

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Characterizing the Effects of N-linked Glycosylations on the IL-33/ST2 Axis through Experiment and Simulation

Permalink

<https://escholarship.org/uc/item/1rd2b09x>

Author

Chan, Joshua C.

Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Characterizing the Effects of N-linked Glycosylations on the IL-33/ST2 Axis through
Experiment and Simulation

A Thesis submitted in partial satisfaction of the requirements of for the degree of Master of
Science

in

Chemistry

by

Joshua C. Chan

Committee in charge:

Professor Patricia Jennings, Chair
Professor Rommie Amaro
Professor Timothy Bigby

2015

Copyright

Joshua C. Chan, 2015

All rights reserved

The Thesis of Joshua C. Chan is approved and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2015

DEDICATION

I would like to dedicate this thesis to my grandmother. Thank you for raising me, for always being patient with me, and always believing in me no matter how much I stumbled. I love you, and you will be missed dearly.

EPIGRAPH

“The devil is in the detail”

-unknown

TABLE OF CONTENTS

SIGNATURE PAGE	iii
DEDICATION	iv
EPIGRAPH	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	viii
LIST OF TABLES	xi
ACKNOWLEDGEMENTS	xii
VITA	xiii
ABSTRACT OF THE THESIS	xiv
Chapter 1 : Expression and Purification of Pro-IL-33	1
1.1 Introduction	2
1.2 Materials and Methods	11
1.3 Results	13
1.4 Discussion	21
1.5 Future Directions	23
1.6 References	26
Chapter 2 : Characterization of the N-linked Glycosylations Present on the ST2 Receptor ...	29
2.1 Introduction	30
2.2 Materials and Methods	39
2.3 Results	42
2.4 Discussion	65
2.5 Future Work	70
2.6 References	72
Chapter 3 : Characterization of the N-linked Glycosylations Present on the ST2 Receptor ...	76
3.1 Introduction	77
3.2 Materials and Methods	84
3.3 Results	85
3.4 Discussion	122

TABLE OF CONTENTS

3.5 Future Work 124

3.6 References 127

LIST OF FIGURES

Figure 1-1: Structure of the activating heterotrimeric complex of IL-1 β , IL-1R1 , and IL-1RAcP	4
Figure 1-2: Processing of IL-33 and signaling through the ST2 receptor.	6
Figure 1-3: X-ray crystal structure of the cytokine IL-33 bound to its decoy receptor sST2	8
Figure 1-4: Vector Screen Schematic.	14
Figure 1-6: Comparison of traditional IPTG induced culture and autoinduction culture in Coomassie stained gel.	16
Figure 1-6: Western Blot Verification of 6xHis-SUMO-pro-IL-33	16
Figure 1-7: Chromatogram of nickel column purification observed by A ₂₈₀	18
Figure 1-8: Coomassie stained SDS-PAGE of nickel column chromatogram	19
Figure 1-9: Western blot against IL-33 verification of post-cleavage size exclusion cleanup	19
Figure 1-10: Coomassie verification of post-cleavage size exclusion cleanup	20
Figure 1-11: PONDR Analysis of Pro-IL-33	24
Figure 2-1: Structure of the three classes of N-linked glycans	36
Figure 2-2: Cleavage site of PNGase F on N-linked glycans	37
Figure 2-3: Gel mobility shift anti ST2 western blot of sST2 upon deglycosylation	43
Figure 2-4: PNGase F release of glycan mechanism	45
Figure 2-5: Peptide coverage map of sST2 by Trypsin	47
Figure 2-6: MS2 Spectra of a representative de-glycosylated peptide digested by PNGase F in ¹⁶ O	51
Figure 2-7: MS2 Spectra of a representative de-glycosylated peptide digested by PNGase F in ¹⁸ O	54
Figure 2-8: Overlaid MS2 Spectra of a representative de-glycosylated peptide digested by PNGase F in ¹⁸ O and ¹⁶ O	56
Figure 2-9: Mapping of newly discovered and verification of previously identified N-linked glycosylation sites	58

Figure 2-10: Endo H and Endo F1 substrate recognition and cleavage site	63
Figure 2-11: Endo F2 and F3 substrate recognition and cleavage site	64
Figure 2-12: Gel mobility shift of endoglycosidase digestion of sST2 probed by anti-ST2 western blotting	67
Figure 3-1: Hamiltonian of an all atom structure based models with diagram of corresponding motion	80
Figure 3-2: Schematic of shadow map determination of contacts versus non-contacts	82
Figure 3-3: Structure of the sST2 receptor after loop filling and attachment of tetra-antennary glycans using the GLYCAM webserver	86
Figure 3-4: Trajectory of de-glycosylated sST2 receptor and glycosylated sST2 receptor.	91
Figure 3-5: RMSF of the de-glycosylated sST2 receptor plotted for each residue along the protein	93
Figure 3-6: RMSF levels of the de-glycosylated simulation of sST2 plotted onto the structure of the sST2 receptor.	95
Figure 3-7: RMSF levels of the glycosylated simulation of sST2 plotted onto the structure of the sST2 receptor glycosylated.	97
Figure 3-8: RMSF differences between de-glycosylated and glycosylated simulations plotted onto the structure of sST2	99
Figure 3-9: Comparison of average RMSF for simulations of the sST2 receptor	101
Figure 3-10: Trajectory of de-glycosylated sST2 receptor and glycosylated sST2 receptor bound to IL-33	103
Figure 3-11: RMSF of the de-glycosylated sST2 receptor bound to IL-33 plotted for each residue along the protein	105
Figure 3-12: RMSF levels of the de- glycosylated simulation of sST2 bound to ligand plotted onto the structure of the sST2 receptor complex	107
Figure 3-13: RMSF levels of the glycosylated simulation of sST2 bound to ligand plotted onto the structure of the sST2 receptor complex	109
Figure 3-14: RMSF differences between de-glycosylated and glycosylated simulations plotted onto the structure of sST2	112
Figure 3-15: Eigenvalue analysis of the first 8 eigenvalues derived from a diagonalized covariance matrix comparing de-glycosylated and glycosylated simulations	115

Figure 3-16: Eigenvalue analysis of the first 8 eigenvalues derived from a diagonalized covariance matrix comparing de-glycosylated and glycosylated simulations of sST2 bound to receptor 118

Figure 3-17: Plot of the first 8 eigenvalues derived from simulations of the sST2 receptor 120

LIST OF TABLES

Table 2-1: Table of common sugars moieties and their corresponding structure and symbol.	33
Table 2-2: Table of theoretical fragmentation masses of the peptide AAVLWQLNGTK in the absence of glycosylation and de-glycosylated by PNGase F in ¹⁶ O	50
Table 2-3: Table of theoretical fragmentation masses of the peptide AAVLWQLNGTK in the absence of glycosylation and de-glycosylated by PNGase F in ¹⁸ O	53
Table 2-4: Table of substrate recognition of endoglycosidases PNGase F, Endo H, Endo F1, Endo F2, and Endo F3	66
Table 3-1: Table of the first 8 eigenvalues for sST2, sST2-glycosylated, sST2-bound, and sST2-bound-glycosylated.	121

ACKNOWLEDGEMENTS

This experience has been one of the most challenging and most rewarding times of my entire life. I would like to thank my advisor, Patricia Jennings, for always having faith in my scientific ability and allowing me the intellectual freedom to discover myself as a scientist. Without this, I would never have been able to develop into the scientist I have become today.

I would like to thank all the current and past members of the Jennings lab especially Kaitlin Fisher, Bob Wong, Dr. Kendra Hailey, Nick Tiee, Dr. Sulyman Barkho, Dr. John Zuris, Dr. Michael Jamros, David Burban, Colin Lipper, Jason Stofleth for all the support and scientific discussions throughout the years. They are all truly outstanding people and I will miss working with each of them.

I would also like to thank all of my friends along the way for keeping me grounded in the world and that there is a whole world out there!

Finally, I would like to thank my family for their love and support. My parents in particular have been extremely patient as I have been learning to walk on my own feet in this world. They instilled in me the determination, drive, and scientific vigor to make all of this possible.

VITA

- 2008 B.S., Bioengineering:Biotechnology, University of California, San Diego
- 2015 M.S, Chemistry , University of California, San Diego

Publications

- 1. A Dual-Mode HDAC Prodrug for Covalent Modification and Subsequent Inhibitor Release;** KB Daniel, ED Sullivan, Y Chen, JC Chan, PA Jennings, CA Fierke, SM Cohen; *Journal of Medicinal Chemistry* (Accepted).
- 2. Heterogeneous Dynamics and Conformational Frustration in the Binding Regions of Interleukin-33.”** Fisher KM, Chan JC, Fuglestad B, Jennings PA (In preparation)

ABSTRACT OF THE THESIS

Characterizing the Effects of N-linked Glycosylations on the IL-33/ST2 Axis through
Experiment and Simulation

by

Joshua C. Chan

Master of Science in Chemistry

University of California, San Diego, 2015

Professor Patricia Jennings, Chair

The role and effects of the IL-33/ST2 pathway in disease is still unknown. The work described here centers upon two important members of the IL-33/ST2 axis: pro-IL-33 and

the sST2 receptor. The first portion of the work focuses on the ligand pro-IL-33, an important signaling and transcriptional regulator for the IL-33/ST2 system whose structure and biophysical characteristics have yet to be elucidated due to lack of a high expression system. This need was addressed through the iterative testing and eventual purification of an *Eschericia coli* based expression system utilizing a His-SUMO construct coupled with autoinduction. The second portion of the work centers upon the role of glycosylations on the sST2 receptor. To address the role of glycosylations in the IL-33/ST2 axis, site specific occupancy was analyzed on the sST2 receptor revealing occupancy of 7 of 8 possible sites present. Subsequent enzymatic based characterization of these sites revealed a mixed population of tri- and tetra- antennary complex glycans present. The effects of these glycans were probed through the implementation of a novel add-on to the computational processing program SMOG2 to enable interpretation and simulation using GROMACS. Analysis of the trajectories revealed that N-linked glycosylations reduce dynamics present in sST2 in a global and local manner; and stabilize the structure of an unstructured loop present from Tyr34 to Val65 of the sST2 receptor when bound to IL-33. As such, this work lays the foundation for further exploration into the function and internal regulation abilities of the IL-33/ST2 axis.

Chapter 1 : Expression and Purification of Pro-IL-33

1.1 Introduction

Cytokines As Regulators of Inflammation and Cancer

Inflammation constitutes a majority of the body's immune response and serves as one of the primary methods by which it reacts to cell damage and foreign pathogens. This system consists of a concerted response between multiple cell types requiring hundreds of molecular messengers to mediate their communication and insure tight regulation and rapid response. Due to its inherent ties to wound repair, cell proliferation is one of the responses regulated by inflammation pathways. As such, dysregulation of inflammation results in more diseases aside from those falling under the umbrella of autoimmune diseases; most notably many forms of cancer (1).

One of the most prevalent molecular messengers utilized by the body in inflammation are cytokines. These messengers are commonly small proteins of less than 20kDa that mediate the communication between humoral and cell-based immune responses through receptor based signaling. Unlike hormones that are also common broad spectrum signaling molecules, the concentration of cytokines can increase up to 1,000-fold resulting in the rapid activation commonly associated with the inflammatory response. Each of these cytokines bind to their own unique receptor to help illicit this rapid response be it production of other cytokines or triggering such effects as cell proliferation and antibody production. As such, elucidation of cytokine regulation and signaling is crucial to the understanding of inflammation, autoimmune diseases, and cancer (2, 3).

Cytokines constitute a broad swath of signaling molecules integral to both the adaptive and innate immune system. These signaling molecules are subsequently divided based on their presumed response: lymphokines which are produced by lymphocytes,

chemokines for their ability to mediate chemotaxis, and interleukins for their effect on leukocytes. Of the three classes, interleukins in particular play a particularly large role in regulating the immune system due to their strong ties to autoimmune diseases and immune deficiency. These proteins are typically synthesized by T cell lymphocytes, monocytes, macrophages, and endothelial cells where they help to promote development and differentiation of a large variety of hematopoietic cells. Because of their ability to cause both cell proliferation and stimulate cell death, understanding the manner in which cytokines work crucial to modulating the desired response. Their pivotal ties to both inflammation and their ability to spur on cell growth and differentiation make interleukins promising drug targets for a variety of illnesses ranging from asthma to certain types of cancers (4–6)

The Interleukin-1 Family of Cytokines

Two of the first discovered and well characterized members of the interleukins are interleukin-1 (IL-1) alpha and beta. These founding members of the IL-1 family of cytokines are pro-inflammatory and serve as the basis by which other cytokines are evaluated to be part of the IL-1 family. Their classification is based on amino acid sequence homology and propensity of maintaining the IL-1 fold consisting of a 12 β -strand fold in the shape of a trefoil. These soluble signaling molecules function primarily through signaling of an extracellular receptor unique for each family member (in the case of IL-1 β is IL-1R1) consist of three Ig domains of the soluble portion that upon binding, trigger recruitment of an additional IL-1 receptor accessory protein (IL-1RAcP) that is conserved across all family members to enable signaling through a conserved toll-interleukin receptor domain (TIR

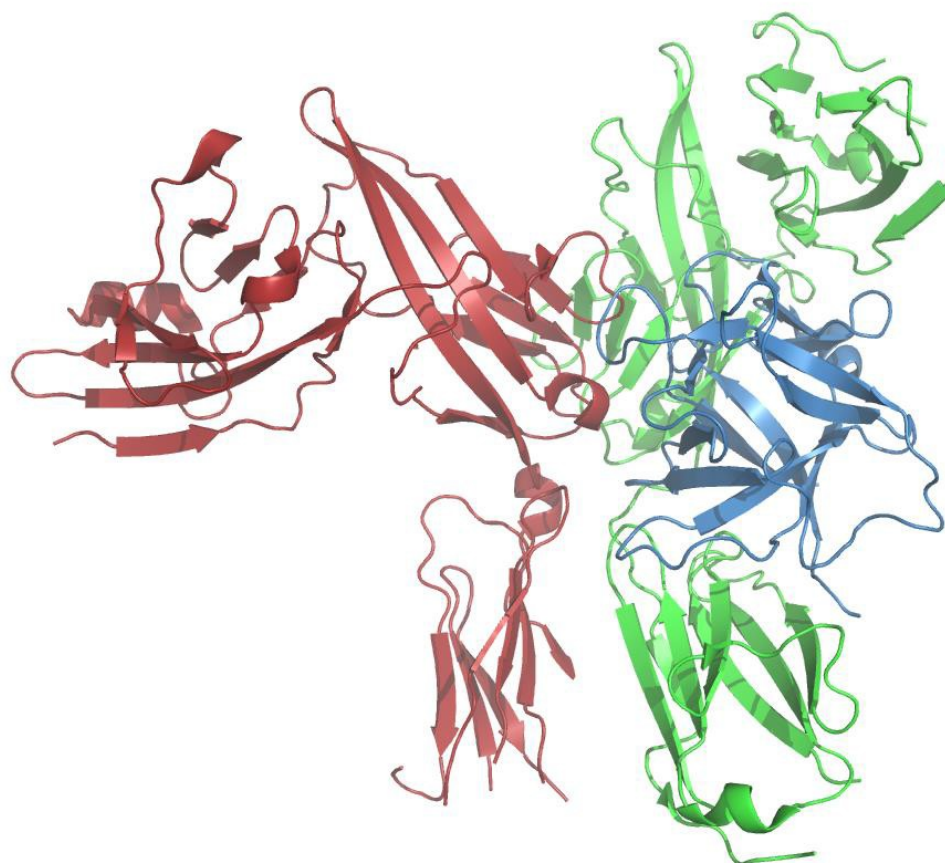


Figure 1-1: Structure of the activating heterotrimeric complex of IL-1 β , IL-1R1 , and IL-1RAcP

Formation of the heterotrimeric complex is initiated by binding between the cytokines IL-1 β and the receptor that consists of 3 Ig domains. Formation of this complex then allows for the recruitment of IL-1RAcP. The formation of this complex can only happen sequentially though the rules governing the order are unknown

domain) (7). These IL-1 receptors are structurally homologous; however, have little sequence homology. Formation of the heterotrimeric complex (Figure 1-1) allows the TIR domain to recruit the TIR domain-containing adaptor, MyD88 and subsequent activation of the IL-1 receptor-associated kinase-4 (IRAK-4) to TLRs that drive the stimulation MAP kinases and the transcription factor NF- κ B (8). Activation of transcription factors force the remodeling of chromatin and transcriptional factor activation which can drive the pro-inflammatory response as well as regulates cell proliferation and survival (9).

Since the discovery of IL-1, much about the function and activity of the IL-1 family of proteins has been elucidated. Although differing in activity from one another, the biological implications of many members of the IL-1 family may hold keys to disease cures. IL-1 for example has strong ties to T helper cells in controlling differentiation and proliferation. Activation of this subset of T cells mediates many autoimmune and chronic inflammatory diseases (10).

The Interleukin-33 System

One of the newest members of the IL-1 family, interleukin-33 (IL-33), was first discovered in 2005 as the ligand for the orphan receptor ST2. Since its discovery, IL-33 has been implicated in a wide range of diseases from asthma and allergy to breast cancer and cardiovascular disease (11–13). In asthma patients, IL-33 is found at elevated levels and its causality has been directly proven through induction of asthmatic attacks via administration of IL-33 (11). Conversely, IL-33 has been shown to have a cardioprotective effect after myocardial infarction through administration of IL-33 (12, 14). These wide range of effects underlie the need for full understanding of the IL-33 system in order to understand how to

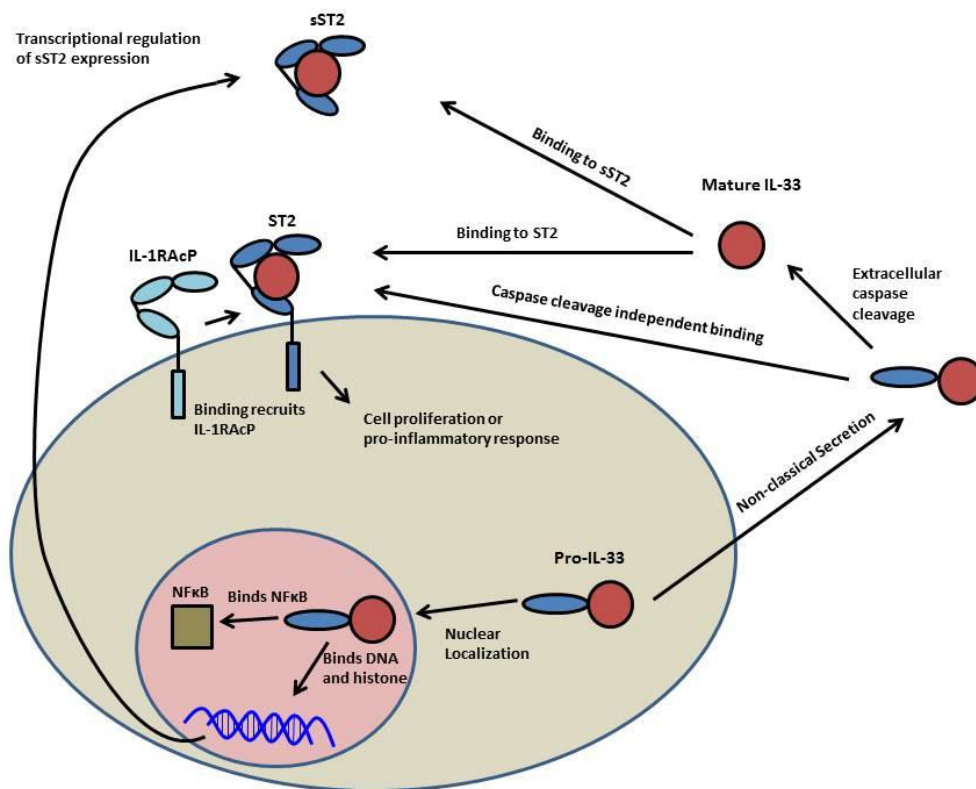


Figure 1-2: Processing of IL-33 and signaling through the ST2 receptor.

IL-33/ST2 system consists of primarily five players: mature IL-33, pro-IL-33, IL-1RAcP, ST2 receptor, and sST2 receptor. The primary signaling complex is created after mature IL-33 binds to ST2 and recruits IL-1RAcP. A natural inhibitor of the system is present in the form of a decoy receptor sST2 which binds extracellular IL-33 to attenuate the effects of IL-33 signaling. Pro-IL-33 is able to bind to ST2 independently of processing into the mature form and also localizes to the nucleus where it directly acts as a transcriptional regulator by binding to DNA, histones, and NFκB

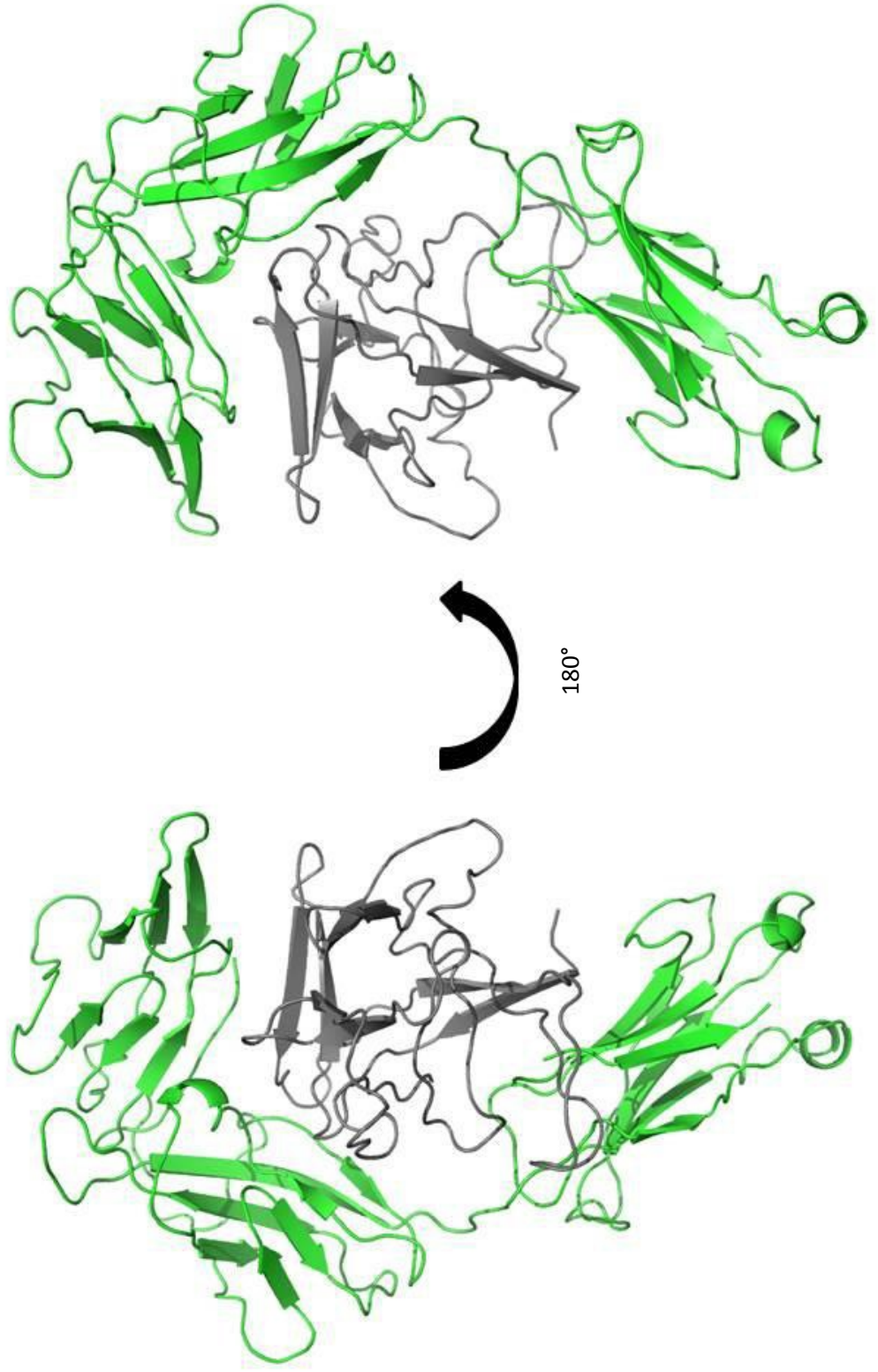
modulate the effects of IL-33 in such a manner to attenuate the harmful responses without knocking out the protective ones.

In the IL-33 signaling system, there are five major components to modulating extracellular signaling: the full length form of IL-33 (pro-IL-33), mature IL-33, the ST2 receptor, a soluble form of the ST2 receptor that serves as a decoy receptor (sST2), and IL-1RAcP (15). Much like the canonical behavior of IL-1 β , IL-33 is first expressed as a 270 amino acid protein where the first 116 residues comprise of a pro-leader sequence that is cleaved by extracellular caspases into the mature form consisting of residues 117 to 270 that fold into the β -trefoil fold (16–18). Subsequent secretion of this mature form through non-classical secretion pathways result in the presence of IL-33 in the extracellular space (19). In the extracellular space, competition for receptor binding is modulated through management of the ratio between ST2 and sST2 forms available for binding. By acting as a decoy receptor, sST2 is able influence an additional level of control over the system (15, 20, 21). IL-33 signals through its cognate receptor ST2 which upon binding of IL-33 is able to recruit IL-1RAcP to enable downstream signaling in a system similar to the traditional IL-1 dependent pathway as shown in Figure 1 (22).

Despite the similarities between the IL-33 and IL-1 β systems, pro-IL-33 represents a unique member of the IL-1 family that potentially holds many interesting aspects both biologically and structurally. Unlike pro-IL-1 β , pro-IL-33 is unique among IL-1 family members in its ability to signal without cleavage of its pro-leader sequence (23). This unique ability suggests that the C-terminal structure of pro-IL-3 retains much of the fold that the mature form does. Despite this, the pro-leader sequence has many functions unique to pro-IL-33 that effects not only the localization but the regulation of the entire IL-33 signaling

Figure 1-3: X-ray crystal structure of the cytokine IL-33 bound to its decoy receptor sST2

Structure of the cytokines IL-33 bound to its decoy receptor sST2 solved by X-ray crystallography and small angle x-ray scattering. The ligand IL-33 is shown in grey while the structure of the receptor is shown in green. IL-33 consists of a beta-trefoil motif consisting of 12 β -strands with unstructured loops connecting them. sST2 consists of 3 Ig domains with Ig2 and Ig3 connected by a long 100 amino acid flexible linker.



system. Present in the leader sequence are four regions of unique functions: nuclear localization sequence, NF- κ B binding region, chromatin binding region, and a helix-turn-helix motif for DNA binding (24–28). These unique abilities of the pro-leader make pro-IL-33 a transcriptional regulator not only through the canonical IL-1 signaling pathway but through exerting direct regulation over the transcription of members of the IL-33 system, most notably the soluble form of ST2. As such, the unique abilities of pro-IL-33 make elucidation of the structure and function pivotal to understanding the ways how the IL-33 system is regulated.

Due to the lack of knowledge concerning the properties of the pro-leader sequence, there is a growing need to elucidate and characterize biophysically the structure and function of pro-IL-33. Previous work on high level expression and purification have yielded some insight into the inherent challenges present in expression and purification of pro-IL-33, but none have yet to achieve a robust method to gain high concentrations of recombinant protein. In the past, mammalian expression systems combined with transient transfection have been able to express low quantities of pro-IL-33, but the resulting cultures have been limited by low yields that are inefficient for biophysical applications. This low yield is due to the rapid ubiquitination of pro-IL-33 and immediate degradation by the proteasome, and usable amounts of pro-IL-33 have only been achieved through the inhibition of the proteasome through the use of MG132 (29). Although this allows for the expression of low levels of pro-IL-33 high level expression is not possible through this method as inhibition of the proteasome subsequently results in high amounts of cell stress and ultimately apoptosis of the transfected cells. Other methods previously used to obtain pro-IL-33 involve the use of *in vitro* expression systems; however, these methods too are limited by their ability to obtain high enough yields to allow for

biophysical characterization. Given the inability to use other systems, the development of a high expressing system independent of mammalian expression is paramount.

1.2 Materials and Methods

Protein Expression and Vector Generation Screening

DNA templates for pro-IL-33 were synthesized by Genscript and subcloned into vectors: pET-20b (EMD Millipore), pET-28a (EMD Millipore), pGEX-6p-1 (GE Healthcare), and pET-SUMO (Life Technologies). Vectors containing pro-IL-33 were then transformed into BL21(DE3) (Novagen), BL21(DE3) pLysS (Novagen), BL21(DE3)RIL (Novagen), Rosetta (EMD Millipore), OverExpress C41(DE3) (Lucigen), OverExpress C43(DE3) (Lucigen), and One Shot DH5 α (Life Technologies) and grown in 2X LB Media to an OD₆₀₀ at 0.6. Growth conditions were screened under three conditions: 4 hours at 37°C after the addition of 1mM isopropyl β -D-1-thiogalactopyranoside (IPTG), 8 hours at 30°C with the addition of 1mM IPTG, and 16 hours at 27°C with 0.5mM IPTG. Cells were evaluated based on final OD₆₀₀ to exclude growths with significant cell death from expression verification by Western Blotting.

Western Blotting of pro-IL-33

1mL aliquots of growths spun at 12,000 x g before aspiration of the media. Cells were then boiled after the addition of 2X Laemmli buffer at 95°C for 20 minutes to ensure complete lysis and full denaturation of protein samples. Samples were then run on a 16% SDS-PAGE gel at 100V for 2 hours before transfer onto Immobilon-P PVDF Membrane (EMD Millipore) at 100V for 45 minutes before blocking in 5% (v/v) Casein-TBST (Tris-buffered saline Tween-20; 0.05% Tween-20 v/v) at 4°C. Blots were then incubated with polyclonal goat anti-human IL33 (R&D Systems) or His-probe-HRP (Santa Cruz Biotech) depending on the tag present in 5%

(v/v) BSA-TBST at dilutions according to manufacturer instructions overnight at 4°C. Blots were then washed by three 5-minute washes in TBST (0.05% v/v) and anti-IL-33 antibodies were incubated with donkey anti-goat-HRP secondary antibodies (Santa Cruz Biotech) for 45 minutes followed by another round of three 5-minute washes. Detection was performed using SuperSignal West Pico Substrate (Pierce) with 10% (v/v) SuperSignal West Femto Substrate (Pierce) and imaged on ChemiDoc XRS+ System with Image Lab Software (Bio-rad).

Autoinduction and Purification of 6xHis-SUMO-pro-IL33 Construct

Autoinduction media was formulated as previously reported (30). Single colony cultures were attained off of LB-Agar plates with appropriate antibiotic added. Starter cultures were done in LB media overnight for a minimum of 14 hours. Cultures were added to complete autoinduction media at 27°C. After 24 hours, cultures were pooled and pelleted at 3,000 x g for 20 minutes at 4°C before removal of media and resuspension in pre-chilled His-tag binding buffer (50mM Na₃PO₄ pH 8, 1M NaCl, 25mM Imidazole, 10% glycerol) supplemented with cOMplete protease EDTA-free tablets (Roche) immediately prior to resuspension. Cells were lysed using an Emulsiflex C-5 Homogenized (Avestin) for 3 passes at 1,000 bar before centrifugation at 35,000 x g. Lysates were applied directly to a gravity driven glass Econo-column (Bio-Rad) with Ni-NTA Agarose beads (Qiagen) before incubation for 1 hour at 4°C. Flow through was then eluted and beads were washed with His-tag wash buffer (50mM Na₃PO₄ pH 8, 1M NaCl, 100mM Imidazole, 10% glycerol) before elution using His-tag elution buffer (50mM Na₃PO₄ pH 7, 1M NaCl, 250mM Imidazole, 10% glycerol). Elution samples were concentrated using 30kDa Vivaspin 20 Concentrator (GE Healthcare) before buffer exchange using PD-10 columns (GE Healthcare) into SUMO cleavage buffer (50mM Tris pH 7, 150mM

NaCl, 0.2% NP-40). Purity and tracking of pro-IL-33 was done using SDS-PAGE and western blotting methods.

1.3 Results

Vector Construction and Screening

Since no high-throughput method was available to screen a large amount of vectors and growth conditions in a time efficient manner, a logistic approach was applied where representative vectors of each type would be utilized in a deterministic manner (Figure 1-4). Prior to cloning into specific *E. coli* vectors, vectors were organized categorically based upon the size of tag that would be attached to the expressed protein: tagless, small protein tag, and large protein tag. Tagless vectors were defined as vectors that did not attach a tag to the protein of interest, small protein tags were those that attached less than 20 amino acids onto the protein such as 6xHis tags, and large protein tags were those constituting larger than 20 amino acids such as thioredoxin and SUMO tags. Via this system, three representative vectors were initially chosen for initial screening based on previous successes on the mature form of IL-33: a tagless form that utilized a pelB leader sequence (pET-20b), a His-tagged leader (pET-28a), and a GST-tagged leader with a PreScission Protease cleavage site (pGEX-6P-1). Each of these vectors was screened against all cell lines and visually evaluated both pre and post induction to determine cytotoxicity of protein expression. Of the vectors tested, only the GST-tagged construct transformed into DH5 α cells consistently grew to high optical density when cultured overnight at 27°C, retained the plasmid, and expressed the protein of interest. Further refinement and progress with the

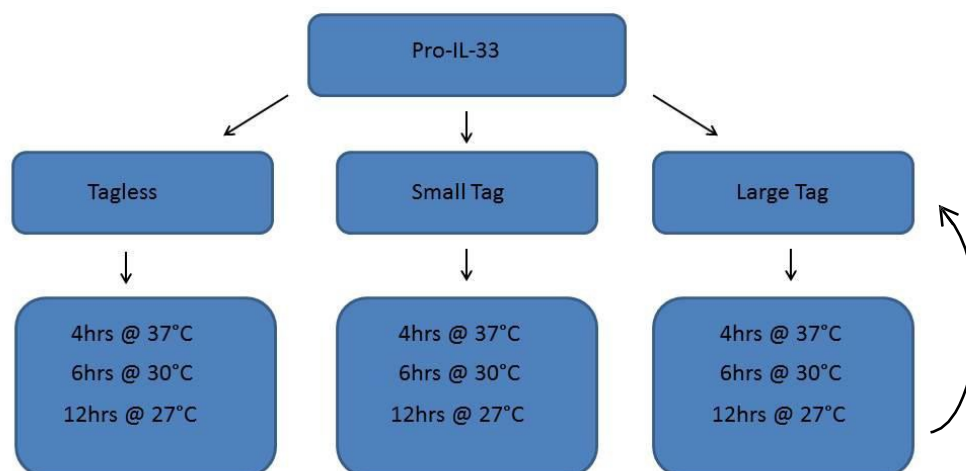


Figure 1-4: Vector Screen Schematic.

Division of vectors into three groups based on tag size. Representative groups for each were then screened under three different growth conditions of 4 hours at 37°C, 6 hours at 30°C, and 12 hours at 27°C to probe for optimal expression conditions

purification of the GST-construct revealed that although possible purification and removal of tag were possible, the final yield did not result in a high enough levels to be useful for biophysical characterization. As a result, further refinement into initial expression levels was required and other vectors were explored.

Since the large tag constructs had the most consistent results, further screens were done within the category consisting of a His-thioredoxin construct (6xHis-Trx) and a His-SUMO construct (6xHis-SUMO) which have been previously reported to have high success with cytotoxic and DNA binding proteins. Replication of the same methods done to the GST-fusion construct showed high expression levels with both C41 and C43 cell lines, with both constructs greater than that previously attained with GST-fusion. Induction times and temperatures both showed the highest levels of expression present in growths utilizing lower IPTG concentrations for upwards of 16 hours (Figure 1-5). Although there were equally good expression level, the His-thioredoxin construct primarily expressed in the insoluble fraction rather than the soluble resulting in further development of the His-SUMO construct for purification purposes rather than proceeding with the His-thioredoxin.

Expression Level Optimization of His-SUMO Construct

Due to positive results with induction at low temperatures and low IPTG concentrations; it was thought that an autoinduction based system may improve the expression levels higher than induction with IPTG alone. Through autoinduction, high density cultures and therefore high yield is attained without the use of IPTG. Because IPTG as lactose mimetic is unable to be degraded, the use

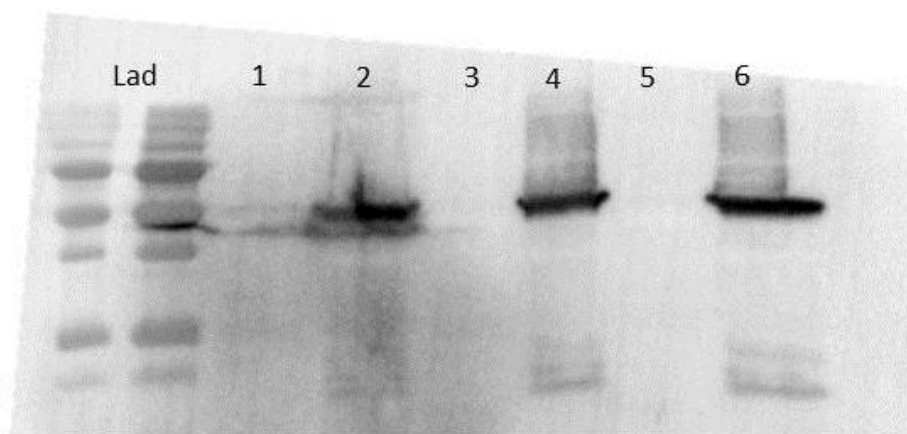


Figure 1-6: Western Blot Verification of 6xHis-SUMO-pro-IL-33.

Verification of expression of 6xHis-SUMO-pro-IL-33 transformed into C43 cells and grown for 12 hours at 27°C using IPTG by blotting against 6xHis. Lanes 1,3,5 are samples extracted from the culture pre-induction while lanes 2,4,6 are extracted from culture post-induction

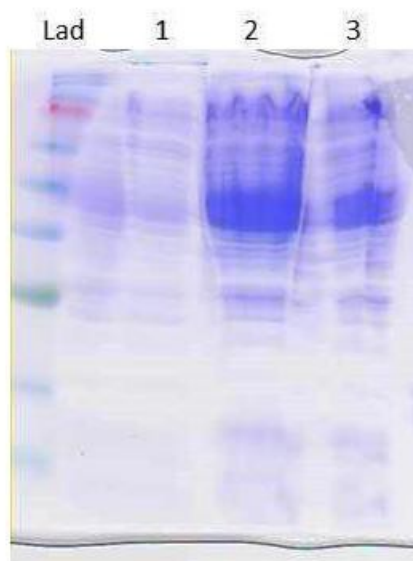


Figure 1-6: Comparison of traditional IPTG induced culture and autoinduction culture in Coomassie stained gel.

Two cultures were inoculated at the same time and with the same amount of starter culture from the same colony. IPTG induced cultures were induced at OD_{600} of 0.6 and transferred to the same shaker as an autoinduction culture. Lane 1 is a sample extracted from the IPTG induced culture after 12 hours of expression at 27°C, Lane 2 is a sample extracted from the autoinduction culture after 12 hours of expression at 27°C, and Lane 3 is a 1:1 dilution of the autoinduction culture sample with Laemmli buffer.

of IPTG can sometimes reduce yield due to increased cytotoxicity inherent in the inability to deactivate the T7 promoter. To alleviate this potential problem, autoinduction uses high levels of both glucose (to stimulate growth and be used as the primary sugar source) and lactose (to allow for activation of the T7 promoter and therefore expression). Unlike traditional expression systems utilizing IPTG, cells grown in autoinduction media are able to attain high density through their own ability to preferentially utilize glucose or lactose as its source of sugar. In this way, cells that undergo potentially apoptotic stress due to protein expression are able to preferentially utilize glucose rather than lactose to avoid excessive cell stress.

Utilization of autoinduction media showed positive results with yields significantly higher than with traditional LB media supplanted with IPTG. Both cell density and yield were compared using SDS-PAGE showing significantly greater yield of at least 2x greater yield as compared to traditional induction with IPTG alone. As such, autoinduction based expression was utilized for all work using the SUMO construct (Figure 1-6).

Nickel Resin Purification of SUMO-Construct and SUMO Protease Digestion

Subsequent lysis, centrifugation, and purification were done as described in the methods (Figure 1-7 and Figure 1-8)). Because the His-tag is only present on one the N-terminus of the protein and on SUMO-protease, it was thought that an on-column cleavage would yield a one step purification scheme; however, upon implementation, this method was unsuccessful. Contrary to what was expected, the cleaved form of the fusion retained a high affinity for the nickel resin causing it to only elute at high imidazole concentrations. As a result of this observation, on-column cleavage was avoided and instead, elution followed by buffer exchange into cleavage buffer was utilized.

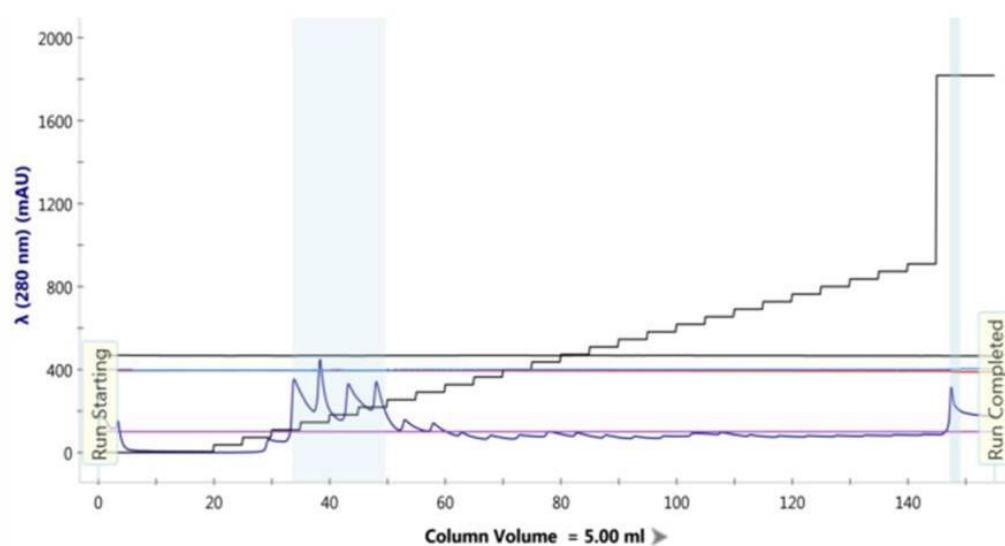


Figure 1-7: Chromatogram of nickel column purification observed by A_{280} .

Absorbance at 280nm is shown in blue while stepwise gradient of imidazole is shown in green where the end point is 100% elution buffer. Results of chromatogram show that the optimal concentration to wash the column with is at 50% buffer B which translates to 125mM imidazole while the 6xHis-SUMO construct elutes at 250mM imidazole.

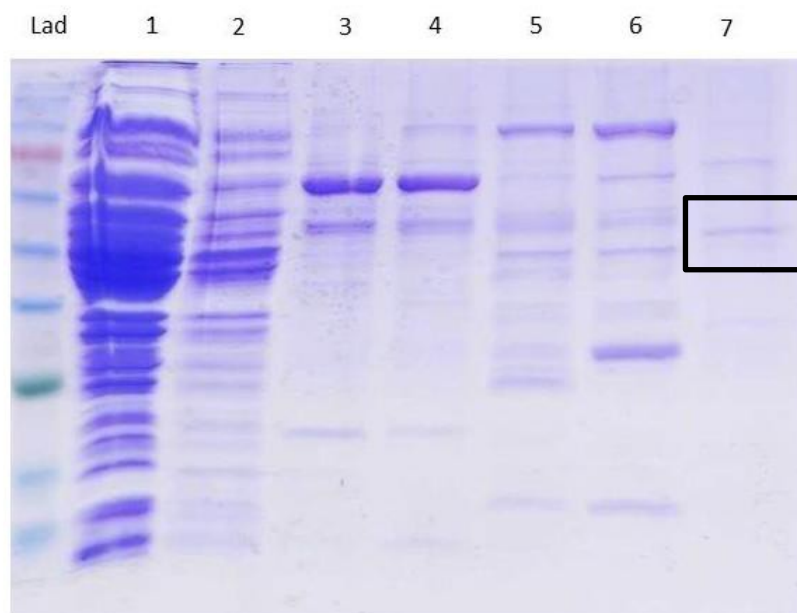


Figure 1-8: Coomassie stained SDS-PAGE of nickel column chromatogram.

Each peak present in the chromatogram was analyzed on SDS-PAGE. Lane 1 corresponds to pre-column, Lane 2 corresponds to flow through, Lane 3 corresponds to the peak corresponding with 15mM imidazole, Lane 4 corresponds to the peak elution with 20mM imidazole, Lane 5 corresponds to the peak elution at 25mM imidazole, Lane 6 corresponds to the peak elution at 30mM imidazole and Lane 7 corresponds to the peak eluting at 250mM imidazole. The band corresponding to pro-IL-33 is boxed in black

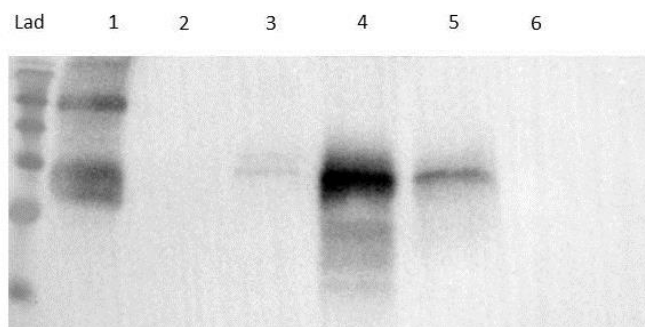


Figure 1-9: Western blot against IL-33 verification of post-cleavage size exclusion cleanup.

Lane 1 corresponds to entire post-cleavage reaction, Lane 2 through 5 each are peaks throughout the size exclusion cleanup. Smearing present in lane 4 suggests the presence of degradation products post-cleavage. Presence of a band in Lane 5 shows incomplete separation of products

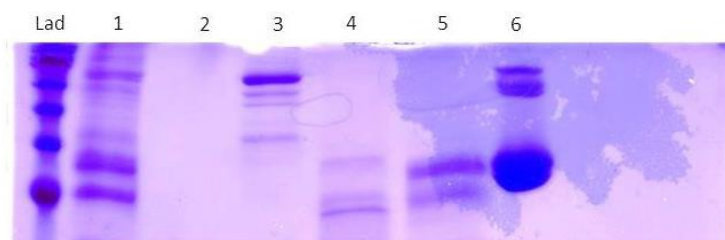


Figure 1-10: Coomassie verification of post-cleavage size exclusion cleanup.

Lane 1 corresponds to entire post-cleavage reaction, Lane 2 through 5 each are peaks throughout the size exclusion cleanup. Smearing present in lane 4 suggests the presence of degradation products post-cleavage. Presence of the band in 5 is due mostly due to SUMO protease though some trace remains of pro-IL-33

Post-digestion attempts to purify pro-IL-33 continued to prove problematic in attempting to achieve pure protein. Despite what was expected, purified pro-IL-33 did not bind to heparin resin nor was it able to purified using ion exchange. As such, size-exclusion chromatography was utilized to remove reaction products from the cleaved pro-IL-33 (Figure 1-9). This method resulted in mixed results that revealed that the size of the tag is too close in size to pro-IL-33 to purify without greatly impacting the yield of the products (Figure 1-10). To add to the challenges present, post digestion blots of pro-IL-33 revealed spontaneous degradation products present, and after 12 hours in cleavage buffer conditions, pro-IL-33 was no longer present in the sample. Although the cause of the degradation is unknown, this time limit great hindered the ability to attempt to purify pro-IL-33.

1.4 Discussion

Without high throughput methods to test vector and cell line screening, attempts to express new proteins is a highly time consuming process limited by a number of factors ranging from time of growth to ability to run multiple conditions in parallel. As such, the work described here attempts to use the properties and observations of each vector in a deterministic manner to iteratively probe for high expression constructs. In this manner, workload and workflow are greatly distilled and allow for logistic analysis behind the behavior of each construct rather than relying upon breadth of conditions tested. The application of this method upon the expression of pro-IL-33 has created the first known high expression conditions for this novel cytokine. Further improvement of conditions through the implementation of autoinduction expression system due to results of low temperature with

long growth times shows promise as an alternative to current methods implemented in the lab currently.

Despite the success found in elucidation of an *E. coli* based expression system, several challenges inherent to the protein arose in the form of high aggregation propensity as well as rapid degradation despite initial purification steps. The high aggregation propensity of the protein in question is thought to be due to disparate charges present between the pro-leader sequence and the mature domain. While the total isoelectric point of the protein balances to 8.6 based on amino acid sequence, the pro-leader sequence has a calculated isoelectric point of 10.27. This high isoelectric point is due to lysine rich regions which are thought to help mediate interactions between DNA. Conversely, the mature domain of IL-33 that consists of the beta trefoil motif has an isoelectric point of 4.93. This highly disparate distribution in charge is suspected to create bivalency of the overall protein and increase aggregation potential between molecules. While this interaction is normally attenuated due to the presence of high salt, SUMO protease digestion conditions require a lower amount of salt present that can have been shown to increase the aggregation potential of pro-IL-33. As such, optimization of protease conditions to ablate the effects of moderate salt conditions possibly through the utilization of ionic detergents that are compatible with SUMO protease should be tested.

The inherent instability found during expression and purification of pro-IL-33 may be a direct influence of the properties of the pro-leader sequence as well as an inherent property of the full length protein in general. Previous works have shown that pro-IL-33 is rapidly degraded in the cell by ubiquitination that is able to target the protein despite localization to the nucleus. In this manner, pro-IL-33 is a highly susceptible target for proteolysis which may

be a result of its regulation in the cell. Likewise, the compaction and overlapping of multiple domains present on the pro-leader sequence is highly indicative of the properties of inherently disordered proteins (IDPs). In order to allow the function transition between properties in overlapping domains, some proteins hold a certain amount of intrinsic disorder. By maintaining a high amount of disorder in the absence of ligand, IDPs are able to maintain function regions across overlapping amino acid sequences. Consistent with this hypothesis, homology analysis of the amount of disorder present in the pro-leader sequence through the use of the PONDR webserver (Predictor of Natural Protein Disorder) shows several regions present in the pro-leader with elevated disorder propensities (Figure 1-11) (31–35). Although this high degree of disorder imparts a high degree of flexibility in functional properties of a protein, it inherently destabilizes the secondary structure of the protein and makes it inherently susceptible for degradation.

1.5 Future Directions

Now that a high expression system is established for the expression and purification of pro-IL-33 as a fusion protein, the future work of the project lies within optimization of digestion and subsequent purification. One point of improvement lies in the digestion efficiency of SUMO protease which, after digestion, still contains a high population of fusion protein. Further optimization of this step may lead to greater yields downstream, and since there is inherent stability issues associated with the final product, optimization of this step may allow for downstream sample stability and therefore, yield. To improve post-cleavage purification of pro-IL-33, one method of purification that can be attempted to purify the post-protease cleavage reaction mixture is hydrophobic exchange chromatography since high salt

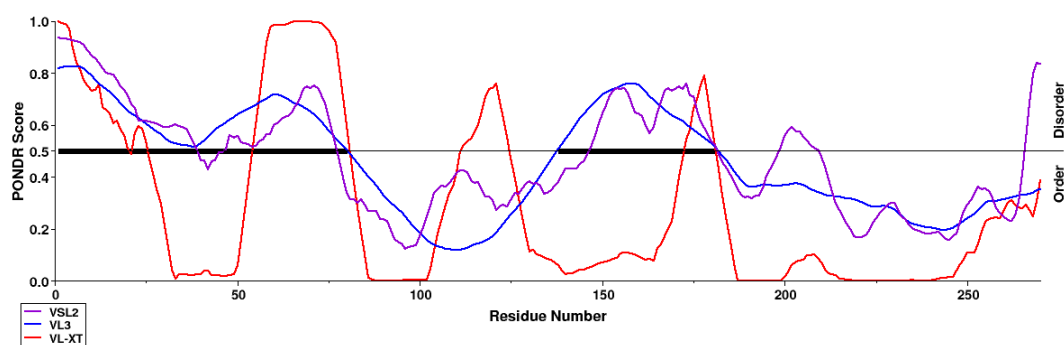


Figure 1-11: PONDOR Analysis of Pro-IL-33

Analysis of pro-IL-33 through input to the PONDOR webserver in which values over 0.5 PONDOR scores represent high propensity of disorder. Each color represents a different algorithm each trained to a database of disordered proteins or domains of varying lengths. Agreement between two of three values above 0.5 is denoted by a bold black line.

concentrations are a known inhibitor to SUMO protease function. Utilizing hydrophobic exchange chromatography, a gradient from high salt to low salt is utilized in order to elute proteins from hydrophobic resin. Due to pro-IL-33's high charged nature, hydrophobic exchange may have allow for purification by binding more strongly to the protease and cleavage products present.

Although not much is known about the pro-leader sequence post cleavage, it may be beneficial to utilize expression of solely the pro-leader sequence to better understand the role of pro-IL-33's role as a transcriptional regulator. Ideally, the ability the express and purify the pro-leader sequence alone is preferred, but may also be worthwhile to create constructs as fusion proteins. One unique fusion protein may be through the use of an alternative family member with similar fold such as IL-1 β . In doing so, the beta trefoil motif found in IL-1 β can be utilized as a pseudo functional tag that may create minimal perturbations to a fusion protein containing the pro-leader sequence. The use of this approach may not only yield insight on the behavior of the pro-leader sequence upon interaction with ligands as well as free, but may be able to help elucidate differences between IL-33 and IL-1 β .

In the event that stability of the final protein proves too much of an issue and is unavoidable, expression and characterization of pro-IL-33 as a fusion protein may be worthwhile. Although there may be perturbations through the addition of a SUMO tag, the construct can be manipulated in such a manner as to allow for a flexible linker to minimize the effect. In this way, isotopically labelled protein can be achieved while still maintaining the pro-leader and mature domain's interactions with one another as closely as possible. Doing so may yield data upon the behavior of the mature domain while the pro-leader sequence is still

attached as well as distal effects upon interactions between pro-leader sequence and binding partners such as DNA, histone, and NF- κ B.

1.6 References

1. Grivennikov SI, Greten FR, Karin M (2010) Immunity, Inflammation, and Cancer. *Cell* 140(6):883–899.
2. Dinarello CA (2000) Proinflammatory cytokines. *Chest* 118(2):503–508.
3. Lee S, Margolin K (2011) Cytokines in cancer immunotherapy. *Cancers (Basel)* 3(4):3856–3893.
4. Freeman BE, Hammarlund E, Raue H-P, Slifka MK (2012) Regulation of innate CD8+ T-cell activation mediated by cytokines. *Proc Natl Acad Sci* 109(25):9971–9976.
5. Yamane H, Paul WE (2012) Cytokines of the γ c family control CD4+ T cell differentiation and function. *Nat Immunol* 13(11):1037–1044.
6. Boyman O, Purton JF, Surh CD, Sprent J (2007) Cytokines and T-cell homeostasis. *Curr Opin Immunol* 19(3):320–326.
7. Thomas C, Bazan JF, Garcia KC (2012) Structure of the activating IL-1 receptor signaling complex. *Nat Struct Mol Biol* 19(4):455–457.
8. Medzhitov R, Preston-Hurlburt P, Kopp E, Stadlen a, Chen C, Ghosh S, Janeway C a (1998) MyD88 is an adaptor protein in the hToll/IL-1 receptor family signaling pathways. *Mol Cell* 2(2):253–258.
9. Weber A, Wasiliew P, Kracht M (2010) Interleukin-1 (IL-1) pathway. *Sci Signal* 3(105):cm1.
10. Dinarello CA (2009) Immunological and inflammatory functions of the interleukin-1 family. *Annu Rev Immunol* 27:519–550.
11. Oboki K, Nakae S, Matsumoto K, Saito H (2011) IL-33 and airway inflammation. *Allergy, Asthma Immunol Res* 3(2):81–88.
12. Sanada S, Hakuno D, Higgins LJ, Schreiter ER, McKenzie ANJ, Lee RT (2007) IL-33 and ST2 comprise a critical biomechanically induced and cardioprotective signaling system. *J Clin Invest* 117(6):1538–1549.

13. Jovanovic IP, Pejnovic NN, Radosavljevic GD, Pantic JM, Milovanovic MZ, Arsenijevic NN, Lukic ML (2014) Interleukin-33/ST2 axis promotes breast cancer growth and metastases by facilitating intratumoral accumulation of immunosuppressive and innate lymphoid cells. *Int J Cancer* 134(7):1669–1682.
14. Miller AM, Xu D, Asquith DL, Denby L, Li Y, Sattar N, Baker AH, McInnes IB, Liew FY (2008) IL-33 reduces the development of atherosclerosis. *J Exp Med* 205(2):339–346.
15. Kakkar R, Lee RT (2008) The IL-33/ST2 pathway: therapeutic target and novel biomarker. *Nat Rev Drug Discov* 7(10):827–840.
16. Ohno T, Oboki K, Kajiwara N, Morii E, Aozasa K, Flavell RA, Okumura K, Saito H, Nakae S (2009) Caspase-1, caspase-8, and calpain are dispensable for IL-33 release by macrophages. *J Immunol* 183(12):7890–7897.
17. Hayakawa M, Hayakawa H, Matsuyama Y, Tamemoto H, Okazaki H, Tominaga S ichi (2009) Mature interleukin-33 is produced by calpain-mediated cleavage in vivo. *Biochem Biophys Res Commun* 387(1):218–222.
18. Lefrançais E, Roga S, Gautier V, Gonzalez-de-Peredo A, Monsarrat B, Girard J-P, Cayrol C (2012) IL-33 is processed into mature bioactive forms by neutrophil elastase and cathepsin G. *Proc Natl Acad Sci* 109(5):1673–1678.
19. Lefrançais E, Cayrol C (2012) Mechanisms of IL-33 processing and secretion: Differences and similarities between IL-1 family members. *Eur Cytokine Netw* 23(4):120–127.
20. Hayakawa H, Hayakawa M, Kume A, Tominaga SI (2007) Soluble ST2 blocks interleukin-33 signaling in allergic airway inflammation. *J Biol Chem* 282(36):26369–26380.
21. Palmer G, Lipsky BP, Smithgall MD, Meininger D, Siu S, Talabot-Ayer D, Gabay C, Smith DE (2008) The IL-1 receptor accessory protein (AcP) is required for IL-33 signaling and soluble AcP enhances the ability of soluble ST2 to inhibit IL-33. *Cytokine* 42(3):358–364.
22. Chackerian AA, Oldham ER, Murphy EE, Schmitz J, Pflanz S, Kastelein RA (2007) IL-1 receptor accessory protein and ST2 comprise the IL-33 receptor complex. *J Immunol* 179(4):2551–2555.
23. Talabot-Ayer D, Lamacchia C, Gabay C, Palmer G (2009) Interleukin-33 is biologically active independently of caspase-1 cleavage. *J Biol Chem* 284(29):19420–19426.
24. Ali S, Mohs A, Thomas M, Klare J, Ross R, Schmitz ML, Martin MU (2011) The dual function cytokine IL-33 interacts with the transcription factor NF- κ B to dampen NF- κ B-stimulated gene transcription. *J Immunol* 187(4):1609–1616.

25. Moussion C, Ortega N, Girard JP (2008) The IL-1-like cytokine IL-33 is constitutively expressed in the nucleus of endothelial cells and epithelial cells in vivo: A novel "Alarmin"? *PLoS One* 3(10).
26. Carriere V, Roussel L, Ortega N, Lacorre D-A, Americh L, Aguilar L, Bouche G, Girard J-P (2007) IL-33, the IL-1-like cytokine ligand for ST2 receptor, is a chromatin-associated nuclear factor in vivo. *Proc Natl Acad Sci U S A* 104(1):282–287.
27. Lingel A, Weiss TM, Niebuhr M, Pan B, Appleton BA, Wiesmann C, Bazan JF, Fairbrother WJ (2009) Structure of IL-33 and Its Interaction with the ST2 and IL-1RAcP Receptors- Insight into Heterotrimeric IL-1 Signaling Complexes. *Structure* 17(10):1398–1410.
28. Shao D, Perros F, Caramori G, Meng C, Dormuller P, Chou PC, Church C, Papi A, Casolari P, Welsh D, Peacock A, Humbert M, Adcock IM, Wort SJ (2014) Nuclear IL-33 regulates soluble ST2 receptor and IL-6 expression in primary human arterial endothelial cells and is decreased in idiopathic pulmonary arterial hypertension. *Biochem Biophys Res Commun*.
29. Lüthi AU, Cullen SP, McNeela EA, Duriez PJ, Afonina IS, Sheridan C, Brumatti G, Taylor RC, Kersse K, Vandenabeele P, Lavelle EC, Martin SJ (2009) Suppression of Interleukin-33 Bioactivity through Proteolysis by Apoptotic Caspases. *Immunity* 31(1):84–98.
30. Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr Purif* 41(1):207–234.
31. Li X, Romero P, Rani M, Dunker A, Obradovic Z (1999) Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome Inform Ser Workshop Genome Inform* 10(I):30–40. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11072340>.
32. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK (2001) Sequence complexity of disordered protein. *Proteins Struct Funct Genet* 42(1):38–48.
33. Romero P, Obradovic Z, Dunker AK (1997) Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family. *Genome Informatics* 8:110–124. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11072311>.
34. Radivojac P, Obradović Z, Brown CJ, Dunker AK (2003) Prediction of boundaries between intrinsically ordered and disordered protein regions. *Pac Symp Biocomput*:216–227.
35. Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK (2005) Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins: Structure, Function and Genetics*, pp 176–182.

Chapter 2 : Experimental Characterization of the N-linked Glycosylations Present on the ST2 Receptor

2.1 Introduction

Glycosylation as a Post-translational Modification

There are a myriad of post-translational modifications that can be added to a protein after transcription off of the ribosome. Some of these modifications are necessary for intracellular signaling or for other use. Two of the most studied types of post-translational modifications are phosphorylation, which is common for signal transduction, and histone methylation which is necessary for transcriptional regulation. Other post-translational modifications can include those required for the activation or inactivation of proteins through cleavage by proteases or targeting for degradation through the attachment of ubiquitin (1, 2). These additional modifications to proteins create a new layer of regulation and an additional source of diversity in proteins.

One critical post-translational modification (PTM) involves the attachment of sugars or sugar moieties onto specific residues of a protein. These modifications are termed glycosylations due to their use of the sugar moieties such as glucose (3). Nearly all transmembrane proteins as well as many secreted proteins are glycosylated totaling to an estimated 60% of all proteins expressed undergoing post-translational glycosylation (4). Their high prevalence has particular influence in reevaluating the model of the extracellular membrane which has traditionally been shown as relatively barren with several proteins embedded into them. Through the years, it has been shown that the extracellular surface of the cell is far more complex (5). Through the use of glycan imaging by the Bertozzi group which has shown that the entirety of the extracellular surface is heavily coated in glycans, the traditional extracellular membrane model has been revised to one in which the surface of the cell is primarily dominated by the presence of glycans (6). In this way, the glycans

extend outward from proteins cell surfaces, creating an effect similar to the canopy of a jungle and has been subsequently termed the “glycan jungle” or glycocalyx (7). Due to their high concentration on the cell surface, glycans and their subsequent glycoproteins form one of the first layers of interactions involved in cell-cell and cell-pathogen interactions (8, 9). As such, glycosylations hold particular importance in understanding cancer metastasis, infection by viruses, and cell signaling. In fact, increases of specific glycoforms can be analyzed as early hallmarks of diseases or markers for cancer metastasis (10, 11).

In addition to the high concentration present on each cell, glycosylations have a high amount of variability in both composition and type, creating a high amount of diversity when attempting to study the biology of glycans (glycobiology). The types of sugars that constitute the building blocks of the sugar chains as well as how these building blocks are attached to one another are highly variable. These sugars are often times crucial to the proper function of both protein and glycan recognition by glycan binding proteins called lectins responsible for cell-cell interactions (3). Different types of glycosylations are classified based on their attachment to residues on transcribed proteins. N-linked glycosylations, O-linked glycosylations, and phosphoglycosylations are named based on their attachment to nitrogens (on asparagine residues), attachment to the oxygen of serine and threonine residues, and attachment to phosphorylated serine and threonine residues, respectively (3, 12). Adding further to the heterogeneity and diversity of glycosylated proteins (glycoproteins), the sugar branching structures attached at each site may not be consistent across all proteins expressed nor will they have the same type of glycan present at a given site (13). This sample heterogeneity greatly convolutes data and presents a distinct challenge in structure elucidation through traditional X-ray crystallography or NMR

based approaches (4). Further adding to the challenge of structural elucidation of glycans is the strong and integral influence of the expression system on the final glycosylation patterns (14). Since glycosylation pathways and subsequent structural patterns found in attached sugars are highly specific to the species expressing the protein, the system by which samples are derived from will invariably lead to varying structures of sugar as well as creating challenges for traditional forms of structural and biophysical characterization methods that require the use of high concentrations of homogenous protein (15). Therefore, understanding the sources of protein expression and their interactions with the addition of glycans is integral to elucidating the structure, function, and dynamics of glycosylated proteins.

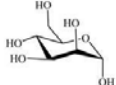

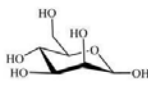

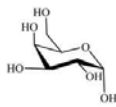

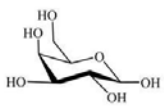

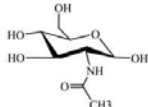

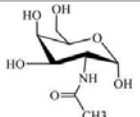

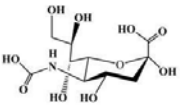

N-linked Glycosylations

Of all the variations in glycosylations, the N-linked varieties constitute the largest and most well understood type of glycosylation. Unlike many other glycosylations, N-linked is governed by the presence of a consensus sequence consisting of Asn –X – Ser/Thr where X can be any amino acid except proline (3, 12). Although the attachment of N-linked glycosylations requires the presence of this consensus sequence, not all sites with this consensus sequence are occupied on a given protein (16). The lack of uniform population of N-linked glycosylations on their conjugate consensus sequences represents yet another layer of complexity in glycosylation events and subsequent patterning.

N-linked glycosylations can be further broken down into three distinct groups: oligomannose, hybrid, and complex type glycans. All N-linked glycans contain a core moiety

Table 2-1: Table of common sugars moieties and their corresponding structure and symbol.

Molecules of the same structure are represented using the same structure regardless of isomer. Atomic composition for each molecule dictates symbol shape while color denotes stereochemistry.

MOLECULE	STRUCTURE	SYMBOL
α -D-mannose		
β -D-mannose		
α -D-galactose		
β -D-galactose		
N-Acetylglucosamine		
N-Acetylgalactosamine		
N-Acetylneuraminic Acid		

consisting of $\text{Man}\alpha 1-6(\text{Man}\alpha 1-3)\text{Man}\beta 1-4\text{GlcNAc}\beta 1-4\text{GlcNAc}\beta 1-\text{Asn-X-Ser/Thr}$ where the mannose groups make up the first branching point referred to as “antennae”. In oligomannose structures, only mannose residues are attached onto the core structure. This differs from complex glycans in which further branching and creation of new antennae are initiated by the presence of GlcNAc residues. Complex glycans also have the ability to exist in highly branched structures with hepta-antennary structures and have been found in birds and fish, but not in mammals. The final group of N-linked glycans is the hybrid groups which blend features from both the oligomannose and complex groups by containing one branch with only mannose residues and a second, separate branch containing GlcNAcs (3, 12).

Despite the complexity inherent to N-linked glycosylations, these N-linked sites can be targeted through the use of the enzyme endoglycosidase peptide-N-glycosidase F (PNGase F). This enzyme recognizes the conserved N,N-diacetylchitobiose core found in all N-linked glycans and cleaves between the first N-acetylglucosamine and the asparagine residue that makes up the initial glycan-amino linkage. Due to the reliability of the PNGase F enzyme and the presence of a consensus sequence, N-linked glycosylations represent the majority of identified glycan groups in glycosylated proteins (17–19).

Glycosylations in the IL-33 system

Compared to other post-translational modifications such as phosphorylation, the functional consequences of glycans on proteins have not been investigated in great detail and to date are poorly understood. Fibroblast Growth Factor (FGF) represents a structural homolog to the IL-1 family and its function has been shown to be highly dependent upon glycosylation sites on its conjugate extracellular receptor (20, 21). The FGF family of proteins is highly homologous to the IL-1 family of proteins where the signaling protein

consists of a β -trefoil fold that binds to an extracellular surface bound receptor composed of three immunoglobulin (Ig) domains. Binding of the FGF ligand to the extracellular conjugate receptor initiates a dimerization event modulated through interactions of a proteoglycan heparin sulfate to achieve full signaling. Through the modulation of interactions by heparin, FGF activity becomes highly modular and enables an extra degree of control over the system (22).

Although the fundamental signaling mechanism by which FGF and the IL-1 family signal are different, there still exist many possible areas by which glycosylations can have a role within the IL-1 family of proteins. Within the IL-1 family of cytokines, many of the conjugate receptors that the members bind and signal through have been shown to be glycosylated. As stated in the introduction, signaling through the IL-1 family occurs through the formation of a heterotrimeric complex consisting of an IL-1 family ligand, its unique receptor, and the receptor accessory protein IL-1RAcP. This binding process occurs sequentially through the binding of ligand to receptor and then bound complex to receptor accessory protein. Previous studies used to elucidate the structure of an IL-1 family heterotrimeric signaling complex by studying the activating complex formed with IL-1 β found glycosylation sites including two O-linked glycans GlcNAc Tyr234 and N-linked glycan at Asn216 of IL-1R1. The placement of these glycosylation sites suggests that they influence the formation of the binding interface between receptor accessory protein called IL-1RAcP and IL-1R1. Samples utilized for crystallization were de-glycosylated with endoglycosidase H (Endo H) resulting in only the base GlcNAc being present in the structure. As a result, the

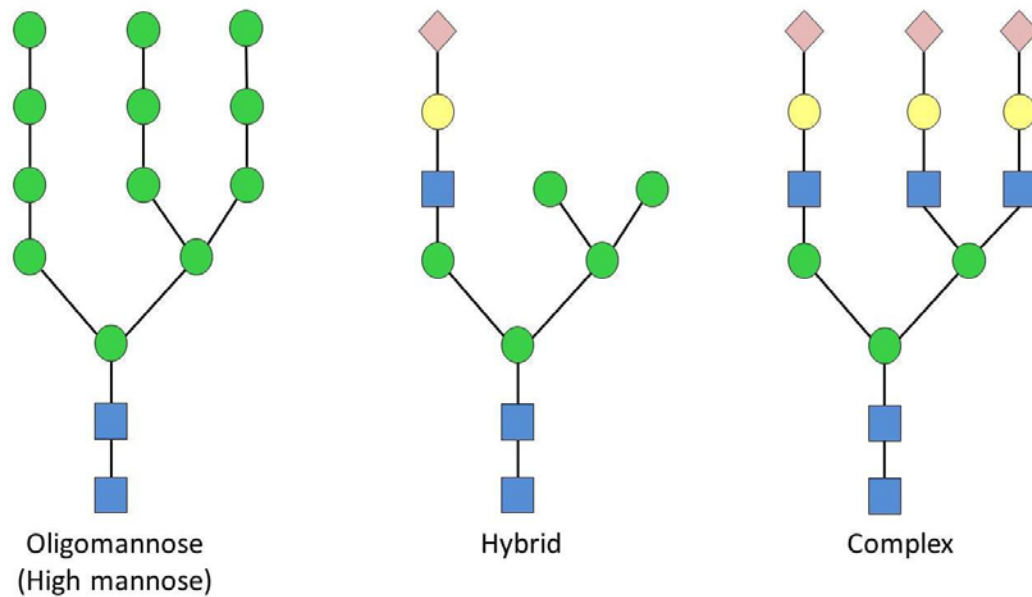


Figure 2-1: Structure of the three classes of N-linked glycans

All N-linked glycans are divided into one of three classes based on their composition. Oligomannose also known as high mannose contain a high percentage of mannose content in their branching and can have large branching structures consisting primarily of mannose aside from the N,N-diacetylchitobiose core. Complex glycans consists of three mannose spanning the first branching segment following the N,N-diacetylchitobiose core and are commonly heavily sialited. Fucosylation of the first GlcNAc is also common. Hybrid glycans are a mix between oligomannose and complex class glycans in that they all retain a 5 mannose structure following the N,N-diacetylchitobiose core. They can also be sialated and fucosylated.

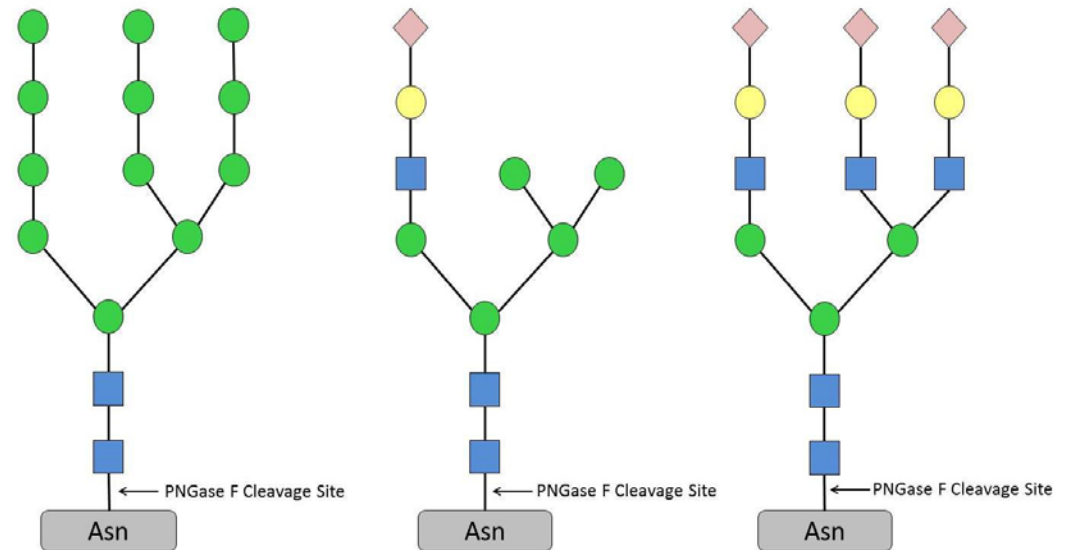


Figure 2-2: Cleavage site of PNGase F on N-linked glycans

Cleavage of N-linked glycans is universal by PNGase F in that the only required structure for cleavage is the conserved N,N-diacetylchitobiose core. Fucosylation of the inner GlcNAc hinders cleavage, but cleavage still occurs.

interaction interface containing fully branched glycosylations may be much more complex than initially proposed via the crystal structure. Likewise, these same interactions involving core GlcNAc residues are no longer present in the binding interface between the decoy receptor IL-1R2. It is likely that these missing sugars could constitute a possible interaction step necessary for full activation of the system, though this interaction has not been explored in detail (23).

In the case of the IL-33 system, there are two known players that are glycosylated: the ST2 receptor and IL-1RAcP. Since the Receptor Accessory Protein is shared in the signaling complex among all IL-1 family members, the same glycosylations key in establishing the IL-1 β heterotrimeric structure will be present when involved in IL-33 based signaling (23, 24). Present in the crystal structure of IL-33 bound to the sST2 receptor are three GlcNAc residues present at Asn95, Asn140, and Asn191, suggesting that N-linked glycosylations are present. Additionally, there are a total of eight Asn-X-Ser/Thr consensus motifs within ST2 (24). To date, the verification of glycan occupancy and their potential functional roles for ST2 has not yet been achieved. The glycosylations present in the crystal structure of the IL-33/ST2 complex were previously shown not to be involved in the binding of IL-33 to ST2; however, samples used for structural and binding studies were obtained from an insect cell line that have inherent differences in both glycosylation patterns as well as in the structures of the glycans attached at each site.(25) Therefore, the model that they propose may have greater complexity in mammalian systems due to the increased variety and branching of glycans possible on these derived receptors.

2.2 Materials and Methods

In Solution PNGase F Digestion

Samples of sST2 (R&D Systems) were resuspended in Denaturing Buffer (20mM Tris pH 7.5, 50mM NaCl, 0.1% SDS, 10mM DTT) before heat denaturation at 100°C for 10 minutes. Denatured samples were then flash frozen and lyophilized before resuspension in PNGase F Digestion buffer (50mM Sodium Phosphate pH 7.5 and 1% NP-40) formulated with either normal abundance water or heavy isotope water (^{18}O) and 1U/uL PNGase F (New England BioLabs) overnight at 37°C. Reactions were quenched using 2X Laemmli Buffer and run on a 8% SDS-PAGE for 2 hours at 100V. Gels were then stained using Coomassie Brilliant Blue and bands corresponding to de-glycosylated protein were excised for in gel tryptic digestion.

In Gel Tryptic Digest

The gel slices were cut to 1mm by 1 mm cubes and destained 3 times by first washing with 100 ul of 100 mM ammonium bicarbonate for 15 minutes, followed by addition of the same volume of acetonitrile (ACN) for 15 minutes. The supernatant was aspirated and samples were dried in a speedvac. Samples were then reduced by mixing with 200 μl of Reducing Buffer (100mM ammonium bicarbonate and 10mM DTT) and incubated at 56°C for 30 minutes. The liquid was removed and 200 ul of Alkylation Buffer (100 mM ammonium bicarbonate, 55mM Iodoacetamide) was added to gel pieces and incubated at 27°C in the dark for 20 minutes. After the removal of the supernatant and one wash with 100 mM ammonium bicarbonate for 15 minutes, same volume of ACN was added to dehydrate the gel pieces. The solution was then removed and samples were dried in a speedvac. For digestion, enough solution of ice-cold Trypsin (0.01 $\mu\text{g}/\text{ul}$) in 50 mM

ammonium bicarbonate was added to cover the gel pieces and set on ice for 30 min. After complete rehydration, the excess protease solution was removed, replaced with fresh 50 mM ammonium bicarbonate, and left overnight at 37°C. The peptides were extracted twice by the addition of 50 µl of 0.2% formic acid and 5 % ACN and vortex mixing at room temperature for 30 min. The supernatant was removed and saved. A total of 50 µl of 50% ACN-0.2% formic acid was added to the sample, which was vortexed again at room temperature for 30 min. The supernatant was removed and combined with the supernatant from the first extraction. The combined extractions are analyzed directly by liquid chromatography (LC) in combination with tandem mass spectroscopy (MS/MS) using electrospray ionization.

LC-MS/MS analysis

Protease-digested peptides were analyzed by ultra high pressure liquid chromatography (UPLC) coupled with tandem mass spectroscopy (LC-MS/MS) using nano-spray ionization. The nanospray ionization experiments were performed using a TripleTof 5600 hybrid mass spectrometer (ABSCIEX) interfaced with nano-scale reversed-phase UPLC (Waters corporation nano ACQUITY) using a 20 cm-75 micron ID glass capillary packed with 2.5-µm C18 (130) CSH™ beads (Waters corporation). Peptides were eluted from the C18 column into the mass spectrometer using a linear gradient (5–80%) of ACN (Acetonitrile) at a flow rate of 250 µl/min for 1h. The buffers used to create the ACN gradient were: Buffer A (98% H₂O, 2% ACN, 0.1% formic acid, and 0.005% TFA) and Buffer B (100% ACN, 0.1% formic acid, and 0.005% TFA). MS/MS data were acquired in a data-dependent manner in which the MS1 data was acquired for 250 ms at m/z of 400 to 1250 Da and the MS/MS data was acquired from m/z of 50 to 2,000 Da. The Independent data acquisition (IDA) parameters

were as follows; MS1-TOF acquisition time of 250 milliseconds, followed by 50 MS2 events of 48 milliseconds acquisition time for each event. The threshold to trigger MS2 event was set to 150 counts when the ion had the charge state +2, +3 and +4. The ion exclusion time was set to 4 seconds. Finally, the collected data were analyzed using Protein Pilot 4.5 (ABSCIEX) for peptide identifications using appropriate settings when utilizing ¹⁸O.

Endoglycosidase Digestion

Samples of sST2 (R&D Systems) were resuspended in Endo H Buffer (20mM Tris pH 7.5, 50mM NaCl, 0.1% SDS, 10mM DTT), Endo-F1 buffer (250mM sodium phosphate buffer pH 5.5, 0.1% SDS, 10mM DTT), or Endo-F2/F3 buffer (250mM sodium acetate buffer pH 4.5, 0.1% SDS, 10mM DTT) before heat denaturation at 100°C for 10 minutes. Enzymes Endo H (New England BioLabs), Endo F1 (Sigma-Aldrich), Endo F2 (Sigma-Aldrich), and Endo F3 (Sigma-Aldrich) were added at 1U/uL for each sample as dictated by the buffer present and allowed to digest overnight at 37°C. Reactions were quenched using 2X Laemmli buffer and resolved on an 8% SDS-PAGE at 100V for 2 hours before western blotting.

Western Blots of Endoglycosidase Digestion Products

SDS-PAGE gels were transferred to Immobilon-P PVDF Membranes (EMD Millipore) at 100V for 45 minutes before blocking in 5% (v/v) Casein-TBST (Tris-buffered saline Tween-20; 0.05% Tween-20 v/v) at 4°C. Blots were then incubated with polyclonal mouse anti-human sST2 (R&D Systems) in 5% (v/v) BSA-TBST at dilutions according to manufacturer instructions overnight at 4°C. Blots were then washed by three 5-minute washes in TBST (0.05% v/v) and anti-IL-33 antibodies were incubated with goat anti-mouse-HRP secondary antibodies (Santa Cruz Biotech) for 45 minutes followed by another round of three 5-minute washes. Detection was performed using SuperSignal West Pico Substrate (Pierce) with 10%

(v/v) SuperSignal West Femto Substrate (Pierce) and imaged on ChemiDoc XRS+ System with Image Lab Software (Bio-rad).

2.3 Results

Gel Mobility Shift of N-linked Glycosylated Proteins

To properly elucidate the functional consequences of glycosylations on the IL-33 heterotrimeric signaling complex, in particular, their influence upon the function of sST2, identification of specific glycosylation sites is necessary for the cell line that the receptor is derived from. Because the regulation of glycosylation attachments can be highly variable and because the underlying mechanisms controlling the attachment of sugars are not fully understood, occupancy of sites can be variable even among samples derived from the same source. To this end, the presence of N-linked glycosylations derived from this source was first probed using a gel mobility shift assay.

Determination of the presence of N-linked glycosylation was achieved through the use of gel mobility assays combined with enzymatic cleavage using PNGase F and Endo H of the N-linked glycans present on a given sample of sST2. Since sodium dodecyl sulfate is only able to bind to the amide backbone of a protein and not the attached glycans, the presence of glycosylations leads to an increase in the estimated molecular weight in SDS-PAGE due to the increased size due to attachment of glycans. However, because of the SDS does not

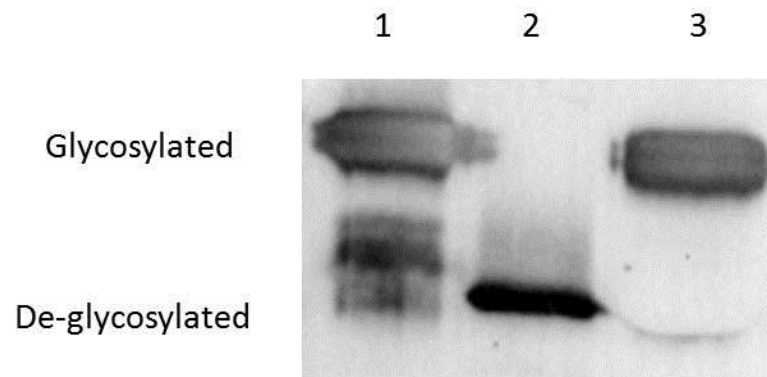


Figure 2-3: Gel mobility shift anti ST2 western blot of sST2 upon deglycosylation

Samples of sST2 were denatured (Lane 1) and then incubated with either PNGase F (Lane 2) or Endo H (Lane 3). Cleavage of PNGase F denotes the presence of N-linked glycans and in the case of sST2, shows as an observed shift of ~30kDa. Samples incubated with Endo H were unaffected suggesting that glycans present fall in the class of complex glycans.

directly interact with the glycan chains, the shift associated with an increase in molecular weight is not universally correlated to the size of the glycan or the amount of sites occupied in a linear manner, the results of the mobility shift can only be used as a rough estimation of the extent of glycosylations present on the sample. In the case of sST2, the observed shifts are large, constituting an equivalent shift of roughly 30kDa which totals to roughly half the weight of sST2 as seen in Figure 2-3. To determine whether the large gel mobility shift was a result of highly occupancy of N-linked glycosylation sites or greater branching structures, a mass spectrometry analysis was used (19, 26).

Site Specific Identification of N-linked Glycosylation Site Occupancy

Having identified the presence of N-linked glycans on the sST2 receptor, site specific occupancy of N-linked glycans were targeted for identification in order to determine the nature of the large molecular weight shift. In order to achieve site specific identification of glycosylations, a mass spectrometry based approach was used that combined the site specific sensitivity of proteomic based methods coupled with enzymatic digestion. During the removal of N-linked glycans by PNGase F, the mechanism of cleavage occurs not between the terminal nitrogen and the first oxygen of the GlcNAc, but between the side chain carbon and terminal nitrogen. Cleavage at this site releases the nitrogen of the asparagine with the glycan and replaces the site with oxygen from locally abundant water Figure 2-4. As a result, amino acids with glycans attached are converted from asparagine to aspartates post-cleavage and contain an observable +1 Da mass shift associated with the loss of nitrogen and replacement with oxygen. In order to observe this mass shift, a data-dependent acquisition (DDAc) method on a liquid chromatography mass spectrometer (LC-

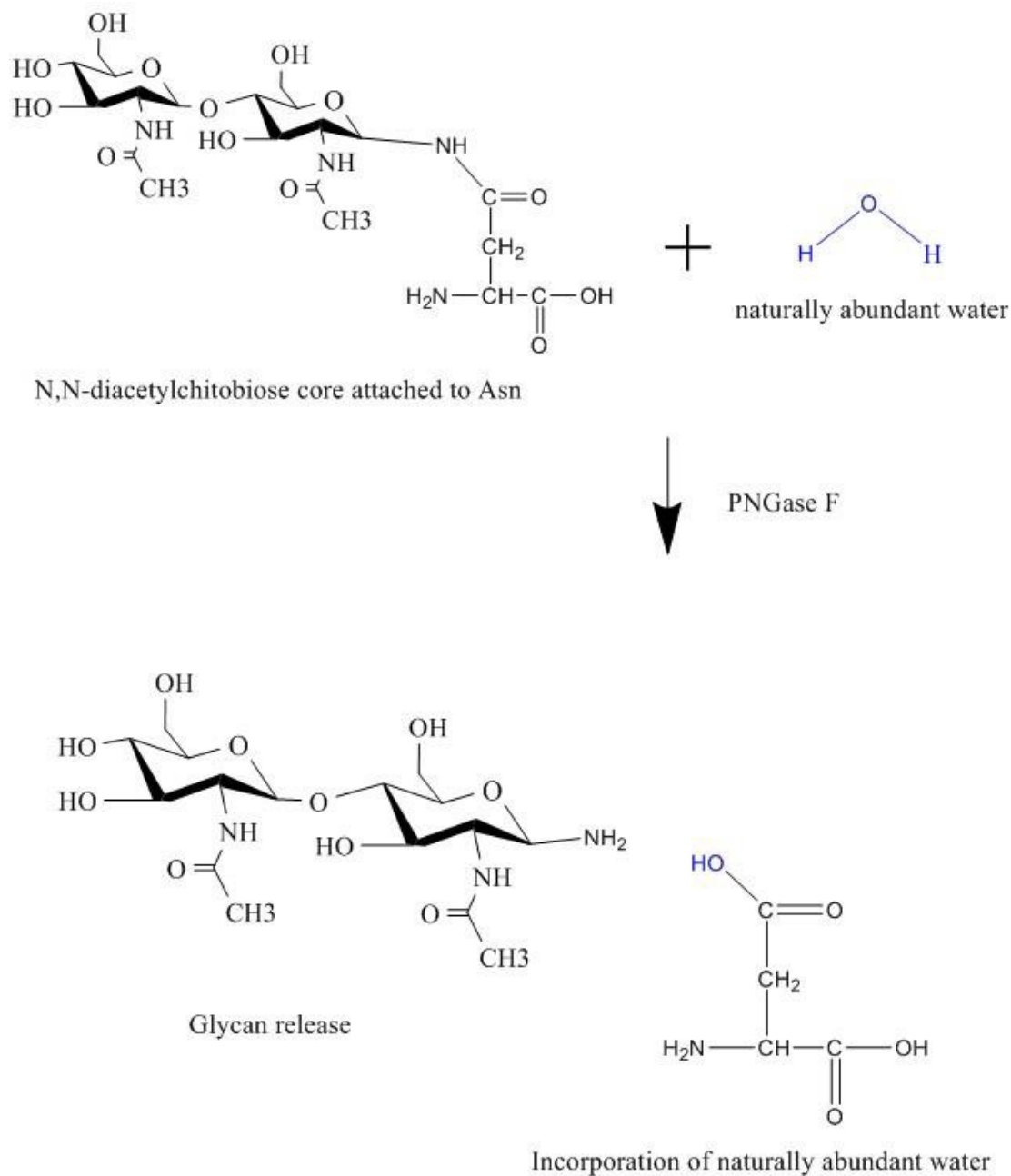


Figure 2-4: PNGase F release of glycan mechanism

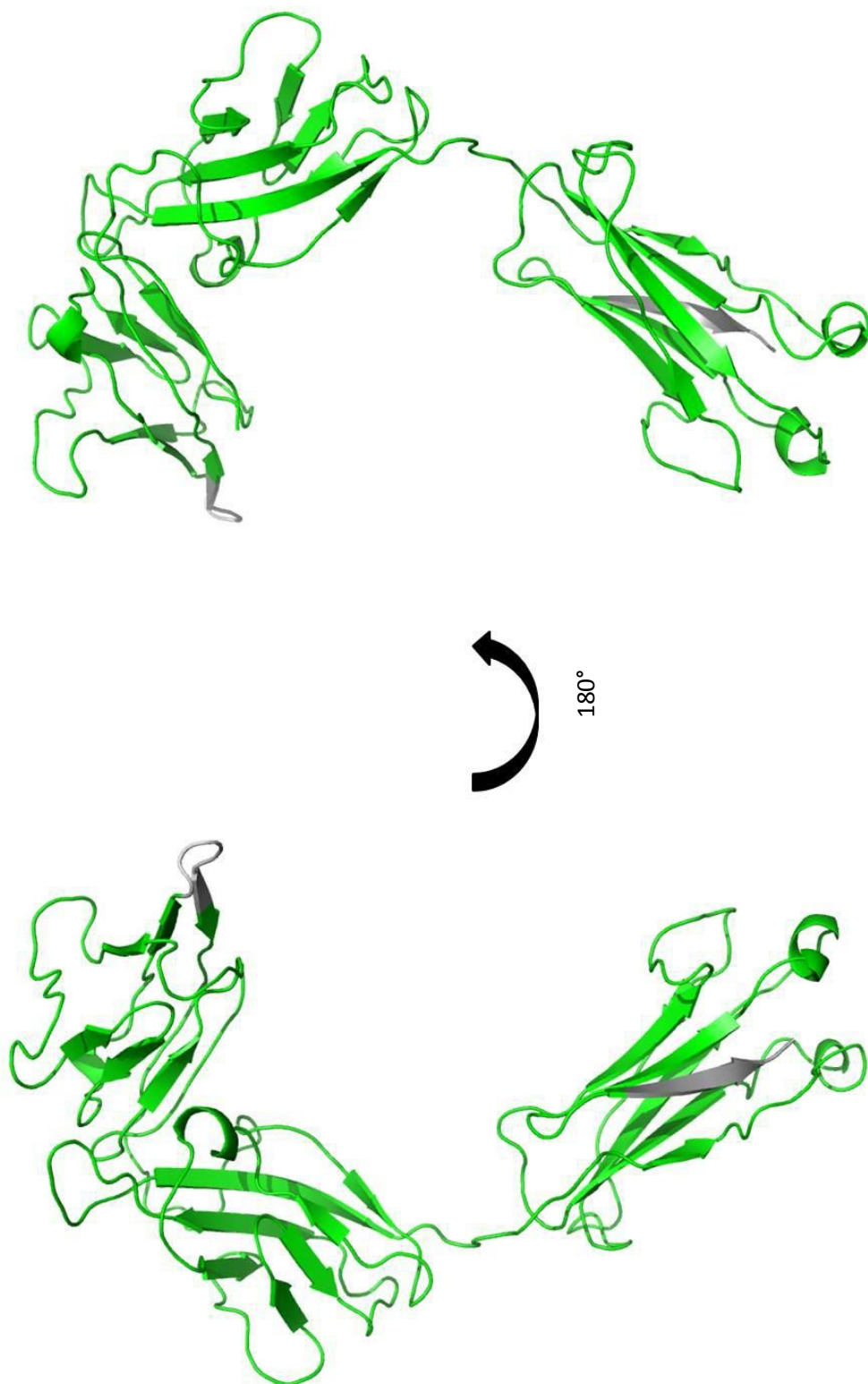
Release of glycans by PNGase F occurs before the N,N-diacetylchitobiose core at the bond between the terminal nitrogen and carbon on asparagine. Cleavage at this point results in incorporation of oxygen from naturally abundant water resulting in the conversion of an asparagine residue to an aspartate.

MS/MS) commonly used for peptide sequencing and proteomic based research was utilized. In this type of method, peptides are applied onto a reverse phase ultra-high pressure liquid chromatography (UPLC) system with the outflow attached to the ionization source. As peptides are eluted off the column, an initial data independent spectra is generated containing all ionized molecules present in the sample. From this data independent spectra termed MS1, the system then chooses all peaks above a specified cutoff intensity to send to collision induced dissociate (CID) chamber and yield a data-dependent spectra that contains the fragmentation pattern of the peptide in question termed MS2. Fragmentation of each peptide occurs from each termini and proceeds inwards resulting in two different ion series: b-ions from the N-termini and y-ions from the C-termini. As a result, fragmentation patterns can be matched with specific amino acid sequences as well as post-translational modifications with fixed mass shifts such as phosphorylations or deaminations through statistical methods. Through the combination of enzymatic digestion to create site specific mass shifts and proteomic based mass spectrometry methods, identification of N-linked glycosylation site is obtainable through tracking of mass shifts associated with deamination (27, 28).

Prior to application to LC-MS/MS, coverage of the sST2 must be achieved through proper digestion and ionization of peptide fragments. To achieve this, nearly any protease with a highly specific cut site can be utilized, though subsequent ionization efficiency and identification is not entirely dependent on the peptide fragment generated. For samples of sST2, trypsin was chosen due to its high cut site avidity and ability to impart an overall charge to the peptide fragment due to its cleavage on the C-terminus of arginine and lysine residues. Ionization efficiency of peptides produced was tested experimentally assuming

Figure 2-5: Peptide coverage map of sST2 by Trypsin

Coverage of nearly all of sST2 was attained by trypsin digestion with only two peptides missing. One peptide constitutes the C-terminal end and does not contain any predicted glycosylation sites by consensus sequence. The second peptide constitutes a peptide in Ig1 and contains one of the glycosylation sites shown to be occupied in insect expressed samples.



deamination (the process by which asparagine is converted to aspartic acid) as a possible post translational modification. The resulting coverage map was plotted onto the structure of sST2 and showed that 285 of 297 residues were covered showing the absence of only two possible peptides Figure 2-5. These two peptides consist of the last fragment on the C-terminus and one possible glycosylation site predicted by the consensus sequence that was previously reported to have been occupied (28). With coverage of nearly all predicted N-linked glycosylation sites achieved with trypsin, de-glycosylated samples were then analyzed for mass shifts resulting from PNGase F digestion. When searching for mass shifts associated with PNGase F digestion, y and b ion series were compared to verify mass shifts present on asparagine residues part of the consensus sequence. The resulting MS2 spectra (Figure 2-6) were plotted based on intensity versus mass over charge and the masses compared to Table 2-2. In the case of Figure 2-6, the charge state of the peptide was found to be +1 Da resulting in masses proportional to those presented. The most notable mass shift present is the peak at 533.2920 Da which corresponds to a y5 ion with theoretical mass of 532.30 Da found in Table 2-2. In order to insure that the deamination event is the result of PNGase F digestion and not an artifact of digestion using trypsin, the digestion of glycans using PNGase F can be done with a high abundance of heavy water (H_2^{18}O) while the tryptic digest can be done in the presence of naturally abundant water (H_2^{16}O). In this way, an additional mass shift is conferred upon PNGase F digestion that is different from possible deamination event of the tryptic digestion. This spectra is verified in Figure 2-7 where the same peptide confirmed earlier is reanalyzed for the additional mass shift of +2 Da due to the presence of heavy water found when comparing Table 2-3 with the results of Figure 2-7. Consistent with earlier results, the spectra at the same y5 ion at 535.2976 Da (Table 2-3) compared to the

Table 2-2: Table of theoretical fragmentation masses of the peptide AAVLWQLNGTK in the absence of glycosylation and de-glycosylated by PNGase F in ^{16}O

Masses of the peptide are given in the table where a +1 Da shift is present in the DeGly columns denoting de-glycosylated in the presence of ^{16}O . Ion series are described by b-ions which fragment from the C terminus and y-ions that fragment from the N-terminus. The Asn residue with that may be de-glycosylated is highlighted in yellow

Amino Acid	b-ion	y-ion	b-ion (DeGly)	y-ion (DeGly)
A	72.044	1200.674	72.044	1201.674
A	143.082	1129.637	143.082	1130.637
V	242.151	1058.6	242.151	1059.6
L	255.235	959.531	255.235	960.531
W	541.314	846.447	541.314	847.447
Q	669.372	660.368	669.372	661.368
L	782.457	532.31	782.457	533.31
N	896.499	419.225	897.499	420.225
G	953.521	305.183	954.521	305.183
T	1054.569	248.161	1055.569	248.161
K	1182.664	147.113	1183.664	147.113

Figure 2-6: MS2 Spectra of a representative de-glycosylated peptide digested by PNGase F in ^{16}O

MS2 spectra extracted from a mass spectrometer post CID with of the de-glycosylated peptide "AAVLWQLNGTK" with a charge state of +1. Intensity units are based by the number of counts or instances a peptide is measured. Fragmentation pattern matches the theoretical masses found in Table 2-2 of a de-glycosylated peptide using PNGase F in the presence of ^{16}O . The peak corresponding to the y5 ion is marked in (*)

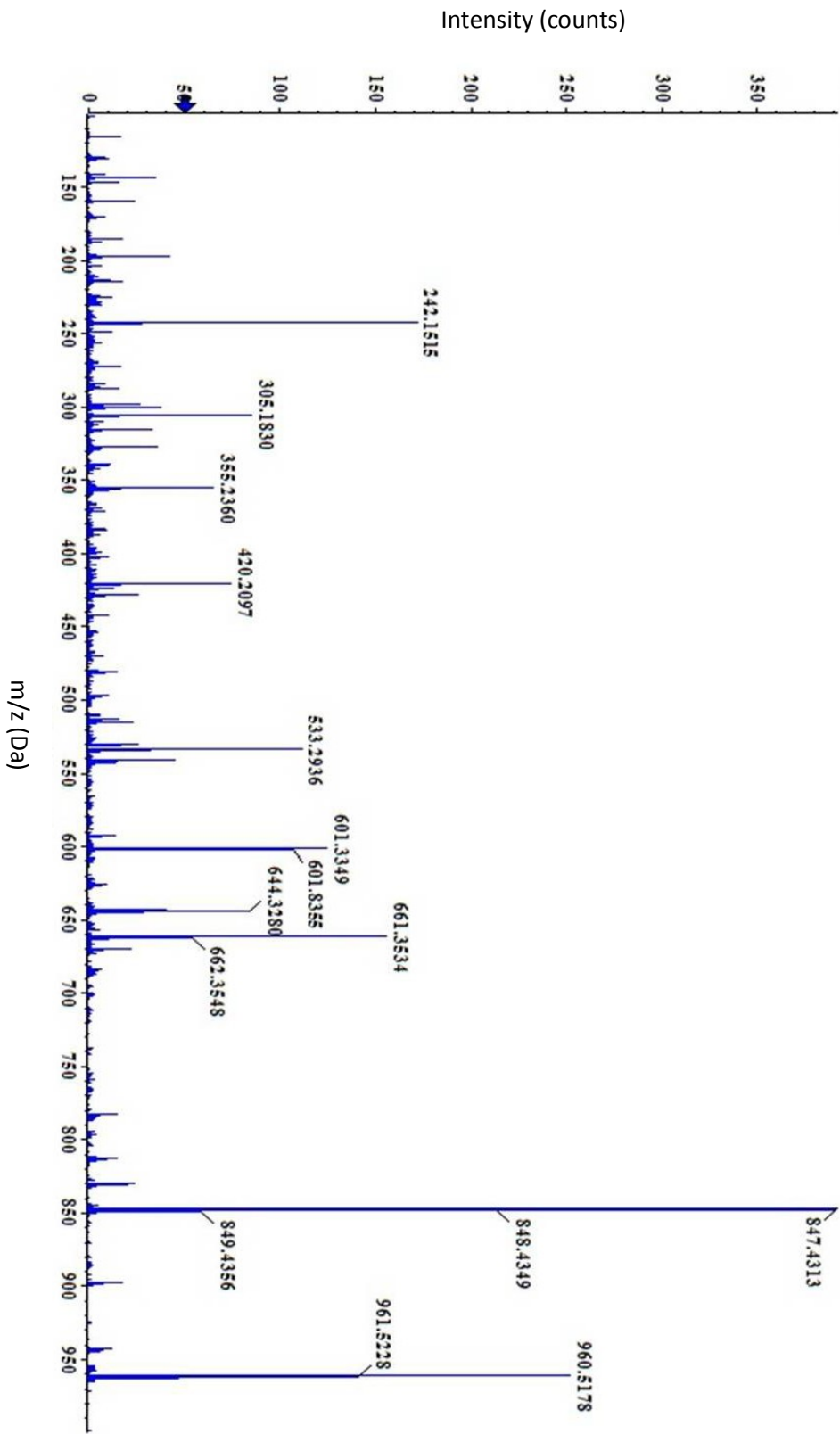


Table 2-3: Table of theoretical fragmentation masses of the peptide AAVLWQLNGTK in the absence of glycosylation and de-glycosylated by PNGase F in ^{18}O

Masses of the peptide are given in the table where a +1 Da shift is present in the DeGly columns denoting de-glycosylated in the presence of ^{18}O . Ion series are described by b-ions which fragment from the C terminus and y-ions that fragment from the N-terminus. The Asn residue with that may be de-glycosylated is highlighted in yellow

Amino Acid	b-ion	y-ion	b-ion (DeGly- ^{18}O)	y-ion (DeGly- ^{18}O)
A	72.044	1200.674	72.044	1203.674
A	143.082	1129.637	143.082	1132.637
V	242.151	1058.6	242.151	1061.6
L	255.235	959.531	255.235	962.531
W	541.314	846.447	541.314	849.447
Q	669.372	660.368	669.372	663.368
L	782.457	532.31	782.457	535.31
N	896.499	419.225	899.499	422.225
G	953.521	305.183	956.521	305.183
T	1054.569	248.161	1057.569	248.161
K	1182.664	147.113	1185.664	147.113

Figure 2-7: MS2 Spectra of a representative de-glycosylated peptide digested by PNGase F in ^{18}O

MS2 spectra extracted from a mass spectrometer post CID with of the de-glycosylated peptide "AAVLWQLNGTK" with a charge state of +1. Intensity units are based by the number of counts or instances a peptide is measured. Fragmentation pattern matches the theoretical masses found in Table 2-3 of a de-glycosylated peptide using PNGase F in the presence of ^{18}O . The peak corresponding to the y5 ion is marked in (*)

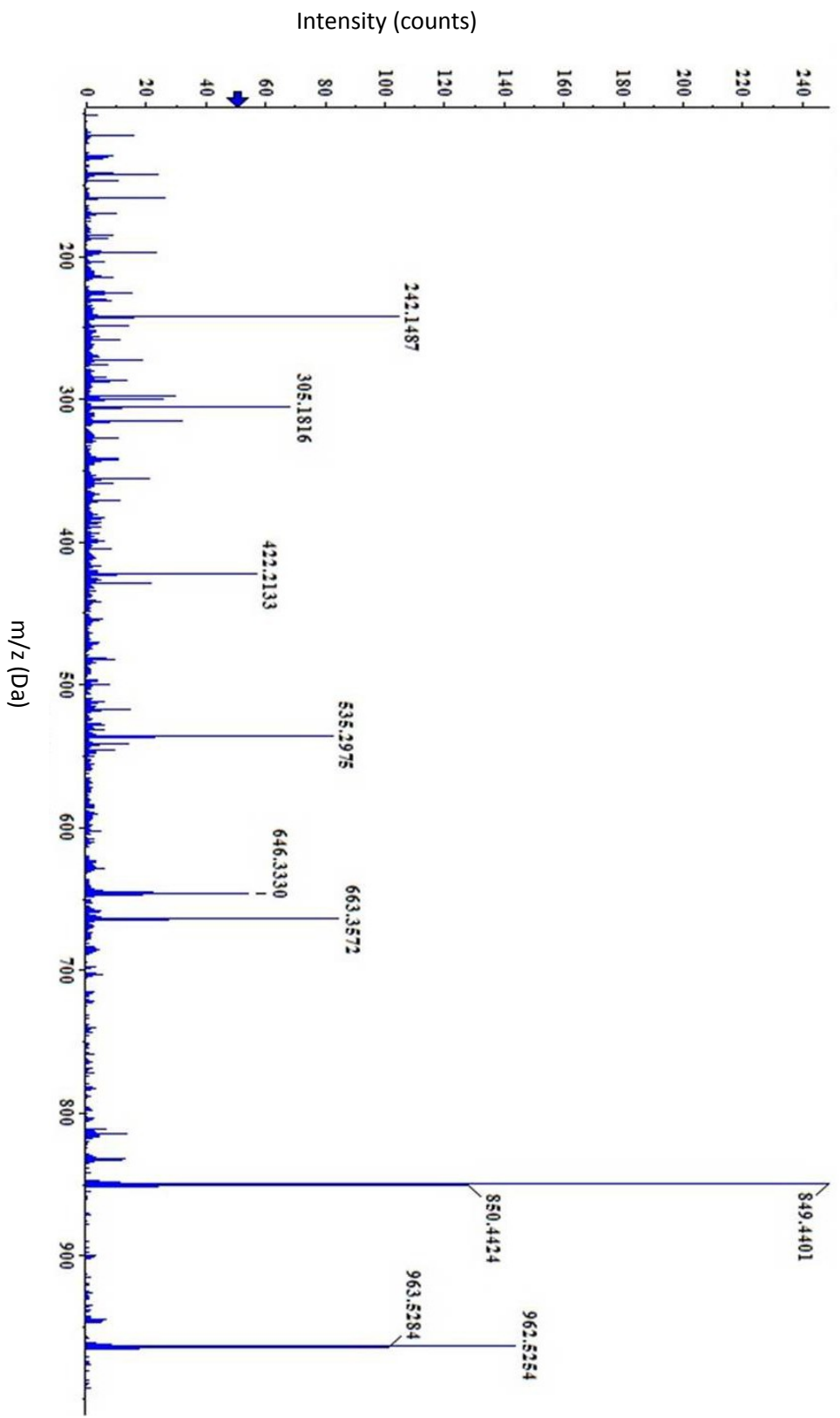


Figure 2-8: Overlaid MS2 Spectra of a representative de-glycosylated peptide digested by PNGase F in ^{18}O and ^{16}O

MS2 spectra extracted from a mass spectrometer post CID with of the de-glycosylated peptide "AAVLWQLNGTK" with a charge state of +1. Intensity units are based by the number of counts or instances a peptide is measured. Peaks have been centroided as peaks based on total area to allow for comparison between spectra. Peaks containing +2Da shift between each other are representative of the same peak depending on digestion by PNGase F in the presence of either ^{18}O (heavier) or ^{16}O (lighter). Presence of both peaks verifies the occupancy of a glycosylation site.

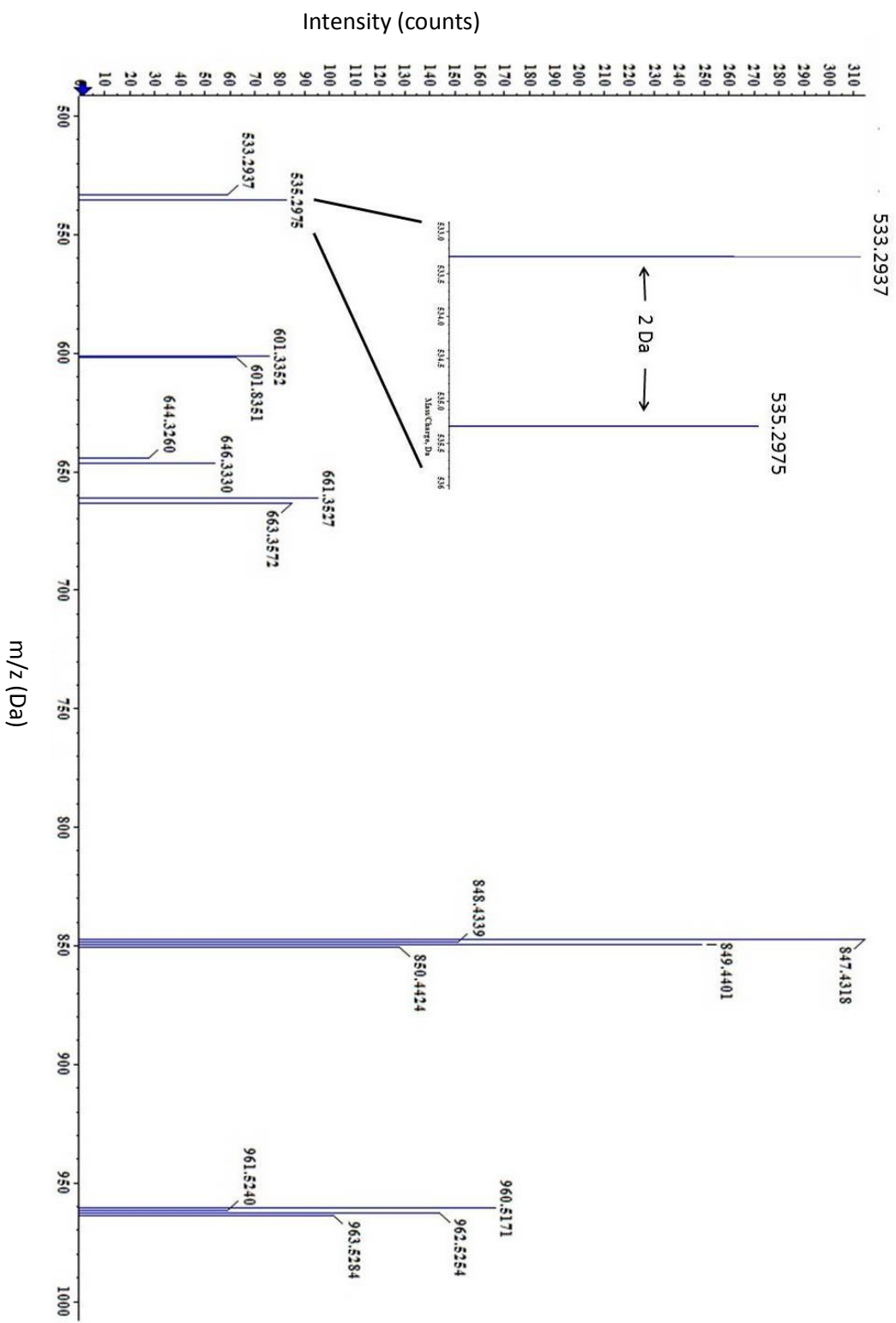
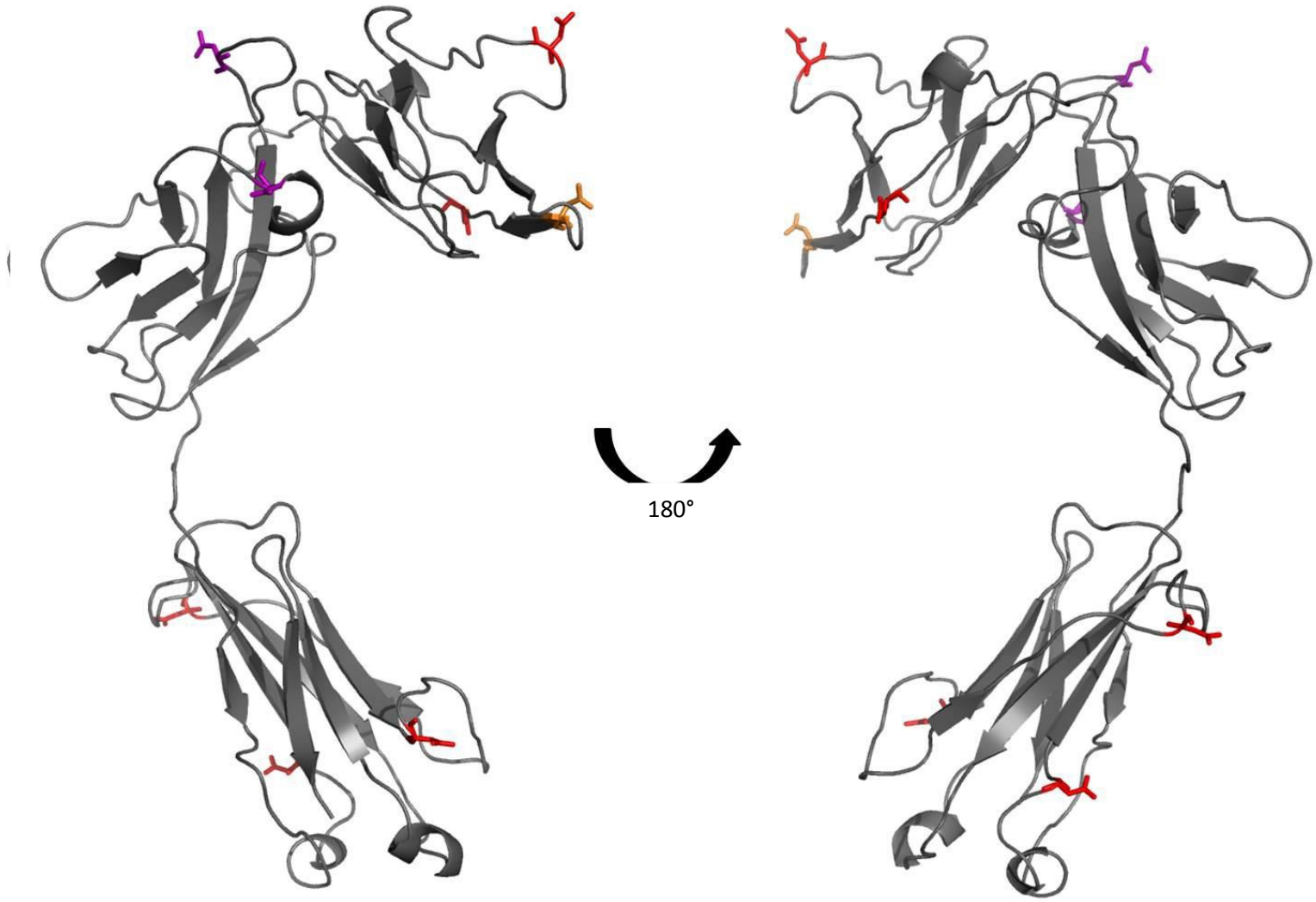


Figure 2-9: Mapping of newly discovered and verification of previously identified N-linked glycosylation sites

N-linked glycosylation sites mapped onto the structure of the sST2 receptor where red colored side chains represent newly verified glycosylation sites, purple colored side chains represent verification of previously identified glycosylation sites, and orange colored side chains represent sites not able to be verified due to poor coverage



533.2920 Da of the normal abundant water (Table 2-2). These same spectra are overlaid for comparison in Figure 2-8. This same analysis was completed for all possible sites, and contrasting the published structure of IL-33 in complex with sST2, analysis of the glycosylation sites by this method revealed that instead of only 3 sites being occupied, 7 of 8 sites predicted through the consensus sequence at Asn54, Asn101, Asn140, Asn191, Asn232, Asn254, and Asn273 showed mass shifts consistent with the presence of N-linked glycosylations (27, 29, 30).

Characterization of Glycosylation State using Endoglycosidases

While determination of glycan occupancy at a given site is relatively straightforward, determination of a glycan structure at a specific residue is much more challenging. Typical approaches are focused on one of two methods: one in which a single glycan attached to a specific peptide is released and analyzed separately from the peptide by mass spectrometry or a method in which the entire glycosylated peptide is analyzed as one bulk sample. When isolating specific peptides post-digest, there are several challenges due to the nature of the mass-spectrometry methodology. These challenges lie in the difficulty in isolating a specific peptide with a glycan attached due to the biochemical properties of the peptide which may or may not be altered by the glycan present at this site as well as possible glycan heterogeneity and variability that may prevent binding or isolation. This challenge is typically overcome using broad spectrum glycan binding proteins called lectins to separate glycosylated peptides from non-glycosylated peptides followed by elution and further purification to achieve homogeneity of the peptide sample(31–33). Following purification, glycans are released and sequenced either through a CID based experiment or sequential

digestion with exoglycosidases that have high specificity for cleaving specific sugar-sugar bonds, allowing a greater degree of confidence in the determination of the glycan structure. The second method, though more efficient, is not without its own challenges. Because sugar-sugar bonds are weaker than peptide bonds, glycans degrade in a sequence specific manner from the terminal sugar to the core $\text{Man}\alpha 1-6(\text{Man}\alpha 1-3)\text{Man}\beta 1-4\text{GlcNAc}\beta 1-4\text{GlcNAc}\beta 1-\text{Asn-X-Ser/Thr}$ resulting in the final de-glycosylated peptide. This method allows for possible sequencing of complex samples through mass spectrometry; however, this method becomes quickly convoluted in the event of heterogeneous glycoforms attached to specific peptides and the statistical methods used to match fragmentation maps to glycan structures and peptides. One way to amend this is through the tandem use of two collision sources of CID and electron transfer dissociation (ETD) to achieve specific fragmentation of the glycans on a given peptide and then subsequent fragmentation of the peptide portion. Due to of the high degree of variability present, identification of the glycoform by this method becomes statistically challenging due to a large number of possible glycoforms to match to and highly convoluted spectra without previous knowledge of the glycoform present (34, 35).

To address the issues inherent in the established methodologies for determining glycan structure and to account for the inherent heterogeneity of glycosylated protein samples, a bulk approximation approach of the glycosylation sites and branching structures using endoglycosidases was utilized. Endoglycosidases (which include PNGase F) are able to cleave at core glycan residues rather than cleaving from the terminal end and proceeding inward in the manner that exoglycosidases are limited to. Traditionally, endoglycosidase H (Endo H) isolated from *Streptomyces plicatus* is used to help identify O-linked glycosylations.

Because Endo H is not limited to N-linked glycans, Endo H is able to cleave O-linked glycans between the two GlcNAc residues of the N,N-diacetylchitobiose core; however, the specificity of Endo H is limited to oligomannose and hybrid type glycans due to the required 5 mannose branching recognition structure. The use of a variety of endoglycosidases with varying specificity including Endo H combined with PNGase F and three homologous endoglycosidases Endo F1, Endo F2, and Endo F3 which each have their own glycan recognition specificity, it is possible to analyze the bulk glycosylation state of N-linked glycosylations in ST2 (36).

Although traditionally used to merely release glycans from proteins, the use of a suite of endoglycosidases in parallel knowing their individual specificity elements allows for the structural determination of the bulk glycosylation state of a protein and a novel analytical method for determining the structure. PNGase F is able to cleave all N-linked glycans regardless of the branching structure or type of glycan present while Endo H only able to cleave oligomannose and hybrid type glycanse. Three additional homologs of Endo H: Endo F1, Endo F2, and Endo F3, each have their own specific recognition site. Endo F1, much like Endo H, recognizes the 5 mannose branching structure and therefore cleaves oligomannose and hybrid type glycans though has the additional benefit being able to recognize sulfated structures. Endo F2 and Endo F3 represent a departure from the recognition of the 5 mannose branching structure required by Endo H and Endo F1 in that they are uniquely able to recognize complex glycans. Endo F2 requires the minimum possible branching structure present in only requiring 3 mannose branching structure allowing it to cleave off bi-antennary complex glycans. Endo F3 also recognizes the 3 mannose branching structure but also contains a larger glycan binding pocket than Endo F2

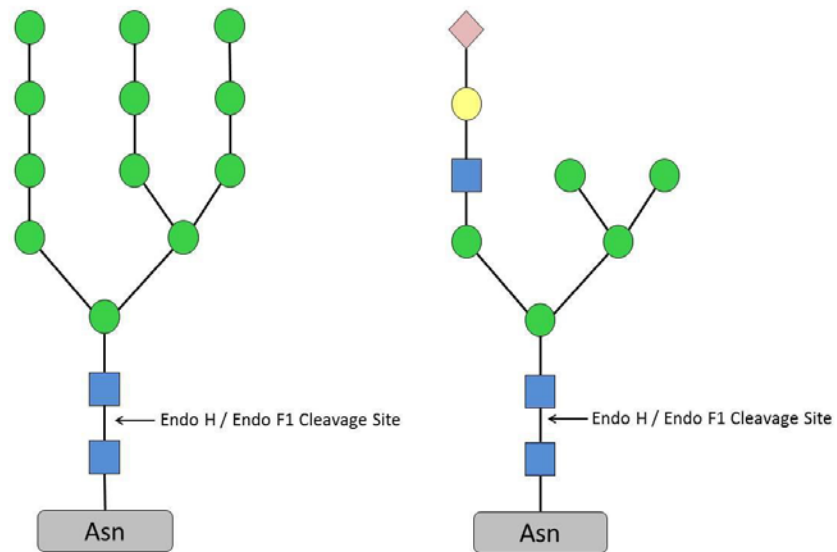


Figure 2-10: Endo H and Endo F1 substrate recognition and cleavage site

Endo H and Endo F1 both have similar substrate recognition sites. Both of these endoglycosidases are able to cleave off oligomannose and hybrid type glycans due to the recognition of the 5 mannose structure present in both. Unique to Endo F1 is the ability to cleave off sulfated residues as well. Both enzymes cleave between the two GlcNAc residues that constitute the N,N-diacetylchitobiose core

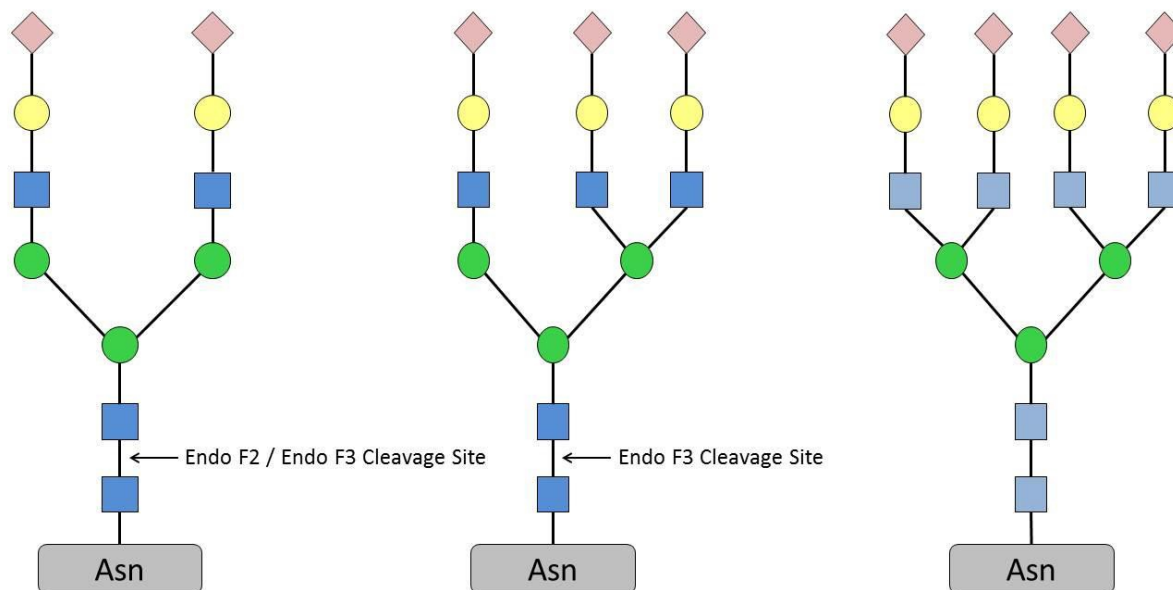


Figure 2-11: Endo F2 and F3 substrate recognition and cleavage site

Endo F2 and Endo F3 have preference for bi- and tri- antennary branched complex glycans. While Endo F2 is only able to recognize bi- antennary branched glycans, Endo F3 can cleave both bi- and tri- antennary glycans. Both cleave between the core GlcNAc residues that form the N,N-diacetylchitobiose core leaving one GlcNAc present. Neither is able to cleave any glycan with branching greater than tri-antennary.

allowing it to cleave of tri-antennary structures. Any larger branching structure is unable to be recognized by any of endoglycosidases other than PNGase F (36, 37).

Using the ability for endoglycosidases to recognize specific structures, a generalized chart can be utilized that determines glycosylations present in a bulk manner. In the case of sST2, samples are already known to be susceptible to PNGase F cleavage; however, these same samples are not able to be digested using Endo H. Upon incubation with Endo H, Endo F1, Endo F2, and Endo F3, it is shown that the only endoglycosidase that has activity is Endo F3. Despite the fact that Endo F3 reacts with the sample, blotting of the digested sample shows heavy amounts of smearing as well as incomplete digestion to the molecular weight of a PNGase F digested sample suggesting that although there are tri-antennary structures present, there are also branching structure greater than tri-antennary present in the sample suggesting heterogeneity of the glycoform distribution. Based on these results, we show for this first time that the sST2 sample contains primarily tetra- or greater antennary complex glycans with some tri-antennary complex glycans present at some sites in some molecules.

2.4 Discussion

Glycosylations Present

The results presented here show that the sST2 receptor exists in glycoforms different from previously published. Contrasting the current model, the results show that sST2 is highly glycosylated with high branching structures and occupancy of nearly all predicted N-linked glycosylation sites. By combining site specific identification through mass spectrometry and a western blot/enzymatic assay for sequencing based on endoglycosidase specificity, the methods utilized allow for the determination of bulk glycosylation and site

Table 2-4: Table of substrate recognition of endoglycosidases PNGase F, Endo H, Endo F1, Endo F2, and Endo F3

While PNGase F allows for the determination of any N-linked glycosylations present, the Endo H, F1, F2, and F3 group allow for the determination of the structure of glycans present. Substrate recognition and cleavage efficiency can be tracked using a gel-mobility shift assay.

Glycan	PNGase F	Endo H	Endo F1 (includes sulfated)	Endo F2	Endo F3
	✓	✓	✓		
	✓	✓	✓		
	✓			✓	✓
	✓				✓
	✓				

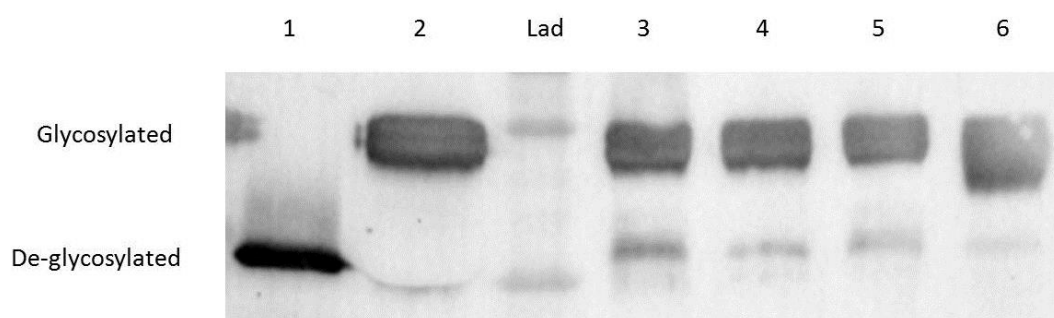


Figure 2-12: Gel mobility shift of endoglycosidase digestion of sST2 probed by anti-ST2 western blotting

Samples of sST2 were denatured and then incubated with either PNGase F (Lane 1), Endo H (Lane 2 and 3), Endo F1 (Lane 4), Endo F2 (Lane 5), and Endo F3 (Lane 6). Cleavage of PNGase F denotes the presence of N-linked glycans and in the case of sST2, shows as an observed shift of ~30kDa. Samples incubated with Endo F3 were the only samples that showed shifts due to loss of glycans. Smearing of the band suggests partial occupancy across several sites while retention of the top most bound of the lane suggests that a population of entirely tetra- or greater antennary complex glycans exists.

specific occupancy of sites in a manner more time-efficient than purely mass spectrometry based ones. Likewise, flexibility of conditions that endoglycosidases are active under (detergent denaturing conditions in particular) allow for highly flexible application of this method compared to mass spectrometry based ones that would either require strong statistical methods coupled with a large database to predict the proper glycoform at a site or pre-existing knowledge of the glycoform by lectin binding.

Biological Implications of Glycosylations

The detection of a wide range of glycosylations present with higher branching structures represents a departure from the traditional model involved with IL-33/ST2 signaling and can have a wide range of implications upon the regulation of the system (24, 38). The presence of sST2 in the system exists as an important regulator of IL-33 levels in the body which can mitigate excessive expression of IL-33 inducing pro-inflammatory effects in the lungs and gut. Both of these areas are well known to contain a mucus layer implying potential interactions between mucins and members of the IL-33/ST2 system, though only the gut has been proven so far (39, 40). In this manner, highly glycosylated molecules are able to interact with the glycosylations of mucins and sequester to the mucus layer of the heart and lung. By doing so, the effective concentration and potential pharmacokinetics of sST2 is increased dramatically and act like a filter to increase the threshold required for IL-33 dependent signaling (10, 41). As such, the number of glycosylations as well as mutations in the glycosylation profile of sST2 could be an important underlying aspect of disease states where excessive IL-33 expression causes dysregulation.

Although the role of glycosylations in systems is highly variable and largely unknown, one of the most well classified effects of glycosylations has been shown in the

case of antibodies where glycosylations increase the protein stability and half-life in serum as well as protein dynamics. Previous work into the effects of N-linked glycosylations have shown that the presence of N-linked glycosylations reduce protein dynamics through the stabilization of native contacts through water interactions as well as steric hindrance of energetically unfavorable conformations or unfolding events. In this manner, it is possible for sST2 to become stabilized through the presence of N-linked glycosylations to further regulate IL-33 levels in the body (42, 43).

The presence of glycosylations in proteins has many strong implications in folding, trafficking, and secretion. Due to the unique environment present within the endoplasmic reticulum, many proteins are expressed within the lumen. As such, glycosylation of these proteins plays an important role in the folding route and has been shown that by stabilizing interactions with water, they can stabilize hydrophobic interactions and facilitate creation of hydrophobic cores. Structure of the glycans attached to proteins also hold particular importance in regard to protein trafficking both intracellularly and extracellularly through the direction of proteins to vesicles or to exosomes for secretion or display. As such, glycosylations act as tags for protein direction and can even have implications on receptor recycling and display times. Through these methods, it is possible that dysregulation of the IL-33/ST2 pathway occurs through alterations of the glycosylation state of either the sST2 levels or the glycosylation state of the insoluble receptor ST2. In the event of a mutation in the glycosylation state that may alter the receptor half-life, protein folding, or secretion, disease states that arise through dysregulation of IL-33 can be traced to alterations to the glycoform compared to healthy models (3, 9).

2.5 Future Work

To probe the patient variability and compare glycosylation states with disease states, analysis of derived samples is necessary. Since the underlying guidelines behind glycosylation and what small changes lead to perturbations in glycosylation state is still unknown, it is necessary to observe the glycosylation state of protein produced using both patient cells and corresponding patient serum in order to avoid unexpected perturbations to the glycosylation pathways. Likewise, utilization of patient samples opens the possibility of analyzing the differences in glycosylation state between sST2 and the insoluble receptor form responsible for signaling, ST2. Although not directly linked yet, the IL-33/ST2 axis is involved in many highly glycosylation implicated systems. Of these systems, the implications with Crohn's disease where much of the sensitivity to stimuli can be attributed due to dysregulation of the glycans lining the gut and asthma where airway mucins have been shown to be important markers and regulators for the system. It is highly suspected that the glycosylation state of the sST2 receptor, in particular, functions in terms of its ability to localize to the mucosal fraction where it can be localized to the area that IL-33 is expressed as well as help to increase the serum half-life. By doing so, the pharmacokinetic properties of sST2 are vastly increased and create an additional method by which the system is able to exert control over IL-33 signaling. When looking at the membrane bound form of ST2, it is highly possible that the glycosylation state has great importance in the receptor half-life in vivo as well as possible localization signal to be displayed on the membrane. Glycosylation states have been heavily tied to the ability for proteins to be either secreted or displayed on the surface of the cell membrane and dysregulation of this process in ST2 can possibly lead to reduced inhibition of IL-33 or enhanced signaling through IL-33 (39–41).

To further probe the effects of N-linked glycosylation, the isolation of the last possible N-linked glycosylation site through digestion with a different protease is necessary. This site holds particular importance due to its presence in the current crystal structure. To achieve coverage of this site, a different protease must be utilized to gain coverage of the peptide fragment. Although there is no guarantee for peptide ionization efficiency, there are several possible choices to attain coverage of the site. The most likely proteases to utilize for this would be LysC and LysN. These protease are very similar to trypsin in that they have high thermal stability and highly compatible with denaturant. Cut-sites for these proteases have high fidelity and cleave on the N or C end of lysine residues and impart a partial charge in the same way that trypsin does. It is expected that use of these protease may achieve coverage over this site.

In order to probe the site specific effects of many of these glycoforms on ST2 *in vitro*, site directed mutagenesis is required at the sites identified. By modifying the consensus sequence of Asn – X – Ser/Thr, it is possible to artificially remove the glycosylation without adversely perturbing the overall structure. Since previous studies have shown that the glycosylation state is not required for proper folding of the final structure (24), it is possible that the glycosylation may have an effect in the heterotrimeric complex with glycosylated IL-1RAcP. With site specific data on the occupancy of glycosylations at a specific site, the ability to probe the effects of glycosylations on the system become much more discrete and can be done both *in vivo* and *in silico*.

To determine the full extent of glycosylation upon the system, further analysis into the glycosylation state of IL-1RAcP is required. Although previously shown to be glycosylated, IL-1RAcP holds three additional possible N-linked glycosylation sites

determined through consensus sequence(23). The effects of the occupancy of these sites upon receptor complex binding may lead to additional insight into how the system is regulated *in vivo*.

2.6 References

1. Zhang Y, Reinberg D (2001) Transcription regulation by histone methylation: Interplay between different covalent modifications of the core histone tails. *Genes Dev* 15(18):2343–2360.
2. Hochstrasser M (1996) Ubiquitin-dependent protein degradation. *Annu Rev Genet* 30:405–439.
3. Ferguson MA, Kinoshita T, Hart GW (2009) *Essentials of Glycobiology. 2nd edition.*
4. Wong C-H (2005) Protein glycosylation: new challenges and opportunities. *J Org Chem* 70(11):4219–4225.
5. Simons K, Vaz WLC (2004) Model systems, lipid rafts, and cell membranes. *Annu Rev Biophys Biomol Struct* 33:269–295.
6. Baskin JM, Prescher JA, Laughlin ST, Agard NJ, Chang P V, Miller IA, Lo A, Codelli JA, Bertozzi CR (2007) Copper-free click chemistry for dynamic *in vivo* imaging. *Proc Natl Acad Sci U S A* 104(43):16793–16797.
7. Reitsma S, Slaaf DW, Vink H, Van Zandvoort MAMJ, Oude Egbrink MGA (2007) The endothelial glycocalyx: Composition, functions, and visualization. *Pflugers Arch Eur J Physiol* 454(3):345–359.
8. Tao SC, Li Y, Zhou J, Qian J, Schnaar RL, Zhang Y, Goldstein IJ, Zhu H, Schneck JP (2008) Lectin microarrays identify cell-specific and functionally significant cell surface glycan markers. *Glycobiology* 18(10):761–769.
9. Helenius A, Aebi M (2001) Intracellular functions of N-linked glycans. *Science* 291(5512):2364–2369.
10. Peracaula R, Barrabés S, Sarrats A, Rudd PM, de Llorens R (2008) Altered glycosylation in tumours focused to cancer diagnosis. *Dis Markers* 25(4-5):207–218.
11. Tuccillo FM, De Laurentiis A, Palmieri C, Fiume G, Bonelli P, Borrelli A, Tassone P, Scala I, Buonaguro FM, Quinto I, Scala G (2014) Aberrant glycosylation as biomarker for cancer: Focus on CD43. *Biomed Res Int* 2014.

12. Kornfeld R, Kornfeld S (1985) Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem* 54:631–664.
13. Schachter H (2000) The joys of HexNAc. The synthesis and function of N- and O-glycan branches. *Glycoconj J* 17(7-9):465–483.
14. Raju TS, Briggs JB, Borge SM, Jones AJ (2000) Species-specific variation in glycosylation of IgG: evidence for the species-specific sialylation and branch-specific galactosylation and importance for engineering recombinant glycoprotein therapeutics. *Glycobiology* 10(5):477–486.
15. Muchmore EA, Diaz S, Varki A (1998) A structural difference between the cell surfaces of humans and the great apes. *Am J Phys Anthropol* 107(2):187–198.
16. Senger RS, Karim MN (2005) Variable site-occupancy classification of N-linked glycosylation using artificial neural networks. *Biotechnol Prog* 21(6):1653–1662.
17. Tarentino AL, Plummer TH (1994) Enzymatic deglycosylation of asparagine-linked glycans: purification, properties, and specificity of oligosaccharide-cleaving enzymes from *Flavobacterium meningosepticum*. *Methods Enzymol* 230:44–57.
18. Tretter V, Altmann F, März L (1991) Peptide-N4-(N-acetyl-beta-glucosaminy)asparagine amidase F cannot release glycans with fucose attached alpha 1----3 to the asparagine-linked N-acetylglucosamine residue. *Eur J Biochem* 199(3):647–652.
19. Plummer TH, Elder JH, Alexander S, Phelan AW, Tarentino AL (1984) Demonstration of peptide:N-glycosidase F activity in endo-beta-N-acetylglucosaminidase F preparations. *J Biol Chem* 259(17):10700–10704.
20. Zhang JD, Cousens LS, Barr PJ, Sprang SR (1991) Three-dimensional structure of human basic fibroblast growth factor, a structural homolog of interleukin 1 beta. *Proc Natl Acad Sci U S A* 88(8):3446–3450.
21. Liu X, Hammel M, He Y, Tainer JA, Jeng U-S, Zhang L, Wang S, Wang X (2013) Structural insights into the interaction of IL-33 with its receptors. *Proc Natl Acad Sci U S A* 110(37):14918–23. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3773798&tool=pmcentrez&rendertype=abstract>.
22. Brown A, Robinson CJ, Gallagher JT, Blundell TL (2013) Cooperative heparin-mediated oligomerization of fibroblast growth factor-1 (FGF1) precedes recruitment of FGFR2 to ternary complexes. *Biophys J* 104(8):1720–1730.
23. Thomas C, Bazan JF, Garcia KC (2012) Structure of the activating IL-1 receptor signaling complex. *Nat Struct Mol Biol* 19(4):455–457.

24. Lingel A, Weiss TM, Niebuhr M, Pan B, Appleton BA, Wiesmann C, Bazan JF, Fairbrother WJ (2009) Structure of IL-33 and Its Interaction with the ST2 and IL-1RAcP Receptors-Insight into Heterotrimeric IL-1 Signaling Complexes. *Structure* 17(10):1398–1410.
25. Shi X, Jarvis DL (2007) Protein N-glycosylation in the baculovirus-insect cell system. *Curr Drug Targets* 8(10):1116–1125.
26. Viitala J, Carlsson SR, Siebert PD, Fukuda M (1988) Molecular cloning of cDNAs encoding lamp A, a human lysosomal membrane glycoprotein with apparent Mr approximately equal to 120,000. *Proc Natl Acad Sci U S A* 85(11):3743–3747.
27. Morelle W, Michalski J-C (2007) Analysis of protein glycosylation by mass spectrometry. *Nat Protoc* 2(7):1585–1602.
28. Shevchenko A, Wilm M, Vorm O, Mann M (1996) Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal Chem* 68(5):850–858.
29. Angel PM, Lim JM, Wells L, Bergmann C, Orlando R (2007) A potential pitfall in 18O-based N-linked glycosylation site mapping. *Rapid Commun Mass Spectrom* 21(5):674–682.
30. Palmisano G, Melo-Braga MN, Engholm-Keller K, Parker BL, Larsen MR (2012) Chemical deamidation: A common pitfall in large-scale N-Linked glycoproteomic mass spectrometry-based analyses. *Journal of Proteome Research*, pp 1949–1957.
31. Zhou Y, Aebersold R, Zhang H (2007) Isolation of N-linked glycopeptides from plasma. *Anal Chem* 79(15):5826–5837.
32. Tian Y, Zhou Y, Elliott S, Aebersold R, Zhang H (2007) Solid-phase extraction of N-linked glycopeptides. *Nat Protoc* 2(2):334–339.
33. Chen J, Shah P, Zhang H (2013) Solid phase extraction of N-linked glycopeptides using hydrazide tip. *Anal Chem* 85(22):10670–10674.
34. Ashline DJ, Lapadula AJ, Liu YH, Lin M, Grace M, Pramanik B, Reinhold VN (2007) Carbohydrate structural isomers analyzed by sequential mass spectrometry. *Anal Chem* 79(10):3830–3842.
35. Mechref Y (2012) Use of CID/ETD mass spectrometry to analyze glycopeptides. *Curr Protoc Protein Sci* 1(SUPPL.68).
36. Freeze HH, Kranz C (2010) Endoglycosidase and glycoamidase release of N-linked glycans. *Curr Protoc Protein Sci* (SUPPL.62).

37. Trimble RB, Tarentino AL (1991) Identification of distinct endoglycosidase (endo) activities in *Flavobacterium meningosepticum*: Endo F1, endo F2, and endo F3: Endo F1 and endo H hydrolyze only high mannose and hybrid glycans. *J Biol Chem* 266(3):1646–1651.
38. Kakkar R, Lee RT (2008) The IL-33/ST2 pathway: therapeutic target and novel biomarker. *Nat Rev Drug Discov* 7(10):827–840.
39. Lopetuso LR, Scaldaferrri F, Pizarro TT (2012) Emerging role of the interleukin (IL)-33/ST2 axis in gut mucosal wound healing and fibrosis. *Fibrogenesis Tissue Repair* 5(1):18.
40. Oboki K, Nakae S, Matsumoto K, Saito H (2011) IL-33 and airway inflammation. *Allergy, Asthma Immunol Res* 3(2):81–88.
41. Erle DJ, Sheppard D (2014) The cell biology of asthma. *J Cell Biol* 205(5):621–631.
42. Huhn C, Selman MHJ, Ruhaak LR, Deelder AM, Wuhrer M (2009) IgG glycosylation analysis. *Proteomics* 9(4):882–913.
43. Zheng K, Bantog C, Bayer R (2011) The impact of glycosylation on monoclonal antibody conformation and stability. *MAbs* 3(6).

**Chapter 3 : Characterization of the Effects of N-linked Glycosylations
Present on the ST2 Receptor through MD Simulation**

3.1 Introduction

Use of MD Simulations to Model Proteins

Molecular Dynamics (MD) computer simulations have helped to revolutionize how small molecules are perceived by generating information at the atomistic level of atomic position and velocities for a given molecule. This unique ability is achieved through iterative calculation of the forces enacted upon a given molecule. With the advent of stronger and more robust computational tools, most notably the central processing unit and graphical processing unit of modern computers and specifically the usage of multiple in parallel, it is now possible to compute the highly taxing equations required to model the forces on a given atom. Through the use of MD simulations, studies of single molecule protein movement and behavior at a residue or atomistic level are relatively easy to attain compared to experimental methods. While many experimental methods are limited by their observation of bulk averaged states, the recreation of a single molecule *in silico* circumvents many of these issues to observe to states of a single protein molecule (1).

Much of the ability to accurately observe the motions and behavior of molecules in MD simulations is derived through the accurate calculation of forces enacted upon a single molecule through classical Newtonian mechanics. In traditional full MD simulations, each molecule requires the balancing of electrostatic forces, Van Der Waals forces, and Potential Energy inherent to the bonds and atoms associated. For each electrostatic term, attractive or repulsive forces are defined by Coulomb's law while all Van Der Waals forces are defined by a Lennard-Jones potential. Additional energy terms are derived out of the equilibrium bond length (bonds), equilibrium bond angles (angles), and torsional rotations about a central bond (dihedrals) based on a given set of two atoms. The sum of all equations and

constants derived to reproduce the effects and behaviors of molecules is subsequently termed the force field. There are a large variety of force fields possible to use for a given simulation, with each force field calculating its terms slightly different from the next due to its specific subset of trained molecules resulting in specific advantages or disadvantages depending on the use. Despite this, the large majority of simulations can use force fields interchangeably to achieve the same result (1).

MD Simulations of Glycoproteins

Due to the challenges present in structural elucidation of glycosylated protein using traditional methods (NMR and X-ray crystallography), MD simulations are an attractive tool to probe the effects of glycosylations on structure and dynamics of proteins. In this manner, studies of glycoproteins *in silico* have yielded new tools to studying the effect of glycans on sugars ranging from protein structure by *in silico* attachment of glycans to folding studies as well as glycan specific force fields (2–5). Although an analysis of glycosylated proteins in the Protein Data Bank (PDB) have shown that the attachment of N-linked glycosylations does not adversely change the protein structure, the dynamics of proteins are significantly changed by the presence of N-linked glycosylations (6). Countering a system wide comparison of glycosylated versus de-glycosylated proteins, studies on the effects of these glycosylations in the context of gp120 have yielded data suggesting that conformation of disordered loops as well as the dynamics are significantly perturbed by the presence of glycosylation sites (7). As such, analysis of the effects of glycosylations on proteins *in silico* should be conducted on a protein to protein specific basis to account for the location of the glycosylation site.

Coarse Grained Simulations and Structure Based Models

Although MD simulations represent an enormous leap forward in the ability to observe single molecule motions, they are not without their own challenges. First and foremost, MD simulations require the use of robust and accurate force fields in order to make their predictions which are derived from experimental observations. In this way, MD simulations and experimental techniques work in a synergistic manner to simultaneously improve the strength of each method and only truly attain their full strength when done in tandem (8). The second major hurdle to MD simulations lies in the computational strength required to observe the motions of macromolecules and macromolecular complexes. As the size of a simulation grows, the number of forces that must be calculated grows exponentially due to interactions each atom present in the system. Likewise, protein dynamics and interactions (particularly protein binding events) typically occur in the long time scale range of ns to ms. Using traditional methods, it is difficult to sample time scales long enough to sample these motions without the use of a supercomputer. In these cases, the limiting factor is the rate at which vectors can be computed and without the use of a supercomputer, would not be achievable in a time efficient manner (9). Due to the high computational costs of running MD simulations on large macromolecules in a traditional manner, it is necessary to utilize other methods compared to traditional MD simulations for large complexes.

To achieve simulations on longer time scales as well as observe the motions of macromolecules and macromolecular complexes, there are several different options to reduce the computing power necessary. Perhaps the biggest leap forward for computational chemistry arrived through the use of structure based models (SBMs). Unlike completely *in silico* based methods, structure based models rely on the input of a base structure to arrive

$$\begin{aligned}
 V = & \sum_{bonds} \epsilon_r (r - r_0)^2 + \sum_{angles} \epsilon_\theta (\theta - \theta_0)^2 + \sum_{impropers / planar} \epsilon_\chi (\chi - \chi_0)^2 \\
 & + \sum_{backbone} \epsilon_{BB} F_D(\phi) + \sum_{sidechain} \epsilon_{SC} F_D(\phi) \\
 & + \sum_{contacts} \epsilon_C \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r} \right)^6 \right] + \sum_{non-contacts} \epsilon_{NC} \left(\frac{\sigma_{ij}}{r} \right)^{12}
 \end{aligned}$$

Figure 3-1: Hamiltonian of an all atom structure based models with diagram of corresponding motion

Structure based models are only defined by the equilibrium bond distance, the angles between three molecules, the improper or proper dihedral that describes the angel between two planes that four corresponding atoms fall upon, backbone angles about the central bond of four atoms, the angles of the atoms in the side chain, interactions between contacts, and non-contacts. Contacts are defined by the a 12-6 Lennard-Jones potential while non-contact terms are defined by the repulsive portion of a 12-6 Lennard Jones potential.

at in the native basin. In this way, the forces generated from the structure present are interpreted to be the lowest energy state (or at least near the lowest energy state) such that it heavily favors the interactions present in the given structure (10, 11). The second most common action to reduce the computational power, without taking away from the behavior of a simulation, lies in removing explicit water or other solvent molecules (explicit solvent) and merely changing the dielectric constant of the vacuum to resemble that of the solvent (implicit solvent). In this manner, the simulation does not calculate explicit interactions between solvent molecules and atoms of the macromolecule, but still retains the general electrostatic properties of the solvent (12). The second action to reduce computational power is to generalize the properties of the atoms themselves in what is known as coarse-graining. Though coarse-graining can be interpreted in many ways, the general principle is that the atoms or residues present in a macromolecule become generalized in some way (13–15). One way in which to coarse-grain a simulation is to remove generalize the properties of a specific residue into a bead; in this way, each residue of a protein acts like a unique bead rather than being composed of specific atoms by reducing the number of vectors calculated per residue to one rather than the number of atoms present (13, 16). A different way to coarse-grain a simulation, though still retain the number of atoms, is to remove electrostatic properties from the a simulation; by removing the effects of electrostatics, only potential energy as a result of bonds and Van Der Waal's forces are calculated. This method strongly relies on finely tuned strengths of contacts between residues as well as the assumption that the starting structure is in the lowest energy state in order to prevent unfeasible conformations from arising (17). Yet another way to observe the motions of macromolecules and macromolecular complexes while reducing the necessary

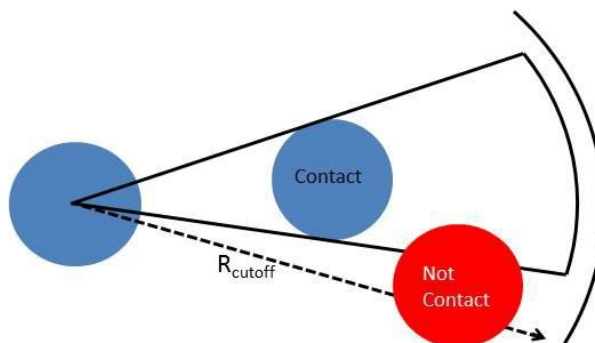


Figure 3-2: Schematic of shadow map determination of contacts versus non-contacts

Determination of contact versus non-contact is based off the application of a cutoff distance which in this case is 6 Å. Contacts are determined based on being within an a defined cutoff distance where contacts that have an intervening atom are considered to be “shadowed” and therefore not included.

computational power is to run several different simulations in parallel to sample as wide of a range of conformations as feasible. This sampling method known as umbrella sampling relies on the statistical averaging of motions that would be monitored in a bulk averaged experiment (18).

One way to achieve coarse-graining is through the use of the SMOG webtool for biomolecular simulations, which is able to process PDB files into coarse-grained files for simulation using the simulation program GROMACS. Using a Protein Data Bank input file, a structure based model (SBM) simulation can be run. Utilizing the conformation defined in the PDB file as being in the native basin, the dynamics of the native basin can be probed without having to utilize the computationally intensive method of *in silico* folding (19). To create the necessary input files for GROMACS, SMOG records the position and atoms present of all residues in the PDB file, and depending on the settings either interprets all atoms present of a residue (all-atom model) or represents the motions of each residue by the alpha-carbon of each amino acid (C-alpha model) (17, 20). To account for contacts between residues and atoms, SMOG utilizes a distance cut-off based algorithm called shadow mapping. Using shadow mapping, SMOG then records the contacts of atoms that fall within a defined distance R to approximate the effects of electrostatics and hydrophobic interactions between residues of the same protein or those present in protein-protein interactions. Once these properties are interpreted by SMOG, the program is able to output the required files for GROMACS to interpret for a MD simulation run. An updated downloadable form of SMOG named SMOG2 that utilizes open-source editing allows for addition and manipulation for a wide variety of new molecules previously unrecognized including nucleic acids and sugars (19).

3.2 Materials and Methods

In Silico Addition of Glycans and Structure Preparation

The crystal structure for the ST2 receptor bound to ligand IL33 (PDB ID 4KC3) was used as the starting structure for simulations. The published structure was crystallized in complex with the ligand IL-33 which was removed prior to addition of glycans. Missing loops were filled in and minimized based on end point distances using LOOPRED web server (21). Side chains for residues containing N-linked glycosylations at newly identified and verified sites were manually manipulated to increase solvent accessibility and decrease the possibility of steric clash upon addition of glycan chain using the sculpting tool in PyMol. Glycans were added to the manipulated structure using the GLYCAM web server with the tetraantennary glycan chosen as number 463 from the CFG glycan array version 5.0 and obtained without the presence of hydrogens in the final structure (22). Bisecting GlcNAc structures were removed to obtain a representative tetraantennary structure topped with sialic acids.

SMOG System Preparation

Using the newly generated PDB files, each glycan residue of the glycosylated form of ST2 was altered to be its own free standing residue through the addition of TER after each residue. Bonds between each glycan residue as well as disulfide bonds between cysteine residues were manually dictated numerically based on the atom number at the end of the PDB file. Manual bonding of cysteine residues to generate disulfide bridges was done to deglycosylated structure as well. Prior to processing by SMOG2, each residue of the glycosylated structure was converted from GLYCAM naming scheme into a new generalized name based on atom composition: GLC, NAG, and OSA. SMOG2 was then used for both the

glycosylated and de-glycosylated structures to generate coarse grained all atom “.gro” and “.top” files with Lennard-Jones potentials and a Shadow Map contact cut-off of 6 Å and a shadow cut-off of 1 Å determination to be run using GROMACS in a 100Å by 100Å by 100Å box. Structures were then minimized using a steepest descent energy minimization until energy fell below 100 kJ/mol/nm. The lowest energy file was then used for production run simulations (19).

Molecular Dynamics Simulation Run

Production run simulations in GROMACS 4.5.3 for 1,500,000 steps with a step size of 0.0005 femtoseconds at 95 K using a md integrator with electrostatic interactions and Van Der Waals interactions both approximated to 1.5. Temperature coupling was done using Nose-Hoover for the system and protein with a tau of 0.025. Coordinates were written to trajectory every 1000 steps. Simulations were run in parallel until end when they were concatenated into a single trajectory with the first 500,000 steps excluded to account for equilibration time using built-in GROMACS tools. Analysis of RMSF and eigenvalues of C- α were done using built-in GROMACS programs `g_rmsf`, `g_covar`, and `g_anaeig`.

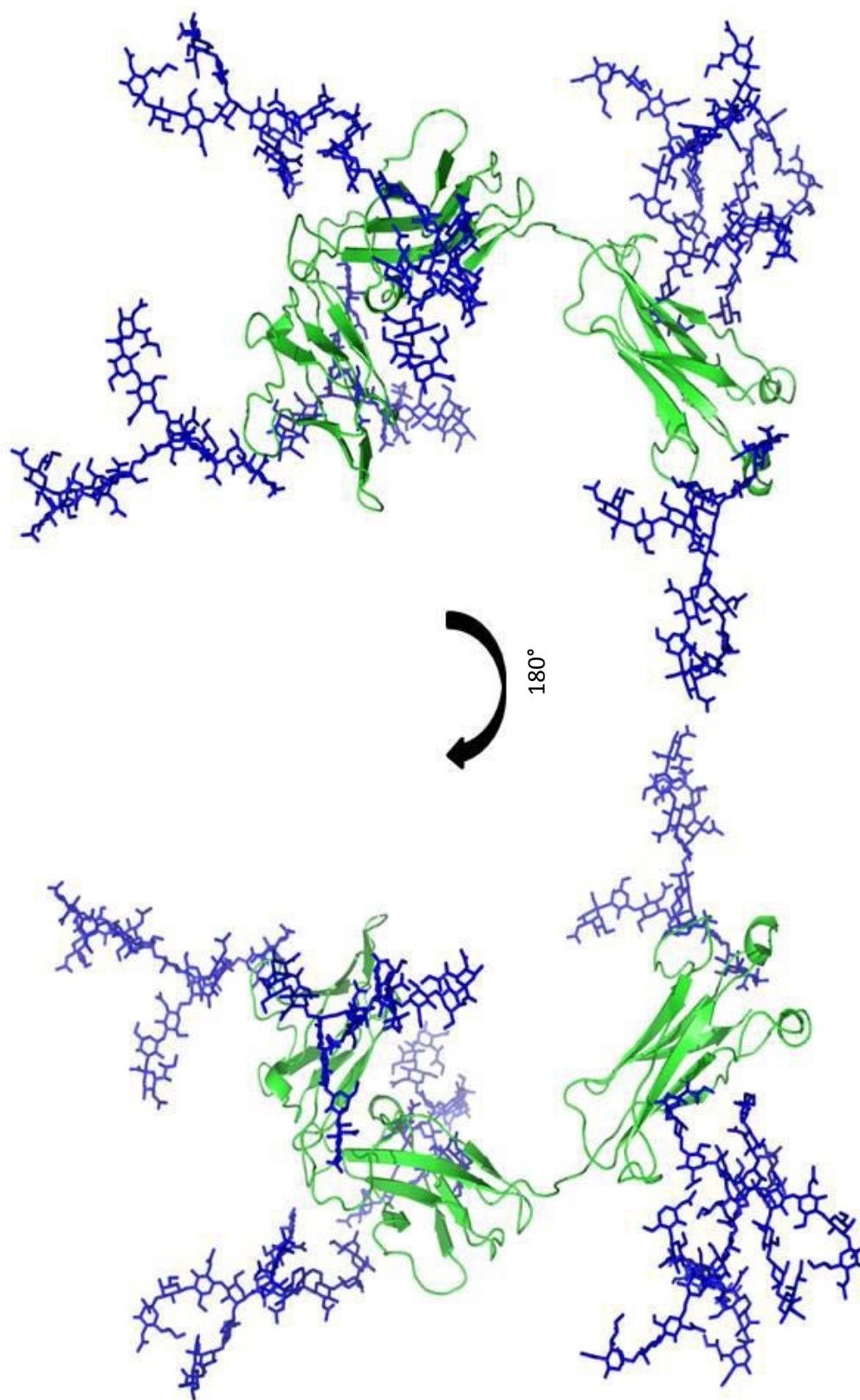
3.3 Results

Attachment of Glycosylations to sST2 Structure

Due to heterogeneity found during the digestion results, as well as the suggestion that the majority of the glycans present were not able to be de-glycosylated by Endo F3, a representative tetra-antennary complex glycan was utilized for creation of the glycosylated protein models and future studies utilizing the newly refined model. Using the GLYCAM

Figure 3-3: Structure of the sST2 receptor after loop filling and attachment of tetra-antennary glycans using the GLYCAM webserver

Missing loops in the crystal structure (4KC3) were added through the use of LOOPRED and glycans attached using the GLYCAM webserver of a representative tetra-antennary glycan with bisecting sugar removed. Structure is minimized using a tleap algorithm built into GLYCAM



web server, PDB files can be processed to allow for the attachment of large glycan structures to specified sites on the protein. These newly glycosylated structures are then energy minimized in a vacuum to generate structures for MD simulations using the tleap minimization program integrated in the GLYCAM web server (22).

SMOG2 Interpretation of Sugar Residues

In order to allow for interpretation of glycosylations by SMOG2, several additions to the program were required that were recently added onto the newest addition of SMOG, SMOG2. Unlike SMOG, SMOG2 allows for open source alterations, and through the use of ad-hoc bonding and ad-hoc dictation of improper dihedrals, is now amenable to processing a greater number of molecules. Interpretation and recognition of molecules begins in the creation of a new group of sugars titled GLYCAN that contains template files that dictate the atomic composition of residues, bonding patterns within molecules, improper dihedrals, contact energy groups, and bond energy groups. Contacts and bond groups were treated in the same manner as AMINO/AMINO contacts and bond strengths. Because SMOG2 only interprets atomic composition, position, and connectivity, the SMOG2 is stereochemically unbiased and as such does not treat molecules two molecules that are stereochemically different. In this way, mannose, glucose, and galactose which all contain the same number of atoms, bonds, and dihedrals, and were generalized as GLC residues while N-acetyl glucosamine and N-acetylgalactosamine were generalized into NAG residues. Sialic acid of Neu5Ac residues retains its GLYCAM naming scheme as OSA. Currently, the work described here as an add-on to SMOG2 contains all-atom template files to interpret mannose, galactose, glucose, GlcNAc, GalNAc, and Neu5Ac residues.

In order for SMOG2 to properly interpret PDB files that have been altered using GLYCAM, files must be prepped prior to SMOG2 interpretation. To do this, all connections between GLYCAN/GLYCAN groups and AMINO/GLYCAN must be specified ad-hoc with improper dihedrals where applicable. Likewise, all GLYCAM named residues with the exception of sialic acid (OSA) must be replaced with generalized naming consistent with SMOG2 template files created. Upon proper interpretation of molecules by SMOG2, five files are created for GROMACS use: temp.gro, temp.top, temp.ndx, shadow.contacts, shadow.log.

Dynamics of Glycosylated and De-glycosylated ST2 Unbound

Since glycans have never been previously interpreted using the methodology surrounding all-atom simulations in SMOG2 before, several short run-time simulations were done initially. These runs were evaluated for maintenance of proper bonds and conformations of glycan residues and compared to previous works in order to evaluate for proper interpretation of glycosylated sST2. Having established proper interpretation of glycans through the newly developed SMOG2 add-on, preliminary simulations were run for 1ns to gain some exploratory insight into the possible dynamic and structural perturbations that may arise due to the presence of N-linked glycosylations.

Although secondary structure showed little perturbation due to the presence of N-linked glycans, the dynamic motions had significant perturbations due to the post-translational modification. The dynamics of sST2 glycosylated and de-glycosylated were interpreted by root mean square fluctuation (RMSF) in angstroms compared to the average structure over the course of the simulation (Figure 3-4). In order to solely compare perturbations on the protein due to glycan conformation, only α -carbons of each residue in

sST2 were compared. The results of the simulation per frame were plotted versus RMSF for both glycosylated and de-glycosylated simulations where conformations with a high deviation between the average structure and the current frame are indicative of a highly dynamic state. Over the course of the simulation, the sST2 is shown to be highly dynamic and samples a wide range of conformations ranging from an “open” to a “closed” state. Simulations also uniquely show that the high flexibility of the hinge region between Ig2 and Ig3 of sST2 allow for a wide range of lateral and rotational motions previously unreported occurring. This wide range of motions manifests as a high average RMSF value of 8 Å; taken from the entire simulation. When N-linked glycosylations are added to the simulation, many of these motions are attenuated resulting in a lower RMSF than the de-glycosylated simulation. This reduction in the RMSF may be due to a combination of steric hindrance and secondary interactions between glycans that limit the degrees of freedom of the Ig3 domain.

To compare the residue specific effects of N-linked glycosylations upon protein dynamics, average RMSF values of each α -carbon over time were plotted versus residue number (Figure 3-6). This comparison elucidates domains of high stability versus domains of high dynamics. Consistent with states previously reported for ST2, sST2 showed the greatest variability in Ig3 most likely due to the presence of the flexible linker. Despite this, there are still several regions with elevated dynamics present in the first two Ig domains. Comparison of the glycosylated and de-glycosylated residue RMSF showed that although protein dynamics were attenuated, the trends per residue largely remained the

Figure 3-4: Trajectory of de-glycosylated sST2 receptor and glycosylated sST2 receptor.

Trajectory of the last 1000 frames of the simulation plotted versus RMSF in angstroms. Glycosylated trajectory is represented in green while the de-glycosylated trajectory is shown in purple. Average RMSF for each trajectory are shown in corresponding colors.

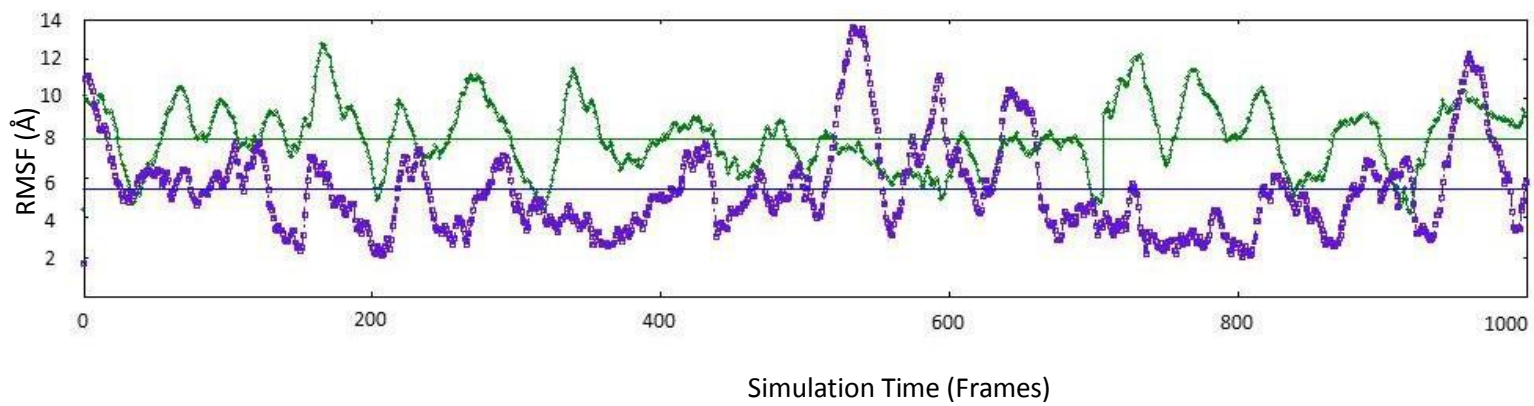


Figure 3-5: RMSF of the de-glycosylated sST2 receptor plotted for each residue along the protein

RMSF in angstroms plotted versus residue number of de-glycosylated sST2 is shown in green by itself (A), glycosylated sST2 is shown in purple by itself (B), and the two plots overlaid (C). Secondary structure elements are shown at the bottom with blue arrows representing β -sheet elements, red arrows representing α -helix elements, glycosylation sites denoted by a blue N, domain constituting Ig 1 is shown in yellow, domain constituting Ig 2 is shown in orange, and domain constituting Ig 3 is shown in blue.

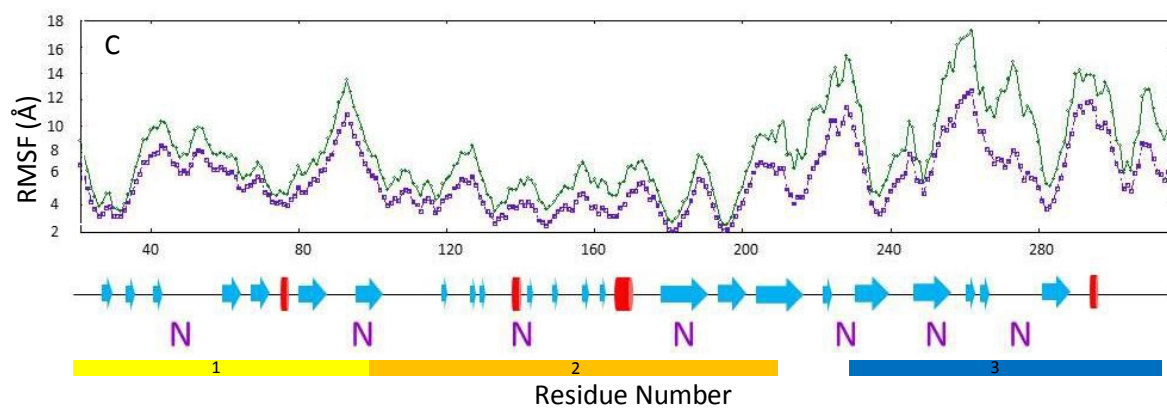
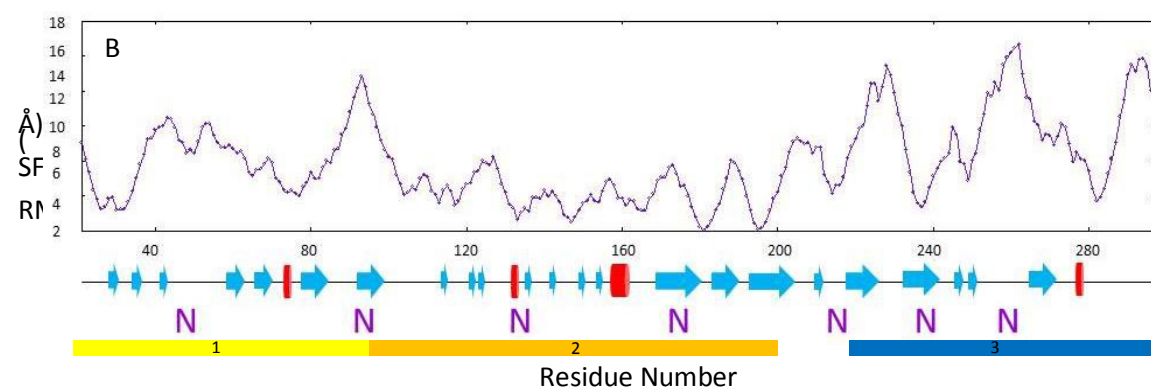
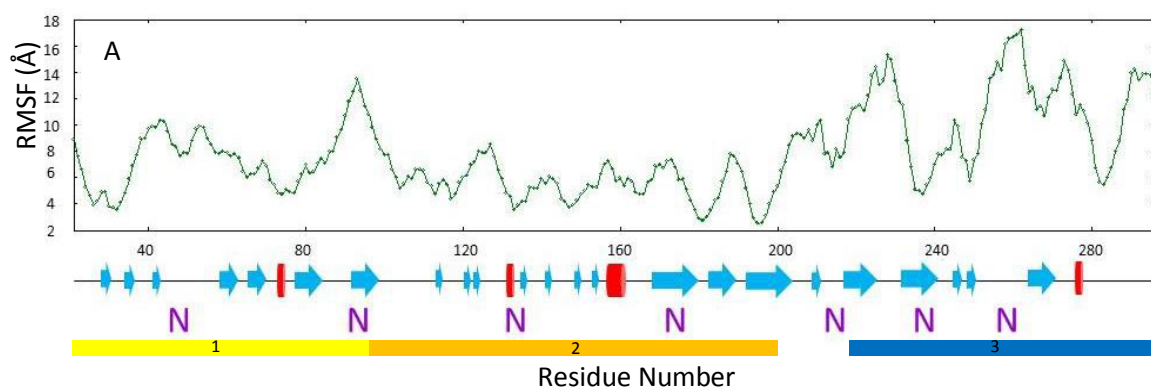


Figure 3-6: RMSF levels of the de-glycosylated simulation of sST2 plotted onto the structure of the sST2 receptor.

RMSF levels plotted onto the structure of the sST2 from the de-glycosylated simulation where red indicates levels above 6 Å, levels between 6 Å and 4 Å are shown in grey, and levels below 4 Å are shown in blue

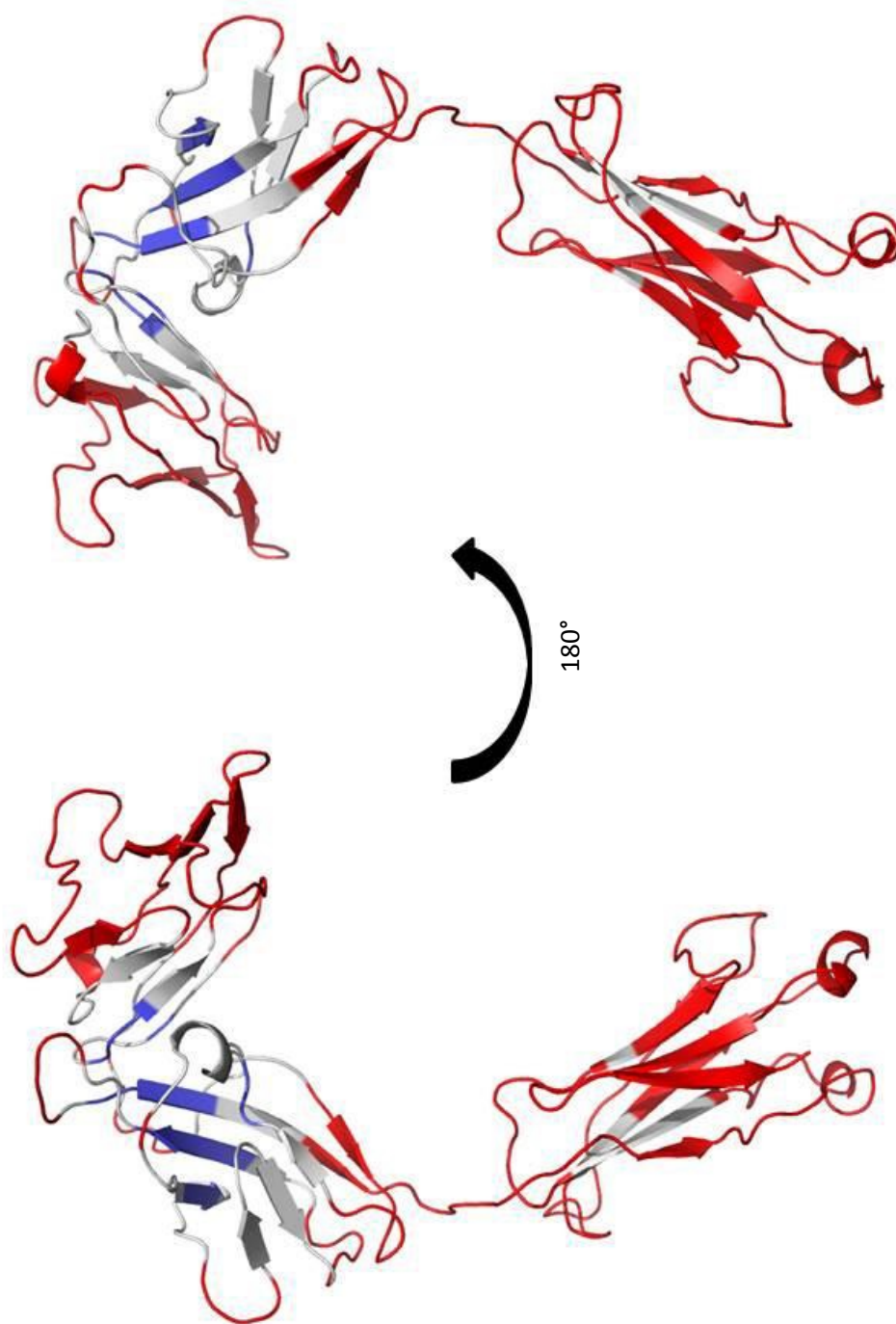


Figure 3-7: RMSF levels of the glycosylated simulation of sST2 plotted onto the structure of the sST2 receptor glycosylated.

RMSF levels plotted onto the structure of the sST2 from the de-glycosylated simulation where red indicates levels above 6 Å, levels between 6 Å and 4 Å are shown in grey, and levels below 4 Å are shown in blue

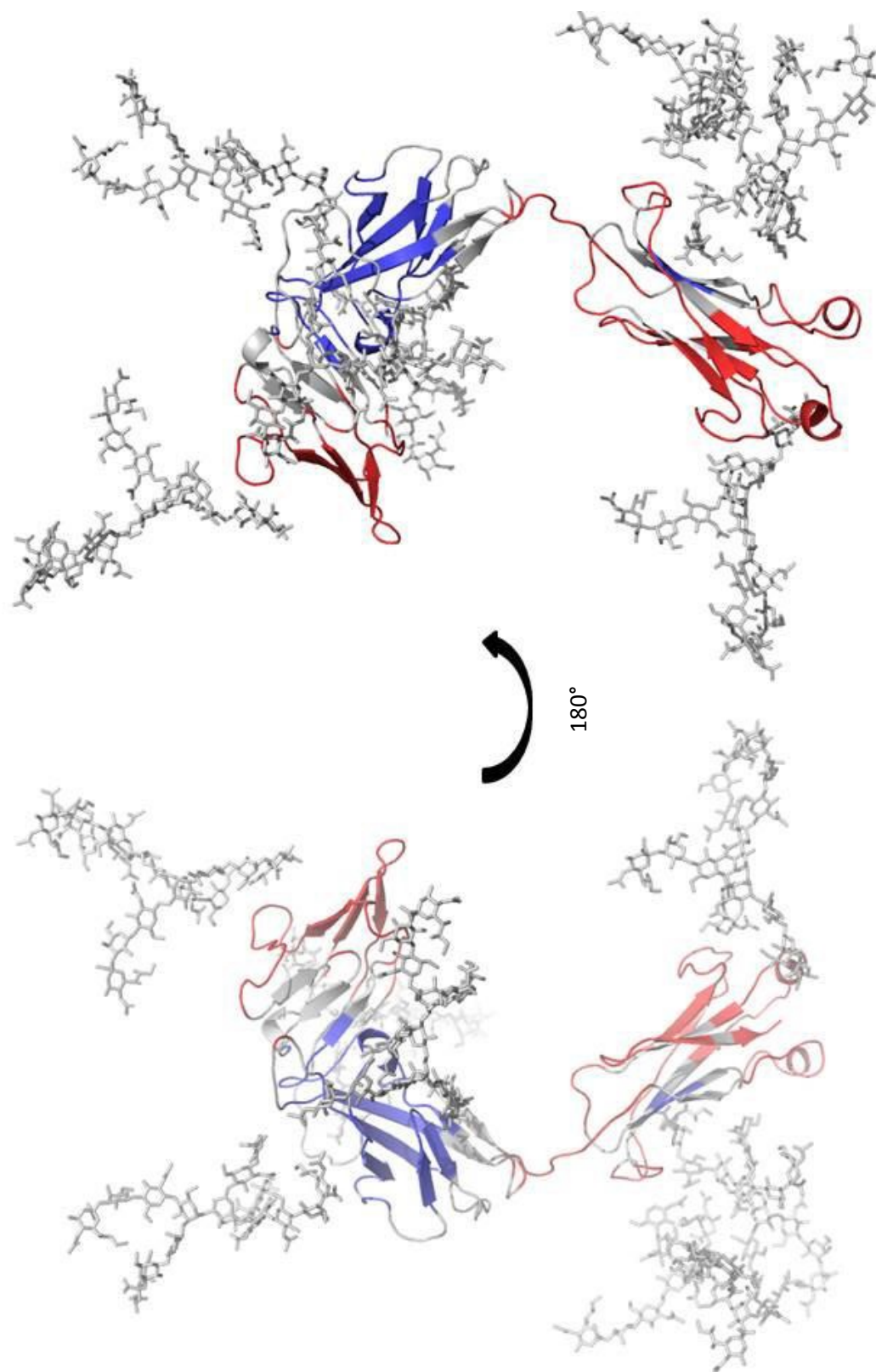
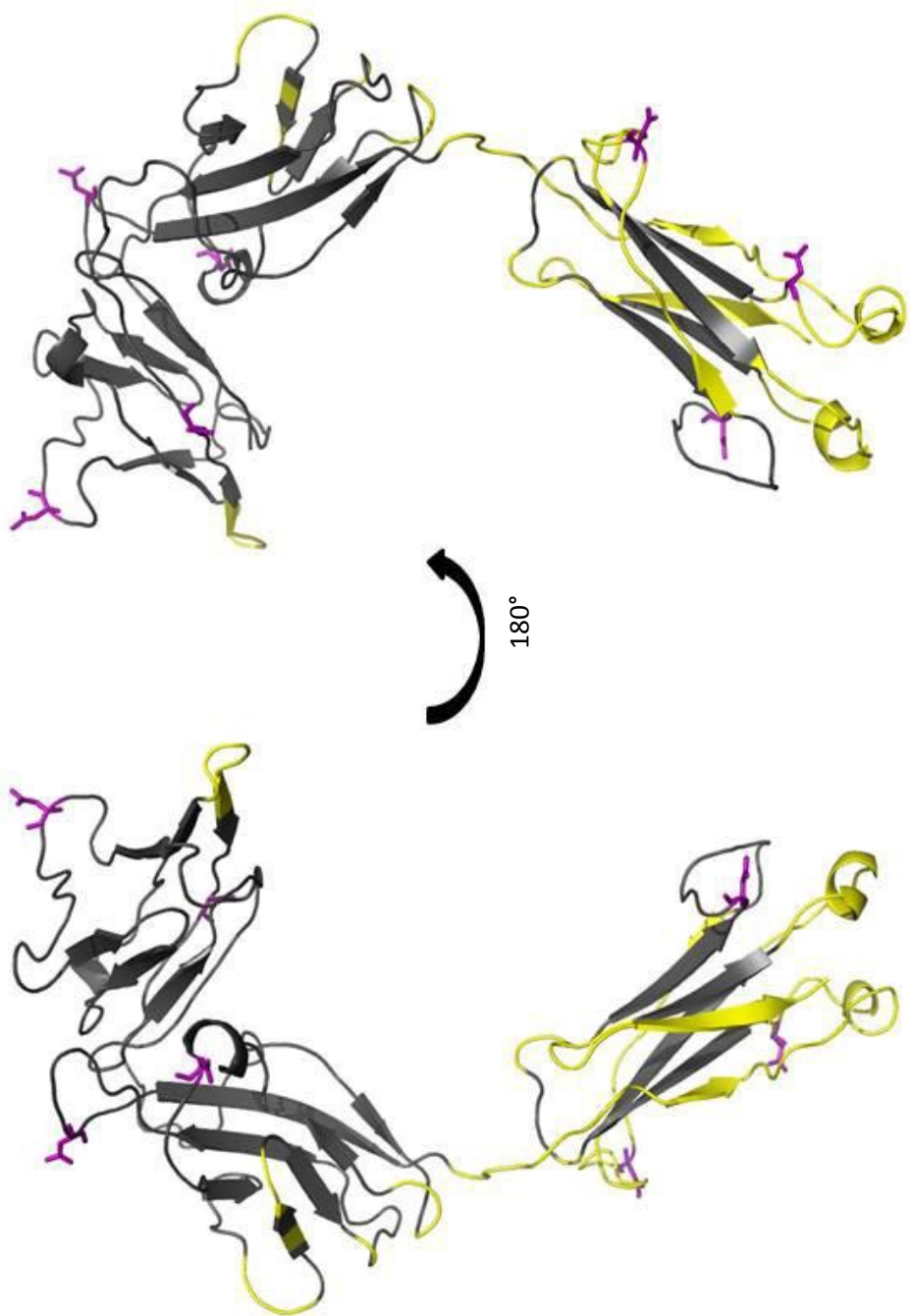


Figure 3-8: RMSF differences between de-glycosylated and glycosylated simulations plotted onto the structure of sST2

RMSF differences between the de-glycosylated and glycosylated simulation per residue of greater than 2 Å are shown in yellow where dark grey are differences less than 2 Å. Glycosylation sites are shown by the side chain they are attached to in purple.



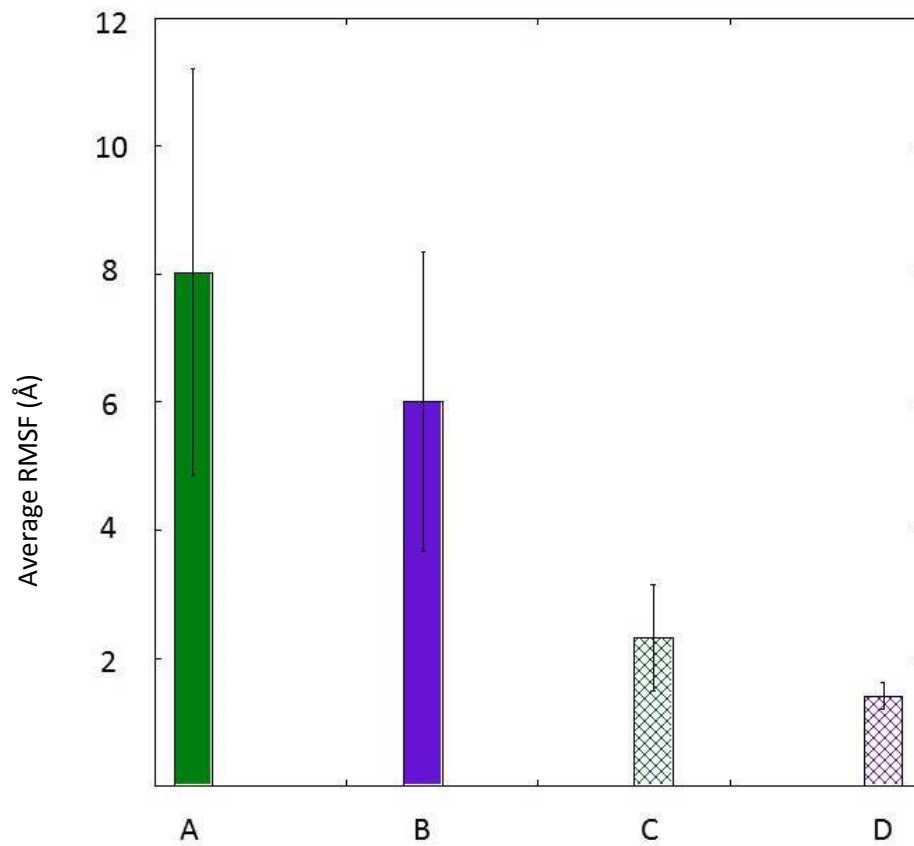


Figure 3-9: Comparison of average RMSF for simulations of the sST2 receptor

Comparison of the average RMSF values for the sST2 receptor de-glycosylated (A), glycosylated (B), de-glycosylated bound to ligand (C), and glycosylated bound to receptor (D)

same, suggesting a global and not site specific based attenuation due to N-linked glycosylations.

Plotting of the RMSF to the structure of sST2 shows that the two greatest areas of dynamics are found in Ig1 and Ig3 (Figure 3-6 and Figure 3-7). When correlated to conformations seen in the simulation, the dynamics of Ig1 are based on two events: one the twisting of the entire Ig1 domain and high dynamics in the disordered loop consisting of residues Tyr34 to Val65. Upon glycosylation, both of these areas show significant reduction in dynamics as well as significant stabilization in the flexible linker region. Although many of these regions contain glycosylations, the wide ranging effects of N-linked glycosylations imply that glycosylations are able to globally reduce the protein dynamics independent of relative proximity to glycosylation site.

Dynamics of Glycosylated and De-glycosylated sST2 Receptor Bound to Ligand

To gain a better understanding of the role of glycosylations upon receptor behavior, additional simulations and analyses were done upon the bound vs unbound complex. Compared to the free receptor, the bound sST2 receptor is much more stable and much of the dynamics have been eliminated through occupation of the binding site Figure 3-9 which manifest with a greatly reduced global RMSF of 2.3 Å. In much the same way that the receptor is highly dynamic prior to binding, the ligand IL-33 is also highly dynamic and undergoes a dynamic transition upon binding with highly disparate dynamic character free versus bound as seen when comparing Figure 3-9 and Figure 3-5. The entire bound complex loses much of its dynamics through stabilizing interactions and steric hindrance of the ligand in the pocket. Since much of the original dynamics present were from opening and closing of the binding pocket, it stands to reason that occupancy of the binding pocket largely hinders

Figure 3-10: Trajectory of de-glycosylated sST2 receptor and glycosylated sST2 receptor bound to IL-33.

Trajectory of the last 1000 frames of the simulation plotted versus RMSF in angstroms. Glycosylated trajectory is represented in green while the de-glycosylated trajectory is shown in purple. Average RMSF for each trajectory are shown in corresponding colors.

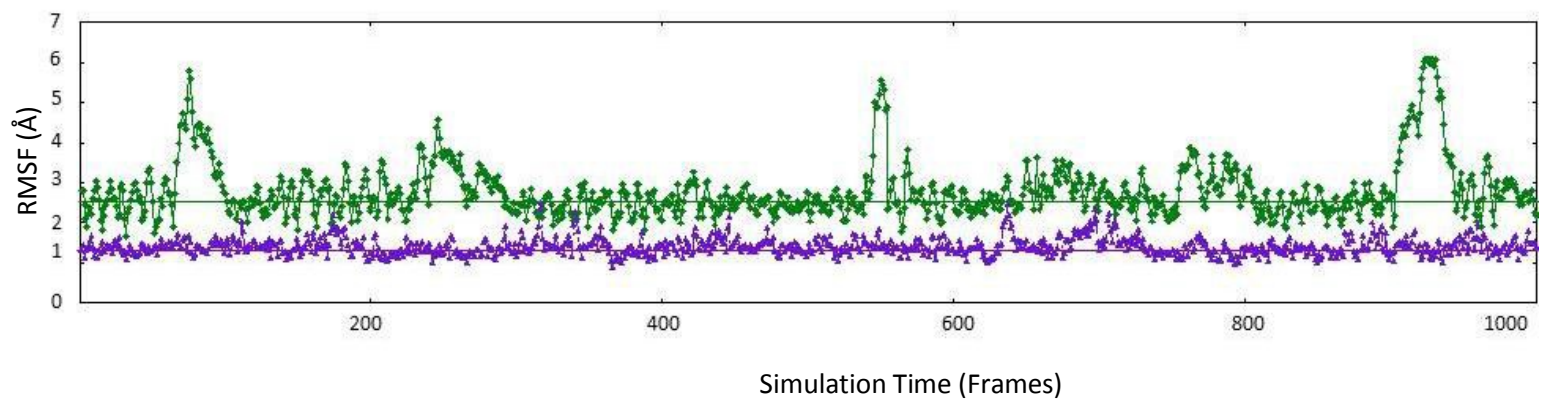


Figure 3-11: RMSF of the de-glycosylated sST2 receptor bound to IL-33 plotted for each residue along the protein

RMSF in angstroms plotted versus residue number of de-glycosylated sST2 is shown in green by itself (A), glycosylated sST2 is shown in purple by itself (B), and the two plots overlaid (C). Secondary structure elements are shown at the bottom with blue arrows representing β -sheet elements, red arrows representing α -helix elements, glycosylation sites denoted by a blue N, domain constituting Ig 1 is shown in yellow, domain constituting Ig 2 is shown in orange, and domain constituting Ig 3 is shown in blue.

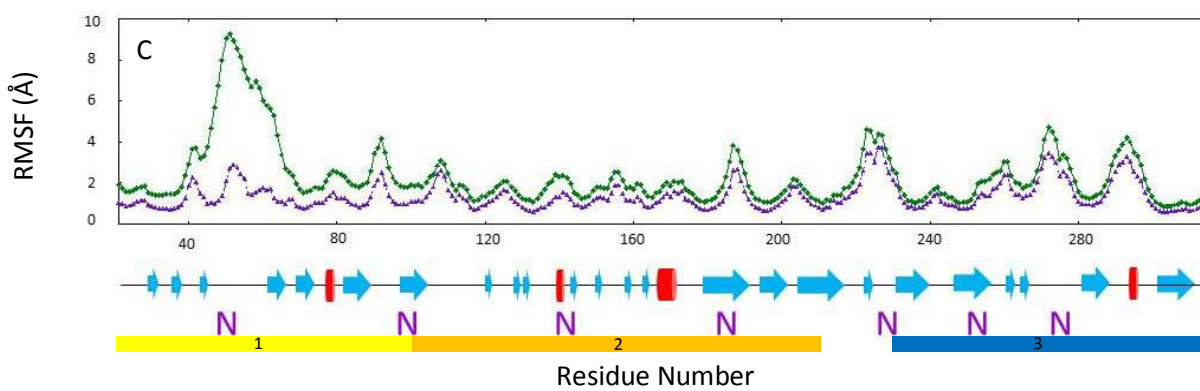
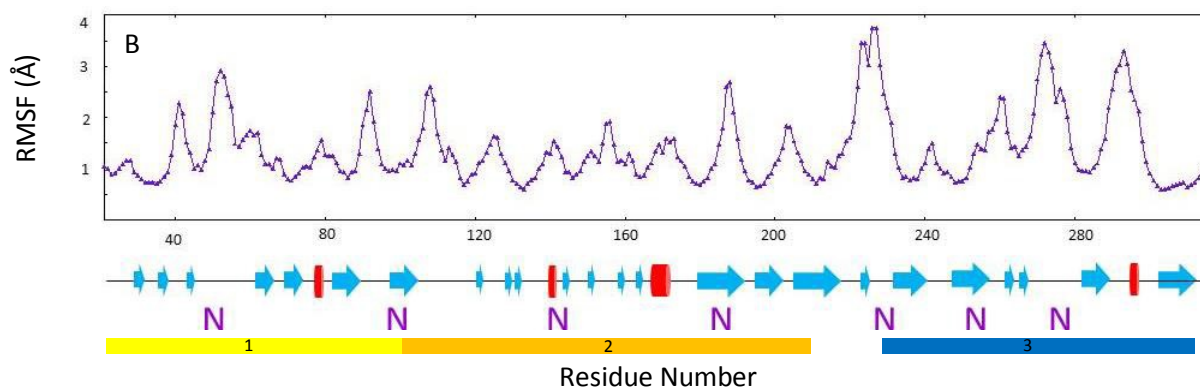
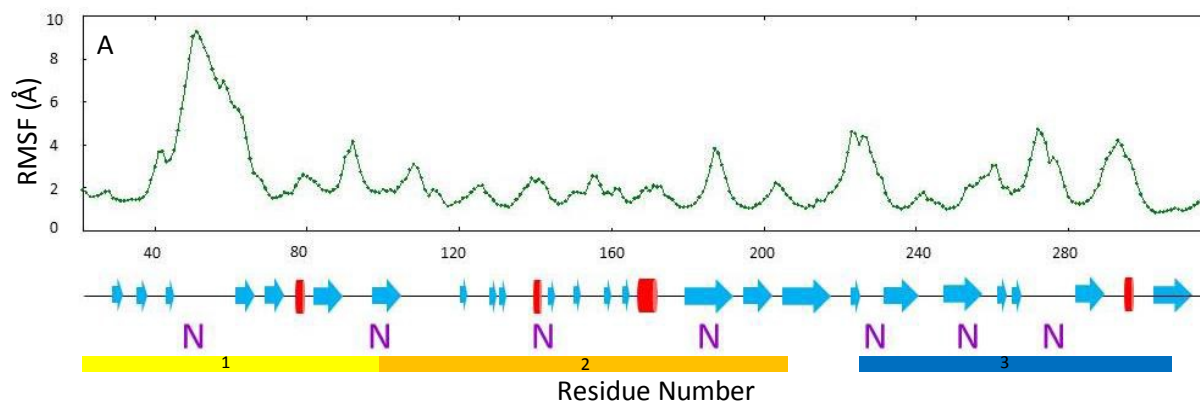


Figure 3-12: RMSF levels of the de- glycosylated simulation of sST2 bound to ligand plotted onto the structure of the sST2 receptor complex.

RMSF levels plotted onto the structure of the sST2 from the de-glycosylated simulation where red indicates levels above 6 Å, levels between 6 Å and 4 Å are shown in grey, and levels below 4 Å are shown in blue. The ligand IL-33 is depicted in dark grey

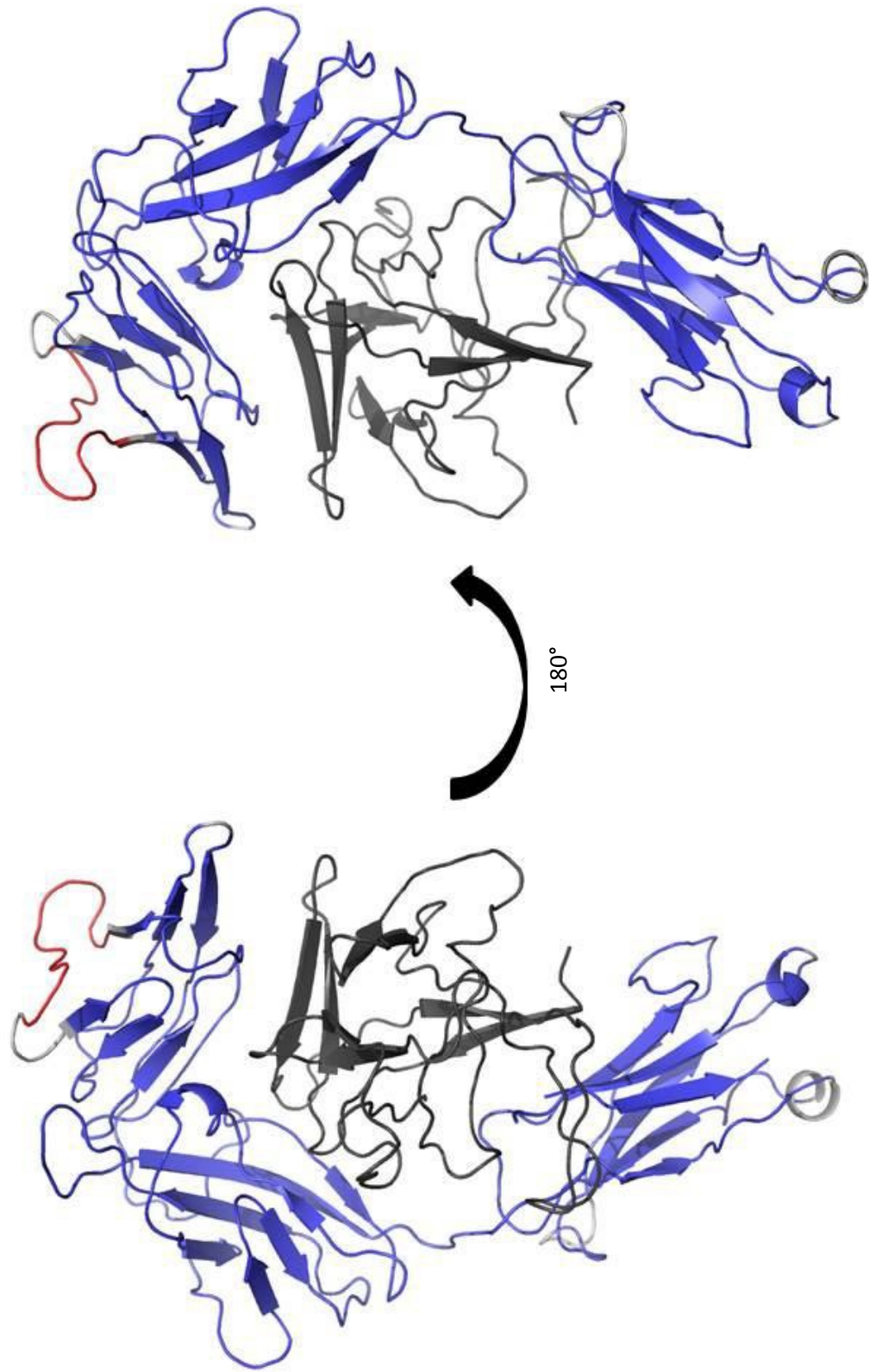
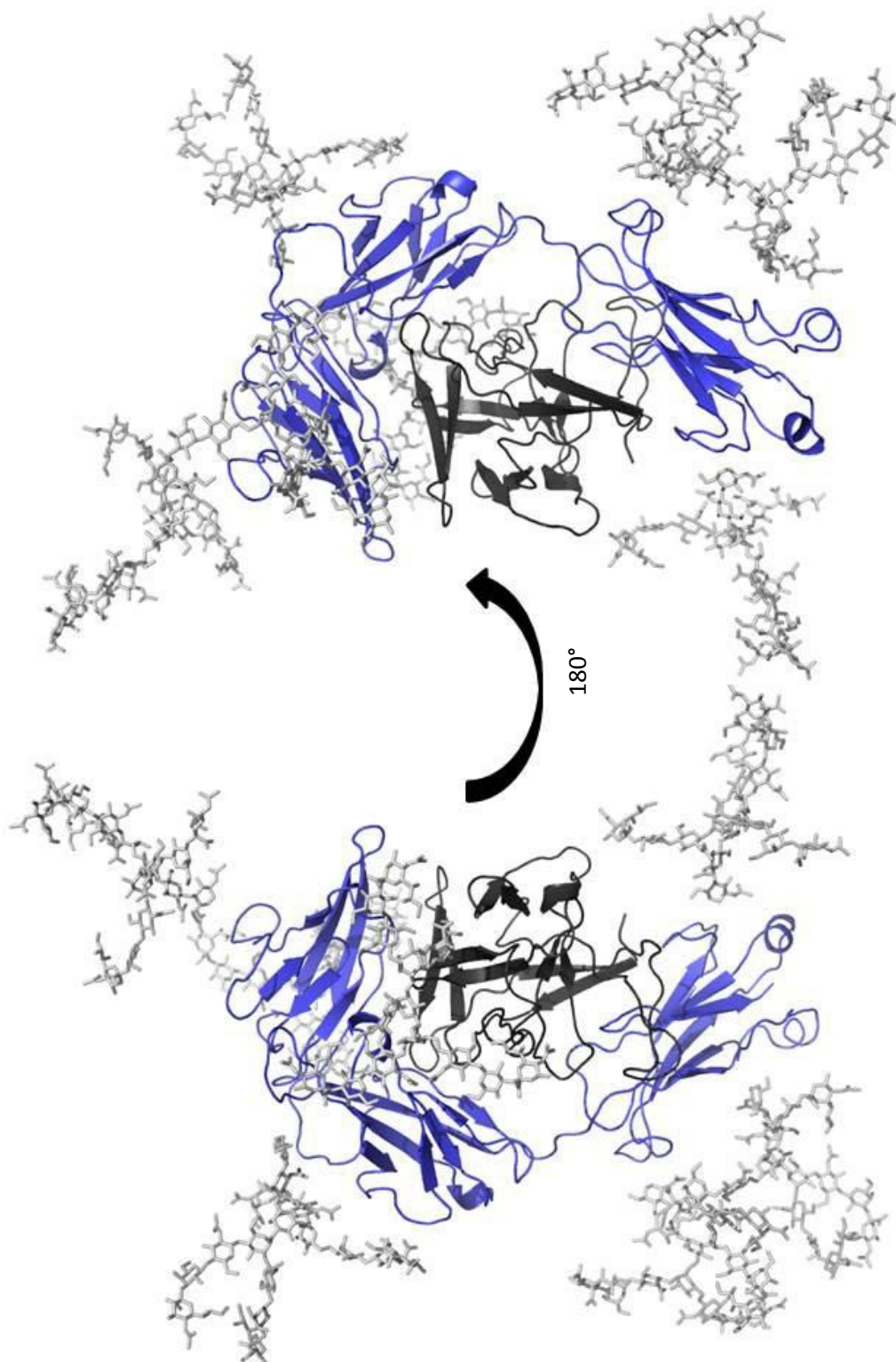


Figure 3-13: RMSF levels of the glycosylated simulation of sST2 bound to ligand plotted onto the structure of the sST2 receptor complex.

RMSF levels plotted onto the structure of the sST2 from the de-glycosylated simulation where red indicates levels above 6 Å, levels between 6 Å and 4 Å are shown in grey, and levels below 4 Å are shown in blue. The ligand IL-33 is depicted in dark grey

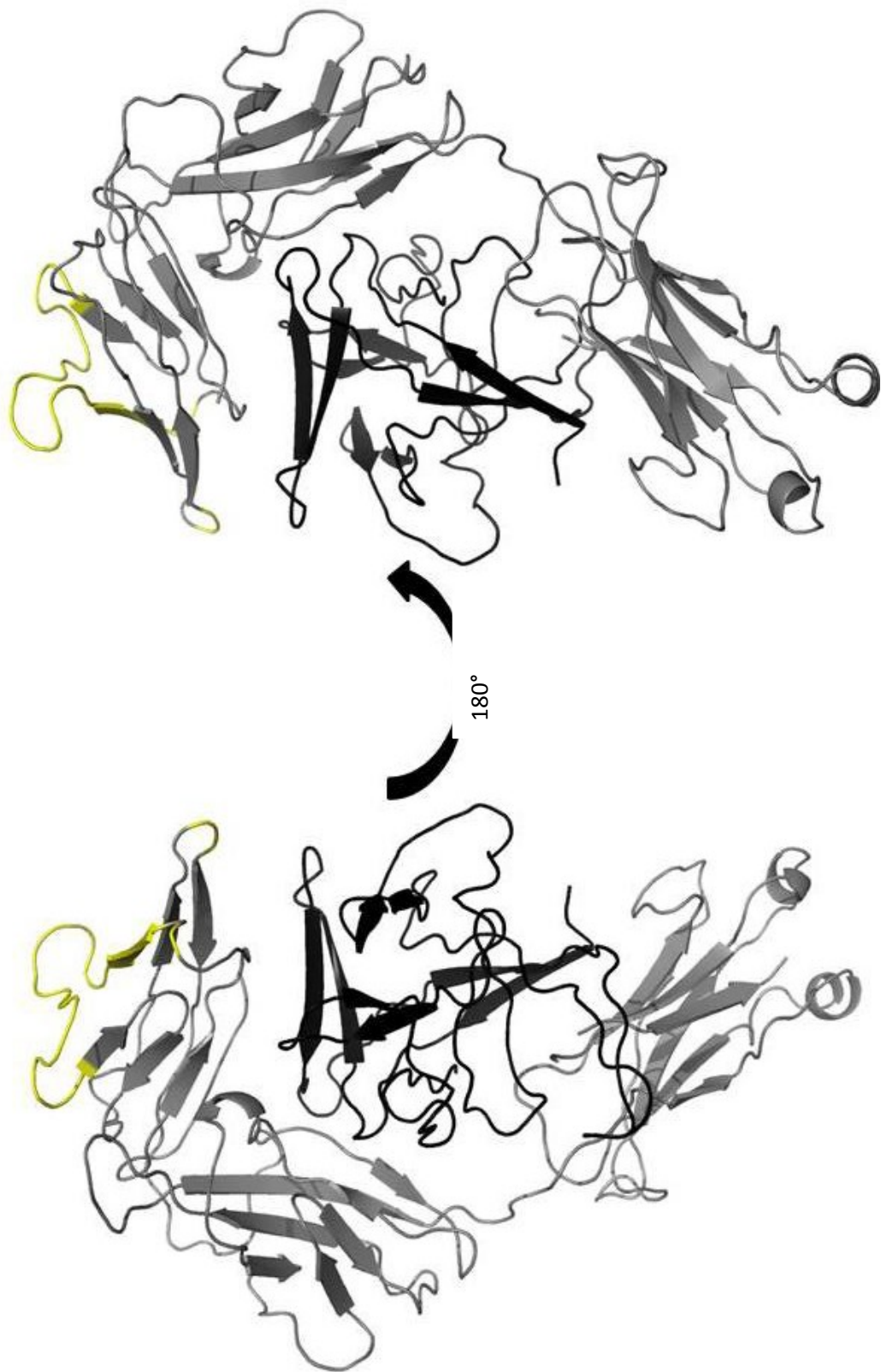


the ability for the receptor to move. Although the majority of the protein undergoes a large dynamic transition, one region retains the majority of its dynamics. This unstructured loop from Tyr34 to Val65 undergoes large loop rearrangement and local unfolding events that are only observed upon binding to the ligand. Much like the character of the de-glycosylated receptor complex, the glycosylated complex also undergoes a large scale dynamic transition into a much more stable structure. Interestingly though, the loop that previously retained much of its dynamics and unfolding character upon binding contains a glycosylation site, and upon glycosylation, the dynamics of the loop become attenuated. This stabilization of unstructured loops is hypothesized to be due to additional contacts that are mediated by the N-linked glycosylation present in order to stabilize a highly dynamic unstructured loop.

Analysis of residue specific dynamics in sST2 reveals a widespread loss of dynamics and stabilization throughout the protein Figure 3-10 consistent with global analysis of RMSF. Although the majority of the protein has lost much of its dynamics, the behavior of the unstructured loop spanning Tyr34 to Val65 maintained its dynamics at values similar to free sST2 receptor Figure 3-10. Aside from this peak, the landscape of the dynamics across the receptor is still rugged with small regions displaying elevated dynamics compared to the rest of the protein. In comparison to the glycosylated receptor, many of these locations of elevated dynamics are reduced through global stabilization of the protein in a manner similar to the free receptor. Due to the receptor being less dynamic, the differences in dynamics between glycosylated and de-glycosylated are much less suggesting that the stabilization of N-linked glycans is, although a global event, dependent on the existing stability of the receptor.

Figure 3-14: RMSF differences between de-glycosylated and glycosylated simulations plotted onto the structure of sST2

RMSF differences between the de-glycosylated and glycosylated simulation per residue of greater than 2 Å are shown in yellow where dark grey are differences less than 2 Å. Glycosylation sites are shown by the side chain they are attached to in purple. Structure of IL-33 is shown in black



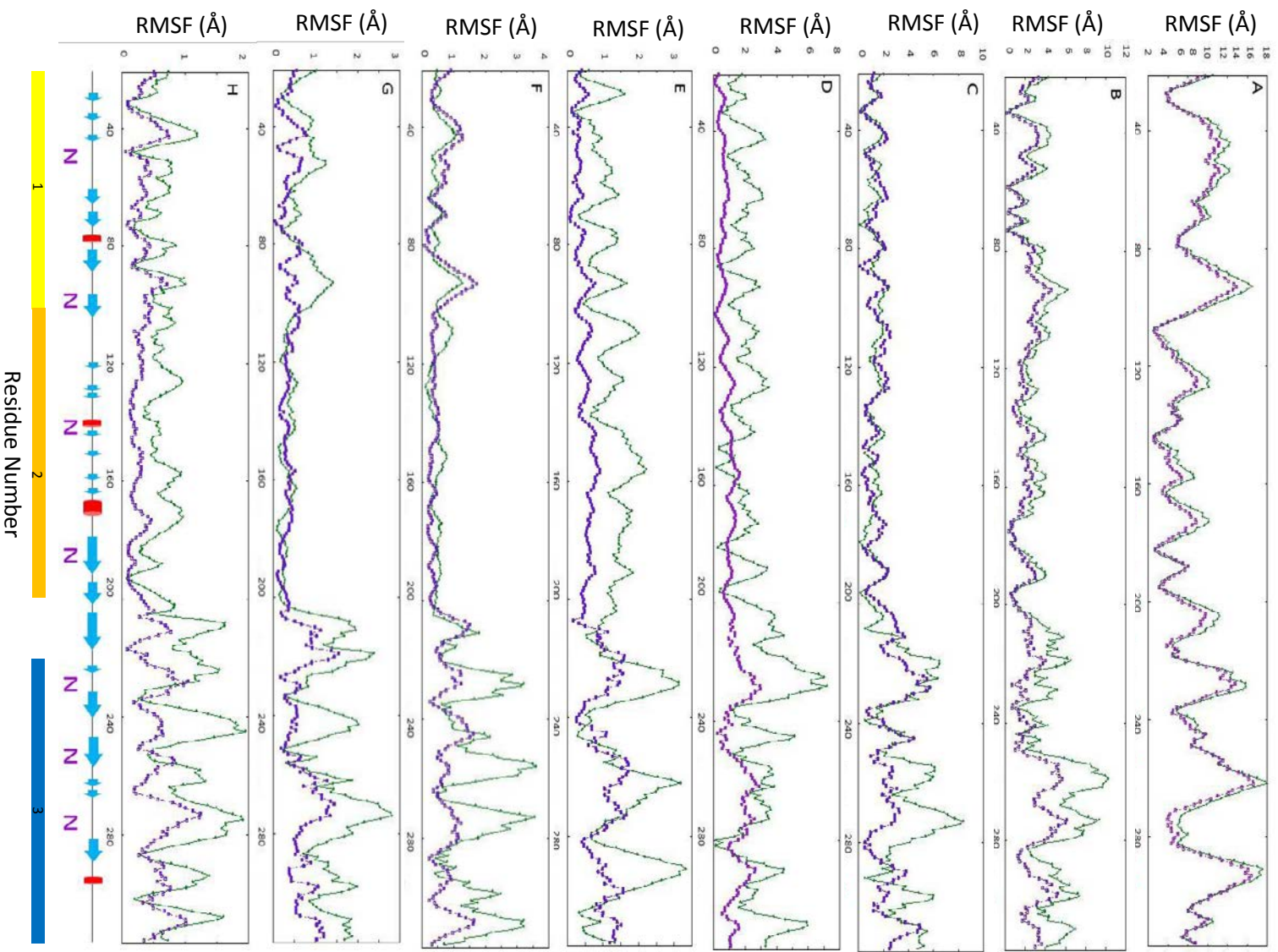
Eigenvalue Analysis of Protein Dynamics

In order to determine the specific motions that are attenuated due to the presence of N-linked glycosylations, covariance analysis was utilized to de-convolute the differences between motions. By correlating the motions of one residue to the next across all frames in the trajectory, covariance analysis divides the motions of the entire simulation into discrete groups based on eigenvalues. As such, each discrete set of motions is assigned an eigenvalue that represents the percentage of the sum of eigenvalues and describe the distribution of motion from the slowest time scales to the fastest time scales. Subsequently, the RMSF calculated using each eigenvalue can be considered as the dynamics of one particular set of motions. Utilizing this type of analysis, the behavior of the first eight eigenvalues for each simulation was analyzed to ascertain the character of each motion as well as the distribution of the motions in order to determine the distribution of motions across all available time scales (23–25).

Analysis of each eigenvalue for both the de-glycosylated and glycosylated receptors reveals unique aspects of the correlation between certain motions and the effect of glycans. Although comparison of the motions of the largest eigenvalue between the de-glycosylated and glycosylated forms of the sST2 receptor show a small degree of change, the eigenvalues for the next seven show much greater change. This single eigenvalue accounts for roughly 40% of the total motions present for the de-glycosylated receptor and 60% of the motions for the de-glycosylated receptor suggesting that the single most dynamic motion of the de-glycosylated free sST2 receptor remains unchanged by the presence of glycosylations (Figure 3-16). This shift in weighting suggests that the shorter time scales of the free receptor lose much of their dynamics upon glycosylation. Consistent with the global RMSF analysis, the largest and nearly universal deviations occur in Ig3. These eigenvalues most likely define the motions

Figure 3-15: Eigenvalue analysis of the first 8 eigenvalues derived from a diagonalized covariance matrix comparing de-glycosylated and glycosylated simulations

Eigenvalue analysis of the RMSF for eigenvalue 1 (A), eigenvalue 2 (B), eigenvalue 3 (C), eigenvalue 4 (D), eigenvalue 5 (E), eigenvalue 6 (F), eigenvalue 7 (G), and eigenvalue 8 (H) where the de-glycosylated simulation RMSF for each eigenvalue is shown in green while the glycosylated RMSF for each eigenvalue is shown in purple across the amino acid sequence of sST2. Secondary structure elements are shown at the bottom with blue arrows representing β -sheet elements, red arrows representing α -helix elements, glycosylation sites denoted by a blue N, domain constituting Ig 1 is shown in yellow, domain constituting Ig 2 is shown in orange, and domain constituting Ig 3 is shown in blue.

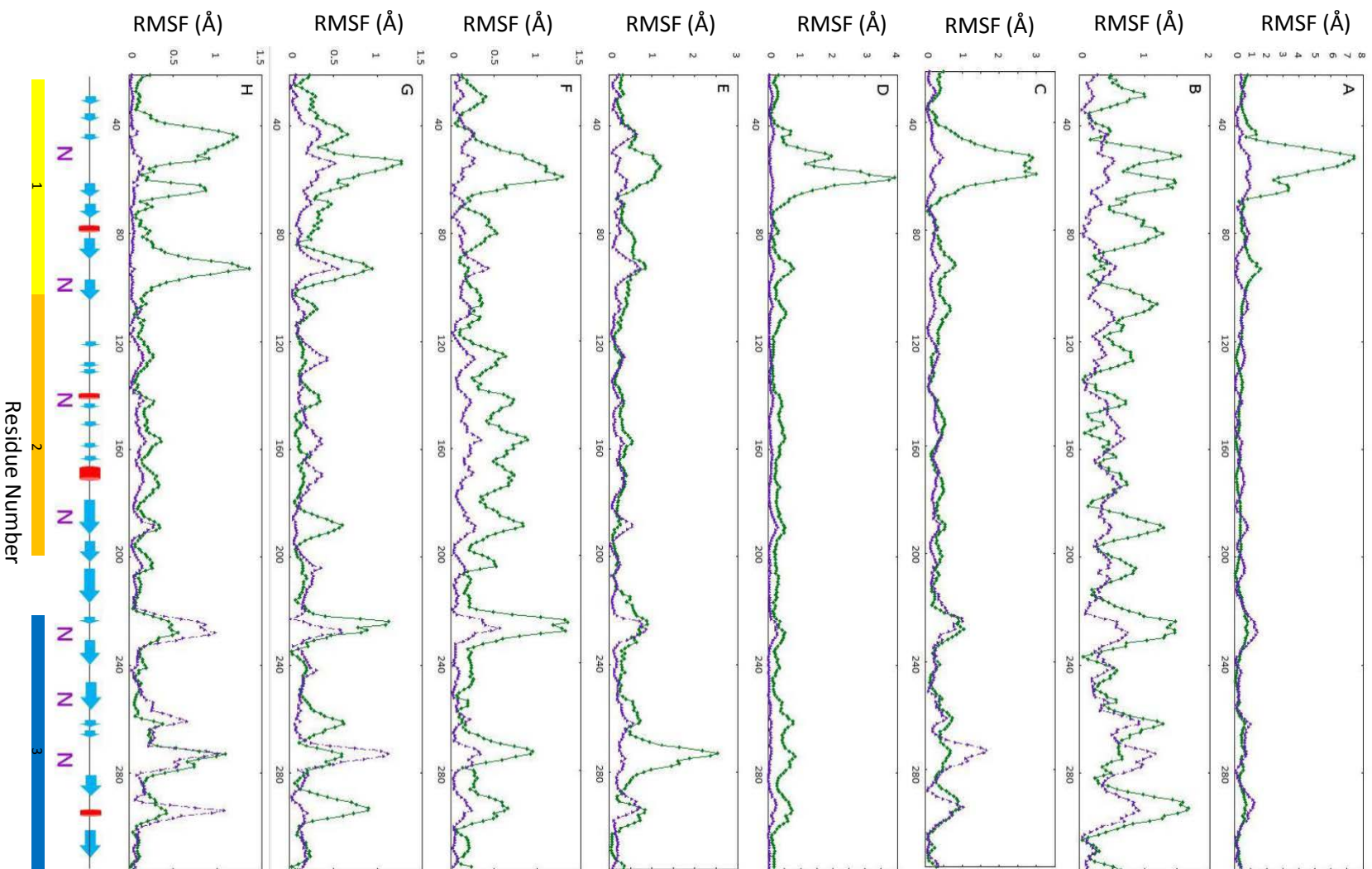


that create the global attenuation of dynamics while eigenvalues 3 and 6 are those that inhibit the dynamics of Ig3 through reduction of the degrees of freedom of the flexible linker. These eigenvalues show attenuation of dynamics present in eigenvalues 3 and 6 occur independently of motions present in other domains. Eigenvalues 4, 5, and 8 in particular show significant change across the entire protein suggesting that N-linked glycosylations are changing the dynamics of the shorter time scale motions of the backbone (Figure 3-14). The sum of the first eight eigenvalues for both the glycosylated and de-glycosylated receptor show that the motions described by the first eight eigenvalues account for over 90% of all motions present, suggesting that the distribution of motions for the distribution of the motions across time scales is largely maintained despite the presence of glycosylations (Figure 3-16).

When analyzing the eigenvalues of the bound ST2 receptor in the de-glycosylated and glycosylated state, a marked difference can be seen in both the eigenvalue distribution as well as the character of each motion. To compare the relative transition in dominant motion, the distribution of eigenvalues across the first eight eigenvalues were compared in Figure 3-11 and shows a large shift towards the higher numbered eigenvalues shown by the occupancy of larger percentages in higher eigenvalue numbers. As such, the large scale motions previously described in the free ST2 receptors are shown to be largely attenuated resulting in population of shorter time-scale motions (Figure 3-17). Unlike the comparison between the free ST2 receptor glycosylated vs de-glycosylated, comparison of the RMSF for each of the first eight eigenvalues shows significant differences between RMSF per residue of all eigenvalues with the greatest differences seen in eigenvalue numbers 2, 6, 7, and 8. Likewise, eigenvalue numbers 1, 3, and 4 describe the set of motions of unstructured loop

Figure 3-16: Eigenvalue analysis of the first 8 eigenvalues derived from a diagonalized covariance matrix comparing de-glycosylated and glycosylated simulations of sST2 bound to receptor

Eigenvalue analysis of the RMSF for eigenvalue 1 (A), eigenvalue 2 (B), eigenvalue 3 (C), eigenvalue 4 (D), eigenvalue 5 (E), eigenvalue 6 (F), eigenvalue 7 (G), and eigenvalue 8 (H) where the de-glycosylated simulation RMSF for each eigenvalue is shown in green while the glycosylated RMSF for each eigenvalue is shown in purple across the amino acid sequence of sST2. S Secondary structure elements are shown at the bottom with blue arrows representing β -sheet elements, red arrows representing α -helix elements, glycosylation sites denoted by a blue N, domain constituting Ig 1 is shown in yellow, domain constituting Ig 2 is shown in orange, and domain constituting Ig 3 is shown in blue.



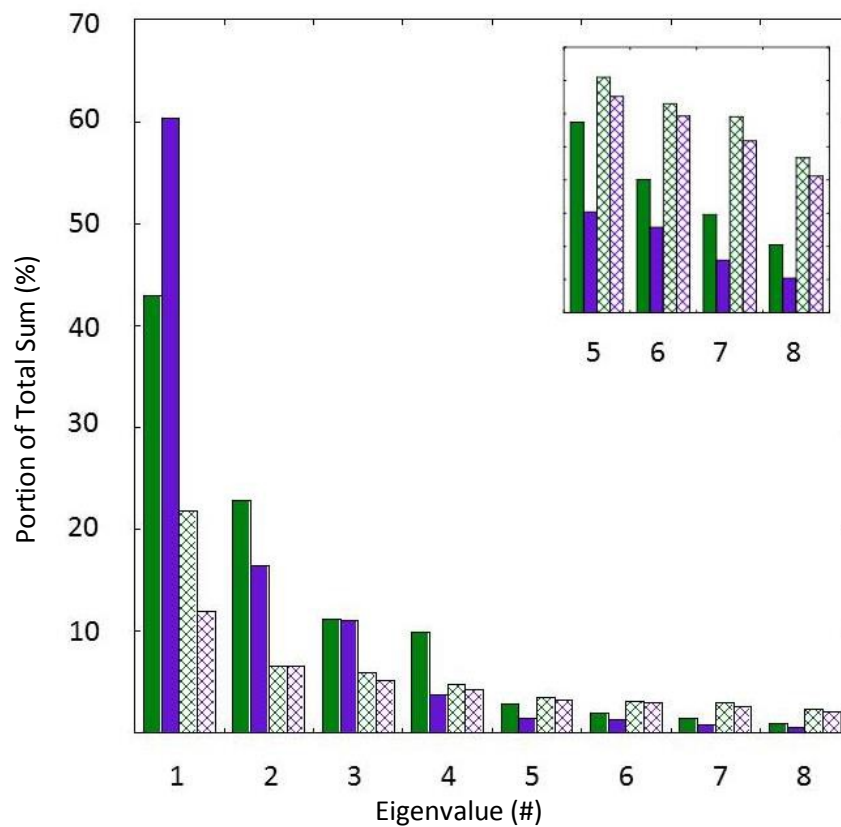


Figure 3-17: Plot of the first 8 eigenvalues derived from simulations of the sST2 receptor

First 8 eigenvalues for each simulation are converted into percentages of the total sum of eigenvalues for each simulation. Eigenvalues for de-glycosylated sST2 are shown in green, eigenvalues for glycosylated sST2 are shown in purple, eigenvalues for de-glycosylated sST2 bound to receptor are shown in lightly shaded green, and eigenvalues for glycosylated sST2 bound to receptor are shown in lightly shaded purple. A zoomed in portion of the last 4 eigenvalues is shown in the top right corner.

Table 3-1: Table of the first 8 eigenvalues for sST2, sST2-glycosylated, sST2-bound, and sST2-bound-glycosylated.

	sST2	sST2-Glycosylated	sST2-Bound	sST2-Bound-Glycosylated
Eigenvalue Number	Eigenvalue	Eigenvalue	Eigenvalue	Eigenvalue
1	94.5472	74.3552	6.02015	1.12412
2	50.2558	20.2184	1.80639	0.622885
3	24.7946	13.626	1.65729	0.4885
4	21.8459	4.59786	1.33599	0.397166
5	6.3305	1.87043	0.979127	0.306666
6	4.4225	1.5866	0.869563	0.279241
7	3.26328	0.977519	0.814792	0.243656
8	2.24796	0.638899	0.644229	0.194028

unfolding that have been eliminated as a result of N-linked glycosylation. Through the eigenvalue analysis of the bound ST2 receptor glycosylated and de-glycosylated, it can be concluded that N-linked glycosylations not only have a greater effect upon the bound complex with IL-33, but attenuate a specific subset of motions found in the loop from Tyr34 to Val65.

3.4 Discussion

Attenuation of Dynamics and Protein Stability

From the preliminary data presented here, N-linked glycosylations do not seem to significantly affect the secondary structure but are able to directly influence the dynamics of the sST2 receptor. Despite the presence of glycosylations occurring at specific sites, the effects of N-linked glycosylations on dynamics are not limited to the loci surrounding each glycosylation site and can be felt globally. The cause of this could be due to steric hindrance of protein motion and stabilization mediated through glycan chains. This effect seems to be consistent behavior bound complex versus the free receptor; though the amplitude of the effect seems to be proportional to the amount of dynamics intrinsically present in the receptor. Additionally, glycosylations smooth the landscape of the bound receptor and limit the unique dynamics that arise upon binding of IL-33 to receptor.

Although much of the behavior of the receptor is consistent with previously suspected models and experimental work done upon the ST2 receptor, the work here describes dynamics previously unreported (26, 27). The unique behavior of the unstructured

loop spanning Tyr34 to Val65 is one that only arises as a result of ligand binding consistent with the fact that the loop was unresolvable by X-ray crystallography when crystallized in complex with the receptor. This unique discovery opens the possibility of this loop as a allosteric regulation of ligand binding and may hold some importance in the behavior of the heterotrimeric complex though more testing is required to probe this event (26, 27).

Dynamics hold particular importance for soluble receptors in their ability to stabilize protein dynamics to increase serum half-life as well as in localization to certain areas of the body. Although more well characterized in the case of antibodies, the presence of N-linked glycosylations stabilize protein dynamics and increase the serum half-life (28). Despite the fact that the simulations run here are coarse-grained, the general behavior is consistent with other work probing the effects of N-linked glycosylations showing the validity of this method in the stabilizing effect of N-linked glycosylations (6). Likewise, it is interesting to note that the effects of N-linked glycosylations are felt throughout the protein and not just localized to one domain. When taken in the context of a mucous environment, it is possible that the dynamics of the receptor *in vivo* or even more attenuated than our current model suggests. Due to the high amount of glycosylations, the receptor would most likely localize to the mucous layer where it can have glycan-glycan interactions with other highly glycosylated proteins as well as mucins (29). In this manner, crowding effects are increased in this unique environment where water molecules are highly ordered and stabilized by glycan orientation.

The dynamics present in the ST2 receptor before and after ligand binding represent a never before observed event present in the IL-1 family. As a result, the ramifications and underlying rules governing this transition have yet to be elucidated; however, the results of

the IL-1RAcP simulation provide some interesting insight into the underlying aspects of the system. One possible rationale behind the dynamic transition lies in the requirement for ST2 to bind to IL-1RAcP. It is known previously that ligand binding must occur prior to recruitment of IL-1RAcP, and in this manner, poses the hypothesis that the ST2 receptor is unable to bind IL-1RAcP due to its high dynamic nature. As such, ligand binding by IL-33 which is also a highly dynamic protein (Data not shown), may allow for the stabilization of ST2 such that it now populates a less dynamic state that is more amenable to ligand binding independent of a perturbations to secondary structure (30).

3.5 Future Work

Based on the success of the preliminary data in properly interpreting glycosylated residues and the correlation with results shown in previously published work, our observation suggest that glycosylations may have a large impact upon the dynamic character of sST2. In order to explore these results further, longer simulations should be run to corroborate the observations shown. Since the suspected perturbations due to glycosylations occur at longer time scales, simulations run along the same time scale are desired to further strengthen the observations shown here.

As with any simulation work, all observed phenomena must be verified experimentally; however, due to the challenges associated with producing high concentrations of homogenous glycosylated protein, this remains a challenge. To address these challenges, there are two approaches: one to utilize an *in vitro* translation or synthesis method to obtain isotopically labelled protein or to obtain an isotopically labelled sample from mammalian lineage and purify all homogenous glycoforms out from others. Regardless

of the method used to obtain the sample, isotopically labeled sample for use in NMR is required. Because NMR is able to monitor all time scales of motion through a standardized suite of experiments, it is the preferred method of choice. Likewise, NMR is able to yield residue specific data that can be directly compared to the results shown here. As such, the results from the NMR experiments can be directly compared to the motions probed via eigenvalues and correlated into discrete time scales (31).

Consistent with previous work shown between antibody stability and the presence of N-linked glycans, the presence of glycosylations on the sST2 receptor attenuates protein dynamics globally. In order to determine the ramifications of this, differential scanning calorimetry (DSC) would be utilized. Both of these two methods allow for sensitive testing of stability in solution and can help to elucidate the effects of increased stability. Using DSC, melting curves of sST2 glycosylated and de-glycosylated can be measured to determine thermal stability of the receptor due to presence of N-linked glycosylations. Additional testing to determine the effects of certain sites upon thermal stability can then be probed by mutating asparagine residues at specific sites to aspartates. In this manner, proteins mimic their de-glycosylated form and the global as well as site specific perturbations that arise due to the presence of N-linked glycosylations can be probed.

Despite the biological work done on the system, not much is known regarding the behavior or mechanism of interaction between the three members of the heterotrimeric complex. Because the contacts and conformation of the final heterotrimeric assembly is still under some debate, a full MD docking study can help to elucidate the mechanism and interactions involved with the recruitment of binding partners and final stabilizing interactions of the entire heterotrimeric complex. Likewise, creation of this system to study

the mediating contacts can also be utilized to monitor the dynamic transitions that seem to occur upon binding of IL-33. To achieve this, the structure of the heterotrimeric complex upon binding would most likely need to be solved. With the structure of the heterotrimeric complex solved, a transition path sampling simulation can be run to elucidate the docking event of ligand to receptor and subsequently receptor complex to receptor accessory protein. Further elucidation into the mechanism behind binding and the subsequent stabilization and dynamic transition that occurs may yield important insight into the mechanism behind the order of recruitment (32).

To further probe the effects of glycosylations upon the receptor complex, an in depth analysis of the distribution of contacts created between GLYCAN/AMINO and GLYCAN/GLYCAN can be done. By analyzing the effects of glycan mediated contacts, a distribution and elucidation into the specific interactions that lead to stabilization of unstructured loops can be achieved (33).

As with any creation of a new tool, in particular simulations, there are plenty of improvements that can be applied to the SMOG2 add-on. For the sake of simplicity and proof of principle, current parameters were based on interactions between amino acid residues and their side chains, and as such, further refinement upon the behavior of glycans is possible. Interaction distances between GLYCAN/GLYCAN interaction strengths and contact distance must be probed experimentally in order to determine the proper distance and interaction strength. Glycans, because of their hydrophilic nature often obtain conformations and create contacts mediated by waters. Because the simulation run is in vacuum, a way to mimic this event is required (34). Likewise, interaction strengths between GLYCAN/GLYCAN residues and GLYCAN/AMINO residues hold room for improvement. Since

these contacts are implicitly different in interaction strength than those of AMINO/AMINO side chains, refinement is necessary in order to correctly model the behavior observed in a full MD simulation.

3.6 References

1. Adcock S a, Mccammon JA (2006) Molecular Dynamics : Survey of Methods for Simulating the Activity of Proteins Molecular Dynamics : Survey of Methods for Simulating the Activity of Proteins. 106(February):1589–1615.
2. Shental-Bechor D, Levy Y (2008) Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proc Natl Acad Sci U S A* 105(24):8256–8261.
3. Banks DD (2011) The effect of glycosylation on the folding kinetics of erythropoietin. *J Mol Biol* 412(3):536–550.
4. Price JL, Shental-Bechor D, Dhar A, Turner MJ, Powers ET, Gruebele M, Levy Y, Kelly JW (2010) Context-dependent effects of asparagine glycosylation on pin WW folding kinetics and thermodynamics. *J Am Chem Soc* 132(43):15359–15367.
5. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, Woods RJ (2008) GLYCAM06: A generalizable biomolecular force field. carbohydrates. *J Comput Chem* 29(4):622–655.
6. Lee HS, Qi Y, Im W (2015) Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study. *Sci Rep* 5:8926. Available at: <http://www.nature.com/doi/10.1038/srep08926>.
7. Wood NT, Fadda E, Davis R, Grant OC, Martin JC, Woods RJ, Travers SA (2013) The influence of N-linked glycans on the molecular dynamics of the HIV-1 gp120 V3 loop. *PLoS One* 8(11).
8. Best RB, Buchete N-V, Hummer G (2008) Are current molecular dynamics force fields too helical? *Biophys J* 95(1):L07–L09.
9. Plimpton S (1995) Computational limits of classical molecular dynamics simulations. *Comput Mater Sci* 4(4):361–364.
10. Wu C, Shea J (2012) Computational Modeling of Biological Systems. *Peptides*:215–227. Available at: <http://www.springerlink.com/index/10.1007/978-1-4614-2146-7>.

11. Whitford PC, Sanbonmatsu KY, Onuchic JN (2012) Biomolecular dynamics: order-disorder transitions and energy landscapes. *Reports Prog Phys* 75(7):076601.
12. Feig M, Brooks CL (2004) Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr Opin Struct Biol* 14(2):217–224.
13. Tozzini V (2005) Coarse-grained models for proteins. *Curr Opin Struct Biol* 15(2):144–150.
14. Clementi C (2008) Coarse-grained models of protein folding: toy models or predictive tools? *Curr Opin Struct Biol* 18(1):10–15.
15. Noid WG (2013) Perspective: Coarse-grained models for biomolecular systems. *J Chem Phys* 139(9).
16. Martinetz T, Schulten K (1994) Topology representing networks. *Neural Networks* 7(3):507–522.
17. Whitford PC, Noel JK, Gosavi S, Schug A, Sanbonmatsu KY, Onuchic JN (2009) An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields. *Proteins Struct Funct Bioinforma* 75(2):430–441.
18. Torrie GM, Valleau JP (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J Comput Phys* 23(2):187–199. Available at: <http://www.sciencedirect.com/science/article/pii/0021999177901218>.
19. Noel JK, Whitford PC, Sanbonmatsu KY, Onuchic JN (2010) SMOG@ctbp: Simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res* 38(SUPPL. 2).
20. Clementi C, Nymeyer H, Onuchic JN (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* 298(5):937–953.
21. LOOP-Pred Server Available at: <http://manaslu.aecom.yu.edu/loopred/> [Accessed May 7, 2015].
22. Woods Group. (2005-2015) GLYCAM Web. Complex Carbohydrate Research Center, University of Georgia, Athens, GA. (<http://www.glycam.com>) Available at: <http://glycam.org/tools/molecular-dynamics/glycoprotein-builder/upload-pdb> [Accessed May 3, 2015].
23. Amadei A, Linssen ABM, Berendsen HJC (1993) Essential dynamics of proteins. *Proteins Struct Funct Genet* 17(4):412–425.

24. Hayward S, De Groot BL (2008) Normal modes and essential dynamics. *Methods Mol Biol* 443:89–106.
25. Van Aalten DMF, de Groot BL, Findlay JBC, Berendsen HJC, Amadei A (1997) A comparison of techniques for calculating protein essential dynamics. *J Comput Chem* 18(2):169–181. Available at:
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.106.9508&rep=rep1&type=pdf>
[http://doi.wiley.com/10.1002/\(SICI\)1096-987X\(19970130\)18:2<169::AID-JCC3>3.0.CO;2-T](http://doi.wiley.com/10.1002/(SICI)1096-987X(19970130)18:2<169::AID-JCC3>3.0.CO;2-T).
26. Lingel A, Weiss TM, Niebuhr M, Pan B, Appleton BA, Wiesmann C, Bazan JF, Fairbrother WJ (2009) Structure of IL-33 and Its Interaction with the ST2 and IL-1RAcP Receptors-Insight into Heterotrimeric IL-1 Signaling Complexes. *Structure* 17(10):1398–1410.
27. Liu X, Hammel M, He Y, Tainer JA, Jeng U-S, Zhang L, Wang S, Wang X (2013) Structural insights into the interaction of IL-33 with its receptors. *Proc Natl Acad Sci U S A* 110(37):14918–23. Available at:
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3773798&tool=pmcentrez&rendertype=abstract>.
28. Zheng K, Bantog C, Bayer R (2011) The impact of glycosylation on monoclonal antibody conformation and stability. *MAbs* 3(6).
29. Peracaula R, Barrabés S, Sarrats A, Rudd PM, de Llorens R (2008) Altered glycosylation in tumours focused to cancer diagnosis. *Dis Markers* 25(4-5):207–218.
30. Thomas C, Bazan JF, Garcia KC (2012) Structure of the activating IL-1 receptor signaling complex. *Nat Struct Mol Biol* 19(4):455–457.
31. Ishima R (2011) Recent Developments in (15)N NMR Relaxation Studies that Probe Protein Backbone Dynamics. *Top Curr Chem*. Available at:
<http://www.ncbi.nlm.nih.gov/pubmed/21898206>.
32. E W, Vanden-Eijnden E (2010) Transition-path theory and path-finding algorithms for the study of rare events. *Annu Rev Phys Chem* 61:391–420.
33. Sonnhammer ELL, Wootton JC (1998) Dynamic contact maps of protein structures. *J Mol Graph Model* 16(1):1–5.
34. Nagae M, Yamaguchi Y (2012) Function and 3D structure of the N-glycans on glycoproteins. *Int J Mol Sci* 13(7):8398–8429.