HETEROLOGOUS EXPRESSION OF RAT ST6GAL-I IN PICHIA PASTORIS FOR STRUCTURAL AND FUNCTIONAL STUDIES

by

Narendra Gurnamal Tejwani

(Under the direction of Kelley Moremen)

Abstract

ST6Gal-I plays an important role in immune regulation mediated via the lectin CD22. The enzyme catalyses transfer of sialic acid to form a Siaα2-6Galβ1-4GlcNAc (Sia6LacNAc) tri-saccharide product, which is ligand for CD22. Despite such a critical role in immune function, no structural data is yet available for ST6Gal-I or any other GT29 enzyme. A major limitation in the structural analysis of glycosylation enzymes, including ST6Gal-I, is the large-scale expression and purification of the glycosylated enzymes in forms compatible with structural analysis.

We have successfully expressed a functional, soluble form of the recombinant ST6Gal-I catalytic domain using the *Pichia pastoris* expression system. The enzymatic properties of the ST6Gal-I were comparable to those of rat liver ST6Gal-I with regard to substrate specificity and stability. We have also established conditions for N-glycan removal from the recombinant ST6Gal-I for structural and functional studies. The results suggest that the enzymatic properties of the deglycosylated enzyme were comparable to that of glycosylated enzyme.

Conditions for stable isotope labeling for structural characterization by NMR have been established and initial studies on the spectral assignment have been initiated. Our NMR data supports the theory of substrate induced conformational changes for ST6Gal-I.

Also, here we have demonstrated that any chemical modification of ST6Gal-I, which are essential for immobilization of enzyme for substrate binding studies using SPR, leads to inactivation of enzyme.

INDEX WORDS: ST6Gal-I, Sialyltransferase, Pichia pastoris, NMR, Crystallization, SPR, Biacore

HETEROLOGOUS EXPRESSION OF RAT ST6GAL-I IN PICHIA PASTORIS FOR STRUCTURAL AND FUNCTIONAL STUDIES

by

Narendra Gurnamal Tejwani

B. Pharm, Nashik District Maratha Vidhya Prasarak Samaj Medical College, India, 1997

M.S, Mumbai University Institute of Chemical Technology, India, 1999

Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2008

HETEROLOGOUS EXPRESSION OF RAT ST6GAL-I IN PICHIA PASTORIS FOR STRUCTURAL AND FUNCTIONAL STUDIES

by

Narendra Gurnamal Tejwani

Major Professor:	Kelley Moremen
Committee:	Michael Tiemeyer
	Michael Pierce
	Lance Wells

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2008

DEDICATION

I would like to dedicate this to my parents Durga and Gurnamal Tejwani and my friend Mangesh Chordiya, without whom I wouldn’t have been here.

ACKNOWLEDGMENTS

My sincerest gratitude goes to my advisor, Dr. Kelley Moremen for his patience, guidance, and support during the course of graduate studies and this project. I would also like to thank the members of my committee, Dr. Michael Pierce, Dr. Michael Tiemeyer, and Dr. Lance Wells for their invaluable suggestions and support.

My deepest thanks to Dr. James Prestegard, Dr. John Glushka, and Dr. Fang Tian who collaborated with me on the NMR front. Also, my special thanks to Dr. Xu Wang who helped me to interpret the NMR data. I would also like to thank Dr. Renu Kadirvelraj, Dr. B.C. Wang and Jessi Chang (from Dr. B.C. Wang's lab) for their sincere efforts in crystallizing the proteins. Thanks to Dr. Parastoo Azadi, who did the glycan analysis and Dr. Gerardo Alvarez-Manilla for providing glycoprotiens for SPR studies.

I would also like to thank Ms. Trisha Sheahan for her moral support and Dr. Khanita Karaveg who helped with the SPR analysis. Also, thanks to Dr. Alison Nairn, Dr. Lu Meng, Yong Xiang, Heather Strachan, Charu Kumar, Mitche Dela Rosa and Dr. Steve Mast, and all the other members in Dr. Moremen & Dr. Pierce's lab for their fellowship and encouragement.

I would like to thank my grandmother Hujibai Tejwani for her blessings and love. I am deeply indebted to my parents, Durgabai and Gurnamal Tejwani for their support and encouragement. I am grateful to Neetu and Kartar (my younger siblings) for their belief in me. I would also like to thank my aunt and uncle, Aruna and Mahesh Zambani for their support and understanding when needed the most.

Lastly, I would like to thank my wife for her endless patience, love & care.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.0 Introduction

Mammalian cell surfaces are covered with complex carbohydrates covalently linked to lipids and proteins. Protein linked glycans are attached either via an amide linkage to Asn side-chains within an Asn-X-Ser/Thr conserved sequence (N-glycan sequon) or via an O-linkage to the hydroxyl group of Ser or Thr residues [1]. The biosynthesis of these complex carbohydrates and polysaccharides is of great biological importance. These molecules govern a diverse range of cellular functions, including cell-cell recognition, cell differentiation and various receptor-ligand interactions with extracellular macromolecules [1-3]. Because these recognition functions potentially act as a cellular language, they must rely on precise carbohydrate diversity. Therefore, it is believed that the biosynthesis of oligosaccharides and polysaccharides may involve the action of hundreds of different and selective superfamilies of enzymes, termed as glycosyltransferases (GTs) [4]. The members of these superfamilies are commonly named according to the type of sugars they transfer (i.e. galactosyltransferase, Sialyltransferase etc.). These enzymes transfer sugar moieties from activated sugar nucleotide donor molecules to specific acceptor molecules in a linkage, donor, and acceptor-specific manner. Although there are many important glycosyltransferases needed for production of the extremely complex glycans chains observed in nature, the focus of this dissertation is on sialyltransferases. Sialyltransferases are of particular interest because they transfer sialic acid, a charged sugar,

which provides the "finishing touches" to cap the non-reducing temini on large oligosaccharide structures.

1.1 Sialic acids

In order to understand the significance of studying this class of enzymes, it is important to understand and appreciate the functional importance of sialic acids in a number of biological processes. However, the purpose of this introduction is not to provide a comprehensive review on the subject of sialic acids, but to merely highlight some of the key roles that sialic acid plays in these biological processes. This information will emphasize the importance of how studying the enzymes involved in sialic acid regulation could lead to ways in which a wide array of biological pathways could be monitored and potentially controlled.

Historically, sialic acid is a common name for the *N*- or *O*-substituted derivatives of a nine-carbon carboxylated sugar. It is *also* the name for the most common member of this group, N-acetylneuraminic acid (Neu5Ac or NANA) (Figure 1). Sialic acid was first isolated and named after its discovery by mild acid hydrolysis of saliva (Greek word sialon), by the German biochemist Gunnar Blix in 1936 [5, 6]. Later in 1941, Klenk discovered a similar crystalline material form by mild acid hydrolysis of brain matter (neuramine) and named it as Neuraminic acid [5, 6]. It was Gottschalk who first proposed the structure for sialic acid [7]. All three investigators agreed to use "sialic acid" as the family name covering all of the more than 40 derivatives of neuraminic acid (Neu), with N-acetylneuraminic acid (Neu5Ac) and N-glycolylneuraminic acid (Neu5Gc) forming the most highly abundant core structures [6]. Later in 1950s and 1960s, various characteristics of sialic acids, including their complete structure,

chemistry and biosynthesis, were established by various independent groups [8]. The early history of sialic acid have been well documented in the article by Faillard [6].

1.2 Sialic acid Structure and its diversity

Sialic acids are found widely distributed in nature from higher animal to archaea. Also, it is the most diverse sugar in nature. There are over 50 structures of sialic acid found in nature [7, 9]. These sugars usually appear on the non-reducing termini of N-glycans, O-glycans, and glycosphingolipids. The structural variability in the sialic acid is attributed to modifications at positions -4, -5, -7, -8, and -9 [9, 10] (Figure 2). All the know sialic acids in nature are acylated on the C-5 amino group of neuraminic acid (Neu) giving two main groups, the N-acetyl and N-glycolyl neuraminic acids. However, a free amino sugar Neu has itself never been found in nature. The C-5 deaminated sugar 3-deoxy-D-glycero-D-galacto-2-nonulosonic acid (KDN) is an additional member of the sialic acid family [11]. In addition to the *N*-acetyl and *N*-glycolyl groups of sialic acids, other common modifications to the ring or exocyclic hydroxyls include one to three acetyl groups added at C-4 or to the C-7 to C-9 hydroxyls of the glycerol side chain [9, 12, 13]. Also, other modifications of hydroxyls such as lactyl, methyl, sulfate, and phosphate groups have been found, but they are less common [5, 8]. Some of these different modifications may be combined (e.g., 8-*O*-methyl with 9-*O*-acetyl and *N*-glycolyl) yielding the manifold types of sialic acids found throughout the animal kingdom. Apart from these types, unsaturated Sia as well as anhydro [10] and lactone (KDN) forms have also been identified in various biological sources.

Furthermore, diversity in sialic acid increases several fold by different α-linkages from the 2-carbon to underlying sugar chains (Figure 4). Of these, the most common are to the 3- or 6-

position of Gal residues or to the 6-position of GalNAc residues. In some instances, sialic acids can also occupy internal positions within glycans, the most common situation being another sialic acid residue attached to the 8-position, although sometimes they occupy 9-position.

The aforementioned substitutions coupled with various different glycosidic linkages generate greater possibilities in which sialic acids can attach themselves to the glycoconjugates on the cell surface. This structural diversity of sialic acids can determine and/or modify the recognition by antibodies, as well as by a variety of sialic-acid-binding lectins of endogenous and exogenous origin.

Sialic acids are not omni-present in nature, but their ubiquitous presence in higher animals is well-established. They have also been found in some microorganisms such as fungi, viruses, bacteria and protozoa [14]. In humans, sialic acids are almost exclusively derived from Neu5Ac, although NeuGc has been found in small amounts [15, 16]. Even though sialic acids are usually found as glycosidically linked to the non-reducing terminal or terminal side units on cell surface glycoconjugates, free sialic acid have also been found in body fluids [17, 18].

1.3 Biological Significance of Sialic acids

Sialic acids are present at the terminal positions of glycans on glycoproteins and gangliosides on the outer cell membrane, either alone or in oligo- or polymeric form. The carboxylate group at the position 1 of sialic acids remains inonized at physiological pH (pKa ≈ 2.6) and confers the negative charge that dominates the physicochemical properties of the family of glycans [19]. Thus, their exposure to the outer environment coupled with their physical properties (negative charge, hydrophilicity, size and structural diversity), makes them a great candidate to influence cell biology and physiology. Today sialic acids are implicated in many

complex cell functionalities including various physiological and pathological events, including cellular and molecular recognition processes [19].

In general, irrespective of their great structural diversity, sialic acids are involved with three main functions: protection, masking underlying glycoconjugate structure to prevent biological recognition and act as recognition receptor for binding and adhesion [9]. First, functional protection resuls from the negative charge of the anomeric carboxylate group, and is non-specific in nature. For example, the negative charge on sialic acids provides this molecule with the ability to attract and repel specific cells and biomolecules [20]. On one hand, negatively charged sialic acids attract positively charged molecules, resulting in the binding of positively charged molecules. On the other hand, sialylated cell surfaces have a net negative charge that is essential for the repulsion of other cells or biomolecules [19]. For example, cell surface sialic acids on erythrocytes and blood platelets prevent the aggregation of these cells in the bloodstream [9]. Furthermore, charge repulsion of heavily sialylated mucins increases the viscosity of the mucus lining, thus protecting the underlying endothelia in the intestine. A similar role of mucin structures in protections occurs on the epidermal surface of fish and frog eggs [21]. Sialic acid modification also influences and stabilizes the conformation of glycoproteins, resulting in modified properties such as higher proteolytic stability [22] and resistance to the action of some endoglycosidases [22]. Thus, electrochemical properties of sialic acids modulate the functions of glycoconjugates in a variety of biological functions.

The other two functions, the masking and recognitions of sialic acids, have opposite effects. In masking, the sialic acid prevents the recognition of subterminal galactose residues, which can serve as recognition sites for galactose-specific lectins or cells. For example, the uptake of desialylated serum glycoproteins was increased by hepatocytes [23]. Similarly, there

was a sharp increase in phagocytosis followed by degradation of desialylated red blood cells resulting from old age or under the influence of sialidase activity from microbial infections [24]. Sialidase-treated lymphoctes and thrombocytes are bound and partly phagocytized by marcophages, mediated by a galactose-specific lectin [24, 25].

In some autoimmune diseases, such as glomerulonephritis, bacterial infection causes desialylation of the cell membrane, resulting in the production of auto-antibodies [26]. Nonetheless, sialic acids have also been implicated in the masking of cancer antigens [27]. The high level of sialic acids on tumor cells protect them from immunological attack and in most cases allow them to grow uncontrollably. The sialylation level is directly proportional to the degree of invasion of a cancer[28].

In contrast to masking, sialic acids are directly involved in a variety of recognition processes as well. The most common mechanism for various microorganisms and viruses to bind to host cells is via sialic acid recognition. The classical example of this is sialic acid mediated binding of influenza virus to human erythrocytes and mucins of the respiratory tract [29]. The viral lectin, hemagglutinin, binds to sialic acid bearing cell receptors and binding is reversed by viral sialidase, which releases the terminal sialic acid. Furthermore, viral sialidase helps in the spreading of virus in tissues by preventing their further attachment to cells and to mucus layers protecting epithelia of the respiratory tract [20].

Another example of the recognition functions of sialic acids is their role in neural development, in which poly-sialic acid is attached to neural cell adhesion molecule (NCAM) [30]. The loss of poly-sialic acid function in mice results in decreased neural cell adhesion, defective cell migration, neutrite out growth and defective synapse formation, suggesting its role in regulation of neural development [30, 31]. Furthermore, it has been demonstrated that there is

a direct correlation between loss of poly-sialic acid function in various diseases such as schizophrenia and Alzeimer's disease [32, 33].

The the recognition effect of sialic acid is also reflected in its binding and recognition by the siglecs in the immune system. Siglecs are sialic acid-binding immunoglobulin-like lectins involved in cellular signaling functions and cell-cell interactions in the nervous and immune systems [34]. Siglecs bind to the sialylated ligands on the cell surface in both *cis* and *trans* mode. Siglec receptor-binding sites are masked by *cis* interactions with sialic acid ligands [34, 35]. These *cis* interactions with sialic acids are essential for the regulation of siglec function by preventing or facilitating specific cell-cell interactions when necessary. For example de-masking of CD22, an α2,6-linked sialic acid binding lectin on B cells causes the internalization of CD22 and thus negatively influences the B-cell mediated immune response [36-38]. Similarly Myelin-associated glycoprotein (MAG), an α2,3-linked sialic acid binding lectin, contributes to the prevention of axonal regeneration in the adult central nervous system. However, neuraminidase treatment of axotomized entorhino-hippocampal cultures potentiates axonal regeneration [39].

There are many more physiological processes that are also regulated by sialic acid recognition, and this field of research is rapidly expanding. Moreover, improper processing of sialic acids leads to diseases such as sialidosis, galactosialidosis, sialuria and sialic acid storage disorder (SASD)[40]. There is an enormous literature that corroborates the function of sialic acids in biology, as noted in several reviews by Schauer [9, 20, 41], which provided a very concise and useful summary.

1.4 Sialyltransferases

Sialyltransferases, a family of enzymes, are involved in the transfer of sialic acids on the carbohydrate moieties of glycoproteins and glycolipds. It has been proposed that the human genome alone encodes for more than 20 sialyltransferases to accommodate all the sialylated structures described to date. So far, 22 mammalian sialyltransferase cDNA have been cloned and characterized [42, 43]. Most notably, the same glycosidic linkage can be formed by different sialyltransferase enzymes.

1.4.1 Classification

When sialyltransferases catalyse the transfer of Neu5Ac from Neu5Acβ2CMP onto a terminal sugar of glycocnjugates there is inversion of configuration at the anomeric centre. Therefore, sialyltransferases are classified into four broad categories based upon the type of glycosidic linkage formed and the nature of the sugar acceptor (Figure 3). These enzymes are referred as ST3Gal, ST6Gal, ST6GalNAc and ST8Sia to designate the transfer of Neu5Ac to the 3^{rd}, 6^{th}, 6^{th} or 8^{th} hydroxyl positions of the glycone acceptor, respectively [44]. Sialyltransferase of the ST3Gal subfamily transfers Neu5Ac onto terminal Gal units. However, the nature of the subterminal units varies amongst members [45]. For example, both the ST6Gal and ST6GalNAc subfamily members transfer Neu5Ac onto terminal Gal and GalNAc respectively i.e. they exhibit different acceptor specificities for adjacent sugar residues [45]. The ST8Sia subfamily differs from the other two subfamilies; it transfers Neu5Ac to other sialic acid residues forming short or long homopolymers in glycoproteins and to other sialic acid residues of gangliosides [45].

1.4.2 Commonalities amongst Sialyltransferase

1.4.2.1 Overall structure

Despite the fundamental roles of the sialyltransferases, there is still limited information on their protein structure, mechanism of action, and on the cellular mechanism involved in regulation of their transcript expression patterns. The primary protein sequence of the cloned enzyme shows that all sialyltransferases are type II membrane proteins, similar to other glycosyltransferases [43]. They all have four common domains (1) a short N-terminal cytoplasmic tail (3-11 a.a.) non essential for catalytic activity, (2) a single pass transmembrane domain (13-18 a.a.), (3) a proteolytically sensitive and proline rich stem domain (30-200 a.a.) and (4) large C-terminal catalytic domain (300-350 a.a.) [46] (Figure 4). Primary sequence data of cloned mammalian sialyltransferase shows very little homology, with the exception of three conserved domains named sialylmotifs L (long), S (short), and VS (very short) in their catalytic domain [47-49] (Figure 5). Sialylmotifs are not ubiquitous in the sialyltransferase family. Bacterial sialyltransferases do not contain of any of these motifs, and in fact belong to different glycosyltransferase CAZy families, (e.g. sialyltransferase CstII from *Campylobacter jejuni* [50] and *Pasteurella multocida* [51] belongs to CAZy family GT42 and GT80 respectively), as compared to the mammalian sialyltransferases which belongs to the GT29 family of glycosyltransferases.

1.4.2.2 Sialyl motifs

The contribution of each motif in enzyme function and mechanism has been studied using site-directed mutagenesis of ST6Gal-I as a model enzyme. The L motif, which comprises of 40-60 amino acids, is situated in the middle of catalytic domain and it binds to the donor

substrate Neu5Ac-CMP [47], the S motif is situated toward the C terminal end and consists about 20-30 amino acids, and is involved in binding to both donor and acceptor substrates [48], and the VS motif, which contains glutamic acid and histidine separated by four other amino acid, has been proposed to participate in the catalytic inverting process, although it is yet to be proven. Recently, it was determined, using ST3Gal-I, that a new motif, the peptide motif III (H/y)Y(Y/W/F/h)(D/E/q/g), participates in catalysis [49], where capital letters signify the occurrence and requirement for enzyme activity. Among all known mammalian sialyltransferases sequences there are 5 and 2 absolutely conserved residues in sialylmotifs L and S, respectively (Figure 5B). There is one cysteine amino acid conserved in each motif. Mutations of any of the two conserved cysteines result in inactive enzymes [47, 48, 52]. It was shown, using both ST6Gal-I and the polysialytransferase, ST8Sia-IV, as model proteins, that the two cysteines form an intramolecular disulfide bond that is essential for the enzyme to maintain an active conformation [52, 53].

ST6Gal-I is shown to form dimer, by means of disulfide bond formation [54], and the cysteine residues involved are located in the transmembrane domain [55]. Furthermore, each subfamily of sialyltransferase shares a common sequence near the 3'end of sialylmotif S, ST3Gal: (Y/A)GF(K/G)(Y/A) [56]; ST6 (A/V)YG(F/M) [48] and ST8Sia: (I/L)(F/Y)GFWPF [43].

1.4.3 Structural Commonalities with other GT families:

The sialyltransferases have been extensively characterized in terms of substrate specificities toward synthetic acceptors as well as their glycoprotein and glycolipid acceptor preference (reviewed in [42] and [43]). However, at the present time, there is no structural

information for GT 29 sialyltransferases family members and therefore, the detailed mechanism of action for these sialyltransferases remains unclear. The crystal structures of several glycosyltransferases belonging to distinct sequence-based families have been recently solved and among them eight are of mammalian origin. Even though there is a very little or no sequence similarity among the closely grouped GTs, there is a surprisingly high degree of structural similarity in the catalytic domains of most of the GTs. This might not be too surprising, as they recognize similar sugar nucleotides [57]. Moreover, almost all of these proteins have been crystallized in the presence of their donor substrate enzymatic product, nucleotide, or the complete donor sugar [58]. These structures have provided considerable insights into the structural bases for catalysis.

Almost all GT proteins have a globular catalytic domain and share one of the two types of protein folds, named GT-A and GT-B. The so-called GT-A fold (glycosyltransferase A, catalytic domain) is composed of two domains compromised of a α/β single Rossmann domain, a conserved Asp-X-Asp (DXD) motif [58-61] (with exception of α-2-3-sialyltransferase from *Campylobacter jejuni*, CstII [50]) and a conical active site cleft formed by the two closely associated domains. The GT-B fold (glycosyltransferase B, catalytic domain) consists of two Rossmann domains separated by a deep substrate binding cleft [50, 58]. With few exceptions both folds have a conserved three residue peptide motif, Asp-X-Asp (DXD) or Glu-X-Asp (EXD), or its equivalent motifs. It is usually involved in the metal ion binding (usually Mn^{2+}) and catalysis in GT-A enzymes, with exception of CstII that lacks the metal ion and DXD motif. The GT-B glycosyltransferase, including microbial GT-B enzymes (BGT) (again with exception is CstII which is GT-A) do not have DXD motif or its equivalent, even though some BGT require a metal ion for activity [58]. Although the DXD motif contains two Asp residues, the first

(in the first position of the motif) is relatively variable while the second (in the third position of the motif) is quite well conserved [58]. With few exceptions, there is an absolute requirement of metal ion for the enzymes sharing this motif. Metal ion is bound in an octahedral coordination. It interacts with one or both acidic residues of the DXD or EXD motif and with two oxygen atoms from the α-phosphate and β-phosphate of UDP [3]. To satisfy the octahedral geometry, the three remaining metal ion points of coordination are made either to water molecules or two water molecules in combination with other residues of the protein. After sequence analysis and "fold recognition" or "threading", it has been suggested that there could be several GT families with structural similarities to members of these two structured folds [62]. Unfortunately, sialyltransferase families are among those families in which the "threading" technique could not predict with high confidence which fold they could adopt. However, analysis did predict the existence of a Rossmann fold [62].

1.4.4 Catalytic Mechanism of Sialyltransferases

The mechanisms of sialyltransferase enzymes have been investigated through the application of kinetic isotope effect (KIE) studies [63-65] using rat ST6Gal-I. The mechanism is proposed to proceed via an SN_1-like reaction mechanism via a sialyloxocarobenium ion transition state in which the leaving group (CMP) mostly leaves before the incoming nucleophile (sugar acceptor) attacks [65]. The kinetic studies with isotopic substrates also indicated that the mechanism was steady-state random and that the enzyme had a bell-shape pH versus sialic acid transfer rate [65]. Similar results were demonstrated with the sialyltransferase (CstII) from the bacteria *C. jejuni* which also showed bell-shape profile for pH versus sialic acid rate [50]. There are only two bacterial sialyltransferases whose crystal structures have been elucidated to date [50, 51]. It was

suggested that in CstII, a histidine acts as a general base to assist in the nucleophilic attack of a hydroxyl and a tyrosine acts as a general acid to assist the departure of CMP. Similar residues were seen to be participating in catalysis in sialyltransferase from *P. multocida*, but it was not clear which residues were directly involved in catalysis. Moreover, the loss of enzyme activity was reported during mutagenesis study on mammalian sialyltransferases on the tyrosine residue (motif 3) [49]. Therefore, when considered with the crystal structure of bacterial sialyltransferase, it is possible that this tyrosine may be involved in active-site catalysis.

However, the possibility of another amino acid other than tyrosine act as general acid can not be ruled out, as the sequence similarity between the mammalian sialyltransferases and the two bacterial sialyltransferases is quite low.

1.5 Glycosylation of Sialyltransferase

Like most of the glycosyltransferases, sialyltransferases are very often themselves glycosylated. The role of *N*-glycosylation sites and their glycosylation varies from protein to protein. In case of ST6Gal-I, *N*-glycans are involved in stabilization of proper conformation and is required for activity of the full-length enzyme *in vivo* [66]. Similarly, ST3Gal-I was shown to be active without glycosylation, but the non-glycosylated enzyme was not processed and secreted as efficiently as their glycosylated counterpart [49, 67]. In the case of ST8Sia-I (GD3 synthase), deglycosylation resulted in enzymatically unstable protein and the *N*-glycosylations along with proper trimming appeared to be critical for proper trafficking of ST8Sia-I to the Golgi complex [68]. These results give us an insight into the probable role of *N*-glycosylation in the proper folding and trafficking. Therefore it is important to use a eukaryotic system for the recombinant expression of all the mammalian sialyltransferases. However, recently it was

pointed out [69] that the glycosylation sites for sialyltransferases are not conserved in nature and, with the correct mutations of amino acids at and around the N-glycosylation site, it is possible to give rise to a fully functional sialyltransferase enzymes which raises the possibility of using a prokaryotic host for recombinant expression.

1.6 Substrate specificities of Sialyltransferase subfamilies

The ST6Gal family members catalyze the transfer of sialic acid from CMP-Sia to the terminal galactose residues of glycan structures through the formation of an α2,6-linkage. So far, there are two members in the ST6Gal family, ST6Gal-I and -II in both mouse and humans [70, 71]. The ST6 Gal family can use type 2 (Galβ1-4GlcNAc) disaccharides as acceptors. However, they are unable to use type 1 (Galβ1-3GlcNAc) or type 3 (Galβ1-4GalNAc) structures as acceptors [43, 72]. ST6Gal-II exhibits narrow acceptor specificity toward oligosaccharides as compared to ST6Gal-I which exhibits broad acceptor specificity toward glycoproteins, glycolipids, and oligosaccharides. Unlike the ST6Gal-I gene, which is expressed ubiquitously, ST6Gal-II gene expression is tissue specific (adult brain) and stage-specific (embryonic stage). While, biological functions of ST6Gal-I are well established, playing a vital role in B cell mediated immune responses [73], the function of ST6Gal-II is yet to be explored.

The ST6GalNAc family members catalyze the transfer of sialic acid from CMP-Sia to the terminal galactosamine residues of glycan structures through an α2,6-linkage. So far there are six members of the ST6GalNAc family, ST6GalNAc-I-VI, in both the mouse and humans. The ST6GalNAc family is divided into two subfamilies based on amino acid sequence similarities, substrate specificities, and gene structures. One subfamily consists of two members, ST6GalNAc-I and ST6GalNAc-II which utilizes terminal GalNAc residues *O*-glycosidically

linked to Ser/Thr. The other ST6GalNAc subfamily members, ST6GalNAc-III, ST6GalNAc-IV, ST6GalNAc-V and ST6GalNAc-VI, utilize the terminal GalNAc of gangliosides as acceptors. The ST6GalNAc-I and ST6GalNAc-II exhibit similar substrate specificity, they utilize GalNAc-, Galβ1,3GalNAc-, and Neu5Acα2,3Galβ1,3GalNAc- structures on O-glycans of glycoproteins as acceptor substrates [74-76]. The human ST6GalNAc-II gene is expressed at low levels in heart, skeletal muscle, kidney, and liver.

The other ST6GalNAc subfamily consists of four members, ST6GalNAc-III, ST6GalNAc-IV, ST6GalNAc-V, and ST6GalNAc-VI, which commonly utilize the Neu5Acα2,3Galβ1,3GalNAc structure as an acceptor substrate. However, the substrate specificity for each enzyme is different. For example, ST6GalNAc-III efficiently utilize an O-glycoprotein and ganglioside G_{M1b} as acceptor substrate, while ST6GalNAc-IV utilizes Neu5Acα2,3Galβ1,3GalNAc oligosaccharide and O-linked glycans as acceptor substrates [77]. The ganglioside G_{M1b} is the preferred acceptor substrate for ST6GalNAc-V [78]. ST6GalNAc-VI prefers ganglioside G_{M1b} and G_{D1a}, along with sialyllacto-tetraosylceramide (sialyl Lc4, NeuAcα2,3Galβ1,3GlcNAcβ1,3Galβ1,4Glcβ1-Cer) [79, 80], as acceptor substrates. The ST6GalNAc-III, ST6GalNAc-IV and ST6GalNAc-VI genes are expressed in a wide range of tissues such as such as colon, liver, heart, spleen and brain, but expression of the ST6GalNAc-V gene is restricted to the brain.

The ST3Gal family members catalyse the transfer of sialic acid from CMP-Sia to the terminal galactose residues of glycan structures through an α2,3-linkages. So far there are six members of the ST3Gal family, ST3Gal-I-VI in both the mouse and humans. The ST3Gal family is divided into two subfamilies, based on amino acid sequence similarities, substrate specificities, and gene structures. One subfamily consists of two members, ST3Gal-I and ST3Gal-II. The

ST3Gal-I and ST3Gal-II exhibit similar substrate specificities and they utilize Galβ1,3GalNAc (core 1 O-glycan) structures on O-glycans of glycoproteins and glycolipids as acceptor substrates [43, 81, 82]. While the biological functions of ST3Gal-I have been elucidated, playing vital roles in T lymphocytes homeostasis by modulating sialylation of O-glycans on $CD8^+$ [83], the function of ST6Gal-II is yet to be explored.

The other ST3Gal subfamily members consist ST3Gal-III, -IV, -V, and -VI. Even though ST3Gal-III and -IV utilize both Galβ1,3GlcNAc and Galβ1,4GlcNAc structures as acceptors, mainly on glycoproteins as acceptor substrates, ST3Gal-III prefers Galβ1,3GlcNAc structure while ST3Gal-IV prefers Galβ1,4GlcNAc structures [82, 84]. ST3Gal-IV sialylates placental glycans, which are involved in leukocyte trafficking and inflammation [72]. The only known substrate acceptor for ST3Gal-V, is Galβ1,4GlcNAc-Cer, which means it synthesizes ganglioside GM3e [43, 72, 85]. ST3Gal-VI utilizes the Galβ1,4GlcNAc structure on glycoproteins and glycolipids as an acceptor substrate [80]. Recently, the substrate specificities of all of the human and mouse sialyltransferase family members have been compiled [43, 72].

1.7 Biological role of ST6Gal-I

ST6Gal-I is probably the most extensively studied sialyltransferase amongst the eukaryotic sialyltransferase family members. It catalyses transfer of sialic acid on terminal Galβ1-4GlcNAc structures to form the Siaα2-6Galβ1-4GlcNAc (Sia6LacNAc) tri-saccharide. This structure is predominantly found on *N*-linked glycans, and is scarcely found on some *O*-glycans, glycosphingolipids, and sialyloligosaccharides. The formed terminal disaccharide, Siaα2-6Gal, is the ligand for the B cell-specific lectin, CD22/Siglec-2 [86-88], which plays an important role for B cell function [89-91]. ST6Gal-I is believed to be involved in activation of B-

cell immune responses and expression of serum immunoglobulin (Ig). ST6Gal-I gene knock-out mice are viable but they are deficient in Siaα2,6Galβ1,4GlcNAc structures and show all the characteristics of severe immunosuppression [73]. The loss of ST6Gal-I function, in these mice, yields reduced serum IgM levels, decreased Ca^{+2} mobilization, partially impaired B cell proliferation in response to IgM and CD40 cross-linking, and a low antibody titer induction upon immunization [73, 92]. It is well established that ST6Gal-I can modulate CD22 signaling in B cells. However, the mechanism of how CD22 sialylation by ST6Gal-I achieves B-cell signaling responses has not yet been determined.

CD22 is a transmembrane glycoprotein lectin found exclusively on B cells (lymphocytes) and is known to play a role in the immunologic activation of these cells [93, 94]. The extracellular domain of CD22 contains seven Ig domains, of which the outermost N-terminal domain binds to sialic acid-containing glycan ligands [95]. Its cytoplasmic domain has both positive and negative regulatory elements that mediate its activity as a modulator of B cell signaling[96]. The inhibitory elements are cytoplasmic immunoreceptor tyrosine inhibitory motifs (ITIMs) that are tyrosine phosphorylated by the SRC-family kinase, LYN, which upon activation recruits the Shp-1 tyrosine phosphatase to inhibit immune signaling [35, 89, 97, 98] (Figure 6). The positive regulatory element of CD22 for B-cell signaling includes GRB2 (growth factor receptor bound protein 2), SHC (SH2 domain containing transforming protein C), PI3K (phosphoinositide 3-kinase) and PLCγ2 (Phospholipase Cγ2) [92] (Figure 6A). The impact of CD22 on these negative and positive regulators of B-cell signaling depends on the manner of B-cell activation. For example, in the ligation of the BCR with either antigen or IgM-specific antibody, or simultaneous ligation of the BCR and CD40 (with IgM-specific and CD40-specific antibodies) results in the differential phosphorylation of CD22 tyrosine-based motifs both

quantitatively and qualitatively [92]. Also, when CD22 is activated by IgG it does not affect B-cell signaling. With our current understating of B-cell signaling, delineating the molecular mechanism of CD22-mediated signaling mediated ST6Gal-I would just be an over simplification of the event observed.

Unlike ST6Gal-I knockout mice, B-cells from CD22 knockout mice show immune activation responses [35, 93], which can be explained by the loss of negative regulatory components of CD22, such as ITIMs. Moreover, the phenotype of ST6Gal-I and CD22 double knockout mice is similar to that of the CD22 knockout [36, 37, 99]. This clearly suggests that the immune suppression of CD22 clearly depends on the ligand binding of CD22. However, mutations in CD22, aimed at ablation of the sialic acid binding activity leads to a phenotype similar to that expected for ST6Gal-I knockout [100]. Thus, CD22 modulates B cell function *in vivo* through both ligand-dependent and ligand-independent mechanisms.

The ligand-dependent B-cell signaling mechanism has been characterized by assessing the microdomain localization of CD22 at the plasma membrane relative to that of the BCR. It has been established that CD22 is bound to ligands *in cis* (i.e. to ligands expressed on the same cellular surface) on the majority of B-cells [38, 101, 102]. In resting B-cells most of the CD22 (~80%) is co-localized with clathrin domains, while BCR shows minimal co-localization with either CD22 or clathrin domains. Upon B-cell activation, BCRs are endocytosed via activated lipid rafts and clathrin pathways [37, 103, 104]. Also, since CD22 is excluded from activation rafts, it is probable that CD22 exerts its regulation of BCR signalling in the fused raft/clathrin domains [37, 92, 96]. It is possible that ST6Gal-I produces ligands that restrict the binding of CD22 to other BCR components such as IgM. Therefore, one would expect that in the absence of sialylated ligands the binding and association of CD22 with IgM would increase. However, there

was no change in CD22 levels in ligand deficient (CD22 mutant to ablate binding) mice and cell surface IgM colocalization with CD22 increased two fold (from 20–25% to 40–50%). This shift in BCR localization results in increased endocytosis and reduced half-life of IgM [36, 92]. However, in mice lacking ST6Gal-I activity, increased co-localization of cell surface IgM with CD22 and clathrin domains was restored to normal levels of IgM co-localization with clathrin when both CD22 and ST6Gal-I were absent [36]. Therefore, redistribution and reduced half-life of IgM mediated by CD22 are not due to another effect of the ST6Gal-I deficiency. These results also indicate that *cis* ligands of CD22 reduce BCR localization in raft/clathrin domains of resting B-cells in wild-type mice by an as-yet-undefined mechanism. Apart from BCR modulation, loss of ST6Gal-I function in B- cells results in a reduction in cell surface CD22 and IgM, coupled with increased endocytosis and B-cell turnover (Figure 6B). Also, there is a reduction in the number of marginal zone B-cells [100].

More recently the possibile of existence of an alternative enzyme, other than ST6Gal-I, which could produce low levels of Siaα2,6Galβ1,4GlcNAc structures or have a different substrate specificity than of ST6Gal-I, have been implicated in B cell signaling [71]. Also, there is possibility that sulfation of GlcNAc in the Siaα2,6Galβ1,4GlcNAc structure may be involved in regulation of CD22/siglec-2 functions in humans [105]. Therefore, it is possible that there may be other important roles for ST6Gal-I-created ligands in CD22 function and at the same time there may be other glycosylating enzymes which may play crucial role in B cell signaling.

1.8 Expression of a mammalian ST6Gal-I in *Pichia pastoris*

Heterologous expression of rat ST6Gal-I in *Pichia pastoris* has been carried out successfully [106]. Even though ST6Gal-I was expressed both as intracellular and secreted proteins, ST6Gal-I activity was much higher in case of the secreted form. Moreover, when both types of recombinant forms of the enzymes (internal or secreted ST6Gal-I), were expressed in yeast cells and cell-surface sialic acids were detected using sialic acid specific lectin, SNA. Higher incorporation was found with yeast expressing intracellular ST6Gal-I than the cells which were secreting ST6Gal-I. However, since *P. pastoris* normally produces only high mannose glycan structure [107], which should not act as acceptors for sialylation, the latter data can not be readily explained.

1.9 *Pichia pastoris* as a host for recombinant expression of glycosyltransferases

The popularity of the methylotrophic yeast, *Pichia pastoris*, as a recombinant host for eukaryotic protein expression has increased recently. There is a growing literature indicating that *Pichia* can be an effective expression platform for recombinant expression of biologically important proteins. The main features of this recombinant host includes the ease of recombinant protein induction, high level protein expression, eukaryotic post translational modifications, efficient secretion into the culture media, minimal media requirements, and scalable fermentation capabilities [108].

Typically, there are two different ways to express recombinant protein in *Pichia pastoris*; the inducible expression under the alcohol oxidase (AOX) promoter using methanol as inducer and the constitutive expression under the glyceraldehyde-3-phosphate dehydrogenase (GAP) promoter (Fig) [109]. There are other promoters (e.g. FLD1, PEX8, YPT1, etc.) for constitutive

expression of recombinant proteins [110]. However, there are very few examples of the use of these latter promoters in the literature and they are not available from a commercial source (e.g. Invitrogen Inc.).

In the *Pichia* genome there are two AOX gene products, *AOX1* and *AOX2,* that cells can use to grow on methanol as a carbon source. Typically, *AOX1* is responsible for 85% of alcohol oxidase synthesis[111]. Therefore, there are three different possible phenotypes with respect to the expression of the two AOX gene: 1) the methanol utilizing plus (Mut^+) *P. pastoris* strain expresses both *AOX1* and *AOX2* genes, which utilizes methanol much more efficiently; 2) the methanol utilizing slow (Mut^S) *P. pastoris* strain with the *AOX1* gene deleted, and expressing only the *AOX2* gene; and 3) the methanol utilizing minus (Mut^-) *P. pastoris* in which both the *AOX1* and *AOX2* genes are deleted. However, all three strains grow with same efficiency on glycerol or glucose as a carbon source.

Although there are numerous examples where *P. pastoris* shake flask cultures have been used for efficient expression of the recombinant proteins, typically shake-flasks produce relatively low recombinant protein yields compared to the yields with controlled fermentation. This is because in controlled fermentation it is possible to grow *Pichia* to high cell densities (>100 g/L dry cell weight or 500 OD_{600} U/ml).

Typically fermentation expression is carried out in three phases: 1) batch growth phase where glycerol is used as a carbon source to achieve high cell density before induction 2) fed-growth phase where a bolus of glycerol is fed for another batch growth or it is fed at a constant or increasing flow rate to maximize the cell mass further before induction 3) induction phase where glycerol is replaced as carbon source with methanol, which initiates the expression of the recombinant protein. There are some alternatives available to this typical fermentation strategy

for *P. pastoris*. In some instances, in the fed-batch phase a mix-feed strategy is employed where the glycerol carbon source is gradually replaced with methanol.

During the induction phase it is very critical to monitor the methanol concentration in the media because elevated levels are toxic to cells and can lead to culture death. Therefore, the methanol feeding strategy is very important for increasing the expression yields of recombinant proteins. There are several strategies in the literature for the expression of the recombinant proteins in *P. pastoris,* however they all generally use repeated fed-batch fermentation. The first method employs a dissolved oxygen spike approach since there is a sharp increase in the dissolved oxygen concentration when methanol is consumed [112]. Thus, methanol is fed based on the level of DO in media. In the second method, the methanol is fed to the fermentor based on the estimated methanol consumption rate. The methanol consumption rate is calculated from the specific growth rate on methanol and cell mass at any given time [112-115]. Even though these two methods provide a good estimate of the methanol concentration, there is always a chance that the methanol concentration could reach a potentially toxic level [116] or temporarily expose the toxic levels of the methanol [117]. In another strategy, which utilizes a methanol sensor and a feedback controller, methanol can be fed to the growing culture irrespective of the culture conditions [117]. The advantage of this strategy is that methanol levels can be monitored on-line and there would be no fluctuation in the methanol concentration over the entire duration of the induction phase.

1.10 Optimization of fermentation conditions for expression of recombinant proteins in *Pichia pastoris*

Expression of recombinant proteins in *Pichia pastoris* may be common these days, but few studies describe the systematic optimization of fermentation conditions to enhance the expression of the respective recombinant protein. So far optimization of fermentation conditions in *Pichia pastoris* has been rather arbitrary and usually applicable only to the protein under study. Typically the variable conditions are pH, temperature, and medium components.

1.10.1. pH

Pichia pastoris grows over a wide pH range, from 3 to 7 [118]. There are minimal or no effects of pH on the growth rate in this range. However, in certain instances, the pH of the culture can effect the recovery of secreted recombinant proteins due to protease activity in the fermentation broth [118]. Therefore, optimization of pH commonly means establishing pH conditions which support optimal cell growth; maximize secreted protein stability, and reducing the degradation of the secreted protein. Even though *Pichia* can grow over a wide pH range, inducing the culture near neutral pH often gives better yields for the secreted proteins, which is assumed to be due to the reduced activity of acidic proteases produced during cell growth. For example, the expression yield of human serum albumin was greatly enhanced when the pH of medium was increased from 5.2 to 6.0 [119]. Similarly, the expression yields were highest at pH 6.0 for the expression of bovine β1,4-galactosyltransferase [120] and human cystatin C [121] in *P. pastoris*.

In several instances reports describe the use of pH conditions which are not optimal for growth of the *P. pastoris*, yet post production resulted in the higher expression of the secreted

products compared to the pH conditions for optimal growth. For example, the expression of the recombinant hookworm anticoagulant peptide (AcAP-5) was several fold higher at pH 7.0 when compared to optimal growth pH at 3.5 and 4.5 [122]. Similarly, expression of full length (non-cleaved) haemagglutinin was possible at higher rather than optimal pH. Full length haemagglutinin was seen only at pH 8.0, while at lower pHs of 5.8 and 7.2 only resulted in expression of fragmented haemagglutinin [123]. Also, the full length expression of chimeric protein prourokinase-annexin was possible only at pH above pH 7.7.

Even though it might seem that expression of recombinant proteins in *P. pastoris* would be more effective at higher at near neutral or alkaline pHs it is completely dependant on the individual proteins that are being expressed. In some cases, even extremely acidic pHs have proven to be more beneficial. For example expression of HIV-1ENV (HIV-1 envelope protein) was optimal under acidic conditions. Proteolytic degradation of expressed HIV-1ENV was greatly reduced when the pH was reduced to 3.0 and 5.0 in shake flask and fermentor cultures, respectively [124]. Similarly, optimal expression of the fusion protein CBM-CALB was achieved at pH 4.0 [125].

In light of all the above mentioned examples, it is not possible to establish a universal pH condition for the expression of recombinant proteins using the methylotropic yeast *P. pastoris*. In fact, each protein has its own optimal pH condition for effective expression. There are numerous examples of varied conditions for expressions that have been effectively summarized [126].

1.10.2. Temperature

The optimal temperature for the growth of *P. pastoris* cells is 30°C, which is obviously the first choice for the expression of the recombinant proteins. However, the expression levels of

some recombinant proteins at this temperature are hampered by proteolytic degradation. In fact, in some cases, proteolysis at 30°C leads to minimal protein recovery. For example, expression of laccase enzymes encoded by *Trametes versicolor lcc1*and *lcc2* genes in *S. cerevisiae* at 30°C exhibited a little or no enzyme activity in the medium [127], whereas reduction of temperature to 19°C, while keeping the other conditions same, led to a 16-fold increase in the enzyme activity. Similarly, a four-fold increase in enzyme activity was observed for the expression of galactose oxidase in *P. pastoris* when the temperature was lowered to 25°C from 30°C [128]. There was not only a 10-fold increase in expression of the herring antifreeze protein (hAFP) from *P. pastoris* when temperature was lowered to 23°C, but also the degradation of hAFP was eleminated [129].

There are numerous examples in which lowering the temperature of the expression culture led to a several-fold increase in protein expression and recovery. In addition to reduced proteolysis, increased protein production has been attributed to reduced thermodynamics during protein folding and prolonged time to complete the protein folding process [129].

The other probable mechanism for the higher expression yields of the secreted proteins is that some proteins are less stable at higher temperature [120]. In another example, it was demonstrated that lowering temperature does not reduce the rate of the biosynthesis of recombinant proteins [125]. However, those experiments were carried out using an exponential methanol feed in the induction phase which was methanol-limited.

In the light of the above examples it can be stated that lowering the fermentation temperature over the entire duration of fermentation or only in induction phase can lead to increases in recombinant protein expression, which is accredited to decrease in the proteolytic

degradation. Thus, induction at lower temperature proved to be an effective method to increase protein production in *P. pastoris*.

1.10.3. Medium components

The other important condition used to increase the expression of recombinant proteins in *Pichia* is through addition of components into the culture media. The standard BSM (Basal Salts Medium) recipe recommended in the *Pichia* Expression Kit [109] is widely used for expression of recombinant proteins in *P. pastoris* using fermentation. This media often works effectively for expression of recombinant proteins, but there are a large number of examples in which BSM medium poses severe problems. For instance, in certain cases salt precipitation is observed in the medium, especially when the pH is adjusted to above 5.0, a common practice currently used for the expression of recombinant protein in *Pichia* fermentation. For example during the expression of a potential vaccine for malaria, P30P2MSP1$_{19}$, a lipid-like substance was formed, which interfered with the processing of the media for purification and caused the aggregation of P30P2MSP1$_{19}$ [130]. It was proposed that a lipid-like substance was released as a consequence of the cell lysis which was induced by the high osmotic pressure in the BSM medium.

A modified BSM (1/4 the regular concentration of media components), was introduced, which eliminated the cell lysis and consequently the protein aggregation was significantly decreased. Similar results were seen when modified BSM was used for the expression of a single-chain Fv antibody (A33scFv) through the fermentation of *P. pastoris* [131].

1.11 Proteases in *P. pastoris* Fermentations

As mentioned in earlier sections, secreted proteases can be a serious problem in the expression of recombinant proteins in *P. pastoris*. The yeast produces these enzymes to degrade proteins, including other enzymes, into amino acids and peptides in an effort to use these macromolecules as a carbon source. Unlike *S. cerevisiae* or other fungal systems, there is very limited information available regarding the characteristics of the proteases produced by *P. pastoris*. However, proteases commonly fall into four major families: serine proteases, aspartic proteases, cysteine proteases and metalloproteases.

Even though the identity of the proteases in *Pichia* cultures has been elusive, the existence of proteases in high cell density fermentation has been widely reported [118, 119, 125, 132-134]. Usually cell lysis followed by death during high cell density fermentation results in the release of intracellular proteases into the medium. Therefore, these proteases can pose challenges in an effort to maximize expression of recombinant proteins in *P. pastoris* fermentations.

There are several strategies that have been successfully used for minimizing the proteolytic instability of recombinant proteins secreted into the *P. pastoris* culture medium: 1) addition of amino acid rich supplements such as peptone or Casamino acids into the culture medium, 2) change of pH range from 3.0 to 7.0, 3) lowering culture media temperature, 4) use of protease deficient *P. pastoris* host strains such as SMD1168 (*his4 PEP4*) (PEP4 gene encodes proteinase A, vacuolar aspartyl protease required for the activation of other vacuolar proteases, such as carboxypeptidase Y, and proteinase B), 5) use of protease inhibitors during purification and in the fermentation.

There are several examples where the use of protease inhibitors has been proven to be effective in reducing the proteolysis of recombinant proteins [119, 125, 133]. The proteases

which cause the degradation of recombinant proteins include acidic proteases, serine proteases [125, 133], or non- specific intracellular proteases released by cell lysis [134].

There are various ways to handle protease problems, some of which were mentioned in earlier sections, such as reduction in cultivation temperature or adjusting the fermentation pH. In some cases protease inhibitors, such as PMSF [125], Pefabloc and E-64 [133], EDTA and pepstain [135], have been proved to reduce the proteolysis of recombinant proteins. Protease inhibitors such as PMSF and Pefabloc block action of serine protease, EDTA targets metalloproteases, and Pepstain works against aspartic proteases and E-64 is effective against cysteine protease.

1.12 Glycoengineered *Pichia pastoris*: roles beyond conventional protein production

Pichia pastoris, which provides high protein yields (sometimes >1g/l) compared to their mammalian expression systems is seen as a potential target for the expression of therapeutic proteins. Moreover, there is a wide range of *Pichia* strains available to the scientific community (reviewed in), with each one of them having unique features. However, all of the *Pichia* strains typically yield yeast-like glycans (extended high mannose termed mannans) on the recombinant glycoproteins. Recently, it has been established that the therapeutic efficacies of many glycoproteins are dependent not only on the protein's structure but also by the glycan structures associated with the glycoproteins [136]. For example, human erythropoietin (EPO), a glycoprotein used in the treatment of anemia, contains three N-glycans, which are critical to its therapeutic activity. Removal of the N-glycans from EPO results in a protein with a very short half-life and virtually no erythropoietin function *in vivo* [137]. More recently, it was shown that

the therapeutic efficacy of erythropoietin can be altered not only by deglycosylation but also by differential glycosylation as a consequence of the recombinant production method [138].

Thus expression of such therapeutic glycoproteins in *Pichia pastoris* would likely lead to rapid clearance and minimal therapeutic efficiency. Fortunately, genetic manipulation of *P. pastoris* (yeast in general) has made it possible to engineer the glycan biosynthetic machinery of *Pichia* to produce glycan structures that resemble human glycan structures [139, 140] (recently reviewed in [136]). The advantage of glycoengineered *P. pastoris* is that unlike mammalian cell cultures, expressed glycoproteins exhibit exceptional uniformity in glycan structure [141-143]. However, a large scale commercial production of recombinant therapeutics of these *Pichia* strains has yet to be reported.

1.13 Isotope labeling in *Pichia pastoris* for structural studies

For obvious reasons, such as growth on relatively inexpensive defined media, bacterial expression is the method of choice for generating isotopically labeled materials for structural studies using Nuclear Magnetic Resonance (NMR). However, only in a few instances have glycosylated proteins been successfully expressed in prokaryotic systems, presumably because of the general necessity for posttranslational modifications, disulfide bond formation, or specific chaperone systems that are unique to eukaryotic organisms. Similarly, expression of proteins in *P. pastoris* allows high protein yields in relatively simple and defined cell culture media. The use of minimal media composition coupled with eukaryotic posttranslational modifications makes *P. pastoris* an ideal expression system for isotope labeling of recombinant glycoproteins for NMR-based studies. However, isotope enriched expression of glycoproteins will never be as cost effective as with *E. coli* because of the high cell density required for the yeast expression system

leads to incorporation of NMR-active isotopes predominantly into the cell mass rather than the recombinant product [144]. Nevertheless, *P. pastoris* can be advantageous, since proteins can be efficiently secreted directly into the culture medium, thus simplifying considerably the subsequent protein purification.

Initial studies on isotope labeling of recombinant proteins using *P. pastoris* suggested that it was necessary to incorporate stable isotope-labeled carbon (^{13}C) and nitrogen (^{15}N) sources in both the growth phase and induction phase [145, 146]. Thus, most of the isotope incorporated into the cell mass and the yield of the desired protein per gram of labeling reagent was quiet low. Typically, isotope labeling was carried out in low density shake flask cultures because fermentation using $^{15}NH_4OH$ as a sole N source was exorbitantly expensive. Use of $(^{15}NH_4)_2SO_4$ or $^{15}NH_4Cl$ as a relatively inexpensive alternative resulted in high ionic strength from K_2SO_4 or KCl which inhibited cell growth. However, this problem was eliminated by a stepwise addition of $(^{15}NH_4)_2SO_4$ followed by media exchange before a very high ionic strength is reached [146].

Later, it was shown that that it was not necessary to incorporate isotope labeled carbon (^{13}C) and nitrogen (^{15}N) sources in the entire growth phase, but they could be added at low concentrations a few hours prior to the protein expression period [144]. The method for isotope-labeling of recombinant proteins using *P. pastoris* has developed significantly in the last few years. Today there are new and better methods available for isotope labeling of proteins in *P. pastoris* (for example see [147]). However, such labeling is unselective, and specific amino acids can still not be labeled efficiently with the wild type strain as it is diluted into the pool of endogenously synthesized amino acids.

Even with triple-labeled proteins, structure determination by NMR for proteins larger than 40 kDa is quiet difficult using current methods. However, selective labeling of specific amino acids in recombinant proteins makes it possible to obtain structural information about the particular amino acids in proteins larger than 40 kDa. Recently, specific amino acid labeling as high as 50-80% incorporation was reported [148].

An alternative way to label specific amino acids in the recombinant proteins using *P. pastoris* as expression system is the use of specific amino acid auxotrophic mutants. Currently, there are only a limited number of auxotrophic strains of *P. pastoris* available, including *HIS4*, *ARG4*, *CYS4* . However, none of these auxotrophic strains have been used to specifically label amino acids for NMR-based studies. Recently, engineered *P. pastoris* strains that are auxotrophic for aromatic amino acids such as tyrosine, phenylalanine, and tryptophan were generated using gene disruption techniques and successfully used for specific amino acid labeling of galactose oxidase [149].

Unfortunately, most methods which describe the efficient metabolic labeling of proteins for NMR based studies using *P. pastoris* have generally studied low molecular weight, stable proteins, usually in poorly controlled fermentations or shake-flask cultures. However, in reality most recombinant glycoproteins are expressed in very small quantities and are unstable in the conditions previously described.

Aims and Objective

Given the importance of ST6Gal-I and its role in immune regulation, the overall aim of this dissertation was to examine its structure, function, and mechanism of action. In particular the goals of this dissertation were to examine the use of *P. pastoris* as a host for the expression of ST6Gal-I that is amenable to crystallization for X-ray diffraction studies and isotope labeling for structural characterization using NMR. The specific objectives of this dissertation were.

1. To develop an expression and purification method to generate a highly enriched preparation of the rat ST6Gal-I using *Pichia pasotoris* as recombinant expression host. Subsequently, process the enzyme so that it is amenable for studies using crystallization and X-ray, NMR, and biophysical and enzymatic characterization.
2. To develop a method for generation of isotope-labeled rat ST6Gal-I for structural and functional studies using NMR.
3. To develop a platform for studying kinetics and bind studies between the substrate or substrate analog of ST6Gal-I and the enzyme.

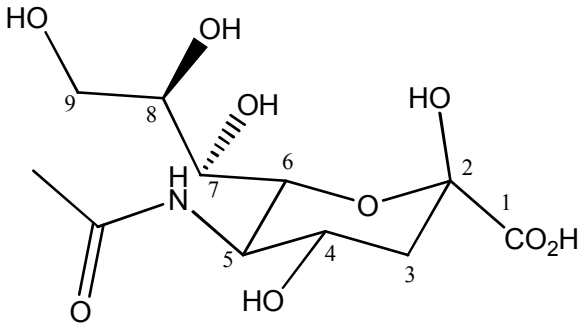


Figure 1: Sialic acid or N-Acetylneuraminic acid (NANA or Neu5Ac)

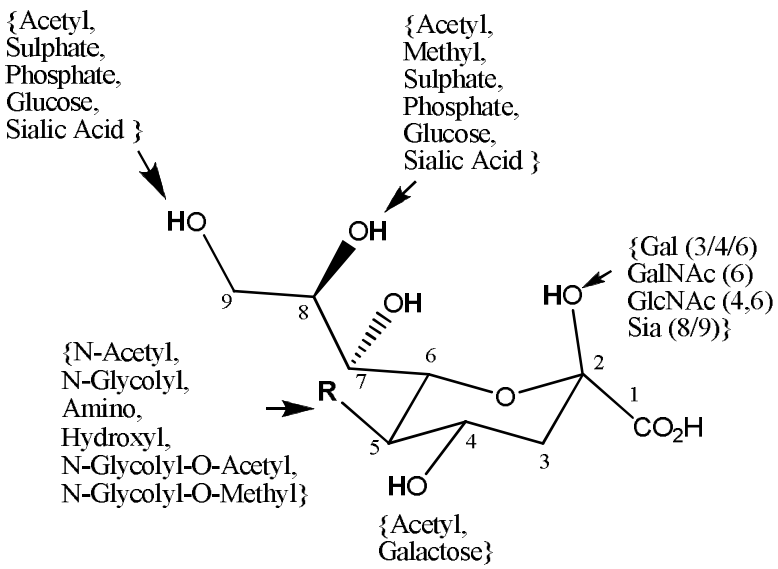


Figure 2: **Diversity of Sialic acids**. Various substitutions, which have been observed in nature, are indicated at various positions. A substitution at position 2 forms a glycosidic linkage with sialic acid and (numbers in brackets) indicates the position on the substitution sugar.

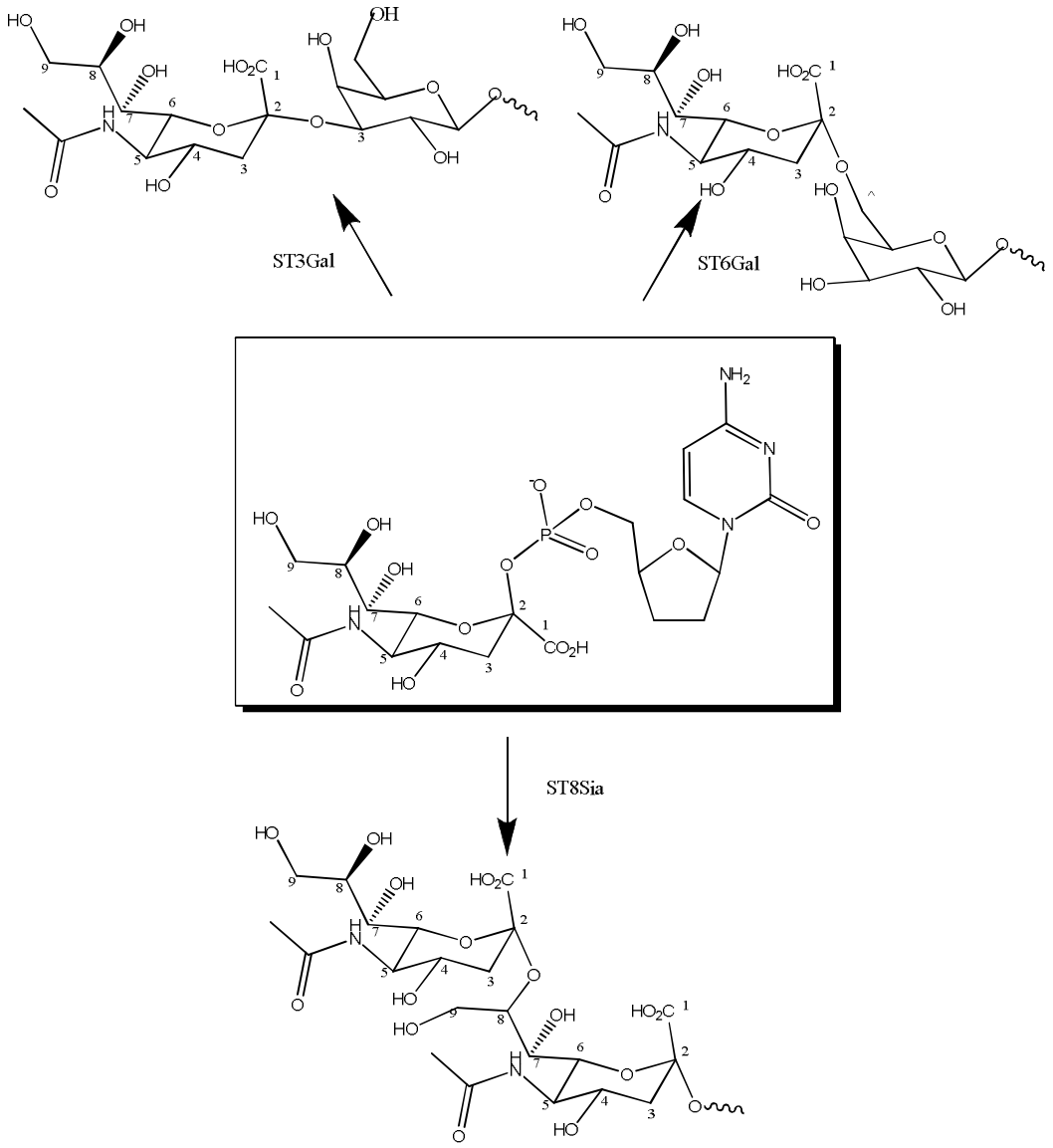


Figure 3: **Synthesis of Sialyl-saccharide by Sialyltransferase family**. Sialyltransferases are classified based upon the type of glycosidic linkage formed

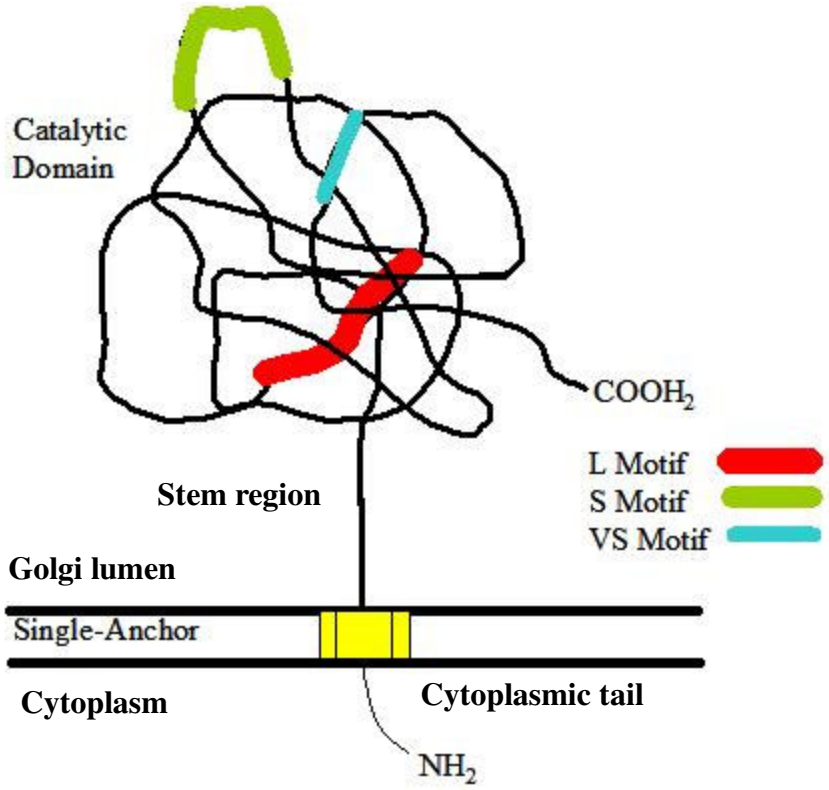


Figure 4: **Proposed topology of Sialyltransferase** [150]. Sialyltransferases are the type II membrane proteins, with a short amino terminus cytoplasmic tail, a single pass transmembrane domain, a stem region in lumen of golgi and a large carboxy terminus catalytic domain.

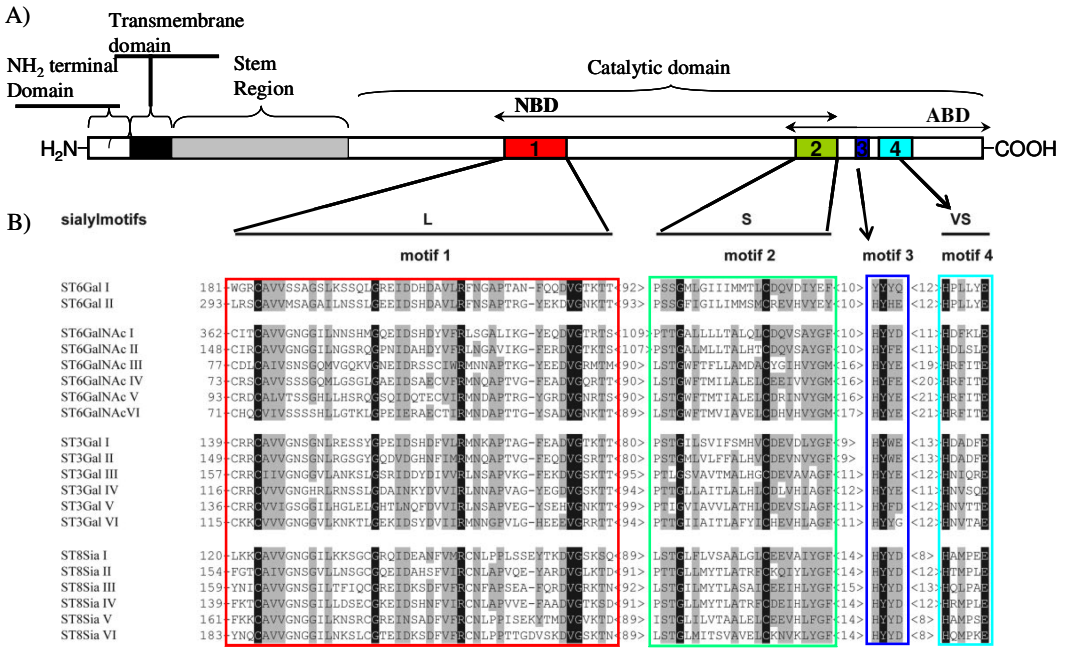


Figure 5: Conserved motifs of sialyltransferases. A) Cartoon diagram of Sialyltransferase depicting the position of individual regions. B) Sequence alignment of the most conserved regions present in human sialyltransferases. Absolutely conserved residues amongst all sialyltransferase sequences are indicated in *white* on a *black background*. The other most conserved amino acid positions are shaded in *gray*, sequence aligment from Jeanneau *et al.*, [49] used with permission of the publisher, Journal of Biological Chemistry.

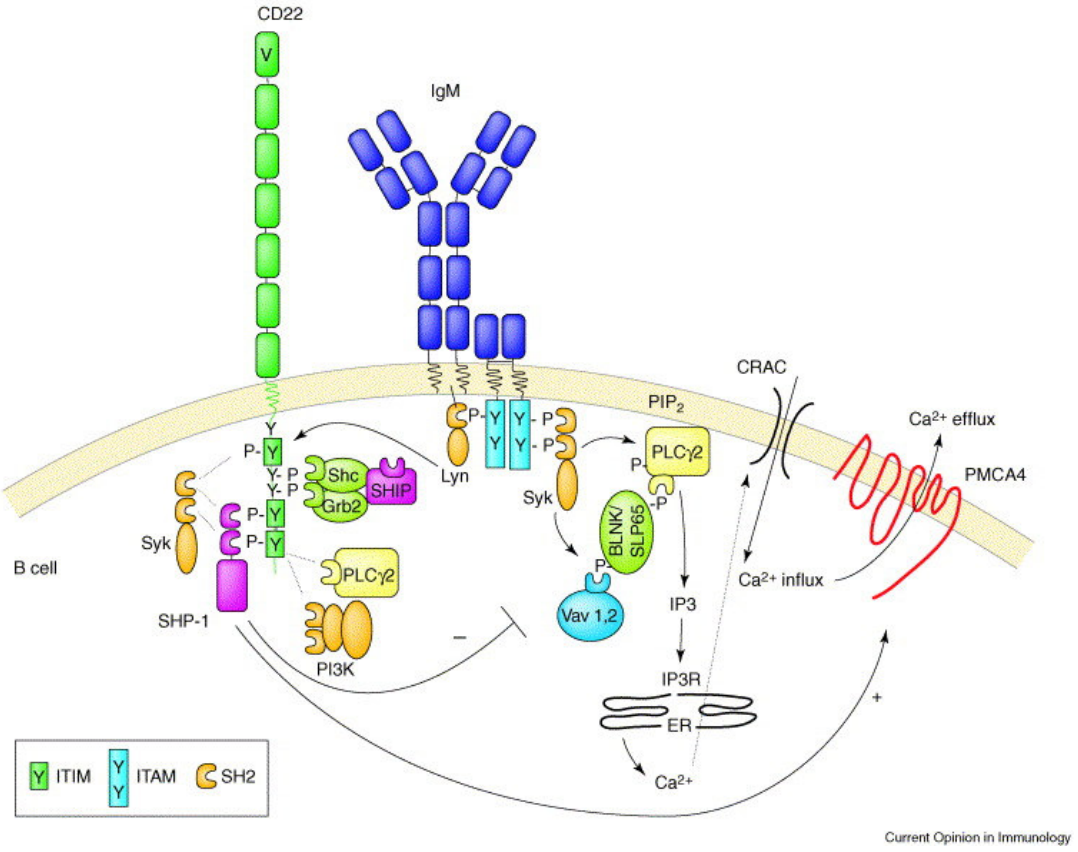
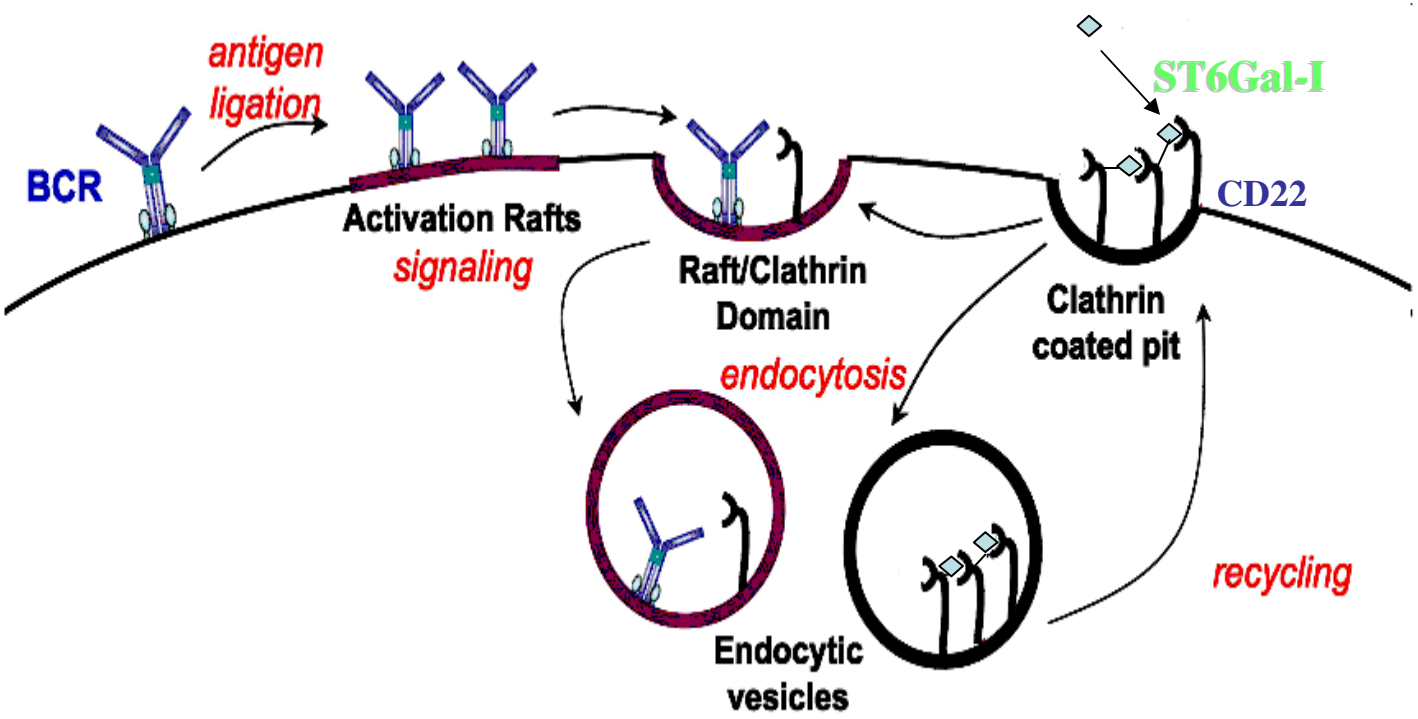


Figure 6A: Regulation of B-cell receptor signalling by CD22. Activation of BCR (ligation of IgM) causes phosphorylation of the CD22 cytoplasmic tail by the BCR-associated kinase Lyn. A clear function has only been shown for binding of tyrosine phosphatase SHP-1. SHP-1 dephosphorylates intracellular signalling molecules (indicated by –). BCR signalling triggers depletion of intracellular Ca^{2+} stores of the endoplasmic reticulum (ER). This activates opening of Ca^{2+}-release-activated channels (CRACs) and triggers Ca^{2+} influx. CD22 (via SHP-1) activates (indicated by +) the Ca^{2+} pump PMCA4 and thereby controls Ca^{2+} efflux. Abbreviations: IP3, inositol-1,4,5-trisphosphate; IP3R, IP3 receptor. (Reprinted from [35] with permission from publisher, Current Opinion in Immunology)



Adopted from http://www.scripps.edu/chemphys/paulson/research.html

Figure 6B: A proposed model for function mechanism of ST6Gal-I in CD22 mediated immune response. It has been established that ST6Gal-I restrains CD22 dependant antigen receptor endocytosis and B-cell turnover. CD22 is constitutively endocytosed and recycled back to B-cells surface in the presence of sialic acid lingands via Lipid Rafts/Clathrin pathway. In absensce of sialic acid ligands it bind to BCR and constriubutes to their internalization via unknown michamis.

CHAPTER 2

MATERIALS AND METHODS

2.1 Cloning of rat α2,6 Sialyltransferase catalytic domain

The wild type cDNA encoding rat α2,6 Sialyltransferase [151] was received from Dr. Karen Colley's Lab. Primers were designed to amplify the catalytic domain within open reading frame 921bp. The 5' primer and 3' primer respectively contain a BamHI site at the 5' end of the catalytic domain and a *NotI* site following the termination codon. The PCR product containing the catalytic domain coding region was resolved on a 1% agarose gel containing 0.5 μg/ml ethidium bromide, purified using the Gel Extraction Kit (Qiagen), and subcloned the bacterial vector pBSSK containing the TCM-His (*T. cruzi* mannosidase secretary sequence) tag domain. The resultant vector was amplified using Top10F' bacterial cells and digested with restriction enzymes *ClaI* and *NotI*. The *ClaI*/*NotI* digested fragment contained the 8xHis tag at NH_2 terminus of the coding region. This fragment was then subcloned into *Pichia pastoris* expression vectors, pPICZαC and pGAPαC (Invitrogen CA).

2.2 Other *Pichia* expression constructs for rat ST6Gal-I

2.2.1 HA-ST6Gal-I_S/pPICZαC construct

The residues which were being clipped from NH_2 terminus were identified by amino acid sequencing. A construct was made using chemical synthesis (Geneart GmbH, Germany) and named HA-ST6Gal-I_S which was 14 amino acid short at the NH_2 terminus compared to ST6Gal-I catalytic domain described above and in addition a hemeagglutinin (HA) tag

followed by TEV protease cleavage site sequence between putative catalytic domain and 8xHis tag. The construct was then digested with restriction enzymes *ClaI* and *NotI*. The *ClaI*/*NotI* digested fragment was then subcloned into *Pichia pastoris* expression vectors, pPICZαC (Appendix 1C).

2.2.2 Strp-ST6Gal-I_S/pPICZαC construct

In Strp-ST6Gal-I_S/pPICZαC, a putative monomeric streptavidin gene [152, 153] was made by chemical synthesis with *EcoRI* restriction site at both 5' and 3' ends. It was digested with *EcoRI* restriction enzyme and then subcloned into ST6Gal-I_S/pPICZαC (Appendix 1D).

2.2.3 MBP-ST6Gal-I_S/pPICZαC construct

For construction of the MBP-ST6Gal-I_S/pPICZαC construct (Appendix 1E), a *Pichia* vector containing maltose binding protein, pMAL/pHILS1 [154], was digested with *SacI* and *NotI* (which included portion of 5' AOX1 promoter, start codon and α-factor secretory sequence followed by MBP sequence). The *SacI*/*NotI* digested fragment was then subcloned into pPICZαC, resultant vector construct was named MBP/pPICZαC. The ST6Gal-I_S/pPICZαC construct was digested with *EcoRI* and *NotI*. The *EcoRI*/*NotI* digest was then subcloned into newly formed MBP/pPICZα (Appendix 1E).

2.2.4 HA-ST6Gal-I_L/pPICZαC construct

The *ApaI* restriction site in the catalytic domain of ST6Gal-I_S/pPICZαC construct was deleted using site directed mutagenesis as described below. Similarly, an *ApaI* site was created towards the 5' end of the coding region close to *EcoRI* site, both times keeping in

mind that the corresponding amino acid is not changed as a consequence of mutagenesis. A small DNA fragment corresponding to the missing 14 amino acids and newly formed *ApaI* site in ST6Gal-I_S/pPICZαC was prepared by chemical synthesis. This fragment was then digested with *EcoRI* and *NotI*. The *EcoRI*/*NotI* digest was then subcloned into newly mutagenized ST6Gal-I_S/pPICZαC (Appendix 1F).

2.2.5 Strp-ST6Gal-I_L/pPICZαC construct

The 5' EcoRI restriction site in Strp-ST6Gal-I_S/pPICZαC, was deleted by site directed mutagenesis as described below. The HA-ST6Gal-I_L/pPICZαC, was digested with *EcoRI* and *NotI*. The *EcoRI*/*NotI* digest was subcloned into newly mutagenized Strp-ST6Gal-I_S/pPICZαC construct (Appendix 1G).

2.2.6 HA-ST6Gal-I_L/pPICZαC construct

The HA-ST6Gal-I_L/pPICZαC, was digested with *EcoRI* and *NotI*. The *EcoRI*/*NotI* digest was subcloned into HA-ST6Gal-I_S/pPICZαC construct construct (Appendix 1G)

2.2.7 Mutagenesis

Site-directed mutagenesis to alter the restrictions sites in various construct described above was performed using QuickChange™ mutagenesis kit (Stratagene Inc., CA). The sense and antisense primers for each mutant were designed based on the sequences for each construct to be made given in Appendix 1 and were synthesized by MWG (MWG Inc.). The desired plasmid (20 ng in 1 µl) and 10 pmole of sense and antisense primers (in 2 µl) were added into into 22 µl of PCR mixture supplied by manufacturer (1 µl of 100 U *Pfu*

polymerase, 2.5 µl 10× buffer, 5 µl of 100 µM dNTP, 13.3 µl H_2O;). The PCR was carried using a thermocycle control unit (MJ Research, Cambridge, MA). The PCR was programmed as follows: 1 cycle of initial denaturation for 1 min at 95 °C, followed by 24 cycles of annealing for 1 min at 58°C followed by elongation from 5-10 min at 68°C (based on the length of plasmid) and denaturation for 45 sec 95°C. The DNA template was digested by the addition of 1 µl *Dpn*I (100 U/ml; Stratagene Inc.) and incubated for 2 h at 37°C. An aliquot of 1-2 µl of *Dpn*I digested mixture was transformed into *E. coli* TOP10F′.

2.3 Transformation of *Pichia pastoris* with pPICZαC-ST6Gal-I

Ten micrograms of the expression vector containing the sense-orientation constructs (e.g. pPICZαC-ST6Gal-I) was linearized by digestion with *SacI* and transformed into *Pichia pastoris* strains KM71H, X-33, SMD1168H (all from Invitrogen Inc) YGly30, RPD361, RPD355 (all from Glycofi Inc.), ARO1, ARO7 and TYR1 (Aromatic auxotophic strains from Dr. James Whittaker [149]) using the EasyComp transformation kit or lithium chloride transformation method as described in the *Pichia* expression manual (Invitrogen). The transformants were selected on YPD agar plates (10 g/l yeast extract, 20 g/l peptone, 2% w/v dextrose and 2% w/v agar) containing 100 µg/ml of Zeocin (Invitrogen). The colonies were isolated following 3 days of growth at 30°C. pGAPαC-ST6Gal-I was expressed only in KM71H and X-33.

2.4 Screening of pPICZαC-ST6Gal-I transformants

The zeocin-resistant colonies were then screened for the expression of the recombinant ST6Gal-I enzyme activity. In each yeast strain, 15-20 colonies were analyzed

for induced enzyme activity in order to screen for the highest expressor. Each of the zeocin resistant colonies was inoculated into 5 ml BMGY (10 g/l yeast extract, 20 g/l peptone, 3.4 g/l YNB (yeast nitrogen base) (DIFCO), 10 g/l ammonium sulfate, 0.1 M phosphate buffer pH 6.0 and 4 µg/l D-biotin) containing 1% v/v glycerol in 50 ml conical tubes.

Overnight cultures were grown in a horizontal shaker at 225 rpm incubated at 30 °C. The cells were collected by centrifugation at 3000×g at 10 °C for 10 min and resuspended in 5 ml of BMMY containing 0.5% v/v methanol. The cultures were continuously shaken at 225-250 rpm for 2-3 days with daily addition of 100 µl of 50% v/v of methanol every 12hr. The expression of wild type recombinant enzyme was monitored by Sialyltransferase activity.

2.5 Shaker flask (1-liter Scale) expression and purification

The *Pichia* transformant expressing highest level of secreted enzyme activity was inoculated from a single colony or frozen glycerol stock into 50 ml BMGY. The culture was grown overnight at 30 ºC prior to transfer into 1L BMGY. After 2 days of incubation at 30 ºC, the culture was then harvested by centrifugation (15000g). The supernatant was discarded and the pellet was then suspended into 1L BMMY containing 0.5% v/v methanol. Additional methanol (10 ml of 50% v/v methanol) was added to the culture every 12 hr for 2-3 days. A culture media was clarified by centrifugation at 6000×g for 15 min followed by filtration through 0.45µm membrane prior to purification step.

2.6 Complex media fermentor expression (1L scale)

Preparation of complex media

Media containing 10 g yeast extract, 20 g peptone, 27.5 g NaH_2PO_4, and 15 ml glycerol were dissolved in water for a total volume of 800 ml. Antifoam 104 (0.05 ml) (Sigma) was added and pH was adjusted to between 5.5-5.7. The media was then autoclaved in a 1.6 L capacity fermentor vessel. Filter sterilized YNB (100 ml of 6 g yeast nitrogen base without ammonium sulphate (DIFCO) and 10 g ammonium sulphate) and 5 ml 0.05% Biotin was added to the autoclaved media.

Fermentation conditions

The enzyme expression was scaled-up by growth of the culture in a 1-liter fermentor using a Bioflow 3000 console (New Brunswick Scientific, Edison, NJ). Before inoculation the fermentor containing complex media was maintained at 30°C with agitation at 300 rpm and air flow of 60 ml/min for 5 hrs. The dissolved air was set as 100%. The fermentor was inoculated with a 50 ml overnight culture of the ST6Gal-I *Pichia* transformant in BMGY. During the growth phase the culture was maintained at 30°C and DO at 50% by varying agitation between 300-1000 rpm. Once agitation reached at 1000 rpm DO was maintained at 50% by partially replacing airflow with pure oxygen. Acidity in the frementor was titrated by addition of 28% NH_4OH. Feeding of glycerol was accomplished by addition of 15 ml of 50%v/v glycerol, injected after 12-15 h, after the batch phase glycerol was consumed. After 18-24 h, when the glycerol in the culture was consumed, ST6Gal-I enzyme expression was then induced by addition of methanol to a final concentration of 0.3%. The methanol level was maintained at 0.3% by feeding methanol containing PTM1 salts (2 ml in 100 ml). Methanol feed was controlled using a methanol probe with a feedback controller (Raven

Biotech, Canada). Also, at the start of the induction phase the culture temperature was lowered to 20°C and the pH was increased to 6.5 and maintained at those levels. After 2 days of induction and 100 ml of methanol consumption, the medium was harvested by centrifugation at 6000xg for 20 min.

2.7 Minimal media Fermentor expression (1 liter expression)

Preparation of minimal media

Minimal media contained 27.5 g NaH_2PO_4 and 15 ml glycerol dissolved in water for a total volume of 800 ml. Antifoam 104 (0.05 ml, Sigma) was added and pH was adjusted between 5.5-5.7. The media was then autoclaved in a 1.6 L fermentor vessel. Filter sterilized YNB (100 ml of 6 g yeast nitrogen base without ammonium sulphate (DIFCO) and 10 g Ammonium sulphate) and 5 ml 0.05% biotin was added to the autoclaved media.

Fermentation conditions

The enzyme expression was scaled-up by growth of the culture in a 1-liter fermentor using Bioflow 3000 console (New Brunswick Scientific, Edison, NJ). Before inoculation the fermentor containg complex media was maintained 30°C with agitation at 300 rpm and air flow of 60 ml/min for 5 h. The dissolved oxygen was set at 100%. The fermentor, was inoculated with a 50 ml overnight culture of the ST6Gal-I *Pichia* transformant in BMGY. During the growth phase the culture was maintained at 30°C and DO at 50% by varying agitation between 300-1000 rpm. Once agitation reached at 1000 rpm the DO was maintained at 50% by partially replacing airflow with pure oxygen. Acidity in the fermentor was titrated by addition of 28% NH_4OH. Feeding of glycerol was accomplished by addition of 15 ml of 50%v/v glycerol, injected after 12-15 h, after the batch phase glycerol was consumed. At this

point titration with NH_4OH was stopped and replaced by 5 M KOH as a base and $(NH_4)_2SO_4$ was used as a nitrogen source. $(NH_4)_2SO_4$ (50% w/v) was added at the rate of 3 ml/hr (total of 30 g) during the second growth phase and entire induction phase. After 18-24 h, when the glycerol in the culture was consumed, ST6Gal-I enzyme expression was then induced by addition of methanol to a final concentration of 0.3%. The methanol level was maintained at 0.3% by feeding methanol containing PTM1 salts (2 ml in 100 ml). Methanol feed was controlled using a methanol probe with a feedback controller (Raven Biotech, Canada). Also, at the start of induction phase the culture temperature was lowered to 20°C and the pH was increased to 6.5 and maintained at those levels. After 2 days of induction and 100 ml of methanol consumption, the medium was harvested by centrifugation at 6000xg for 20 min.

2.8 Screening of protease inhibitors during induction phase

Screening of protease inhibitors was carried out as shown in Table 1. A single colony was picked for inoculation of 80 ml of BMGY medium. The culture was allowed to grow for two days in a horizontal shaker at 225 rpm incubated at 30 °C. The cells were collected by centrifugation at 3000×g at 10°C for 10 min and resuspended in 80 ml of BMMY containing 0.5% v/v methanol. The cell culture medium was then split into 16 equal aliquots (5 ml each). Subsequently, protease inhibitors were added to each tube as described in the Table 1. The final concentration of each of the inhibitor in their respective tubes is described in Table 2. The addition of all the inhibitors was performed at the beginning of induction phase. However, the addition of pepstatin A and chymostatin was repeated every 24 hr. The cultures were continuously shaken at 225-250 rpm for 2-3 days with daily addition of 100 µl of 50%

v/v of methanol every 12 hr. The expression of wild type recombinant enzyme was monitored by Sialyltransferase activity.

2.9 Complex media fermentor expression with Protease inhibitors (1litre scale)

Preparation of complex media

Complex media was composed of 10 g yeast extract, 20 g peptone, 27.5 g NaH_2PO_4 and 15 ml glycerol dissolved in water for a total volume of 800 ml. Antifoam 104 (0.05 ml) (Sigma) was added and pH was adjusted between 5.5-5.7. The media was then autoclaved in the 1.6 L fermentor vessel. Filter sterilized YNB (100 ml of 6 g yeast nitrogen base without ammonium sulphate (DIFCO) and 10 g ammonium sulphate) and 5 ml 0.05% Biotin was added to the autoclaved media.

Fermentation conditions

Fermentation conditions were kept similar to the above described conditions in the "complex media fermentor expression" section. As described earlier before inoculating the complex media in the fermentor vessel, the media was equilibrated at 30°C with agitation at 300 rpm and air flow of 60 ml/min for 5 h. The dissolved oxygen was set as 100%. The fermentor, was inoculated with a 50 ml overnight culture of the ST6Gal-I *Pichia* transformant in BMGY. During the growth phase the culture was maintained at 30°C and the DO at 50% by varying agitation between 300-1000 rpm. Once agitation reached at 1000 rpm DO was maintained at 50% by partially replacing airflow with pure oxygen. Acidity in the frementor was titrated using 28% NH_4OH. Feeding of glycerol was accomplished by addition of 5 ml of 50%v/v glycerol, injected after 12-15h, after the batch phase glycerol was consumed. After 18-24h, when the glycerol in the culture was consumed, ST6Gal-I enzyme

expression was induced by the addition of methanol to a final concentration of 0.3% followed by addition of the following protease inhibitors: filtered sterilized BSA (10 g in 50 ml water), 20 ml EDTA (0.5 M), 1 ml pepstatin A and 1 ml chymostatin (both dissolved in DMSO) with the final concentration of 10 mg/L and 10 μM, respectively). Addition of 1 ml pepstatin A and 1 ml chyomstatin was repeated every 12 h until the end of induction phase. The methanol level was maintained at 0.3% by feeding absolute methanol. Also, at the start of induction phase the culture temperature was lowered to 20°C and pH was increased to 6.5 and maintained at those levels for the duration of the fermentation. After 2 days of induction and 100 ml of methanol consumption, the medium was harvested by centrifugation at 6000xg for 20 min.

2.10 Minimal media Fermenator expression with Protease inhibitors (1litre scale)

Preparation of minimal media

Minimal media contained 27.5 g NaH_2PO_4 and 15 ml glycerol dissolved in water for a total volume of 800 ml. Antifoam 104 (0.05 ml) (Sigma) was added and pH was adjusted between 5.5-5.7. The media was then autoclaved in a 1.6 L fermentor vessel. Filter sterilized YNB (100 ml of 6 g yeast nitrogen base without ammonium sulphate (DIFCO) and 10 g ammonium sulphate) and 5 ml 0.05% biotin was added to the autoclaved media.

Fermentation conditions

The enzyme expression was scaled-up by growth of the culture in a 1 L fermentor using Bioflow 3000 console (New Brunswick Scientific, Edison, NJ). Before inoculation, the fermentor containing minimal media was maintained 30°C with agitation at 300 rpm and air flow of 60 ml/min for 5 h. The dissolved oxygen was set as 100%. The fermentor, was

inoculated with a 50 ml overnight culture of the ST6Gal-I *Pichia* transformant in BMGY. During the growth phase the culture was maintained at 30°C and DO at 50% by varying agitation between 300-1000 rpm. Once agitation reached 1000 rpm the DO was maintained at 50% by partially replacing airflow with pure oxygen. Acidity in the fermentor was titrated by 28% NH_4OH. Feeding of glycerol was accomplished by addition of the 5 ml of 50% v/v glycerol, injected after 12-15h, after the batch phase glycerol was consumed. After 18-24h, when the glycerol in the culture was consumed, ST6Gal-I enzyme expression was induced by the addition of methanol to a final concentration of 0.3% followed by addition of the following protease inhibitors: filtered sterilized BSA (10 g in 50 ml water), 20 ml EDTA (0.5 M), 1 ml pepstatin A and 1 ml chymostatin (both dissolved in DMSO) with the final concentration of 10 mg/L and 10 μM, respectively). Addition of 1 ml pepstatin A and 1 ml chyomstatin was repeated every 12 h until the end of induction phase. The methanol level was maintained at 0.3% by feeding absolute methanol. Also, at the start of induction phase culture temperature was lowered to 20°C and pH was increased to 6.5 and maintained at those levels for the duration of the fermentation. After 2 days of induction and 100 ml of methanol consumption, the medium was harvested by centrifugation at 6000xg for 20 min.

2.11 Isotope labeling in 1L fermentor

Fermentation was carried out similar to that of unlabeled protein expression in minimal media. For single ^{15}N isotope incorporation, $(^{15}NH_4)_2SO_4$ (Cambridge Isotope Ltd) was used in the second growth phase and induction phase instead of unlabeled $(NH_4)_2SO_4$. For double labeling with ^{15}N and ^{2}H, the fermentation was carried out similar to that of the single isotope incorporation except all the media and other components were prepared in

95%, 97%, or 99% D_2O (Cambridge Isotope Ltd). For triple isotope labeling ^{13}C, ^{15}N and ^{2}H fermentation was carried out similar to that of double label incorporation except the culture waced induced with ^{13}C methanol (Cambridge Isotope Ltd) instead of unlabeled methanol. For double and triple isotope labeling experiments the seed culture was prepared by inoculating a single colony in 5 ml of BMGY media and allowed to grow overnight. Then, next day the seed culture was centrifuged at 1500xg for 5 min and resuspended in 50 ml BMG media in 95 %D_2O and allowed to grow for 2-3 days. The isotope-labeled cultures also took longer to grow in the fermentor than unlabeled cultures.

2.12 Purification of ST6Gal-I

2.12.1 IMAC:

The harvested culture medium was filtered through 0.45 μM membrane and diluted with water or 300 mM NaCl (1:3) followed by adjusting the pH to 7.4 using NaOH. The sample was then applied to a Ni^{+2} Superflow resin (Qiagen) column (33 x 10 mm, Millipore), pre-equilibrated with 25 mM HEPES containing 300 mM NaCl, at a flow rate of 8-10 ml/min. The column was washed with 150 ml of equilibration buffer at a flow rate of 3 ml/min. The protein was then eluted with 150 ml of an initial linear gradient of 0-60 mM imidazole followed by 50 ml of 60 mM imidazole isocratic flow and a 75 ml short linear gradient from 60 mM-300 mM imidazole, followed by 200 ml of 300 mM imidazole.

2.12.2 Gel filtration

The second peak from the IMAC column, which contained predominantly ST6Gal-I activity was pooled and concentrated to 3 ml using an ultrafiltration unit (50 ml, Amicon)

containing 10 kDa cutoff membrane. The sample was then applied at 0.5 ml/min onto a Superdex 75 gel filtration column (1.6 × 65 cm, Amersham Pharmacia Biotech) pre-equilibrated with 25 mM MES and 200 mM NaCl at pH 6.5.

2.13 Purification and deglycosylation of of ST6Gal-I

2.13.1 IMAC

The harvested culture medium was filtered through 0.45 µM membrane and diluted with water or 300 mM NaCl (1:3) followed by adjusting the pH to 7.4 using NaOH. The sample was then applied to a Ni^{+2} Superflow resin (Qiagen) column (33 x 10 mm, Millipore), pre-equilibrated with 25 mM HEPES containing 300 mM NaCl, at a flow rate of 8-10 ml/min. The column was washed with 150 ml of equilibration buffer at a flow rate of 3 ml/min. The protein was then eluted with 150 ml of an initial linear gradient of 0-60 mM imidazole followed by 50 ml of 60 mM imidazole isocratic flow and a 75 ml short linear gradient from 60 mM-300 mM imidazole, followed by 200 ml of 300 mM imidazole.

2.13.2 Deglycosylation

The second peak from the IMAC column was then dialyzed against 50 mM citrate buffer pH 5.5 containing 10 mM EDTA by concentration and dilution in a stirred pressure cell (Amicon Inc.) using a 10 kDa cutoff ultrafiltration membrane. The pool from second peak from the IMAC column was concentrated to 50 ml and then diluted back to 150-170 ml and concentrated again. This process of concentration and dilution was repeated three times. The fourth time, when the volume reached to 50 ml a protease inhibitor cocktail (Complete EDTA free, Roche Diagnostics) tablet dissolved in 1 ml of citrate buffer, pH 5.5, was added

and then the sample was further concentrated to 3-4 ml. Endo H_f (15-20 µl : New England Biolab Inc. NE) was added to the sample and it was incubated for 1-1 ½ h.

2.13.3 ConA Sepharose (Removal of non-deglycosylated ST6Gal-I)

To the deglycosylated sample from above a 0.5 ml aliquot of a dilavalent metal ion solution was added (containing 200 mM each of $CaCl_2$, $MnCl_2$ and $CoCl_2$) and diluted to 10-11 ml with equilibration buffer (25 mM Tris-HCl, 1mM of each of $CaCl_2$, $MnCl_2$ and $CoCl_2$) for ConA Sepharose column (ConA Sepharose, Amersham in 11×10 mm column, Millipore).

The sample was applied to the ConA column pre-equilibrated with 5 column volume of equilibration buffer at flow rate 1 ml/min. The flow through was collected and used further. After every run the column was regenerated by bringing the column at room temperature and washing with 5 column volumes of 1 M maltose (in equilibration buffer).

2.13.4 Gel filtration

The flow through material from the ConA sepharose column was concentrated to 3 ml in a stirred pressure cell, using a 10kDa cutoff ultrafiltration membrane. The sample it was then applied at 0.5 ml/min onto a Superdex 75 gel filtration column (1.6×65 cm, Amersham Pharmacia Biotech) pre-equilibrated with 25 mM MES and 200 mM NaCl at pH 6.5.

2.14 N-glycan analysis on ST6Gal-I

2.14.1 Preparation of Glycopeptides and Release of N-Linked Glycans

Following gel filtration, purified ST6Gal-I was pooled dialysed and freeze dried. The protein (1mg) was suspended in 200 µl of 50mM phosphate buffer pH 7.5 and boiled for 5 min. After cooling to room temperature, 5 µl of trypsin solution (2 mg/ml in trypsin buffer: 50mM NH_4HCO_3, pH 7.8) and 25 µl of chymotrypsin solution (2 mg/ml in trypsin buffer) were added. Digestion was allowed to proceed overnight at 37 °C then the mixture was boiled for 5 min to inactivate both proteases. The sample was then dried by vacuum centrifugation. The dried peptide mixture was resuspended in 250 µl of 5% acetic acid (v/v) and loaded onto a Sep-Pak C_{18} cartridge column [155]. The cartridge was washed with 10 column volumes of 5% acetic acid. Glycopeptides were eluted stepwise, first with 3 volumes of 20% isopropyl alcohol in 5% acetic acid and then with 3 volumes of 40% isopropyl alcohol in 5% acetic acid. The 20 and 40% isopropyl alcohol steps were pooled and evaporated to dryness. Dried glycopeptides were resuspended in 50 µl of 20 mM sodium phosphate buffer, pH 7.5, followed by addition 2µl PNGaseF (New England Biolab), the mixture was then incubated overnight at 37 °C. The released oligosaccharides were separated from peptide and enzyme by passage through a Sep-Pak C_{18} cartridge. The digestion mixture was adjusted to 5% acetic acid and loaded onto the Sep-Pak. The column run-through and an additional wash with 3 column volumes of 5% acetic acid, containing released oligosaccharides, were collected together and evaporated to dryness. The dried samples were then methylated as per standard protocol [155].

2.14.2 Matrix-assisted Laser Desorption-Ionization/Time-of-Flight Mass Spectrometry

MALDI-TOF/MS was performed on an Applied Biosystems Voyager System mass spectrometer in positive linear mode. Permethylated glycan solutions were mixed in a 1:1 ratio with matrix (1 µl of glycan solution with 1 µl of 2,5-dihydroxybenzoic acid as a 20 mg/ml solution in 50% methanol), and 1 µl of the mixture was spotted on the stainless steel sample plate. Ions were generated by irradiation with a pulsed nitrogen laser (337nm) with a 4-ns pulse time. The instrument was externally calibrated with permethylated authentic $(Man)_5$-$(GlcNAc)_2$ oligosaccharide. Average masses were obtained from spectra summed over 100-200 shots with a 337 nm nitrogen laser. Ions corresponding to individual glycans, which gave signal intensities greater than 3-fold above background, were quantified relative to the sum of the intensities of all candidate glycan ions to give "% total profile" for each glycan [141, 155].

2.15 NH_2-terminal sequencing:

An aliquot of the purified enzyme (50 µg) was subjected to SDS-PAGE and transferred to a PVDF membrane (Millipore) using a transfer cell as described by manufacturer (BioRad Inc., CA). The blot was stained using Coomassie R-250 and destained by submerging the blot in 50% v/v methanol. The bands corresponding to ST6Gal-I were excised and NH_2-terminal protein sequencing was performed by the Microchemical and Proteomic Facility (Emory University, Georgia) using an Applied Biosystems 494 protein sequencer.

2.16 Purification of MBP-ST6Gal-I with TEV protease digests

2.16.1 IMAC

The harvested culture medium was filtered through 0.45 µM membrane and diluted with water or 300 mM NaCl (1:3) followed by adjusting the pH to 7.4 using NaOH. The sample was then applied to a Ni^{+2} Superflow resin (Qiagen) column (33 x 10 mm, Millipore), pre-equilibrated with 25 mM HEPES containing 300 mM NaCl, at a flow rate of 8-10 ml/min. The column was washed with 150 ml of equilibration buffer at a flow rate of 3 ml/min. The protein was then eluted with a 150 ml of an initial linear gradient of 0-50 mM imidazole followed by 50 ml of 50 mM imidazole isocratic flow and a 75 ml short linear gradient from 50 mM-250 mM imidazole, followed by 200 ml of 250 mM imidazole

2.16.2 Gel filtration:

The second peak from IMAC was then diafiltered by two fold dilution with water followed by concentration to 50 ml by ultrafiltration with a 50 kDa cutoff membrane. A protease inhibitor cocktail tablet (dissolved in 1ml of 20mM HEPES buffer, pH 7.0) was added to the sample before concentrating it further to 3 ml. The sample was then applied at 0.5 ml/min onto a Superdex 75 gel filtration column (1.6×65 cm, Amersham Pharmacia Biotech) pre-equilibrated with 20 mM HEPES and 100 mM NaCl at pH 7.0.

2.16.3 TEV protease digest and Deglycosylation

The peaks from gel filtration with ST6Gal-I enzyme activity were collected and concentrated to 5 ml. ProTEV protease (15 µl, Promega Inc.) was then added to the sample followed by incubation at 4°C for 5 h to overnight with intermittent shaking. After digestion

the pH was lowered to 5.5 by titration with 100 mM citric acid. 50X EDTA free protease inhibitor cocktail tablet (200 µl, 1 tablet dissolved in 1 ml water) was added followed by the addition of 15 µl of Endo H_f (Roche diagnostics) and the sample was incubated for 1-1 ½ h at 37°C.

2.16.4 IMAC:

TEV protease-digested and deglycosylated MBP-ST6Gal-I from the above step was then diluted to 20 ml with 25 mM HEPES, containing with 200 mM NaCl. The sample was then applied to a Ni^{+2} Superflow resin (Qiagen) column (11×10 mm Millipore), pre-equilibrated with 25 mM HEPES containing 200 mM NaCl, at 3 ml/min. The column was washed with 50 ml of equilibration buffer at a flow rate of 3 ml/min. The protein was eluted with an initial 100 ml linear gradient of 0-20 mM imidazole followed a 75 ml linear gradient from 20 mM-250 mM imidazole, followed by 50 ml of 250 mM imidazole.

2.16.5 Gel filtration

The first peak from above IMAC column was collected and concentrated to 3 ml in presence of protease inhibitor cocktail and applied to a Sephadex-75 column pre-equillibrated with 25 mM MES buffer with 100 mM NaCl.

2.17 Sialyltransferase Assay and Kinetics

2.17.1 Enzyme Assay

Enzyme assays were performed in duplicate with slight modifications as described previously [156]. Unless otherwise stated, upto 22 µl of harvested cell medium containing

expressed recombinant sialyltransferases was assayed in a 60 µl reaction volume containing a mixture of 25 nmol of CMP-NeuAc and CMP-[^{14}C]NeuAc (DuPont NEN; 12000 cpm/per assay) and as donor, 2.2 mM N-acetyl lactosamine as an acceptor, 50 µg of bovine serum albumin, 50 mM sodium cacodylate (pH 6.5) with 0.5% Triton X-100. After incubation at 37 °C for 1 h, the reaction was stopped by adding 250 µl of chilled 1mM Phosphate buffer at pH 6.8 and immediately the radio-labeled product was isolated by applying the assay reaction to an anion exchange resin (Dowex resin regenerated for phosphate) column (Bio-Rad). The enzyme product was directly collected into a scintillation vial (7ml). The column was then washed with 250 µl, 500 µl and 1000 µl of chilled 1 mM phosphate buffer (pH 6.5). Scintillation fluid (Fisher Scientific Inc) was added (4.5 ml) and the samples were counted immediately as described previously [157].

2.17.2 Enzyme kinetic analyses

Initial rates (v) for the enzymes were determined in following manner. The reaction mixture in 60 µl contained 15 µl of 4x universal buffer (100 mM succinic acid, 100 mM MES, 100 mM, and 100 mM TRIS adjusted to pH with 10 M NaOH) pH 6.5 and varied substrate concentration. In first set of experiments the donor substrate, CMP-Neu5Ac was varied from 15 µM - 1000 µM with constant CMP-[^{14}C]Neu5Ac (DuPont NEN; 30,000 cpm/per assay), and N-acetyl lactosamine, the acceptor substrate was kept constant 6.0 mM. In the second experiment the donor substrate was kept constant at 1 mM CMP-Neu5Ac with with constant CMP-[^{14}C]Neu5Ac (DuPont NEN; 30,000 cpm/per assay) and N-acetyl lactosamine concentration, the acceptor substrate concentartion, was varied from 0.4mM to 6.0mM. After incubation at 37°C for 15 min, the reaction was stopped by adding 250 µl of

chilled 1 mM Phosphate buffer at pH 6.8 and the radio-labeled product was isolated by chromatography over an anion exchange resin (Dowex resin regenerated for phosphate) column (Bio-Rad) as described above.

The catalytic coefficient (k_{cat}) and Michaelis constant (k_m) values were determined by fitting initial rates to the Michaelis-Menten function (eq. 1) by non-linear regression analysis using SigmaPlot 9.0 (Jandel Scientific, San Rafael, CA).

$$v = \frac{k_{cat}.E_t[S]}{k_m + [S]} \qquad \text{(Eq. 1)}$$

where E_t is total enzyme concentration.

$k_{cat}t/k_m$ values were derived from reciprocal plots of v and [S] where needed. The k_m and k_{cat} was determined from the initial rate (v) using 5-8 substrate concentrations, by Linewear-Burk plots.

2.17.3 pH-rate dependence analysis

The k_{cat}/k_m values were determined from initial rates of enzyme reactions at pH values ranging from 5-9 using the 4x universal buffer described above. Plots of log k_{cat}/k_m versus pH were fitted to the appropriate bell-shaped equation to estimate the pK_a values of possible amino acid residues involved in substrate binding or catalysis.

2.18 Isotope labeling of MBP-ST6Gal-I fusion protein

A single zeocin resistant colony was inoculated into 5 ml a BMGY in 50 ml conical tube. The *Pichia* culture was then incubated at 30°C and allowed to grow overnight in a horizontal shaker at 225 rpm. The next morning the cells were collected by centrifugation at 3000×g for 10 min and resuspended in 50 ml of BMG (6.0 g/l YNB (yeast nitrogen base)

(DIFCO), 8.0 g/l $(^{15}NH_4)_2SO_4$, 0.1 M phosphate buffer pH 6.0 and 4 μg/l biotin and 10g/L glycerin) in 95% D_2O in a 250 ml conical flask then incubated at 30°C and allowed to grow for 24 hr in a horizontal shaker at 225 rpm. The cell culture was then diluted with 1L of BMG media in 95% D_2O which was split into two 500 ml samples in 3L flask and then incubated at 30°C for 24 hrs in a horizontal shaker at 225 rpm.

The cells were then transferred into several 50 ml conical tubes and collected by centrifugation at 1500xg for 5 min. The cells from all the tubes were then resuspended in 1.4 to 1.5L of BMM media (6.0 g/l YNB (yeast nitrogen base) (DIFCO), 8.0 g/l $(^{15}NH_4)_2SO_4$, 0.1 M phosphate buffer pH 6.0 and 4 μg/l biotin, 5 ml/l of (^{12}C or ^{13}C)-methanol and 10 g/l BSA). The culture was then split into several (240-260) 50 ml conical tubes, each containing 6 ml of cell culture. The cultures were continuously shaken at 225-250 rpm for two and half days with addition of 30 μl of methanol (^{12}C for double labeling or ^{13}C for triple labeling) every 12 hr. Similarly, protease inhibitors (1 ml/l pepstatin A (10μM in DMSO) and 1ml/l chymostatin (10 mg/ml in DMSO) were added at 0 hr, 24 hr and 48 hr intervals during the induction. The media was harvested after two and half days of induction and treated with TPCK and PMSF before proceeding with purification.

2.19 NMR Analysis

The efficiency of isotope incorporation in recombinant ST6Gal-I generated in fermenter cultures was carried out using "Inova 900" spectrometer (Varian Inc.). The sample contained about 100 μM of uniformly labeled (^{15}N) ST6Gal-I in 25 mM MES and 150 mM NaCl in H_2O. 1D 1H spectra were collected in such a way that for one spectrum the proton signal was decoupled from ^{15}N during acquisition while in the other the proton was left

coupled to ^{15}N during acquisition. 2D ^{1}H-^{15}N HSQC spectra were collected on the "Inova 800" spectrometer equipped with a high sensitivity cold probe using gNhsqc sequence [158]. While, 1024 complex points were collected in the directly-detected dimension and 128 complex points in the indirect dimension, 64 points were collected in transient dimension for FID.

2D ^{1}H-^{15}N HSQC spectrum of the uniformly labeled (^{15}N) ST6Gal-I and deglycosylated (Endo H treated) enzyme (100μM) in 25 mM MES and 150 mM NaCl was collected on the "Inova 900" spectrometer equipped with a high sensitivity cold probe. The experimental parameters were similar to that used for the HSQC of the untreated sample described above.

Similarly, a 2D ^{1}H-^{15}N HSQC spectrum of the double labeled (^{2}H and ^{15}N) ST6Gal-I and deglycosylated (Endo H treated) sample (100 μM in 25 mM MES and 150 mM NaCl) was collected on the "Inova 900" spectrometer equipped with a high sensitivity cold probe. The experimental parameters for the experiment were similar to that used for the HSQC of untreated sample described above.

A 3D of ^{2}H-^{13}C-^{15}N HNCA spectrum of the triple labeled (^{2}H,^{13}C,^{15}N) ST6Gal-I and deglycosylated sample (60 μM) in 25 mM MES and 150 mM NaCl was collected on "Inova 900" and "Inova 600"spectrometers both equipped with a high sensitivity 3 mm cold probe. A total of 52 complex points were collected in the first indirect dimension and 30 complex points were collected in the second indirect dimension. 1024 points were collected in the direct dimension. 64 transients were accumulated for each FID.

2.20 Surface plasmon resonance study

2.20.1 Instrument and sensor chip preparation

The biosensor assays were performed on a BIACORE T100 instrument and data analyses were manipulated and fit using the BIAevaluation 3.1 software (Biacore AB, Piscataway, NJ). Prior to beginning a binding assay session, the liquid handling portion of the instrument was routinely cleaned utilizing the automated subroutines as previously reported [159]. The system was first subjected to DESORB subroutine, which allows the system to be flushed with 0.5% SDS and 50 mM glycine pH 9.5. New biosensor chips were docked into the instrument and were subjected to preconditioning by applying two consecutive pulses (20 µl) of 50 mM NaOH, followed by one pulse of 10 mM HCl and one pulse of 0.1% SDS followed by a normalization which allows the calibration of the sensor chip surface using a 70% glycerol solution.

2.20.2 Immobilization on CM5 chips by amine coupling

The amine coupling kit (1-ethyl-3-(dimethylamiopropyl)carbodiimide, EDC), (N-hydroxysuccinimide, NHS) and ethanolamine), HBS-EP (20 mM HEPES, pH 7.4, 150 mM NaCl, 3.4 mM EDTA, 0.01% polysorbate P20) and research-grade CM5 sensor chips were purchased from the manufacturer (Biacore Inc.). Recombinant ST6Gal-I immobilized chip surfaces were prepared at 25 ºC by an amine-coupling method using the automated Application Wizard (BIAcore T100 control software, BIAcore AB). The flow cells were activated by injecting a mixture of 50 mM NHS and 200 mM EDC over the CM5 sensor chips surface for 7 min at 5 µl/min. The recombinant ST6Gal-I (10mg/ml) prepared in 25 mM MES and 100 mM NaCl (pH 6.5) was diluted in the 10 mM Acetate buffer (pH 6.5) to

obtain a concentration of 20 µg/ml prior to injection onto the activated surface. The desired immobilization level was achieved by specific contact time. The remaining reactive groups were either blocked by injection of 1 M ethanolamine-HCl at pH 8.5 for 7 min at 10 µl/min or were left untreated. The immobilization efficiency for ST6Gal-I was about 1,500 RU/min at a flow rate of 10 µl/min and 25ºC using HSB-EP as running buffer. Mock derivatized flow cells served as reference surfaces.

Similarly, desialylated glycopeptides from the pepsin digest of asialofeutin proteins were immobilized on CM5 Chip using amine coupling as described above. The maximum immobilization that was achieved was 900 RU.

2.20.3 Immobilization on Ni-NTA chip by Ni-affinity

Immobilization of ST6Gal-I on a Ni-NTA chip was carried out similar to previously reported [160]. Activation of the NTA chip was carried out at room temperature (25°C) on a BIAcore T100 instrument (Biacore, Sweden). NTA chips were loaded with 50 µl of 0.1 mM nickel sulfate in HBS running buffer (10 mM Hepes, 150 mM NaCl, 0.005% (v/v) P20, pH 7.4) for 5 min at 10 µl/min. ST6Gal-I enzyme was diluted in HBS and loaded on to activated Ni-NTA sensorchip flow cell 2 (Fc2) with a 5 ml/min flow until a signal of approx 10000 RU was recorded. After ST6Gal-I loading was completed the instrument was allowed to run in standby mode until the baseline reached equilibrium.

2.21 Immobilization of ST6Gal-I on Affigel

Approximately 100 µl of AffiGel 10 slurry was placed in a 1.5ml microfuge tube was centrifuged for 1 min at 14,000 rpm at 4ºC. The supernatant was removed by pipette. The

resin was then resuspended in 200 µl of ice cold (4°C) and then it was centrifuged and the supernatant was removed by pipette. Resuspending and centrifugation step was repeated 3 times. To this slurry, 25 µl of ST6Gal-I and 175 µl of immobilization buffer (10 mM Sodium acetate) pH 4.5 or 5.5 or 10 mM MES pH 6.5) with or without 1 mM CMP-Sialic acid. The mixture was incubated at 4ºC for 1 hr with shaking. The mixture was centrifuged and supernatant was removed pipetted and assayed for enzyme activity. The immobilized enzyme column resin was washed with the respective immobilization buffer. The remaining active groups on gel were blocked by addition of 100 µl of 50 mM 2-mercaptoethanol (pH 8.0). Unreacted mercaptoethanol was removed by centrifugation as described above. The slurry was resuspended in 50 µl 25 mM cocodylate buffer and 20 µl of this slurry was used in the enzyme assays.

2. 22 Biotinylation of ST6Gal-I

Biotinylation of ST6Gal-I was carried out using EZ-Link® NHS-LC-Biotin kit, as described in instruction manual (Pierce Labs Inc.). The purified ST6Gal-I, derived from the original construct or the TEV protease cleaved MBP-ST6Gal-I, dialyzed in PBS buffer by alternate concentration and dilution using Centricon® 10kDa cutoff membrane concentrator. NH-LC-Biotin reagent (10 µl of 10 mM reagent in cold PBS prepared just before addition adding to enzyme) was added to 200 µl of purified ST6Gal-I (2 mg/L in PBS). The sample was incubated at 4ºC for 2 hrs with agitation. Sample was washed on a 10 kDa ultrafiltration membrane for buffer exchange and removal of unutilized or hydrolyzed NH-LC-Biotin reagent. An aliquot was taken for enzyme assays.

2.23 Methylation and Biotinylation of MBP-ST6Gal-I fusion protein

The purified MBP-ST6Gal-I fusion protein washed and concentrated in 50 mM phosphate, pH 7.4, 100 mM NaCl buffer with final protein concentration of 2 mg/ml. To a 250 ml enzyme solution 6 µl of 1M DMAB and 12 µl formaldehyde was added. The sample was covered in foil and agitated on rocker at 4ºC for 2 h followed by another addition of 6 µl of 1M DMAB and 12 µl formaldehyde. After 2 h incubation at 4ºC with agitation, 3 µl of 1M DMAB was added and allowed to shake overnight at 4ºC. The next day 50 µl of 1 M $(NH_4)_2SO_4$ was added to the sample mixture and allowed to incubate with shaking at 4ºC. An aliquot was taken for enzyme assays.

The sample was washed on a 10 kDa ultrafiltration membrane for buffer exchange and removal of all the unutilized DMAB, formaldehyde, and $(NH_4)_2SO_4$ products. An aliquot was taken for enzyme assays. The sample was then digested with TEV protease by adding 1µl of ProTEV protease (Promega Inc) and incubating sample at 4ºC for 4-5 h. An aliquot was taken for enzyme assays.

Biotinylation of this digested mixture was performed as described above. An aliquot was taken for enzyme assays.

2.24 Screening of pGAPαC-ST6Gal-I transformants

The zeocin-resistant colonies were screened for the expression of recombinant ST6Gal-I enzyme activity. Fifteen to twenty colonies were analysed for expressed enzyme activity in order to screen for the highest expressor. Each of the zeocin resistant colonies was inoculated into 5 ml YPD (10 g/l yeast extract, 20 g/l peptone 20 g/l dextrose, and 4 µg/l D-biotin) in 50 ml conical tubes. The cultures were grown in a horizontal shaker at 225 rpm

incubated for 2-3 days at 30 °C. An aliquot was taken every 24 hrs tp check the expression of recombinant enzyme by Sialyltransferase activity.

Table 1: Protease Inhibitor matrix

	BSA	Chymo.	Pep. A	EDTA	EACA
1	-	-	-	-	-
2	+	-	-	-	-
3	-	+	-	-	-
4	-	-	+	-	-
5	-	-	-	+	-
6	-	-	-	-	+
7	-	+	+	-	-
8	-	+	-	+	-
9	-	+	-	-	+
10	-	-	+	+	-
11	-	-	+	-	+
12	-	-	-	+	+
13	-	+	+	+	-
14	-	-	+	+	+
15	-	+	+	+	+
16	+	+	+	+	+

Table 2: Protease inhibitors final concentration and their mechanism of action

Protea1se inhibitor	**Mechanism**	**Concentration**
BSA/HAS	Substrate for protease	10 g/L
Chymostatin	Serine protease inhibitor	10 μM
Pepstatin A	Aspartic acid protease inhibitor	10 mg/L
EDTA	Metalloprotease inhibitor	10 mM
ε-Amino Caproic Acid	Carboxypeptidase inhibitor	100 mg/L

CHAPTER 3

RESULTS

3.1 Expression of recombinant rat ST6Gal-I

In order to generate protein for structural and functional studies, constructs were prepared for large scale expression in *Pichia*. The first construct, ST6Gal-I/pPICZαC contained the putative catalytic domain coding region along with an 8xHis tag at NH_2 terminus (Figure 7 and Appendix 1A). The soluble protein expressed from this construct was further degraded from NH_2 terminal (details are described later in this section). Therefore we expressed another construct truncated construct, HA-ST6Gal-I/pPICZαC, in which an NH_2-terminal HA epitope was inserted between the 8xHis tag and the ST6Gal-I coding region (Appendix 1B) and this construct had a deletion of 14 amino acids from catalytic domain and four extra amino acids were added as a consequence of mutagenesis.

An alternative construct of HA-ST6Gal-I/pPICZαC, HA-ST6Gal-I_S/pPICZαC, was made by chemical synthesis which did not have the extra residues that were found in the above construct (Appendix 1C). Also this truncated construct was sub-cloned into two other modified pPICZαC vectors and named HA-STRP-ST6Gal-I_S/pPICZαC (Appendix 1D) and MBP-ST6Gal-I_S (Appendix 1E), which were the streptavidin and Maltose Binding Protein fusion proteins, respectively. The HA-STRP-ST6Gal-I_S/pPICZαC construct has NH_2 terminal 8xHis tag followed by HA tag followed by monomeric streptavidin and C terminal ST6Gal-I_S coding region.

Later three more construct were made which retained the NH_2-terminal 14 residues that had been deleted in HA-ST6Gal-I/pPICZαC. HA-ST6Gal-I_L/pPICZαC, was equivalent to HA-ST6Gal-I_S/pPICZαC construct, which retained 14 deleted residues (Appendix 1G). Moreover, HA-ST6Gal-I_L/pPICZαC construct was sub-cloned into fusion protein pPICZαC vectors to make HA-STRP-ST6Gal-I_L/pPICZαC (Appendix 1G) and MBP-ST6Gal-I_L (Appendix 1H) constructs, which also retained the 14 deleted residues.

Transformants were generated using linearlized constructs in the *Pichia* host strains KM71H (unless and otherwise mentioned) by homologous recombination of the construct in the *Pichia* genome at the 5' region of the alcohol oxidase gene (AOX1). Transformants were selected by antibiotic selection to the vector-encoded zeocin resistance gene. Expression was induced by the methanol-inducible promoter of AOX1.

The recombinant enzymes were expressed as secreted soluble proteins into the culture media, which can be induced by addition of methanol to the *Pichia* culture. The level of expression of the recombinant enzyme was monitored by enzyme activity by measuring the transfer of the ^{14}C labeled CMP-NeuAc to the LacNAc disaccharide. The detectable sialyltransferase activity from the expression culture media was determined following 24 h of methanol induction. Activity increased over the first 48 hr and plateaued after 2-3 days of induction. The recombinant enzyme activity in the induced media started to decline after 3 days of induction indicating that the secreted recombinant enzyme was being degraded faster than it was being secreted into the media. The colony expressing the highest level of expression of ST6Gal-I produced 1187 ± 164 U/L where 1 unit of enzyme activity if 1μmole of sialic acid transferred per minute (Figure 8). Western blots of conditioned culture media

showed a major band corresponding to a 38kDa protein with an additional broad diffuse smear of cross-reactive material up to 70kDa.

3.2 Effect of pH on Expression of ST6Gal-I

The expression of the ST6Gal-I was carried out by a method similar to the one described above. A single colony was inoculated into 50 ml BMGY at pH 6.5 media and incubated in rotating platform shaker for 36 hr. The sample was then divided into 5ml culture tubes. The 5 ml culture was subsequently centrifuged and the supernatant was resuspended in 5 ml BMMY at pH 7.5, 6.5 and pH 5.5. Methanol was added every 12 hr to a final concentration of 0.5%. The cells were harvested and the supernatant was assayed for ST6Gal-I activity. The ST6Gal-I activity increased from pH 5.5 to 6.5 and rapidly dropped at pH 7.5 (Figure 9). This result was expected, since *Pichia* cultures do not grow well above pH 7.0 [161]. Therefore, it was decided that a pH between 5.5 and 6.5 would be used for induction in scale-up of ST6Gal-I expression using 1L fermentor or shake flasks cultures.

3.3 Effect of Temperature on the production of ST6Gal-I

In experiments that test the expression of ST6Gal-I at various temperatures, cultures were generated as described above. A single colony was inoculated into 50 ml BMGY at pH 6.5, which was subsequently incubated in a rotation platform shaker for 36 hr. The sample was then divided into 5 ml culture tubes. The 5 ml culture was subsequently centrifuged and the pellet was resuspended in 5 ml BMMY at 6.5 and incubated at three different temperatures, 30°C, 25°C and 20°C. Methanol was added every 12hr to a final concentration

of 0.5%. The cells were then harvested and the supernatant was assayed for ST6Gal-I activity.

Even though the optimal temperature for the growth is 30°C, the secreted ST6Gal-I activity was observed at non optimal temperatures. In fact there was two fold more enzyme activity at 25°C and 20°C compared 30°C (Figure 10), which indicates that either recombinant ST6Gal-I is unstable and/or is proteolyzed at higher temperatures.

3.4 pH profiling of the Proteases during the fermentation

Experiments to probe the effect of pH on recombinant enzyme proteolysis were carried out using growth and induction conditions as described above and samples were drawn after 5 and 48 h of induction. The protease assays were performed on samples at pH 5.0, 6.0 and 7.2 using FTC labeled casein as substrate (Figure 11). Since the protease activity was higher at pH 6 and 7.0, it was decided to use pH 5.0-5.5 for the induction phase.

3.5 Shake flask expression

Initially the highest expressing colony from tube culture was expressed in 50 ml shake flask culture in borosilicate shake flasks. The expression level of ST6Gal-I activity was in a range from 100-250 Units/L and in some instances there was no detectable enzyme activity. These activity levels were much lower than equivalent 5 ml tube cultures. Similar results were seen in 1L shake flask cultures in 3L borosilicate shake flask. Use of polypropylene or polycarbonate material for shake flask cultures did not improve the yield of enzyme activity of induced culture.

3.6 Expression in 1L fermentor

Unlike shake flask experiments, the cell densities achieved in the fermentor were much higher (>200g/L compared to the 20-25g/L in shake flask cultures). Even though there was an 8-fold higher cell density in controlled fermentor, induction at 30°C gave a very little ST6Gal-I enzyme activity (≈ 300-400 U/L, data not shown) even after 72 hr of induction. However, lowering the induction temperature to 20°C resulted in a > 7-fold increase in enzyme activity to 2878 U/L after 72 hr of induction (Table 3).

Similarly, when minimal media was used for the expression of ST6Gal-I there was little or no enzyme activity detected when induction was carried out at 30°C. However, induction at 20°C yielded 2585 U/L enzyme activity after 60 hr of induction (Table 4).

3.7 Screening of protease inhibitor for the production of ST6Gal-I

Expression of ST6Gal-I was carried out using 5 ml culture media as described above (the cells were grown in BMGY media at 30°C followed by induction in BMMY media at 20°C). The effect of various protease inhibitors alone or in combination was carried out by adding the protease inhibitors in concentration shown in Table 1 and 2. BSA (B), EDTA (E) and ε-amino caprioic acid (Ec) were added only once at the beginning of induction phase, while addition of Pepstatin A (P) and Chymostatin (C) was repeated every 24 hr after they were added at the beginning of the induction phase. Samples were drawn after 48 and 72 hrs of induction for enzyme assay. There was no significant increase in enzyme activity when inhibitors were used alone as a single inhibitors or in combinations of any two inhibitors. However, more than two-fold increase in enzyme activity was measured in the harvested media when three or more inhibitors were used. Similarly there was more than a 3-fold

increase in enzyme activity in harvested media when all 5 protease inhibitors were used (Figure 12).

There was no increase in enzyme activity when Ec was added in combination with P and E or in combination with C, P, and E, indicating that Ec was not effective in reducing proteolysis of ST6Gal-I (Figure 12). Therefore, Ecwas not included for the scale-up using the 1L fermentor.

3.8 Expression in 1L fermentor with protease inhibitors

Unlike the tube culture screening of protease inhibitors, only four protease inhibitors were used during the induction phase of ST6Gal-I expression in 1L controlled fermentor (ε-amino caprioic acid was not added). Addition of protease inhibitors resulted in a significant improvement in the expression level of ST6Gal-I within the first 24 hours. There was at least 5-fold more secreted enzyme activity (6377 U/L) in the media in the first 24 hours (Figure 13) compared to the same time frame with no protease inhibitor addition and overall 2.4-fold more activity (6377 U/L, Table 7) for the media induced with no inhibitors (2878 U/L, Table 3). Also when fermentation was carried out with just two protease inhibitors (B and E), the secreted enzyme activity was reduced to 2/3 (5400 U/L) compared to the addition of all four protease inhibitors after 24 hr of induction. However, it was still 2-fold more enzyme activity in comparison to secreted enzyme activity in the absence of protease inhibitors. This suggest that use of combination of protease inhibitors during fermentation reduces the steady state proteolysis of recombinant ST6Gal-I.

3.9 Expression in 1L fermentor with protease inhibitors in minimal media

When all four protease inhibitors (B, C, E, and P) were used for the induction phase of the ST6Gal-I expression in minimal media, there was very little or no enzyme activity detected in the media (data not shown). But when protease inhibitor E was deleted from protease inhibitors combination, the expression of ST6Gal-I in mimimal media was resumed. However, the expression levels were relatively low (5469 U/L, Table 8) compared to expression levels in complex media with four inhibitors (6377 U/L, Table 7). Also, the expression levels were slightly higher than fermentation in complex media with two inhibitors (B and E).

3.10 Purification of ST6Gal-I

The soluble form of the catalytic domain of rat ST6Gal-I was harvested from 1L fermentor cultures after 3 days of induction. The harvested media was diluted with either 300 mM NaCl in water or 25 mM HEPES buffer in a 1:2 ratio, followed by adjusting the pH between 7.2 - 7.4. The diluted sample was then applied on to Ni-NTA column, followed by washing with 300 mM NaCl in HEPES buffer pH 7.4 till UV absorption went down to base level. No activity was detected in the run through fractions of the column. The column was then eluted with a gradient of increasing imidazole concentration. The first UV peak started to elute at about 6 mM imidazole and second peak eluted at 90 mM imidazole. The first peak predominantly contained the contaminating media proteins along with about 23% of total loaded enzyme activity. The second peak appeared to be largely homogeneous ST6Gal-I which was then concentrated and purified using gel filtration using a Superdex-75 column. On Superdex-75 the first peak eluted was broad protein peak with 3 apparent unresolved

peaks, however only the second and third unresolved peak contained enzyme activity (Figure 14). SDS-PAGE and Western blot analysis showed that the a later peak of ST6Gal-I activity eluted with a molecular weight of 37 kDa while the earlier eluting peak had some of the 37kDa component as well as addition smear of cross-reactive material up to 75kDa suggesting the hyperglycosylation at the two N-glycosylation sites. Alternatively, addition of 5 mM imidazole in the loading sample on the Ni-NTA column (harvested media) reduced the abundance of the first peak on the ST6Gal-I activity corresponding with the weekly bound enzyme activity. This activity and protein was subsequently found in the column run-through fractions. However, addition of imidazole to the load buffer did not alter the ST6Gal-I activity in the second peak, which eluted at 90 mM imidazole.

3.11 Analysis of N-glycans on recombinant rat ST6Gal-I from Pichia

N-glycans from a recombinant ST6Gal-I were released by treatment with peptide N-glycosidase F (PNGase F). MALDI-TOF MS analysis was performed on the released premehtylated-N-glycans. The most intense isotopic peak cluster seen in the mass spectrum of the glycans from recombinant ST6Gal-I contained m/z 2395.9, which corresponded to $Man_9GlcNAc_2$, and is denoted as Man_9 (Figure 15). There were other high intensity peaks in the MS profile corresponding to $Man_{10}GlcNAc_2$ and $Man_{12}GlcNAc_2$ suggesting presence of the structures in addition to Man_9. It is common for glycoproteins to be expressed with a considerable microheterogeneity in their glycan structures [162]. As expected, relatively low intensity peaks were also seen in mass spectrum, corresponding to other high mannose-type glycans, ranging from Man_5 to Man_{17} as reported previously for other glycoprotiens [141, 163]. Moreover some glycan structures were found to be modified with negatively charged

phosphate groups (Figure 15), unlike previous reports reporting only non-substituted high mannose N-glycan structures ($Man_{10}GlcNAc_2$ to $Man_{16}GlcNAc_2$) [141].

3.12 Deglycosylation of partially purified ST6Gal-I

Endo-H_f and PNGase F were compared for their ability to de-glycosylate partially purified ST6Gal-I under native conditions. Each de-glycosylating enzyme was added in 10-50 fold excess of enzyme required to deglycosylate 500 μg/ml of ST6Gal-I in 1 hr in denaturing conditions. Under denaturing conditions Endo-H_f was able to deglycosylate ST6Gal-I even at 0 hr because it was found to be resistant to subsequent denaturation (Figure 16A). However, at the non-optimal pH of 6.5 under non-denaturing conditions the deglycosylation was found to take 30 – 60 min. Deglycosylation of ST6Gal-I by PNGase F was not as efficient as by Endo-H_f but it was able to deglycosyate completely over a extended time frame (Figure 16). Unfortunately, deglycosylation of ST6Gal-I with with PNGase F resulted in the precipitation of the protein which could not be resolubilized. Therefore, Endo-H_f was used for deglycosylation of native ST6Gal-I in subsequent experiments.

Moreover, a proteolytic degradation of recombinant ST6Gal-I was observed during the deglycosylation reaction at 37°C. Addition of protease a inhibitor cocktail reduced the proteolysis to great extent.

3.13 ST6Gal-I Purification following deglycosylation

Recombinant ST6Gal-I was harvested and purified using IMAC as described above. The peaks eluting from the Ni-NTA column were then processed by diafiltration to allow a

buffer exchange into 5 mM EDTA in 50 mM citrate buffer at pH 5.5 by repeatedly concentrating and diluting using an Amicon ultrafiltration membrane (10kDa molecular cutoff). Finally, the sample was concentrated to 3-4 ml in presence of protease inhibitor cocktail followed by addition of 15 µl of Endo-H_f and incubated it at 37°C for 60-90 min. The deglycosylated ST6Gal-I was further purified using gel filtration on a Superdex-75 column. Unfortunately, the overall recovery of the purified enzyme was considerably lower, (9%) by comparison to sample generated without deglycosylation (46%) (Table 3 and 5). On Superdex-75 the deglycosyated peak resolved as a much sharper peak compared to the glycosylated ST6Gal-I. However on several occasions deglycosylation of ST6Gal-I was not complete (Figure 17). Therefore, the remaining glycosylated ST6Gal-I was removed by passing the Endo-H_f digested sample over a ConA column, which resulted the removal of contaminating glycosylated ST6Gal-I and elution of a homogeneous sample of deglycosylated ST6Gal-I. However, SDS-PAGE and Western blot analysis showed two bands for the deglycosylated protein with molecular weights of about 32kDa and 27kDa (Figure 18).

When NH_2-terminal sequence analysis was performed on both protein bands, the first 32kDa protein band yielded an amino acid sequence of SIHHHHHH which corresponded to an NH_2-terminus of the protein cleaved at the vector encoded α-factor signal sequence. However the second 27 kDa protein band yielded two amino acid sequences KIWRNYL and RNYLNM which correspond to an NH_2-terminus at amino acid position 108 and 111, which indicated that, in addition to the cleavage of the vector-encoded α-factor signal sequence, an additional 24 and 27 NH_2-terminal amino acids were excised by proteolytic cleavage before or during the purification of the protein.

Alternatively, the harvested media was treated with TPCK, a chymotrypsin specific protease inhibitor, and PMSF followed by Ni-affinity chromatography as mentioned earlier and further purification was carried out as mentioned above. The overall yields from conditional media were 34% (1042 U out of 2988 U) (Table 6), a significant improvement in yield (267 U out of 2808 U, Table 5) from a sample in the absence of protease inhibitors.

Purification of recombinant ST6Gal-I from 1L fermentor media with protease inhibitors (C, P, B, and E) during fermentation resulted in two-fold greater enzyme recovery of enzyme activity following purification (2354 U, Table 7) compared to fermentor media without protease inhibitors (1042 U). However, the yield during purification was unchanged (34% and 37%) as there was two fold more enzyme activity in the harvested media from the fermentation with protease inhibitors compared to the fermentation in the absence of protease inhibitors (6377 U, Table 7 and 2988 U, Table 6), respectively.

3.14 Isotope Labeling

The single isotope labeling with ^{15}N was carried out in controlled fermentation conditions as described in the Materials and Methods section. Labeled $(^{15}NH_4)_2SO_4$ was used as the source of ^{15}N in the late growth phase and during the entire induction phase. The growth characteristics of the recombinant *Pichia* strain were unaffected when it was grown in ^{15}N labeled media (data not shown). Also, the overall expression and purification recoveries were comparable to the expression in complex media (Table 8).

For double (^{2}H and ^{15}N) labeling of the ST6Gal-I in Pichia, the culture was slowly adapted to growth in deuteriated media without ^{15}N as described in Materials and Methods. To examine the effect of deuteraton on culture growth and recombinant enzyme expression

the culture was inoculated in 3 different fermentor conditions containing 95%, 97% and 99% deuterium in minimal media while keeping the other conditions constant. The growth characteristics were altered in each of the deuterium-containing media. It took longer for cells to grow to equivalent cell densities in 99% deuterium than in cultures containing 97% and 95% deuterium. However, there was not much difference in the rate of methanol consumption once the culture reached higher cell densities (data not shown). An induction phase of 48 hr was employed and the expression levels of recombinant ST6Gal-I were compared by ST6Gal-I enzyme assays. The expression of ST6Gal-I was much higher in 95% and 97% D_2O containing media (805 U/L and 505 U/L, respectively), while expression in 99% D_2O containing media was considerably reduced (400 U/L, data not shown).

Moreover, the expression level of recombinant ST6Gal-I in the harvested media from 95% D_2O media was less than 20% compared to the unlabeled media (643 U, Table 9 and 5469 U respectively). However, the final recovery of the deuteriated sample was 103% (669 U) compared to just 38% (2086 U) for the unlabeled sample, indicating that the expression level of the enzyme was underestimated as a result of an isotope effect on the enzyme assay. Moreover, NMR analysis indicated that there was over 90% deuterium incorporation in the Cα- carbons of recombinant ST6Gal-I.

Double-labeling (^2H and ^{15}N) of ST6Gal-I was carried out in 95% D_2O containing media and $(^{15}NH_4)_2SO_4$ was used as nitrogen source in late growth phase and entire induction phase as described in Materials and Methods section. Expression and purification levels were similar to those containing deuterium media alone. Harvested media contained a total of 616 U (Table 10) and final recovery after purification was 662 U (107%).

For the triple-labeling of ST6Gal-I (^{2}H, ^{13}C, and ^{15}N) the conditions used were similar to the double-labeling experiment except that the induction phase employed ^{13}C-methanol instead of unlabeled methanol.

3.15 Enzyme kinetics

The pH optimum of the recombinant ST6Gal-I was determined to be 7.0 in a universal buffer (Figure 19). The k_m for donor substrate CMP-NeuAc and acceptor substrate at the optimum pH of 7.0 were determined to be 177 ± 10μM and 2.6 ± 0.2 mM respectively.

A plot of log (k_{cat}/k_m) versus pH was a bell shaped curve with the maximum peak at pH 7.0 for the donor substrate (Figure 20), while a similar plot for the acceptor yielded a flat line between pH 6.0 to 9.0, indicating reduced effect of varied pH for acceptor substrate binding (Figure 21).

3.16 Surface Plasmon Resonance (SPR) Analysis

The binding affinities of CMP-Neu5Ac, a donor substrate, and LacNac, an acceptor substrate were examined by surface plasmon resonance (SPR). The main purpose of these studies was to establish conditions to examine on- and off-rates for binding of the donor and acceptor substrates for recombinant ST6Gal-I and compare these values with enzyme kinetics. ST6Gal-I was immobilized using amine coupling with standard NHS-EDC chemistry (total immobilization 10,000 RU) on CM5 chips. No real binding was observed with any of the donor or acceptor substrates in any running buffers; commercial buffers such as HBS-EP or self-prepared running buffers such as 20mM MES or HEPES with 150mM NaCl, pH 6.5 and 7.4 with or without EDTA or P20 (data not shown). Similarly,

immobilization of ST6Gal-I using amine coupling in presence of donor substrate (CMP-Neu5Ac), acceptor substrate (LacNac), and the inhibitor CMP-3F-Sia, alone or in combination gave us similar results as the enzyme immobilized without additives.

Efforts to immobilize recombinant ST6Gal-I on a Ni^{+2}-NTA chip using the 8xHIS affinity tag were unsuccessful. Even though a high level of immobilization (7000 RU) was achieved for recombinant ST6Gal-I, the enzyme was rapidly washed off the chip surface and a stable baseline was never achieved. Under all the conditions tested the immobilized ST6Gal-I would dissociate at a stable rate from the surface until the original baseline was achieved.

Similarly, efforts to achieve a high level of affinity immobilization of recombinant ST6Gal-I using mannan affinity on CM5 chip, pre-immobilized with ConA protein (10,000 RU), were also unfruitful as only a low level of immobilization was achieved (1,500 RU).

Also, no binding was observed in a reverse experiment in which bi-antennary glycopeptides from asialofetuin containing terminal LacNAc glycan structures were immobilized on the CM5 chip surface and ST6Gal-I was used as analyte.

3.17 Immobilization of ST6Gal-I on Affigel 10

ST6Gal-I was immobilized using the same chemistry as that of SPR immobilization on Affigel 10, pre-activated with NHS-EDC chemistry at 3 different pHs (4.5, 5.5 and 6.5). Following binding, the beads were collected by centrifugation and washed with enzyme buffer. Enzyme assays were performed on the first supernatant following centrifugation as well as the wash fractions and immobilized beads. There was very little immobilization of recombinant ST6Gal-I at pH 4.5 and pH 5.5, as more than 70% enzyme activity was found in

the supernatant and a very little activity assay was detected on the immobilized Affigel beads (Figure 22). Also, the sum of the supernatant, wash, and immobilized activity did not add up to 100% at both pHs, with or without donor substrate addition. At pH 6.5, which is more suitable for amine coupling using NHS-EDC chemistry, there was not much enzyme activity in the supernatant, wash, or on the immobilized gel beads. The activity on the immobilized gel beads in presence of CMP-NeuAc was slightly better than in the absence of CMP-Neu5Ac (Figure 22), indicating that the immobilization of ST6Gal-I using amine coupling even in the presence of donor substrate results in loss of enzyme activity.

3.18 Chemical modification of ST6Gal-I

To investigate the effect of immobilization chemistry on enzyme activity we decided to biotinylate the enzyme using the same NHS-EDC chemistry in the liquid state using pre-activated biotin reagent. As expected from the immobilization results, biotinylation of ST6Gal-I resulted in complete loss of enzyme activity (data not shown). Similar results were seen with the MBP-ST6Gal-I fusion protein (data not shown). However, when the MBP-ST6Gal-I fusion protein was methylated using reductive methylation chemistry there was a loss of about 60% enzyme activity. However, TEV protease digestion of the fusion protein resulted in 50% increase in enzyme activity, yielding about 60% of original enzyme activity. But once again biotinylation of the methylated sample resulted in inactive enzyme (Figure 23).

3.19 Comparison of S6GalI expression constructs

The expression and screening of the other constructs (HA-ST6Gal-I/pPICZαC in SMD1168H and HA-ST6Gal-I_S/pPICZαC, HA-STRP-ST6Gal-I_S/pPICZαC, MBP-ST6Gal-I_S/pPICZαC, HA-STRP-ST6Gal-I_L/pPICZαC and MBP-ST6Gal-I_L/pPICZαC in KM71H) was performed similar to that of the ST6Gal-I/pPICZαC construct in KM71H strain.

During the screening of the HA-ST6Gal-I/pPICZαC construct in SMD1168H or HA-ST6Gal-I_S/pPICZαC construct in KM71H there was very little enzyme activity detected (310 U/L and 270 U/L respectively, compared to 1187 U/L for the ST6Gal-I/pPICZαC construct in KM71H (Figure 24)). However, from Western blot and SDS-PAGE analysis the expression levels of ST6Gal-I were similar for all three constructs.

In contrast, the HA-STRP-ST6Gal-I_S/pPICZαC, a streptavidin fusion with the truncated ST6Gal-I, construct gave absolutely no enzyme activity during the screening of the transformants. Despite the lack of enzyme activity in the streptavidin fusion construct Western blots of conditioned media showed a small amount of protein band at 32kDa corresponding to clipped ST6Gal-I (Figure 24).

In contrast to the HA-STRP-ST6Gal-I_S/pPICZαC construct, HA-STRP-ST6Gal-I_L/pPICZαC, a streptavidin fusion with the longer ST6Gal-I, construct exhibited a trace amount of ST6Gal-I activity (Figure 24). However, Western blot analysis of both the streptavidin constructs showed similar results (only a trace amount of clipped ST6Gal-I equivalent band and no Strevptavidin-ST6Gal-I fusion protein band (Figure 24)).

The MBP-ST6Gal-I fusion protein construct in MBP-ST6Gal-I_S/pPICZαC in KM71H yielded a twice as much activity and twice as much protein compared to HA-ST6Gal-I/pPICZαC and HA-ST6Gal-I_S/pPICZαC constructs. Western blot analysis showed a minor

band at 37kDa equivalent to a clipped ST6Gal-I, in addition to a major band of expected MBP-ST6Gal-I fusion protein at 80 kDa. However, the enzyme activity of purified MBP-ST6Gal-I_S, a MBP fusion with truncated ST6Gal-I was significantly less (≈10%) than enzyme activity of the equivalent longer ST6Gal-I, when same amount of protein was assayed (data not shown).

In contrast, MBP-ST6Gal-I_L/pPICZαC, an MBP fusion with the longer form of the ST6Gal-I catalytic domain, showed two-fold more enzyme activity and two-fold more protein when compared to ST6Gal-I/pPICZαC, the longer His-tagged ST6Gal-I construct. Also, Western blots showed a greater degradation of the fusion protein for MBP-ST6Gal-I_L/pPICZαC, than MBP-ST6Gal-I_S/pPICZαC, MBP fusion with truncated ST6Gal-I.

3.20 Expression and Purification of MBP-ST6Gal-I with TEV protease digests

In 5ml tube culture expression experiments, the MBP-ST6Gal-I_L/pPICZαC construct expressed at least two-folds higher ST6Gal-I activity (2960 U/L) compared to the original ST6Gal-I/pPICZαC construct (1187 U/L (Figure 24)). However, when MBP-ST6Gal-I_L/pPICZαC was expressed in a 1L fermentor using the same conditions as in controlled fermentation for using the original ST6Gal-I/pPICZαC construct, the expression levels were much lower. Even with all the protease inhibitors, the maximum ST6Gal-I activity from the harvested media of the fusion protein construct was between 1038 and 1483 U/L (data not shown), much lower than the equivalent ST6Gal-I/pPICZαC construct (6377 U/L).

Therefore isotope labeling of the MBP-ST6Gal-I_L fusion protein was carried out in multiple 5 ml tube cultures rather than controlled fermentation. Only double labeled (^2H, ^{15}N) and triple labeled (^2H, ^{13}C, ^{15}N) MBP-ST6Gal-I protein was expressed. The expression of the

double labeled fusion protein was carried out in 95% D_2O, and $(^{15}NH_4)_2SO_4$ was used as nitrogen source in the entire growth phase and the entire induction phase as described in Chapter 2. The expression levels of the doubled labeled MBP-ST6Gal-I fusion proteins from the tube cultures (565 U per 1.4 L, Table 11) were less than the expression level for the original ST6Gal-I in controlled fermentation (616 U per 0.9 L, Table 10). The final recovery the double labeled deuteriated sample was 105% (596 U), thus indicating that the expression level of the enzyme was underestimated as result of an isotope effect on the enzyme assay, similar to the original ST6Gal-I construct.

For the triple labeling of MBP-ST6Gal-I (^2H, ^{13}C and ^{15}N) the conditions used were similar to the double experiment except ^{13}C-methanol was used in the induction phase instead of unlabeled methanol. However, very little enzyme activity was detected from triple labeled conditioned media which was lost during the purification.

3.21 NMR Analysis

The efficiency of isotope incorporation in recombinant ST6Gal-I was estimated by ^{15}N-coupled and -decoupled one-dimensional ^1H spectra (Figure 25). The effect of decoupling was clearly visible when the signal from the down-field (11.4 ppm) Trp side chain was examined: the side chain signal was a singlet in the decoupled spectrum whereas it was a triplet in the coupled spectrum, indicating a mixture of ^{15}N and ^{14}N containing NH pairs. The estimates of the area of the ^{14}NH singlet in the coupled spectrum were compared with the area of the singlet in the decoupled spectrum indicating a loss of one third of the total signal. Therefore, the total ^{15}N incorporation efficiency is believed to be in the range of 70-90%.

The 2D ^{1}H-^{15}N Heteronuclear Single Quantum Spectroscopy (HSQC) spectrum of 100μM ^{15}N-labeled ST6Gal-I showed only a low degree of dispersion of peaks (Figure 26). A large number of peaks were aggregated between 7.5 and 8.5 ppm, which indicates that the protein is either unstructured or contains a large number of helices.

In contrast to the 2D ^{1}H-^{15}N HSQC spectrum of untreated recombinant ^{15}N-labeled ST6Gal-I, the HSQC spectrum of endo-H_f treated, deglycosylated, ^{15}N-labeled ST6Gal-I demonstrated a much greater dispersion of the peaks. The overlapping peak region seen in untreated samples between between 7.5 and 8.5 ppm largely disappeared after deglycosylation (Figure 26). The HSQC spectrum of the deglycosylated enzyme showed roughly 180 major peaks with uniform peak intensities and very good dispersion. In addition there were about 100 minor peaks in the spectrum. Moreover, the peak width did not change significantly before and after deglycosylation.

When the HSQC spectra of single labeled (^{15}N) deglycosylated ST6Gal-I was compared with the HSQC spectra of the double labeled (2H and ^{15}N) ST6Gal-I, even greater dispersion of peaks was detected. Moreover, peak widths in double labeled samples were much narrower than in single labeled; as a result many more peaks were resolved. There were as many as 240 well resolved peaks in the double labeled sample (Figure 27). Peak intensities of the deuterated samples were also much higher than the equivalent protonated sample as the contour level of the HSQC spectrum of the protonated sample could only be lowered to just above the noise threshold. In constrast, contour level and background adjustment of the HSQC spectrum for the deuterated sample could be displayed in a form that would significantly reduce the contribution from background noise.

In addition, binding studies employed the substrate analog CMP-3F-Sia, with the double labeled sample demonstrating a sharp decrease in the intensities of few a peaks and increases in intensities of others, indicating that the in the absence of substrate the enzyme exhibits conformational flexibility and that binding of the substrate induces a conformation change (Figure 29).

The 3D ^{2}H-^{13}C-^{15}N, HNCA experiment was carried out on 60 μM triple labeled (^{2}H, ^{13}C, ^{15}N) ST6Gal-I. A number of three dimensional NMR experiments were collected on both the Varian Inova 900 MHz spectrometer as well as the Varian Inova 600 MHz spectrometer equipped with a high sensitivity 3 mm cold probe. The HNCA data shown in the Figure 30 was collected on the 900 MHz spectrometer. A total of 52 complex points were collected in the first indirect dimension and 30 complex points were collected in the second indirect dimension. 1024 points were collected in the direct dimension. 64 transients were accumulated for each FID.

In the HNCA spectrum (Figure 30) shown is the projection of all planes along the ^{15}N axis after Fourier transformation of all three dimensions. Each peak in the spectrum represents either a Cα-HN pair from within a residue or between Cα of a residue and HN of the next residue. Therefore, it is theoretically possible to see twice the number of peaks in the HNCA projection as the number of residues excluding proline. However, in the HNCA spectrum about 200 peaks were seen, a majority of which correspond to Cα-HN within the same residue. A quick estimation of the completeness of the data can also be made: the Cα chemical shifts of glycines possesses unique chemical shifts around 44 ppm. Since there are 16 glycines in ST6Gal-I, 16 peaks are expected to be seen in that region. However, only 6 are visible in the spectrum.

A 2D TROCY spectrum of the double labeled TEV-protease digested and deglycosylated ST6Gal-I from MBP-ST6Gal-I fusion protein was overlayed on the 2D HSQC spectrum of deglycosylated double labeled ST6Gal-I. In overlay, almost all the peaks overlap exactly (Figure 31) indicating that both proteins have the same overall protein structure. The additional peaks in 2D TROCY around 6.75ppm and 7.75ppm in proton and 110ppm in nitrogen are because of the difference in the experimental conditions in which side chain signals were not suppressed and are expected to be the side chain NH_2 groups of Asn and Gln.

3.22 Expression Purification of rat ST6Gal-I from glyco-engineered Pichia strain YGly30

Transformants of the YGly30 strain (Glycofi Inc.) were generated as in above by homologous recombination of ST6Gal-I/pPICZαC construct in the *Pichia* genome at the 5' region of the alcohol oxidase gene (AOX1). Transformants were selected by growth on antibiotic based on the vector encoded zeocin-resistance gene. Expression was induced by growth on methanol as carbon source.

When transformants were expressed using a 1 L controlled fermentor, the expression pattern of recombinant ST6Gal-I using the YGly30 strain was similar to that of KM71H strain. Induction of the culture at 30°C yielded a considerably reduced expression level compared 5 ml tube cultures. However, there was a sharp increase in expressed enzyme activity when induced at 20°C. Induction of the YGly30 strain at 20°C for 60 hr produced 2950 U/L.

When ST6Gal-I was purified using an Ni-NTA column and gel filtration, the overall yields were similar to that of KM71H (data not shown). However, there was a difference in the stability of the purified ST6Gal-I from the YGly30 cultures. The ST6Gal-I protein from YGly30 strain lost all of its enzyme activity within 3-4 days compared a greater stability to over 7 days for the KM71H strain (data not shown). The purification profile of ST6Gal-I from the Ygly30 culture is shown below (Figure 32).

3.23 Expression rat ST6Gal-I using engineered Pichia strain Auxotrophic for aromatic amino acid residues

Transformants of engineered *Pichia* strains ARO1, ARO7, and TYR1 were generated as mentioned in Whittaker *et al.*, 2005 [149], by homologous recombination of the ST6Gal-I/pPICZαC construct into the *Pichia* genome at the 5' region of the alcohol oxidase gene (AOX1). Transformants were selected by growth on antibiotic based on the vector encoded zeocin-resistance gene as mentioned in the Chapter 2.

The screening of transformants for recombinant enzyme activity was carried out similar to the methods described for KM71H strain except that the media composition was modified as described in Chapter 2 [149]. The highest expressing colonies in all the three strains were equivalent low levels of expression of recombinant ST6Gal-I (≈210 U/L) compared to the KM71H strain (1187 U/L). Efforts to scale-up the expression of ST6Gal-I in fermentor cultures was unfruitful as no enzyme activity was detected in the conditioned media.

3.24 Crystallization trials

Deglycosylated recombinant ST6Gal-I was purified and concentrated to 10 mg/ml in 25mM MES buffer (pH 6.5) and the substrate analog, CMP-3FNeuAc was added to a final concentration of 1.5 mM. Several standard screens were used to identify lead conditions for crystallization including, Personal structure I, II, MemSys (Molecular Dimensions, UK), Wizard I, II (Emerald Biostructures, USA), Crystal Screen, Crystal Screen 2, Crystal Screen Cryo, PEG/Ion, PEG4000, 6000 and 8000, Low ionic strength PEG3350, MembFac, (NH4)2SO4 (Hampton Research, USA and Jena, Germany), as well as custom screening conditions.

Few promising conditions were found in the crystal screening trials. Several conditions showed a brown precipitation, and a few conditions showed phase separation, but most conditions were clear of precipitation or micro-crystals. One condition from the screen is illustrated in Figure 33 where the left side of the panel was dyed with Izit Crystal Dye, and the blue color of the aggregation indicates that they are protein precipitates and not salt. However, no conditions were identified leading to a crystal formation (Figure 33).

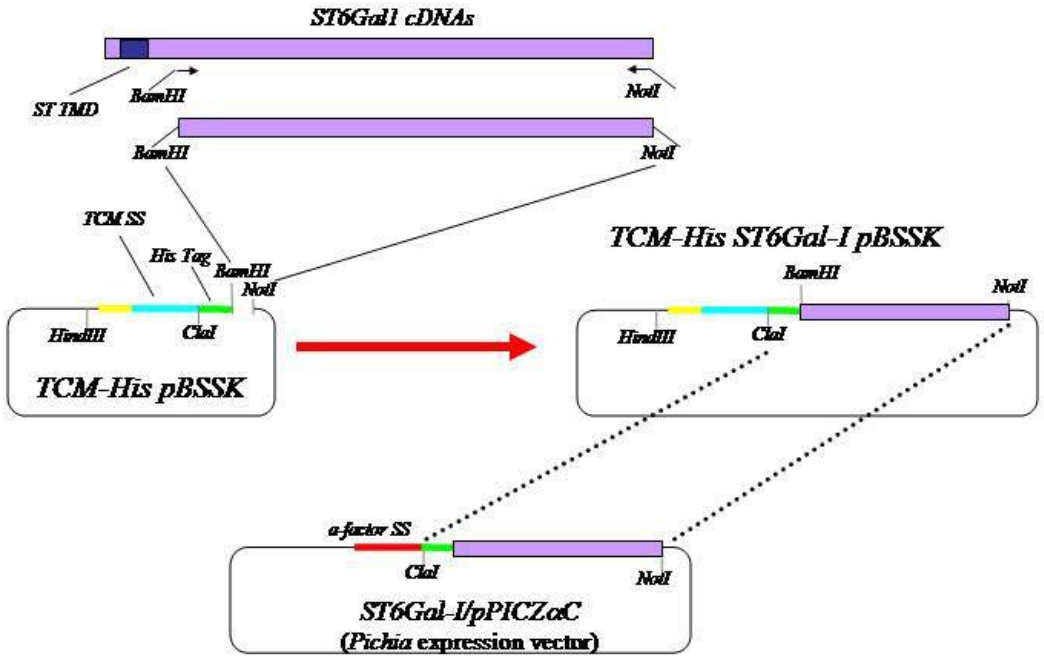


Figure 7: Overview of the steps involved in the cloning of rat ST6Gal-I catalytic domain in the *Pichia* expression vector pPICZαC. cDNA of rat ST6Gal-I was mutagenized to incoportate the amino acid residues from 97-403. It was then sub-cloned into TCM-His-PBSSK vector to aquire an N-terminal His tag which was then subsequently sub-cloned into pPICZαC

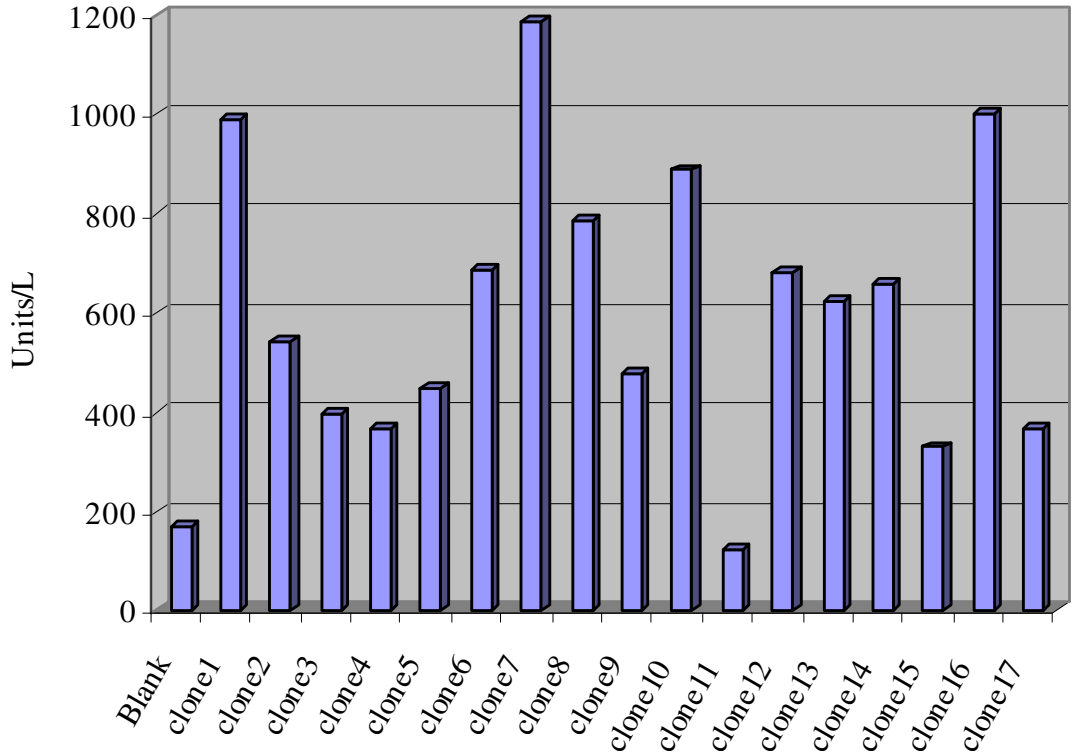


Figure 8: Screening of ST6Gal-I/pPICZαC transformants: Each of the zeocin-resistant clones were inoculated in BMGY media and incubated at 30°C for 24 hrs followed by induction in BMMY media for 3 days. The cells were centrifuged and the supernatant media was assayed for the secreted ST6Gal-I enzyme activity

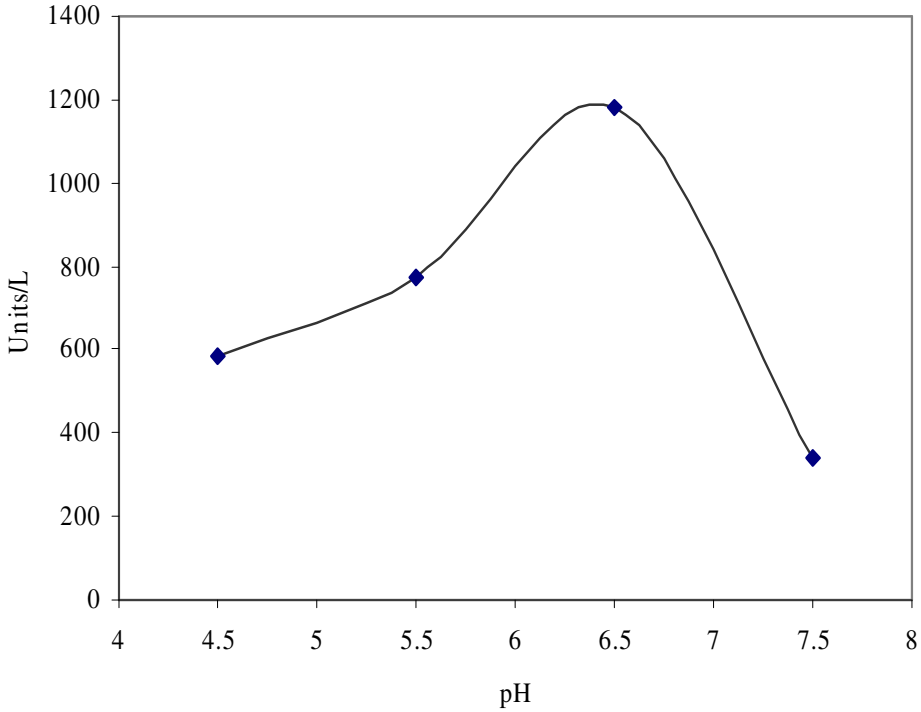


Figure 9: Effect of pH on expression level of ST6Gal-I. A single zecocin-resistant clone was inoculated in BMGY incubated at 30°C for 24 hrs. The cells were equally divided into 4 tubes and centrifuged and resuspended in BMMY media at various pHs. The cultures were induced for 3 days by addition of 0.5% methanol every 12 hours. The cells were centrifuged and the supernatant media was assayed at pH 6.5 for secreted ST6Gal-I enzyme activity.

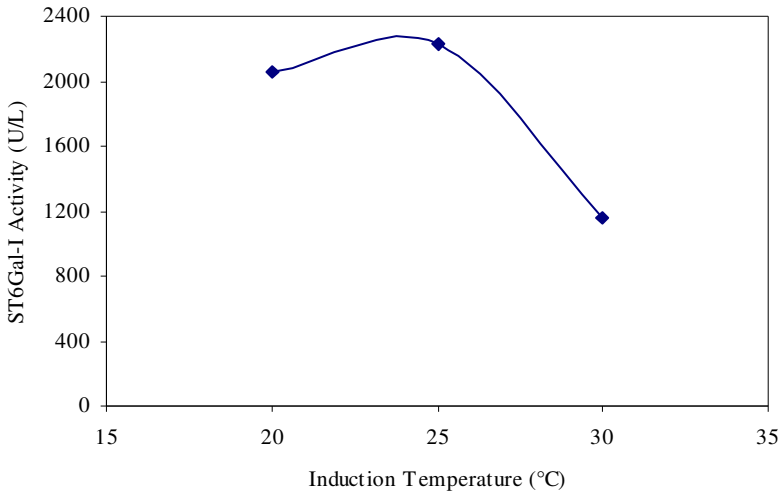


Figure 10: Effect of temperature on the expression level of ST6Gal-I. A single zecocin-resistant clone was inoculated in BMGY incubated at 30°C for 24 h. The cells were equally divided into 3 tubes and centrifuged and resuspended in BMMY media at pH 6.0 and incubated at various temperatures. The cultures were induced for 3 days by addition of 0.5% methanol every 12 h. The cells were centrifuged and the supernatant media was assayed for the secreted ST6Gal-I enzyme activity.

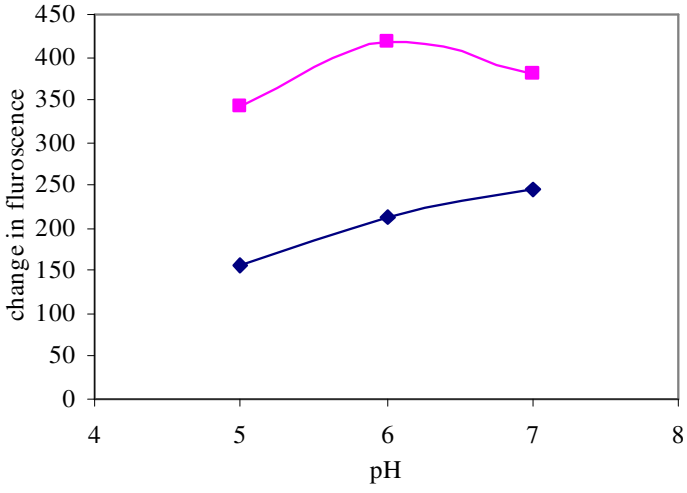


Figure 11: pH profile of the proteases in pichia fermentation. Samples were drawn at 5 hr and 48 h of induction during the recombinant ST6Gal-I expression in 1 L fermentor. Protease assays were performed at various pHs using FTC-Casein

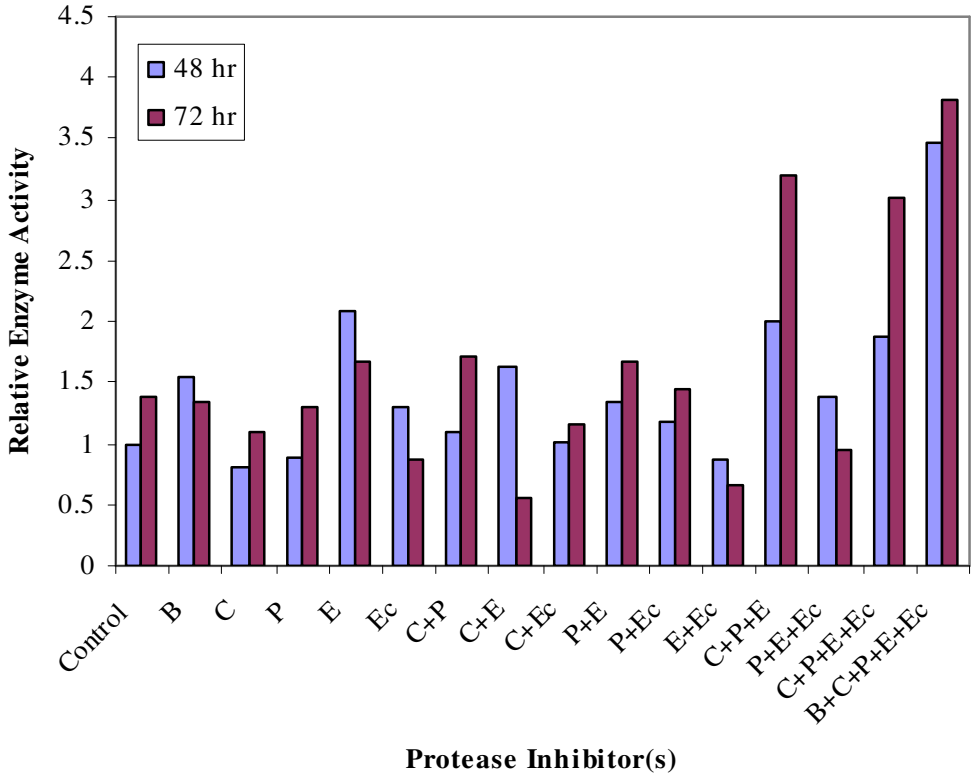
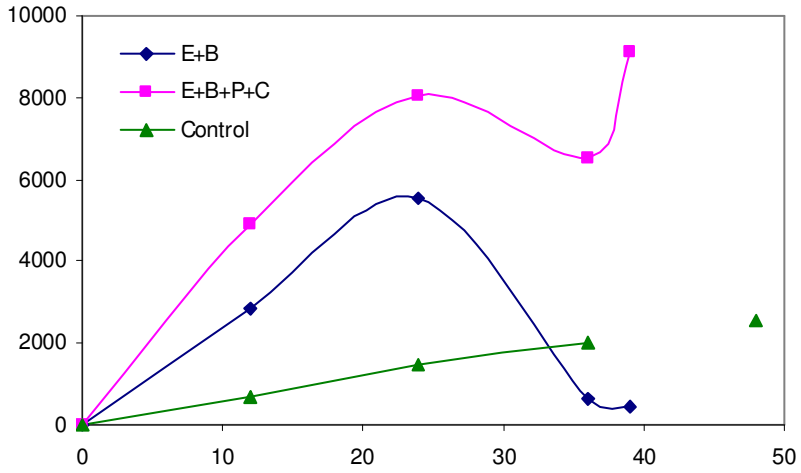


Figure 12: Screening of protease inhibitor for the production of ST6Gal-I. The effect of protease inhibitors was studied similar conditions of growth and induction as above. Additionally protease inhibitors were added to investigate their effect. BSA (B, 10 g/L) and EDTA (E, 10 mM) were added one time at the beginning of the induction phase. Addition of pepstatatin A (P, 10 mg/L), chymostatin (C, 10 μM) and ε-amino caprioic acid (Ec, 100 mg/L) were repeated

A)



B)

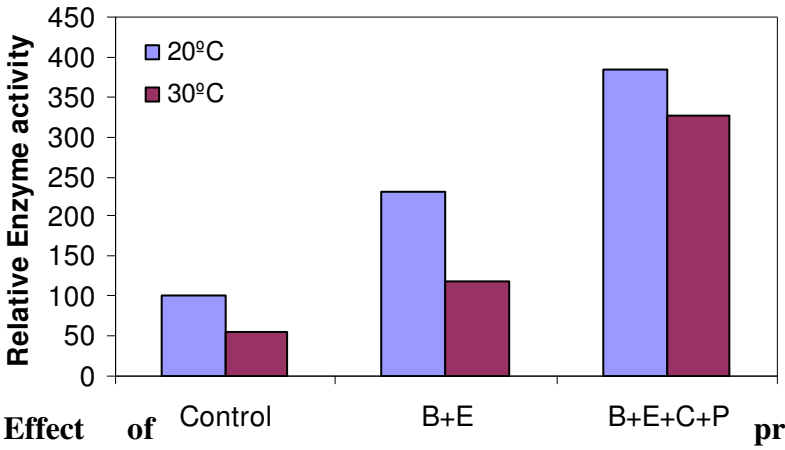


Figure 13: Effect of protease inhibitors on the expression of ST6Gal-I. A) The effect of protease inhibitors was studied using controlled fermentation in complex media as described in materials and methods, 10 g/L BSA (B), and 10mM EDTA (E), were injected at the beginning of the induction phase and 10 mg/L Pepstatin A (P), and 10µM Chymostatin (C) were injected every 12 h during the induction phase. B) The effect of protease inhibitor on the expression of ST6Gal-I in 5 ml tube culture was was studied similar to described in Figure 12 at 20°C and 30°C.

A)

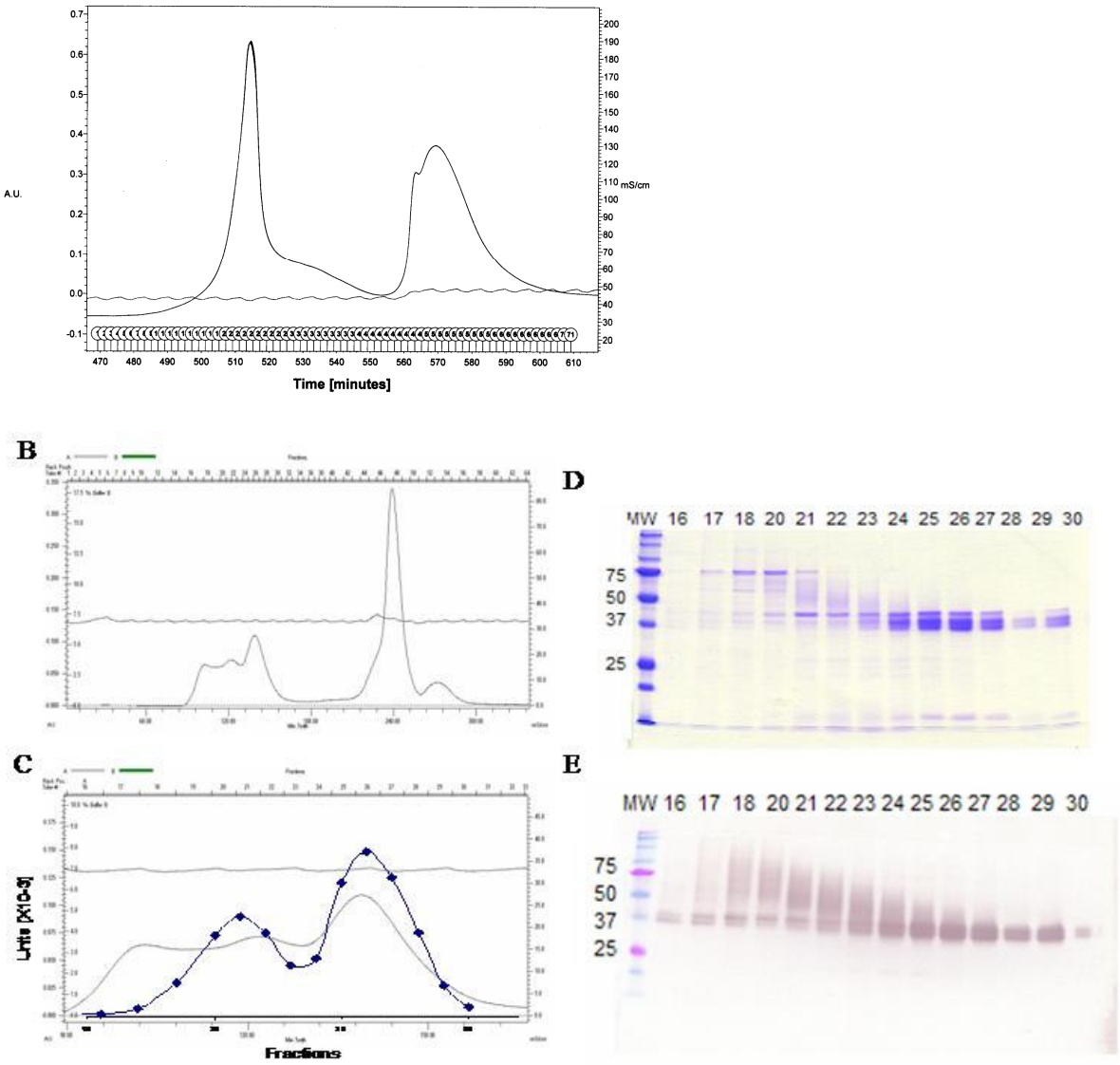


Figure 14: Purification profile of ST6Gal-I. A) Chromatogram from Ni-NTA purification step, where absorbance at 280nm was plotted versus elution time. Fraction collected from the column elution are indicated at the bottom of the plot conductivity is also indicated on the right axis in the eluted solution, B) Chromatogram of gel filtration step, the second peak from Ni-NTA column was loaded Superdex G-75 column, where absorbance at 280nm was plotted versus elution time similar to Ni-NTA purification step, C) Zoom-in on the peak fraction from gel filtration and enzyme activity overlayed, an aliquot was taken from each fraction and assayed for ST6Gal-I enzyme activity, D) SDS-PAGE of fractions derived from the gel filtration profile shown in panel C, and E) Western-blot of fractions derived from the gel filtration profile in panel C.

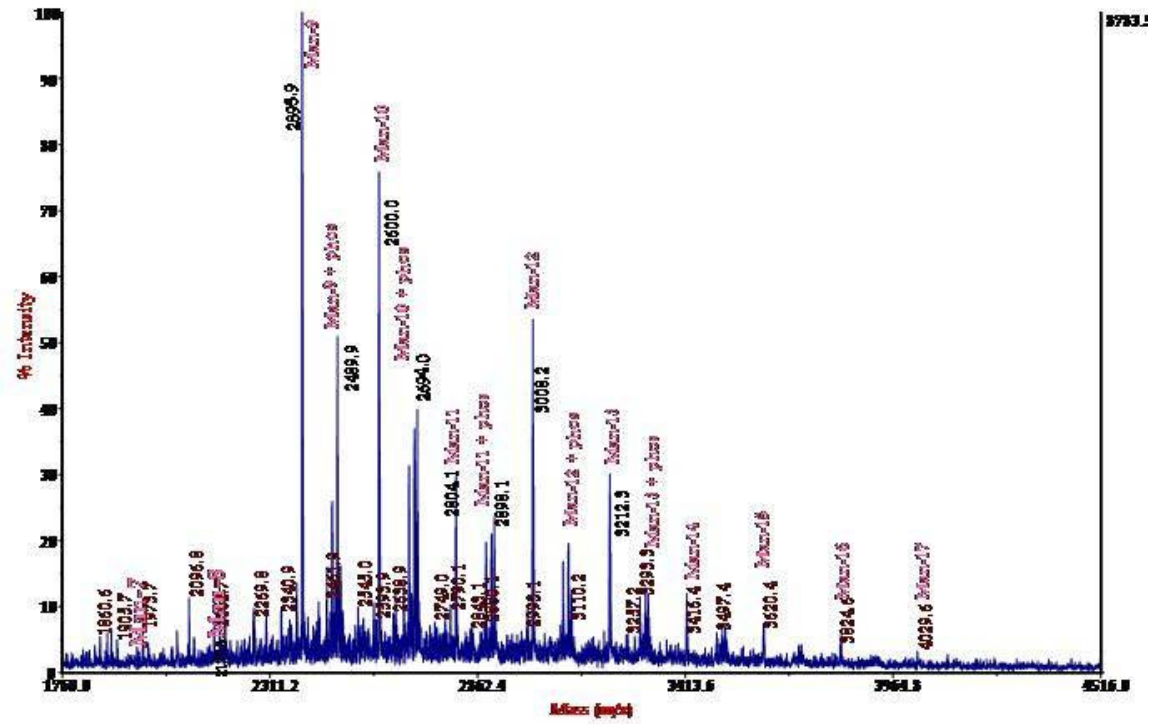


Figure 15: Analysis of N-glycan on recombinant rat ST6Gal-I from *Pichia*. Glycans from recombinant ST6GalI were released using PNGase F and analyzed using a MALDI-TOF mass spectrophotometer as described in chapter 2. The sample showed high mannose structures from $Man_{17}GlcNAc_2$ structrures (Man17) to $Man_7GlcNAc_2$ (Man7), however, the most abundant structures were from Man9-Man12 as indicated by their relative peak intensities. Some of these high mannose structures were found to be substituted with negatively charged phosphate groups.

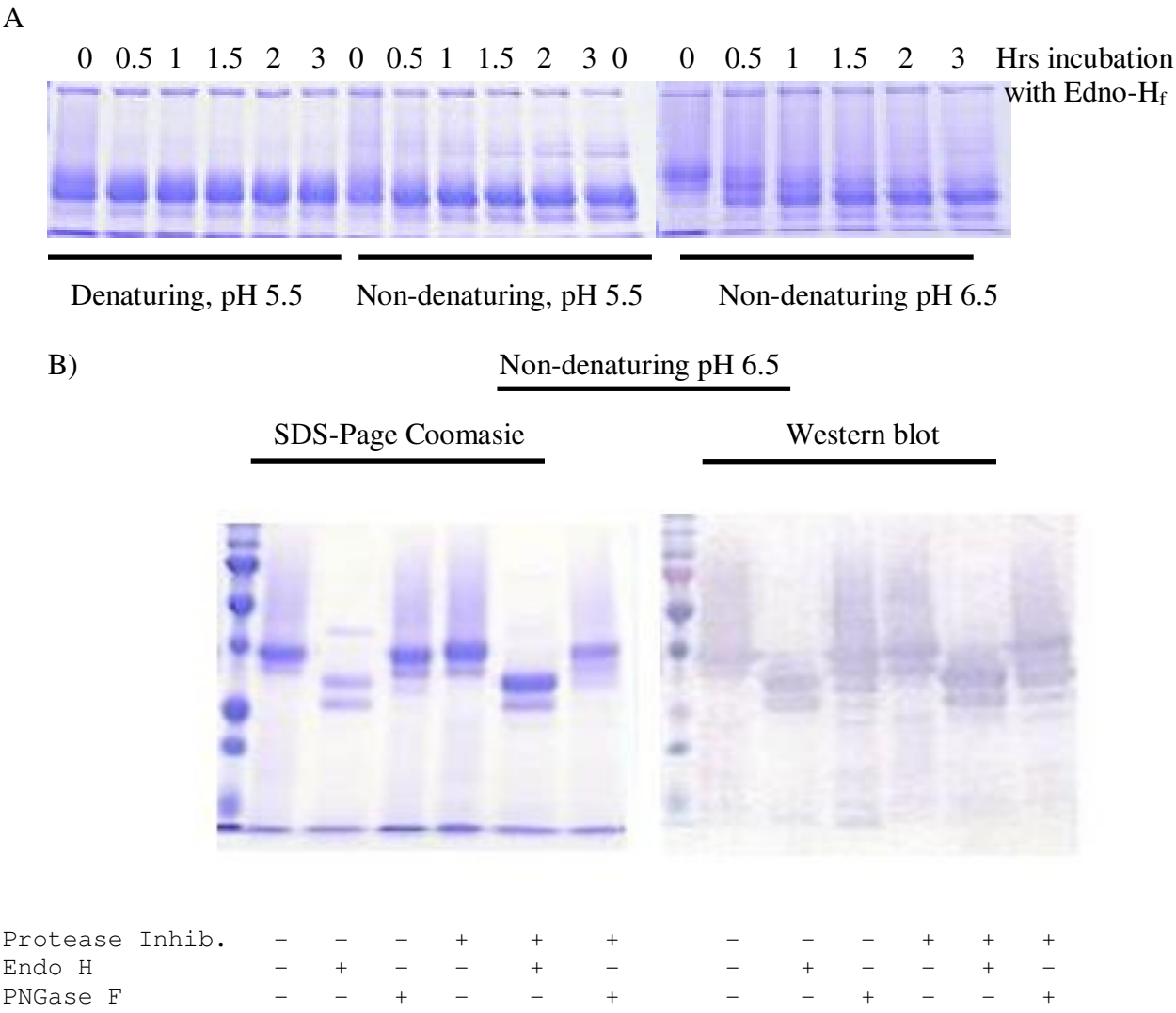


Figure 16: Optimization of deglycosylation conditions. (A) SDS-PAGE showing time profile of deglycosylation of of partially-purified ST6Gal-I using Enod-H_f under denaturing and non-denaturing conditions (The recombinant ST6Ga-I sample, after gel-filtration, was concentrated to 200 µg/ml. Three aliquots (150 µl each) were derived from the concentrated sample. To the first aliquot, 15 µl of 500 mM citrate buffer, pH 5.5, and 5X SDS-PAGE sample buffer was added and sample was denatured by heating at 95°C for 5 min. To the second aliquot, 15 µl of 500 mM citrate buffer pH 5.5 was added. To the third aliquot, 15 µl of 200 mM MES buffer pH 6.5 was added. To each of these samples 10 µl of endo H_f was added and all the three samples were incubated at 37°C. Small 25 µl aliquots were drawn from each of the samples at 0, 0.5, 1, 1.5, 2, and 3 hr intervals followed by immediate addition of 5X SDS-PAGE buffer and denaturation by heating at 95°C for 5 min. (B) A comparision of deglycosylation of ST6Gal-I in native state by Endo-H_f or PNGase F. with or without the addition of a protease inhibitor cocktail. (The samples were prepared as described for panel A).

A) **B)**

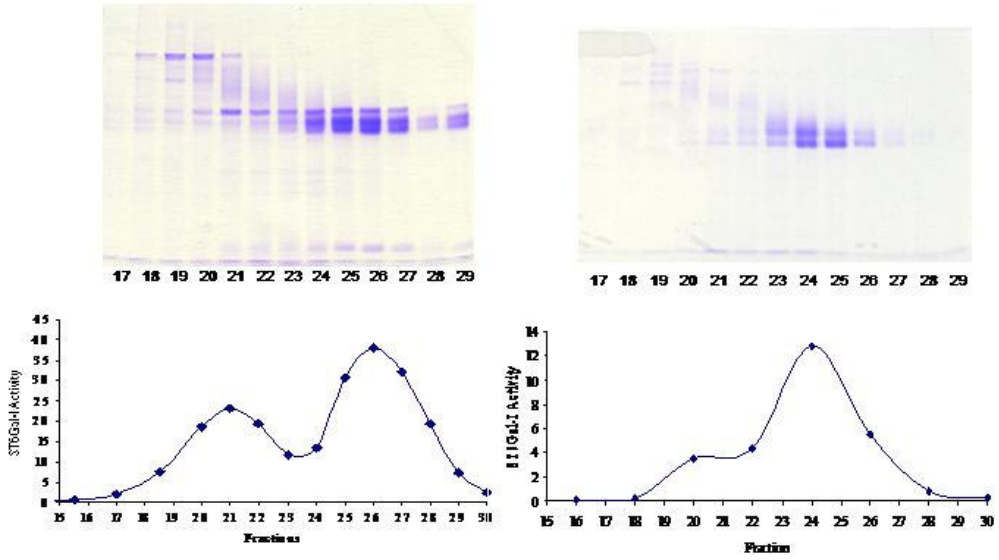


Figure 17: Comparison of the gel filtration profiles of glycosylated and deglycosylated recombinant ST6Gal-I. A) SDS-PAGE (top panel) and enzyme activity (bottom panel) profile of gel filtration fractions of glycosylated ST6Gal-I. (The second peak from Ni-NTA column (Figure 14) was loaded on Superdex G-75 gel-filtration column and 3 ml fractions were collected over the entire run. An aliquot was taken from each fraction and assayed for ST6Gal-I enzyme activity), B) SDS-PAGE (top panel) and enzyme activity (bottom panel) profile of gel filtration fractions of deglycosylated ST6Gal-I. The second peak from Ni-NTA column (Figure 14) was dialysed against 50 mM sodium citrate buffer pH 5.5 and concentrated to 3-5ml using ultrafiltration device (10kDa cutoff) followed by deglycosylation for 1 hr at 37 °C using endo H_f (15 µl). The sample was then loaded on Superdex G-75 gel filtration column and 3 ml fractions were collected over entire run, an aliquot was taken from each fraction and assayed for ST6Gal-I enzyme activity.

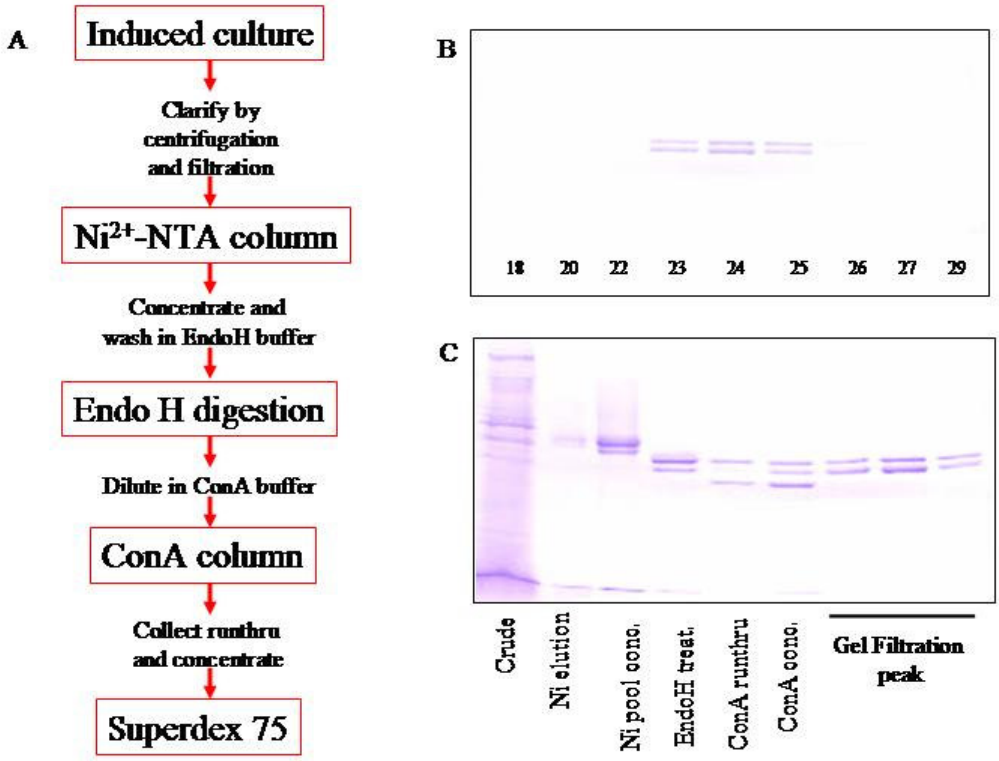


Figure 18: Purification of ST6Gal-I incorporating deglycosylation with endo H_f. A) Purification scheme with deglycosylation of recombinant ST6Gal-I, followed by chromatography over ConA and Superdex 75, B) SDS-PAGE of gel filtration profile of the deglycosylated sample (the second peak from Ni-NTA column was dialysed against 50 mM sodium citrate buffer pH 5.5 and concentrated to 3-5 ml using an ultrafiltration device (10kDa cutoff) followed by deglycosylation for 1 hr at 37 °C using endo H_f (15 µl). The sample was then loaded on ConA sepharose. A runthrough sample from ConA was then then loaded on Superdex G-75 gel filtration column and 3 ml fractions were collected over entire run. C) SDS-PAGE of purification profile of deglycosylated ST6Gal-I. The aliquots were taken at each step of purification and used for SDS-PAGE as shown in panel C.

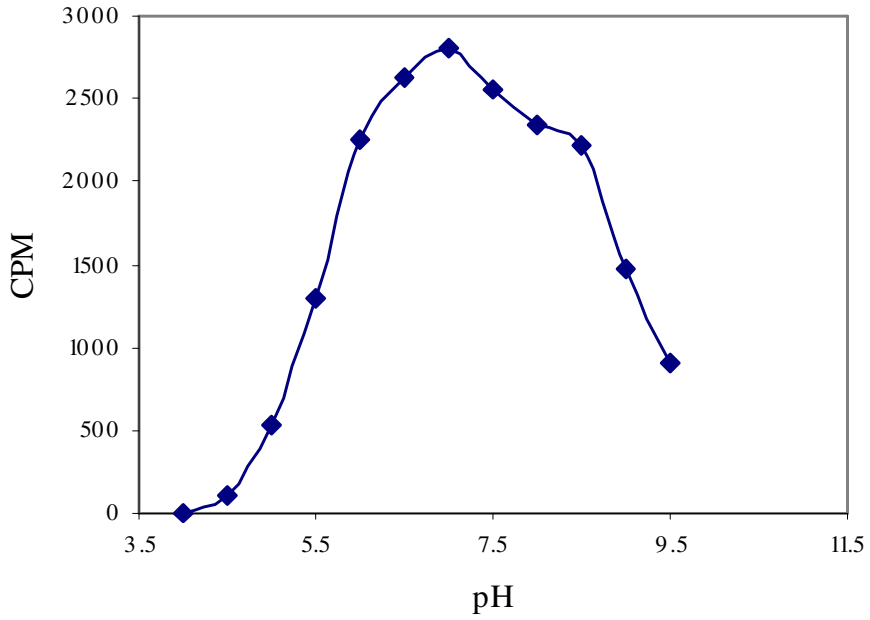
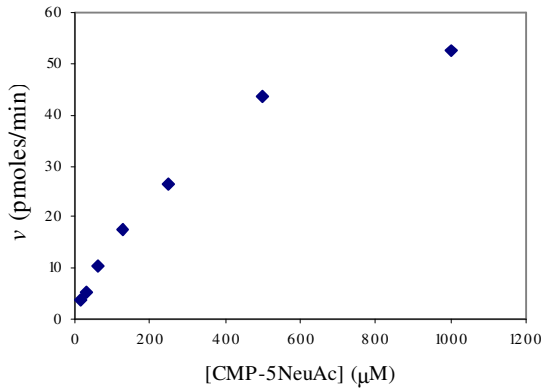
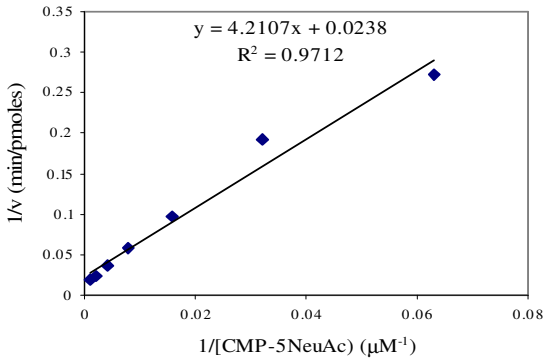


Figure 19: pH profile of recombinant ST6Gal-I. Enzyme assays were performed at 37 °C for 1 h, in a universal buffer at pH intervals of 0.5 as described in Chapter 2. In all assays a constant amount of donor substrates, CMP-Neu5Ac (400 µM) and [^{14}C]-CMP-Neu5Ac (12,500 CPM/assay) and acceptor substrate, LacNAc (3 mM) was used.

A)



B)



C)

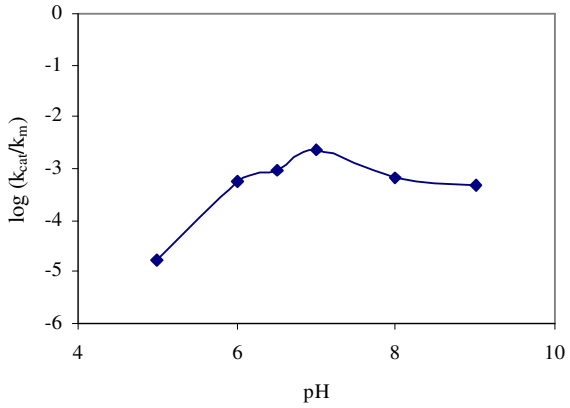
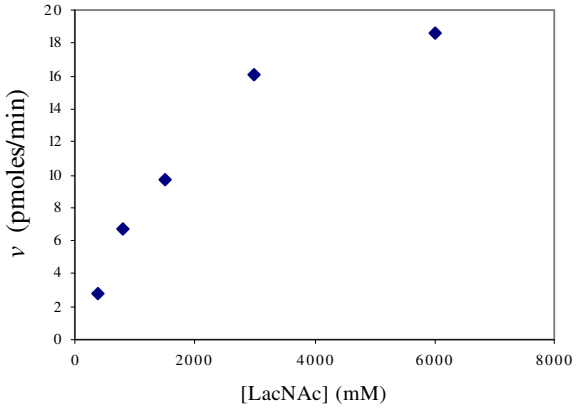
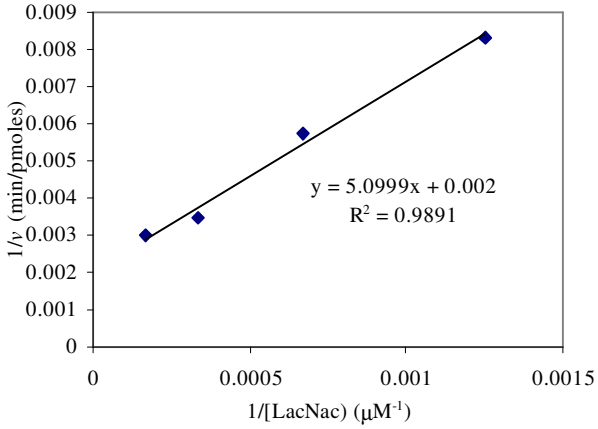


Figure 20: Kinetic parameters for CMP-Neu5Ac donor substrate. Acceptor substrate LacNAc was kept constant at 6 mM and total [^{14}C]-CMP-Neu5Ac was maintained at 30,000 cpm/per assay. A) Sample of saturation kinetics of donor substrate, CMP-Neu5Ac, in universal buffer in an optimal pH of 7.0 at 37 °C for 15 min, B) Lineweaver-Burk plot at the optimal pH of 7.0, and C) plot of log (k_{cat}/k_m) vs pH plot (the V_{max} was used instead of k_{cat} and the same enzyme quantity was used for all assay conditions).

A)



B)



C)

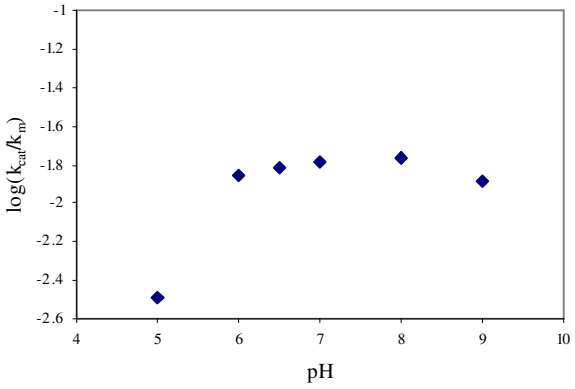


Figure 21: Kinetic parameters for LacNAc acceptor substrate. Donor substrate CMP-Neu5Ac was kept constant at 1 mM and total [^{14}C]-CMP-Neu5Ac was maintained at 30,000 cpm/per assay. A) Sample of saturation kinetics of acceptor substrate LacNAc in universal buffer in the optimal pH of 7.0 at 37 °C for 15 min., B) Lineweaver-Burk plot at the optimal pH of 7.0, and C) plot of log (k_{cat}/k_m) vs pH plot (the V_{max} was used instead of k_{cat} and the same enzyme quantity was used for all assay conditions).

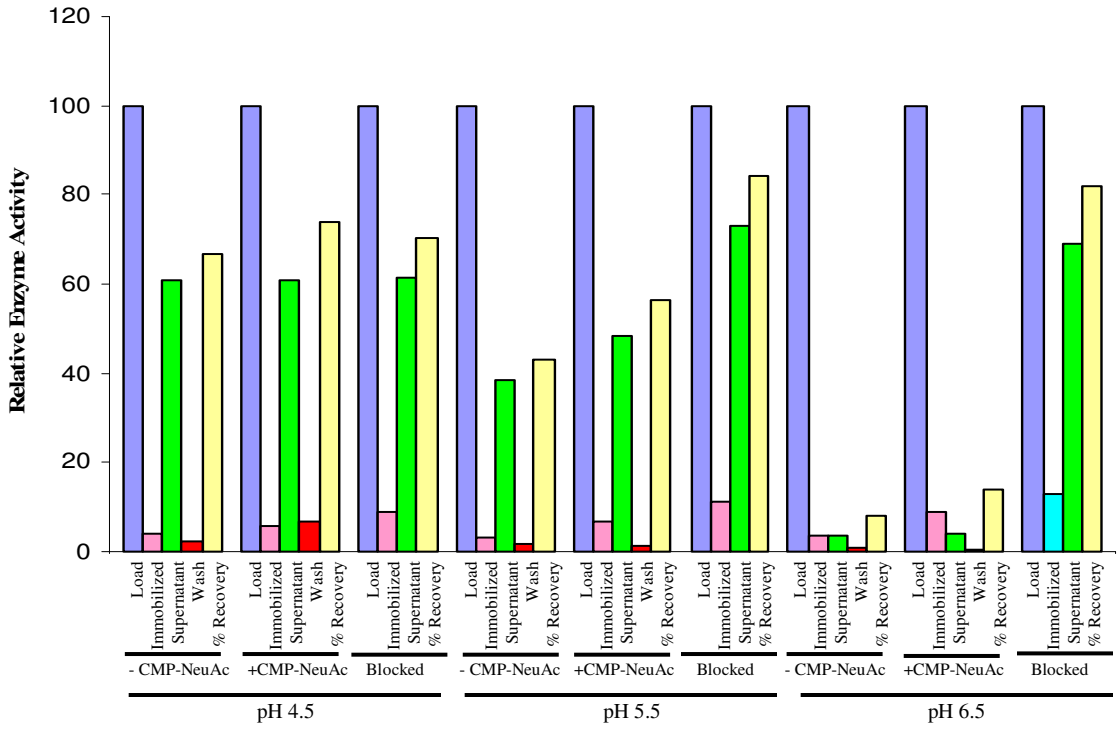


Figure 22: Immobilization of recombinant ST6Gal-I on Affigel-10. Recombinant ST6Gal-I (50 µg) was diluted in 200 µl of immobilization buffer at pH 4.5, 5.5 and 6.5 with or without donor substrate, CMP-Neu5Ac (500 µM) and remaining active groups were blocked by 50 mM ethanolamine. The beads were collected by centrifugation or washed with respective immobilization buffers. The orginal supernantant from the bead immobilization, the immobilization buffuer washes, and immobilization beads were then assayed for ST6Gal-I activity. The load activity in each fraction and the % recovery based on the sum of the activity in each fraction are plotted. Control experiments were run where the resin beads were pre-blocked with ethanolamine as indicated with "blocked". In addition, the effect of treatment with donor substrate is inidicated with ± CMP-Neu5Ac

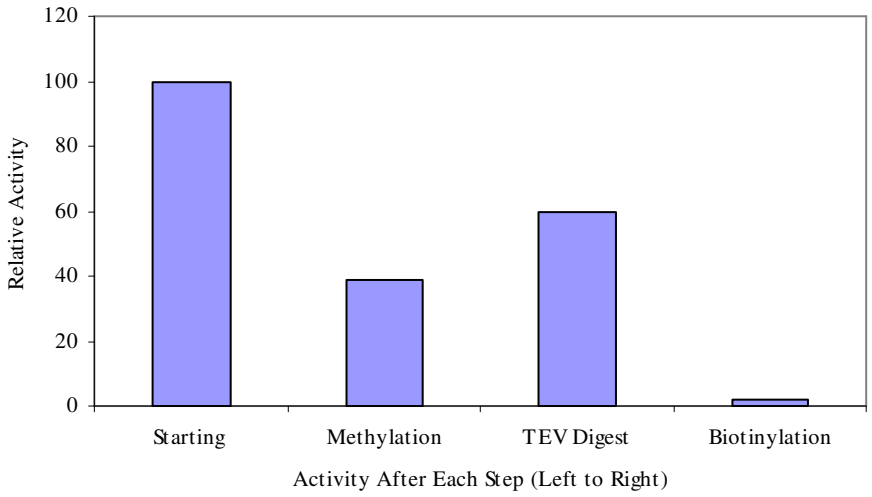
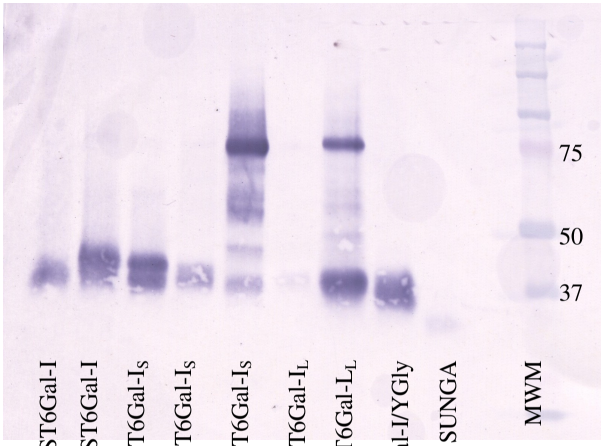
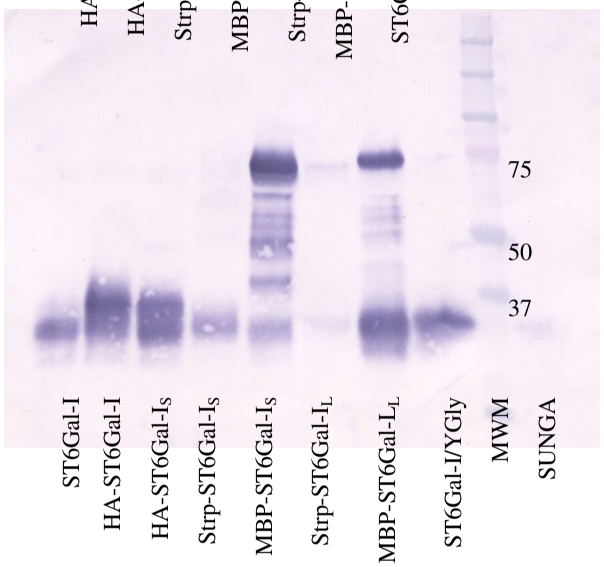


Figure 23: Chemical modification of ST6Gal-I. MBP-ST6Gal-I (500µg) was methylated using reductive methylation followed by a wash step using a Centricon®, an ultrafiltration device (10kDa cutoff), to remove unreacted reagent. The sample was then digested with the TEV protease. The sample was then biotinylated using pre-activated biotin followed by a wash step using ultrafiltration device (10kDa cutoff). Aliquots of each sample were then assayed for ST6Gal-I activity as indicated.

A)


ST6Gal-I
HA-ST6Gal-I
HA-ST6Gal-I$_S$
Strp-ST6Gal-I$_S$
MBP-ST6Gal-I$_S$
Strp-ST6Gal-I$_L$
MBP-ST6Gal-L$_L$
ST6Gal-I/YGly
SUNGA
MWM
75
50
37

B)


ST6Gal-I
HA-ST6Gal-I
HA-ST6Gal-I$_S$
Strp-ST6Gal-I$_S$
MBP-ST6Gal-I$_S$
Strp-ST6Gal-I$_L$
MBP-ST6Gal-L$_L$
ST6Gal-I/YGly
MWM
SUNGA
75
50
37

C)

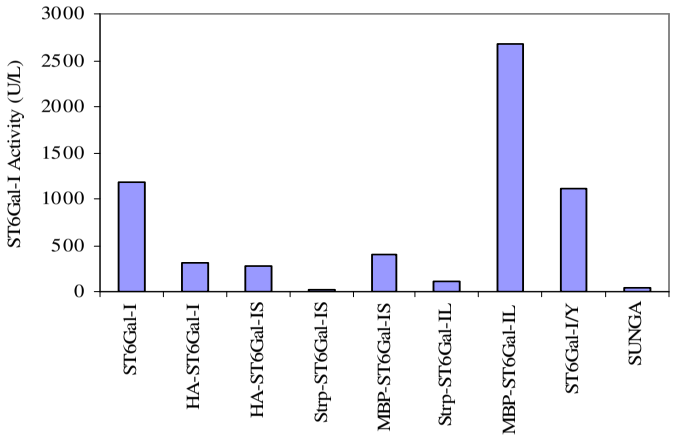

ST6Gal-I Activity (U/L)
3000
2500
2000
1500
1000
500
0
ST6Gal-I
HA-ST6Gal-I
HA-ST6Gal-IS
Strp-ST6Gal-IS
MBP-ST6Gal-IS
Strp-ST6Gal-IL
MBP-ST6Gal-IL
ST6Gal-I/Y
SUNGA

Figure 24: Comparison of various ST6Gal-I constructs for the level of expressed ST6Gal-I activity and protein form by Western blot in *Pichia* strain KM71H. A) Western blot of the crude secreted media from the respective cultures, B) Western blot of endo H_f-treated crude secreted media from the respective cultures, and C) Enzyme activities found in the secreted media from the respective ST6Gal-I transformants. The highest expressing zeocin-resistant c0lony of each construct was inoculated in BMGY media and incubated at 30°C for 24 hrs followed by induction in BMMY media for 2 days. The cells were centrifuged and the supernatant media was assayed for the secreted ST6Gal-I enzyme activity. Aliquots (50 µl)of each of the sample were taken followed by immediate addition of 5X SDS-PAGE buffer and denaturation by heating at 95°C for 5 min. The denautured samples were divided into two sets. The first set of samples was used directly for Western blot analysis. To each sample of second set 1 µl of endo H_f was added and all the three samples were incubated at 37°C for 1 h. Later they were used for Westrn blot analysis.

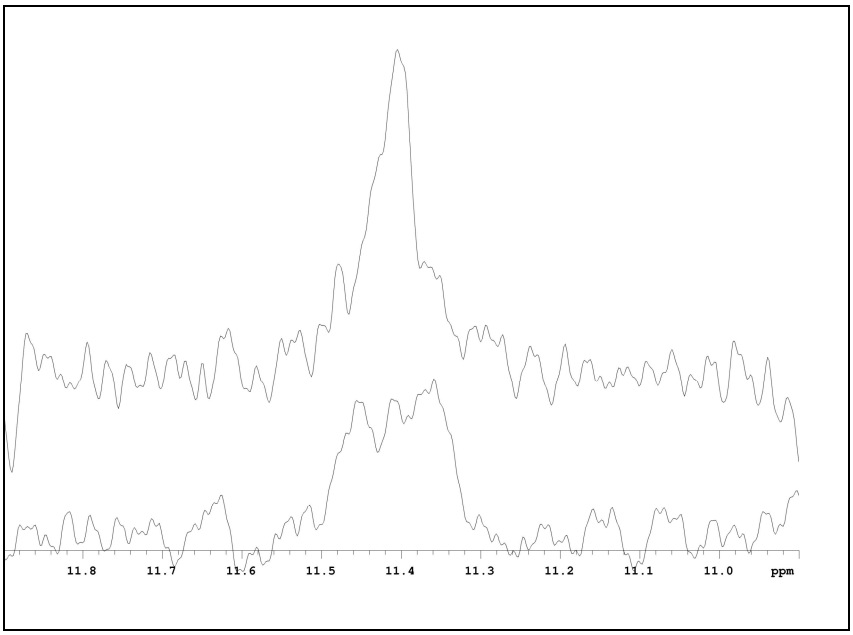


Figure 25: One dimensional ^{1}H spectra of ^{15}N labeled ST6Gal-I. Shown here is the signal from Trp side chain amide proton (11.4ppm), the bottom peaks (triplet) are as a consequence of ^{15}N coupled signal while the top peak is ^{15}N decoupled signal (Singlet). The efficiency of ^{15}N incorporation was then calculated based on the area under curve of the peaks in triple vs the singlet peak. It was assumed that backbone amide incorporates ^{15}N with same degree of efficiency as Trp side chain amides.

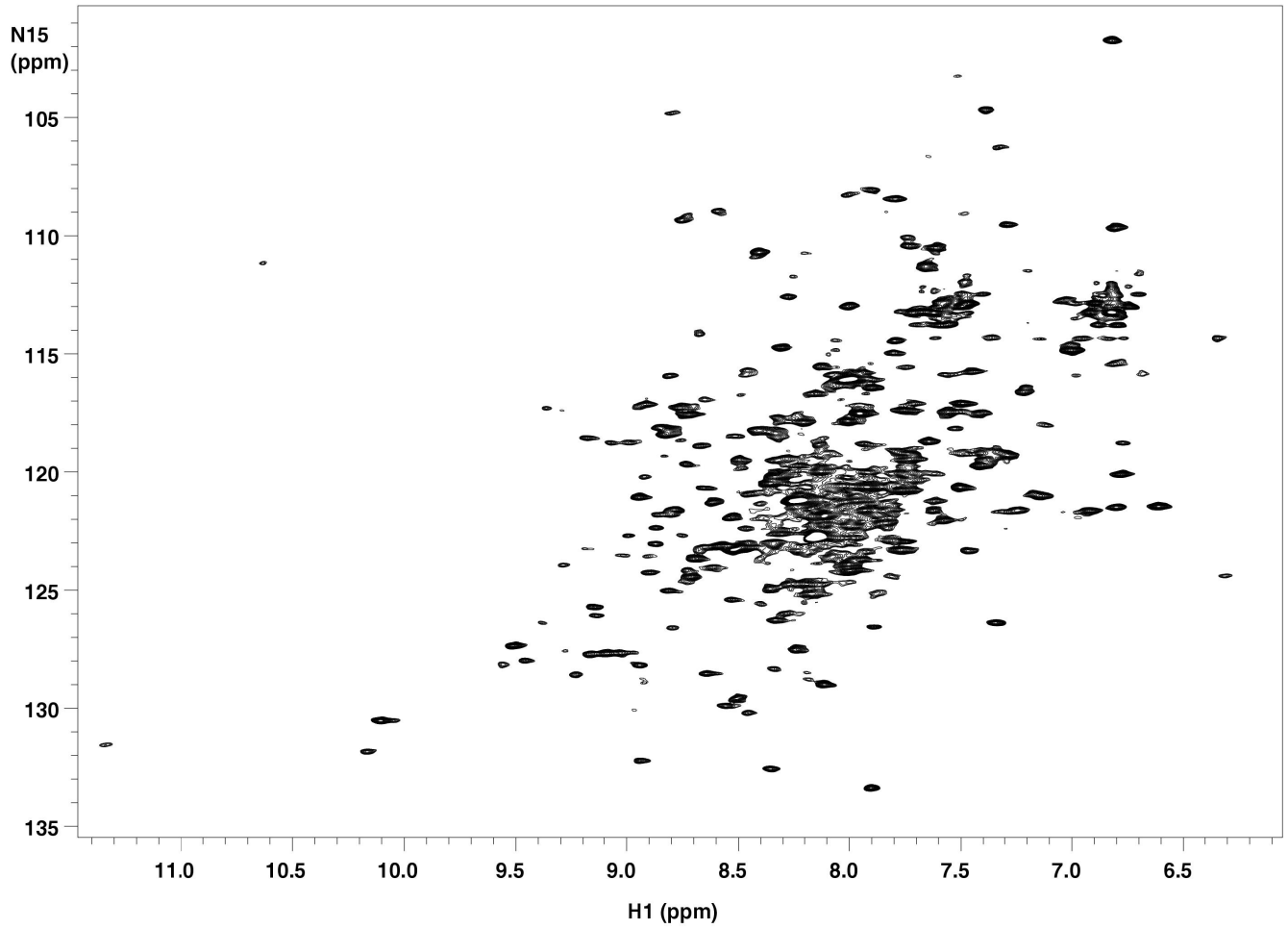


Figure 26: Two dimensional HSQC of ^{15}N labeled ST6Gal-I. The HSQC spectrum of ST6Gal-I obtained without deglycosylation. The spectrum was collected on 100 μM ^{15}N-labeled ST6Gal-I. A large number of peaks were aggregated between 7.5 and 8.5 ppm, which indicates that the protein is either unstructured or contains a large number of helices.

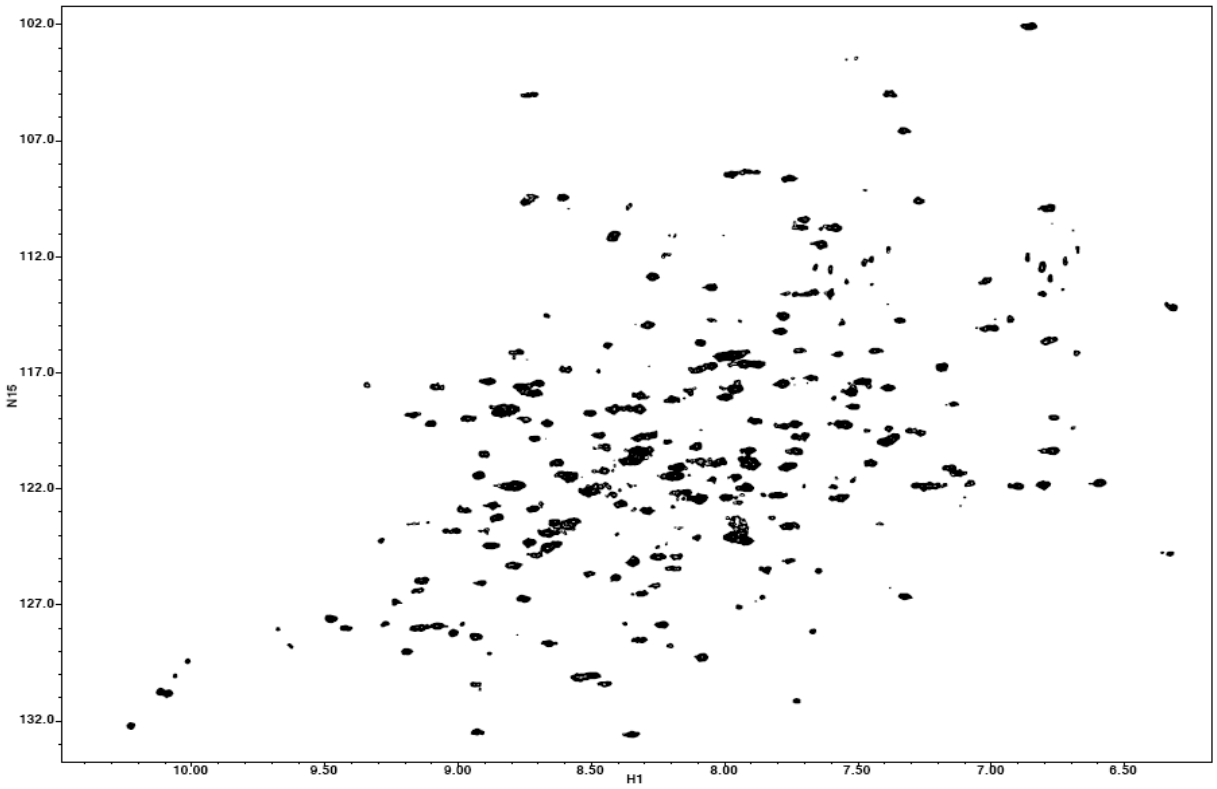


Figure 27: Two dimensional HSQC of deglycosylated ^{15}N labeled ST6Gal-I. The HSQC spectrum of deglycosylated ST6Gal-I was collected on a 100 μM sample. Unlike glycosylated ST6Gal-I, the sample peaks between 7.5 and 8.5 ppm, were well dispersed indicating a well structured protein. There are roughly 180 major peaks and 100 minor peaks with uniform peak intensities and very good dispersion.

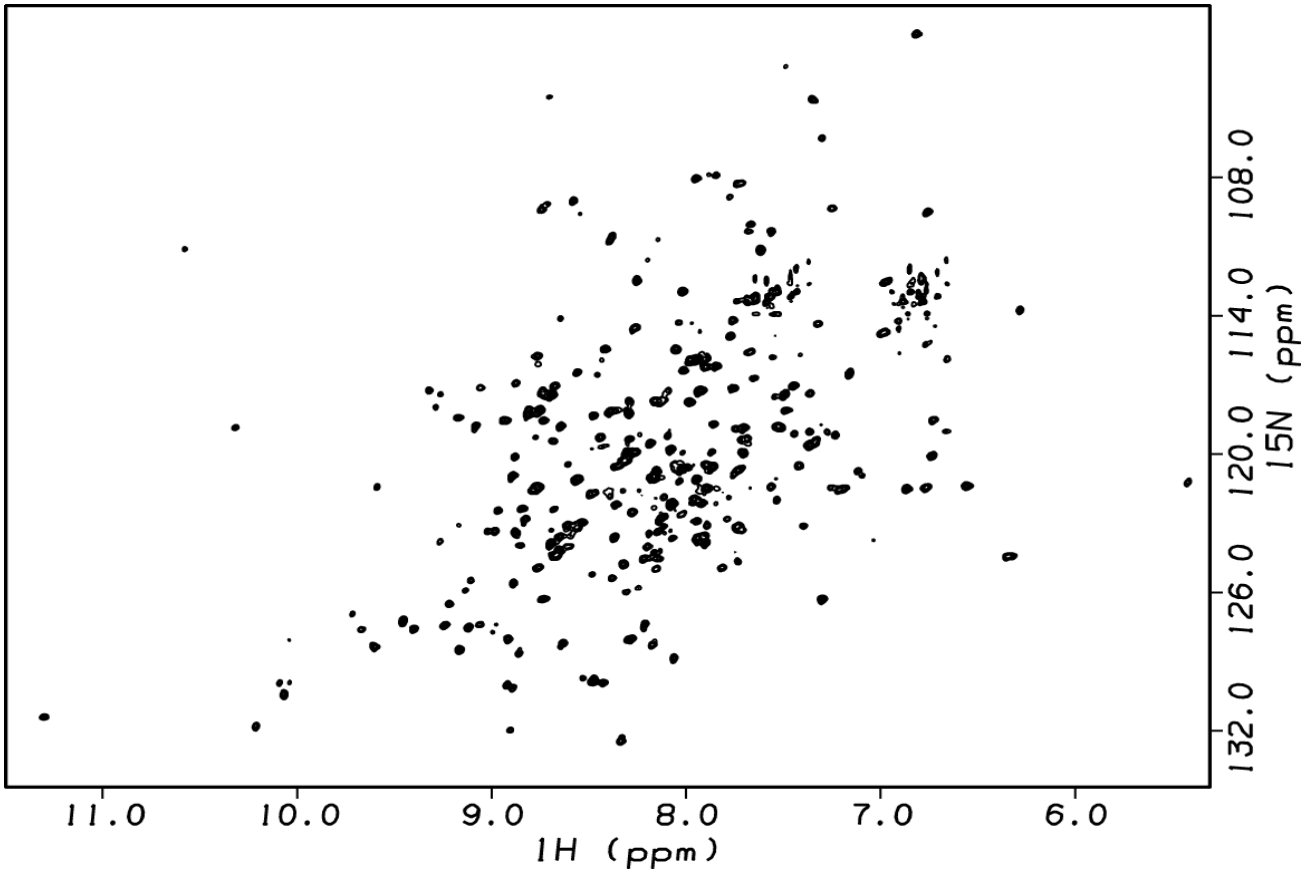


Figure 28: Two dimensional HSQC of deglycosylated double-labeled ^{1}H and ^{15}N labeled ST6Gal-I. HSQC spectrum of deglycosylated, double-labeled (^{2}H & ^{15}N) ST6Gal-I was collected on a 100 μM sample. The spectrum shows the well dispersed and sharp peaks. There are roughly 240 peaks with uniform peak intensities and very good dispersion.

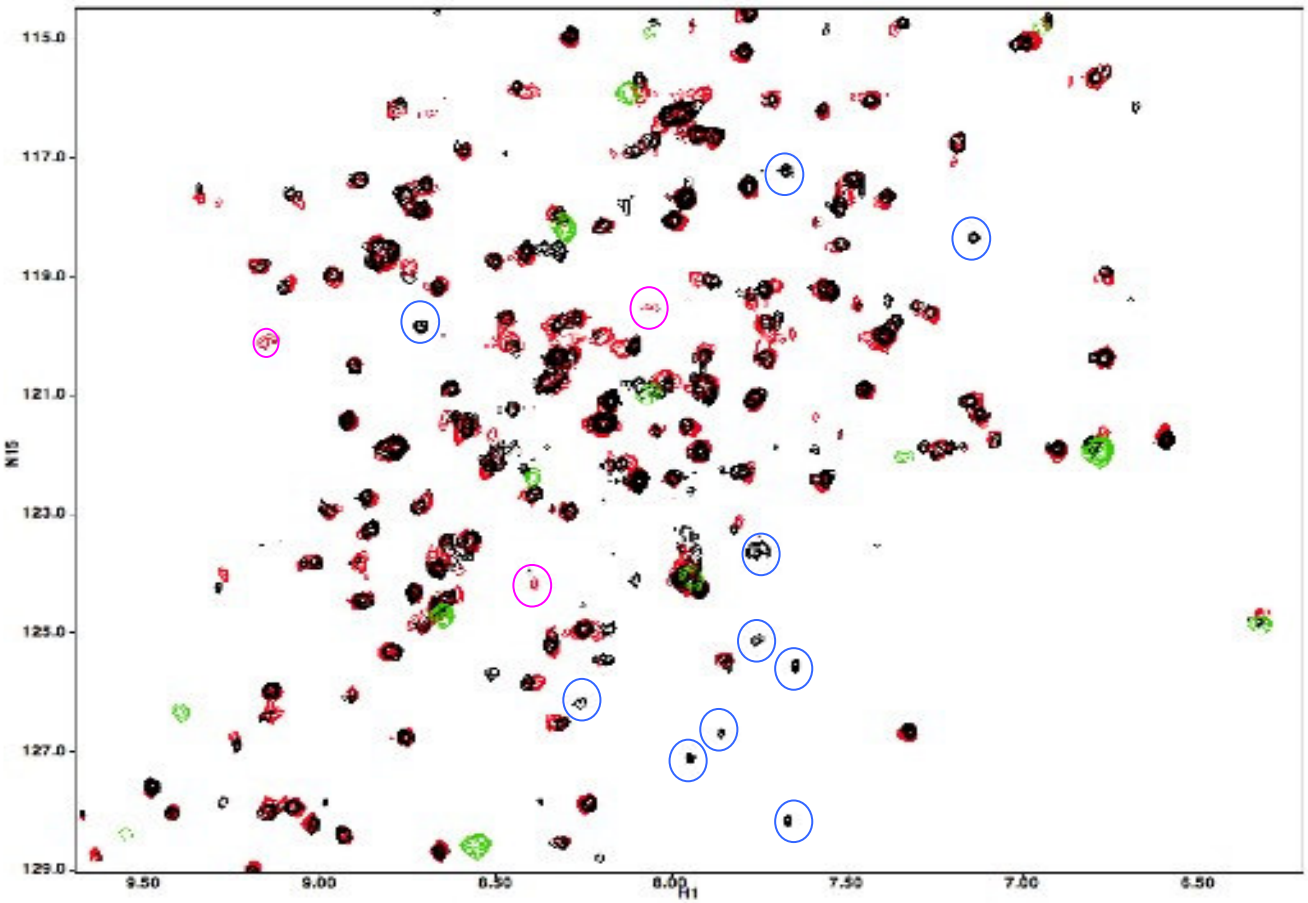


Figure 29: Substrate analog binding studies. The HSQC spectrum of deglycosylated ST6Gal-I was collected on 100 µM sample (Black peaks), Later, the donor substrate analog CMP-3FSia was added to a final concentration of 1 mM and an HSQC spectrum was collected (Red peaks). Intensities of Peaks with blue circles decreased with addition of substrate analog and intensities of peaks with pink circles increased with addition of substrate analog indicating the protein undergoes a substrate induced conformational changes. Green peaks are from ^{15}N Phe labeled recombinant ST6GalI expressed from HEK cells.

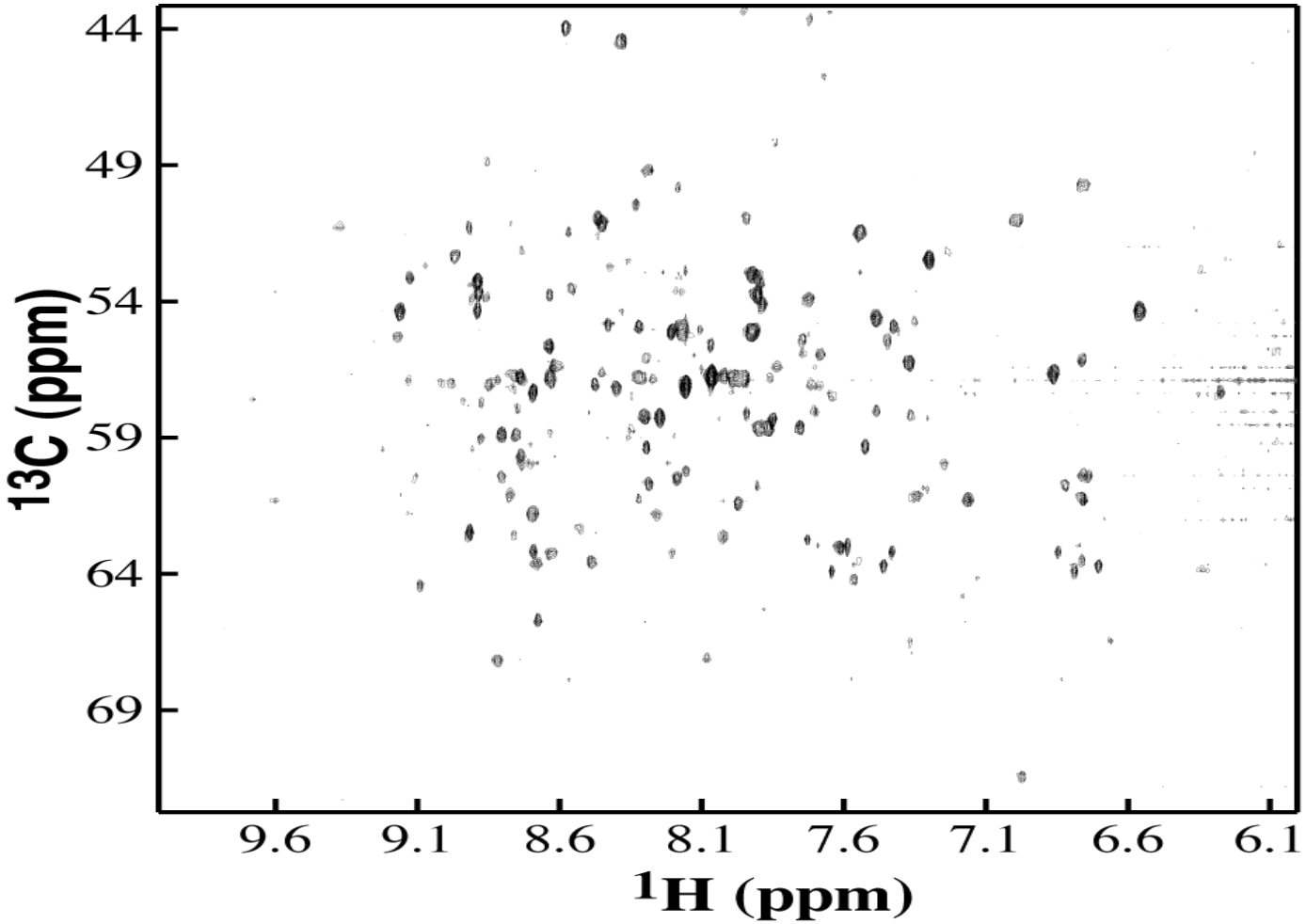


Figure 30: Three dimensional HNCA spectrum of triple-labeled ST6Gal-I. Shown here is the projection of all planes along the ^{15}N axis after Fourier transform of all three dimensions. Each peak in the spectrum represents either a Cα-HN pair from within a residue or between Cα of a residue and HN of the next residue. Therefore, it is theoretically possible to see twice the number of peaks in the HNCA projection as the number of residues excluding proline. However, in the HNCA spectrum about 200 peaks were seen, a majority of which correspond to Cα-HN within the same residue.

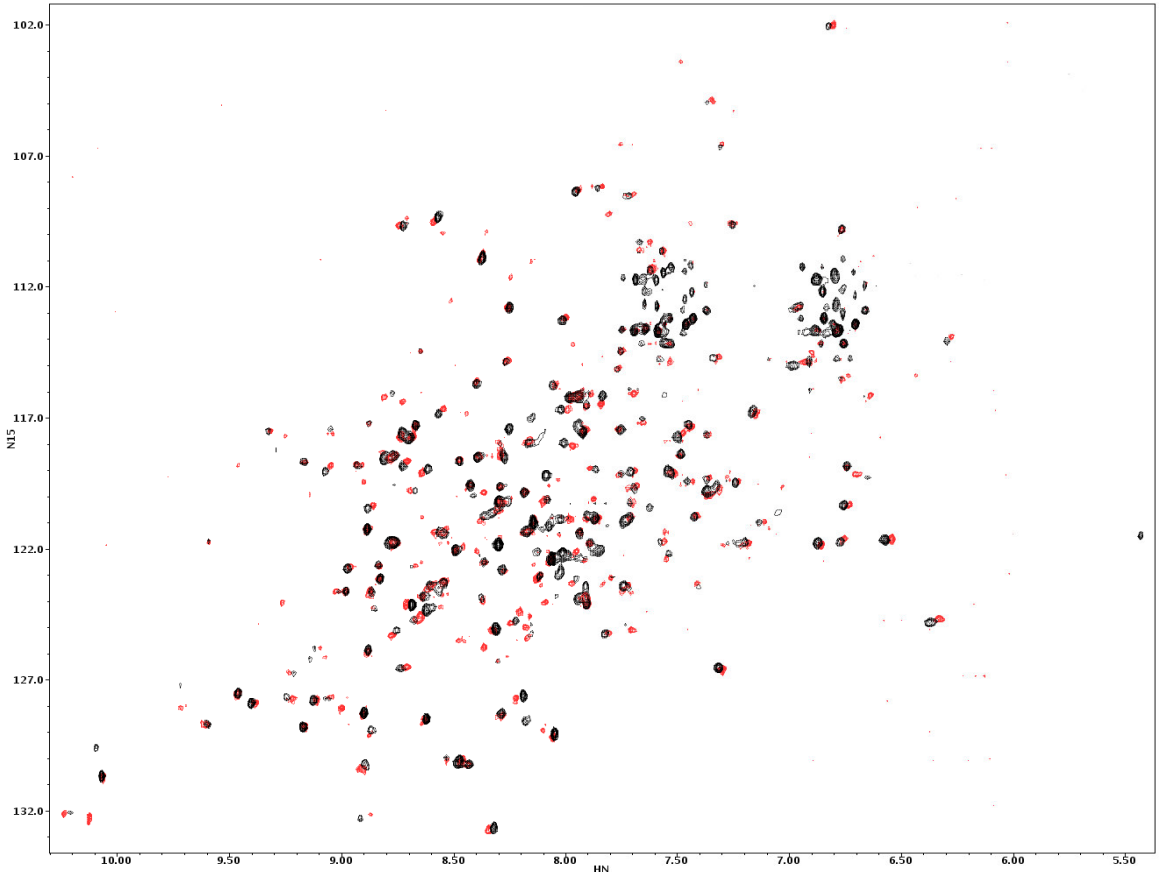


Figure 31: Overlay of the HSQC spectra of deglycosylated ST6Gal-I and TROCY spectra of deglycosylated and TEV-Protease digested MBP-ST6Gal-I. The HSQC of degycosylated double-labeled (^{2}H and ^{15}N) ST6Gal-I (Red peaks), expressed using ST6Gal-I/pPICZαC construct is compared to TROCY spectrum of deglycosylated and TEV protease-digested double labeled (^{2}H and ^{15}N) MBP-ST6Gal-I (Black peaks), expressed using MBP-ST6Gal-I_L/pPICZαC construct. Almost all the amide peaks overlap except 6.75ppm and 7.75ppm in proton and 110ppm in nitrogen peaks in from later sample in which amino acid side chains aminde peaks were not suppressed, indicating that both samples have the same overall structure.

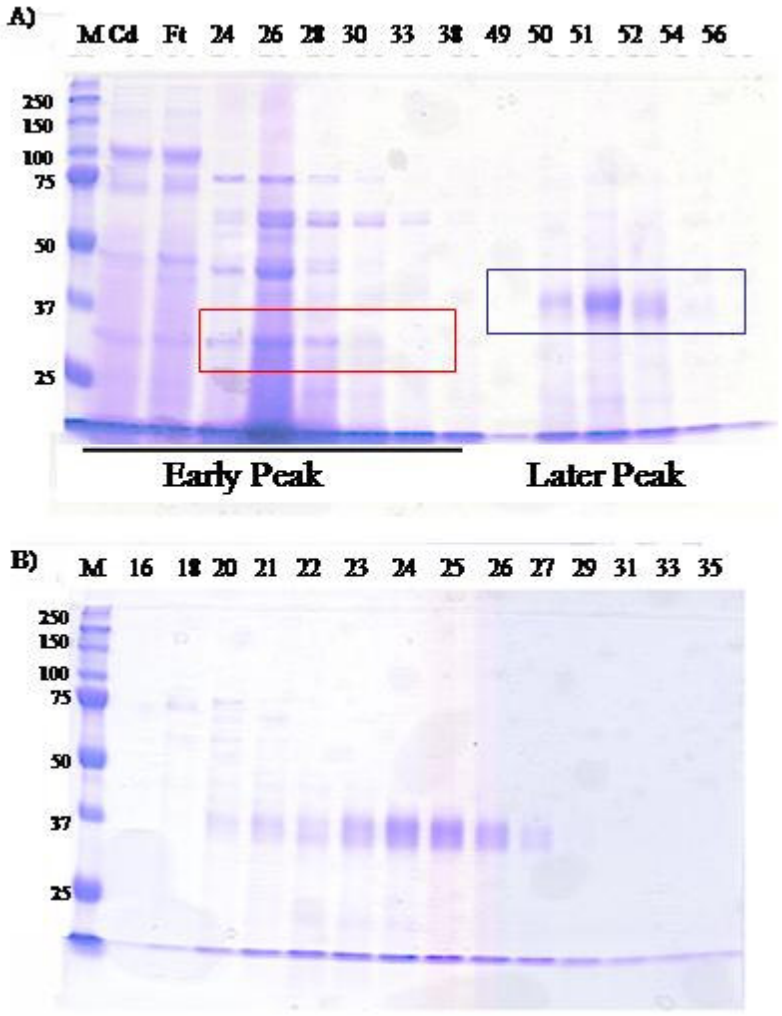


Figure 32: SDS-PAGE purification profile of recombinant ST6Gal-I from the YGly30 strain. Recombinant ST6Gal-I was expressed from a glycoengineered Pichia strain YGly30 in the fermentor similar to the wild type KM71H strain as described above. The harvested sample was purified by a combinationof Ni-NTA column chromatography and gel filtration chromatography. Shown here is A) SDS-PAGE of fractions derived from the Ni-NTA column purification, the bands in the red box in early fractions were the N-terminal degraded ST6Gal-I. B) SDS-PAGE of fractions derived from the gel filtration profile

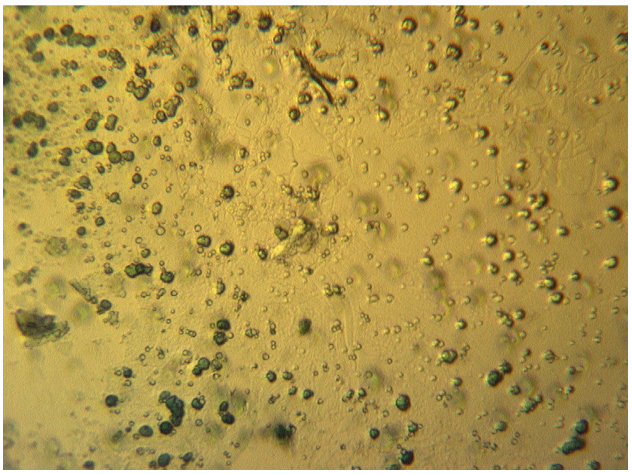


Figure 33: Protein precipitation during crystal screening. All the crystal screeing trials were performed with purified, deglycosylated recombinant ST6Gal-I (10 mg/ml in 25mM MES buffer, pH 6.5) in the presence of substrate analog, CMP-3FNeuAc (1.5 mM). Shown here is a picture from one of the original crystal screening conditions. Protein precipitates on the left side were dyed with the Izit Crystal Dye. The aggregates turned blue indicating that they are protein precipitation and not salt precipitatess.

Table 3: Purification table for ST6Gal-I expressed using controlled fermentation in complex media. Shown in the table are the enzyme activity and protein assays averages of at least 3 different fermentor runs of the ST6Gal-I/pPICZαC construct.

	Activity (μmoles/min)	Volume (ml)	Protein (mg)	Specific Activity μmoles/min/mg	Fold Purification	Yields (%)
Crude	2880 ± 210	950 ± 50	10800	000.3	1	100
Ni NTA	1630 ± 131	140 ± 10	3.82	426	1610	56.5 ± 4.9
Gel filtration	1350 ± 102	40 ± 3	2.19	615	2320	46.8 ± 3.5

Table 4: **Purification table for ST6Gal-I expressed using controlled fermentation in minimal media.** Shown in the table are the values for enzyme activity and protein recovery for the enzyme purification from the controlled fermentation of the ST6Gal-I/pPICZαC construct in minimal media.

	Activity (μmoles/min)	Volume (ml)	Protein (mg)	Specific Activity (μmoles/min/mg)	Fold Purification	Yields (%)
Crude	2590	910	7860	0.33	1	100
Ni NTA	1560	131	3.61	433	1320	60.5
Gel filtration	1140	36	1.82	625	1900	44.0

Table 5: Purification table for deglycosylated ST6Gal-I. Shown in the table are the values for enzyme activity and protein recovery for the enzyme purification including deglycosylation using endo H_f of recombinant ST6Gal-I that was expressed using controlled fermentation of the ST6Gal-I/pPICZαC construct in complex media.

	Activity (μmoles/min)	Volume (ml)	Protein (mg)	Specific Activity (μmoles/min/mg)	Fold Purification	Yields (%)
Crude	2810	950	11100	0.25	1	100
Ni NTA	1590	150	3.85	414	1640	56.8
Endo H	604	4.5	2.89	209	826	21.5
ConA	598	55	2.20	278	1100	21.3
Gel filtration	268	35	0.61	439	1730	9.5

Table 6: Purification table for deglycosyated ST6Gal-I, with protease inhibitor in purification (TPCK). Shown in the table are the values for enzyme activity and protein recovery for the enzyme purification including deglycosylation using endo H_f of recombinant ST6Gal-I using controlled fermentation of the ST6Gal-I/pPICZαC construct in complex media. The harvest media was treated with TPCK before proceeding for purification.

	Activity (μmoles/min)	Volume (ml)	Protein (mg)	Specific Activity (μmoles/min/mg)	Fold Purification	Yields (%)
Crude	2990	950.0	10900	0.27	1	100
Ni NTA	1980	145.0	3.78	524	1910	66.3
Endo H	1220	4.5	3.20	382	1390	40.7
ConA	1060	55.0	3.00	356	1300	35.5
Gel filtration	1040	35.0	1.76	592	2160	34.9

Table 7: Purification table for deglycosylated ST6Gal-I with protease inhibitors during both expression and purification (TPCK). Shown in the table are the values for enzyme activity and protein recovery for the enzyme purification including deglycosylation using endo H_f of recombinant ST6Gal-I using controlled fermentation of the ST6Gal-I/pPICZαC construct in complex media with protease inhibitors (BSA, EDTA, Pepstain A, and Chymostatin). Also, the harvest media was treated with TPCK before proceeding for purification.

	Activity (μmoles/min)	Volume (ml)	Protein (mg)	Specific Activity (μmoles/min/mg)	Fold Purification	Yields (%)
Crude	6380	950	36000	0.18	1	100
Ni NTA	4160	150	4.11	1010	5710	65.2
Endo H	2800	5	3.90	724	4090	44.0
ConA	2590	55	3.78	684	3860	40.6
Gel filtration	2350	40	2.92	806	4550	36.9

Table 8: Purification table for ^{15}N labeled and deglycosylated ST6Gal-I with protease inhibitors in both expression (BSA, EDTA, Pepstain A, and Chymostatin) and purification (TPCK). Shown in the table are the values for enzyme activity and protein recovery for the enzyme purification including deglycosylation using endo H_f of recombinant ST6Gal-I using controlled fermentation of the ST6Gal-I/pPICZαC construct in ^{15}N-labeled minimal media with protease inhibitors (BSA, EDTA, Pepstain A, and Chymostatin). Also, the harvest media was treated with TPCK before proceeding for purification.

	Activity (μmoles/min)	Volume (ml)	Protein (mg)	Specific Activity (μmoles/min/mg)	Fold Purification	Yields (%)
Crude	5470	950	17900	0.31	1	100
Ni NTA	3460	150	4.00	873	2850	63.2
Endo H	2250	5	3.13	719	2350	41.2
ConA	2240	55	3.01	751	2450	40.9
Gel filtration	2090	40	2.54	821	2680	38.1

Table 9: Purification table for ^{2}H labeled and deglycosylated ST6Gal-I with protease inhibitors in both expression and purification. Shown in the table are the values for enzyme activity and protein recovery for the enzyme purification including deglycosylation using endo H_f of recombinant ST6Gal-I using controlled fermentation of the ST6Gal-I/pPICZαC construct in deuterated minimal media with protease inhibitors (BSA, EDTA, Pepstain A, and Chymostatin). Also, the harvest media was treated with TPCK before proceeding for purification.

	Activity (μmoles/min)	Volume (ml)	Protein (mg)	Specific Activity (μmoles/min/mg)	Fold Purification	Yields (%)
Crude	643	900	17500	0.037	1	100
Ni NTA	1200	150	3.65	328	890	186.1
Endo H	960	5	2.74	350	9500	149.2
ConA	790	55	2.23	353	9580	122.4
Gel filtration	670	40	1.45	461	12500	104.0

Table 10: Purification table for double-labeled (^{2}H and ^{15}N) and deglycosylated ST6Gal-I with protease inhibitors in both expression and purification. Shown in the table are the values for enzyme activity and protein recovery for the enzyme purification including deglycosylation using endo H_f of recombinant ST6Gal-I using controlled fermentation of the ST6Gal-I/pPICZαC construct in ^{15}N-labeled deuterated minimal media with protease inhibitors (BSA, EDTA, Pepstain A, and Chymostatin). Also, the harvest media was treated with TPCK before proceeding for purification.

	Activity (μmoles/min)	Volume (ml)	Protein (mg)	Specific Activity (μmoles/min/mg)	Fold Purification	Yields (%)
Crude	616	900	17700	0.036	1	100
Ni NTA	1170	150	3.70	316	8750	189.7
Endo H	942	5	2.69	350	9700	152.9
ConA	770	55	2.12	363	10100	124.8
Gel filtration	662	40	1.52	435	12100	107.4

Table 11: Purification table for double-labeled (^{2}H and ^{15}N), deglycosylated and TEV protease digested MBP-ST6Gal-I with protease inhibitors during both expression and purification. Shown in the table are the values for enzyme activity and protein recovery for the enzyme purification including deglycosylation using endo H_f and TEV protease digest using proTEV protease of recombinant MBP-ST6Gal-I that was expressed using 5 ml tube cultures of MBP-ST6Gal-I_L/pPICZαC construct in ^{15}N-labeled deuterated minimal media with protease inhibitors (BSA, EDTA, Pepstain A, and Chymostatin). Also, the harvest media was treated with TPCK before proceeding for purification.

	Activity (μmoles/min)	Volume (ml)	Protein (mg)	Specific Activity (μmoles/min/mg)	Fold Purification	% Yields
Crude	565	1400	17482	0.032	1	100
Ni NTA	1570	190	9.61	163	5030	276.8
Gel filtration	1010	10	6.67	151	4680	178.4
TEV digest	951	9	6.89	138	4280	168.1
Endo H	918	9	7.17	128	3950	162.2
NiNTA	637	6	2.71	235	7260	112.6
Gel filtration	597	6	2.10	284	8780	105.4

CHAPTER 4

DISCUSSION

The aim of our current study was to evaluate *P. pastoris* as an expression host for the expression of rat ST6Gal-I for structural and functional studies. Our study shows the suitability of *P. pastoris* for both X-ray crystallization and NMR based structural analysis. Even though suitability for *P. pastoris* as a heterologous expression system for ST6Gal-I has been demonstrated by expressing a soluble and functional form of rat ST6Gal-I [106, 164], further purification and characterization has never been reported. Similarly recombinant expression of sialyltransferases in prokaryotes yields insoluble inlusion bodies of enzyme which cannot be solubilized with mild detergents [80, 165] or soluble enzyme with a miniscule yields [166]. Also, the expression of the soluble and functional form of ST6Gal-I has been very well established in a mammalian expression system [47]. Our present data indicates that the expression levels of the ST6Gal-I were sufficently high for use in the crystallization trials as well as structural and functional studies.

Like most of the glycosyltransferases, ST6Gal-I is a type II membrane protein with short NH_2 terminal cytoplasmic domain, single pass transmembrane domain, stem region, and large COOH terminal catalytic domain. Here, we examined the expression of both active and soluble forms of recombinant ST6Gal-I, a rat-derived sialyltransferase, in a yeast host system. Therefore, for our purposes, both the cytosolic region and transmembrane domain were removed, resulting in a soluble form of the catalytic domain of the enzyme. Moreover, two different isoforms (STtyr or STcys) of ST6Gal-I exist in rat which differ in a single amino acid residue at position 123 (i.e.

there is either Tyr or Cys in that position [151]). Our present studies exclusively used the STtyr isoform, as it catalytically more active than the STcys isoform.

The commercially available vector from Invitrogen used for the expression of ST6Gal-I in *Pichia pastoris* provides a wide variety of options for generating fusion proteins, including NH_2-terminal α-factor signal sequence as well as a COOH-terminal His tag for affinity purification and Myc tag for Western blot analysis. Unfortunately, any modification on the COOH-terminus of ST6Gal-I results in an inactive enzyme. Therefore, an 8xHis tag was incorporated at the NH_2-terminus of truncated ST6Gal-I. We also made additional constructs with the Streptavidin and MBP fusion proteins to test their effectiveness in protein production.

Initial screening of transformants in 5 ml tube cultures (using BMGY for growth and BMMY for induction at 30°C) achieved a cell density of 19-21g/L and produced 1187 ± 164 Units/L of ST6Gal-I activity. Similarly, when ST6Gal-I was expressed in shake flask cultures (50 ml culture in 250 ml flask or 1 L culture in 3 L shake flask), the cells grew to same cell densities. However, little or no enzyme activity was observed in either scale of shake flask cultures. These results were highly unexpected since the cells grew to the same cell densities and were induced using equivalent concentrations of methanol as in the tube cultures. This result was highly unlikely since the shake flasks are better designed for aeration of the cultures compared to the cylindrical tube cultures.

Similarly, when the *Pichia* construct was grown using the same temperature (30°C) and pH conditions in a 1L fermentor in a controlled environment and dissolved oxygen was maintained at 50% saturation, there was a little or no improvement in enzyme expression level compared to the shake flask cultures, even though the cell densities were several-fold higher than

in the tube cultures. These data suggest that the recombinant enzyme expressed was unstable in these conditions [120] and was degraded as soon as it was being produced.

When induction was carried out at lower temperature 20°C, there was 4-5 fold increase in recombinant enzyme production compared to induction at 30°C. There could be numerous reasons to explain these results, in addition to reduced proteolysis at lower temperatures, increased protein production has been attributed to reduced thermodynamics during protein folding and a prolonged time frame to complete the protein folding process [129]. Subsequent optimization of pH and media conditions resulted in expression levels of 2988 U/L, nearly 3 times higher expression level than the 5 ml tube cultures, but surprisingly low productivity considering the high cell densities of the controlled fermentation conditions by comparison to the low cell densities of the tube culture. However, these expression levels were high enough to carry out enzyme production for structural and functional studies.

The expressed ST6Gal-I protein was effectively purified by a combination of IMAC and gel filtration column chromatography. In the protein fraction purified by gel filtration column chromatography, a few minor bands with faster mobility on the gel could be detected, with cross-reactivity to the anti- ST6Gal-I antibody, indicating that they represented proteolytic degradation during purification.

A previous study has demonstrated that removal of up to 96 amino acid residues from NH_2-terminus of rat ST6Gal-I does not alter any enzyme functionality [66]. Similarly, ST6Gal-I/pPICZaC lacks of 96 amino acid residues from the NH_2-terminus of wild type protein. As expected, the enzymatic properties of the soluble ST6Gal-I were comparable to those of rat liver ST6Gal-I, with regard to substrate specificity and stability. The k_m value for the recombinant enzyme (177 ± 10 μM) against the donor substrate, CMP-Neu5Ac, and acceptor substrate (2.6 ±

0.2 mM) was similar to published values for ST6Gal-I and truncated ST6Gal-I expressed in mammalian or yeast cells [164, 167, 168].

In any eukaryotic expression system glycoproteins are expressed with a considerable microheterogeneity in their glycan structure [162], and yeast protein glycosylation in the *Pichia pastoris,* is no exception. Typically, N-glycan structures in *Pichia* are high mannose structures on the secreted glycoproteins ranging from $Man_{10}GlcNAc_2$ to $Man_{16}GlcNAc_2$. Therefore, as expected, MALDI-TOF mass analysis of the released N-glycan from recombinant ST6Gal-I yielded high mannose structures ranging from $Man_5GlcNAc_2$ to $Man_{17}GlcNAc_2$. However, the majority of the glycans were found as $Man_7GlcNAc_2$ to $Man_{12}GlcNAc_2$ structures. Unlike previously demonstrated, [141] some of the N-glycans released from recombinant protein were substituted with negatively charged phosphate.

Hyperglycosylation of the glycoproteins can pose serious difficulties for structural and functional analysis by crystallization or NMR based studies (described in detail below). Moreover, it has been previously established that the N-glycans of the ST6Gal-I STtyr isoform are not essential for enzyme activity *in vivo* or *in vitro* [66]. However, they are necessary for proper folding and transportation of the enzyme through secretary pathway [66]. Therefore, we decided to deglycosylate recombinant ST6Gal-I. Typically, there are two options for removing N-glycans, PNGase F and Endo H. PNGase can remove all of N-glycans, irrespective of their type, and will result in deamination of Asn (N) to form Asp (D), thus changing the character of the amino acid side chain. However, Endo-H removes high mannose type glycans and it cleaves between the two core GlcNAc residues to result in the retention of the Asn-GlcNAc linkage. When we compared the action of both enzymes to remove glycans from recombinant ST6Gal-I, both were able to remove glycans from denatured ST6Gal-I, but only Endo H was able to

remove glycans from a native ST6Gal-I. Moreover, even partial deglycosylation by PNGase F caused precipitation of recombinant ST6Gal-I. Endo-H_f, a MBP-fusion protein of Endo-H, was used in later purification steps and its larger size allowed it to be easily separated from St6Gal-I by gel filtration.

The expressed ST6Gal-I protein was subsequently purified and deglycosylated to near homogeneity by a combination of IMAC, dialysis, deglycosylation by Endo-H_f, and gel filtration column chromatography. The yield of enzyme activity through this procedure was less than 10 percent. Moreover, in the final steps of protein purification, showed a pair of bands, one with anticipated size and a second faster migrating band that suggested proteolytic degradation during purification. NH_2-terminal sequencing of the two bands indicated that the top band was the intact protein with the 8xHis tag and lower band was a mixture of proteolytic products cleaved after residues 24 (residue 108 of full length) and 27 (residue 111 of full length). Sequence analysis suggested that the cleavage was being carried by an enzyme with chymotrypsi-like specificity. Therefore, when harvested media was treated with the chymotrypsin inhibitor, TPCK, the recovery of enzyme increase 4-fold, and the lower band was still detected by Western blots, but with much lower the intensity.

Thus, as mentioned above, secreted proteases can be a serious problem in the purification and expression of recombinant proteins in *P. pastoris*. Even though the identities of such proteases in *Pichia* cultures have been elusive, the existence of proteases in high cell density fermentation has been widely accepted [118, 119, 125, 132-134]. Usually cell lysis followed by death during high cell density fermentation results in the release of intracellular protease into the medium, which could explain the low yields of recombinant ST6Gal-I in high density fermentation at 30°C.

Therefore, to increase the yields of the recombinant ST6Gal-I protein, we decided to use a collection of protease inhibitors during the fermentation process in addition to the use of TPCK during purification. The protease substrates and protease inhibitors included BSA (10g/L, protease substrate), chymostatin (110 μM serine protease inhibitor), pepstatin A (10 mg/L aspartic acid protease inhibitor), EDTA (10 mM, metalloprotease inhibitor) and ε-amino caproic acid (100 mg/L, carboxypeptidase inhibitor). When these protease inhibitors were used either singly or in dual combination they did not significantly improve the expression yields of ST6Gal-I. However, the combination of three or more of the inhibitors resulted in improved expression yields by two- to four -fold compared to no protease inhibitor addition during expression. Even in high cell-density fermentations, the expression yields increased by two-fold following the addition of four protease inhibitors.

Following the harvest of conditioned media follow remaineder of the induction the percent recoveries of recombinant ST6Gal-I through the protein purificatioin scheme were the same regardless of wheather the preparation was generated with or without protease inhibitors. Thus, the controlled fermentation with protease inhibitors yielded two-fold more purified ST6Gal-I enzyme activity compared to the controlled fermentation without protease inhibitors.

Even though in previous study [66] the minimal length for the catalytic domain was established to be between residues 97-403, in our hands the purified recombinants ST6Gal-I was further degraded at the NH_2 terminus to position 111 . This clipped protein was active and only slightly compromised in catalytic activity. Therefore, we generated another construct termed ST6Gal-I_S, that started at position 111 and contained the remaining catalytic domain sequence. We also synthesized two more constructs with monomeric streptavidin and maltose binding

protein fusions, Strp-ST6Gal-I_S and MBP-ST6Gal-I_S, respectively, that were fused upstream of the position 111 start of the catalytic domain.

While, ST6Gal-I_S construct transformants with NH_2–terminal MBP or HA tag showed less than 10% of the ST6Gal-I enzyme activity and two-fold more protein as compared to the longer ST6Gal-I transformants, containing only an NH_2 His tag, the Strp-ST6Gal-I_S fusion transformants did not show any enzyme activity. Moreover, in the Strp-ST6Gal-I_S transformants, Western blots indicated that the induced cultures produced only a small amount of clipped ST6Gal-I and no fusion protein. In some instances of bacterial expression of monomeric streptavidin fusion proteins the fusion proteins become unstable and are degraded by proteases in the media (communication with the author of ref [152]). Also, since biotin is essential for the growth of *Pichia pastoris*, expression of a Strp-ST6Gal fusion protein might be toxic to the cells and potentially lead to the degradation of the expressed protein. In contrast, the MBP-ST6Gal-I_S construct transformants expressed the fusion protein at comparable or higher levels than the His-tagged longer ST6Gal-I construct, but an appreciable cleavage of the tag sequence was detected. These results suggest that a larger catalytic domain (more than 112-403 residues) is required for the proper initial folding and functioning of ST6Gal-I even though the pre-folded longer enzyme can be subsequently trimmed at the NH_2 terminus and remain catalytically active.

Similarly, the Strp-ST6Gal-I_L construct (a construct with catalytic domain from 97-403 residues), showed essentially no enzymatic activity and no detectable fusion protein by SDS-PAGE, thus verifying that the monomeric streptavidin fusion expression in *Pichia* destabilizes the recombinant fusion protein, which is then degraded.

In 5ml tube cultures the MBP-ST6Gal-I_L construct transformants expressed at least two-fold more enzyme activity compared to the same sized catalytic domain of the His-tagged

ST6Gal-I construct without an MBP fusion. When the same construct was expressed using controlled fermentation a much lower yield, (<1/6th of the level of the His-tagged ST6Gal-I construct) were observed. These lower yields maybe attributed to the precipitation of the fusion protein under high shear stress of impellers in the controlled fermentation as precipitation of the fusion protein was observed under high stirring during ultrafiltration.

Isotope Labeling

Even though the *Pichia pastoris* expression system provides several advantages, labeling in this system is far from being economically feasible. Unfortunately, most methods which have been previously described for the efficient labeling of recombinant proteins using *P. pastoris* have generally utilized the expression of low molecular weight, stable proteins, whose yields were several hundred milligrams per liter. However, in practice most glycoproteins are rarely expressed at even a few milligram/L in *Pichia* levels and are commonly unstable in most of the previously published conditions.

The efficient labeling of ST6Gal-I with ^{13}C or ^{15}N isotopes does not require isotopes in the initial glycerol and glycerol fed-batch steps. Despite this benefit of reduced isotope usage, the efficiency of labeling is generally much lower than in other recombinant host systems such as bacteria. *Pichia* cultures were allowed to grow to high cell density such as 150-200 g/L (wet cell mass) before inducing the culture with methanol. The initial growth phase was started by inoculating seed cultures in the unlabeled minimal media with a low concentration (2.5 g/L) of unlabeled nitorgen source in the form of $(NH_4)_2SO_4$. In normal fermentation conditions, during the exponential growth phase NH_4OH was used as the sole nitrogen source, which serves two purposes. First, as the organism grows, it produces acid and addition of NH_4OH functions as a

base to adjust the pH of the growth medium. Second, the added NH_4OH acts as a nitrogen source that does not lead to increased ionic strength of the medium, because the counter ion is –OH. However, just before the end of growth phase $(^{15}NH_4)_2SO_4$ was used to replace NH_4OH as a nitrogen source and the culture pH was maintained using KOH as the titrating base. This protocol eliminates the harmful built-up of high salt concentration if $(NH_4)_2SO_4$ or NH_4Cl were used as nitrogen source for the entire growth and induction phases [146]. By following this approach we were able to achieve about 70% to 90% ^{15}N isotope incorporation using controlled fermentation. The recovery and purification of the single isotope ^{15}N labeled ST6Gal-I was comparable to unlabeled ST6Gal-I.

For double ^{2}H and ^{15}N labeling of ST6Gal-I in *Pichia* cultures the cells were adapted to the isotopically-labeled media as described earlier [169]. The cells were slowly adapted to deuterated media with a gradual increment of deuterium content untill 95% D_2O was obtained. Subsequent inoculation into 3 different fermentors with 95%, 97% and 99% deuterium containing minimal media revealed unique growth characteristics in each of the three different media. It took longer for cells to grow to same cell density in 99% deuterium compared to 97% and 95% deuteration. Moreover, there was little difference in the rate of methanol consumption once reached to same cell density. However, ST6Gal-I enzyme expression was considerably higher in 95% D_2O compared to 97% or 99% D_2O, but still much lower than in non-deuterated media. Further purification of expressed ST6Gal-I in 95% D_2O yielded activity levels that exceeded 100% compared to recoveries of 35-40% for unlabeled enzyme, suggesting that there was isotope effect not only for the production of ST6Gal-I, but also for catalysis by ST6Gal-I in enzyme assays. As expected, the total yield from the deuterated media decreased from 2.54 mg/liter in unlabeled media to 1.45 mg/liter in deuterated media.

The isotope effect was even more severe in triple labeled experiments where the yield of the triple labeled protein (^{2}H, ^{13}C and ^{15}N) was significantly lower (0.6 mg/ liter) than the double labeled (^{2}H and ^{15}N) protein (1.45 mg/liter).

NMR analysis

The efficiency of isotope (^{15}N) incorporation in recombinant ST6Gal-I, using *P. pastoris* as an expression host was estimated using ^{15}N-coupled and -decoupled one-dimensional ^{1}H spectra. The signal from Trp residues appear downfield (11.4) compared to other signals. It is a singlet in the decoupled spectrum whereas it is a triplet in the coupled spectrum indicating that there was a mixture of ^{15}N and ^{14}N containing NH pairs. For the ^{15}NH pair, the lack of decoupling during acquisition resulted in the signal being split into a doublet. This superimposed with the singlet signal originating from ^{14}NH pairs resulting in the triplet signal seen in the coupled spectrum. In the decoupled spectrum, both ^{15}NH and ^{14}NH signals are in the form of a singlet. Therefore, to estimate the ^{15}N incorporation percentage, the volume of the ^{14}NH singlet in the coupled spectrum is compared with the volume of the singlet in the decoupled spectrum. Despite the low sensitivity-to-noise ratio of the spectra, it is obvious that the size of the ^{14}NH component of the signal is less than one third of the total signal. Therefore, the total ^{15}N incorporation efficiency is believed to be in the range of 70 to 90%. One caveat of such a method for estimating ^{15}N incorporation efficiency is that the method implicitly assumes backbone amides incorporate ^{15}N with same degree of efficiency as Trp side chain amides. So far we have no data that contradicts this hypothesis.

Even though the 2D ^{1}H-^{15}N HSQC spectrum of uniformly ^{15}N-labeled ST6Gal-I showed only a low degree of dispersion of peaks, a large number of peaks were aggregated between 7.5

and 8.5 ppm, which at first look indicates that the protein is either unstructured or contains a large number of helices. However, from sequence analysis it has been estimated that, like most of the glycosyltrasferases, ST6Gal-I contains a "Rosmann-type" fold which contains alternating of β-sheets and α-helices [62].

The other possible reasons for such aggregated peaks can be attributed to N-glycosylation induced protein heterogeneity [170]. ST6Gal-I has two N-glycosylation sites, which are heterogeneously glycosylated *in vivo*. Analysis of N-glycan structures released from recombinant ST6Gal-I by MALDI-TOF MS showed a heterogeneous high mannose type structure. Glycan structures as high as Man_{17} were detected and some of the residues were found to be phosphorylated. As N-glycosylation on ST6Gal-I is very heterogeneous, the resulting heterogeneity might have affected the chemical shift of the backbone amides, producing broad-weak peaks. Moreover, the location of the glycosylation is important. If glycosylation site is located in the helical or unstructured regions then peak broadening would be centered on peaks in a helical region, which would explain the aggregation seen in the HSQC.

Deglycosylation of ^{15}N-labeled ST6Gal-I using Endo H would remove all the glycosylation heterogeneity by cleaving between the two core GlcNAc residues leaving a single GlcNAc attached to the protein Asu residue. The 2 D (^{1}H-^{15}N)-HSQC spectrum of deglycosylated ^{15}N-labeled ST6Gal-I showed no signs of an aggregated region as was seen in the 2D (^{1}H-^{15}N)-HSQC of glycosylated ST6Gal-I. Thus, it was confirmed that the aggregation of peaks was likely caused by heterogeneous glycosylation. With deglycosylated sample there were about 180 major peaks and their intensities were very uniform with great dispersion of the roughly 180 major peaks. However, peak widths did not change significantly before and after deglycosylation, indicating that despite the reduction in molecular weight, the dynamic

properties of deglycosylated ST6Gal1 is not more favorable towards NMR data collection than its glycosylated counterpart.

Given the size of ST6Gal-I and its tendency to dimerize at high concentrations, it became imperative to develop strategies for deuteration if the data quality of the NMR spectra was to be improved. The mechanism behind the deuteration-induced increase in sensitivity-to-noise ratio lies in the fact that NMR signals can be degraded by the presence of nuclei possessing strong magnetic properties. Since protons produce one of the strongest nuclear magnetic dipoles seen in any atom, the replacement of protons with its isotopic counter part, deuterium, should improve the sensitivity of NMR signals. In the case of ST6Gal-I, deuteration produced a dramatic improvement in the NMR spectra as there was very effective and uniform deuteration at the Cα position with growth in 95% D_2O and protonated carbon sources.

A comparison of the spectra with and without deuteration showed that the deuterated sample was able to produce peak widths that were much narrower than the equivalent spectra of protonated samples. As a consequence of the improved resolution many more peaks were resolved. There were as many as 240 out of 319 expected peaks that were very well resolved as compared to just 180 in the protonated sample. Moreover, many peaks seen in the deuterated HSQC originated from residues that have well dispersed chemical shifts and were invisible in the spectrum of the protonated sample. The peak intensities of the deuterated sample were also much higher than equivalent peaks in the protonated samples as the contour level of the HSQC of the protonated sample could barely be lowered to just above the noise threshold, whereas the background threshold of the HSQC spectra of the deuterated sample could be raised to produce spectra that were free from background noise. Therefore, deuteration has clearly increased the

possibility of collecting three dimensional NMR data that are high enough quality to be used in the backbone assignment of ST6Gal-I.

Unlike single- and double-labeled HSQC experiments, which were performed on 100 μM solution of ST6Gal-I, the 3D ^{1}H-^{13}C-^{15}N, HNCA experiment was carried out on only 60 μM of triple labeled ST6Gal-I as a result of lower recovery of the triply labeled enzyme. The HNCA spectrum was analyzed by a projection of all planes along the ^{15}N axis after Fourier transform of all three dimensions. Each peak in the spectrum represents either a Cα-HN pair from within a residue or between Cα of a residue and HN of the next residue, which theoretically increases the possibility that there would be twice as many peaks in the HNCA projection as the number of residues excluding proline. However, in the HNCA spectrum only about 200 peaks were seen, the majority of which correspond to Cα-HN resources within the same residue. A quick estimation of the completeness of the data can also be made: the Cα chemical shifts of glycines possesses unique chemical shifts around 44 ppm. Since there are 16 Glycines in ST6Gal-I, 16 peaks are expected to be seen in that region. However, only 6 are visible in the spectrum. To make sequential assignment, the HNCOCA experiment as well as the HNCO experiment has been collected. The HNCOCA correlates exclusively the Cα of the residue with HN of the next residue. Thus, a combination of HNCA and HNCOCA will allow the assignment of the ST6Gal-I backbone atom an attainable possibility. However, the data from these experiments is yet to be analyzed to reach this very goal.

A comparison of the 2D (^{1}H-^{15}N)-HSQC spectrum of deglycosylated ST6Gal-I from the ST6Gal-I construct and the 2D (^{1}H-^{15}N)-TROCY spectrum of deglycosylated and TEV-protease digested ST6Gal-I enzyme from MBP-ST6Gal-I construct showed a significant overlap of the peaks. This indicates that the fusion protein did not alter any structural characteristics of the

enzyme after cleavage of the MBP component. Moreover, the use of the MBP fusion in multiple small scale tube cultures makes the process of isotope labeling more laborious, but much more economical at the same time.

Surface Plasmon Resonance (SPR) Analysis

Binding affinity of CMP-Neu5Ac, the donor substrate and LacNAc, the small acceptor analog, was examined by surface plasmon resonance (SPR). The main purpose of the experiments was to establish the conditions to study and generate on- and off-rates for binding of donor (CMP-Neu5Ac) and acceptor (LacNAc) substrate for recombinant ST6Gal-I and compare these values with enzyme kinetics. However, we could not detect any real binding with either of the substrate as the analyte in any of the running buffers tested, including the HBS-EP commercial buffer or in any of the custom running buffers made from a combination of 20 mM MES or HEPES, with 150 mM NaCl, pH 6.5 or 7.4, with or without EDTA and P20 (data not shown). Similarly, immobilization using amine coupling in the presence of donor substrate (CMP-Neu5Ac), acceptor substrate (LacNac), and inhibitor CMP-3F-Sia, alone or in combination, gave us similar results as the enzyme immobilized alone.

In all our experiments, the only promising binding results seen were with the acceptor (LacNAc) binding in a very low salt concentrations, less than 25 mM NaCl or salt free buffers, similar to that previously reported for the acceptor, asialofetuin [171]. However, the binding kinetic parameters did not match with the enzyme kinetic parameters; they differed by at least two orders of magnitude, suggesting a possible non-specific interaction with the acceptor at low salt concentration.

Similarly, we did not see any binding when glycopeptides, obtained by pepsin digest of asialofetuin, were immobilized on the chip surface and ST6Gal-I was injected in running buffers containing 150 mM NaCl.

In order to check whether the enzyme was still active after immobilization, ST6Gal-I was immobilized using the same chemistry as that of SPR immobilization on Affigel 10, pre-activated with NHS-EDC chemistry at 3 different pHs (4.5, 5.5 and 6.5). The ST6Gal-I enzyme assays were performed on the supernatant solution after immobilization, the wash solution after immobilization, and the immobilized beads. There was not much immobilization of recombinant ST6Gal-I at pH 4.5 and pH 5.5 as there was more than 70% of the input enzyme activity found in the supernatant and little or no activity associated with the immobilized Affigel 10 beads. In addition the sum of the supernatant, wash, and bead immobilized activity did not add up to 100% in all four conditions the (i.e. two pHs with or without donor substrates). At pH 6.5, which is more suitable for amine coupling using NHS-EDC chemistry, there was reduced enzyme activity in the supernatant, wash, and on the immobilized gel beads. The enzyme activity associated with the immobilized beads in presence of CMP-Neu5Ac was slightly greater than when the immobilization conditions were carried out without CMP-Neu5Ac, indicating that the immobilization of ST6Gal-I using amine coupling results in loss of enzyme activity.

In order to support our hypothesis that immobilization eliminated catalytic activity of the ST6Gal-I, we carried out another set of experiments. We decided to biotinylate the enzyme using the same NHS-EDC chemistry in solution using a pre-activated biotin reagent. As expected there was no detectable ST6Gal-I enzyme activity following biotinylation. This would suggest that immobilization of ST6Gal-I using amine coupling via available free amino-groups resulted in the

inactivation of ST6Gal-I. However, previous reports suggested that blocking of the free primary amine groups by reductive methylation would have no effect on catalysis by ST6Gal-I [172].

However, in our hands reductive lysine methylation of the MBP-ST6Gal-I fusion protein yields an enzyme with a compromised ST6Gal-I activity. Even though cleavage of the methylated MBP fusion protein with TEV protease recovers some of the enzyme activity, biotinylation of methylated sample with NHS-EDC chemistry resulted in an inactive enzyme. These results suggest that immobilization of ST6Gal-I on the surface of a SPR chip or resin bead yields an inactive enzyme. The probable explanation is that any such chemical modification of ST6Gal-I restrict the confirmation dynamics of the enzyme induced by substrate binding, which is required to detect any binding events of the donor or substrate molecule by SPR.

Crystallization of ST6Gal-I

We have used several standard screening conditions for the crystallization trails. There were several conditions in preliminary screening, which showed a protein precipitation. Further screening around those conditions never demonstrated productive crystrallization. The discrepancy in these results could be attributed to the NH_2-terminal heterogeneity. As mentioned earlier, at any given point following purification the sample contained three different peptides. Analysis of those samples suggested that there was a change in the ratio of these terminal peptide sequences over time, so if a given condition showed a precipitation with one ratio of protein tremini, it might not consistently show the same precipitation at a different time where a different ratio would be found. Expression of the NH_2 terminal MBP-fusion protein should be beneficial for future studies as it may eliminate the problem of NH_2-terminal heterogeneity.

Expression of sugar-free ST6Gal-I

It has been previously established that the N-glycans of the ST6Gal-I are not essential for enzyme activity *in vivo* or *in vitro*. However, they are necessary for proper folding and transportation of the enzyme through secretary pathway [66]. It has also been demonstrated that the function of N-glycans in glycosyltranferases can be substituted by a single amino acids [69]. Moreover, the N-glycosylation sites amongst the ST6Gal-I isoforms from various species are not conserved. We decided to take advantage of these observations to generate a sugar free ST6Gal-I mutant protein (S148R-T159K-T160L) that was created using the SUNGA method and by replacing the glycosylation acceptor sequence. The ST6Gal-I activity from transformants containing the sugar-free mutant coding region was less than 10% compared to the wild type sequence transformants. At the same time the expression level of the protein was less than 10%, suggesting that the sugar free-mutant was equally active if expressed as the same level. The lower expression level of the sugar free mutant may result from higher proteolysis of the enzyme in the absence N-glycan structures. Now we are attempting to express the sugar-free mutant protein in bacteria. If successful, it will be a great step towards economical isotope labeling of ST6Gal-I and thus subsequent structure determination.

Conclusion

We have successfully expressed a functional soluble catalytic domain of rat ST6Gal-I gene using the yeast expression system *P. pastoris* which displays properties similar to those described for the native enzymes. We have successfully purified this enzyme with a yield of a few milligrams per liter of the medium.

Even though we did not achieve the overall objective of solving the structure of rat ST6Gal-I, we have achieved several milestones which were never achieved before. First, the main hurdle in the scale-up of recombinant protein expression in *P. pastoris* was the steady state proteolysis of recombinant ST6Gal-I by the secreted proteases. We were able to address this issue by two different approaches 1) use of a selective combination of protease inhibitors in fermentation and and purification which improved the yields considerably, and 2) using a fusion protein construct which gave a significantly greater enzyme yield and reduced N- terminal heterogeneity.

Second, even though we were not able to obtain the crystals of ST6Gal-I, we have established a few lead conditions which were promising and with few refinements it could result in protien crystalization. A major limitation in any crystallization trial is achieving a homogenious sample. In eukaryotic glycoproteins, glycan structures are always expressed with microheterogeneity, which we have been able to control by ferforming deglycosylation with endo H_f. However, in all crystallization trials, we have had additional NH_2-terminal protein heterogeneity as a consequence of proteolysis during fermentation and purification. We have now successfully removed the NH_2-terminal heterogeneity by using fusion protein constructs. This reduction in NH_2-terminal heterogeneity in combination with enzymatic deglycosylation may allow the isolation of recombinant protein compatible with X-ray crystallography studies.

Third, we have also successfully optimized the fermentation conditions for the expression of ST6Gal-I in a form suitable for uniform stable isotope labeling, which is the foundation for structural determination of proteins by NMR spectroscopy. Even with problems of proteolysis and isotope effect, reducing expression levels of ST6Gal-I, we were successful in producing recombinant enzyme that had a significant incorporation of all three isotopes (^{2}H, ^{13}C, and ^{15}N) for NMR studies. However, the low yields and large quantities of expensive reagents required for the expression of triple-labeled ST6Gal-I using the fermentor was discouraging for the routine preparation of samples necessary for structural elucidation by NMR. We have also been successful in expressing a double-labeled (^{2}H and ^{15}N) fusion protein (MBP-ST6Gal-I) in tube cultures rather than in the fermentor. Preliminary NMR data on ST6Gal-I from a TEV protease digested preparation of MBP-ST6Gal-I suggested that the enzyme had the same overall structure as that of the untagged enzyme and the cost of expression is significantly reduced. We are currently exploring the conditions for triple-labeling of this fusion protein. We assume that this latter strategy will be an effective approach to solve the structure of ST6Gal-I.

Fourth, we have clearly demonstrated that the current methods, which utilize the covalent linking of ST6Gal-I on the SPR chip surface, are not adequate for substrate binding studies, as any such chemical modification renders this enzyme inactive. In light of these results, a serious question is raised regarding previous reports on subsrate binding studies using SPR. These results imply that a new approach is required for substrate binding studies using SPR.

Fifth, we have successfully demonstrated the application of the SUNGA method for expression of the deglycosylated form of ST6Gal-I in *P. pastoris* and even in *E. coli*. Even though the expression levels of sugar free ST6Gal-I was reduced compared to glycosylated form when expressed in *Pichia*, it was a soluble and active. The low yields may be because of

increase proteolysis in abasence of glycans in the *Pichia* host system. These results are highly promising and now we are exploring the possibility of expressing these proteins in bacterial expression system, which if successful will reduce the cost of isotope labeling enormously. As a result of our generation of labeled ST6Gal-I in *Pichia* and the preliminary NMR data obtained from that set of samples, we are now in position to evaluate the utility of the bacterial expression system in in generating samples suitable for structural studies.

REFRENCES

1. Gagneux, P. and Varki, A., *Evolutionary Considerations in Relating Oligosaccharide Diversity to Biological Function.* Glycobiology, 1999. **9**. 747-755.

2. Coutinho, P. M., Deleury, E., Davies, G. J. and Henrissat, B., *An Evolving Hierarchical Family Classification for Glycosyltransferases.* J. Mol. Biol., 2003. **328**. 307-317.

3. Unligil, U. M. and Rini, J. M., *Glycosyltransferase Structure and Mechanism.* Curr. Opin. Struct. Biol., 2000. **10**. 510-517.

4. Breton, C., Mucha, J. and Jeanneau, C., *Structural and Functional Features of Glycosyltransferases.* Biochimie, 2001. **83**. 713-718.

5. Vimr, E. R., Kalivoda, K. A., Deszo, E. L. and Steenbergen, S. M., *Diversity of Microbial Sialic Acid Metabolism.* Microbiol. Mol. Biol. Rev., 2004. **68**. 132-153.

6. Faillard, H., *The Early History of Sialic Acids.* Trends Biochem. Sci., 1989. **14**. 237-241.

7. Schauer, R., *Achievements and Challenges of Sialic Acid Research.* Glycoconjugate J., 2000. **17**. 485-499.

8. Varki, A., *Sialic Acids*, in *Essentials of Glycobiology*, A. Varki, Cummings R, Esko, J, Freeze, H, Hart G, Marth, J, Editor. 1999, CSHL Press. p. 195-210.

9. Schauer, R., *Sialic Acids: Fascinating Sugars in Higher Animals and Man.* Zoology, 2004. **107**. 49-64.

10. Schauer, R., Kelm, S, Reuter, G, Roggentin, P, and Shaw, L, *Biochemistry and Role of Sialic Acids*, in *Biology of Sialic Acids*, A. Rosenberg, Editor. 1995, Plenum Press: New York. p. 7-16

11. Nadano, D., Iwasaki, M., Endo, S., Kitajima, K., Inoue, S. and Inoue, Y., *A Naturally Occurring Deaminated Neuraminic Acid, 3-Deoxy-D-Glycero-D- Galacto-Nonulosonic Acid (KDN). Its Unique Occurrence at the Nonreducing Ends of Oligosialyl Chains in Polysialoglycoprotein of Rainbow Trout Eggs.* J. Biol. Chem., 1986. **261**. 11550-11557.

12. Cornfield, A., P Schauer, R, *Sialic Acids - Chemistry, Metabolism and Function*, ed. R. Schauer. 1982, New York: Springer Verlag Wien. 5-55.

13. Varki, A., *Diversity in the Sialic Acids* Glycobiology, 1992. **2**. 25-40.

14. Angata, T. and Varki, A., *Chemical Diversity in the Sialic Acids and Related α-Keto Acids: An Evolutionary Perspective.* Chem. Rev., 2002. **102**. 439-470.

15. Malykh, Y. N., Schauer, R. and Shaw, L., *N-Glycolylneuraminic Acid in Human Tumours.* Biochimie, 2001. **83**. 623-634.

16. Varki, A., *N-Glycolylneuraminic Acid Deficiency in Humans.* Biochimie, 2001. **83**. 615-622.

17. Hayakawa, K., De Felice, C., Watanabe, T., Tanaka, T., Iinuma, K., Nihei, K., Higuchi, S., Ezoe, T., Hibi, I. and Kurosawa, K., *Determination of Free N-Acetylneuraminic Acid in Human Body Fluids by High-Performance Liquid Chromatography with Fluorimetric Detection.* J. Chromatogr. (B), 1993. **620**. 25-31.

18. Siskos, P. A. and Spyridaki, M.-H. E., *Determination of Sialic Acids in Biological Fluids Using Reversed-Phase Ion-Pair High-Performance Liquid Chromatography.* J. Chromatogr. (B), 1999. **724**. 205-212.

19. Traving, C. and Schauer, R., *Structure, Function and Metabolism of Sialic Acids.* Cell. Mol. Life Sci., 1998. **54**. 1330-1349.

20. Schauer, R., *Victor Ginsburg's Influence on My Research of the Role of Sialic Acids in Biological Recognition.* Arch. Biochem. Biophys., 2004. **426**. 132-141.

21. Schauer, R., *Sialic Acids Regulate Cellular and Molecular Recognition*, in *Carohydrates Synthetic Methods and Applications in Medicinal Chemistry*, H. Ogura, Hasegawa, A, Suami, T, Editor. 1992, Kodnsha Ltd: Tokyo. p. 340-354.

22. Schauer, R., Kamerling, J. P., J. Montreuil, J. F. G. V. and Schachter, H., *Biochemistry and Biology of Sialic Acids*, in *New Comprehensive Biochemistry*. 1997, Elsevier. p. 243-402.

23. Ashwell, G. and Morell, A.G., *The Role of Surface Carbohydrates in the Hepatic Recognition and Transport of Circulating Glycoproteins.* Adv. Enzymol., 1974. **41**. 99-128.

24. Klunge, A., Reuter, G, Lee, H, Ruch-Heeger, B, Schauer R, *Interaction of Rat Peritoneal Macrophages with Homologous Sialidase-Treated Thrombocytes in Vitro: Biochemical and Morphological Studies. Detection of N-(O-Acteyl)Glycoloylneuraminic Acid.* . Eur. J. Cell Biol., 1992. **59**. 12-20.

25. Fischer, C., Kelm, S., Ruch, B. and Schauer, R., *Reversible Binding of Sialidase-Treated Rat Lymphocytes by Homologous Peritoneal Macrophages.* Carbohydrate Research, 1991. **213**. 263-273.

26. Corfield, T., *Bacterial Sialidases - Roles in Pathogenicity and Nutrition.* Glycobiology, 1992. **2**. 509-521.

27. Shen, R., Su, Z.-z., Olsson, C. A., Goldstein, N. I. and Fisher, P. B., *Surface-Epitope Masking: A Strategy for the Development of Monoclonal Antibodies Specific for Molecules Expressed on the Cell Surface.* J. Natl. Cancer Inst., 1994. **86**. 91-98.

28. Dall'Olio, F. and Chiricolo, M., *Sialyltransferases in Cancer.* Glycoconjugate J., 2001. **18**. 841-850.

29. Hirst, G. K., *Adsorption of Influenza Hemagglutinins and Virus by Red Blood Cells.* J. Exp. Med., 1942. **76**. 195-209.

30. Rutishauser, U. and Landmesser, L., *Polysialic Acid in the Vertebrate Nervous System: A Promoter of Plasticity in Cell-Cell Interactions.* Trends Neurosci., 1996. **19**. 422-427.

31. Angata, K. and Fukuda, M., *Polysialyltransferases: Major Players in Polysialic Acid Synthesis on the Neural Cell Adhesion Molecule.* Biochimie, 2003. **85**. 195-206.

32. Barbeau, D., Liang, J. J., Robitalille, Y., Quirion, R. and Srivastava, L. K., *Decreased Expression of the Embryonic Form of the Neural Cell Adhesion Molecule in Schizophrenic Brains.* PNAS, 1995. **92**. 2785-2789.

33. Mikkonen, M., Soininen, H., Alafuzoff, I. and Miettinen, R., *Hippocampal Plasticity in Alzheimer's Disease.* Rev. Neurosci., 2001. **12**. 311.

34. Crocker, P. R., *Siglecs: Sialic-Acid-Binding Immunoglobulin-Like Lectins in Cell-Cell Interactions and Signalling.* Curr. Opin Struct. Biol., 2002. **12**. 609-615.

35. Nitschke, L., *The Role of CD22 and Other Inhibitory Co-receptors in B-Cell Activation.* Curr. Opin. Immunol., 2005. **17**. 290-297.

36. Collins, B. E., Smith, B. A., Bengtson, P. and Paulson, J. C., *Ablation of CD22 in Ligand-Deficient Mice Restores B Cell Receptor Signaling.* Nat Immunol., 2006. **7**. 199-206.

37. Grewal, P. K., Boton, M., Ramirez, K., Collins, B. E., Saito, A., Green, R. S., Ohtsubo, K., Chui, D. and Marth, J. D., *ST6Gal-I Restrains CD22-Dependent Antigen Receptor Endocytosis and Shp-1 Recruitment in Normal and Pathogenic Immune Signaling.* Mol. Cell. Biol., 2006. **26**. 4970-4981.

38. Brian, E. C., Ola, B., Nicolai, V. B., Claus-Peter, D., Daniel, C., Jamey, D. M., Lars, N. and James, C. P., *Constitutively Unmasked CD22 on B Cells of ST6Gal I Knockout Mice: Novel Sialoside Probe for Murine CD22.* Glycobiology, 2002. **12**. 563-571.

39. Mingorance, A., Fontana, X., Soriano, E. and del Río, J. A., *Overexpression of Myelin-Associated Glycoprotein after Axotomy of the Perforant Pathway.* Mol. Cell. Neurosci., 2005. **29**. 471-483.

40. Strehle, *Sialic Acid Storage Disease and Related Disorders.* Genetic Testing 2003. **7**. 113-121.

41. Schauer, R., *Sialic Acids and Their Role as Biological Masks.* Trends Biochem. Sci., 1985. **10**. 357-360.

42. Harduin-Lepers, A., Recchi, M.-A. and Delannoy, P., *1994, the Year of Sialyltransferases.* Glycobiology, 1995. **5**. 741-758.

43. Harduin-Lepers, A., Vallejo-Ruiz, V., Krzewinski-Recchi, M.-A., Samyn-Petit, B., Julien, S. and Delannoy, P., *The Human Sialyltransferase Family.* Biochimie, 2001. **83**. 727-737.

44. Tsuji, S., Datta, A. K. and Paulson, J. C., *Letters to the Glyco-Forum: Systematic Nomenclature for Sialyltransferases.* Glycobiology, 1996. **6**. 647-647.

45. Lau, J. T. Y. and Wuensch, S. A., in *Carbohydrates in Chemistry and Biology*, B. Ernst, Hart, G., W, and Sinay, P., Editors. 2000. p. 213.

46. Tsuji, S., *Molecular Cloning and Functional Analysis of Sialyltransferases.* J Biochem, 1996. **120**. 1-13.

47. Datta, A. K. and Paulson, J. C., *The Sialyltransferase ``Sialylmotif'' Participates in Binding the Donor Substrate CMP-NeuAc.* J. Biol. Chem., 1995. **270**. 1497-1500.

48. Datta, A. K., Sinha, A. and Paulson, J. C., *Mutation of the Sialyltransferase S-Sialylmotif Alters the Kinetics of the Donor and Acceptor Substrates.* J. Biol. Chem., 1998. **273**. 9608-9614.

49. Jeanneau, C., Chazalet, V., Auge, C., Soumpasis, D. M., Harduin-Lepers, A., Delannoy, P., Imberty, A. and Breton, C., *Structure-Function Analysis of the Human Sialyltransferase ST3Gal I: Role of N-Glycosylation and a Novel Conserved Sialylmotif.* J. Biol. Chem., 2004. **279**. 13461-13468.

50. Chiu, C. P. C., Watts, A. G., Lairson, L. L., Gilbert, M., Lim, D., Wakarchuk, W. W., Withers, S. G. and Strynadka, N. C. J., *Structural Analysis of the Sialyltransferase CstII from Campylobacter Jejuni in Complex with a Substrate Analog.* Nat. Struct. Mol. Biol., 2004. **11**. 163-170.

51. Ni, L., Chokhawala, H. A., Cao, H., Henning, R., Ng, L., Huang, S., Yu, H., Chen, X. and Fisher, A. J., *Crystal Structures of Pasteurella Multocida Sialyltransferase Complexes with Acceptor and Donor Analogues Reveal Substrate Binding Sites and Catalytic Mechanism.* Biochem., 2007. **46**. 6288-6298.

52. Datta, A. K., Chammas, R. and Paulson, J. C., *Conserved Cysteines in the Sialyltransferase Sialylmotifs Form an Essential Disulfide Bond.* J. Biol. Chem., 2001. **276**. 15200-15207.

53. Angata, K., Yen, T.-Y., El-Battari, A., Macher, B. A. and Fukuda, M., *Unique Disulfide Bond Structures Found in ST8Sia IV Polysialyltransferase Are Required for Its Activity.* J. Biol. Chem., 2001. **276**. 15369-15377.

54. Ma, J. and Colley, K. J., *A Disulfide-Bonded Dimer of the Golgi β-Galactoside α2,6-Sialyltransferase Is Catalytically Inactive yet Still Retains the Ability to Bind Galactose.* J. Biol. Chem., 1996. **271**. 7758-7766.

55. Qian, R., Chen, C. and Colley, K. J., *Location and Mechanism of α2,6-Sialyltransferase Dimer Formation. Role of Cysteine Residues in Enzyme Dimerization, Localization, Activity, and Processing.* J. Biol. Chem., 2001. **276**. 28641-28649.

56. Fukumoto, S., Miyazaki, H., Goto, G., Urano, T., Furukawa, K. and Furukawa, K., *Expression Cloning of Mouse cDNA of CMP-NeuAc:Lactosylceramide α2,3-Sialyltransferase, an Enzyme That Initiates the Synthesis of Gangliosides.* J. Biol. Chem., 1999. **274**. 9271-9276.

57. Hu, Y. and Walker, S., *Remarkable Structural Similarities between Diverse Glycosyltransferases.* Chem.Biol., 2002. **9**. 1287-1296.

58. Qasba, P. K., Ramakrishnan, B. and Boeggeman, E., *Substrate-Induced Conformational Changes in Glycosyltransferases.* Trends Biochem. Sci., 2005. **30**. 53-62.

59. Ramakrishnan, B., Balaji, P. V. and Qasba, P. K., *Crystal Structure of β1,4-Galactosyltransferase Complex with UDP-Gal Reveals an Oligosaccharide Acceptor Binding Site.* J. Mol. Biol., 2002. **318**. 491.

60. Unligil, U. M., Zhou, S., Yuwaraj, S., Sarkar, M., Schachter, H., Rini, J. M., *X-Ray Crystal Structure of Rabbit N-Acetylglucosaminyltransferase I: Catalytic Mechanism and a New Protein Superfamily.* The EMBO J., 2000. **19**. 5269-80.

61. Patenaude, S. I., Seto, N. O. L., Borisova, S. N., Szpacenko, A., Marcus, S. L., Palcic, M. M. and Evans, S. V., *The Structural Basis for Specificity in Human ABO(H) Blood Group Biosynthesis.* Nat. Struct. Mol. Biol., 2002. **9**. 685-690.

62. Breton, C., Heissigerova, H., Jeanneau, C., MoravcovÃ¡, J. and Imberty, A., *Comparative Aspects of Glycosyltransferases.* Biochem. Soc. Symp., 2002. 23-32.

63. Bruner, M. and Horenstein, B. A., *Use of an Altered Sugar-Nucleotide to Unmask the Transition State for α(2→6) Sialyltransferase.* Biochem., 2000. **39**. 2261-2268.

64. Horenstein, B. A. and Bruner, M., *The N-Acetyl Neuraminyl Oxecarbenium Ion Is an Intermediate in the Presence of Anionic Nucleophiles.* J. Am. Chem. Soc., 1998. **120**. 1357-1362.

65. Horenstein, B. A. and Bruner, M., *Quantum Mechanical Analysis of an Carboxylate-Substituted Oxocarbenium Ion. Isotope Effects for Formation of the Sialyl Cation and the Origin of an Unusually Large Secondary ^{14}C Isotope Effect.* J. Am. Chem. Soc., 1997. **119**. 1101-1107.

66. Chen, C. and Colley, K. J., *Minimal Structural and Glycosylation Requirements for St6gal I Activity and Trafficking.* Glycobiology, 2000. **10**. 531-583.

67. Vallejo-Ruiz, V., Haque, R., Mir, A.-M., Schwientek, T., Mandel, U., Cacan, R., Delannoy, P. and Harduin-Lepers, A., *Delineation of the Minimal Catalytic Domain of*

Human Galβ1-3galnac α2,3-Sialyltransferase (hST3Gal I). Biochim Biophys. Acta, 2001. **1549**. 161-173.

68. Martina, J. A., Daniotti, J. L. and Maccioni, H. J. F., *Influence of N-Glycosylation and N-Glycan Trimming on the Activity and Intracellular Traffic of GD3 Synthase.* J. Biol. Chem., 1998. **273**. 3725-3731.

69. Uemura, S., Kurose, T., Suzuki, T., Yoshida, S., Ito, M., Saito, M., Horiuchi, M., Inagaki, F., Igarashi, Y. and Inokuchi, J.-i., *Substitution of the N-Glycan Function in Glycosyltransferases by Specific Amino Acids: ST3Gal-V as a Model Enzyme.* Glycobiology, 2006. **16**. 258-270.

70. Takashima, S., Tsuji, S. and Tsujimoto, M., *Comparison of the Enzymatic Properties of Mouse β-Galactoside α2,6-Sialyltransferases, ST6Gal I and II.* J Biochem, 2003. **134**. 287-296.

71. Takashima, S., Tsuji, S. and Tsujimoto, M., *Characterization of the Second Type of Human β-Galactoside α2,6-Sialyltransferase (ST6Gal II), Which Sialylates Galβ1,4GlcNAc Structures on Oligosaccharides Preferentially. Genomic Analysis of Human Sialyltransferase Genes.* J. Biol. Chem., 2002. **277**. 45719-45728.

72. Takashima, S., *Characterization of Mouse Sialyltransferase Genes: Their Evolution and Diversity.* Biosci. Biotechnol. Biochem., 2008. **72**. 1155-1167.

73. Hennet, T., Chui, D., Paulson, J. C. and Marth, J. D., *Immune Regulation by the ST6Gal Sialyltransferase.* PNAS, 1998. **95**. 4504-4509.

74. Ikehara, Y., Kojima, N., Kurosawa, N., Kudo, T., Kono, M., Nishihara, S., Issiki, S., Morozumi, K., Itzkowitz, S., Tsuda, T., Nishimura, S. I., Tsuji, S. and Narimatsu, H., *Cloning and Expression of a Human Gene Encoding an N- Acetylgalactosamine-α2,6-Sialyltransferase (ST6GalNAc I): A Candidate for Synthesis of Cancer-Associated Sialyl-Tn Antigens.* Glycobiology, 1999. **9**. 1213-1224.

75. Samyn-Petit, B., Krzewinski-Recchi, M.-A., Steelant, W. F. A., Delannoy, P. and Harduin-Lepers, A., *Molecular Cloning and Functional Expression of Human ST6GalNAc II. Molecular Expression in Various Human Cultured Cells.* Biochim. Biophys. Acta, 2000. **1474**. 201-211.

76. Kono, M., Tsuda, T., Ogata, S., Takashima, S., Liu, H., Hamamoto, T., Itzkowitz, S. H., Nishimura, S. and Tsuji, S., *Redefined Substrate Specificity of ST6GalNAc II: A Second Candidate Sialyl-Tn Synthase.* Biochem.Biophys. Res. Commun., 2000. **272**. 94-97.

77. Lee, Y.-C., Kaufmann, M., Kitazume-Kawaguchi, S., Kono, M., Takashima, S., Kurosawa, N., Liu, H., Pircher, H. and Tsuji, S., *Molecular Cloning and Functional Expression of Two Members of Mouse NeuAcα2,3galβ1,3GalNAc GalNAcα2,6-Sialyltransferase Family, ST6GalNAc III and IV.* J. Biol. Chem., 1999. **274**. 11958-11967.

78. Ikehara, Y., Shimizu, N., Kono, M., Nishihara, S., Nakanishi, H., Kitamura, T., Narimatsu, H., Tsuji, S. and Tatematsu, M., *A Novel Glycosyltransferase with a Polyglutamine Repeat; a New Candidate for GD1α Synthase (ST6GalNAc V).* FEBS Lett., 1999. **463**. 92-96.

79. Okajima, T., Chen, H.-H., Ito, H., Kiso, M., Tai, T., Furukawa, K., Urano, T. and Furukawa, K., *Molecular Cloning and Expression of Mouse GD1α/GT1aα /GQ1bα Synthase (ST6GalNAc VI) Gene.* J. Biol. Chem., 2000. **275**. 6717-6723.

80. Okajima, T., Fukumoto, S., Miyazaki, H., Ishida, H., Kiso, M., Furukawa, K., Urano, T. and Furukawa, K., *Molecular Cloning of a Novel α2,3-Sialyltransferase (ST3Gal VI) That Sialylates Type II Lactosamine Structures on Glycoproteins and Glycolipids.* J. Biol. Chem., 1999. **274**. 11479-11486.

81. Lee, Y. C., Kojima, N., Wada, E., Kurosawa, N., Nakaoka, T., Hamamoto, T. and Tsuji, S., *Cloning and Expression of cDNA for a New Type of Galβ1,3GalNAcα2,3-Sialyltransferase.* J. Biol. Chem., 1994. **269**. 10028-10033.

82. Kono, M., Ohyama, Y., Lee, Y.-C., Hamamoto, T., Kojima, N. and Tsuji, S., *Mouse β-Galactoside α2,3-Sialyltransferases: Comparison of in Vitro Substrate Specificities and Tissue Specific Expression.* Glycobiology, 1997. **7**. 469-479.

83. Priatel, J. J., Chui, D., Hiraoka, N., Simmons, C. J. T., Richardson, K. B., Page, D. M., Fukuda, M., Varki, N. M. and Marth, J. D., *The ST3Gal-I Sialyltransferase Controls $CD8^+$ T Lymphocyte Homeostasis by Modulating O-Glycan Biosynthesis.* Immunity, 2000. **12**. 273-283.

84. Sasaki, K., Watanabe, E., Kawashima, K., Sekine, S., Dohi, T., Oshima, M., Hanai, N., Nishi, T. and Hasegawa, M., *Expression Cloning of a Novel Gal β(1-3/1-4) GlcNAc α 2,3- Sialyltransferase Using Lectin Resistance Selection.* J. Biol. Chem., 1993. **268**. 22782-22787.

85. Preuss, U., Gu, X., Gu, T. and Yu, R. K., *Purification and Characterization of CMP-N-Acetylneuraminic Acid:Lactosylceramide (α2-3) Sialyltransferase (GM3-Synthase) from Rat Brain.* J. Biol. Chem., 1993. **268**. 26273-26278.

86. Powell, L. D., Sgroi, D., Sjoberg, E. R., Stamenkovic, I. and Varki, A., *Natural Ligands of the B Cell Adhesion Molecule CD22 Beta Carry N- Linked Oligosaccharides with α-2,6-Linked Sialic Acids That Are Required for Recognition.* J. Biol. Chem., 1993. **268**. 7019-7027.

87. Sgroi, D., Varki, A., Braesch-Andersen, S. and Stamenkovic, I., *CD22, a B Cell-Specific Immunoglobulin Superfamily Member, Is a Sialic Acid-Binding Lectin.* J. Biol. Chem., 1993. **268**. 7011-7018.

88. Powell, L. D. and Varki, A., *The Oligosaccharide Binding Specificities of CD22 Beta, a Sialic Acid- Specific Lectin of B Cells.* J. Biol. Chem., 1994. **269**. 10628-10636.

89. Doody, G. M., Justement, L. B., Delibrias, C. C., Matthews, R. J., Lin, J., Thomas, M. L. and Fearon, D. T., *A Role in B Cell Activation for CD22 and the Protein Tyrosine Phosphatase Shp.* Science, 1995. **269**. 242-244.

90. O'Keefe, T. L., Williams, G. T., Davies, S. L. and Neuberger, M. S., *Hyperresponsive B Cells in CD22-Deficient Mice.* Science, 1996. **274**. 798-801.

91. Schulte, R. J., Campbell, M. A., Fischer, W. H. and Sefton, B. M., *Tyrosine Phosphorylation of CD22 During B Cell Activation.* Science, 1992. **258**. 1001-1004.

92. Crocker, P. R., Paulson, J. C. and Varki, A., *Siglecs and Their Roles in the Immune System.* Nat. Rev. Immunol., 2007. **7**. 255-266.

93. Tedder, T. F., Poe, J. C. and Haas, K. M., *Cd22: A Multifunctional Receptor That Regulates B Lymphocyte Survival and Signal Transduction.* Adv. Immunol., 2005. **88**. 1-50.

94. Campana, D., Janossy, G., Bofill, M., Trejdosiewicz, L. K., Ma, D., Hoffbrand, A. V., Mason, D. Y., Lebacq, A. M. and Forster, H. K., *Human B Cell Development. I. Phenotypic Differences of B Lymphocytes in the Bone Marrow and Peripheral Lymphoid Tissue.* J. Immunol., 1985. **134**. 1524-1530.

95. van der Merwe, P. A., Crocker, P. R., Vinson, M., Barclay, A. N., Schauer, R. and Kelm, S., *Localization of the Putative Sialic Acid-Binding Site on the Immunoglobulin Superfamily Cell-Surface Molecule CD22.* J. Biol. Chem., 1996. **271**. 9273-9280.

96. Collins, B. E., Blixt, O., Han, S., Duong, B., Li, H., Nathan, J. K., Bovin, N. and Paulson, J. C., *High-Affinity Ligand Probes of CD22 Overcome the Threshold Set by Cis Ligands to Allow for Binding, Endocytosis, and Killing of B Cells.* J. Immunol., 2006. **177**. 2994-3003.

97. Cornall, R. J., Cyster, J. G., Hibbs, M. L., Dunn, A. R., Otipoby, K. L., Clark, E. A. and Goodnow, C. C., *Polygenic Autoimmune Traits: Lyn, CD22, and Shp-1 Are Limiting Elements of a Biochemical Pathway Regulating BCR Signaling and Selection.* Immunity, 1998. **8**. 497-508.

98. Tedder, T. F., Tuscano, J., Sato, S. and Kehrl, J. H., *CD22, a B Lymphocyte-Specific Adhesion Molecule That Regulates Antigen Receptor Signaling.* Annu. Rev. Immunol., 1997. **15**. 481-504.

99. Ghosh, S., Bandulet, C. and Nitschke, L., *Regulation of B Cell Development and B Cell Signalling by CD22 and Its Ligands {Alpha}2,6-Linked Sialic Acids.* Int. Immunol., 2006. **18**. 603-611.

100. Poe Jonathan C , F. Y., Hasegawa Minoru, Haas Karen M, Miller Ann S, Sanford Isaac G, Bock Cheryl B, Fujimoto Manabu, Tedder Thomas F *CD22 Regulates B Lymphocyte Function in Vivo through Both Ligand-Dependent and Ligand-Independent Mechanisms.* Nat. Immunol., 2004. **5**. 1078-1087.

101. Danzer, C.-P., Collins, B. E., Blixt, O., Paulson, J. C. and Nitschke, L., *Transitional and Marginal Zone B Cells Have a High Proportion of Unmasked CD22: Implications for BCR Signaling.* Int. Immunol., 2003. **15**. 1137-1147.

102. Brian, E. C., Ola, B., Alexis, R. D., Nicolai, B., Jamey, D. M. and James, C. P., *Masking of CD22 by Cis Ligands Does Not Prevent Redistribution of CD22 to Sites of Cell Contact.* PNAS, 2004. **101**. 6104-9.

103. Stoddart, A., Dykstra, M. L., Brown, B. K., Song, W., Pierce, S. K. and Brodsky, F. M., *Lipid Rafts Unite Signaling Cascades with Clathrin to Regulate BCR Internalization.* Immunity, 2002. **17**. 451-462.

104. Stoddart, A., Jackson, A. P. and Brodsky, F. M., *Plasticity of B Cell Receptor Internalization Upon Conditional Depletion of Clathrin.* Mol. Biol. Cell, 2005. **16**. 2339-2348.

105. Kimura, N., Ohmori, K., Miyazaki, K., Izawa, M., Matsuzaki, Y., Yasuda, Y., Takematsu, H., Kozutsumi, Y., Moriyama, A. and Kannagi, R., *Human B-Lymphocytes Express α2-6-Sialylated 6-Sulfo-N-Acetyllactosamine Serving as a Preferred Ligand for CD22/Siglec-2.* J. Biol. Chem., 2007. **282**. 32200-32207.

106. Chotigeat, W., Chayanunnukul, W. and Phongdara, A., *Expression of a Mammalian α2,6-Sialyltransferase Gene in Pichia pastoris.* J. Biotechnol., 2000. **81**. 55-61.

107. Herscovics, A. and Orlean, P., *Glycoprotein Biosynthesis in Yeast.* FASEB J., 1993. **7**. 540-550.

108. Rachel Daly, M. T. W. H., *Expression of Heterologous Proteins in Pichia pastoris, A Useful Experimental Tool in Protein Engineering and Production.* J. Mol. Recognition, 2005. **18**. 119-138.

109. *Pichia Expression Kit : A Manual of Methods for Expression of Recombinant Proteins in Pichia pastoris Version M, Invitrogen Inc.* . 2002. p. 55-58.

110. Cereghino, J. L. and Cregg, J. M., *Heterologous Protein Expression in the Methylotrophic Yeast Pichia pastoris.* FEMS Microbiol. Rev., 2000. **24**. 45-66.

111. Zang Wenhui, I. M., Meagher Michael M, *Fermentation Strategies for Recombinant Protein Expression in the Methylotrophic Yeast Pichia pastoris.* Biotechnol. Bioprocess Engg., 2000. **5**. 275-287.

112. Lim, H. K., Choi, S. J., Kim, K. Y. and Jung, K. H., *Dissolved-Oxygen-Stat Controlling Two Variables for Methanol Induction of rGuamerin in Pichia pastoris and Its Application to Repeated Fed-Batch.* Appl. Microbiol. Biotechnol., 2003. **62**. 342-348.

113. Wenhui, Z., Mark, A. B., Bradley, A. P., Leonard, A. S. and Michael, M. M., *Modeling Pichia pastoris Growth on Methanol and Optimizing the Production of a Recombinant*

Protein, the Heavy-Chain Fragment C of Botulinum Neurotoxin, Serotype A. Biotechnol. Bioeng., 2000. **70**. 1-8.

114. País, J. M., Varas, L., Valdés, J., Cabello, C., Rodríguez, L. and Mansur, M., *Modeling of Mini-Proinsulin Production in Pichia pastoris Using the AOX Promoter.* Biotechnol. Lett., 2003. **25**. 251-255.

115. Zhang, W., Smith, L. A., Plantz, B. A., Schlegel, V. L. and Meagher, M. M., *Design of Methanol Feed Control in Pichia pastoris Fermentations Based Upon a Growth Model.* Biotechnol. Prog., 2002. **18**. 1392-1399.

116. http://www.ravenbiotech.com/products.html.

117. Zhou, X.-S., Lu, J., Fan, W.-M. and Zhang, Y.-X., *Development of a Responsive Methanol Sensor and Its Application in Pichia pastoris Fermentation.* Biotechnol. Lett., 2002. **24**. 643-646.

118. Cregg, J. M., Vedvick, T. S. and Raschke, W. C., *Recent Advances in the Expression of Foreign Genes in Pichia pastoris.* Nat. Biotech., 1993. **11**. 905-910.

119. Sreekrishna K, Barr K A and SA, H., *Expression of Human Serum Albumin in Pichia Pastroris*, in *15th Int. Congr. Yeast Genetics and Molecular Biology, Hague, Netherlands Yeast 6 (Special Issue)*, O. S. a. W. R, Editor. 1990, Wiley: New York. p. S447.

120. Bencúrová, M., Rendic, D., Fabini, G., Kopecky, E.-M., Altmann, F. and Wilson, I. B. H., *Expression of Eukaryotic Glycosyltransferases in the Yeast Pichia pastoris.* Biochimie, 2003. **85**. 413-422.

121. Pritchett, J. and Baldwin, S. A., *The Effect of Nitrogen Source on Yield and Glycosylation of a Human Cystatin C Mutant Expressed in Pichia pastoris.* J. Ind. Microbiol. Biotechnol., 2004. **31**. 553-558.

122. Inan, M., Chiruvolu, V., Eskridge, K. M., Vlasuk, G. P., Dickerson, K., Brown, S. and Meagher, M. M., *Optimization of Temperature-Glycerol-pH Conditions for a Fed-Batch Fermentation Process for Recombinant Hookworm (Ancylostoma Caninum) Anticoagulant Peptide (ACAP-5) Production by Pichia pastoris.* Enzyme Microb. Technol., 1999. **24**. 438-445.

123. Saelens, X., vanlandschoot, P., Martinet, W., Maras, M., Neirynck, S., Contreras, R., Fiers, W. and Jou, W. M., *Protection of Mice against a Lethal Influenza Virus Challenge after Immunization with Yeast-Derived Secreted Influenza Virus Hemagglutinin.* Eur. J. Biochem., 1999. **260**. 166-175.

124. Clare, J., Scorer, C., Buckholz, R. and Romanos, M., *Expression of EGF and HIV Envelope Glycoprotein*, in *Pichia Protocols*. Ed. D. R. Higgins and J. M. Cregg., 1998, Humana Press. p. 209-225.

125. Jahic, M., Gustavsson, M., Jansen, A.-K., Martinelle, M. and Enfors, S.-O., *Analysis and Control of Proteolysis of a Fusion Protein in Pichia pastoris Fed-Batch Processes.* J. Biotechnol., 2003. **102**. 45-53.

126. Jiang, F., *Expression of Recombinant Manganese Peroxidase in Pichia pastoris: Biosynthesis, Stability and Kinetics Analysis*. 2006, Syracuse University: United States - New York.

127. Cassland, P. and Jönsson, L. J., *Characterization of a Gene Encoding Trametes Versicolor Laccase A and Improved Heterologous Expression in Saccharomyces Cerevisiae by Decreased Cultivation Temperature.* Appl. Microbiol. Biotechnol., 1999. **52**. 393-400.

128. Whittaker, M. M. and Whittaker, J. W., *Expression of Recombinant Galactose Oxidase by Pichia pastoris.* Protein Expression Purif., 2000. **20**. 105-111.

129. Li, Z., Xiong, F., Lin, Q., d'Anjou, M., Daugulis, A. J., Yang, D. S. C. and Hew, C. L., *Low-Temperature Increases the Yield of Biologically Active Herring Antifreeze Protein in Pichia pastoris.* Protein Expression Purif., 2001. **21**. 438-445.

130. Brady, C. P., Shimp, R. L., Miles, A. P., Whitmore, M. and Stowers, A. W., *High-Level Production and Purification of P30P2MSP119, an Important Vaccine Antigen for Malaria, Expressed in the Methylotropic Yeast Pichia pastoris.* Protein Expression Purif., 2001. **23**. 468-475.

131. Damasceno, L. M., Pla, I., Chang, H.-J., Cohen, L., Ritter, G., Old, L. J. and Batt, C. A., *An Optimized Fermentation Process for High-Level Production of a Single-Chain Fv Antibody Fragment in Pichia pastoris.* Protein Expression Purif., 2004. **37**. 18-26.

132. Clare, J. J., Romanos, M. A., Rayment, F. B., Rowedder, J. E., Smith, M. A., Payne, M. M., Sreekrishna, K. and Henwood, C. A., *Production of Mouse Epidermal Growth Factor in Yeast: High-Level Secretion Using Pichia pastoris Strains Containing Multiple Gene Copies.* Gene, 1991. **105**. 205-212.

133. Shi, X., Karkut, T., Chamankhah, M., Alting-Mees, M., Hemmingsen, S. M. and Hegedus, D., *Optimal Conditions for the Expression of a Single-Chain Antibody (scFv) Gene in Pichia pastoris.* Protein Expression Purif., 2003. **28**. 321-330.

134. Sinha, J., Plantz, B. A., Zhang, W., Gouthro, M., Schlegel, V., Liu, C. P. and Meagher, M. M., *Improved Production of Recombinant Ovine Interferon-t by Mut+ Strain of Pichia pastoris Using an Optimized Methanol Feed Profile.* Biotechnol. Prog., 2003. **19**. 794-802.

135. Jayanta, S., Bradley, A. P., Mehmet, I. and Michael, M. M., *Causes of Proteolytic Degradation of Secreted Recombinant Proteins Produced in Methylotrophic Yeast Pichia pastoris: Case Study with Recombinant Ovine Interferon-t.* Biotechnol. Bioeng., 2005. **89**. 102-112.

136. Gerngross, T. U., *Advances in the Production of Human Therapeutic Proteins in Yeasts and Filamentous Fungi.* Nat. Biotech., 2004. **22**. 1409-1414.

137. Takeuchi, M., Inoue, N., Strickland, T. W., Kubota, M., Wada, M., Shimizu, R., Hoshi, S., Kozutsumi, H., Takasaki, S. and Kobata, A., *Relationship between Sugar Chain Structure and Biological Activity of Recombinant Human Erythropoietin Produced in Chinese Hamster Ovary Cells.* PNAS, 1989. **86**. 7819-7822.

138. Yuen, C.-T., Storring, P. L., Tiplady, R. J., Izquierdo, M., Wait, R., Gee, C. K., Gerson, P., Lloyd, P. and Cremata, J. A., *Relationships between the N-Glycan Structures and Biological Activities of Recombinant Human Erythropoietins Produced Using Different Culture Conditions and Purification Procedures.* Br. J. Haematol., 2003. **121**. 511-526.

139. Chiba, Y., Suzuki, M., Yoshida, S., Yoshida, A., Ikenaga, H., Takeuchi, M., Jigami, Y. and Ichishima, E., *Production of Human Compatible High Mannose-Type (Man5GlcNAc2) Sugar Chains in Saccharomyces Cerevisiae.* J. Biol. Chem., 1998. **273**. 26298-26304.

140. Martinet, W., Maras, M., Saelens, X., Jou, W. M. and Contreras, R., *Modification of the Protein Glycosylation Pathway in the Methylotrophic Yeast Pichia pastoris.* Biotechnol. Lett., 1998. **20**. 1171-1177.

141. Choi, B.-K., Bobrowicz, P., Davidson, R. C., Hamilton, S. R., Kung, D. H., Li, H., Miele, R. G., Nett, J. H., Wildt, S. and Gerngross, T. U., *Use of Combinatorial Genetic Libraries to Humanize N-Linked Glycosylation in the Yeast Pichia pastoris.* PNAS, 2003. **100**. 5022-5027.

142. Hamilton, S. R., Bobrowicz, P., Bobrowicz, B., Davidson, R. C., Li, H., Mitchell, T., Nett, J. H., Rausch, S., Stadheim, T. A., Wischnewski, H., Wildt, S. and Gerngross, T. U., *Production of Complex Human Glycoproteins in Yeast.* Science, 2003. **301**. 1244-1246.

143. Bobrowicz, P., Davidson, R. C., Li, H., Potgieter, T. I., Nett, J. H., Hamilton, S. R., Stadheim, T. A., Miele, R. G., Bobrowicz, B., Mitchell, T., Rausch, S., Renfer, E. and Wildt, S., *Engineering of an Artificial Glycosylation Pathway Blocked in Core Oligosaccharide Assembly in the Yeast Pichia pastoris: Production of Complex Humanized Glycoproteins with Terminal Galactose.* Glycobiology, 2004. **14**. 757-766.

144. Rodriguez, E. and Krishna, N. R., *An Economical Method for $^{15}N/^{13}C$ Isotopic Labeling of Proteins Expressed in Pichia pastoris.* J. Biochem., 2001. **130**. 19-22.

145. Laroche, Y., Storme, V., Meutter, J. D., Messens, J. and Lauwereys, M., *High-Level Secretion and Very Efficient Isotopic Labeling of Tick Anticoagulant Peptide (Tap) Expressed in the Methylotrophic Yeast, Pichia pastoris.* Nat. Biotech., 1994. **12**. 1119-1124.

146. Wood, M. J. and Komives, E. A., *Production of Large Quantities of Isotopically Labeled Protein in Pichia pastoris by Fermentation.* J. Biomol. NMR, 1999. **13**. 149-159.

147. Pickford, A. R. and O'Leary, J. M., *Isotopic Labeling of Recombinant Proteins from the Methylotrophic Yeast Pichia pastoris*, in *Protein NMR Techniques*. 2004. p. 17-33.

148. Chen, C.-Y., Cheng, C.-H., Chen, Y.-C., Lee, J.-C., Chou, S.-H., Huang, W. and Chuang, W.-J., *Preparation of Amino-Acid-Type Selective Isotope Labeling of Protein Expressed in Pichia pastoris.* Proteins: Struct. Funct. Bioinform., 2006. **62**. 279-287

149. Whittaker, M. M. and Whittaker, J. W., *Construction and Characterization of Pichia pastoris Strains for Labeling Aromatic Amino Acids in Recombinant Proteins.* Protein Expression Purif., 2005. **41**. 266-274.

150. Wang, X., Zhang, L.-H. and Ye, X.-S., *Recent Development in the Design of Sialyltransferase Inhibitors.* Med. Res. Rev., 2003. **23**. 32-47.

151. Ma, J., Qian, R., Rausa, F. M., III and Colley, K. J., *Two Naturally Occurring α2,6-Sialyltransferase Forms with a Single Amino Acid Change in the Catalytic Domain Differ in Their Catalytic Activity and Proteolytic Processing.* J. Biol. Chem., 1997. **272**. 672-679.

152. Wu, S.-C. and Wong, S.-L., *Engineering Soluble Monomeric Streptavidin with Reversible Biotin Binding Capability.* J. Biol. Chem., 2005. **280**. 23225-23231.

153. Laitinen, O. H., Nordlund, H. R., Hytönen, V. P. and Kulomaa, M. S., *Brave New (Strept)Avidins in Biotechnology.* Trends Biotechnol., 2007. **25**. 269-277.

154. Fritz, T. A., Hurley, J. H., Trinh, L.-B., Shiloach, J. and Tabak, L. A., *The Beginnings of Mucin Biosynthesis: The Crystal Structure of UDP-GalNAc:Polypeptide α-N-Acetylgalactosaminyltransferase-T1.* PNAS, 2004. **101**. 15307-15312.

155. Aoki, K., Perlman, M., Lim, J.-M., Cantu, R., Wells, L. and Tiemeyer, M., *Dynamic Developmental Elaboration of N-Linked Glycan Complexity in the Drosophila Melanogaster Embryo.* J. Biol. Chem., 2007. **282**. 9127-9142.

156. Paulson, J. C., Rearick, J. I. and Hill, R. L., *Enzymatic Properties of β-D-Galactoside $\alpha 2 \rightarrow 6$ Sialytransferase from Bovine Colostrum.* J. Biol. Chem., 1977. **252**. 2363-2371.

157. Weinstein, J., de Souza-e-Silva, U. and Paulson, J. C., *Purification of a Galβ1,4GlcNAc α2, 6 Sialyltransferase and a Galβ1,3(4)GlcNAc α2,3 Sialyltransferase to Homogeneity from Rat Liver.* J. Biol. Chem., 1982. **257**. 13835-13844.

158. Kay, L., Keifer, P. and Saarinen, T., *Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity.* J. Am. Chem. Soc., 1992. **114**. 10663-10665.

159. Karaveg, K., *Structure and Function of ER Class 1 Alpha-Mannosidase*, in *Biochemistry and Molecular biology*. 2005, University of Georgia: Athens. p. 259.

160. Steinrücke, P., Aldinger, U., Hill, O., Hillisch, A., Basch, R. and Diekmann, S., *Design of Helical Proteins for Real-Time Endoprotease Assays.* Anal. Biochem., 2000. **286**. 26-34.

161. Cregg, J. M., Barringer, K. J., Messier, A. Y. and Madden, K. R., *Pichia pastoris as a Host System for Transformations.* Mol. Cell. Biol., 1985. **5**. 3376-3385.

162. Papac, D. I., Wong, A. and Jones, A. J. S., *Analysis of Acidic Oligosaccharides and Glycopeptides by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry.* Anal. Chem., 1996. **68**. 3215-3223.

163. Davidson, R. C., Nett, J. H., Renfer, E., Li, H., Stadheim, T. A., Miller, B. J., Miele, R. G., Hamilton, S. R., Choi, B.-K., Mitchell, T. I. and Wildt, S., *Functional Analysis of the ALG3 Gene Encoding the Dol-P-Man: Man5GlcNAc2-PP-Dol Mannosyltransferase Enzyme of P. pastoris.* Glycobiology, 2004. **14**. 399-407.

164. Malissard, M., Zeng, S. and Berger, E. G., *Expression of Functional Soluble Forms of Human β-1,4-Galactosyltransferase I, α-2,6-Sialyltransferase, and α-1,3-Fucosyltransferase VI in the Methylotrophic Yeast Pichia pastoris.* Biochem. Biophys. Res. Commun., 2000. **267**. 169-173.

165. Hamamoto, T., Lee, Y.-C., Kurosawa, N., Nakaoka, T., Kojima, N. and Tsuji, S., *Expression of Mouse Gal,ß14GlcNAc α2,6-Sialyl-Transferase in an Insoluble Form in Escherichia coli and Partial Renaturation.* Bioorg. Med. Chem., 1994. **2**. 79-84.

166. Hidari, K. I. P. J., Horie, N., Murata, T., Miyamoto, D., Suzuki, T., Usui, T. and Suzuki, Y., *Purification and Characterization of a Soluble Recombinant Human ST6Gal I Functionally Expressed in Escherichia coli.* Glycoconjugate J., 2005. **22**. 1-11.

167. Galan, M. C., Dodson, C. S., Venot, A. P. and Boons, G.-J., *Glycosyltransferase Activity Can Be Selectively Modulated by Chemical Modifications of Acceptor Substrates.* Bioorg. Med. Chem. Lett., 2004. **14**. 2205-2208.

168. Legaigneur, P., Breton, C., El Battari, A., Guillemot, J.-C., Auge, C., Malissard, M., Berger, E. G. and Ronin, C., *Exploring the Acceptor Substrate Recognition of the Human β-Galactoside α2,6-Sialyltransferase.* J. Biol. Chem., 2001. **276**. 21608-21617.

169. Morgan, W. D., Kragt, A. and Feeney, J., *Expression of Deuterium-Isotope-Labelled Protein in the Yeast Pichia pastoris for NMR Studies.* J. Biomol. NMR, 2000. **17**. 337-347.

170. Poon, D. K. Y., Withers, S. G. and McIntosh, L. P., *Direct Demonstration of the Flexibility of the Glycosylated Proline-Threonine Linker in the Cellulomonas Fimi Xylanase Cex through NMR Spectroscopic Analysis.* J. Biol. Chem., 2007. **282**. 2091-2100.

171. Laroy, W., Ameloot, P. and Contreras, R., *Characterization of Sialyltransferase Mutants Using Surface Plasmon Resonance.* Glycobiology, 2001. **11**. 175-182.

172. Macnaughtan, M., A., Tian, F., Liu, S., Meng, L., Park, S., Azadi, P., Moremen, K. W. and Prestegard, J. H., *^{13}C-Sialic Acid Labeling of Glycans on Glycoproteins Using ST6Gal-I*. J. Am. Chem. Soc., 2008. **130**. 1186-5.

APPENDIX

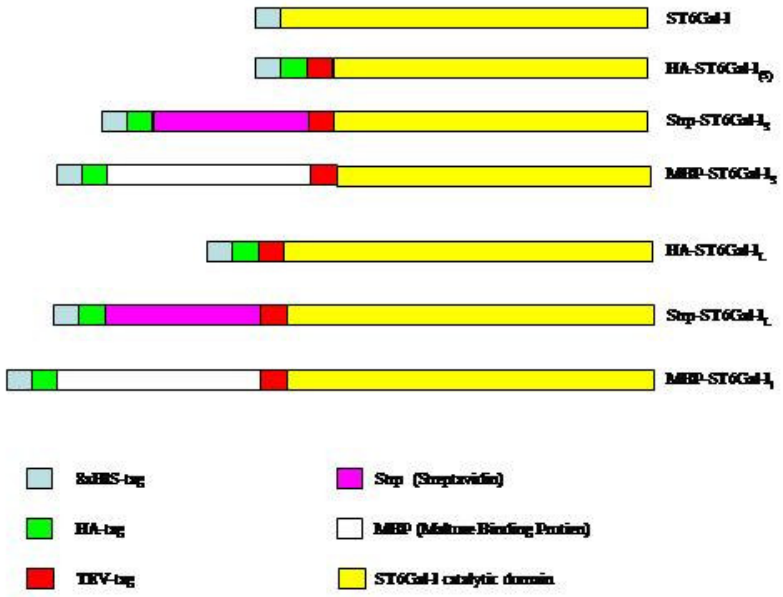


Figure 34: Overview of various constructs for expression of ST6Gal-I as fusion proteins in *P. pastoris*. Positions of various fusion fragments in generation of of ST6Gal-I expression constructs along with restriction sites or mutation used to make these constructs is given below with the sequences of each construct.

A) ST6Gal-I

ClaI BamHI

```
atcgatgcatcatcatcatcatcatcataaggatcctccacatactcaaaacttaac
  S  M  H  H  H  H  H  H  H  H  K  D  P  S  T  Y  S  K  L  N
cccaggctgctgaagatctggagaaactatctgaacatgaacaaatataaagtatcctac
 P  R  L  L  K  I  W  R  N  Y  L  N  M  N  K  Y  K  V  S  Y
aagggaccggggccaggagtcaagttcagcgtagaagcactgcgttgccaccttcgagac
 K  G  P  G  P  G  V  K  F  S  V  E  A  L  R  C  H  L  R  D
catgtgaacgtgtctatgatagaggccacagattttcccttcaacaccactgagtgggag
 H  V  N  V  S  M  I  E  A  T  D  F  P  F  N  T  T  E  W  E
ggttacctgcccaaggagaactttagaaccaaggttgggccttggcaaaggtgtgccgtc
 G  Y  L  P  K  E  N  F  R  T  K  V  G  P  W  Q  R  C  A  V
```

```
gtctcttctgcaggatctctgaaaaactcccagcttggtcgagagattgataatcatgat
 V  S  S  A  G  S  L  K  N  S  Q  L  G  R  E  I  D  N  H  D
gcagttctgaggtttaatggggcccctaccgacaacttccaacaggatgtgggctcaaaa
 A  V  L  R  F  N  G  A  P  T  D  N  F  Q  Q  D  V  G  S  K
actaccattcgcctaatgaactctcagttagtcaccacagaaaagcgcttcctcaaggac
 T  T  I  R  L  M  N  S  Q  L  V  T  T  E  K  R  F  L  K  D
agtttgtacaccgaaggaatcctaattgtatgggacccatccgtgtatcatgcagatatc
 S  L  Y  T  E  G  I  L  I  V  W  D  P  S  V  Y  H  A  D  I
ccaaagtggtatcagaaaccagactacaatttcttcgaaacctataagagttaccgaagg
 P  K  W  Y  Q  K  P  D  Y  N  F  F  E  T  Y  K  S  Y  R  R
ctgaaccccagccagccattttatatcctcaagccccagatgccatgggaactgtgggac
 L  N  P  S  Q  P  F  Y  I  L  K  P  Q  M  P  W  E  L  W  D
atcattcaggaaatctctgcagatctgattcagccaaatcccccatcctccggcatgctg
 I  I  Q  E  I  S  A  D  L  I  Q  P  N  P  P  S  S  G  M  L
ggtatcatcatcatgatgacgctgtgtgaccaggtagatatttacgagttcctcccatcc
 G  I  I  I  M  M  T  L  C  D  Q  V  D  I  Y  E  F  L  P  S
aagcgcaagacggacgtgtgctattatcaccaaaagttctttgacagcgcttgcacgatg
 K  R  K  T  D  V  C  Y  Y  H  Q  K  F  F  D  S  A  C  T  M
ggtgcctaccacccgctcctcttcgagaagaatatggtgaagcatctcaatgagggaaca
 G  A  Y  H  P  L  L  F  E  K  N  M  V  K  H  L  N  E  G  T
gatgaagacatttatttgtttgggaaagccacccttctggcttccggaacattcgttgt
 D  E  D  I  Y  L  F  G  K  A  T  L  S  G  F  R  N  I  R  C
taagcggccgc
 -    NotI
```

B) HA-ST6Gal-I

```
gcagagcatcatcatcatcatcatcattacccctacgacgtccccgactacgccgaa
 A  E  H  H  H  H  H  H  H  H  Y  P  Y  D  V  P  D  Y  A  E
                AscI
aacctgtacttccaaagcgggcgcgccagaaactatctgaacatgaacaaatataaagta
 N  L  Y  F  Q  S  G  R  A  R  N  Y  L  N  M  N  K  Y  K  V
tcctacaagggaccggggccaggagtcaagttcagcgtagaagcactgcgttgccacctt
 S  Y  K  G  P  G  P  G  V  K  F  S  V  E  A  L  R  C  H  L
cgagaccatgtgaacgtgtctatgatagaggccacagattttcccttcaacaccactgag
 R  D  H  V  N  V  S  M  I  E  A  T  D  F  P  F  N  T  T  E
tgggagggttacctgcccaaggagaactttagaaccaaggttgggccttggcaaaggtgt
 W  E  G  Y  L  P  K  E  N  F  R  T  K  V  G  P  W  Q  R  C
gccgtcgtctcttctgcaggatctctgaaaaactcccagcttggtcgagagattgataat
 A  V  V  S  S  A  G  S  L  K  N  S  Q  L  G  R  E  I  D  N
catgatgcagttctgaggtttaatggggcccctaccgacaacttccaacaggatgtgggc
 H  D  A  V  L  R  F  N  G  A  P  T  D  N  F  Q  Q  D  V  G
tcaaaaactaccattcgcctaatgaactctcagttagtcaccacagaaaagcgcttcctc
 S  K  T  T  I  R  L  M  N  S  Q  L  V  T  T  E  K  R  F  L
aaggacagtttgtacaccgaaggaatcctaattgtatgggacccatccgtgtatcatgca
 K  D  S  L  Y  T  E  G  I  L  I  V  W  D  P  S  V  Y  H  A
gatatcccaaagtggtatcagaaaccagactacaatttcttcgaaacctataagagttac
```

D I P K W Y Q K P D Y N F F E T Y K S Y
cgaaggctgaaccccagccagccattttatatcctcaagccccagatgccatgggaactg
R R L N P S Q P F Y I L K P Q M P W E L
tgggacatcattcaggaaatctctgcagatctgattcagccaaatcccccatcctccggc
W D I I Q E I S A D L I Q P N P P S S G
atgctgggtatcatcatcatgatgacgctgtgtgaccaggtagatatttacgagttcctc
M L G I I I M M T L C D Q V D I Y E F L
ccatccaagcgcaagacggacgtgtgctattatcaccaaaagttctttgacagcgcttgc
P S K R K T D V C Y Y H Q K F F D S A C
acgatgggtgcctaccacccgctcctcttcgagaagaatatggtgaagcatctcaatgag
T M G A Y H P L L F E K N M V K H L N E
ggaacagatgaagacatttatttgtttgggaaagccacccttctggcttccggaacatt
G T D E D I Y L F G K A T L S G F R N I
cgttgttaagcggccgc
R C -

C) HA-ST6GalI$_s$

ClaI EcoRI

atcgatacatcatcatcatcatcatcattacccatacgacgtcccagactacgcagaa
S I H H H H H H H H Y P Y D V P D Y A E
ttctcttctgaaaacttgtactttcaatctaactatctgaacatgaacaaatataaagta
F S S E N L Y F Q S N Y L N M N K Y K V
tcctacaagggaccggggccaggagtcaagttcagcgtagaagcactgcgttgccacctt
S Y K G P G P G V K F S V E A L R C H L
cgagaccatgtgaacgtgtctatgatagaggccacagattttcccttcaacaccactgag
R D H V N V S M I E A T D F P F N T T E
tgggagggttacctgcccaaggagaactttagaaccaaggttgggccttggcaaaggtgt
W E G Y L P K E N F R T K V G P W Q R C
gccgtcgtctcttctgcaggatctctgaaaaactcccagcttggtcgagagattgataat
A V V S S A G S L K N S Q L G R E I D N
catgatgcagttctgaggtttaatggggcccctaccgacaacttccaacaggatgtgggc
H D A V L R F N G A P T D N F Q Q D V G
tcaaaaactaccattcgcctaatgaactctcagttagtcaccacagaaaagcgcttcctc
S K T T I R L M N S Q L V T T E K R F L
aaggacagtttgtacaccgaaggaatcctaattgtatgggacccatccgtgtatcatgca
K D S L Y T E G I L I V W D P S V Y H A
gatatcccaaagtggtatcagaaaccagactacaatttcttcgaaacctataagagttac
D I P K W Y Q K P D Y N F F E T Y K S Y
cgaaggctgaaccccagccagccattttatatcctcaagccccagatgccatgggaactg
R R L N P S Q P F Y I L K P Q M P W E L
tgggacatcattcaggaaatctctgcagatctgattcagccaaatcccccatcctccggc
W D I I Q E I S A D L I Q P N P P S S G
atgctgggtatcatcatcatgatgacgctgtgtgaccaggtagatatttacgagttcctc
M L G I I I M M T L C D Q V D I Y E F L
ccatccaagcgcaagacggacgtgtgctattatcaccaaaagttctttgacagcgcttgc
P S K R K T D V C Y Y H Q K F F D S A C

```
acgatgggtgcctaccacccgctcctcttcgagaagaatatggtgaagcatctcaatgag
 T  M  G  A  Y  H  P  L  L  F  E  K  N  M  V  K  H  L  N  E
ggaacagatgaagacatttatttgtttgggaaagccacccttctggcttccggaacatt
 G  T  D  E  D  I  Y  L  F  G  K  A  T  L  S  G  F  R  N  I
cgttgttaataagcggccgc
 R  C  -  -     NotI
```

D) Strp-ST6GalI$_s$

```
ClaI                                                        EcoRI
atcgatacatcatcatcatcatcatcatcattacccatacgacgtcccagactacgcagaa
  S  I  H  H  H  H  H  H  H  H  Y  P  Y  D  V  P  D  Y  A  E
ttcaccggtacgtggtataaccagctgggcagcacctttattgtgaccgccggtgcggat
 F  T  G  T  W  Y  N  Q  L  G  S  T  F  I  V  T  A  G  A  D
ggtgcgctgaccggcacctatgaaagcgcggtgggcaacgcggaaagccgttataccctg
 G  A  L  T  G  T  Y  E  S  A  V  G  N  A  E  S  R  Y  T  L
accggccgttatgatagcgcgccggcgaccgatggatccggcaccgcgctgggttggcgt
 T  G  R  Y  D  S  A  P  A  T  D  G  S  G  T  A  L  G  W  R
gtggcgtggaaaaacaactatcgtaacgcgcatagcgcgaccacctggagcggccagtat
 V  A  W  K  N  N  Y  R  N  A  H  S  A  T  T  W  S  G  Q  Y
gtgggtggtgcggaagcgcgtattaacacccagtggaccctgaccagcggcaccaccgaa
 V  G  G  A  E  A  R  I  N  T  Q  W  T  L  T  S  G  T  T  E
gcgaacgcgtggaaaagcaccctgcgtggccatgataccttaccaaagtgaaaccgagc
 A  N  A  W  K  S  T  L  R  G  H  D  T  F  T  K  V  K  P  S
gcggcgagcattgatgcggcgaaaaaagcgggcgtgaacaacggcaatccgctggatgcg
 A  A  S  I  D  A  A  K  K  A  G  V  N  N  G  N  P  L  D  A
        EcoRI
gtgcagcaagaattctcttctgaaaacttgtactttcaatctaactatctgaacatgaac
 V  Q  Q  E  F  S  S  E  N  L  Y  F  Q  S  N  Y  L  N  M  N
aaatataaagtatcctacaagggaccggggccaggagtcaagttcagcgtagaagcactg
 K  Y  K  V  S  Y  K  G  P  G  P  G  V  K  F  S  V  E  A  L
cgttgccaccttcgagaccatgtgaacgtgtctatgatagaggccacagattttcccttc
 R  C  H  L  R  D  H  V  N  V  S  M  I  E  A  T  D  F  P  F
aacaccactgagtgggagggttacctgcccaaggagaactttagaaccaaggttgggcct
 N  T  T  E  W  E  G  Y  L  P  K  E  N  F  R  T  K  V  G  P
tggcaaaggtgtgccgtcgtctcttctgcaggatctctgaaaaactcccagcttggtcga
 W  Q  R  C  A  V  V  S  S  A  G  S  L  K  N  S  Q  L  G  R
gagattgataatcatgatgcagttctgaggtttaatggggcccctaccgacaacttccaa
 E  I  D  N  H  D  A  V  L  R  F  N  G  A  P  T  D  N  F  Q
caggatgtgggctcaaaaactaccattcgcctaatgaactctcagttagtcaccacagaa
 Q  D  V  G  S  K  T  T  I  R  L  M  N  S  Q  L  V  T  T  E
aagcgcttcctcaaggacagtttgtacaccgaaggaatcctaattgtatgggacccatcc
 K  R  F  L  K  D  S  L  Y  T  E  G  I  L  I  V  W  D  P  S
gtgtatcatgcagatatcccaaagtggtatcagaaaccagactacaatttcttcgaaacc
 V  Y  H  A  D  I  P  K  W  Y  Q  K  P  D  Y  N  F  F  E  T
tataagagttaccgaaggctgaaccccagccagccattttatatcctcaagccccagatg
 Y  K  S  Y  R  R  L  N  P  S  Q  P  F  Y  I  L  K  P  Q  M
ccatgggaactgtgggacatcattcaggaaatctctgcagatctgattcagccaaatccc
```

P W E L W D I I Q E I S A D L I Q P N P
ccatcctccggcatgctgggtatcatcatcatgatgacgctgtgtgaccaggtagatatt
P S S G M L G I I I M M T L C D Q V D I
tacgagttcctcccatccaagcgcaagacggacgtgtgctattatcaccaaaagttcttt
Y E F L P S K R K T D V C Y Y H Q K F F
gacagcgcttgcacgatgggtgcctacgacccgctcctcttcgagaagaatatggtgaag
D S A C T M G A Y D P L L F E K N M V K
catctcaatgagggaacagatgaacacatttatttgtttgggaaagccacccttctggc
H L N E G T D E H I Y L F G K A T L S G
ttccggaacattcgttgttaataagcggccgc
F R N I R C - - NotI

E) MBP-ST6Gal-I$_S$

gcttaccatcatcatcatcatcattacgtaaaaatcgaagaaggtaaactggtaatctgg
A Y H H H H H H Y V K I E E G K L V I W
attaacggcgataaaggctataacggtctcgctgaagtcggtaagaaattcgagaaagat
I N G D K G Y N G L A E V G K K F E K D
accggaattaaagtcaccgttgagcatccggataaactggaagagaaattcccacaggtt
T G I K V T V E H P D K L E E K F P Q V
gcggcaactggcgatggccctgacattatcttctgggcacacgaccgctttggtggctac
A A T G D G P D I I F W A H D R F G G Y
gctcaatctggcctgttggctgaaatcaccccggacaaagcgttccaggacaagctgtat
A Q S G L L A E I T P D K A F Q D K L Y
ccgtttacctgggatgccgtacgttacaacggcaagctgattgcttacccgatcgctgtt
P F T W D A V R Y N G K L I A Y P I A V
gaagcgttatcgctgatttataacaaagatctgctgccgaacccgccaaaaacctgggaa
E A L S L I Y N K D L L P N P P K T W E
gagatcccggcgctggataaagaactgaaagcgaaaggtaagagcgcgctgatgttcaac
E I P A L D K E L K A K G K S A L M F N
ctgcaagaaccgtacttcacctggccgctgattgctgctgacgggggttatgcgttcaag
L Q E P Y F T W P L I A A D G G Y A F K
tatgaaaacggcaagtacgacattaaagacgtgggcgtggataacgctggcgcgaaagcg
Y E N G K Y D I K D V G V D N A G A K A
ggtctgaccttcctggttgacctgattaaaaacaaacacatgaatgcagacaccgattac
G L T F L V D L I K N K H M N A D T D Y
tccatcgcagaagctgcctttaataaaggcgaaacagcgatgaccatcaacggcccgtgg
S I A E A A F N K G E T A M T I N G P W
gcatggtccaacatcgacaccagcaaagtgaattatggtgtaacggtactgccgaccttc
A W S N I D T S K V N Y G V T V L P T F
aagggtcaaccatccaaaccgttcgttggcgtgctgagcgcaggtattaacgccgccagt
K G Q P S K P F V G V L S A G I N A A S
ccgaacaaagagctggcaaaagagttcctcgaaaactatctgctgactgatgaaggtctg
P N K E L A K E F L E N Y L L T D E G L
gaagcggttaataaagacaaaccgctgggtgccgtagcgctgaagtcttacgaggaagag
E A V N K D K P L G A V A L K S Y E E E
ttggcgaaagatccacgtattgccgccactatggaaaacgcccagaaaggtgaaatcatg

```
 L  A  K  D  P  R  I  A  A  T  M  E  N  A  Q  K  G  E  I  M
ccgaacatcccgcagatgtccgctttctggtatgccgtgcgtactgcggtgatcaacgcc
 P  N  I  P  Q  M  S  A  F  W  Y  A  V  R  T  A  V  I  N  A
                                          EcoRI
gccagcggtcgtcagactgtcgatgaagccctgaaagacgcggaattctcttctgaaaac
 A  S  G  R  Q  T  V  D  E  A  L  K  D  A  E  F  S  S  E  N
ttgtactttcaatctaactatctgaacatgaacaaatataaagtatcctacaagggaccg
 L  Y  F  Q  S  N  Y  L  N  M  N  K  Y  K  V  S  Y  K  G  P
gggccaggagtcaagttcagcgtagaagcactgcgttgccaccttcgagaccatgtgaac
 G  P  G  V  K  F  S  V  E  A  L  R  C  H  L  R  D  H  V  N
gtgtctatgatagaggccacagattttcccttcaacaccactgagtgggagggttacctg
 V  S  M  I  E  A  T  D  F  P  F  N  T  T  E  W  E  G  Y  L
cccaaggagaactttagaaccaaggttgggccttggcaaaggtgtgccgtcgtctcttct
 P  K  E  N  F  R  T  K  V  G  P  W  Q  R  C  A  V  V  S  S
gcaggatctctgaaaaactcccagcttggtcgagagattgataatcatgatgcagttctg
 A  G  S  L  K  N  S  Q  L  G  R  E  I  D  N  H  D  A  V  L
aggtttaatggggcccctaccgacaacttccaacaggatgtgggctcaaaaactaccatt
 R  F  N  G  A  P  T  D  N  F  Q  Q  D  V  G  S  K  T  T  I
cgcctaatgaactctcagttagtcaccacagaaaagcgcttcctcaaggacagtttgtac
 R  L  M  N  S  Q  L  V  T  T  E  K  R  F  L  K  D  S  L  Y
accgaaggaatcctaattgtatgggacccatccgtgtatcatgcagatatcccaaagtgg
 T  E  G  I  L  I  V  W  D  P  S  V  Y  H  A  D  I  P  K  W
tatcagaaaccagactacaatttcttcgaaacctataagagttaccgaaggctgaacccc
 Y  Q  K  P  D  Y  N  F  F  E  T  Y  K  S  Y  R  R  L  N  P
agccagccattttatatcctcaagccccagatgccatgggaactgtgggacatcattcag
 S  Q  P  F  Y  I  L  K  P  Q  M  P  W  E  L  W  D  I  I  Q
gaaatctctgcagatctgattcagccaaatcccccatcctccggcatgctgggtatcatc
 E  I  S  A  D  L  I  Q  P  N  P  P  S  S  G  M  L  G  I  I
atcatgatgacgctgtgtgaccaggtagatatttacgagttcctcccatccaagcgcaag
 I  M  M  T  L  C  D  Q  V  D  I  Y  E  F  L  P  S  K  R  K
acggacgtgtgctattatcaccaaaagttctttgacagcgcttgcacgatgggtgcctac
 T  D  V  C  Y  Y  H  Q  K  F  F  D  S  A  C  T  M  G  A  Y
gacccgctcctcttcgagaagaatatggtgaagcatctcaatgagggaacagatgaacac
 D  P  L  L  F  E  K  N  M  V  K  H  L  N  E  G  T  D  E  H
atttatttgtttgggaaagccacccttctggcttccggaacattcgttgttaataagcg
 I  Y  L  F  G  K  A  T  L  S  G  F  R  N  I  R  C  -  -   NotI
gccgc
```

F) HA-ST6GalI$_L$

```
ClaI                                                    EcoRI
atcgatacatcatcatcatcatcatcattacccatacgacgtcccagactacgcagaa
  S  I  H  H  H  H  H  H  H  H  Y  P  Y  D  V  P  D  Y  A  E
ttctcttctgaaaacttgtactttcaatctaaggactccacatactcaaaacttaacccc
 F  S  S  E  N  L  Y  F  Q  S  K  D  S  T  Y  S  K  L  N  P
aggctgctgaagatctggagtaactatctgaacatgaacaaatataaagtatcctacaag
```

```
 R  L  L  K  I  W  S  N  Y  L  N  M  N  K  Y  K  V  S  Y  K
       ApaI
ggaccgggcccaggagtcaagttcagcgtagaagcactgcgttgccaccttcgagaccat
 G  P  G  P  G  V  K  F  S  V  E  A  L  R  C  H  L  R  D  H
gtgaacgtgtctatgatagaggccacagattttcccttcaacaccactgagtgggagggt
 V  N  V  S  M  I  E  A  T  D  F  P  F  N  T  T  E  W  E  G
tacctgcccaaggagaactttagaaccaaggttgggccttggcaaaggtgtgccgtcgtc
 Y  L  P  K  E  N  F  R  T  K  V  G  P  W  Q  R  C  A  V  V
tcttctgcaggatctctgaaaaactcccagcttggtcgagagattgataatcatgatgca
 S  S  A  G  S  L  K  N  S  Q  L  G  R  E  I  D  N  H  D  A
gttctgaggtttaatggcggcccctaccgacaacttccaacaggatgtgggctcaaaaact
 V  L  R  F  N  G  A  P  T  D  N  F  Q  Q  D  V  G  S  K  T
accattcgcctaatgaactctcagttagtcaccacagaaaagcgcttcctcaaggacagt
 T  I  R  L  M  N  S  Q  L  V  T  T  E  K  R  F  L  K  D  S
ttgtacaccgaaggaatcctaattgtatgggacccatccgtgtatcatgcagatatccca
 L  Y  T  E  G  I  L  I  V  W  D  P  S  V  Y  H  A  D  I  P
aagtggtatcagaaaccagactacaatttcttcgaaacctataagagttaccgaaggctg
 K  W  Y  Q  K  P  D  Y  N  F  F  E  T  Y  K  S  Y  R  R  L
aaccccagccagccattttatatcctcaagccccagatgccatgggaactgtgggacatc
 N  P  S  Q  P  F  Y  I  L  K  P  Q  M  P  W  E  L  W  D  I
attcaggaaatctctgcagatctgattcagccaaatcccccatcctccggcatgctgggt
 I  Q  E  I  S  A  D  L  I  Q  P  N  P  P  S  S  G  M  L  G
atcatcatcatgatgacgctgtgtgaccaggtagatatttacgagttcctcccatccaag
 I  I  I  M  M  T  L  C  D  Q  V  D  I  Y  E  F  L  P  S  K
cgcaagacggacgtgtgctattatcaccaaaagttctttgacagcgcttgcacgatgggt
 R  K  T  D  V  C  Y  Y  H  Q  K  F  F  D  S  A  C  T  M  G
gcctaccacccgctcctcttcgagaagaatatggtgaagcatctcaatgagggaacagat
 A  Y  H  P  L  L  F  E  K  N  M  V  K  H  L  N  E  G  T  D
gaagacatttatttgtttgggaaagccacccttctggcttccggaacattcgttgttaa
 E  D  I  Y  L  F  G  K  A  T  L  S  G  F  R  N  I  R  C  -
taagcggccgc
 -     NotI
```

G) Strp-ST6GalI$_L$

```
ClaI                                                        EcoRI
atcgatacatcatcatcatcatcatcattacccatacgacgtcccagactacgcagaa
  S  I  H  H  H  H  H  H  H  H  Y  P  Y  D  V  P  D  Y  A  E
tttaccggtacgtggtataaccagctgggcagcacctttattgtgaccgccggtgcggat
 F  T  G  T  W  Y  N  Q  L  G  S  T  F  I  V  T  A  G  A  D
ggtgcgctgaccggcacctatgaaagcgcggtgggcaacgcggaaagccgttataccctg
 G  A  L  T  G  T  Y  E  S  A  V  G  N  A  E  S  R  Y  T  L
accggccgttatgatagcgcgccggcgaccgatggatccggcaccgcgctgggttggcgt
 T  G  R  Y  D  S  A  P  A  T  D  G  S  G  T  A  L  G  W  R
gtggcgtggaaaaacaactatcgtaacgcgcatagcgcgaccacctggagcggccagtat
 V  A  W  K  N  N  Y  R  N  A  H  S  A  T  T  W  S  G  Q  Y
```

gtgggtggtgcggaagcgcgtattaacacccagtggaccctgaccagcggcaccaccgaa
V G G A E A R I N T Q W T L T S G T T E
gcgaacgcgtggaaaagcaccctgcgtggccatgatacctttaccaaagtgaaaccgagc
A N A W K S T L R G H D T F T K V K P S
gcggcgagcattgatgcggcgaaaaaagcgggcgtgaacaacggcaatccgctggatgcg
A A S I D A A K K A G V N N G N P L D A
EcoRI
gtgcagcaagaattcttcttctgaaaacttgtactttcaatctaaggactccacatactca
V Q Q E F S S E N L Y F Q S K D S T Y S
aaacttaaccccaggctgctgaagatctggagtaactatctgaacatgaacaaatataaa
K L N P R L L K I W S N Y L N M N K Y K
ApaI
gtatcctacaagggaccgggcccaggagtcaagttcagcgtagaagcactgcgttgccac
V S Y K G P G P G V K F S V E A L R C H
cttcgagaccatgtgaacgtgtctatgatagaggccacagattttcccttcaacaccact
L R D H V N V S M I E A T D F P F N T T
gagtgggagggttacctgcccaaggagaactttagaaccaaggttgggccttggcaaagg
E W E G Y L P K E N F R T K V G P W Q R
tgtgccgtcgtctcttctgcaggatctctgaaaaactcccagcttggtcgagagattgat
C A V V S S A G S L K N S Q L G R E I D
aatcatgatgcagttctgaggtttaatggcgcccctaccgacaacttccaacaggatgtg
N H D A V L R F N G A P T D N F Q Q D V
ggctcaaaaactaccattcgcctaatgaactctcagttagtcaccacagaaaagcgcttc
G S K T T I R L M N S Q L V T T E K R F
ctcaaggacagtttgtacaccgaaggaatcctaattgtatgggacccatccgtgtatcat
L K D S L Y T E G I L I V W D P S V Y H
gcagatatcccaaagtggtatcagaaaccagactacaatttcttcgaaacctataagagt
A D I P K W Y Q K P D Y N F F E T Y K S
taccgaaggctgaaccccagccagccattttatatcctcaagccccagatgccatgggaa
Y R R L N P S Q P F Y I L K P Q M P W E
ctgtgggacatcattcaggaaatctctgcagatctgattcagccaaatcccccatcctcc
L W D I I Q E I S A D L I Q P N P P S S
ggcatgctgggtatcatcatcatgatgacgctgtgtgaccaggtagatatttacgagttc
G M L G I I I M M T L C D Q V D I Y E F
ctcccatccaagcgcaagacggacgtgtgctattatcaccaaaagttctttgacagcgct
L P S K R K T D V C Y Y H Q K F F D S A
tgcacgatgggtgcctaccacccgctcctcttcgagaagaatatggtgaagcatctcaat
C T M G A Y H P L L F E K N M V K H L N
gagggaacagatgaagacatttatttgtttgggaaagccacccttctggcttccggaac
E G T D E D I Y L F G K A T L S G F R N
attcgttgttaataagcggccgc
I R C - - NotI

H) MBP-ST6Gal-I$_L$

gcttaccatcatcatcatcatcattacgtaaaaatcgaagaaggtaaactggtaatctgg
A Y H H H H H H Y V K I E E G K L V I W

attaacggcgataaaggctataacggtctcgctgaagtcggtaagaaattcgagaaagat
 I N G D K G Y N G L A E V G K K F E K D
accggaattaaagtcaccgttgagcatccggataaactggaagagaaattcccacaggtt
 T G I K V T V E H P D K L E E K F P Q V
gcggcaactggcgatggccctgacattatcttctgggcacacgaccgctttggtggctac
 A A T G D G P D I I F W A H D R F G G Y
gctcaatctggcctgttggctgaaatcaccccggacaaagcgttccaggacaagctgtat
 A Q S G L L A E I T P D K A F Q D K L Y
ccgtttacctgggatgccgtacgttacaacggcaagctgattgcttacccgatcgctgtt
 P F T W D A V R Y N G K L I A Y P I A V
gaagcgttatcgctgatttataacaaagatctgctgccgaacccgccaaaaacctgggaa
 E A L S L I Y N K D L L P N P P K T W E
gagatcccggcgctggataaagaactgaaagcgaaaggtaagagcgcgctgatgttcaac
 E I P A L D K E L K A K G K S A L M F N
ctgcaagaaccgtacttcacctggccgctgattgctgctgacgggggttatgcgttcaag
 L Q E P Y F T W P L I A A D G G Y A F K
tatgaaaacggcaagtacgacattaaagacgtgggcgtggataacgctggcgcgaaagcg
 Y E N G K Y D I K D V G V D N A G A K A
ggtctgaccttcctggttgacctgattaaaaacaaacacatgaatgcagacaccgattac
 G L T F L V D L I K N K H M N A D T D Y
tccatcgcagaagctgcctttaataaaggcgaaacagcgatgaccatcaacggcccgtgg
 S I A E A A F N K G E T A M T I N G P W
gcatggtccaacatcgacaccagcaaagtgaattatggtgtaacggtactgccgaccttc
 A W S N I D T S K V N Y G V T V L P T F
aagggtcaaccatccaaaccgttcgttggcgtgctgagcgcaggtattaacgccgccagt
 K G Q P S K P F V G V L S A G I N A A S
ccgaacaaagagctggcaaaagagttcctcgaaaactatctgctgactgatgaaggtctg
 P N K E L A K E F L E N Y L L T D E G L
gaagcggttaataaagacaaaccgctgggtgccgtagcgctgaagtcttacgaggaagag
 E A V N K D K P L G A V A L K S Y E E E
ttggcgaaagatccacgtattgccgccactatggaaaacgcccagaaaggtgaaatcatg
 L A K D P R I A A T M E N A Q K G E I M
ccgaacatcccgcagatgtccgctttctggtatgccgtgcgtactgcggtgatcaacgcc
 P N I P Q M S A F W Y A V R T A V I N A
                                                ApaI
gccagcggtcgtcagactgtcgatgaagccctgaaagacgcggaattctcttctgaaaac
 A S G R Q T V D E A L K D A E F S S E N
ttgtactttcaatctaaggactccacatactcaaaacttaaccccaggctgctgaagatc
 L Y F Q S K D S T Y S K L N P R L L K I
tggagtaactatctgaacatgaacaaatataaagtatcctacaagggaccgggccccagga
 W S N Y L N M N K Y K V S Y K G P G P G
gtcaagttcagcgtagaagcactgcgttgccaccttcgagaccatgtgaacgtgtctatg
 V K F S V E A L R C H L R D H V N V S M
atagaggccacagattttcccttcaacaccactgagtgggagggttacctgcccaaggag
 I E A T D F P F N T T E W E G Y L P K E
aactttagaaccaaggttgggccttggcaaaggtgtgccgtcgtctcttctgcaggatct
 N F R T K V G P W Q R C A V V S S A G S

ctgaaaaactcccagcttggtcgagagattgataatcatgatgcagttctgaggtttaat
L K N S Q L G R E I D N H D A V L R F N
ggcgcccctaccgacaacttccaacaggatgtgggctcaaaaactaccattcgcctaatg
G A P T D N F Q Q D V G S K T T I R L M
aactctcagttagtcaccacagaaaagcgcttcctcaaggacagtttgtacaccgaagga
N S Q L V T T E K R F L K D S L Y T E G
atcctaattgtatgggacccatccgtgtatcatgcagatatcccaaagtggtatcagaaa
I L I V W D P S V Y H A D I P K W Y Q K
ccagactacaatttcttcgaaacctataagagttaccgaaggctgaaccccagccagcca
P D Y N F F E T Y K S Y R R L N P S Q P
ttttatatcctcaagccccagatgccatgggaactgtgggacatcattcaggaaatctct
F Y I L K P Q M P W E L W D I I Q E I S
gcagatctgattcagccaaatcccccatcctccggcatgctgggtatcatcatcatgatg
A D L I Q P N P P S S G M L G I I I M M
acgctgtgtgaccaggtagatatttacgagttcctcccatccaagcgcaagacggacgtg
T L C D Q V D I Y E F L P S K R K T D V
tgctattatcaccaaaagttctttgacagcgcttgcacgatgggtgcctaccacccgctc
C Y Y H Q K F F D S A C T M G A Y H P L
ctcttcgagaagaatatggtgaagcatctcaatgagggaacagatgaagacatttatttg
L F E K N M V K H L N E G T D E D I Y L
tttgggaaagccacccttctggcttccggaacattcgttgttaataagcggccgc
F G K A T L S G F R N I R C - - NotI

I) SUNGA-ST6GalI

ClaI

atcgatgcatcatcatcatcatcatcataaggatccctccacatactcaaaacttaac
S M H H H H H H H H H K D P S T Y S K L N
cccaggctgctgaagatctggagaaactatctgaacatgaacaaatataaagtatcctac
P R L L K I W R N Y L N M N K Y K V S Y
aagggaccggggccaggagtcaagttcagcgtagaagcactgcgttgccaccttcgagac
K G P G P G V K F S V E A L R C H L R D
catgtgaacgtgcgtatgatagaggccacagattttcccttcaacaagctggagtgggag
H V N V R M I E A T D F P F N K L E W E
ggttacctgcccaaggagaactttagaaccaaggttgggccttggcaaaggtgtgccgtc
G Y L P K E N F R T K V G P W Q R C A V
gtctcttctgcaggatctctgaaaaactcccagcttggtcgagagattgataatcatgat
V S S A G S L K N S Q L G R E I D N H D
gcagttctgaggtttaatggggcccctaccgacaacttccaacaggatgtgggctcaaaa
A V L R F N G A P T D N F Q Q D V G S K
actaccattcgcctaatgaactctcagttagtcaccacagaaaagcgcttcctcaaggac
T T I R L M N S Q L V T T E K R F L K D
agtttgtacaccgaaggaatcctaattgtatgggacccatccgtgtatcatgcagatatc
S L Y T E G I L I V W D P S V Y H A D I
ccaaagtggtatcagaaaccagactacaatttcttcgaaacctataagagttaccgaagg
P K W Y Q K P D Y N F F E T Y K S Y R R
ctgaaccccagccagccattttatatcctcaagccccagatgccatgggaactgtgggac

```
 L  N  P  S  Q  P  F  Y  I  L  K  P  Q  M  P  W  E  L  W  D
atcattcaggaaatctctgcagatctgattcagccaaatcccccatcctccggcatgctg
 I  I  Q  E  I  S  A  D  L  I  Q  P  N  P  P  S  S  G  M  L
ggtatcatcatcatgatgacgctgtgtgaccaggtagatatttacgagttcctcccatcc
 G  I  I  I  M  M  T  L  C  D  Q  V  D  I  Y  E  F  L  P  S
aagcgcaagacggacgtgtgctattatcaccaaaagttctttgacagcgcttgcacgatg
 K  R  K  T  D  V  C  Y  Y  H  Q  K  F  F  D  S  A  C  T  M
ggtgcctaccacccgctcctcttcgagaagaatatggtgaaactgtcgcgaacgtgctac
 G  A  Y  H  P  L  L  F  E  K  N  M  V  K  L  S  R  T  C  Y
ccacggatgctgggcgaggagaagctcttcttataccacttgcatctcaatgagggaaca
 P  R  M  L  G  E  E  K  L  F  L  Y  H  L  H  L  N  E  G  T
gatgaagacatttatttgtttgggaaagccacccttctggcttccggaacattcgttgt
 D  E  D  I  Y  L  F  G  K  A  T  L  S  G  F  R  N  I  R  C
taataagcggccgc
 -  -     NotI
```