ABSTRACT

DETERMINATION OF THE SOLUTION STRUCTURE OF ANTIFREEZE GLYCOPROTEIN FRACTION 8 (AFGP8) IN DEUTERATED DIMETHYL SULFOXIDE (DMSO) USING NUCLEAR MAGNETIC RESONANCE (NMR) SPECTROSCOPY

Fish native to the polar regions produce biological antifreezes called antifreeze glycoproteins (AFGPs) to prevent themselves from freezing at supercool temperatures (-1.9˚C). AFGP lowers the freezing point of water in a *non-colligative manner*. AFGP's ability to induce thermal hysteresis (change in freezing point with little to no effect on the melting point) makes them about 500 times more effective than traditional colligative antifreeze, such as salts and sugars. By understanding the antifreeze mechanism of AFGP, it is possible to model molecules that can mimic AFGP. The structure-function paradigm states that the function of a protein can be related to its structure. The primary sequence of AFGP is highly degenerate consisting of multiple repeats of the same tripeptide Ala-Ala-Thr*, in which Thr* is glycosylated with the disaccharide *beta-D-galactosyl-(1,3)-alpha-N-acetyl-D-galactosamine*. The smallest number of tripeptide repeats called AFGP fraction 8 (AFGP8) was studied. In addition to its highly degenerate primary sequence, AFGP seems to function as an intrinsically disordered protein which presents challenges in determining their native structure. In this thesis work, a different approach was used to elucidate the three-dimensional structure of AFGP8 from Arctic cod *Boreogadus saida* and Antarctic notothenioid *Trematomus borchgrevinki*. Dimethyl sulfoxide, a non-native solvent, was used to make AFGP8 less dynamic in solution and induced a non-native structure which was determined via nuclear magnetic resonance (NMR)

spectroscopy. The overall three-dimensional structure between the two AFGP8, from two different natural sources, was different, but their “compactness” were very similar due to the same diffusion coefficient measured with NMR spectroscopy. In addition to the similar compactness, the conserved motifs, Ala-Thr*-Pro-Ala and Ala-Thr*-Ala-Ala, presented in both AFGP8 seemed to have very similar three-dimensional structure. The determination of the three-dimensional structure of AFGP8 in dimethyl sulfoxide sets a foundation for a future experiment in performing a solvent titration scheme. The solvent titration scheme will allow for tracking structural changes of the non-native structure of AFGP (in DMSO) to the native, disordered structure of AFGP (in water) in hope that the perturbation of the structure will provide insight on the antifreeze mechanism of an antifreeze glycoprotein.

Cheenou Her
May 2018

DETERMINATION OF THE SOLUTION STRUCTURE OF ANTIFREEZE GLYCOPROTEIN FRACTION 8 (AFGP8) IN DEUTERATED DIMETHYL SULFOXIDE (DMSO) USING NUCLEAR MAGNETIC RESONANCE (NMR) SPECTROSCOPY

by

Cheenou Her

A thesis

submitted in partial

fulfillment of the requirements for the degree of

Master of Science in Chemistry

in the College of Science and Mathematics

California State University, Fresno

May 2018

APPROVED

For the Department of Chemistry:

We, the undersigned, certify that the thesis of the following student meets the required standards of scholarship, format, and style of the university and the student's graduate degree program for the awarding of the master's degree.

Cheenou Her
Thesis Author

Viswanathan V. Krishnan (Chair) Chemistry

Santanu Maitra Chemistry

Cory L. Brooks Chemistry

For the University Graduate Committee:

Dean, Division of Graduate Studies

AUTHORIZATION FOR REPRODUCTION

OF MASTER'S THESIS

X I grant permission for the reproduction of this thesis in part or in its entirety without further authorization from me, on the condition that the person or agency requesting reproduction absorbs the cost and provides proper acknowledgment of authorship.

_____ Permission to reproduce this thesis in part or in its entirety must be obtained from me.

Signature of thesis author:_____

ACKNOWLEDGMENTS

I thank Dr. Krishnan for giving me the opportunity to do my thesis research under his guidance. Also, I am extremely grateful for Dr. Krishnan for the support and advice throughout my thesis research and career at California State University, Fresno. I have learned so much researching in your group and enjoy my time as part of the Krish Research Group. THANK YOU Dr. Krishnan.

I thank Dr. Yin Yeh from UC Davis for providing antifreeze glycoproteins. Without you, this thesis research would not have been possible.

I thank Dr. Andrew H. Lane and Dr. Shin-Ichiro Nishimura for providing their solution structure of AFGP for comparison in this thesis research.

To my committee members, Dr. Cory L. Brooks and Dr. Santanu Maitra, I appreciate and thank you for your advices and discussions with me from time to time whether it's about my thesis, my career in science, or just random thoughts.

To Dr. Alam Hasson, Dr. Joy J. Goto, Dr. Jai-Pil Choi, and Dr. Kalyani Maitra, I owe you my deepest gratitude for helping in my transition into the Chemistry Graduate program here at California State University, Fresno.

As for my friends: Justin Vang, Jaideep Singh, Candice C. Cortney, Candido Breceda, Angham Ahmed, Catalina Olea, Ryan Dougherty, Pedro Diaz-Parga, and Aaron Alonso I thank you and value your friendship very much.

I thank Salvador Vazquez for helping with the methylation of AFGP.

From the bottom of my heart, I thank my family and friends. They are the unsung heroes that made it possible for me to continue my education.

To my mom, Yeng Moua, and dad, Seng Her, I am extremely grateful for the support that you provided and thank you for always being there. **"You're the Real MVP."**

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

Page

Page

I. INTRODUCTION

I.1 Intrinsically Disordered Proteins

The structure-function paradigm linked a well-defined, three-dimensional structure to the function of the protein. This idea was the leading application toward studying proteins until the discovery that some natural proteins are not folded into well-defined, three-dimensional structures and yet still functional. These proteins can contain disordered regions or are entirely disordered. The disorder is crucial for their biological function.[1] Proteins that are disordered are called intrinsically disordered proteins (IDP).[1] It is expected that IDP does not adopt a well-defined three-dimensional fold (tertiary or quaternary structure), but IDP might adopt and interchange between various secondary structures for very brief moments.[2] One proposal for describing IDP is that IDP can exist as multiple different conformations while in solution due to the low energy barriers between the different possible conformations. With the low free-energy barriers between the multiple structures, IDP can rapidly interconvert between the different conformations.[3] This dynamic behavior of IDP present one of the most challenging barrier in studying and understanding the structure of IDP using the currently available structural determination techniques.[1]

I.2. Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a physical technique that can use to determine the structure and dynamics of proteins in the solution state. Protein structural determination using NMR spectroscopy is possible because all proteins are made up of atoms, such as hydrogen, carbon, nitrogen, and oxygen, which are NMR active nuclei.[1] NMR active nuclei are any nucleus that has a non-zero spin quantum number (e.g., 1/2, 1, 3/2, etc.) which allows the

nucleus to be perturbed with an external magnetic field.[4-5] In the absence of an external magnetic field, the energy level of the two possible spin states of a nuclear spin, spin up (α) state and spin down (β) state, are degenerate.[4-5] However, in the presence of an applied magnetic field, the energy level of the α and β states are not degenerate. The α state is lower in energy due to this spin state aligning with the applied magnetic field, and the β states is higher in energy due to this spin aligning against the applied magnetic field. The nuclear spin of a single nucleus can be in either of this two spin states. In an experimental study, the number of nuclei involved in the experiment is close to Avogadro's number (6.022×10^{23} nucleus). The distribution of the nuclei favors that slightly more than 50% of the population is in the α states and slightly less than 50% of the population is in the β states.[4-5] This little difference in population makes it possible to detect the energy difference between the α and β states corresponding to the resonating frequency of the nucleus.

Under the influence of a magnetic field, each type of NMR active nucleus resonates at its unique Larmor frequency; the Larmor frequency is the precession rate of a nucleus within a given magnetic field strength.[4-5] While the same type of nuclei has the same Larmor frequency, the chemical environment, particularly the electron density of nearby nuclei, influences the resonating frequency that is detected. The nearby nuclei act as magnetic fields that are very localized affecting any nuclei that are in the immediate vicinity resulting in the same type of nuclei resonating at a frequency slightly different from the Larmor frequency.

These changes in resonating frequency are commonly observed in a molecule. Within a molecule, the same type of nuclei can experience different chemical environments as a result of the molecular structure of the molecule. The different chemical environments within the molecule influence each nucleus to

resonate at a slightly different frequency from the Larmor frequency. The resonating frequency of each nucleus is converted into a normalized value called chemical shift (δ) value in the unit of parts per million (ppm) using equation (1) below.[4-5] The normalization makes it so that the actual chemical shift of a nucleus

$$\delta(ppm) = \frac{\nu - \nu_{ref}}{\nu_{ref}} \times 10^6 \qquad (1)$$

would be the same across all NMR spectrometer frequencies.
(ν is the frequency of a particular nucleus, and ν_{ref} is the spectrometer frequency.)

I.2.A. Typical Chemical Shift Ranges for Proton (^{1}H) and Carbon-13 (^{13}C) in Protein

Proteins are mainly made up of hydrogen (proton, ^{1}H), carbon, nitrogen, and oxygen atoms, which at least one isotope form of each is NMR active. Proton is the most commonly used nucleus due to its natural abundance. The typical chemical shift range for the proton is about 0-10 ppm.[4] For protein studies, the proton chemical shift range is categorized into three main regions: amide proton region (~6-9 ppm), alpha proton region (~3-5 ppm), and methyl protons region (~0.5-2 ppm).[6] Another nucleus, in natural abundance, used in protein structural studies is the carbon-13 isotope (^{13}C), but the major setback for ^{13}C is it has a natural abundance of one percent. Unlike the proton chemical shift range, the ^{13}C chemical shift has a wider range going from about 0-220 ppm.[4] Similar to a proton, in protein studies, the common ^{13}C chemical shift regions are carbonyl carbon (~180-200 ppm), alpha carbon (~40-60 ppm), and methyl carbon (~10-20 ppm). Except for the methyl proton/carbon regions, all chemical shift regions mentioned focus primarily on the chemical shift that each nucleus would resonate at regarding the protein backbone.

II. ANTIFREEZE GLYCOPROTEINS (AFGP) AND ANTIFREEZE PROTEINS (AFP)

Fish that live in the polar regions experience temperature (~-1.8 °C) well below the freezing point of their blood serum (~-1.0 °C).[7-8] Through evolution, the fish developed a defense against ice nucleation in the forms of biological antifreezes called antifreeze protein (AFP) and antifreeze glycoprotein (AFGP), which inhibit the growth of ice crystal within the fishes.[9-10] The biological antifreezes were divided into two groups, non-glycosylated (AFP) and glycosylated (AFGP).[7, 11] The natural sources of AFGP are from two polar species of fish, the Arctic cod and Antarctic notothenioid.[8] AFP, on the other hand, is found in both fish and other organisms such as amphibians, plants and insects.[12]

AFP and AFGP cause a freezing point depression of water in a non-colligative manner unlike traditional antifreeze such as salts, sugars, and polyol, which act in a colligative manner. Using the colligative freezing point definition, the physiological concentration (~10-30 mg/mL) of AFGP within the fish exhibit a freezing point depression about 500 times than what is expected in terms of molal concentration.[11] Because of their non-colligative property, AFP and AFGP exhibit a unique phenomenon called thermal hysteresis; thermal hysteresis is a difference between the melting point and freezing point. The thermal hysteresis exhibit by AFGP ranges between 1-2 °C below the normal melting/freezing point (0 °C) of water.[7] Since AFP is diversely distributed between among different organisms, which lives on land and/or water, the thermal hysteresis has a wider range from 1 °C to about 7 °C below that of the normal freezing point (0 °C) of water.[13-14]

In the presence of AFP and AFGP, water freezes, forming ice crystal, at a lower temperature (hysteresis freezing point) than 0 °C but the ice crystal would melt at a higher temperature (hysteresis melting point) compared to the hysteresis

freezing point. While AFP and AFGP were known to affects the freezing point of water, it was thought that AFP and AFGP do not affect the melting point of water. It was previously suggested and shown by Knight and DeVries[15] that AFGP inhibits the melting of ice crystal at the normal melting temperature of water at 0 °C. As a result, a higher temperature than 0 °C was needed to completely melt all the ice crystals formed during the cooling process. As for AFP, it was currently shown that AFP exhibits the same melting inhibition of ice crystal as AFGP.[16-17] AFP and AFGP inhibit both the freezing and melting process of ice crystal, but the freezing point depression (difference between the normal freezing point and hysteresis freezing point) is greater than the melting point elevation (difference between the normal melting point and hysteresis melting point).[15-17]

In addition to the thermal hysteresis observed, the presence of AFP and AFGP also structure the ice crystal differently as compared to the structure of normal ice crystal at ambient condition.[7, 18] Pure water, upon cooling, will form hexagonal shape ice crystal; this is considered the normal structure of ice crystal at ambient condition.[19] In the presence of AFP and AFGP, upon cooling, water will form bipyramidal shape ice crystal.[7, 18] Upon further cooling below the thermal hysteresis gap, rapid ice crystal growth will occur resulting in the ice crystals to resemble needle-like structure.[7]

AFP and AFGP are structurally diverse but exhibits common properties, such as ice crystal restructuring and thermal hysteresis effect. The thermal hysteresis observed is the result of AFP and AFGP binding onto the ice surface affecting the way ice crystal normally grows thus resulted in the proposed adsorption inhibition model.[20-22] The basis for explaining the adsorption inhibition model is the Kelvin effect.[11, 20] Upon binding to the ice surface, AFP and AFGP restructure the ice surface resulting in a curvature surface instead of a flat surface

thus promoting the ice crystals to grow in a way that is unfavorable within the thermal hysteresis gap.[11, 15, 22] For the ice crystals to grow in that fashion, a lower temperature will be needed.

A different mechanism for ice crystal growth inhibition, other than adsorption inhibition, was proposed by recent studies using Terahertz spectroscopy on AFP and AFGP.[23-25] Terahertz absorption spectroscopy can be used to study the dynamics and interactions between biomolecules and the solvent (water molecules). Therefore, it can be used to probe how AFP and AFGP interact with the water molecules upon cooling. Both AFP and AFGP were predicted to have a larger hydration layer than that of normal protein due to their antifreeze properties. Experimentally, the hydration shell of AFP and AFGP increases in size upon lowering the temperature of the AFP and AFGP solutions.[23, 26] The findings lead to the proposal that AFP and AFGP inhibit the ice crystal growth via long-range perturbation of the bulk water molecules. The long-range perturbation of the water molecules enables AFP and AFGP to disturb the water molecules at a longer distance than normal proteins, breaking the hydrogen bonds necessary for addition of water molecules to the ice crystal surface for ice crystal growth. At supercool temperature, the hydration layers of AFP and AFGP at physiological concentration were predicted to be large enough to cover the entire solution thus allowing for complete perturbation of the entire hydrogen-bonding network of the water molecules.

Even though both proposed antifreeze mechanisms have theoretical and experimental data to support them, neither has been the accepted explanation for the antifreeze mechanism because each mechanism does not fully explain the observed antifreeze effect. For the adsorption inhibition model, it was shown by Brown et al.[27] that AFGP does adsorb onto the ice surface, but it is still unclear on

whether the adsorption is reversible or not, as discussed in the review by Yeh and Feeney[11]. The recent mechanism proposed by Terahertz absorption spectroscopy did explain how AFP and AFGP can inhibit ice crystal growth, but does not include explanation for why do ice the ice crystals grow differently compared to normal ice.[23, 26] Also, AFP and AFGP are structurally diverse, so that also brought up the question of if the antifreeze mechanism of AFP and AFGP can be explained with one common mechanism.

One key difference between AFP and AFGP is that AFP is known to have well-defined three-dimensional structure in solution while the three-dimensional structure of AFGP is still undetermined.[28-31] There are four different types of AFP found in fish; each type of AFP possesses a unique trait that structurally differentiate them from one another.[8] Type 1 AFP is an alanine rich α-helix protein with 4-5 threonine residues with 11 residues in between spacing each threonine residue apart.[28] Type 2 AFP possess cysteine residues which form disulfide bonds and can also be classified as calcium-dependent or calcium-independent depending on if a Ca^{2+} binding site is needed in order to funtion.[29] Type 3 AFP is the structure of a β-sandwich, consisting of multiple beta sheets, and does not have a large unequal distribution of a single amino acid like type 1 AFP.[30] Type 4 AFP is also an alanine rich, helical bundle AFP.[8]

The primary sequence of AFGP, shown in **Figure 1**, consists of a series repeating tripeptide sequence $(Ala\text{-}Ala\text{-}Thr^*)_n$ in which the Thr* is glycosylated on the hydroxyl group with the disaccharide *beta-D-galactosyl-(1,3)-alpha-N-acetyl-D-galactosamine*. There are eight distinct fractions (isoforms) of AFGP identified as AFGP1-8; each fraction is not referring to one distinct molecular mass but a mixture of molecular masses.[8] For each fraction, the number of repeating tripeptide varies from 4 to 50 repeating units, making the four repeating

units AFGP8 (2.6 kDa) the shortest and the 50 repeating units AFGP1 (33.7 kDa) the longest.[7, 12] In the shorter AFGP fraction such as AFGP6-8, there are minor variations in the primary sequence $(Ala\text{-}Ala\text{-}Thr^*)_n$ in which the alanine at certain positions is replaced with a proline.[32] The minor variations are specific to the natural sources, the species of fish.[32] Of all the AFGP fractions, AFGP1 is the most effective and AFGP8 is the least effective at inhibiting ice crystal growth.[8] One speculation for this trend is the number of repeating tripeptide; with more repeating tripeptide, AFGP1 can occupy a larger surface volume of the ice crystal than AFGP8 thus inhibiting the ice crystal growth better per AFGP molecule.

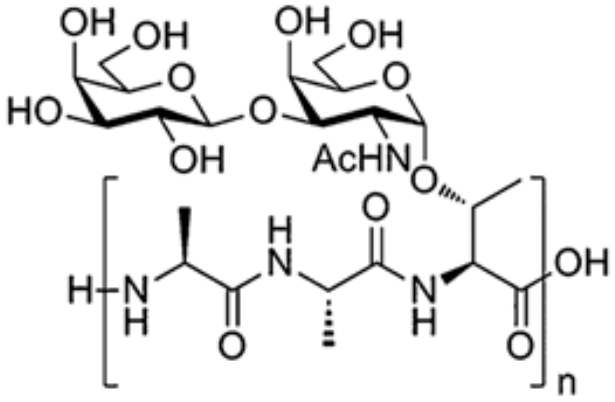


Figure 1. The repeating tripeptide $(Ala\text{-}Ala\text{-}Thr^*)_n$ of antifreeze glycoproteins (AFGP). Thr* is glycosylated with the disaccharide *beta-D-galactosyl-(1,3)-alpha-N-acetyl-D-galactosamine*.

Both AFP and AFGP exhibit thermal hysteresis and control ice crystal growth, but the antifreeze mechanism of both is still not well known. Since AFP have well-defined structures determined by both X-ray crystallography and nuclear magnetic resonance spectroscopy, the proposed antifreeze mechanisms of AFP have better experimental support as compared to AFGP. As an example, Type 1 AFP is the most well-studied biological antifreeze out of both AFP and AFGP. It is known that Type 1 AFP is an alpha helix structure with 4-5 threonine

residues spacing out by 11 other amino acid residues.[28, 33] This three-dimensional structure leads to all the threonine residue facing one side resulting in Type 1 AFP to have a hydrophobic side and a hydrophilic side. The distance between the hydroxyl groups of the threonine residues is similar to the distance of certain segment of water molecules on the ice crystal surface. The similarity in the distance leads to the proposal that Type 1 AFP function by binding to the ice surface via the threonine residues and exposing the backbone and hydrophobic residues toward the bulk water.[33] Additions of the bulk water molecules to the ice crystals are hindered thus inhibits further growth of the ice crystals. The proposed mechanism makes use of the amphiphilic nature, containing hydrophilic and hydrophobic residues, that is present in AFP and AFGP. Also, the proposed mechanism is a perfect example of the structure function paradigm. From the three-dimensional structure, a mechanism was predicted that might explain how Type 1 AFP functions. By determining the three-dimensional structure of AFGP, one might be able to propose an antifreeze mechanism that may explain how the glycosylated biological antifreeze, AFGP, function.

II.1. Aim: Structure of AFGP8 in DMSO-d_6

Even though both AFP and AFGP inhibit ice crystal growth, it is expected that their antifreeze mechanisms could be different due to the glycosylation of AFGP. Unlike AFP, the three-dimensional structure of AFGP has not been determined. The book chapter by Krishnan and Yeh[31] reviewed studies attempting to determine the native structure of AFGP. The lack of consistency in the structures determined lead to the authors proposing that AFGP function as an intrinsically disordered protein.[31] Therefore, the hypothesis is that AFGP function as an intrinsically disordered protein (IDP) in its native environment thus having

the potential to simultaneously bind to the ice crystal surface and interact with the bulk water molecules to inhibit further ice crystal growth.

The high degeneracy in the primary sequence (repeating tri-peptides) of AFGP results in poor spectral resolution and large regions of resonance overlap on the NMR spectrum of AFGP. To minimize the spectral overlaps, the shortest fraction, antifreeze glycoprotein 8 (AFGP8), was used for the structural study. The primary sequence of antifreeze glycoproteins fraction 8 (AFGP8) from two different natural sources, *Boreogadus saida* (BS) and *Trematomus borchgrevinki* (TB), are shown in **Figure 2**.[34] For both types of AFGP8, two alanines are replaced with prolines while all the glycosylated threonine (T*) remain unchanged. AFGP8-BS has its alanine at position 4 and 10 replaced with prolines, and AFGP8-TB has its alanine at position 7 and 13 replaced with prolines. It is important to note that the primary sequences shown in **Figure 2** are the major portion of AFGP fraction 8 because the fractions of AFGP are not one molecular weight but a mixture of molecular weight, due to the possible mutation of certain alanine to proline or threonine to arginine.[8, 34] There are other proline mutations in the smaller portions of AFGP fraction 8 as discussed by Geoghegan et al.[34]

AFGP8	1	2	3	4	5	6	7	8	9	10	11	12	13	14
BS	A	A	T*	P	A	T*	A	A	T*	P	A	T*	A	A
TB	A	A	T*	A	A	T*	P	A	T*	A	A	T*	P	A

Figure 2. The primary sequence of antifreeze glycoprotein fraction 8 (AFGP8) from *Boreogadus saida* (BS) and *Trematomus borchgrevinki* (TB).
Note: The amino acids were color coded as red (alanine, A), blue (glycosylated threonine, T*), and light blue (proline, P). The primary sequence has repeats of AAT* or PAT* highlighted in yellow or blue, respectively. Both BS and TB have two units of AAT* and PAT*, but at different location in the primary sequence. The N-terminus was indicated with the underlined alanine at position 1.

Additional to using the AFGP with the smallest number of repeats, AFGP8, chemically modifying the N-terminus of AFGP with carbon-13 isotope (^{13}C) enriched methyl groups can enhance the sensitivity in the ^{13}C spectrum. Since the natural abundance of the ^{13}C nucleus is about 1%, the ^{13}C-methyl groups (~100% ^{13}C) greatly increased the sensitivity by about ~100 times for ^{13}C NMR spectroscopy. The study by Geoghegan et al. confirmed that the added methyl groups on the N-terminus have little to no effect on the antifreeze property of antifreeze glycoproteins fraction 8 (AFGP8).[34] Since the antifreeze property remained the same or very similar, it was expected that the methyl groups do not cause a significant change in the native structure of AFGP8. The chemically modified AFGP allows for observing any interaction between the ^{13}C-labeled methyl group of the N-terminus with the rest of AFGP8. Being able to detect protons on the N-terminus provides valuable structural and dynamics information on AFGP8 because the amine protons on the N-terminus of peptides and proteins are very difficult to detect in NMR spectroscopy. The first part of the aim focuses on chemically modifying the N-terminus of AFGP8 with ^{13}C enriched methyl groups for sensitivity enhancement.

In combination with the sensitivity enhancement described above, IDP can be studied using a solvent perturbation scheme with NMR spectroscopy.[35] In the solvent perturbation scheme, the non-native structure of a protein is induced using a non-native solvent. The protein then goes through several structural changes using various in-between ratio mixture of the non-native solvent and native solvent before introducing the protein back into the native solvent (no mixture). The titration aims to study a protein going from the denatured state to the native, functional state. By going through this titration process, it will be possible to see how the protein folds or unfolds while going to its functional states. While not as

useful for well-structured protein, IDP or proteins with disordered regions are highly dynamic in its native, solution state which makes it difficult to track how the proteins fold or move using NMR spectroscopy. The titration process will somewhat "slow down" the protein at varying folding/unfolding stages so each stage may be investigated in detail. The second part of the aim is introducing AFGP8 from the Arctic cod *Boreogadus saida* and Antarctic notothenioid *Trematomus borchgrevinki* into dimethyl sulfoxide (DMSO), a non-native solvent, to induce some non-native structures. These non-native structures will result in better chemical shift resolution and less resonance overlaps. The chemical shifts of the protons of AFGP8 can then be assigned in DMSO, and structures can be generated based on the NMR data in DMSO. **Determination of the structure of AFGP8 in DMSO concludes the first step of the solvent perturbation scheme.** Finishing the first step sets the foundation for future studies using the solvent perturbation scheme which involves titration of AFGP8 from pure DMSO (non-native solvent) to pure water (native solvent) following the changes in its three-dimensional structure.

III. MATERIALS AND METHODS

III.1. Antifreeze Glycoproteins Fraction 8 (AFGP8)

III.1.A. NMR Sample Preparation

Both antifreeze glycoproteins fraction 8 (AFGP8) from *Boreogadus saida* and *Trematomus borchgrevinki* were provided by Professor Yin Yeh from UC Davis and used as received. The deuterated dimethyl sulfoxide (DMSO-d_6) was purchased from Sigma Aldrich and used as purchased. For both AFGP8 NMR samples, 18.0 mg of the AFGP8 was dissolved in 600 µL of DMSO-d_6 in an Eppendorf tube, resulting in a concentration of 30.0 mg/mL. The AFGP8 solutions were transferred to an NMR tube. After the solutions were transferred into an NMR tube, the NMR tube was degassed and sealed to perform all the NMR experiments.

The disaccharide *beta-D-galactosyl-(1,3)-alpha-N-acetyl-D-galactosamine* (5 mg, 98 % purity, CAS 3554-90-3, Cat # A152000) was purchased from Toronto Research Chemicals Inc. and used as purchased. Directly into the purchased tube containing the disaccharide, 620 µL of DMSO-d_6 was added. After the disaccharide fully dissolved, 600 µL of the disaccharide solution was transferred into an NMR tube. The concentration of the disaccharide was estimated to be 21 mM.

III.1.B. Carbon-13 (^{13}C) Isotope Enriched Methylation of N-terminus

Reagents: AFGP8-TB was provided by Professor Yin Yeh from UC Davis. Borane Dimethylamine Complex (CAS 74-94-2), Formaldehyde (13C-labeled,

99% atom pure; 20% by weight, CAS 3228-27-1), and dialysis tubing (M.W. cutoff of 2000 Da, Lot # SLBK7102V) were purchased from Sigma Aldrich.

The methylation of AFGP8-TB with two ^{13}C-isotope enriched methyl groups was done following the procedure published by Means and Feeney.[36-37] For every mole of AFGP8-TB, at least two moles of each formaldehyde and borane dimethylamine complex (reducing agent) needs to be used.

The reaction was performed at 4 ˚C in a phosphate buffer with pH 7.4 using Nanopure distilled water. The volumes of the solutions used were small enough (less than 2 mL total) that the reaction was run inside an Eppendorf tube. All solutions were made using the phosphate buffer with pH 7.4. The AFGP8-TB solution (20.2 mg, ~7.77 µmol) and borane dimethylamine complex solution (~98.8 µmol) were made separately. The ^{13}C-isotope enriched (~99%) formaldehyde was used as purchased. For the volume of formaldehyde and borane dimethylamine complex used, 60 µL was added twice totaling to a volume of 120 µL for each (~79.9 µmol for formaldehyde and ~98.8 µmol for borane dimethylamine complex). The borane dimethylamine complex and formaldehyde solutions were added to the AFGP8-TB solution in two parts. The first addition was to start the reaction. After 3 hours of incubation at 4 ˚C, a second addition of the borane dimethylamine complex and formaldehyde was added to the AFGP8-TB solution. The reaction was incubated at 4 ˚C for an additional 21 hours. The Eppendorf tube was gently inverted a few times every couple of hours. After incubating for 21 hours, the products were purified using dialysis. The purification was performed for 48 hours at 4 ˚C, changing the Nanopure distilled water every 8-12 hours.

After purification, the solution inside the dialysis tubing was lyophilized. The final product was a white, flaky powder. The percent yield was not

determined. According to the references[36-37], the percent yield of this reaction is about 95%. The product was dissolved in 600 µL of DMSO-d_6 and transferred into a NMR tube. The success of the methylation was determined using proton and carbon-13 (^{13}C) NMR spectroscopy.

III.2. Nuclear Magnetic Resonance Spectroscopy Experiments

All nuclear magnetic resonance (NMR) spectroscopy experiments were run at 25 ˚C in deuterated dimethyl sulfoxide (DMSO-d_6) unless indicated otherwise.

III.2.A. Two-Dimensional Nuclear Magnetic Resonance Techniques Used

Two-dimensional **d**ouble **q**uantum **f**iltered **c**orrelation spectroscopy (DQFC) is a type of homonuclear **Co**rrelation **S**pectroscop**y** (COSY) that uses ^3J coupled protons to provide structural information on the connectivity, through covalent bonds, of all the protons within a molecule. The "double quantum filtered" filters out any proton signal that does not have a double quantum transition probability, in other words, the signal of protons that are not *J* coupled to any other proton.[38] For example, in a peptide, the amino acid alanine has an amide proton that is ^3J coupled to the alpha proton, and the alpha proton is ^3J coupled to the beta protons. On the DQFC spectrum, there will be a cross-peak (through bond correlation) between the chemical shift of the amide proton to the alpha proton and a cross-peak between the alpha proton and beta protons. The DQFC spectrum will not show a cross-peak between the amide proton and beta protons because the protons are not ^3J coupled or the protons have a long-range *J* coupling constant that cannot be measured.

Two-dimensional (2D) **To**tal **C**orrelation **S**pectroscop**y** (TOCSY) is a homonuclear NMR technique that uses the ^{3}J coupled protons of a molecule to identify the spin system. Spin system is a network of protons in which each proton is ^{3}J coupled to at least one other proton within the network.[39] This NMR technique provides information similar to a two-dimensional correlation spectroscopy (COSY), but the 2D TOCSY can show cross-peak (through bond correlation) of one proton to protons that are more than three covalent bonds distance away within the same spin system. For example, using the amino acid alanine, the DQFC spectrum will not show a cross-peak between the amide proton and beta protons, but the 2D TOCSY spectrum will show a cross-peak between the amide proton and beta protons. In a peptide/protein, each amino acid has a unique spin system that is independent of the other amino acids due to the peptide bonds. The unique spin systems result in unique patterns, in the 2D TOCSY spectrum, that act as fingerprints for the amino acids. In the amide proton to alpha proton and methyl proton region, the 2D TOCSY spectrum exhibits the unique patterns that can be used to identify the amino acids, except proline; the proline within a peptide/protein does not have an amide proton.

Two-Dimensional (2D) **N**uclear **O**verhauser **E**ffect (NOE) **S**pectroscop**y** (NOESY) is a homonuclear NMR technique that provides through-space (NOE) structural information between two protons that are in close proximity (5 Å $\geq$) to each other.[40] The NOE cross peak (through-space correlation) intensity is inversely proportional to the distance of two protons to the six power. Unlike the 2D COSY and TOCSY spectra, the 2D NOESY spectrum shows NOE cross-peaks between both ^{3}J coupled protons and non-^{3}J coupled protons. The distance constraints obtained from the NOE cross-peak serves as the main basis for folding the protein/peptide.

Two-Dimensional (2D) **H**eteronuclear **M**ultiple **Q**uantum **C**oherence (HMQC) Spectroscopy is a heteronuclear NMR technique that provides structural information on the connectivity of the protons with any adjacent NMR active nucleus other than protons.[41] For this thesis, the other NMR active nucleus is carbon-13 (^{13}C).

III.2.B. NMR Instruments

The NMR experiments were performed using the Varian/Agilent VNMRS-400 MHz spectrometer at the Chemistry Department of Fresno State University. All the experiments were performed using a One-NMR probe with a single axis (along-z) pulsed field gradient. Additional experiments were performed at the UC Davis NMR facility using the Varian 600 MHz (triple resonance probe) and Bruker 800 MHz (triple resonance cold probe) spectrometers.

III.2.C. One-Dimensional NMR Experiments

Standard pulse calibration method was used to calculate the 90˚ pulse width for the one-dimensional NMR experiment. The transmitter offset was set to the center of the spectrum. A spectral width that covers all the proton resonances of the sample was used. All other parameters (number of points, relaxation delay, and transients) were adjusted accordingly to obtain each spectrum. **Table 1** shows the parameters used at different spectrometer frequency for each 1D and 2D NMR technique. Below are examples of how the 1D proton and ^{13}C NMR experiments were performed on the 400 MHz NMR spectrometer and similarly on the 600 and 800 MHz NMR spectrometers.

For the 400 MHz proton NMR experiments, a spectral width of 4800 Hz was used over 16384 number of points totaling to an acquisition time of 3.41

seconds. A 1-second relaxation delay was used which resulted in a total time of 4.41 seconds per transient. The NMR data was collected using 256 transients.

For the 400 MHz ^{13}C one-dimensional NMR experiment, a spectral width of 22000 Hz was used over 16384 number of points totaling to an acquisition time of 0.74 seconds. A 1-second relaxation delay was used which resulted in a total time of 1.74 seconds per transient. The NMR data was collected using 64 transients. (See **Table 1** for specific parameters used at different NMR spectrometer frequencies.)

III.2.D. Two-Dimensional NMR Experiments

The 1D NMR parameters (90˚ pulse width, spectral width, transmitter offset, and relaxation delay) were used for 2D NMR experiments. The same spectral width was used for both direct and indirect dimensions. The additional parameters (number of increments and mixing time) for the 2D NMR techniques were adjusted as needed. See **Table 1** for specifics on the parameters used for each 2D NMR experiment. Below is an example of how the 2D NMR experiments were performed using the 2D TOCSY experiment and similarly on the 600 and 800 MHz NMR spectrometers.

Using the 400 MHz NMR spectrometer, the 2D TOCSY spectral data were collected with 64 transients using 4096 points in the direct dimension and 256 points (number of increments) in the indirect dimension. A spectral width of 4800 Hz was used for both direct and indirect time dimension. The DIPSI pulse sequence[42] was used with a spinlock field of 7000.0 Hz during the mixing time of 80 milliseconds. A 1-second relaxation delay was used after each transient for the TOCSY experiments.

Table 1: NMR Parameters

NMR Experiment	SW (Hz)	SW1 (Hz)	Points Size	t1 Increments	Mixing Time (ms)	Relaxation Delay (s)	Transients
400 MHz							
1D Proton	4800	-	16384	-	-	1	256
1D Carbon-13	22000	-	16384	-	-	1	16
2D DQFC	4800	4800	8192	512	-	1	32
2D TOCSY	4800	4800	4096	256	80	1	64
2D NOESY	4800	4800	4096	256	150, 300	1	64
2D HMQC	4800	22000	4096	128	-	1	32
600 MHz							
1D Proton	8000	-	16384	-	-	1	64
2D DQFC	8000	8000	2048	512	-	1	
2D TOCSY	8000	8000	2048	256	40, 80	1.5	64
2D NOESY	8000	8000	2048	256	50, 100, 150, 300	1.5	128
2D HMQC	8000	22600	2048	128	-	1.5	256
800MHz							
2D TOCSY	9600	9600	1024	128	80	1	64
2D NOESY	9600	9600	2048	128	200	1	64

The Varian/Agilent 400 and 600 MHz 2D NMR spectral data were collected in complex mode while the Bruker 800 MHz 2D NMR spectral data were collected in States-TPPI mode.

III.2.E. Diffusion Measurement of AFGP8 Using NMR Spectroscopy

The diffusion coefficient of AFGP8 in DMSO was measured with Diffusion-order spectroscopy (DOSY), using the one-shot pulse sequence developed by Michelle et al.[43]. The parameters (90˚ pulse, spectral width, transmitter offset, number of points) were transferred from the 1D proton NMR experiment. For the diffusion delay, 0.2-second was used. A relaxation delay of 2-seconds was used. The gradient range used for the experiment was 1000 to 27000 (Varian NMR DAC unit), collecting 1D proton spectral data at every 1000 DAC unit increment. The gradient calibration of 1 DAC unit is equal to 0.001744 G/cm.

A total of 27 1D proton spectral data were collected in an array fashion. Each spectral data was collected using 16 transients.

III.2.F. NMR Spectral Data Processing and Analysis

For the 1D NMR experimental data, the computer software *MNova*© (Version: 11.0.1-17801) was used to process the data. The free induction decay (FID) was multiplied with sine square 90˚ weighting function. The spectrum was phase corrected (zeroth order) by setting the pivot to the center of the most intense peak. Frequency independent (zero-order) phasing was used to phase correct the center peak region. Frequency-dependent (first-order) phasing was then used to phase correct all other peaks. After phasing the spectrum to in-phase absorptive peaks, the spectrum was baseline corrected with a polynomial baseline correction and reference to the DMSO peak (solvent peak) at 2.5 ppm for proton and 40 ppm for ^{13}C.

The DOSY spectral data were processed in *MNova*© (Version: 11.0.1-17801). The FID was multiplied by sine square 90˚ weighting function. Since each dataset consists of 27 1D proton spectra, the first spectrum was phase and baseline corrected using the same process as mentioned previously for the 1D proton spectral data processing. All processing on the first spectrum was applied automatically by *MNova*© to the rest of the DOSY spectra. The area under the DMSO solvent peak and methyl peak region (0.9 – 1.5 ppm) were integrated using *MNova*©. For both the solvent and methyl peaks, the peak areas from the integration were normalized to the peak area of the first spectrum, making the normalized peak area of the first spectrum "1". The normalized peak areas were plotted against the gradient level following the procedure by Harmon et al.[44] The diffusion coefficient of DMSO and AFGP8 were determined from the slope of the

exponential fitting function applied to the plots in the software *SigmaPlot*© (Version 12.5, Build 12.5.0.38).

The Varian and Bruker two dimensional NMR data were processed using NMRPipe[45]/NMRDraw[45]. NMRPipe is a UNIX-based software that was developed to assist in processing 2D NMR spectral data. All the 2D NMR spectral data (DQFC/TOCSY/NOESY/HMQC) were processed using NMRPipe. An example of the macro script used to process each 2D NMR spectral data is shown in **Appendix A**. As an example of the data processing, using NMRDraw, the 400 MHz TOCSY data, FID in both direct and indirect dimensions, was multiplied with a cosine-bells weighting function. Linear-prediction with an additional 64 number of points on top of the number of increments (points) was applied to the indirect dimension. Zero-filling was applied to both dimension to double the point size and round the number of points to the closes power of 2 before the data was Fourier transform. The phasing of both dimensions was adjusted as needed to get the in-phase, absorptive line shape in the 2D TOCSY spectrum. After proper phasing, a polynomial baseline correction was applied to both dimensions.

The 2D TOCSY spectrum in Sparky[46] format was created using the NMRPipe/NMRDraw command ‘pipe2ucsf’; Sparky is a program that is used to analyze multi-dimensional NMR spectra, such as chemical shift assignments and NOE cross peak integration. All chemical shifts of the 2D TOCSY spectrum were referenced to the DMSO-d_6 solvent peak at 2.5 ppm.

All chemical shift assignments were done in Sparky using the standard notation compatible with the program *Combined Assignment and Dynamics Algorithm for NMR Application* (CYANA)[47]. Since AFGP have glycosylated threonine (T*), which is not a standard residue in the CYANA library, there is no standard notation for each atom. The glycosylated threonine (Thr* or T*) residue

library was created following the CYANA website guideline. The notation for each atom of the disaccharide was self-set. **Figure 3** shows the notation used for the glycosylated threonine while assigning the chemical shift in Sparky. For example, the amide proton (atom number 56) of the disaccharide of threonine position 3 used the following notation "THG3 H56" when assigning the chemical shift in Sparky.

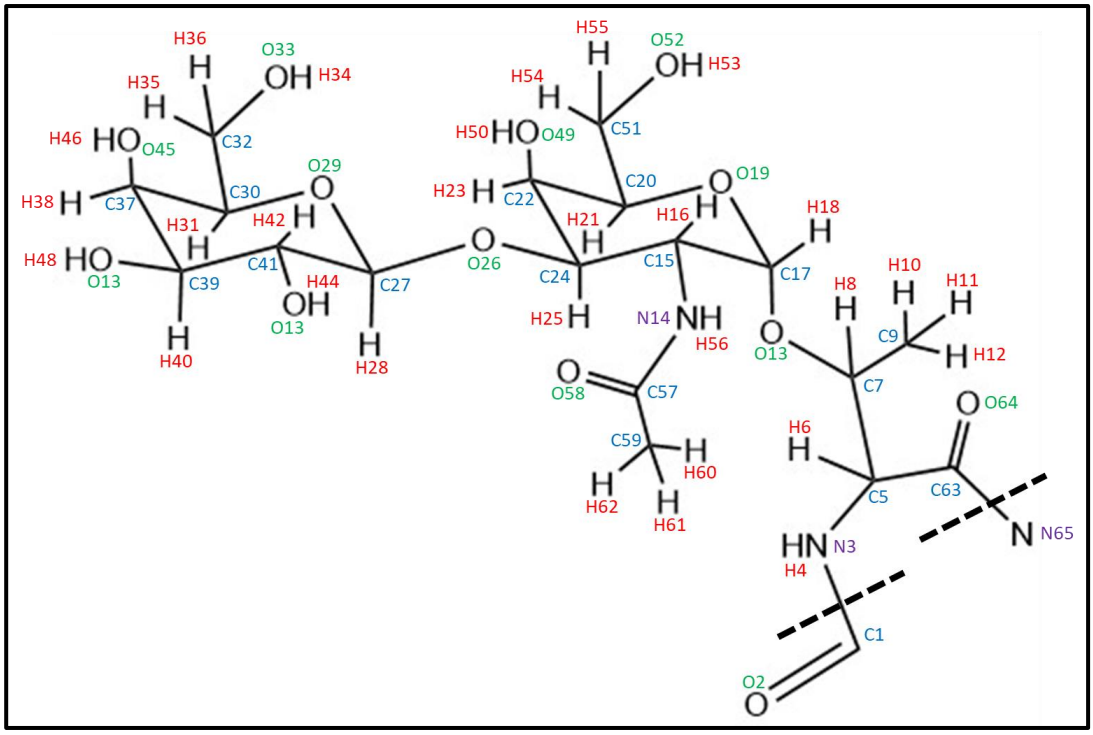


Figure 3. The glycosylated threonine created into CYANA library format.
Note: The notation "THG" was used for the glycosylated threonine. The color codes for the types of atom are as follow: red (hydrogen), blue (carbon), green (oxygen), and purple (nitrogen).

The integrations of all NOE cross-peak volume were done in Sparky. Any NOE cross-peak that was well-resolved used the Gaussian integration method and integrated with the automatic integration command "pi." The integrated volume of an NOE cross-peak that was well-resolved but was a result of two or more cross-peak overlaps was divided by the number of overlaps; that volume was then

assigned to all the NOE cross-peak involved in the overlaps. The box integration method was used for regions where NOE cross-peak overlaps were not well-resolved. All cross-peak assignments were locked in place then the box integration method was dragged over that overlapping region to assign each cross-peak with a volume.

III.3. Two-dimensional NMR Spectra Simulation of the Disaccharide

The two-dimensional NMR spectra (COSY/TOCSY) simulations of the disaccharide *beta-D-galactosyl-(1,3)-alpha-N-acetyl-D-galactosamine* were performed using the web service *Glycan Optimized Dual Empirical Spectrum Simulation (GODESS)*[48] (http://csdb.glycoscience.ru/database/index.html). The "Predict NMR" option was used. The disaccharide structure was input into *GODESS* using the "Input using Structure Wizard" option. The parameters used to make the disaccharide on *GODESS* is shown in **Appendix B**. After the disaccharide was made in *GODESS*, the 2D COSY and TOCSY spectra were simulated in water at 303 K. The spectra were simulated at two NMR spectrometer frequency, 400 MHz and 600 MHz. Specific parameters used during the simulation is also shown in **Appendix B**.

III.4. Chemical Shift-Based Secondary Structure Prediction

The secondary structure of both AFGP8 was predicted using the method developed by Wishart et al. [49], for the proton chemical shifts, and Wishart and Sykes [50], for the carbon-13 (^{13}C) chemical shifts. The proton and ^{13}C random coil chemical shift (RCCS) index generated by Tremblay et al. [51] were used as the reference chemical shifts when assigning the score of negative one (-1), zero (0), and positive one (1). For predicting the secondary structure using the proton

chemical shift, only the alpha proton (Hα) chemical shift was used. The secondary structure prediction with the ^{13}C chemical shifts used the alpha carbon (Cα) and beta carbon (Cβ) chemicals shifts. A negative one was assigned to the chemical shifts value less than the RCCS index. A zero was assigned to the chemical shifts with the same as the RCCS index. A positive one was assigned to the chemical shifts value greater than the RCCS index.

III.5. Structural Models

Three-dimensional coordinate in protein database (PDB) format of each AFGP8-BS and AFGP8-TB primary sequence were created. The 3D structure of each was a linear peptide, and the linear peptide was the starting structure used for the *Yet Another Scientific Artificial Reality Application* (YASARA) structure calculation. The CYANA format of the distance constraints was not compatible with the program (YASARA). The CYANA format of the distant constraints and atom labeling was converted into the Xplor format. The Xplor format was then converted into a format compatible with YASARA using a script written generated in R-Statistical program (R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.)

The structure calculation was ran using YASARA (version 16.9.23). The upper and lower distance limit (distance deviation) allowed during the structure calculation was 30% of the actual distance calculated from the integrated NOE volume. A total of 100 structures, with the lowest potential energy, were generated. After the 100 structures were generated, the 100 structures were refined in DMSO using YASARA. A macro, in YASARA, was used to combine the 100 structures into a single PDB file and a text file was generated containing the

following: RMSD of backbone for each of the 100 structures, an average RMSD of the 100 structures, potential energy of each structure, and distance violation of the distance constraints used. Out of the 100 structures generated, the ten structures with the least number of distance violations along with lowest potential energy were chosen for the AFGP8 structures.

A summary of the structure generating process is shown in **Figure 4** as a flow chart. Each box lists the overview of the actions performed at each step. The molecular graphics system *PyMol*™ (Version 1.7.0.0.) was used to align the three-dimensional structure of AFGP8-BS and AFGP8-TB for comparison as shown in the Results section.

The PDB file of antifreeze glycoprotein fraction 8 from *Trematomus borchgrevinki* (AFGP8-TB), in water at 5 ˚C, was provided by Dr. Andrew N. Lane[52] (Lane personal communication to Krishnan). The structures of synthetic antifreeze glycoproteins analog (sAFGP), in water at 5 ˚C, was provided by Dr. Shin-Ichiro Nishimura[53] (Nishimura personal communication to Krishnan). The primary sequence of sAFGP is Ala-Thr*-Ala-Ala-Thr*-Ala-Ala-Thr*-Ala. The threonine (Thr*) is glycosylated with the disaccharide *beta-D-galactosyl-(1,3)-alpha-N-acetyl-D-galactosamine* just like natural AFGP.

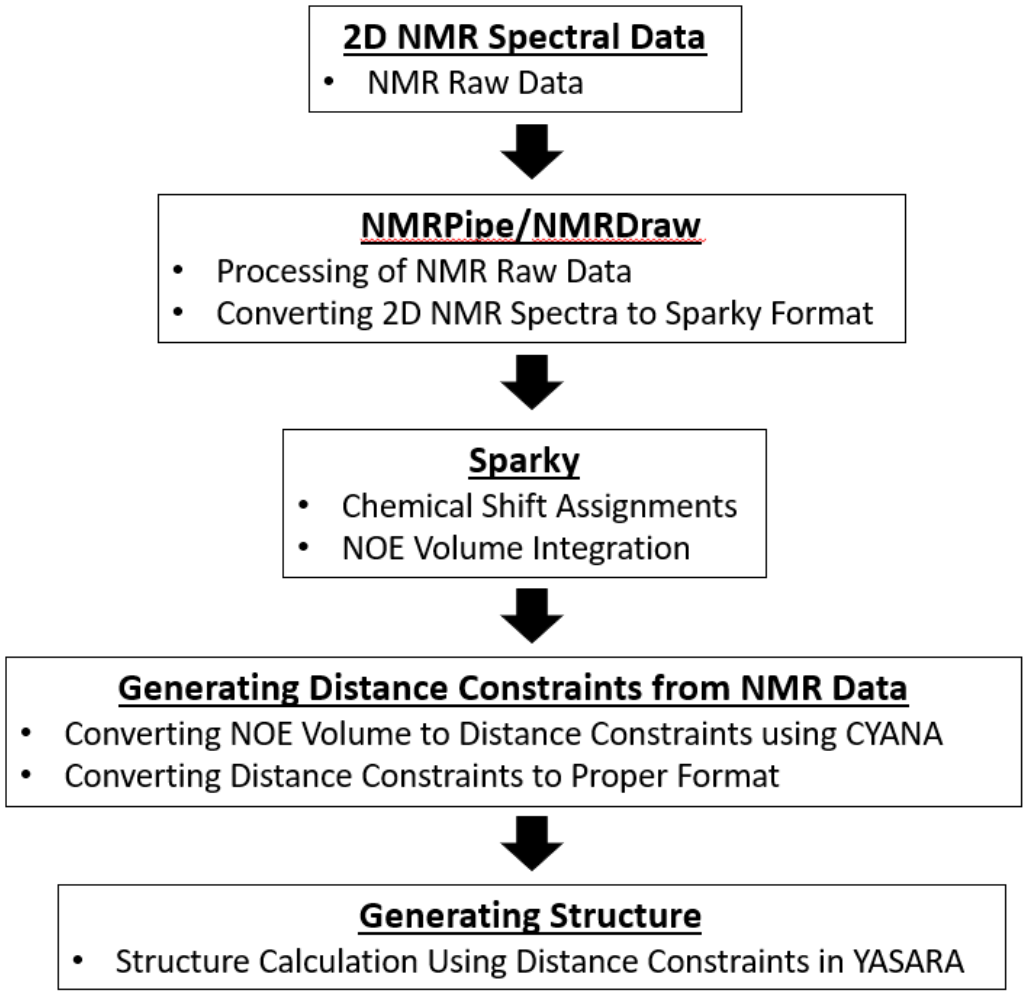


Figure 4. The flow chart summarizing each step involved in generating the protein structures from the NMR spectral data.

IV. RESULTS

Overview

The ensemble of structures of antifreeze glycoproteins fraction 8 (AFGP8) were determined in dimethyl sulfoxide (DMSO) using nuclear magnetic resonance (NMR) spectroscopy. All NMR data collected and analyzed were presented in the following sections for the AFGP8 structures in DMSO. However, the structures of AFGP8-TB and synthetic AFGP analog were not determined experimentally as part of the thesis work but provided by Dr. Andrew H. Lane and Dr. Shin-Ichiro Nishimura.

IV.1. One-Dimensional NMR Spectra

IV.1.A. One-Dimensional Proton NMR Spectra

The proton resonance distribution of peptides and proteins falls into three major regions: amide proton region (~6-9 ppm), alpha proton region (~3-5 ppm), and methyl protons region (~0.5-2 ppm), despite being folded or not. **Figure 5** shows the 1D proton NMR spectrum of AFGP8-BS (**Figure 5A**) and AFGP8-TB (**Figure 5B**) and the distribution of the proton resonances in the regions mentioned previously. While the three major regions were distinct on the 1D proton NMR spectra, there were proton resonance overlaps in each of the regions which are the characteristic of peptides and proteins

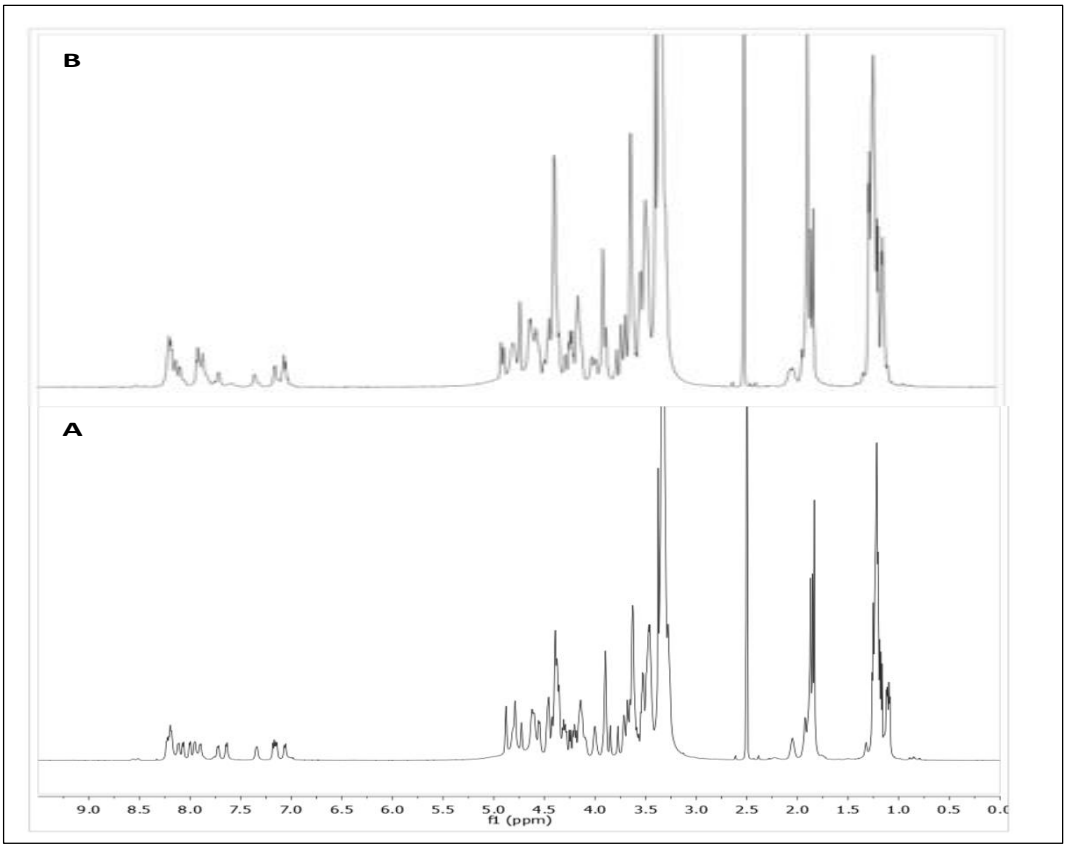


Figure 5. The 600 MHz 1D proton NMR spectra of antifreeze glycoprotein fraction 8 (AFGP8) from Arctic cod *Boresogadus saida* (A) and from Antarctic notothenioids *Trematomus borchgrevinki* (B).

IV.1.B. ^{13}C-enriched N-Terminus Methylated AFGP8-TB

To enhance the ^{13}C signal for structural studies regarding the N-terminus, AFGP8-TB was methylated with two ^{13}C-isotope enriched methyl groups as shown in the NMR spectra of **Figure 6**. **Figure 6A** shows the 1D proton NMR spectra of N-terminus ^{13}C-isotope enriched methylation of AFGP8-TB (*m*AFGP8-TB) ran without decoupling the ^{13}C signal (bottom spectrum) and with decoupling the ^{13}C signal (top spectrum). The DMSO solvent peak was at 2.50 ppm for both proton NMR spectra. The resonance overlaps between the chemical shift range of 1.75 – 1.95 ppm were the proline beta and gamma protons along with the N-acetyl methyl protons of the disaccharides. The spectrum, without ^{13}C decoupling,

resulted in two doublet peaks with the chemical shifts of 2.03 and 2.35 ppm. The two peaks centered at 2.03 ppm and 2.35 ppm were separated by 128 Hz (0.32 ppm) which was the result of the methyl protons ^{1}J coupled to the directly bond ^{13}C nucleus. The additional splitting of the peaks into doublets was the result of an additional proton to ^{13}C ^{3}J coupling of the ^{13}C nucleus three bond distance away. The proton to ^{13}C ^{3}J coupling was confirmed by running another 1D proton NMR experiment on the same mAFGP8-TB sample decoupling the ^{13}C signal as shown in the top spectrum of **Figure 6A**. The two doublet peaks collapsed into a singlet peak at 2.18 ppm due to the decoupling averaging out the ^{13}C signal of the two possible spin states, spin up (α) and spin down (β). **Figure 6B** shows the ^{13}C NMR spectra of regular AFGP8-TB (bottom spectrum) and *m*AFGP8-TB (top spectrum). The septet peak on both spectra was the result of the ^{13}C nuclei of the DMSO solvent ^{1}J coupled to three deuterium nuclei. With just 16 transients, a carbon peak appeared on the top 1D ^{13}C spectrum for *m*AFGP8-TB in **Figure 6B**. Regular AFGP8-TB with the same concentration was tested with 16 scans, and no carbon peak appeared as shown in the bottom 1D ^{13}C spectrum of **Figure 6B**. The comparison of the two ^{13}C NMR spectra confirmed the success of the N-terminus methylation process of AFGP8-TB.

IV.2. NMR-Based Strategy on Spectral Data Analysis

The sequential amino acid assignment of AFGP8 was done following the NMR concept developed by Kurt Wüthrich.[6] The 2D TOCSY and COSY spectra can be used to identify different spin systems; in this case, the spin systems were the amino acids. The peptide bond prevents TOCSY and COSY transfer between different amino acid of primary sequence. On the 2D TOCSY spectrum, there will not be any cross-peak between two or more different amino acids unless it is the

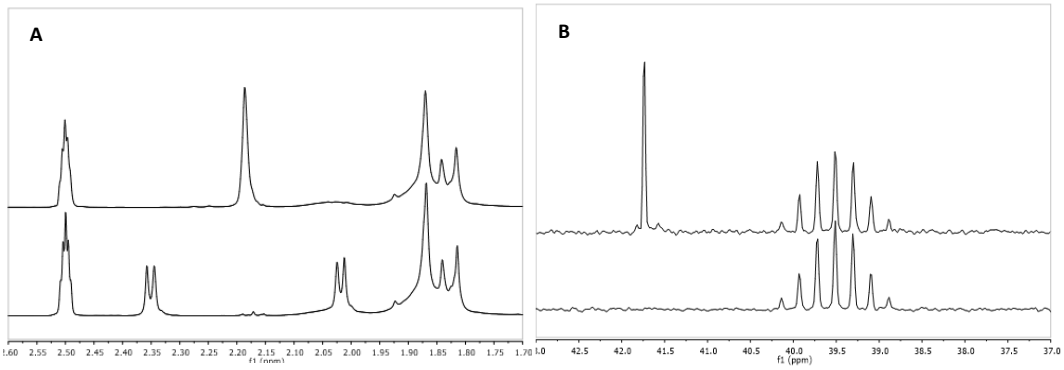


Figure 6. The 400 MHz 1D proton (A) and ^{13}C (B) NMR spectra of ^{13}C-isotope enriched methylation of AFGP8-TB and regular AFGP8-TB (the bottom spectrum in (B)).

Note: For Figure 6A, the bottom 1D proton NMR spectrum was ran without decoupling the ^{13}C signal. The top 1D proton NMR spectrum was ran with decoupling the ^{13}C signal. For Figure 6B, the bottom ^{13}C NMR spectrum is the regular AFGP8-TB and top ^{13}C NMR spectrum is the ^{13}C-methylated AFGP8-TB.

result of resonance overlaps. The 2D NOESY spectrum can be used to identify inter-residue NOE which are between different amino acids through space. There are also intra-residue NOE cross-peaks, but those will be ignored during the sequential amino acid assignment. For inter-residue NOE, there are proton distances between adjacent amino acids that will always have NOE cross-peak with each other appearing on the 2D NOESY spectrum, except when the product of the correlation time with the spectrometer frequency is equal to 1. This exception occurs due to the transition of the NOE cross peak intensity flips its sign as it goes from negative to positive and vice versa. One of the inherent, inter-residue NOE cross peaks that all peptides/proteins exhibit is the alpha proton the immediate amino acid to the amide proton of the adjacent amino acid on the right as shown as the solid green line in **Figure 7** (going from N-terminus to C-terminus). **Figure 7** shows the sequencing concept and nomenclature used to trace out the peptide/protein backbone. The figure consists of alanine (i), threonine (i+1,

without the hydroxyl group on the side chain), and proline (i+2). The green dashes show the proton cross-peak that would appear in the TOCSY/COSY spectrum. The solid green lines show the inter-residue NOE cross peak that would appear in the NOESY spectrum. These two types of cross-peak can be used to trace out the peptide/protein backbone to determine and confirm the primary sequence. After the primary sequence is confirmed, the chemical shift of the protons can be assigned.

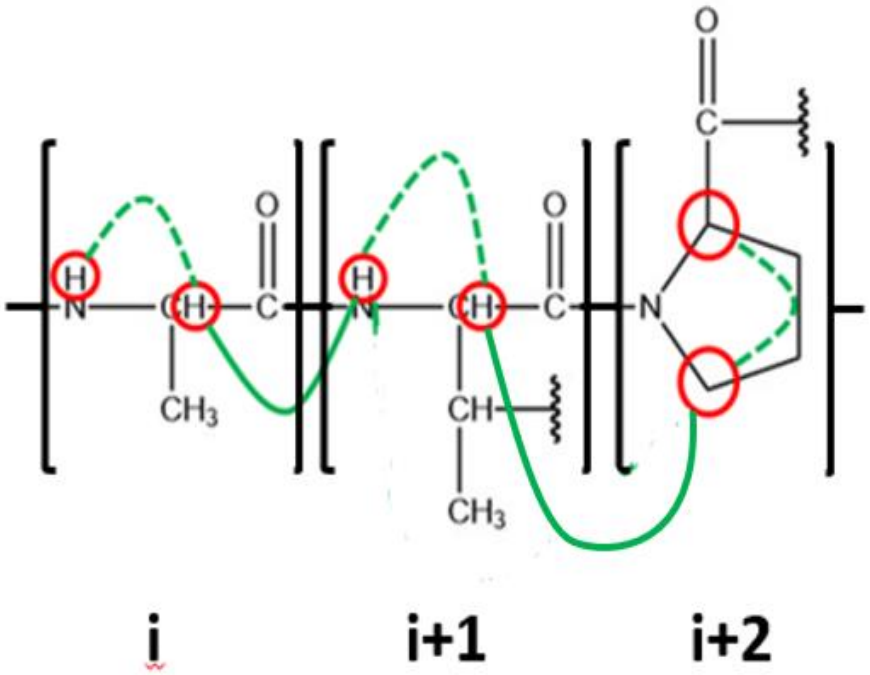


Figure 7. Nomenclature for sequential assignment using the tripeptide Ala-Thr*-Pro as an example.
Note: The green dashes are proton-proton through-bond correlation that would show up on a 2D TOCSY spectrum. The green solid lines are proton-proton NOE (through-space) correlation that would show up on a 2D NOESY spectrum.

IV.3. Sequence Specific Assignment of AFGP8-BS and AFGP8-TB

IV.3.A TOCSY-NOESY Crosswalk of Primary Sequence

On the amide proton chemical shift to the alpha proton and methyl protons chemical shift regions of the TOCSY spectrum, there were unique patterns matching the alanine and threonine in the primary sequence of AFGP8. Using a mixing time of 80 milliseconds during the TOCSY experiment, there were two cross-peaks: amide proton to alpha proton and amide proton to beta protons, corresponding to alanine. For threonine, there were three cross-peaks: amide proton to alpha proton, amide proton to beta proton, and amide proton to gamma protons. As shown in **Figure 8**, for both AFGP8-BS (A) and AFGP8-TB (B), four threonines and seven alanines were identified. For both AFGP8 fractions, only seven out of the eight total alanines were identified because the first alanine at the N-terminus does not have an amide proton. The chemical shifts of the N-terminus amine protons were not expected to show up in the amide chemical shift region, or rapid exchange with the residual water could have occurred to prevent detection of the amine (NH_2) protons. The two prolines at position four and ten for AFGP8-BS and position seven and thirteen for AFGP8-TB were not identified in this region because the prolines do not have amide protons. Even without TOCSY signal in this region, proline has a distinct pattern that shows up between the 2-5 ppm chemical shift region as shown in **Figure 9**. The distinct region of the proline on a TOCSY spectrum consists of the cross peaks between the alpha proton to the beta/gamma protons and delta protons to the beta/gamma protons. Only one proline pattern was seen for AFGP8-BS in **Figure 9A**; it was determined that the proton chemical shifts of the two prolines were very close to being degenerate so the chemical shifts were not resolved. On the contrary, AFGP8-TB, as shown in

Figure 9B, had two distinct alpha proton chemical shifts for the two prolines while the delta protons chemical shifts were nearly degenerate. Using the DQFC spectrum, it was determined that one of the beta protons has a very similar chemical shift with the gamma protons and cannot be resolved on the TOCSY spectrum. All but one amino acid (alanine 1) were confirmed using the TOCSY and DQFC spectrum.

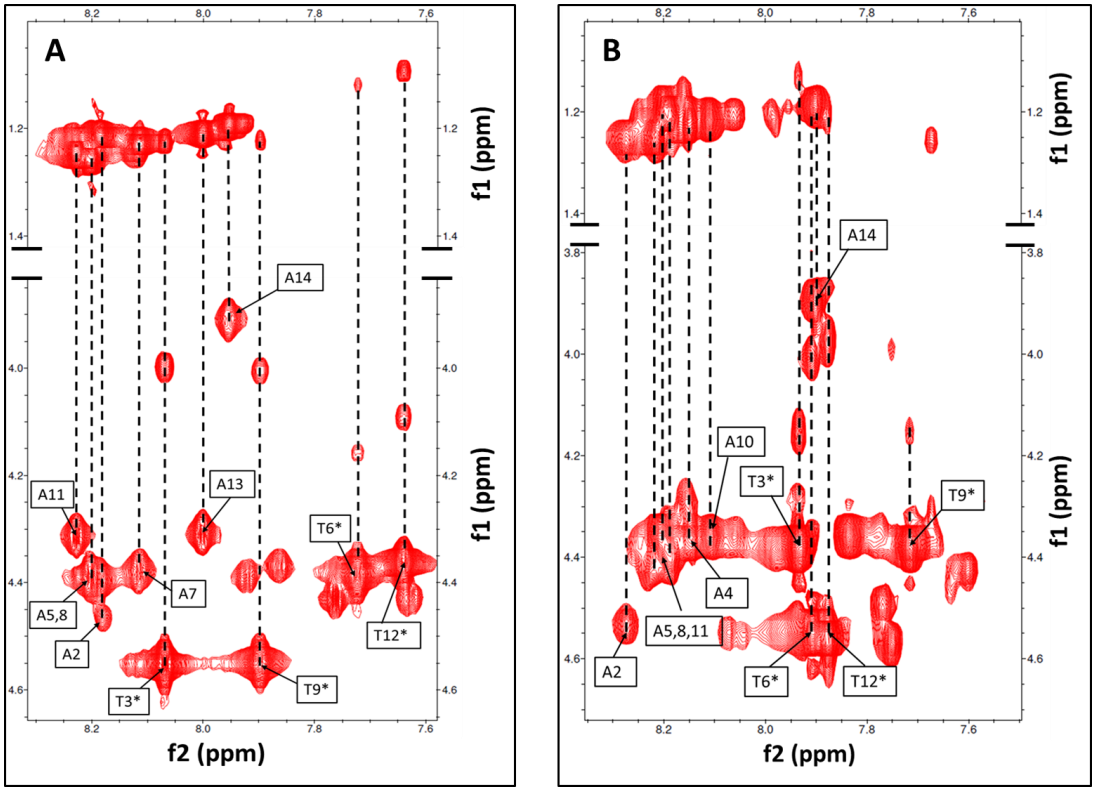


Figure 8. The 2D TOCSY NMR spectra of AFGP8-BS (A, 600 MHz) and AFGP8-TB (B, 800MHz) showing the amide to alpha and methyl proton region.

Note: Both spectrum show the TOCSY patterns associated with all the amino acid of the primary sequence except for alanine at position 1, proline at position 4 and 10 in AFGP8-BS, and position 7 and 13 in AFGP8-TB.

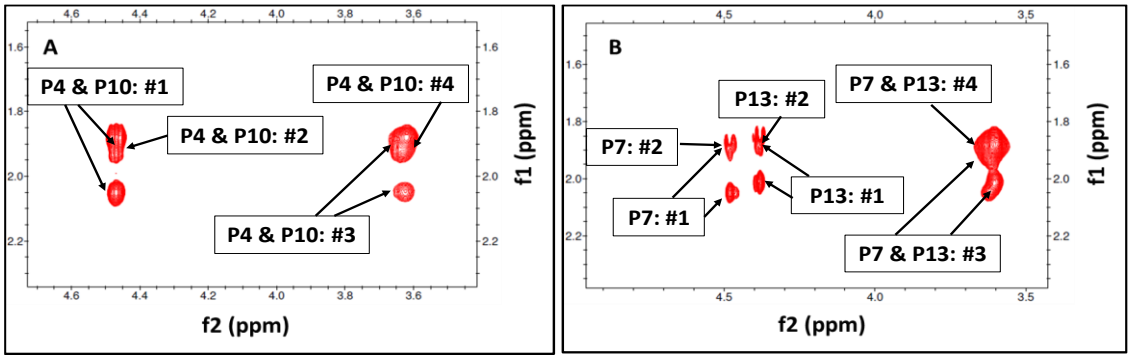


Figure 9. The 2D TOCSY NMR spectra of AFGP8-BS (A) and AFGP8-TB (B) showing the distinct pattern that proline exhibit between its four types of protons: alpha proton (Hα), beta protons (Hβ), gamma protons (Hγ), and delta protons (Hδ).
Note: The notations are as follows: #1 is Hα/Hβ, #2 is Hα/Hγ, #3 is Hβ/Hδ, and #4 is Hγ/Hδ. Multiple arrows may point to one cross peak due to overlaps in proton chemical shifts.

The final alanine, for both AFGP8-BS and AFGP8-TB, was identified using the combination of the TOCSY and HMQC spectra. While the alanine did not have an amide to alpha proton cross-peaks on the TOCSY spectrum, there was a cross-peak between a proton (3.37 ppm) in the alpha proton chemical shift range and a proton (1.17 ppm) in the methyl proton chemical shift range. From the HQMC spectrum, the proton in the alpha proton chemical shift range was covalently bonded to a carbon (49.61 ppm) in the alpha carbon chemical shift range; the proton in the methyl proton chemical shift range was covalently bonded to a carbon (20.91 ppm) in the methyl carbon chemical shift range. From these observations, the chemical shifts of the two protons were assigned as the alpha proton and beta protons to an alanine. As shown in **Figure 10** for AFGP8-BS and **Figure 11** AFGP8-TB, the TOCSY spectrum was coded in red and NOESY spectrum was coded in blue. The TOCSY-NOESY crosswalk started at alanine 2 (green rectangle) since the chemical shifts of the amine protons on alanine 1 could not be identified. Also, the chemical shift of the alpha proton on alanine 1 was

very close to the residual water chemical shift, so the NOE cross-peak of alanine 1 alpha proton to alanine 2 amide proton was overlapping with a possible exchange cross-peak of the water to alanine 2 amide proton. Going from alanine 2 to alanine 14 (C-terminus), the intra-residue TOCSY cross peak of the amide proton (i) to alpha proton (i) and inter-residue NOESY cross peak of alpha proton (i) to amide proton (i+1) were used to traced and assigned chemical shifts to the protons of the peptide backbone.

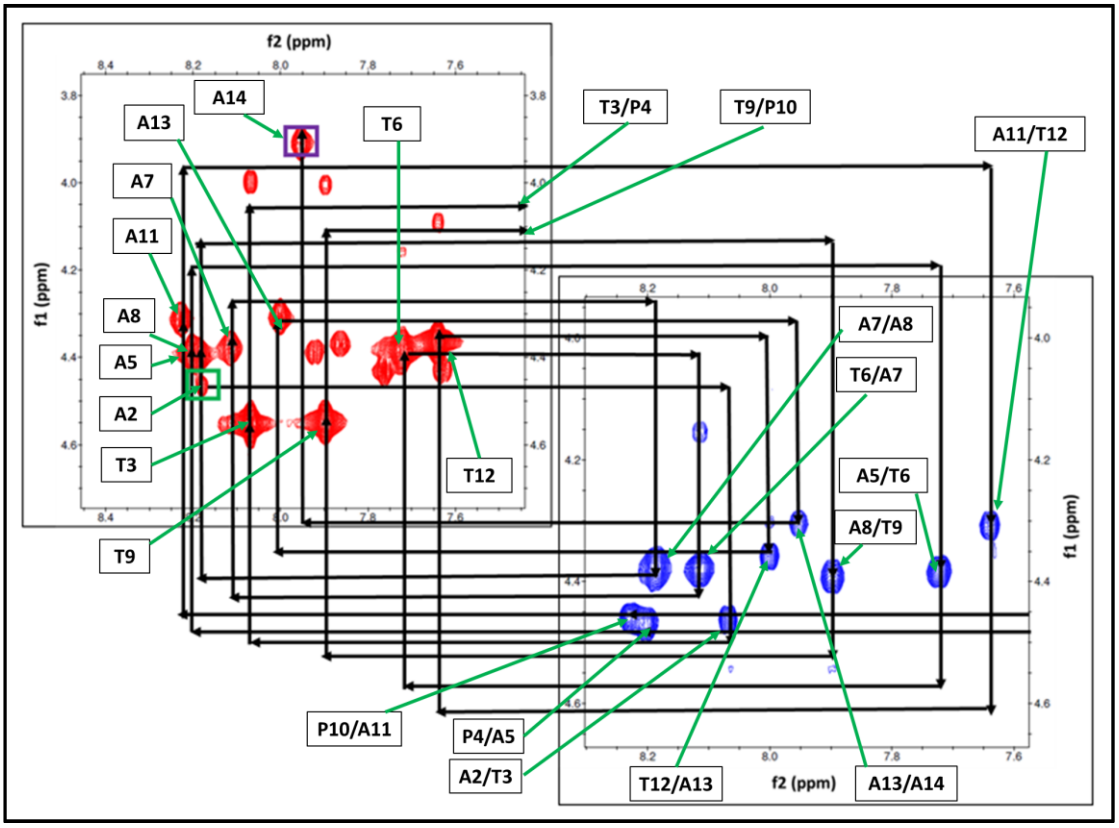


Figure 10. The TOCSY-NOESY crosswalk of AFGP8-BS. The 2D TOCSY spectrum was in red, and the 2D NOESY spectrum was in blue.

Note: The crosswalk started on the alanine at the second position of the primary sequence (green box) and ends on the alanine at the fourteenth position (C-terminus) of the primary sequence (purple box). The TOCSY correlation (red) used one amino acid code (e.g., A4). The NOESY correlation (blue) used two amino acids code (e.g., A4/A5).

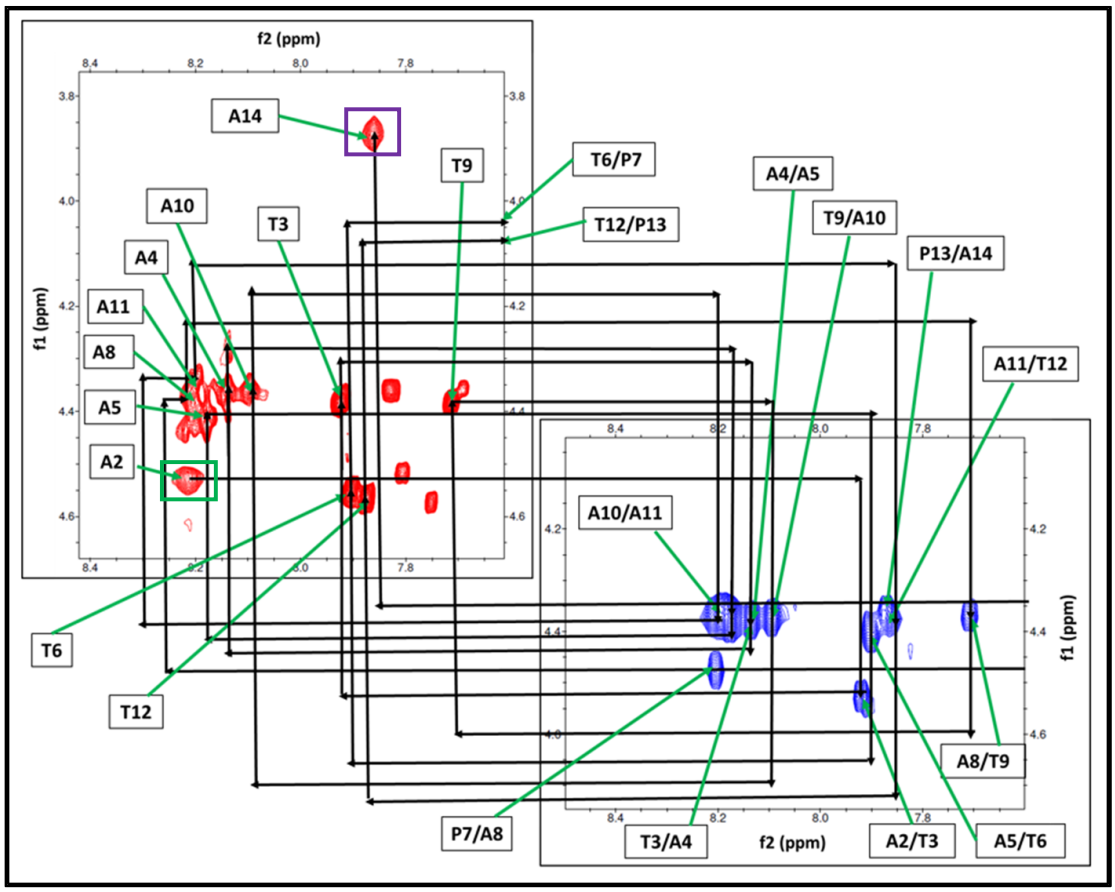


Figure 11. The TOCSY-NOESY crosswalk of AFGP8-TB. The TOCSY spectrum was in red, and the NOESY spectrum was in blue.

Note: The crosswalk started on the alanine at the second position of the primary sequence (green box) and ends on the alanine at the fourteenth position (C-terminus) of the primary sequence (purple box). The TOCSY correlation (red) used one amino acid code (e.g., A5). The NOESY correlation (blue) used two amino acids code (e.g., A5/T6).

The individual black boxes in **Figure 10** and **Figure 11** with one amino acid notation (e.g., A4) stands for TOCSY correlation (through bond correlation) of the amide proton to alpha proton. The individual black box with a slash between two amino acids (e.g., A4/A5) stands for NOESY correlation (through space). Since the prolines do not have amide protons, the delta protons were used for the inter-residue NOE on the TOCSY-NOESY crosswalk as shown in **Figure 10**, using the notations "T3/P4" and "T9/P10" and **Figure 11**, using the notations "T6/P7" and "T12/P13" (cross peaks involving the prolines were not shown on the

TOCSY-NOESY crosswalk). This confirmed the primary sequence of AFGP8 due to the amino acids being in the correct positions.

IV.4. Chemical Shift Assignment of the Disaccharide

IV.4.A. Simulated and Experimental 2D NMR Spectra of Free Disaccharide

Two-dimensional NMR experiments were run on the free disaccharide (not attached to AFGP), in a DMSO-d_6 solvent, to assist in the chemical shift assignment of the actual disaccharides that are present in AFGP. Along with the NMR experiments ran on the free disaccharide, a TOCSY simulation of the disaccharide was also run in deuterated water. While the simulation and experimental spectra were run in different solvents, the TOCSY patterns do show similarity between the experimental and simulated TOCSY spectra, as shown in **Figure 12**, even though the chemical shift do not line up. The numbering system used for the protons follows the standard numbering system used for carbohydrate. For example, the number "1" will stand for the proton attached the carbon 1 of the carbohydrate. The simulated TOCSY spectrum, as shown in **Figure 12A**, shows TOCSY correlation between all protons within a spin system. However, the experimental TOCSY spectrum, as shown in **Figure 12B**, was only able to show TOCSY correlation (80 milliseconds mixing time) between two protons up to 4 covalent bond distances away from each other. The proton chemical shifts of the disaccharides are shown in **Table 2**.

IV.4.B. Chemical Shift Assignment of the Disaccharides in AFGP8

The amino acid positions of the four disaccharides, attached to the threonine residues, were determined by the NOE cross-peak between the amide

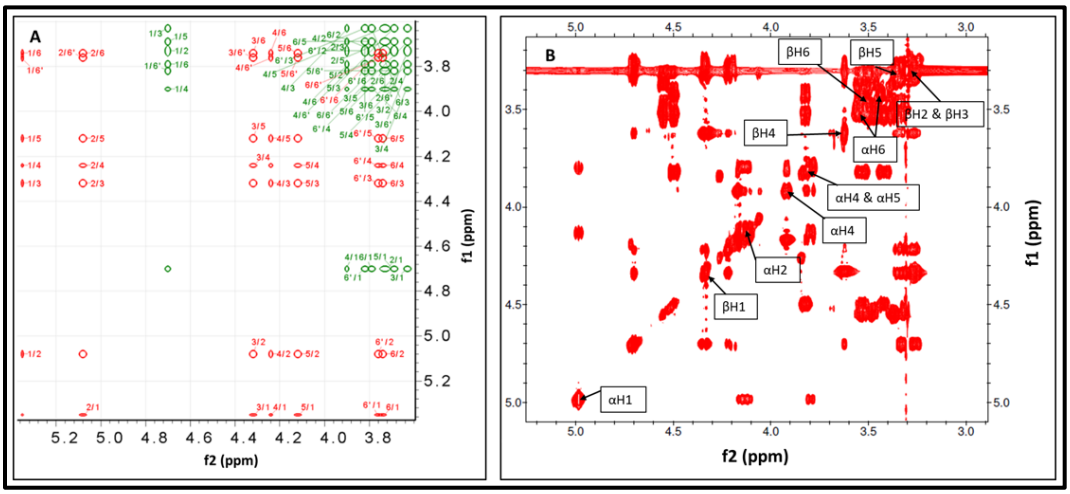


Figure 12. The comparison of the simulated (A) and experimental (B) 400MHz 2D NMR TOCSY spectra of the disaccharide *beta-D-galactosyl-(1,3)-alpha-N-acetyl-D-galactosamine*.

Note: The simulated NMR TOCSY spectrum was done in water while the experimental NMR TOCSY spectrum was done in DMSO-d_6. The red open circles in the simulated TOCSY spectrum (A) were corresponding the TOCSY correlation of *alpha-N-acetyl-D-galactosamine* of the disaccharide, and the green open circles were corresponding the TOCSY correlation of *beta-D-galactosyl* of the disaccharide. The experimental TOCSY spectrum (B) had no color differentiation, but was differentiated by the symbol α and β as in the disaccharide name.

Table 2. ^{1}H and ^{13}C Chemical Shifts of the Disaccharide

Proton	NH	Methyl	H1	H2	H3	H4	H5	H6	O1	O2	O3	O4	O6
α-GalNAc	7.47	1.81	4.99	4.13	3.80	3.92	3.82	3.42; 3.53	6.45	-	-	4.17	4.50
β-Gal3	-	-	4.33	3.32	3.27	3.62	3.37	3.52	-	4.22	4.70	4.33	4.55
Carbon-13	C1	C2	C3	C4	C5	C6							
α-GalNAc	91.27	49.20	75.89	67.69	70.57	60.66							
β-Gal	104.00	71.02	73.49	68.36	62.89	60.66							

proton of the disaccharide and the amide proton of threonine as shown in **Figure 13**. The chemical shifts of the first six-member ring (*alpha-N-acetyl-D-galactosamine*) of the disaccharide *beta-D-galactosyl-(1,3)-alpha-N-acetyl-D-galactosamine* were assigned starting from the amide proton next to the acetyl group. The amide protons of the disaccharides were determined by the TOCSY pattern in the amide proton to alpha proton chemical shift regions. The TOCSY patterns did not match any of the known amino acids (alanines and threonines) in the primary sequence, as shown in **Figure 14A**. With the TOCSY spectrum alone, it was not clear as to which chemical shift belongs to the proton (C2 proton) adjacent to the amide proton. Using the DQFC spectrum, as shown in **Figure 14B**, it was possible to determine the C2 proton chemical shift out of the four chemical shifts in the TOCSY spectrum in **Figure 14A**. Starting from the amide proton, the rest of the protons on the first six-member ring of the disaccharide were assigned. The two C6 protons were not degenerate, but it was not possible to unambiguously determine the chemical shift of each. One chemical shift (the center of the overlaps) was assigned to both C6 protons.

For the second six-member ring of the disaccharide, there was no amide proton to assist in the chemical shift assignment. The 2D TOCSY NMR technique can only show more than 3 bonds distance (more than 3 covalent bonds away) correlation between two protons if there is an intermediate proton that is ^{3}J coupled to both the protons. Since the nearest protons between the first and the second six-member ring were 4 covalent bonds away from each other, each six-member ring of the disaccharide was its independent spin system. The NOE correlation between the C3 proton of the first six-member ring and C1’ proton of the second six-member ring was used to assist in assigning the proton chemical shifts of the second six-member ring. For the chemical shift assignment of the

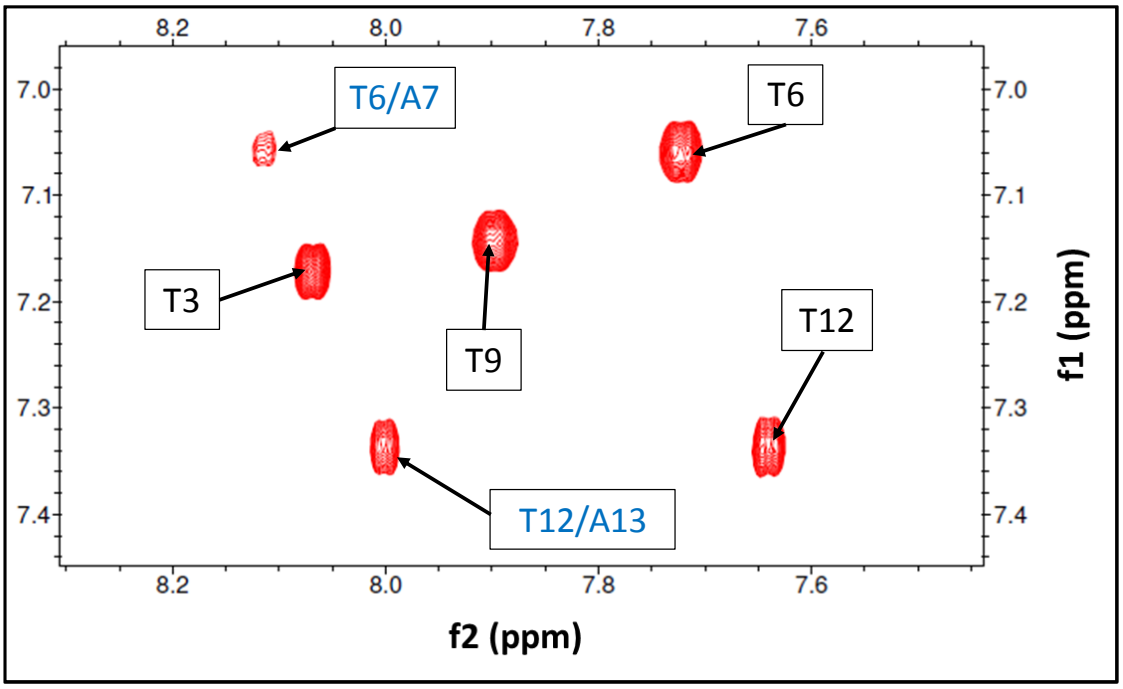


Figure 13. The 600 MHz 2D NOESY spectrum of AFGP8-BS focusing on the amide of the backbone to amide of the disaccharide NOE correlation.
Note: The position of the disaccharides and the threonines it's bonded to in the primary sequence were numbered in the black boxes. The blue numbered amino acids indicated the amide proton NOE cross peak with an adjacent amino acid (alanine) instead of the threonine it's bonded to.

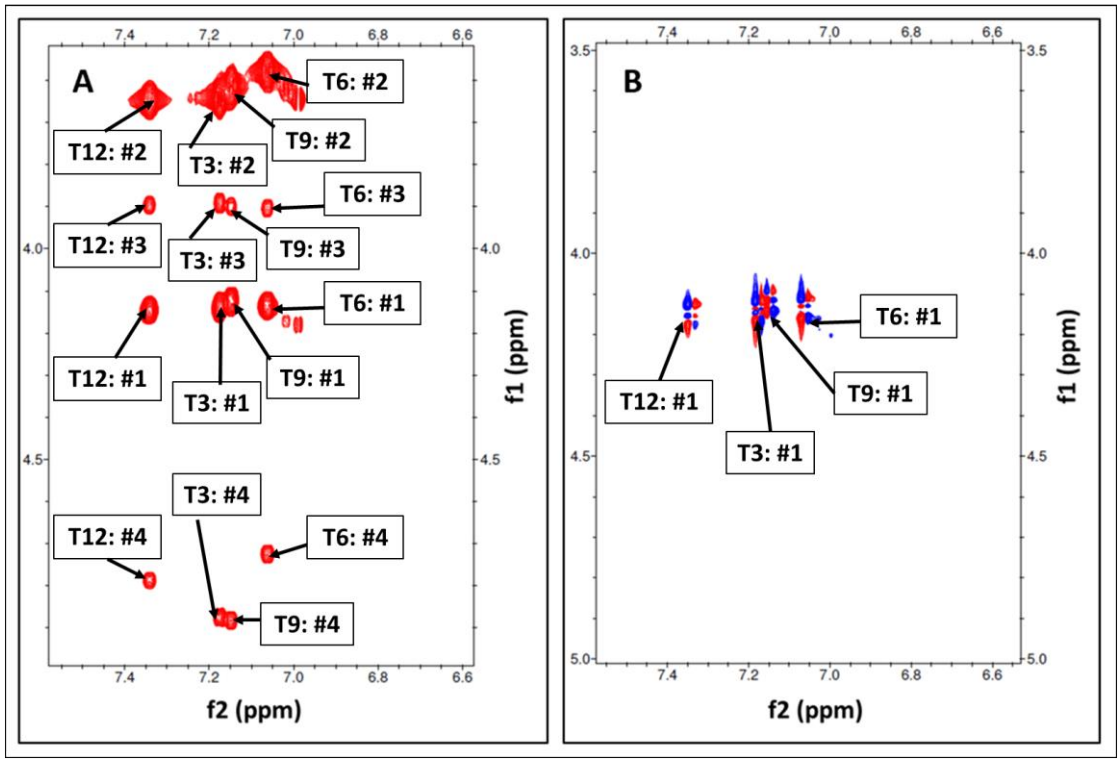


Figure 14. The 600 MHz 2D NMR TOCSY (A) and DQFC (B) spectra of AFGP8-BS.

Note: The spectra show the through bond correlation of the amide proton to the neighboring protons of the first six-member ring (*alpha-N-acetyl-D-galactosamine*) of the disaccharide *beta-D-galactosyl-(1,3)-alpha-N-acetyl-D-galactosamine*. The notations are as follows: #1 is NH/H2, #2 is NH/H3, #3 is NH/H4, and #4 is NH/H1.

two C6’ protons, one chemical shift (the center of the overlaps) was assigned to both protons since it was not possible to unambiguously assign a distinct chemical shift to each proton. The chemical shifts and amino acid positions of the methyl protons on the acetyl group were assigned using the NOE cross-peaks between the acetyl methyl protons and the amide proton of the disaccharides.

The chemical shift of the hydroxyl protons of the disaccharides was assigned using the combination of the TOCSY and DQFC spectra. As shown in **Figure 15**, the chemical shift for all the non-labile protons was identified in both simulated and experimental TOCSY spectra. There was a similarity between the simulated TOCSY pattern of the free disaccharide (**Figure 15A**) and experimental TOCSY pattern of the disaccharides in AFGP8 (**Figure 15 B**), which was similar to what was shown in **Figure 12** for the free disaccharide, not part of AFGP. Using the DQFC spectrum, the chemical shift of the hydroxyl proton was assigned using the non-labile ^{3}J coupled proton of the disaccharide. However, it was not possible to unambiguously assign the amino acid position of each C4’ hydroxyl group on AFGP8-BS and C4 and C4’ hydroxyl proton on AFGP8-TB.

Figure 16 shows an example of how the hydroxyl protons were unambiguously or ambiguously assigned. The four labeled TOCSY correlations on **Figure 16** were between the amide proton and C4 proton of the disaccharide. The four non-labeled TOCSY correlations were between the C4 proton and C4 hydroxyl proton. As shown in **Figure 16A**, the C4 proton had a very slight difference chemical shift between all four threonine residues, so it was possible to assign each C4 hydroxyl proton unambiguously in the 4.30-4.50 ppm region. For AFGP8-TB, as shown in **Figure 16B**, the chemical shift of the C4 proton between all the threonine residues seems to be overlapping and not resolved. Since there was not a slight difference in chemical shift, it was not possible to unambiguously

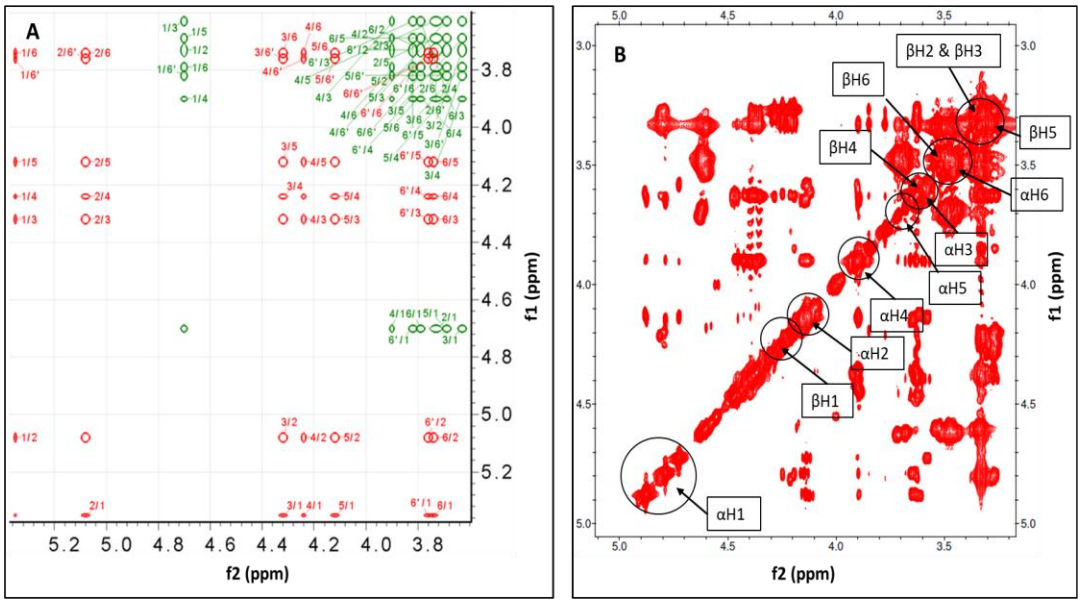


Figure 15. The comparison of the simulated (A) 2D TOCSY spectrum of the free disaccharide in water and the experimental (B) 2D TOCSY spectrum of AFGP8-BS focusing on the chemical shift region of the disaccharides in DMSO-d_6.

Note: For the simulated 2D TOCSY spectrum, the α-N-acetyl-D-galactosamine is represented by red number and β-D-galactosyl is represented by the green number. For the experimental 2D TOCSY spectra, the symbol α and β were used to represent α-N-acetyl-D-galactosamine and β-D-galactosyl of the disaccharide.

assign the amino acid position of the C4 hydroxyl proton using the known amino acid position of the C4 proton. The C4' hydroxyl proton amino acid position of both AFGP8-BS and TB was not unambiguously assigned due to the same reasoning. Even though the residue position of the C4 and C4' hydroxyl proton was unclear, it was certain that the chemical shifts assigned belong to a hydroxyl proton. Also, the 2D HMQC spectrum did not show any carbon to proton correlation involving the chemical shifts assigned to the hydroxyl protons. In addition, with the temperature dependent TOCSY, the chemical shifts of these protons constantly shifted upfield as the temperature increases which is a clear sign of label proton capable of exchanging with the residual water present in the AFGP8 NMR samples.

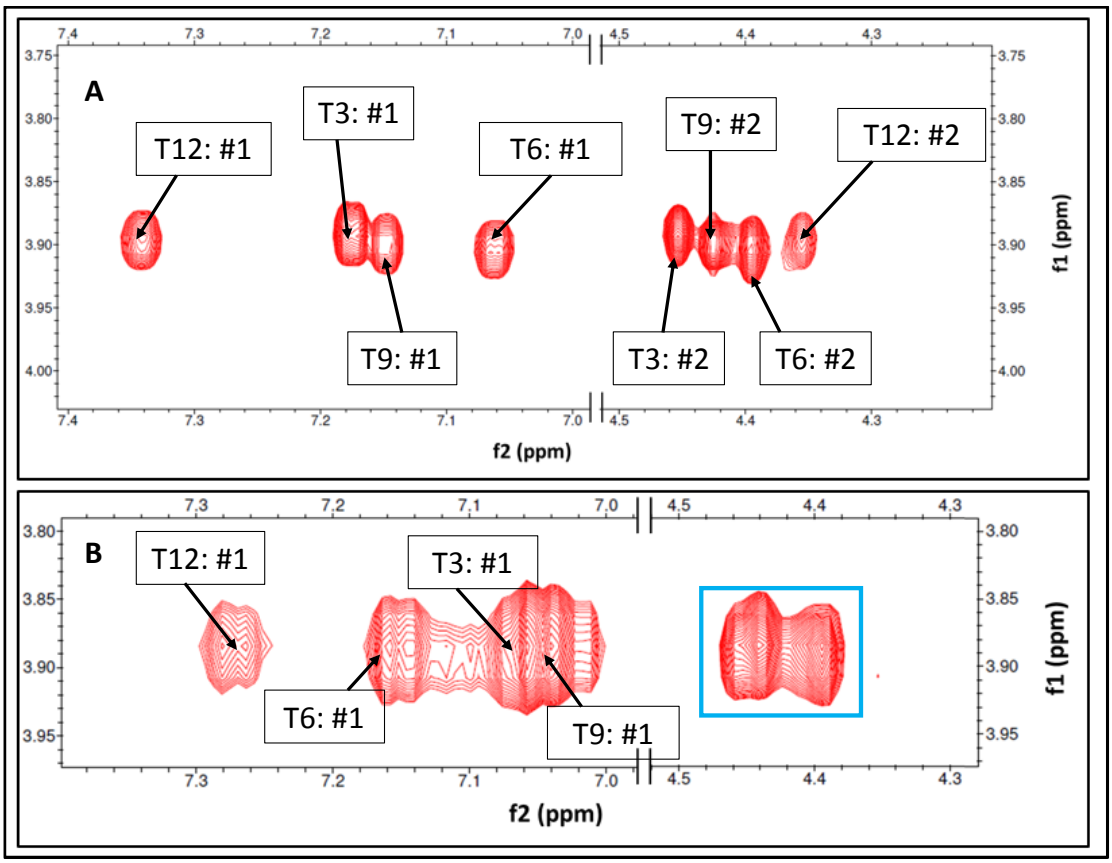


Figure 16. The 600 MHz 2D TOCSY spectrum of AFGP8-BS (A) and 800 MHz 2D TOCSY spectrum of AFGP8-TB (B).
Note: The regions focus on the TOCSY correlations involving the following protons: amide proton (NH), C4 proton (H4), and C4 hydroxyl proton (HO4) of the disaccharides on AFGP8.The notations are as follow: #1 is NH/H4, and #2 is H4/HO4. The blue box in B indicated the ambiguous amino acid position of the C4 hydroxyl protons in AFGP8-TB

The chemical shifts of the C2' hydroxyl proton of all four threonine residues were not assigned for AFGP8-TB because no TOCSY or DQFC correlations were hinting toward that particular hydroxyl proton. For AFGP8-BS, the C2' hydroxyl protons of the threonine residues at position 6 and 9 were assigned but not assigned to position 3 and 12. Since the C2' hydroxyl proton was a labile proton, fast exchange with the residual water or other labile protons could have resulted in the cross-peaks not being present when analyzing the 2D spectrum. Another possibility for not being able to identify the hydroxyl proton of C2' was chemical shift overlaps. Since this chemical shift region does contain a

significant amount of chemical shift overlaps, the cross peak involving the C2' hydroxyl proton could be overlapping with another cross peak thus preventing possible identification of the chemical shift.

IV.5. ^{1}H and ^{13}C Chemical Shift Table of AFGP8-BS and AFGP8-TB

The chemical shifts of all non-labile proton chemical shifts were confirmed using the 600 MHz 2D HMQC spectrum, as shown in **Figure 17** and **Figure 18** for AFGP8-BS. All the chemical shifts assigned to the amide and hydroxyl protons did not show a cross-peak with a carbon chemical shift. For both AFGP8, the C1 carbon chemical shift was not shown because the spectral width used on the carbon dimension did not cover that chemical shift.

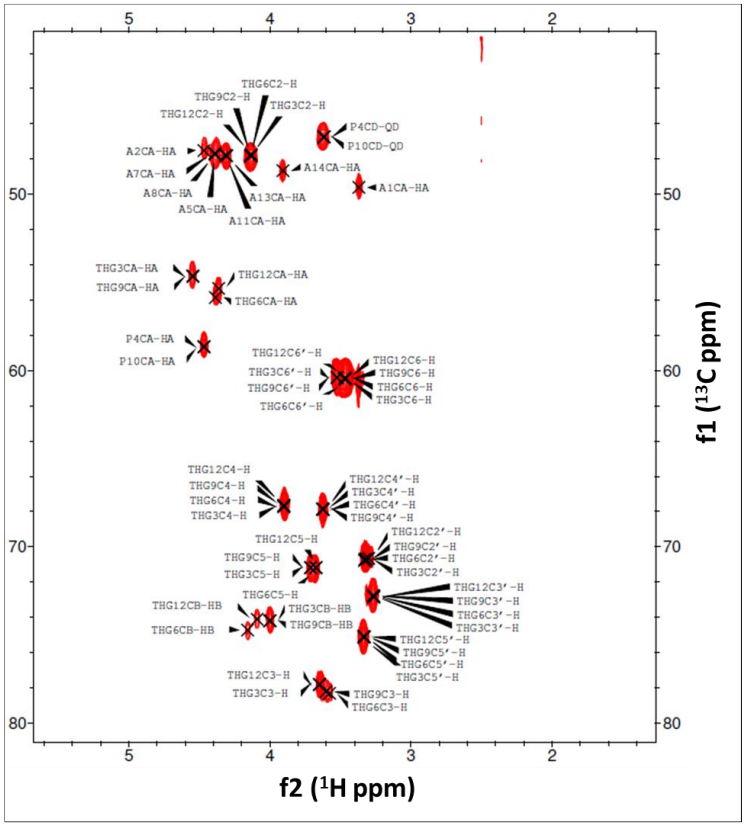


Figure 17. The 600 MHz 2D HMQC spectrum of AFGP8-BS showing the cross peak region between the alpha proton to alpha carbon.

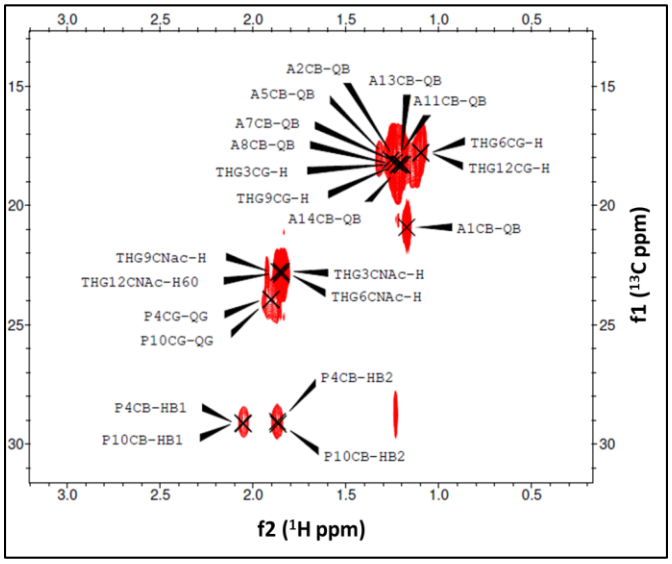


Figure 18. The 600 MHz 2D HMQC spectrum of AFGP8-BS showing the cross peak region between the methyl protons to methyl carbon.

The proton chemical shifts of AFGP8-BS are shown in **Table 3** with the exceptions of the amine protons of alanine 1 and the two C2' hydroxyl protons of the disaccharide on threonine 3 and 12. The "Methyl" column of the disaccharide corresponds to the methyl protons of the acetyl group. The notation "O#" was used for the hydroxyl protons of the disaccharide. The ^{13}C chemical shift assignment table is shown in **Table 4**. The C1 chemical shift was not assigned because the spectral width used on the ^{13}C dimension on the 2D HMQC was not large enough to cover the C1 chemical shift.

For AFGP8-TB, the proton and ^{13}C chemical shifts are shown in **Table 5** and **Table 6**. The chemical shifts of the amine protons of alanine 1 were not assigned. None of the C2' hydroxyl proton (O2) chemical shifts were assigned since it was unclear where the chemical shifts would be. The C1 carbon chemical shifts of AFGP8-TB were not assigned due to the same reasoning as mentioned in AFGP8-BS above.

Table 3: ^{1}H Chemical Shift of AFGP8-BS

Residue	NH	Hα	Hβ	Hγ	Hδ							
Ala 1	-	3.37	1.17	-	-							
Ala 2	8.18	4.47	1.22	-	-							
Thr 3*	8.07	4.55	4.00	1.23	-							
Pro 4	-	4.68	2.05; 1.88	1.90	3.63							
Ala 5	8.20	4.38	1.26	-	-							
Thr 6*	7.72	4.39	4.16	1.12	-							
Ala 7	8.12	4.38	1.23	-	-							
Ala 8	8.19	4.40	1.22	-	-							
Thr 9*	7.90	4.55	4.00	1.22	-							
Pro 10	-	4.70	2.05; 1.88	1.90	3.63							
Ala 11	8.23	4.31	1.24	-	-							
Thr 12*	7.64	4.36	4.09	1.09	-							
Ala 13	8.00	4.31	1.21	-	-							
Ala 14	7.95	3.91	1.20	-	-							
Disaccharide	NH	Methyl	H1	H2	H3	H4	H5	H6	O2	O3	O4	O6
α-GalNAc3	7.18	1.83	4.88	4.18	3.64	3.90	3.71	3.47	-	-	4.45	4.63
α-GalNAc6	7.06	1.87	4.73	4.13	3.59	3.90	3.68	3.46	-	-	4.40	4.62
α-GalNAc9	7.15	1.83	4.88	4.13	3.61	3.90	3.71	3.47	-	-	4.42	4.63
α-GalNAc12	7.34	1.85	4.79	4.14	3.65	3.90	3.68	3.46	-	-	4.35	4.62
β-Gal3	-	-	4.24	3.31	3.27	3.62	3.34	3.53	-	4.80	4.39	3.53
β-Gal6	-	-	4.19	3.32	3.27	3.62	3.34	3.47	3.85	4.80	4.39	3.47
β-Gal9	-	-	4.21	3.32	3.27	3.62	3.34	3.53	3.77	4.80	4.37	3.53
β-Gal12	-	-	4.29	3.30	3.27	3.62	3.34	3.47	-	4.80	4.35	3.47

Table 4: ^{13}C Chemical Shifts of AFGP8-BS

Residue	Cα	Cβ	Cγ	Cδ		
Ala 1	49.61	20.91	-	-		
Ala 2	47.51	18.28	-	-		
Thr 3*	54.63	74.19	18.29	-		
Pro 4	58.63	29.10	23.94	46.74		
Ala 5	47.69	18.11	-	-		
Thr 6*	55.86	74.73	17.77	-		
Ala 7	47.68	18.28	-	-		
Ala 8	47.73	18.28	-	-		
Thr 9*	54.68	74.20	18.29	-		
Pro 10	58.63	29.10	23.94	46.74		
Ala 11	47.82	18.20	-	-		
Thr 12*	55.34	74.12	17.80	-		
Ala 13	47.82	18.26	-	-		
Ala 14	48.65	18.30	-	-		
Disaccharide	C1	C2	C3	C4	C5	C6
α-GalNAc3	-	47.82	77.79	67.69	71.19	60.40
α-GalNAc6	-	47.82	78.31	67.69	71.19	60.40
α-GalNAc9	-	47.82	78.22	67.69	71.19	60.40
α-GalNAc12	-	47.82	77.84	67.69	71.19	60.40
β-Gal3	-	70.70	72.81	67.87	75.12	60.40
β-Gal6	-	70.70	72.81	67.87	75.12	60.40
β-Gal9	-	70.70	72.81	67.87	75.12	60.40
β-Gal12	-	70.70	72.81	67.87	75.12	60.40

Table 5: ^{1}H Chemical Shifts of AFGP8-TB

Residue	NH	Hα	Hβ	Hγ	Hδ							
Ala 1	-	3.38	1.18	-	-							
Ala 2	8.21	4.53	1.25	-	-							
Thr 3*	7.92	4.38	4.16	1.13	-							
Ala 4	8.19	4.37	1.23	-	-							
Ala 5	8.18	4.41	1.21	-	-							
Thr 6*	7.90	4.55	4.01	1.22	-							
Pro 7	-	4.48	2.05; 1.88	1.88	3.62							
Ala 8	8.21	4.38	1.26	-	-							
Thr 9*	7.71	4.38	4.16	1.13	-							
Ala 10	8.10	4.47	1.22	-	-							
Ala 11	8.19	4.36	1.21	-	-							
Thr 12*	7.88	4.57	3.97	1.23	-							
Pro 13	-	4.38	2.01; 1.86	1.86	3.61							
Ala 14	7.86	3.87	1.20	-	-							
Disaccharide	NH	Methyl	H1	H2	H3	H4	H5	H6	O2	O3	O4	O6
α-GalNAc3	7.05	1.87	4.72	4.15	3.62	3.90	3.68	3.47	-	-	4.43	4.62
α-GalNAc6	7.15	1.84	4.88	4.13	3.61	3.90	3.72	3.48	-	-	4.38	4.63
α-GalNAc9	7.07	1.87	4.72	4.14	3.58	3.90	3.68	3.47	-	-	4.45	4.62
α-GalNAc12	7.35	1.81	4.91	4.11	3.64	3.90	3.72	3.48	-	-	4.38	4.63
β-Gal3	-	-	4.23	3.28	3.27	3.62	3.34	3.47	-	4.79	4.33	4.59
β-Gal6	-	-	4.21	3.28	3.27	3.62	3.34	3.53	-	4.79	4.38	4.60
β-Gal9	-	-	4.20	3.28	3.27	3.62	3.34	3.57	-	4.79	4.38	4.59
β-Gal12	-	-	4.27	3.30	3.30	3.62	3.34	3.53	-	4.79	4.38	4.60

Table 6: ^{13}C Chemical Shifts of AFGP8-TB

Residue	Cα	Cβ	Cγ	Cδ		
Ala 1	49.53	20.80	-	-		
Ala 2	47.55	18.01	-	-		
Thr 3*	55.87	74.82	18.29	-		
Ala 4	47.53	18.36	-	-		
Ala 5	47.79	17.68	-	-		
Thr 6*	54.58	74.02	18.64	-		
Pro 7	58.62	28.98	24.22	46.82		
Ala 8	47.69	18.01	-	-		
Thr 9*	55.87	74.82	18.29	-		
Ala 10	47.53	18.69	-	-		
Ala 11	47.51	17.65	-	-		
Thr 12*	54.00	73.99	18.64	-		
Pro 13	59.08	28.98	24.22	46.82		
Ala 14	48.65	18.30	-	-		
Disaccharide	C1	C2	C3	C4	C5	C6
α-GalNAc3	-	47.86	78.24	67.67	71.27	60.40
α-GalNAc6	-	47.96	78.16	67.67	71.27	60.40
α-GalNAc9	-	47.86	78.27	67.67	71.27	60.40
α-GalNAc12	-	47.96	77.91	67.67	71.27	60.40
β-Gal3	-	70.63	72.81	67.92	75.12	60.40
β-Gal6	-	70.63	72.81	67.92	75.12	60.40
β-Gal9	-	70.63	72.81	67.92	75.12	60.40
β-Gal12	-	70.63	72.81	67.92	75.12	60.40

IV.6. Secondary Structure Prediction for AFGP8 Using ^{1}H and ^{13}C Random Coil Chemical Shift Index

The proton and ^{13}C chemical shifts of both AFGP8 were compared to the random coil chemical shift (RCCS) index generated by Tremblay et al. [51]. The secondary structure prediction of AFGP8-BS and AFGP8-TB, using the alpha proton (Hα) chemical shift, were shown in **Table 7** and **Table 8**. The second column contains the alpha proton (Hα) chemical shift obtained experimentally in this thesis work. The third column contains the Hα chemical shift generated by Tremblay et al.[51] The value (1, 0, or -1) on the scoring column was assigned in respect to the value of "Hα (Ref*)" ± 0.10 ppm following the method described in Materials and Methods section III.5. The ± 0.10 ppm uncertainty was used because it was used by Wishart et al.[49]

Table 7: ^{1}H Secondary Structure Prediction of AFGP8-BS

Residue	Hα (Expt)	Hα (Ref*)	Score
Ala 1	3.37	4.47	-1
Ala 2	4.47	4.47	0
Thr 3*	4.55	4.41	1
Pro 4	4.68	4.50	1
Ala 5	4.38	4.47	0
Thr 6*	4.39	4.41	0
Ala 7	4.38	4.47	0
Ala 8	4.40	4.47	0
Thr 9*	4.55	4.41	1
Pro 10	4.70	4.50	1
Ala 11	4.31	4.47	-1
Thr 12*	4.36	4.41	0
Ala 13	4.31	4.47	-1
Ala 14	3.91	4.47	-1

Table 8: ^{1}H Secondary Structure Prediction of AFGP8-TB

Residue	Hα (Expt)	Hα (Ref*)	Score
Ala 1	3.38	4.47	-1
Ala 2	4.53	4.47	0
Thr 3*	4.38	4.41	1
Ala 4	4.37	4.47	0
Ala 5	4.41	4.47	0
Thr 6*	4.55	4.41	1
Pro 7	4.48	4.50	0
Ala 8	4.38	4.47	0
Thr 9*	4.38	4.41	0
Ala 10	4.47	4.47	1
Ala 11	4.36	4.47	-1
Thr 12*	4.57	4.41	1
Pro 13	4.38	4.50	-1
Ala 14	3.87	4.47	-1

Table 9 and **Table 10** show the secondary structure prediction for AFGP8-BS and AFGP8-TB, using the alpha carbon (Cα) and beta carbon (Cβ) ^{13}C chemical shifts. The second and fifth columns are the Cα and Cβ chemical shifts obtained experimentally. The second and fifth columns are the Cα and Cβ chemical shifts generated by Tremblay et al.[51] The value (1, 0, or -1) on the scoring columns were assigned in respect to the chemical shift value of "Cα (Ref*)" ± 0.7 ppm and "Cβ (Ref*)" ± 0.7 ppm following the method describe in Materials and Methods section III.5. The ± 0.7 ppm uncertainty was used because it was used by Wishart et al.[50]

Table 9: ^{13}C Secondary Structure Prediction for AFGP8-BS

Residue	Cα (Expt)	Cα (Ref*)	Score	Cβ (Expt)	Cβ (Ref*)	Score
Ala 1	49.61	51.4	-1	20.91	20.9	0
Ala 2	47.51	51.4	-1	18.28	20.9	-1
Thr 3*	54.63	61.3	-1	74.19	69.6	1
Pro 4	58.63	62.9	-1	29.1	32.2	-1
Ala 5	47.69	51.4	-1	18.11	20.9	-1
Thr 6*	55.86	61.3	-1	74.73	69.6	1
Ala 7	47.68	51.4	-1	18.28	20.9	-1
Ala 8	47.73	51.4	-1	18.28	20.9	-1
Thr 9*	54.68	61.3	-1	74.2	69.6	1
Pro 10	58.63	62.9	-1	29.1	32.2	-1
Ala 11	47.82	51.4	-1	18.2	20.9	-1
Thr 12*	55.34	61.3	-1	74.12	69.6	1
Ala 13	47.82	51.4	-1	18.26	20.9	-1
Ala 14	48.65	51.4	-1	18.3	20.9	-1

Table 10: ^{13}C Secondary Structure Prediction for AFGP8-TB

Residue	Cα (Expt)	Cα (Ref*)	Score	Cβ (Expt)	Cβ (Ref*)	Score
Ala 1	49.53	51.4	-1	20.80	20.9	0
Ala 2	47.55	51.4	-1	18.01	20.9	-1
Thr 3*	55.87	61.3	-1	74.82	69.6	1
Ala 4	47.53	51.4	-1	18.36	32.2	-1
Ala 5	47.79	51.4	-1	17.68	20.9	-1
Thr 6*	54.58	61.3	-1	74.02	69.6	1
Pro 7	58.62	62.9	-1	28.98	20.9	-1
Ala 8	47.69	51.4	-1	18.01	20.9	-1
Thr 9*	55.87	61.3	-1	74.82	69.6	1
Ala 10	47.53	51.4	-1	18.69	32.2	-1
Ala 11	47.51	51.4	-1	17.65	20.9	-1
Thr 12*	54.00	61.3	-1	73.99	69.6	1
Pro 13	59.08	62.9	-1	28.98	20.9	-1
Ala 14	48.65	51.4	-1	18.30	20.9	-1

IV.7. Distance Constraints for Structure Calculation

The distance constraints were generated in CYANA using the integrated NOE volume obtained from the 2D NOESY spectrum of AFGP8-BS and AFGP8-TB. The distribution of distance constraints generated was shown in **Figure 19** for AFGP8-BS (A) and AFGP8-TB (B).

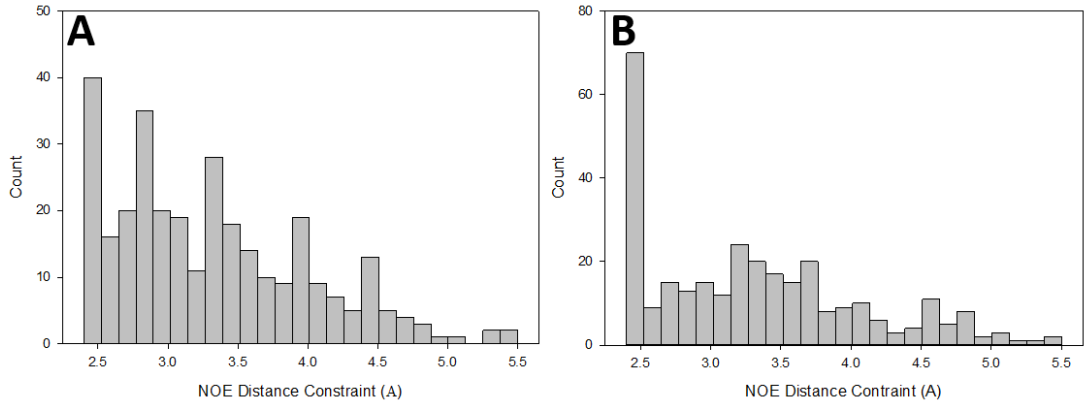


Figure 19. The distribution of distance constraints generated from the integrated NOE cross peak volume using CYANA for AFGP8-BS (A) and AFGP8-TB (B).

The numbers of distance constraints obtained for AFGP8-BS and AFGP8-TB were shown in **Table 11** and **Table 12**. The distance constraints were listed in three NOE categories: intra-residue, short range, and medium range. The intra-residue NOE involved any NOE cross peak observed that was the result of any protons from one amino acid with itself. The short range NOE involved NOE cross peaks that were the result of a particular amino acid to the adjacent amino acid, the amino acid one peptide bond over to the right or left. The medium range NOE involved NOE cross peaks that were the result of a particular amino acid to any amino acid 2 to 5 amino acid position away in the primary sequence. The number in parenthesis indicated the number of distance constraints involved with the disaccharide of AFGP8.

Table 11: Numbers of NOE for AFGP8-BS

NOE Range	Numbers of NOE (Disaccharide)
Intra-Residue	237 (181)
Short Range	56 (15)
Medium Range	18 (12)
Total	311 (208)

Table 12: Numbers of NOE for AFGP8-TB

NOE Range	Numbers of NOE (Disaccharide)
Intra-Residue	237 (180)
Short Range	57 (19)
Medium Range	9 (4)
Total	303 (203)

A total of 311 distance constraints was assigned for AFGP8-BS. Of the 311 distance constraints, 238 resulted from intra-residue NOE. For short range and medium range NOE, 56 and 18 distance constraints were obtained. For AFGP8-TB, a total of 303 distance constraints was assigned for AFGP8-TB. Of the 303 distance constraints, 237 resulted from intra-residue NOE. For short range and medium range NOE, 57 and 9 distance constraints were obtained. There was no long range (i to i+6 or greater) NOE detected between the amino acids, including the disaccharides, for both AFGP8-BS and AFGP8-TB.

IV.8. Three-Dimensional Structures of AFGP8-BS and AFGP8-TB

IV.8.A. Comparing the Three-Dimensional Structures of AFGP8-BS and AFGP8-TB

From the 100 total structures generated in YASARA, the ten structures with the least number of distance violation and lowest potential energy of AFGP8-BS and AFGP8-TB were overlaid and shown in **Figure 20**. The color codes for both

AFGP8-BS, shown in **Figure 20A**, and AFGP8-TB, shown in **Figure 20B**, were blue (alanine), red (glycosylated threonine), and green (proline). For AFGP8-BS and AFGP8-TB, the backbone RMSD of the 10 structures from the average was 2.72 Å ± 0.49 Å and 3.87 Å ± 0.72 Å. The average number of distance violation obtained from YASARA at the end of the structure calculation was about 40 violations for AFGP8-BS and 60 violations for AFGP8-TB. The largest distance violation was about 0.5 Å for AFGP8-BS and 0.7 Å for AFGP8-TB after the allowed 30% plus and minus distance allowed from the distant constraints including the solvent refinement process in DMSO.

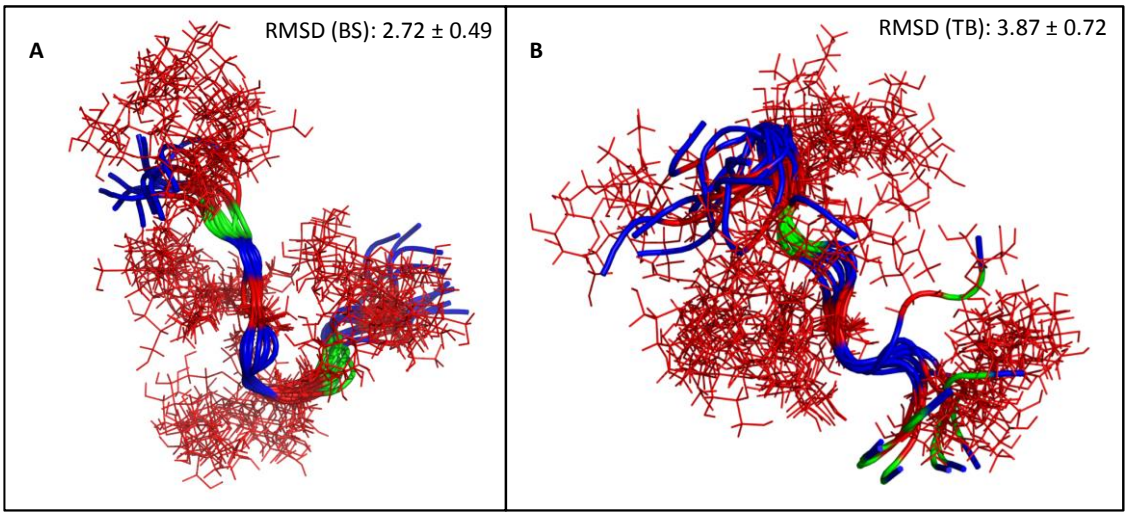


Figure 20. The 10 YASARA generated structures of AFGP8-BS (A) and AFGP8-TB (B), going from N-terminus (left) to C-terminus (right). The color codes were blue (alanine), red (glycosylated threonine), and green (proline).

From **Figure 21**, the overall structure seemed different from one another. The protein backbone of the ten structures of AFGP8-BS (red) and AFGP8-TB (blue) were overlaid from using *PyMol* to see how much the backbone deviated from each other. From aligning the backbone of the 10 structures of AFGP8-BS and AFGP8-TB, the RMSD was 2.22 Å ± 0.49 Å and 3.57 Å ± 0.78 Å. The RMSD obtained by aligning the one structure, with the least number and distance

violation and lowest potential energy, of AFGP8-BS and AFGP8-TB with each other was 5.06 Å.

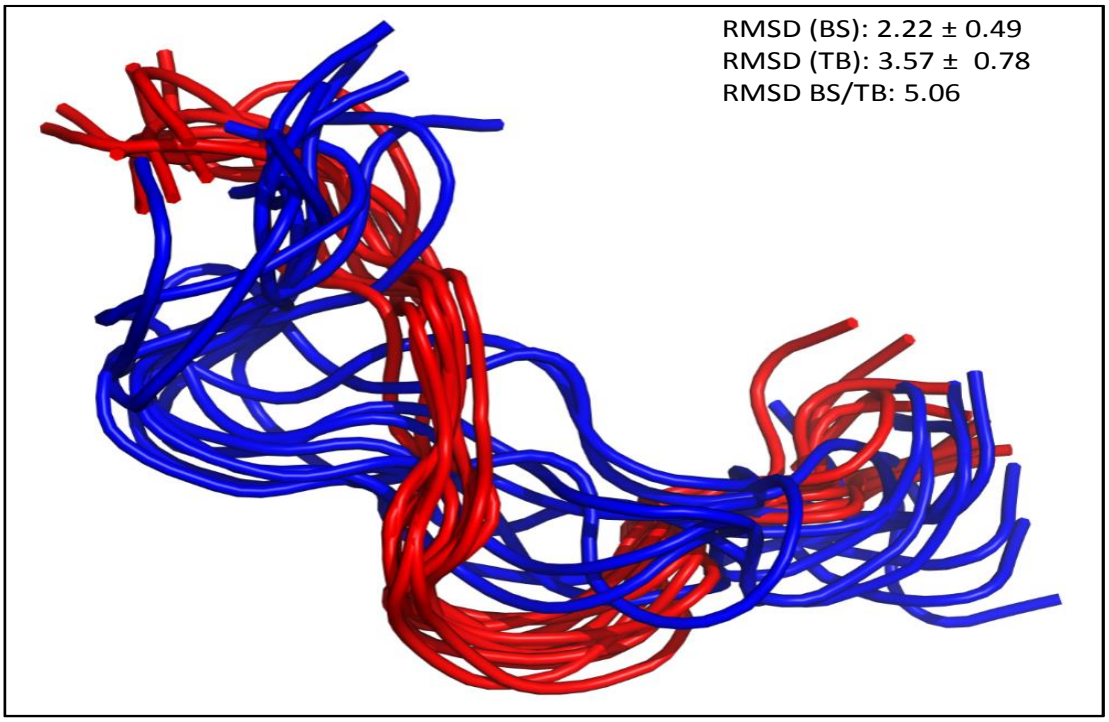


Figure 21. The overlaid of the YASARA backbone structures of AFGP8-BS (red) and AFGP8-TB (blue) going from N-terminus (left) to C-terminus (right).

While the positions of the prolines in the primary sequence were different between AFGP8-BS and AFGP8-TB, there were two motifs, consisting of *AT*PA* and two *AT*AA*, conserved in both primary sequence. As suggested by the study done by Lane et al.[52], the three-dimensional structure of the motifs of AFGP8-BS and AFGP8-TB in DMSO were overlaid, as shown in **Figure 22** and **Figure 23**. The focus for **Figure 22** was aligning the two motifs *AT*PA* present in each AFGP8-BS and AFGP8-TB while **Figure 23** focused on aligning the two motifs *AT*AA*. The structures of AFGP8-TB were color coded as blue (backbone) and red (glycosylated threonine) for DMSO and purple (backbone) and green (glycosylated threonine) for water. The motif of *AT*AA* was shown in part A and C, and the motif of *AT*PA* were shown in part B and D.

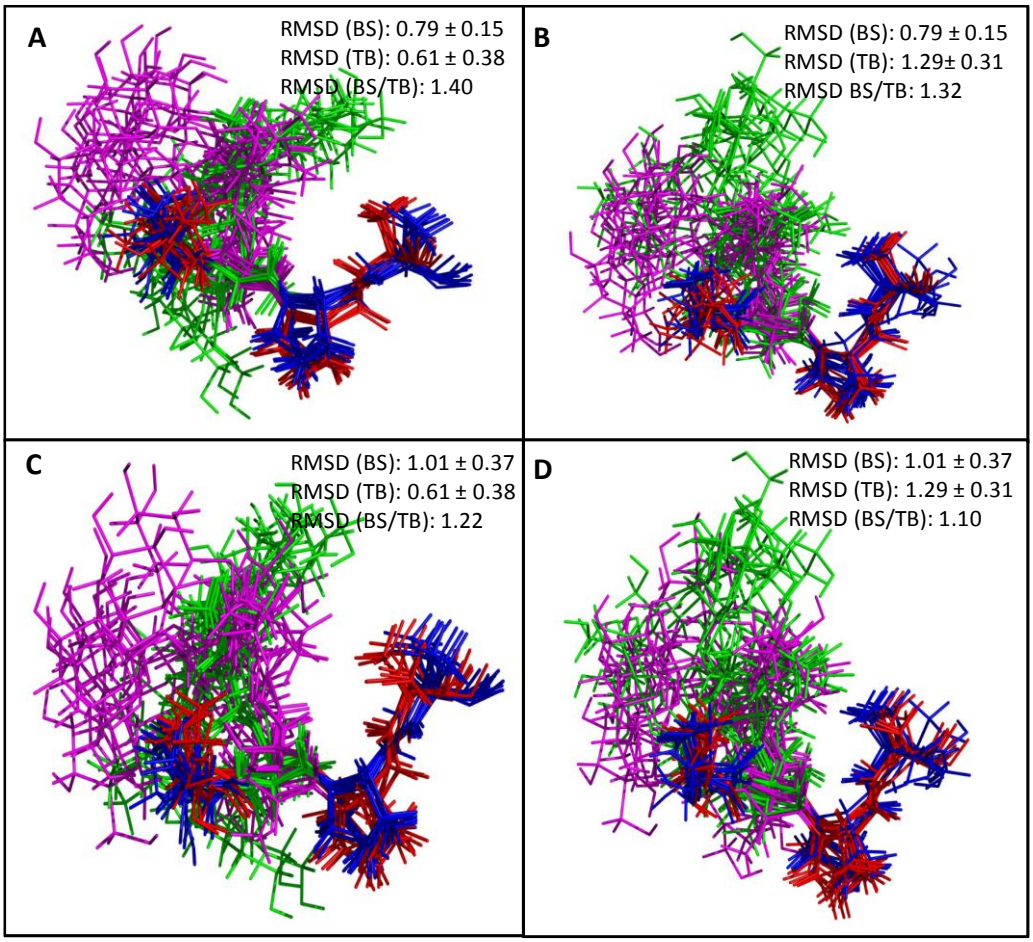


Figure 22. Overlaid of two motif repeats, *Ala-Thr*-Pro-Ala*, present in the primary sequence of AFGP8-BS and AFGP8-TB.

Note: The backbone and glycosylated threonine were shown as red and purple for AFGP8-BS and shown as blue and green for AFGP8-TB. The segments of AFGP8-BS and AFGP8-TB were overlaid as listed: AFGP8-BS segment 2-5 to AFGP8-TB segment 5-8 (part A), AFGP8-BS segment 2-5 to AFGP8-TB segment 11-14 (part B), AFGP8-BS segment 8-11 to AFGP8-TB segment 5-8 (part C), and AFGP8-BS segment 8-11 to AFGP8-TB segment 11-14 (part D).

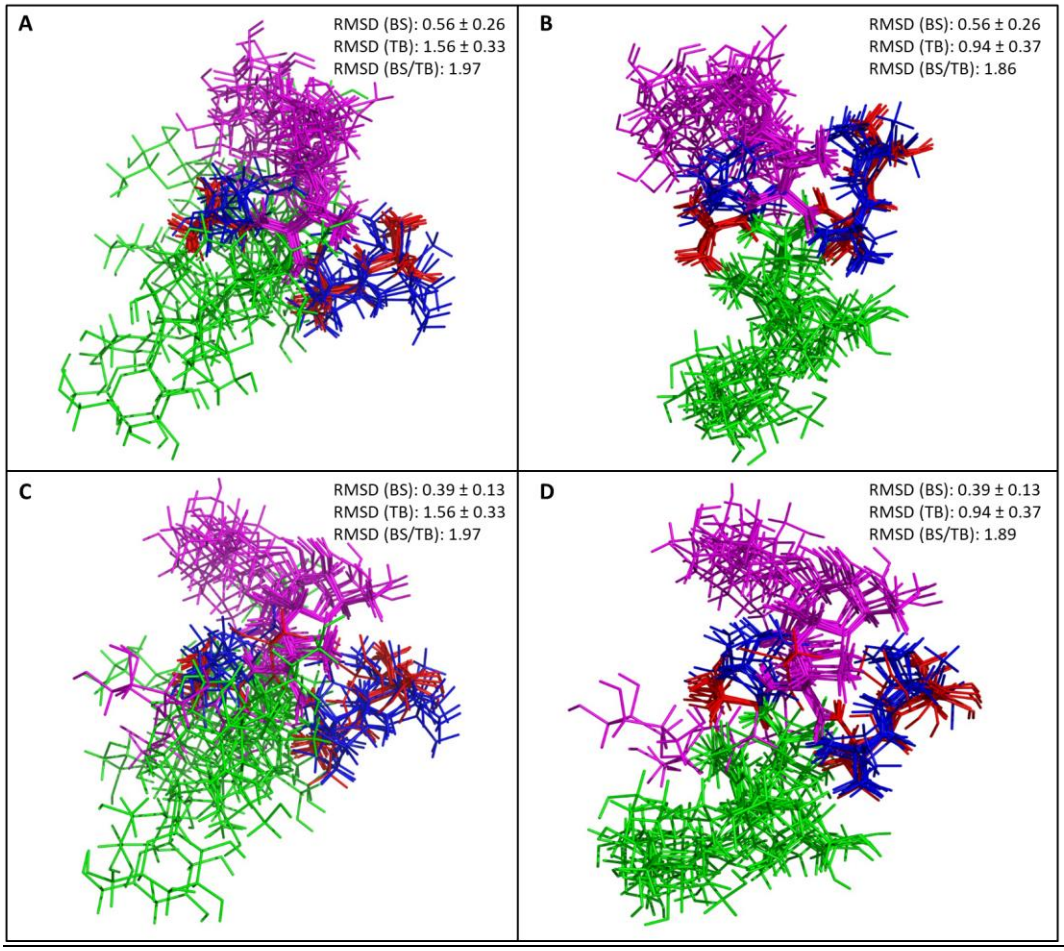


Figure 23. Overlaid of two motif repeats, *Ala-Thr*-Ala-Ala*, present in the primary sequence of AFGP8-BS and AFGP8-TB.

Note: The backbone and glycosylated threonine were shown as red and purple for AFGP8-BS and shown as blue and green for AFGP8-TB. The segments of AFGP8-BS and AFGP8-TB were overlaid as listed: AFGP8-BS segment 5-8 to AFGP8-TB segment 2-5 (part A), AFGP8-BS segment 5-8 to AFGP8-TB segment 8-11 (part B), AFGP8-BS segment 11-14 to AFGP8-TB segment 2-5 (part C), and AFGP8-BS segment 11-14 to AFGP8-TB segment 8-11 (part D).

IV.8.B. Structure Alignment of AFGP8-TB in DMSO and Water

From **Figure 24**, the structural ensemble of AFGP8-TB (**Figure 24A**) in DMSO was used as generated by YASARA without any further alignment done in *PyMol*. The structural ensemble of AFGP8-TB in water, as shown in **Figure 24B**, was aligned in *PyMol* using the fitting script. No particular residue or segment was chosen as the alignment focus; the structures determined by Lane et al.[52] were aligned to the first structure as obtained. Both structural ensembles used the same color code: blue (alanine), green (proline), and red (threonine).

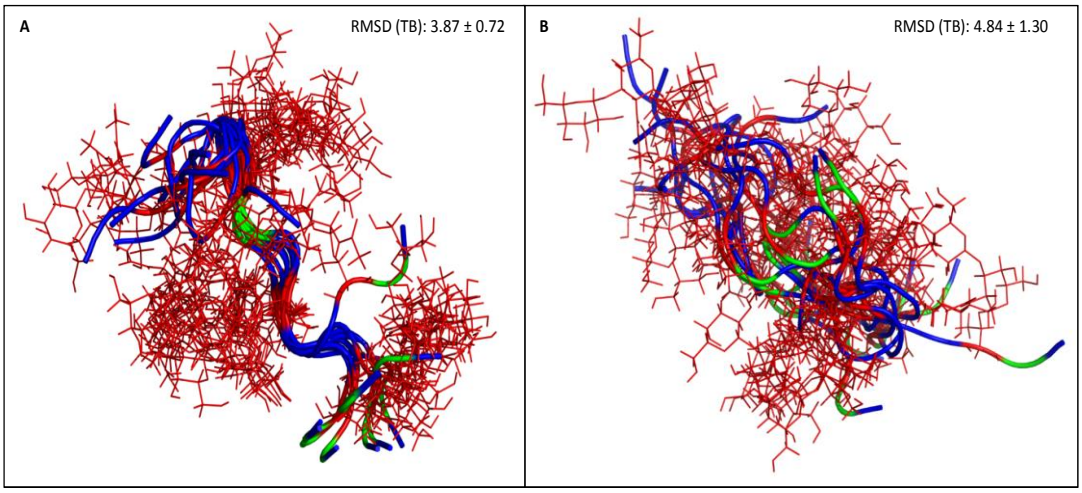


Figure 24. The structure of AFGP8-TB in DMSO (A) generated using YASARA and water (B) generated by Lane et al.[52] The color codes were blue (alanine), red (glycosylated threonine), and green (proline).

The overall structure of AFGP8-TB in DMSO and water were different from one another. Since the disaccharide fluctuated much more in the structures determined in water as compared to DMSO, only the backbone of the structure in DMSO and water were overlaid using *Pymol* for additional comparison as shown in **Figure 25**.

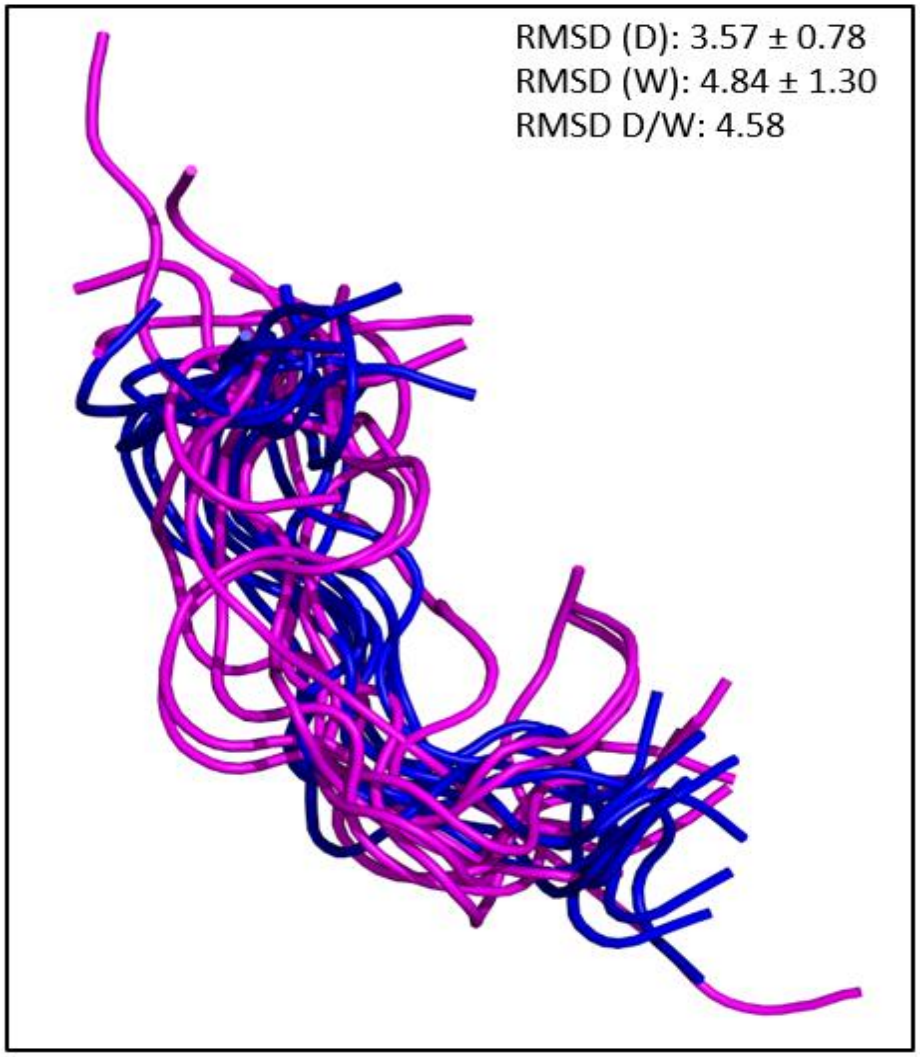


Figure 25. The overlaid the YASARA backbone structures of AFGP8-TB (blue) in dimethyl sulfoxide and AFGP8-TB (purple) in water.

To compare if the three-dimensional structure of the motifs were conserved between water and DMSO, the motifs of the same segment of AFGP8-TB in DMSO and water were overlaid. As shown in **Figure 26**, the structures of AFGP8-TB were color code as blue (backbone) and red (glycosylated threonine) for DMSO and purple (backbone) and green (glycosylated threonine) for water. The motif of *AT*AA* in position 2-5 and 8-11 was shown in part A and C. The motif of *AT*PA* in position 5-8 and 11-14 was shown in part B and D.

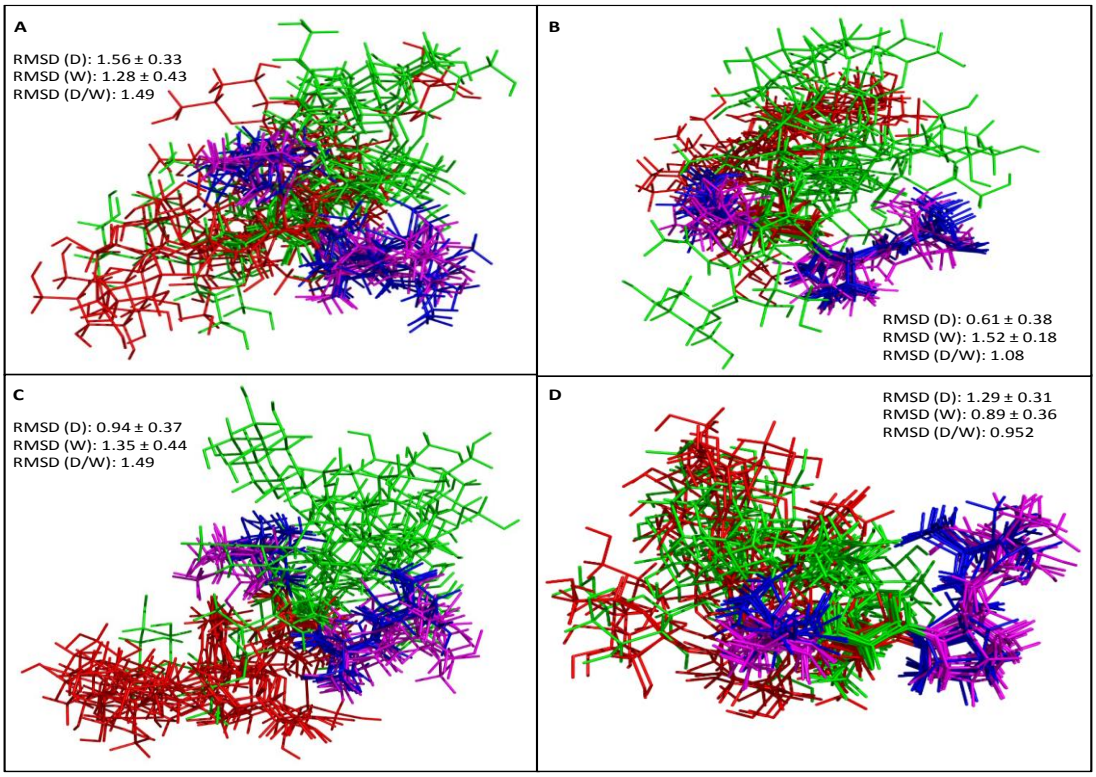


Figure 26. Comparison of motif repeats, *Ala-Thr*-Pro-Ala* and *Ala-Thr*-Ala-Ala*, present in the primary sequence of AFGP8-TB.

Note: The backbone and glycosylated threonine were shown as purple and green for structure determined in water (W). For the structure determined in DMSO (D), the backbone and glycosylated threonine were shown as blue and red. The motif Ala-Thr*-Ala-Ala was shown in part A (residues 2-5) and part C (residues 8-11). The motif Ala-Thr*-Pro-Ala was shown in part B (residues 5-8) and part D (residues 11-14).

The structure of synthetic AFGP (sAFGP), as shown in **Figure 27**, showed that sAFGP adopts a three-dimensional structure consisting of a hydrophilic side and hydrophobic side. The backbone was shown in cyan and disaccharide was shown in gray. Unlike natural AFGP, sAFGP was well-structured, but contains the conserved motif *AT*AA* like natural AFGP. The two motifs *AT*AA* of sAFGP were aligned to the natural AFGP motifs in water and DMSO as shown in **Figure 28** (AFGP8-BS in DMSO), **Figure 29** (AFGP8-TB in DMSO), and **Figure 30** (AFGP8-TB in water). There were four parts (A-D) because each sAFGP and natural AFGP contains two motifs *AT*AA*, so there were four possible alignments when comparing.

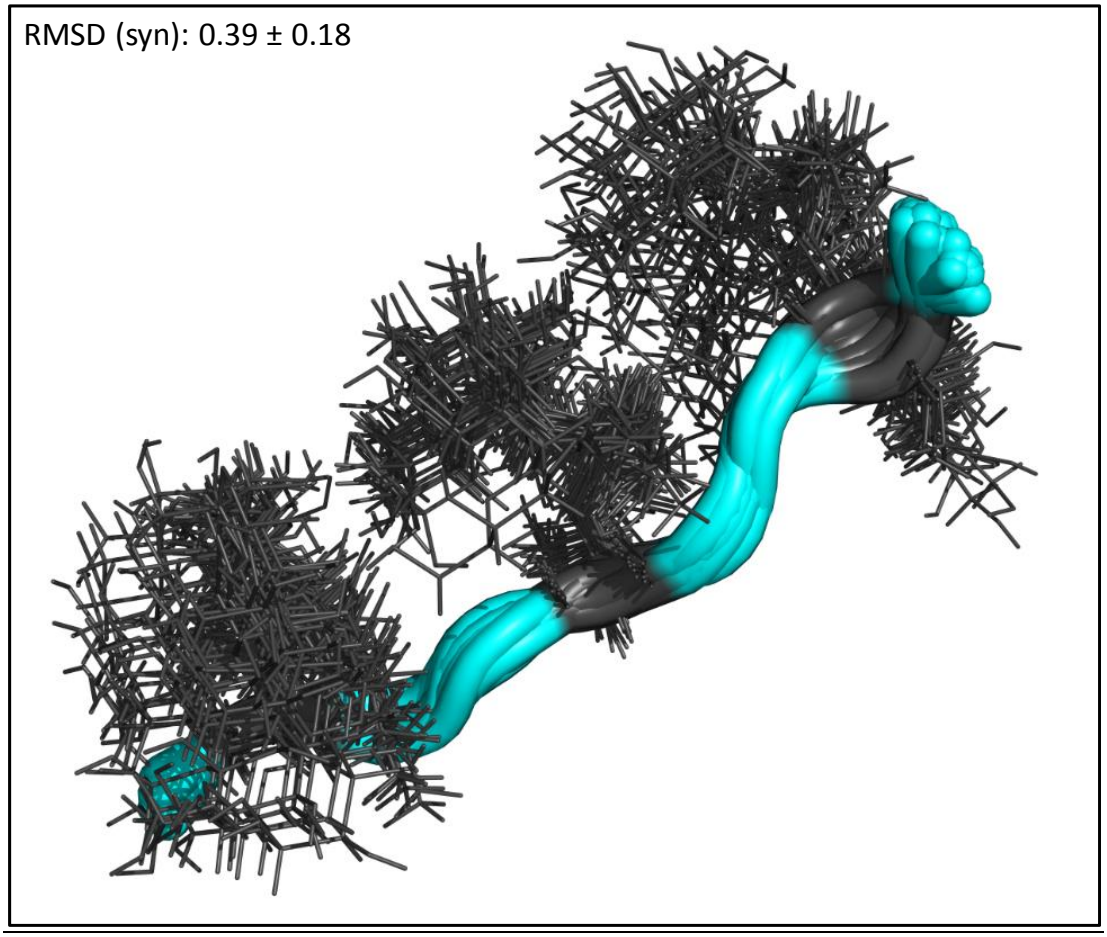


Figure 27. The structural ensemble of synthetic antifreeze glycoproteins (sAFGP) at 5 ˚C.

Note: The primary sequence consists of Ala-Thr*-Ala-Ala-Thr*-Ala-Ala-Thr*-Ala in which the threonine (Thr*) is glycosylated with the disaccharide *beta-D-galactosyl-(1,3)-alpha-N-acetyl-D-galactosamine*. The backbone and glycosylated threonines were shown as cyan and gray. The sAFGP structural ensemble was provided by Shin-Ichiro Nishimura.

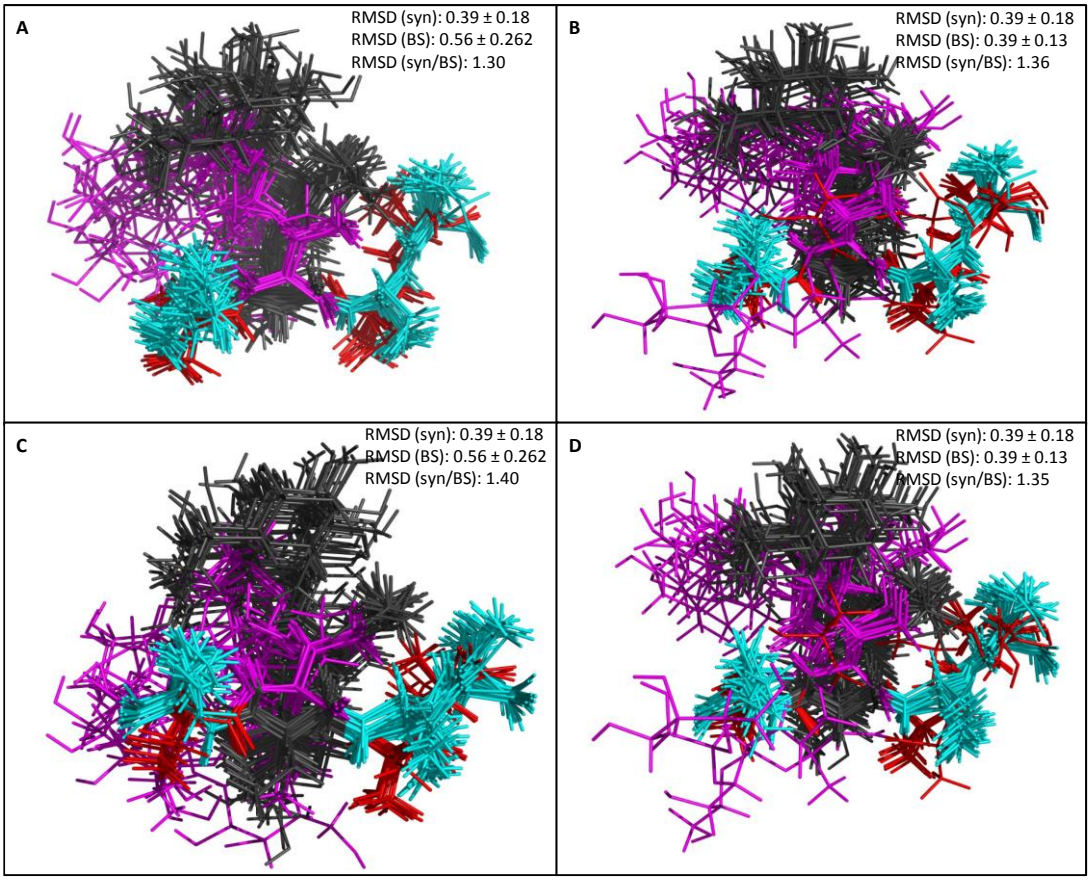


Figure 28. Overlaid of the two motif repeats *Ala-Thr*-Ala-Ala* present in synthetic antifreeze glycoproteins (sAFGP) and AFGP8-BS (DMSO structure).
Note: The backbone and glycosylated threonines were shown as cyan and gray for sAFGP (syn). The backbone and glycosylated threonines were shown as red and purple for AFGP8-BS (BS). Segment 1-4 of sAFGP were overlaid with segment 5-8 and 11-14 of AFGP8-BS in part A and B. Segment 4-7 of sAFGP were overlaid with segment 5-8 and 11-14 of AFGP8-BS in part C and D.

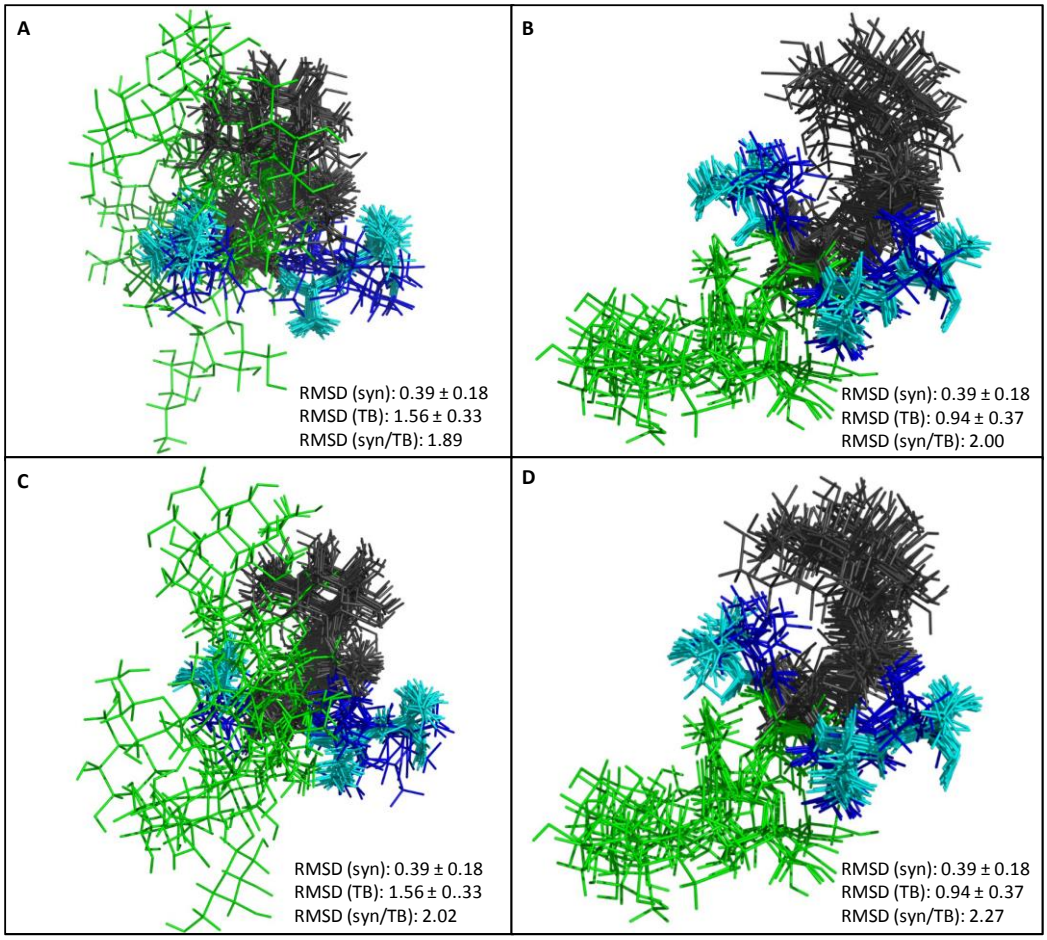


Figure 29. Overlaid of the two motif repeats *Ala-Thr*-Ala-Ala* present in synthetic antifreeze glycoproteins (sAFGP) and AFGP8-TB (DMSO structure).
Note: The backbone and glycosylated threonines were shown as cyan and gray for sAFGP (syn). The backbone and glycosylated threonines were shown as blue and green for AFGP8-TB (TB). Segment 1-4 of sAFGP were overlaid with segment 2-5 and 8-11 of AFGP8-TB in part A and B. Segment 4-7 of sAFGP were overlaid with segment 2-5 and 8-11 of AFGP8-TB in part C and D.

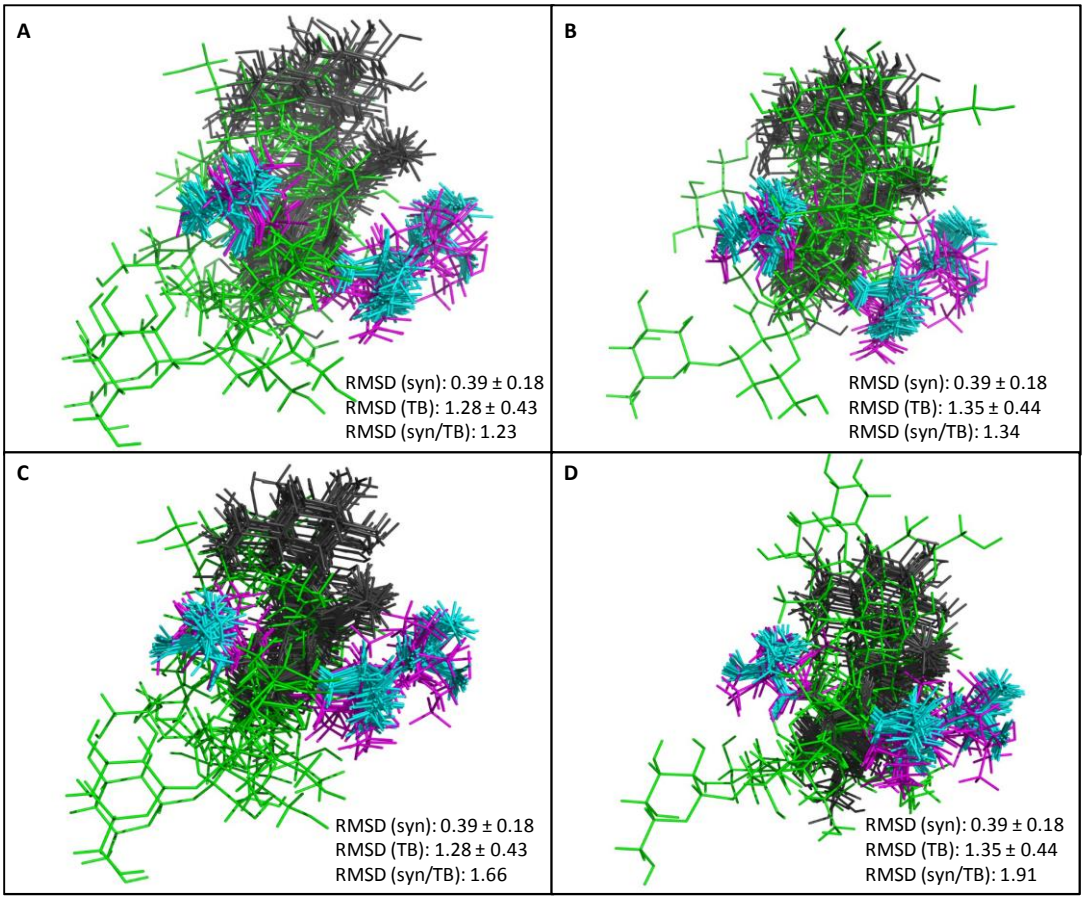


Figure 30. Overlaid of the two motif repeats *Ala-Thr*-Ala-Ala* present in synthetic antifreeze glycoproteins (sAFGP) and AFGP8-TB, using the structural ensemble generated by Andrew Lane.

Note: The backbone and glycosylated threonines were shown as cyan and gray for sAFGP (syn). The backbone and glycosylated threonines were shown as blue and green for AFGP8-TB (TB). Segment 1-4 of sAFGP were overlaid with segment 2-5 and 8-11 of AFGP8-TB in part A and B. Segment 4-7 of sAFGP were overlaid with segment 2-5 and 8-11 of AFGP8-TB in part C and D.

IV.9. Diffusion Coefficient of AFGP8-BS and TB in Dimethyl Sulfoxide

The diffusion coefficient of AFGP8 was measured using NMR spectroscopy to see if the proline positions significantly impact their three-dimensional structure. Generated from the NMR diffusion data, the plot of AFGP8-BS (A) and AFGP8-TB (B) diffusion data was shown in **Figure 31** by plotting the normalized peak area of the methyl proton region (0.9 – 1.2 ppm) versus the magnetic gradient used along the Z-axis. Direct comparison of the plots was done by overlaying both AFGP8 data. From the NMR spectral data, the diffusion coefficients determined for the DMSO solvent were 6.69×10^{-10} m^2/s (within AFGP8-BS sample) and 6.63×10^{-10} m^2/s (within AFGP8-TB sample). The diffusion coefficients determined for AFGP8-BS and AFGP8-TB were 6.46×10^{-11} m^2/s and 6.47×10^{-11} m^2/s. The AFGP8 diffusion coefficients in DMSO were compared to the diffusion coefficients determined by Krishnan et al.[54] and Inglis et al.[55] in water at 0 ˚C and 4 ˚C as shown in **Table 13**.

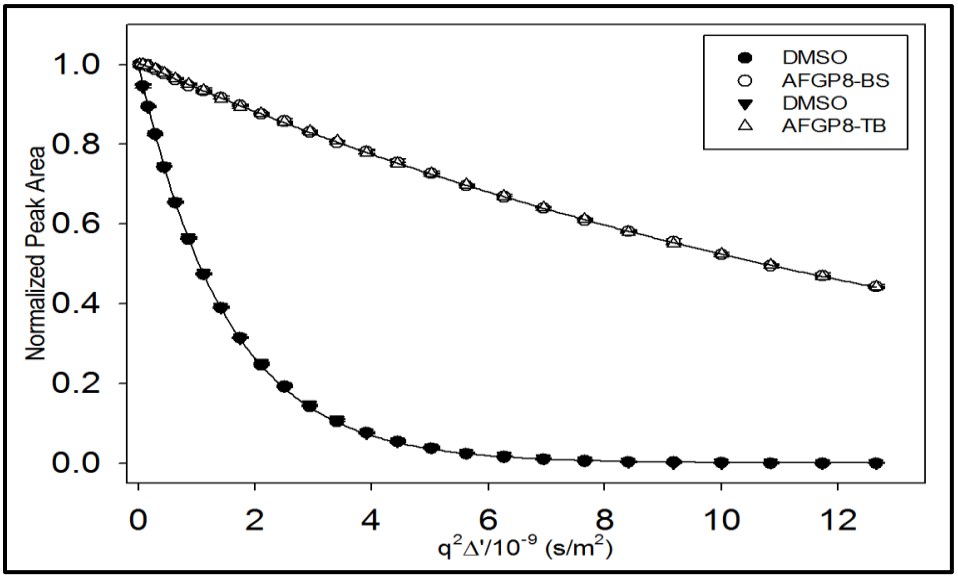


Figure 31. The diffusion plot overlaying the AFGP8-BS and AFGP8-TB data.
Note: The DMSO, from AFGP8-BS and AFGP8-TB, are shown as filled circle and filled triangle symbols. AFGP8-BS and AFGP8-TB are shown as unfilled circle and unfilled triangle symbol. The data were fitted to an exponential trend.

Table 13. NMR Determined Diffusion Coefficient (D) of AFGP

AFGP8	Temperature (˚C)	D (10^{-11} m^2/s)
DMSO (BS & TB)	25	6.46 ± 0.12 & 6.47 ± 0.12
Water (BS) (Krishnan et al.[54])	0	8.16 ± 0.34
Water (Inglis et al.[55])	4	7.31 ± 0.00

V. DISCUSSION AND CONCLUSION

V.1. Comparison of the Structures of AFGP8-BS and AFGP8-TB

V.1.A. Prediction on the Three-Dimensional Structure of AFGP8

In DMSO, AFGP8 was expected to form a more static three-dimensional structure than when in its native solvent, water. Unlike water, which is both a hydrogen bond donor and acceptor, DMSO is only a hydrogen bond acceptor.[56] The difference in the hydrogen bonding network of the two solvents might have disrupted the intrinsically disordered nature of AFGP8, from being highly dynamic in solution, and turned AFGP8 into a more static three-dimensional structure. Comparing the amide proton chemical shifts obtained in DMSO to the chemical shift obtained in water by Lane et al.[52], the amide proton chemical shift of the alanines and threonines in DMSO were more distributed, over a wider chemical shift range, than in water. In water, the clustering of the amide proton of the threonines into one very similar chemical shift region and alanine into another resembles the character of a random coil's amide proton chemical shifts. On the contrary to the amide chemical shift in water, the amide proton chemical shifts in DMSO were distributed over a wider chemical shift range. The distribution of amide proton chemical shifts over a wider range indicated a folded three-dimensional structure. Judging from the amide proton chemical shift region, it was expected that AFGP8 would exhibit a well-folded, three-dimensional structure in DMSO.

In addition to the distribution of chemical shifts, the secondary structure prediction method (SSPM) can predict structures based on the Hα, Cα, and Cβ chemical shifts of proteins in water.[49-50] However, the study done by Tremblay et

al.[51] suggested that it is possible to use SSPM for predicting the structure of proteins in DMSO. From **Table 7** and **Table 8**, using the Hα proton chemical shifts, the SSPM predicted the secondary structures of both AFGP8 as a coil. There were no helix or β-strand region since the score did not show any region in the primary sequence with consecutive of at least four negative ones (-1), for helix, or positive ones (1), for β-strand. Unlike the proton prediction, using the Cα chemical shift shown in **Table 9** and **Table 10**, the SSPM predicted AFGP8-BS and AFGP8-TB to be β-strand throughout its entire primary sequence because a negative one was assigned to all the Cα chemical shift.

When using the secondary structure prediction method, the Cβ chemical shift shows an opposite trend to the Cα chemical shift as mentioned by Wishart and Sykes.[50] The β-strand is the result of a positive one for the Cβ chemical shift and not a negative one like the Cα chemical shift. Wishart and Sykes[50] also noted that the Cβ chemical shift should only be used to predict β-strand region only; any other regions that are not β-strand should be marked as a coil. From **Table 9** and **Table 10**, the Cβ chemical shifts did not label any region in both AFGP8 as β-strand since the score did not have consecutive of at least four positive ones for the amino acids in the primary sequence. Judging from the Cβ chemical shifts, the secondary structure of both AFGP8 were a coil.

The SSPM, using the Hα, Cα, and Cβ chemical shifts, did not yield similar result so it was inconclusive on which secondary structure the method predicted for both AFGP8. The inconsistency in predicting the secondary structure of AFGP8 in DMSO could be because AFGP8 still retain some characteristic of an intrinsically disordered protein. Multiple different conformations of AFGP8 can still exist in DSMO resulting in the observed chemical shifts as an average of all the chemical shifts of the different conformations. While the average of the

ensemble of structures of AFGP8 was defined enough for structural studies via NMR spectroscopy, the chemical shifts obtained might have been an average of multiple different configurations of AFGP in DMSO. If the chemical shifts obtained were the average, then the SSPM will not give an accurate prediction resulting in the observed difference between the SSPM using the Hα, Cα, and Cβ chemical shifts.

V.1.B. Prolines Influence on the Three-Dimensional Structure of AFGP8

For both AFGP8 fractions, two alanines were replaced by prolines but at different positions of the primary sequence while all the position of the glycosylated threonine (*T*/Thr**) remained the same. Differences in the alpha proton chemical shift were observed between AFGP8-BS and AFGP8-TB TOCSY spectra, as shown in **Figure 9**. Even though both AFGP8 contained the same number of motif repeats, such as *Ala-Thr*-Pro-Ala* and *Ala-Thr*-Ala-Ala*, the difference in the overall conformation might have resulted from the difference of the two proline residues within the two AFGP8. Since the prolines of AFGP8-BS were spaced out toward both N-terminus and C-terminus in the primary sequence, that could explain why the overall structural ensemble was more static, regarding lower RMSD value, compared to AFGP8-TB as shown in **Figure 20** and **Figure 21**. Going from N-terminus to C-terminus, the proline residues seemed to stabilize the segment to the right of the proline more resulting in a more static region within the peptide after each proline residues. In addition to the different conformations, in AFGP8-BS, each proline at position 4 and 10 was between two disaccharides. In AFGP8-TB, the proline at position 7 was between two disaccharides while the proline at position 13 only had a disaccharide to the left of it. The disaccharides

could have caused additional shielding/deshielding effect resulting in the chemical shift difference between AFGP8-BS and AFGP-TB.

The proline positions seemed to stabilize the ensemble of structures of AFGP8-BS better compared to AFGP8-TB. Due to the difference in the primary sequence, it was difficult to make a direct comparison of the overall ensemble of structures of both AFGP8 because the backbones were not directly on top of each other. Instead of studying the overall structural ensemble, Lane et al.[52] studied the structure of AFGP8-TB using the conserved motif repeats, two *AT*PA* and two *AT*AA*, present in the primary sequence. The overlaid of the *AT*PA* motifs of AFGP8-BS and AFGP8-TB in **Figure 22** showed that these segments aligned. Not only did the backbone aligned, but also the disaccharide clustered to a common region. From **Figure 23**, the overlaid of the *AT*AA* motifs did not align as well as the *AT*PA* motifs. One reason why the *AT*PA* motifs aligned better between AFGP8-BS and AFGP8-TB (regarding to "RMSD (BS/TB)" values on **Figure 22** and **Figure 23**) was positions of the prolines. Proline has less bond rotational freedom compared to the alanine. This restriction holds the motifs to fewer degrees of freedom resulting in the *AT*PA* motifs between AFGP8-BS and AFGP8-TB to align better regarding the RMSD value. For aligning the *AT*AA* motifs between AFGP8-BS and AFGP8-TB, the clustering of the disaccharides was into two regions, one region for AFGP8-BS and the other for AFGP8-TB, but the clustering of the disaccharide into two different regions was consistent throughout the alignment.

The three-dimensional structure of the motifs seemed to be conserved throughout the ensemble of structures generated for both AFGP8-BS and AFGP8-TB in DMSO. The motifs might contain partial secondary structure characteristics that cannot be observed from the protein as a whole. From a comparison of the

motifs present in both AFGP8, the difference in the overall structure was due to the difference in the prolines position, but it seemed that the motifs structures were conserved in both AFGP8 in DMSO.

V.1.C. Comparison between structures of AFGP8 in DMSO and Water

Comparing the diffusion coefficient of AFGP8-BS and AFGP8-TB in DMSO, there was not a significant difference. From the diffusion coefficient, it seems that both AFGP8 had the same "compactness" regarding their three-dimensional structure. The YASARA generated structures of both AFGP8 agreed with the idea of similar compactness between the two structural ensembles. As shown in **Figure 20**, there were differences in the clustering regions of the disaccharides and alignment of the backbone within each AFGP8-BS and AFGP8-TB, but neither structural ensembles of AFGP8 showed some compact folding such as grouping the hydrophilic disaccharides or hydrophobic amino acids (alanine and proline) to a cluster or region. Also, with the alignment of the backbone of both AFGP8, as shown in **Figure 21**, the difference between the backbone was that AFGP8-BS deviated less as compared to AFGP8-TB, but the compactness of the overall structural ensembles did not seem too different.

Using the AFGP8-TB structures determined in water as a comparison to the AFGP8-TB structures determined in DMSO, the AFGP8-TB structure in water was much more dynamic, regarding the structural alignment, than AFGP8-TB in DMSO. The disaccharides were more stable, as clustering to a more defined region in DMSO than water. The overall more ordered AFGP8-TB could also be a contribution of the DMSO solvent. DSMO and water have different hydrogen bonding potential that could influence different intra hydrogen bonding between the protein with itself and inter hydrogen bonding between the protein and solvent.

DMSO is known to be only hydrogen bond acceptor while water can be both a hydrogen bond acceptor and donor. The AFGP8-TB in DMSO could have used more of its potential intra protein hydrogen bond to form inter hydrogen bonding with the DMSO solvent and unfolded the protein a little more compared to the structure in water. This can be seen in **Figure 24** as there was a clear difference between the overall structural ensemble of AFGP8-TB in DMSO and water.

To further confirm unfolding (less compactness), comparing the AFGP8-BS diffusion coefficient 8.16×10^{-11} m^2/s, determined in water (0 °C) by Krishnan et al.[54], to the AFGP8-BS diffusion coefficient 6.46×10^{-11} m^2/s determined in DMSO, there was a difference even though the viscosities of the two solutions were very similar. The viscosity of DMSO at 25 °C is 1.99 centipoises[57], and the viscosity of water in the presence of AFGP at 0 °C is ~2 centipoises[58]. In both instances, the key difference was the hydrogen bonding potential of the solvents. This helps support the idea that the structure of AFGP8 might be more unfolded or less compacted in DMSO than in water. It is not conclusive from the results that the structure in DMSO is less compacted compared to the water since the hydration of water around AFGP8 can be completely different when compared to DMSO. The hydration of the water molecule around the protein does influence the overall radius while in solution, so the difference in the diffusion coefficient might just be due to the difference in hydration water and "hydration" DMSO.

While the compactness of AFGP8 in water and DMSO may be different, the local structure of the motifs seemed conserved in both solvents. As shown in **Figure 26**, the overlaid of the AFGP8-TB motifs aligned the backbone and sugar into clustering region except for the case in **Figure 26C** where the disaccharide clustered into different regions between the water and DMSO motif.

V.1.D. Comparison of Synthetic and Natural Antifreeze Glycoproteins

Unlike the natural antifreeze glycoproteins (AFGP), which seems to be high disordered in water, the synthetic AFGP (sAFGP) forms a well-defined structure having structural features similar to polyproline II helix as reported by Tachibana et al.[53] The structural features allowed sAFGP to take full advantage of the amphipathic characteristic inherent to both natural and synthetic AFGP as to form a hydrophilic side and hydrophobic side as shown in **Figure 27**. When the motifs *AT*AA* were overlaid between sAFGP to AFGP8-BS (**Figure 28**), AFGP8-TB in DMSO (**Figure 29**), and AFGP8-TB in water (**Figure 30**), the overlaid of the motifs varied from one another. AFGP8-BS seemed to overlap better regarding the backbone and clustering of the disaccharide with sAFGP. AFGP8-TB, in DMSO, does not seem to be in good agreement for neither the backbone and clustering of the disaccharide. Unlike AFGP8-TB, in DMSO, the backbone of AFGP8-TB, in water, seemed to overlap with the backbone of sAFGP. However, the disaccharides in water fluctuate and do not cluster into a defined region. There seemed to be some common conserved motifs between sAFGP and natural AFGP, but still inconclusive overall.

V.1.A. Role of the Conformations of the Disaccharides of AFGP

It is well-known that a pyranose can alter between the chair and boat conformations. Similar to the chair and boat conformation of a pyranose, the hydrogen bonding network of hexagonal ice (normal ice) mimics that of a chair and boat conformation (**Figure 32**).[19] The prism faces of hexagonal ice have a hydrogen bonding network that mimics the boat conformation, and the hydrogen bonding network of the basal plane of hexagonal ice mimics that of a chair conformation. From experiments, it is known that AFGP binds onto the prism

faces of ice crystals and alters the shapes of growing ice crystals.[20] From this knowledge, it is speculated that one of the possible mechanism for inhibiting ice crystal growth is that the disaccharides of AFGP alter between the chair and boat conformations. The disaccharides of AFGP could act as a sink for the latent heat release from the bulk water molecules trying to add to the ice crystals surface. By absorbing the latent heat released by the bulk water molecules, the disaccharides can alter from the chair conformation to the boat conformation. The boat conformation might allow for AFGP to bind onto the prism face via hydrogen bonds mimicking water molecules. Due to the chair conformation being more thermodynamically favorable, the disaccharides convert from the boat conformation back to the chair conformation releasing the latent heat that the disaccharides absorbed back to the bulk water molecules. This process prevents the addition of water molecules to the ice crystal surface within the thermal hysteresis gap. This could explain why natural AFGP and synthetic AFGP analog still inhibit ice crystal growth, causing a thermal hysteresis, even though the three-dimensional structure of the two are drastically different.

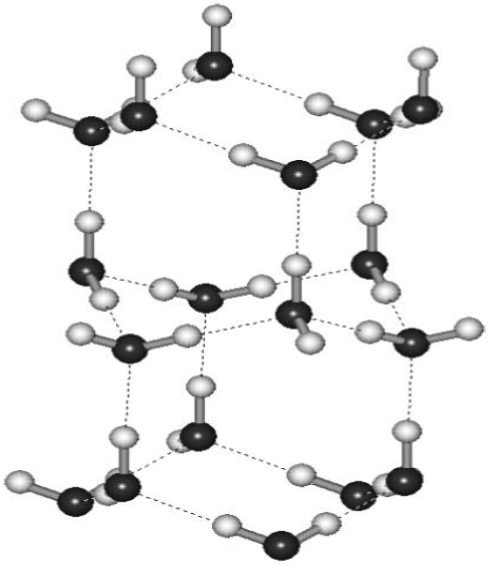


Figure 32. The hydrogen bonding network of hexagonal ice (normal ice). Figure was borrowed[19].

V.2. Conclusion

Determination of the three-dimensional structure of antifreeze glycoproteins (AFGP) in their native, water solvent proved to be very challenging as evident from previous structural studies. Introducing AFGP into a non-native solvent induces structural features that were determined using nuclear magnetic resonance (NMR) spectroscopy. Comparison of the three-dimensional structure determined by Lane et al.[52] in water and the DMSO structure showed that in the presence of water AFGP seemed highly dynamic overall as compared to the more static structure determined in DMSO. This structural difference suggested that native structure of AFGP is not well-defined and fits into the category of an intrinsically disordered protein.

The compactness of the structural ensemble for AFGP in both water and DMSO seemed different regarding their different diffusion coefficient and the three-dimensional structure determined using NMR spectroscopy. Studying AFGP as an ensemble of structures might not provide much insight since nature AFGP seemed to function without a well-defined structure while synthetic AFGP functions with a well-defined structure. Using the short, conserved motifs present in both natural and synthetic AFGP as the focus of structural studies might be of significance due to the similarity between the structures of natural AFGP8 in DMSO and water and synthetic AFGP in water. Maybe, the overall three-dimensional structure might not be crucial, but the local structure that each conserved motif, *AT*PA* or *AT*AA,* takes might be more important. AFGP might just function as an intrinsically disordered protein taking on local secondary structures and being able to bind to multiple different types of ice surface while simultaneously interacting with the bulk water molecules.

If AFGP functions as an intrinsically disordered protein, AFGP might be able to satisfy both proposed ice crystal inhibition model, ice crystal adsorption inhibition and long-range perturbation of water molecule. It had been shown with Terahertz absorption spectroscopy and molecular dynamic simulation that the hydration layer of AFGP increase with decreasing temperature.[23-24] Also, AFGP increases the viscosity of water upon cooling.[58] The combination of the mentioned properties might provide a good explanation of the often-asked questions, "Why do AFP and AFGP wait until ice crystal formation to start inhibiting the ice crystal growth? Wouldn't it be more effective to prevent ice crystal from forming at all?" Maybe, AFGP is inhibiting ice crystal formation by increasing the viscosity of the solution and increasing its hydration layer to increase the range of protein-bulk water interaction. The increase in viscosity decreases the probability of the bulk water molecule to encounter a heterogeneous growing ice crystal thus inhibiting the ice crystal from reaching a critical size or radius which will allow nucleation to occur spontaneously. This inhibition of ice crystal formation with increasing viscosity is assisted by the longer than normal protein to bulk water interaction through the increase in hydration layer. When the temperature gets low enough to promote spontaneous ice nucleation, AFGP will bind onto the ice crystals surface thus matching the adsorption inhibition model. It would make sense that AFGP needs to be able to adopt multiple functional conformation since the ice crystal surface would be restructured by AFGP resulting in multiple different binding sites on the surface of the ice crystal. Within the thermal hysteresis gap there exist both liquid water and small ice crystals. For AFGP to function in both the solution and ice-water interface, AFGP has to have multiple different function conformation. The argument of AFGP functioning as an intrinsically disordered protein satisfy the multiple different, functional conformations since intrinsically

disordered protein can exist as multiple different conformation in solution while all conformations are still functional.

REFERENCES

REFERENCES

1. Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z., Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling* **2001,** *19* (1), 26-59.

2. Radivojac, P.; Iakoucheva, L. M.; Oldfield, C. J.; Obradovic, Z.; Uversky, V. N.; Dunker, A. K., Intrinsic Disorder and Functional Proteomics. *Biophysical Journal 92* (5), 1439-1456.

3. Turoverov, K. K.; Kuznetsova, I. M.; Uversky, V. N., The protein kingdom extended: Ordered and intrinsically disordered proteins, their folding, supramolecular complex formation, and aggregation. *Progress in Biophysics and Molecular Biology* **2010,** *102* (2–3), 73-84.

4. Jacobsen, N. E., *NMR Spectroscopy Explained: Simplified Theory, Application and Examples for Organic Chemistry and Structural Biology*. John Wiley & Sons, Inc.: Hoboken, New Jersey, 2007.

5. Keeler, J., *Understanding NMR Spectroscopy*. Second Edition ed.; John Wiley & Sons, Ltd: 2010.

6. Wüthrich, K., *NMR of Proteins and Nucleic Acids*. John Wiley & Sons, Inc.: 1986.

7. Davies, P. L.; Hew, C. L., Biochemistry of fish antifreeze proteins. *The FASEB Journal* **1990,** *4* (8), 2460-8.

8. Harding, M. M.; Anderberg, P. I.; Haymet, A. D. J., 'Antifreeze' glycoproteins from polar fish. *European Journal of Biochemistry* **2003,** *270* (7), 1381-1392.

9. Chen, L.; DeVries, A. L.; Cheng, C.-H. C., Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proceedings of the National Academy of Sciences* **1997,** *94* (8), 3811-3816.

10. Chen, L.; DeVries, A. L.; Cheng, C.-H. C., Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. *Proceedings of the National Academy of Sciences* **1997,** *94* (8), 3817-3822.

11. Yeh, Y.; Feeney, R. E., Antifreeze Proteins: Structures and Mechanisms of Function. *Chemical Reviews* **1996,** *96* (2), 601-618.

12. Bang, J. K.; Lee, J. H.; Murugan, R. N.; Lee, S. G.; Do, H.; Koh, H. Y.; Shim, H.-E.; Kim, H.-C.; Kim, H. J., Antifreeze Peptides and Glycopeptides, and Their Derivatives: Potential Uses in Biotechnology. *Marine Drugs* **2013,** *11* (6), 2013-2041.

13. Marshall, C. B.; Fletcher, G. L.; Davies, P. L., Hyperactive antifreeze protein in a fish. *Nature* **2004,** *429* (6988), 153-153.

14. Marshall, C. B.; Daley, M. E.; Sykes, B. D.; Davies, P. L., Enhancing the Activity of a β-Helical Antifreeze Protein by the Engineered Addition of Coils†. *Biochemistry* **2004,** *43* (37), 11637-11646.

15. KNIGHT, C. A.; DEVRIES, A. L., Melting Inhibition and Superheating of Ice by an Antifreeze Glycopeptide. *Science* **1989,** *245* (4917), 505-507.

16. Celik, Y.; Graham, L. A.; Mok, Y.-F.; Bar, M.; Davies, P. L.; Braslavsky, I., Superheating of ice crystals in antifreeze protein solutions. *Proceedings of the National Academy of Sciences* **2010,** *107* (12), 5423-5428.

17. Cziko, P. A.; DeVries, A. L.; Evans, C. W.; Cheng, C.-H. C., Antifreeze protein-induced superheating of ice inside Antarctic notothenioid fishes inhibits melting during summer warming. *Proceedings of the National Academy of Sciences of the United States of America* **2014,** *111* (40), 14583-14588.

18. Hew, C. L.; Yang, D. S. C., Protein interaction with ice. *European Journal of Biochemistry* **1992,** *203* (1-2), 33-42.

19. George, M., Liquid water and ices: understanding the structure and physical properties. *Journal of Physics: Condensed Matter* **2009,** *21* (28), 283101.

20. Raymond, J. A.; DeVries, A. L., Adsorption inhibition as a mechanism of freezing resistance in polar fishes. *Proceedings of the National Academy of Sciences of the United States of America* **1977,** *74* (6), 2589-2593.

21. Kristiansen, E.; Zachariassen, K. E., The mechanism by which fish antifreeze proteins cause thermal hysteresis. *Cryobiology* **2005,** *51* (3), 262-280.

22. Bouvet, V.; Ben, R., Antifreeze glycoproteins. *Cell Biochemistry and Biophysics* **2003,** *39* (2), 133-144.

23. Ebbinghaus, S.; Meister, K.; Born, B.; DeVries, A. L.; Gruebele, M.; Havenith, M., Antifreeze Glycoprotein Activity Correlates with Long-Range Protein−Water Dynamics. *Journal of the American Chemical Society* **2010,** *132* (35), 12210-12211.

24. Mallajosyula, S. S.; Vanommeslaeghe, K.; MacKerell, A. D., Perturbation of Long-Range Water Dynamics as the Mechanism for the Antifreeze Activity of Antifreeze Glycoprotein. *The Journal of Physical Chemistry B* **2014,** *118* (40), 11696-11706.

25. Drori, R.; Celik, Y.; Davies, P. L.; Braslavsky, I., *Ice-binding proteins that accumulate on different ice crystal planes produce distinct thermal hysteresis dynamics*. 2014; Vol. 11.

26. Meister, K.; Ebbinghaus, S.; Xu, Y.; Duman, J. G.; DeVries, A.; Gruebele, M.; Leitner, D. M.; Havenith, M., Long-range protein–water dynamics in hyperactive insect antifreeze proteins. *Proceedings of the National Academy of Sciences of the United States of America* **2013,** *110* (5), 1617-1622.

27. Brown, R. A.; Yeh, Y.; Burcham, T. S.; Feeney, R. E., Direct evidence for antifreeze glycoprotein adsorption onto an ice surface. *Biopolymers* **1985,** *24* (7), 1265-1270.

28. Harding, M. M.; Ward, L. G.; Haymet, A. D. J., Type I 'antifreeze' proteins. *European Journal of Biochemistry* **1999,** *264* (3), 653-665.

29. Cheng, Y.; Yang, Z.; Tan, H.; Liu, R.; Chen, G.; Jia, Z., Analysis of ice-binding sites in fish type II antifreeze protein by quantum mechanics. *Biophysical Journal* **2002,** *83* (4), 2202-2210.

30. Sönnichsen, F. D.; DeLuca, C. I.; Davies, P. L.; Sykes, B. D., Refined solution structure of type III antifreeze protein: hydrophobic groups may be involved in the energetics of the protein–ice interaction. *Structure* **1996,** *4* (11), 1325-1337.

31. Krishnan, V.; Yeh, Y., STRUCTURE AND FUNCTIONAL DYNAMICS OF ANTIFREEZE GLYCOPROTEINS. In *Biochemistry and Function of Antifreeze Proteins*, Graether, S. P., Ed. Nova Science: 2010.

32. Burcham, T. S.; Osuga, D. T.; Rao, B. N.; Bush, C. A.; Feeney, R. E., Purification and primary sequences of the major arginine-containing antifreeze glycopeptides from the fish Eleginus gracilis. *Journal of Biological Chemistry* **1986,** *261* (14), 6384-6389.

33. Wen, D.; Laursen, R. A., A model for binding of an antifreeze polypeptide to ice. *Biophysical Journal* **1992,** *63* (6), 1659-1662.

34. Geoghegan, K. F.; Osuga, D.; Ahmed, A. I.; Yeh, Y.; FeeneySII, R. E., Antifreeze Glycoproteins from Polar Fish. *J. Biol. Chem.* **1980,** *255*, 663-667.

35. Heisel, K. A.; Krishnan, V. V., NMR based solvent exchange experiments to understand the conformational preference of intrinsically disordered proteins using FG-nucleoporin peptide as a model. *Peptide Science* **2014,** *102* (1), 69-77.

36. Means, G. E.; Feeney, R. E., Reductive alkylation of amino groups in proteins. *Biochemistry* **1968,** *7* (6), 2192-2201.

37. Means, G. E.; Feeney, R. E., Reductive Alkylation of Proteins. *Analytical Biochemistry* **1995,** *224* (1), 1-16.

38. Piantini, U.; Sorensen, O. W.; Ernst, R. R., Multiple quantum filters for elucidating NMR coupling networks. *Journal of the American Chemical Society* **1982,** *104* (24), 6800-6801.

39. Davis, D. G.; Bax, A., Assignment of complex proton NMR spectra via two-dimensional homonuclear Hartmann-Hahn spectroscopy. *Journal of the American Chemical Society* **1985,** *107* (9), 2820-2821.

40. Macura, S.; Ernst, R. R., Elucidation of cross relaxation in liquids by two-dimensional N.M.R. spectroscopy. *Molecular Physics* **1980,** *41* (1), 95-117.

41. Bodenhausen, G.; Ruben, D. J., Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters* **1980,** *69* (1), 185-189.

42. Rucker, S. P.; Shaka, A. J., Broadband homonuclear cross polarization in 2D N.M.R. using DIPSI-2. *Molecular Physics* **1989,** *68* (2), 509-517.

43. Pelta, M. D.; Morris, G. A.; Stchedroff, M. J.; Hammond, S. J., A one-shot sequence for high-resolution diffusion-ordered spectroscopy. *Magnetic Resonance in Chemistry* **2002,** *40* (13), S147-S152.

44. Harmon, J.; Coffman, C.; Villarrial, S.; Chabolla, S.; Heisel, K. A.; Krishnan, V. V., Determination of Molecular Self-Diffusion Coefficients Using Pulsed-Field-Gradient NMR: An Experiment for Undergraduate Physical Chemistry Laboratory. *Journal of Chemical Education* **2012,** *89* (6), 780-783.

45. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR 6* (3), 277-293.

46. T. D. Goddard, D. G. K., SPARKY 3. University of California, San Francisco.

47. Güntert, P., Automated NMR Structure Calculation With CYANA. In *Protein NMR Techniques*, Downing, A. K., Ed. Humana Press: Totowa, NJ, 2004; pp 353-378.

48. Kapaev, R. R.; Toukach, P. V., Simulation of 2D NMR Spectra of Carbohydrates Using GODESS Software. *Journal of Chemical Information and Modeling* **2016,** *56* (6), 1100-1104.

49. Wishart, D. S.; Sykes, B. D.; Richards, F. M., The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **1992,** *31* (6), 1647-1651.

50. Wishart, D. S.; Sykes, B. D., The 13C Chemical-Shift Index: A simple method for the identification of protein secondary structure using 13C chemical-shift data. *Journal of Biomolecular NMR* **1994,** *4* (2), 171-180.

51. Tremblay, M.-L.; Banks, A. W.; Rainey, J. K., The predictive accuracy of secondary chemical shifts is more affected by protein secondary structure than solvent environment. *Journal of Biomolecular NMR* **2010,** *46* (4), 257-270.

52. Lane, A. N.; Hays, L. M.; Crowe, L. M.; Crowe, J. H.; Feeney, R. E., Conformational and dynamic properties of a 14 residue antifreeze glycopeptide from antarctic cod. *Protein Science* **1998,** *7* (7), 1555-1563.

53. Tachibana, Y.; Fletcher, G. L.; Fujitani, N.; Tsuda, S.; Monde, K.; Nishimura, S.-I., Antifreeze Glycoproteins: Elucidation of the Structural Motifs That Are Essential for Antifreeze Activity. *Angewandte Chemie* **2004,** *116* (7), 874-880.

54. Krishnan, V. V.; Fink, W. H.; Feeney, R. E.; Yeh, Y., Translational dynamics of antifreeze glycoprotein in supercooled water. *Biophysical Chemistry* **2004,** *110* (3), 223-230.

55. Inglis, S. R.; McGann, M. J.; Price, W. S.; Harding, M. M., Diffusion NMR studies on fish antifreeze proteins and synthetic analogues. *FEBS Letters* **2006,** *580* (16), 3911-3915.

56. Laurence, C.; Brameld, K. A.; Graton, J.; Le Questel, J.-Y.; Renault, E., The pKBHX Database: Toward a Better Understanding of Hydrogen-Bond Basicity for Medicinal Chemists. *Journal of Medicinal Chemistry* **2009,** *52* (14), 4073-4086.

57. LeBel, R. G.; Goring, D. A. I., Density, Viscosity, Refractive Index, and Hygroscopicity of Mixtures of Water and Dimethyl Sulfoxide. *Journal of Chemical & Engineering Data* **1962,** *7* (1), 100-101.

58. Eto, T. K.; Rubinsky, B., Antifreeze Glycoproteins Increase Solution Viscosity. *Biochemical and Biophysical Research Communications* **1993,** *197* (2), 927-931.

APPENDICES

APPENDIX A: NMRPIPE SCRIPTS

NOESY NMRPipe Script

```
#!/bin/csh

var2pipe -in ./fid \
 -noaswap  \
  -xN              4096  -yN              512  \
  -xT              2048  -yT              256  \
  -xMODE       Complex  -yMODE       Complex  \
  -xSW        8000.000  -ySW        8000.000  \
  -xOBS        599.754  -yOBS        599.754  \
  -xCAR          4.773  -yCAR          4.773  \
  -xLAB            H1x  -yLAB            H1y  \
  -ndim              2  -aq2D         States  \
  -out ./test.fid -verb -ov

sleep 5

#!/bin/csh

nmrPipe -in test.fid \
| nmrPipe  -fn SP -off 0.5 -end 1.00 -pow 1 -c 1.0    \
| nmrPipe  -fn ZF -auto                          \
| nmrPipe  -fn FT -auto                          \
| nmrPipe  -fn PS -p0 -43.6 -p1 0.00 -di -verb       \
| nmrPipe -fn POLY -auto                           \
| nmrPipe  -fn TP                                \
| nmrPipe -fn LP -fb                              \
| nmrPipe  -fn SP -off 0.5 -end 1.00 -pow 1 -c 1.0    \
| nmrPipe  -fn ZF -auto                           \
| nmrPipe  -fn FT -auto                           \
| nmrPipe  -fn PS -p0 0.00 -p1 0.00 -di -verb       \
| nmrPipe -fn POLY -auto                           \
| nmrPipe -fn TP                                 \
   -ov -out afgp8_bs_600_noe100_test1.ft2
pipe2ucsf afgp8_bs_600_noe100_test1.ft2 afgp8_bs_600_noe100_test1.ucsf
```

TOCSY NMRPipe Script

```
#!/bin/csh

var2pipe -in
/home/cheeher/afgp_600_data/AFGP8_dmso_600/BS/BS_tocsy_840ms.fid/fid \
 -noaswap  \
  -xN            4096  -yN             512  \
  -xT            2048  -yT             256  \
  -xMODE       Complex  -yMODE       Complex  \
  -xSW        8000.000  -ySW        8000.000  \
  -xOBS        599.754  -yOBS        599.754  \
  -xCAR          4.773  -yCAR          4.773  \
  -xLAB            H1x  -yLAB            H1y  \
  -ndim              2  -aq2D         States  \
  -out /home/cheeher/afgp_600_data/AFGP8_dmso_600/BS/BS_tocsy_840ms.fid/test.fid
-verb -ov

sleep 5

#!/bin/csh



nmrPipe -in test.fid \
| nmrPipe  -fn SP -off 0.5 -end 1.00 -pow 1 -c 1.0    \
| nmrPipe  -fn ZF -auto                          \
| nmrPipe  -fn FT -auto                          \
| nmrPipe  -fn PS -p0 -39.4 -p1 0.00 -di -verb       \
| nmrPipe -fn POLY -auto                           \
| nmrPipe  -fn TP                               \
| nmrPipe -fn LP -fb                             \
| nmrPipe  -fn SP -off 0.5 -end 1.00 -pow 1 -c 1.0    \
| nmrPipe  -fn ZF -auto                          \
| nmrPipe  -fn FT -auto                          \
| nmrPipe  -fn PS -p0 0.00 -p1 0.00 -di -verb       \
| nmrPipe -fn POLY -auto                           \
| nmrPipe -fn TP                               \
   -ov -out afgp8_bs_600_toc840_test1.ft2

pipe2ucsf afgp8_bs_600_toc840_test1.ft2 afgp8_bs_600_toc840_test1.ucsf
```

DQFC NMRPipe Script

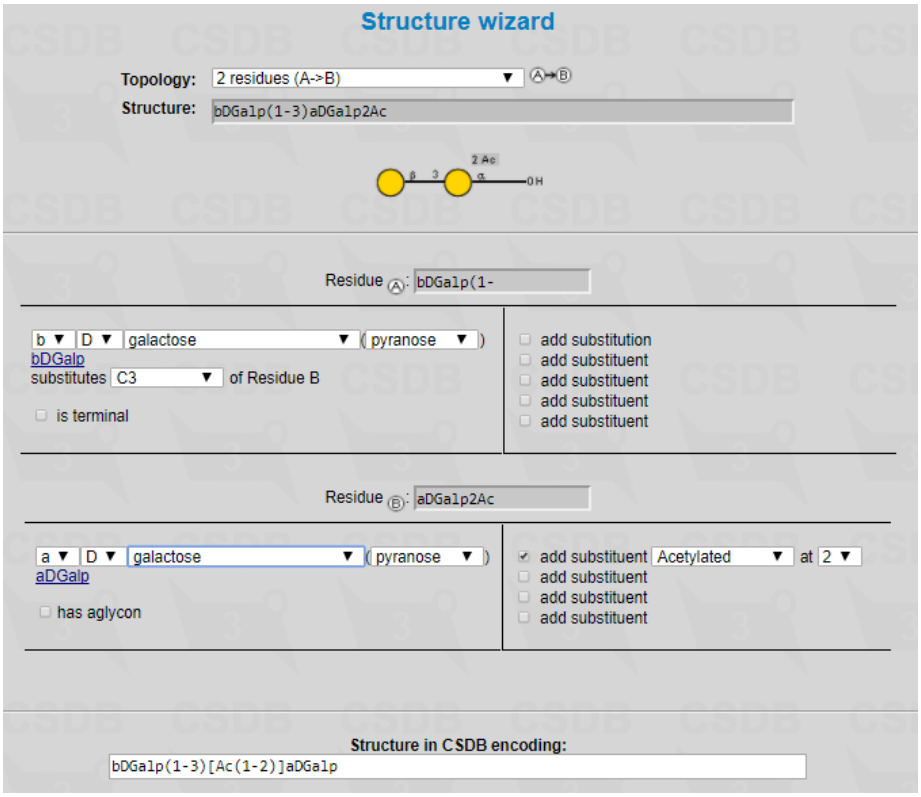
```
#!/bin/csh

var2pipe -in ./fid \
 -noaswap  \
  -xN              4096  -yN              1024  \
  -xT              2048  -yT              512  \
  -xMODE        Complex  -yMODE        Complex  \
  -xSW         8000.000  -ySW         8000.000  \
  -xOBS         599.753  -yOBS         599.753  \
  -xCAR           4.773  -yCAR          4.773  \
  -xLAB             1H  -yLAB           H1y  \
  -ndim              2  -aq2D         Complex  \
  -out ./test.fid -verb -ov

sleep 5

#!/bin/csh

nmrPipe -in test.fid \
| nmrPipe  -fn SP -off 0.5 -end 0.98 -pow 2 -c 0.5    \
| nmrPipe  -fn ZF -auto                          \
| nmrPipe  -fn FT -auto                          \
| nmrPipe  -fn PS -p0 0.00 -p1 0.00 -di -verb        \
| nmrPipe -fn POLY -auto                           \
| nmrPipe  -fn TP                                \
| nmrPipe -fn LP -fb                              \
| nmrPipe  -fn SP -off 0.5 -end 0.98 -pow 2 -c 0.5    \
| nmrPipe  -fn ZF -auto                          \
| nmrPipe  -fn FT -auto                          \
| nmrPipe  -fn PS -p0 0.00 -p1 0.00 -di -verb        \
| nmrPipe -fn POLY -auto                           \
| nmrPipe -fn TP                                \
   -ov -out afgp8_bs_600_dqfc_test.ft2

pipe2ucsf afgp8_bs_600_dqfc_test.ft2 afgp8_bs_600_dqfc_test.ucsf
```

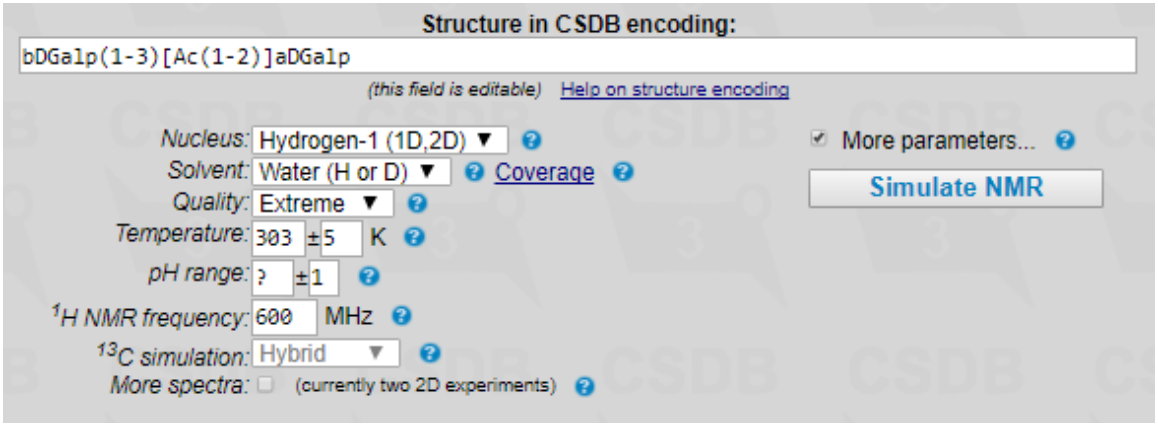
HMQC NMRPipe Script

```
#!/bin/csh

var2pipe -in ./fid \
 -noaswap  \
  -xN              4096  -yN              256  \
  -xT              2048  -yT              128  \
  -xMODE        Complex  -yMODE        Complex  \
  -xSW         8000.000  -ySW        22624.021  \
  -xOBS         599.754  -yOBS         150.811  \
  -xCAR           4.773  -yCAR         26.391  \
  -xLAB              1H  -yLAB           C13  \
  -ndim               2  -aq2D         Complex  \
  -out ./test.fid -verb -ov

sleep 5

#!/bin/csh

nmrPipe -in test.fid \
| nmrPipe  -fn SP -off 0.5 -end 1.00 -pow 1 -c 1.0    \
| nmrPipe  -fn ZF -auto                          \
| nmrPipe  -fn FT -auto                          \
| nmrPipe  -fn PS -p0 141.6 -p1 0.00 -di -verb        \
| nmrPipe  -fn TP                                \
| nmrPipe  -fn SP -off 0.5 -end 1.00 -pow 1 -c 1.0    \
| nmrPipe  -fn ZF -auto                          \
| nmrPipe  -fn FT -auto                          \
| nmrPipe  -fn PS -p0 0.00 -p1 0.00 -di -verb        \
| nmrPipe -fn TP                                 \
   -ov -out afgp8_bs_600_HMQC_test1.ft2

pipe2ucsf afgp8_bs_600_HMQC_test1.ft2 afgp8_bs_600_HMQC_test1.ucsf
```

APPENDIX B: GLYCAN OPTIMIZED DUAL EMPIRICAL SPECTRUM SIMULATION (GODESS) PARAMETERS

Online Parameter Used in GODESS





The web service *Glycan Optimized Dual Empirical Spectrum Simulation (GODESS)*[48]. (http://csdb.glycoscience.ru/database/index.html)

Fresno State

Non-Exclusive Distribution License (Keep for your records)
(to archive your thesis/dissertation electronically via the library's eCollections database)

By submitting this license, you (the author or copyright holder) grant to Fresno State Digital Scholar the non-exclusive right to reproduce, translate (as defined in the next paragraph), and/or distribute your submission (including the abstract) worldwide in print and electronic format and in any medium, including but not limited to audio or video.

You agree that Fresno State may, without changing the content, translate the submission to any medium or format for the purpose of preservation.

You also agree that the submission is your original work, and that you have the right to grant the rights contained in this license. You also represent that your submission does not, to the best of your knowledge, infringe upon anyone's copyright.

If the submission reproduces material for which you do not hold copyright and that would not be considered fair use outside the copyright law, you represent that you have obtained the unrestricted permission of the copyright owner to grant Fresno State the rights required by this license, and that such third-party material is clearly identified and acknowledged within the text or content of the submission.

If the submission is based upon work that has been sponsored or supported by an agency or organization other than Fresno State, you represent that you have fulfilled any right of review or other obligations required by such contract or agreement.

Fresno State will clearly identify your name as the author or owner of the submission and will not make any alteration, other than as allowed by this license, to your submission. **By typing your name and date in the fields below, you indicate your agreement to the terms of this distribution license.**
Publish/embargo options (type X in one of the boxes).

[X] **Make my thesis or dissertation available to eCollections immediately upon submission.**

[ ] Embargo my thesis or dissertation for a period of 2 years from date of graduation. After 2 years, I understand that my work will automatically become part of the university's public institutional repository unless I choose to renew this embargo here: fsdr-discuss@csufresno.edu

[ ] Embargo my thesis or dissertation for a period of 5 years from date of graduation. After 5 years, I understand that my work will automatically become part of the university's public institutional repository unless I choose to renew this embargo here: fsdr-discuss@csufresno.edu

Cheenou Her

Type full name as it appears on submission

January 10, 2018

Date