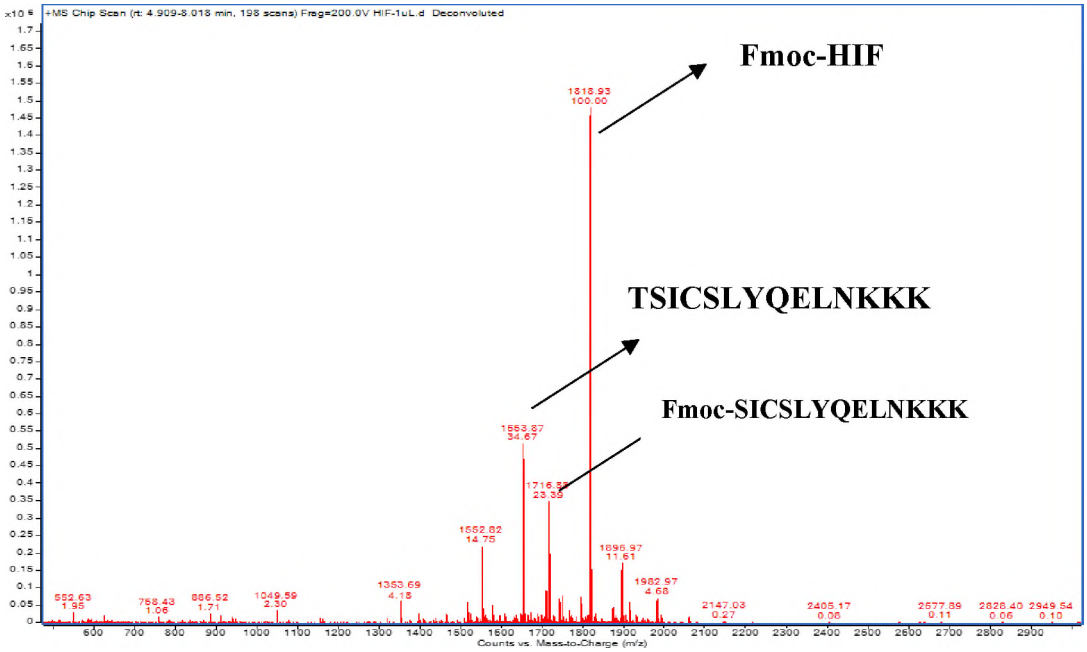

Figure 4. LC-MS of native HIF showing the native HIF, truncated HIF (with Fmoc) and impurity (Fmoc-HIF)

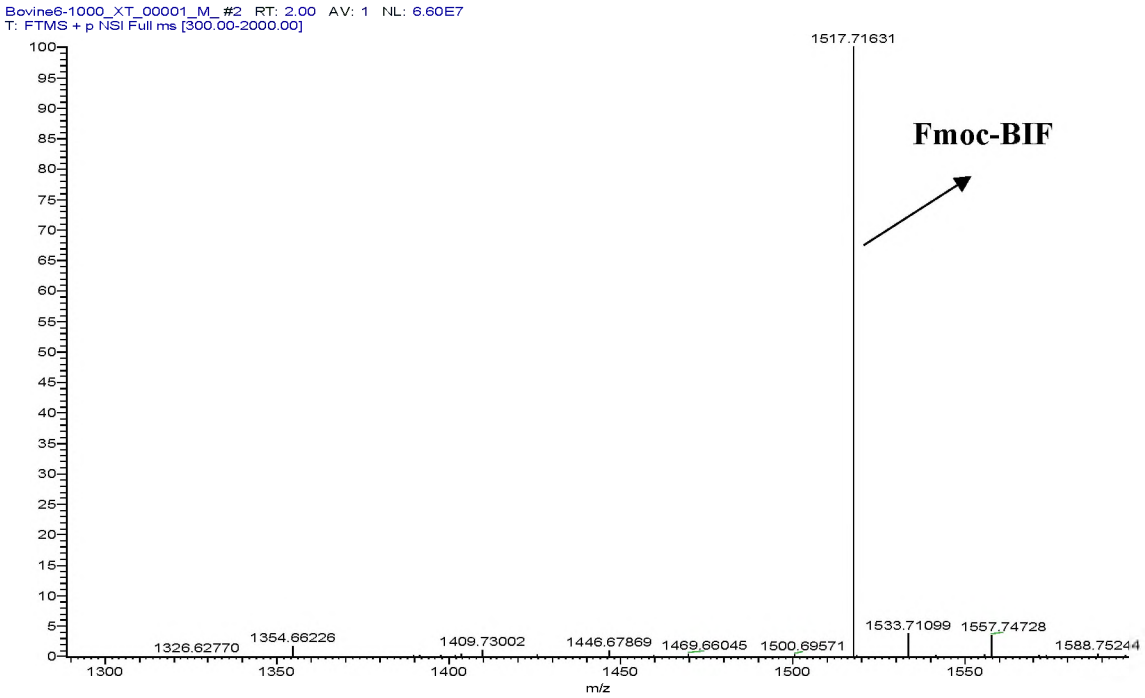

Figure 5. LC-MS graph of fibrils obtained after the experiment for BIF at pH 5, 25°C, 1M urea, 0.02M NaCl showing the relative abundance of Fmoc-BIF

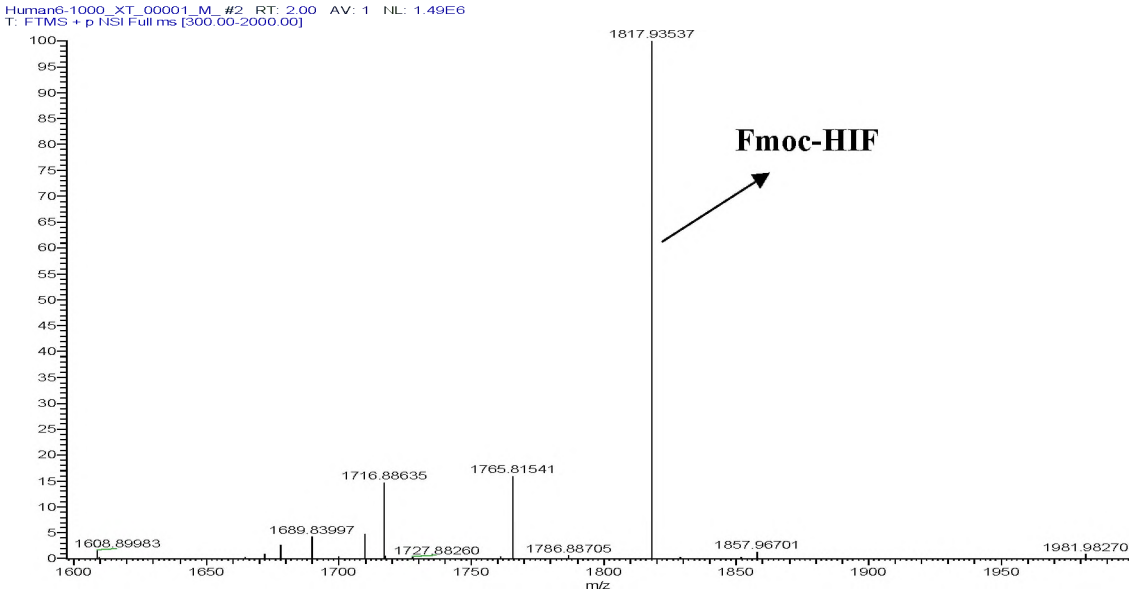

Figure 6. LC-MS graph of fibrils obtained after the experiment for HIF at pH 5, 25°C, 1M urea, 0.02M NaCl showing the relative abundance of Fmoc-HIF

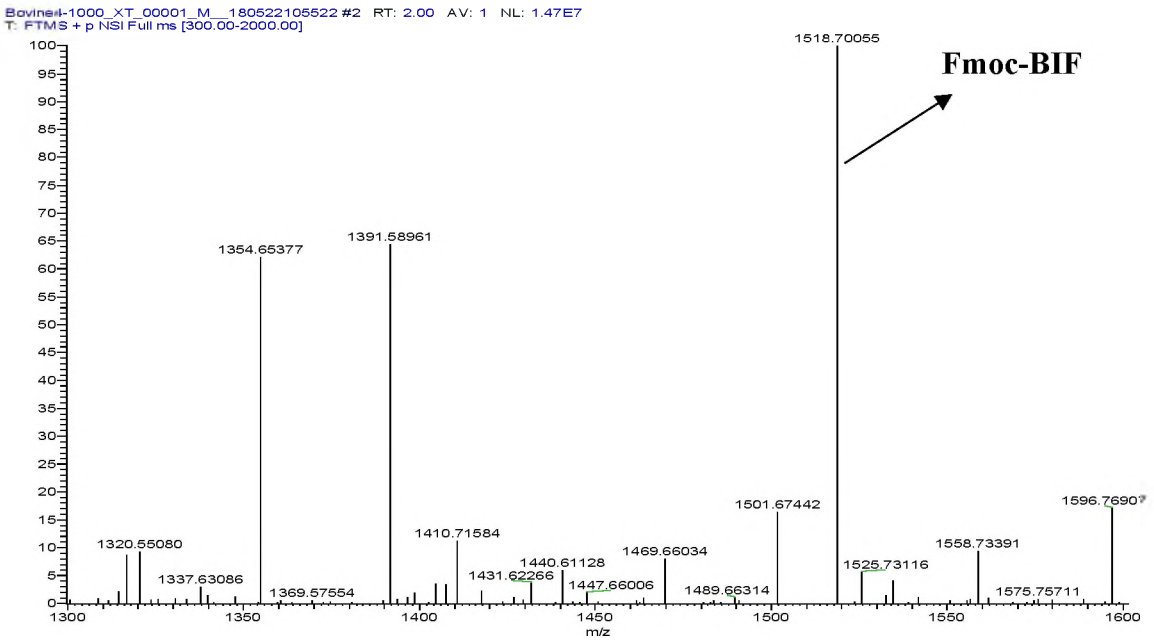

Figure 7. LC-MS graph of fibrils obtained after the experiment for BIF at pH 1.6, 60°C, 1M urea, 0.02M NaCl showing the relative abundance of Fmoc-BIF

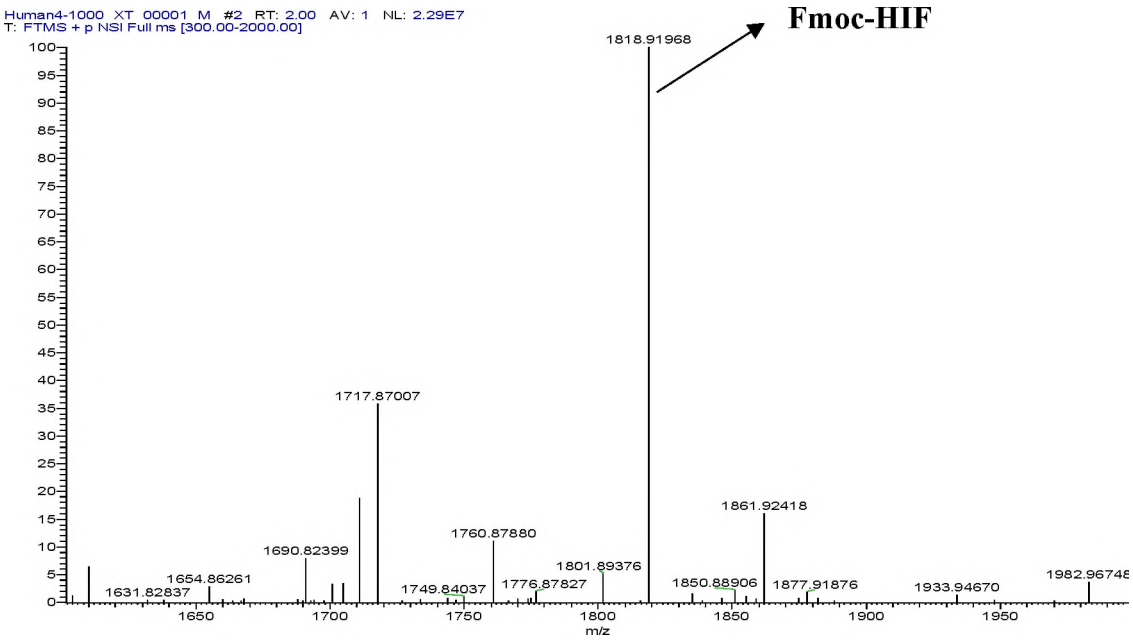

Figure 8. LC-MS graph of fibrils obtained after the experiment for HIF at pH 1.6, 60°C, 1M urea, 0.02M NaCl showing the relative abundance of Fmoc-HIF

While at pH 1.6 and 60°C, fibrils in BIF consist of equal amounts of native and Fmoc attached peptide but HIF fibrils are made of 74% Fmoc HIF, 22% truncated Fmoc HIF and only 4% HIF (Figure 7 and 8). Finally, both BIF and HIF at pH 1.6 and 60 oC shows deamidation of Asn to Asp, but this is not observed at pH 5 and 25 °C. Deamidation is favored by low pH and high temperature. It is not clear if deamidation occurred before, during or after the formation of the aggregates. Deamidation may occur by acid/base catalysis as well as by intramolecular nucleophilic catalysis.

One commonly used mechanism to explain fibril formation consists of three phases: 1) nucleation/lag, 2) elongation/exponential growth and 3) equilibrium/ precipitation [24]. The process begins with a lag phase where there is no significant change in the soluble peptide concentration. Nuclei/oligomer structures are formed by partly denatured peptides [25,26]. The lag phase is followed by an elongation or the exponential growth phase where the size of the fibrils increases by addition of preformed oligomer structures to protofilaments [27,28]. Finally, it reaches an equilibrium/precipitation phase where most of the soluble peptide has been converted into fibrils.

The process of fibril formation was studied using ThT fluorescence assay and the changes in the secondary structure were simultaneously determined using ATR-FTIR. ThT bound to amyloid-like fibrils show a strong fluorescent signal [29]. ThT dye does not bind to either soluble peptides or to amorphous aggregates [30]. ThT-induced spectral changes for various amyloidogenic proteins are qualitatively identical irrespective of the primary structure, this feature is important when comparing two slightly different peptides [31].

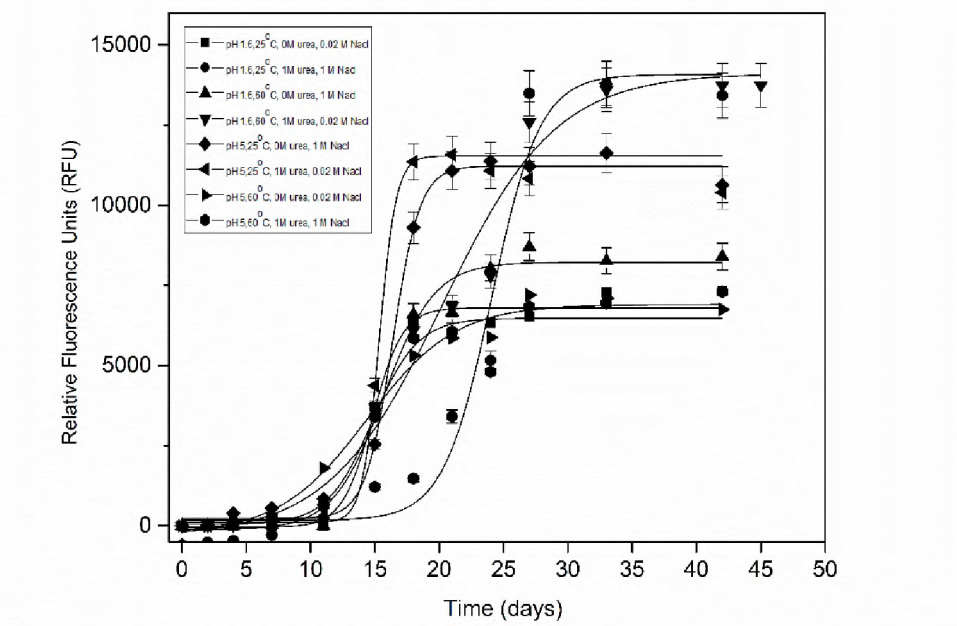

Figure 9. ThT fluorescence plots showing the fibril formation process with time for BIF at all the runs

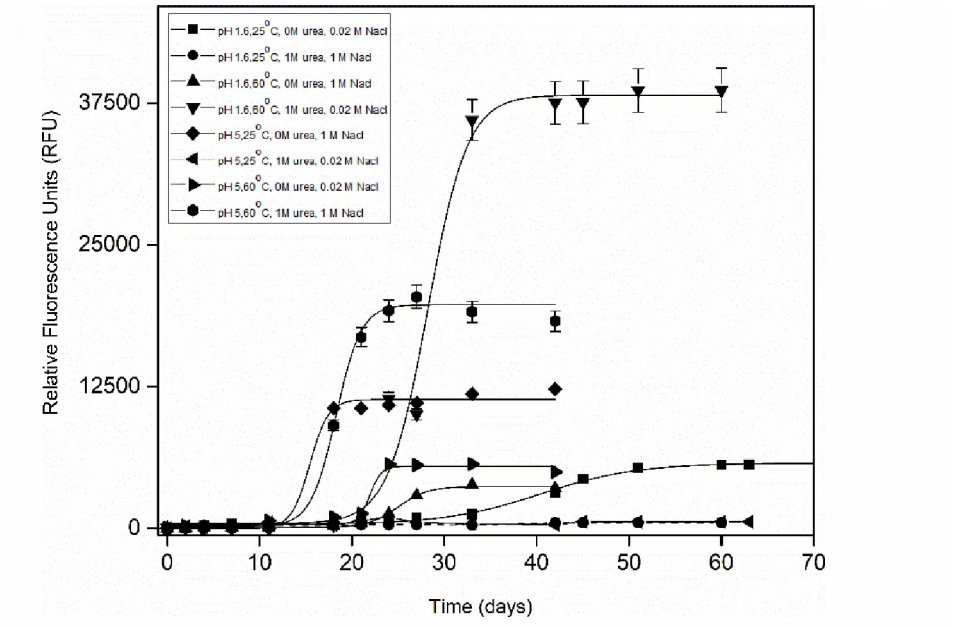

Figure 10. ThT fluorescence plots showing the fibril formation process with time for HIF at all the runs

Figures 9 and 10 show relative fluorescence vs time at all experimental conditions for BIF and HIF respectively. Representative error bars are included in both figures. Most curves show a characteristic sigmoidal shape consisting of three regions: a lag phase, an exponential phase and an equilibrium phase. Therefore, the kinetics of fibril formation can be characterized by a lag time (t_{lag}) and apparent aggregation growth rate constant (K_{app}). The obtained sigmoidal curves for each run were curve fitted using non-linear curve fit in Origin Pro data analysis software to determine t_{lag} and K_{app} by [32]:

$$y = \frac{1}{1+e^{(x-x_0)/dx}} \quad (1)$$

where y is the fluorescence intensity for the normalized data, x is the time in days, x_o is the time at 50% fluorescence intensity and dx is the time constant. From Eq.1 the lag time (t_{lag}) and apparent aggregation growth rate constant (K_{app}) can be calculated using $t_{lag} = (x_0 - 2\,dx)$ and $K_{app} = 1/dx$. The calculated lag time (t_{lag}) and apparent aggregation growth rate constants (K_{app}) for both BIF and HIF are shown in Table 5.

All the runs for BIF formed fibrils while 2 of the 8 runs for HIF (pH 1.6, 25°C, 1M urea, 1M NaCl; pH 5, 25°C, 1M urea, 0.02M NaCl) are still in the lag phase after 8 weeks. This finding suggests that HIF does not form nuclei at room temperature in the presence of urea irrespective of the other conditions. In contrast, HIF at room temperature in the absence of urea (pH 1.6, 25°C, 0M urea, 0.02M NaCl; pH 5, 25°C, 0M urea, 1M NaCl) form fibrils.

Table 5. Lag times (t_{lag}) and apparent aggregation growth rate constants (K_{app}) calculated for both BIF and HIF

Run #	Bovine		Human	
	t_{lag} (days)	K_{app} (day^{-1})	t_{lag} (days)	K_{app} (day^{-1})
1	12 ± 1	0.85 ± 0.37	30 ± 2	0.18 ± 0.02
2	16 ± 1	0.53 ± 0.16	-	-
3	12 ± 1	0.51 ± 0.10	22 ± 1	0.62 ± 0.07
4	11 ± 2	0.22 ± 0.05	24 ± 2	0.47 ± 0.16
5	14 ± 1	0.94 ± 0.13	13 ± 2	0.99 ± 0.63
6	14 ± 2	1.51 ± 0.15	-	-
7	7 ± 1	0.29 ± 0.04	20 ± 1	0.70 ± 0.17
8	11 ± 2	0.57 ± 0.22	15 ± 1	0.73 ± 0.18

BIF has shorter lag times than HIF under all incubation conditions. The lag times are shorter at higher temperatures under all conditions for both BIF and HIF. This dependence on temperature was also observed in native bovine and human insulin [17]. Though the lag times are shorter for BIF than HIF (at the same conditions), the rate of aggregation is faster for HIF than BIF except for run #1 (Table 1). The lag time is the shortest for BIF at pH 5, 60°C, no urea, 0.02M NaCl (7 ± 1 days) and it is the longest for BIF at pH 1.6, 25°C, 1M urea, 1M NaCl (16 ± 1 days).

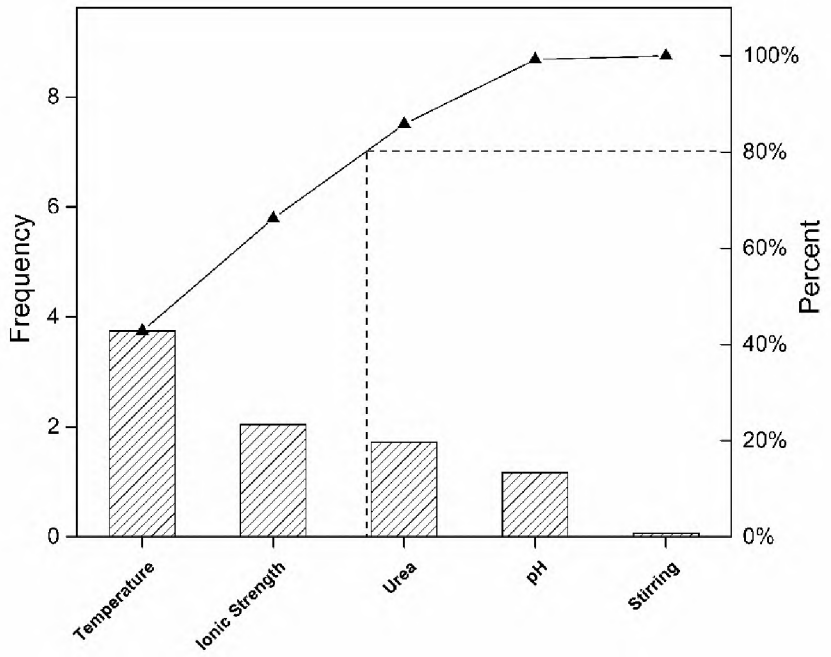

Figure 11. Pareto charts showing the effect of various factors based on T_{lag} as an outcome from the experimental design FF0508 for BIF

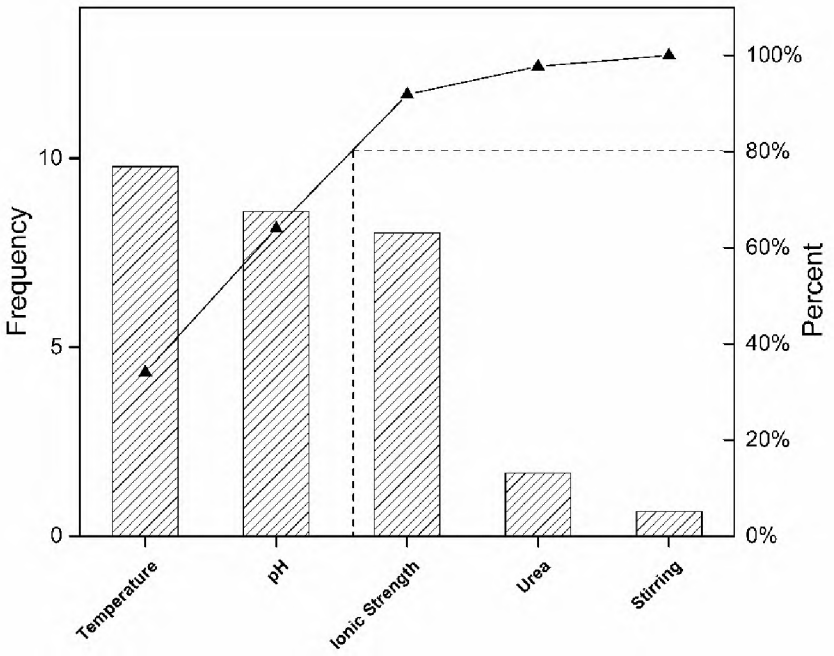

Figure 12. Pareto charts showing the effect of various factors based on T_{lag} as an outcome from the experimental design FF0508 for HIF

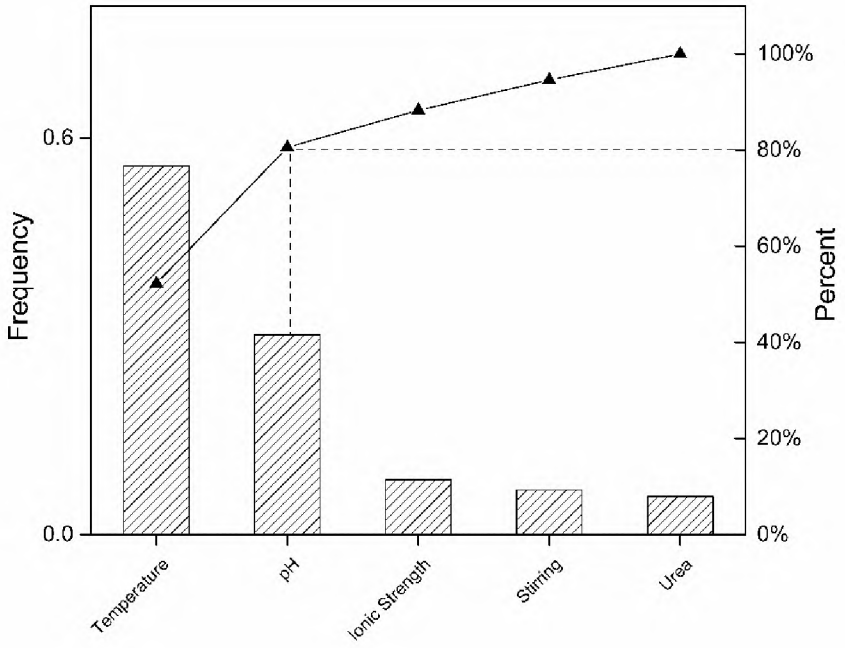

Figure 13. Pareto charts showing the effect of various factors based on K_{app} as an outcome from the experimental design FF0508 for BIF

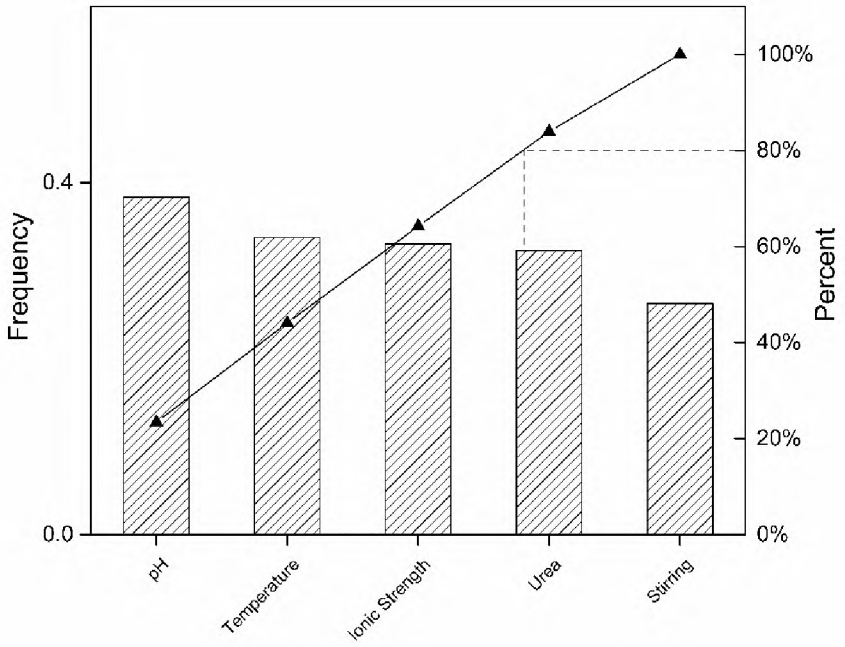

Figure 14. Pareto charts showing the effect of various factors based on K_{app} as an outcome from the experimental design FF0508 for HIF

For HIF, lag time is the shortest at pH 5, 25°C, 0M urea, 1M NaCl (13 ± 2 days) while it is the longest at pH 1.6, 25°C, 0M urea, 0.02M NaCl (30 ± 2 days). At 1M urea and 0.02M NaCl, samples at the lowest pH and highest temperature (BIF and HIF at pH 1.6, 60°C) have a shorter lag time than samples at the highest pH and lowest temperature (BIF and HIF at pH 5, 25°C).

Pareto charts using t_{lag} or K_{app} as an outcome are shown in Figures 11,12 and 13,14 respectively. The important factors using each outcome were identified using the Pareto principle [33]. The important factors during lag phase are temperature, ionic strength and urea for BIF while for HIF temperature, pH and ionic strength are significant. Although all the factors seem to be important during the elongation phase for HIF, temperature and pH are the most important ones which is also true for BIF. The pH is of importance during elongation phase as it affects the interactions among the side chains of amino acids leading to the formation of fibrils. The presence of gas/liquid interfaces has a negligible effect on fibrillation.

Exposure of hydrophobic surfaces to water plays an important role in the formation of fibrils [34]. Hydrophobicity of each amino acid residue can be assigned from the native structure according to the hydrophobicity scale of Kyte and Doolittle in the UCSF Chimera package [35,36]. The native peptide structures were generated using a web server PEP-FOLD 3 [37]. The structures shown in Figures 15 and 16 correspond to BIF and HIF respectively. Models were visualized using the UCSF Chimera package [36]. The residues in each peptide are shown in various colors according to their hydrophobicity. Colors range from blue for the most hydrophilic to white (neutral) to orange for the most hydrophobic.

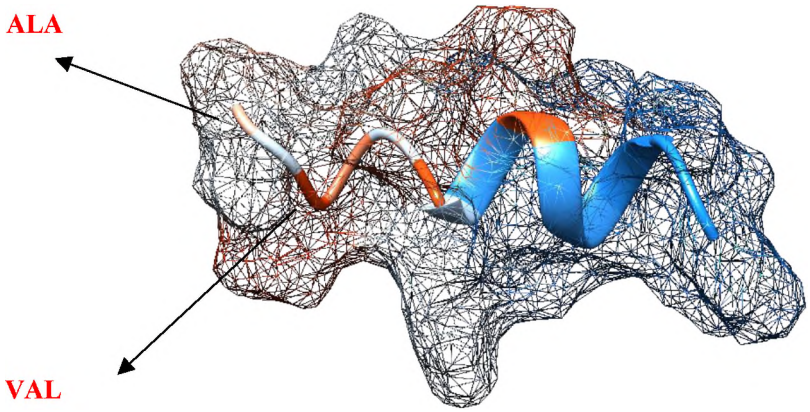

Figure 15. BIF fragment visualized using UCSF chimera package with variations in colors according to their hydrophobicity

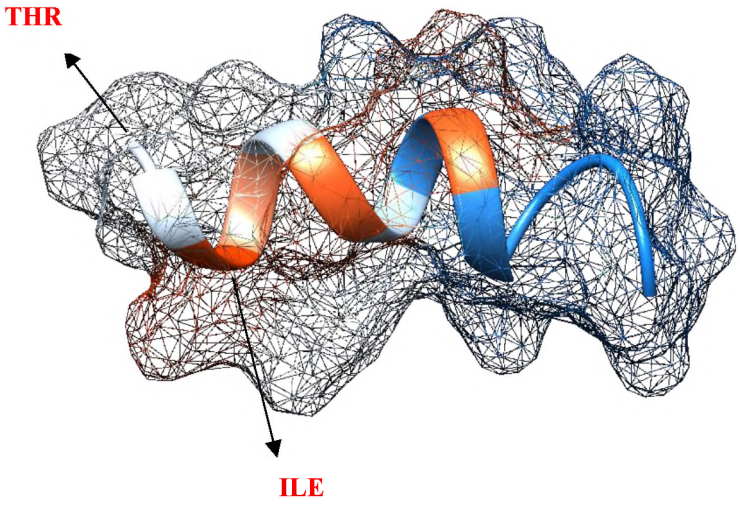

Figure 16. HIF fragment visualized using UCSF chimera package with variations in colors according to their hydrophobicity

BIF has alanine and valine which are replaced by threonine and isoleucine in HIF at their respective positions. Comparing the hydrophobicity indexes, alanine (47), valine (79) and isoleucine (99) are very hydrophobic amino acids at both acidic and basic pH compared to threonine which is a neutral amino acid (13) [38]. The presence of an extra hydrophobic residue for BIF over HIF could be the reason for shorter lag times observed with BIF. However, BIF shows a slower growth rate than HIF. This suggests that the nuclei of both peptides are structurally different.

The formation of intermolecular β-sheet secondary structure is one criterion used to label an aggregate as an "amyloid fibril" [22]. The changes in the secondary structure of peptides can be followed by FTIR [39]. Moreover, ATR-FTIR plays a prominent role in differentiating the oligomers (anti-parallel β-sheets) from fibrils (parallel β-sheets) [40]. Therefore, IR spectroscopy is an alternative method to complex techniques like X-ray diffraction [41]. Second derivative analysis of the spectra allows the determination of protein secondary structure [42]. The number of peaks from the deconvolved spectra should be chosen carefully. It is necessary to confirm that the maximum frequency of the curve-fit peaks correspond to the maxima evident in the raw data to ensure that the peak is assigned to its specific secondary structure [43]. The precision of the calculated peak areas from the second derivative spectra largely depends on the baseline chosen while fitting [39]. Origin Pro peak fit analysis has various baseline modes for this purpose.

The relevant region in the spectra is the Amide I region (1600-1700 cm^{-1}), which is caused by the carbonyl group stretching vibrations from the peptide backbone [44]. Each structural conformation has a unique C=O vibration dependent on its hydrogen bonding pattern [45]. One major limitation with the spectra is that the Amide I peak has a maximum

at ~1650 cm^{-1} that coincides with the water maximum absorption at ~1643 cm^{-1}. The water background needs to be carefully subtracted from the spectra. The use of an ATR attachment avoids this problem and allows data collection with minimal use of sample [46]. A flat baseline obtained between 2200 to 1800 cm^{-1} is often considered a standard to confirm proper water subtraction [47]. Absorption of the side-chains also cause spectral deviation during the analysis. Carbonyl group vibrations for asparagine and glutamine occur at around 1678 cm^{-1} and 1670 cm^{-1} which can affect the spectra [48]. Urea also has an absorption peak in the amide I region which can be deceptive [49]. Hence during the background subtraction for the samples in the presence of urea, 20 µl of 1M urea was added to the blank to subtract the undesired peaks.

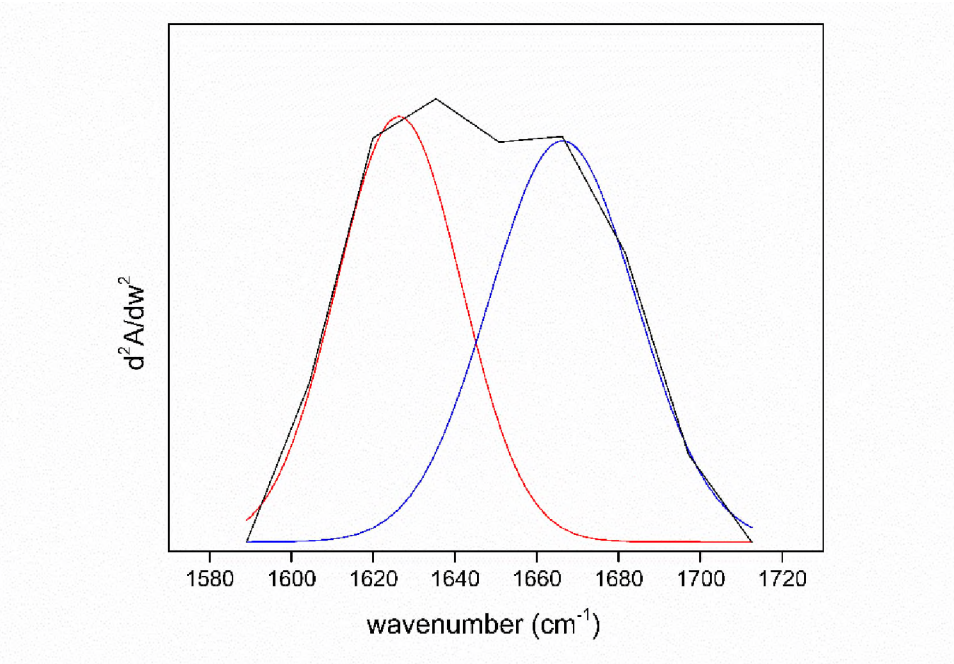

Figure 17. Second Derivative spectra for native BIF and its curve fitted peaks

FTIR analysis of the native peptide fragments shows an α-helix content of 52.6 % and 53.9 % for BIF and HIF respectively while the rest of the structure are random coils (Figures 17 and 18). The calculated secondary structures for both peptides predicted by PEPFOLD3 are 42.9 % and 64.3 % α-helix for both BIF and HIF (the rest being random coils). The secondary structure is calculated using DSSP, a hydrogen bond estimation-based algorithm by the secondary structure server [50]. PSIPRED and Porter 4.0 protein servers also confirmed that both the peptides consist of α-helices and random coils [51]. Comparing the experimental and theoretical values of α-helix content, BIF has ~82 % Q3 accuracy while HIF has ~81% Q3 accuracy, which are acceptable [52].

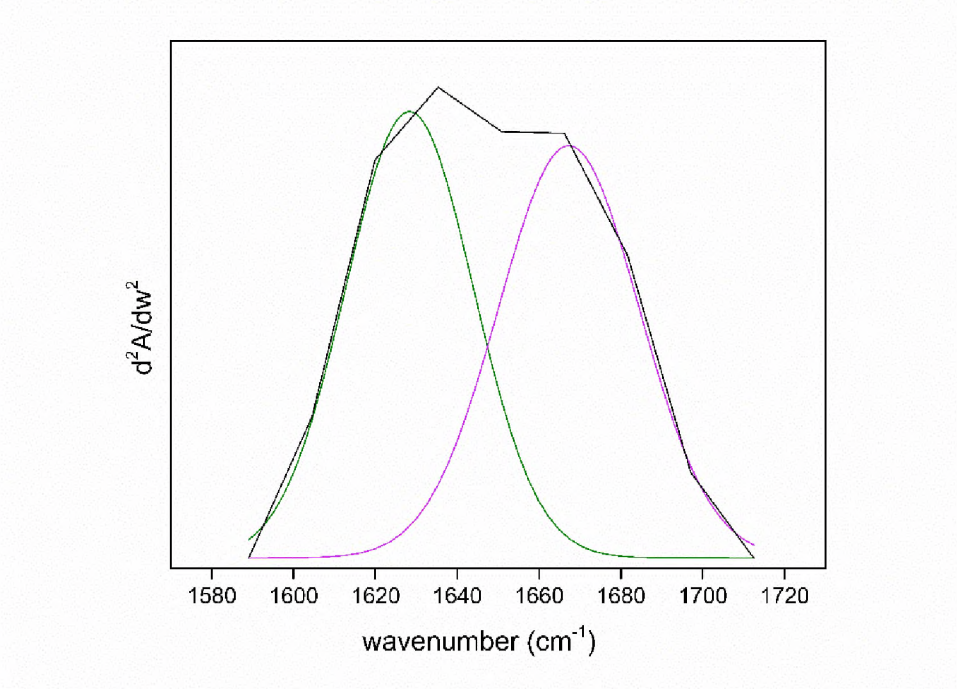

Figure 18. Second Derivative spectra for native BIF and its curve fitted peaks

Figure 19 shows the FTIR spectra of BIF at pH 1.6, 60°C, 1M urea and 0.02M NaCl collected during weeks 1 and 6. At time zero the amide I peak is centered at 1650 cm^{-1} but after 42 days the peak shifted to 1681 cm^{-1}. This net band shift corresponds to an α-helix to β-sheet transition with time [53]. By curve-fitting the deconvoluted spectra the amount of each secondary structure was determined as shown in Figure 20, which correspond to BIF at the above conditions during week-1. The integration of each curve-fitted peak yields the total secondary structure content. At the end of week-1 (Figure 20), three peaks are found around 1635, 1663 and 1689 cm^{-1}. The bands around 1635 and 1689 cm^{-1} correspond to β-sheets/random coil and turns/anti-parallel β-sheets respectively whereas the peak around 1663 cm^{-1} can be assigned to α-helix.

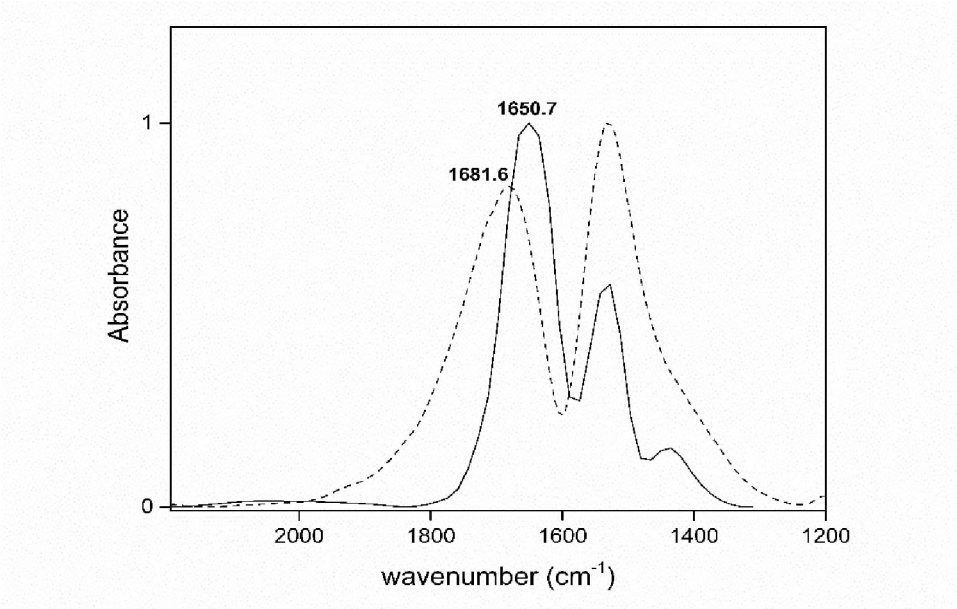

Figure 19. FTIR spectra for BIF at pH 1.6, 60°C, 1M urea, 0.02M NaCl collected at t = 0 days (Solid line) and t = 42 days (dashes) showing the shift in the amide I region

At the end of week-6, four peaks around 1620, 1651, 1674 and 1681 cm^{-1} are found. The peak at 1635 cm^{-1} during week-1 vanishes with time and the presence of a peak around 1620 cm^{-1} during week-6 corresponds to an increase in β-sheet content. Researchers previously reported that the fibrils show a new β band at ~1620 cm^{-1} as opposed to the native FTIR β band at ~1635 cm^{-1} [54]. The peaks at 1620 and 1651 cm^{-1} can be assigned to extended β-sheets and α-helix whereas the other two peaks correspond to β-turns/ anti-parallel β-sheets.

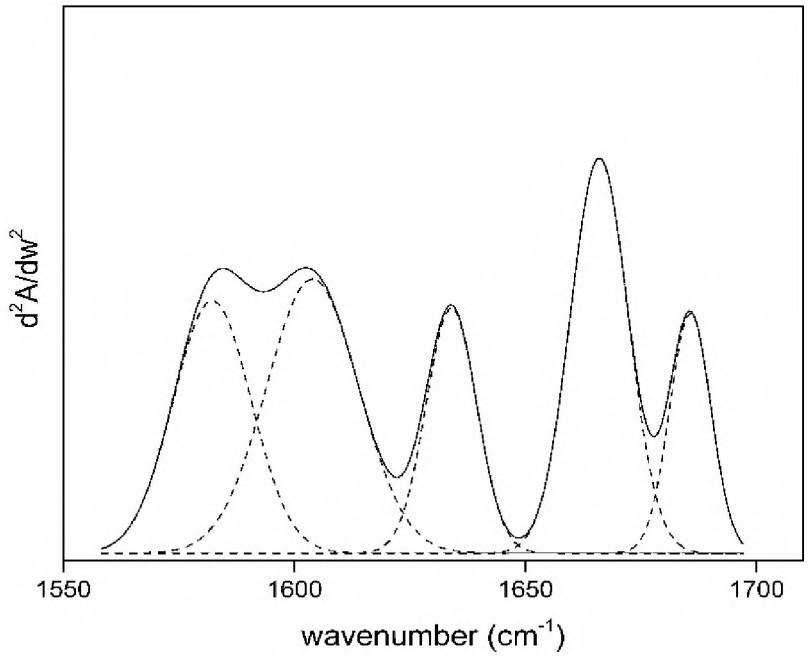

Figure 20. Second derivative spectra for BIF at pH 1.6, 60°C, 1M urea, 0.02M NaCl collected at the end of week 1 and its curve fitted peaks

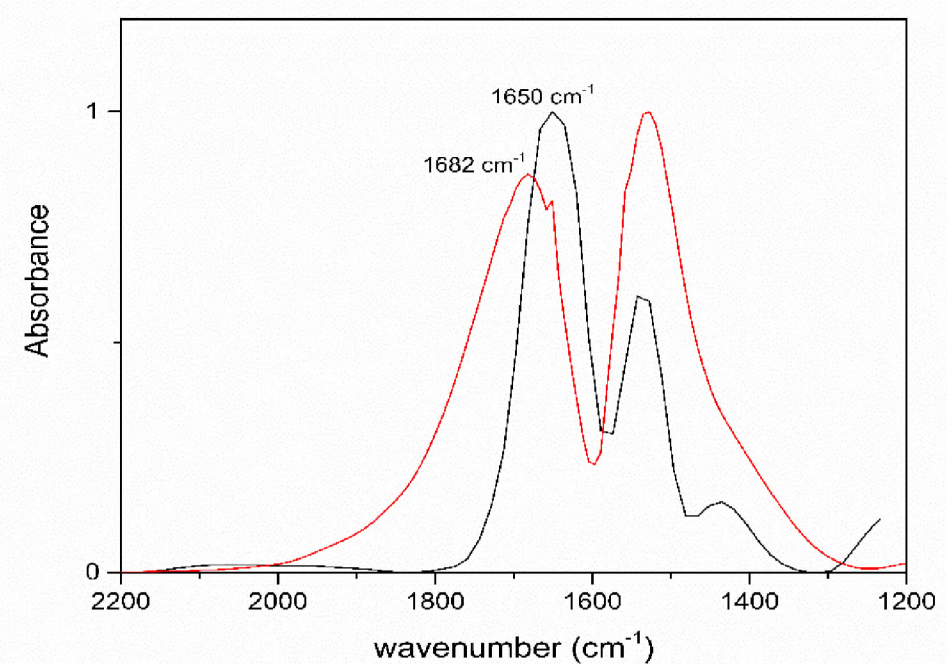

Figure 21. FTIR spectra for HIF at pH 1.6, 60°C, 1M urea, 0.02M NaCl collected at t = 0 days (black) and t = 42 days (red) showing the shift in the amide I region

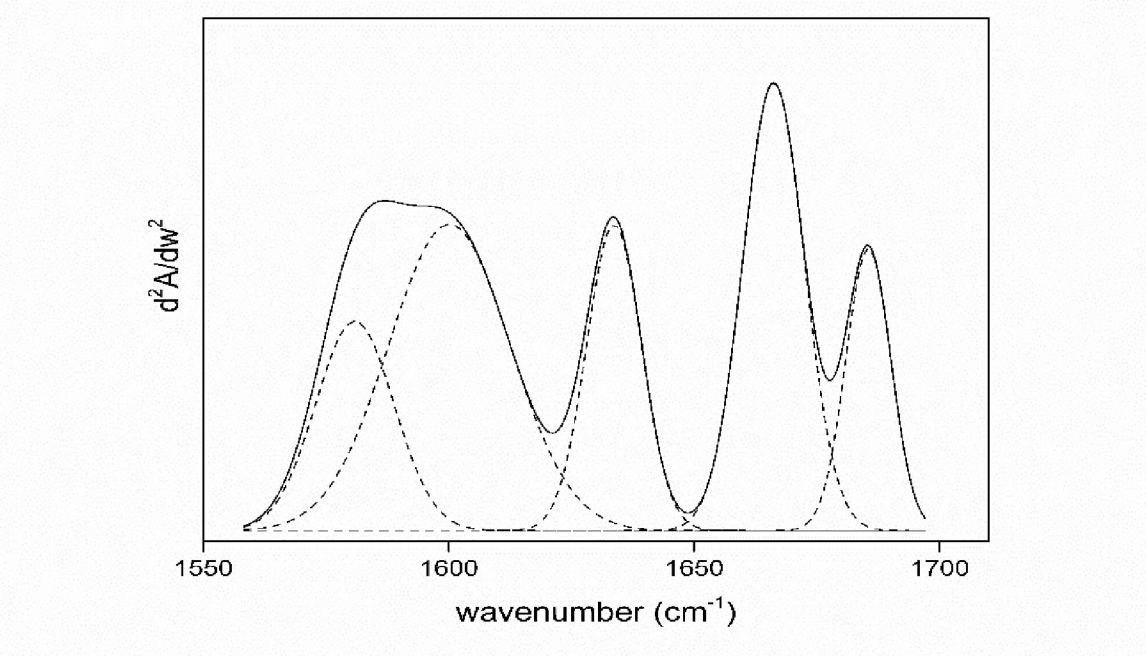

Figure 22. Second derivative spectra for HIF at pH 1.6, 60°C, 1M urea, 0.02M NaCl collected at the end of week 1 and its curve fitted peaks

Similar analysis was also conducted on HIF and the secondary structure content was determined (Figures 21 and 22). HIF peaks were located around same wavenumbers as that of BIF (Table 6).

Table 6. FTIR deconvoluted peaks corresponding to HIF at pH 1.6, 60 °C, 1M urea, 0.02M NaCl

Wavenumber (cm^{-1})	Week-1	Week-6
HIF	1633	1619
	1667	1649
	1686	1696

Figures 23 and 24 show plots for the change in the intermolecular β-sheet content with time for both BIF and HIF respectively. All the runs show an increase in the percentage of β-sheets with time. All BIF samples have a higher percentage of β-sheets than HIF when incubated at the same conditions. The highest percentage of β-sheets was found in BIF (34%) and HIF (31%) at pH 1.6, 60°C, 1M urea, 0.02M NaCl. HIF at 60 °C for all runs has a higher percentage when compared to the one at 25°C. As it is evident from the plots, most of the increase in β-sheets occurred in the first two weeks of the experiment for BIF and in the first three weeks for HIF. A comparison with ThT data shows that the β-sheet increase occurs mostly during the lag times (12 days for BIF and 21 days for HIF). Thus, the progression of lag phase to the elongation phase in both the fragments can be assigned to the increase in the number of β-sheets and their extension to form fibril.

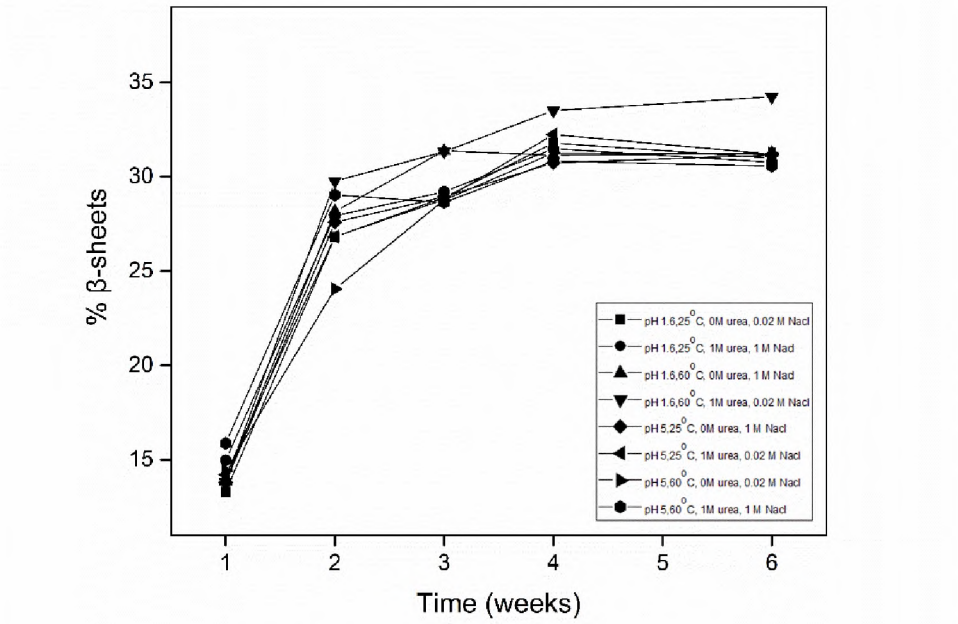

Figure 23. Evolution of the intermolecular β-sheets (%) with time (weeks) for BIF at all the runs

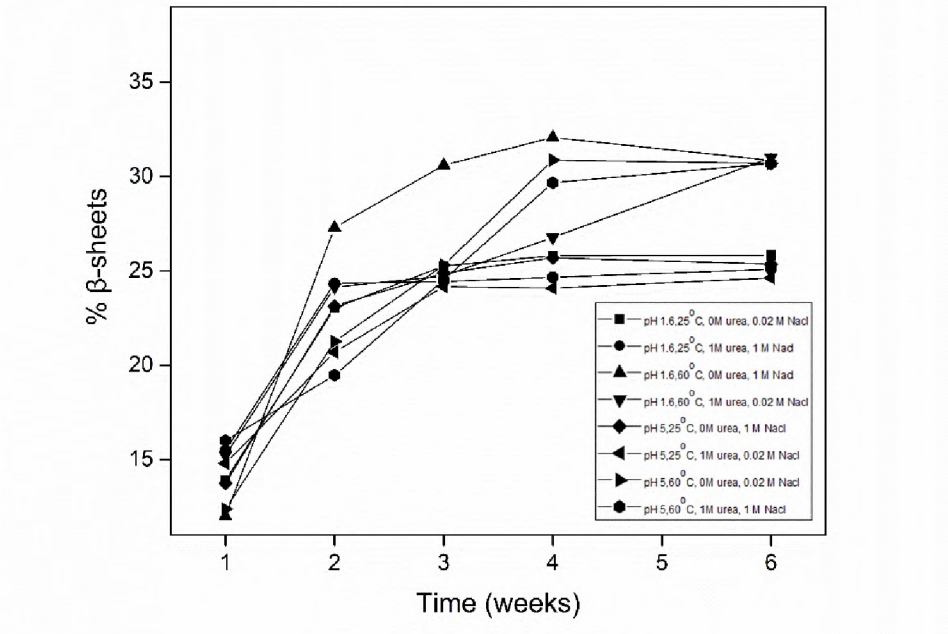

Figure 24. Evolution of the intermolecular β-sheets (%) with time (weeks) for HIF at all the runs

The samples HIF at pH 1.6, 25°C, 1M urea, 1M NaCl and at pH 5, 25°C, 1M urea, 0.02M NaCl are still in the lag phase at the end of the experiment, they have a β-sheet content of 25.1 % and 24.6 % respectively. There is no specific criterion as to how much of β-sheet percentage qualifies to be an amyloid fibril. According to ThT data, HIF at pH 5, 25°C, 0M urea, 1M NaCl forms fibrils whereas HIF at pH 1.6, 25°C, 1M urea, 1M NaCl and at pH 5, 25°C, 1M urea, 0.02M NaCl do not; despite the fact that both peptides reach an almost identical percentage of β-sheets. While there is an increase in the β-sheet content with time for the rest of the runs, these two runs have a constant β-sheet content. These results are consistent with the ThT data (Figure 10).

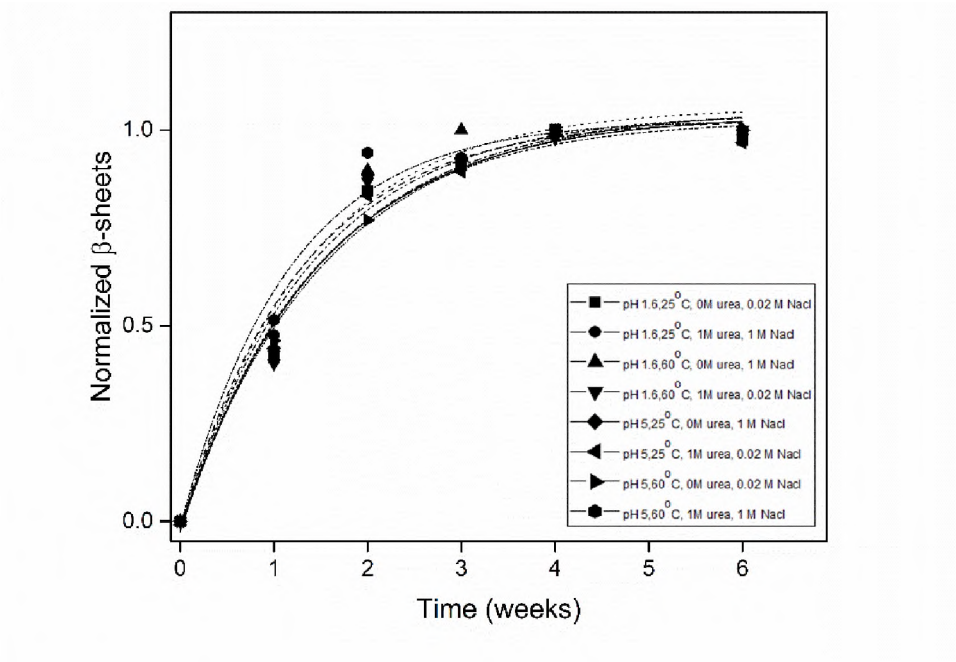

Figure 25. Curve fitted plots for the evolution of the intermolecular β-sheets (%) with time (weeks) for BIF at all the runs

The data shown in Figures 23 and 24 was normalized and fitted with,

$$y = 1 - (K_\beta)^t \quad (2)$$

where y is the normalized β-sheet percentage, t is the time in weeks and K_β is the rate of increase in β-sheets. Figures 25 and 26 show the curve fitted β-sheet plots for BIF and HIF respectively. All the fitted curves have an R-squared value of at least 97%. The K_β values calculated for BIF and HIF are shown in Table 7. K_β values for BIF are similar for all the runs but the K_β values for HIF are larger at 60°C than at 25°C.

Table 7. Rate of increase in β-sheets (K_β) calculated for both BIF and HIF

Run #	Bovine K_β (week^{-1})	Human K_β (week^{-1})
1	0.50 ± 0.08	0.43 ± 0.05
2	0.46 ± 0.07	0.34 ± 0.06
3	0.47 ± 0.09	0.51 ± 0.11
4	0.51 ± 0.09	0.49 ± 0.06
5	0.48 ± 0.08	0.42 ± 0.06
6	0.49 ± 0.07	0.40 ± 0.02
7	0.52 ± 0.04	0.60 ± 0.05
8	0.42 ± 0.08	0.58 ± 0.07

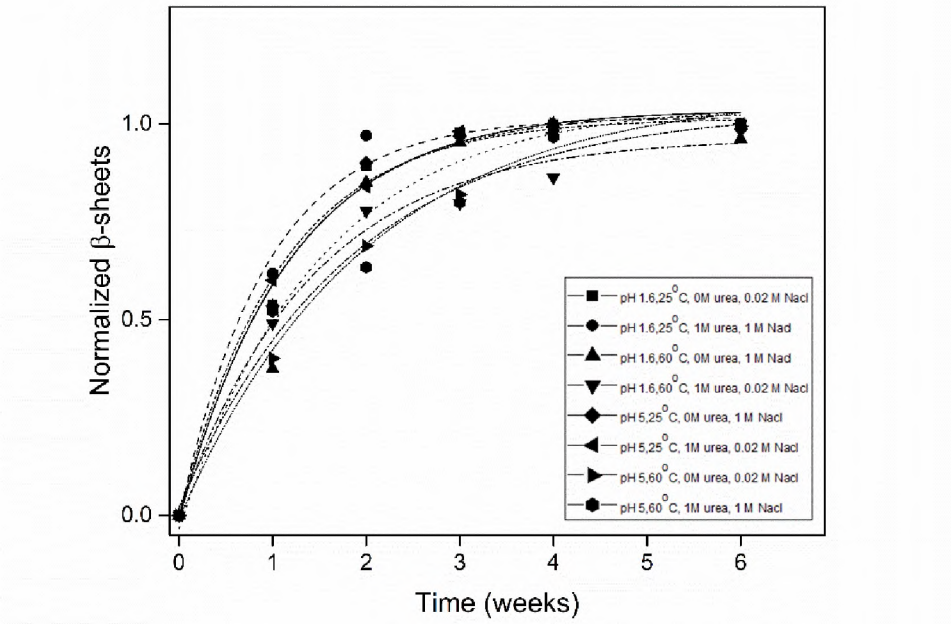

Figure 26. Curve fitted plots for the evolution of the intermolecular β-sheets (%) with time (weeks) for HIF at all the runs

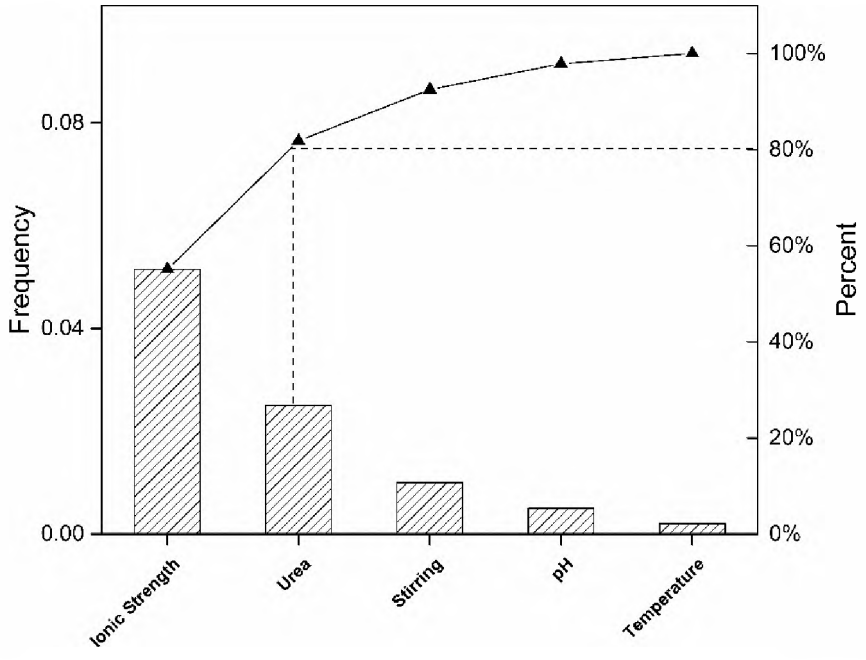

Figure 27. Pareto charts showing the effect of various factors based on K_β as an outcome from the experimental design FF0508 for BIF

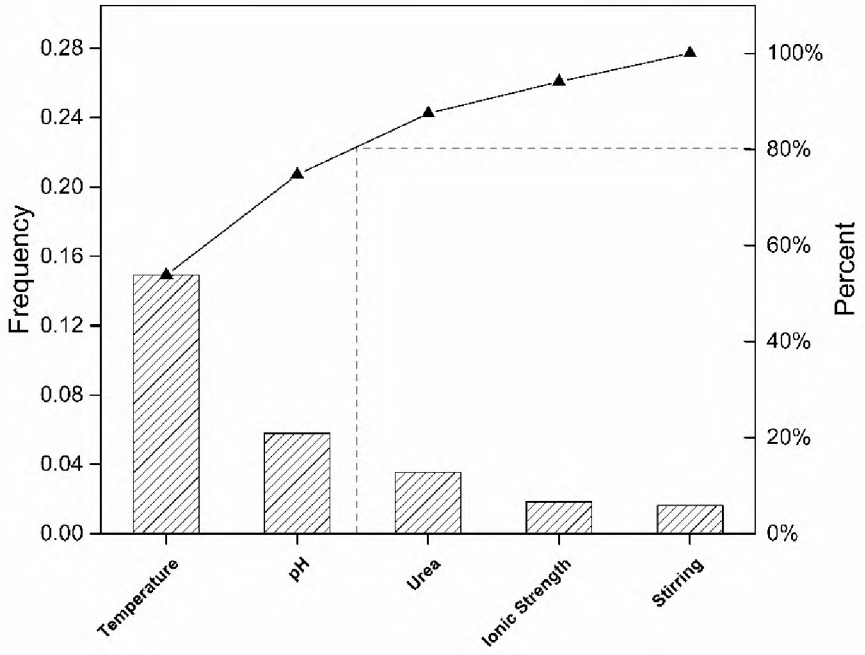

Figure 28. Pareto charts showing the effect of various factors based on K_β as an outcome from the experimental design FF0508 for HIF

At 25°C the rate of increase in β-sheets was higher in BIF than HIF while at 60°C the rates were similar. The K_β value decreased with an increase in both urea and NaCl concentration irrespective of other conditions. Using K_β as an outcome Pareto analysis was conducted to determine the important factors (Figures 27 and 28). Temperature and pH are the most significant factors for HIF which is also true for the lag and elongation phases from ThT data (Figures 12, 14). Ionic strength and urea are the most important factors for BIF. This is also true for BIF during lag phase but during the elongation phase, temperature and pH are more important.

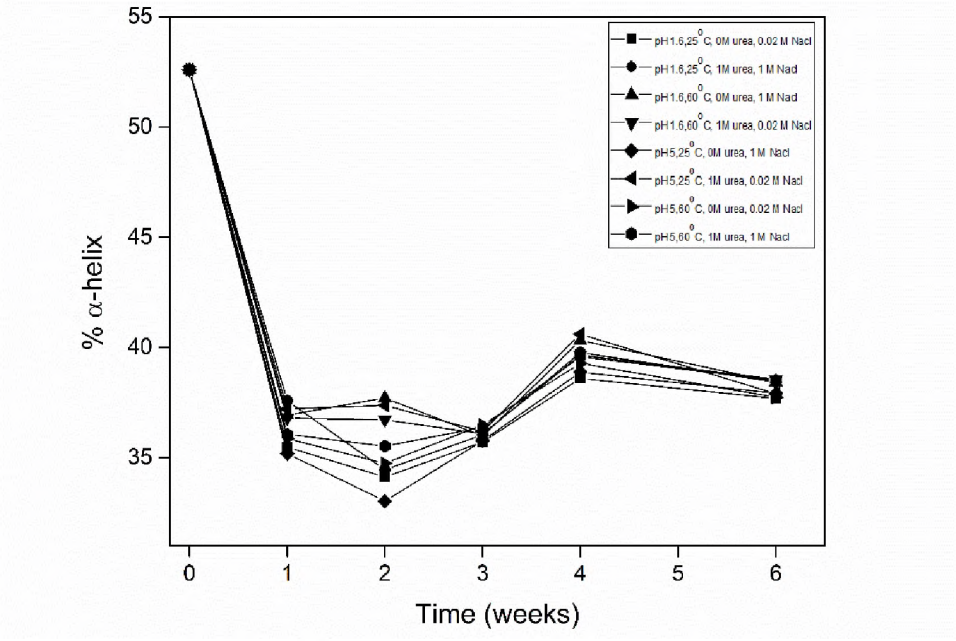

Figure 29. α-Helix percentage change for BIF at all the runs

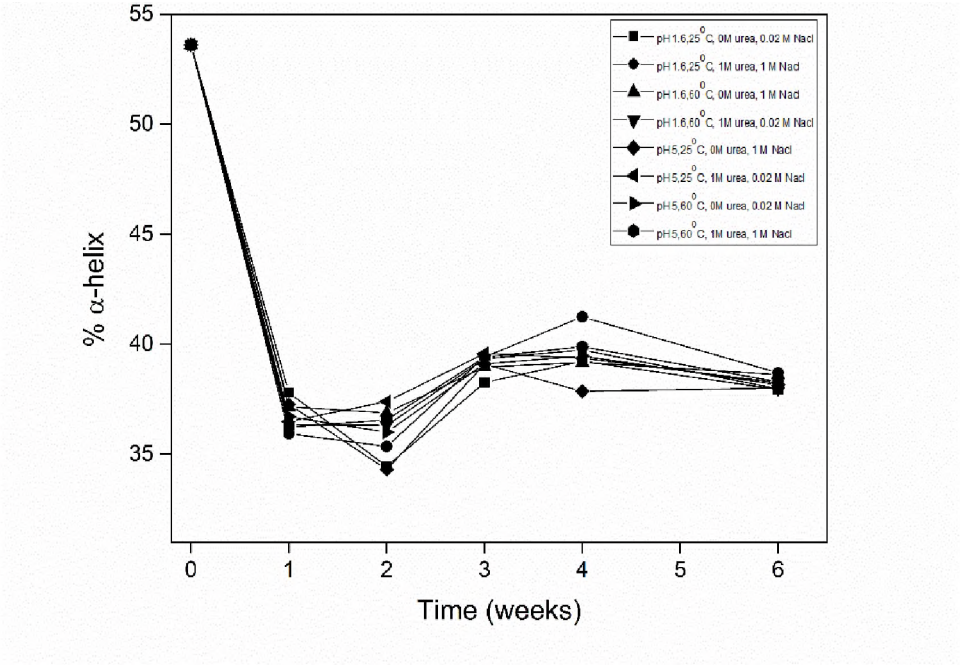

Figure 30. α-Helix percentage change for HIF at all the runs

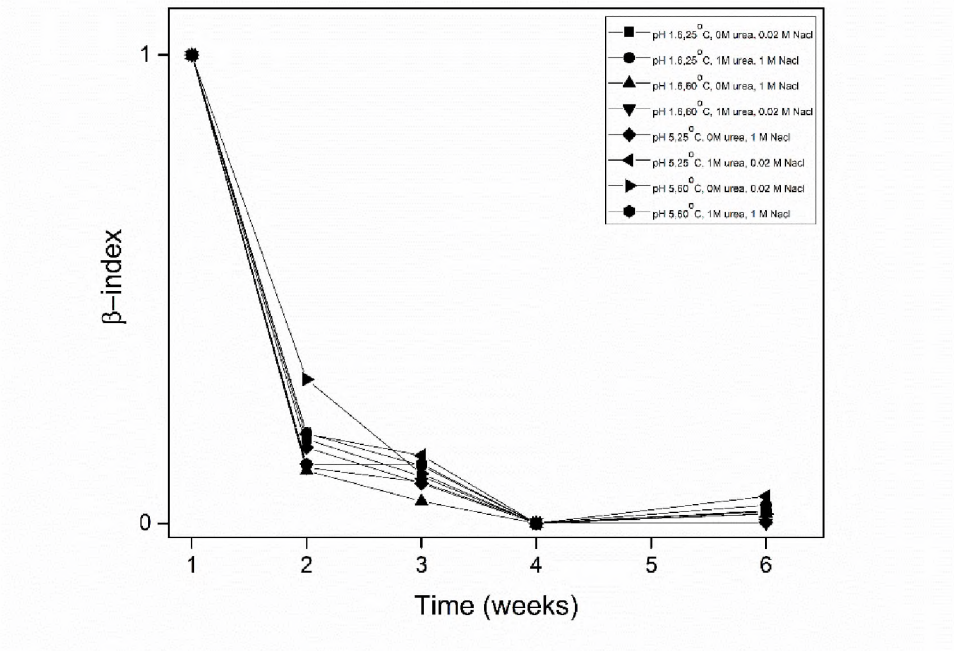

Figure 31. β-index plotted as a function of time at all the conditions for BIF

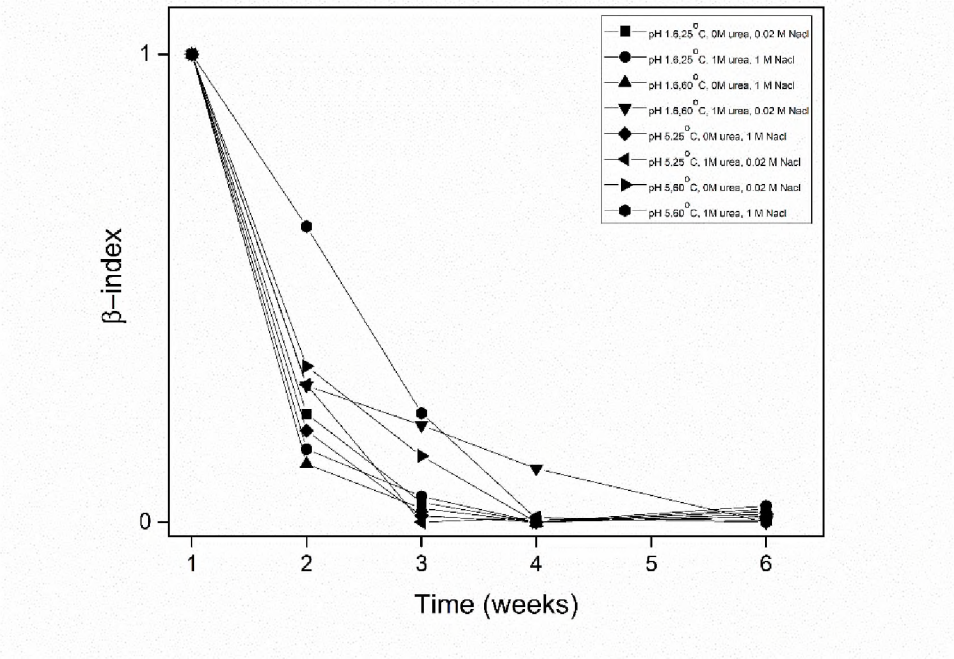

Figure 32. β-index plotted as a function of time at all the conditions for HIF

Figures 29 and 30 show a plot of percentage change in the α-helix content with time for BIF and HIF respectively for all runs. Both fragments show a decrease in α-helices with time. Though there is consistent decrease in the α-helix content in both fragments during the lag phase, a slight increase in α-helices can be seen during the transition from lag phase to elongation phase. Formation of α-helical rich intermediates was also reported previously in the fibrillation of α-Synuclein [55,56].

FTIR analysis was also used to distinguish oligomers from fibrils. Oligomers are rich in anti-parallel β-sheets (1695-1680 cm^{-1}) while fibrils are rich in parallel β-sheets (1635-1620 cm^{-1}) [40]. We calculated the β-index i.e., the ratio of (% of turns + anti−parallel β−sheets) to % of parallel β−sheets using a similar protocol as previously mentioned [57]. The β-index as a function of time was plotted for both the fragments as shown in Figures 31 and 32 respectively. Its value decreases with increasing time, which shows that samples rich in anti-parallel β-sheets (oligomers) convert to parallel β-sheet rich fibrils. That transition from anti-parallel to parallel β-sheets confirms that oligomers are the base unit of fibrillation.

The presence of fibrils was further confirmed using optical microscope. A recent NMR study shows that Congo red binding is specific to fibrils [58]. The combination of ThT and Congo red assays minimizes confounding because they are independent assays [59]. Micrographs shown in Figure 33 contain irregular proteinaceous deposits when stained with Congo red. These micrographs demonstrate the amyloid properties for BIF and HIF at pH 1.6, 60°C, 1M urea, 0.02M salt (Figure 33 A, B) and pH 5, 60°C, 0M urea, 0.02M salt (Figure 33 C, D) respectively. At pH 1.6, BIF forms amyloid plaques (Figure 33A) whereas HIF forms distinct annular amyloid aggregates (Figure 33B) [60]. The

maximum fluorescence intensity represents the maximum aggregation growth. This is higher in HIF compared to BIF at pH 1.6 which explains the morphological differences. Similarly, it is the same at pH 5 for both fragments showing aggregates of similar morphology.

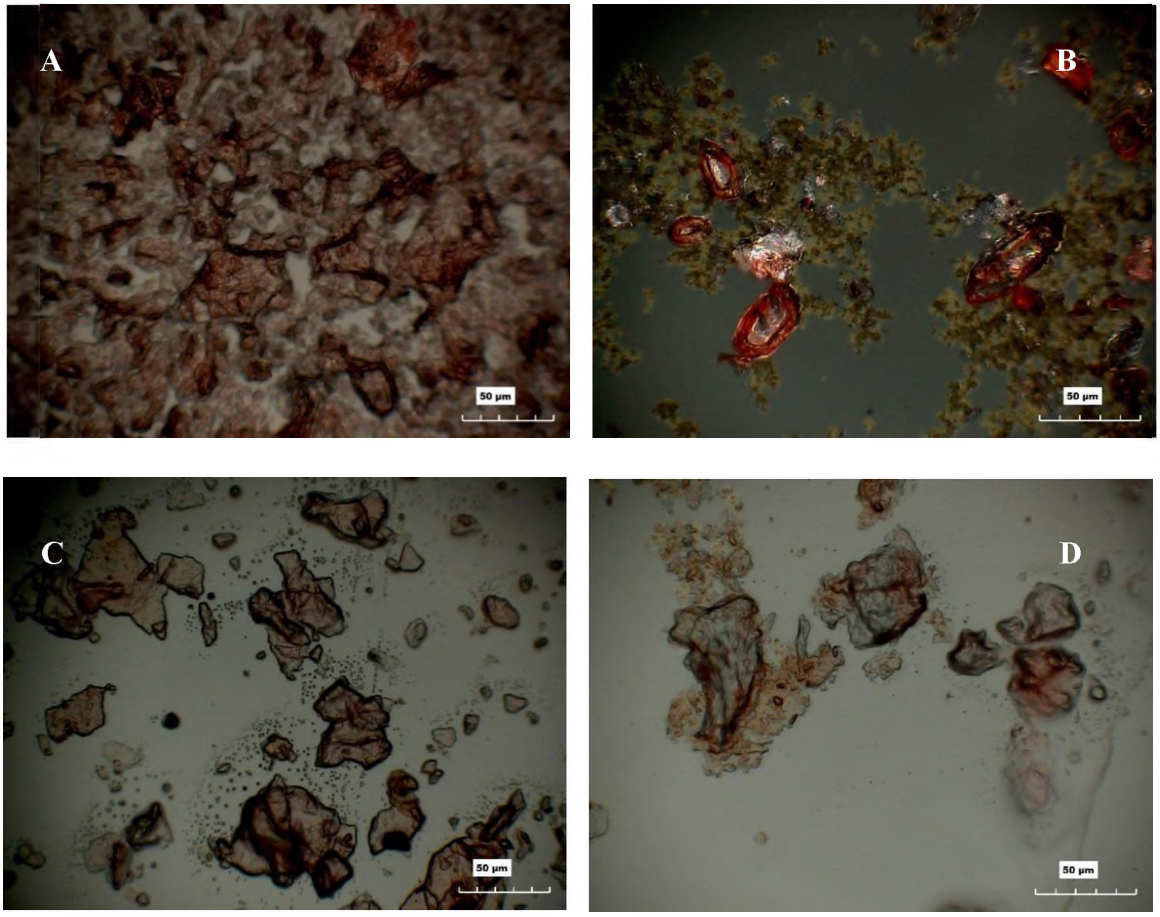

Figure 33. Micrographs showing the presence of fibrils for A) BIF at pH 1.6, 60°C, 1M urea, 0.02M NaCl B) HIF at pH 1.6, 60°C, 1M urea, 0.02M NaCl C) BIF at pH 5, 60°C, 0M urea, 0.02M NaCl D) HIF at pH 5, 60°C, 0M urea, 0.02M NaCl

4. CONCLUSIONS

Two peptide fragments of insulin A-chain with subtle differences in their primary structure were synthesized and their fibrillation propensity was studied using fluorescence and infrared spectroscopy. We found that the bovine insulin fragments have a shorter lag time in comparison with the human insulin ones but, for most samples, the exponential phase rate was faster for HIF than for BIF. The presence of an Fmoc attached peptide has an impact on the fibrillation kinetics which is dependent on pH and temperature. An increase in β-sheets content with time was observed in both fragments. The increase in β-sheets was preceded by an initial decrease in α-helices followed by an intermediate increase during the transition from the lag phase to elongation phase. The conversion of oligomer structures rich in anti-parallel β-sheets to fibrils rich in parallel β-sheets is a characteristic of amyloid. The observed α-helix to β-sheet transition as well as the micrographs obtained using Congo red staining confirmed the formation of ring-like oligomer structures rich in anti-parallel β-sheets, which then form fibrils rich in parallel β-sheets.

ACKNOWLEDGEMENTS

The work done on this project was funded in part by theNational Science Foundation (CBET 0933468).

REFERENCES

(1) Cromwell, M. E. M.; Hilario, E.; Jacobson, F. Protein Aggregation and Bioprocessing. AAPS J. 2006, 8 (3), E572-9.

(2) Sipe, J. D.; Benson, M. D.; Buxbaum, J. N.; Ikeda, S.; Merlini, G.; Saraiva, M. J. M.; Westermark, P. Amyloid Fibril Proteins and Amyloidosis: Chemical Identification and Clinical Classification International Society of Amyloidosis 2016 Nomenclature Guidelines. Amyloid 2016, 23 (4), 209–213.

(3) de Groot, N. S.; Sabate, R.; Ventura, S. Amyloids in Bacterial Inclusion Bodies. Trends Biochem. Sci. 2009, 34 (8), 408–416.

(4) Selivanova, O. M.; Suvorina, M. Y.; Dovidchenko, N. V.; Eliseeva, I. A.; Surin, A. K.; Finkelstein, A. V.; Schmatchenko, V. V.; Galzitskaya, O. V. How to Determine the Size of Folding Nuclei of Protofibrils from the Concentration Dependence of the Rate and Lag-Time of Aggregation. II. Experimental Application for Insulin and LysPro Insulin: Aggregation Morphology, Kinetics, and Sizes of Nuclei. J. Phys. Chem. B 2014, 118 (5), 1198–1206.

(5) Dovidchenko, N. V.; Glyakina, A. V.; Selivanova, O. M.; Grigorashvili, E. I.; Suvorina, M. Y.; Dzhus, U. F.; Mikhailina, A. O.; Shiliaev, N. G.; Marchenkov, V. V.; Surin, A. K.; Galzitskaya, O. V. One of the Possible Mechanisms of Amyloid Fibrils Formation Based on the Sizes of Primary and Secondary Folding Nuclei of Aβ40 and Aβ42. J. Struct. Biol. 2016, 194 (3), 404–414.

(6) Galzitskaya, O. V; Galushko, E. I.; Selivanova, O. M. Studies of the Process of Amyloid Formation by Aβ Peptide. Biochem. 2018, 83 (S1), S62–S80.

(7) Selivanova, O. M.; Glyakina, A. V.; Gorbunova, E. Y.; Mustaeva, L. G.; Suvorina, M. Y.; Grigorashvili, E. I.; Nikulin, A. D.; Dovidchenko, N. V.; Rekstina, V. V.; Kalebina, T. S.; Surin, A. K.; Galzitskaya, O. V. Structural Model of Amyloid Fibrils for Amyloidogenic Peptide from Bgl2p–glucantransferase of S. Cerevisiae Cell Wall and Its Modifying Analog. New Morphology of Amyloid Fibrils. Biochim. Biophys. Acta - Proteins Proteomics 2016, 1864 (11), 1489–1499.

(8) Selivanova, O. M.; Surin, A. K.; Ryzhykau, Y. L.; Glyakina, A. V.; Suvorina, M. Y.; Kuklin, A. I.; Rogachevsky, V. V.; Galzitskaya, O. V. To Be Fibrils or to Be Nanofilms? Oligomers Are Building Blocks for Fibril and Nanofilm Formation of Fragments of Aβ Peptide. Langmuir 2018, 34 (6), 2332–2343.

(9) McBride, S. A.; Sanford, S. P.; Lopez, J. M.; Hirsa, A. H. Shear-Induced Amyloid Fibrillization: The Role of Inertia. Soft Matter 2016, 12 (14), 3461–3467.

(10) Maji, S. K.; Perrin, M. H.; Sawaya, M. R.; Jessberger, S.; Vadodaria, K.; Rissman, R. A.; Singru, P. S.; Nilsson, K. P. R.; Simon, R.; Schubert, D.; Eisenberg, D.; Rivier, J.; Sawchenko, P.; Vale, W.; Riek, R. Functional Amyloids as Natural Storage of Peptide Hormones in Pituitary Secretory Granules. Science 2009, 325 (5938), 328–332.

(11) Fändrich, M.; Fletcher, M. A.; Dobson, C. M. Amyloid Fibrils from Muscle Myoglobin. Nature 2001, 410 (6825), 165–166.

(12) Bermudez, O.; Forciniti, D. Aggregation and Denaturation of Antibodies: A Capillary Electrophoresis, Dynamic Light Scattering, and Aqueous Two-Phase Partitioning Study. In Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences; 2004; Vol. 807, pp 17–24.

(13) Png, G. M.; Falconer, R. J.; Abbott, D. Tracking Aggregation and Fibrillation of Globular Proteins Using Terahertz and Far-Infrared Spectroscopies. IEEE Trans. Terahertz Sci. Technol. 2016, 6 (1), 45–53.

(14) Krebs, M. R. .; Wilkins, D. K.; Chung, E. W.; Pitkeathly, M. C.; Chamberlain, A. K.; Zurdo, J.; Robinson, C. V; Dobson, C. M. Formation and Seeding of Amyloid Fibrils from Wild-Type Hen Lysozyme and a Peptide Fragment from the β-Domain. J. Mol. Biol. 2000, 300 (3), 541–549.

(15) Liu, R.; He, M.; Su, R.; Yu, Y.; Qi, W.; He, Z. Insulin Amyloid Fibrillation Studied by Terahertz Spectroscopy and Other Biophysical Methods. Biochem. Biophys. Res. Commun. 2010, 391 (1), 862–867.

(16) McBride, S. A.; Tilger, C. F.; Sanford, S. P.; Tessier, P. M.; Hirsa, A. H. Comparison of Human and Bovine Insulin Amyloidogenesis under Uniform Shear. J. Phys. Chem. B 2015, 119 (33), 10426–10433.

(17) Barranco Morales, Paulina. Kinetics of human and bovine insulin amyloid fibril formation in the presence of solid/liquid interfaces, Masters Theses. 7193; Missouri University of Science and Technology; 2013; pp 1-148.

(18) Ivanova, M. I.; Sievers, S. A.; Sawaya, M. R.; Wall, J. S.; Eisenberg, D. Molecular Basis for Insulin Fibril Assembly. Proc. Natl. Acad. Sci. 2009, 106 (45), 18990–18995.

(19) Selivanova, O.; Suvorina, M.; Surin, A.; Dovidchenko, N.; Galzitskaya, O. Insulin and Lispro Insulin: What Is Common and Different in Their Behavior? Curr. Protein Pept. Sci. 2016, 18 (1), 57–64.

(20) Babenko, V.; Piejko, M.; Wójcik, S.; Mak, P.; Dzwolak, W. Vortex-Induced Amyloid Superstructures of Insulin and Its Component A and B Chains. Langmuir 2013, 29 (17), 5271–5278.

(21) Frommeyer, M.; Steinbüchel, A. Increased Lysine Content Is the Main Characteristic of the Soluble Form of the Polyamide Cyanophycin Synthesized by Recombinant Escherichia Coli. Appl. Environ. Microbiol. 2013, 79 (14), 4474–4483.

(22) Nilsson, M. R. Techniques to Study Amyloid Fibril Formation in Vitro. Methods 2004, 34 (1), 151–160.

(23) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Kent Wenger, R.; Yao, H.; Markley, J. L. BioMagResBank. Nucleic Acids Res. 2007, 36 (Database), D402–D408.

(24) Lee, C.-C.; Nayak, A.; Sethuraman, A.; Belfort, G.; McRae, G. J. A Three-Stage Kinetic Model of Amyloid Fibrillation. Biophys. J. 2007, 92 (10), 3448–3458.

(25) Selivanova, O. M.; Surin, A. K.; Marchenkov, V. V.; Dzhus, U. F.; Grigorashvili, E. I.; Suvorina, M. Y.; Glyakina, A. V.; Dovidchenko, N. V.; Galzitskaya, O. V. The Mechanism Underlying Amyloid Polymorphism Is Opened for Alzheimer's Disease Amyloid-β Peptide. J. Alzheimer's Dis. 2016, 54 (2), 821–830.

(26) Galzitskaya, O. V.; Selivanova, O. M. Rosetta Stone for Amyloid Fibrils: The Key Role of Ring-Like Oligomers in Amyloidogenesis. J. Alzheimer's Dis. 2017, 59 (3), 785–795.

(27) Jiménez, J. L.; Nettleton, E. J.; Bouchard, M.; Robinson, C. V; Dobson, C. M.; Saibil, H. R. The Protofilament Structure of Insulin Amyloid Fibrils. Proc. Natl. Acad. Sci. U. S. A. 2002, 99 (14), 9196–9201.

(28) Dovidchenko, N. V.; Finkelstein, A. V.; Galzitskaya, O. V. How to Determine the Size of Folding Nuclei of Protofibrils from the Concentration Dependence of the Rate and Lag-Time of Aggregation. I. Modeling the Amyloid Protofibril Formation. J. Phys. Chem. B 2014, 118 (5), 1189–1197.

(29) Naiki, H.; Higuchi, K.; Hosokawa, M.; Takeda, T. Fluorometric Determination of Amyloid Fibrils in Vitro Using the Fluorescent Dye, Thioflavin T1. Anal. Biochem. 1989, 177 (2), 244–249.

(30) Maskevich, A. A.; Stsiapura, V. I.; Kuzmitsky, V. A.; Kuznetsova, I. M.; Povarova, O. I.; Uversky, V. N.; Turoverov, K. K. Spectral Properties of Thioflavin T in Solvents with Different Dielectric Properties and in a Fibril-Incorporated Form. J. Proteome Res. 2007, 6 (4), 1392–1401.

(31) LeVine, H. Thioflavine T Interaction with Amyloid β-Sheet Structures. Amyloid 1995, 2 (1), 1–6.

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

(33) Fotopoulos, C.; Kafetzopoulos, D.; Gotzamani, K. Critical Factors for Effective Implementation of the HACCP System: A Pareto Analysis. Br. Food J. 2011, 113 (5), 578–597.

(34) Münch, C.; Bertolotti, A. Exposure of Hydrophobic Surfaces Initiates Aggregation of Diverse ALS-Causing Superoxide Dismutase-1 Mutants. J. Mol. Biol. 2010, 399 (3), 512–525.

(35) Kyte, J.; Doolittle, R. F. A Simple Method for Displaying the Hydropathic Character of a Protein. J. Mol. Biol. 1982, 157 (1), 105–132.

(36) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera?A Visualization System for Exploratory Research and Analysis. J. Comput. Chem. 2004, 25 (13), 1605–1612.

(37) Lamiable, A.; Thévenet, P.; Rey, J.; Vavrusa, M.; Derreumaux, P.; Tufféry, P. PEP-FOLD3: Faster de Novo Structure Prediction for Linear Peptides in Solution and in Complex. Nucleic Acids Res. 2016, 44 (W1), W449–W454.

(38) Sereda, T. J.; Mant, C. T.; Sönnichsen, F. D.; Hodges, R. S. Reversed-Phase Chromatography of Synthetic Amphipathic Alpha-Helical Peptides as a Model for Ligand/Receptor Interactions. Effect of Changing Hydrophobic Environment on the Relative Hydrophilicity/Hydrophobicity of Amino Acid Side-Chains. J. Chromatogr. A 1994, 676 (1), 139–153.

(39) Dong, A.; Huang, P.; Caughey, W. S. Protein Secondary Structures in Water from Second-Derivative Amide I Infrared Spectra. Biochemistry 1990, 29 (13), 3303–3308.

(40) Sarroukh, R.; Goormaghtigh, E.; Ruysschaert, J. M.; Raussens, V. ATR-FTIR: A "Rejuvenated" Tool to Investigate Amyloid Proteins. Biochim. Biophys. Acta - Biomembr. 2013, 1828 (10), 2328–2338.

(41) Moran, S. D.; Zanni, M. T. How to Get Insight into Amyloid Structure and Formation from Infrared Spectroscopy. J. Phys. Chem. Lett. 2014, 5 (11), 1984–1993.

(42) Yang, W.-J.; Griffiths, P. R.; Byler, D. M.; Susi, H. Protein Conformation by Infrared Spectroscopy: Resolution Enhancement by Fourier Self-Deconvolution. Appl. Spectrosc. 1985, 39 (2), 282–287.

(43) Byler, D. M.; Susi, H. Examination of the Secondary Structure of Proteins by Deconvolved FTIR Spectra. Biopolymers 1986, 25 (3), 469–487.

(44) Susi, H.; Timasheff, S. N.; Stevens, L. Infrared Spectra and Protein Conformations in Aqueous Solutions I. THE AMIDE I BAND IN Hz0 AND DzO SOLUTIONS. J. Bro~oorcn~ Chem. 1967, 242 (23), 6460–5466.

(45) Kong, J.; Yu, S. Fourier Transform Infrared Spectroscopic Analysis of Protein Secondary Structures Protein FTIR Data Analysis and Band Assign- Ment. 2007, 39 (8), 549–559.

(46) Goldberg, M. E.; Chaffotte, A. F. Undistorted Structural Analysis of Soluble Proteins by Attenuated Total Reflectance Infrared Spectroscopy. Protein Sci. 2005, 14, 2781–2792.

(47) Dong, a C.; Huang, P.; Caughey, W. S. Redox-Dependent Changes in Beta-Extended Chain and Turn Structures of Cytochrome c in Water Solution Determined by Second Derivative Amide I Infrared Spectra. Biochemistry 1992, 31 (1), 182–189.

(48) Venyaminov, S. Y.; Kalnin, N. N. Quantitative IR Spectrophotometry of Peptide Compounds in Water (H2O) Solutions. I. Spectral Parameters of Amino Acid Residue Absorption Bands. Biopolymers 1990, 30 (13–14), 1243–1257.

(49) Speare, J. O.; Rush, T. S. IR Spectra of Cytochrome c Denatured with Deuterated Guanidine Hydrochloride Show Increase in ?? Sheet. Biopolym. - Biospectroscopy Sect. 2003, 72 (3), 193–204.

(50) Klose, D. P.; Wallace, B. A.; Janes, R. W. 2Struc: The Secondary Structure Server. Bioinformatics 2010, 26 (20), 2624–2625.

(51) Buchan, D. W. A.; Minneci, F.; Nugent, T. C. O.; Bryson, K.; Jones, D. T. Scalable Web Services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. 2013, 41 (Web Server issue), W349-57.

(52) Wang, S.; Peng, J.; Ma, J.; Xu, J. Protein Secondary Structure Prediction Using Deep Convolutional Neural Fields. Sci. Rep. 2016, 6 (1), 18962.

(53) Dzwolak, W.; Muraki, T.; Kato, M.; Taniguchi, Y. Chain-Length Dependence of α-Helix to β-Sheet Transition in Polylysine: Model of Protein Aggregation Studied by Temperature-Tuned FTIR Spectroscopy. Biopolymers 2004, 73 (4), 463–469.

(54) Shivu, B.; Seshadri, S.; Li, J.; Oberg, K. A.; Uversky, V. N.; Fink, A. L. Distinct β-Sheet Structure in Protein Aggregates Determined by ATR-FTIR Spectroscopy. Biochemistry 2013, 52 (31), 5176–5183.

(55) Juszczyk, P.; Kolodziejczyk, A. S.; Grzonka, Z. FTIR Spectroscopic Studies on Aggregation Process of the β-Amyloid 11-28 Fragment and Its Variants. J. Pept. Sci. 2009, 15 (1), 23–29.

(56) Ghosh, D.; Singh, P. K.; Sahay, S.; Jha, N. N.; Jacob, R. S.; Sen, S.; Kumar, A.; Riek, R.; Maji, S. K. Structure Based Aggregation Studies Reveal the Presence of Helix-Rich Intermediate during α-Synuclein Aggregation. Sci. Rep. 2015, 5, 9228.

(57) Ruysschaert, J.-M.; Raussens, V. ATR-FTIR Analysis of Amyloid Proteins. Methods Mol. Biol. 2018, 1777.

(58) Gowda, C.; Zandomeneghi, G.; Zimmermann, H.; Schütz, A. K.; Böckmann, A.; Ernst, M.; Meier, B. H. The Conformation of the Congo-Red Ligand Bound to Amyloid Fibrils HET-s(218–289): A Solid-State NMR Study. J. Biomol. NMR 2017, 69 (4), 207–213.

(59) Girych, M.; Gorbenko, G.; Maliyov, I.; Trusova, V.; Mizuguchi, C.; Saito, H.; Kinnunen, P. Combined Thioflavin T-Congo Red Fluorescence Assay for Amyloid Fibril Detection. Methods Appl. Fluoresc. 2016, 4 (3), 34010.

(60) Clement, C. G.; Truong, L. D. An Evaluation of Congo Red Fluorescence for the Diagnosis of Amyloidosis. Hum. Pathol. 2014, 45 (8), 1766–1772.

(61) Mitchell, A. R. Bruce Merrifield and Solid-Phase Peptide Synthesis: A Historical Assessment. Biopolymers 2008, 90 (3), 175-184.

(62) Wang Su, S. P-Alkoxybenzyl Alkohol Resin and p-Alkoxybenzylcarbonhydrazide Resin for Solid Phase Synthesis of Protected Peptide Fragments. Jacs 1972, 21 (1971), 1328–1333.

III. EFFECT OF GLYCOSYLATION ON THE AGGREGATION OF INSULIN A-CHAIN FRAGMENTS

ABSTRACT

The fibrillation of the amyloidogenic region of bovine and human insulin A-chain was study at pH 1.6 and 60°C in the presence and absence of DTT. A bovine insulin fragment (BIF) and a human insulin fragment (HIF) were glycosylated to form gBIF and gHIF. A full factorial experimental design with three factors at two levels each was conducted. The glycosylated peptides exhibited a much longer lag time than the non-glycosylated counterparts. BIF and HIF in the presence and absence of DTT show a biphasic kinetics when the fibrillation is followed by turbidity but not by the ThT assay. The fibrillation process consists of an initial transition from monomers or dimers to globular oligomers followed by a transition from oligomers to curvilinear fibrils that become rigid fibrils. Both BIF and HIF show prominent peaks for amyloid fibrils around 1620 cm^{-1} while the glycosylated peptides have aggregates still in the pre-fibrillar stage after 42 days. Congo Red stained fibrils further confirm the presence of fibrils both in the presence and absence of DTT.

1. INTRODUCTION

Proteins fold into a compact structure to carry out their specific biological functions. However, under environmental stresses or because of mutations, they may misfold leading to the formation of aggregates. Furthermore, loss of biological activity in

protein therapeutics and problems during patient administration of drugs are associate to the formation of aggregate [1]. Protein aggregation can happen at any stage during a production process, from the recovery of intracellular recombinant proteins by an unfolding/ folding reaction, to lyophilization, and long-term storage. Protein and peptide aggregates are also manifestations of various pathological conditions ranging from disordered neurodegenerative patterns to systemic amyloidosis [2].

Amyloid fibril deposits and degeneration of neural networks are the most prominent features found in many neurodegenerative diseases like Parkinson's disease, multiple sclerosis, amyotrophic lateral sclerosis and Alzheimer's disease [3]. The number of neurodegenerative cases have increased exponentially in the past decade [2]. Fibrils are linear aggregates of proteins rich in intermolecular β-sheet structures that are formed from circular oligomers as the building blocks [4,5]. It has been argued that any protein or peptide can form fibrillar structures in vitro [6,7].

Although most proteins and peptides may form amyloid fibrils, the aggregation pathways may be different from protein to protein and from species to species for the same protein. At least ten different mechanisms were proposed over the years to explain the formation of amyloid fibrils [2]. Three of the most popular are (1) monomer addition, (2) two steps model of nucleation and (3) secondary nucleation mechanism [8-10]. The **Monomer addition** mechanism is observed in the aggregation of insulin. The existence of self-complementary surfaces in insulin facilitates the formation of reversible oligomers. These oligomers, over time, form irreversible aggregates [11]. Although some proteins do not have the tendency to self-associate, conformational changes may induce the formation of oligomers [12]. According to this mechanism, unfolding of a protein is not a necessary

criterion for the proteins to aggregate. Instead, side-by-side or end-to-end union of molecules causes aggregation. In the **two-step model of nucleation**, a primary nucleation phase is followed by sigmoidal growth. In this aggregation process, monomers form critical nuclei during the lag phase followed by an elongation/growth phase where fibril size increases upon addition of pre-formed nuclei to form protofilaments [13-15]. As the equilibrium phase is reached, most of the soluble protein has been converted into amyloid fibrils. In the **secondary nucleation mechanism**, the fibrils formed during the primary nucleation mechanism and the monomers undergo a secondary nucleation reaction to form toxic oligomeric species [8]. This results in the formation of denser and longer fibrils.

The inhibition of amyloidogenesis is poorly understood despite the volume of research done [16]. Post translational modifications like glycosylation, phosphorylation, and lipidation affect the stability of a protein or a peptide [17,18]; and therefore, they may prevent or retard fibrillation. There are a few examples about that in the open literature. For example, fibrillogenesis was greatly retarded upon modification of α-helical hairpin peptide using glycosylation or phosphorylation [19]. The glycosylated peptide gR4 of Tau greatly showed a slower aggregation kinetics in comparison with its non-glycosylated counterpart [20]. The size and number of glycan groups seems to affect different aspects of protein stability differently. The glycan groups in native proteins may have a stabilizing role because they increase the hydrophilicity of the peptide surface, which in turn increases its solubility and inhibits aggregation [21]. The anomeric configuration of the sugar is one of the most important factors affecting fibrillogenesis. For example, fibrillation of the prion peptide was better inhibited by α-linked sugars than by β-linked sugars [22].

Thermodynamic stability is also altered upon glycosylation. For example, the endothermic denaturation temperature of whey protein isolate increased upon glycosylation [23].

The effect of subtle differences in the primary structure of insulin A-chain from bovine and human species was previously reported by our group [24]. The insulin A-chain is a 21-residue peptide that differs at only 2 positions for bovine and human species [25]. We used fragments of the A-chain, containing the protective group Fmoc, where those differences are located. The peptide BIF (bovine insulin fragment) has the sequence **A**S**V**CSLYQLENK$_3$ and HIF (human insulin fragment) has the sequence **TSI**CSLYQLENK$_3$. Letters in bold identify the differences in the amino acid residues in both the fragments. At pH 1.6 and 60°C, both the fragments formed fibrils in the presence of 1M urea and 20mM salt. Our previous work showed that temperature and ionic strength are the most important factors during the lag phase whereas pH is more prominent during the elongation phase during the aggregation of both fragments. We reported that the lag times are shorter for BIF while the exponential phase rate was faster for HIF. At pH 1.6 and 60°C, the values of t_{lag} and K_{app} are 11±2 days and 0.22±0.05 day^{-1} for BIF and 24±2 days and 0.47±0.16 day^{-1} for HIF respectively whereas at 25°C, they are 12±1 days and 0.85±0.37 day^{-1} for BIF and 30±2 days and 0.18±0.02 day^{-1} for HIF respectively. In addition, an increase in the percentage of β-sheets with time was observed in both the fragments. Increase in β-sheets was accompanied by a simultaneous decrease in α- helices observed during the transition from lag phase to exponential phase.

Four peptides BIF, HIF, gBIF and gHIF are studied in this article. BIF and HIF are the non-glycosylated peptides whereas gBIF and gHIF are their glycosylated counterparts. To explore the effect of subtle differences in the primary structure in the absence of Fmoc, the aggregation of pure BIF and HIF was studied at favorable fibrillation conditions [24]. Also, to study the effect of glycosylation on the fibrillation kinetics of insulin A-chain fragments from both bovine and human species, the peptides were further modified using α-D-glucose. Experiments were conducted at pH 1.6, 60°C, 1M urea and 20mM NaCl both in the presence and in the absence of dithiothreitol (DTT). DTT is a reducing reagent which can break the disulfide linkages in between peptides. We have found that in the absence of DTT, both BIF and HIF form dimers through a disulfide linkage. The effect of DTT on fibrillation, if any, was also explored in this work.

2. EXPERIMENTAL SECTION

2.1. MATERIALS

Pure peptides BIF and HIF were purchased from Biobasic Inc. ThT was purchased from Acros Organics (USA). DTT, urea, sodium chloride, Congo red dye and deuterium oxide; 99.9 atom % D were purchased from Sigma-Aldrich (USA). All other solvents were purchased from Fischer Scientific (USA).

2.2. GLYCOPEPTIDE SYNTHESIS

The pure peptides BIF and HIF were used for the synthesis of glycopeptides gBIF and gHIF respectively. The method used by Broersen *et al.* [26] was modified by increasing

the incubation time in order to glycosylate the peptides completely. 4.83 mg of BIF or 4.96 mg of HIF and 10.53 mg of D-Glucose were added to 20 ml of DI water. The pH was adjusted to 8. The solution was frozen at -20°C and lyophilized at -50°C. The lyophilized powder was then transferred into a chamber which was already maintained at a relative humidity (RH) of 65% and incubated at 60°C for 24 hours. A RH of 65% was obtained using a saturated $NaNO_2$ solution. After incubation for 24 hours, the final product was dissolved in 20 ml of DI water and then dialyzed against water for 48 hours and lyophilized to obtain the final glycosylated peptide. The glycosylated peptides tend to dimerize upon prolonged exposure to a humid environment. 1mM of glycosylated peptide was treated with 10 mM DTT in 5mM ammonium acetate buffer (pH 5.5) for 30 minutes at 37°C to break the dimers into monomers.

2.3. NMR SPECTROSCOPY

All the 4 peptides BIF, HIF, gBIF and gHIF were analyzed using liquid state one-dimensional (1D) ^1H NMR spectroscopy. To prepare the respective peptide samples for the analysis, 10 mg of the peptide was dissolved in 750 µl D_2O (99.9 % deuterium). A Bruker 400-MHz spectrometer equipped with a BBO probe was used to obtain the spectra. ^1H NMR data were obtained with a sweep width of 3997.8 MHz, 0.256 s acquisition time, 16 scans and a 2s recycle delay.

2.4. MASS SPECTROMETRY

The mass of the non-glycosylated peptides BIF and HIF was obtained using a LTQ Orbitrap. 1mg/ ml samples were diluted in 70% acetonitrile, 0.1% formic to obtain a final

concentration of 10 μM. The sample was loaded onto a static nano spray ECONO 12 tip (Proxeon) and analyzed using nano-electrospray ionization in positive-ion mode on a Thermo Scientific LTQ Orbitrap XL mass spectrometer. Typical flow rates from these tips are estimated to be 50 nL/min. FTMS data were collected in the Orbitrap (30,000 resolving power, 100-2000 m/z, 1 micro scan, maximum inject time of 100 ms) over 2 minutes of infusion. Xcalibur Qual Browser was used to analyze the data.

The purity and mass of the glycopeptides was determined using LC-MS. Data were acquired on an Agilent 6520 Q-TOF LCMS system. Lyophilized glycopeptides were suspended in 0.1% formic acid in water to obtain a final concentration of 1mg/ml. 5 μL of the sample was loaded on an Agilent Technologies C8 protein column. An Agilent Technologies 1200 Series Quaternary HPLC was used. The following elution protocol was used: initial conditions 3% B (A: 0.1% formic acid in water; B: 99.9% acetonitrile, 0.1% formic acid), rapid ramp to 50% B over 2 min, gradient 50 to 90% B over 5 min, hold at 90% B for 7 min, ramp back to (0.5 min) and hold at (2.5min) initial conditions. An Agilent Technologies 6520A Accurate Mass QTOF MS with integrated Chip Cube source was used to collect data. Data across a total of 15 minutes of elution were collected. MS data were collected between 295 and 3200 m/z.

2.5. EXPERIMENTAL DESIGN

All the experiments were carried out at pH 1.6, 60°C, 1M urea, and 20 mM NaCl for over a span of 6 weeks initially. While the runs with the non-glycosylated peptides were stopped after 6 weeks, the ones with the glycosylated peptides were restarted for further

analysis. A two- level full factorial design as shown in Table 1 consisting of three factors at two levels each with a total of 8 runs was chosen.

Table 1. Experimental design FF0308

Run	Type of peptide	Glycosylation	DTT	Notation
1	-	-	+	BIF+
2	-	-	-	BIF-
3	-	+	+	gBIF+
4	-	+	-	gBIF-
5	+	-	+	HIF+
6	+	-	-	HIF-
7	+	+	+	gHIF+
8	+	+	-	gHIF-

Table 2. Experimental factors with low and high levels

Factor	-	+
Type of Peptide	BIF	HIF
Glycosylation	No	Yes
DTT	0	1

Type of peptide, glycosylation and DTT are the three experimental factors. The lower and higher levels chosen for the factors are shown in Table 2.

2.6. IN VITRO FIBRILLATION

Samples were prepared with a peptide concentration of 10 mg/ml. All the samples were prepared in 0.22 μm filtered phosphate buffer at pH 1.6. Urea and sodium chloride were added at 1M and 20 mM concentrations respectively. Samples were incubated with ThT in the presence and absence of DTT according to the experimental design at 60°C under stirring conditions. The evolution of the fibrillation process with time was followed by turbidity and the ThT assay. Changes in secondary structure were studied using FT-IR. The water used for the preparation of buffers was purified to a resistivity of 17.5 MΩ.cm.

2.7. TURBIDITY ASSAY

The turbidity assay was conducted using a Thermo Spectronic Genesis 5 UV/Visible spectrophotometer. Absorbance at 600 nm was recorded for all the samples. All the reported data were an average of 5 replicates.

2.8. THT FLUORESCENCE ASSAY

A stock solution for the ThT assay was prepared by adding 0.8 mg of ThT dye to 50 ml of 0.22 μm filtered phosphate buffer (pH 7) and stored at 4 °C in a dark place to prevent quenching. 0.05 mM of ThT solution was added to each sample at the beginning of incubation. Fluorescence for each sample was recorded once a week. Increase in viscosity or the appearance of gel/ precipitate with time upon visual inspection is a characteristic of the ThT assay [26]. All the reported fluorescence readings were an average of 5 replicates.

2.9. FTIR ANALYSIS

Analysis was carried out with a Nicolet 6700 FT-IR spectrometer from Thermoscientific equipped with a KBr beam-splitter and a DTGC detector cooled by liquid nitrogen. Samples were dried on a CaF_2 window which is used to obtain the spectra. Drying time was kept constant from sample to sample in order to avoid any discrepancies in the spectra. Interferograms were obtained for each sample at a resolution of 8 cm^{-1} and 32 scans were collected for each run to obtain a good signal-to-noise ratio.

2.10. CONGO RED STAINING

Congo red dye was used to stain and obtain the micrographs of the fibrils. A stock solution was prepared by adding 7 mg of Congo red to 1 ml of 0.22 μm filtered water. Samples containing fibrils were first centrifuged using a micro centrifuge at 9450 x g and then washed thoroughly with water. The pellet was then resuspended in 1 ml of water and 5 μL of Congo red solution was added and incubated at 25°C for one hour. After obtaining a red precipitate, the sample was centrifuged again. The pellet was washed twice with water to remove the unattached dye. The fibrils were then resuspended in a small amount of water. 30 μL of Congo red stained fibrils were placed on a microscopic slide and air dried. Micrographs were obtained at 1000x using Hirox KH-8700 digital microscope.

3. RESULTS AND DISCUSSION

3.1. SYNTHESIS OF GLYCOPEPTIDES

The peptides bovine insulin fragment (BIF) and human insulin fragment (HIF) were used to synthesize the glycopeptides gBIF and gHIF respectively. Mass spectrometry was used to determine the purity of the peptides BIF and HIF. Figure 1 and 2 shows the deconvoluted mass spectra obtained using LTQ Orbitrap.

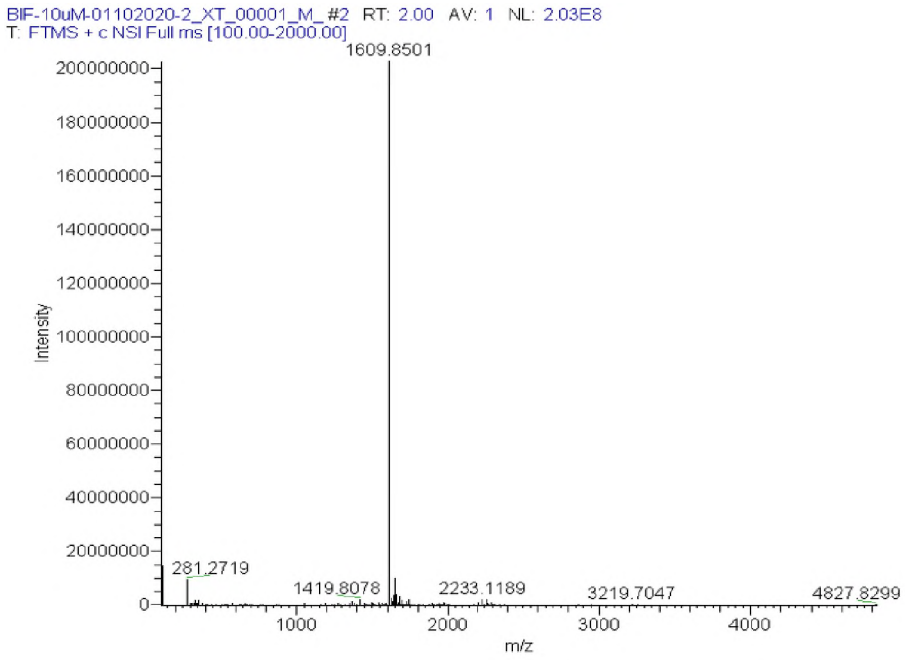

Figure 1. Deconvoluted mass spectra for BIF

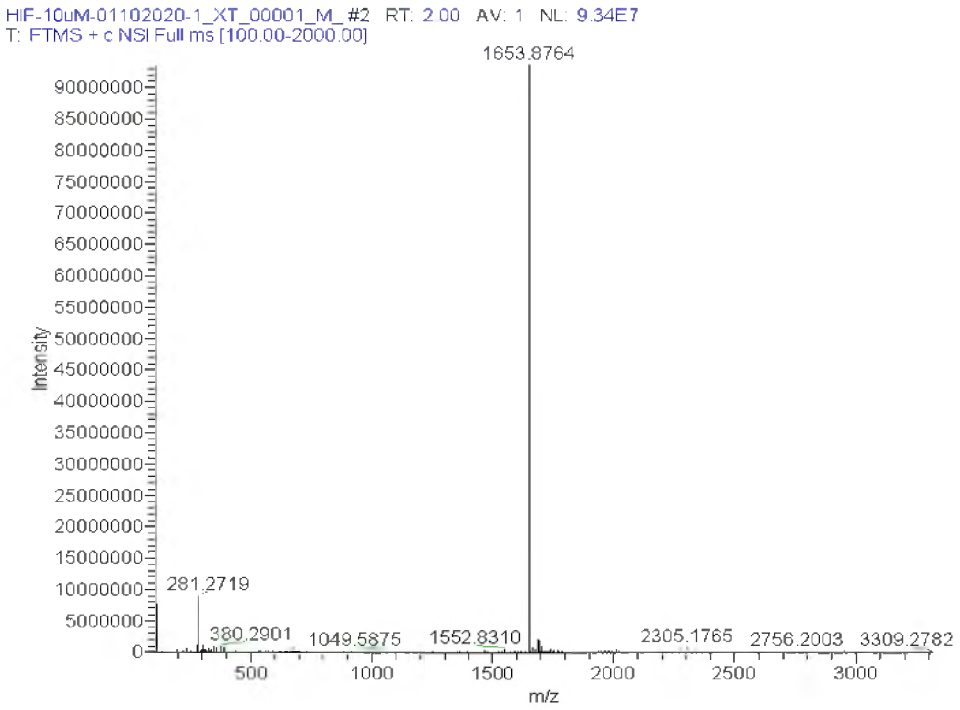

Figure 2. Deconvoluted mass spectra for HIF

BIF and HIF were identified at masses of 1609.85 Da and 1653.88 Da with a mass error of < 1 ppm. Both peptides have the same amino acid sequence except for two amino acids. Each of them has two serine residues and three lysine residues which can undergo O-glycosylation upon reacting with α-D-Glucose following the Maillard reaction [27]. Both glycosylated peptides were analyzed using a C8 column on a 1200 series LC attached to an Agilent Technologies 6520A QTOF mass spectrometer. Figures 3 and 4 shows the chromatograms for both gBIF and gHIF respectively. The samples were well resolved with a major peak around 14 min.

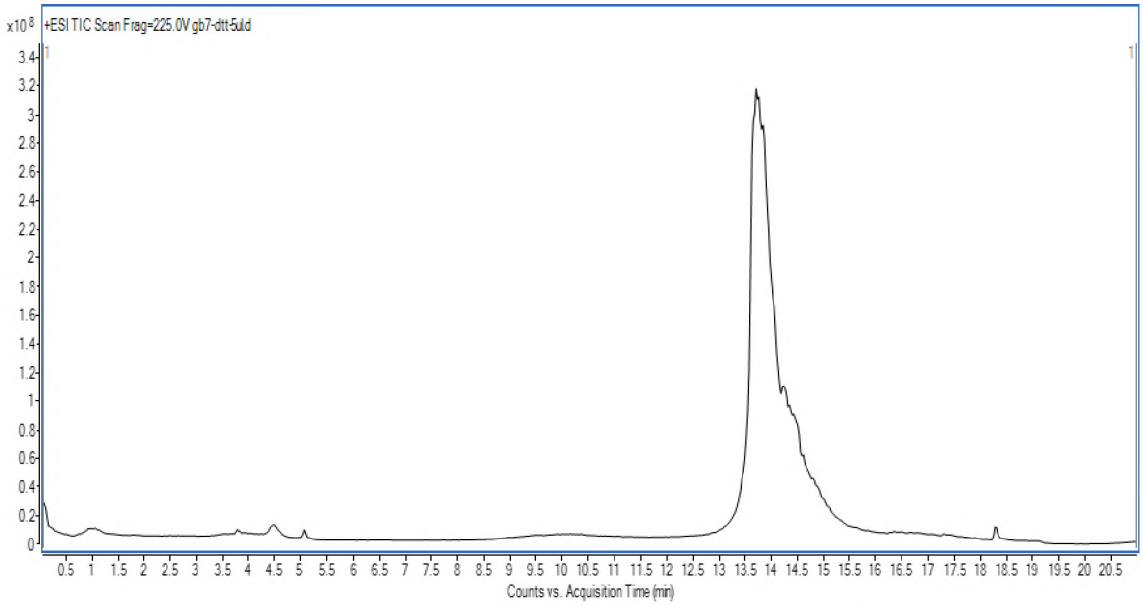

Figure 3. Chromatograms showing the elution of gBIF at around 14 min

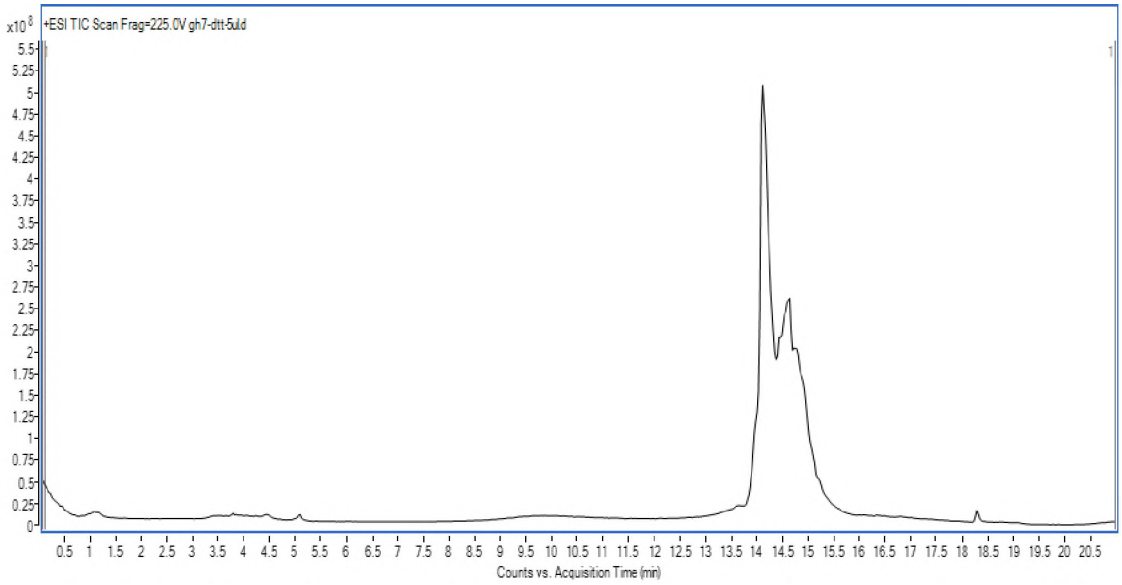

Figure 4. Chromatograms showing the elution of gHIF at around 14 min

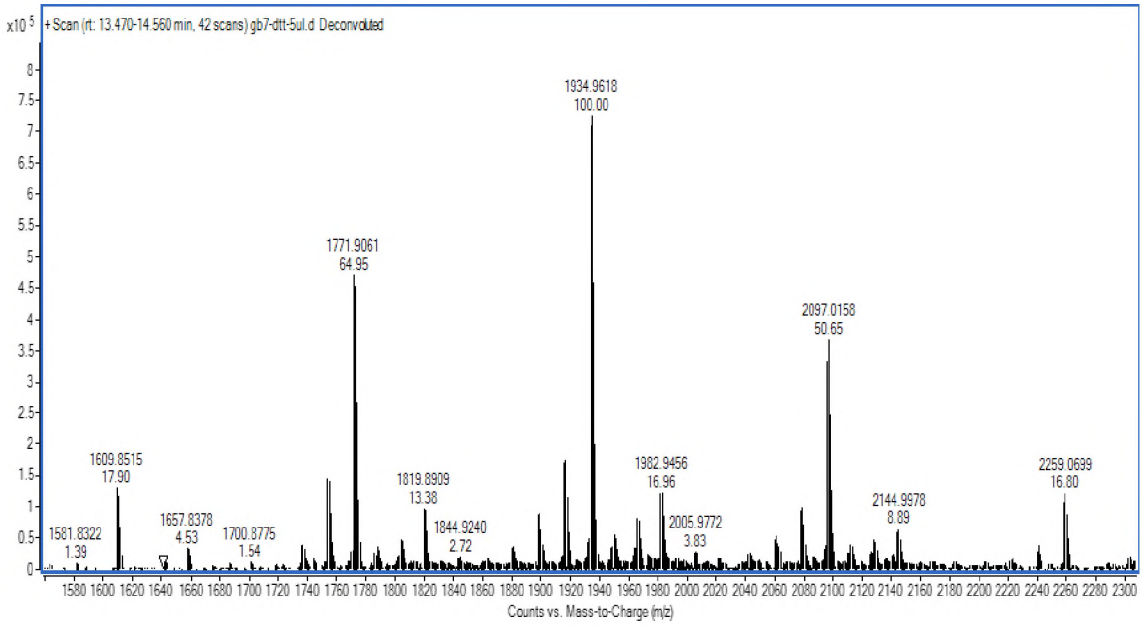

Figure 5. Deconvoluted mass spectra for gBIF

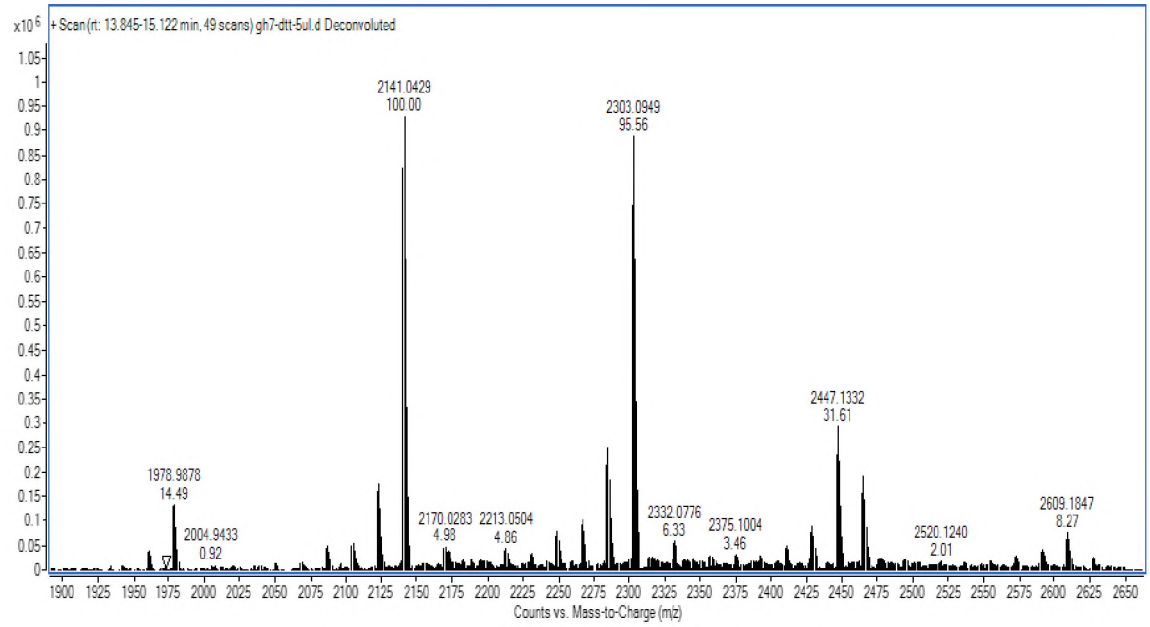

Figure 6. Deconvoluted mass spectra for gBIF

Deconvoluted mass spectra for gBIF and gHIF are shown in Figures 5 and 6 respectively. The peptide gBIF is 93% glycosylated whereas gHIF is fully glycosylated. About 86% of gBIF has 1-3 glucoses attached to the peptide with an average mass of 1934.62 Da and 87% of gHIF has 2-4 glucoses attached to the peptide with an average mass of 2141.04 Da while rest of the 13% has 5 glucoses attached. About 7% of gBIF is non-glycosylated monomer and rest of the 7% has 4 glucoses.

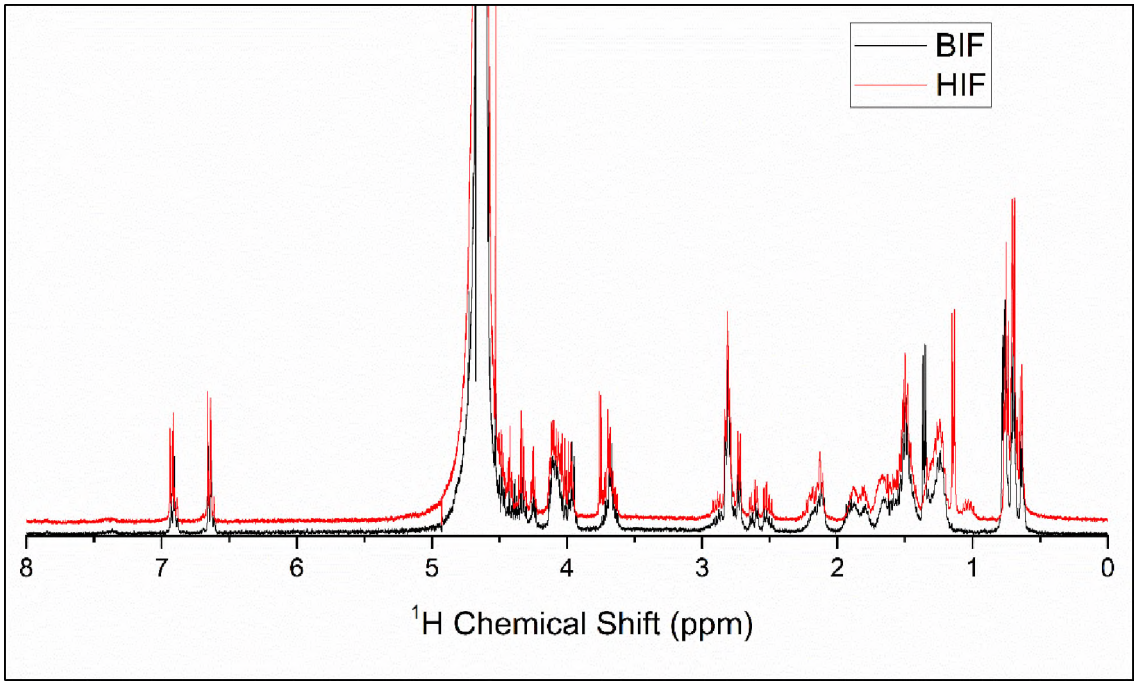

Figure 7. Full ^{1}H NMR spectra showing the comparison of BIF and HIF

3.2. NMR ANALYSIS

To determine the amino acid composition and to confirm the glucose attachment to the peptides, ^{1}H NMR spectroscopy was carried out on all the four peptides. Based on the proton chemical shifts for each amino acid, the ^{1}H resonance assignments were done.

Standards for the proton chemical shifts were taken from the Biological Magnetic Resonance Data Bank [28]. ^{1}H NMR spectra for the side chain, back bone and aromatic regions are shown separately along with the full spectra for better understanding. A comparison of the full spectra is shown in Figure 7 for BIF and HIF which shows clearly that there are only subtle differences in the spectra.

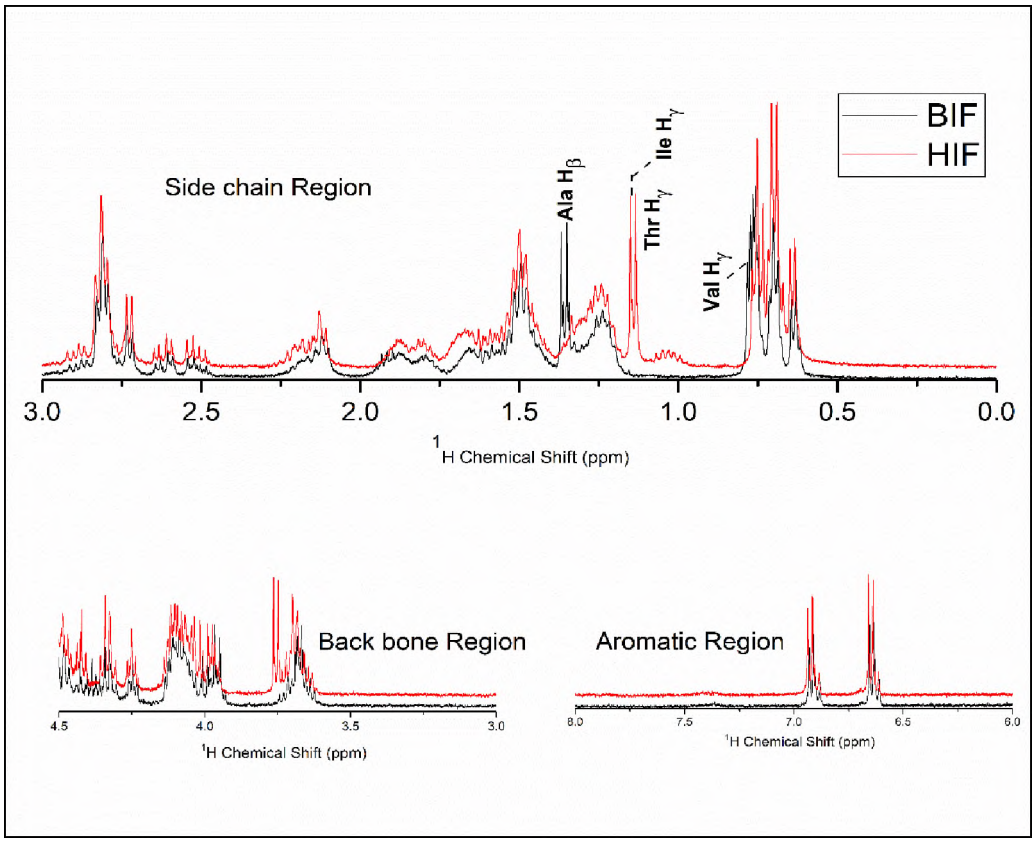

Figure 8. ^{1}H NMR spectra showing the comparison of BIF and HIF for back bone, side chain and aromatic regions

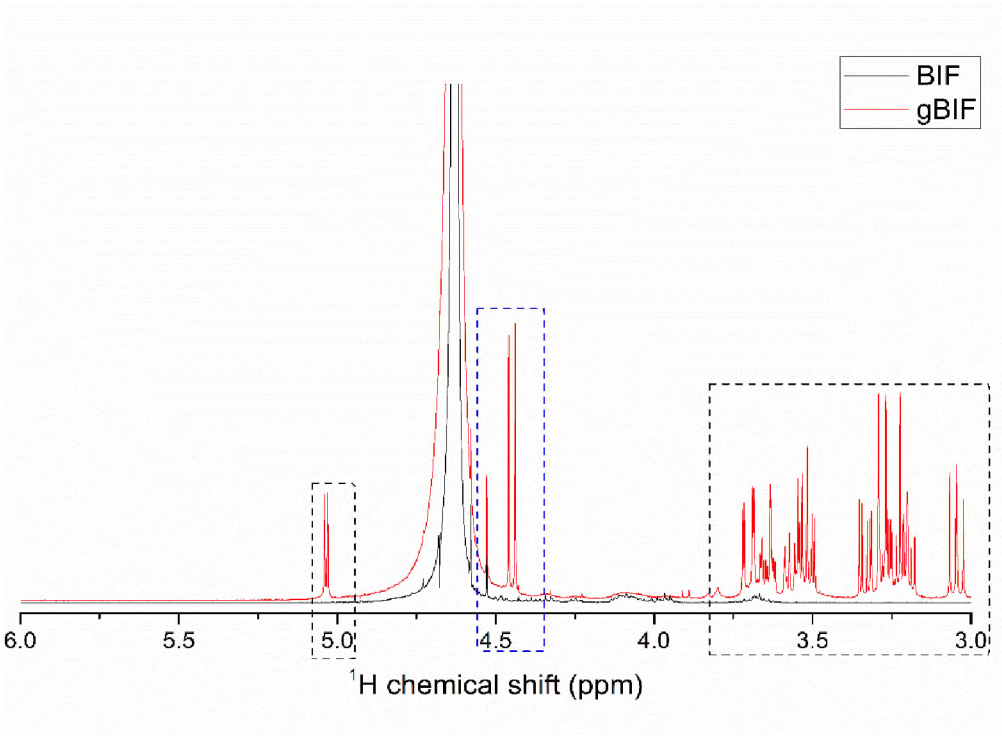

Figure 9. NMR spectra showing the attachment of glucose in BIF

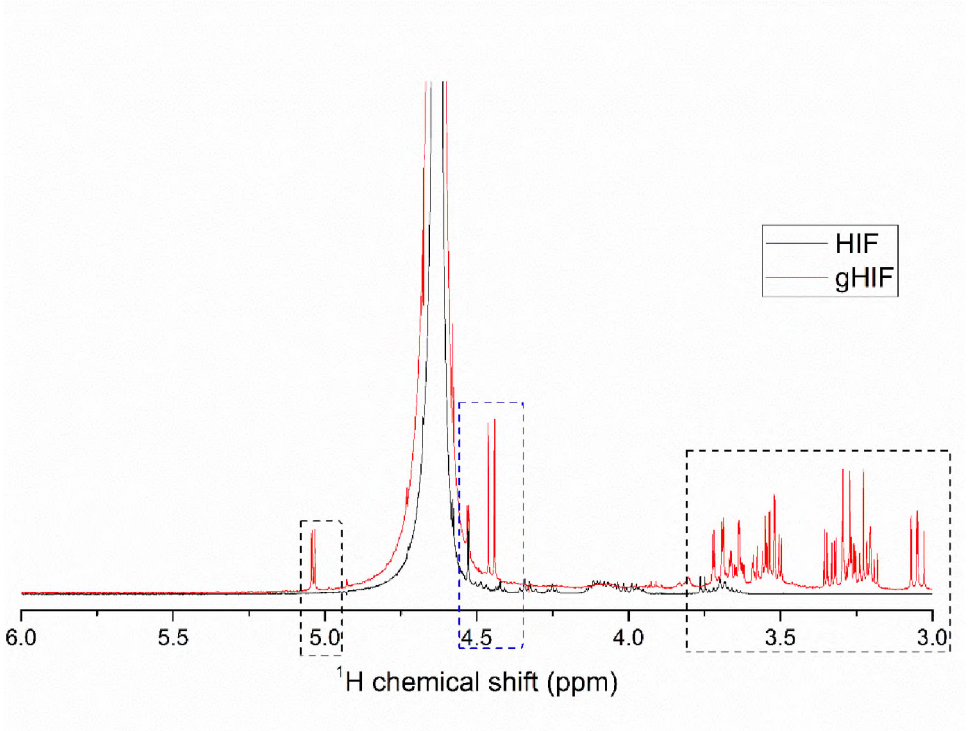

Figure 10. NMR spectra showing the attachment of glucose in HIF

These subtle differences are highlighted in Figure 8 which shows the backbone region, side chain and aromatic region in an expanded view. Assignments for rest all the amino acids were discussed in detail in our previous work [24].

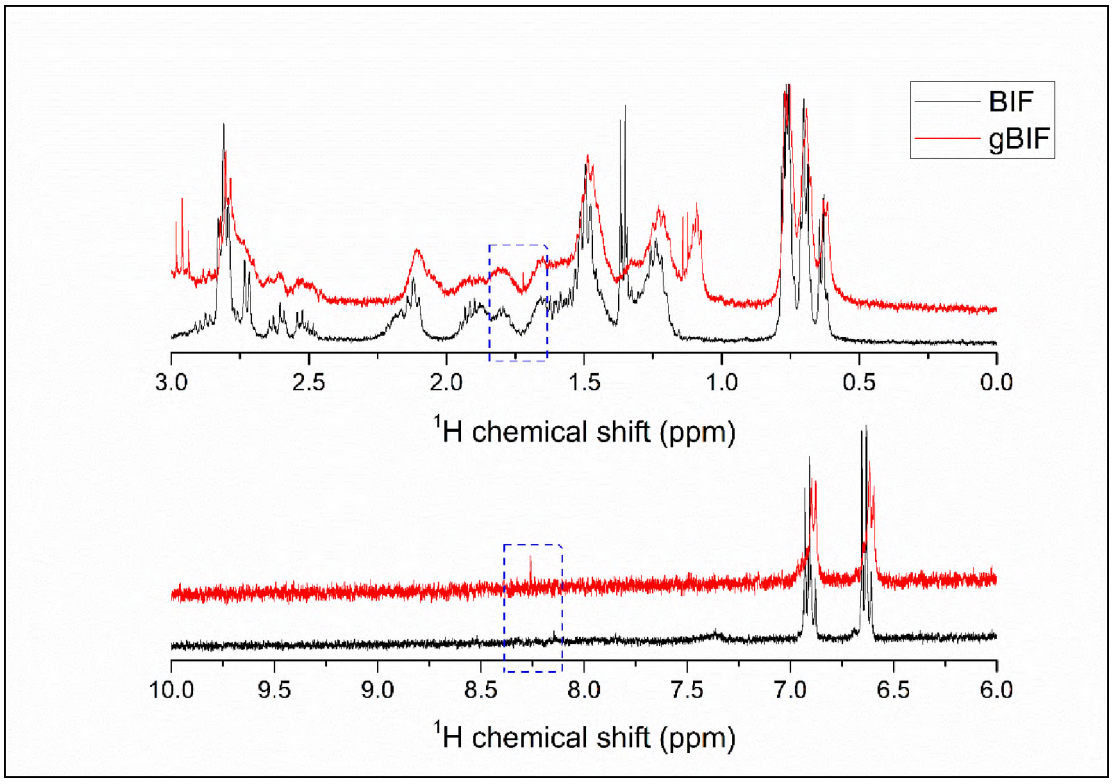

Figure 11. Comparison of various regions for glycosylated and non-glycosylated peptides in BIF

Figures 9 and 10 shows the comparison of gBIF and gHIF with its respective non-glycosylated counterparts in the backbone region. The peaks between 3- 3.75 ppm and at around 5 ppm (highlighted in black dotted line) are evident only in gBIF and gHIF which are absent in BIF and HIF. These peaks correspond to the presence of glucose attachment in the glycosylated peptides [29]. Also, the peaks at around 4.5 ppm (highlighted in blue

dotted line) correspond to lysine and serine which has higher intensity in the glycosylated ones as compared to the non-glycosylated ones. Lysine and serine being the glycosylation sites and the attachment of glucose to these amino acids could have resulted in higher intensity [30].

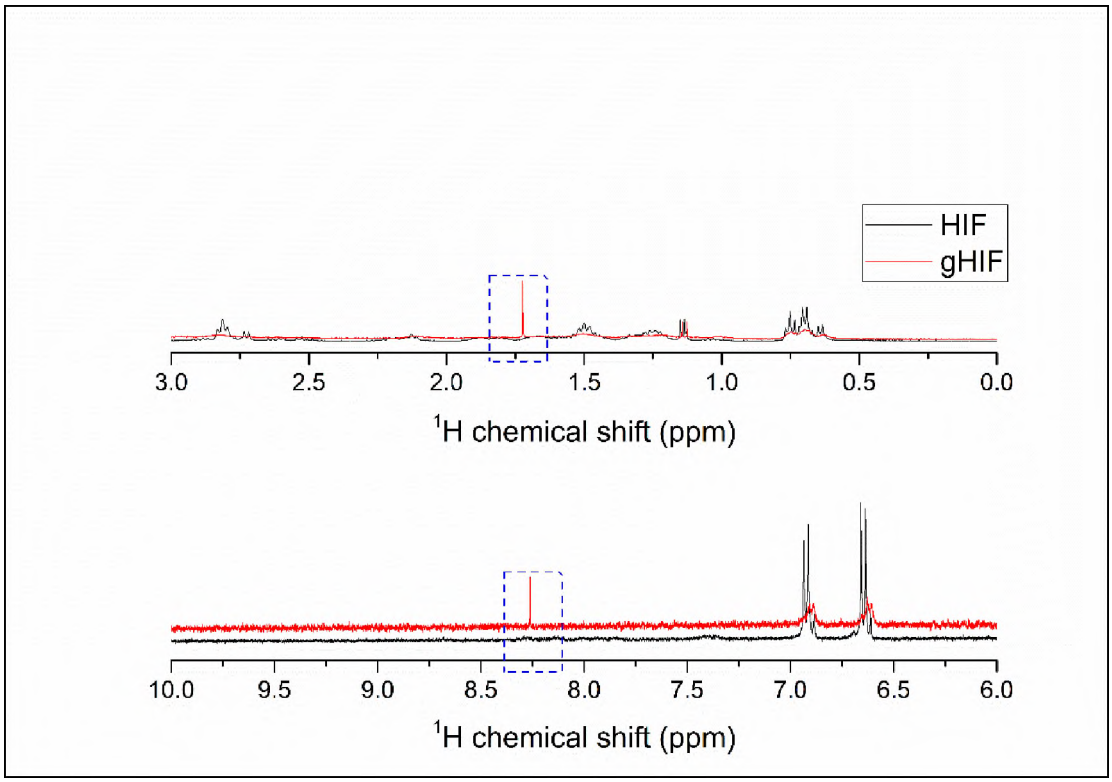

Figure 12. Comparison of various regions for glycosylated and non-glycosylated peptides in HIF

A comparison of the side chains and aromatic regions is shown for gBIF and gHIF in Figures 11 and 12 respectively. All other peaks except for the peaks around 1.75 ppm and 8.25 ppm show an increase in the intensity in the glycosylated ones. These peaks

correspond to Lys H_β and Ser H_1 respectively. This clearly shows that both the peptides BIF and HIF are glycosylated at Lys and Ser residues to form gBIF and gHIF respectively.

3.3. KINETIC STUDIES USING DIFFERENT ASSAYS

The formation of fibrils can follow different kinetic pathways. Different assays are used to track the progress of fibril formation with time. Popular among them are the ThT and the turbidity assays. Fluorescence can be monitored in the presence of a ThT dye at an emission wavelength of 485 nm whereas turbidity can be monitored by measuring the absorbance at 600 nm [30,31]. ThT dye binds to amyloid fibrils. Thus, it can only track the fibrillation once the fibrils start to form. It cannot monitor the process during the pre-fibrillar stage where monomers convert into globular oligomers.

This can be monitored using crystal violet dye which has a strong absorbance at 590 nm and is a strong oligomer selective dye [33,34]. The use of a particular method may hinder the true kinetics. For example, fibril formation monitored using ThT dye follows sigmoidal kinetics. On the contrary, by using crystal violet dye monitoring a biphasic kinetic pathway was observed [33]. This biphasic behavior corresponds to the presence of a secondary nucleation mechanism on top of the primary nucleation where fibrils undergo further oligomerization. Globular oligomers formed in the initial lag phase of 2-step kinetics elongate into curvilinear fibrils in the primary nucleation step. These curvilinear fibrils upon further oligomerization convert into rigid fibrils in the secondary nucleation step.

All the non-glycosylated peptides formed fibrils over a span of six weeks. Nevertheless, the glycosylated peptides were still in the lag phase both in the presence and

absence of DTT. Figure 13 shows the turbidity plots (absorbance at 600 nm vs time in days) and fittings for BIF+, BIF-, HIF+ and HIF- along with its glycosylated counterparts. A similar biphasic kinetics was observed in our experimental data which can be fit using the following model [34].

$$y = A1 + (A2 - A1)\left[\frac{p}{1+10^{(\log x01 - x)k1}} + \frac{1-p}{1+10^{(log x02 - x)k2}}\right] \quad (1)$$

where A1 = 0, A2 = 1, p =0.5 which leaves us with four parameters to fit the data.

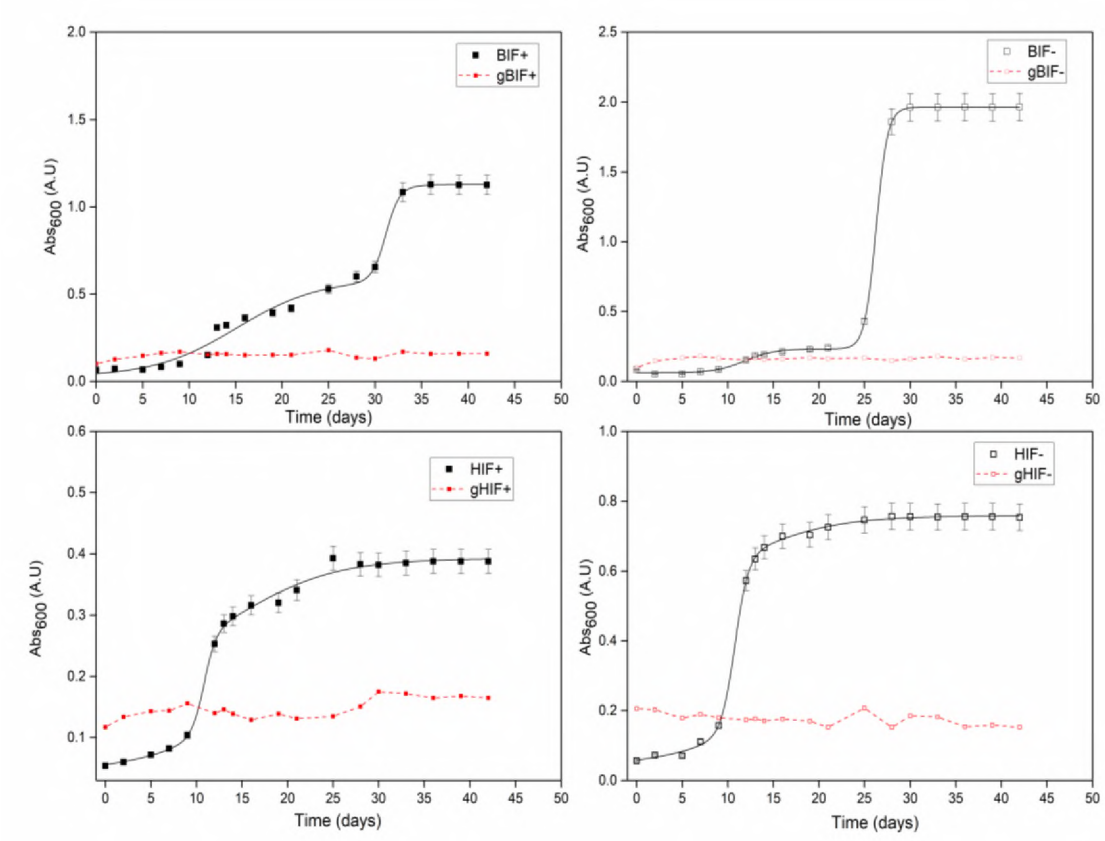

Figure 13. Turbidity plots along with its respective fits for each peptide

Table 3. Lag time and aggregation growth rate constant for all the runs

Run#	Notation	t_{lag1} (days)	t_{lag2} (days)	K_1 (day^{-1})	K_2 (day^{-1})
1	BIF+	15±1	16±1	0.09±0.03	0.59±0.22
2	BIF-	12±1	14±1	0.27±0.05	0.69±0.01
3	gBIF+	-	-	-	-
4	gBIF-	-	-	-	-
5	HIF+	11±1	4±2	0.65±0.27	0.08±0.01
6	HIF-	10±1	2±1	0.59±0.05	0.09±0.01
7	gHIF+	-	-	-	-
8	gHIF-	-	-	-	-

The parameters k_1 and k_2 correspond to the growth rate constants and log x01 = t_{lag1} and log x02 = t_{lag2}. k_1 and t_{lag1} are the growth rate and the lag phase for the conversion of globular oligomers into curvilinear fibrils while k_2 and t_{lag2} are the growth rate and the lag phase for the conversion of curvilinear fibrils into rigid fibrils. The details of all the runs in our experimental design are shown in Tables 1 and 2 along with its detailed notations. The prefix 'g' to any peptide refers to its glycosylated counterpart and the suffix '+' indicates the presence of DTT and '-' the absence of DTT.

The fitted parameters (t_{lag1}, t_{lag2}, k_1, and k_2) are reported in Table 3. For the conversion of oligomers to curvilinear fibrils the lag times (t_{lag1}) for HIF+, HIF- and BIF- are an average of 11 days whereas it is slower for BIF+ at 15 days. In both the fragments t_{lag1} is faster in the absence of DTT (BIF- and HIF-) in comparison with those in the presence of DTT (BIF+ and HIF+). Because the peptides remain monomeric in the

presence of DTT but they are dimers in its absence we can conclude that dimers aggregate faster than monomers. The rate of conversion of oligomers to curvilinear fibrils (k_1) is faster for human fragments in both the runs than their respective bovine fragments.

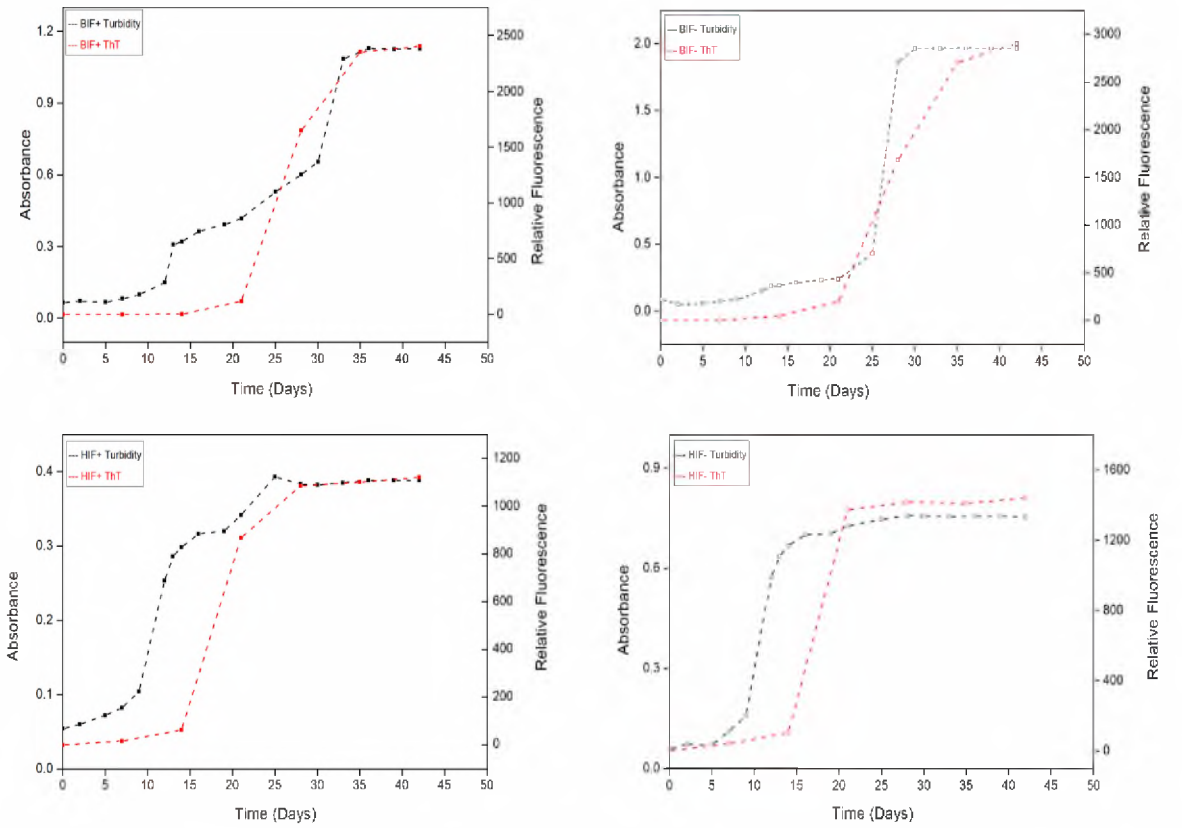

Figure 14. Turbidity and ThT plots comparison for each peptide

This is opposite for the rate of conversion of curvilinear fibrils into rigid fibrils (k_2). The value of k_2 is higher for BIF+ and BIF- in comparison with HIF+ and HIF-. But the lag time (t_{lag2}) for the formation of rigid fibrils from curvilinear fibrils is shorter for human

fragments and in case of bovine fragments t_{lag2} is even longer than t_{lag1}. Although, a clear 2-step mechanism is not evident for HIF-, the data fits well with the biphasic model.

Table 4. Lag time and aggregation growth rate constant for non-glycosylated peptides using ThT assay

Run#	Notation	t_{lag} (days)	K_{app} (day^{-1})
1	BIF+	23±1	0.53±0.03
2	BIF-	22±1	0.41±0.05
5	HIF+	16±1	0.62±0.07
6	HIF-	15±1	0.87±0.05

A similar analysis was conducted using ThT data. We observed a sigmoidal kinetics as opposed to biphasic kinetics which clearly shows that ThT assay cannot distinguish between oligomer intermediates and rigid fibrils. The two datasets for each peptide are compared in Figure 14. It clearly shows a sigmoidal kinetics with ThT assay and a biphasic kinetics with the turbidity assay. The data obtained using ThT assay were analyzed and fit using the model described in our previous work [24]. The values for t_{lag} and k_{app} obtained using the model were tabulated in Table 4. The lag times are 23±1 days and 22±1 days for BIF+ and BIF- respectively while they are 16±1 days and 15±1 days for HIF+ and HIF- respectively. In the absence of DTT at pH 1.6 and 60°C, these values can be compared with those from our previous work for Fmoc-BIF and Fmoc-HIF. The lag time for BIF is longer in comparison with Fmoc-BIF (11±2 days) while for HIF it is shorter in comparison with Fmoc-HIF (24±2 days). Also, the growth rate constants for both BIF

(0.41±0.05 day^{-1}) and HIF (0.87±0.05 day^{-1}) are faster in comparison with Fmoc-BIF (0.22±0.05 day^{-1}) and Fmoc-HIF (0.47±0.16 day^{-1}) respectively.

3.4. SEEDING EXPERIMENTS

The fibril samples BIF- and HIF- at the end of experiment are used to further conduct seeding experiments. A fresh 10 mg/ml of the respective peptide sample was added to the existing fibrillar sample and monitored using turbidity. These fibrils act as seeds to the monomers added which can accelerate the process of fibrillation.

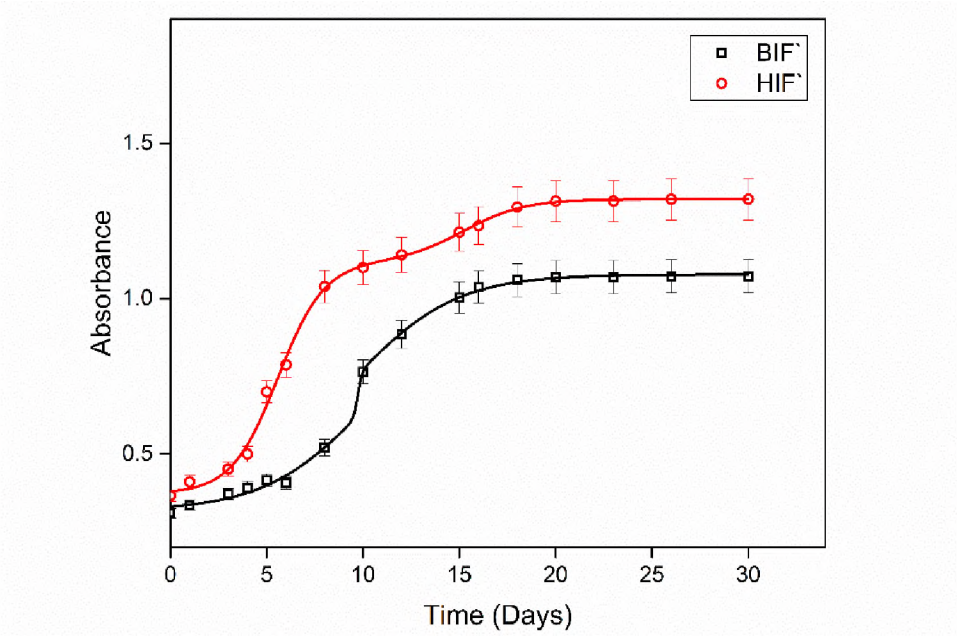

Figure 15. Turbidity plots comparison for each peptide in seeding experiments

Figure 15 shows the turbidity plots along with its respective fits using biphasic model for BIF` and HIF` respectively. The lag time t_{lag1} is shorter for BIF` in comparison with HIF` while it is the opposite for t_{lag2}. Similarly, the rate constant k_1 is faster and k_2 is slower for BIF`.

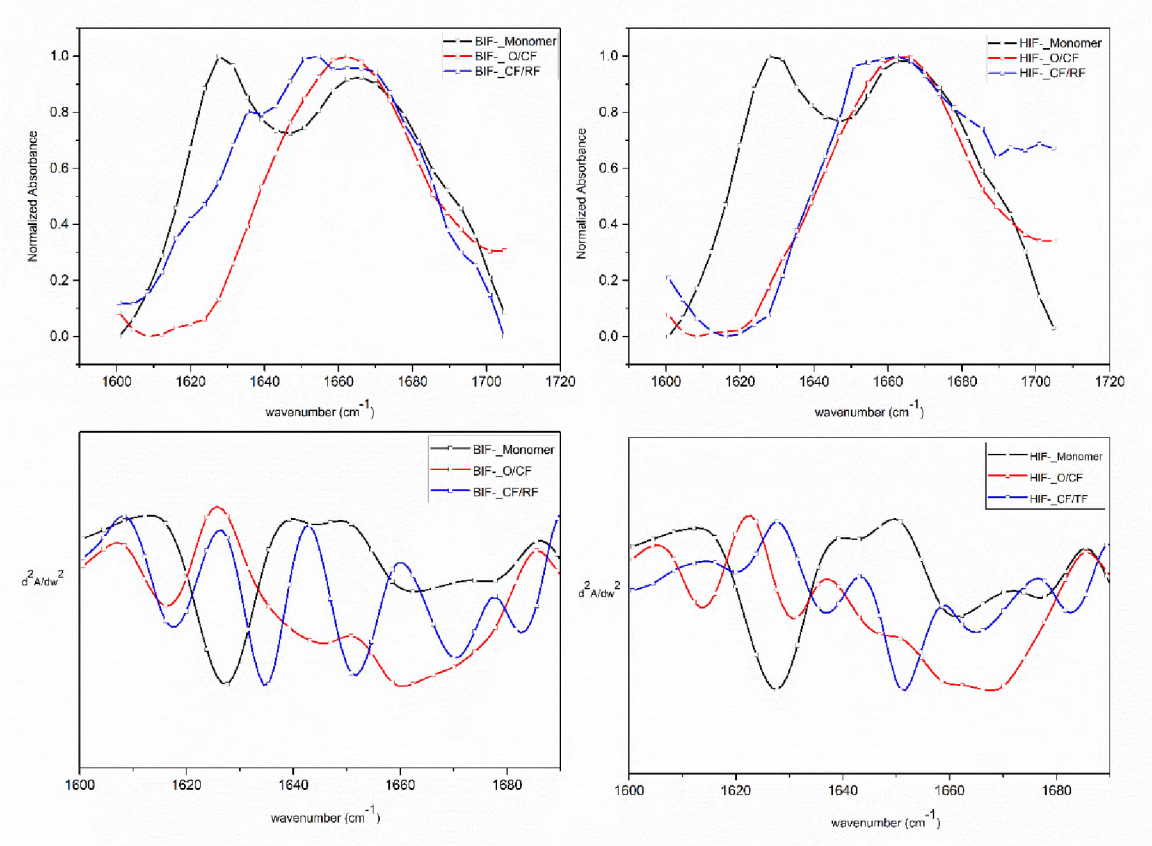

Figure 16. FTIR spectra showing the amide I peaks and their corresponding deconvoluted spectra for BIF and HIF in the absence of DTT

3.5. SECONDARY STRUCTURE ANALYSIS

Another important characteristic of an amyloid fibril is the formation of intermolecular β-sheets [26]. Secondary structure of a protein can be monitored using Fourier Transform Infrared spectroscopy (FTIR). The amide I peak obtained using an FTIR ranges between 1700–1600 cm^{-1}. With change in the secondary structure of a protein/peptide the amide I peak shifts its position [36]. Second derivative analysis of the spectra deconvolutes the band to highlight the underlying peaks which corresponds to specific secondary structure [37,38] which can be used to characterize both oligomers (anti-parallel β-sheets) and amyloid fibrils (parallel β-sheets).

Figure 16 shows the comparison of amide I peaks (top panel) along with its deconvoluted spectra (bottom panel) for BIF and HIF in the absence of DTT. The amide I spectra are shown at three different stages of the fibril formation as mentioned above i.e., during the initial stage when the peptides are monomers or dimers, during the conversion of oligomers to curvilinear fibrils and then during the conversion of curvilinear fibrils to rigid fibrils. A clear shift in the amide I peak with the progress of fibril formation can be seen in Figure 16 for both the peptides. From the second derivative spectra, during the initial stage both the peptides BIF and HIF in the absence of DTT show prominent peaks at around 1630, 1660 and 1680 cm^{-1} respectively which correspond to native β-sheets, α-helices and turns. During the transition of oligomers to curvilinear fibrils, the peak at 1630 cm^{-1} shifts to around 1618 cm^{-1} in BIF which is an indicative of amyloid fibrils. But in HIF, a peak at 1630 cm^{-1} still persists along with the peak at 1618 cm^{-1}, which indicates some amount of native β-sheets. In addition, an intermediate increase in the peak at 1660 cm^{-1} indicates the presence of increase in α-helices in both the peptides as evident from Figure

16 (bottom panel). During the final stage of Curvilinear fibril/ Rigid fibril transition, only amyloid fibrils are present in both the peptides while the peak at 1660 cm^{-1} which corresponds to α-helices decreases at the expense of random coils and turns in both the peptides. Although the glycosylated peptides are still in the lag phase after six weeks, there are some changes in the secondary structure as shown in Figure 17. A plot of amide I peak and their respective deconvoluted spectra at three different times are shown in Figure 17 for both gBIF and gHIF respectively.

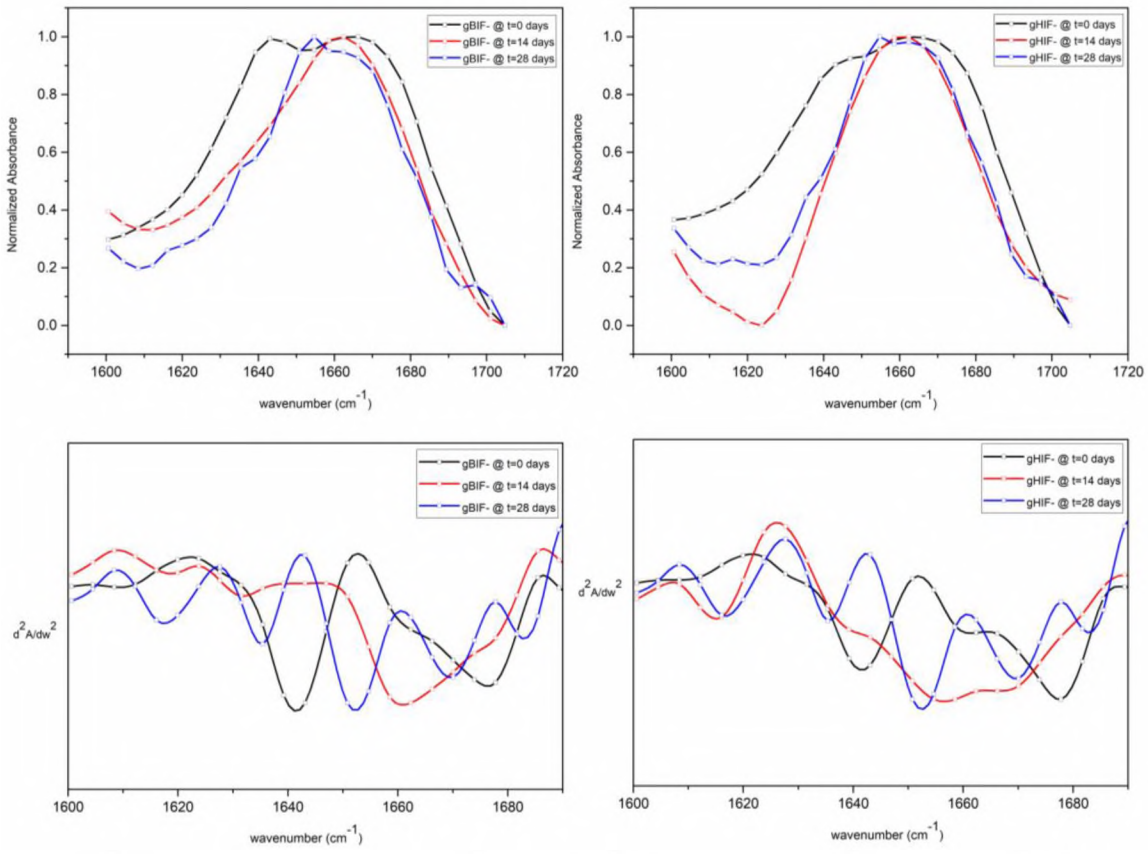

Figure 17. FTIR spectra showing the amide I peaks and their corresponding deconvoluted spectra for gBIF and gHIF in the absence of DTT

At t=0 days, both the glycopeptides are rich in α-helices, random coils and turns while at t=14 days both the peptides show increase in α-helices with a small peak at around 1620 cm^{-1}. At t=28 days, a peak at around 1635 cm^{-1} is prominent which corresponds to parallel β-sheets along with a peak around 1690 cm^{-1} which corresponds to anti-parallel β-sheets which shows that the glycosylated peptides are still in oligomer phase. After 28 days, no change in the amide I peak was observed for both gBIF and gHIF; therefore, they stay as oligomers.

A similar analysis was also conducted on all the four peptides in the presence of DTT. Figure18 shows the deconvoluted amide I peaks at different stages for BIF+, HIF+, gBIF+ and gHIF+ respectively. BIF and gBIF show transitions in its secondary structure similar to those observed in the absence of DTT. This is not true in case of HIF and gHIF. For HIF in the presence of DTT, the intensity of peak at 1620 cm^{-1} decreased as it reaches final aggregation stage with a prominent peak close to 1640 cm^{-1}; this suggests that HIF+ is rich in parallel β-sheets but not in fibrils.

The effect of DTT on both the peptides BIF and HIF can be explained as follows. For HIF in the presence of DTT, there is no significant shift in the peak around 1620 cm^{-1} during the curvilinear fibril/rigid fibril transition. Hence in both the peptides, absence of DTT favors the formation of rigid fibrils while its presence slows down the process resulting in only curvilinear fibrils for HIF.

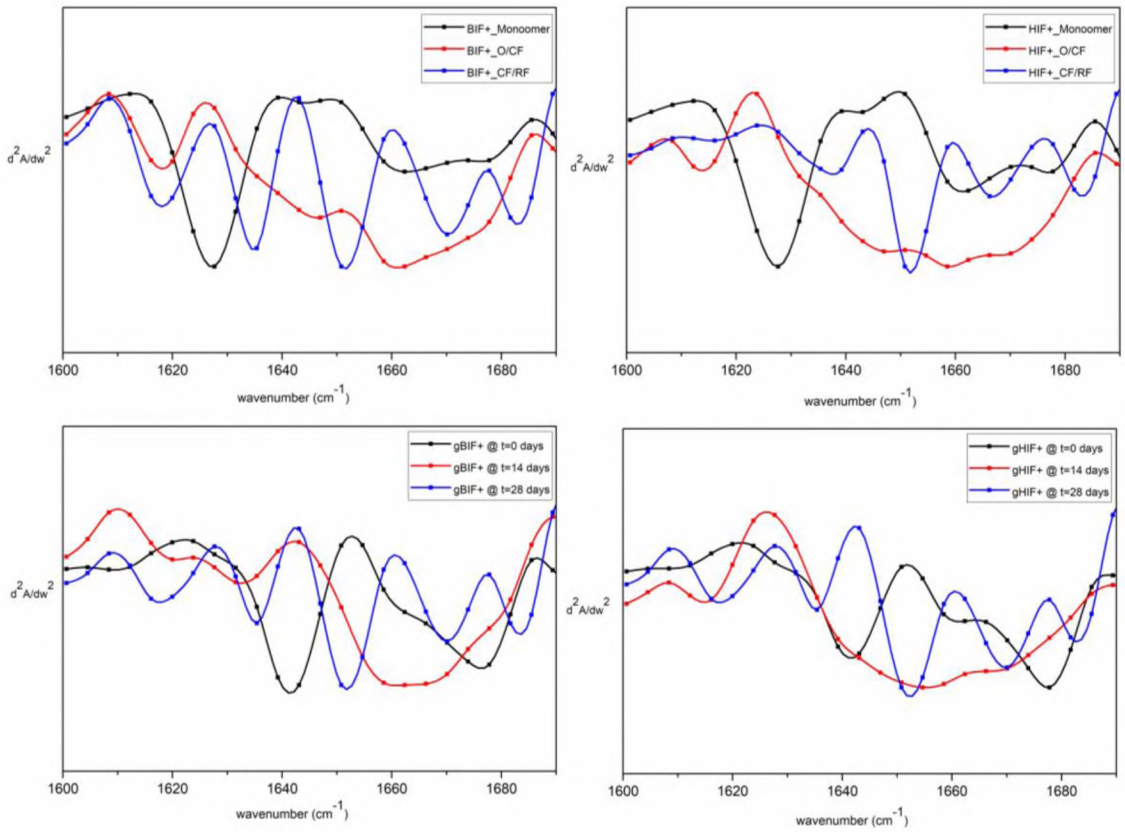

Figure 18. FTIR spectra showing the corresponding deconvoluted spectra for BIF and HIF in the presence of DTT

3.6. CONGO RED ANALYSIS

It is well known that Congo red binds to amyloid fibrils [41]. Free Congo red shows a peak maxima at around 500 nm while Congo red bound to the fibrils shows a peak shift to around 540 nm [42]. The fibrils obtained at the end of experiment were suspended in water and incubated with Congo red for 1 hr. Wavelength scans from 400 to 700 nm were done immediately after the dye was added and after 1 hr has elapsed. Figures 19 and 20 show a wavelength scan of fibrils between 400 – 700 nm for BIF and HIF respectively both in the presence and absence of DTT.

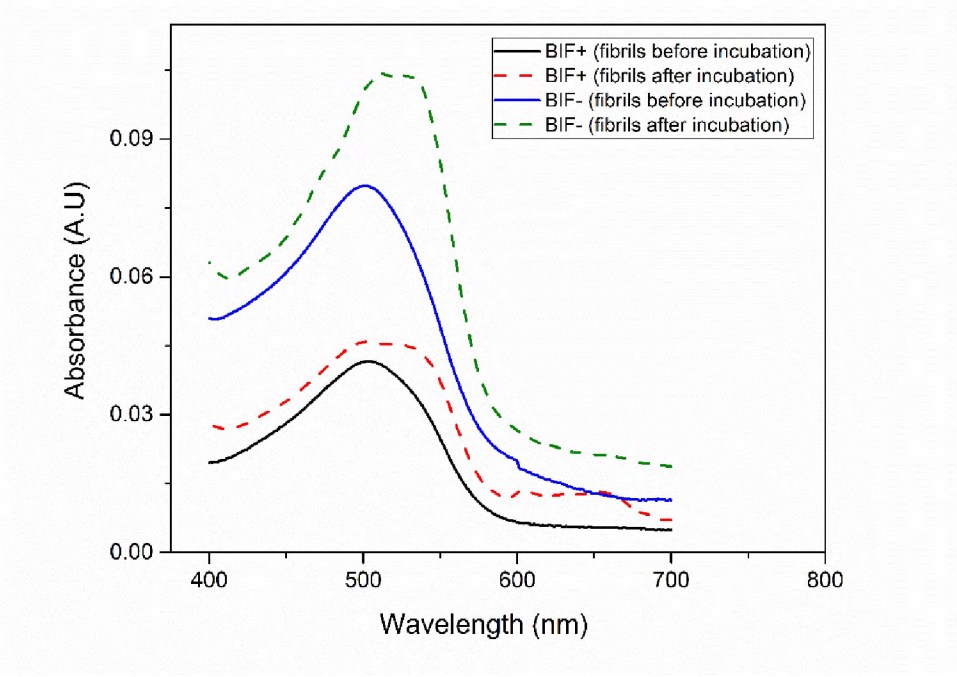

Figure 19. Congo Red spectroscopy revealing the shift in the peak for BIF at different conditions

Before the incubation starts, a peak at around 500 nm is persistent in all the four samples (BIF+, BIF-, HIF+ and HIF-), that corresponds to the free dye. After 1 hour, a shift in the peak maxima and an increase in the absorbance is observed in all the samples which corresponds to the presence of fibrils. Microscopy was used to provide a visual confirmation of the presence of fibrils. Staining the samples with Congo red dye and inspection under a digital microscope shows the morphology of both the fibrils formed by peptides BIF and HIF both in the presence and absence of DTT (Figures 23-26). All the micrographs were obtained at a magnification of 1000x.

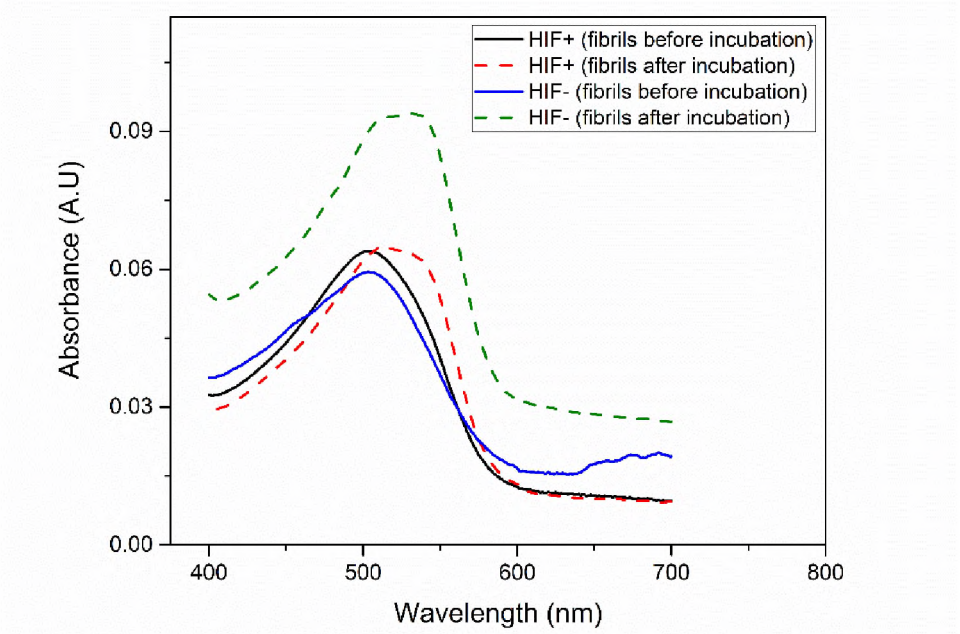

Figure 20. Congo Red spectroscopy revealing the shift in the peak for HIF at different conditions

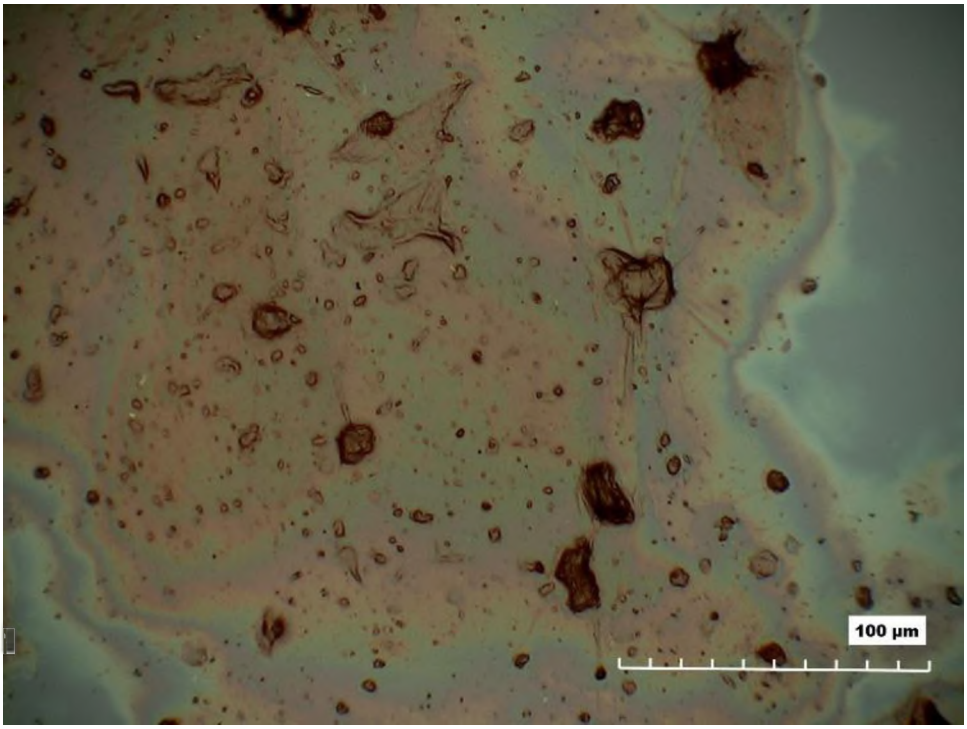

Figure 21. Micrographs showing the presence of fibrils for BIF+

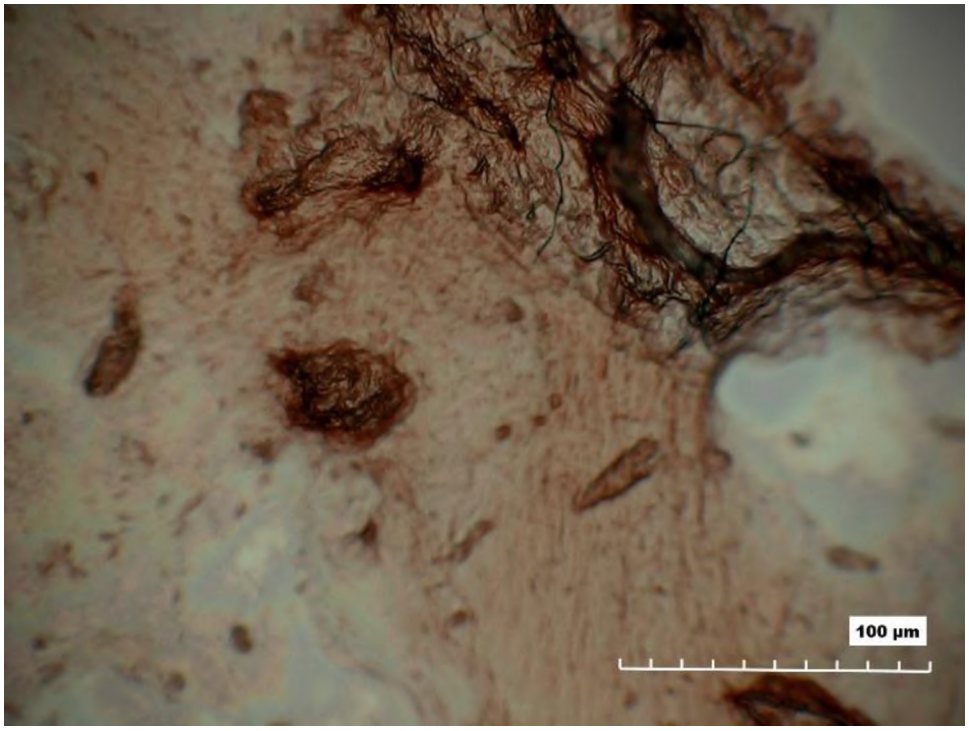

Figure 22. Micrographs showing the presence of fibrils for BIF in absence of DTT

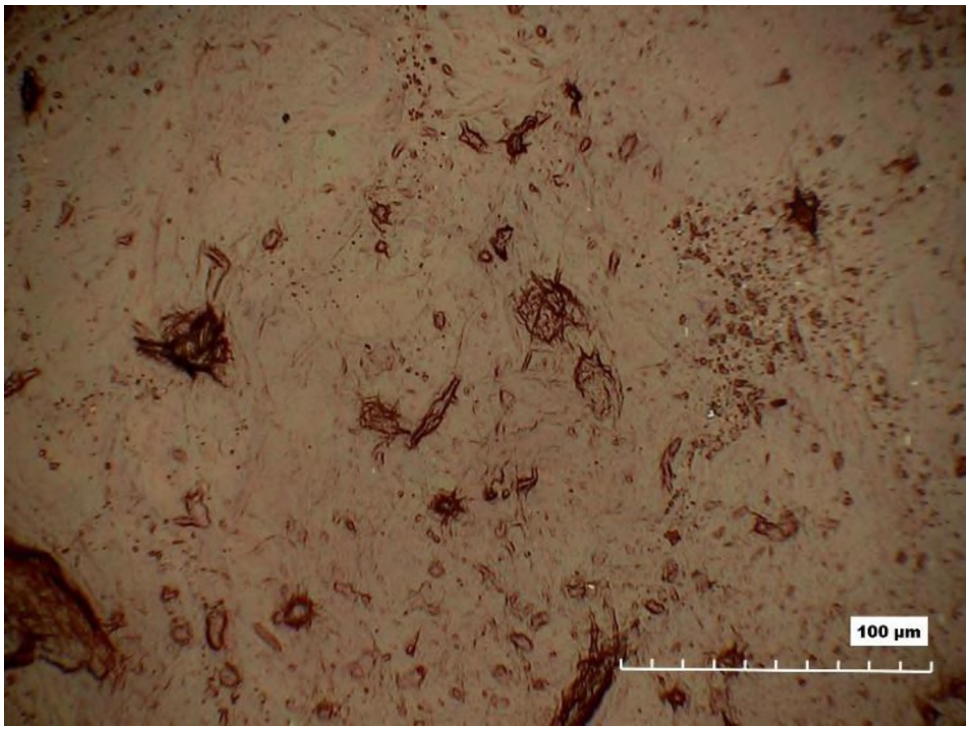

Figure 23. Micrographs showing the presence of fibrils for HIF+

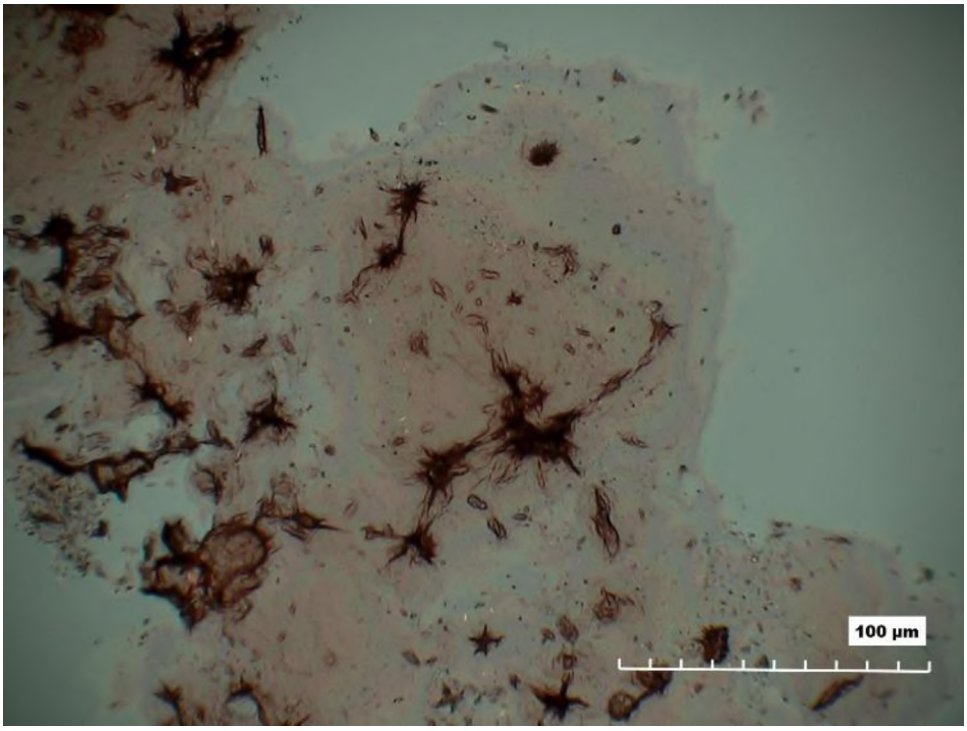

Figure 24. Micrographs showing the presence of fibrils for HIF in absence of DTT

Figures 21 and 22 show the Congo red stained fibrils for BIF+ and BIF- whereas Figures 23 and 24 shows the fibrils for HIF+ and HIF-. Annular amyloid aggregates are observed in all of the samples which is an indication of amyloidosis but denser fibrillar structures are observed in the absence of DTT (Figure 22 and 24) which is not seen in its presence (Figure 21 and 23).

4. CONCLUSIONS

Two bovine and human insulin A-chain fragments (BIF and HIF) were successfully glycosylated using α-D-glucose. Using LC-MS and ^{1}H NMR, the attachment of the sugar to the peptide was confirmed. Delay in the kinetics of fibrillation was observed in the

glycosylated peptides in comparison with their non-glycosylated counterparts. We found a biphasic kinetics for both BIF and HIF in the presence and absence of DTT, which was not observed previously using the ThT assay. A progression in fibril formation from monomer to globular oligomers to curvilinear fibrils to rigid fibrils was observed for both peptides. Changes in the secondary structure of both peptides were observed during fibrillation. Both BIF and HIF show prominent peaks corresponding to amyloid fibrils (1620 cm^{-1}) while the glycosylated peptides have aggregates still in the pre-fibrillar stage. Congo red stained fibrils further confirm the presence of fibrils both in the presence and absence of DTT.

REFERENCES

(1) Cromwell, M. E. M.; Hilario, E.; Jacobson, F. Protein Aggregation and Bioprocessing. AAPS J. 2006, 8 (3), E572-9.

(2) Siddiqi, M. K.; Alam, P.; Chaturvedi, S. K.; Shahein, Y. E.; Khan, R. H. Mechanisms of Protein Aggregation and Inhibition. Frontiers in Bioscience - Elite. Frontiers in Bioscience January 1, 2017, pp 1–20.

(3) Vaquer-Alicea, J.; Diamond, M. I. Propagation of Protein Aggregation in Neurodegenerative Diseases. Annu. Rev. Biochem. 2019, 88 (1), 785–810.

(4) Selivanova, O. M.; Surin, A. K.; Ryzhykau, Y. L.; Glyakina, A. V.; Suvorina, M. Y.; Kuklin, A. I.; Rogachevsky, V. V.; Galzitskaya, O. V. To Be Fibrils or to Be Nanofilms? Oligomers Are Building Blocks for Fibril and Nanofilm Formation of Fragments of Aβ Peptide. Langmuir 2018, 34 (6), 2332–2343.

(5) Galzitskaya, O. V; Galushko, E. I.; Selivanova, O. M. Studies of the Process of Amyloid Formation by Aβ Peptide. Biochem. 2018, 83 (S1), S62–S80.

(6) Bermudez, O.; Forciniti, D. Aggregation and Denaturation of Antibodies: A Capillary Electrophoresis, Dynamic Light Scattering, and Aqueous Two-Phase Partitioning Study. In Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences; 2004; Vol. 807, pp 17–24.

(7) Png, G. M.; Falconer, R. J.; Abbott, D. Tracking Aggregation and Fibrillation of Globular Proteins Using Terahertz and Far-Infrared Spectroscopies. IEEE Trans. Terahertz Sci. Technol. 2016, 6 (1), 45–53.

(8) Cohen, S. I. A.; Linse, S.; Luheshi, L. M.; Hellstrand, E.; White, D. A.; Rajah, L.; Otzen, D. E.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P. J. Proliferation of Amyloid-B42 Aggregates Occurs through a Secondary Nucleation Mechanism. Proc. Natl. Acad. Sci. U. S. A. 2013, 110 (24), 9758–9763.

(9) Watzky, M. A.; Finke, R. G. Transition Metal Nanocluster Formation Kinetic and Mechanistic Studies. A New Mechanism When Hydrogen Is the Reductant: Slow, Continuous Nucleation and Fast Autocatalytic Surface Growth. J. Am. Chem. Soc. 1997, 119 (43), 10382–10400.

(10) Hofrichter, J.; Ross, P. D.; Eaton, W. A. Kinetics and Mechanism of Deoxyhemoglobin S Gelation: A New Approach to Understanding Sickle Cell Disease. Proc. Natl. Acad. Sci. U. S. A. 1974, 71 (12), 4864–4868.

(11) Pekar, A. H.; Frank, B. H. Conformation of Proinsulin. A Comparison of Insulin and Proinsulin Self-Association at Neutral PH. Biochemistry 1972, 11 (22), 4013–4016.

(12) Raso, S. W.; Abel, J.; Barnes, J. M.; Maloney, K. M.; Pipes, G.; Treuheit, M. J.; King, J.; Brems, D. N. Aggregation of Granulocyte-Colony Stimulating Factor In vitro Involves a Conformationally Altered Monomeric State. Protein Sci. 2005, 14 (9), 2246–2257.

(13) Lee, C.-C.; Nayak, A.; Sethuraman, A.; Belfort, G.; McRae, G. J. A Three-Stage Kinetic Model of Amyloid Fibrillation. Biophys. J. 2007, 92 (10), 3448–3458.

(14) Jiménez, J. L.; Nettleton, E. J.; Bouchard, M.; Robinson, C. V; Dobson, C. M.; Saibil, H. R. The Protofilament Structure of Insulin Amyloid Fibrils. Proc. Natl. Acad. Sci. U. S. A. 2002, 99 (14), 9196–9201.

(15) Dovidchenko, N. V.; Finkelstein, A. V.; Galzitskaya, O. V. How to Determine the Size of Folding Nuclei of Protofibrils from the Concentration Dependence of the Rate and Lag-Time of Aggregation. I. Modeling the Amyloid Protofibril Formation. J. Phys. Chem. B 2014, 118 (5), 1189–1197.

(16) Ehrnhoefer, D. E.; Bieschke, J.; Boeddrich, A.; Herbst, M.; Masino, L.; Lurz, R.; Engemann, S.; Pastore, A.; Wanker, E. E. EGCG Redirects Amyloidogenic Polypeptides into Unstructured, off-Pathway Oligomers. Nat. Struct. Mol. Biol. 2008, 15 (6), 558–566.

(17) Hu, J.; Guo, Y.; Li, Y. Research Progress in Protein Post-Translational Modification. Chinese Sci. Bull. 2006, 51 (6), 633–645.

(18) Jenkins, N.; Murphy, L.; Tyther, R. Post-Translational Modifications of Recombinant Proteins: Significance for Biopharmaceuticals. Mol. Biotechnol. 2008, 39 (2), 113–118.

(19) Liang, F. C.; Chen, R. P. Y.; Lin, C. C.; Huang, K. T.; Chan, S. I. Tuning the Conformation Properties of a Peptide by Glycosylation and Phosphorylation. Biochem. Biophys. Res. Commun. 2006, 342 (2), 482–488.

(20) Yu, C. H.; Si, T.; Wu, W. H.; Hu, J.; Du, J. T.; Zhao, Y. F.; Li, Y. M. O-GlcNAcylation Modulates the Self-Aggregation Ability of the Fourth Microtubule-Binding Repeat of Tau. Biochem. Biophys. Res. Commun. 2008, 375 (1), 59–62.

(21) Oliveira, L. M.; Lages, A.; Gomes, R. A.; Neves, H.; Família, C.; Coelho, A. V.; Quintas, A. Insulin Glycation by Methylglyoxal Results in Native-like Aggregation and Inhibition of Fibril Formation. BMC Biochem. 2011, 12 (1), 41.

(22) Ho, C. C.; Lee, L. Y. L.; Huang, K. T.; Lin, C. C.; Ku, M. Y.; Yang, C. C.; Chan, S. I.; Hsu, R. L.; Chen, R. P. Y. Tuning the Conformational Properties of the Prion Peptide. Proteins Struct. Funct. Bioinforma. 2009, 76 (1), 213–225.

(23) Liu, G.; Zhong, Q. Thermal Aggregation Properties of Whey Protein Glycated with Various Saccharides. Food Hydrocoll. 2013, 32 (1), 87–96.

(24) Nakka, P. P.; Li, K.; Forciniti, D. Effect of Differences in the Primary Structure of the A-Chain on the Aggregation of Insulin Fragments. ACS Omega 2018, 3 (8), 9636–9647.

(25) Babenko, V.; Piejko, M.; Wójcik, S.; Mak, P.; Dzwolak, W. Vortex-Induced Amyloid Superstructures of Insulin and Its Component A and B Chains. Langmuir 2013, 29 (17), 5271–5278.

(26) Nilsson, M. R. Techniques to Study Amyloid Fibril Formation in Vitro. Methods 2004, 34 (1), 151–160.

(27) Ogasawara, M.; Katsumata, T.; Egi, M. Taste Properties of Maillard-Reaction Products Prepared from 1000 to 5000 Da Peptide. Food Chem. 2006, 99 (3), 600–604.

(28) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Kent Wenger, R.; Yao, H.; Markley, J. L. BioMagResBank. Nucleic Acids Res. 2007, 36 (Database), D402–D408.

(29) Bagno, A.; Rastrelli, F.; Saielli, G. Prediction of the 1H and 13C NMR Spectra of α-D-Glucose in Water by DFT Methods and MD Simulations. J. Org. Chem. 2007, 72 (19), 7373–7381.

(30) Naganagowda, G. A.; Gururaja, T. E. L.; Satyanarayana, J.; Levine, M. J. NMR Analysis of Human Salivary Mucin (MUC7) Derived O-Linked Model Glycopeptides: Comparison of Structural Features and Carbohydrate-Peptide Interactions. J. Pept. Res. 1999, 54 (4), 290–310.

(31) Maskevich, A. A.; Stsiapura, V. I.; Kuzmitsky, V. A.; Kuznetsova, I. M.; Povarova, O. I.; Uversky, V. N.; Turoverov, K. K. Spectral Properties of Thioflavin T in Solvents with Different Dielectric Properties and in a Fibril-Incorporated Form. J. Proteome Res. 2007, 6 (4), 1392–1401.

(32) Pourjabbar, B.; Hassani, L.; Sajedi, R. Investigating the Effect of Structural Transition on Aggregation of β-Lactoglobulin. Protein Pept. Lett. 2015, 22 (12), 1089–1097.

(33) Barton, J.; Sebastian Arias, D.; Niyangoda, C.; Borjas, G.; Le, N.; Mohamed, S.; Muschol, M. Kinetic Transition in Amyloid Assembly as a Screening Assay for Oligomer-Selective Dyes. Biomolecules 2019, 9 (10).

(34) Dimitrijevic, N. M.; Takahashi, K.; Jonah, C. D. Visible Absorption Spectra of Crystal Violet in Supercritical Ethane-Methanol Solution. J. Supercrit. Fluids 2002, 24 (2), 153–159.

(35) Yuede, C. M.; Lee, H.; Restivo, J. L.; Davis, T. A.; Hettinger, J. C.; Wallace, C. E.; Young, K. L.; Hayne, M. R.; Bu, G.; Li, C. zhong; Cirrito, J. R. Rapid in Vivo Measurement of SS-Amyloid Reveals Biphasic Clearance Kinetics in an Alzheimer's Mouse Model. J. Exp. Med. 2016, 213 (5), 677–685.

(36) Dong, A.; Huang, P.; Caughey, W. S. Protein Secondary Structures in Water from Second-Derivative Amide I Infrared Spectra. Biochemistry 1990, 29 (13), 3303–3308.

(37) Juszczyk, P.; Kolodziejczyk, A. S.; Grzonka, Z. FTIR Spectroscopic Studies on Aggregation Process of the β-Amyloid 11-28 Fragment and Its Variants. J. Pept. Sci. 2009, 15 (1), 23–29.

(38) Byler, D. M.; Susi, H. Examination of the Secondary Structure of Proteins by Deconvolved FTIR Spectra. Biopolymers 1986, 25 (3), 469–487.

SECTION

2. CONCLUSIONS AND FUTURE WORK

2.1. CONCLUSIONS

Protein stability is of great concern not only to sustain life but also for the biopharmaceutical business, which uses live organisms to make protein-based therapeutics and vaccines. This dissertation is a humble effort to better understand and control protein stability. We recognize that the lack of stability may translate into different phenomena ranging from protein aggregation to loss of biological activity caused by subtle changes in the protein's three-dimensional shape. Because of that, our studies ranged from experiments probing molecular level phenomena to experiments at the macroscopic scale. The molecule that we selected for the molecular level experiments was human serum albumin (HSA) because it is abundant, cheap, and well characterized. The molecule(s) chosen for the macroscopic-scale experiments were the amylogenic regions of bovine and human insulin. That choice was our attempt to prove the universality of the aggregation process as well as testing means to inhibit it. The molecular level experiments consisted of using SANS to detect changes in structure and in the intermolecular forces of HSA molecules as the environment was changed from physiological conditions to situations of stress. The macroscopic experiments consisted of inducing the formation of aggregates for the two fragments and monitoring the kinetics of aggregation using various experimental techniques. In the following paragraphs, the main conclusions of the three chapters of this dissertation are discussed in more detail.

In Paper I, the effect of cosolvents, pH and temperature on HSA conformation and protein-protein interactions was explored using SANS. The shape of HSA changed from a prolate to an oblate ellipsoid in the absence of intermolecular interactions with an increase in protein concentration. The addition methanol and glycerol affected the strength and range of intermolecular interactions. Methanol makes the solution of HSA more gas-like (disordered) whereas the addition of glycerol makes the solution more liquid-like (ordered). The effect of temperature on the intermolecular potential was negligible.

In Paper II, two peptide fragments of insulin A-chain with subtle differences in their primary structure were synthesized and their fibrillation propensity was studied using fluorescence and infrared spectroscopy. The presence of an Fmoc group attached to the peptide has an impact on the fibrillation kinetics, which is dependent on pH and temperature. We found that in the presence of Fmoc, the bovine insulin fragments have a shorter lag time than the human insulin ones but, for most samples, the exponential phase rate was faster for HIF than for BIF. An increase in β-sheets content with time was observed in both fragments. The conversion of oligomer structures rich in anti-parallel β-sheets to fibrils rich in parallel β-sheets is a characteristic of amyloid formation.

In Paper III, the fragments of insulin A-chain BIF and HIF were successfully glycosylated using α-D-glucose. The glycosylated peptides (gBIF and gHIF) were highly stable and they did not show signs of fibril formation during the course of the experiment. Some changes in the secondary structure were observed. In the absence of Fmoc, HIF has a shorter lag time than BIF. Turbidity measurements showed a relatively more complex kinetic than ThT measurements. A two phases kinetics was observed in the presence and

absence of DTT. A clear progress in the fibril formation from monomer to globular oligomers to curvilinear fibrils and finally rigid fibrils was observed in both the peptides.

2.2. FUTURE WORK

Several questions were answered in this dissertation but an equal number has arose as a consequence of this work. A better characterization of the intermediates formed during the aggregation process as well as other means of inhibiting the formation of fibrils are worth to be pursued. The use of a β-sheet specific dye like ThT or Congo Red helped in monitoring the kinetics of fibril formation but they do not provide the same insight into the process as oligomer-specific assays. Also, visual monitoring of the aggregates in real time would advance the understanding of the fibril formation process. Finally, a better quantification of the sizes and shape of these aggregates is needed to better understand the difference between amorphous and amyloid aggregates.

Techniques like Nano IR spectroscopy, flow cytometry, multiple angle light scattering (MALS) and X-ray diffraction (XRD) help in differentiating amyloid aggregates from amorphous ones. Nano IR, a combination of atomic force microscopy and infra-red spectroscopy can help in analyzing the type of aggregates in real time [37]. Flow cytometry can characterize protein aggregates of different size and composition. A MALS instrument coupled with HPLC system can determine absolute molar mass, size and shape of the aggregates directly [38]. Finally, XRD in combination with solid state NMR can help in identifying the atomic resolution structures of amyloid fibrils [39,40].

Our results showed that glycopeptides of both insulin fragments typically had 2-4 sugars attached to the peptide. Glycopeptides seem to be stable but the effect of the number

of carbohydrate moieties on each peptide was not explored. Moreover, the type of sugar and the type of anomeric configuration of each sugar have a potential impact on the fibril formation kinetics. Exploring these factors with a comprehensive experimental design in combination with selective glycosylation of the peptides along with the use of the new techniques mentioned in the previous paragraph can take the study to a whole new level and make substantial progress in the control of protein aggregation.

Phosphorylation and lipidation are other important posttranslational modifications that can effect amyloidogenesis. Phosphorylation consists of the addition of phosphoryl groups to the serine amino acid during peptide synthesis and lipidation consists of the attachment of aliphatic hydrocarbons typically on a cysteine residue. Literature shows that phosphorylation [41,42] and lipidation [43] can affect the molecular stability of β-amyloid deposits. Exploring these two posttranslational modifications along with glycosylation would have an impact in the understanding (and prevention) of amyloidogenesis at a molecular level.

BIBLIOGRAPHY

(1) Deller, M. C.; Kong, L.; Rupp, B. Protein Stability: A Crystallographer's Perspective. Acta Crystallographica Section:F Structural Biology Communications. International Union of Crystallography 2016, pp 72–95.

(2) Gromiha, M. M. Protein Stability. In Protein Bioinformatics; Elsevier, 2010; pp 209–245.

(3) Nick Pace, C.; Scholtz, J. M.; Grimsley, G. R. Forces Stabilizing Proteins. FEBS Lett. 2014, 588 (14), 2177–2184.

(4) Lesser, G. J.; Rose, G. D. Hydrophobicity of Amino Acid Subgroups in Proteins. Proteins Struct. Funct. Bioinforma. 1990, 8 (1), 6–13.

(5) Myers, J. K.; Pace, C. N. Hydrogen Bonding Stabilizes Globular Proteins. Biophys. J. 1996, 71 (4), 2033–2039.

(6) Dombkowski, A. A.; Sultana, K. Z.; Craig, D. B. Protein Disulfide Engineering. FEBS Letters. No longer published by Elsevier January 21, 2014, pp 206–212.

(7) Li, R.; Wu, Z.; Wangb, Y.; Ding, L.; Wang, Y. Role of PH-Induced Structural Change in Protein Aggregation in Foam Fractionation of Bovine Serum Albumin. Biotechnol. Reports 2016, 9, 46–52.

(8) Thomas, T.; Cavicchioli, R. Effect of Temperature on Stability and Activity of Elongation Factor 2 Proteins from Antarctic and Thermophilic Methanogens. J. Bacteriol. 2000, 182 (5), 1328–1332.

(9) Przybycien, T. M.; Bailey, J. E. Secondary Structure Perturbations in Salt-Induced Protein Precipitates. Biochim. Biophys. Acta 1991, 1076 (1), 103–111.

(10) Allen, S. J.; Schwartz, A. M.; Bush, M. F. Effects of Polarity on the Structures and Charge States of Native-like Proteins and Protein Complexes in the Gas Phase. Anal. Chem. 2013, 85 (24), 12055–12061.

(11) Gonnelli, M.; Strambini, G. B. Glycerol Effects on Protein Flexibility: A Tryptophan Phosphorescence Study. Biophys. J. 1993, 65 (1), 131–137.

(12) Canchi, D. R.; García, A. E. Cosolvent Effects on Protein Stability. Annu. Rev. Phys. Chem. 2013, 64 (1), 273–293.

(13) Patel, R.; Kumari, M.; Khan, A. B. Recent Advances in the Applications of Ionic Liquids in Protein Stability and Activity: A Review. Applied Biochemistry and Biotechnology. Humana Press Inc. March 6, 2014, pp 3701–3720.

(14) Gordon, C. M. New Developments in Catalysis Using Ionic Liquids. Appl. Catal. A Gen. 2001, 222 (1–2), 101–117.

(15) Satish, L.; Millan, S.; Sahoo, H. Sustained Activity and Stability of Lysozyme in Aqueous Ionic Liquid Solutions Containing Carboxymethylcellulose and Polyethylene Glycol. J. Mol. Liq. 2019, 278, 329–334.

(16) Azuaga, A. I.; Dobson, C. M.; Mateo, P. L.; Conejero-Lara, F. Unfolding and Aggregation during the Thermal Denaturation of Streptokinase. Eur. J. Biochem. 2002, 269 (16), 4121–4133.

(17) Dong, A.; Prestrelski, S. J.; Allison, S. D.; Carpenter, J. F. Infrared Spectroscopic Studies of Lyophilization- and Temperature-induced Protein Aggregation. J. Pharm. Sci. 1995, 84 (4), 415–424.

(18) Pace, C. N.; Grimsley, G. R.; Scholtz, J. M. Protein Ionizable Groups: PK Values and Their Contribution to Protein Stability and Solubility. Journal of Biological Chemistry. American Society for Biochemistry and Molecular Biology May 15, 2009, pp 13285–13289.

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

(20) de Groot, N. S.; Sabate, R.; Ventura, S. Amyloids in Bacterial Inclusion Bodies. Trends Biochem. Sci. 2009, 34 (8), 408–416.

(21) Galzitskaya, O. V; Galushko, E. I.; Selivanova, O. M. Studies of the Process of Amyloid Formation by Aβ Peptide. Biochem. 2018, 83 (S1), S62–S80.

(22) Selivanova, O. M.; Surin, A. K.; Marchenkov, V. V.; Dzhus, U. F.; Grigorashvili, E. I.; Suvorina, M. Y.; Glyakina, A. V.; Dovidchenko, N. V.; Galzitskaya, O. V. The Mechanism Underlying Amyloid Polymorphism Is Opened for Alzheimer's Disease Amyloid-β Peptide. J. Alzheimer's Dis. 2016, 54 (2), 821–830.

(23) Galzitskaya, O. V.; Selivanova, O. M. Rosetta Stone for Amyloid Fibrils: The Key Role of Ring-Like Oligomers in Amyloidogenesis. J. Alzheimer's Dis. 2017, 59 (3), 785–795.

(24) Selivanova, O. M.; Glyakina, A. V.; Gorbunova, E. Y.; Mustaeva, L. G.; Suvorina, M. Y.; Grigorashvili, E. I.; Nikulin, A. D.; Dovidchenko, N. V.; Rekstina, V. V.; Kalebina, T. S.; Surin, A. K.; Galzitskaya, O. V. Structural Model of Amyloid Fibrils for Amyloidogenic Peptide from Bgl2p–Glucantransferase of S. Cerevisiae Cell Wall and Its Modifying Analog. New Morphology of Amyloid Fibrils. Biochim. Biophys. Acta - Proteins Proteomics 2016, 1864 (11), 1489–1499.

(25) Selivanova, O. M.; Surin, A. K.; Ryzhykau, Y. L.; Glyakina, A. V.; Suvorina, M. Y.; Kuklin, A. I.; Rogachevsky, V. V.; Galzitskaya, O. V. To Be Fibrils or to Be Nanofilms? Oligomers Are Building Blocks for Fibril and Nanofilm Formation of Fragments of Aβ Peptide. Langmuir 2018, 34 (6), 2332–2343.

(26) McBride, S. A.; Sanford, S. P.; Lopez, J. M.; Hirsa, A. H. Shear-Induced Amyloid Fibrillization: The Role of Inertia. Soft Matter 2016, 12 (14), 3461–3467.

(27) Vaquer-Alicea, J.; Diamond, M. I. Propagation of Protein Aggregation in Neurodegenerative Diseases. Annu. Rev. Biochem. 2019, 88 (1), 785–810.

(28) Fändrich, M.; Fletcher, M. A.; Dobson, C. M. Amyloid Fibrils from Muscle Myoglobin. Nature 2001, 410 (6825), 165–166.

(29) Maji, S. K.; Perrin, M. H.; Sawaya, M. R.; Jessberger, S.; Vadodaria, K.; Rissman, R. A.; Singru, P. S.; Nilsson, K. P. R.; Simon, R.; Schubert, D.; Eisenberg, D.; Rivier, J.; Sawchenko, P.; Vale, W.; Riek, R. Functional Amyloids as Natural Storage of Peptide Hormones in Pituitary Secretory Granules. Science 2009, 325 (5938), 328–332.

(30) Bermudez, O.; Forciniti, D. Aggregation and Denaturation of Antibodies: A Capillary Electrophoresis, Dynamic Light Scattering, and Aqueous Two-Phase Partitioning Study. In Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences; 2004; Vol. 807, pp 17–24.

(31) Png, G. M.; Falconer, R. J.; Abbott, D. Tracking Aggregation and Fibrillation of Globular Proteins Using Terahertz and Far-Infrared Spectroscopies. IEEE Trans. Terahertz Sci. Technol. 2016, 6 (1), 45–53.

(32) Krebs, M. R. .; Wilkins, D. K.; Chung, E. W.; Pitkeathly, M. C.; Chamberlain, A. K.; Zurdo, J.; Robinson, C. V; Dobson, C. M. Formation and Seeding of Amyloid Fibrils from Wild-Type Hen Lysozyme and a Peptide Fragment from the β-Domain. J. Mol. Biol. 2000, 300 (3), 541–549.

(33) Shental-Bechor, D.; Levy, Y. Folding of Glycoproteins: Toward Understanding the Biophysics of the Glycosylation Code. Current Opinion in Structural Biology. October 2009, pp 524–533.

(34) Hu, J.; Guo, Y.; Li, Y. Research Progress in Protein Post-Translational Modification. Chinese Sci. Bull. 2006, 51 (6), 633–645.

(35) Ogasawara, M.; Katsumata, T.; Egi, M. Taste Properties of Maillard-Reaction Products Prepared from 1000 to 5000 Da Peptide. Food Chem. 2006, 99 (3), 600–604.

(36) Cudic, M.; Ertl, H. C. J.; Otvos, L. Synthesis, Conformation and T-Helper Cell Stimulation of an O-Linked Glycopeptide Epitope Containing Extended Carbohydrate Side-Chains. Bioorganic Med. Chem. 2002, 10 (12), 3859–3870.

(37) Ruggeri, F. S.; Longo, G.; Faggiano, S.; Lipiec, E.; Pastore, A.; Dietler, G. Infrared Nanospectroscopy Characterization of Oligomeric and Fibrillar Aggregates during Amyloid Formation. Nat. Commun. 2015, 6, 7831.

(38) Gentiluomo, L.; Schneider, V.; Roessner, D.; Frieß, W. Coupling Multi-Angle Light Scattering to Reverse-Phase Ultra-High-Pressure Chromatography (RP-UPLC-MALS) for the Characterization Monoclonal Antibodies. Sci. Rep. 2019, 9 (1), 1–8.

(39) Nielsen, J. T.; Bjerring, M.; Jeppesen, M. D.; Pedersen, R. O.; Pedersen, J. M.; Hein, K. L.; Vosegaard, T.; Skrydstrup, T.; Otzen, D. E.; Nielsen, N. C. Unique Identification of Supramolecular Structures in Amyloid Fibrils by Solid-State NMR Spectroscopy. Angew. Chemie - Int. Ed. 2009, 48 (12), 2118–2121.

(40) Morris, K. L.; Serpell, L. C. X-Ray Fibre Diffraction Studies of Amyloid Fibrils. Methods Mol. Biol. 2012, 849, 121–135.

(41) Liang, F. C.; Chen, R. P. Y.; Lin, C. C.; Huang, K. T.; Chan, S. I. Tuning the Conformation Properties of a Peptide by Glycosylation and Phosphorylation. Biochem. Biophys. Res. Commun. 2006, 342 (2), 482–488.

(42) Rezaei-Ghaleh, N.; Amininasab, M.; Kumar, S.; Walter, J.; Zweckstetter, M. Phosphorylation Modifies the Molecular Stability of β-Amyloid Deposits. Nat. Commun. 2016, 7 (1), 1–9.

(43) Dal Magro, R.; Simonelli, S.; Cox, A.; Formicola, B.; Corti, R.; Cassina, V.; Nardo, L.; Mantegazza, F.; Salerno, D.; Grasso, G.; Deriu, M. A.; Danani, A.; Calabresi, L.; Re, F. The Extent of Human Apolipoprotein A-I Lipidation Strongly Affects the β-Amyloid Efflux Across the Blood-Brain Barrier in Vitro. Front. Neurosci. 2019, 13 (MAY), 419.

VITA

Paul Praveen Nakka was born in Tenali, Andhra Pradesh, India. In 2009, Paul relocated to Hyderabad, India to complete his bachelor's degree in chemical engineering. Upon securing his bachelor's degree in 2013, Paul decided to expand his knowledge and acquire skills to conduct his own research. He moved to United States in 2014 and got admitted into Missouri University of Science and Technology, Rolla, MO. He pursued his doctoral studies in the department of Chemical Engineering under Dr. Daniel Forciniti and obtained his PhD in Chemical Engineering in August 2020.