Effect of glycosylation on the structure and dynamics of
interferon alpha-2a by NMR studies

by

Pascal Belcourt

A thesis submitted to the Faculty of Graduate and
Postdoctoral Affairs in partial fulfillment of the requirements
for the degree of

Master of Science

In

Biology

Carleton University
Ottawa, Ontario

© 2011

Pascal Belcourt

Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-83174-8
Our file *Notre référence*
ISBN: 978-0-494-83174-8

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

■ ■ ■
Canada

Abstract

Glycosylation is a post-translational modification of proteins which modulates their biological properties. The impact on the structure and dynamics of a simple O-glycan on interferon alpha-2a has been studied. Interferon alpha-2a was produced *in vivo* with *E. coli*, in a medium enriched with NMR active isotopes (^{13}C , ^{15}N) and subsequently glycosylated *in vitro* with a single N-acetyl galactosamide sugar unit. NMR spectroscopy was used to obtain information on the dynamics and the structure of the protein. Assignment of the backbone resonances of glycosylated interferon alpha-2a confirmed that Thr-106 is the sole glycosylation site. Comparison between the unglycosylated and glycosylated interferon suggested that glycosylation did not have an important impact on the structure of the protein. Measurements of backbone dynamics showed a significant variation in the glycosylated loop dynamics.

Acknowledgements

I would like to thank my supervisor Yves Aubin for the opportunity to work at his laboratory and the help he provided to me during my time there. I also wish to thank Simon Sauvé for providing great help during the collection and analysis of the NMR data, as well as mentoring me a great deal during those two years. This project was made possible by the collaboration between the National Research Council of Canada and Health Canada. Great many thanks to Michel Gilbert and Denis Brochu at the Institute for Biological Sciences at the NRC for the glycosylation of interferon alpha-2a. Finally, I would like to thank NSERC for providing financial support.

This thesis is dedicated to my parents and brothers who have never failed to give me support when needed. I would like to give my thanks to my mother, Lise Laferrière, who gave me unconditional help in anything I needed and corrected all my terrible English writing mistakes.

Table of Contents

Title page	i
Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Abbreviations	vii
List of Tables	viii
List of Figures	ix
List of Appendices	xi
1. <u>Introduction</u>	1
1.1. Interferon	2
1.1.1. Interferon alpha-2a	2
1.2. Glycobiology	5
1.2.1. Biological and biophysical properties due to glycosylation	6
1.2.1.1. Molecular recognition	
1.2.1.2. Glycans immunogenicity	
1.2.2. Glycosylation process	9
1.2.2.1. Saccharides are the building blocks of glycans	
1.2.2.2. N-glycosylation mechanism	
1.2.2.3. O-glycosylation mechanism	
1.2.3. Heterogeneity	17
1.2.4. Effect on structure and dynamics	17
1.3. Interferon used as a biotherapeutic	18
1.3.1. Hypothesis	20
1.3.2. Strategy	20
1.4. NMR background on techniques used	21
1.4.1. Basic principles of NMR	21
1.4.1.1. Relaxation and molecular motion	
1.4.1.2. Protein dynamics	
1.4.1.2.1. Time-scale of high resolution NMR	
1.4.2. Multidimensional NMR	28
1.4.3. ^1H - ^{15}N Heteronuclear Single Quantum Correlation (HSQC)	29
1.4.4. Extension to the third dimension, the 'HNCO' experiment	30
1.4.5. Resonance assignment: First step toward structure determination	31
1.4.5.1. Sequential assignment	
1.4.5.2. HNCACB experiment	
1.4.5.3. CBCACONH experiment	

1.5. Sequential assignment: the strategy	34
1.5.1. Sequential assignment of side-chain resonance	35
1.5.2. Filtered COSY spectra	36
1.5.3. Secondary structure determination with chemical shifts assignment	37
1.5.4. Backbone torsion angles	37
1.5.4.1. TALOS	
1.5.5. The Nuclear Overhauser effects and the measurement of the internuclear distances	39
1.5.6. Backbone motion via the $\{^1\text{H}\}$ - ^{15}N Heteronuclear NOE effects	40
1.6. Tertiary structure calculation	42
1.6.1. Semi-automatic software CYANA	42
1.6.2. Ramachandran plots	43
2. <u>Materials and Methods</u>	45
2.1. Protein expression and purification	46
2.2. Protein glycosylation	46
2.3. NMR spectroscopy	47
2.4. ^{15}N Backbone dynamics	49
2.5. Data processing	49
2.6. Structure calculation using semi-automated software CYANA	50
3. <u>Results and Discussion</u>	52
3.1. ^1H - ^{15}N HSQC Fingerprint	
3.1.1. Effect glycosylation on the HSQC spectrum	53
3.2. Glycosylated IFN alpha-2a resonance assignments	
3.2.1. Chemical shift correlations	58
3.2.2. Comparison of T106 and T108 HNCACB spectrum	62
3.2.3. Threonine 106 is the glycosylated residue on IFN alpha-2a	64
3.2.4. Chemical shift perturbation and structure	64
3.3. Backbone dynamics of interferon alpha-2a	
3.3.1. $\{^1\text{H}\}$ ^{15}N Heteronuclear NOE backbone dynamics pattern	66
3.3.2. Glycosylation reduces backbone motions of interferon alpha-2a	68
3.4. Secondary structure	
3.4.1. Chemical shift index	69
3.4.2. Chemical shift index comparison with unglycosylated IFN	72
3.4.3. TALOS torsion angle prediction	74
3.4.4. Short and medium range NOE for α -helices	74
3.4.5. Side chain and methyl group assignments	77
3.5. Three-dimensional structure obtained with CYANA	78
3.6. Impact of results obtained	80

4. <u>Conclusion</u>	76
5. <u>References</u>	78
6. <u>Appendix</u>	92

List of Abbreviations

Asp	Asparagine
Cys	Cysteine
Dol-P	Dolichol precursor lipid
ER	Endoplasmic reticulum
Gal	Galactose
GalNAc	N-Acetylgalactosamine
GalNAcT	N-Acetylgalactosamine transferase
GalT	Galactosyl transferase
GBP	Glycan-Binding Protein
Gln	Glutamine
HSQC	Heteronuclear Single Quantum Correlation
IFN	Interferon
MHC	Major Histocompatibility Complex
Neu5Ac	N-acetylneuraminic acid
NMR	Nuclear magnetic resonance
OST	Oligosaccharyl transferase
Ser	Serine
Thr	Threonine

List of Tables

Table 2.3: Parameters used for the NMR experiments done at 600 MHz	48
--	----

List of Figures

Figure 1: Three-dimensional structure of interferon alpha-2a	4
Figure 2: Glycosidic bond orientations for three possible sugars moieties (GalNAc, Gal and Neu5Ac)	11
Figure 3: Diagram of the N-glycosylation process and the various glycan chain modifications	12
Figure 4: Diagram of the O-glycosylation process	15
Figure 5: Glycosylation process leading to the native interferon alpha-2a O-glycosylation and the specific structures of the glycan at each step	16
Figure 6: Effect of the applied magnetic field on the orientation of the nuclei spins, splitting into two populations for the NMR active nuclei (^1H , ^{13}C , ^{15}N)	23
Figure 7: Effect of a perpendicular radio-frequency pulse on the Z component and its gradual shift toward the transverse plane	24
Figure 8: Spin-lattice and spin-spin relaxations effects on the coherence of the signal	24
Figure 9: Time scales of High-resolution NMR of different components which can be measured at precise frequency ranges	27
Figure 10: Diagram of a dipeptide backbone, with the label for the atoms positions and the possible coupling constant between each atom	28
Figure 11: Magnetization transfer diagram for the HNCACB experiment	32
Figure 12: Magnetization transfer diagram for the CBCA(CO)NH experiment	34
Figure 13: Three-dimensional display of the ^{15}N dimension separated into different layers corresponding to a specific amide chemical shift and the ^{13}C - ^1H chemical shift range extracted from one of these layers	35
Figure 14: Diagram showing the phi and phi backbone torsion angle of proteins	38
Figure 15: Correlation graph between the correlation time and the heteronuclear NOE ratios	41
Figure 16: Overall method used by CYANA for automated NOE assignment and structure calculation	51

Figure 17: HSQC spectrum of the unglycosylated IFN alpha-2a with amino acid labels	54
Figure 18: HSQC spectrum of the glycosylated IFN alpha-2a with amino acid labels	55
Figure 19: Overlay of both HSQC spectra of the unglycosylated and glycosylated protein, with labelled shifted residues	57
Figure 20: Overlay for the T106-T108 region of the unglycosylated and glycosylated IFN alpha-2a HSQC spectrum	58
Figure 21: ^{13}C - $^1\text{H}^{\text{N}}$ strips of CBCA(CO)NH and HNCACB 3D-experiments for the V103-T108 part of the glycosylated loop of interferon alpha-2a	61
Figure 22: ^{13}C - $^1\text{H}^{\text{N}}$ strips of the HNCACB experiments of T106 and T108 for the glycosylated and unglycosylated interferon alpha-2a	63
Figure 23: Attachment of a single N-acetyl galactosamine to T106 on one of the 20 structures for the bundle of structures calculated by NMR constraints of unglycosylated interferon alpha-2a	65
Figure 24: Overlay of the distribution of $\{^1\text{H}\}$ - ^{15}N Heteronuclear values for the interferon alpha-2a from Klaus <i>et al</i> and the unglycosylated protein prior to glycosylation	67
Figure 25: Overlay of the distribution of $\{^1\text{H}\}$ - ^{15}N Heteronuclear values for the unglycosylated protein prior to glycosylation and the glycosylated protein with a single N-acetyl galactosamine unit on T106	69
Figure 26: Chemical shift index of backbone atoms C_α , C_β , CO and H_α	71
Figure 27: Comparison of the C_α chemical shifts differences between glycosylated and unglycosylated IFN alpha-2a, including their overlay differences	73
Figure 28: Short and medium-range NOE connectivities and helical regions of the protein	76
Figure 29: A: Three-dimensional structure of glycosylated interferon alpha-2a, obtained by the CYANA	79

List of Appendices

Appendix A: Chemical shift table for the backbone atoms	92
Appendix B: Chemical shift table for the carbon atoms of the side chains	98
Appendix C: Chemical shift table for the proton and nitrogen atoms of the side chains	104
Appendix D: SDS-PAGE of the glycosylation process	110

1-Introduction

1.1. Interferon

Interferons (IFN) are members of the cytokine family which are produced by specific cells from the immune system in response to pathogens, such as viruses. These proteins are often glycosylated prior to being released into the environment to accomplish their functions (DeMaeyer 1988). Some may have important regulatory roles on cellular processes and control cell division cycles as well as cellular growth (Pitha 2007).

Different types of interferons have been classified by their functions. The two major type I interferon groups are the IFN- α which are produced by leukocytes and the IFN- β which are produced by fibroblasts, both having mainly antiviral functions. The only type II interferon is IFN- γ and is only produced by lymphocytes and has more functions than the type I interferon. The main function of type I interferon is to inhibit virus replication in infected cells, while interferon type II will not only inhibit virus replication but can also stimulate antigen-presenting cell Major Histocompatibility Complex (MHC) expression on macrophages. There are 14 human genes coding for the corresponding subtypes of interferon alpha; each having close to 80% homology between the amino acid sequences (Pestka 2004, Pitha 2007).

1.1.1.

Interferon alpha-2a

Interferon alpha-2 has three different subtypes differing by only a few amino acid residues. Interferon alpha-2a (IFN- α 2a) differs only by a lysine replaced with an arginine at position 23 when compared to the IFN- α 2b subtype. IFN- α 2b possesses a histidine at

position 34 which is replaced by an arginine in the IFN- α 2c subtype (Günther 1991). IFN- α 2a binds to its receptor on the target cell membrane being the type I receptor complexes consisting of two different subunits, IFN- α / β R and IFN- α R1 (Novick 1994). IFN- α / β R can bind both IFN- α and IFN- β subtypes and is probably the major ligand-binding component of the receptor complex. The other subunit of the complex is IFN- α R1 which can only bind to IFN- α 8. This subunit, although not essential, helps increase the affinity of the receptor complex for IFN by tenfold, making it essential for signal transduction (Colamonici 1993). Binding of interferon causes a dimerization of the IFN- α receptor, which is an important step in transmembrane signaling similar to what is observed in other characterized cytokine receptor systems (Wells, 1994). Interferon alpha-2a plays an important role in the immune response against both cancerous and viral diseases. Even though all functions are not known, it is always present in small concentrations in blood and can trigger the immune system if given from an external source. The intra-cellular cascade triggered by binding of the cellular complex activates further production of chemokines or activation/suppression of certain cellular pathways leading to cell death of cancerous cells (Wu 2005).

The structure and dynamics of interferon alpha-2a has been previously determined in solution NMR spectroscopy by Klaus and coworkers. This interferon is composed of 5 alpha helices (red) labelled from A to E. From Figure 1, we notice that four of these helices align to form a helix bundle (A, B, C and E) to make up the stable core of the protein. Also note the presence of two disulfide bonds between both Cys1-Cys98 and Cys29-Cys138 (yellow). There are four regions with a high degree of flexibility (white

and green), which are the C-terminal loop (Cys1-Ser8), N-terminal loop (Ser160-Glu165) and the two amino acid stretches of Gly44-Ala50 and Ile100-Lys112 (Klaus *et al.* 1997). This creates a stable and rigid scaffold for that protein allowing it to be stable outside the cells for it to properly reach its target cells.

Figure 1: A: Three-dimensional structure of interferon alpha-2a, obtained by Klaus *et al.* 1997 (PDB: 1ITF). There are five helices (red), two disulfide bonds (yellow) and loops (white and green). The green loop represents the region natural glycosylation occurs on. B: 3D structure of interferon alpha-2a obtained by Klaus *et al.* 1997, showing all 20 structures obtained by NMR.

In vivo, interferon alpha-2a is *O*-glycosylated at threonine-106, as are the 2b and 2c subtypes, from mass spectrometry analysis done on a mixture of these interferon extracted from serum (Adolf 1991). While the glycan is not required for biological function, it may help to stabilize the protein by interacting with the surrounding amino

acid side-chains or even creating a pocket increasing interaction with water molecules.

By looking at Figure 1B, we can see that although the helices of the structure are well defined, the loops, including the one where the glycosylation occurs is not well defined for these 20 NMR structures. In this figure, we can also see treonine-106, which shows a very wide area of position, depending of the structure.

1.2. Glycobiology

Glycobiology is an extension of molecular biology, which incorporates the effects and functions of carbohydrate chains covalently attached to a variety of macromolecules in the molecular workings of the cell. It also represents the study of the structure, biological function and biogenesis of the wide range of saccharide chains available in nature. Carbohydrate chains are created by biosynthesis from a complex array of sugar units. These sugar chains, called oligosaccharides or glycans, can either be linked onto glycoproteins, or be freely available in the cellular matrix. Glycoproteins are synthesized by the cell by a process called glycosylation by the action of a wide variety of selective enzymes named glycosyltransferases (Varki 2009). Glycosylation is an important post-translational modification of proteins. Many of these glycosylated proteins act as important intracellular and intercellular communicators and have a broad array of important functions in developmental or functional areas of the cell (Houde 2010). Glycans serve different purposes such as inducing proper protein folding, trafficking of the glycoprotein within the cell and attributing the correct protein function (Varki 1993). Glycans may modulate in some cases either the protein stability, protein folding process, or even help keep the protein in an active form. During the folding process of certain

proteins, stable intermediaries may form causing an improper protein fold which is thermodynamically stable and unable to be properly refolded (Ohgushi 1983). In order to bypass these structurally stable folds, glycans are added to certain amino acids, allowing the proper protein folding to occur. Of course some proteins called chaperones usually help in the folding process, however glycans can also perform a similar role in certain cases. An example of this would be the addition of high-mannose N-glycans on proteins, which act as chaperones and help properly fold the protein (Jitsuhara 2002), in certain cases hiding hydrophobic residues which would be sent forced within the core of the protein causing an improper fold to result if the glycan is absent. This also means that glycans can have an important role in modifying the structure and the dynamics of the protein (Varki 2009). In other cases, the glycan chain may cause a modification in the bioactivity or ligand interaction of the protein, especially if the glycan itself binds to the receptor and ligand, in the case of lectins (Lis 1993). Even if the glycans may not always affect the structure and function of the proteins, some may increase the half life of the protein in the cellular or outer cellular environment, by blocking parts of the protein which are prone to degradation. In order to produce glycoproteins, cells must build particular glycan chains from the array of saccharides available. The chain can either be produced prior to the linkage to the protein or it may be built directly onto the protein.

1.2.1.

Biological and biophysical properties due to glycosylation

The addition of the glycan chain on the protein modulates several biological properties, depending on the glycosylation site. Not only do they sometimes act in the recognition

process and binding to the receptor, but they can also have an important role in protein folding. This is especially true for multiple N-glycosylated proteins. Although in many proteins the hydrophobic core of the protein is able to form instantly, in some cases, it is essential for a glycan to be added to some amino acid side-chains, hiding them from the environment and allowing the core of the protein to properly form (Molinari, 2007). Finally, the presence of glycans can induce structural and dynamic changes to the glycosylated protein (Mandal, 2001). These changes are due to local environment changes from the presence of the glycan within close proximity of other amino acids. This allows for low affinity interactions, whether ionic or hydrogen bonds with for example the acetyl group or the hydroxyl groups on the sugar moieties, to occur and modify the local structure and dynamics. When adding these glycan chains to the proteins, the newly formed glycoproteins may have altered biophysical properties (Varki 1993). For example, the addition of the glycan may increase or decrease the solubility of the protein by creating a water shell around the glycan. In some cases the addition of the glycan will serve to hide certain amino acid sequences which are more prone to degradation, which in turn will increase the half life within the body of the protein. The presence of the N-acetyl and hydroxyl, as well as, other chemical groups on the sugar moieties may increase or decrease the hydrophobic or hydrophilic interactions between both the protein and the water medium surrounding it (Varki 2009).

1.2.1.1

Molecular recognition

The principle of glycan recognition lies on the fact that specific glycan chains will bind to different glycan-binding proteins (GBP), such as lectins, or have different binding affinities (Lis 1993). The binding site will only allow for a certain variety of glycan chains, depending on their structure, alternative saccharides or even the presence of certain modifications done on the sugar units. Due to the heterogeneity, discussed further, of the glycan chains present on glycoproteins, the variety of glycans may bind differently to the GBP in terms of their affinity and possibly binding site. These differences in binding affinity are affected by factors such as the conditions at which the glycan will bind to the site.

In some cases, glycans may provide steric hindrance by blocking either binding or access to a catalytic site. For example, RNase B, which is simply glycosylated RNase A, is catalytically inactive because its glycan blocks the active site. Upon removal of the glycan RNA hydrolysis can occur (Williams 1987).

1.2.1.2.

Glycans immunogenicity

Since many chemokines are glycoproteins, this implies an important role of glycan in the field of immunology. Some glycans play a role in the immune response from T cells by contributing either to the structure or profile of the epitope binding to the MHC. Glycans often do not contribute to the primary biological function, such as receptor binding but

may confer higher half lives. Absence of these may expose epitopes that will elicit the generation of antibodies. For example GM-CSF is heavily O-glycosylated on the N-terminal region and has two N-glycans at residues N27 and N37. Recombinant GM-CSF produced in *E. coli* that is used as therapeutics elicits the production of non-neutralizing antibodies at the unstructured N-terminal and at the globular domain near the two glycosylation sites. (Ragnhammar 1996). In general, antibodies will recognize specific glycan chains which are found on pathogenic bodies, such as viruses. This can also lead to a variety of auto-immune diseases due to the self-recognition of certain self-belonging glycan chains, if the glycans shared similarities with a recently encountered pathogen. In the case of interferon alpha-2a, the glycosylation does not play a role in recognition of the binding site, but may increase its half-life in blood (Bailon 2001).

1.2.2.

Glycosylation

Glycoproteins are created by the fusion of a protein with glycan chains in a process called glycosylation. It is an important post-translational modification which can occur multiple times on the same protein before the glycan chains are mature, thus allowing the glycoprotein to carry out its function. Various factors such as the enzymatic activity of the transferase and the glycans available in proximity of the reaction play a crucial role in the modulation of the glycan chain elongation process. In eukaryotes, protein glycosylation takes place via two processes, the N-glycosylation and O-glycosylation. A brief description of both of these processes will be given as well as examples of the glycans produced.

1.2.2.1.

Saccharides are the building blocks of glycans

The basic units of the glycan chains are monosaccharides. Monosaccharides are first activated at their anomeric center by the addition of an activated phosphonucleotide (UDP or CMP activated depending on the saccharide) prior to linking to the acceptor by the action of a glycosyltransferase. These linkages are called glycosidic bonds and link the anomeric carbon of a sugar moiety to either an amino acid, initiating the glycan chain, or to a saccharide on the glycan chain. There are two possible orientations for anomeric carbons; they are either forming α - or β -glycosidic bonds. The differences between them are based on the relative stereochemistry of the anomeric position and the stereocentre furthest from C1. In the case of the axial orientation, where we find an α -glycosidic bond, the bond is pointing under the plane, while when the anomeric position is equatorially oriented, then we find a β -glycosidic bond, which means the glycosidic bond is pointing above the plane. This can be seen in Figure 2.A, where three sugar moieties are shown in both their axial and equatorial glycosidic bond orientations. The most common glycan chain for interferon alpha-2a is shown in Figure 2.B. Considering that many possible glycosidic linkage can be formed, this will generate heterogeneity of the glycan chain population, as we will be discussed later.

Figure 2: Example glycosidic bond orientations (red) for the three possible sugars moieties (Neu5Ac, Gal and GalNAc) for the interferon alpha-2a glycan (A). In A, all three glycans are shown at first in their axial form, which can be determined by the orientation of the anomeric carbons glycosidic bond which points below the plane. Afterwards, the three glycans are shown in their equatorial forms, where the anomeric carbon glycosidic bond is pointing above the plane of the glycan structure. The proper glycosidic bond orientations (red) for the glycan are showed in the structure (B).

1.2.2.2.

Figure 3: Diagram of the N-glycosylation process showing translation of mRNA into a polypeptide by the ribosome, followed by nascent glycosylation with a preformed N-glycan chain to aid in the proper folding in the ER lumen and by a series of modifications done on the chain in the Golgi apparatus. The lower section shows the evolution of the N-glycan structure during its attachment to the nascent protein, protein folding and trafficking.

The first step during N-glycosylation requires the assembly of the oligosaccharide precursor. This is biosynthesized within the cytosol by sequentially adding sugars on the dolichol-phosphate (Dol-P) lipid. This precursor chain is a complex glycan construct attached to the oligosaccharyltransferase (OST) complex and continues to grow in the luminal compartment of the endoplasmic reticulum (ER) until it is required for glycosylation. Glycosylation with the glycan precursor is done via the OST complex onto nascent proteins (Figure 3A), typically targeting asparagine amide side-chains (Figure 3B). As the nascent protein exits the ribosome, the oligosaccharide precursor is enzymatically attached to the nitrogen of the amide side-chain group of an Asn or Gln. The minimal sequence required for a particular region to be a candidate for N-glycosylation is Asn-X-Ser/Thr (Medzihradszky 2008), where X represents any amino acid. Occasionally, although rare, N-glycosylation may also occur at the Asn-X-Cys sequence. Once the oligosaccharide is transferred, a series of enzymatic reactions occur which are essential for both the proper protein folding process with specific chaperones and lysosomal trafficking of the glycosylated protein (Figure 3C); the latter is modulated by glucosidase I and II (Glcase I, II). These two enzymes ensure that the protein is properly folded by trimming off glucose units from the glycan, allowing the glycan to be trafficked properly. If the structure is not properly folded, the glucosidase will not trim the glucose from the glycan chain and the glycoprotein will return to the chaperone for proper folding. The N-glycan goes through a process which trims off additional glucose or mannose units from the chain branches before being released from the ER, leading to a diversification of the glycan chains (Figure 3D). In certain cases, glycosidase reactions in the ER and Golgi create important mannose chains on the N-glycan; these are known as

high-mannose subtype N-glycans usually found on lysosomal hydrolases (Figure 3D¹).

For other proteins, a series of enzymatic reactions will trim and modify the chain until the glycoprotein is released or secreted and ready to perform its function (Figure 3D²).

1.2.2.3.

O-glycosylation mechanism

In contrast to N-glycosylation, O-glycosylation proceeds by assembling the glycan one unit at a time on the folded protein. Figure 4 shows that the glycan is attached onto the protein after the protein is properly folded and has entered the subcellular compartments, which include the ER and the Golgi apparatus. The glycan chains are built by attaching one sugar unit at a time with specific glycotransferases onto the chain. When the folded protein arrives in the Golgi, a single sugar moiety is linked to the protein on the hydroxyl group of either a serine or a threonine residue. Each O-glycan are less branched than N-glycans, however, there can be a large quantity of O-glycan on the same protein (Toribara 1991). This can be seen with many mucin-type glycoproteins which can have dozens of O-glycans over regions of the protein.

Figure 4: Diagram of the O-glycosylation process showing translation of mRNA into the amino acids by ribosome, followed by the proper protein fold in the ER lumen and finally serial glycosylation of select serine and threonine residues once in the Golgi apparatus.

The first glycosylation step is carried out by a N-acetyl-galactosamine (GalNAc) transferase which takes a nucleotide activated GalNAc and reacts to form an alpha linkage with the serine or threonine residues, creating a $\text{GalNAc}\alpha\text{-Ser/Thr}$ linkage, as shown in Figure 5. Chain elongation of the GalNAc moiety linked to the protein is done by a glycosyltransferase which adds a nucleotide activated galactose in a $\beta 1-3$ linkage to the $\text{GalNAc}\alpha\text{-Thr}$ by Core 1 GalT. For interferon alpha-2a, the $\text{Gal}\beta\text{-3GalNAc}\alpha\text{-Thr}$ glycan chain can be extended by adding two N-acetylneuraminic acid units sequentially. The first reaction is done by the enzyme ST3Gal I linking in $\alpha 2-3$ the Neu5Ac unit on the galactose, while the enzyme ST6GalNAc links in $\alpha 2-6$ a second Neu5Ac on the GalNAc.

Figure 5: Glycosylation process leading to the native interferon alpha-2a (A) O-glycosylation under the Core 1 glycan structure (B and C) including the enzymes responsible for the glycan chain elongation. The most common interferon alpha-2a glycan moiety linked to Thr-106 is encased in the red box (labelled D). Based upon the book “Essential of Glycobiology, Second Edition, 2009, by Varki and others”, Chapter 9, Figure 9.3. The structure for glycoforms A to D are shown.

1.2.3.

Heterogeneity

Protein glycosylation results from a series of enzymatic reactions that are modulated by various factors and often result in a range of structure variations for the glycan chain at any glycosylation site. The main factors modulating glycosylation are the availability of the nucleotide activated monosaccharide, the specificity of the glycosyltransferases with the amino acid and glycan sequence and their enzymatic activity. A single glycoprotein can have a multitude of different glycoforms, which may take part in different functions, such as the differential regulation of CD43, where two glycoforms are subject to regulation depending on the active cells function (Ellies 1994). This heterogeneity may either be an effect of the timeframe on which the sequential glycosylation process is based, or simply due to precise functions which require the presence of a specific glycan chain variation. However, this aspect creates great challenges when analyzing the structure of these glycoproteins. In the case of interferon alpha-2, this causes different glycan chains to be present on the native glycoprotein produced *in vivo* which is observed when isolating the proteins from serum.

1.2.4.

Effect on structure and dynamics

Previous studies (Mandal 2001, Krapp 2003) have reported that the presence of glycan chains attached to the protein may affect the structure and dynamics of the protein, depending mostly on the nature of the glycan chain. A study on the effect of a small glycan chain on an MHC class I glycoprotein by Mandal (2001), showed no important

effect on the overall structure of the protein although local changes to structure and dynamics depended on the nature of the glycan. It was also shown that there was an increase of flexibility that had occurred for the loop region near the glycosylated residue. In other instances, presence of a glycan near a binding site reduced the binding affinity without affecting the structure or dynamics. This occurs in the case of an anti-p145 antibody fragment which had a reduced binding activity to its antigen (Kubetzko 2005). For interferon alpha-2a, this glycosylation is located quite far from the receptor binding surfaces; no significant changes to the activity of the protein are observed (Itri 1987). In fact, non-glycosylated interferon alpha-2a shows similar antiviral and antitumor activity as naturally glycosylated interferon alpha-2a (Trown 1986).

1.3.

Interferon used as a biotherapeutic

As knowledge of the biological function of interferon was developing, these cytokines were identified as potential therapeutic agents. They have been found to act on the immune system by reducing tumor growth and viral infections. Since the mid 1990's, a number of products have been marketed, where Roferon-A and Intron-A are the most important ones. Both result from recombinant DNA technology and are produced in *E.coli*. Roferon-A[®] is a protein produced by a bioengineered *E.coli* bacterium expressing the human interferon alpha-2a gene. It has the same activity and structure as the native interferon alpha-2a produced by humans; however it lacks the glycosylation found on Thr106 (Adolf 1991). One of the most common glycan chains found is Neu5Ac-Gal-

GalNAc, although shorter chain variants can be found due to the heterogeneity of the chains.

Pegylation is a synthetic route to mimic certain biophysical properties exhibited by glycan chains; it may confer a longer bioavailability to the protein within the blood. One of these biotherapeutic forms of interferon alpha-2a which has been pegylated is called Pegasys. The pegylation process adds to certain amino acids, such as lysines, the polyethylene glycol (PEG) chains onto the protein surface. Due to the size of these chains, the binding site of interferon may be blocked in some cases. Pegylation itself does not have any anti-viral or anti-tumoral effects. When comparing the efficacy between Roferon A and its pegylated form against hepatitis B (Yu 2010), it was found that the pegylated form had similar viral inhibition. Pegylation increased the half-life of the protein in the blood allowing a lower dosage of interferon to be administered, thus improving safety and efficacy during clinical treatments and trials. During a clinical trial by a Chinese group (Zhang 2010), treatment with pegylated interferon alpha-2a on haemophilic patients with both HIV and Hepatitis C was found to cause reduction in the antiviral response but also reduced platelet production although platelet levels were still within the normal range. Thus pegylation may improve biophysical properties probably in a similar fashion to O-glycosylation.

1.3.1.

Hypothesis

We are proposing that the overall structure of interferon alpha-2a will not be significantly affected by the O-glycosylation, although local changes to the structure may occur from the interactions of the residues with the GalNAc unit. In addition, the backbone dynamics of the glycosylated protein may be altered by the presence of the glycan.

1.3.2.

Strategy

Understanding the effects of glycosylation on the protein can be achieved using NMR spectroscopy. This technique can measure a variety of parameters that are affected by the structure and the motion of the protein. The approach used consists at preparing a doubly labelled interferon with a bioengineered *E.coli* that use ^{13}C -glucose and ^{15}N -ammonium chloride as sole carbon and nitrogen sources. ^{13}C , ^{15}N -interferon alpha-2a is then O-glycosylated *in vitro* by adding successively the unlabelled oligosaccharide units one at a time. This strategy produces a ^{13}C , ^{15}N -labelled protein with a covalently bound glycan comprised of NMR-active ^{13}C and ^{15}N . This will allow NMR to measure properties that are either protein or glycan specific, and glycan-protein interactions such as glycan-protein internuclear distances for structure calculation. A protein NMR sample must meet some conditions to yield reasonable spectra: the protein must be soluble and monodispersed at higher concentration (0.2-1.0 mM). The protein must also remain stable and properly folded at temperatures around 25 °C to 37 °C for extended period of time,

from days to months. The sample is kept at pH 3.5 to respect conditions of sample from Klaus *et al.*, resulting in the protonation of the amide side chains.

1.4. NMR background on techniques used

Nuclear magnetic resonance (NMR) spectroscopy is a well established technique to determine the three dimensional structure of biomolecules in solution. Compared to optical spectroscopy techniques, such as UV-vis and infrared spectroscopies, NMR is not very sensitive. However, it has the ability to select particular types of information and spread the latter in multiple dimensions. In addition, NMR can study molecular motions over a wide range of timescale (picoseconds to seconds). An important feature of NMR is the ability to select a particular nucleus and gain information on its local chemical environment. Although NMR can be performed with natural abundance ^{13}C and ^{15}N , enrichment of these isotopes strongly increases the signal and can give better resolution for the spectrum. However, NMR is less sensitive than other techniques, such as mass spectrometry, and hence requires a higher concentration of sample. NMR experiments are able to provide various type of data, such as the chemical shifts of specific atoms, the scalar couplings (J coupling observed between two nuclei), the relaxation times and nuclear Overhauser effects.

1.4.1.

Basic principles of NMR

NMR spectroscopy measures the frequency of precession of active nuclei in the presence of a strong magnetic field following excitation from a radio-frequency pulse. In order to

be NMR active, a nucleus must have a non-zero spin angular momentum. Nuclei such as ^1H , ^{13}C , ^{15}N and ^{31}P have a spin of $\frac{1}{2}$ and are commonly found in biomolecules. While ^1H and ^{31}P are present at nearly 100%, ^{13}C and ^{15}N have a lower natural abundance of 1% and 0.3%, respectively. In many protein studies, these must be enriched by the use of recombinant technology in *E.coli* using defined growth media to produce uniformly labelled ^{13}C - ^{15}N -labelled proteins. In this section, a brief description of the principle of NMR spectroscopy will be given.

In the presence of a strong magnetic field (B_0), nuclear spins align either parallel (low energy state α , spin $\frac{1}{2}$) or anti-parallel (high energy state β , spin $-\frac{1}{2}$) to the field. This is the Zeeman interaction. Each nucleus will possess a component that is aligned with the magnetic field (the Z axis), while also having a component in the transverse plane (X Y axis). The nuclei will precess about the Z-axis due to the torque produced by interaction of the angular moment and the magnetic field. The rate at which the nuclei precess is called the Larmor frequency (ω_0) and depends on the magnetic field strength experienced and the gyromagnetic ratio (γ) for a particular type of spin as defined by equation 1:

$$\text{Equation 1: } \omega = \gamma B_0$$

The sum of all Z components, i.e. the sum of the parallel minus the sum of antiparallel, forms the bulk magnetization (M_0), as seen in Figure 6. On the other hand, the sum of all transverse components, i.e. individual XY projection, will be zero since the transverse components are distributed randomly in the XY plane. Since detection occurs in the

transverse plane and no transverse component is present, no NMR signal can be observed.

Figure 6: Effect of the applied magnetic field on the orientation of the nuclei spins, splitting into two populations for the NMR active nuclei (^1H , ^{13}C , ^{15}N , ^{31}P). Due to the precessing movement of the spin which are not oriented identically, no transverse component is detected.

A simple NMR experiment consists of a relaxation delay followed by a single radio-frequency pulse and a detection period. The radio-frequency pulse (B_1) is applied perpendicularly to the bulk magnetization, e.g. the X axis. The Z component precesses about this B_1 field. Thus the effect of the pulse is to tilt the bulk magnetization on the transverse plane. This is illustrated in Figure 7. At the end of the pulse, the nuclei are on the transverse plane and precess about the magnetic field B_0 . All nuclei that oscillate at ω_0 do so coherently, thus giving a coherent signal. This precessing transverse magnetization induces an oscillating voltage in the instrument detection circuit that is called the field induction decay (FID). The FID is a spectrum of signal intensity versus time. A spectrum of intensity versus frequency is obtained using a Fourier transformation.

Figure 7: Effect of a perpendicular radio-frequency pulse on the Z component and its gradual shift toward the transverse plane.

1.4.1.1.

Relaxation and molecular motion

The NMR signal will not remain forever but it will return to its initial state due to relaxation processes. These are the longitudinal and transverse relaxations, which are also known as the spin-lattice and spin-spin relaxations, respectively.

Figure 8: Spin-lattice and spin-spin relaxation effects on the coherence of the signal and its transfer back toward the initial state parallel to the applied field, once the radio-frequency pulse is no longer applied.

The longitudinal relaxation is characterized by the spin lattice relaxation time constant T_1 . The rate R_1 (the reciprocal of T_1) corresponds to the speed at which the bulk magnetization returns to the Z axis as seen in Figure 8. The transverse relaxation is characterized by the relaxation time constant T_2 . The rate R_2 describes the speed at which the transverse magnetization, i.e. the coherence is lost. In other words, it describes the

gradual dephasing of the signal until no net magnetization is observable in the transverse plane.

1.4.1.2

Protein dynamics

Proteins are not static molecules; they are in constant motion. A range of motions are taking place over a wide timescale, the fastest being vibration of chemical bonds up to the slowest conformational change such as unfolding. During any motion, a spin moves in space and experiences different magnetic environments. In addition, the spin of a nucleus simulated by a small magnet will create, by its motion, magnetic fluctuations. These will affect the NMR signal in various ways because the frequency is directly related to the magnetic field experienced at a particular site.

From the protein tumbling to the inter-atomic interactions, these dynamic interactions strongly affect the results obtained during NMR experiments. At a smaller scale, modification of the local environment of each residue can cause the dynamics of the regions to be changed, since the side-chains and backbone of local residues are in constant interaction with each other. An example of this would be the reduction of freedom of movement of the residues when a fairly dynamic binding site binds to its ligand, causing a very restrictive positioning of the side-chains. The same can occur in the case of glycoproteins, where the addition of the glycan chains onto the protein would cause a reduction of movement in the region interacting with the glycan, since the glycan creates and potentially destroys certain interactions with the surrounding amino acids

(Houde 2010). All these variations would be observable with NMR experiments, due to changes in the relaxation times (T_1 and T_2) as well as potential changes in the dynamic Heteronuclear NOE patterns.

As the molecular weight increases, broadening of the line width occurs in the spectra due mainly to the shorter T_2 relaxation times. Since these larger proteins often tumble at a slower rate, their orientation within the field also becomes a factor. Chemical-shift anisotropy occurs because of the dependence of the orientation of the molecules in the external magnetic field. This is an important aspect to consider for aromatic amino acids, or even the amide linkage between the amino acids of the proteins. Since smaller molecules in solution tend to tumble faster, this causes no orientation preferences of the protein within the magnetic field and sharper peaks are obtained. On the contrary, larger molecules which tumble slowly will cause an unequal distribution of the different orientations of the chemical shifts, causing the chemical shift peaks to broaden.

1.4.1.2.1.

Time-scale of high resolution NMR

Figure 9: Time scales of High-resolution NMR of different components which can be measured at precise frequency ranges.

Various movements in a protein will occur in completely different time-scales. High-resolution NMR is able to differentiate between these time-scales and give information specific to each of these movements or interactions as shown in Figure 9. Very slow movement, which occur by a time-scale of seconds, could arise from the folding of the protein, which is relatively slow compared to other movements, and requires real time experiments in order to be able to examine such movements. Domain motions, which represent movement of multiple proteins interacting together, are on a millisecond scale. The secondary structure motions are on a microsecond scale while the peptide bond motions are in a nanosecond scale. The fastest movement present in the protein would be the local vibrations of each atom which would be on a picoseconds scale. A variety of NMR experiments have been developed to measure relaxation parameters thus allowing the description of protein dynamics

1.4.2.

Multidimensional NMR

NMR spectroscopy is unique in its ability to transfer polarization or coherence of a group of nuclei to another via a coupling mechanism. When ^{13}C , ^{15}N -labelled proteins are used a network of coupling constants can be used (see Figure 10) that will allow the correlation of all backbone and side-chain nuclei.

Figure 10: Diagram of a dipeptide backbone, with the label for the atoms positions and the possible coupling constant between each atom. J couplings values for $^1J_{\text{NH}} \sim 91$ Hz, $^1J_{\text{NC}_\alpha} \sim 7-11$ Hz, $^2J_{\text{NC}_\alpha} \sim 4-9$ Hz, $^1J_{\text{NCO}} \sim 15$ Hz, $^1J_{\text{C}_\alpha\text{CO}} \sim 55$ Hz, $^1J_{\text{CH}}$ ($^1J_{\text{C}_\alpha\text{H}_\alpha}$, $^1J_{\text{C}_\beta\text{H}_\beta}$) ~ 140 Hz, $^1J_{\text{C}_\alpha\text{C}_\beta} \sim 35$ Hz.

The simplest experiment that illustrates this ability is the correlation spectroscopy (COSY) experiment. If spin A is coupled (scalar coupling) with spin B a COSY experiment can transfer the coherence of spin A to spin B. If appropriate data collection and processing are carried out, a new resonance carrying both frequencies can be measured. This new resonance is referred as a correlation because it correlates spin B with spin A. In order to achieve this in a protein, it is desirable that all atoms be ^{13}C and ^{15}N labelled. In addition, these nuclei have large spectral windows, thus providing a significant increase in resolution over 1D experiments.

1.4.3.

^1H - ^{15}N Heteronuclear Single Quantum Correlation (HSQC)

One of the most utilized two-dimensional NMR experiments is the Heteronuclear Single Quantum Correlation experiment, which correlates protons covalently bound to nitrogens. The HSQC spectrum of a protein shows a single peak for each backbone amide pair (H^{N} -N) of each residue. In addition, all amides and amines H-N from the side-chains of asparagine, glutamine, lysine, arginine, histidine and tryptophan residues will also show correlations on the spectrum. The spread of resonances of the HSQC spectrum gives a direct indication whether the protein is folded or in a random coil conformation. If the protein were to be completely unfolded or be in a state called molten globule, which is a partially folded state often an intermediary fold, then the distribution of the chemical shifts would be very poor (Ohgushi 1983). This can be rationalized by the fact that all backbone amides experience a similar magnetic environment, mainly a solvent environment. On the other end, backbone amides in a folded protein are not all in the

same environment. Some are deep into the hydrophobic core; some are involved in hydrogen bonding within secondary structure elements such as helices and sheets while others are on the surface and exposed to the solvent. For these reason, an HSQC spectrum is often referred to the fingerprint spectrum of the protein.

1.4.4.

Extension to the third dimension, the 'HNCO' experiment

In most cases, solving large protein structures requires more than simple 2D experiments since severe peak overlaps often occur simply by similar residues having similar chemical environments resulting in close chemical shifts that are difficult to distinguish. For this reason, a variety of three-dimensional experiments are used to measure all chemical shifts of the backbone and side-chain nuclei of larger protein. Although there may still be resonance overlap, the combination of all these experiments provides the necessary resolution to perform a complete assignment. The simplest of these experiments is the HNCO, which in addition to the H^N and N dimensions found in the HSQC, uses a third carbon dimension. In the HNCO experiment, the carbon dimension separates the resonances by their carbonyl chemical shifts. This experiment provides a higher resolution than the HSQC, as well as correlating atoms between two residues, being the correlation of a residue to carbonyl of the preceding residue. This is required to identify which resonances belong to which residue in the amino acid primary sequence.

1.4.5.

Resonance assignment: First step toward structure determination

Structure determination of a protein requires multiple steps. Initially, a complete assignment of all the chemical shifts is required for all atoms. The procedure starts with the sequential assignment. This consists at assigning all backbone nuclei to the amino acids in the primary sequence. Then experiments that are specific for side-chain resonances are collected and analyzed in order to assign the remaining nuclei. Once the chemical shifts of all nuclei are known, measurement of inter-proton distances can be performed via NOESY-type experiments.

1.4.5.1

Sequential assignment

Two key experiments are used to perform this step: the HNCACB and the CBCA(CO)NH. In an ideal world, these two would suffice to correlate backbone nuclei but in real life, data from other experiments are often needed to complete the task. However, description of these two experiments will suffice to illustrate the principle of sequential assignment.

1.4.5.2.

HNCACB experiment

The HNCACB is one of the key experiments that are used to determine the sequential assignment of the backbone atoms (Grzesiek¹ 1992). It correlates the amide ^1H and ^{15}N pair with its side-chain $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ (residue i) as well as those of the preceding residues

(residue $i-1$) (Figure 11). In the experiment, the magnetization originates on $^1\text{H}^{\text{N}}$ and is transferred to the ^{15}N and subsequently to $^{13}\text{C}_\alpha$ then $^{13}\text{C}_\beta$, where both will evolve simultaneously, followed by the transfer from $^{13}\text{C}_\beta$ to $^{13}\text{C}_\alpha$, causing the inverted phase between both. On the spectrum, the C_α and C_β correlation will have opposite signs from each other due to the way the magnetization was transferred. This experiment is therefore useful for the sequential assignment of residues, as well as obtaining the chemical shifts of these carbon atoms. In principle, this experiment could provide all the data needed to perform sequential assignment. However, other experiments are also required since the HNCACB experiment sometimes suffers from a low sensitivity such that weaker inter-residue correlations are not detected, as well as being subject to resonance overlap. Therefore, the HNCACB data must be complemented by other experiments.

Figure 11: HNCACB magnetization transfer diagram for the experiment shows that the magnetic transfer is initiated at the amide proton and ends on that nuclei for detection. There are two paths the magnetization can be transferred, because of the presence of a similar coupling constant between the intra-residue and inter-residue (red) nitrogen-carbon alpha coupling (blue).

1.4.5.3.

CBCA(CO)NH experiment

The CBCA(CO)NH complements the HNCACB because it exclusively correlates the amide of one residue with the previous residue $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ shifts (Grzesiek² 1992). In the experiment, the magnetization originates on the $^1\text{H}_\alpha$ and $^1\text{H}_\beta$ and is transferred to their attached carbons, $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$, respectively, where both will evolve simultaneously, and remain on the same phase. The magnetization is then transferred to the amide nitrogen via carbonyl (CO) where no chemical evolution occurs and finally to $^1\text{H}^{\text{N}}$ of residue *i* (Figure 12). When compared to the HNCACB experiment, this experiment is another example of selection: only the $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ from previous residues are observed. The shorter transfer path in the CBCA(CO)NH gives more intense inter-residues correlations. The latter are usually weaker on the HNCACB, and sometimes they are not observed. Therefore the CBCA(CO)NH experiment is a useful complement of the former experiment.

Figure 12: CBCA(CO)NH magnetization transfer diagram for the experiment shows that the magnetic transfer initiates at both $^1H_\alpha$ and $^1H_\beta$, while ending at the amide proton.

1.5. Sequential assignment: the strategy

Three-dimensional spectra such as the HNCACB are best analyzed with the help of appropriate software such as NMRView (Johnson 1994). The software reduces the dimensionality from three to two-dimensions by generating a strip plot. The idea is to plot a carbon sub-spectrum centered at the amide proton-nitrogen chemical shifts. This is shown in Figure 13 where each layer represents a specific ^{15}N chemical shift corresponding to a backbone amide and the red band represents a selected range of chemical shifts from the $^1H^N$ and ^{13}C dimensions on which the chemical shifts of the proton and carbons can be extracted and associated with that particular amide. Because all datasets collected for sequential assignment correlate various carbon chemical shifts to

the amide pair, strip plots from different experiments can be combined to facilitate data analysis. Examples of strip plots will be shown in the results section.

Figure 13: Three-dimensional display of the ^{15}N dimension separated into different layers corresponding to a specific amide chemical shift and the ^{13}C - ^1H chemical shift range extracted from one of these layers.

The assignment procedure is comparable to assembling a jigsaw puzzle: one must start with easily recognizable features. Some residues have unique chemical shifts values, such as the absence of a C_β for glycine. Other residues which can be easily identified are serine/threonine with their C_β located in the 60-70 ppm range and alanine with a C_β at around 19 ppm. Therefore the analysis starts with these residues and the neighbours are identified by using the correlations with the previous residue in the chain.

1.5.1.

Assignment of side-chain resonances

Once the assignment of the backbone is complete, assignment of the aliphatic side-chain atoms requires several other experiments to be used such as HCCH-TOCSY and HCCH-COSY. These correlate protons to their attached carbons and to their neighbour carbons. Since, alpha and beta carbon chemical shifts are known from the sequential assignment, the side-chain experiments provide the remaining correlation to measure and assign all

remaining chemical shifts. This is the most tedious, time-consuming and important step because side-chain protons provide long-range internuclear distances. These are essential to obtain a properly folded protein. In fact, hydrophobic interactions that stabilize the core of most globular proteins such as interferons involve amino acid side-chains. Once most chemical shifts (i.e. >80%) have been assigned, structural data can be measured, namely torsion angles and internuclear distances.

1.5.2

Filtered COSY spectra

The present study deals with a doubly labelled protein that is glycosylated with a non-labelled glycan. Therefore, the above assignment procedure will not provide any chemical shifts for the saccharide moiety. Here, the use of use of filtered experiments comes to the rescue. These can isolate either the protein or the glycan signals by selecting the signals of protons attached on heteronuclei, such as ^{13}C and ^{15}N . Thus experiments such as the filtered COSY are used to filter out all proton signals from the protein while observing only the protons of the glycans. This approach is especially useful to selectively measure glycan-protein interproton distances in filtered-NOESY experiments via the measurement of structurally-dependent parameters. These can be observed when dynamic and structural changes occur by the presence of the glycan. Since none of the resonances from amino acids will appear on the filtered spectrum, the chemical shift of the glycan can be assigned and the chemical shifts obtained can provide valuable structure information.

1.5.3.

Secondary structure determination with chemical shift assignment

Chemical shifts of backbone atoms are influenced by the local environment. It is possible to predict the secondary structure of the protein once the sequential assignment is obtained. By comparing the chemical shifts obtained for an amino acid in the protein sequence to those of the same amino acid preceded in a random coil (unstructured) values by an alanine residue (Wishart 1995), a pattern can be observed for different atoms of the backbone. For example, in an alpha helix, all C_α and CO atoms would have chemical shifts which are higher than the random coil values, while the C_β and H_α would have lower chemical shifts than the random coil values. In addition, alpha helices and beta sheets have different atom orientations and torsion angles, which cause the chemical shifts to vary in the opposite way. In a beta sheet, all C_α and CO atoms would have chemical shifts lower than random coil values, while the C_β and H_α would have higher chemical shifts than random coil values. This provides an easy way to quickly predict secondary structure elements.

1.5.4.

Backbone torsion angles

The backbone of the protein is composed of a series of unbendable planes (the amide group) linked by torsion points that can be rotated. These torsion points will vary the angle between the two planes, causing the backbone of the protein to adopt different conformation. The peptide bond linking two amino acids called omega (ω), cannot be rotated since the plane of the peptide bonds restricts this angle to 180° , unlike the torsion

angles surrounding C_α , which are less restricted. These angles, which are shown in Figure 14 named phi (ϕ) defined by the rotation of the C_α -N point and psi (ψ) defined by the rotation of the C_α -C, are found along the backbone of the protein. Secondary structures such as alpha helix and beta sheets have well-defined dihedral angles for both phi and psi.

Figure 14: Diagram showing the phi and psi backbone torsion angle of proteins.

1.5.4.1.

TALOS

As discussed previously, the secondary structures have defined torsion angle ranges and secondary structure elements affect the chemical shifts of the backbone atoms. An empirical method named TALOS+, (Torsion Angle Likelihood Obtained from Shift and sequence similarity) has been developed to predict phi and psi torsion angles from a database of protein structures with the C_α , C_β , C, N, H_α , H^N chemical shifts. TALOS establishes an empirical relationship between the chemical shifts obtained and known protein torsion angles for sequences of three amino acids and chemical shifts (Shen 2009). The ten highest rated hits are then returned for each one of these triplets, as well as a Ramachandran plot showing the angle pairs compared to each others. Good predictions are determined by the consistency of the predictions; the closer the prediction angles are,

the better rated is the prediction. TALOS predicts torsion angles with a high consistency for around 90% of the residues, while often predicting a very small range of possible angles. This allows for a reliable prediction of the phi and psi angles for structure calculations.

1.5.5.

The Nuclear Overhauser effects and the measurement of internuclear distances

The Nuclear Overhauser effect (NOE) occurs due to the dipolar interactions of magnetic moments of different nuclear spins in close proximity to each other. These NOE contain important distance information, which can be extracted and used as distance constraints when calculating three-dimensional structures. Briefly, the NOE is a measure of the effect caused by the closely neighbouring spins on the relaxation of the observed spin. This effect, also called cross-relaxation, is proportional to the reciprocal of the internuclear distance to the sixth power. The internuclear distance relationship of the NOE intensity allows it to be used to obtain structure information. In practice, an NOE can be observed between two spins separated by 5 Å or less. These NOE give an upper limits of acceptable distance values, while the van der Waals atom radius gives the lower distance limits. Because of this strong dependence, a strong signal indicates relatively short distances between the protons (1.8 – 2.7 Å), while weaker NOE signals indicate a longer distance between protons (1.8 – 5.0 Å). Experiments of the type NOESY (NOE SpectroscopY) are used to measure NOE derived internuclear distances.

1.5.6.

Backbone motion via the $\{^1\text{H}\}$ - ^{15}N Heteronuclear NOE

The Heteronuclear NOE experiment, in its simplest form, revolves around the ^{15}N - $\{^1\text{H}\}$ steady-state NOE factor obtainable by doing two separate experiments. In one of these experiments, the NOE are built up while continuously irradiating the protons and therefore saturating their spins and varying the NOE signal. This signal reduction or increase depends on the NOE factor which affects the signal obtained during the experiment. The region of the protein with faster backbone motions will tend to give a negative signal in ^{15}N - $\{^1\text{H}\}$ experiments. For the other experiment, no irradiation takes place and maximum NOE signal is obtained. NOE from both experiments are then compared for each residue by dividing the NOE from the spectra which was submitted for a 3 second saturation versus the one where no saturation occurred. In the case of proteins, the most common atoms studied in this type of experiment are ^{15}N - $\{^1\text{H}\}$. For larger proteins, their structured regions will have almost no NOE factors hence the NMR signals will almost be identical on both experiments. The presence of nitrogen in the backbone allows a direct determination of slower or faster backbone motions.

The experiment can distinguish backbone motions that are faster or slower than 0.5 nanoseconds (Parella 1998). In figure 15, fast motions of the backbone are reflected as a negative ^{15}N - $\{^1\text{H}\}$ value, shown in blue, compared to the slower motions of the backbones which will have correlation times over 0.5 nanoseconds, shown in red. Since the heteronuclear NOE values are directly related to the movement of the protein backbone, comparison of all residues for a given protein in graphical form would allow a

pattern to be observed. Regions of low mobility are usually an indication of rigid structures which would have entire regions with values approaching 1. In contrast, regions of high movement such as loops would appear as dips on the pattern allowing secondary structure verification to be done (Poon 2007). An important feature would be the impact on the pattern of chemical modification of the protein. For example the PT-linker of a Family 10 Xylase (Poon 2007), showed a significant increase of 0.4 HetNOE values for multiple residues (from -0.4 to 0 and 0 to 0.4 in average) on a very flexible long loop upon addition of a glycan (≈ 4 kDa). This caused the partial dampening of the flexibility of that region of the linker, which could also be observed for interferon alpha-2a, although a less intense variation may be observed, since a GalNAc is only 0.22 kDa.

Figure 15: Relationship between the correlation time and the heteronuclear NOE ratios observed for these. The ratio varies most near 0.5 nanosecond and shows the ratio for slower motions (red) and faster motions (blue).

1.6. Tertiary structure calculation

In the not so distant past, NMR structure calculations depended on entirely manual measurements of NOE-derived distances by trained spectroscopists. These measurements were used as constraints in a molecular modeling program to generate structures that were consistent with the measurements. In this approach, the treatment of ambiguous distance restraints was not efficiently handled. Ambiguous restraints results from the presence of a NOE peak that can contain contributions from more than one proton pair. This was a time-consuming process. With the development of semi-automated programs the extraction of distance constraints from NOESY datasets is performed automatically and part of the structure calculation step in a bootstrap fashion. This means that structures generated in the first round of calculations with loosened distance constraints are used to reanalyze the NOESY datasets to extract more restraints. The treatment of ambiguous restraints is thus much more efficient and removes human bias since the algorithm can consider all possibilities.

1.6.1.

Semi-automatic software CYANA

One of these programs is called CYANA, which stands for Combined Assignment and Dynamics Algorithm for NMR Applications. The program is semi-automatic; it will function automatically and give complete tertiary structures but will still require user intervention to verify and correct problems leading to the violations or incorrect structures. The problems are most often the result of peak overlap leading to mis-assignment or error in the chemical shift list. The procedure can start without any manual

assignment of the NOE datasets. In this case, CYANA will automatically assign the NOE list following the “Combined automated NOE assignment and structure determination module” (CANDID) method (Herrmann 2002). From these assignments, CYANA uses NOE peak intensities and calibrates them to interproton distances and builds a list of unambiguous constraints and a list of ambiguous constraints. CYANA uses these constraints to minimize an energy function with a torsion angle dynamics algorithm. Initially, the assignment procedure yields a large number of ambiguous constraints that may not be resolved, i.e. multiple pairs of protons contribute to the NOE intensity, causing ambiguity for the program. CYANA uses a control system named network anchoring to eliminate superfluous constraints in the initial rounds of structure calculation. The basic idea relies on the creation of a network of constraints. If any constraints are isolated, CYANA will temporarily remove them from the calculations to reduce the effect of such constraints thus facilitating convergence. In the refinement steps of structure calculation, CYANA repeats the NOE assignment and if isolated NOEs still remain, they are added to the subsequent rounds of calculations. This process ends after a set amount of rounds, determined within the program specifications, usually seven rounds.

1.6.2.

Ramachandran plots

Helices and sheets are both secondary structure elements that have well defined sets of phi and psi torsion angles. Ramachandran plots display psi-phi angles allowed, mostly allowed and disallowed regions of the Ramachandran map. A precise structure is

characterized with secondary structure elements that display all the phi-psi angles in the allowed regions corresponding to the structural elements (helix or strand). In other words, the Ramachandran plots are measure the precision of a structure, by comparing the region where the phi-psi angles are located. This means if you have an alpha helix structure, all the angle pairs should appear within the area usually delimited for such a structure on the plot.

2-Materials and Methods

2.1.

Protein expression and purification

Interferon alpha-2a was produced doubly labelled with both ^{13}C and ^{15}N by COOP student, Naim Panjwani (Panjwani 2010).

2.2.

Protein glycosylation

The purified interferon alpha-2a was sent to the National Research Council (NRC) for N-acetyl galactosylation of the doubly labelled protein with a single unlabelled GalNAc unit by our collaborators Michel Gilbert and Denis Brochu (Gilbert 2005). See Appendix D for SDS-PAGE of glycosylation process over time, showing the slight variation in molecular weight between the unglycosylated and glycosylated forms.

Sample received from the NRC was exchanged to a NaAcO buffer (pH 3.5, 25 mM) by centrifugation (10 kDa cutoff mass, three times 25 minutes at 4000 RPM (3250 xg) at 4 °C). Final volume was concentrated to 0.55 ml and 1% of DSS with 5% of D₂O were added for a final concentration of 10 mg/ml of interferon alpha-2a glycosylated. The sample was kept at 4 °C until NMR experiments were done.

2.3.

NMR spectroscopy

The experiments were carried out on the glycosylated protein using a Bruker AVANCE III 600 MHz spectrometer equipped with a three channels cryoprobe. Sample temperature was kept at 298 Kelvin inside the magnet.

All pulse sequences used were taken from the Bruker library without modifications. Acquisition parameters for all 3D experiments are summarized in Table 2.3. All three-dimensional experiments had an initial relaxation delay of 1 second and 8 or 16 transients were collected per FID. The mode used during the acquisition is also identified in the table. The 3D HCCH-TOCSY and HCCH-COSY experiments were the only experiments which the F2 spectral window acquisition mode was done in State-TPPI instead of Echo-AntiEcho. The F3 window acquisition mode for all 3D experiments was done in State-TPPI. Carbohydrate-protein NOEs were obtained from 3D ^{13}C , ^{15}N -half-filtered NOESY experiments with a mixing time of 120 ms. A ^{13}C , ^{15}N -Filtered COSY experiment that selectively identify the unlabelled sugar protons without presence of the labelled protein was also carried out. 2D HSQC experiments were recorded as 2048 complex point by 128 complex point matrices with t1 and t2 acquisition times of 106 ms and 40.5 ms and spectral widths F1 and F2 of 9615 Hz and 1581 Hz. The same relaxation delay of 1 second as the 3D experiments was used and 8 scans were taken per FID.

Table 2.3: Parameters used for the NMR experiments done at 600 MHz.

Experiments	F1-Spectral width (Hz)	NP	F2-Spectral width (Hz)	NP	F3-Spectral width (Hz)	NP	Scans per FID
Mode	DQD		Echo-AntiEcho		State-TPPI		
	H		N		C		
HNCACB	8417.51	2048	1824.82	64	11325.03	128	8
CBCA(CO)NH	8417.51	2048	1824.82	64	11325.03	128	8
HNCO	8417.51	2048	1824.82	64	3321.16	64	8
HNCA	9615.39	2048	1824.82	64	4828.59	128	8
	H		N		H		
HNHA	9615.39	2048	1824.82	40	4800.77	128	16
HBHA(CO)NH	9615.39	2048	1824.82	40	4800.77	128	16
N-NOESY	9615.39	2048	1946.28	40	5402.49	256	16
	H		C		H		
HCCH-TOCSY*	8417.51	2048	10559.66*	64	5402.49*	128	16
HCCH-COSY*	8417.51	2048	10559.66*	64	5402.49*	128	16
C-NOESY	8417.51	2048	11325.028	64	6002.40	256	16
	H		N				
HSQC	9615.39	2048	1581.28	128			8

* For the HCCH-TOCSY and HCCH-COSY, the F2 spectral width was not in mode Echo-AntiEcho but in State-TPPI.

2.4.

¹⁵N Backbone dynamics

In order to obtain the heteronuclear NOEs, two experiments were done on the glycosylated and unglycosylated interferon sample with a presaturation delay of 3 seconds. Both the saturated and normal experiments were recorded with a 2048 complex point by 256 complex point matrices and the same spectral width as the HSQC spectrum. The two spectra obtained resembled an HSQC spectrum, allowing for easy assignment of the peaks in NMRView. With NMRView's HetNOE analysis tool, the spectrum with the NOE saturation and the one without any saturation were selected for calculation of the HetNOE effect ratio for each residue.

2.5.

Data processing

The raw Bruker time-domain data was first converted to NMRPipe-format data. Afterwards, data sets from all 3D experiments were processed with NMRpipe (Delaglio 1995) and visualized with NMRViewJ (Johnson 1994). Data processing will be briefly described for a typical 3D HNC experiment such as HNCO or HNCACB. In most cases, data acquisition proceeds in the following order: proton, nitrogen, carbon. But data visualization requires the nitrogen dimension to be along the z axis. The 3D spectrum was initially transformed by processing the x-axis, for example, the direct detection of the axis for the proton dimension. A residual solvent filtration (SOL) was done using NMRpipe to the data prior to a squared sine bell apodization shifted by 63° which was applied to the proton dimension to enhance resolution. An automatic zero filling of the

complex points was followed by Fourier transformation on the x-axis of the spectrum.

The spectrum was then phased with a zero-order phase correction. The left side of the spectrum was extracted from 11.75 ppm to 4.75 ppm for the proton dimension, the right side contains no data and is discarded to reduce the size of the data matrix. The data were transposed to allow processing of the carbon dimension. A 90° shifted sine bell apodization was applied prior to zero filling and Fourier transformation followed by phasing. The data were transposed again to process the nitrogen dimension. In order to increase resolution, the data were extended by linear prediction. Apodization with a 90° shifted sine bell window function was followed by a zero filling and Fourier transformation prior to the phasing of the spectrum.

2.6.

Structure calculation using semi-automated software CYANA

Specific data was given to CYANA in order to produce a 3D structure as shown in Figure 16, including the amino acid sequence, a list of assigned chemical shift and lists of cross-peak positions and volumes from both a carbon- and nitrogen-edited NOESY spectra from which interproton distance constraints will be derived. Additional non-NOE constraints were included such as lists of phi and psi angles obtained from TALOS+ as well as disulfide bonds. The chemical shift variation tolerance was individually set for each dimension of the NOESY spectra.

The NOEs from the peak files were assigned automatically with the CANDID module, which may not give single assignments for every NOE cross-peak, resulting in a array of

ambiguous NOEs. Once the program determined all possible assignments which may correspond to a certain NOE, CANDID then evaluated the quality of each of the NOE assignment and created a network of constraints. This will initially cause a distinction between the non-ambiguous and ambiguous NOEs. Finally, a series of 100 structures was produced from the constraints provided to CYANA using a torsion angle dynamics algorithm. This was done in 50,000 torsion steps and only the 20 best structures were kept. These selected structures were those whose target functions were lowest and served as a template for the subsequent NOE assignment cycle. After the 7th cycle a final structure was calculated and the corresponding NOE assignment used for it were kept for manual inspection.

Figure 16: Flow chart showing the various step performed by CYANA for automated NOE assignment and structure calculation. After structures are calculated, a new cycle begins which allows further modification of the NOE used for subsequent structures.

3-Results and Discussion

3.1. ^1H - ^{15}N HSQC Fingerprint

3.1.1

Effect of glycosylation on the HSQC spectrum.

The HSQC spectrum recorded for Met-IFN alpha-2a shows a very similar pattern to the published spectrum by Klaus and coworkers (Klaus 1997). The chemical shift perturbations are limited to residues at the N-terminal part of the protein. In contrast, the HSQC of the glycosylated Met-IFN alpha-2a display more chemical shift perturbations. The HSQC spectrum of unglycosylated interferon alpha-2a (black peaks) and glycosylated interferon alpha-2a (red peaks) are shown with their respective amino acid assignments, in Figure 17 and 18 respectively. The spread of the peaks across the spectra indicates the protein is folded properly. All expected backbone amides for the glycosylated protein are labelled in the figures, except for the five proline residues (Pro4, Pro39, Pro54, Pro109, Pro137), Lys23 and Met0. Prolines do not appear on an HSQC spectrum because they do not have a proton on their backbone amide. The unassigned resonances on both spectra belong to the side-chain amides of specific amino acids, which were not indicated on the HSQC.

Figure 17: HSQC spectrum of the unglycosylated IFN alpha-2a with amino acid assignments for backbone amides. The only missing amino acids are the prolines (Pro4, Pro39, Pro54, Pro109, Pro137) and Lys23, the former was located at a chemical shift for ^1H of 10.24 and ^{15}N 128.26 which was cut off the window to shorten the spectrum. The initial methionine M0 does not appear on the HSQC spectrum since it is the first amino acid. This spectrum was taken previously by Naim Panjwani.

Comparison of the HSQC spectra is best accomplished by overlaying both HSQC contour maps. Figure 19 shows an overlay of the glycosylated (red) and the unglycosylated (black) HSQC spectra. Overall, both the ^1H and ^{15}N chemical shifts for the glycosylated IFN alpha-2a did not vary significantly for the majority of the peaks. This suggests that the structure of interferon alpha-2a did not induce significant structural differences in the backbone. The most important variation of the chemical shift values were found in the loop regions of the protein, more specifically the glycosylated loop area. These are indicated by the black arrows in Figure 19. The residues that were most affected by this change were Thr106 and Thr108, which the former would be glycosylated with an N-acetyl galactosamine sugar unit. The glycosylation can therefore be detected using only a simple HSQC experiment, due to the backbone amide chemical environment changes. In order to determine the glycosylation site, the sequential assignment of the glycosylated sample had to be carried out. The procedure described previously was applied here and three experiments were sufficient to obtain most of the assignments. These are the HNCACB and CBCA(CO)NH and HNCA.

Figure 19: Overlay of the HSQC spectra of both the unglycosylated IFN alpha-2a (black peaks) and glycosylated IFN alpha-2a (red peaks). Chemical shift that seem to have varied are identified by an arrow and the two resonances are linked with a green line.

Figure 20 shows an enlargement of a region of the spectrum with the resonance assignment, which will be described in the next section in more details. By comparing both spectra, it is shown that both Thr106 and Thr108 no longer have the same ^1H and ^{15}N chemical shifts for the backbone amides.

Figure 20: Overlay for the Thr106-Thr108 region of the unglycosylated (black) and glycosylated (red) IFN alpha-2a HSQC spectrum. In this figure, only Thr106 and Thr108 have different chemical shifts between both spectrums. For these two residues, the labels for their location on the unglycosylated spectrum are green.

3.2 Glycosylated IFN alpha-2a resonance assignments

3.2.1

Chemical shift sequential assignment

The previous section showed that no major chemical shift perturbations were observed.

This indicated that no major structural changes were to be expected for glycosylated

interferon alpha-2a. While this may have suggested that the assignment reported by Klaus *et al.* could have been used, we opted for a *de novo* assignment in order to avoid any bias that could miss small conformational changes induced by the GalNAc moiety.

The sequential assignment was initiated at resonances which had unique carbon chemical shifts. For example, serines and threonines both have a C_β chemical shift with the highest value, i.e. near 60-70 ppm, while alanines have C_β which are near 18 ppm and glycines have a C_α near 45 ppm with no C_β . Since the glycosylation loop encompassing residues 103 to 108 contained two threonines, it was easy to first assign this portion of the protein backbone resonances. By identifying these residues, multiple chains of residues were successfully assigned. From these assigned chemical shifts, the chemical shifts for the entire sequence could be obtained for both the C_α and C_β by using different NMR experiments such as HNCACB, HNCA and CBCA(CO)NH.

Since the chemical shift of each backbone amide pairs remains the same across all experiments, it is possible to combine data from various experiments to greatly facilitate sequential assignment. Figure 21 shows a strip plot of the HNCACB and CBCA(CO)NH experiments for residues Val103 to Thr108 of glycosylated interferon alpha-2a. To obtain the information and compare it between experiments, both the HNCACB and CBCA(CO)NH experiments can be described as being books, where each page is a defined ^{15}N chemical shift. On these pages every proton and carbon, both correlated to that amide will also be correlated together and appear at their own chemical shifts. In Figure 21, only a selected range of chemical shifts were taken from the proton dimension,

while the entire carbon spectral window was kept. Strip plots are used to organize the data in order to simplify the comparison between the experiments. In most cases, these strips contain a single backbone amide unless overlapping of the residues occurs. In other words, a strip is a carbon spectra extracted from a 3D spectrum at the proton-nitrogen amide chemical shifts. In Figure 21, the positive signals are shown in black, while the negative signals are shown in red. While in the CBCA(CO)NH strips, both carbon alpha and beta signals are positive; in the HNCACB strips, the alpha carbon have a positive signal while the beta carbons have a negative signal. Since the CBCA(CO)NH strips give the carbon chemical shifts of the *i*-1 residue and the HNCACB strips give both the *i*-1 and the *i* residues, this allowed for the proper sequential assignment.

To illustrate the methodology, the sequential assignment of the glycosylation loop will be described. The loop contained two threonine residues, T106 and T108. Initially, strips corresponding to threonine or serine residues were identified in the dataset, then strips or any residues that correlated with a previous threonine or serine were also identified. Val105 precedes Thr106 and appears in the Thr106 CBCA(CO)NH strip with a C_α of 62.1 ppm and C_β of 32.8 ppm, compared to Glu107 which has a C_α 55.7 ppm and C_β of 30.5 ppm which appears in the Thr108 CBCA(CO)NH strip. Glu107 also has Thr106 C_α and C_β chemical shifts in the Glu107 CBCA(CO)NH. Thus matching the correlations could unambiguously identify and assign the resonances of Glu 107.

3.2.2

Determination of the glycosylation site

Glycosylation chemically modifies the side-chain of serine-threonine residues. The primary effect of this is a change of the electronic density of the nearby atom, the carbon beta. In addition, the presence of a glycan may cause the surrounding structure to be altered and this can modify the chemical shift of the backbone atoms, as previously observed in the HSQC spectra. The spectra suggested that either one or both Thr-106 and Thr-108 could be glycosylated. To determine which and how many residues of interferon alpha-2a was glycosylated, the carbon chemical shifts of all serine and threonines were analyzed. All but one threonine residues maintained very similar chemical shifts. Figure 22, shows the comparison between the HNCACB strips of Thr106 and Thr108 for both the glycosylated and the unglycosylated interferon alpha-2a. Analysis of the C_α and C_β chemical shifts of Thr106 shows a slight variation of 2.5 ppm between the glycosylated (59.5 ppm) and unglycosylated protein (62 ppm), while the variation for Thr106 C_β between the glycosylated (78.5 ppm) and unglycosylated protein (69.6 ppm) is 8.9 ppm. On the other end, no significant change was noticed on the Thr108 HNCACB strips. This indicates that a significant chemical environment change must be disturbing the C_β of Thr106 to cause such a large variation in chemical shift, even affecting mildly the C_α of Thr106. The chemical shift of the C_β of Thr106 is found at a surprising low value. This results from a C_β chemical shift with a higher value than the edge of the spectral window, thus appearing at the low chemical shift on the spectrum. This process is known as spectral aliasing.

Figure 22: ^{13}C - ^1H strips of the HNCACB experiments of T106 and T108 for the glycosylated and unglycosylated interferon alpha-2a.

3.2.3

Threonine 106 is the glycosylated residue on IFN alpha-2a

From the comparison of the C_α and C_β chemical shifts in the HNCACB strips, it can be determined that Thr106 was the only amino acid that was glycosylated. This is shown by the significant variation in the C_α (2.5 ppm) and C_β (8.9 ppm) chemical shifts for this residue. These chemical shift variations indicate that the GalNAc moiety unit is linked to the C_β of Thr106. On the unglycosylated protein, a hydroxyl group is found on the threonine, while the glycosylated protein finds the hydroxyl group modified with a large GalNAc unit. When comparing the other serine or threonine, none showed significant variation of their C_β chemical shifts. Therefore interferon alpha-2a was not glycosylated anywhere other than Thr106.

3.2.4

Chemical shift perturbations and structure

The sequential assignment can be used to visualize the chemical shift perturbations induced upon addition of the GalNAc unit. A model of the glycoprotein was built using the program Chimera (Pettersen 2004). The coordinates of the first model of the NMR structure and a GalNAc unit was connected manually (PDB identifier: NGA). In addition, the side-chains of residues experiencing significant chemical shift perturbations ($\delta_{HN} \pm 0.1$ ppm and $\delta_N \pm 0.5$ ppm) were displayed on the model shown in Figure 23. A quick analysis of the model reveals that all perturbations affect residues which are in close proximity to the glycan.

Figure 23: Model of GalNAc interferon alpha-2a built by attaching a N-acetyl galactosamine to T106 on the first model of 1ITF from the ensemble of 20 structures determined by NMR (Klaus et al 1997). Residues experiencing the most chemical shift perturbations have been highlighted in green as well as showing their side-chain atoms. In addition, disulfide bonds are shown in yellow, indicating the cysteins.

3.3. Backbone dynamics of interferon alpha-2a

3.3.1.

{1H} 15N Heteronuclear NOE backbone dynamics pattern

Since the addition of a single GalNAc unit occurred on Thr106, we then proceeded to study its effect on the backbone dynamics of the protein. In order to achieve this, we compared heteronuclear NOE results between our glycosylated and unglycosylated protein. Since the group that previously characterized the structure of the unglycosylated interferon alpha-2a also did these experiments, we were able to also compare with their own heteronuclear NOE analysis. The heteronuclear NOE experiment was performed on the unglycosylated protein by Klaus *et al.* (Figure 24A), where the intensity of the Het-NOE for each amino is represented by a bar in the graphic which can vary from -0.6 (more dynamic) to 1 (less dynamic). This means the faster the motion of the backbone, the lower the heteronuclear values which will be obtained for the saturated experiment, which will result in a lower saturated over unsaturated intensity ratio. These were calculated from the ratio of the intensity of the peaks between the Het-NOE, which represents the saturated experiment, and No-NOE experiments, representing the unsaturated experiment. The overall average value obtained for the unglycosylated protein was of 0.72 and compares very well with the values reported by Klaus and coworkers (Figure 24B). The N-terminal end of the protein is more rigid than the C-terminal end which is the more flexible part of the protein. This results from the fact that the first residue is a cysteine involved in a disulfide bridge. The other regions of the protein with faster backbone motion is loop 1, which goes from Arg22 to Ile53 and contains a small helix named B1 (Gln40 to Gly44); the second is loop 2 which goes from

Ile100 to Leu110 and is parallel to the end region of loop 1, Gly44 to Gly50. This is observed by the depression on the values of the graphic for these regions giving an average value of 0.55 for part of loop 1 close to loop 2 and 0.39 for loop 2. The remainder of the protein display slower backbone motions with average values of 0.78. Some values were missing due to overlapping peaks, which can be seen as only a black or red band in the graph.

A

B

Figure 24: Overlay of the distribution of $\{^1\text{H}\}$ - ^{15}N heteronuclear values for the interferon alpha-2a from Klaus *et al* (black in A and B) and the unglycosylated protein prior to glycosylation (red in B).

3.3.2.

Glycosylation reduces backbone motions of interferon alpha-2a

Comparison of the heteronuclear-NOE values between the unglycosylated protein (black and red) and the glycosylated protein (blue) is shown in Figure 25. Overall, the pattern is conserved between the two proteins. However, an increase of the heteronuclear-NOE values is observed in the loop 2 of the glycosylated protein. For the segment Ile100-Leu110, a sum average of over 0.2 is measured for the glycosylated loop versus the unglycosylated interferon alpha-2a. If for a stretch of at least four residues a sum average of more than 0.15 was found, that increase was said to be a small loss of mobility. Since in our case, we had a sum average of 0.2 over nine residues, we can say that our heteronuclear NOE value increase was significant, although not as much as that observed for the PT-linker of a Family 10 Xylase (Poon 2007) where they observed twice our increase. This increase that we observed in heteronuclear NOE values indicates that the loop displays slower motions in the glycosylated protein. The rest of the protein does not seem to be significantly affected by the presence of the GalNAc moiety. The helix averages and loop averages have very similar values in both proteins. This increase of heteronuclear NOE intensity in loop 2 can be correlated with a reduced mobility. A smaller change was also seen for parts of loop 1, indicating that there may be some interactions between the two loops in presence of the glycan, however we cannot say we had a significant increase of rigidity for loop 1.

Figure 25: Overlay of the distribution of $\{^1\text{H}\}\text{-}^{15}\text{N}$ heteronuclear NOE values for the unglycosylated protein prior to glycosylation (red) and the glycosylated protein with a single N-acetyl galactosamine unit on T106 (blue).

3.4. Secondary structure

3.4.1.

Chemical shift index

The sequential assignment yields chemical shifts of all backbone atoms. These can be analysed with empirical tools such as the chemical shift index to predict if glycosylation has induced changes in the secondary structure elements. Here, these are alpha-helices only. The chemical shift index is an empirical system that consists at subtracting the chemical shift of a residue in a structure, say alanine, and its chemical shift in a random coil situation. Random coil values have been determined using tri-peptide studies in water (Wishart 1995). Since other factors could perturb the chemical shifts of residues in secondary structure elements, the chemical shift index must be regarded as a predictive measure of these structural elements. Nevertheless, the CSI may reveal the presence of significant structural perturbations, if they exist. Figure 26, shows the chemical shift index for the C_α , C_β , CO and H_α for all residues. If a residue has a positive CSI value

over 1 for the C_α and CO, while having a negative CSI value for the C_β and H_α , this indicates that this residue is located in an alpha helix. While one or two residues may have CSI values that are inconclusive, it is the pattern of all residues that help to locate helices and beta strands along the primary sequence. The CSI plot shows the presence of five alpha helices in the structure of the glycosylated protein (Helix A: Ser11-Met21; Helix B: Pro54-Ser68, B1: Lys70-Ala75; Helix C: Glu78-Ile100; Helix D: Leu110-Glu132; Helix E: Pro137-Leu157). This corresponds to structural elements found for unglycosylated interferon alpha-2a. After examining the four chemical shift index plots, it is noted that only a few residues in the helix regions have negative values chemical shift differences for the C_α (E113 and I147) backbone atom while some have positive values for the C_β and H_α . However when compared to results obtained by Klaus *et al.* with the unglycosylated interferon alpha-2a, we noticed that most of these amino acids have similar values.

Figure 26: Chemical shift index of backbone atoms Ca , $\text{C}\beta$, CO and Ha to assess the secondary structure of the folded glycosylated protein. Regions containing the five main alpha helices are highlighted in grey, as well as the two minor alpha helix regions. See Appendix A for values.

3.4.2.

Chemical shift index comparison with unglycosylated interferon alpha-2a

The backbone secondary chemical shifts of the glycosylated product are broadly consistent with the backbone secondary chemical shifts of the unglycosylated interferon alpha-2a determined by Klaus *et al.* We note negligible differences between the different chemical shifts for all backbone resonances, supporting that glycosylation did not cause structural changes in the secondary structure of the protein. This can be observed in Figure 27, where the chemical shift index of all carbon alpha atoms is compared between the glycosylated and unglycosylated proteins. These two plots can be superimposed with only minor differences, indicating no changes in the secondary structure of the protein due to the glycosylation.

It is however important to remind that the CSI is an empirical measurement; it is not sufficiently sensitive to detect small structural differences induced by glycosylation. For this reason, a complete three-dimensional structure determination is required to assess the effects of glycosylation on the structure of interferon alpha-2a.

Figure 27: Comparison of the C_{α} chemical shifts variations from the random coil values for glycosylated and unglycosylated interferon alpha-2a. Differences between both values are shown in red on the overlay.

3.4.3.

TALOS torsion angle prediction

The information from the chemical shift index is essentially an indication of the structural elements. It is not possible to use it directly as torsion angle constraints. Torsion angle constraints can be obtained using the software TALOS. As discussed previously, TALOS predicts psi and phi angles by using backbone chemical shifts and comparing these shifts with chemical shifts of the same residues in known structures. Out of the 165 residues, TALOS predicted 129 torsion angles. These were in agreement with the corresponding secondary structure of the unglycosylated protein and were used in the structure calculations with CYANA.

3.4.4.

Short and medium range NOE for α -helices

The ^{15}N and ^{13}C -edited NOESY experiments give distance information between protons that can be first used to confirm or complete the sequential assignment. In addition, medium range NOE connectivities are used to establish the location of secondary structure elements. Sequential NOE connectivity are denoted d_{NN} , $d_{\alpha\text{N}}$ and $d_{\beta\text{N}}$, where the subscript refers to the protons giving rise to the NOE, i.e. 'N' is the HN, α is the alpha proton. Medium-range range connectivities are denoted $d_{\alpha\text{N}}(i, i+3)$, $d_{\alpha\text{N}}(i, i+4)$ and $d_{\alpha\beta}(i, i+3)$, where the 'i' indices refers to the position of the second label with respect with the first. In Figure 28, the sequential d_{NN} , $d_{\alpha\text{N}}$ and $d_{\beta\text{N}}$ NOE connectivities network was almost complete along the protein sequence. NOE connectivities were useful in the assignment of certain ambiguous residues such as R12 and R13. The repetitive pattern of

medium-range NOE connectivities is characteristic of a helical conformation. These show the presence of five main helical regions; represented as Helix A to Helix E in Figure 28 (Helix A: Gly10-Arg22; Helix B: Pro54-Ser68, B1: Lys70-Ala75; Helix C: Glu78-Ile100; Helix D: Leu110-Glu132; Helix E: Pro137-Leu157). These helices correspond to the same regions identified by TALOS and chemical shift differences showed earlier, confirming the idea that glycosylation did not modify the secondary structure of the protein.

Figure 28: Short and medium-range NOE connectivities and helical regions of the protein. Greyed out NOE connectivities indicate important overlapping of peaks causing ambiguity.

3.4.5.

Side-chains and methyl group assignments

Assignment of the side-chain proton resonances is the last and an essential step prior to the calculation of the three-dimensional structure. The HCCH-TOCSY and HCCH-COSY experiments were used to perform this task. These experiments correlate side-chain protons via the network of carbons atoms of aliphatic side-chains, allowing a near complete assignment of all proton chemical shifts. These experiments provide almost all aliphatic side-chains carbon and proton chemical shifts, including almost all methyl groups for the aliphatic side-chains. Some of the longer side-chains do not have complete assignments, due to severe resonance overlap. This was the case for the terminal regions of lysine and arginine where only some isolated resonances could be distinguished. Other residues such as glutamic acid or glutamine had severe overlap of their H γ and C γ resonances and could not be assigned specifically. The total percentage of side-chain assignment was over 90% for the methyl groups and 70% for the remainder of the side-chain proton resonances. Due to the ambiguity of the overlapped resonances and the fact that certain resonances may belong to more than a single nucleus, causes the process of side-chain chemical shift assignment to be time consuming. The inability to obtain all side-chain resonances may cause assignment errors that will create constraint violations during the structure calculation process. These can be corrected at that stage, if they exist. The side chain atom chemical shifts obtained can be observed in both Appendix B for the carbon chemical shifts and Appendix C for the proton and nitrogen chemical shifts.

3.5.

Toward the calculation of the three-dimensional structure with CYANA

At this time, the calculations had not produced a high-resolution structure of GalNAc-interferon alpha-2a. A number of issues have to be resolved and certain NOESY datasets will need to be collected again with greater resolution. In addition ^{13}C -edited NOESY 3D and 4D datasets should be done in D_2O , while a larger spectral width of the ^{15}N -edited NOESY should be used to include all protons without folding. These two issues caused CYANA to not be able to create a proper network of NOE and resulted in structures similar to those seen in Figure 29A. It can be observed that the secondary structures seem to have been generally correctly formed, however the three-dimensional folding was not proper. Since the program was able to obtain adequate secondary structures, we couldn't be sure where the problem was located. However more recently our group retook the two NOESY spectra in D_2O , and the resulting 3D-fold obtained can be seen in Figure 29B. The folds of these structures are closer to that of interferon alpha-2a, indicating that our issues were caused by the improper NOESY spectra.

Figure 29: A: Three-dimensional structure of glycosylated interferon alpha-2a, obtained by the CYANA structure calculation and visualized with Chimera . There are five helices (red), three small helices (red), and loops (white and green). The green loop represents the region natural glycosylation occurs on. The structure obtained does not represent entirely the correct secondary structure and its tertiary structure is completely improper. B: 3D structure of glycosylated interferon alpha-2a obtained by Houman at our laboratory, showing all 20 structures obtained by NMR. The structure shows a better defined secondary and tertiary structure and was obtained by retaking the NOESY experiment in D₂O.

3.6

Impact of results obtained

The chemical shift variations we observed due to the presence of a GalNAc on Thr-106, were also observed by other groups studying other side-chain modifications. Since the carbon beta of the threonine was directly linked to the oxygen linked to the GalNAc, the entire chemical environment for carbon beta changed. This resulted in the major chemical shift perturbation we observed only for that atom, although the carbon alpha of the threonine also had a slight variation. Only the backbone amide and amide proton of closely located residues suffered from slight chemical shift variations. When comparing this to other groups, we observe similar results, where they also see changes in their chemical shifts. A good example of this was the study done on the glycosylation of a threonine located on the PT-linker of a Family 10 Xylase (Poon 2007), where the glycosylation caused perturbation of the chemical shift of the amide and amide proton located within the PT-linker, having no effects on residues located within the two domains at each end of the linker were not affected by such variations. In this case, since no carbon chemical shifts were taken, it cannot be known if the effect was similar to ours. However it would be expected to similarly affect the carbon beta of that threonine.

Since chemical shift perturbations were observed for our glycosylated interferon, we measured if these changes were caused either by the backbone motions or if the glycan was having a direct interaction with each of these atoms. Having determined that there was an effect from the O-glycan on the backbone motion of interferon alpha-2a and not the direct interaction with the atoms, we can compare our heteronuclear NOE results to

other studies. In our case, we added a single GalNAc on Thr-106 of interferon alpha-2a, causing the backbone dynamics of the residues located near the glycosylation site to have been partially rigidified. By comparing this to another study, which was on the addition of a large glycan to a PT-linker of the Family 10 Xylase (Poon 2007), we observed similar result. Although in the case of the PT-linker a decrease in mobility was also noted, that decrease in mobility was twice as more important as for glycoprotein. This was probably due in part to the size of the glycans, where our glycan chain was smaller and therefore having a lesser motion dampening effect. This means that by further increasing the size of our glycan, it may be possible to further increase the rigidity of the glycosylated loop.

Since we noticed this reduction of mobility of our glycosylated loop and the potential for the formation of a water shell around the glycan, we can predict the increase of the solubility of the protein caused by the glycan. In addition, the shielding of hydrophobic side chain caused by their interaction with the glycan may have occurred since multiple amino acids such as valines are located closely to our glycan. In one study, it was shown that the glycan linked to an asparagine caused shielding of the protein surface by the interaction of the glycan with some hydrophobic residues (Erbel 2000). This shielding would increase the solubility of the protein and could mean that the protein remains longer within the blood, as was shown by research done on the pegylation of protein, in particular on interferon alpha-2a. The research done showed that pegylated interferon had a reduced degradation rate compared to the unglycosylated interferon alpha-2a (Bailon 2001). However compared to the pegylated product, our glycosylated interferon alpha-2a

does not have a reduced activity caused by the PEG chains sometimes closely located to the binding site, reducing the binding affinity with the receptor. This may allow the glycosylated interferon alpha-2a to have both a high binding affinity to its receptor as well as longer bioavailability in the blood. However these conclusions would require enzymatic experiments, such as experiments of the binding activity with the receptor, to be carried out in order to verify if the activity is affected by the glycosylation.

Although we did not complete the tertiary structure of our glycosylated interferon alpha-2a, by looking at other studies and our secondary structure information, we can predict that very little structural changes should be caused by the addition of the glycan. For example, in the study where they observed the effects of an N-glycan on the structure of their protein (Erbel 2001), they observed no 3D-structural changes, although at the local level there were some slight alterations. In our case, we would expect no real structural change of our protein, since the overall backbone dynamics as well as chemical shifts had no changes. Structural differences would cause variations in these which we could of measured and have had an indication prior to the initiation of the structural calculations. However, there may be some local changes in the structure near the glycan, since the side chains may interact with some chemical groups on the glycan, such as what was observed in the previous study with the N-glycan.

4-Conclusion

The study of GalNAc-interferon alpha-2a by high-resolution NMR techniques has shown that the *in vitro* glycosylation of the protein occurred exclusively at Thr106 as expected. Analysis of the chemical shift perturbations induced by the presence of the GalNAc unit, and of secondary structure elements does suggest that no significant structural changes were induced with glycosylation. While no conclusions regarding the effect of glycosylation can be drawn until a high-resolution structure is obtained; we were able to observe a small loss of mobility of the glycosylation loop. Considering that the addition of the single N-acetyl galactosamine on Thr106 was sufficient to affect of the dynamics of this region, future experiments will focus on the effect of additional sugar moieties in the glycan chain to determine their contributions to the conformational behaviour of interferon alpha-2a.

5-References

- Adolf GR, Kalsner I, Ahorn H, Maurer-Fogy I and Cantell K, Natural Human Interferon- α 2 is O-glycosylated, *Biochemical Journal*, 1991, 276, 511-518
- Bailon P, Palleroni A, Schaffer CA, Spence CL, Fung W-J, Porter JE, Ehrlich GK, Pan W, Xu ZX, Modi MW, Farid A, Berthold W and Graves M, Rational Design of a Potent, Long-lasting Form of Interferon: a 40 kDa Branched Polyethylene Glycol-conjugated Interferon alpha-2a for the Treatment of Hepatitis C, *Bioconjugate Chemistry*, 2001, 12:195-202
- Becker ED, *High Resolution NMR: Theory and Chemical Applications*, 3rd Edition, 2000, Academic Press
- Borden E C, Interferons-expanding Therapeutic Roles, *The New England Journal of Medicine*, 1992, 326:1491-1493
- Cavanagh J, *Protein NMR Spectroscopy: Principle and Practice*, 1996, Academic Press
- Colamonici O R and Domansk P, Identification of a Novel Subunit of the Type I Interferon Receptor Localized to Human Chromosome 21, *Journal of Biological Chemistry*, 1993, 265(15):10895-10899
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J and Bax A, NMRPipe: a Multidimensional Spectral Processing System Based on UNIX Pipes, *Journal of Biomolecular NMR*, 1995, 6:277-293
- DeMaeyer E M, *Interferon and Other Regulatory Cytokines*, 1988, Wiley
- Downing AK, *Automated NMR Structure Calculation With CYANA*, by Güntert P, *Protein NMR techniques*, Volume 278, 2004, Humana Press, Totowa, NJ
- Edwards AJ, Reid D, *Introduction to NMR of Proteins*, *Current Protocols in Protein Science*, 2001, Wiley
- Ellies LG, Jones AT, Williams MJ and Ziltener HJ, Differential Regulation of CD43 Glycoforms on CD4⁺ and CD8⁺ T Lymphocytes in Graft-versus-host Disease, *Glycobiology*, 1994, 4(6):885-893
- Erbel P, Karimi-Nejad Y, van Kuik A, Boelens R, Kamerling J and Vliegthart J, Effects of the N-Linked Glycans on the 3D Structure of the Free α -Subunit of Human Chorionic Ganadotropin, *Biochemistry* 2000, 39:6012-6021

- Furukawa K, Kobata A, Protein glycosylation, *Current Opinion in Biotechnology* 1992, 3:554–559
- Gilbert M, Wakarchuk W, Polypeptides Having beta-1,4-GalNAc Transferase Activity, National Research Council of Canada, US patent entry: 20050227248
- Grzesiek S, Bax A, An Efficient Experiment for Sequential Backbone Assignment of Medium-sized Isotopically Enriched Proteins, *Journal of Magnetic Resonance*, 1992, 99:201–207
- Grzesiek S, Bax A, Correlating Backbone Amide and Side-chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR, *Journal of the American Chemical Society*, 1992, 114(16):6291–6293
- Hart GW, Dynamic O-linked Glycosylation of Nuclear and Cytoskeletal Proteins, *Annual Review of Biochemistry*, 1997, 66:315–335
- Hauschild A, Weichenthal M, Rass K, Linse R, Berking C, Böttjer J, Vogt T, Spieth K, Eigentler T, Brockmeyer NH, Stein A, Näher H, Schadendorf D, Mohr P, Kaatz M, Tronnier M, Hein R, Schuler G, Egberts F and Garbe C, Efficacy of Low-Dose Interferon $\alpha 2a$ 18 Versus 60 Months of Treatment in Patients With Primary Melanoma of ≥ 15 mm Tumor Thickness: Results of a Randomized Phase III DeCOG Trial, *Journal of Clinical Oncology*, 2010, 28(5):841–846
- Hennet T, Chui D, C Paulson JC and Marth JD, Immune Regulation by the ST6Gal Sialyltransferase, *Proceedings of the National Academy of Sciences*, 1998, 95:4504–4509
- Herrmann T, Güntert P and Wüthrich K, Protein NMR Structure Determination with Automated NOE Assignment Using the New Software CANDID and the Torsion Angle Dynamics Algorithm DYANA, *Journal of Molecular Biology*, 2002, 319(1)
- Holzgrabe U, Wawer I, Diehl B, *NMR Spectroscopy in Drug Development and Analysis*, 1999, Wiley
- Houde D, Berkowitz SA, Investigating the Influence of Glycosylation on Protein Conformation and Dynamics, *Pharmaceutical Technology*, 2010, 34:12–17

- Itri LM, Campion M, Dennin RA, Palleroni AV, Gutterman JU, Groopman JE and Trown PW, Incidence and Clinical Significance of Neutralizing Antibodies in Patients Receiving Recombinant Interferon Alfa-2a by Intramuscular Injection, *Cancer*, 1987, 59:668-674
- Jacobsen NE, *NMR Spectroscopy Explained: Simplified Theory, Applications and Examples for Organic Chemistry and Structural Biology*, 2007, Wiley
- Johnson BA, and Blevins RA, NMR View: A Computer Program for the Visualization and Analysis of NMR Data, *Journal of Biomolecular NMR*, 1994, 4:603-614.
- Jorgenson L, Nielson HM, *Delivery Technologies for Biopharmaceuticals: Peptides, Proteins, Nucleic Acids and Vaccines*, 2010, Wiley
- Jitsuvara Y, Toyoda T, Itai T, and Yamaguchi II, Chaperone-Like Functions of High-Mannose Type and Complex-Type *N*-Glycans and Their Molecular Basis, *Journal of Biochemistry*, 2002, 132(5): 803-811
- Karlberg H, Lindegren G, and Mirazimi A, Comparison of Antiviral Activity of Recombinant and Natural Interferons Against Crimean-Congo Hemorrhagic Fever Virus, *The Open Virology Journal*, 2010, 4: 38–41
- Klaus W, Gsell B, Labhardt AM, Wipf B, Senn H, The Three-dimensional High Resolution Structure of Human Interferon alpha-2a Determined by Heteronuclear NMR Spectroscopy in Solution, *Journal of Molecular Biology*, 1997, 274:661–675
- Kowalewski J, Mäler L, *Nuclear Spin Relaxation in Liquids: Theory, Experiments, and Applications*, 2006, CRC Press, Boca Raton, FL
- Krapp S Mimura Y, Jefferis R, Huber R, Sonderrmann P, Structural Analysis of Human IgG-Fc Glycoforms Reveals a Correlation Between Glycosylation and Structural Integrity, *Journal of Molecular Biology*, 2003, 325(5): 979–989
- Kubetzko S, Sarkar CA and Plückthun A, Protein PEGylation Decreases Observed Target Association Rates via a Dual Blocking Mechanisms, *Molecular Pharmacology*, 2005, 68(5):1439-1454
- Lis H, Sharon N, *Protein Glycosylation—Structural and Functional Aspects*, *European Journal of Biochemistry*, 1993, 218:1–27

- Mandal TK, Mukhopadhyay C, Effect of Glycosylation on Structure and Dynamics of MHC class I Glycoprotein: a Molecular Dynamics Study, *Biopolymers*, 2001, 59(1):11-23
- Marshall SA, Lazar GA, Chirino AJ, Desjarlais JR, Rational Design and Engineering of Therapeutic Proteins, *Drug Discovery Today*, 2003, 8(5):212-215
- Medzihradszky KF, Characterization of Site-specific N-Glycosylation, *Methods in Molecular Biology*, 2008, 446
- Molinari M, N-glycan Structure Dictates Extension of Protein Folding or Onset of Disposal, *Nature Chemical Biology*, 2007, 3(6):313-20
- Novick D, Cohen B and Rubinstein M, The Human Interferon α/β Receptor: Characterization and Molecular Cloning, *Cell*, 1994, 77:391-400
- Nyman T A, Tölö H, Parkkinen J, and Kalkkinen N, Identification of Nine Interferon- α Subtypes Produced by Sendai Virus-induced Human Peripheral Blood Leucocytes, *Biochemistry Journal*, 1998, 329(Pt 2): 295–302
- Ohgushi M, Wada A, Molten-globule State—a Compact Form of Globular-proteins with Mobile Side-chains, *Federation of European Biochemical Societies Letters*, 1983, 164:21–24
- Panjwani N, Hodgson DJ, Sauvé S, Aubin Y, Assessment of the Effects of pH, Formulation and Deformulation on the Conformation of Interferon α -2 by NMR, *Journal of Pharmaceutical Sciences*, 2010, 99(8):3334-42
- Parella T, Protein Dynamics NMR Time Scale in Relaxation, *Magnetic Resonance in Chemistry*, 1998, 36:467-495
- Pestka S, Krause C D, Walter M R, Interferons, Interferon-like Cytokines, and their Receptors, *Immunological Reviews*, 2004, 202(1)
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE, UCSF Chimera- A Visualization System for Exploratory Research and Analysis. *Journal of Computational Chemistry*, 2004, 13:1605-12.
- Pitha P M, Interferon: the 50th Anniversary, 2007, Springer

- Poon DKY, Withers SG, and McIntosh LP, Direct Demonstration of the Flexibility of the Glycosylated Proline-Threonine Linker in the *Cellulomonas fimi* Xylanase Cex through NMR Spectroscopic Analysis, *Journal of Biological Chemistry*, 2007, 282(3)
- Purcell AW, van Driel IR and Gleeson PA, Impact of Glycans on T-cell Tolerance to Glycosylated Self-antigens Glycans and T-cell responses, *Immunology and Cell Biology*, 2008, 86:574-579
- Ragnhammar P and Wadhwa M, Neutralising Antibodies to Granulocyte-macrophage Colony Stimulating Factor (GM-CSF) in Carcinoma Patients Following GM-CSF Combination Therapy, *Medical Oncology*, 1996 13(3): 161-166
- Reid DG, Protein NMR techniques, *Methods in Molecular Biology*, Volume 60, 1997, Humana Press
- Rosner MR, Hubbard SC, Ivatt RJ, Robbins PW, N-asparagine-linked Oligosaccharides: Biosynthesis of the Lipid-linked Oligosaccharides, *Methods in Enzymology*, 1982; 83: 399–408
- Rule GS, Hitchens TK, *Fundamentals of Protein NMR Spectroscopy*, 2006, Springer
- Seng GC and Lengyel P, The Interferon System, *Journal of Biological Chemistry*, 1992, 267:5017-5020
- Sheehan D, *Physical Biochemistry: Principles and Applications*, Second Edition, 2009, Wiley
- Shen Y, Delaglio F, Cornilescu G and Bax A, TALOS+: A Hybrid Method for Predicting Protein Backbone Torsion Angles from NMR Chemical Shifts, *Journal of Biomolecular NMR*, 2009
- Toribara NW, Gum JR, Culhane PJ, Lagace RE, Hicks JW, Petersen GM, and Kim YS, MUC-2 Human Small Intestinal Mucin Gene Structure Repeated Arrays and Polymorphism, *The Journal of Clinical Investigation*, 1991, 88(3)
- Trown PW, Wills RJ and Kamm JJ, The Preclinical Development of Roferon-A, *Cancer*, 1986, 57:1648-1656
- Varki A, Biological Roles of Oligosaccharides: All of the Theories are Correct, *Glycobiology*, 1993, 3: 97–130

- Varki A, Cummings RD, Esko JD, Freeze HH, Stanley J, Bertozzi CR, Hart GW and
Etzler ME, *Essentials of Glycobiology*, Second Edition, 2009, Cold Spring Harbor
Laboratory Press
- Webb GA, *Nuclear Magnetic Resonance*, 2005, Wiley
- Wells JA, *Structural and Functional Basis for Hormone Binding and Receptor
Oligomerization*, *Current Opinion in Cell Biology*, (1994) 6:63-173
- Williams RL, Greene SM, and Pherson AM, *The Crystal Structure of Ribonuclease B
at 2.5-Å Resolution*, *Journal of Biological Chemistry*, 1987, 262:3
- Wu WZ, Sun HC, Shen YF, Chen J, Wang L, Tang ZY, Iliakis G, Liu KD, *Interferon
alpha 2a Down-regulates VEGF Expression Through PI3 Kinase and MAP Kinase
Signaling Pathways*, *Journal of Cancer Research and Clinical Oncology*, 2005,
131(3):169-78
- Yua H, Liua E, Lud S and Zhao S, *Treatment with Peginterferon versus Interferon in
Chinese Patients with Hepatitis B*, *Biomedicine and Pharmacotherapy*, 2010
- Zhang RF, Sun HQ, Huang Q, Wang JR, Zhang XX, Liu XN, Ma Q and Lu HZ,
*Efficacy and Safety of Pegylated Interferon alpha-2a Therapy for Chronic
Hepatitis C in HIV-infected Patients with Haemophilia*, *Haemophilia*, 2010,
16:502–507

Appendix A: Chemical shift table for the backbone atoms.

Residue	$\sigma^{13}\text{C}$	$\sigma^{13}\text{Ca}$	$\sigma^{13}\text{C}\beta$	$\sigma^{15}\text{N}$	$\sigma^1\text{H}^{\text{N}}$	$\sigma^1\text{H}\alpha$	$\sigma^1\text{H}\alpha_2$	$\sigma^1\text{H}\beta$	$\sigma^1\text{H}\beta_2$
1 CYS	172,97	55,95	41,68	121,13	8,91	4,60	3,89	3,09	
2 ASP	174,76	52,84	39,32	125,51	8,77	4,71		2,70	2,59
3 LEU		53,65	40,56	123,91	8,16	4,40		1,62	1,42
4 PRO	176,21	63,90	31,89			4,35		2,25	1,83
5 GLN	177,00	57,71	28,95	121,61	8,78	4,23		2,10	2,05
6 THR	174,70	62,66	69,26	111,42	7,87	4,15		4,18	
7 HIS	174,98	56,00	28,23	119,14	8,57	4,61		3,34	3,15
8 SER	175,31	58,97	63,44	115,70	8,18	4,36		3,99	3,89
9 LEU	178,57	56,97	41,80	123,82	8,40	4,16		1,61	1,61
10 GLY	174,59	46,91		107,58	8,43	3,93			
11 SER	175,75	61,30	62,65	118,81	8,24	4,16		3,87	
12 ARG	178,28	59,83	29,58	121,96	8,03	3,95		1,80	1,80
13 ARG	178,48	60,49	30,08	119,43	8,44	3,84		1,43	1,69
14 THR	175,58	68,07	67,97	116,83	7,94	3,54		4,19	
15 LEU	178,65	58,09	41,55	120,19	7,54	3,95		1,81	1,58
16 MET	178,72	58,51	32,15	118,79	7,99	4,02		2,10	2,10
17 LEU	178,24	57,73	42,09	120,64	7,81	3,97		1,97	0,92
18 LEU	175,74	57,73	42,25	118,18	7,49	3,52		1,58	1,12
19 ALA	180,74	55,07	18,08	121,89	7,95	3,84		1,43	
20 GLN	176,95	58,02	28,47	118,60	7,96	3,94		2,23	1,98
21 MET	174,67	57,11	34,11	116,65	7,42	3,86		1,60	1,49
22 ARG	175,95	58,45	31,65	115,40	6,48	3,85		1,71	1,42
23 LYS	176,64	57,29	36,25	128,56	10,22	4,54		1,88	1,69
24 ILE	173,88	59,82	41,46	118,60	8,48	4,53		1,79	
25 SER	176,36	56,98	64,76	113,57	8,10	4,51		3,89	3,76
26 LEU	178,95	57,68	41,72	127,51	8,80	3,88		1,67	1,41
27 PHE	177,07	60,11	38,28	117,18	7,99	4,27		3,15	3,04
28 SER	174,53	59,45	63,79	115,19	7,93	4,53		4,05	4,03

Residue	$\sigma^{13}\text{C}$	$\sigma^{13}\text{Ca}$	$\sigma^{13}\text{C}\beta$	$\sigma^{15}\text{N}$	$\sigma^{1}\text{H}^{\text{N}}$	$\sigma^{1}\text{H}\alpha$	$\sigma^{1}\text{Ha}_2$	$\sigma^{1}\text{H}\beta$	$\sigma^{1}\text{HB}_2$
29 CYS	176,48	53,43	37,21	119,93	7,87	5,08		3,42	2,58
30 LEU	179,77	58,30	41,58	122,27	7,80	3,96		1,74	1,46
31 LYS	176,69	54,08	31,59	117,65	8,59	4,69		1,77	1,77
32 ASP	175,51	54,24	41,82	117,74	7,81	4,69		2,40	2,80
33 ARG	175,29	58,08	30,01	120,85	7,39	4,04		1,79	1,79
34 HIS	170,49	54,61	31,33	120,66	7,97	4,22		2,65	2,65
35 ASP	175,94	52,46	40,82	124,93	7,79	4,65		2,51	2,30
36 PHE	176,64	59,45	40,69	122,83	8,97	4,35		3,28	2,77
37 GLY	175,19	48,14		111,74	8,73	3,92	3,68		
38 PHE		55,80	38,66	123,05	9,19			2,72	2,94
39 PRO	174,63	61,54	28,70			4,57		1,99	1,72
40 GLN	176,96	58,07	28,82	123,59	7,96	3,62		2,07	2,02
41 GLU	177,35	57,93	27,83	119,45	9,04	4,14		1,97	1,96
42 GLU	174,89	56,38	27,56	117,10	7,65	4,06		1,64	1,44
43 PHE	175,62	57,59	39,89	115,51	7,60	4,49		3,33	2,72
44 GLY	174,03	45,61		107,24	7,74	4,08	4,04		
45 ASN	175,89	53,85	38,42	117,46	8,35	4,55		2,81	2,81
46 GLN	175,85	56,65	28,44	118,92	8,65	4,08		1,76	1,76
47 PHE	176,26	57,92	39,86	119,49	7,92	4,58		3,14	2,87
48 GLN	181,27	55,58	29,47	120,13	8,26	4,38		2,23	1,91
49 LYS	177,97	58,15	32,39	121,78	8,43	4,10		1,77	1,70
50 ALA	178,70	54,15	18,83	120,46	8,57	4,09		1,36	
51 GLU	177,00	56,99	28,91	115,00	7,91	4,27		2,08	2,08
52 THR	175,27	63,61	69,67	111,21	7,86	4,11		4,27	
53 ILE		66,89	34,95	119,07	8,34	3,89		2,10	
54 PRO	179,66	66,01	31,89			4,31		2,35	1,77
55 VAL	177,65	65,77	31,47	116,42	7,12	3,70		2,02	1,56
56 LEU	178,64	58,09	42,32	123,05	8,54	4,31		1,80	3,40
57 HIS	175,64	60,70	27,91	116,98	9,06	3,79		3,09	2,12
58 GLU	179,03	58,97	28,29	119,52	7,84	4,01		2,12	
59 MET	177,03	59,41	32,53	119,51	8,44	4,07		2,64	2,50

Residue	$\sigma^{13}\text{C}$	$\sigma^{13}\text{Ca}$	$\sigma^{13}\text{C}\beta$	$\sigma^{15}\text{N}$	$\sigma^{1}\text{H}^N$	$\sigma^1\text{Ha}$	$\sigma^1\text{Ha}_2$	$\sigma^1\text{H}\beta$	$\sigma^1\text{H}\beta_2$
60 ILE	177,52	65,33	36,63	117,12	8,35	3,43		1,67	
61 GLN	178,27	58,79	28,67	120,19	8,23	3,70		1,96	1,79
62 GLN	179,69	58,01	24,34	116,67	8,42	3,63		0,94	0,51
63 ILE	177,28	66,84	37,39	122,80	8,16	3,65		1,99	
64 PHE	178,40	62,17	38,42	120,82	8,41	3,72		2,75	3,30
65 ASN	177,94	55,87	37,26	119,89	8,82	4,06		2,85	2,78
66 LEU	177,78	58,27	42,61	121,53	8,18	4,22		1,70	2,20
67 PHE	177,97	59,40	38,60	110,69	8,16	4,93		3,94	3,39
68 SER	173,47	59,13	63,11	119,13	8,02	4,61		3,62	3,11
69 THR	175,66	61,10	72,41	111,31	6,98	4,38		4,70	
70 LYS	179,26	57,96	31,75	121,03	9,07	4,13		1,78	1,78
71 ASP	178,37	57,07	40,60	116,36	7,51	4,18		2,30	2,30
72 SER	176,90	60,22	62,61	117,28	8,07	3,73		4,06	4,06
73 SER	174,86	61,39	62,80	113,22	7,98	4,11		3,93	3,93
74 ALA	177,63	52,81	18,76	119,38	7,09	4,19		1,31	
75 ALA	177,20	53,01	18,12	121,36	7,16	3,71		-0,30	
76 TRP	177,62	53,70	33,10	116,71	7,08	5,10		3,02	2,69
77 ASP	176,88	54,82	42,62	119,93	8,26	4,37		2,68	2,56
78 GLU	177,44	59,38	28,64	127,14	8,93	3,89		2,23	2,07
79 THR	177,33	65,96	68,00	115,40	8,13	3,94		4,23	
80 LEU	179,02	57,24	41,65	121,80	7,74	3,89		1,75	1,05
81 LEU	177,06	57,77	40,95	119,28	8,34	3,48		1,70	1,04
82 ASP	179,02	56,82	38,22	115,25	8,00	4,29		2,85	2,85
83 LYS	178,62	59,75	33,11	119,49	7,34	3,98		1,85	1,68
84 PHE	177,21	60,34	39,51	123,91	8,47	4,06		3,08	2,74
85 TYR	178,10	58,38	36,50	118,31	9,01	4,09		3,24	3,24
86 THR	176,71	66,76	68,89	114,03	7,61	3,97		4,27	
87 GLU	179,61	58,40	28,29	120,86	7,70	4,05		2,09	1,84
88 LEU	178,70	58,25	41,59	120,67	8,39	3,73		1,73	1,03
89 TYR	178,44	61,42	37,63	118,57	8,64	3,98		2,99	2,99
90 GLN	178,41	58,78	28,15	118,28	7,78	3,99		2,16	2,16

Residue	$\sigma^{13}\text{C}$	$\sigma^{13}\text{Ca}$	$\sigma^{13}\text{C}\beta$	$\sigma^{15}\text{N}$	$\sigma^{1\text{H}^N}$	$\sigma^{1\text{H}\alpha}$	$\sigma^{1\text{H}\alpha_2}$	$\sigma^{1\text{H}\beta}$	$\sigma^{1\text{H}\beta_2}$
91 GLN	178,71	58,82	29,17	117,96	7,73	4,08		2,05	
92 LEU	178,98	58,49	41,38	119,98	8,48	3,73		1,74	1,36
93 ASN	178,31	56,40	38,64	117,64	8,08	4,40		2,76	2,76
94 ASP	178,33	56,55	38,65	121,09	8,37	4,41		2,70	2,91
95 LEU	178,83	57,03	42,21	119,12	8,15	4,04		1,43	1,98
96 GLU	177,64	57,71	28,14	118,81	8,11	3,96		2,06	2,06
97 ALA	179,13	54,03	17,98	120,55	7,80	4,15		1,44	
98 CYS	175,36	57,17	41,88	115,51	7,60	4,47		3,24	3,19
99 VAL	177,36	64,58	32,03	121,59	7,84	3,79		2,08	
100 ILE	176,83	62,24	38,31	122,35	8,08	3,95		1,84	
101 GLN	176,44	56,32	29,01	121,74	8,07	4,23		2,04	2,12
102 GLY	173,86	45,16		109,22	8,09	4,03	3,84		
103 VAL	176,30	62,50	32,85	118,94	8,08	4,06		2,02	
104 GLY	173,69	45,21		111,95	8,45	4,03	3,82		
105 VAL	175,96	62,14	32,81	120,42	7,87	4,19		2,08	
106 THR	173,73	59,50	78,52	116,32	8,57	4,58		4,26	
107 GLU	175,64	55,71	30,50	119,74	8,44	4,46		1,94	1,95
108 THR		60,35	68,78	114,36	8,16	4,47		4,48	
109 PRO	178,67	65,38	31,85			4,24		2,30	1,85
110 LEU	178,29	56,69	41,97	118,57	8,02	4,12		1,53	1,53
111 MET	178,72	58,14	32,37	118,89	7,71	4,21		2,52	2,52
112 LYS	178,00	56,31	32,37	121,31	8,28	4,11		1,87	1,72
113 GLU	178,29	58,25	28,14	119,03	8,26	4,07		2,08	2,08
114 ASP	178,37	56,11	38,24	119,20	8,60	4,32		2,71	2,71
115 SER	175,99	61,59	62,65	118,14	8,19	4,20		3,79	3,79
116 ILE	178,37	64,82	37,42	123,27	7,79	3,63		2,01	
117 LEU	178,75	58,14	41,59	120,97	8,39	3,91		1,84	1,65
118 ALA	181,63	55,31	18,14	120,15	7,89	4,11		1,47	
119 VAL	177,90	67,07	31,89	118,86	7,59	3,77		2,47	
120 ARG	179,85	61,27	29,70	121,41	8,62	4,03		2,08	2,08
121 LYS	178,33	60,44	32,11	119,55	9,02	4,07		1,94	1,90

Residue	$\sigma^{13}\text{C}$	$\sigma^{13}\text{Ca}$	$\sigma^{13}\text{C}\beta$	$\sigma^{15}\text{N}$	$\sigma^{1\text{H}^N}$	$\sigma^{1\text{H}\alpha}$	$\sigma^{1\text{H}\alpha_2}$	$\sigma^{1\text{H}\beta}$	$\sigma^{1\text{H}\beta_2}$
122 TYR	176,53	60,40	37,73	121,52	7,49	4,35		3,58	3,13
123 PHE	179,01	63,53	39,33	115,97	7,93	3,74		3,20	2,92
124 GLN	178,77	59,32	27,63	122,99	8,46	4,18		2,38	2,20
125 ARG	180,34	61,30	30,33	120,34	8,14	4,08		2,57	2,57
126 ILE	176,86	65,75	37,24	121,84	7,53	3,37		1,67	
127 THR	176,65	67,17	68,83	115,62	8,30	3,64		4,24	
128 LEU	178,33	58,09	41,57	122,60	8,26	3,96		1,71	1,47
129 TYR	176,04	61,25	38,90	120,89	7,51	3,95		3,25	2,79
130 LEU	177,95	58,11	41,59	119,09	7,72	2,86		1,31	0,28
131 LYS	181,27	59,67	32,67	116,78	7,49	3,44		1,78	1,76
132 GLU	178,38	58,89	28,67	121,92	8,43	3,84		2,03	1,88
133 LYS	175,28	54,30	31,98	116,02	7,40	3,97		1,01	1,70
134 LYS	176,27	57,22	28,80	117,76	7,72	3,60		1,85	1,85
135 TYR	174,15	59,60	34,58	110,53	8,43	2,87		3,34	3,34
136 SER		57,11	62,50	113,65	6,74	4,55		4,37	4,01
137 PRO	180,68	65,66	31,98			4,20		2,09	2,09
138 CYS	176,02	57,14	36,78	113,86	8,57	4,68		2,92	2,78
139 ALA	180,15	54,78	18,88	125,25	8,09	4,45		1,04	
140 TRP	179,14	60,53	29,26	116,48	8,50	4,61		3,39	2,92
141 GLU	177,89	58,03	28,72	121,56	8,18	5,25		2,21	2,14
142 VAL	180,25	66,98	31,82	121,16	8,62	3,63		2,69	
143 VAL	176,41	67,88	31,66	120,08	45,61	3,31		2,41	
144 ARG	178,13	61,00	29,94	123,99	9,62	3,70		2,46	2,22
145 ALA	180,76	55,51	18,12	119,37	8,87	3,87		1,50	
146 GLU	177,30	58,38	28,13	121,82	8,13	4,21		2,24	2,07
147 ILE	178,28	61,25	35,30	119,50	8,72	3,97		2,19	
148 MET	178,21	60,18	32,91	119,50	8,44	3,80		2,17	2,01
149 ARG	178,98	59,86	30,28	119,74	7,64	3,95		2,07	1,86
150 SER	177,29	61,32	60,12	118,11	9,19	3,34		2,76	2,19
151 PHE	178,70	61,42	39,23	125,16	9,47	4,04		3,01	3,01
152 SER	175,86	62,39	62,52	115,83	7,84	3,88		3,87	3,87

Residue	$\sigma^{13}\text{C}$	$\sigma^{13}\text{Ca}$	$\sigma^{13}\text{C}\beta$	$\sigma^{15}\text{N}$	$\sigma^1\text{H}^{\text{N}}$	$\sigma^1\text{Ha}$	$\sigma^1\text{Ha}_2$	$\sigma^1\text{H}\beta$	$\sigma^1\text{H}\beta_2$
153 LEU	179,52	57,88	42,34	121,34	7,77	4,08		1,72	1,43
154 SER	174,31	61,37	63,25	113,93	7,88	3,78		3,38	3,31
155 THR	174,80	62,28	69,39	109,34	7,18	4,05		4,03	
156 ASN	175,53	54,16	38,34	119,64	7,74	4,54		2,92	2,70
157 LEU	174,70	55,67	42,35	120,55	8,10	4,26		1,51	1,51
158 GLN	176,43	56,90	29,01	119,86	8,23	4,18		1,98	1,98
159 GLU	175,25	56,70	28,98	120,46	8,30	4,22		2,05	1,94
160 SER	174,19	58,48	63,73	118,36	8,12	4,37		3,82	
161 LEU	177,29	55,19	42,28	123,79	8,03	4,29		1,62	1,54
162 ARG	176,29	57,05	30,94	120,72	8,09	4,17		1,81	1,70
163 SER	174,18	58,48	63,75	116,78	8,20	4,37		3,82	3,82
164 LYS	175,62	56,31	32,97	123,10	8,25	4,28		1,82	1,70
165 GLU		56,90	29,51	125,64	8,02	4,15		2,04	1,87

Chemical shifts of the backbone atoms were obtained by a combination of HNCACB, CBCA(CO)NH, HBHA(CO)NH, HNCO, HNCA, HSQC and HNHA experiments. These chemical shifts perfectly fitted the peaks for those various experiments. From these values, attribution of the most aliphatic side chain atoms was possible with HCCH-COSY and HCCH-TOCSY, with the exception of some end of the terminal amino acids. The chemical shift values of all carbon chemical shifts from the side chains are found in Appendix B, while the proton and nitrogen chemical shifts are found in Appendix C, due to the amount of atoms included in the table.

Appendix B: Chemical shift table for the carbon atoms of the side chains.

Residues	$\sigma^{13}\text{C}\gamma$	$\sigma^{13}\text{C}\gamma_2$	$\sigma^{13}\text{C}\delta$	$\sigma^{13}\text{C}\delta_2$	$\sigma^{13}\text{C}\epsilon$	$\sigma^{13}\text{C}\text{H}$	$\sigma^{13}\text{C}\zeta$	$\sigma^{13}\text{C}\zeta_2$
1 CYS								
2 ASP								
3 LEU	26,88		24,98	24,40				
4 PRO	27,61		50,44					
5 GLN	33,92							
6 THR	22,05							
7 HIS				120,20	136,00			
8 SER								
9 LEU	27,05		24,25	24,34				
10 GLY								
11 SER								
12 ARG	27,90		43,31					
13 ARG			43,70					
14 THR	21,20							
15 LEU	26,71		25,04	24,05				
16 MET	31,66				16,94			
17 LEU	23,18		26,26	22,98				
18 LEU	27,44		25,29	23,74				
19 ALA								
20 GLN	34,65							
21 MET	32,70				15,31			
22 ARG	26,49		43,99					
23 LYS	25,20		29,77		41,84			
24 ILE	26,49	17,53	14,43					
25 SER								
26 LEU	27,08		24,73	25,05				
27 PHE								
28 SER								
29 CYS								

Residues	$\sigma^{13}\text{C}\gamma$	$\sigma^{13}\text{C}\gamma_2$	$\sigma^{13}\text{C}\delta$	$\sigma^{13}\text{C}\delta_2$	$\sigma^{13}\text{C}\epsilon$	$\sigma^{13}\text{C}\text{H}$	$\sigma^{13}\text{C}\zeta$	$\sigma^{13}\text{C}\zeta_2$
30 LEU	25,16		25,09	23,11				
31 LYS	24,79		29,03		42,62			
32 ASP	175,64							
33 ARG	27,19		43,17					
34 HIS				120,20				
35 ASP								
36 PHE								
37 GLY								
38 PHE			131,75		130,55			
39 PRO			50,52					
40 GLN	33,07							
41 GLU	34,30							
42 GLU	33,06							
43 PHE				132,10				
44 GLY								
45 ASN								
46 GLN	33,08							
47 PHE			132,10		130,20			
48 GLN	33,97							
49 LYS	24,96		30,46		41,94			
50 ALA								
51 GLU	34,75							
52 THR	22,61							
53 ILE	29,74	17,85	12,69					
54 PRO	28,71		49,61					
55 VAL	22,34	22,42						
56 LEU	24,31		24,10	24,44				
57 HIS				119,70				
58 GLU								
59 MET					17,99			
60 ILE	29,60	19,50	12,50					

Residues		$\sigma^{13}\text{C}_\gamma$	$\sigma^{13}\text{C}_{\gamma_2}$	$\sigma^{13}\text{C}_\delta$	$\sigma^{13}\text{C}_{\delta_2}$	$\sigma^{13}\text{C}_\epsilon$	$\sigma^{13}\text{C}_\text{H}$	$\sigma^{13}\text{C}_\zeta$	$\sigma^{13}\text{C}_{\zeta_2}$
61	GLN	33,93	18,89	14,60	27,40	130,90	124,30	127,62	122,40
62	GLN	32,92							
63	ILE	29,13							
64	PHE								
65	ASN								
66	LEU	27,14							
67	PHE								
68	SER								
69	THR	21,34							
70	LYS	24,69							
71	ASP		129,75	24,49	42,03	122,95	124,30	114,20	122,40
72	SER								
73	SER								
74	ALA								
75	ALA								
76	TRP								
77	ASP								
78	GLU	34,44							
79	THR	21,25							
80	LEU	26,07							
81	LEU	26,48							
82	ASP		26,24	25,66	22,99	23,29	38,25	130,90	128,40
83	LYS	29,52							
84	PHE								
85	TYR								
86	THR	21,89							
87	GLU	33,97							
88	LEU	25,86							
89	TYR								
90	GLN	33,97							
91	GLN	35,30							

Residues	$\sigma^{13}\text{C}_\gamma$	$\sigma^{13}\text{C}_{\gamma_2}$	$\sigma^{13}\text{C}_\delta$	$\sigma^{13}\text{C}_{\delta_2}$	$\sigma^{13}\text{C}_\epsilon$	$\sigma^{13}\text{C}_\text{H}$	$\sigma^{13}\text{C}_\zeta$	$\sigma^{13}\text{C}_{\zeta_2}$
92 LEU	26,45		24,25	23,67				
93 ASN								
94 ASP								
95 LEU	26,93		26,22	24,00				
96 GLU	33,72							
97 ALA								
98 CYS								
99 VAL	21,09	21,69						
100 ILE	27,77	17,53	12,74					
101 GLN	33,16							
102 GLY								
103 VAL	21,06	20,95						
104 GLY								
105 VAL	21,00	20,25						
106 THR	21,25							
107 GLU	32,94							
108 THR	22,15							
109 PRO	28,02		50,60					
110 LEU	27,00		24,18	24,21				
111 MET	32,75				17,26			
112 LYS			29,61		41,99			
113 GLU	33,90							
114 ASP								
115 SER								
116 ILE	29,12	16,81	12,48					
117 LEU	27,05		24,93	24,02				
118 ALA								
119 VAL	22,27	23,73						
120 ARG								
121 LYS	29,49							
122 TYR								

Residues	$\sigma^{13}\text{C}\gamma$	$\sigma^{13}\text{C}\gamma_2$	$\sigma^{13}\text{C}\delta$	$\sigma^{13}\text{C}\delta_2$	$\sigma^{13}\text{C}\epsilon$	$\sigma^{13}\text{C}\text{H}$	$\sigma^{13}\text{C}\zeta$	$\sigma^{13}\text{C}\zeta_2$
123 PHE								
124 GLN	33,88							
125 ARG	30,87							
126 ILE	28,51	18,72	12,10					
127 THR	21,98							
128 LEU	27,31		24,42	24,32				
129 TYR			133,20		118,20			
130 LEU	26,51		25,07	21,58				
131 LYS	25,23		29,52		41,99			
132 GLU	33,99							
133 LYS	23,17		27,75		42,68			
134 LYS	25,30		29,45		42,35			
135 TYR			133,30		118,50			
136 SER								
137 PRO			50,49					
138 CYS								
139 ALA								
140 TRP			127,10		121,90	124,57	114,65	120,40
141 GLU	33,86							
142 VAL	23,37	20,60						
143 VAL	22,13	25,28						
144 ARG								
145 ALA								
146 GLU	33,04							
147 ILE	28,28	18,02	10,99					
148 MET	32,95				17,30			
149 ARG	27,38		43,02					
150 SER								
151 PHE								
152 SER								
153 LEU	25,52		25,24	23,55				

Residues	$\sigma^{13}\text{C}_\gamma$	$\sigma^{13}\text{C}_{\gamma_2}$	$\sigma^{13}\text{C}_\delta$	$\sigma^{13}\text{C}_{\delta_2}$	$\sigma^{13}\text{C}_\epsilon$	$\sigma^{13}\text{C}_\text{H}$	$\sigma^{13}\text{C}_\zeta$	$\sigma^{13}\text{C}_{\zeta_2}$
154 SER								
155 THR	21,68							
156 ASN								
157 LEU	25,33		25,09	23,79				
158 GLN	31,81							
159 GLU	34,71							
160 SER								
161 LEU	26,69		25,21	23,14				
162 ARG	27,15		43,38					
163 SER								
164 LYS	24,97		29,04		42,04			
165 GLU	34,06							

Chemical shifts of the carbon side chain atoms were obtained by extending the chemical shifts assignment from $\text{C}\alpha/\text{H}\alpha$ and $\text{C}\beta/\text{H}\beta$ chemical shifts obtained initially from the HNCACB, CBCA(CO)NH, HBHA(CO)NH, HNCO, HNCA, HSQC and HNHA experiments. By using the HCCH-COSY and HCCH-TOCSY experiments, it was possible to obtain almost all carbon chemical shifts from aliphatic side chains, although some end of the terminal amino acids and aromatic side chains were not obtained due do severe overlapping chemical shifts. This was especially true for the GLU and PHE, where their C_γ and C_δ atoms were almost always in the same area.

Appendix C: Chemical shift table for the proton and nitrogen atoms of the side chains.

Residues	$\sigma^1\text{H}\gamma$	$\sigma^1\text{H}\gamma_2$	$\sigma^1\text{H}\delta$	$\sigma^1\text{H}\delta_2$	$\sigma^1\text{H}\epsilon$	$\sigma^1\text{H}\epsilon_2$	$\sigma^1\text{H}\text{H}$	$\sigma^1\text{H}\zeta$	$\sigma^1\text{H}\zeta_2$	$\sigma^{15}\text{N}\delta$	$\sigma^{15}\text{N}\epsilon$
1 CYS											
2 ASP											
3 LEU	1,61		0,81	0,84							
4 PRO	1,95		3,60	3,72							
5 GLN	2,37										
6 THR	1,12										
7 HIS			7,33		8,55						
8 SER											
9 LEU	1,55		0,86	0,83							
10 GLY											
11 SER											
12 ARG	1,54	1,66	3,12	3,13							
13 ARG	1,40		3,13	3,15							
14 THR	0,51										
15 LEU	1,71		0,90	0,87							
16 MET	2,43	2,61			2,01						
17 LEU	1,69		0,82	0,78							
18 LEU	1,12		0,39	0,17							
19 ALA											
20 GLN	2,33	2,54									
21 MET	2,04	2,12			0,42						
22 ARG	1,44	2,08	3,05	3,33							
23 LYS	1,51	1,98	1,66	1,76	3,05						
24 ILE	1,24/1,25	0,85	0,70								
25 SER											
26 LEU	1,62		0,83	0,78							
27 PHE			7,25		7,36			7,32			
28 SER											
29 CYS											

Residues	$\sigma^1\text{H}\gamma$	$\sigma^1\text{H}\gamma_2$	$\sigma^1\text{H}\delta$	$\sigma^1\text{H}\delta_2$	$\sigma^1\text{H}\epsilon$	$\sigma^1\text{H}\epsilon_2$	$\sigma^1\text{H}\text{H}$	$\sigma^1\text{H}\zeta$	$\sigma^1\text{H}\zeta_2$	$\sigma^{15}\text{N}\delta$	$\sigma^{15}\text{N}\epsilon$
30 LEU	1,73		0,90	0,79							
31 LYS	1,37		1,34		2,93						
32 ASP											
33 ARG	1,66	1,73	3,11								
34 HIS			7,38								
35 ASP											
36 PHE			6,98		6,80						
37 GLY											
38 PHE			6,17		6,88						
39 PRO	1,28	1,55	2,64	3,17							
40 GLN	2,30	2,34			6,80	7,56					111,07
41 GLU	2,20										
42 GLU	2,05	2,30									
43 PHE			7,11		6,97						
44 GLY											
45 ASN			6,90	7,55						111,81	
46 GLN	1,96				6,77	7,23					111,76
47 PHE			7,19		7,26						
48 GLN	2,34				6,85	7,51					111,83
49 LYS	1,39		1,62		2,93						
50 ALA											
51 GLU	2,36										
52 THR	1,23										
53 ILE	1,22	1,73	0,85								
54 PRO	1,92	2,13	3,58	3,62							
55 VAL	0,79	0,64									
56 LEU	1,73		0,89	0,91							
57 HIS			7,20		8,62						
58 GLU	2,36										
59 MET					2,01						
60 ILE	1,18/1,72	0,77	0,86								

Residues	$\sigma^1\text{H}\gamma$	$\sigma^1\text{H}\gamma_2$	$\sigma^1\text{H}\delta$	$\sigma^1\text{H}\delta_2$	$\sigma^1\text{H}\epsilon$	$\sigma^1\text{H}\epsilon_2$	$\sigma^1\text{H}\text{H}$	$\sigma^1\text{H}\zeta$	$\sigma^1\text{H}\zeta_2$	$\sigma^{15}\text{N}\delta$	$\sigma^{15}\text{N}\epsilon$
61 GLN	2,26				6,73	7,19					113,17
62 GLN	1,88	2,63			7,08	8,24					114,00
63 ILE	1,17/1,60	0,90	0,74								
64 PHE			6,93		6,89			6,98			
65 ASN											
66 LEU	1,24		0,85	0,87							
67 PHE			6,84	7,49	6,84			6,79			
68 SER											
69 THR	1,55										
70 LYS	1,39		1,33	1,33	2,96						
71 ASP											
72 SER											
73 SER											
74 ALA											
75 ALA											
76 TRP			7,33		6,79	7,63	7,17	7,82	7,33		121,68
77 ASP											
78 GLU	2,43	2,58									
79 THR	1,18										
80 LEU	1,60		0,73	0,64							
81 LEU	1,47		0,91	0,74							
82 ASP											
83 LYS	1,48				3,04						
84 PHE			6,68	6,69	6,46			6,20			
85 TYR			6,88		6,49						
86 THR	1,23										
87 GLU	2,38										
88 LEU	1,14		0,28	0,07							
89 TYR			6,89		6,45						
90 GLN	2,36	2,44			6,84	7,49					112,71
91 GLN	2,27	2,30			6,82	7,56					111,80

Residues	$\sigma^1\text{H}\gamma$	$\sigma^1\text{H}\gamma_2$	$\sigma^1\text{H}\delta$	$\sigma^1\text{H}\delta_2$	$\sigma^1\text{H}\epsilon$	$\sigma^1\text{H}\epsilon_2$	$\sigma^1\text{H}\text{H}$	$\sigma^1\text{H}\zeta$	$\sigma^1\text{H}\zeta_2$	$\sigma^{15}\text{N}\delta$	$\sigma^{15}\text{N}\epsilon$
92 LEU	1,47		0,33	0,30							
93 ASN			6,42	7,35						111,58	
94 ASP											
95 LEU	1,71		0,88	0,89							
96 GLU	2,40	2,44									
97 ALA											
98 CYS											
99 VAL	0,81	0,84									
100 ILE	1,18/1,39	0,85	0,78								
101 GLN	2,43				6,82	7,46					111,71
102 GLY											
103 VAL	0,90	0,89									
104 GLY											
105 VAL	0,87	0,90									
106 THR	1,19										
107 GLU	2,38										
108 THR	1,35										
109 PRO	1,96	2,08	3,85								
110 LEU			0,87	0,82							
111 MET	2,54				2,07						
112 LYS			1,49		2,84						
113 GLU	2,35	2,38									
114 ASP											
115 SER											
116 ILE	1,18/1,60	0,78	0,74								
117 LEU	1,70		0,92	0,85							
118 ALA											
119 VAL	1,22	1,32									
120 ARG	1,94										
121 LYS					3,41						
122 TYR			6,98								

Residues	$\sigma^1\text{H}\gamma$	$\sigma^1\text{H}\gamma_2$	$\sigma^1\text{H}\delta$	$\sigma^1\text{H}\delta_2$	$\sigma^1\text{H}\epsilon$	$\sigma^1\text{H}\epsilon_2$	$\sigma^1\text{H}\text{H}$	$\sigma^1\text{H}\zeta$	$\sigma^1\text{H}\zeta_2$	$\sigma^{15}\text{N}\delta$	$\sigma^{15}\text{N}\epsilon$
123 PHE			7,58		7,44						
124 GLN	2,40	2,57			6,80	7,48					110,40
125 ARG											
126 ILE	-0,6/0,73	0,51	-0,50								
127 THR	1,24										
128 LEU	1,67		0,90	0,87							
129 TYR			6,43		6,45						
130 LEU	1,18		-0,06	-0,50							
131 LYS	1,34	1,50	1,62		2,92						
132 GLU	2,26	2,42									
133 LYS	0,39	0,74	1,09	1,24	2,74	2,64					
134 LYS	1,26		1,62		2,95						
135 TYR			6,68		6,56						
136 SER											
137 PRO			3,88								
138 CYS											
139 ALA											
140 TRP			6,81		9,63	8,24	7,02	7,01	6,46		126,56
141 GLU	2,37										
142 VAL	1,00	0,97									
143 VAL	0,88	0,83									
144 ARG											
145 ALA											
146 GLU	2,20										
147 ILE	1,50/1,78	0,95	0,81								
148 MET	2,35	2,55			2,06						
149 ARG	1,32	1,63	3,13	3,28							
150 SER											
151 PHE			6,92		7,24	7,42		7,31			
152 SER											
153 LEU			0,79	0,80							

Residues	$\sigma^1\text{H}\gamma$	$\sigma^1\text{H}\gamma_2$	$\sigma^1\text{H}\delta$	$\sigma^1\text{H}\delta_2$	$\sigma^1\text{H}\epsilon$	$\sigma^1\text{H}\epsilon_2$	$\sigma^1\text{H}\zeta$	$\sigma^1\text{H}\zeta_2$	$\sigma^{15}\text{N}\delta$	$\sigma^{15}\text{N}\epsilon$
154 SER										
155 THR	0,62									
156 ASN			6,83	7,48					111,54	
157 LEU	1,62		0,76	0,76						
158 GLN	2,32				6,74	7,44				111,08
159 GLU	2,37	2,38								
160 SER										
161 LEU	1,61		0,85	0,78						
162 ARG	1,53		3,13							
163 SER										
164 LYS	1,35		1,64		2,95					
165 GLU	2,30									

Chemical shifts of the carbon side chain atoms were obtained by extending the chemical shifts assignment from $\text{C}\alpha/\text{H}\alpha$ and $\text{C}\beta/\text{H}\beta$ chemical shifts obtained initially from the HNCACB, CBCA(CO)NH, HBHA(CO)NH, HNCO, HNCA, HSQC and HNHA experiments. By using the HCCH-COSY and HCCH-TOCSY experiments, it was possible to obtain almost all proton chemical shifts from aliphatic side chains, although some end of the terminal amino acids and aromatic side chains were not obtained due do severe overlapping chemical shifts. This was especially true for the GLU and PHE, where their $\text{C}\gamma$ and $\text{C}\delta$ atoms were almost always in the same area, causing proton chemical shifts to be very ambiguous. In the case of isoleucines, two values for $\text{H}\gamma$ are found, since there are two different protons on $\text{C}\gamma$, while a methyl group is found on $\text{C}\gamma_2$.

Appendix D: SDS-PAGE of the glycosylation process

Glycosylation SDS-PAGE showing the protein size before (column 2) and after (column 3-10) glycosylation of interferon alpha-2a with GalNAc. The paler band above the protein represents a small population of slower migrating protein due to the presence of SDS.

This figure shows the SDS-PAGE of the glycosylation product at different intervals of time since glycosylation started. The lighter band over the initial dark protein band before glycosylation occurred represents a certain amount of the protein which had a disulfide bond breakage, varying their migration through the gel (Column 3). As the glycosylation process takes place inside the reaction mix, the unglycosylated band becomes lighter, indicating it was being glycosylated. It completely vanishes between 3-4 hours, with more than 97% of the protein having been glycosylated. This also shows the high purity of the protein, where no impurities are being detected on the SDS-PAGE. It was obtained by our collaborators at the NRC.