Solution-Phase Nuclear Magnetic Resonance Studies of a Nonribosomal Peptide Synthetase Adenylation Domain, of a Bacterial Glycosyltransferase, and the Rational Design of Inhibitors and Mutants of Glycosyltransferases

by

John Samuel MacMaster

Department of Computer Science
Duke University

Date:_______________________
Approved:

____
Bruce R. Donald, Supervisor

____
Carlo Tomasi

____
David Richardson

____
Thomas LaBean

Dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy
in the Department
of Computer Science
in the Graduate School
of
Duke University

2015

ABSTRACT

Solution-Phase Nuclear Magnetic Resonance Studies of a Nonribosomal Peptide Synthetase Adenylation Domain, of a Bacterial Glycosyltransferase, and the Rational Design of Inhibitors and Mutants of Glycosyltransferases

by

John Samuel MacMaster

Department of Computer Science
Duke University

Date:_______________________
Approved:

____
Bruce R. Donald, Supervisor

____
Carlo Tomasi

____
David Richardson

____
Thomas LaBean

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of
Computer Science in the Graduate School of
Duke University

2015

Copyright by
John Samuel MacMaster
2015

Abstract

A molecule’s biological function is determined by its chemical structure and its three dimensional (3D) shape. While a molecule’s chemical structure is fairly static its physical 3D structure is typically very dynamic and thus more difficult to determine. A protein’s 3D structure is actually an ensemble of shapes that it can assume depending on its immediate surroundings. The two main methods of determining a protein’s 3D structure at high resolution are X-ray crystallography and nuclear magnetic resonance (NMR). These two methods complement each other by allowing for a protein’s 3D shape to be studied in a wider variety of environments than either one alone can do. We are working to develop new methods for determining the 3D structures of proteins in solution by NMR, with and without ligands present that may bind to them. In particular we are developing NMR methods for studying the solution-phase 3D structures of large, biologically important, enzymes.

We are interested in determining the solution-phase 3D structures of enzymes at the atomic level so that we can understand their biological functions and how they accomplish them, and thus how to control them in order to treat diseases and improve human health. We are also interested in using high resolution structures of enzymes to do structure-based reengineering of them. Redesigning enzymes enhances our understanding of how they function in their native environment and leads to redesigned versions of them that can be used to chemoenzymatically synthesize clinically important drugs.

This dissertation begins with our studies, by NMR, of the solution-phase structures of two bacterial enzymes involved in the biosynthesis of antibiotics. In particular we studied the solution-phase structures of the adenylation domain responsible for selectively activating the amino acid phenylalanine in the biosynthetic pathway for the antibiotic gramicidin S. Next, we present our studies of two glycosylation enzymes involved in the final phase of biosynthesis of the antibiotic vancomycin. We compared two approaches to determine the amino acids involved in substrate binding by these two enzymes, a solution-phase NMR approach and an *in silico* protein modeling, with ligand docking, approach. These enzymes are each quite large for current NMR solution-phase techniques and we present the lessons we learned from studying them and our plans for future work. Finally, we present a review of the use of small-molecule inhibitors and enzyme redesign in the study of the function of glycosyltransferases, with applications in the treatment of glycosylation disorders in humans and the chemoenzymatic synthesis of homogeneously glycosylated molecules.

Dedication

This work is dedicated in loving memory to my father, who passed away way too soon. You taught me that no challenge is too big to take on. If you're watching, I hope you're proud of what I've done so far.

Contents

Abstract iv

Dedication vi

Contents..... vii

List of Tables.....xi

List of Figures xii

Acknowledgements xvii

1. Introduction 1

  1.1 Where Are The Wonder Drugs?.....1

    1.1.1 Antimicrobials.....5

      1.1.1.1 Penicillins 10

      1.1.1.2 Cephalosporins.....14

    1.1.2 Antimicrobial-Resistant Microbes.....18

  1.2 Structure-based Protein Redesign 19

    1.2.1 3D Protein Structures from X-ray Crystallography.....20

    1.2.2 Methods of Protein Redesign 23

  1.3 Protein Redesign Algorithms.....23

    1.3.1 Protein Redesign Challenges 23

    1.3.2 OSPREY Suite of Protein Redesign Algorithms.....24

  1.4 Summary of Chapters 25

2. Solution-Phase NMR Studies of Wild-Type and Mutant PheA 27

2.1 Introduction 27

2.1.1 NRP Antibiotics 27

2.1.2 NRP Antibiotic Biosynthesis 28

2.1.3 Modifying Antibiotic Biosynthesis Machinery 35

2.1.4 Structure-Based Computational Protein Redesign 37

2.1.5 Structure Determination, by Solution-Phase NMR, of wild type and mutant PheA 41

2.2 Methods 43

2.3 Results 44

2.4 Discussion 49

3. Using Computational Protein Redesign to Develop Novel Antibiotics to Combat Vancomycin Resistance 52

3.1 Introduction to Vancomycin 52

3.1.1 Vancomycin-Resistant Microbes 57

3.1.2 Novel Approaches to Combat Vancomycin Resistance 60

3.1.2.1 Computational Redesign of Nonribosomal Peptide Antibiotics 60

3.1.2.2 Computational Redesign of Vancomycin Glycosyltlransferases to Chemoenzymatically Produce Glycovariants of Vancomycin 61

3.1.3 Biosynthesis of Wild-Type Vancomycin 62

3.1.4 Vancomycin Glycosyltlransferases 66

3.2 Materials and Methods 71

3.2.1 Materials 71

3.2.2 Protein Expression and Purification 71

3.3 Results 72

3.3.1 GtfB 72

3.3.2 GtfD 74

3.3.2.1 Determination of the active site residues of GtfD that bind the sugar-donor substrate 75

3.4 Discussion 77

3.5 Future Work 79

3.5.1 Determining the GtfB amino acids involved in binding both substrates 79

3.5.2 Finding optimal growth conditions for *Amycolatopsis orientalis* 82

3.5.3 Purifying vancomycin and glycovariants of vancomycin 84

3.5.4 Determining the GtfD amino acids involved in binding both substrates 84

3.5.5 Selecting a non-native GtfB sugar-donor substrate to be our first redesign target 85

3.5.6 Selecting a non-native GtfD sugar-donor substrate to be our first redesign target 85

3.5.7 Finding or adapting a gene modification protocol for *Amycolatopsis orientalis* 86

3.5.8 Determining the enzyme kinetics for wild type and mutant GtfB and GtfD87

4. Rationally Designed Enzymes and Enzyme Inhibitors, to Understand and Control Glycosyltransferases 88

4.1 Introduction 88

4.1.1 Carbohydrates and Glycans in Biology 88

4.1.2 The Glycan Processing Enzymes (GPEs) 89

4.1.3 Understanding How Glycosyltransferases Work 96

4.2 Inhibitors of Glycosyltransferases 100

4.2.1 Tunicamycin 100

4.2.2 Inhibitors of β1-4 Galactosyltransferase (β1-4 Gal T) 106

4.2.2.1 Acceptor-based Inhibitors of β1-4 Galactosyltransferase (β1-4 Gal T) .. 110

4.2.2.2 Donor-based Inhibitors of β1-4 Galactosyltransferase (β1-4 Gal T) 112

4.2.3 Inhibitors of Sialyltransferases 119

4.2.4 Inhibitors of The ABO Blood Group Glycosyltransferases 133

4.3 Redesigns of Glycosyltransferases 143

4.3.1 Redesigns of The ABO Blood Group Glycosyltransferases 143

4.3.2 Redesigns of The Glycopeptide-Antibiotic Glycosyltransferases 152

4.4 Conclusion 165

5. Conclusions and Future Work 169

5.1 Develop a novel protein solubility enhancement tag 169

5.2 Use wild type PheA as a model system of a large protein with low solubility . 178

5.3 Use set of wild type and redesigned enzymes to study the effects of engineered glycosylation on their enzymatic activities 182

References 184

Biography 203

List of Tables

Table 1: Definitions of the four cephalosporin generations, by spectra of activity 15

Table 2: Some early NRP antibiotics that became clinically useful 28

Table 3: Amino acid sequence comparison of two NRP antibiotics 31

Table 4: Amino acid sequence comparison of Gramicidin S and the Tyrocidines A-D 33

Table 5: Amino acid sequences of all six isoforms of the NRP antibiotic Gramicidin D 34

Table 6: Summary of Successful wt-PheA Preparations 44

Table 7: Summary of Solubility Testing With wt-PheA-His6 47

Table 8: Summary of Solubility Testing With His6-ThrombinSite-wt-PheA 49

Table 9: Amino acids in the heptapeptide core of vancomycin-group glycopeptides 56

Table 10: Computational Redesign Targets 61

Table 11: Mutability of GtfD amino acids within 4 Å of the sugar-donor substrate 76

Table 12: Growing *Amycolatopsis orientalis* to Produce Vancomycin 83

Table 13: Inhibition of GalTs, by compounds in Fig. 40, or Uridine Diphosphate (UDP) 116

Table 14: Kinetics of inhibition of ST6 Gal-I, for inhibitors shown in Figure 43 124

Table 15: Kinetics of inhibition of ST3 Gal members, inhibitors in Figures 45 and 46 128

Table 16: Kinetics of inhibition of human ABO blood group system glycosyltransferases (compounds in Figure 50) 138

Table 17: Kinetics of inhibition of nonionic mimics of UDP and UDP-Gal (with Mg^{2+}) against human ABO blood group glycosyltransferases (inhibitors in Figure 51) 141

Table 18: Nomenclature system for hybrid and mutant glycosyltransferases, GTA and GTB, of the human ABO blood group system. 144

List of Figures

Figure 1: Penicillin Class antimicrobials: (a) natural penicillins, (b) penicillin nomenclature, (c) some semi-synthetic penicillins ... 11

Figure 2: Cephalosporin Class antimicrobials: (a) starting material for semi-synthetic cephalosporins, (b) synthetic intermediates, (c) some semi-synthetic cephalosporins (first and second generation) ... 16

Figure 3: The NRPS assembly line for Gramicidin S [Reprinted from Stevens, et al., 2006a] .. 29

Figure 4: Chemical reactions catalyzed by the core domains of the NRPS enzymes [Reprinted from Kopp and Marahiel, 2007] ... 30

Figure 5: Tyrocidine NRPS enzymes [Reprinted from Linne, et al., 2003] ... 32

Figure 6: K*-predicted structure of the lowest-energy conformation of T278L/A301G binding with the redesign-target, Leu, in the active site [Reprinted from Chen, et al. 2009] ... 40

Figure 7: Specificity ratios for WT and mutant PheA redesigned to adenylate Leu [Reprinted from Chen, et al. 2009] ... 41

Figure 8: [^{1}H, ^{15}N] TROSY-HSQC of perdeuterated [^{15}N]-labeled wt-PheA, at 0.25mM. ... 45

Figure 9: [^{1}H, ^{15}N]TROSY-HNCO of perdeuterated [^{13}C, ^{15}N]-labeled wt-PheA, at 1.0 mM. (as in Figure 8, ^{1}H ppm is along the horizontal axis, ^{15}N ppm is along the vertical axis) ... 46

Figure 10: Some glycosylated secondary metabolites of pharmacological importance. [Reprinted from Barton, et al., 2001] ... 52

Figure 11: Some glycosylated secondary metabolites of pharmacological importance produced by actinomycetes. [Reprinted from Salas and Mendez, 2007] ... 54

Figure 12: Hydrogen bonds between vancomycin and normal PG-crosslinking peptide analog. Asterisk indicates the nitrogen that is replaced by oxygen in D-Ala-D-Lactate terminus. [Reprinted from Nitanai, et al., 2009] ... 56

Figure 13: Chemical structures of two members of the vancomycin group of glycopeptide antibiotics. [Reprinted from Groves, et al., 1994] ... 58

Figure 14: Phase 1 of biosynthesis of vancomycin. Three NRPS enzymes are responsible for selecting, activating, and linking seven amino acids into the heptapeptide chain shown. [Reprinted from Hubbard and Walsh, 2003] ... 63

Figure 15: The four known types of glycopeptides are distinguished by the amino acids at positions #1 and #3 in the heptapeptide core. [Reprinted from Rao, et al., 1995] 63

Figure 16: Phase 2 of biosynthesis of vancomycin. The production of the crosslinked heptapeptide core by the oxidative coupling of three pairs of aromatic rings. [Reprinted from Hubbard and Walsh, 2003] 65

Figure 17: Phase 3 of biosynthesis of vancomycin. Characteristic glycosylation of the crosslinked heptapeptide core shown for two members of the vancomycin group of glycopeptides; vancomycin (left) and chloroeremomycin (right). Each sugar molecule is labeled with its name and the abbreviated name of the glycosyltransferase that catalyzes its covalent attachment. [Reprinted from Walsh, et al., 2003] 65

Figure 18: Summary of the chemical reactions that are catalyzed by glycosyltransferases (GTs). The product molecules will have a new glycosidic bond with either inversion or retention of the anomeric configuration of the glycosidic bond broken in the sugar-donors, depending on the GT. [Reprinted from Coutinho, et al., 2003] 67

Figure 19: Mechanism of catalysis for an inverting glycosyltransferase with TDP-α-glucose as its sugar-donor substrate, creating a glycosidic bond with β configuration. [Reprinted from Walsh, et al., 2003] 68

Figure 20: The sequence-based classification system CAZy, for glycosyltransferases (GTs). [Reprinted from Coutinho, et al., 2003] 70

Figure 21: [^{1}H, ^{15}N] HSQC spectrum of ^{15}N-labeled GtfB-His$_6$ 73

Figure 22: Two glycosyltransferases with the GT-B fold [Reprinted from Walsh, et al., 2003] 80

Figure 23: The most common forms of the nine monosaccharides commonly found in vertebrates. 90

Figure 24: The glycolipid used as the donor-substrate by mammalian oligosaccharyltransferase (OST). [Reprinted from Varki, et al., 2009, Ch. 8] 92

Figure 25: The monosaccharide symbols used here to represent glycan structures. [Reprinted from Varki, et al., 2009] 92

Figure 26: Typical mature GlcNAcβ1-Asn N-glycans. [Reprinted from Varki, et al., 2009, Ch. 8] 93

Figure 27: The reaction catalyzed by Glucosylceramide Synthase (GCS), and the inhibitor miglustat. 95

Figure 28: A typical nucleotide sugar donor substrate, and the products of sugar transfer catalyzed by either a retaining GT or an inverting GT. 97

Figure 29: Possible reaction mechanisms for inverting and retaining GTs. The M^{2+} metal ion shown is required by all GT-A fold enzymes, but not GT-B fold GTs. 99

Figure 30: The reaction catalyzed by the bacterial transferase MraY, and the inhibitor tunicamycin. 101

Figure 31: The reaction catalyzed by the bacterial transferase MraY, and inhibited by tunicamycin, in context of the membrane steps of peptidoglycan biosynthesis. [Reprinted from Bouhss, et al., 2008] 102

Figure 32: The reaction catalyzed by the eukaryotic GlcNAc-1-Phosphotransferase, ALG7, is at the beginning of the membrane-associated biosynthesis of the dolichol-linked 14-sugar precursor of N-linked glycans. [Reprinted from Varki, et al., 2009, Ch. 8] 104

Figure 33: The reaction catalyzed by GlcNAc-1-Phosphotransferase, and the inhibitor tunicamycin. 105

Figure 34: The biosysnthetic maturation of N-glycans produces diversity in the branching structure and in the composition of the terminal sugar residues, depending on the cellular environment of these N-linked oligosaccharides. Symbol key on next page. [Reprinted from Varki, et al., 2009, Ch. 13] 106

Figure 35: The modification of a terminal GlcNAc residue by a β1-4 glactosyltransferase (β1-4 Gal T) yields a Type-2 unit (LacNAc, or LN), while modification by a β1-3 glactosyltransferase (β1-3 Gal T) yields a Type-1 unit. R stands for an N-glycan, O-glycan, or glycolipid. [Reprinted from Varki, et al., 2009, Ch. 13] 108

Figure 36: The reaction catalyzed by β 1-4 Galactosyltransferase (β1-4 Gal T). 110

Figure 37: Acceptor-based inhibitors of β1-4 Galactosyltransferase (β1-4 Gal T). [Brown, et al., 2009] 111

Figure 38: Some Lewis blood group antigens. [Reprinted from Varki, et al., 2009, Ch. 13] 111

Figure 39: Early donor-based inhibitors of β1-4 Galactosyltransferase (β1-4 Gal T). [Takaya, et al., 2005] 113

Figure 40: Recent donor-based inhibitors of (β1-4 Gal T). [from Pesnot, et al., 2010 and Wang, et al., 2013] 115

Figure 41: The Lewis blood group antigens (Type-1 and Type-2 Lewis determinants). Type-1 and Type-2 units differ only in the linkage of the outermost galactose. [Reprinted from Varki, et al., 2009, Ch. 13] 120

Figure 42: Reaction catalyzed by ST6 Gal-I 122

Figure 43: Donor- and Transition State-based inhibitors of ST6 Gal I 123

Figure 44: The α2-3 sialyltransferases, the ST3Gal family, all use CMP-Neu5Ac as their donor substrate. Each family member has its own preferred acceptor substrate. [Reprinted from Varki, et al., 2009, Ch. 13] 127

Figure 45: Competitive inhibitors of ST3Gal-I, found in large-scale screening efforts. 128

Figure 46: Inhibitors of ST3 Gal members 129

Figure 47: Fluorinated analogues of the sialic acid Neu5Ac 132

Figure 48: Biosynthesis of Type-2 H, A, and B blood-group antigens. R represents an N-glycan, O-glycan, or a glycolipid. [Reprinted from Varki, et al., 2009, Ch. 13] 135

Figure 49: Reactions catalyzed by the ABO blood group system glycosyltransferases GTA and GTB. R represents either an N-glycan, O-glycan, or a glycolipid. 137

Figure 50: Human blood group ABO system H antigen-mimic (37) and C-3-modified analogues 138

Figure 51: Nonionic mimics of UDP and UDP-Gal 141

Figure 52: Chemical structures of the first clinically useful glycopeptide and lipoglycopeptide antibiotics: vancomycin (natural), teicoplanin (natural), and telavancin (synthetic derivative of vancomycin) 154

Figure 53: The glycosyltransferases, and the monosaccharides that they each transfer, in the biosynthesis of the natural GPAs vancomycin and chloroeremomycin 156

Figure 54: FDA-approved GCS inhibitors for the treatment of Type 1 Gaucher's disease 166

Figure 55: Lipoglycopeptides approved by the FDA in 2014 167

Figure 56: Combined chemoenzymatic result of NGT and α6GlcT, in the presence of UDP-glucose (this is Figure 5, reprinted from Schwarz, et al., 2011) The number of glucose units added by α6GlcT depends on the concentration of UDP-glucose. 177

Figure 57: The three N-X-S/T sequences in wt PheA, shown in a ribbon diagram (top) and in a surface model (bottom) of the crystal structure of wtPheA (PDB ID = 1AMU). The unordered N-terminal tail is not shown beyond G17. Drawn with PyMOL (edu, 1.3). 180

Acknowledgements

Thank you to all of my friends at Duke, in the Computer Science Department, and in the Donald Lab.

Thank you to my very patient thesis committee and the Computer Science Department for giving me the opportunity to finish my dissertation after numerous personal and family health emergencies. Thank you in particular to my advisor Bruce Donald for all the years of support and advice and to Dave Richardson for many hours of supportive conversations and thesis writing advice.

Thank you Michelle. Your friendship and phone calls, from California, during all of the difficult times, and especially the last few years of grad school, enabled me to keep my head “above water” so that I could write this dissertation and prepare my defense seminar.

1. Introduction

1.1 Where Are The Wonder Drugs?

People, of a sufficient age, living in "first world" countries have witnessed decades of increasingly amazing technological accomplishments, some of which seemed like science fiction. After the United States successfully sent men to land on the moon and return safely to earth, more than once, manned space travel was taken for granted by many. Expectations grew for one "next great thing" after another, as a high-tech future without limits seemed inevitable. Now, in an era of realizing our limitations, the phrase "Where's my jetpack?" uttered with comic indignation signifies an acceptance that some were led to adopt unrealistic expectations for the future, or at least for the near future. During those same decades of very public technological advances there was a less-visible, but equally fantastic, period of advances in biochemistry and medical research sped up by advances in technology, and by interdisciplinary collaborations, in numerous research labs. The results, in the first world, include longer life expectancies, greater health expectancies, and the expectation that every successive generation will enjoy more of these benefits than the preceding one. One consequence of decades of biomedical research was an impressive, and still growing, set of "wonder drugs" to combat many of the pathogenic microbes that infect man. Now, in this era of new limitations, the sobering realization that motivates my thesis is embodied in the ominous question: "Where are the Wonder Drugs?"

Penicillin was the first wonder drug of the 20th century. It was discovered by Alexander Fleming in 1928, who showed its breadth of "bacteriolytic" powers, against many pathogenic bacteria, at least a decade before the terms "antibiotic" or "wonder drug" had been given there present meanings. Penicillin was the first clinically useful antibiotic, but it was not readily available for over a decade when several chemists finally worked out a method to mass produce it in a highly purified form. After some initial human testing in the early 1940s purified penicillin was finally released and rushed to the battlefields and hospitals of World War II in 1944. Penicillin worked so well that the 1945 Nobel Prize in Medicine or Physiology went to Fleming and the two primary chemists who produced it in purified form.

Penicillin's nickname, "wonder drug," was soon shared by other antibiotics discovered in the 1940s such as the first aminoglycosides, streptomycin (1943) [Schatz and Waksman, 1944] and neomycin (1948), discovered by Selman Waksman. The term "antibiotic" was first used in 1942, by Waksman and his collaborators, to refer to natural products such as penicillin [Waksman, 1947] and was formally defined by Waksman in 1947, at the request of the editor of Biological Abstracts. The meaning of "antibiotic" has changed in two major ways over the years since then. In the 1940s, an antibiotic was not limited to only antibacterials, but referred to all antimicrobials. Also, the old definition of antibiotic was limited to include only natural products "produced by micro-organisms" [Waksman, 1947]. To avoid confusion, the meaning for "Wonder Drug" that I will use is any of the known antimicrobial drugs which, according to the Centers for Disease Control and Prevention (CDC), now includes antibacterial drugs, antiviral drugs, antifungal drugs, and antiparasitic drugs. Our current set of Wonder Drugs faces two main challenges: (i) the arrival of new pathogenic microbes and (ii) the arrival of new versions of old pathogenic microbes that have evolved to become resistant to the Wonder Drug that used to stop them.

The currently untamed Ebola outbreak in West Africa is an example of a "new" pathogenic microbe, a virus first encountered in an outbreak in Zaire in 1976 [An International Commission, 1978], for which there is still neither an effective antimicrobial drug nor a vaccine available. In April [Baize, et al., 2014] it was announced that this current outbreak began in Guinea, in December 2013, and was due to the *Zaire ebolavirus* strain, which was the original strain of Ebola found to infect humans 37 years ago in Zaire. The first three victims were two young girls and their mother, in a village in Guinea that is approximately adjacent to Guinea's borders with both Liberia and Sierra Leone. The 2-year old girl died on December 6th, her mother died on December 13th, and her 3-year old sister died on December 29th 2013. Then the grandmother of these two girls died January 1st 2014, followed by a nurse and a village midwife, who both died on February 2nd 2014, after the village midwife had spread Ebola to two other villages. As of August 22, 2014, the CDC reports [CDC web site] that this current Ebola outbreak has now spread to four countries: Guinea, Liberia, Sierra Leone, and Nigeria and the number of suspected and confirmed cases of Ebola infection is 2615. Of these, 1427 are believed to have died from the Ebola virus disease (EVD). Clearly, the difficulty in controlling the spread of this Ebola virus in these four countries is exacerbated by factors such as local funeral and burial traditions and the under-developed, or under-repair, public health infrastructure there. But, the lack of a Wonder Drug, or vaccine, for this "new" virus, Ebola, has made it difficult for developed countries to provide rapid and effective assistance to the people in West Africa who are either suffering from EVD or living in fear of it.

My thesis will address the other problem that threatens to make us ask "Where are the Wonder Drugs?," the arrival of antimicrobial-resistant versions of old pathogenic microbes that are now undeterred by any of our Wonder Drugs. In this first chapter, I will present a brief overview of antimicrobials, followed by a summary of the current status of the problem of antimicrobial-resistant microbes. Then I will give a brief overview of structure-based protein redesign, followed by a description of the suite of algorithms that we use in the Donald Lab to computationally redesign the enzymes that make up the machinery by which an antimicrobial is biosynthesized. In my thesis I will focus on efforts to develop new methods to develop new antimicrobials that are derivatives of antimicrobials already in our set of Wonder Drugs.

1.1.1 Antimicrobials

We live in a world where single-celled microbes can live almost anywhere and in quantities that far exceed the total number of all metazoans. Considering the differences in size, this is hardly surprising. What is surprising, and known as a result of ongoing large-scale research, is the size, diversity, and biochemical role of the human microbiome, the collection of all the microbes living on and inside a typical healthy human being. A microbiome includes bacteria, archaea, viruses, fungi, and single-celled eukaryotes such as yeast, all residing on or in one host [American Academy of Microbiology, 2014]. One human microbiome includes about 100 trillion bacterial cells, which is about three times the total number of human cells in the body. With the number of viruses in a microbiome estimated to be five times the number of bacteria, all the microbes in one healthy human microbiome would weigh more than one kilogram and far outnumber all the human cells [American Academy of Microbiology, 2014]. Thus, humans are actually supraorganisms, consisting of more microbial cells than human cells, carrying a microbiogenome that is many times larger than the "human genome" first sequenced in 2003 [Pennisi, 2003], after 12.5 years of work at 20 sequencing centers in six countries [Austin, 2004].

The size and diversity of the human microbiome illustrates why human health depends on healthy immune systems. One healthy human microbiome contains about one thousand different species [American Academy of Microbiology, 2014]. Different regions of the human body support different subsets of these 1,000 species and every human being has unique and characteristic distributions of these 1,000 species that remain stable during good health [The Human Microbiome Project Consortium, 2012]. Most of these microbes are essential for human health and only become pathogenic when they invade the wrong part of the body. Humans have two immune systems, innate and adaptive, that kill and remove invading pathogenic microbes and possibly work to maintain well balanced ecosystems of microbes throughout a healthy body. Medical intervention can help to combat pathogenic microbes in two different ways. Vaccines can prepare the adaptive immune system before an infection occurs. Antimicrobials can directly kill, or stop the growth of, microbes after an infection of microbes that were not detected by, or that has overwhelmed, the immune systems. Medical research into the development of vaccines began in the 1800s, after Edward Jenner, in 1798, coined the term vaccination and demonstrated that inoculation with cowpox virus led to immunity to smallpox [Riedel, 2005], long before any laboratory research and development of antimicrobials.

Many microbes are equipped to engage in chemical warfare with their neighbors, in order to kill off any nearby microbes, different from themselves, that are competing with them for a limited food supply. The first recorded observation of such antagonistic relationships between microbes was made by Louis Pasteur in 1877 when he observed "that saprophytic bacteria may, under certain circumstances, decrease the infective power of the anthrax bacillus" [Dubos, 1941]. Two years later Pasteur discovered, by accident, how to attenuate the chicken cholera bacteria (*Pasteurella multocida*) such that when it was injected into chickens they survived, after showing only mild symptoms of cholera, and afterwards were immune from injections of the live bacteria and thus, in 1879, produced the first laboratory-developed vaccine [The College of Physicians of Philadelphia, 2014]. Pasteur then "created several veterinary vaccines before his development of the rabies vaccine in 1885 for use in humans" [The College of Physicians of Philadelphia, 2014].

As the 1800s came to an end, research into vaccine development was producing good results, while the search for antimicrobials had so far only turned up compounds like pyocyanase, extracted from cultures of *Pseudomonas pyocyanea* [Emmerich and Loew, 1899]. Pyocyanase showed "marked bacteriostatic and bactericidal effect against many unrelated microbial species" [Dubos, 1941] but suffered from great toxicity for animal cells. In 1906 Alexander Fleming joined the lab of a successful and powerful immunologist, Almroth Wright, who believed that developing vaccines was the only viable way to combat the pathogenic microbes that infect man [Lightman, 2005]. Fleming was still working under Almroth Wright's supervision in 1928 when he famously observed, by accident, that "around a large colony of a contaminating mould the staphylococcus colonies became transparent and were obviously undergoing lysis" [Fleming, 1929]. Fleming, a self-described "immunologist working in a laboratory almost entirely devoted to immunology" [Fleming, 1945b], continued to be interested in the "bacteriolytic substance" formed in that mold culture, though his boss was not. Fleming said in his 1945 Nobel Lecture that he "referred again to penicillin in one or two publications up to 1936 but few people paid any attention" [Fleming, 1945a].

It took 16 years for the first wonder drug, a microbial natural product, to go from Fleming's discovery to it being a clinically useful antimicrobial. Fleming was not a chemist, so he tried several times to find one who could figure out how to purify penicillin from cultures of the *Penicillium* mold. The chemists he got to work on this were unsuccessful [Lightman, 2005]. Finally, in 1938 Fleming got two chemists at Oxford, Ernst Chain and Howard Florey, to work on purifying penicillin from filtrates of broth cultures of the mold. Chain and Florey and five collaborators at Oxford succeeded in a few years, were able to concentrate penicillin by 40,000 fold by producing it in pure crystalline form [Lightman, 2005], and published their preliminary results in 1940 [Chain, et al., 1940]. The medical research community in the late 1930s had taken more interest in chemotherapy and efforts to find and develop antibacterial drugs due to several important events that occurred during the decade following Fleming's discovery of penicillin.

In the early 1930's chemists at the Bayer Company in Germany, already making synthetic dyes and some drugs, such as Aspirin, were making and testing dyes and related compounds, directed by Gerhard Domagk, to find any with antibacterial properties. This led to the discovery in 1932 of a sulfonamide azo dye, later named Prontosil, which became the world's first synthetic antimicrobial drug to be marketed. After successful clinical trials of Prontosil, begun in 1933, the age of sulfa drugs began. Domagk published the work on Prontosil in 1935 [Otten, 1986; Raju, 1999; Bentley, 2009] and was awarded the Nobel prize in Physiology or Medicine in 1939. Fleming remarked in his 1945 Nobel lecture that "the introduction of sulphonamide had completely changed the medical mind in regard to chemotherapy of bacterial infections" [Fleming, 1945a]. In the late 1930's the American microbiologist Rene Dubos was publishing accounts of an antimicrobial natural product that he soon named Gramicidin [Dubos, 1939; Hotchkiss and Dubos, 1940a; Hotchkiss and Dubos, 1940b]. Fleming said in his

Nobel lecture that it was “after Dubos had shown that a powerful antibacterial agent, gramicidin, was produced by certain bacteria that my co-participators in this Nobel Award, Dr. Chain and Sir Howard Florey, took up the investigation” [Fleming, 1945a]. I will have more to say about gramicidin in Chapter 2.

1.1.1.1 Penicillins

The penicillin submitted for clinical trials in 1941 was purified by Chain et al. [Chain, et al., 1940] from Fleming’s strain of the mold *Penicillium notatum* when grown in surface culture. The penicillin that was finally available in large quantities for clinical use in 1944, on the battlefields of World War II, was purified from an improved strain of *Penicillium notatum* that was grown in massively scaled up submerged fermentation conditions. The chemical structure of these penicillin molecules could not be determined until 1945 [Bentley, 2009]. Only then was it realized that the penicillin that passed clinical trials in 1941 had a different chemical structure than the penicillin that was mass-produced and used in 1944. Luckily, that didn’t matter. The penicillin made and tested in 1941 was then named Penicillin F (for Florey, or Fleming), and the penicillin that was first used clinically, beginning in 1944, was named Penicillin G (simply because that comes after F) [Bentley, 2009]. The structures of these two penicillins are shown in Figure 1.

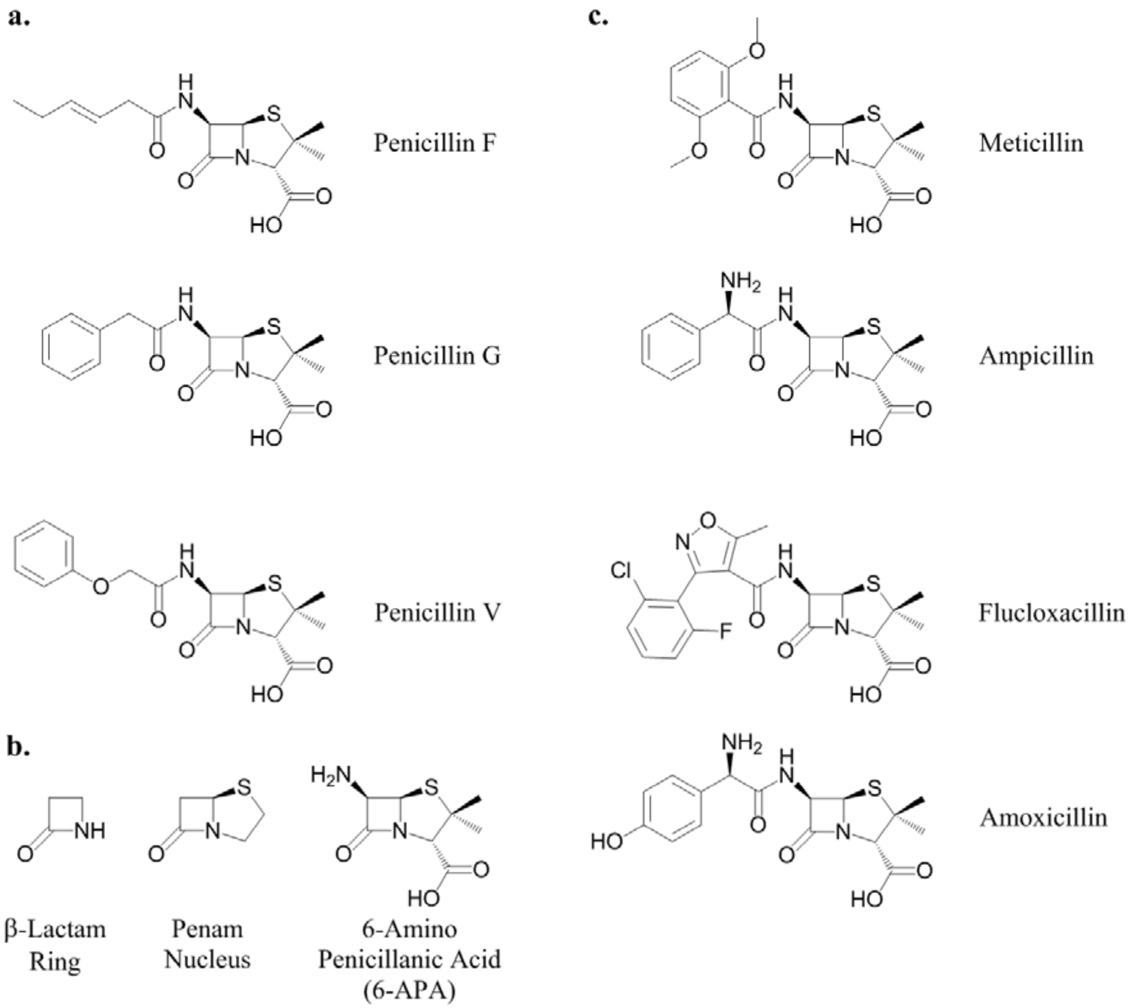

a.
Penicillin F
Penicillin G
Penicillin V
b.
β-Lactam Ring
Penam Nucleus
6-Amino Penicillanic Acid (6-APA)
c.
Meticillin
Ampicillin
Flucloxacillin
Amoxicillin

Figure 1: Penicillin Class antimicrobials: (a) natural penicillins, (b) penicillin nomenclature, (c) some semi-synthetic penicillins

As soon as the chemical structures of penicillins F and G were known in 1945, structure-activity studies could begin. However, the wonder drug penicillin gave the Allied forces a huge military advantage during the end of World War II. Thus, research on penicillin in the mid 1940s, including a large transatlantic "secret project" from 1943-1945 aimed at discovering a purely synthetic route to produce it, was conducted without any results being published in the scientific journals [Curtis and Jones 2007]. Instead, the research results of hundreds of scientists were published together in 1949 in one book, *The Chemistry of Penicillin*, of over one thousand pages [Clarke, et al., 1949]. That book summarized what has been called "the most intensive investigation of any chemical problem ever undertaken" [Curtis and Jones 2007]. This massive effort failed to discover a chemical synthesis of penicillin. Instead, it was found that growing the *Penicillium* mold in fermentation culture with various side-chain-precursor molecules added to the growth medium resulted in variants of Penicillin G being produced [Rolinson and Geddes, 2007]. Many structural variants of penicillin were purified from such cultures and screened for their antibacterial properties, but none were found to be superior to Penicillin G. One of these many variants was named Penicillin V (see Figure 1), but its valuable properties, acid stability and absorption when taken orally, were not realized until scientists at an Austrian company published their discovery of this new penicillin in 1954 [Brandl and Margreiter, 1954]. In 1957, Penicillin V became the first natural penicillin to finally be made by total chemical synthesis [Sheehan and Henery-Logan, 1957].

Another breakthrough in 1957 was the discovery that 6-amino penicillanic acid (6-APA, Figure 1) could be purified from penicillin fermentations, without added side chain precursors, using an isolate of the mold *Penicillium chrysogenum* [Batchelor, et al., 1959]. The availability of highly purified 6-APA led to the era of making semi-synthetic variants of penicillin, by chemically attaching to the 6-amino group of 6-APA practically any side chain that terminated in a carboxylic acid. In 1957, the penicillin class of antimicrobials being used clinically had only two members, penicillin G and penicillin V, but that "grew to more than 20 different compounds in clinical use by the end of the 1970s" [Rolinson and Geddes, 2007]. The four semi-sysnthetic penicillins shown in Figure 1 were some of the first, and most successful, variants of penicillin to make it through clinical trials and into the clinic after 6-APA became available. Meticillin (originally called Methicillin until 2005) was introduced in 1960 [Rolinson, et al., 1960], followed by Ampicillin in 1961, the Meticillin-replacement Flucloxacillin in 1970, and the improved-ampicillin Amoxicillin in 1972. Amoxicillin "became the most widely prescribed antibiotic in clinical practice" [Rolinson and Geddes, 2007].

All of the antimicrobials shown in Figure 1 are in the Penicillin class because they each contain the "penam" nucleus, coined in 1953 [Sheehan, et al., 1953]. The penicillin class is one member of the β-Lactam Family of antimicrobials. All β-Lactam antimicrobials have a β-Lactam ring (see Figure 1) and the same general mechanism of killing microbes. Each drug interferes with microbial cell wall biosynthesis by binding to penicillin-binding proteins (PBPs).

1.1.1.2 Cephalosporins

Another very large class of antimicrobials in the β-Lactam Family is the Cephalosporin Class. The first cephalosporin was discovered in the late 1940s by Giuseppe Brotzu on the island of Sardinia. Brotzu found a fungus, in seawater near a sewage outlet, with mild activity against both gram-positive and gram-negative bacteria. He sent samples of this fungus, initially identified as a species of *Cephalosporium*, to collaborators at Oxford where several antibacterial substances were purified from these fungal cultures. One of these was named Cephalosporin C, and its structure is shown in Figure 2. While cephalosporin had a good broad spectrum of activity, it had only moderate activity. The success in the 1960s with making semi-synthetic variants of penicillin G motivated similar efforts to make variants of cephalosporin C. This required a lot more effort than making variants of penicillin G because the synthetic intermediate analogous to 6-APA for penicillin could not be made by fermentation. Eventually enzymes were found and developed that would hydrolyze the peptide bond involving the 7-amino nitrogen and produce 7-Amino Cephalosporanic Acid (7-ACA) shown in Figure 2, below cephalosporin C. Below 7-ACA are shown two examples of semi-synthetic cephalosporins made from 7-ACA. At the top of the right hand column of Figure 2 is cephalosporin G, which must be made by chemically modifying penicillin G. Cephalosporin G can be enzymatically converted to 7-Amino Deacetoxy

Cephalosporanic Acid (7-ADCA), which is a synthetic intermediate for semi-synthetic cephalosporins such as the two shown below 7-ADCA in Figure 2.

The cephalosporin class is currently divided into four generations. These generations are based on each drug's spectra of activity and the currently accepted definitions [Laudano, 2011] are summarized in Table 1. Three of the four semisynthetic cephalosporins shown in Figure 2 are considered to be first generation, and the remaining one, Cefoxitin, is considered to be in the second generation.

Table 1: Definitions of the four cephalosporin generations, by spectra of activity

Cephalosporin Generation	Activity against Gram-positive bacteria	Activity against Gram-negative bacteria	Activity against Other microbes
First	++	−	
Second	++	+	
Third	+	++	
Fourth	++	+++	*Pseudomonas* spp Enterobacteriaceae

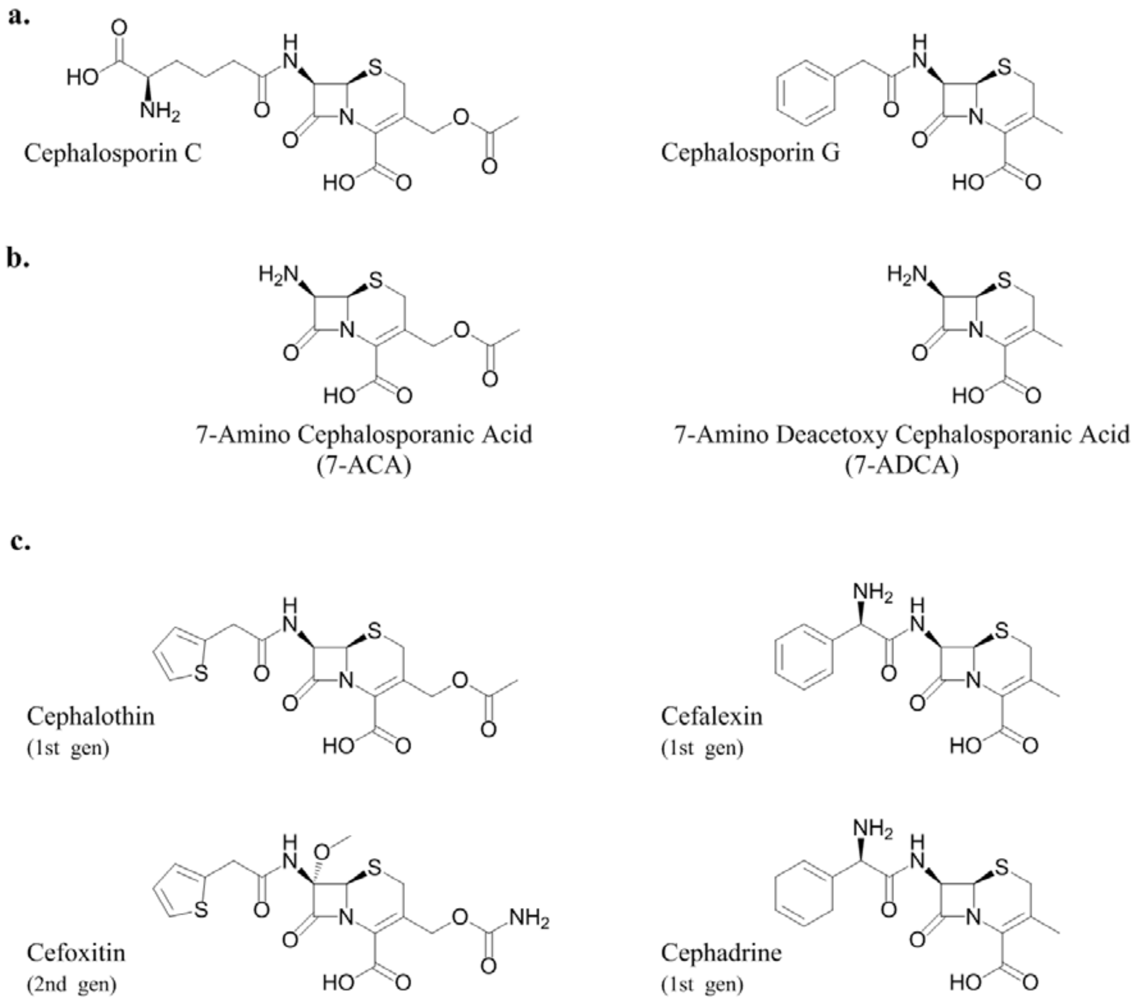

a.
Cephalosporin C
Cephalosporin G
b.
7-Amino Cephalosporanic Acid
(7-ACA)
7-Amino Deacetoxy Cephalosporanic Acid
(7-ADCA)
c.
Cephalothin
(1st gen)
Cefalexin
(1st gen)
Cefoxitin
(2nd gen)
Cephadrine
(1st gen)

Figure 2: Cephalosporin Class antimicrobials: (a) starting material for semi-synthetic cephalosporins, (b) synthetic intermediates, (c) some semi-synthetic cephalosporins (first and second generation)

Finally, international disasters have affected, and will affect, the efforts to combat microbial infections. Both World Wars had significant effects on the development of the first antimicrobials, both penicillin and Prontosil. During World War I, Alexander Fleming (1881-1955) served in England's Royal Army Medical Corps, in a wound research lab [Lightman, 2005], while Gerhard Domagk (1895-1964) served in the German Army, and was wounded [Bentley, 2009]. When World War II began in 1939, efforts were still underway to purify and concentrate penicillin. In 1941, the purified penicillin that Chain and Florey had produced was tested successfully in clinical trials. World War II motivated scaling up the production of penicillin but made it difficult to obtain the necessary materials in England. As a result, in mid 1941 Florey and colleague N.G. Heatley went to the United States and Canada and catalyzed the development of new large scale fermentation techniques and facilities to accomplish the massive production of penicillin needed on the battlefields of Europe [Bentley, 2009]. Penicillin made its clinical debut in World War II in 1944 and saved the lives of uncounted wounded soldiers. Now, in 2014, we are seeing a multinational effort develop to combat the world's largest ebola outbreak, which has now been growing in West Africa for nearly one year. This highly visible disaster will hopefully inspire more interest in developing new methods to prevent and combat deadly microbial infections.

1.1.2 Antimicrobial-Resistant Microbes

In 1929, Fleming coined the name "penicillin," to refer to a "filtrate of a broth culture" of what he believed to be the mold *Penicillium rubrum*, and determined that it had a "bacteriolytic action" [Fleming, 1929]. Fleming also discovered another important molecule that he described as a "bacteriolytic element" long before he discovered penicillin. Fleming discovered this "bacteriolytic element" in human nasal mucus when he, or someone in his lab, was suffering with a cold. Fleming noticed that this nasal mucus, diluted 1:5 in normal salt water, would inhibit the growth of a gram-positive cocci that he named *Micrococcus lysodeikticus*. Fleming published this finding in 1922, and named the bacteriolytic compound "lysozyme" [Fleming, 1922]. While lysozyme never became a clinically useful antimicrobial, Fleming commented in his 1945 Nobel banquet speech that lysozyme "paved the way for penicillin for me and I think also for my partners in this Nobel Award" [Fleming, 1945b] and Chain and Florey say in their 1940 paper that they decided to systematically investigate available antibacterial substances, including penicillin, "following the work on lysozyme in this laboratory" [Chain, et al., 1940].

Finally, we should not be surprised that microbes develop resistance to antimicrobials that they are exposed to at low doses. In his 1945 Nobel Lecture Fleming stated that there is the danger that underdosing with a non-lethal quantity of an antimicrobial may lead to the drug making the microbes become resistant to that drug. He even published just such experimental findings in 1922, with the "anti-bacterial substance" lysozyme [Fleming and Allison, 1922], just six months after publishing his discovery of lysozyme [Fleming, 1922].

1.2 Structure-based Protein Redesign

Hopes for an atomic-level understanding of the mechanisms and biochemical functions of proteins grew in 1953 after the publication of Watson and Crick's elegant, but "unproved," helical structure of double-stranded deoxyribonucleic acid (DNA) based on published experimental data and "stereochemical arguments" [Watson and Crick, 1953a]. New X-ray diffraction data [Wilkins, et al., 1953; Franklin and Gosling, 1953] published simultaneously with the Watson and Crick article supported the helical nature of their proposed model for the double-stranded structure of fibers of the salt of DNA. One month later Watson and Crick published [Watson and Crick, 1953b] their description of how the mechanisms of the biologically important functions of DNA could be understood by the structural features of their proposed model for DNA fibers as an anti-parallel helix of two complementary strands. The ability of this proposed three dimensional (3D) structure of DNA in 1953 to potentially reveal the mechanisms of heredity at an atomic-level must have motivated those already working to obtain X-ray crystal diffraction data on members of a more diverse family of biopolymers, the proteins.

1.2.1 3D Protein Structures from X-ray Crystallography

Five years after Watson and Crick published their famous model for the helical structure of double-stranded DNA, the first 3D structure of a 17 kDa protein, myoglobin from sperm-whale, was published at 6 Å resolution by Kendrew and colleagues [Kendrew, et al., 1958]. The low resolution electron-density map could only reveal the 3D locations of regions of the myoglobin molecule with high electron density. But, it was enough to show a complete lack of the structural symmetry seen in the structure of the DNA double helix. It was now apparent that getting an atomic-level understanding of the function of proteins was going to be much more difficult than doing the same was for DNA.

Two years later, 1960, the age of protein X-ray crystallography truly began [Richardson and Richardson, 2014] when two protein X-ray crystal structures were published simultaneously in Nature. One was a higher resolution, 2 Å, structure of the same sperm-whale myoglobin, also from Kendrew and colleagues [Kendrew, et al., 1960]. The other was the first structure of another, oxygen-carrying molecule, horse haemoglobin, at 5.5 Å resolution from Perutz and colleagues [Perutz, et al., 1960]. The much larger haemoglobin protein had a molecular weight of 67 kDa, and was a complex of four polypeptide chains of roughly equal length and mass, each being roughly equal in mass to the 17 kDa myoglobin molecule. This low resolution structure of horse haemoglobin revealed enough of the global fold of its four individual polypeptide chains that Perutz and his collaborators concluded that these subunits each have the same “polypeptide chain-fold which Kendrew and his collaborators first discovered in sperm whale myoglobin” and led to the speculation that “all haemoglobins and myoglobins of vertebrates follow the same pattern” [Perutz, et al., 1960]. At that time, the amino acid sequences of those different polypeptide chains with similar global folds was largely unknown.

Five years later protein sequencing had advanced a great deal and the sequence of all 129 amino acids in hen egg-white lysozyme was known when, in 1965, it became the first enzyme to have its 3D structure revealed by X-ray crystallography [Blake, et al., 1965]. Blake and his colleagues reported that, at 2 Å resolution, they could “find a continuous ribbon of high density with characteristic features at regular intervals” for each of the 129 amino acids, and they found regions of this ribbon with α-helical conformation, and “the four disulphide bridges can be identified unambiguously in the X-ray image” (B.2-9). Because competitive inhibitors of lysozyme were available, Johnson and Phillips simultaneously published their 6 Å structure of lysozyme-inhibitor complexes, that revealed "a part of the molecule which may be responsible for its enzymatic activity"[Johnson and Phillips, 1965].

Following these 3D structures of hen egg-white lysozyme (HEWL) in 1965 it became a popular model system and analogues of HEWL were sought for comparison. The lysozyme from bacteriophage T4 was purified, characterized, and sequenced in 1968, and found to resemble HEWL in many ways [Tsugita, et al., 1968]. Both lysozymes are muramidases, though HEWL hydrolyzed 64% of the N-acetylmuramide linkages in the standard *M. lysodeikticus* cell wall preparation, while T4 lysozyme only hydrolyzed 20% of those linkages. The T4 lysozyme 3D crystal structure, at 2.5 Å resolution, was published in 1974 [Matthews and Remington, 1974]. They concluded that while HEWL and T4 lysozyme hydrolyze the same linkage, their 3D structures are quite different, and "it is not clear at this time whether or not their respective mechanisms of catalysis may be related." Studying the atomic-level details of the active sites of these lysozyme molecules by making rational point mutations in them would need to wait for many technical advances to come after the lifting of the voluntary moratorium on the use and development of recombinant DNA techniques in the 1970s, which were put in place partially because of fears of accidentally converting an innocuous microbe into a pathogenic antibiotic-resistant microbe [Berg and Singer, 1995].

1.2.2 Methods of Protein Redesign

Since the advent of detailed 3D structures of enzymes nearly 50 years ago, beginning with egg-white lysozyme, hopes grew that knowledge of their structures and catalytic mechanisms would lead to the ability to redesign them and modify the reactions they catalyze. In 1995, one of the first enzymes to successfully have its active site rationally redesigned was the T4 lysozyme [Kuroki, et al., 1995].

1.3 Protein Redesign Algorithms

1.3.1 Protein Redesign Challenges

Doing computational protein redesign presents four main challenges to the limits of computational power, even with current high-power processors [Gainza et al., 2013]. First, there is the fact that as the number of amino acid positions being allowed to differ from the wild type increases, the number of protein sequences to be considered (the size of the sequence space to be searched) increases exponentially. Next, changing even one amino acid in the wild type sequence can give the new protein a new global minimum energy conformation (GMEC) which must be searched for in the protein's astronomically large conformation space. Third, when redesigning a protein to bind a target ligand, that ligand has a conformation space that increases exponentially with the number of rotatable bonds that it has, and to find the GMEC of the ligand-protein complex requires searching a space that is proportional to the product of the sizes of the conformational spaces of the ligand and the protein alone. Finally, an energy function is required in order to rank the conformations of each protein and ligand, alone or in complex, and the most accurate functions require quantum mechanical modeling of these molecules which is computationally impossible. Amazingly, with clever algorithms and carefully chosen simplifying assumptions, solution-based computational protein redesign (SCPR) is possible and does work.

1.3.2 OSPREY: Suite of Protein Redesign Algorithms

OSPREY is a free, open-source, suite of algorithms for doing SCPR developed by the Donald Lab [Gainza et al., 2013]. The OSPREY algorithms are built on three main design principles. First, proteins are conformationally dynamic molecules with flexible backbones and amino acid side-chains. Many SCPR algorithms make the simplifying approximation that flexible side-chains can be represented by a small library of conformations, observed in X-ray crystal protein structures, called discrete rotamers. OSPREY uses continuous rotamers instead, which more accurately represent the conformation space. OSPREY has the ability to model flexible protein backbones instead of the traditional rigid backbones, or as a set of discrete backbones. Second, proteins, and protein-ligand complexes, exist as ensembles of all possible conformations and OSPREY uses an ensemble-based algorithm, called K*, to more accurately approximate the association constant of a protein-ligand complex. K* uses ensembles limited to only the most probable conformations, based on their energy, and accurately ranks all possible protein sequences by Boltzman-weighting each conformation's contribution. Finally, OSPREY uses only provable algorithms guaranteeing that it finds all of the lowest-energy conformations for the particular input model. The benefit of this, in contrast to using heuristic methods, is that any error observed in the output must be due to error in the input model, which contains various simplifying assumptions, as well as measured data.

1.4 Summary of Chapters

In Chapter 2, I will describe our work on determining the solution-phase structure of a nonribosomal peptide synthetase adenylation domain called PheA, by nuclear magnetic resonance (NMR). PheA is a large protein, 563 amino acids, by NMR standards and we had good success with it until we found that PheA suffers from low solubility at temperatures above 25 °C. In Chapter 3, I will describe two approaches to generate variants of the powerful antibiotic vancomycin. We chose to pursue the approach of making glycovariants of vancomycin by redesigning the glycosyltlransferase enzymes involved in its biosynthesis. To this end we compared two methods of determining the active site residues of their substrate-binding sites that we would allow to be mutated *in silico* by the protein redesign algorithms in OSPREY. In Chapter 4, I will present a review of published efforts to understand and control the function of glycosyltransferases by small molecule inhibitors, with and without the benefit of high-resolution 3D structures of these enzymes, and by structure-based redesigning of these enzymes. In Chapter 5, I will describe a recent idea for using a redesigned glycosyltransferase to solve the low solubility problem of many large proteins being studied by solution-phase NMR, like what we experienced in the project described in Chapter 2, and the interesting findings that came out of investigating the feasibility of using PheA as a model protein with low solubility.

2. Solution-Phase NMR Studies of Wild-Type and Mutant PheA

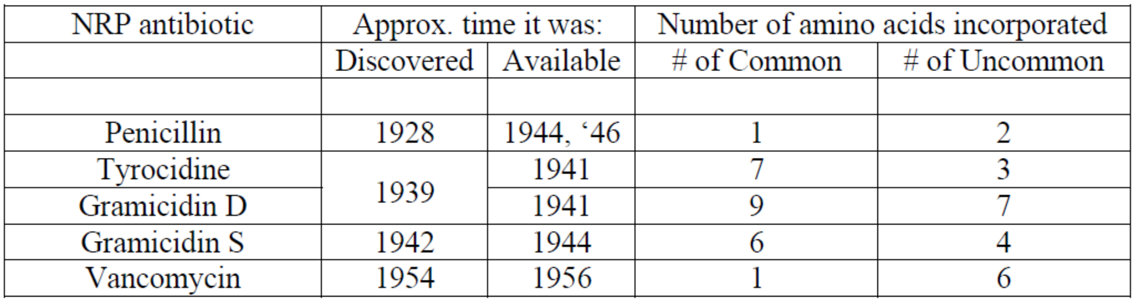
2.1 Introduction

2.1.1 NRP Antibiotics

We will be focusing on those antibiotics that are nonribosomal peptides (NRPs). The NRP antibiotics are a large family of secondary metabolites produced by single-celled microbes as a means of defending their territory. They are finely tuned molecules with the ability to kill, or inhibit the growth of, a variety of other microbes that are nearby and thus competing for the available nutrients. Their production is triggered by environmental cues.

The first commercially available antibiotics were NRP antibiotics. Table 2 contains information about penicillin and four NRP antibiotics that will be discussed further. The number of amino acids incorporated into each of these antibiotic molecules is shown there and divided into two categories. The “common” amino acids are those 20 that are incorporated into proteins by the ribosomal pathway. The large number of “uncommon” or nonproteinogenic amino acids used in these molecules is just one way in which the nonribosomal peptide synthetase (NRPS) pathway can produce small molecules with tremendous structural diversity.

Table 2: Some early NRP antibiotics that became clinically useful



NRP antibiotic	Approx. time it was:		Number of amino acids incorporated	
	Discovered	Available	# of Common	# of Uncommon
Penicillin	1928	1944, ‘46	1	2
Tyrocidine	1939	1941	7	3
Gramicidin D		1941	9	7
Gramicidin S	1942	1944	6	4
Vancomycin	1954	1956	1	6

2.1.2 NRP Antibiotic Biosynthesis

Molecules biosynthesized by a nonribosomal peptide synthetase (NRPS) pathway consist entirely, or only partially, of a chain of “n” amino acids connected by n-1 peptide bonds. The formation of each of these n-1 bonds is catalyzed by only one of n-1 different enzymatic active sites, each labeled as a condensation (C) domain. Before two amino acids can be bonded together by their unique C domain, each amino acid must first be adenylated by a unique A domain and then transferred to a thiol group attached through a molecular arm to a T domain. These C, A, and T domains are all part of one or more multi-domain proteins called NRPS enzymes. For each amino acid that is incorporated into a nonribosomal peptide, there is a unique set of domains, called a module, that are arranged in a linear fashion in an NRPS enzyme. Some NRPS enzymes contain many modules. The NRPS system that biosynthesizes gramicidin S consists of two NRPS enzymes and is illustrated in Figure 3. One of these two enzymes is called GrsA and contains just one module, consisting of three domains. GrsA also contains an epimerization (E) domain and lacks a C domain. GrsA must contact the four-module GrsB molecule so that it can use the first C domain on GrsB. Figure 3 also shows the growing chain of amino acids, the growing peptide, that would be attached to each T domain after a new peptide bond is formed at the preceding C domain in the NRPS assembly line.

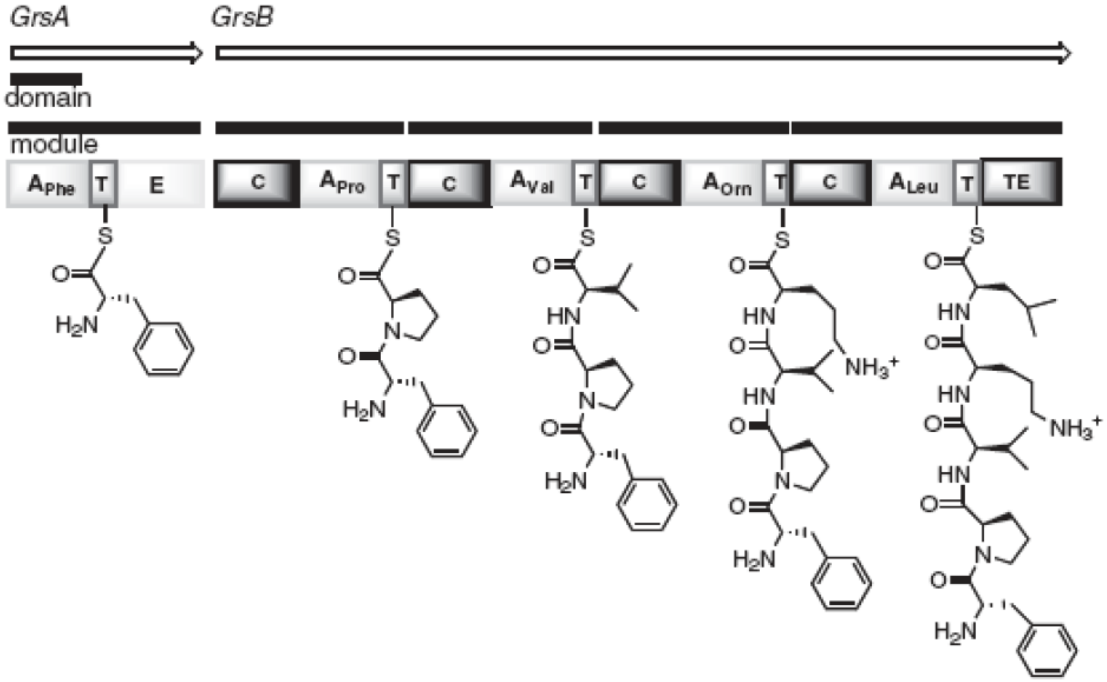

GrsA
GrsB
domain
module
APhe
T
E
C
APro
T
C
AVal
T
C
AOrn
T
C
ALeu
T
TE
S
NH3+

Figure 3: The NRPS assembly line for Gramicidin S [Reprinted from Stevens, et al., 2006a]

Before peptide bond formation can take place at a C domain, there must be an amino acid bonded to the two T domains on either side of that C domain. And, before that can happen, the specific amino acids must be activated by adenylation at their A domains, “upstream” of their T domains. The details of the adenylation reaction, transfer to the T domain, and peptide bond formation are shown in Figure 4, in which “PCP” (peptidyl carrier protein) is used instead of the synonymous “T” in this context.

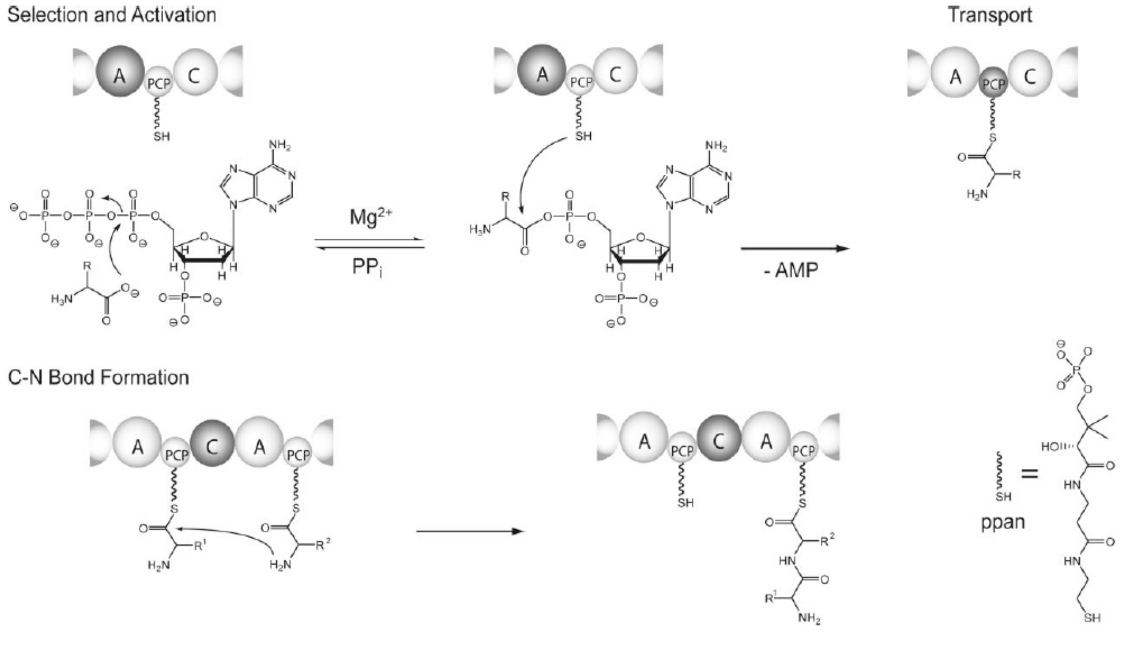
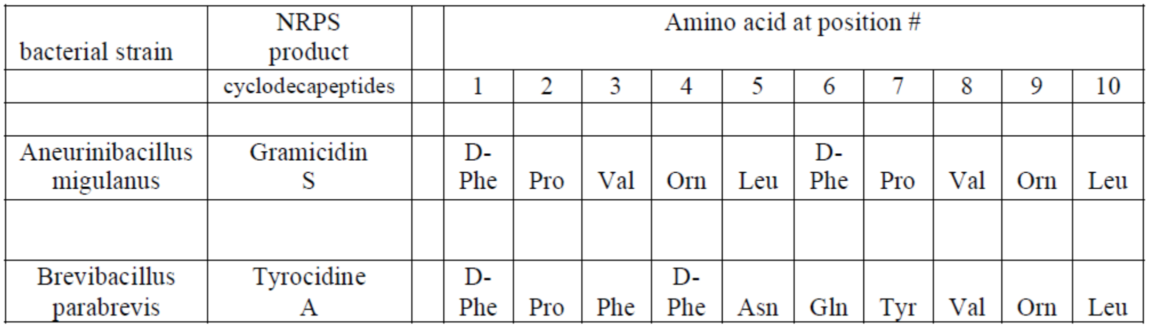

Selection and Activation
Transport
A
PCP
C
SH
NH₂
Mg²⁺
PPᵢ
- AMP
R
H₃N
C-N Bond Formation
R¹
R²
H₂N
NH₂
ppan
HO
HN
=

Figure 4: Chemical reactions catalyzed by the core domains of the NRPS enzymes [Reprinted from Kopp and Marahiel, 2007]

After the growing peptide has been "handed" from one T (aka, PCP) domain to the next in an NRPS assembly line, and it reaches the final T domain on the final NRPS enzyme, it must be released from the enzyme. That is accomplished by the thioesterase (TE) domain which is immediately downstream from the fourth and final T domain on GrsB (Figure 3). In order for the cyclic decapaptide gramicidin S molecule to be made, this TE domain catalyzes both the release of the finished pentapeptide and the head-to-tail dimerization with an identical pentapeptide, made on another GrsB molecule, and possibly the cyclization of the resulting decapeptide. Tyrocidine A is also a cyclodecapeptide, but it is not composed of two identical pentapeptides. The amino acid sequences of these two similar NRP antibiotics are compared in Table 3.

Table 3: Amino acid sequence comparison of two NRP antibiotics



bacterial strain	NRPS product		Amino acid at position #									
	cyclodecapeptides		1	2	3	4	5	6	7	8	9	10
Aneurinibacillus migulanus	Gramicidin S		D-Phe	Pro	Val	Orn	Leu	D-Phe	Pro	Val	Orn	Leu
Brevibacillus parabrevis	Tyrocidine A		D-Phe	Pro	Phe	D-Phe	Asn	Gln	Tyr	Val	Orn	Leu

The cyclodecapeptide molecule tyrocidine A is made on an NRPS assembly line that has ten modules, distributed on three NRPS enzymes, and ends with a TE domain which releases the finished decapeptide and cyclizes it. This is shown in Figure 5.

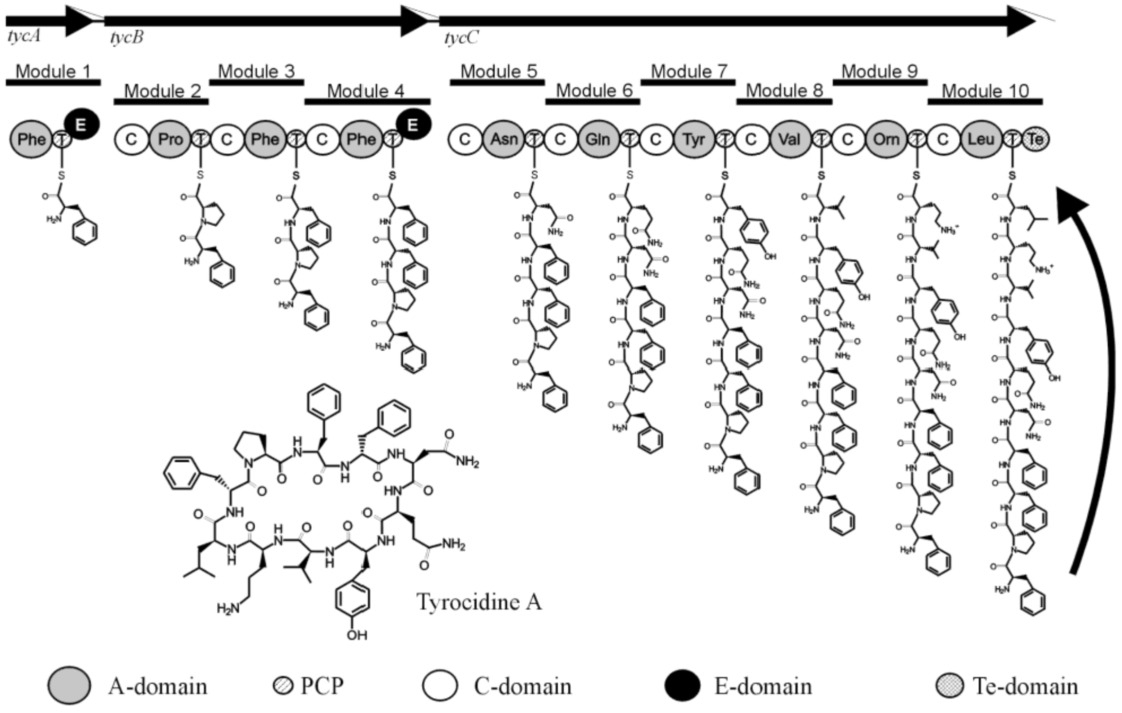
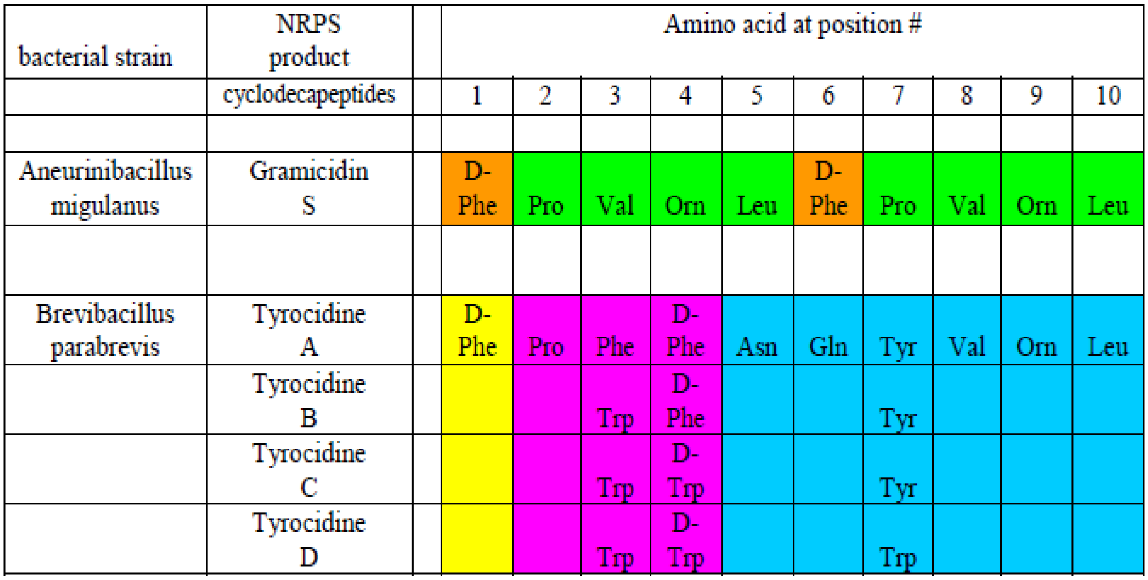

tycA
tycB
tycC
Module 1
Module 2
Module 3
Module 4
Module 5
Module 6
Module 7
Module 8
Module 9
Module 10
Phe
T
E
C
Pro
T
C
Phe
T
C
Phe
T
E
C
Asn
T
C
Gln
T
C
Tyr
T
C
Val
T
C
Orn
T
C
Leu
T
Te
Tyrocidine A
A-domain
PCP
C-domain
E-domain
Te-domain

Figure 5: Tyrocidine NRPS enzymes [Reprinted from Linne, et al., 2003]

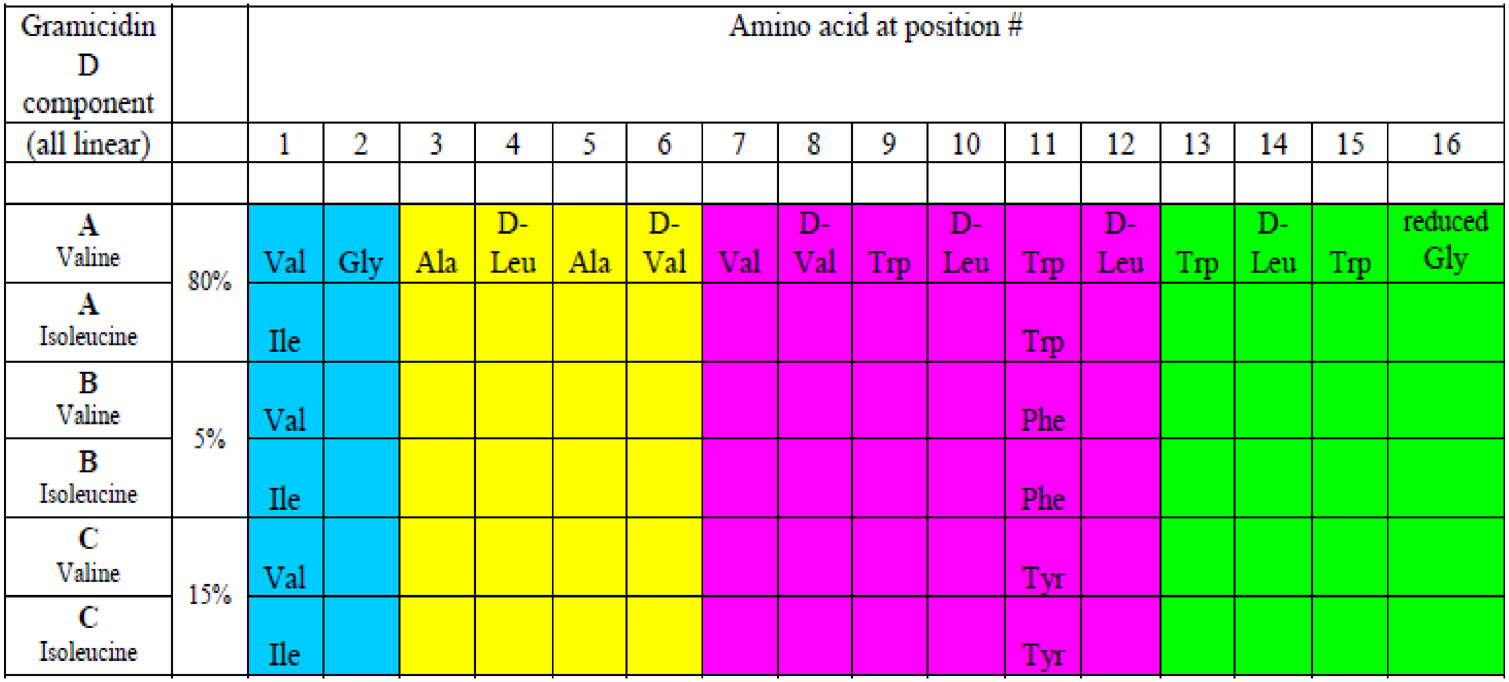
Another difference between tyrocidine A and gramicidin S is that three of the ten tyrocidine modules have the ability to each incorporate two different amino acids into the decapeptide. Only four of the eight possible decapaptides are made in quantities that can be detected. Table 4 shows the amino acid sequences for these four isoforms of tyrocidine A, compared with gramicidin S. Also, Table 4 is colored to indicate which amino acids in gramicidin S, as well as all four versions of tyrocidine, are incorporated by modules on the same NRPS enzyme, for each system. Finally, gramicidin D is different than gramicidin S in several ways. Gramicidin D is a mixture of six nearly identical isoforms of a linear hexadecapeptide. Gramicidin D is another NRP antibiotic which, like tyrocidine, is made by a system of NRPS enzymes in which some modules have the ability to incorporate more than one amino acid. Table 5 shows the amino acid sequences of all six components of gramicidin D, and is colored to indicate which amino acids are incorporated by the same NRPS enzyme.

Table 4: Amino acid sequence comparison of Gramicidin S and the Tyrocidines A-D



bacterial strain	NRPS product		Amino acid at position #									
	cyclodecapeptides		1	2	3	4	5	6	7	8	9	10
Aneurinibacillus migulanus	Gramicidin S		D-Phe	Pro	Val	Orn	Leu	D-Phe	Pro	Val	Orn	Leu
Brevibacillus parabrevis	Tyrocidine A		D-Phe	Pro	Phe	D-Phe	Asn	Gln	Tyr	Val	Orn	Leu
	Tyrocidine B				Trp	D-Phe			Tyr			
	Tyrocidine C				Trp	D-Trp			Tyr			
	Tyrocidine D				Trp	D-Trp			Trp			

Table 5: Amino acid sequences of all six isoforms of the NRP antibiotic Gramicidin D



Gramicidin D component		Amino acid at position #															
(all linear)		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A Valine	80%	Val	Gly	Ala	D-Leu	Ala	D-Val	Val	D-Val	Trp	D-Leu	Trp	D-Leu	Trp	D-Leu	Trp	reduced Gly
A Isoleucine		Ile										Trp					
B Valine	5%	Val										Phe					
B Isoleucine		Ile										Phe					
C Valine	15%	Val										Tyr					
C Isoleucine		Ile										Tyr					

2.1.3 Modifying Antibiotic Biosynthesis Machinery

A wealth of detailed knowledge about the function of various NRPS enzymes has now accumulated over several decades. This has led to novel approaches for discovering novel antibiotics. Instead of searching for new strains of microbes that produce new antibiotics there has been a steady effort for more than a decade to use this knowledge and "teach" old antibiotic-producing bugs to make new drugs, by re-engineering one or more of their NRPS enzymes. This approach, called combinatorial biosynthesis by some, has been recently reviewed [Sieber and Marahiel, 2005; Fischbach and Walsh, 2006; Stevens, et al., 2006a]. Because of the modular nature of the NRPS enzymes, many groups have tried to replace domains, or entire modules, in the target enzyme with an analogous domain, or module, with different amino acid specificity from another NRPS enzyme. The goal is to modify the microbe so that it produces a simple variant of the natural NRP product, which differs only by the amino acid(s) at the targeted position(s) of the peptide. In 2008, Doekel, et al. [Doekel, et al., 2008] reported being able to swap full modules and produce, *in vivo*, variants of daptomycin, a lipopeptide antibiotic, with the desired non-natural amino acids in the C-terminal position # 13. In spite of this outstanding success, the domain or module swapping approach is very difficult and will require much work before it will be an easy thing to do. Even then, it might not allow any chosen common amino acid to be placed in any position in the NRP peptide chain.

An alternative is the finer scale approach of making only individual amino acid changes in the target NRPS enzyme. This domain redesign approach initially focused on changing the specificity of adenylation domains. This approach requires knowing which amino acids in the target adenylation domain are responsible for determining the substrate specificity of its active site. This approach also requires being able to predict which amino acids in the wild type enzyme should be changed, and to what, in order to achieve the desired change in substrate specificity. One approach is to analyze the database of amino acid sequences of known NRPS adenylation domains, there were 1,230 in 2005 [Rausch, et al., 2005], to find a substrate specificity conferring code. This sequence-based approach was used by Eppelmann et al. [Eppelmann, et al., 2002] to redirect surfactin biosynthesis. Alternatively, a structure-based approach has been developed in the Donald lab, using the K* algorithm, in their OSPREY suite [Lilien, et al., 2005, Georgiev, et al., 2006; Georgiev and Donald, 2007; Georgiev, et al., 2008a; Georgiev, et al., 2008b; Gainza, et al., 2013], to predict which amino acids to change, and to what, in order to get the desired change in substrate specificity.

2.1.4 Structure-Based Computational Protein Redesign

The Donald lab has developed an ensemble-based protein-redesign algorithm, called K* ("K-star"), which is at the heart of their suite of protein redesign software called OSPREY [Gainza, et al., 2013]. The K* algorithm calculates a provably-accurate approximation of the protein-ligand binding constant for each pairing of the user-specified ligand with all of the possible mutant versions of one specific protein. These mutants differ from the wild type version of the user-specified protein in only a few amino acid substitutions that are usually restricted to just those residues that form non-covalent interactions with the ligand.

OSPREY uses a suite of recently developed extensions of Dead-End Elimination (DEE) [Desmet, et al., 1992], as initial pruning filters, for computational efficiency. OSPREY then uses a branch-and-bound search, derived from the A* algorithm [Leach and Lemon, 1998], to create a gap-free list of mutant protein sequences, ordered by the predicted energy of their predicted 3-D conformations. From this gap-free list, the K* algorithm calculates an ε-approximation of the protein-ligand binding constant (the K* score) for only the user-specified number of all of the top scoring redesigned versions of the protein. The K* algorithm uses a ratio of partition functions over ensembles of conformations, weighted by Boltzman probabilities, to compute the K* score for each mutant protein, binding with the one target ligand [Gainza, et al., 2013]. OSPREY outputs a list of mutant sequences, in order of their calculated K* scores. The mutant proteins with higher K* scores are predicted to bind more tightly to the target ligand.

OSPREY [Gainza, et al., 2013] has been used to redesign the active site of an NRPS enzyme's adenylation domain. Specifically, Chen, et al. [Chen, et al. 2009], used MinDEE with K* to change the amino acid substrate specificity of the adenylation domain of the first NRPS module for gramicidin S, called GrsA-PheA or PheA. Using a crystal structure of wild type (wt) PheA complexed with AMP and its cognate amino acid substrate Phe [Conti, et al., 1997], OSPREY was applied to switch the substrate specificity from Phe to Leu. The top-ranking active-site redesign output by OSPREY had two amino acid changes, T278L and A301G, in the active site (Chen, et al. 2009). The OSPREY-predicted lowest-energy conformation of this T278L/A301G double mutant binding to the target substrate Leu is shown in Figure 6. This redesigned adenylation domain was able to adenylate the target substrate Leu. The adenylation activity of the wt and mutant versions of PheA was determined by measuring the rate of PPi release.

To improve the binding specificity and adenylation activity of this mutant PheA domain, OSPREY was further used to discover distal bolstering mutations, outside of the active site. It was found that by adding a third mutation, V187L, the resulting triple mutant PheA had greater activity (Chen, et al. 2009). This triple mutant, T278L/A301G/V187L, had a Specificity Ratio of 9.4 (see Figure 7) and 1/6 of the specificity constant (k_{cat}/K_M) of the wt PheA with the wt substrate Phe. The specificity ratio for each version of PheA is the specificity constant of that version with Leu divided by that same constant with Phe instead.

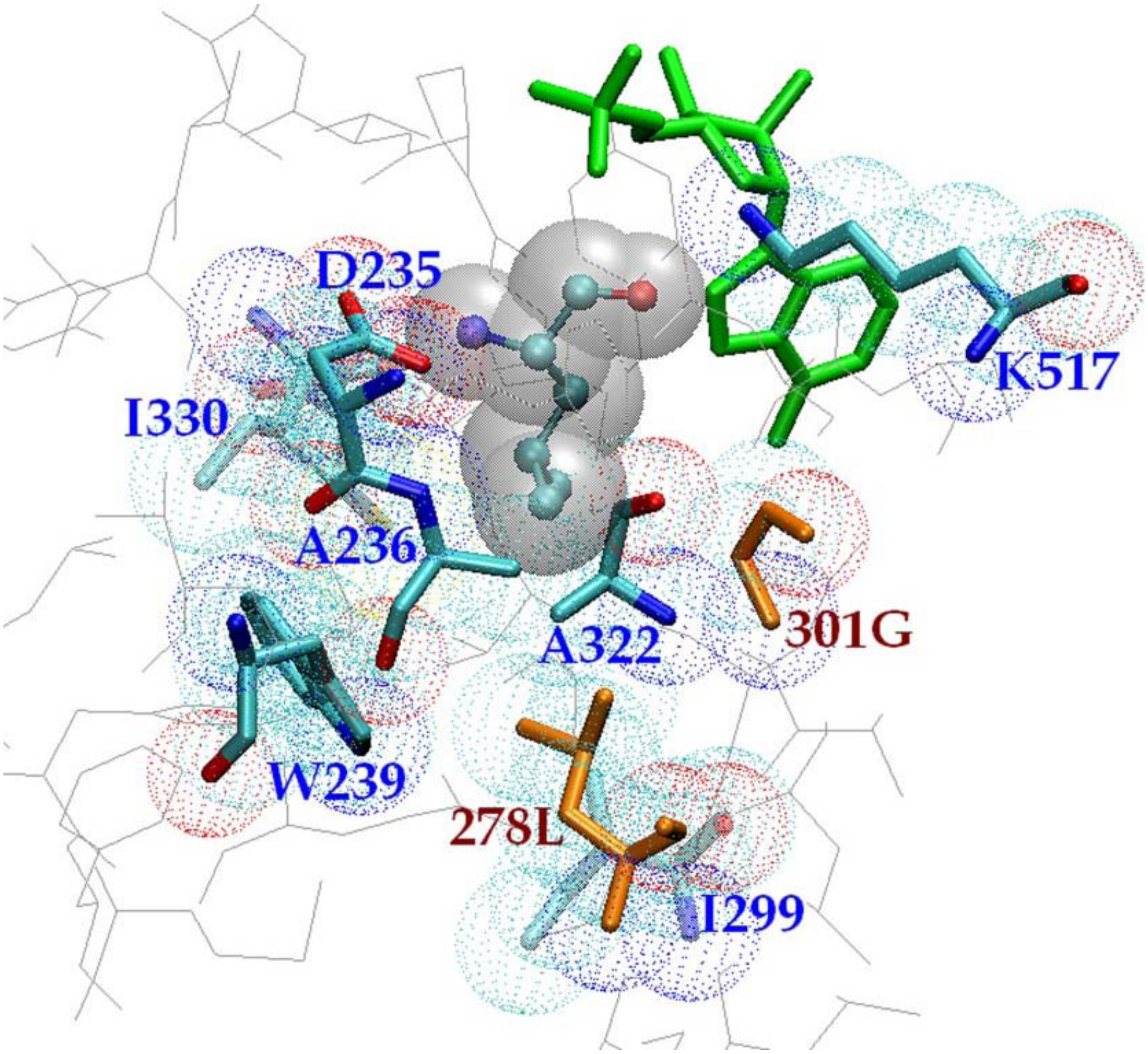

D235
I330
A236
W239
A322
301G
K517
278L
I299

Figure 6: K*-predicted structure of the lowest-energy conformation of T278L/A301G binding with the redesign-target, Leu, in the active site [Reprinted from Chen, et al. 2009]

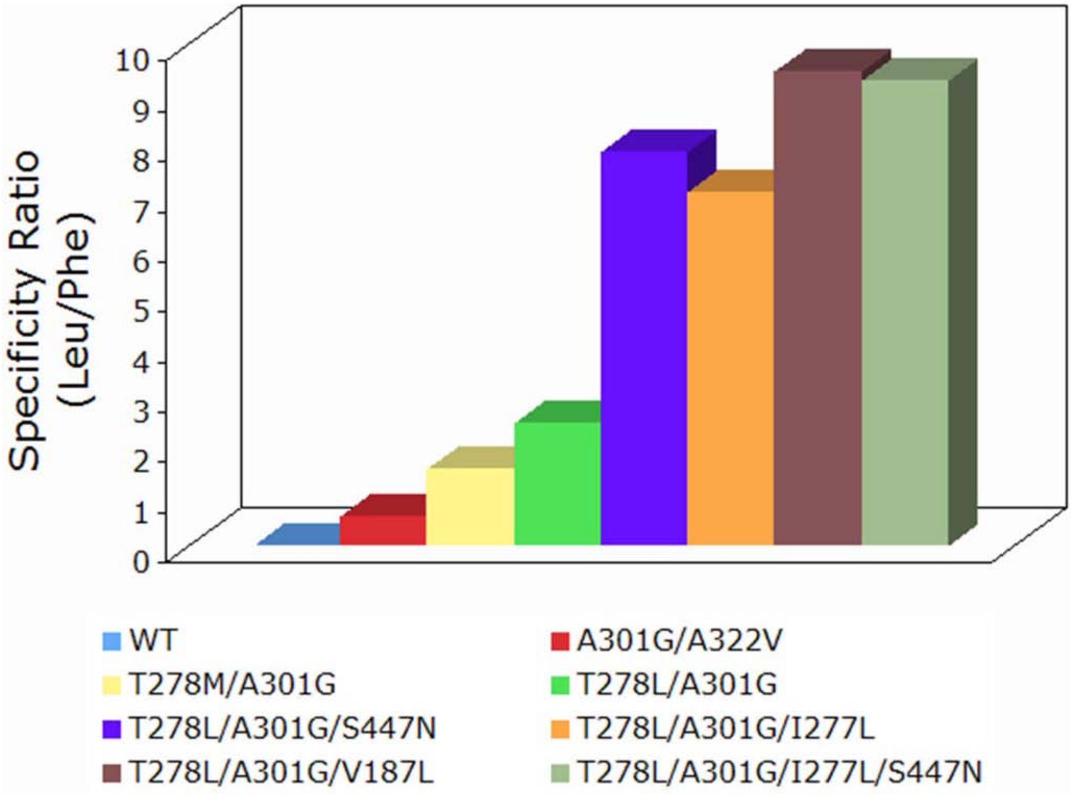


Figure 7: Specificity ratios for WT and mutant PheA redesigned to adenylate Leu [Reprinted from Chen, et al. 2009]

2.1.5 Structure Determination, by Solution-Phase NMR, of wild type and mutant PheA

The Donald lab has been studying the 65kDa NRPS adenylation domain of Gramicidin synthetase A, GrsA-PheA, called PheA, (Figure 3) as a model system for our lab's protein redesign work [Chen, et al. 2009, Stevens, et al., 2006b]. PheA has been a model system for NRPS adenylation domains since it's X-ray crystal structure was published more than a decade ago [Conti, et al., 1997]. From 2006-2009 I worked on collecting NMR data to determine the solution structure of wild type (wt) PheA. Our hypothesis was that the solution-phase NMR techniques that I would develop while studying the structure of wt-PheA could easily be applied to perform identical or similar NMR experiments to study the solution-phase structures of some of the OSPREY-predicted mutant versions of PheA made by another graduate student, Cheng-Yu Chen, in our lab [Chen, et al. 2009].

The goals of my PheA structure-determination project included: (1) determining the NMR solution structure of wt-PheA without either of the two substrates binding to it; (2) understanding the dynamics of the wt-PheA protein; (3) detecting conformational changes in the 3-D structure of wt-PheA when one or both of the substrates bind to it; (4) understanding which residues of wt-PheA are involved in binding the substrates, or adjacent NRPS domains; and (5) all of the previous four goals applied to some of the mutant versions of PheA with altered substrate specificity, predicted by Ivelin Georgiev, using MinDEE and K*, and produced by Cheng-Yu Chen in our lab. We hypothesized that this new structural knowledge of wt-PheA, and mutants of PheA, could be used to improve the protein redesign algorithms in OSPREY, as well as our general understanding of the structures and mechanisms of NRPS adenylation domains.

2.2 Methods

The experimental steps to determining these solution structures by NMR include: (1) expressing and purifying isotopically labeled versions of wt PheA; (2) collecting [^{1}H, ^{15}N] HSQC data; (3) collecting [^{1}H, ^{13}C, ^{15}N] triple resonance data; (4) collecting NOESY spectra; and (5) collecting residual dipolar couplings (RDCs). The data analysis steps include: (1) evaluating the initial [^{1}H, ^{15}N] HSQC data from perdueterated ^{15}N-labeled WT PheA to determine whether the cross peaks are dispersed well enough to make success in subsequent steps likely; (2) using the triple resonance data to assign the resonances of the backbone and then the side-chain atoms; and (3) using the RDCs and NOEs to calculate the protein's backbone and side-chain structure using Donald Lab software, such as RDC-Panda [Zeng, et al., 2009].

We express His-tagged wt-PheA in *E. coli* bearing an expression plasmid containing the gene construct. Over expression is induced by Isopropyl Thiogalactopyranoside (IPTG), with Mg2+ ions present to assist with the proper folding of wt-PheA. We typically purify wt-PheA from cleared lysates of *E. coli* collected from one or two liters (L) of culture induced with IPTG for about 18 hours, at 18 C. Purification involves Ni-affinity chromatography, followed by size-exclusion chromatography. The preparations summarized in Table 6 were done over a couple of years and the yield shown is from the Ni affinity column, for comparison purposes. A large amount of unlabeled wt-PheA was made to optimize the purification process and the solution conditions for wt-PheA before making the more costly versions that are labeled with the stable isotopes shown. The double labeled (2H, 15N) wt-PheA was used to obtain the [1H, 15N] TROSY-HSQC spectrum shown in Figure 8. The triple labeled wt-PheA was used to obtain the TROSY-HNCO spectrum shown in Figure 9.

2.3 Results

Table 6: Summary of Successful wt-PheA Preparations

Protein	Residues	Labeling	Average yield	Total of cultures
PheA – His6	565	None	75 mg/ L	14 L
		^{2}H, ^{15}N	10 mg/ L	2 L
		^{2}H, ^{13}C, ^{15}N	37 mg/ L	1 L
His6-thrombin site-PheA	564	None	62 mg/ L	4 L

The [^{1}H, ^{15}N] TROSY-HSQC spectrum, in Figure 8, showed excellent dispersion of cross peaks and we decided to continue our structural studies with this particular protein. The next steps included more protein preps of unlabeled wt-PheA to further optimize our protocol for purification and the solution conditions. We obtained a good yield of triple labeled wt-PheA from one liter of culture. However, the three dimensional

HNCO data we collected, shown in Figure 9 projected onto the [^{1}H, ^{15}N] plane, was missing many peaks that were in the HSQC spectrum (Figure 8). Thus, we realized that

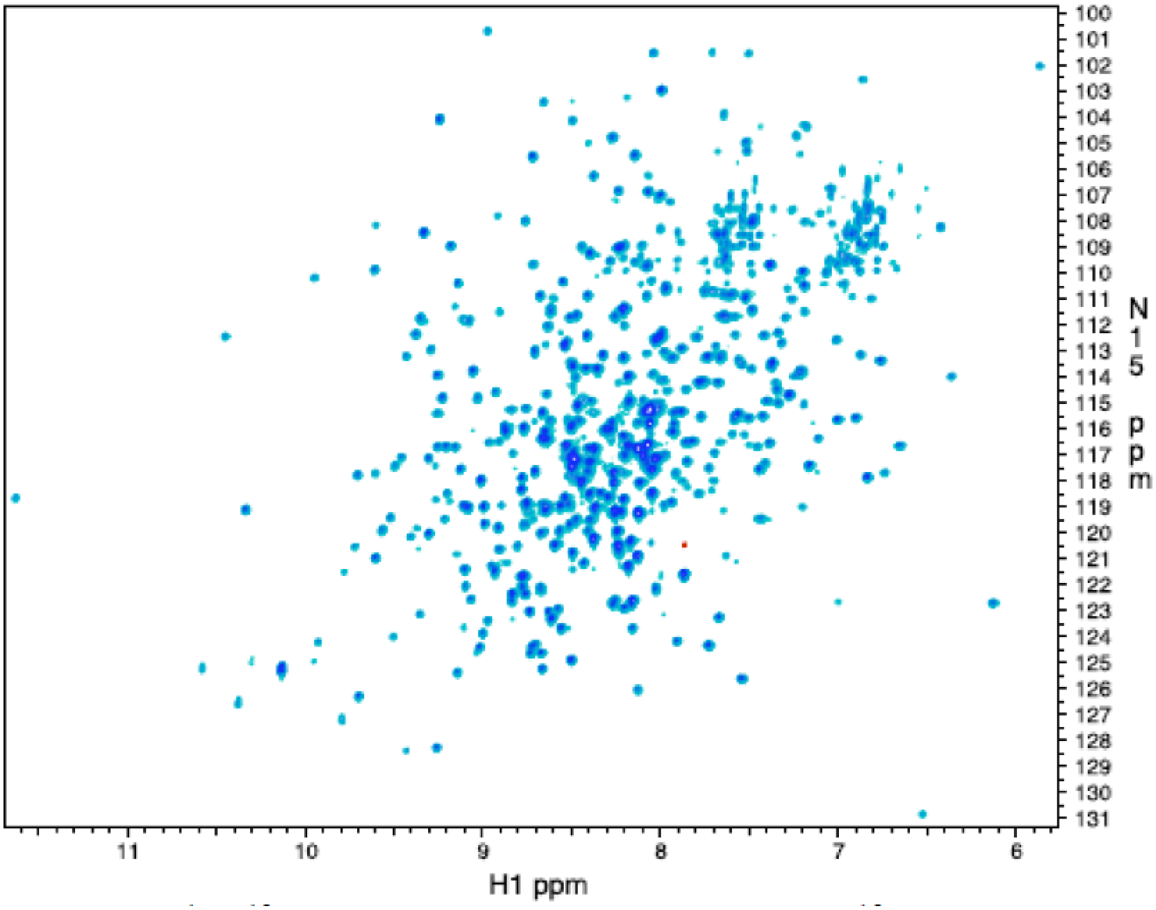


Figure 8: [^{1}H, ^{15}N] TROSY-HSQC of perdeuterated [^{15}N]-labeled wt-PheA, at 0.25mM.

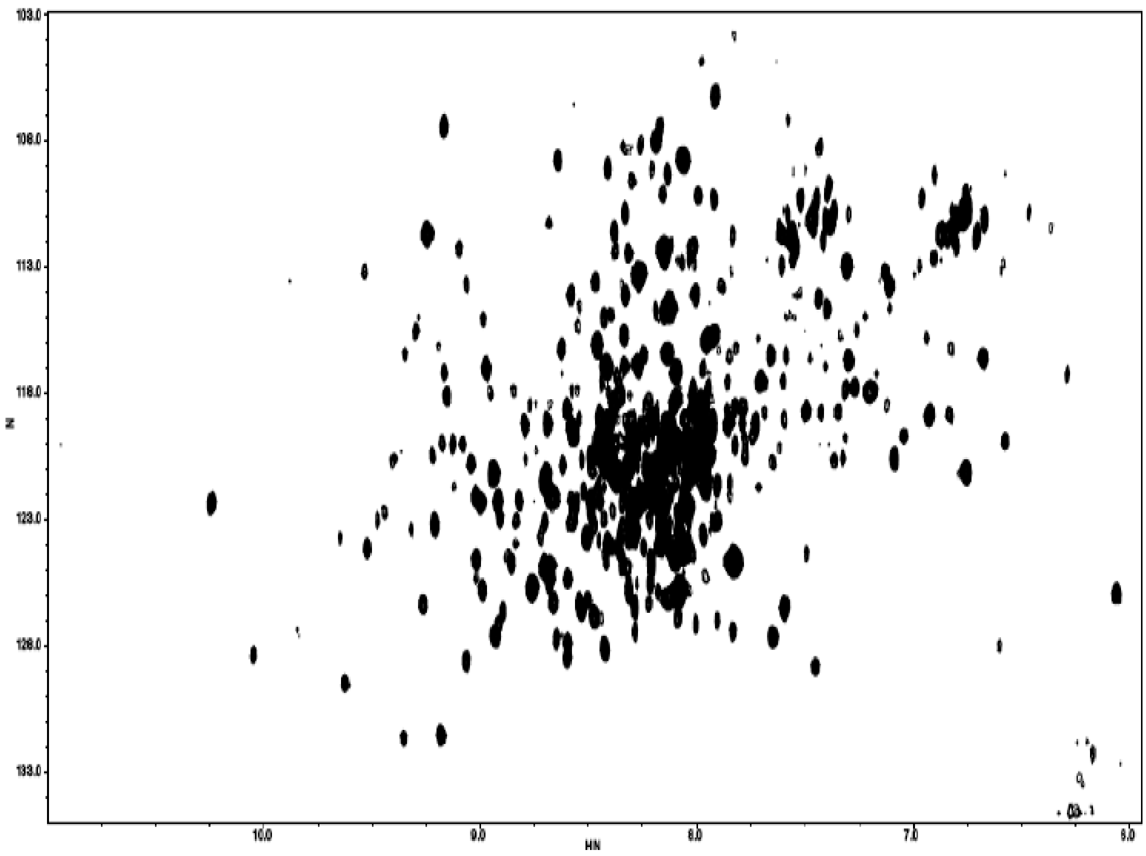
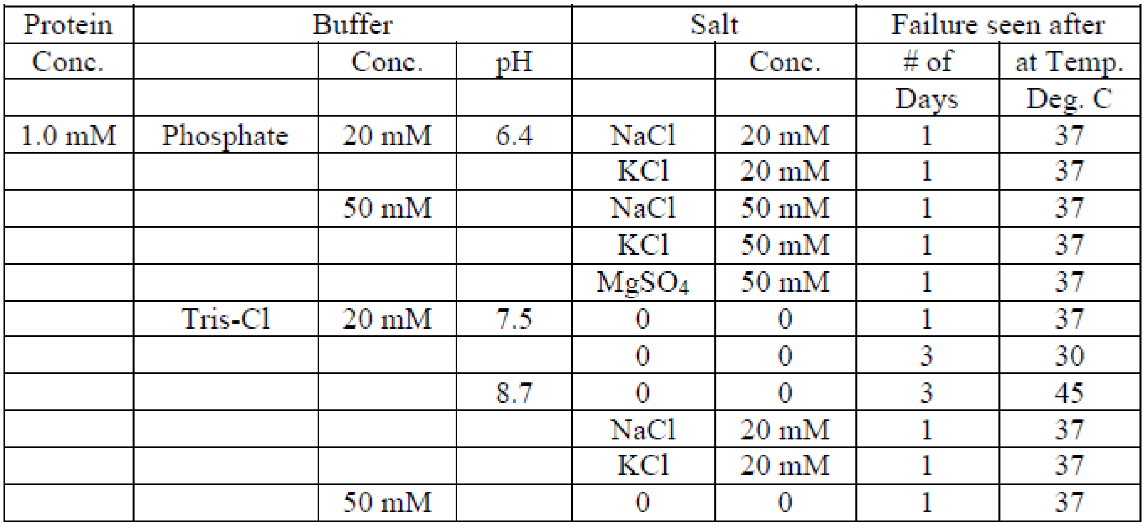


Figure 9: [^{1}H, ^{15}N]TROSY-HNCO of perdeuterated [^{13}C, ^{15}N]-labeled wt-PheA, at 1.0 mM. (as in Figure 8, ^{1}H ppm is along the horizontal axis, ^{15}N ppm is along the vertical axis)

we needed to get stronger NMR signals. The best way to do that is to increase the tumbling rate of the protein in the NMR sample. The easiest way to do that is to increase the temperature of the solution during data collection.

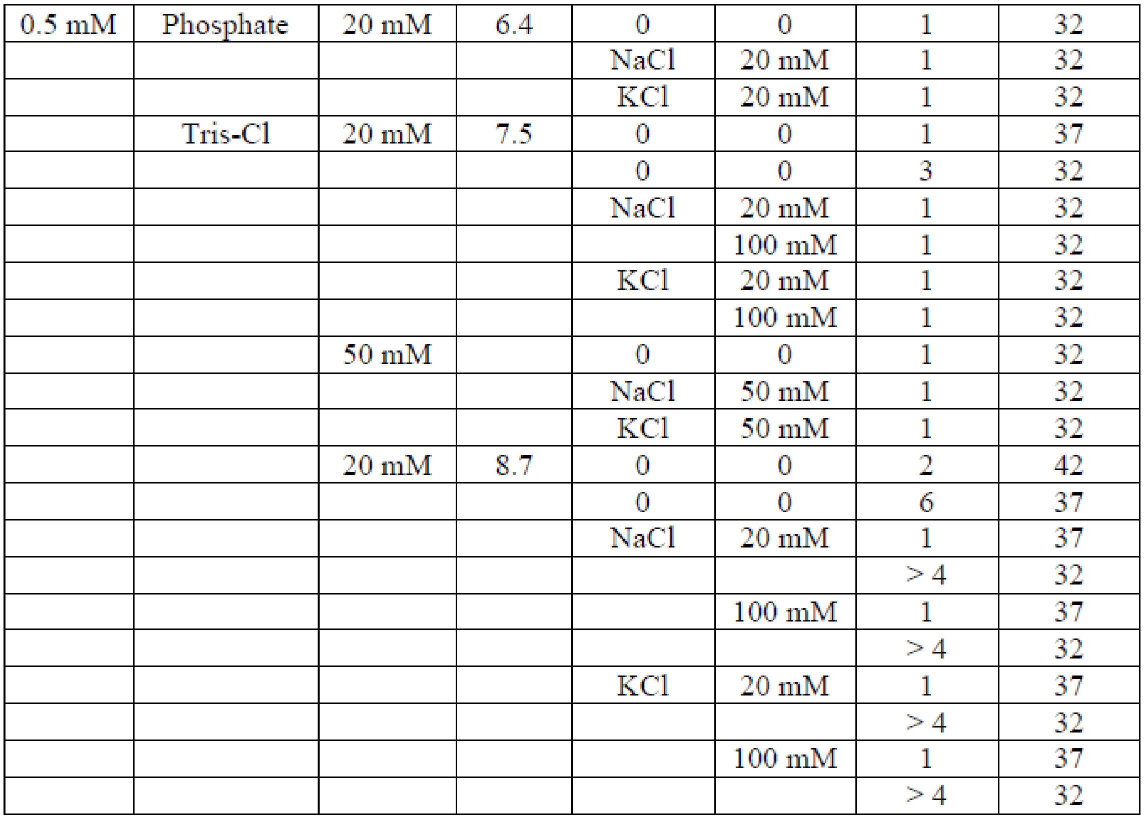
Since the triple resonance experiments that we have planned will require data collection times of two to seven days, we need to find solution conditions that will allow wt-PheA to stay in solution for up to seven days at temperatures of at least 35 °C, and hopefully 45 °C. The results of my buffer and salt testing are summarized in Tables 7 and 8. When I didn't find suitable solution conditions after the testing shown in Table 7, we decided to remake the wt-PheA gene construct so that it would be possible to remove the His tag after the Ni affinity purification step. This often helps solve solubility problems. The buffer and salt testing with this His-cleavable version of wt-PheA is summarized in Table 8.

Table 7: Summary of Solubility Testing With **wt-PheA-His6**
(Any blank cell contains the same parameter value as in the cell above it.)



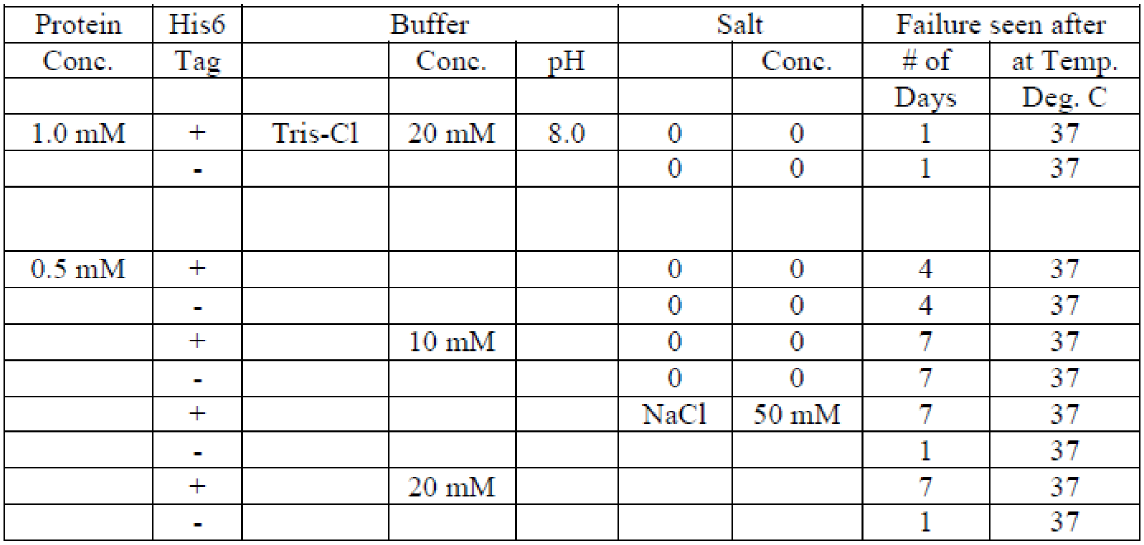
Protein	Buffer			Salt		Failure seen after	
Conc.		Conc.	pH		Conc.	# of	at Temp.
						Days	Deg. C
1.0 mM	Phosphate	20 mM	6.4	NaCl	20 mM	1	37
				KCl	20 mM	1	37
		50 mM		NaCl	50 mM	1	37
				KCl	50 mM	1	37
				$MgSO_4$	50 mM	1	37
	Tris-Cl	20 mM	7.5	0	0	1	37
				0	0	3	30
			8.7	0	0	3	45
				NaCl	20 mM	1	37
				KCl	20 mM	1	37
		50 mM		0	0	1	37

Table 7: (continued) Summary of Solubility Testing With **wt-PheA-His6**
(Any blank cell contains the same parameter value as in the cell above it.)



0.5 mM	Phosphate	20 mM	6.4	0	0	1	32
				NaCl	20 mM	1	32
				KCl	20 mM	1	32
	Tris-Cl	20 mM	7.5	0	0	1	37
				0	0	3	32
				NaCl	20 mM	1	32
					100 mM	1	32
				KCl	20 mM	1	32
					100 mM	1	32
		50 mM		0	0	1	32
				NaCl	50 mM	1	32
				KCl	50 mM	1	32
		20 mM	8.7	0	0	2	42
				0	0	6	37
				NaCl	20 mM	1	37
						> 4	32
					100 mM	1	37
						> 4	32
				KCl	20 mM	1	37
						> 4	32
					100 mM	1	37
						> 4	32

Table 8: Summary of Solubility Testing With **His6-ThrombinSite-wt-PheA**



Protein	His6	Buffer			Salt		Failure seen after	
Conc.	Tag		Conc.	pH		Conc.	# of	at Temp.
							Days	Deg. C
1.0 mM	+	Tris-Cl	20 mM	8.0	0	0	1	37
	-				0	0	1	37
0.5 mM	+				0	0	4	37
	-				0	0	4	37
	+		10 mM		0	0	7	37
	-				0	0	7	37
	+				NaCl	50 mM	7	37
	-						1	37
	+		20 mM				7	37
	-						1	37

2.4 Discussion

The great utility, in protein NMR, of the [^{1}H, ^{15}N] HSQC experiment derives from the following facts. In a ^{15}N-labeled protein, every bonded pair of ^{1}H and ^{15}N atoms can produce one cross-peak in the frequency spectrum and all amino acids contain at least one ^{15}N atom. Because the chemical shifts of the NMR-active ^{1}H and ^{15}N nuclei are determined by the 3D location of their nearest-neighboring atoms, a stably folded protein yields an [^{1}H, ^{15}N] HSQC frequency spectrum with well-dispersed cross-peaks, while a protein, or just a portion of a protein, that is randomly changing its 3D conformation will yield an [^{1}H, ^{15}N] HSQC frequency spectrum in which most of its cross-peaks have the same location, due to averaging of their ^{1}H and ^{15}N electronic environments and thus their chemical shifts are all mostly the same. In addition to being able to distinguish between well-folded and poorly-folded proteins, an [^{1}H, ^{15}N] HSQC frequency spectrum reveals whether a large well-folded protein has well-dispersed cross-peaks, which can reduce the time-consuming problems of disambiguating overlapping cross-peaks.

Our [^{1}H, ^{15}N] TROSY-HSQC spectrum of wt-PheA, in Figure 8, showed excellent dispersion of cross peaks for a protein of this considerable size, 563 amino acids. The quality of this [^{1}H, ^{15}N] TROSY-HSQC spectrum of wt-PheA, and advice from protein-NMR experts here at Duke, gave us confidence that this large protein could be amenable to solution-phase NMR studies. To continue these NMR structural studies of wt-PheA I further optimized our protocols for expression and purification of isotopically-labeled wt-PheA. To perform three dimensional NMR experiments, I produced and purified triple-labeled (^{1}H, ^{13}C, ^{15}N) wt-PheA. Though we obtained a good amount of this triple-labeled wt-PheA, the three dimensional HNCO data that we collected from it was not good. In this HNCO data, shown in Figure 9 projected onto the [^{1}H, ^{15}N] plane, many [^{1}H, ^{15}N] cross-peaks that we could see in our HSQC data (Figure 8) were not visible. This meant that we needed to get a stronger NMR signal, with sharper peaks. The best way to do that is to increase the tumbling rate of the protein molecules in the NMR sample by increasing the temperature of the sample solution during data collection.

The triple resonance experiments that we needed to do would require data collection periods of two to seven days. Thus I tried to find solution conditions that would allow our His-tagged wt-PheA to stay in solution for up to seven days at temperatures of 35 °C, and preferably 45 °C. After all of the buffer and salt solution testing that I did, shown in Table 7, failed I re-engineered our wt-PheA gene construct so that I could make His-tagged wt-PheA with a removable His tag. Removing the His tag after the Ni-affinity column purification step is known to sometimes solve protein solubility problems. The buffer and salt testing that I did with our His-tag-cleaved version of wt-PheA, shown in Table 8, also failed. Unfortunately, solution conditions to solve the solubility problems with wt-PheA were never found. Protein solubility is actually a very common problem in protein NMR, where high protein concentrations are needed. Recently it has been estimated that 75% of soluble proteins have low solubility [Zhou and Wagner, 2010]. The low solubility of wt-PheA at temperatures greater than 25 °C made it impossible for us to collect the multidimensional NMR data that is required for structure determination of such a large protein. In 2009 I began to look for another interesting protein to determine the solution structure of by solution-phase NMR.

3. Using Computational Protein Redesign to Develop Novel Antibiotics to Combat Vancomycin Resistance

3.1 Introduction to Vancomycin

Vancomycin is one of several glycosylated natural products, microbial secondary metabolites, produced by a variety of microbes as chemical weapons that they release to kill nearby microbes competing with them for a scarce food supply. The chemical structure of vancomycin is shown in Figure 10 with the glycan portion shaded in red. Vancomycin was discovered and developed by Eli Lilly and Company in the 1950s as a

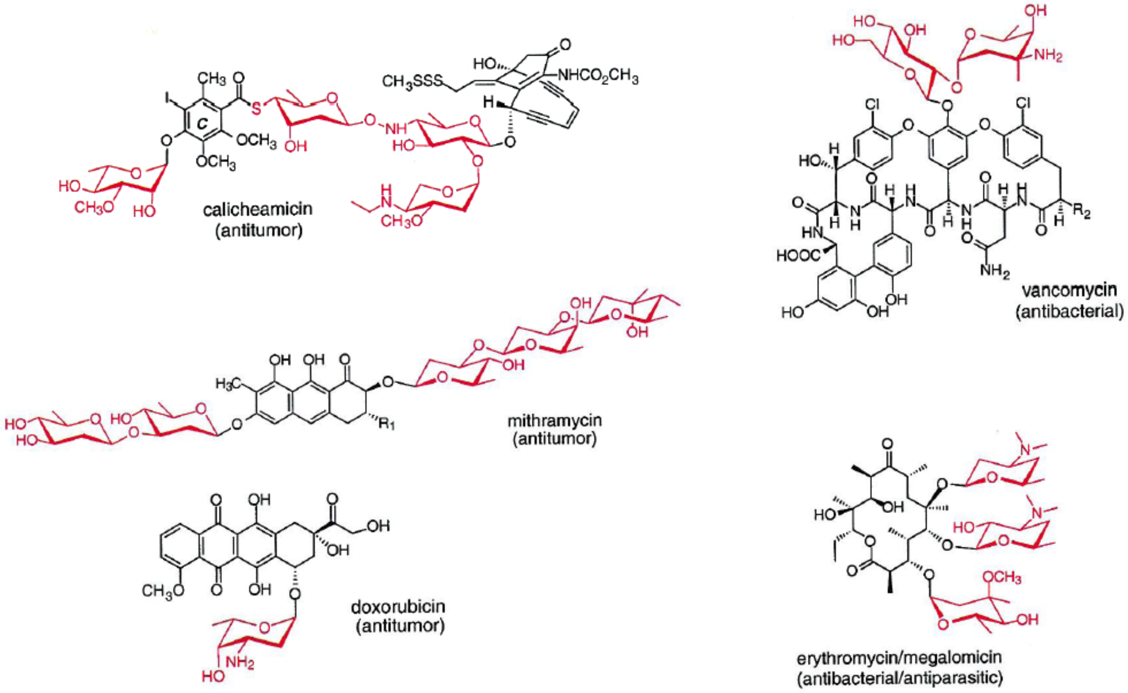

calicheamicin
(antitumor)
vancomycin
(antibacterial)
mithramycin
(antitumor)
doxorubicin
(antitumor)
erythromycin/megalomicin
(antibacterial/antiparasitic)

Figure 10: Some glycosylated secondary metabolites of pharmacological importance. [Reprinted from Barton, et al., 2001]

result of their efforts to discover new antibiotics to combat the increasingly prevalent penicillin-resistant staphylococci. *In vitro* testing done at Eli Lilly showed that staphylococci developed resistance to penicillin much faster than to vancomycin; 100,000-fold increase in penicillin resistance compared to less than a 10-fold increase in vancomycin resistance, after 20 serial passages [McGuire, et al., 1955-1956]. Before clinical trials could begin, difficulties in purifying vancomycin needed to be resolved. The active preparation from bacterial culture was originally called “Mississippi mud” due to its brown color. Vancomycin is produced by an actinomycete (a microbe in the order *Actinomycetales*) originally classified as *Streptomyces orientalis* that was isolated from a sample of dirt sent to an Eli Lilly chemist, Dr. E. C. Kornfield, from a missionary friend in Borneo [Levine, 2006]. When the purification protocol was worked out, the active ingredient was named “vancomycin,” from the word “vanquish,” and clinical trials began [Levine, 2006].

After Waksman discovered streptomycin, the first successful treatment for tuberculosis, from an actinomycete in 1943 [Schatz and Waksman, 1944] the actinomycetes became the target of many searches for new antimicrobial natural products. The actinomycetes are the most abundant microbes in the soil. The chemical structure of vancomycin, along with some other glycosylated natural products of pharmacological importance made by various actinomycetes, is shown in Figure 11, where the glycans are shaded in red. Tremendous variation in glycan complexity is illustrated by the eight examples of glycosylated secondary metabolites shown in Figures 10 and 11. The biochemical function of the glycan portions of these molecules is not well understood.

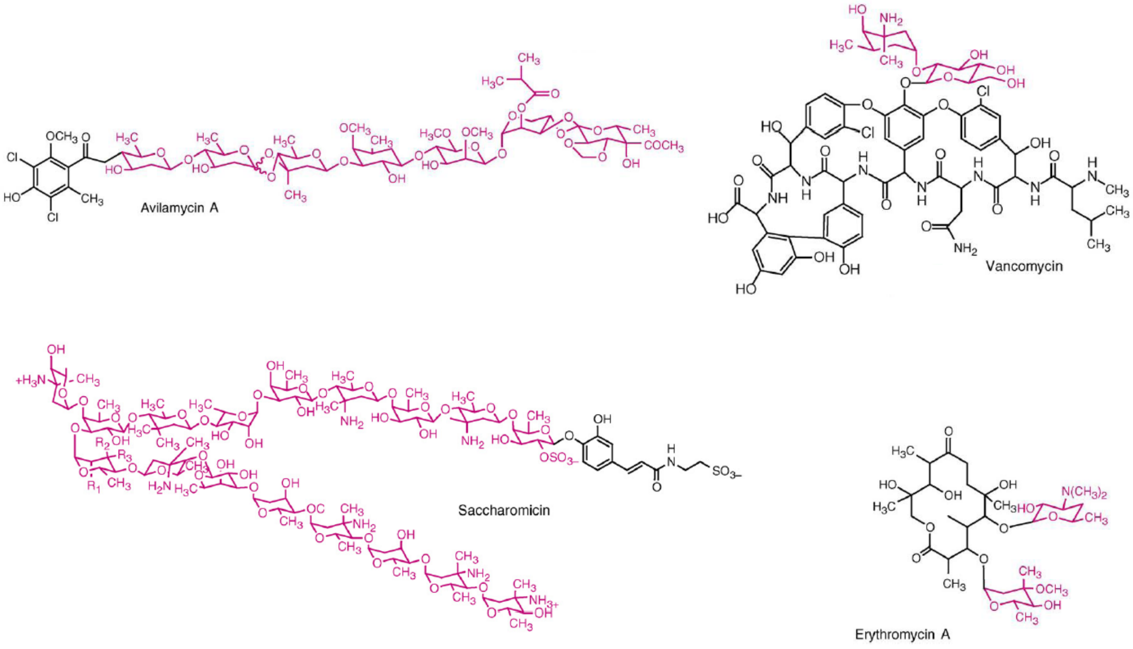


Figure 11: Some glycosylated secondary metabolites of pharmacological importance produced by actinomycetes. [Reprinted from Salas and Mendez, 2007]

Vancomycin exerts its antibiotic activity by interfering with crosslinking in the normal biosynthesis of the peptidoglycan (PG) layer in the cell wall of Gram positive bacteria. Normal crosslinking of the PG layer in bacteria is critical for creating a cell wall of sufficient strength to withstand normal osmotic pressures in the cell's environment.

Vancomycin prevents normal PG crosslinking by binding tightly to the C-terminal DAlanine-D-Alanine of the PG-crosslinking peptide, by making five hydrogen bonds, as shown in Figure 12. The PG layer is normally 20-80nm thick in the cell wall of Gram positive bacteria, while only 7-8nm in Gram negative bacteria. The PG layer in Gram negative cells is located between their inner and outer membranes, while Gram positive bacteria have a thicker PG layer and no outer membrane to physically block antibiotics from interfering with PG crosslinking. Vancomycin prevents normal Gram positive cells from successfully dividing, but has little impact on Gram negative bacteria.

Eremomycin and chloroeremomycin are members of the vancomycin group of glycopeptides because they all have the same heptapeptide core, shown in Table 9. Eremomycin and chloroeremomycin differ from vancomycin in their glycans, as shown in Figure 13. Three-dimensional structures of eremomycin, by solution-phase NMR [Groves, et al., 1994], and vancomycin, by X-ray crystallography [Schafer, et al., 1996], both showed that these glycopeptides antibiotics form an asymmetric back-to-back dimer. The conformation of each of these homodimers leaves the “front” of each monomer, where a C-terminal D-Ala - D-Ala peptide can bind (see Figure 12), exposed and available to bind to two nearby PG-crosslinking peptides, in antiparallel orientation to each other. Such glycopeptides homodimers have been proposed to allow a

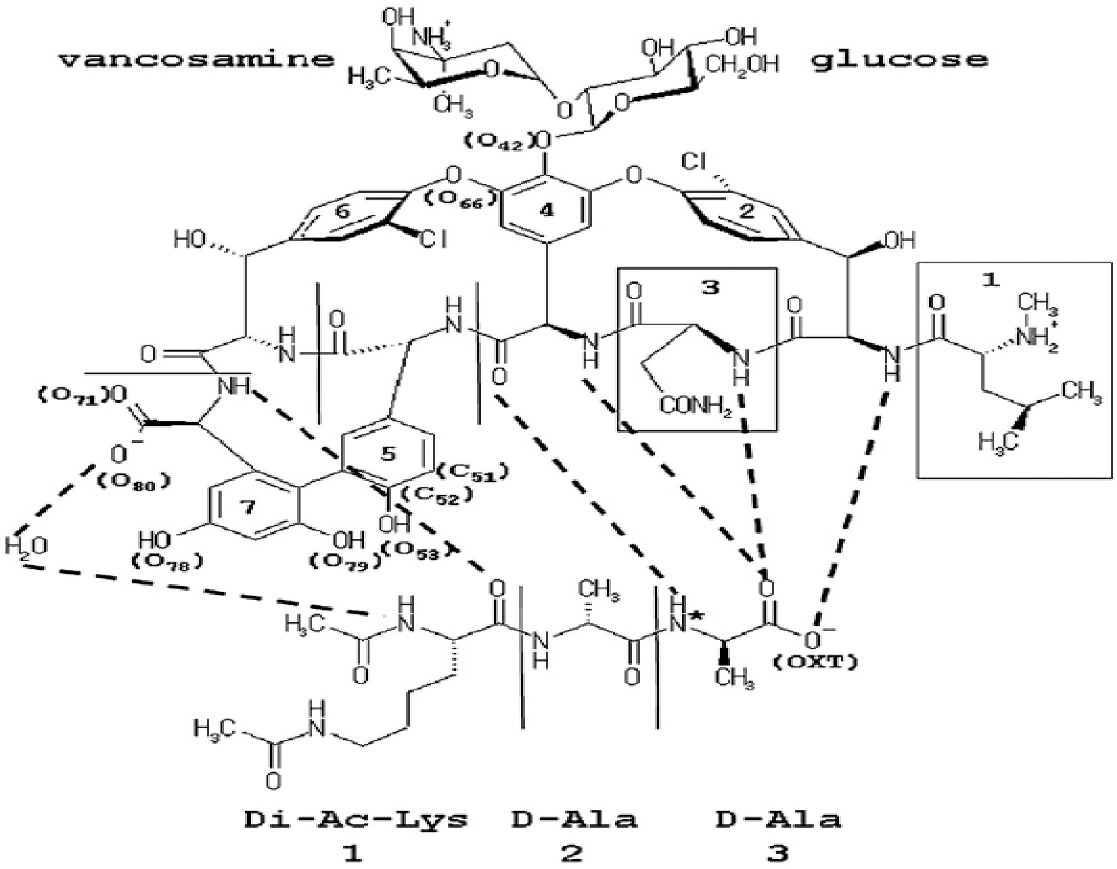

vancosamine
glucose
(O42)
(O66)
(O71)
(O80)
(O78)
(O79)
(O53)
(C51)
(C52)
(OXT)
1
2
3
4
5
6
7
H2O
CONH2
Di-Ac-Lys
1
D-Ala
2
D-Ala
3

Figure 12: Hydrogen bonds between vancomycin and normal PG-crosslinking peptide analog. Asterisk indicates the nitrogen that is replaced by oxygen in D-Ala-D-Lactate terminus. [Reprinted from Nitanai, et al., 2009]

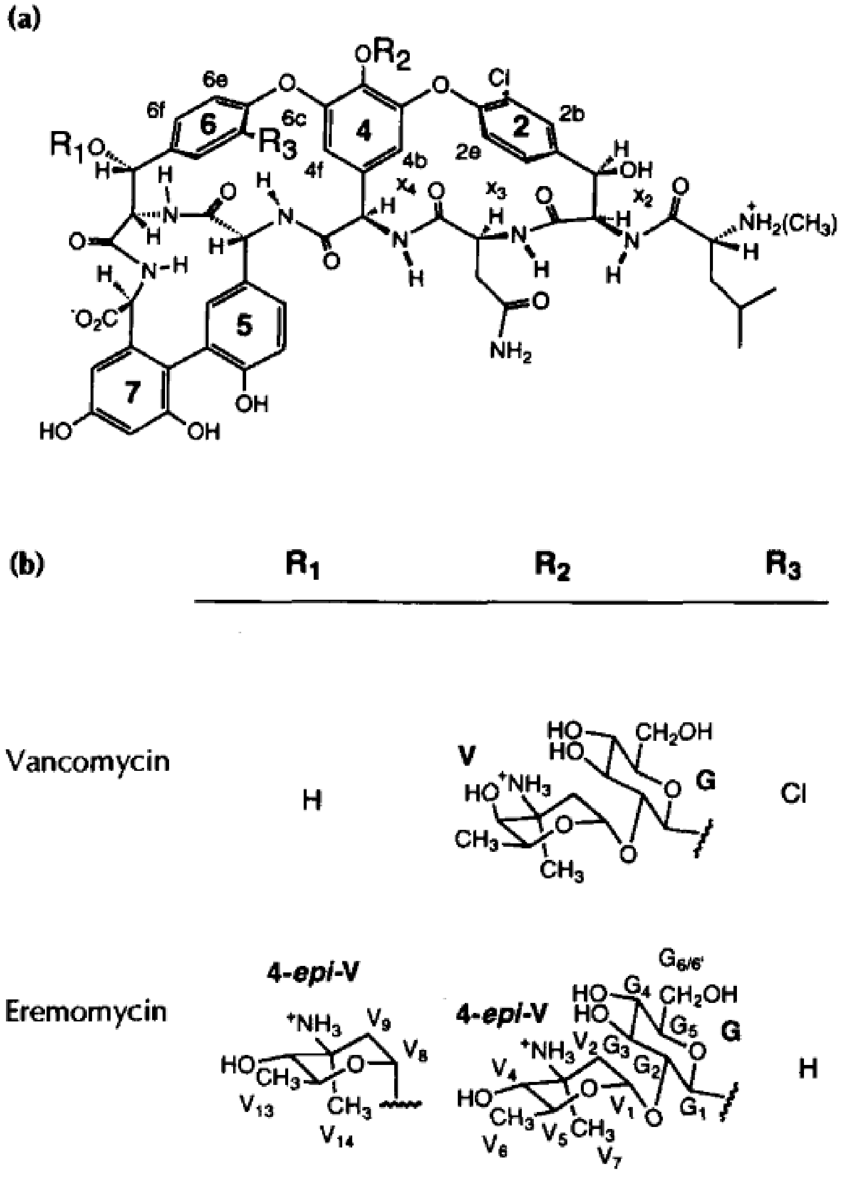
Table 9: Amino acids in the heptapeptide core of vancomycin-group glycopeptides

NRPS product		Amino acid at position #						
heptapeptide		1	2	3	4	5	6	7
unmodified heptapeptide		D-Leu	D-Tyr	Asn	D-HP-Gly	D-HP-Gly	Tyr	diHP-Gly
Vancomycin core		N-Me-D-Leu	CH-D-Tyr	Asn	D-HP-Gly	D-HP-Gly	CH-Tyr	diHP-Gly

cooperative binding effect, that enhances the binding between glycopeptides and PG-crosslinking peptides. But, the role of the glycans remains unclear. The authors of the NMR structure [Groves, et al., 1994] concluded that the glycans were important in forming the eremomycin homodimer, while the authors of the X-ray crystal structure [Schafer, et al., 1996] concluded that, at least for vancomycin, the glycans are not important in stabilizing the homodimer, but may have a role in its formation.

3.1.1 Vancomycin-Resistant Microbes

There are several different mechanisms by which bacteria develop or acquire resistance to a particular antibiotic in their environment. Briefly these are: (1) enzymatic inactivation of the drug once it enters the bacterial cell; (2) enzymatic removal of the drug from the cell; (3) change of the cell's permeability to reduce the amount of drug that can enter; (4) change the drug's target molecule by mutation of the bacteria's wild type gene; (5) change the drug's target molecule by expressing different genes which alter the targeted pathway; (6) change the cell's metabolism so as to "ignore" the pathway that is disrupted by the drug. All of these mechanisms, except #4 are usually



(a)
(b)
R1
R2
R3
Vancomycin
H
Cl
Eremomycin
H
4-epi-V
V
G

Figure 13: Chemical structures of two members of the vancomycin group of glycopeptide antibiotics. [Reprinted from Groves, et al., 1994]

due to the bacterial cell expressing drug-resistance conferring genes that bacteria often acquire by horizontal gene transfer (HGT). It's worth noting that all of these general mechanisms that bacteria can use to develop resistance to an antibiotic are the same mechanisms which cancer cells can use to develop resistance to an anticancer drug.

Vancomycin had been in clinical use for more than 30 years before the emergence of marked resistance to it [Gold and Moellering, 1996]. Until the 1990s, vancomycin was the main drug which could be relied upon for the treatment of infections caused by antibiotic-resistant *enterococci* and *Staphylococcus aureus.* Vancomycin is now the last line of defense against methicillin-resistant *Staphylococcus aureus* (MRSA) and many other multiple-antibiotic-resistant infections [Roper, et al., 2000]. The rates of nosocomial infections due to vancomycin-resistant *enterococci* (VRE) rose from 0.5% in 1989 to 10% in 1995 [Gold and Moellering, 1996].

In the 20 years since the first occurrence of VRE was reported in 1986, six distinct patterns of vancomycin resistance have been reported, designated "VanA" through VanE", and "VanG" [Levine, 2006]. All six phenotypes are due to multiple acquired genes. These vancomycin-resistance genes permit normal PG crosslinking, and thus bacterial growth, in the presence of vancomycin by two slightly different mechanisms. In both cases an alternative PG-crosslinking peptide is used by the cell. When the alternative PG-crosslinking peptide has a D-Alanine-D-Lactate C-terminus (in the VanA, VanB, and VanD phenotypes) a high level of vancomycin resistance results. When the alternative is a D-Alanine-D-Serine C-terminus (in VanC, VanE, and VanG) a low level of resistance is the result.

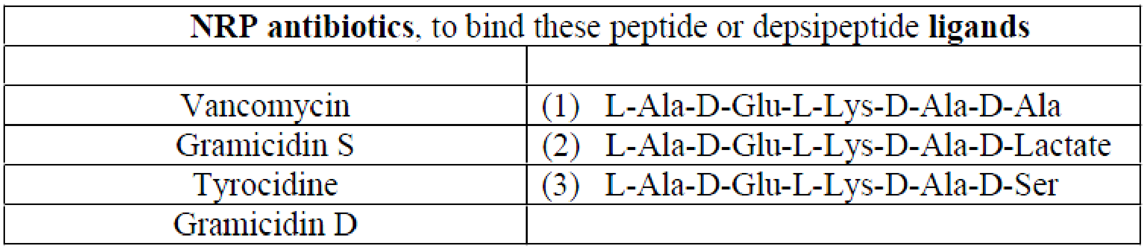
3.1.2 Novel Approaches to Combat Vancomycin Resistance

3.1.2.1 Computational Redesign of Nonribosomal Peptide Antibiotics

The first approach that we considered was to extend the Donald Lab's *in silico* protein-redesign suite of algorithms known as OSPREY [Gainza, et al., 2013], creating and adding software to redesign nonribosomal peptide (NRP) antibiotics. We would use OSPREY to redesign several NRP antibiotics (described in section 2.1.2) to find one or more that bind tightly enough to one of the two clinically relevant ligands, #2 and #3 in Table 10, along with redesign-target NRP antibiotics, that it can inhibit PG crosslinking, and cell wall biosynthesis, in a vancomycin-resistant microbe. We considered redesigning vancomycin first because it's a small, fairly rigid, molecule for which extensive NMR structural analysis has been published [Harris, et al., 1983; Groves, et al., 1994; Pearce and Williams, 1995; Pearce, et al., 1995] along with crystal structures for it alone [Schafer, et al., 1996] and in complexes [Nitanai, et al., 2009] with analogs of the C-termini of ligands #1 and #2 (Table 10).

Any novel NRP antibiotics that we found to have sufficient potency against vancomycin-resistant microbes would lead us to then use OSPREY to redesign one or more of the NRP enzymes that biosynthesize the wild-type version of that NRP antibiotic, in order to discover mutant versions of those NRP enzymes that could then be used to chemoenzymatically mass-produce this novel antibiotic.

Table 10: Computational Redesign Targets



NRP antibiotics, to bind these peptide or depsipeptide **ligands**	
Vancomycin	(1) L-Ala-D-Glu-L-Lys-D-Ala-D-Ala
Gramicidin S	(2) L-Ala-D-Glu-L-Lys-D-Ala-D-Lactate
Tyrocidine	(3) L-Ala-D-Glu-L-Lys-D-Ala-D-Ser
Gramicidin D	

3.1.2.2 Computational Redesign of Vancomycin Glycosyltlransferases to Chemoenzymatically Produce Glycovariants of Vancomycin

We also considered using the power of OSPREY [Gainza, et al., 2013] to redesign one of the glycosyltransferases that catalyze one of the final steps of the biosynthesis of wild-type vancomycin so that we could make glycovariants of this powerful NRP antibiotic. This is the approach that we chose to pursue.

3.1.3 Biosynthesis of Wild-Type Vancomycin

The biosynthesis of vancomycin, by the actinomycete *Amycolatopsis orientalis*, occurs in three discrete phases, each of which includes multiple enzyme-catalyzed chemical transformations. The first phase is the assembly of the heptapeptide core by a nonribosomal peptide synthetase (NRPS) pathway, as described previously in Chapter 2, section 2.1.2. The NRPS pathway for Phase 1 of vancomycin biosynthesis consists of three NRPS enzymes. The cartoon in Figure 14 indicates which catalytic domains each of these three NRPS enzymes contains, with the first two enzymes containing three modules each, and the third NRPS enzyme having only one module [Hubbard and Walsh, 2003]. As described in Chapter 2, each module selects, activates, and links one specific amino acid into the growing chain.

The final product of Phase 1 of vancomycin biosynthesis is the heptapeptide chain shown in Figure 14. This heptapeptide chain is exactly the same, at this stage of biosynthesis, for all vancomycin-group, also called vancomycin-type, glycopeptides. Over 200 glycopeptides are known, and they all have a heptapeptide core, differing only in the amino acid identity at positions #1 and #3 [Rao, et al., 1995]. The identity of the amino acids at positions #1 and #3 is used to classify all known glycopeptides into only four types, as shown in Figure 15.

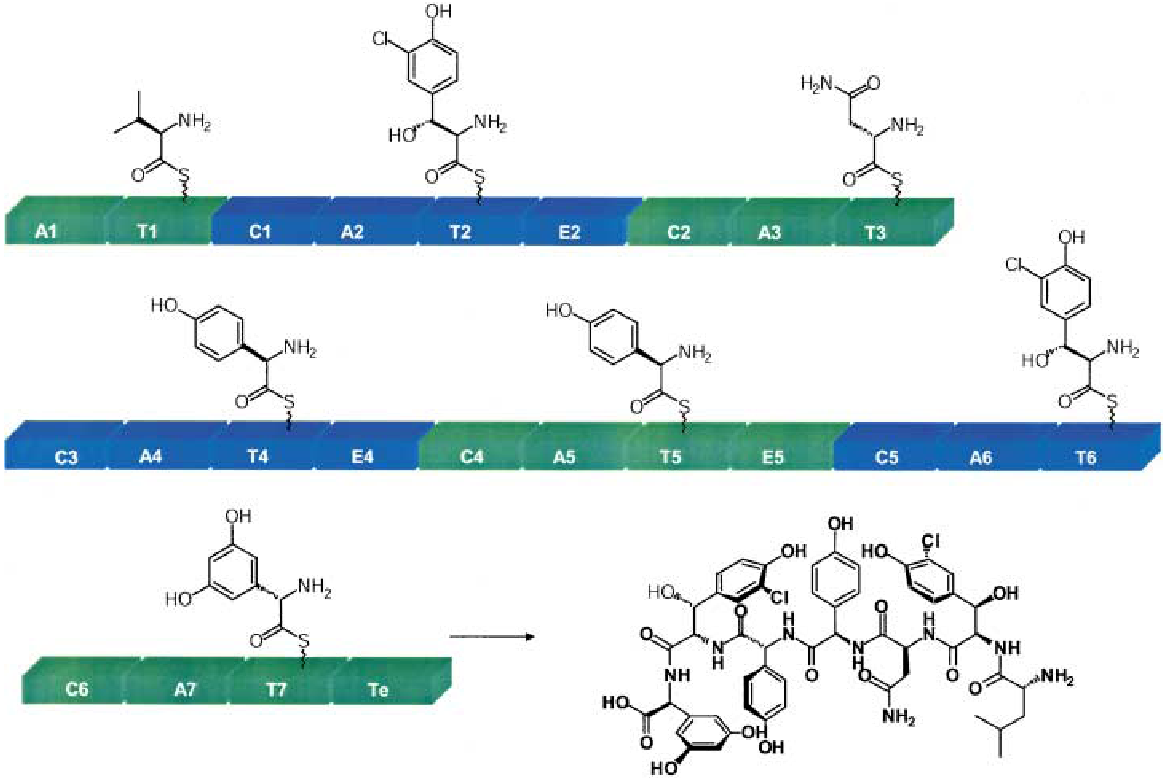
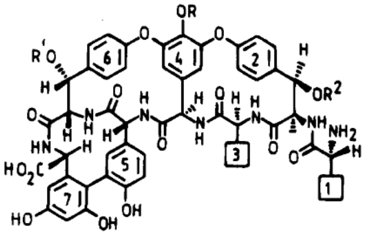


Figure 14: Phase 1 of biosynthesis of vancomycin. Three NRPS enzymes are responsible for selecting, activating, and linking seven amino acids into the heptapeptide chain shown. [Reprinted from Hubbard and Walsh, 2003]



type	1	2	3	4	5	6	7
vancomycin	Leu	β-OH Tyr	Asn	p-OHPhg	p-OHPhg	β-OHTyr	3,5-OHPhg
ristocetin[b]	p-OHPhg	β-OHTyr	3,5-OHPhg	p-OHPhg	p-OHPhg	β-OHTyr	3,5-OHPhg
avoarcin	p-OHPhg	β-OHTyr	p-OHPhg	p-OHPhg	p-OHPhg	β-OHTyr	3,5-OHPhg
synmonicin[c]	p-OHPhg	β-OHTyr	Met	p-OHPhg	p-OHPhg	β-OHTyr	3,5-OHPhg

[a] Leu = Leucine; β-OHTyr = β-hydroxytyrosine; Asn = aspargine; p-OHPhg = (p-hydroxyphenyl)glycine; 3,5-OHPhg = 3,5-dihydroxyphenylglycine; Met = Methionine. [b] Terminl carboxyl is $COOCH_3$. [c] Terminal amino group is NCH_3.

Figure 15: The four known types of glycopeptides are distinguished by the amino acids at positions #1 and #3 in the heptapeptide core. [Reprinted from Rao, et al., 1995]

In the second phase of the biosynthesis of vancomycin, three of the seven amino acid side chains in the heptapeptide chain are covalently linked to each other. This crosslinking occurs only between three particular pairs of the aromatic rings in five of the side chains. The end result is the oxidative coupling of the aromatic rings in residues #6 and #4, #4 and #2, and #7 and #5, and possibly in that exact order [Hubbard and Walsh, 2003]. These three covalent cross links are highlighted in Figure 16, and are shown in the order in which it is believed they occur based on gene-knockout or gene-knockdown experiments. Regardless of the order in which these three cross links are formed, the final result is a conformationally less-flexible crosslinked heptapeptide core.

The third and final phase of the biosynthesis of vancomycin is the glycosylation phase, where a glycan is covalently attached to the crosslinked heptapeptide core [Walsh, et al., 2003]. This is where the diversity within a particular glycopeptide type is created. The many members of the vancomycin group of glycopeptides differ by the characteristic glycans that become attached to their common crosslinked heptapeptide core in this final phase of biosynthesis. Vancomycin-group member chloroeremomycin differs from vancomycin by the type of sugars in its glycans and in the location of its glycans on its heptapeptide core, as shown in Figure 17. Also shown in Figure 17, next to each sugar, are the abbreviated names of the glycosyltransferases that catalyze the covalent attachment of that sugar during the glycosylation phase of biosynthesis.

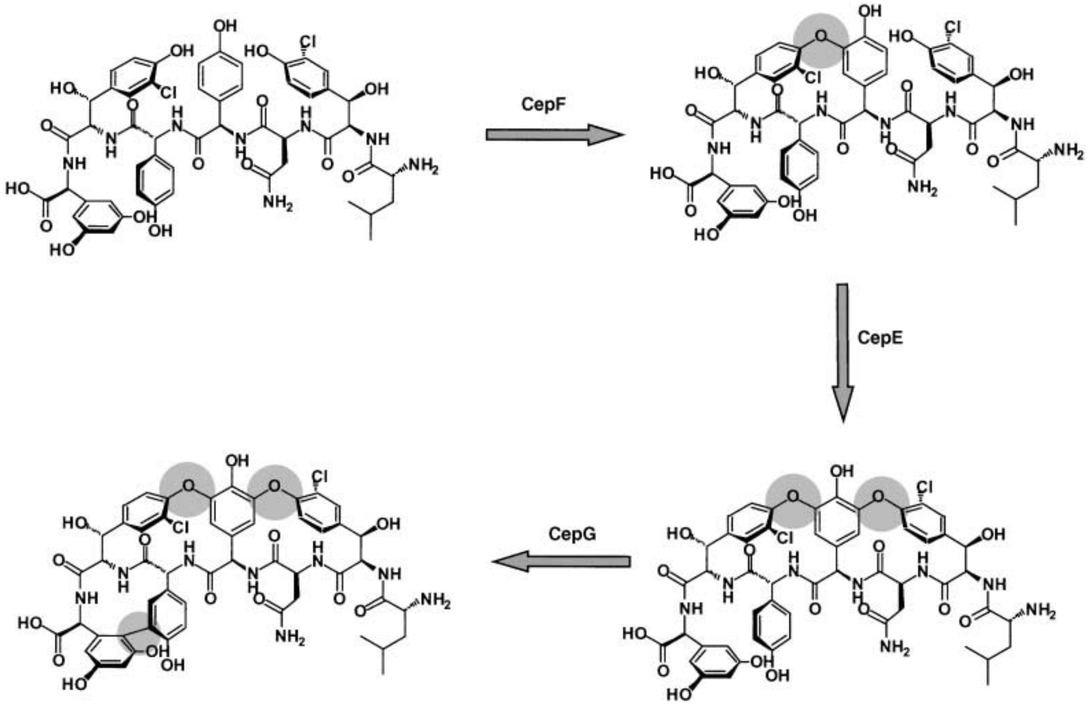

CepF
CepE
CepG

Figure 16: Phase 2 of biosynthesis of vancomycin. The production of the crosslinked heptapeptide core by the oxidative coupling of three pairs of aromatic rings. [Reprinted from Hubbard and Walsh, 2003]

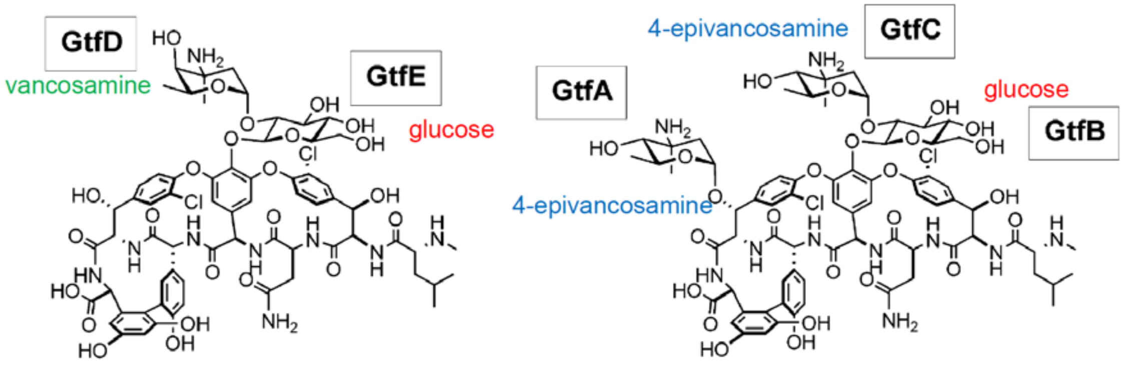

GtfD
vancosamine
GtfE
glucose
4-epivancosamine
GtfC
GtfA
glucose
GtfB
4-epivancosamine

Figure 17: Phase 3 of biosynthesis of vancomycin. Characteristic glycosylation of the crosslinked heptapeptide core shown for two members of the vancomycin group of glycopeptides; vancomycin (left) and chloroeremomycin (right). Each sugar molecule is labeled with its name and the abbreviated name of the glycosyltransferase that catalyzes its covalent attachment. [Reprinted from Walsh, et al., 2003]

3.1.4 Vancomycin Glycosyltlransferases

Glycosyltransferases (GTs) are enzymes that are required for the biosynthesis of glycans of all sizes, ranging from one single sugar molecule to polysaccharides containing thousands of sugar monomers linked together to create linear or branched polymers. Glycosylated molecules typically consist of one or more glycans attached to a non-sugar component called the aglycone. In the case of the two glycopeptides antibiotics shown in Figure 17, vancomycin has one glycan, chloroeremomycin has two glycans, and these two glycopeptides have the same aglycone, called the heptapeptide core. GTs act by transferring one sugar molecule at a time, from an activated sugar-donor molecule to a sugar-acceptor molecule, by breaking one glycosidic bond in the donor and creating a new glycosidic bond between the released sugar and the acceptor (aka, aglycone) molecule. The general GT-catalyzed sugar-transfer reaction is summarized in Figure 18, showing the two products that are possible, depending on the 3D structure of the active site of the GT that catalyzed the reaction.

In most sugar-donors for GT-catalyzed sugar transfer, the sugar is activated by having a phosphate group bonded to its anomeric carbon, as shown in Figure 18, giving it a high-energy glycosidic bond. The **R** bonded to the phosphate group in Figure 18 represents many possible chemical structures depending on whether the sugar-donor represented is a nucleotide diphosphate sugar (NDP-sugar), a nucleotide monophosphate sugar (NMP-sugar), or a lipid phosphate sugar. The configuration of the glycosidic bond in the sugar-donor, between the phosphate group and the anomeric carbon of the sugar, is fixed in one of two possibilities, is called the anomeric configuration of that sugar, and is designated either α or β. The anomeric configuration of the sugar-donor affects the shape of the molecule and most GTs will only bind productively to a sugar-donor that has the correct sugar, bonded to the correct phosphate-containing compound, by an activated glycosidic bond with the correct anomeric configuration. However, there are some GTs that have less stringent requirements for these three parameters that determine the shape of the sugar-donor molecule that they will bind to productively. Most GTs can be characterized by

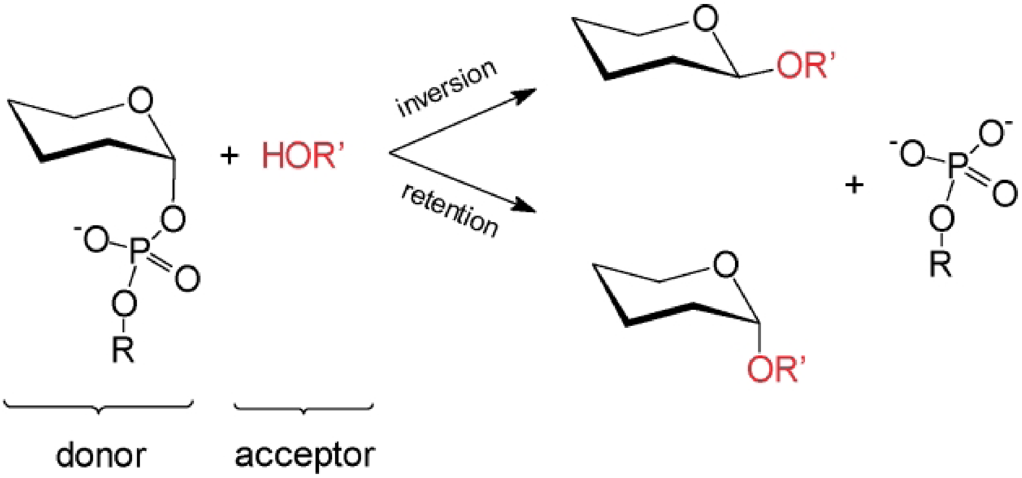


Figure 18: Summary of the chemical reactions that are catalyzed by glycosyltransferases (GTs). The product molecules will have a new glycosidic bond with either inversion or retention of the anomeric configuration of the glycosidic bond broken in the sugar-donors, depending on the GT. [Reprinted from Coutinho, et al., 2003]

whether the new glycosidic bond they catalyze the formation of has either the opposite (aka, inverted) or the same (aka, retained) anomeric configuration as the activated glycosidic bond in the sugar-donor. A GT is usually either an inverting GT or a retaining GT, but not both. The general mechanism of an inverting GT, acting on a sugar-donor with the *α* anomeric configuration, is shown in Figure 19.

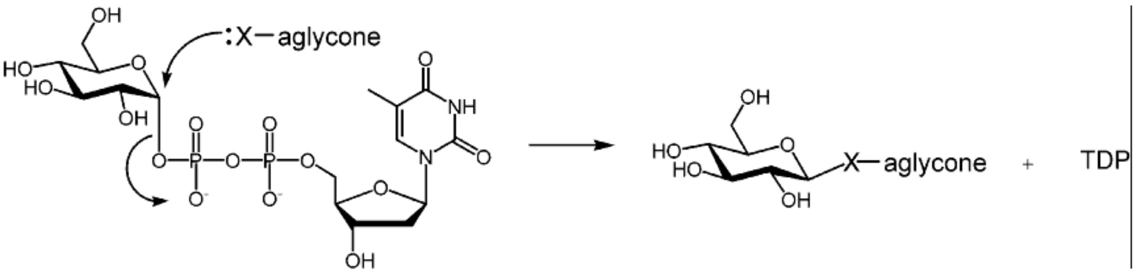


Figure 19: Mechanism of catalysis for an inverting glycosyltransferase with TDP-α-glucose as its sugar-donor substrate, creating a glycosidic bond with β configuration. [Reprinted from Walsh, et al., 2003]

Interest in GTs has been growing steadily for over twenty years now. In 1997 a classification system for GTs based on their amino acid sequence was proposed [references in Coutinho, et al., 2003] and the 600 GT sequences available then were organized into 27 families. By 2003 there were over 7,200 sequences available and they were divided into 65 families, available on the Carbohydrate-Active enZymes (CAZy) database [Coutinho, et al., 2003]. As November 2014, the CAZy database (www.cazy.org) contains over 157,000 GT sequences that are classified into 96 families, and another 3,000 sequences that have not been classified yet.

The five vancomycin-group GTs shown in Figure 17, Gtf's A through E, have a lot in common. Based on their amino acid sequences they are all in the same CAZy family: GT1. The GT1 family is one of the larger families, currently with 7,335 members. As members of GT1, all five of these vancomycin-group enzymes are identified as being inverting GTs, and they all have the same global fold, designated GT-B. Most all of the GTs with known 3D structures can be classified as having one of two common global folds, designated GT-A and GT-B, and shown in Figure 20. Three of these five vancomycin-group GTs have had their X-ray crystal structures published: GtfB in 2001 [Mulichak, et al., 2001], GtfA in 2003 [Mulichak, et al., 2003], and GtfD in 2004 [Mulichak, et al., 2004].

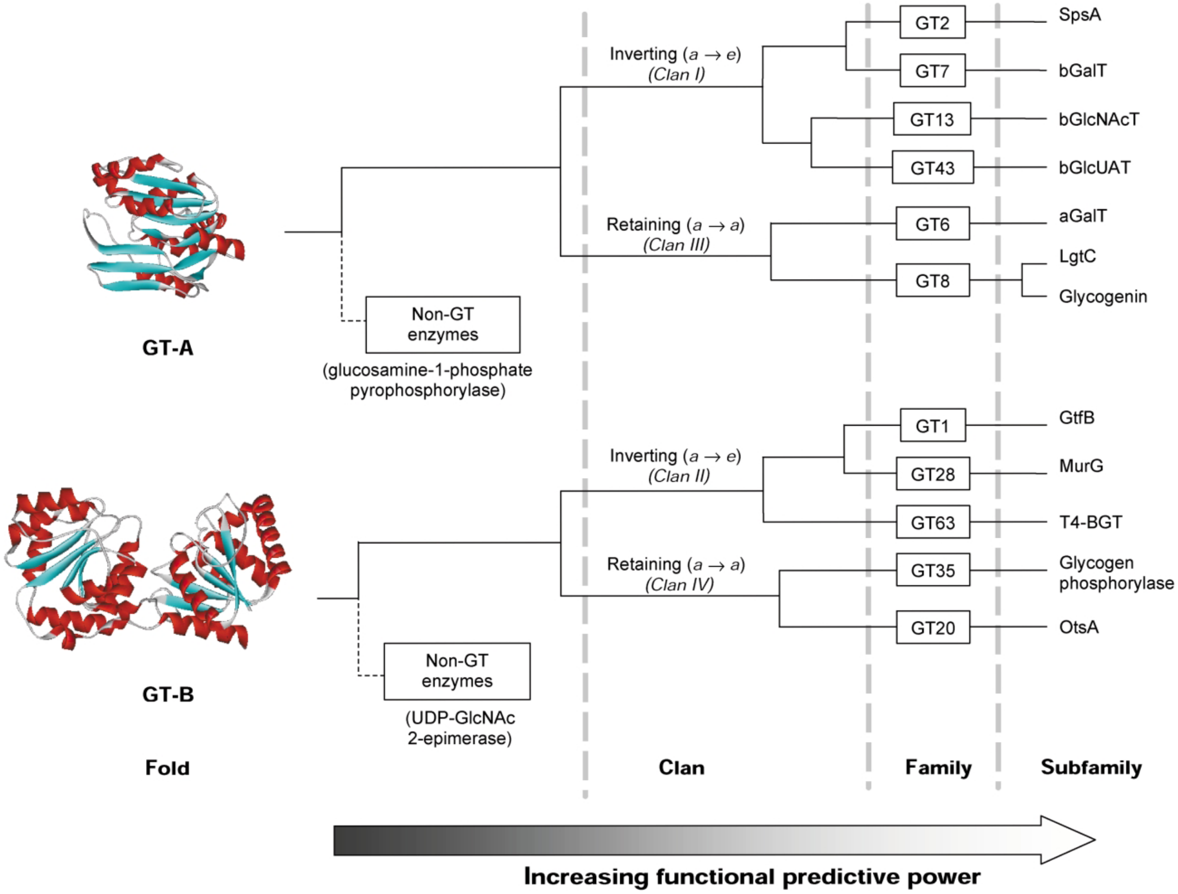

Inverting (a → e)
(Clan I)
Retaining (a → a)
(Clan III)
Inverting (a → e)
(Clan II)
Retaining (a → a)
(Clan IV)
GT2
GT7
GT13
GT43
GT6
GT8
GT1
GT28
GT63
GT35
GT20
SpsA
bGalT
bGlcNAcT
bGlcUAT
aGalT
LgtC
Glycogenin
GtfB
MurG
T4-BGT
Glycogen
phosphorylase
OtsA
GT-A
Non-GT
enzymes
(glucosamine-1-phosphate
pyrophosphorylase)
GT-B
Non-GT
enzymes
(UDP-GlcNAc
2-epimerase)
Fold
Clan
Family
Subfamily
Increasing functional predictive power

Figure 20: The sequence-based classification system CAZy, for glycosyltransferases (GTs). [Reprinted from Coutinho, et al., 2003]

3.2 Materials and Methods

3.2.1 Materials

In 2011, we obtained the GtfB gene with a C-terminal His tag, GtfB-His_6, in the expression vector pET22b, in *Escherichia coli* BL21(DE3), from Dr. Christopher Walsh [Mulichak, et al., 2001]. In 2013, we obtained the GtfD gene with a C-terminal His tag, GtfD-His_6, also in the expression vector pET22b, in *Escherichia coli* BL21(DE3), from Dr. Christopher Walsh [Mulichak, et al., 2004].

3.2.2 Protein Expression and Purification

BL21(DE3) *E. coli* cells containing GtfB-His_6 in the expression vector pET22b were grown in Luria Broth to OD = 0.7 and then induced with 0.5 mM IPTG for 15 hours at 25 C. The *E. coli* cells were harvested by centrifugation, resuspended in lysis buffer (wash buffer (400 mM NaCl, 25 mM Tris-Cl (pH=8.0), 20 mM Imidazole, 0.5 mM TCEP) plus a cocktail of three protease inhibitors) and then lysed by a French press. The cell debris was pelleted by centrifugation at 38,000 x g for 30 min. and the supernatant was incubated with Ni-NTA beads, equilibrated in wash buffer, for 90 min at 4 C. These Ni-NTA beads were then washed with 80-100 volumes of wash buffer and then eluted with a step gradient of 40, 80, 120, and finally 160 mM Imidazole in 400 mM NaCl, 25 mM

Tris-Cl (pH=8.0), 0.5 mM TCEP. Eluted proteins were further purified by size exclusion chromatography over Superdex 200.

3.3 Results

3.3.1 GtfB

We produce purified proteins in our lab using standard molecular biology methods, bacteria, and reagents. Briefly, we chemically induce the over expression of an engineered heterologous gene on a plasmid in a laboratory-modified strain of E. coli. During gene expression, proteins are synthesized, by the machinery of the E. coli, using amino acids made from compounds available in the growth medium. After gene expression, we lyse the bacteria and purify our protein of interest, from the protein-enriched cell lysate, by the ability of its engineered terminal Histidine residues to bind strongly to an immobilized divalent metal cation. We can control the isotopic content of our protein of interest by controlling the isotopic content of the molecules in the bacterial growth medium.

To make proteins in which all of their nitrogen atoms are the rare, nuclear magnetic resonance (NMR) -active, stable isotope ^{15}N instead of the naturally more abundant, but NMR-inactive, stable isotope ^{14}N we use a bacterial growth medium in which all of the nitrogen atoms are the rare isotope ^{15}N. After making several test preparations of unlabeled GtfB-His$_6$, using bacterial growth medium with only naturally occurring nitrogen, mostly ^{14}N, and the naturally occurring NMR-active isotope of hydrogen, ^{1}H, I made one preparation of ^{15}N-labeled GtfB-His$_6$. From a two-liter culture I was able to purify 12 mg of ^{15}N-labeled GtfB-His$_6$, and used that to collect our first NMR data on GtfB, which gave us the [^{1}H, ^{15}N] HSQC spectrum shown in Figure 21.

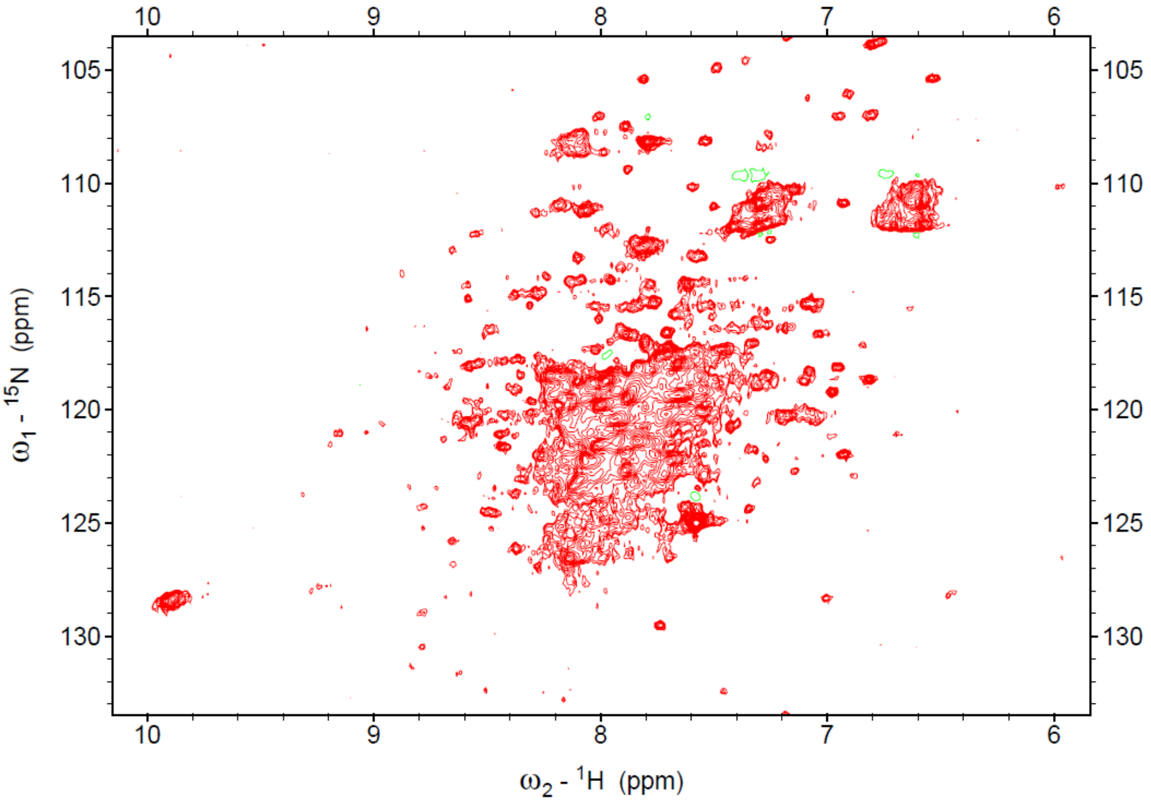


Figure 21: [^{1}H, ^{15}N] HSQC spectrum of ^{15}N-labeled GtfB-His$_6$.

To assess the suitability of a protein for 3D structure analysis by solution-phase NMR we usually begin by collecting [^{1}H, ^{15}N] HSQC data. One reason for this choice is that the bacterial growth medium needed for making ^{15}N-labeled proteins is much less expensive than the growth medium needed for making ^{13}C-labeled proteins. Another reason is that the data obtained from an [^{1}H, ^{15}N] HSQC experiment is relatively easy to interpret quickly, and provides essential information for an initial assessment of the protein. The heteronuclear single-quantum correlation (HSQC) NMR experiment produces, in the resulting frequency-domain spectrum (e.g., Figure 21), one cross-peak for each pair of NMR-active nuclei that are correlated with each other by either a scalar coupling through a chemical bond or by a dipolar coupling through space. Each cross-peak in an [^{1}H, ^{15}N] HSQC spectrum represents a unique pair of one ^{1}H and one ^{15}N that are chemically bonded to each other. The location of each cross-peak in an [^{1}H, ^{15}N] HSQC frequency-domain spectrum identifies the chemical shifts of the ^{1}H and the ^{15}N nuclei that are bonded to each other in the ^{15}N-labeled protein sample.

3.3.2 GtfD

A computational approach to analyzing the substrate-binding residues of GtfD was begun in collaboration with Ryan Muraglia, a rotation student in Bruce's lab in the Spring semester of 2013. Ryan took on the rotation project of using OSPREY to determine which amino acids in the active site of GtfD are involved in binding the sugar-donor substrate. I gave Ryan project guidance, answering any questions he had about the aims of the project and selecting the first redesign target, while Jonathon Jou gave Ryan technical assistance with using OSPREY.

3.3.2.1 Determination of the active site residues of GtfD that bind the sugar-donor substrate

The published X-ray crystal structure of GtfD is as a complex with two ligands: the complete sugar-acceptor substrate, desvancosaminyl vancomycin (DVV), and part of the sugar-donor substrate, specifically only the TDP of the sugar-donor TDP-β-vancosamine [Mulichak, et al., 2004]. The crystallized GtfD had 416 amino acids and Ryan decided to focus on those 25 amino acids that had one or more of their atoms within 4 Å of the sugar-donor substrate when bound to the active site. Ryan combined information from the following three sources to reach a decision about the mutabililty of each of those 25 amino acids, as listed in Table 11. First, the authors of this X-ray crystal structure of GtfD proposed the likely position of the missing β-vancosamine moiety of the sugar-donor substrate, and show it modeled it into their 3D structure [Mulichak, et al., 2004] for GtfD, attached to the TDP that is complexed there. Second, Ryan could not find any published 3D coordinates for a TDP-β-vancosamine molecule, as a ligand, in the PDB. Thus, he decided to model in TDP-α-glucose by taking sets of coordinates for TDP-α-glucose that he could find, as ligands, in the PDB and overlaying the TDP in those TDP-α-glucose ligands with the TDP bound in the sugar-donor binding site of

GtfD [Mulichak, et al., 2004]. Then MolProbity [Chen, et al., 2010] was used to analyze the contacts between TDP-α-glucose and the active site residues of GtfD. Finally, visual inspection of the location and proximity of each of the atoms of the TDP-α-glucose to the active site residues of GtfD was used to reach the conclusions listed in Table 11.

Table 11: Mutability of GtfD amino acids within 4 Å of the sugar-donor substrate

Residue	Comment	Verdict
Thr10	Catalytic conserved	Not mutable
Arg11	Oriented towards TDP moiety	Low potential
Gly12	Contacts tail and glucose	Mutable
Asp13	Catalytic conserved	Not mutable
Glu15	Only contacts TDP moiety	Not mutable
Pro126	Changed to His in glucose transferring GtfB/E	Promising mutation to His
Tyr130	Currently no contacts: would need an even bigger sc	Low potential
Leu218	Only contacts TDP moiety	Not mutable
Arg223	Only contacts TDP moiety	Not mutable
Gly245	No reason to rule out	Mutable
Ser246	Conserved, but different rotamers observed	Mutable
Glu293	Only contacts TDP moiety	Not mutable
Val294	Only contacts TDP moiety	Not mutable
Asn295	Only contacts TDP moiety	Not mutable
Phe296	Only contacts TDP moiety	Not mutable
Hie309	No reason to rule out	Mutable
Ser311	Changed to Gly in glucose transferring GtfB/E	Promising mutation to Gly
Ala312	No reason to rule out	Mutable
Gly313	No reason to rule out	Mutable
Thr314	Somewhat distant to tail-glucose	Low potential
Val317	Only contacts TDP moiety	Not mutable
Asn331	Somewhat distant to tail-glucose	Low potential
Thr332	Somewhat distant to tail-glucose	Low potential
Asp333	No reason to rule out	Mutable
Gln334	No reason to rule out	Mutable

3.4 Discussion

The great utility, in protein NMR, of the [^{1}H, ^{15}N] HSQC experiment derives from the following facts. In a ^{15}N-labeled protein, every bonded pair of ^{1}H and ^{15}N atoms can produce one cross-peak in the frequency spectrum and all amino acids contain at least one ^{15}N atom. Because the chemical shifts of the NMR-active ^{1}H and ^{15}N nuclei are determined by the 3D location of their nearest-neighboring atoms, a stably folded protein yields an [^{1}H, ^{15}N] HSQC frequency spectrum with well-dispersed cross-peaks, while a protein, or just a portion of a protein, that is randomly changing its 3D conformation will yield an [^{1}H, ^{15}N] HSQC frequency spectrum in which most of its cross-peaks have the same location, due to averaging out of their ^{1}H and ^{15}N electronic environments and thus their chemical shifts are all mostly the same. In addition to being able to distinguish between well-folded and poorly-folded proteins, an [^{1}H, ^{15}N] HSQC frequency spectrum reveals whether a large well-folded protein has well-dispersed cross-peaks, which can reduce the time-consuming problems of disambiguating overlapping cross-peaks.

The initial NMR data collected with the protein GtfB for this research project has been encouraging. The first HSQC frequency spectrum of ^{15}N-labeled GtfB-His$_6$, shown in Figure 21, looks encouraging in terms of cross-peak dispersion for a protein of this size, 415 residues. This first HSQC spectrum was collected at 25 °C. Since then I found that this protein will remain in solution at 42 °C for more than four weeks. This suggests that we can get sharper cross-peaks, and do higher multidimensional NMR experiments with it, just by raising the temperature during NMR data collection.

Ryan Muraglia's work on predicting the residues of GtfD that are involved in binding the native sugar-donor TDP-β-vancosamine, was a very useful start. Ryan uncovered some difficulties in doing *in silico* docking of a ligand to a protein. Of course, Ryan's rotation project conclusions, summarized in Table 11 will need to be confirmed by doing more *in silico* docking experiments using different ligands than he used. For example, a means to dock the actual native sugar-donor, TDP-β-vancosamine, in wild type GtfD, instead of TDP-α-glucose, needs to be found and used. Doing that may confirm Ryan's conclusions, or it may lead to a slightly different set of mutable residues, and thus the need to do even more *in silico* docking to decide which set is most likely to be correct. With the uncertainties and technical difficulties of *in silico* docking, it's still unclear whether the set of active-site amino acids involved in binding the sugar-donor substrate in GtfD can be found more quickly, using *in silico* docking, than the analogous set of amino acids can be found in GtfB using NMR ligand titration experiments, as described in the next section.

3.5 Future Work

3.5.1 Determining the GtfB amino acids involved in binding both substrates

The X-ray crystal structure of GtfE has not been published yet. GtfB, from the chloroeremomycin biosynthetic pathway, is the functional analogue of GtfE in the vancomycin biosynthetic pathway. The X-ray crystal structure of GtfB was determined by analysis at 1.8 Å resolution and published in 2001 [Mulichak, et al., 2001]. Therefore, we wish to redesign GtfB and use it to replace GtfE *in vitro* and *in vivo* in the synthesis of glycovariants of vancomycin. Because the crystal structure for GtfB did not contain either of its two native substrates we plan to do NMR ligand titration experiments, with the sugar-donor and the sugar-acceptor substrates, first individually and then together, to determine which amino acids are involved in binding these two substrates.

These NMR ligand titration experiments would begin with measuring [^{1}H, ^{15}N] HSQC data for ^{15}N-labeled GtfB alone and then with increasing amounts of substrate included. We would expect to see a small subset, maybe 20-40 of these HSQC crosspeaks shifted in proportion to the amount of substrate added. Next we would assign as many crosspeaks as we could, with the primary goal being to assign all of the peaks that are shifted by the presence of substrate. Assigning these peaks will be made easier since we have a published X-ray crystal structure for GtfB. When we have assigned all of the shifted peaks, we can locate those amino acids in the 3D crystal structure of GtfB and determine which of them are in the active site area and which, if any, are not. Only those in the active site are likely to be involved in binding the substrate that caused those HSQC crosspeaks to shift. Once we have a set of active site residues that are likely to be involved in binding substrate, we can try different docking modes of that substrate in the 3D crystal structure and then use OSPREY software to compare those docking modes to find the most probable one.

This process of determining which residues are involved in binding substrates will be assisted by the fact that there is a wealth of published results in the literature that suggest that in GTs with the GT-B global fold (Figure 22), the sugar-acceptor binds to the N-terminal domain while the sugar-donor binds to the C-terminal domain, and that each substrate binds its domain independently.

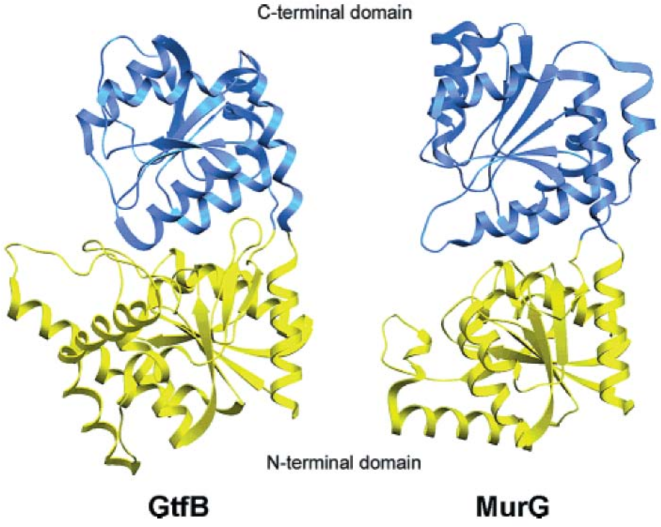


Figure 22: Two glycosyltransferases with the GT-B fold [Reprinted from Walsh, et al., 2003]

This independent binding of GT-B fold GTs, one substrate to each of the two domains, suggests the alternative approach of doing these NMR ligand titration experiments on each of these two domains separately. These single-domain ligand titration experiments would be almost identical to those described above for whole GtfB, except that each of the two GtfB domains would be expressed separately as a ^{15}N-labeled protein, and then [^{1}H, ^{15}N] HSQC data collected from it, alone and then with each substrate titrated in, separately and then together, and even the other GtfB domain, unlabeled. The advantage of doing NMR experiments on each domain separately is that cutting the mass of the ^{15}N-labeled protein roughly in half will result in much more than a two-fold reduction in the time to analyze the data because there will be fewer crosspeaks and they will be sharper because of the smaller mass. One technical problem with doing this with GTs with the GT-B fold, like GtfB, is that they have one α-helix at the C-terminus that is part of the N-terminal domain as shown in Figure 22. This is not an insurmountable technical challenge, and OSPREY software can be used to redesign the individual N- and C-terminal domains to be soluble as separate domains.

3.5.2 Finding optimal growth conditions for *Amycolatopsis orientalis*

There are several reasons to develop, early on, the ability to grow *Amycolatopsis orientalis*, the microbe that sysnthesizes and secretes vancomycin. In the long term, we will need to have expertise growing *A. orientalis* when we begin to develop the ability to knock in the mutant GT genes for specific aim #3. In the short term, we can produce the vancomycin aglycone core (VAC) that is the GtfB sugar-acceptor substrate by knocking out the GtfE gene in *A. orientalis.* Of course, to purify that VAC we will need to also develop a standard affinity column that also captures vancomycin by binding to the same binding site that the VAC has.

There are several publications, over the past twenty years, describing different growth media and protocols to induce *A. orientalis* to produce good amounts of vancomycin. I have summarized the best ones, for our purposes, that I have found so far, in Table 12.

Table 12: Growing *Amycolatopsis orientalis* to Produce Vancomycin

ATCC product information	Reference = McIntyre, et al., 1996	Reference = Jung, et al., 2007	Reference = Ayar-Kayali, 2011	Reference = Zeng, et al., 2013
#1877 Broth	Seed	Growth	M65 (spore-growth)	Gause's I synthetic
ISP Medium 1 (g/ L)	Medium (g/ L)	Medium (g/ L)	Solid Medium (g/ L)	Solid Medium (g/ L)
Pancreatic digest	Glucose (5)	Glucose (17)	Glucose (4)	?
of casein (5)	Peptone (5)	Peptone (11)	$CaCO_3$ (2)	
Yeast extract (3)	Yeast extract (2)	Yeast extract (3)	Yeast extract (4)	28 C/ 120-192 hr
	Soluble starch (10)	Malt extract (3)	Malt extract (10)	
			Starch (20)	
	30 C/ 48 hr	30 C/ 48 hr	Agar (12)	Seed
				Medium (g/ L)
#196 Agar	Semidefined	Seed		Glucose (17)
ISP Medium 2 (g/ L)	Medium (g/ L)	Medium (g/ L)		Tryptone (11)
				Yeast extract (3)
Dextrose (4)	Glucose (20)	Dextrin (50)		Malt extract (3)
Yeast extract (4)	Peptone (5)	Soybean flour (5)		
Malt extract (10)	$MgSO_4$ $6H_2O$ (0.75)	Potato protein (5)	Defined fermentation	30 C/ 24-36 hr
Agar (20)	NaCl (1)		Medium (g/ L)	
	KCl (0.5)	30 C/ 60 hr		
	trace metals (10 mL)		$MgSO_4$ $7H_2O$ (0.6)	
			KH_2PO_4 (3.5)	
	30 C/ 120 hr	Production	Asparagine (2)	Production
		Medium (g/ L)	Glycerol (10)	Medium (g/ L)
			MOPS (21)	
		Dextrin (140)	trace salts (1 mL)	Soluble starch (20)
		Soybean flour (30)		Glycerol (53)
		Potato protein (25)	28 C/ 96 hr	Potato protein (18)
		NaCl (1.2)		K_2HPO4 (0.1)
				NaCl (1)
		34 C/ 120 hr		KNO_3 (2)
				$MgSO_4$ $7H_2O$ (0.5)
				$FeSO_4$ $7H_2O$ (0.01)
				30 C/ 120 hr

3.5.3 Purifying vancomycin and glycovariants of vancomycin

Vancomycin, as well as the vancomycin aglycone core (VAC) and all planned glycol-variants of vancomycin, can be easily purified by its ability to bind to D-Ala–D-Ala dipeptides attached to a solid support by a suitable-length linker. This is the affinity chromatography matrix reported in 1987 to be optimal for purification of glycopeptide antibiotics [Folena-Wasserman, et al., 1987]. Such a chromatography matrix can easily be made from the commercially available components Affi-Gel 10 [Bio-Rad web site] and D-Ala–D-Ala dipeptide.

3.5.4 Determining the GtfD amino acids involved in binding both substrates

The X-ray crystal structure of the vancosaminyltransferase GtfD, from the vancomycin biosynthetic pathway, was determined at 2.0 Å resolution and published in 2004 [Mulichak, et al., 2004]. The published crystal structure for GtfD is as a complex with TDP, part of the sugar-donor substrate, and the complete sugar-acceptor, desvancosaminyl vancomycin [Mulichak, et al., 2004]. With that much of the two native substrates bound to GtfD in the crystal structure, we plan to use OSPREY to determine which amino acids are involved in binding the native sugar-donor TDP-β-Vancosamine.

3.5.5 Selecting a non-native GtfB sugar-donor substrate to be our first redesign target

We know that GtfB (PDB ID = 1iir) is in CAZy Family GT1, and that it is an inverting GT. We know that the native substrates for GtfB are as follows. The GtfB sugar-acceptor is the vancomycin aglycone core (VAC), and the sugar-donor is believed to be either UDP-α-Glucose or TDP-α-Glucose. Our first redesign goal for GtfB is to alter its sugar-donor binding site so that it will productively bind a non-native sugar-donor, but we have not decided which one.

3.5.6 Selecting a non-native GtfD sugar-donor substrate to be our first redesign target

We know that GtfD (PDB ID = 1rrv) is in CAZy Family GT1, and that it is an inverting GT. We know that the native substrates for GtfD are as follows. The GtfD sugar-acceptor is desvancosaminyl vancomycin (DVV), and the sugar-donor is believed to be TDP-β-Vancosamine. Our first redesign goal for GtfD, like for GtfB, is to alter its sugar-donor binding site so that it will productively bind a non-native sugar-donor. However, in the case of GtfD we have chosen our first redesign target for its sugar-donor binding site. We plan to redesign its sugar-donor binding site so it will bind either UDP-α-Glucose or TDP-α-Glucose. We selected this as our first redesign target for GtfD because one or both of those two non-native sugar-donors are present in the cell, because they are the native sugar-donors for GtfB and its analogue GtfE. Thus we expect to be able to screen these redesigned GtfD proteins *in vivo* without needing to supply an additional sugar-donor for them inside the *Amycolatopsis orientalis* cell.

3.5.7 Finding or adapting a gene modification protocol for *Amycolatopsis orientalis*

There is a fair amount of published work in this area that can help us to find or develop protocols for disrupting and replacing a wild-type gene in *Amycolatopsis orientalis*. This is because the *Amycolatopsis* species and the *Streptomyces* species are actinomycetes, members of the same order (*Actinomycetales*), and in 1978 the *Streptomyces* species were the source of more than 60% of the known antibiotics [Bibb, et al., 1978]. Though laborious and difficult, the preparation of protoplasts was a required part of the early successful protocols for transforming actinomycetes. A 1987 paper [Matsushima, et al., 1987] describes a successful protocol for transforming *A. orientalis* protoplasts. A 1991 paper [Lal, et al., 1991] reports the development of a hybrid plasmid and a protocol to transform it into two related *Amycolatopsis* species, *A. orientalis* and *A. mediterranei* by electroporation, with 100-fold greater efficiency than by the protoplast method, and with an additional 100-fold greater efficiency into *A. orientalis* compared to *A. mediterranei*. Three more papers published in the 1990s [Madon and Hutter, 1991;

Vrijbloed, et al., 1995; Pelzer, et al., 1997] describe methods of directly transforming the mycelium of *A. mediterranei* and *Amycolatopsis methanolica*, and the production of plasmids for gene disruption and for gene replacement [Pelzer, et al., 1997].

3.5.8 Determining the enzyme kinetics for wild type and mutant GtfB and GtfD

Measurements to determine enzyme kinetics of these wild type and mutant GTs would be done by the same methods as described for GtfB [Mulichak, et al., 2001] and GtfD [Mulichak, et al., 2004], using HPLC and MALDI mass spectrometry to analyze the reaction products at various time points.

4. Rationally Designed Enzymes and Enzyme Inhibitors, to Understand and Control Glycosyltransferases

4.1 Introduction

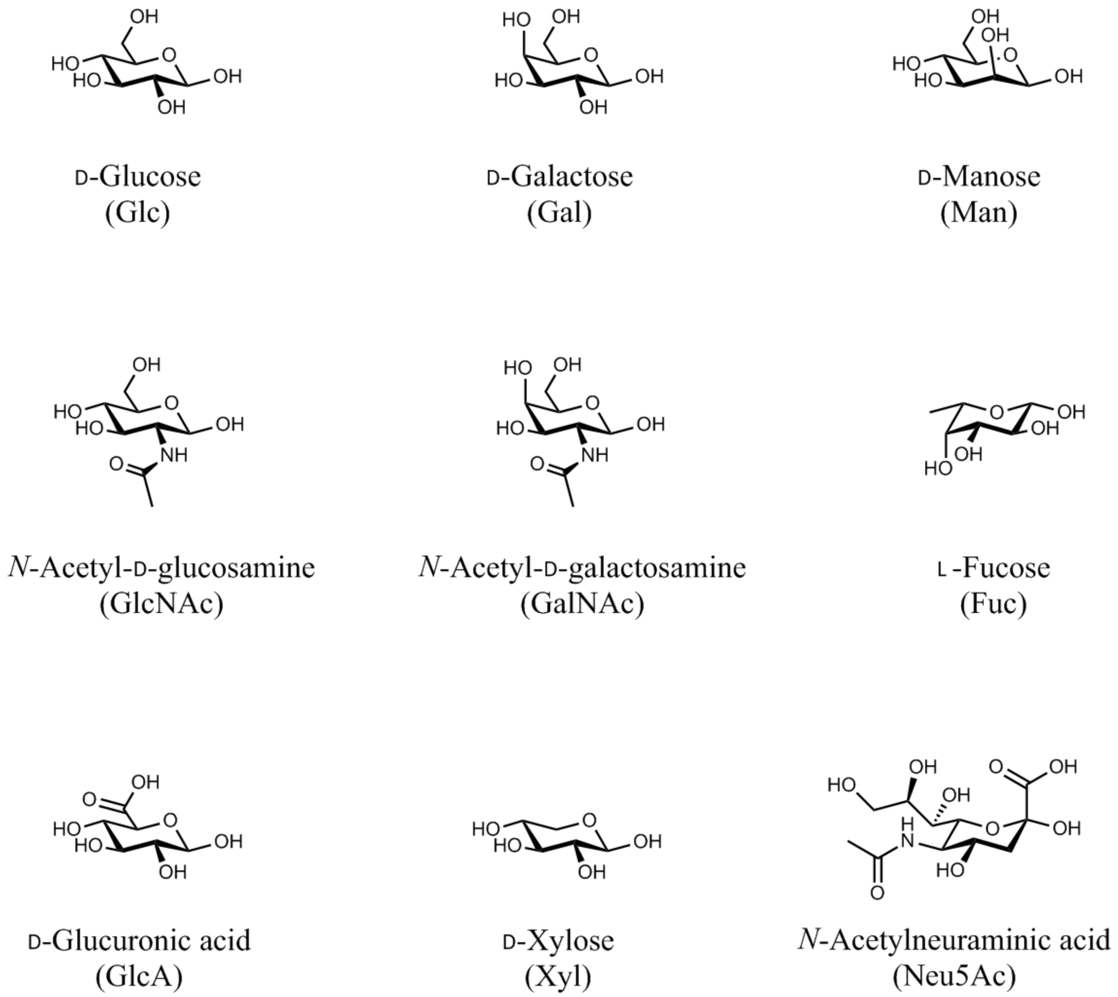
4.1.1 Carbohydrates and Glycans in Biology

Carbohydrates are the third major biopolymer, along with proteins and nucleic acids. However, their structures can be more complex than the other two biopolymers because there are many more available monomers, monosaccharides, and because carbohydrate chains, polysaccharides, can be linear or branched. Carbohydrates can be found in molecules composed of only carbohydrates. Or, they can be found covalently attached to other molecules in the form of glycoconjugates, which are most often either glycoproteins or glycolipids. Understanding the biological role of the carbohydrate in a glycoconjugate is one of the fundamental goals of the field of glycobiology. Because they have been studied for a long time, there are many synonyms for carbohydrates of different sizes, or in different biochemical situations. The field of glycobiology has chosen to use the term “glycan” to mean the carbohydrate portion, of any size, in any glycoconjugate. Then, the non-carbohydrate portion of any glycoconjugate is called the aglycone. In addition, the term glycan has come to be used as a generic term to refer to any carbohydrate, mono- or polysaccharide, either free or covalently bound to another molecule.

The monosaccharide is the smallest possible carbohydrate. It can vary in the number of carbon atoms it contains, but it cannot be hydrolyzed into a smaller carbohydrate. The monosaccharides found in Nature rarely have a carbon-chain with more than nine carbons, and most have either six or five. A monosaccharide's carbon-chain backbone is poly-hydroxylated and must contain a carbonyl group at one end, in the form of either an aldehyde or a ketone. With this broad definition, thousands of different monosaccharides could occur naturally, but in fact only a few hundred do. Because monosaccharides in Nature must be biosynthesized and metabolized by enzymes, every organism has a relatively small set of monosaccharides that it contains or is exposed to. There are only nine monosaccharides that are commonly found in vertebrates, and those are shown in their most common forms in Figure 23.

4.1.2 The Glycan Processing Enzymes (GPEs)

Another important difference between carbohydrates, or glycans, and proteins and nucleic acids is that their structures are not directly encoded in an organism's genome. Instead, glycans are biosynthesized by one or more glycan processing enzymes (GPEs) which are encoded in the genome. Thus glycan structures are indirectly encoded in the organism's genome and their biosynthesis in that organism depends on the concurrent availability of the appropriate GPE and its substrates. Thus, there is a tremendous amount of heterogeneity in the glycosylation patterns of naturally occurring


D-Glucose
(Glc)

D-Galactose
(Gal)

D-Manose
(Man)

N-Acetyl-D-glucosamine
(GlcNAc)

N-Acetyl-D-galactosamine
(GalNAc)

L-Fucose
(Fuc)

D-Glucuronic acid
(GlcA)

D-Xylose
(Xyl)

N-Acetylneuraminic acid
(Neu5Ac)

Figure 23: The most common forms of the nine monosaccharides commonly found in vertebrates.

glycoconjugates. The GPEs are a huge category of enzymes, with roughly 2% of the typical mammalian genome used to encode them. The two major kinds of GPEs are the anabolic glycosyltransferases (GTs) and the catabolic glycoside hydrolases (GHs). The GTs catalyze the transfer of one monosaccharide at a time to build up a glycan, and the

GHs break down glycans by catalyzing either the cleavage of a glycosidic bond in the middle of a polysaccharide, or at the end to release a single monosaccharide.

A glycan covalently bonded to a protein via a glycosidic bond to the nitrogen of an asparagine (Asn) side-chain is called an N-glycan. Of the five different types of N-glycan linkages observed in natural glycoproteins, the *N*-acetylglucosamine to Asn (GlcNAcβ1-Asn) is the most common. A brief look at the biosynthesis and maturation of GlcNAcβ1-Asn N-glycans provides an illustration of the interconnected roles of GTs and GHs, and an example of the health consequences when one of them is defective.

N-glycosylation of a protein occurs in the lumen of the endoplasmic reticulum (ER) as the nascent protein chain leaves the ribosome, on the cytoplasmic face of the ER, and enters the ER lumen, and is catalyzed by a complex of proteins called the oligosaccharyltransferase (OST). The OST transfers a 14-sugar glycan to the Asn nitrogen of the nascent protein from a glycolipid, $Glc_3Man_9GlcNAc_2$-P-P-Dolichol, shown in Figure 24. (The glycan in this and several other figures in this chapter will be drawn using the symbols recommended in Varki, et al., 2009, shown in Figure 25, to represent the common monosaccharides.) The biosynthesis of the 14-sugar glycolipid, Glc3Man9GlcNAc2-P-P-Dolichol, begins on the cytoplasmic face of the ER with the transfer of GlcNAc-Phosphate, from UDP-GlcNAc to P-Dolichol. After six different GTs each catalyze the transfer of one more sugar the 7-sugar glycolipid is flipped into the ER lumen where seven more GTs each transfer one more sugar to complete the biosynthesis

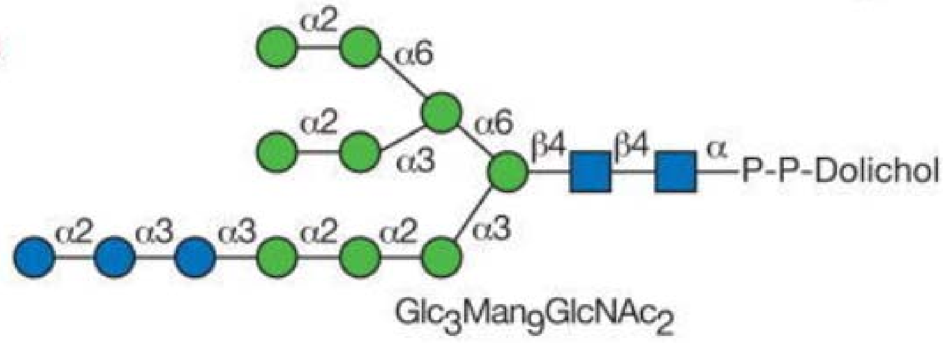

α2
α6
α2
α6
α3
β4
β4
α
P-P-Dolichol
α2
α3
α3
α2
α2
α3
Glc3Man9GlcNAc2

Figure 24: The glycolipid used as the donor-substrate by mammalian oligosaccharyltransferase (OST). [Reprinted from Varki, et al., 2009, Ch. 8]

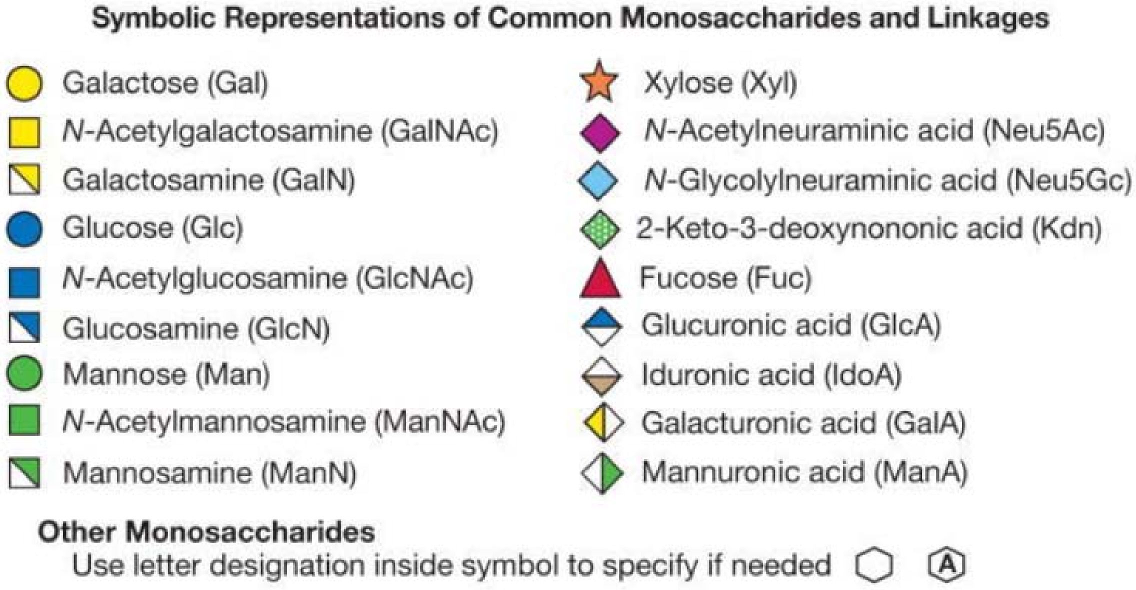

Symbolic Representations of Common Monosaccharides and Linkages
Galactose (Gal)
N-Acetylgalactosamine (GalNAc)
Galactosamine (GalN)
Glucose (Glc)
N-Acetylglucosamine (GlcNAc)
Glucosamine (GlcN)
Mannose (Man)
N-Acetylmannosamine (ManNAc)
Mannosamine (ManN)
Xylose (Xyl)
N-Acetylneuraminic acid (Neu5Ac)
N-Glycolylneuraminic acid (Neu5Gc)
2-Keto-3-deoxynononic acid (Kdn)
Fucose (Fuc)
Glucuronic acid (GlcA)
Iduronic acid (IdoA)
Galacturonic acid (GalA)
Mannuronic acid (ManA)
Other Monosaccharides
Use letter designation inside symbol to specify if needed
A

Figure 25: The monosaccharide symbols used here to represent glycan structures. [Reprinted from Varki, et al., 2009]

of the 14-sugar glycan. GlcNAcβ1-Asn N-glycosylation, of any glycoprotein, always begins with this same 14-sugar glycan being installed on each asparagine residue destined to be N-glycosylated. Then, GHs begin to do their job as the glycan "matures" while the glycoprotein folds and is trafficked from the lumen of the ER, through the different compartements of the Golgi, to its final destination. GHs do all of the early processing of the 14-sugar glycan, removing all three glucose and four of the manose residues. Then both GHs and more GTs are involved in the late stages of glycan maturation. Some examples of typical complex N-glycans found on mature glycoproteins are shown in Figure 26.

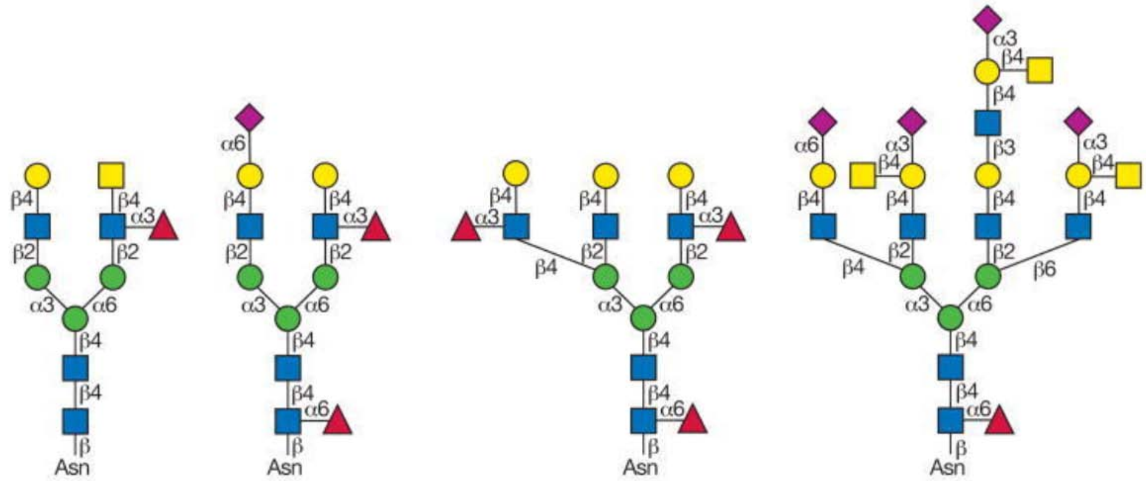
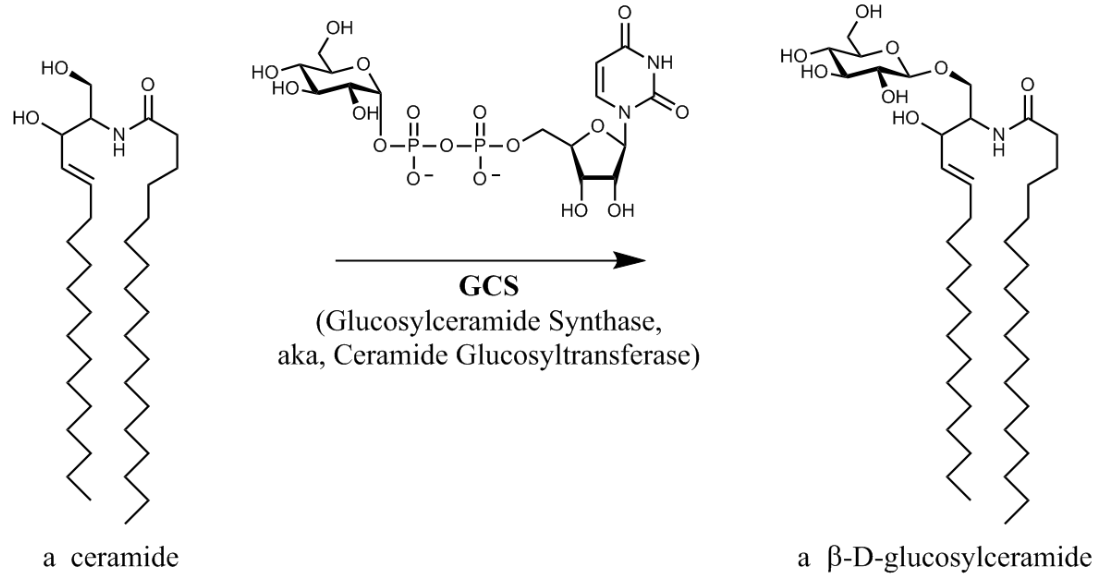
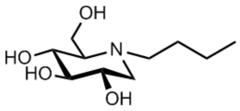

β4
β4
α3
β2
β2
α3
α6
β4
β4
β
Asn
α6
β4
β4
α3
β2
β2
α3
α6
β4
β4
α6
β
Asn
β4
β4
β4
α3
α3
β4
β2
β2
α3
α6
β4
β4
α6
β
Asn
α3
β4
β4
β3
α6
α3
β4
α3
β4
β4
β4
β4
β4
β4
β2
β2
β4
β6
α3
α6
β4
β4
α6
β
Asn

Figure 26: Typical mature GlcNAcβ1-Asn N-glycans. [Reprinted from Varki, et al., 2009, Ch. 8]

Lysosomes are membrane bound compartments where many macromolecules are degraded by a collection of 50-60 hydrolytic enzymes, including GHs to disassemble glycans. These hydrolytic lysosomal enzymes are N-glycoproteins that are trafficked from the ER to the lysosome as a result of being recognized in the *cis*-Golgi by a GlcNAc-Phospho-Transferase and then recognized in the *trans*-Golgi by a phosphodiester glycosidase, resulting in their N-glycans having one or more Manose residues converted to Manose-6-Phosphate (M6P) residues. M6P residues then bind to M6P receptors which traffic them to the lysosomes. When one or more of these lysosomal enzymes doesn't get M6P residues on its N-glycan, it never makes it the lysosomes and one of many lysosomal storage diseases may develop as a result of undegraded molecules accumulating in the lysosomes.

When the lysosomal GH called β-glucoceramidase doesn't make it to the lysosomes, β-glucosylceramides accumulate and Gaucher's Disease results. One treatment for Gaucher's Disease is enzyme replacement therapy (ERT) and this is done and it works in many cases. But, it is extremely expensive. An alternative is substrate reduction therapy (SRT) in which the GT that catalyzes the biosynthesis of β-glucosylceramides is inhibited so that there is less of it made and less of it going to the lysosomes for degradation. For over a decade, SRT has been done, with some success, using miglustat (Zavesca) to inhibit the reaction catalyzed by Glucosylceramide


GCS
(Glucosylceramide Synthase,
aka, Ceramide Glucosyltransferase)
a ceramide
a β-D-glucosylceramide



***N*-butyl-deoxynojirimycin**
(NB-DNJ, or NB-dNM
miglustat (Zavesca))

Figure 27: The reaction catalyzed by Glucosylceramide Synthase (GCS), and the inhibitor miglustat.

Synthase (GCS), shown in Figure 27, for a typical ceramide. Miglustat is an analogue of glucose, and it also has a short hydrophobic chain similar to the acceptor-substrate of GCS. Miglustat is a rare example of a GT inhibitor that has made it into the clinic. The authors of a recent review of inhibitors of GHs and GTs wrote that "… progress in the area of glycosyltransferase inhibitors has been slower than for glycoside hydrolases, and GCS remains the only well-validated therapeutic target among mammalian glycosyltransferases. No inhibitors have been rigorously tested and found to lack class promiscuity" [Gloster and Vocadlo, 2012].

This system of using an N-glycan with M6P residues as the signal to traffick a glycoprotein to a certain destination was the first accepted example of a biological role for the glycans on glycoproteins, and one that explained the cause of a disease and enabled efforts to find a cure or treatment for that disease [Varki, et al., 2009, Ch. 30]. Since then, many other roles for glycans have been elucidated, and there are many links between abnormal glycans and human diseases. Because there are many glycans on a cell's surface, they are known to be involved in how cells respond to their environment and migrate in it. In fact, because many integral membrane proteins are glycosylated many cells are covered with a glycocalyx, like the canopy of tree tops in a tropical rainforest. Changes in the structures of normal cell surface glycans have been found in cancer cells [Dall'Olio and Chiricolo, 2001, Peracaula, et al., 2005, Jankovic, 2011, Adamczyk, et al., 2012, Perez-Garay, et al., 2013].

4.1.3 Understanding How Glycosyltransferases Work

The majority of characterized glycosyltlransferases (GTs) require a donor-substrate that is a nucleotide diphosphate sugar (NDP-sugar), such as the UDP-Glucose shown in Figure 28, and are called Leloir GTs. Most of the Leloir GTs with solved 3D

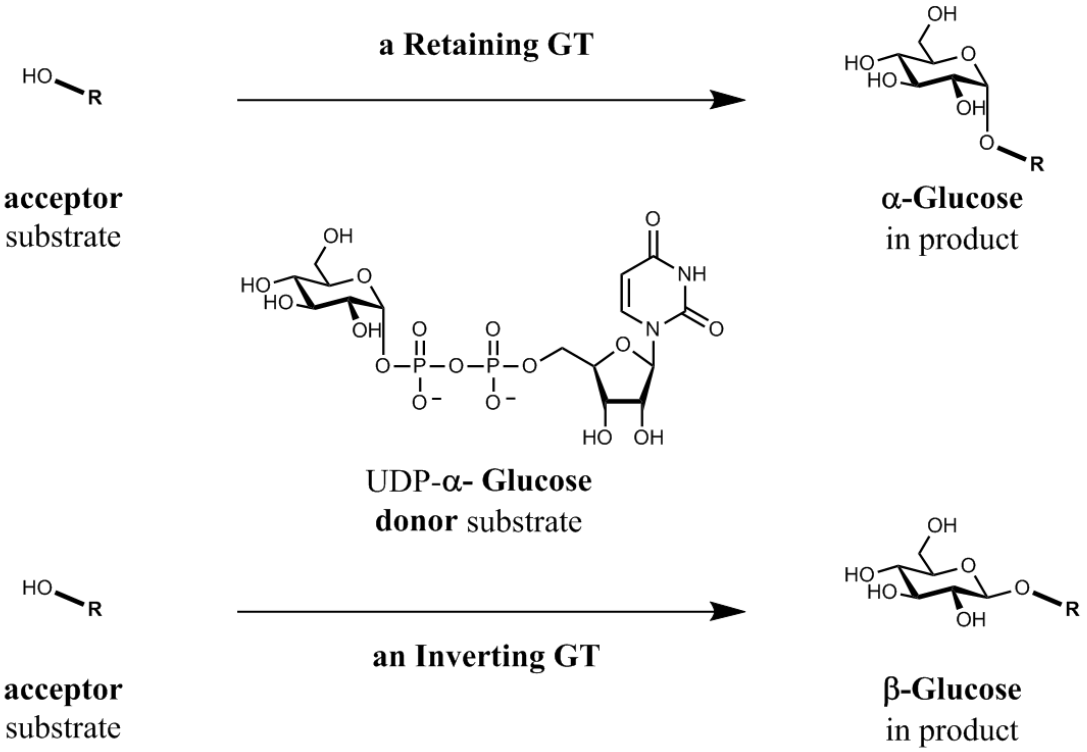

a Retaining GT
acceptor
substrate
α-Glucose
in product
UDP-α- Glucose
donor substrate
an Inverting GT
acceptor
substrate
β-Glucose
in product

Figure 28: A typical nucleotide sugar donor substrate, and the products of sugar transfer catalyzed by either a retaining GT or an inverting GT.

structures have been found to have one of two global folds, called GT-A and GT-B, with some recent evidence for a possible third fold called GT-C. The Leloir GTs all have at least one nucleotide-binding domain of the Rossmann type. The GT-A and GT-B folds both contain a pair of β/α/β Rossmann domains. In the GT-A fold, the two Rossman domains are closely abutting, and therefore the GT-A fold has been described as a single domain fold. In the GT-B fold the pair of Rossmann domains face each other and are connected by a flexible linker, giving the molecule a clear two-domain appearance [Lairson, et al., 2008].

Because of the binary nature of the configuration (either α or β) of the anomeric carbon in all glycosidic bonds, all GTs can be categorized into one of just two categories, retaining or inverting, depending on whether the anomeric carbon configuration in the sugar donor is retained or inverted in the product, as shown in Figure 28. The catalytic mechanism of inverting GTs is generally accepted to be an S_N2-like reaction that begins with an enzyme aspartic acid (Asp) or glutlamic acid (Glu) side-chain carboxylate acting as a general base and deprotonating the hydroxyl group of the acceptor substrate which then attacks the anomeric carbon of the sugar-donor, displacing the phosphate of the NDP-sugar donor-substrate, as shown in Figure 29. Another characteristic of GT-A fold GTs is that they require a divalent metal cation (M^{2+}) for activity. GT-A fold GTs have a conserved Asp-Xxx-Asp (DXD) motif, and GT-B fold GTs do not. This DXD motif coordinates the required divalent metal cation which then stabilizes one or or both of the negatively charged phosphate groups of the NDP-sugar donor-substrate.

The catalytic mechanism of retaining GTs remains controversial, but may be a double-displacement reaction in which an enzyme-nucleophile, an Asp or Glu side-chain, displaces the phosphate of the NDP-sugar forming a glycosyl-enzyme covalent intermediate, which is then attacked from the other face of the sugar by the acceptor's deprotonated hydroxyl oxygen [Lairson, et al., 2008, Breton, et al., 2012], as shown in Figure 29. One important aspect of the mechanism of GTs is that during their catalytic

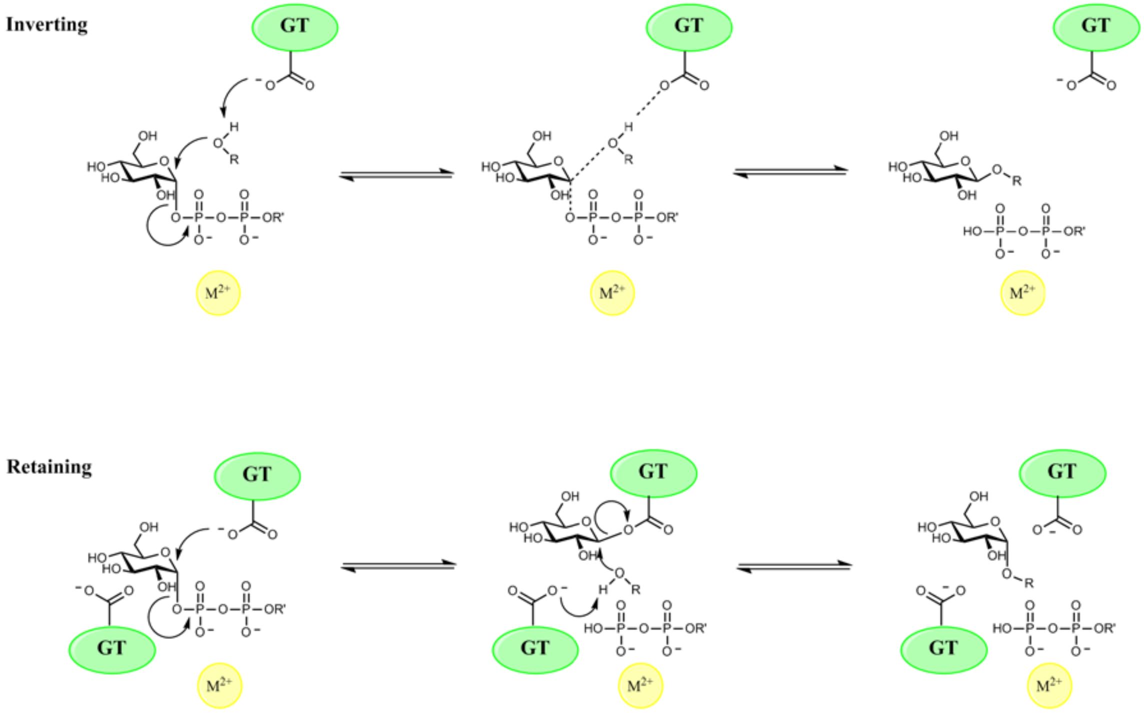


Figure 29: Possible reaction mechanisms for inverting and retaining GTs. The M^{2+} metal ion shown is required by all GT-A fold enzymes, but not GT-B fold GTs.

cycle they undergo significant conformational changes, and that, in general, the donor-substrate must bind before the acceptor-substrate does. Details of the conformational changes that GTs go through will be described in the next section of this chapter where the important contributions of studies with GT inhibitors, with or without high resolution 3D structures, will be seen in building a detailed understanding of how GTs work.

4.2 Inhibitors of Glycosyltransferases

4.2.1 Tunicamycin

Tunicamycin (**1**, Figure 30) is a natural product that was discovered in the 1970s. It is now known to be produced by several species of *Streptomyces*, and always as a mixture of at least ten homologues that differ in the structure of their fatty acid chains, having 14-17 carbons. Tunicamycin is now available as a mixture of the major homologues, with the amount of each homologue varying from lot to lot. As a result of this variability, kinetic measurements for inhibition by tunicamycin are rarely published.

Tunicamycin is a nucleoside antibiotic that interferes with bacterial wall biosynthesis. It got its name from the Latin word for coat, tunica, and contains a rare disaccharide called tunicamine, and the nucleoside uridine. Tunicamycin inhibits the bacterial transferase MraY (aka, translocase) that catalyzes the reaction shown in Figure 30, which is an early step in the biosynthesis of bacterial peptidoglycan as shown in Figure 31. A commercially available mixture of four homologues of tunicamycin (A, 3%; B, 36%; C, 38%; and D, 20%) was found to be a reversible inhibitor of *E. coli* MraY, with $IC_{50} = 2\ \mu M$ and $K_i = 0.55\ \mu M$, competitive with respect to the donor substrate and noncompetitive with respect to the acceptor [Brandish, et al., 1996]. Comparing the structures of the natural substrates of MraY (Fig. 30) to that of tunicamycin (**1**, Fig. 30)

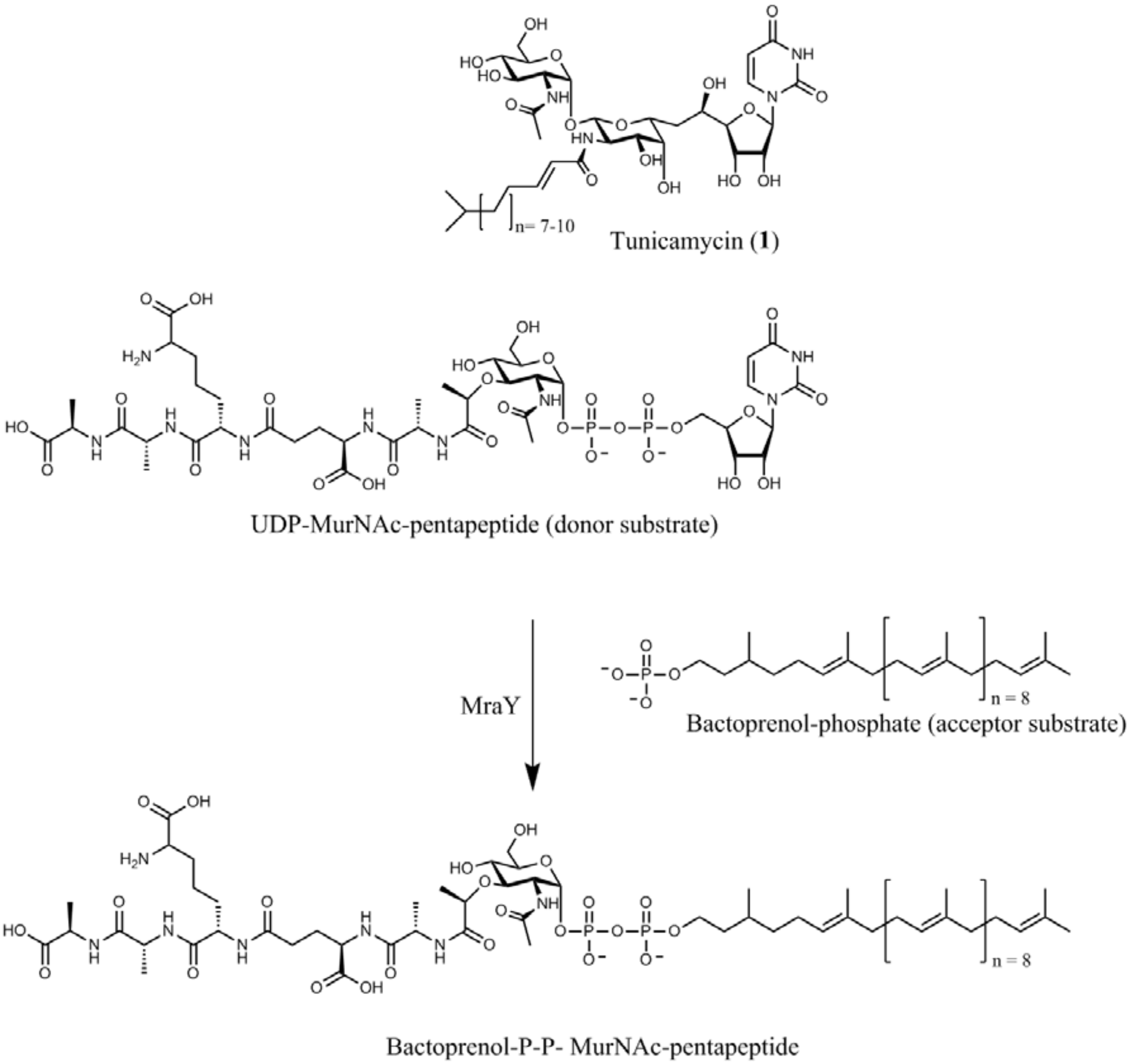

Tunicamycin (**1**)
n= 7-10
UDP-MurNAc-pentapeptide (donor substrate)
MraY
Bactoprenol-phosphate (acceptor substrate)
n = 8
Bactoprenol-P-P- MurNAc-pentapeptide

Figure 30: The reaction catalyzed by the bacterial transferase MraY, and the inhibitor tunicamycin.

it's likely that tunicamycin binds to MraY in its donor substrate binding pocket, mimicking the UDP-MurNAc portion of the natural donor substrate.

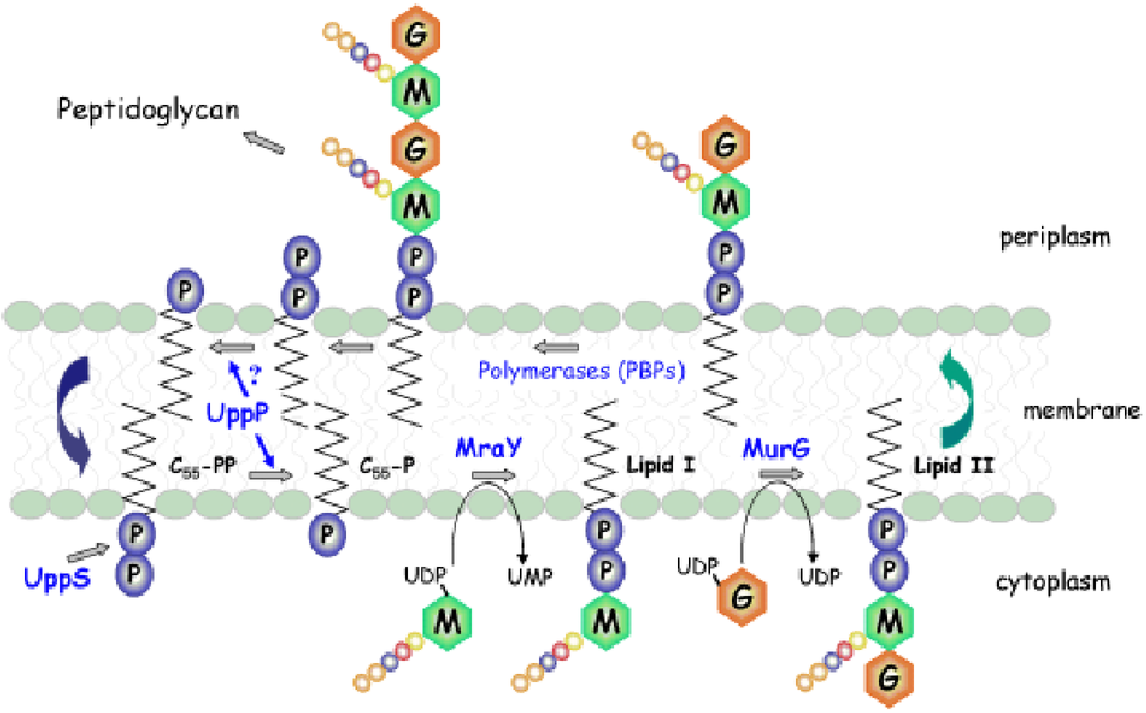

Peptidoglycan
G
M
periplasm
membrane
cytoplasm
P
UppP
?
Polymerases (PBPs)
C_{55}-PP
C_{55}-P
MraY
MurG
Lipid I
Lipid II
UppS
UDP
UMP
UDP
UDP

Figure 31: The reaction catalyzed by the bacterial transferase MraY, and inhibited by tunicamycin, in context of the membrane steps of peptidoglycan biosynthesis. [Reprinted from Bouhss, et al., 2008]

Tunicamycin is not a clinically useful antibiotic because it's toxic to eukaryotes due to the similarities between the biosynthesis of the precursors of bacterial cell walls and the first steps of eukaryotic N-linked glycan biosynthesis. However, it was one of the first compounds found that could block the biosynthesis of N-linked glycans and thus it has been a valuable research tool for many decades. Tunicamycin prevents N-glycosylation by blocking the first step in the biosynthesis of the N-glycan precursor, Dolichol-P-P-GlcNAc$_2$ Man$_9$ Glc$_3$, catalyzed by the GlcNAc-1-Phosphotransferase, ALG7, as shown in the biosynthetic pathway in Figure 32. Tunicamycin does this by mimicking the shape of the sugar donor Uridine Diphosphate *N*-Acetyl-D-Glucosamine (UDP-GlcNAc) for the eukaryotic enzyme GlcNAc-1-Phosphotransferase (Figure 33). Tunicamycin's structural resemblance to UDP-sugars makes it a sugar-donor mimic for many glycosyltransferases (GTs) that use UDP-sugars, which is called class promiscuity and is a lack of specificity that many such carbohydrate-based inhibitors suffer from. Finding inhibitors that are not promiscuous is usually a goal when designing enzyme inhibitors. Also, the two enzymes inhibited by tunicamycin that have been covered here are not, strictly speaking, GTs because they also transfer one phosphate group along with the glycosyl group. The next examples covered will be inhibitors of true GTs.

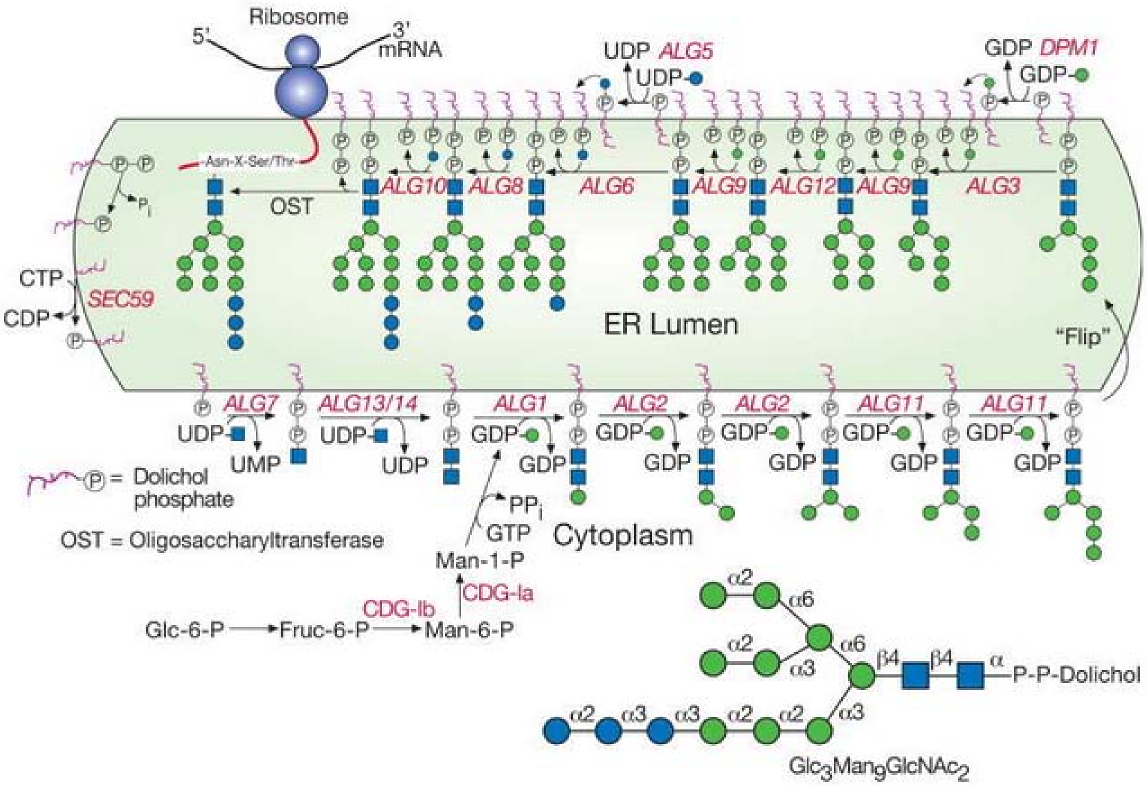

Ribosome
5'
3'
mRNA
UDP ALG5
UDP
GDP DPM1
GDP
Asn-X-Ser/Thr
OST
ALG10
ALG8
ALG6
ALG9
ALG12
ALG9
ALG3
P
P_i
CTP
SEC59
CDP
ER Lumen
"Flip"
ALG7
UDP
UMP
ALG13/14
UDP
UDP
ALG1
GDP
GDP
ALG2
GDP
GDP
ALG2
GDP
GDP
ALG11
GDP
GDP
ALG11
GDP
GDP
= Dolichol phosphate
OST = Oligosaccharyltransferase
PP_i
GTP
Man-1-P
Cytoplasm
CDG-Ia
CDG-Ib
Glc-6-P
Fruc-6-P
Man-6-P
α2
α6
α6
α2
α3
β4
β4
α
P-P-Dolichol
α2
α3
α3
α2
α2
α3
$Glc_3Man_9GlcNAc_2$

Figure 32: The reaction catalyzed by the eukaryotic GlcNAc-1-Phosphotransferase, ALG7, is at the beginning of the membrane-associated biosynthesis of the dolichol-linked 14-sugar precursor of N-linked glycans. [Reprinted from Varki, et al., 2009, Ch. 8]

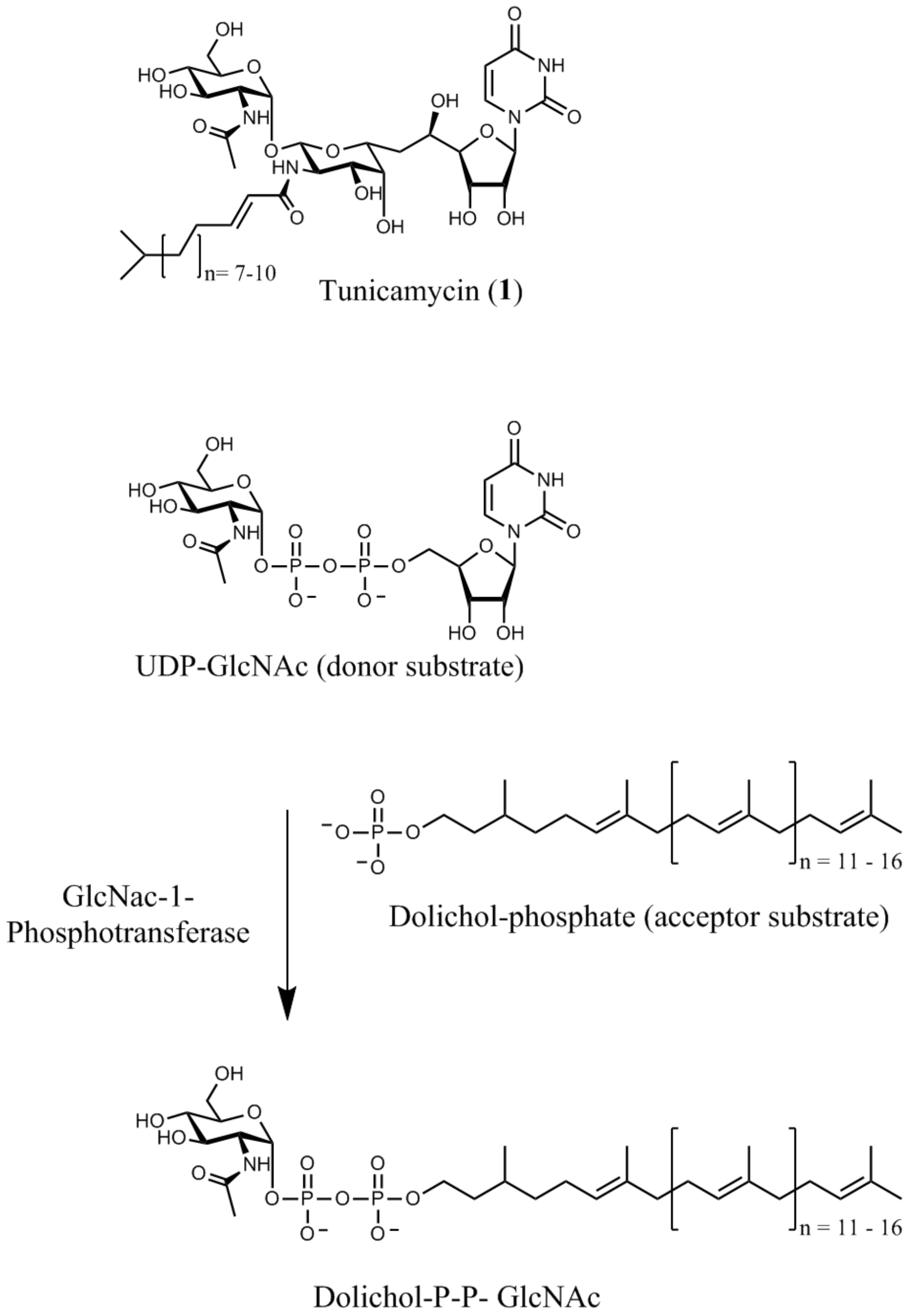

n= 7-10
Tunicamycin (1)
UDP-GlcNAc (donor substrate)
GlcNac-1-
Phosphotransferase
n = 11 - 16
Dolichol-phosphate (acceptor substrate)
n = 11 - 16
Dolichol-P-P- GlcNAc

Figure 33: The reaction catalyzed by GlcNAc-1-Phosphotransferase, and the inhibitor tunicamycin.

4.2.2 Inhibitors of β1-4 Galactosyltransferase (β1-4 Gal T)

Eukaryotic N-linked glycans begin with a common core oligosaccharide, of 14 monosaccharides, transferred to the side-chain Nitrogen of an asparagine: $GlcNAc_2$ Man_9 Glc_3 . As a glycoprotein moves along its biosynthetic path all of its N-linked oligosaccharide cores undergo a series of remodeling steps: removal of some monosaccharides, catalyzed by glycoside hydrolases (GHs), and addition of new monosaccharides, catalyzed by glycosyltransferases (GTs). The large diversity of possible glycan structures is due to the large number of available monosaccharides, the variety of possible glycosidic linkages, and the occurrence of branching as shown in Figure 34.

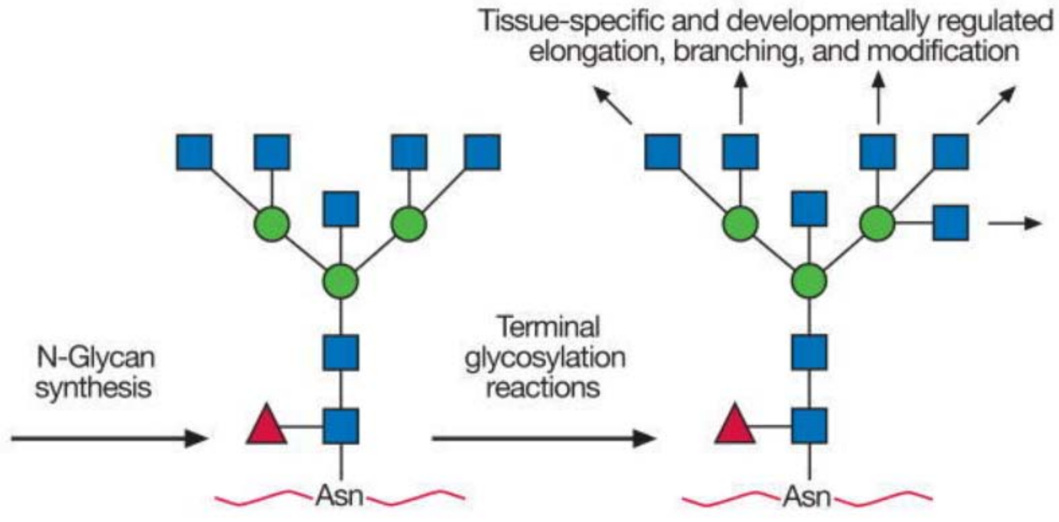
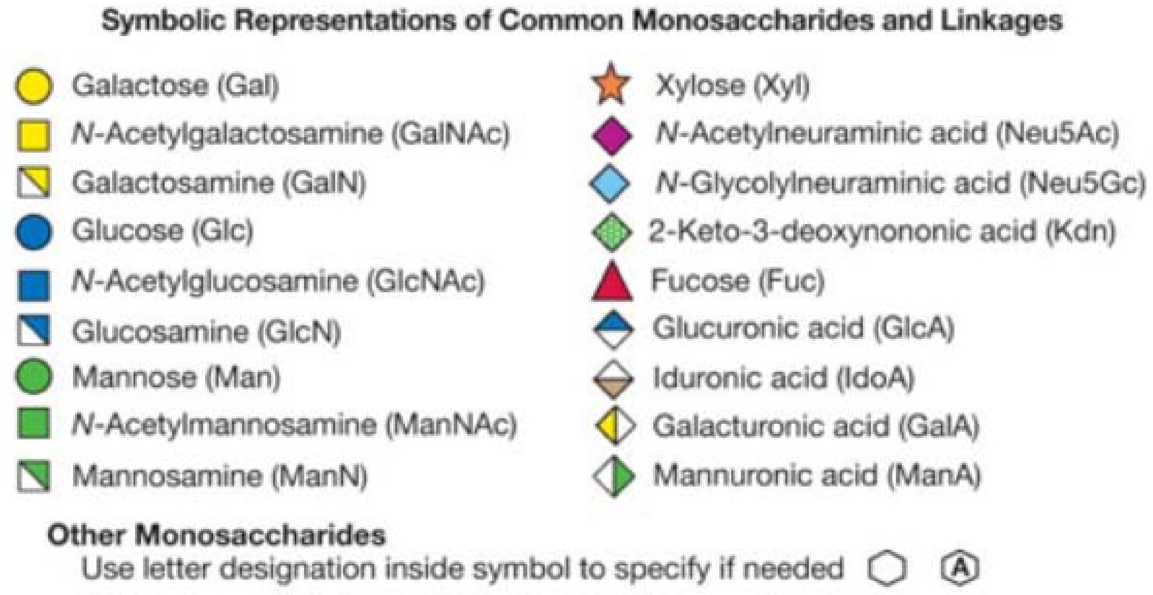

Tissue-specific and developmentally regulated elongation, branching, and modification
N-Glycan synthesis
Terminal glycosylation reactions
Asn
Asn

Figure 34: The biosysnthetic maturation of N-glycans produces diversity in the branching structure and in the composition of the terminal sugar residues, depending on the cellular environment of these N-linked oligosaccharides. Symbol key on next page. [Reprinted from Varki, et al., 2009, Ch. 13]


Symbolic Representations of Common Monosaccharides and Linkages
Galactose (Gal)
N-Acetylgalactosamine (GalNAc)
Galactosamine (GalN)
Glucose (Glc)
N-Acetylglucosamine (GlcNAc)
Glucosamine (GlcN)
Mannose (Man)
N-Acetylmannosamine (ManNAc)
Mannosamine (ManN)
Xylose (Xyl)
N-Acetylneuraminic acid (Neu5Ac)
N-Glycolylneuraminic acid (Neu5Gc)
2-Keto-3-deoxynononic acid (Kdn)
Fucose (Fuc)
Glucuronic acid (GlcA)
Iduronic acid (IdoA)
Galacturonic acid (GalA)
Mannuronic acid (ManA)
Other Monosaccharides
Use letter designation inside symbol to specify if needed
A

Symbol Key: Recommended symbols and conventions for drawing glycans. [Reprinted from Varki, et al., 2009, Ch. 13]

It is often the terminal residues on the various branches of a glycan that carry the critical biochemical information, which is decoded by receptor molecules that bind specifically to them. Terminal *N*-acetylglucosamine (GlcNAc) residues on all types of glycans (N-linked, O-linked, and glycolipids) are frequently modified by the addition of a galactose (Gal) residue in either a β1-3 or a β1-4 linkage, depending only on which GT catalyzes the addition. Both β1-3 glactosyltransferase (β1-3 Gal T) and β1-4 glactosyltransferase (β1-4 Gal T) use the same sugar donor substrate, UDP-Gal. When Gal is added in a β1-4 linkage, a Galβ1-4GlcNAc disaccharide is produced, called a type-2 unit, or *N*-acetyllactosamine (LacNAc, or LN), as shown in Figure 35. These

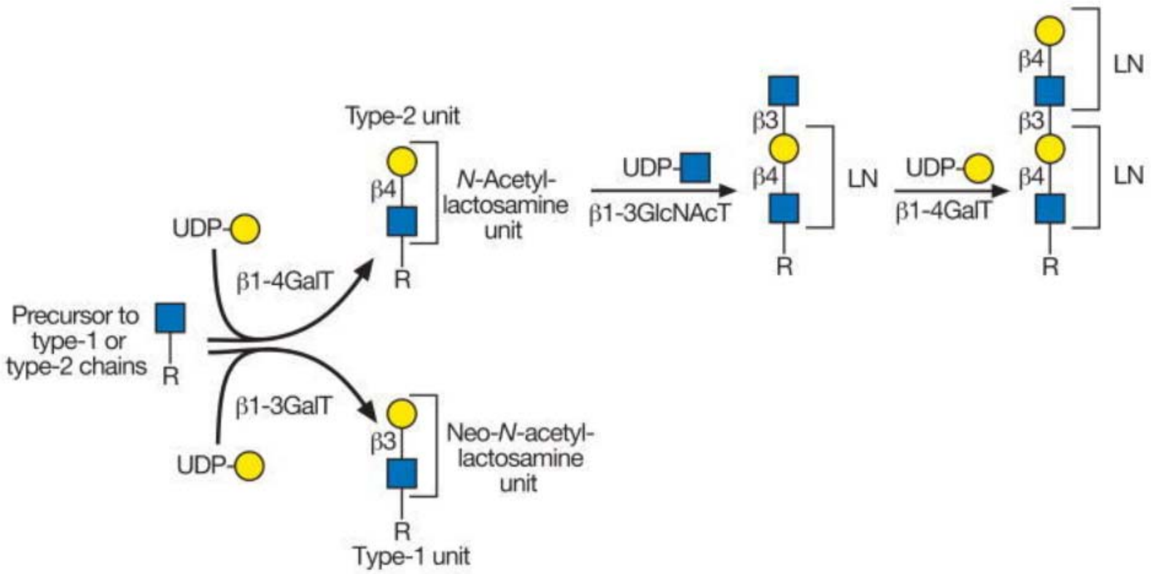

Type-2 unit
UDP-
β4
N-Acetyl-lactosamine unit
β1-4GalT
R
Precursor to type-1 or type-2 chains
R
β1-3GalT
UDP-
β3
Neo-N-acetyl-lactosamine unit
R
Type-1 unit
UDP-
β1-3GlcNAcT
β3
β4
LN
R
UDP-
β1-4GalT
β4
LN
β3
β4
LN
R

Figure 35: The modification of a terminal GlcNAc residue by a β1-4 glactosyltransferase (β1-4 Gal T) yields a Type-2 unit (LacNAc, or LN), while modification by a β1-3 glactosyltransferase (β1-3 Gal T) yields a Type-1 unit. **R** stands for an N-glycan, O-glycan, or glycolipid. [Reprinted from Varki, et al., 2009, Ch. 13]

LacNAc disaccharides are found on glycans in all mammalian tissues, either as single terminal disaccharides or in poly-*N*-acetyllactosamine (Galβ1-4GlcNAc)$_n$ chains.

The galactosyltransferases (Gal Ts) have been called "the most important and ubiquitous superfamily of sugar elongation enzymes in eukaryotes" [Takaya, et al., 2005]. There are five known subfamilies, β1-4, β1-3, α1-3, α1-4, and α1-6, all of which use the same sugar donor substrate, UDP-α-galactose (UDP-Gal) to transfer galactose to a sugar acceptor, which depends on the enzyme, at the O-4, O-3 or O-6 position in the β- or α-anomeric configuration [Ramakrishnan, et al., 2004]. Fifty years of biochemical studies on β1-4 Gal T have revealed many interesting details about its catalytic mechanism. The β1-4 Gal T enzyme changes conformation during its catalytic cycle, only creating its sugar acceptor substrate binding site after a divalent metal ion and the sugar donor substrate, UDP-Gal, have bound first [Ramakrishnan, et al., 2004].

Since the β1-4 Gal T acceptor substrate binding site doesn't exist until after the donor substrate has bound, most efforts to design competitive inhibitors have focused on donor analogues that will bind tightly to the donor substrate binding site. The chemical structures of the natural donor and acceptor substrates of β1-4 Gal T are shown in Figure 36.

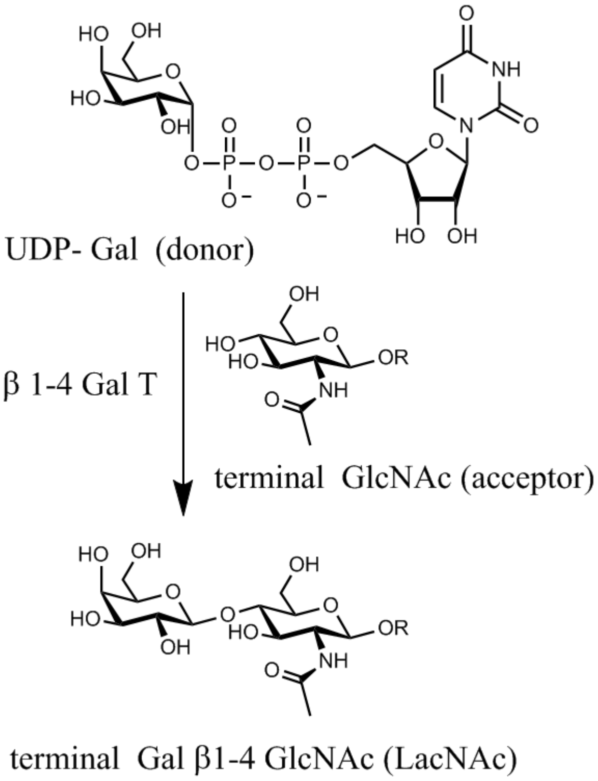

UDP- Gal (donor)
β 1-4 Gal T
terminal GlcNAc (acceptor)
terminal Gal β1-4 GlcNAc (LacNAc)

Figure 36: The reaction catalyzed by β 1-4 Galactosyltransferase (β1-4 Gal T).

4.2.2.1 Acceptor-based Inhibitors of β1-4 Galactosyltransferase (β1-4 Gal T)

When compound **2** (Figure 37) was co-crystalized with human M340H β1-4 Gal T, with Mn^{2+} and UDP-hexanolamine to induce the acceptor-binding (closed) conformation of the enzyme, the GlcNAc residue of **2** was found tightly bound in the acceptor binding site and the β-linked Gal residue was hydrophobically packed against the side chain of Tyr282 [Brown, et al., 2009]. When compounds **4** or **5** were individually added to the growth medium for U937, human lymphoma-derived, cells surface expression of Sialyl Lewis^X (SLe^X) (Figure 38) was diminished [Brown, et al., 2009].

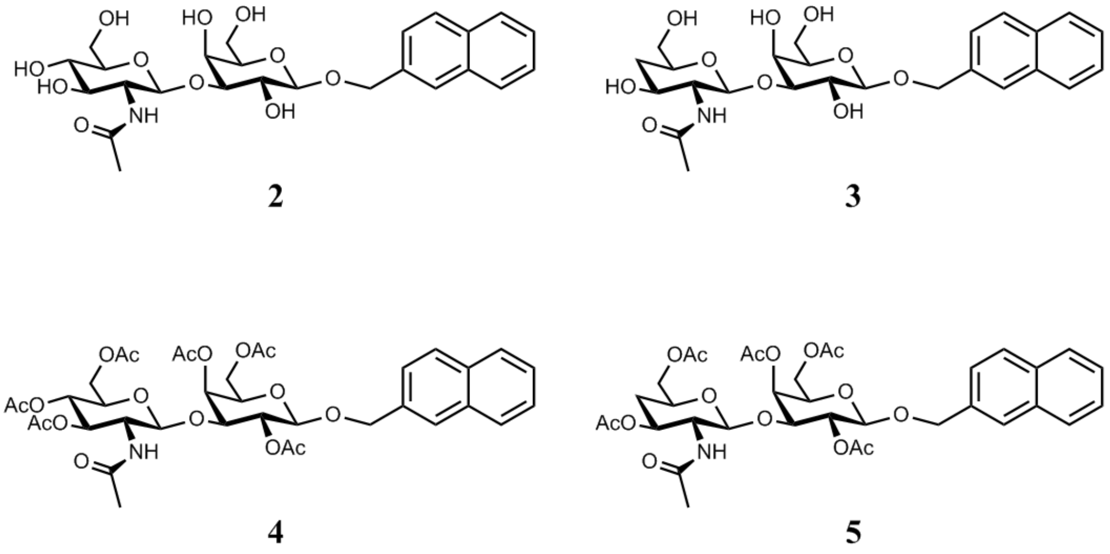

OH
HO
OH
HO
O
HO
NH
O
O
OH
O
2
3
OAc
AcO
OAc
AcO
NH
OAc
4
5

Figure 37: Acceptor-based inhibitors of β1-4 Galactosyltransferase (β1-4 Gal T). [Brown, et al., 2009]

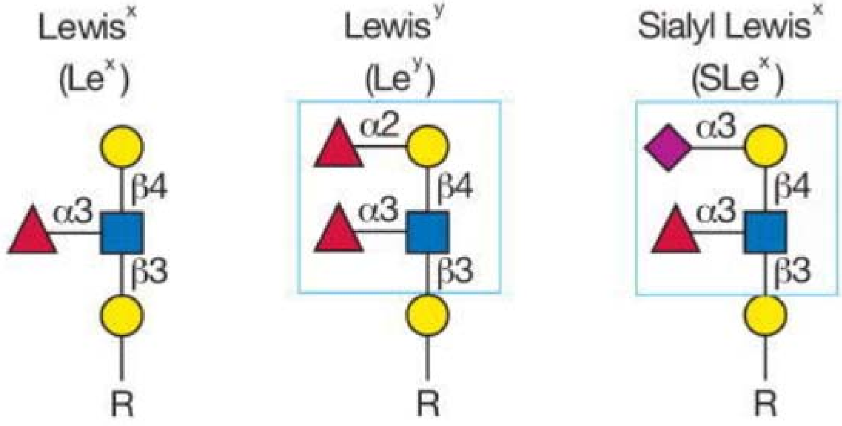
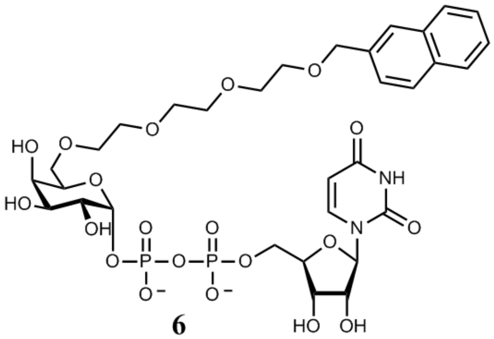

Lewis^x
(Le^x)
Lewis^y
(Le^y)
Sialyl Lewis^x
(SLe^x)
α2
α3
β4
β3
R

Figure 38: Some Lewis blood group antigens. [Reprinted from Varki, et al., 2009, Ch. 13]

However, the putative mechanisms of how compounds **4** and **5** reduced the formation of SLe^X antigens differed because compound **4** became sialylated and sulfated in the U937 cells while compound **5** did not. The authors concluded that compound **4** is acting as a decoy acceptor substrate, causing unproductive glycosylation, possibly slowing down β1-4 Gal T, which must act to produce the precursor to the SLe^X antigen, and compound **5** is acting as a true inhibitor, possibly of β1-4 Gal T, since **5** cannot be galactosylated by β1-4 Gal T. Finally, in *in vitro* studies with compound **3** and bovine β1-4 Gal T, with compound **2** as the acceptor substrate, compound **3** was found to be a competitive inhibitor with K_i = 192 μM [Brown, et al., 2009].

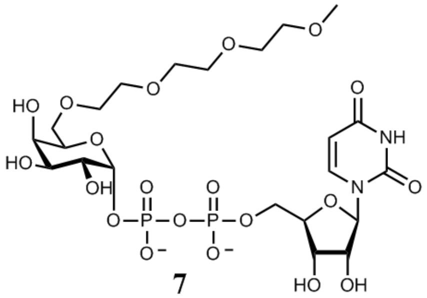
4.2.2.2 Donor-based Inhibitors of β1-4 Galactosyltransferase (β1-4 Gal T)

There are at least three general ways to modify the natural sugar donor substrate, UDP-Gal, for β1-4 Gal T. Those are by modifying the sugar, by modifying the nucleoside, or by modifying the diphosphate that links together those other two components (Figure 36). The first class-specific inhibitor for human β1-4 Gal T [Hosoguchi, et al., 2010] was a sugar-modified analogue of UDG-Gal, compound **6**, with K_i = 1.86 μM [Takaya, et al., 2005]. Figure 39 shows compound **6** and three closely related UDP-Gal analogues that were tested under similar conditions and reported in 2005. The K_i values in Figure 39 are against UDP-Gal (K_m = 4.91 μM), using recombinant



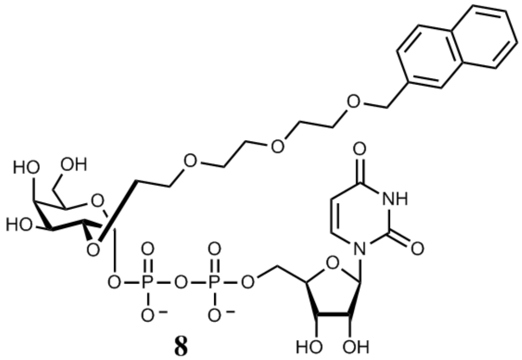
6

$K_i = 1.86\ \mu M$



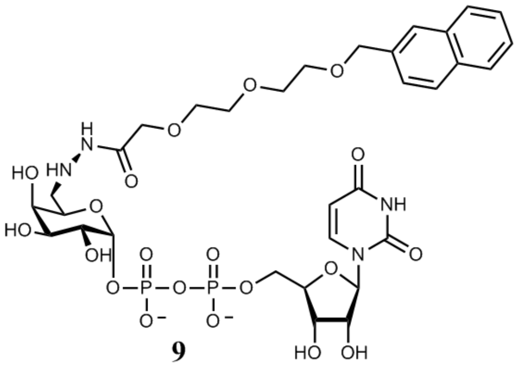
7

$K_i = 149\ \mu M$



8

$K_i = 11.7\ \mu M$



9

$K_i = 20.6\ \mu M$

Figure 39: Early donor-based inhibitors of β1-4 Galactosyltransferase (β1-4 Gal T). [Takaya, et al., 2005]

human β1-4 Gal T, and compound **6** was reported to be a competitive inhibitor. The much weaker binding of compound **7** (K_i = 149 μM) illustrates the importance of the naphthalene group for binding to β1-4 Gal T. The inhibitors in Figure 39 were designed after affinity labeling experiments with a variant of compound **6**, with a bromomethyl on its naphthalene group, that covalently bound to Trp310, in the short loop, of human β1-4 Gal T [Takaya, et al., 2005].

Compound **10** (Figure 40) is a rare example of a nucleoside-modified analogue of UDP-Gal, the natural sugar donor for β1-4 Gal T, and all Leloir-type galactosyltransferases (GalTs) [Pesnot, et al., 2010]. This inhibitor was designed to bind in the donor substrate binding pocket of any GalT and has a formylthienyl group on uridine's base to prevent binding of any acceptor substrate by preventing the conformational change that GalTs undergo after donor substrate binding, and which is required for acceptor substrate binding. Compound **10** was found to be a competitive inhibitor of the chimeric human blood group GalT AA(Gly)B, and then tested against four other GalTs (K_i values shown in Table 13). High resolution X-ray crystal structures of AA(Gly)B in various forms, apo, in a complex with Uridine Diphosphate (UDP), and in a complex with compound **10**, revealed details of how its formylthienyl group blocks the enzyme from making the conformational change to the fully closed conformation that is required for the acceptor substrate to bind. The authors state that compound **10** "locks the enzyme in an unproductive conformation and effectively inhibits glycosyl transfer" [Pesnot, et al., 2010].

A major problem for any inhibitors that are analogues of UDP-Gal is that the diphosphate group gives the molecule a 2- charge which makes it difficult for it to pass through a cell's phospholipid bilayer outer membrane. This cell membrane effectively prevents highly charged inhibitors from reaching enzymes inside the cell. To address

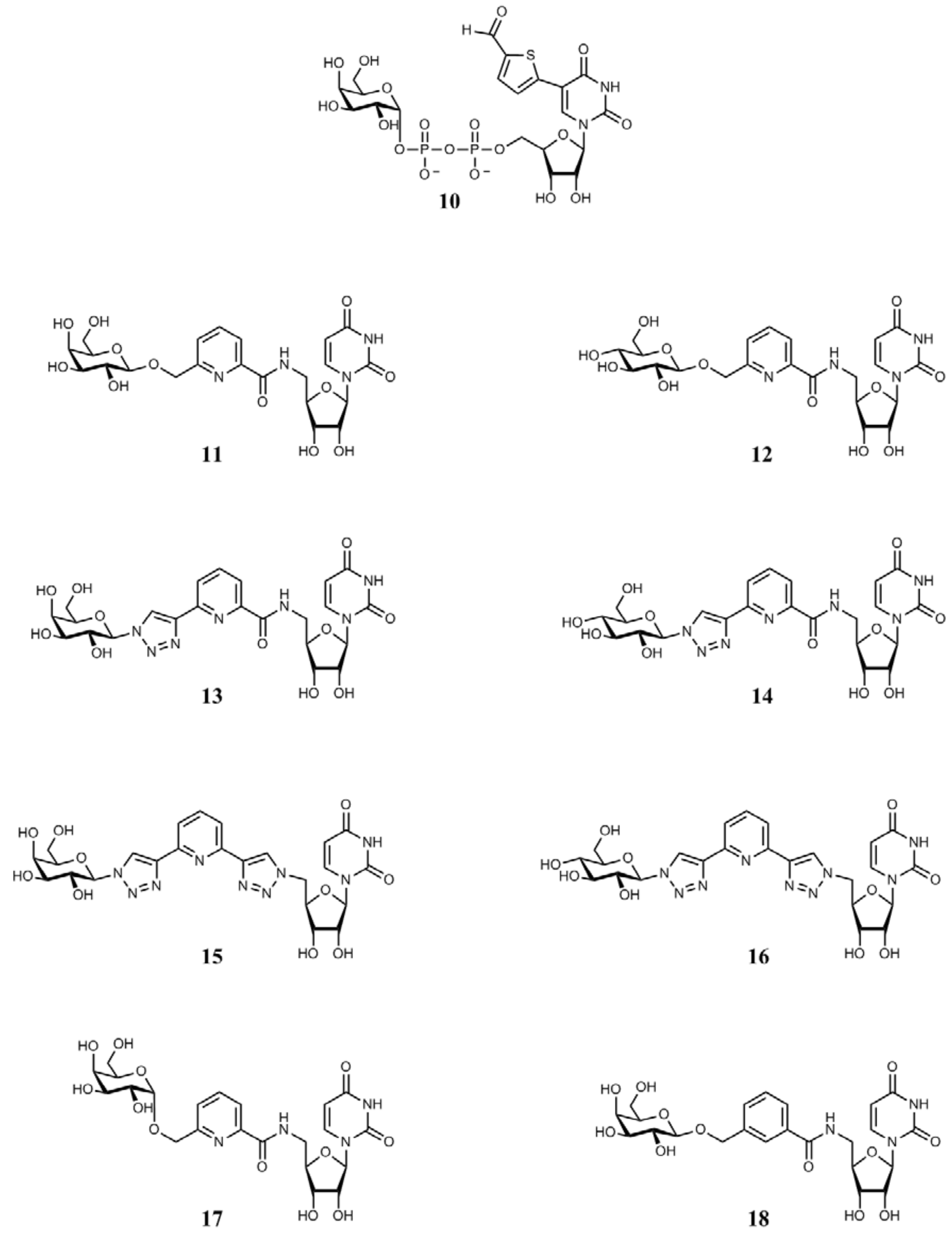

10
11
12
13
14
15
16
17
18

Figure 40: Recent donor-based inhibitors of (β1-4 Gal T). [from Pesnot, et al., 2010 and Wang, et al., 2013]

Table 13: Inhibition of GalTs, by compounds in Fig. 40, or Uridine Diphosphate (UDP)

Cmp #	**β 1-4 Gal T**		**α 1-4 Gal T**		**α 1-3 Gal T**		**GTB**		**AA(Gly)B**	
	Bos taurus		Neisseria meningitidis		Bos taurus		Homo sapiens		Homo sapiens	
10	$K_i = 38.8\ \mu M$		$K_i = 0.45\ \mu M$		$K_i = 9.8\ \mu M$		$K_i = 2.4\ \mu M$		$K_i = 0.53\ \mu M$	
	Bos taurus		Neisseria meningitidis		Bos taurus		Homo sapiens		Murine	
	Inhibit. % at 1mM	IC$_{50}$ (μM)	Inhibit. % at 1mM	IC$_{50}$ (μM)	Inhibit. % at 1mM	IC$_{50}$ (μM)	Inhibit. % at 1mM	IC$_{50}$ (μM)	Inhibit. % at 1mM	IC$_{50}$ (μM)
11	100	152	80	546	99	320	84	262	93	493
12	13	- -	23	- -	19	- -	11	- -	47	1,220
13	34	- -	2	- -	11	- -	10	- -	22	- -
14	12	- -	0	- -	11	- -	1	- -	14	- -
15	15	- -	11	- -	22	- -	34	- -	19	- -
16	100	634	33	1,962	30	1,061	36	1,767	36	1,417
17	97	334	37	1,597	79	602	53	725	80	584
18	78	573	27	- -	39	1,020	23	- -	19	- -
UDP	- -	25	- -	62	- -	53	- -	5	- -	1

Compound **10**: ref. = [Pesnot, et al., 2010], Compounds **11-18,** UDP: ref. = [Wang, et al., 2013] this problem, several analogues of NDP-sugars have been made with neutral replacements for the diphosphate group, that should still interact well with the divalent metal cation that binds to the enzyme before the NDP donor substrate binds, as required by the GTs having the GT-A fold. Compounds **11** – **18**, in Figure 40, are neutral donor analogues with various diphosphate-group replacements which, at 1mM, gave a variety of inhibition patterns (Table 13) against a panel of five well-studied GalTs [Wang, et al., 2013]. In this panel of five recombinant GalTs (Table 13) only the bovine β1-4 Gal T is an inverting enzyme and the other four are retaining GTs.

The inhibition patterns for these diphosphate-modified UDP-Gal analogues **11** – **18** (Figure 40), reveal which diphosphate-replacements work better than others. Surprisingly, analogues that have the same sugar, galactose, as the natural substrate, UDP-Gal, didn't always make the best inhibitors. Compound **16** was a more potent inhibitor of all five GalTs relative to its galactose-containing counterpart (**15**), being a very potent inhibitor of the inverting enzyme β1-4 Gal T, but only a weak inhibitor of all four of the retaining GalTs. The most potent inhibitor of all five of these GalTs was compound **11**, which, as with all but one of these compounds, has its sugar in the β-anomeric configuration which is the opposite anomeric configuration of the natural donor substrate, UDP-Gal. Detailed kinetics analysis revealed that both compounds **11**, and **17**, were mixed inhibitors of β1-4 Gal T.

To make sense of these inhibition patterns, crystals of "the well characterized AA(Gly)B enzyme" [Wang, et al., 2013] were soaked with each of three inhibitors, **11**, **12**, and **17**, and their structures were solved and refined. All three inhibitors were found to bind to the active site of AA(Gly)B in nearly the same mode, though the entire molecules were not well ordered. The uracil and ribose groups, of **11**, **12**, and **17**, were visible and bound to the enzyme in nearly the identical positions that those groups occupy in the 3D crystal structure (PDB ID: 2RJ7) of AA(Gly)B with UDP-Gal in the donor-binding site. Not surprisingly, the positions of the pyridine ring were different from that of the diphosphate group in UDP-Gal, and this caused considerable disorder of the attached galactose or glucose. For compounds **11** and **12**, these sugars were not visible in the electron density. For compound **17**, two conformations of the anomeric oxygen of galactose were observed, and only one of these revealed any, though weak, electron density for galactose [Wang, et al., 2013]. The authors concluded, for these three inhibitors, that the diphosphate-replacements on each inhibitor chelated the enzyme-bound manganese ion slightly differently than diphosphate does and that the sugar portion of each inhibitor was in a much different position than galactose of UDP-Gal would be, causing the sugars to have only limited contact with the enzyme.

4.2.3 Inhibitors of Sialyltransferases

A major reason that galactosyltransferases (Gal Ts) are biologically important is that they biosynthesize terminal disaccharide units that are then modified by other GTs, creating important glycan-branch terminal structures. The two β galactosyltransferases, β1-3 GalT and β1-4 Gal T, add a galactose to a terminal *N*-acetylglucosamine (GlcNAc) to create the terminal disaccharides known as the Type-1 unit (neo-LacNAc) and the Type-2 unit (LacNAc), respectively, as shown in Figure 35. Type-1 and Type-2 units are then modified by fucosyltransferases and sialyltransferases, adding fucose and sialic acid monosaccharides, respectively, and sometimes sulfotransferases, to create the set of terminal structures known as the Lewis blood group antigens, shown in Figure 41.

In addition to the Lewis blood group antigens, the full range of biological roles for sialic acid on glycans has yet to be discovered. Some other known functions include "capping" the growth of glycan branches. If a GlcNAc is added to a terminal Gal residue, then a galactosyltransferase can add another Gal to that new terminal GlcNAc. After a sialyltransferase adds a sialic acid to a terminal Gal, that oligosaccharide chain cannot continue to be elongated, unless a sialidase cuts it off. Also, when a sialic acid is attached to a terminal Gal residue, it "hides" it from galactose-binding receptors, until a sialidase cuts off the sialic acid and exposes the Gal residue.

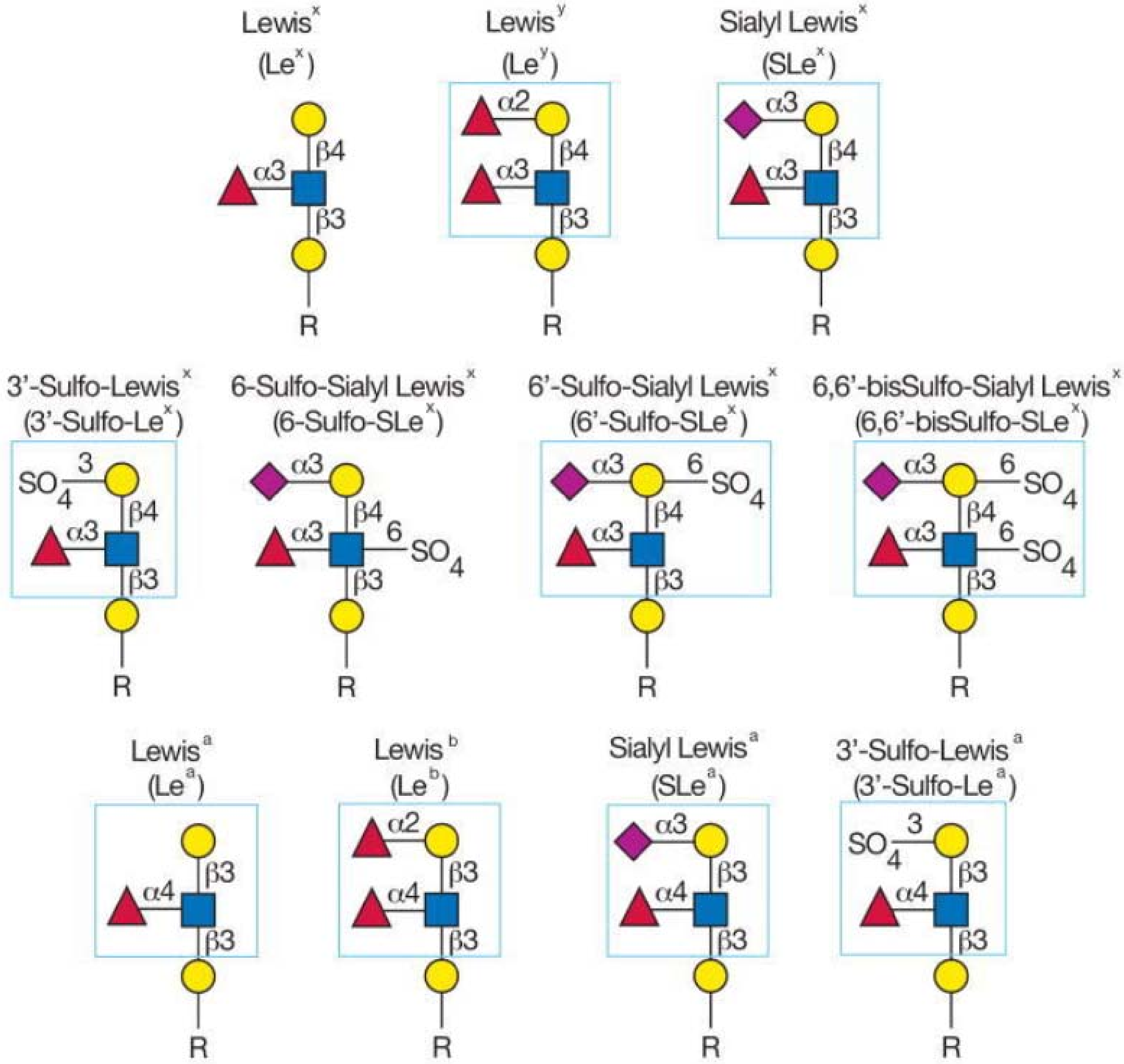

Lewis^x
(Le^x)
Lewis^y
(Le^y)
Sialyl Lewis^x
(SLe^x)
3'-Sulfo-Lewis^x
(3'-Sulfo-Le^x)
6-Sulfo-Sialyl Lewis^x
(6-Sulfo-SLe^x)
6'-Sulfo-Sialyl Lewis^x
(6'-Sulfo-SLe^x)
6,6'-bisSulfo-Sialyl Lewis^x
(6,6'-bisSulfo-SLe^x)
Lewis^a
(Le^a)
Lewis^b
(Le^b)
Sialyl Lewis^a
(SLe^a)
3'-Sulfo-Lewis^a
(3'-Sulfo-Le^a)
α2
α3
α4
β3
β4
3
6
SO_4
R

Figure 41: The Lewis blood group antigens (Type-1 and Type-2 Lewis determinants). Type-1 and Type-2 units differ only in the linkage of the outermost galactose. [Reprinted from Varki, et al., 2009, Ch. 13]

All vertebrate sialyltransferases (STs) are type II transmembrane glycoproteins that predominantly are found in the *trans*-Golgi compartment, have a short N-terminal tail, one transmembrane domain, and a variable length stem region of 20 to 200 amino acids linking the C-terminal catalytic domain to the luminal side of the Golgi membrane [Harduin-Lepers, et al., 2005]. Twenty ST genes have been found in the human genome and each of these 20 putative STs is believed to use the same activated sialic acid donor substrate, CMP-β-Neu5Ac, to add an α sialic acid to a glycan [Harduin-Lepers, et al., 2005]. These 20 STs differ by the acceptor substrate that they add sialic acid (Sia) to, and they have been categorized into four main families, containing a total of 20 subfamilies. The names of the four main families, ST3 Gal, ST6 Gal, ST6 GalNAc, and ST8 Sia, indicate the acceptor monosaccharide and the position where Sia is linked to that acceptor, by that family of STs. The ST6 Gal family has two subfamilies; ST6Gal-I and ST6Gal-II. Members of the ST6Gal-I subfamily use a terminal Galβ1-4GlcNAc (LacNAc) as their acceptor substrate, and attach the sialic acid Neu5Ac in an α linkage to the C-6 OH group, as shown in Figure 42.

For decades, the most potent inhibitor of an α2-6 sialyltransferase, specifically bovine ST6 Gal-I, was Cytidine Diphosphate (CDP), shown in Figure 43, a competitive inhibitor with a K_i of 13 μM published in 1977 [E.2.12]. Clearly, CDP is competing for

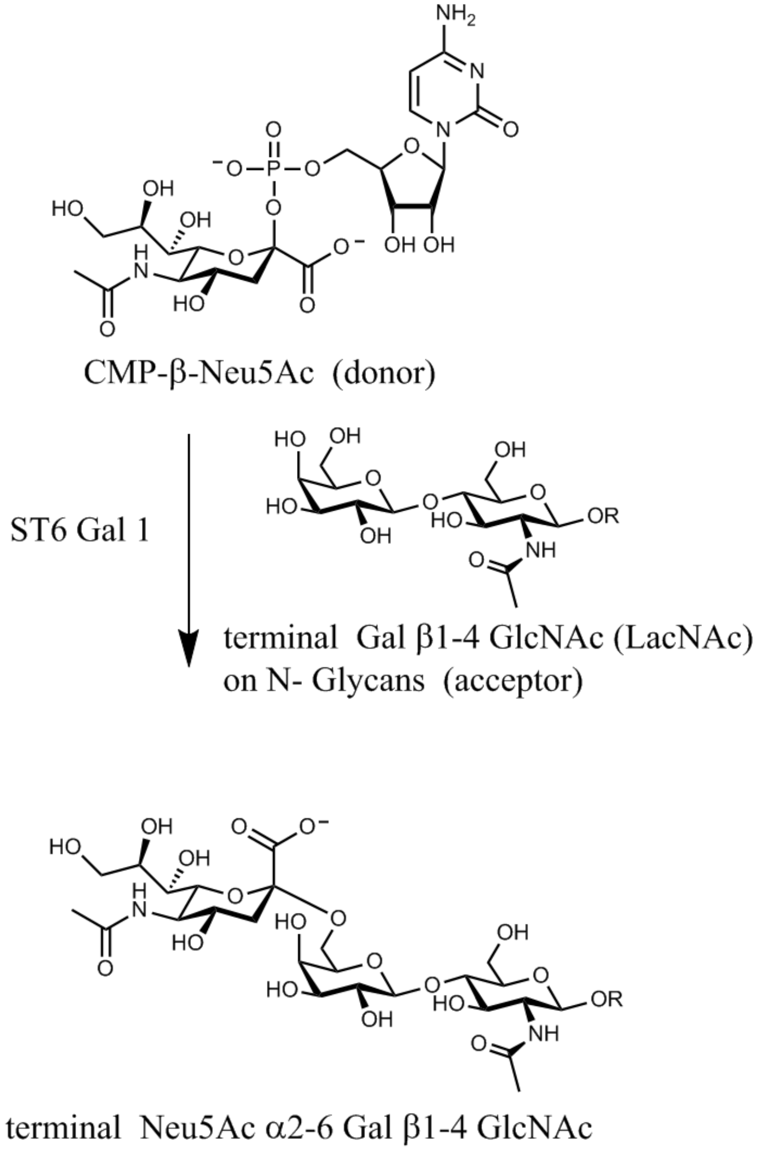
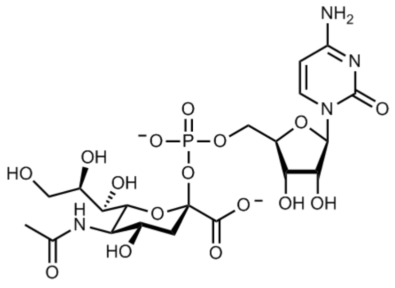
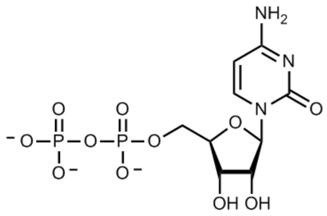

CMP-β-Neu5Ac (donor)
ST6 Gal 1
terminal Gal β1-4 GlcNAc (LacNAc) on N- Glycans (acceptor)
terminal Neu5Ac α2-6 Gal β1-4 GlcNAc

Figure 42: Reaction catalyzed by ST6 Gal-I

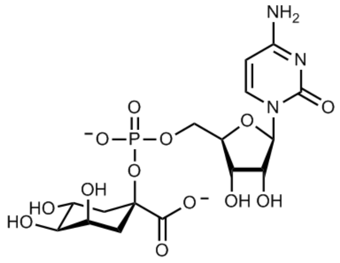
the donor-binding site by mimicking the natural donor substrate, CMP-Neu5Ac, (Figure 43) and binding more tightly to the donor-binding site than the



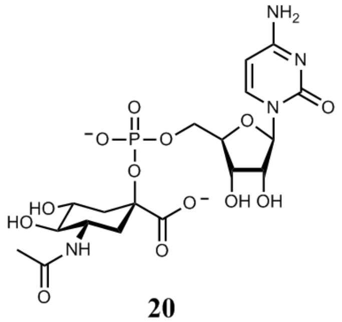
CMP-β-Neu5Ac (donor)



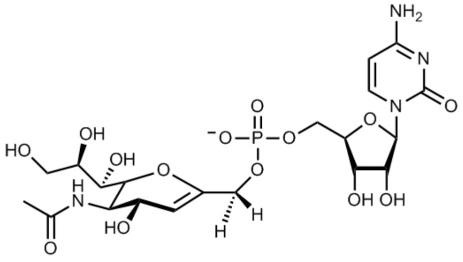
Cytidine Diphosphate (**CDP**)



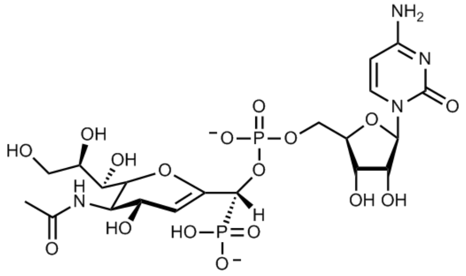
19



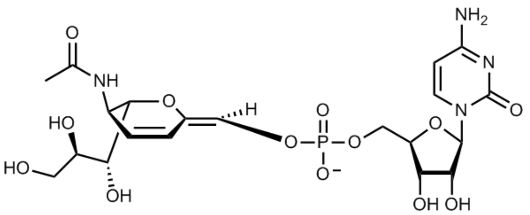
20



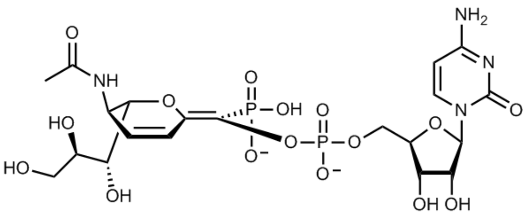
21



22



23



24

Figure 43:   Donor- and Transition State-based inhibitors of ST6 Gal I

Table 14: Kinetics of inhibition of ST6 Gal-I, for inhibitors shown in Figure 43

Compound	**ST6 Gal-I source**	**K_m (μM)**	**K_i (μM)**	**Type of inhibition**	**Reference #**
CMP-Neu5Ac	Rat liver	46		Competitive	11
CDP	Bovine colostrum		13	Competitive	12
19	Rat liver		44	Competitive	11
20	Rat liver		84	Competitive	11
21	Rat liver		< 2,000	Competitive	13
22	Rat liver		0.35	Competitive	13
23	Rat liver		6.0	Competitive	13
24	Rat liver		0.04	Competitive	14

Ref. # 11: [Schaub, et al., 1998] Ref. # 12: [Paulson, et al., 1977]

Ref. # 13: [Amann, et al., 1998] Ref. # 14: [Muller, et al., 1998]

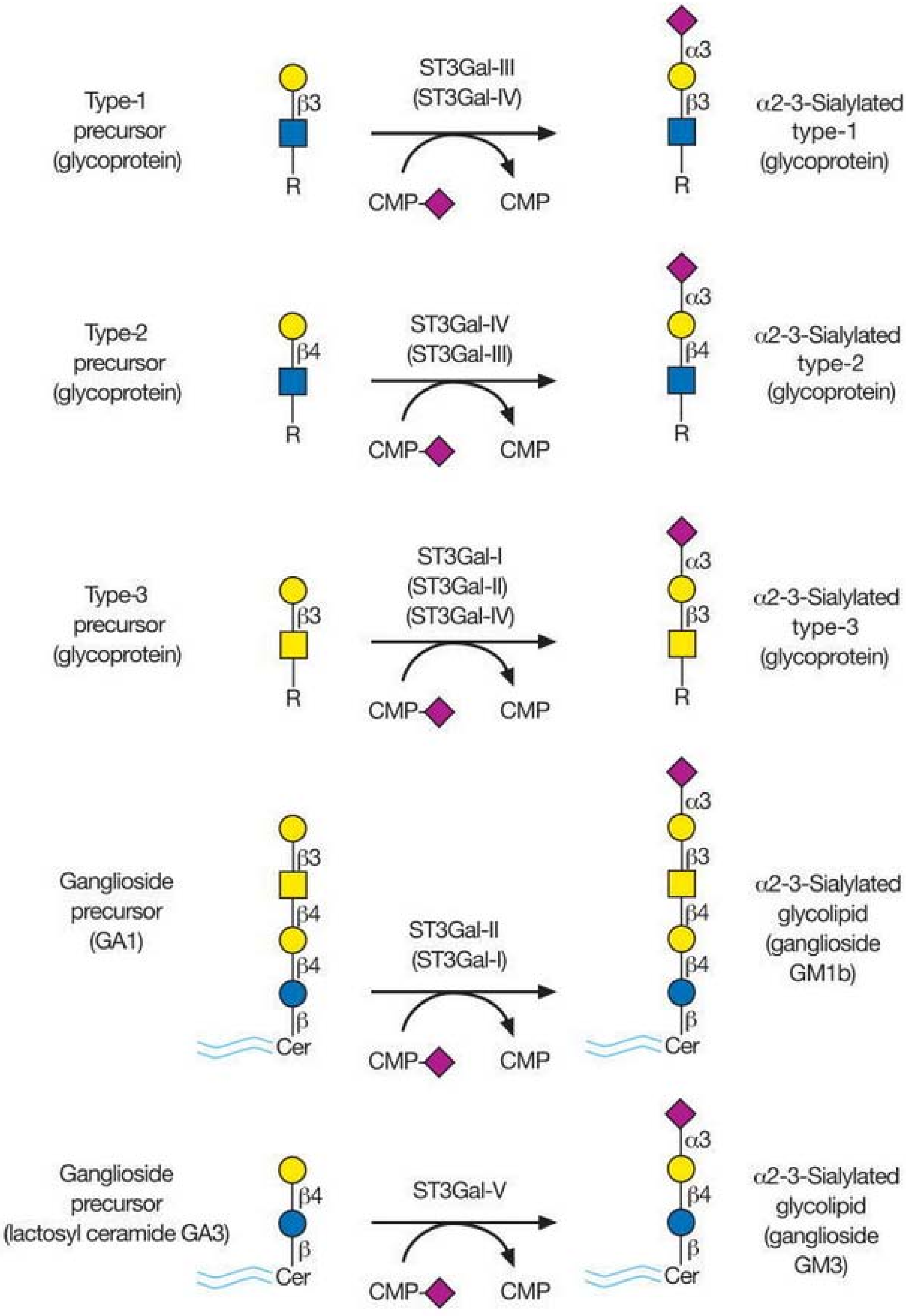
natural donor substrate, Km = 46 μM (Table 14). Though CDP is considerably smaller than the natural donor substrate, it has in common with CMP-Neu5Ac the cytidine moiety and a pair of negatively-charged oxygen atoms. This tight biding of CDP suggests, in the absence of an X-ray crystal structure of this ST in a complex with CMP-Neu5Ac, what parts of CMP-Neu5Ac are important for it to bind tightly in this ST's donor-binding site.

In 1998, a series of three papers were published by the Schmidt lab, describing a series of synthetic donor-based analogues that they made and tested against an ST6 Gal-I from rat liver, culminating in a transition-state analogue that binds 300-fold more tightly than CDP. In the first paper, they reported that compound **19** (Figure 43), a CMP-Neu5Ac analogue with Neu5Ac replaced with quinic acid, bound as tightly as CMP-Neu5Ac [Schaub, et al., 1998]. Replacing the quinic acid's axial C4-OH with an equatorial NH-Acetyl group (**20**) only reduced the binding strength. In their next paper, a series of transition-state analogues of CMP-Neu5Ac were made and tested against the same rat liver ST6 Gal-I [Amann, et al., 1998]. The pair of compounds **21** and **22** (Figure 43) reinforced the importance of having a pair of negatively charged oxygen atoms, separated by five bonds, just the same as they are in CMP-Neu5Ac. Compound **22** was found to bind 30-fold more tightly to the ST than CDP (Table 14). Compound **23** was a minor by-product of their chemical syntheses, which had a different ring system in place of Neu5Ac, which bound well to the ST (Table 14) considering it didn't have two negatively charged oxygen atoms. In their third paper they reported that compound **24**, a variant of **23** with two negatively charge oxygen atoms, binds 300-fold more tightly to the ST than CDP [Muller, et al. 1998].

Biosynthesis of a glycan with a terminal Sia α2-3 Gal disaccharide requires a member of the ST3Gal family of STs. This main family has six subfamilies [Harduin-

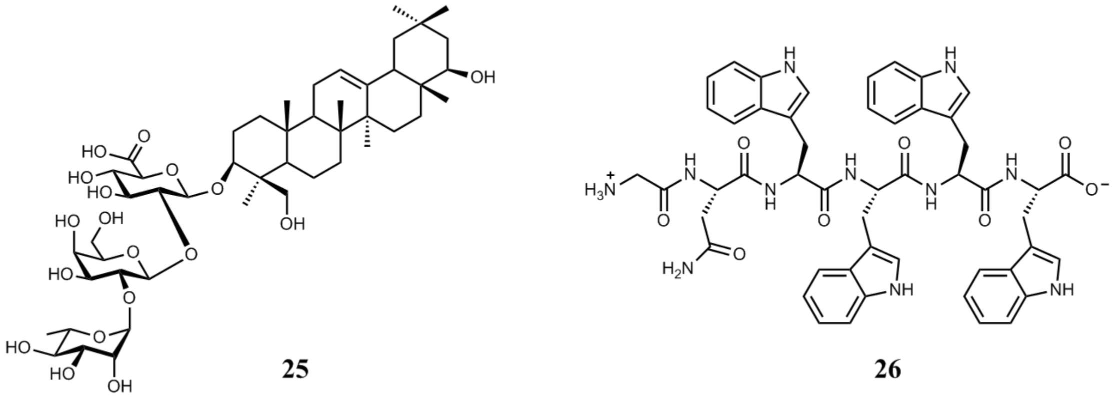
Lepers, et al., 2005], ST3Gal-I through ST3Gal-VI. Each ST3Gal subfamily is defined by its preferred acceptor substrate, as they all use the same donor substrate CMP-Neu5Ac. Figure 44 shows the preferred acceptor substrate for each of the subfamilies ST3Gal-I through ST3Gal-V, and the product. There is some acceptor substrate promiscuity. The subfamilies named in parentheses above the reaction arrows will transfer Neu5Ac to that acceptor *in vitro*, at relatively low levels [Varki, et al., 2009, Ch. 13].

Intense interest in finding small molecule inhibitors of sialyltransferases has fueled some large-scale screening efforts. Screening 7500 natural products and extracts, with the criteria of a molecular weight below 1 kDa and "significant hydrophobicity," revealed soyasaponin I (**25**, Figure 45) from soy beans [Wu, et al., 2001]. Soyasaponin I was found to be a CMP-Neu5Ac competitive inhibitor, against ST3Gal-I, with *Ki* = 2.3 μM (Table 15). Surprisingly, it doesn't appear to have any of the attributes of CDP that are believed to make CDP a tight-binding inhibitor of an ST6Gal-I (Table 14). Another similarly surprising inhibitor of an ST3Gal-I was found by screening a library containing all possible hexapeptides, excluding cysteine. The hexapaptide GNWWWW (**26**, Figure 45) was found to be an even better CMP-Neu5Ac competitive inhibitor [Lee, et al., 2002], with *Ki* = 1.1 μM (Table 15).



Type-1 precursor (glycoprotein)
β3
R
ST3Gal-III
(ST3Gal-IV)
CMP-
CMP
α3
β3
R
α2-3-Sialylated type-1 (glycoprotein)
Type-2 precursor (glycoprotein)
β4
R
ST3Gal-IV
(ST3Gal-III)
CMP-
CMP
α3
β4
R
α2-3-Sialylated type-2 (glycoprotein)
Type-3 precursor (glycoprotein)
β3
R
ST3Gal-I
(ST3Gal-II)
(ST3Gal-IV)
CMP-
CMP
α3
β3
R
α2-3-Sialylated type-3 (glycoprotein)
Ganglioside precursor (GA1)
β3
β4
β4
β
Cer
ST3Gal-II
(ST3Gal-I)
CMP-
CMP
α3
β3
β4
β4
β
Cer
α2-3-Sialylated glycolipid (ganglioside GM1b)
Ganglioside precursor (lactosyl ceramide GA3)
β4
β
Cer
ST3Gal-V
CMP-
CMP
α3
β4
β
Cer
α2-3-Sialylated glycolipid (ganglioside GM3)

Figure 44: The α2-3 sialyltransferases, the ST3Gal family, all use CMP-Neu5Ac as their donor substrate. Each family member has its own preferred acceptor substrate. [Reprinted from Varki, et al., 2009, Ch. 13]



25

26

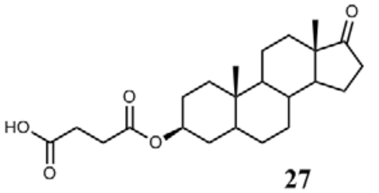
Figure 45: Competitive inhibitors of ST3Gal-I, found in large-scale screening efforts.

Table 15: Kinetics of inhibition of ST3 Gal members, inhibitors in Figures 45 and 46

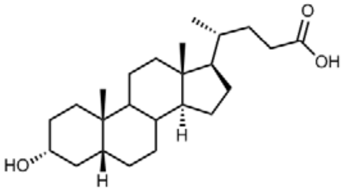
Compound	**ST3 Gal enzyme**	***K_i*** **(μM)**	**IC_{50}** **(μM)**	**Type of inhibition** (for which substrate)	**Ref. #**
25	ST3 Gal-I	2.3		Competitive (donor)	15
26	ST3 Gal-I	1.1		Competitive (donor)	16
27	ST3 Gal-V *		350	Noncompetitive (NA)	17
28	ST3 Gal-V *		21	Noncompetitive (NA)	17
29	ST3 Gal-V *		12	Noncompetitive (NA)	17
30	ST3 Gal-V *		6	Noncompetitive (NA)	17
31	ST3 Gal-V *		7	Noncompetitive (NA)	17
32	ST3 Gal-V *	2.2	5	Noncompetitive (donor)	17
33	ST3 Gal-I		0.88	NA	18
34	ST3 Gal-III		8.2	NA	8

*The authors (Ref. #17) did not explicitly state which ST3Gal family member they used. They only implied ST3 Gal-V by showing the acceptor substrate (Scheme 3) used by their ST3Gal enzyme, which is the acceptor preferred by ST3 Gal-V. NA = information Not Available from the paper referenced

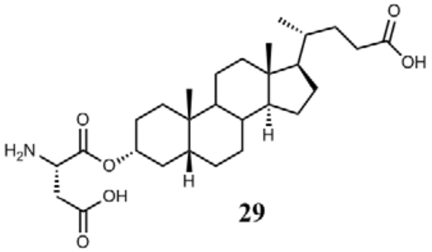
Ref. #8: [Hosoguchi, et al., 2010], Ref. #15: [Wu, et al., 2001], Ref. #16: [Lee, et al., 2002], Ref. #17: [Chang, et al., 2006], Ref. #18: [Chiang, et al., 2010].



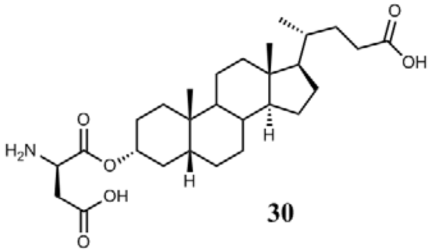
27



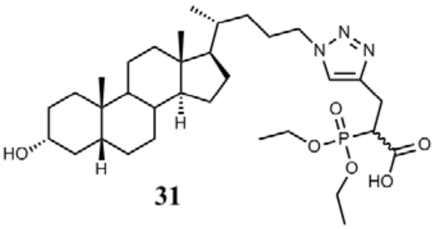
28



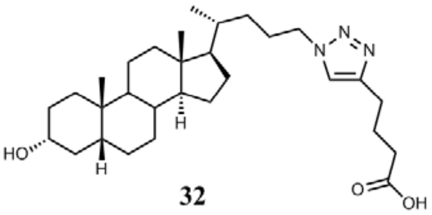
29



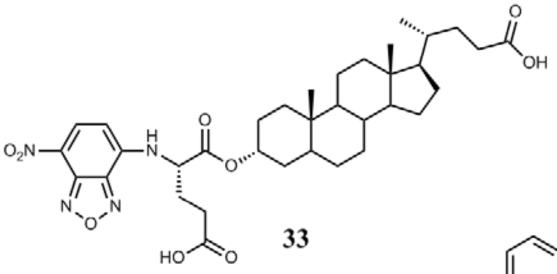
30



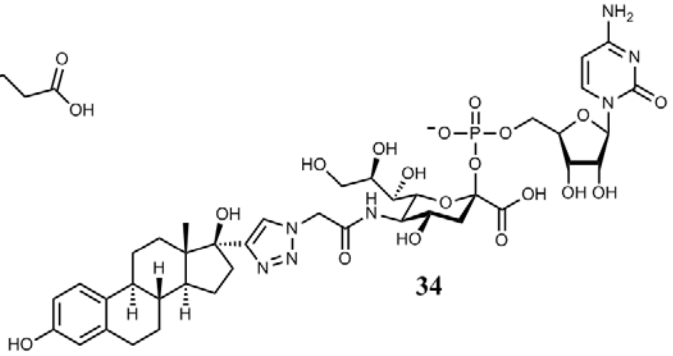
31



32



33



34

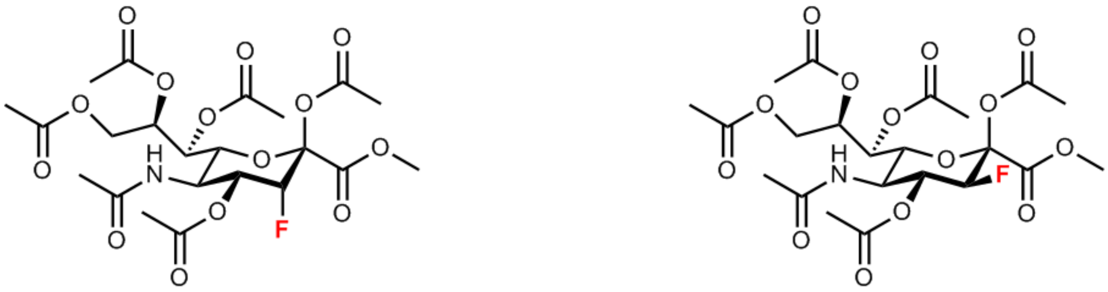
Figure 46: Inhibitors of ST3 Gal members

Inspired by the ST3 Gal-I inhibiting potency of soyasaponin I (**25**, Figure 45), a screening of a library of compounds with steroid-related structures revealed compounds **27** and **28** [Chang, et al., 2006]. It appears that this work, reported in 2006 [Chang, et al., 2006], was done with an ST3 Gal-V though the authors never explicitly state which ST they used (see Table 15 foot note). They synthesized and tested sixteen derivatives of compound **28**, lithocholic acid, and the most potent four of those are shown in Figure 46, compounds **29** – **32**, and in Table 15. Four years later, an even more potent lithocholic acid derivative, compound **33**, was reported [Chiang, et al., 2010]. Then, in 2011, the results of many *in vivo* tests of compound **30** were published, revealing many important results and suggesting that this inhibitor (now called Lith-O-Asp) be tested as "a novel antimetastasis drug for cancer treatment" [Chen, et al., 2011].

For a glycosyltransferase inhibitor to be useful clinically, or to study the role of a particular glycan in the biochemistry inside a living cell, it must not only be potent and specific for the target enzyme, but it must also be able to get inside the cell where that enzyme resides. Inhibitors that are small and not too hydrophilic have the best chance of being able to easily pass through the cell's lipid bilayer membranes. The molecular characteristics of being "small" and "not too hydrophilic" are intentionally vague, as what will actually work for a particular inhibitor or cell type must be determined empirically. Because cells typically have endogenous esterases in their cytosol, a molecule with a lot of hydroxyl groups can be made less hydrophilic by acetylating all of its hydroxyl groups. The peracetylated molecule will be able to pass through the lipid interior of a lipid bilayer membrane than its unacetylated version. Once inside the cell, cytosolic esterases will convert the peracetylated molecule back to its unacetylated form. Examples of such peracetylated inhibitors include compounds **4** and **5** (Figure 37) which were used for tests inside cells where they are presumably converted into **2** and **3**, respectively.

It has been known for over a decade that putting a fluorine atom on the carbon adjacent to a monosaccharide's anomeric carbon, will produce a ring system that mimics a flattened half-chair transition state. By peracetylating fluorinated analogues of the sialic acid Neu5Ac they become cell permeable. Two peracetylated fluorinated analogues of Neu5Ac (compounds **35** and **36**, Figure 47) were tested in a several cell lines for their ability to enter the cell and, after deacetylation by an endogenous cytosolic esterase, enter into the cell's sialic acid salvage pathway and become converted into fluorinated analogues of CMP-Neu5Ac (Figure 42), and then inhibit the biosynthesis of a variety of sialylated glycans [Rillahan, et al., 2012].

Compound **35** was found to be a global metabolic inhibitor, abolishing all sialylation of N-linked glycans and most sialylation of O-linked glycans [Rillahan, et al.,



35 **36**

Figure 47: Fluorinated analogues of the sialic acid Neu5Ac

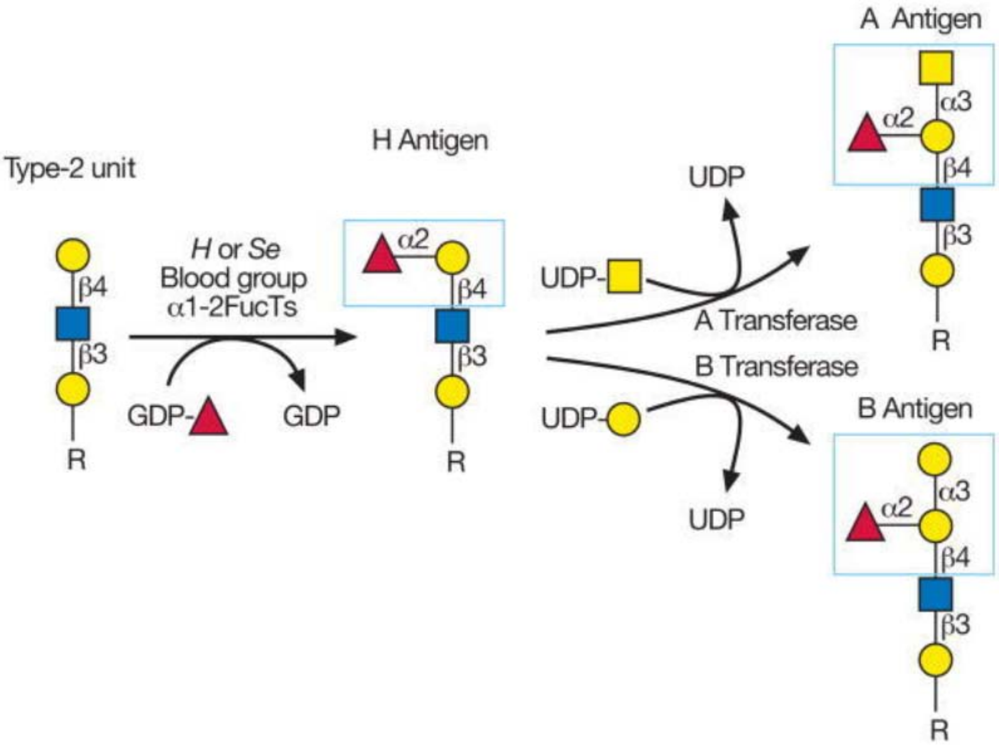
2012]. Compound **35** was found to substantially inhibit, in the human myeloid cell line HL-60, the biosynthesis of the tetrasaccharide SLe^x (Figure 41) which is a ligand for selectins and regulates the movement of leukocytes out of blood vessels and into inflamed tissue. While **35** was found to have an inhibitory effect on all of the cell's STs, after entering a cell's sialic acid salvage pathway, its epimer, compound **36**, was found to have no inhibitory effect in any of the cell lines in which it was tested. The authors hypothesize that this is due to the cells' CMP-Neu5Ac synthases not being able to use the deacetylated form of compound **36** as a substrate [Rillahan, et al., 2012].

4.2.4 Inhibitors of The ABO Blood Group Glycosyltransferases

In the early years of the 20th century, Landsteiner and colleagues found that people could be categorized into a few groups based on the presence or absence of substances in their blood serum that could agglutinate red blood cells taken from individuals from a different group. Before these "substances" in the serum were identified as antibodies, or anything was known about the structure of the antigens that they were binding to, this discovery immediately led to blood-typing and successful blood transfusions. What Landsteiner discovered is now known as the ABO blood group system, and is due to three different alleles, called A, B, and O, at the ABO locus. The antigens of the ABO system are called A, B, and H, and are found on the surfaces of red blood cells as well as many others such as the epithelial cells lining certain exocrine glands, and the gastrointestinal, pulmonary, urinary, and reproductive tracts [Lowe, 1993].

In the 1950s the chemical structures of these A, B, and H antigens were discovered to be oligosaccharides, and later the subtle structural differences between these antigenic glycans were defined. The core structure of these three antigens varies slightly based on the cell type that they are expressed on the surface of, but the differences between the A, B, and H antigens are the same. The core structure of the A, B, and H antigens on red blood cells is the type-2 unit (Galβ1-4GlcNAc, aka, LacNAc, or

LN, in Figure 35) [Lowe, 1993]. The ABO locus A allele encodes the glycosyltranferase (GTA) required to biosynthesize the A antigen. Similarly, the B allele encodes a glycosyltranferase called GTB or B transferase. The O allele differs from the A allele by a single-base deletion in the coding region, causing a frame shift that makes it a non-functional gene [Yamamoto, et al., 1990]. Individuals who are homozygous for the O allele have blood-type O and have only the H antigen. The A and B antigens are terminal trisaccharides biosynthesized from the terminal disaccharide H antigen, by the action of either GTA (A Transferase) or GTB (B Transferase), respectively, as shown for type-2 A, B, and H antigens in Figure 48.



Type-2 unit
β4
β3
R
H or Se
Blood group
α1-2FucTs
GDP-
GDP
H Antigen
α2
β4
β3
R
UDP-
UDP
A Transferase
B Transferase
UDP-
UDP
A Antigen
α3
α2
β4
β3
R
B Antigen
α3
α2
β4
β3
R

Figure 48: Biosynthesis of Type-2 H, A, and B blood-group antigens. **R** represents an N-glycan, O-glycan, or a glycolipid. [Reprinted from Varki, et al., 2009, Ch. 13]

The substrates used by the glycosyltransferases GTA and GTB, in Figure 49, are remarkably similar, differing only by the presence of a 2-acetamido group in place of a 2-OH group on the donor substrate sugar. This leads to the only difference between the blood-group A and B antigens being just this 2-acetamido group, on the A antigen, in place of a 2-OH, on the B antigen, on the terminal monosaccharide (Figure 49). There are 2-3x10^6 A, B, or H antigens on a typical red blood cell [Lowe, 1993].

The ABO blood group system glycosyltransferases, GTA and GTB, are possibly the most highly homologous natural pair of enzymes which utilize different substrates. Cloning and sequencing of their genes in 1990 revealed that GTA and GTB are each made up of 354 amino acids and are identical except for the four residues at positions 176, 235, 266, and 268 [Yamamoto, et al., 1990]. Before the first published 3D structures of these two enzymes in 2002 [Patenaude, et al., 2002] efforts to understand how they work included searching for inhibitors that could specifically inhibit one over the other. The H antigen, disaccharide L-Fucose α1-2 D-Galactose, (Figure 49) is the minimal acceptor substrate recognized by both GTA and GTB. A derivative of the H antigen with an n-octyl group glycosidically linked to the galactose residue (**37**, Figure 50) was used as the acceptor substrate and variants of compound **37** were made and tested for their ability to inhibit GTA and GTB. These compounds were used to probe the binding requirements of the active sites.

Modifying the substituents at each of C-3, C-4, or C-6 of galactose revealed that the C-4 OH group is required for binding to the active site of either GTA or GTB, while modifying the hydroxyl groups at C-3 or C-6 was tolerated. The galactose-modified variants of the H antigen with the most inhibitory character were only those with C-3 modifications, and are shown in Figure 50. Compounds **38** and **41** each inhibit GTB

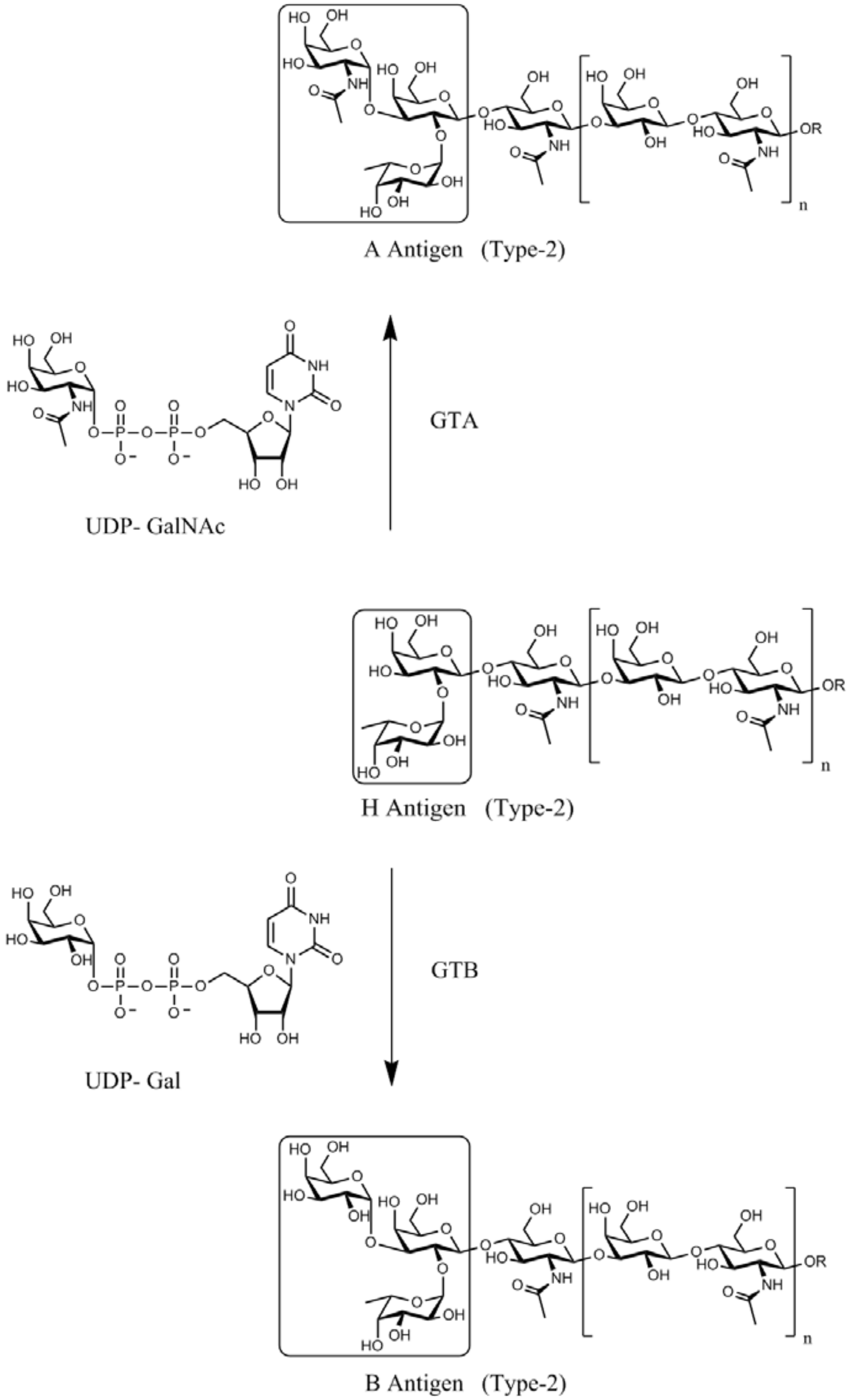

A Antigen (Type-2)
UDP- GalNAc
GTA
H Antigen (Type-2)
UDP- Gal
GTB
B Antigen (Type-2)

Figure 49: Reactions catalyzed by the ABO blood group system glycosyltransferases GTA and GTB. **R** represents either an N-glycan, O-glycan, or a glycolipid.

37 R^1 = OH, R^2 = H
38 R^1 = H, R^2 = H
39 R^1 = F, R^2 = H
40 R^1 = OCH_3, R^2 = H
41 R^1 = H, R^2 = OH
42 R^1 = NH_2, R^2 = H

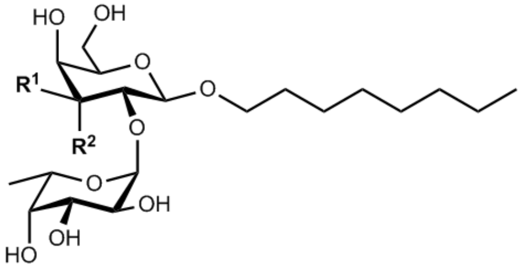


Figure 50: Human blood group ABO system H antigen-mimic (37) and C-3-modified analogues

Table 16: Kinetics of inhibition of human ABO blood group system glycosyltransferases (compounds in Figure 50)

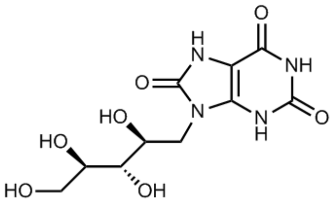
Compound	**A Transferase (GTA)**		**B Transferase (GTB)**		**Type of inhibition**	**Ref. #**
	Inhibition	K_m **(μM)**	**Inhibition**	K_m **(μM)**		
37	- - -	1.50	- - -	21.9	- - -	25
		K_i **(μM)**		K_i **(μM)**		
38	30 %	68	85 %	14	Competitive	25
39	22 %	48	24 %	110	Competitive	25
40	4 %	n.d.	15 %	n.d.	n.d.	26
41	36 %	22	88 %	7.8	n.d. for GTA Competitive for GTB	26
42	98 %	0.2	93 %	5	n.d.	26

n.d. = not determined
Ref. #25: [Lowary and Hindsgaul, 1993] , Ref. #26: [Lowary and Hindsgaul, 1994]

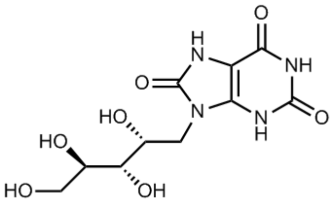
better than they inhibit GTA and have K_i values below the K_m value for the acceptor **37** (Table 16), suggesting that removing the equatorial hydroxyl group at C-3 makes for a better GTB-specific inhibitor. Compound **42**, with an equatorial NH_2 group at C-3 is the best inhibitor of this group, for both enzymes, though inhibits GTA better than it does GTB. These results led the authors to conclude that there might be a negatively charged residue, in both enzymes' acceptor binding sites that is close to the C-3 hydroxyl group of galactose [Lowary and Hindsgaul, 1994]. They also tested the synthetic precursor to **42** that is identical to **42** except that it lacks the fucose residue in **42**, and found it to be inactive as an inhibitor.

Nine years later, the importance of the fucose residue in **42** and the identity of the nearby negatively charged amino acid in the acceptor binding site were revealed when the 3D structures of GTA and GTB in complex with compounds **38** or **42** were published [Nguyen, et al., 2003]. The 3-amino analog, compound **42**, was found to form an intramolecular hydrogen bond between the amino group on the galactose residue and the C-2 hydroxyl group on the fucose residue when **42** binds to the acceptor binding site without UDP bound in the donor binding site. This intramolecular hydrogen bond changes the low-energy conformation of **42** and pushes the fucose residue into the UDP-bindidng site. The authors concluded that acceptor-analog **42** is competing for binding with both the natural acceptor and the UDP portion of the natural donor and that this explains the complex mode of inhibition that prevented determining its type of inhibition earlier [Lowary and Hindsgaul, 1994].

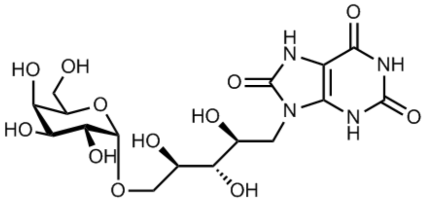
With a pair of enzymes as similar in sequence [Yamamoto, et al., 1990], structure [Patenaude, et al., 2002], and substrate specificities as are GTA and GTB, it is a daunting challenge to design an inhibitor that is a potent inhibitor for only one of them and not both. Compound **42** has been shown to inhibit GTA in a cell line derived from a human colorectal carcinoma (HT29), which is known to be blood group A positive [Laferte, et al., 2000]. These authors reported that HT29 cells grown with **42** in the medium showed a decrease in the expression of blood group A determinants, inside the cells and on their surfaces. They reported that **42** had an IC_{50} of 280 uM, for intracellular inhibition of GTA in HT29 cells. But, they did not report testing **42** against GTB in any way, and previous publications from the same lab reported significant, though less than for GTA, inhibitory potency against GTB [Lowary and Hindsgaul, 1994] as shown in Table 16. Recently reported inhibitors of GTB, compounds **10 – 18**, described earlier in this chapter and shown in Figure 40 with their inhibition data in Table 13, have not been reported to have been tested against GTA.



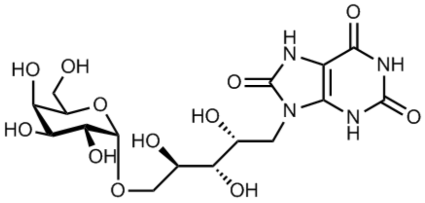
43



44



45



46

Figure 51: Nonionic mimics of UDP and UDP-Gal

Table 17: Kinetics of inhibition of nonionic mimics of UDP and UDP-Gal (with Mg^{2+}) against human ABO blood group glycosyltransferases (inhibitors in Figure 51)

Compound	**GTA (A Transferase)**	**GTB (B Transferase)**	**Type of inhibition**	**Ref. #**
	K_i (μM)	**K_i (μM)**		
43	n.d.	361	Competitive	29
44	n.d.	571	Competitive	29
46	No Inhibition	565	Competitive	30
	K_m (μM)	**K_m (μM)**		
UDP-Gal	115	260		30

n.d. = not determined

Ref. #29: [Schaefer, et al., 2012] ; Ref. #30: [Schaefer, et al., 2013]

Two nonionic derivatives of uric acid, compounds **43** and **44** (Figure 51), have been reported to bind to GTB by mimicking uridine diphosphate (UDP) [Schaefer, et al., 2012]. These compounds have a pentityl group which serves to mimic the diphosphate group of UDP without giving the compounds any ionic charge and thus enhancing their cell permeability. These two compounds differ only in the stereochemistry of their pentityl groups, which was found to make **43** a slightly better inhibitor of GTB, as shown in Table 17. The same lab reported the following year that these two compounds inhibit both GTA and GTB [Schaefer, et al., 2013]. In that paper they report modifying one of these uric acid derivatives by adding a galactose residue to the end of the pentityl group, to make a GTB-specific inhibitor. They chose not to modify the more potent of these two UDP mimics, compound **43**, because synthesizing its galactose derivative, compound **45**, was much more difficult. They made compound **46** instead and reported that **46** did not inhibit GTA at all, while it inhibited GTB as well as its parent compound, **44**, did (Table 17). They commented that NDP-sugar donor-analog inhibitors of glycosyltransferases derive most of their enzyme-binding affinity from their base and their diphosphate group, while enzyme specificity is due to the sugar group [Schaefer, et al., 2013]. This makes a lot of sense except that, probably, the whole nucleoside is important for enzyme-binding affinity.

4.3 Redesigns of Glycosyltransferases

4.3.1 Redesigns of The ABO Blood Group Glycosyltransferases

The usual primary goal of redesigning an enzyme is to create an altered version of the original which then catalyzes a different reaction, using a different substrate, by changing only a few of the amino acids in the original enzyme. This is currently very hard to do in the lab, but the ABO blood group glycosyltransferases, GTA and GTB, described in subsection E.2.4, provide an example of how well enzyme redesign can be done in Nature. In this subsection I'll briefly describe the work done to understand precisely the roles of each of the donor specificity-determining amino acids in this pair of naturally redesigned enzymes, before and after the 3D structures of these two highly homologous enzymes were published in 2002 [Patenaude, et al., 2002].

As soon as the genes for GTA and GTB had been cloned and sequenced, and the difference of only four amino acids out of 354 revealed [Yamamoto, et al., 1990], efforts to understand how these four critical residues determined each enzyme's donor-substrate specificity began with making hybrids of these two glycosyltransferases. GTA/B hybrid genes were made, encoding all 14 possible GTA/B hybrid enzymes, and, along with the two wild type genes, expressed in HeLa cells [Yamamoto and Hakomori, 1990], which are homozygous for the blood group O allele. ABO blood group antigens

Table 18: Nomenclature system for hybrid and mutant glycosyltransferases, GTA and GTB, of the human ABO blood group system.

Human ABO blood group Glycosyltransferase		**UDP-Sugar Donor Used**	**Amino acids at the four critical positions**				**Ref. #**
Name	**Description**		**176**	**235**	**266**	**268**	
AAAA	(wild type) GTA	UDP-GalNAc	Arg	Gly	Leu	Gly	
BBBB	(wild type) GTB	UDP-Gal	Gly	Ser	Met	Ala	
AABA	hybrid GTA/B	Either	Arg	Gly	Met	Gly	31
ABAB	hybrid GTA/B	Either	Arg	Ser	Leu	Ala	31
ABBA	hybrid GTA/B	Either	Arg	Ser	Met	Gly	31
BABA	hybrid GTA/B	Either	Gly	Gly	Met	Gly	31
BBAB	hybrid GTA/B	Either	Gly	Ser	Leu	Ala	31
BBBA	hybrid GTA/B	Either	Gly	Ser	Met	Gly	31
AA(Gly)B	mutant GTA/B	Either	Arg	Gly	Gly	Ala	38

Either = UDP-GalNAc or UDP-Gal (aka, dual-specificity)
Ref. #31: [Yamamoto and Hakomori, 1990] ; Ref. #38: [Yamamoto, et al., 2001]

were then detected on these transfected HeLa cells, using A- and B-antigen-specific antibodies. Six of these 14 GTA/B hybrid enzymes had dual specificity, catalyzing the final step of biosynthesis of both the A and B blood group antigens, and are listed in Table 18. Of the other eight hybrid GTA/B enzymes five had only GTA activity and three had only GTB activity. These authors concluded from these experiments that the amino acids at positions 266 and 268 were the most important for determining the enzyme's donor-substrate specificity, with the residue at position 235 being of some minor importance [Yamamoto and Hakomori, 1990].

They went on to determine partial sequences of the ABO genes of several primates, found conserved differences between these A and B alleles only at the positions corresponding to amino acids 266 and 268 in the human GTA and GTB enzymes and concluded that those two residues were the most important for determining donor-substrate specificity [Kominato, et al., 1992]. The discovery of a second type of O allele in the Danish population, without the single nucleotide deletion in all other known O alleles, and then the comparison of human GTA and GTB enzymes with bovine and murine α1-3 galactosyltransferase, led to a thorough testing of the hypothesis that the amino acid at position 268, not that at position 266, was responsible for determining the donor-substrate specificity of GTA and GTB. Mutants of GTA and GTB, with every possible amino acid at position 268 of each, were made and tested, and it was concluded that the residue at position 268 was in fact responsible for determining both the activity and the donor-substrate specificity of these two homologous glycosyltransferases [Yamamoto and McNeill, 1996]. Years later, this same group was preparing to create a mouse model of the human blood group ABO system when they cloned the murine equivalent of the human ABO locus and discovered that the predominant murine allele there encodes an enzyme with both GTA and GTB activities. Sequencing of this gene revealed that this murine dual-specificity enzyme was highly homologous to human GTA and GTB and was identical to GTA at its critical positions corresponding to 176 and 235, identical to GTB at its corresponding critical position 268, and identical to neither at its corresponding critical position 266 where it had a glycine [Yamamoto, et al., 2001]. This set of four critical amino acids would later be created in the human GTA background, given the name AA(Gly)B in the new system of nomenclature (Table 18), and become a well-studied dual-specificity glycosyltransferase.

Soluble versions of wild type GTA and GTB and three hybrid enzymes (BAAA, BBAA, and BBBA) were engineered and then expressed and purified from *E. coli* [Seto, et al., 1997]. Kinetic characterization *in vitro*, using the hydrophobic derivative of the H antigen (**37**, Figure 50) as the acceptor-substrate, gave some different results than when the activities of the same enzymes were measured *in vivo* by transfection of HeLa cells [Yamamoto and Hakomori, 1990]. Most surprisingly, the hybrid enzyme BAAA had only half the activity of wild type GTA, when measured in transfected HeLa cells, but it had an 11-fold increase in the turnover number (k_{cat}) and a 4-fold increase in specificity constant with respect to the acceptor-substrate (k_{cat} / K_{mA}) relative to the activity of wild type GTA [Seto, et al., 1997]. Detailed kinetic characterization of these three hybrid enzymes led to the conclusions that the amino acid at position 176 has little effect on the binding of the acceptor-substrate, but does affect k_{cat} , residue 235 could affect binding of the acceptor, and "residues 266 and 268 could be most critical for binding of the nucleotide sugar donor" [Seto, et al., 1997]. Similar *in vitro* kinetic characterizations were done by the same researchers, two years later, with an additional four hybrid enzymes (ABAA, AABA, ABBA, and BABA). This time they concluded that one amino acid difference was the most important for "determining the A vs. B donor specificity" and that was the amino acid at position 266 [Seto, et al., 1999].

In 2002, the first X-ray crystal structures of GTA and GTB, alone and in complex with the disaccharide H-antigen acceptor and UDP, were published [Patenaude, et al., 2002]. After the sugar moieties GalNAc and Gal were modeled into the donor-substrate binding sites of GTA and GTB, respectively, the roles of the four critical amino acids, which differ between these two enzymes, could be proposed. Only two of these four residues are located in the active sites of GTA and GTB were they can contact the distinguishing sugar moiety of each sugar-donor substrate. Residue 268 is located where it can only contact the C-3 and C-4 hydroxyl groups, of the GalNAc or Gal sugar moiety, which are identical in the natural sugar-donor substrates for GTA and GTB. Only residue 266 is able to contact the C-2 substituents, acetamido on GalNAc or hydroxyl on Gal, which are the only structural differences between the two natural sugar-donor substrates for GTA and GTB, respectively. This supports the earlier suggestion that only residue 266 is responsible for determining the donor-substrate specificity of GTA and GTB. These 3D structures also showed that the conserved glutamate at position 303 (E303) is a good candidate to be the enzyme-nucleophile that is believed to be required in the mechanism of retaining GTs (shown in Figure 29) such as GTA and GTB. To test this hypothesis, they expressed and purified the single mutant E303A GTB, showed it was homogeneous and formed crystals with the same space group and unit cell as wild type GTB, but had a 30,000 fold reduction in specific activity relative to wild type GTB [Patenaude, et al., 2002].

A detailed understanding of this model system for naturally redesigned enzymes, human GTA and GTB, grew as more natural mutants of them were discovered, kinetically characterized, and their 3D structures solved and published. A natural dual specificity enzyme was found to encode GTB with a single point mutation causing the conserved Proline 234 to be replaced by Serine (P234S). The mutant P234S GTB was expressed and purified and found to have reversed its donor-substrate preference [Marcus, et al., 2003]. The donor-substrate specificity constants (k_{cat}/K_m), in units of s^{-1} mM^{-1}, for P234S GTB were found to be 2.3 for UDP-Gal, the natural donor-substrate for wild type GTB, and 86.2 for UDP-GalNAc, the natural donor-substrate for wild type GTA. The same authors reported an X-ray crystal structure of P234S GTB, with and without the acceptor-substrate H-antigen bound. These 3D structures showed that replacing Pro-234 with a Serine created a void where the gamma carbon of Proline would normally be in van der Walls contact with Met-266, allowing the side chain of Met-266 to adopt a different conformation which then allows C-2 acetamido group on UDP-GalNAc to fit. This also creates a void that is left empty when UDP-Gal is in the donor-binding pocket [Marcus, et al., 2003], thus creating a preference for binding UDP-GalNAc, the natural donor-substrate of GTA.

Three other natural mutants of GTB were found in blood banking labs, all with mutations replacing conserved Methionine 214 with either Valine, Threonine, or Arginine. Met-214 is adjacent to the conserved DXD motif, present in all GTs with the GT-A fold, which coordinates the required Mn^{2+} ion (shown as M^{2+} in Figure 29). In GTA and GTB the DXD motif consists of Asp-211, Val-212, and Asp-213. Mutations to GTB's Met-214, being next to GTB's DVD motif, could be expected to have an effect on catalysis. Kinetic characterization of these three Met-214 mutations revealed two different types of changes. M214R GTB showed extremely low activity and extremely low turnover rates (k_{cat}) with either UDP-Gal or UDP-GalNAc, while M214T GTB and M214V GTB had both GTA and GTB activity [Persson, et al., 2007]. In the same paper, X-ray crystal structures of M214R GTB and M214T GTB showed a structural basis for both of these kinetic results. The 3D structure of M214R GTB showed that Arg-214 interferes with the catalytic mechanism by, together with Asp-211 and Asp-213, trapping a water molecule where the Mn^{2+} ion normally binds to the two aspartic acid residues in the DVD motif. The 3D structures of M214T GTB, with and without UDP bound, showed that replacing Met-214 with Thr-214 allows the donor-substrate-determining side-chain of Met-266 to adopt a conformation that creates enough additional space in the donor-substrate binding site to allow the larger donor-substrate UDP-GalNAc to bind there [Persson, et al., 2007].

Analysis of numerous X-ray crystal structures of GTA, GTB, and various hybrids, alone and in complexes with various donor-substrates, acceptor substrates, and analogues thereof, have revealed that the conformation of these enzymes changes as these enzymes go through the steps required for catalysis to occur. One set of 12 crystal structures, comprising three versions of GTB (BBBB, ABBB, and AABB), both natural donor-substrates, and three versions of the H-antigen acceptor substrate, showed that GTB goes through at least three distinct conformations as it prepares for catalysis. These conformations are called “open form,” semi-closed,” and “closed form” and are defined by how ordered the amino acids are that make up an internal loop (residues 176-195) and the C-terminus (residues 346-354). Before either a donor-substrate or an acceptor-substrate has bound in the active site, the enzyme is in the “open form,” after binding to UDP the internal loop moves to achieve the “semi-closed” conformation and create the acceptor-substrate binding site, and when bound to UDP or UDP-Gal and acceptor the enzyme's C-terminus becomes ordered and the enzyme is in the "closed form [Alfaro, et al., 2008]. Confirmation of the importance, to the enzyme's activity, of these conformational changes came from a crystal structure of AA(Gly)B in complex with the base-modified UDP-Gal inhibitor **10** (Figure 40) showing that this inhibitor's formylthienyl substituent is blocking the normal movement of the "internal loop" and thus blocking catalysis [Pesnot, et al., 2010].

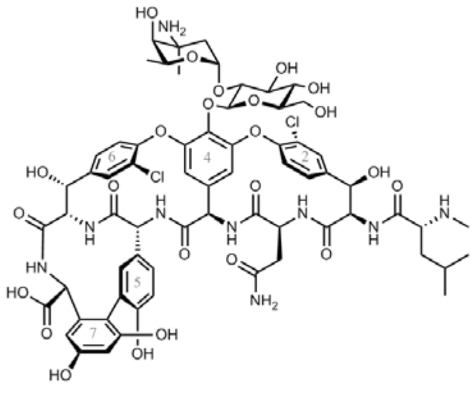
Important electrostatic interactions are believed to exist between the bound UDP-sugar donor-substrate and these highly dynamic portions, this 20-residue internal loop and the C-terminal nine residues, of these glycosyltransferases. Both wild type GTB and the hybrid ABBB, in the absence of any substrate, crystallized in the "open form" where the nine C-terminal residues and a major portion of the internal loop are disordered [Alfaro, et al., 2008]. These authors proposed that this open form is maintained, in the absence of a UDP-sugar donor-substrate, by the mutual repulsion of the positively charged side chains in the internal loop (Lys-179, Arg-180, and Arg-188) and the C-terminus (Lys-346 and Arg-352), and that the conformational change to the semi-closed form becomes possible after a UDP-sugar binds in the donor-binding site and the negatively charged diphosphate of UDP interacts with some of these positively charged side chains in the vicinity. A recently published set of 13 X-ray crystal structures of

GTA, GTB, and the hybrid ABBA, at various pH values from 5.0 to 10.0, support this hypothesis [Johal, et al., 2014]. This series of 3D structures at increasing pH showed a gradual increase in the order of this internal loop and the C-terminus leading to a pH-induced semi-closed conformation in GTB and ABBA. These authors conclude that this conformational change takes place with increasing pH because the positively charged side chains are becoming more completely neutralized, reducing their mutual repulsion that had been proposed to maintain the open conformation.

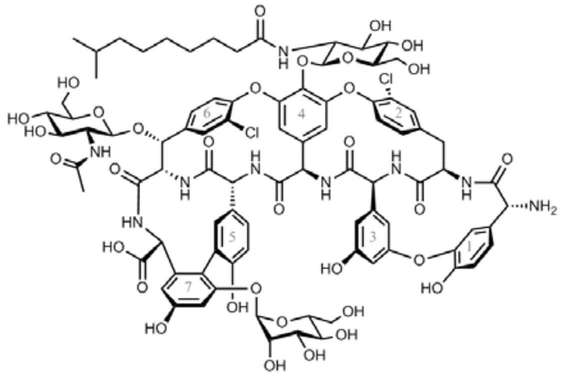
4.3.2 Redesigns of The Glycopeptide-Antibiotic Glycosyltransferases

All glycopeptide antibiotics are built on a crosslinked heptapeptide core that binds to the C-terminal dipeptide D-Alanyl-D-Alanine (D-Ala-D-Ala). By binding to the D-Ala$_4$–D-Ala$_5$ of the pentapeptide on uncrosslinked peptidoglycan (PG) chains on bacterial lipid II molecules, glycopeptides block the formation of a crosslinked PG layer covering the outer surface of the cytoplasmic membrane of gram-positive bacteria, and thus stop them from growing by blocking their cell wall biosynthesis. There are two subfamilies of glycopeptide antibiotics (GPAs), exemplified by the microbial natural products vancomycin and teicoplanin [Kruger, et al., 2005], also called a lipoglycopeptide, shown in Figure 52. The acyl chain on teicoplanin makes it lipophilic, likely to be cell membrane-bound, and is considered to be responsible for its antimicrobial properties that distinguish it from the vancomycin subfamily of glycopeptides.

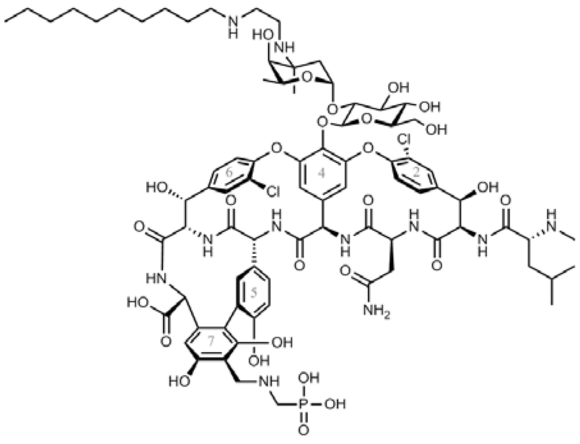
At the end of the 20th century vancomycin and teicoplanin were "the antibiotics of choice against methicillin-resistant *Staphylococcus aureus* (MRSA)" [van Wageningen, et al., 1998]. Vancomycin was discovered in the early 1950s, produced by microbes in a soil sample from Borneo, and was developed and brought to the clinic by the U.S. company Eli Lilly in the late 1950s. The natural lipoglycopeptide teicoplanin was developed in Europe and brought into clinical use there in the mid-1980s [Yim, et al., 2014], but has never been available in the U.S. Telavancin (Vibactiv), shown in Figure 52, is a semisynthetic lipoglycopeptide that was developed in 2004 and approved by the Food and Drug Administration (FDA) in 2009.



Vancomycin



Teicoplanin

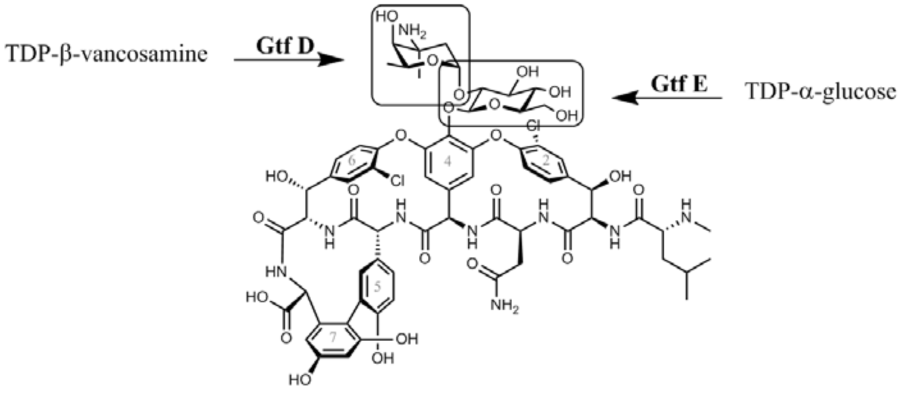


Telavancin

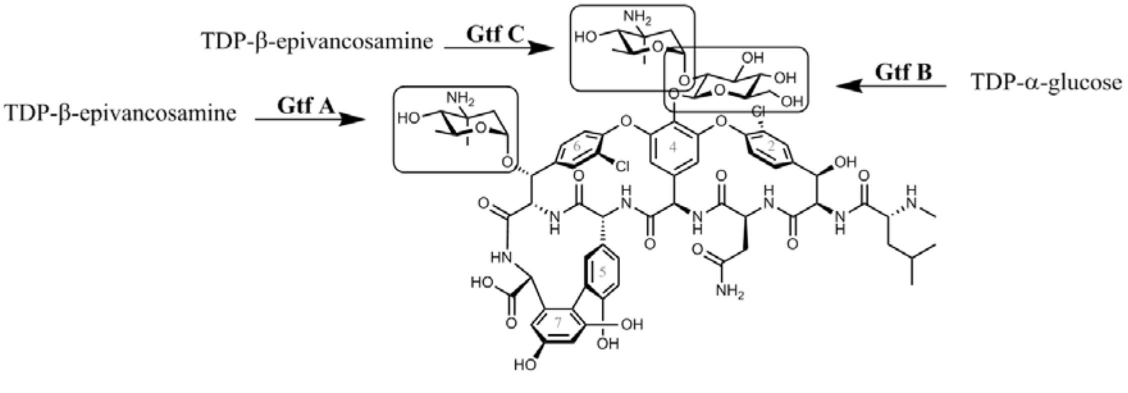
Figure 52: Chemical structures of the first clinically useful glycopeptide and lipoglycopeptide antibiotics: vancomycin (natural), teicoplanin (natural), and telavancin (synthetic derivative of vancomycin).

Telavancin is synthesized from vancomycin in two steps. First, a decylaminoethyl group is added to the nitrogen of the vancosamine residue. Then, a phosphonomethylaminomethyl group is added to the resorcinol-like C-4′ of the side chain of amino acid 7 [Leadbetter, et al., 2004]. Telavancin's superior bactericidal activity, relative to vancomycin, has been attributed to it being able to harm bacterial cells in two ways. As with all GPAs, telavancin binds to peptidoglycan (PG) precursor molecules and blocks PG crosslinking, thus interfering with biosynthesis of the bacterial cell wall. Also, as a lipoglycopeptide, telavancin molecules can insert their fatty acyl chains into bacterial lipid bilayer membranes, locally damage the cell membrane's integrity, and cause a loss of membrane potential by causing the cell membrane to leak [Higgins, et al., 2005; Hegde and Janc, 2014].

The first glycopeptide glycosyltransferase genes to be cloned were those for the vancomycin subfamily GPA chloroeremomycin, shown in Figure 53. Because chloroeremomycin has three monosaccharides there should be three distinct glycosyltransferases encoded in the genome of the microbe, *Amycolatopsis orientalis*, which naturally produces it. Analysis of the chloroeremomycin biosynthetic operon revealed 39 putative genes which included three glycosyltlransferase genes, which were then named *gtfA*, *gtfB*, and *gtfC*, which encode glycosyltransferases that were thus named GtfA, GtfB, and GtfC [van Wageningen, et al., 1998]. Biosynthesis of vancomycin


TDP-β-vancosamine
Gtf D
Gtf E
TDP-α-glucose

Vancomycin


TDP-β-epivancosamine
Gtf C
TDP-β-epivancosamine
Gtf A
Gtf B
TDP-α-glucose

Chloroeremomycin

Figure 53: The glycosyltransferases, and the monosaccharides that they each transfer, in the biosynthesis of the natural GPAs vancomycin and chloroeremomycin.

requires two glycosyltransferases named GtfD and GtfE, which are structurally and functionally homologous to GtfC and GtfB, respectively. All five of these glycosyltlransferases are shown in Figure 53, with an arrow pointing from the name of their natural sugar-donor substrate, and to the monosaccharide that they attach. Three of these glycosyltransferases have had their X-ray crystal structures solved: GtfB was the first [Mulichak, et al., 2001], followed by GtfA [Mulichak, et al., 2003], and most recently GtfD [Mulichak, et al., 2004].

Soon after these five glycosyltlransferases could be heterologously expressed, purified from *E. coli*, and characterized *in vitro*, it was discovered that they differed from each other in the degree of substrate selectivity that they had, some of them were more promiscuous than others, regarding what donor and acceptor substrates they could use. GtfE was found to be able to use the teicoplanin aglycone (AGT) core (teicoplanin without its three monosaccharides, or the acyl chain attached to its glucosamine) as its acceptor substrate and transfer a glucose to the C4′-OH of the 4-hydroxyphenylglycine. That glucose-AGT was then used as the acceptor substrate by GtfD, which then used UDP-β-epivancosamine as its donor substrate and transferred epivancosamine to the C2′-OH of the glucose-AGT forming a disaccharide and producing a novel hybrid glycopeptide [Losey, et al., 2001]. The next year, this same group reported that GtfE has even more donor-substrate flexibility by showing that it could use any of a set of ten synthetic deoxy- and amino-derivatives of UDP- or TDP-glucose [Losey, et al., 2002].

One year later the space of hybrid glycopeptides reachable, by this method of *in vitro* glycorandomization (IVG) using promiscuous Gtfs, increased when it was reported that GtfE could use as donor-substrate an additional 21 synthetic nucleotide-sugars, all different than glucose, the natural donor-substrate sugar for GtfE, bringing the total of monoglycosylated vancomycins produced by IVG to 31 [Fu, et al., 2003]. A subsequent study of GtfA, GtfC, and GtfD, found that the high promiscuity of GtfE was not typical and that donor-substrate specificity depends on the acceptor-substrate being used, with an unnatural acceptor-substrate usually reducing the promiscuity possible for the donor-substrate [Oberthur, et al., 2005]. In addition, they reported that the faster enzymes, GtfC and GtfD, with higher turnover rates for their natural pair of substrates, are the most promiscuous. Finally, they concluded that for chemoenzymatically producing unnatural glycovariants of vancomycin, GtfD is the most promising glycosyltransferase because it "showed the most relaxed substrate specificity" [Oberthur, et al., 2005].

The primary reason to study and redesign the glycosyltlransferases involved in the biosynthesis of GPAs is to discover, and/or mass-produce chemoenzymatically, novel GPAs to combat vancomycin-resistant pathogenic microbes. The observation that the lipoglycopeptide teicoplanin (Figure 52) was effective against some vancomycin-resistant microbes motivated research into the structure-activity relationships (SARs) of different acyl chains on GPAs. The current understanding of the mechanisms of vancomycin-resistance in pathogenic microbes, and why some lipoglycopeptides are effective against some vancomycin-resistant microbes, has been reviewed [Kahne, et al., 2005; Yim, et al., 2014]. Briefly, all GPA-producing microbes studied, except one, possess the same *vanHAX* cassette of GPA-resistance genes, expressed only when they are producing their GPA, allowing them to divide and produce offspring that will not be killed by their own GPA. The Van H protein is a D-lactate (D-Lac) dehydrogenase that reduces pyruvate to generate the D-Lac precursors needed to biosynthesize the alternative C-terminus, D-Ala$_4$ -D-Lac$_5$, of the pentapeptide on uncrosslinked PG chains on bacterial lipid II. Van A is a D-Ala-D-Lac ligase, that catlayzes the attachment of D-Lac, instead of D-Ala, to D-Ala$_4$. VanX is a dipeptidase that cleaves D-Ala$_4$ -D-Ala$_5$, while leaving D-Ala$_4$ -D-Lac$_5$ intact. The wide-spread GPA-resistance in enterococci is due to these microbes using this alternative C-terminus, D-Ala$_4$ -D-Lac$_5$, on their uncrosslinked PG chains as a result of a version of these *vanHAX* genes that they appear to have acquired, by horizontal gene transfer, from GPA-producing microbes [Kahne, et al., 2005]. GPAs bind only weakly to this D-Ala$_4$ -D-Lac$_5$, and cannot block cell wall biosynthesis of microbes using it in place of the usual D-Ala$_4$ -D-Ala$_5$. Microbes that possess GPA-resistance genes only express them when their VanS protein “senses” that a GPA is blocking their cell wall biosynthesis, by binding to it. Some lipoglycopeptides, because of their acyl chain, do not bind to VanS, will not trigger the expression of the GPA-resistance genes, and thus can block cell wall biosynthesis.

Several novel lipoglycopeptides, with unnatural combinations of heptapeptide scaffolds, sugars, and acyl chains, have been reported recently, made using combinations of enzymatic and chemical synthesis. Because teicoplanin and other lipoglycopeptides are often found in Nature as mixtures of related compounds that vary only in the length and structure of their acyl chains, two acyltransferases (Atfs) that catalyze acyl chain attachment to an amine substituent of a sugar, were investigated for their substrate promiscuity. As expected, these Atfs were able to use a variety of acyl-donor substrates. Together with the promiscuous vancomycin GTs GtfD and GtfE, several novel lipoglycopeptides were made, starting with teicoplanin. The teicoplanin aglycone heptapeptide core was generated chemically from teicoplanin, and then GtfE as used to attach either glucose, 2-amino glucose, or 6-amino glucose to its phenolic hydroxyl group of the phenylglycine residue at position 4. These three different compounds each served as the acceptor-substrate for the next set of enzymatic reactions, acylation by Atf, with a variety of different acyl-donor substrates. Finally, vancomycin GtfD was used to attach vancosamine to these different unnatural acceptor-substrates, creating novel lipoglycopeptides, with the vancomycin-like disaccharide, acylated, on the teicoplanin heptapeptide scaffold [Kruger, et al., 2005].

To advance the systematic study of the SARs of lipoglycopeptides, a set of eight liponeoglycopeptides, including all four possible N′-decanoylglucopyranose and all four possible N′-biphenoylglucopyranose regioisomers, were synthesized and tested for their potency against a panel of vancomycin-resistant *Enterococci* (VRE) clinical isolates. The

N-decanoyl series of compounds had better activity than the N-biphenoyl series against these VRE, while the glucose C-3′ or C-4′ *N*-acylated compounds gave the best activity against VRE, within each separate series [Griffith, et al., 2007]. It's worth noting that this work uses purely chemical means and comes from the lab of Jon Thorson, one of the leading experts in manipulating GTs, from many different biosynthetic systems, to catalyze unnatural reactions. Perhaps it's also an indication of the difficulty of rationally redesigning GTs to alter one or both of their substrate specificities that two other papers from The Thorson lab include the following thought, almost copied verbatim from each paper. Despite the wealth of GT structural and biochemical information, attempts to alter GT donor/acceptor specificities via rational engineering have been largely unsuccessful [Williams, et al., 2007, Williams, et al., 2008].

The most successful examples of GT redesign are all domain-swap chimeras from different pairs of GTs in the family of GTs with the GT-B global fold. A pair of papers published in 2009, from the same group, reported making chimeric GTs composed of the N-terminal domain of the kanamycin GT KanF, and the C-terminal domain of the vancomycin GT GtfE. These two papers differed mainly in how they chose the linker polypeptide between these two domains. In the first paper they screened a library of chimeric GTs with randomly created linkers [Park, et al., 2009a], while in the second paper they rationally designed a linker by using portions of each

GT's linker region and comparing possible hybrid linkers by structural modeling [Park, et al., 2009b]. In both cases their chimeras catalyzed the transfer of glucose from TDP-glucose, the natural donor-substrate of GtfE, to 2-deoxystreptamine (2-DOS), the natural acceptor-substrate of KanF. The single best chimeric GT from each paper were compared in activity assays and both had increased donor-substrate promiscuity. For using TDP-glucose to glucosylate 2-DOS they both did this better than either wild type KanF or GtfE. For glucosylating 2-DOS, the chimeric GT with the randomly created linker had a conversion rate of about 4-fold greater than that of the chimera with the rationally designed linker [Park, et al., 2009a].

Another paper, published in 2009, reported making and characterizing three different chimeric GTs composed of N- and C-domains from GT-B fold GTs [Truman, et al., 2009]. The first pair of chimeric GTs were made from a pair of glycopeptide GTs using very similar natural substrates: GtfB (acceptor: 4-OH of PheGly$_4$ of vancomycin aglycone (AGV), donor: TDP-glucose) and Orf10* (acceptor: 4-OH of PheGly$_4$ of teicoplanin aglycone (AGT), donor: NDP-GlcNAc). The chimera GtfBH1 was made from the N-terminal domain of GtfB and the C-terminal domain of Orf10*, while the chimera GtfBH2 was made from the other pairing of swapped domains. Activity assays for soluble versions of each of these chimeras confirmed the generally expected roles for the N- and C-domains, with some fine tuning. They found that the N-terminal domain entirely controls the specificity of acceptor-substrate binding, while the C-terminal domain is the major determinant of the donor-substrate specificity, but there are some unresolved portions of the enzyme involved. The third chimera made, called GtfAH1, was composed of the N-terminal domain of GtfA (acceptor: benzylic OH of β-OH-Tyr_6 of desvancosaminyl vancomycin, donor: TDP-epivancosamine), and the C-terminal domain of Orf1 (acceptor: benzylic OH of β-OH-Tyr_6 of teicoplanin glucosaminyl-pseudoaglycone, donor: UDP-GlcNAc). Activity assays of GtfAH1 with pairings of five different acceptor-substrates and UDP-Glc and UDP-GlcNAc as the possible donor-substrates, revealed unexpected activity, promiscuity, with all acceptor-donor pairings considering that the two parent GTs have strict substrate specificity.

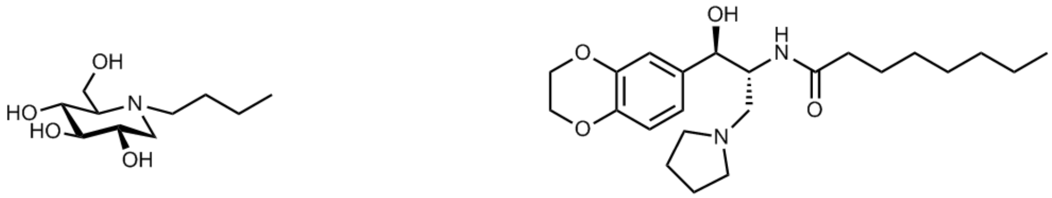
The X-ray crystal structure of GtfAH1, with UDP bound, was solved at 1.15 Å resolution, the highest resolution achieved for a GT-B fold GT. GtfAH1 was found to have the characteristic GT-B bilobal architecture with two facing Rossman-like αβ domains. The GtfAH1:UDP complex was found to be in the closed conformation, as expected after UDP or UDP-sugar has bound in the donor-substrate binding-pocket. This study demonstrated that functional chimeras can be made from closely related GTs [Truman, et al., 2009]. Concerns remain about how well chimeras will behave when made from more distantly related GTs. And, there is the obvious limitation of domain swapping, without redesign, that you're limited to swapping only naturally occurring domains which only have substrate binding pockets for naturally occurring substrates.

Related to the quest to find novel antibiotics to combat VREs are the recent reports of the chemical synthesis of a glycopeptide with a dual-binding variant of the heptapeptide core, such that it can bind tightly to either the dipeptide D-Ala$_4$ -D-Ala$_5$, or the depsipeptide D-Ala$_4$ -D-Lac$_5$ at the C-terminus of the pentapeptide on uncrosslinked PG chains. [Xie, et al., 2011]. By replacing residue 4 (4-hydroxyphenylglycine (Hpg)) in the vancomycin heptapeptide core with the sulfur analogue of Hpg (Tpg), which has a sulfur atom in place of the carbonyl oxygen of Hpg, they created a vancomycin heptapeptide core with residue 4 thioamide-bonded to residue 5. This thoamide is chemically converted to an amidine, putting an NH in place of the residue 4 carbonyl oxygen which can be either an H-bond donor or acceptor, and thus can bind to either the dipeptide D-Ala$_4$ -D-Ala$_5$, or the depsipeptide D-Ala$_4$ -D-Lac$_5$ [Xie, et al., 2012]. Another recent publication reports the chemical synthesis of new lipoglycopeptides with novel acyl chains that are attached, through a quaternary ammonium linker, to the C-terminus of the vancomycin heptapeptide core. These vancomycin analogues have "broad-spectrum antibacterial activity and are about 1000-fold more effective than vancomycin against VRE" [Yarlagadda, et al., 2014].

4.4 Conclusion

In 2014 the Food and Drug Association (FDA) approved 41 new molecular entities, the most in nearly two decades [Jarvis, 2015]. Three of these new drugs are relevant to this chapter. First, there was a GT inhibitor approved by the FDA last year, eliglustat (Cerdelga), for the treatment of Type 1 Gaucher's disease [Shayman, 2010]. Eliglustat, like miglustat (both are shown in Figure 54) is an inhibitor of Glucosylceramide Synthase (GCS, the reaction GCS catalyzes is shown in Figure 27) and is a substrate reduction therapy (SRT) to reduce the toxic accumulation of glucosylceramides in lysosomes by reducing their biosynthesis. Structurally, eliglustat is more similar to ceramide, the acceptor-substrate of GCS, than miglustat, while miglustat is more similar to UDP-glucose, the donor-substrate of GCS. Miglustat inhibits a wide range of glucosidases, which is believed to be why it causes a high rate of gastrointestinal and neurologic side effects, and poor tolerability in clinical studies. Eliglustat has shown fewer side effects in clinical testing [Mistry, et al., 2015].

There were two lipoglycopeptides approved by the FDA in 2014, for the treatment of skin infections [Jarvis, 2015]. Oritavancin (Orbactiv), Figure 55, is a semisynthetic lipophilic derivative of chloroeremomycin (Figure 53), which has had a *p*-chlorophenylbenzyl group chemically added to the nitrogen atom of the epivancosamine in its disaccharide group [Zhanel, et al., 2010].

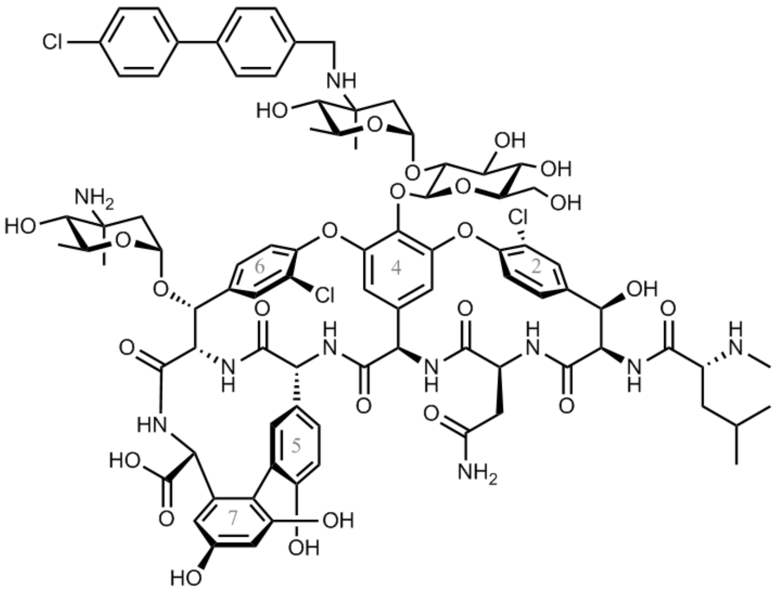


miglustat (Zavesca) eliglustat (Cerdelga)

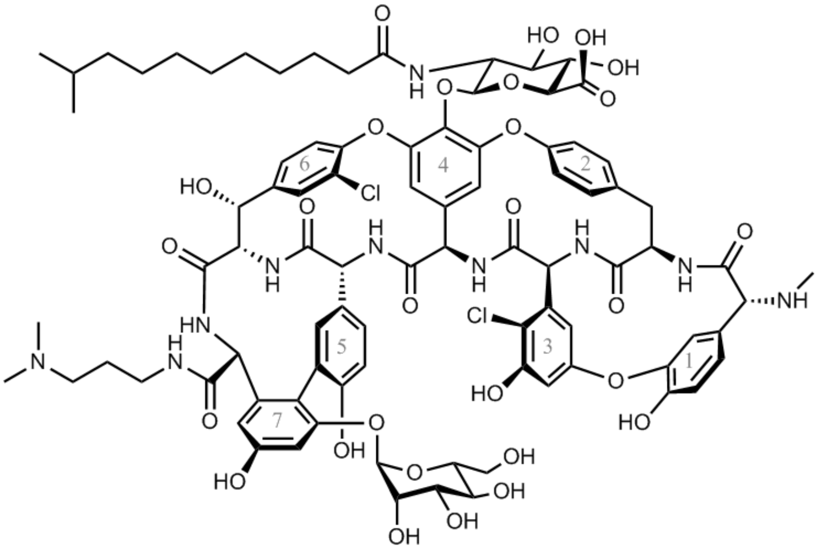
Figure 54: FDA-approved GCS inhibitors for the treatment of Type 1 Gaucher's disease

Dalbavancin (Dalvance), Figure 55, is a semisynthetic derivative of a naturally occurring teicoplanin-subfamily (Figure 52) antibiotic. Dalbavancin is made in three chemical synthetic steps from a naturally produced mixture of lipoglycopeptides with slightly varying acyl chains [Zhanel, et al., 2010].

The recent success of novel variants of naturally occurring glycopeptides that are all chemically produced, without the benefit of enzymes, suggests that there could be many new opportunities now for protein redesigners to create novel enzymes which can be used in chemoenzymatic synthesis of these novel glycopeptides. Large-scale production of any antimicrobial is more efficient, and less expensive, when several of the synthetic steps can be carried out chemoenzymatically. There is certainly more interest now in finding and testing new antimicrobials that can be brought into the



oritavancin (Orbactiv)



dalbavancin (Dalvance)

Figure 55: Lipoglycopeptides approved by the FDA in 2014

clinic. Oritavancin has a surprising history, having been bought and sold three times between when it was developed in the mid-1990s by Eli Lilly and when it finally got FDA approval two decades later [Karaoui, et al., 2013]. There is a market now for new lipoglycopeptide antimicrobials. A few months after Durata Therapeutics received FDA approval for its antibiotic, dalbavancin, it was purchased by Actavis for $675 million [Jarvis, 2015].

5. Conclusions and Future Work

Doing scientific research offers many rewards and challenges, many surprises good and bad. In this final chapter of my dissertation I'll describe a new idea to address an old problem, one of the bad surprises in my thesis research, and a good surprise that resulted from thinking about possible ways to tackle that problem.

5.1 Develop a novel protein solubility enhancement tag

In general, many proteins are difficult to study by solution-phase NMR due to their having low solubility especially at temperatures above 37 °C, such that are required for multidimensional NMR experiments. In a recent review on solubility enhancement tags (SETs) for use in protein NMR studies, by Pei Zhou and Gerhard Wagner [Zhou and Wagner, 2010], the authors state that "an estimated 75% of soluble proteins and many biologically important macromolecules are characterized by low solubility and instability." In particular, I found in 2009 that wt-PheA has serious solubility problems, at temperatures above 25 C. In 2014, while writing my thesis, I came up with an idea for an alternative type of SET which I describe below.

Hypothesis 1: Adding, to a protein of interest (POI), engineered N- and/or C-terminal tails containing one or more recognition sequences for an appropriate wild type or redesigned glycosyltransferase (GT) and then allowing such GT to catalyze the covalent attachment of one or more monosaccharides to the tails of this POI will create a glycan-based SET on this POI which will significantly improve the solubility of this POI without interfering with its function (eg, its enzymatic activity).

The number of glycans in a glycan-based SET on any protein can be controlled by the number of recognition sequences built into the engineered tails, while the size and complexity of each individual glycan added to these tails can be controlled by the choice of GTs used to catalyze the attachment of the first and subsequent monosaccharides to these engineered tails. The size and monosaccharide content of some glycan-based SETs may have some effect on the function of the POI, requiring some empirical guidelines to be developed for what type of glycan-based SET to attach to what type of POI. In general, I expect that relatively small glycans (consisting of only one to three monosaccharides each) will provide sufficient improvement in solubility without interfering with the POI's function.

Such glycan-based SETs offer a large amount of solubility enhancement from a very small tag added to one or both termini of the protein. This is because the GT-recognition sequence can be as short as just three amino acids (eg, for N-linked glycans). Thus a short tail added to the POI can contain several recognition sequences, to each of which will be attached a glycan with tremendous water-binding capacity. The small size of such glycan-based SETs is extremely important for improving the solubility of large proteins to be studied in solution-phase NMR experiments. Increasing the mass of the molecules being studied in an NMR sample will slow down their tumbling rate, which causes the data collected from them to lose resolution due to peak broadening. For this reason, most all peptide-based SETs are not appropriate for very large proteins because they add too much mass to an already massive protein.

In addition to enabling solution-phase NMR studies of proteins with low solubility glycan-based SETs, just like peptide-based SETs, will usually increase the amount of the POI that can be purified after expressing the gene if the glycan-based SET is added co-translationally. This is due to the extra water molecules, that are bound to the SET, decreasing the chances of the protein molecules binding to each other as soon as they are translated and ending up in inclusion bodies. In general, SETs allow more POI to be purified from smaller culture volumes, which reduces the cost of the very expensive growth media that contain expensive NMR-active isotopes.

There are several questions regarding this hypothesis that I have considered and will discuss here.

(1) Is such a glycan-based SET truly a novel idea for solution-phase NMR studies?

My literature searching, though not exhaustive, has not revealed any published reports of this type of SET. This appears to be a novel idea [Pei Zhou, personal communication].

(2) Why has no one yet developed and reported such a glycan-based SET for NMR?

My feeling is that it could just be that no single lab has had simultaneous interests in GTs and NMR, and been struggling with solution-phase NMR of a protein with solubility problems. Perhaps it is because peptide-based SETs are so much easier to use [Pei Zhou, personal communication].

(3) Is such a glycan-based SET likely to work?

In my literature searching I found that for more than 50 years now, protein scientists have been covalently attaching carbohydrates to enzymes to improve their characteristics. In a recent review of this field of neoglycoenzymes [Villalonga, et al., 2014] the authors state that "carbohydrate units confer important physiochemical and biological properties to glycoproteins such as conformational stability, protease resistance, hydrophilicity, charge, aqueous solubility, cell and biomolecular recognition, and reduced immunogenicity."

I have not found any published examples of the use of a glycosyltlransferase (GT) to synthetically glycosylate a protein of interest in order to improve its thermal solubility, its solubility as the temperature is increased. The field of neoglycoenzymes is dominated by examples of the chemical attachment of a non-natural glycan to an enzyme of interest in order to improve its thermal stability, its activity as the temperature is increased. One example was found of the use of an enzyme, though not a GT, to covalently attach a non-natural glycan to an enzyme [Villalonga, et al., 2003] and it will be described below. These synthetically glycosylated enzymes are often used in industrial applications where increased thermal stability will increase the rate of the reactions they catalyze and their product yield before they become inactivated and need to be replaced. To the extent that improvements in an enzyme's thermal stability and improvements in its thermal solubility are both due to its ability to maintain its native conformation at increasing temperatures, the following four representative examples from the literature strongly suggest that glycan-based SETs will work as hypothesized.

An early example of the benefits of covalently attaching non-natural glycans to enzymes was the increased thermal stability of α-amylase when conjugated to dextran polymers with an average molecular mass of 1,000 kDa (approx. 5,500 glucose monomers). This α-amylase-dextran conjugate had an extended half-life at 65 °C, 63 minutes compared to 2.5 minutes for native α-amylase, but only 30% of the activity of native α-amylase. This decrease in activity was most likely due to the α-amylase-dextran conjugate being a crosslinked complex of many α-amylase and dextran molecules, with some of the native enzyme molecules becoming inactivated by the crosslinking [Marshall, 1976].

Invertase from *Saccharomyces cerevisiae* (SInv) is an octameric protein, with molecular mass of 428 kDa, described as a tetramer of dimers [Sainz-Polo, et al., 2013], that exists in two forms, a non-glycosylated intercellular form and a heavily glycosylated secreted form. The glycosylated form of SInv was chemically modified by the covalent attachment of pectin, with average molecular mass of 103 kDa (approx. 550 monomers, mostly galacturonic acid), to its ethylenediamine-activated glycan chains. This pectin-modified SInv showed increased thermal stability, with a half-life at 65 °C of 48 hours compared to 5 minutes for the native SInv [Gomez and Villalonga, 2000]. These authors propose that the addition of pectin to the already glycosylated SInv may improve its thermal stability by reducing aggregation of the pectin-modified SInv by electrostatic repulsion, due to the anionic pectin molecules.

An example of the chemoenzymatic attachment of a non-natural glycan to an enzyme, to improve its thermal stability, used the enzyme transglutaminase (TGase) to covalently attach chemically modified cyclodextrins (CDs) to trypsin [Villalonga, et al., 2003]. TGase was used to catalyze the formation of an isopeptide bond between the primary amine of an amino-CD and the side-chain carboxamide group of one of the nine solvent-accessible glutamines in bovine pancreatic trypsin, molecular mass 23 kDa. Three different trypsin-CD conjugates were synthesized from three different amino-CDs with different numbers of D-glucose units: αCDNH$_2$ (six D-glucose), βCDNH$_2$ (seven D-glucose), and γCDNH$_2$ (eight D-glucose). The three trypsin-CD conjugates synthesized and characterized were determined to, on average, each have 3 CD moieties attached to each trypsin molecule, though which glutamine side chains of trypsin these CDs were bonded to was not reported. While the amino-CDs used did not differ much in size, the three different trypsin-CD conjugates compared showed a difference in their thermal stability when measured at temperatures ranging from 45 °C to 70 °C. This difference was most striking at 45 °C, where their active half-life times were 60 min for native trypsin, 235 min for trypsin- αCDNH$_2$, 151 min for trypsin- βCDNH$_2$, and 123 min for trypsin- γCDNH$_2$ [Villalonga, et al., 2003].

As the above examples illustrate, as time progressed there were more reports of synthetic glycoenzymes made using increasingly smaller non-natural glycans. More recently, a group reported [Sola and Griebenow, 2006] chemically glycosylating the side-chain amines of some of the 14 surface accessible lysine residues of α-chymotrypsin (α-CT) with either the disaccharide Lactose or a Dextran with average molecular mass of 10 kDa (approx. 55 glucose monomers). They produced two series of α-CT conjugates, with the average number of non-natural glycans per molecule of α-CT varying from about 2 to about 8, and compared them with native α-CT for their molecular dynamics, thermal stability, and catalytic function. They found that, in general, for either the α-CT-Lactose series or the α-CT-Dextran series, as the average number of glycans per α-CT molecule increased, the structural dynamics of the α-CT conjugate and its catalytic efficiency both decreased, while the thermal stability of the α-CT conjugate increased [Sola and Griebenow, 2006]. The size of the non-natural glycan appeared to have less effect than the quantity of those glycans. The thermal stability of α-CT-Lactose, with 7.4 Lactose per α-CT, had increased by 7 °C, while the thermal stability of α-CT-Dextran, with 7.6 Dextran per α-CT, had increased by 8 °C.

(4) Have any glycosyltransferases (GTs) been published that have desirable qualities for enzymatically attaching such glycan-based SETs to an engineered tail on a POI ?

Many GTs are naturally membrane proteins and thus are difficult to work with *in vitro* due to their transmembrane domain(s). Thus, it caught my attention when I read of a soluble GT in a recent review article. Schwarz, et al., reported [Schwarz, et al., 2011] enzymatic characterization of a cytoplasmic bacterial GT, specifically an N-glycosyltransferase (NGT), and their discovery of another GT, called α6GlcT, in the same bacteria which adds glucoses to an N-linked glucose. They report that both of these GTs use UDP-glucose as their sugar donor substrate and that the NGT uses as its acceptor substrate the side-chain Nitrogen of asparagine (Asn, N) residues in the consensus recognition triplet of NX(S/T), where X is any amino acid except proline. Working together NGT and α6GlcT, in the presence of UDP-glucose, will covalently attach a chain of three to seven glucose units to the side chain of an Asn (N) residue, in a recognition triplet of NX(S/T), as shown in Figure 56.

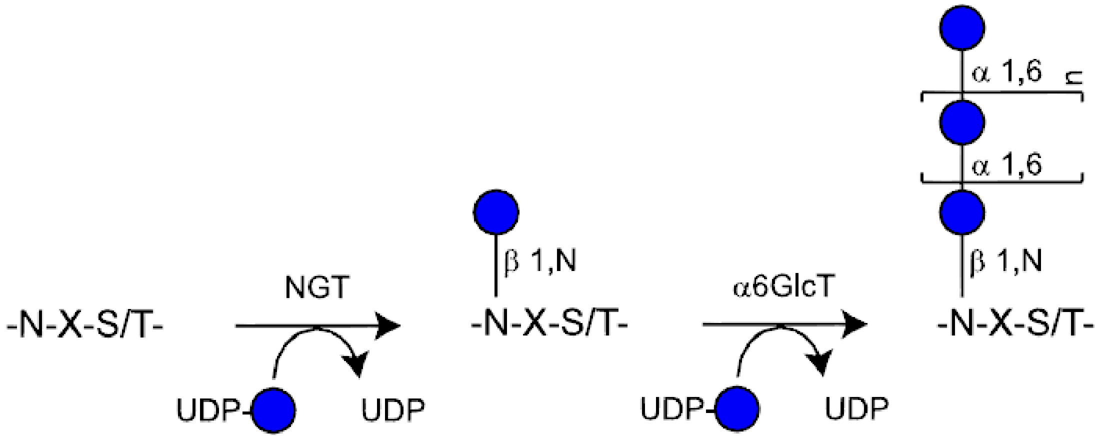


Figure 56: Combined chemoenzymatic result of NGT and α6GlcT, in the presence of UDP-glucose (this is Figure 5, reprinted from Schwarz, et al., 2011) The number of glucose units added by α6GlcT depends on the concentration of UDP-glucose.

5.2 Use wild type PheA as a model system of a large protein with low solubility

If we could obtain this pair of GTs, NGT and α6GlcT, or something similar, then we would only need a test protein with low solubility for measuring the solubility enhancements due to different glycan-based SETs. Maybe the low solubility of wild type PheA at temperatures above 25 C could be used to our advantage, by using it to test the effects of glycan-based SETs of different sizes, composition, and location.

Hypothesis 2: Wild type PheA is a good model enzyme, with low solubility, for studying the effects of prototypes of a new type of SET consisting of glycans of varying sizes attached to the side-chains of Asn residues in an engineered tail added to either its N- or C-terminus.

Assuming for the moment that we could obtain an N-glucosylating GT, like NGT, raised the question in my mind of whether there were any Asn residues in NX(S/T) sequences in wt PheA, before any tail is added that contains such recognition consensus sequences. I was wondering if an N-glucosylating GT would add glucose residues to PheA and break it, by making it unable to fold into its native conformation. I checked the published sequence of wtPheA in the PDB (ID: 1amu) and found that there are three NX(S/T) sequences in wtPheA. The three Asn residues that I found in consensus recognition triplets had these positions: 3, 16, and 219, and were in these amino acid triplets: NSS, NGT, and NVT, respectively. I then looked into where these three Asn residues were in the 3D structure of wtPheA. I found that N3 and N16 are in the unordered N-terminal tail, while N219 is in a linker between a helix and a strand in the X-ray crystal structure of wtPheA published in 1997 [Conti, et al., 1997], (PDB ID: 1amu). The locations of these three Asn residues in wtPheA are shown in Figure 57. It seemed likely that attaching a glucose to any or all of these three Asn residues would not do any harm to the folding of wtPheA. Then, this finding raised the following interesting possibility.

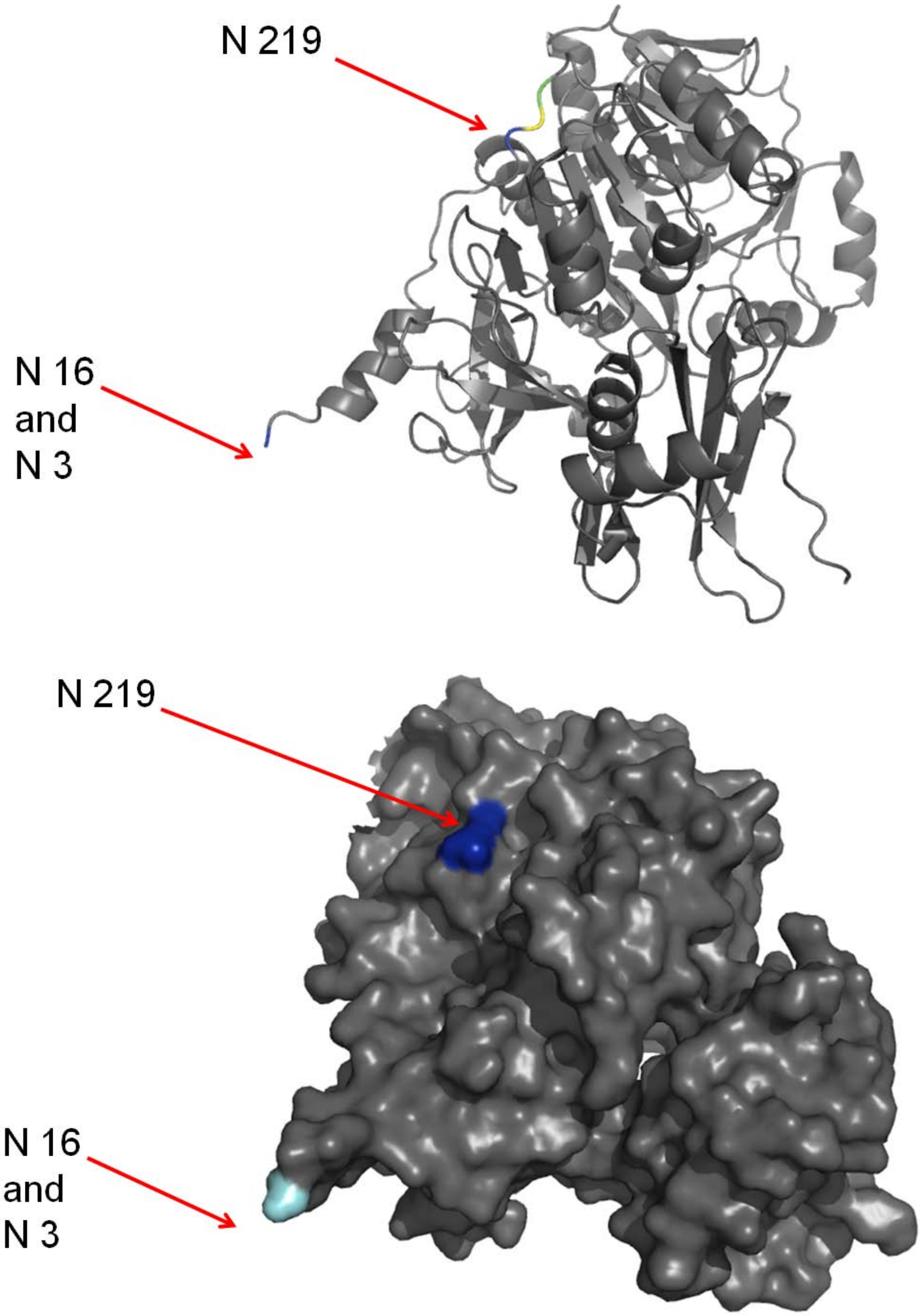

N 219
N 16
and
N 3
N 219
N 16
and
N 3

Figure 57: The three N-X-S/T sequences in wt PheA, shown in a ribbon diagram (top) and in a surface model (bottom) of the crystal structure of wtPheA (PDB ID = 1AMU). The unordered N-terminal tail is not shown beyond G17. Drawn with PyMOL (edu, 1.3).

Hypothesis 2.b: Wild type PheA is actually a glycoprotein, when it is in its natural environment, inside a *Brevibacillus brevis* cell.

I found some evidence that supports this hypothesis. I found that the genome of *Brevibacillus brevis* had been published in late 2012 [Chen, et al., 2012] and that it had been processed into open reading frames (ORFs) and those ORFs identified as encoding genes for Carbohydrate Active enzymes have been entered into the CAZy database (www.cazy.org). In the CAZy database I found that the *Brevibacillus brevis* genome has 75 ORFs that are in four different enzyme classes in the CAZy database. Of these 75 putative genes, there are 14 carbohydrate esterases, 20 glycoside hydrolases, and 40 glycosyltransferases. This doesn't prove that wt PheA is naturally a glycoprotein, but it sure doesn't disprove it either. Now, it's interesting to wonder if some, most, or all, other NRPS enzymes are actually glycoproteins and if ignoring that fact could explain some of the difficulties people have had with attempts to reengineer them, such as domain swapping experiments. Also, if wt PheA is a glycoprotein, it very likely would explain the solubility problems that I had with it. And, of the 75% of soluble proteins which have low solubility [Zhou and Wagner, 2010], how many of those low-solubility proteins are actually glycoproteins in their natural environments? If wt PheA is actually a glycoprotein it could still be used as a test protein for developing glycan-based SETs, by just mutating each of the three N-linked glycan recognition sequences, so those three Asn residues are either replaced or not in a consensus triplet anymore.

5.3 Use set of wild type and redesigned enzymes to study the effects of engineered glycosylation on their enzymatic activities

Hypothesis 3.a: The size and composition of glycan-based SETs, on engineered tails, can affect the altered function that is given to a protein by redesigning the wild type version of it.

With protocols developed to be able to attach glycan-based SETs to proteins, it would be very interesting to explore how such SETs affect the new functions of redesigned enzymes. The Donald Lab's set of mutant PheA could be used to study the effects of glycan-based SETs, on engineered tails, of varying composition, on the new enzyme activities of redesigned versions of PheA. Such data from the wet lab would assist protein-design algorithm designers to move towards developing glycan-based-SET design algorithms.

Hypothesis 3.b: The size and composition of glycans at natural N-linked glycosylation sites can affect the altered function that is given to a protein by redesigning the wild type version of it.

With protocols developed to be able to control the attachment of glycans to proteins, it would be very interesting to explore how differing glycans affect the new functions of redesigned enzymes. The Donald Lab's set of mutant PheA could be used to study the effects of glycans at natural N-linked glycosylation sites, of varying composition, on the new enzyme activities of redesigned versions of PheA. Such data from the wet lab would assist protein-design algorithm designers to move towards developing **glyco**protein design algorithms.

References

Adamczyk, B.,, T. Tharmalingam, and P. M. Rudd (2012) Glycans as Cancer Biomarkers, *Biochimica et Biophysica Acta*, **1820**, 1347-1353.

Alfaro, J. A.,, R. B. Zheng, M. Persson, J. A. Letts, R. Polakowski, Y. Bai, S. N. Borisova, N. O. L. Seto, T. L. Lowary, M. M. Palcic, and S. V. Evans (2008) ABO(H) Blood Group A and B Glycosyltransferases Recognize Substrate via Specific Conformational Changes, *J. Biol. Chem.*, **283**, 10097-10108.

Amann, F.,, C. Schaub, B. Muller, and R. R. Schmidt (1998) New Potent Sialyltransferase Inhibitors – Synthesis of Donor and of Transition-State Analogues of Sialyl Donor CMP-Neu5Ac, *Chem. Eur. J.*, **4**, 1106-1115.

American Academy of Microbiology (2014) Human Microbiome, *American Society of Microbiology FAQ*

An International Commission (1978) Ebola Haemorrhagic Fever in Zaire, 1976, *Bulletin of the World Health Organization* **56**, 271-293.

Austin, C.P. (2004) The Impact of the Completed Human Genome Sequence on the Development of Novel Therapeutics for Human Disease, *Annu. Rev. Med.* **55**, 1-13.

Ayar-Kayali, H. (2011) Pentose phosphate pathway flux analysis for glycopeptides antibiotic vancomycin production during glucose-limited cultivation of *Amycolatopsis orientalis*, *Preparative Biochemistry and Biotechnology*, **41**, 94-105.

Baize, S.,, et al. (2014) Emergence of Zaire Ebola Virus Disease in Guinea – Preliminary Report, *New England J. Medicine*, **xx**, DOI: 10.1056/NEJMoa1404505, pub. ahead of print

Barton, W. A.,, J. Lesniak, J. B. Biggins, P. D. Jeffrey, J. Jiang, K. R. Rajashankar, J. S. Thorson, and D. B. Nikolov (2001) Structure, mechanism and engineering of a nucleotidyltransferase as a first step toward glycorandomization, *Nature Structural Biology*, **8**, 545-551.

Batchelor, F. R.,, F. P. Doyle, J. H. C. Nayler, and G. N. Rolinson (1959) Synthesis of Penicillin : 6-Aminopenicillanic Acid in Penicillin Fermentations, *Nature*, **183**, 257-258.

Bentley, R. (2009) Different roads to discovery; Prontosil (hence sulfa drugs) and Penicillin (hence β-lactams), *J. Ind. Microbiol. Biotechnol.*, **36**, 775-786.

Berg, P., and M. F. Singer (1995) The recombinant DNA controversy: Twenty years later, *Proc. Nat. Acad. Sci. USA*, **92**, 9011-9013.

Bibb, M. J.,, J. M. Ward, and D. A. Hopwood (1978) Transformation of plasmid DNA into *Streptomyces* at high frequency, *Nature*, **274**, 398-400.

Bio-Rad, http://www.bio-rad.com/webroot/web/pdf/lsr/literature/LIT156.pdf. (Bio-Rad) Activated Immunoaffinity Supports: Affi-Gel 10, Affi-Gel 15.

Blake, C. C. F.,, D. F. Koenig, G. A. Mair, A. C. T. North, D. C. Phillips, and V. R. Sarma (1965) Structure of Hen Egg-White Lysozyme – A Three-Dimensional Fourier Synthesis at 2 Angstrom Resolution, *Nature*, **206**, 757-761.

Bouhss, A.,, A. E. Trunkfield, T. D. H. Bugg, and D. Mengin-Lecreulx (2008) The biosynthesis of peptidoglycan lipid-linked intermediates, *FEMS Microbiol. Rev.*, **32**, 208-233.

Brandish, P. E.,, K-I. Kimura, M. Inukai, R. Southgate, J. T. Lonsdale, and T. D. H. Bugg (1996) Modes of Action of Tunicamycin, Liposidomycin B, and Mureidomycin A: Inhibition of Phospho-*N*- Acetylmuramyl-Pentapeptide Translocase from *Escherichia coli*, *Antimicrob. Agents Chemother.*, **40**, 1640-1644.

Brandl, E., and H. Margreiter (1954) Ein saurestabil biosynthetisches Penicillin, *Osterreichische Chemiker-Zeitung* , **55**, 11. This is reference #2 from G. N. Rolinson and A. M. Geddes (2007) The 50th anniversary of the discovery of 6-

amino penicillanic acid (6-APA), *International Journal of Antimicrobial Agents*, **29**, 3-8.

Breton, C.,, S. Fournel-Gigleux, and M. M. Palcic (2012) Recent Structures, Evolution and Mechanisms of Glycosyltransferases, *Curr. Opin. Struct. Biol.*, **22**, 540-549.

Brown, J. R.,, F. Yang, A. Sinha, B. Ramakrishnan, Y. Tor, P. K. Qasba, and J. D. Esko (2009) Deoxygenated Disaccharide Analogs as Specific Inhibitors of β1-4 Galactosyltransferase 1 and Selectin-mediated Tumor Metastasis, *J. Biol. Chem.*, **284**, 4952-4959.

CDC, *http://www.cdc.gov/vhf/ebola/outbreaks/guinea/ 2014 Ebola Outbreak in West Africa (CDC)*

Chain, E.,, H.W. Florey, A.D. Gardner, N.G. Heatley, M.A. Jennings, J. Orr-Ewing, and A.G. Sanders (1940) Penicillin as a Chemotherapeutic Agent, *The Lancet*, **2***, 226-228. * Volume # **was 2 (**in 1940**) (is 236** now**)**

Chang, K-H.,, L. Lee, J. Chen, and W-S. Li (2006) Lithocholic acid analogues, new and potent α-2,3-sialyltransferase inhibitors, *Chem. Commun*, 2006, 629-631.

Chen, C-Y.,, I. Georgiev, A. C. Anderson, and B. R. Donald (2009) Computational Structure-Based Redesign of Enzyme Activity, *Proc. Natl. Acad. Sci.* **106**, 3764-3769.

Chen, V. B.,, W. B. Arendall III, J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson (2010) *MolProbity*: all-atom structure validation for macromolecular crystallography, *Acta Crystallographica, D*, **66**, 12-21.

Chen, J-Y.,, Y-A. Tang, S-M. Huang, H-F. Juan, L-W. Wu, Y-C. Sun, S-C. Wang, K-W. Wu, G. Balraj, T-T. Chang, W-S. Li, H-C. Cheng, and Y-C. Wang (2011) A Novel Sialyltransferase Inhibitor Suppresses FAK/Paxillin Signaling and Cancer Angiogenesis and Metastasis Pathways, *Cancer Res.*, **71**, 473-483.

Chen, W.,, Y. Wang, D. Li, L. Li, Q. Xiao, and Q. Zhou (2012) Draft Genome Sequence of *Brevibacillus brevis* Strain X23, a Biocontrol Agent against Bacterial Wilt, *J. Bacteriol*. **194**, 6634-6635.

Chiang, C-H.,, C-H. Wang, H-C. Chang, S. V. More, W-S. Li, and W-C. Hung (2010) A Novel Sialyltransferase Inhibitor AL10 Suppresses Invasion and Metastasis of Lung Cancer Cells by Inhibiting Integrin-Mediated Signaling, *J. Cell Physiol.*, **223**, 492-499.

Clarke, H. T.,, J. R. Johnson, and R. Robinson (eds) (1949) The Chemistry of Penicillin, Princeton University Press. This is reference #2 from: R. Curtis and J. Jones (2007) Robert Robinson and penicillin – An unnoticed document in the saga of its structure, *Journal of Peptide Science*, **13**, 769-775.

Conti, E.,, T. Stachelhaus, M. A. Marahiel, and P. Brick (1997) Structural Basis For the Activation of Phenylalanine in the Nonribosomal Biosynthesis of Gramicidin S, *EMBO J.* **16**, 4174-4183.

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

Curtis, R., and J. Jones (2007) Robert Robinson and penicillin – An unnoticed document in the saga of its structure, *Journal of Peptide Science*, **13**, 769-775.

Dall'Olio, F., and M. Chiricolo (2001) Sialyltransferases in Cancer, *Glycoconjugate Journal*, **18**, 841-850.

Desmet, J.,, M. De Maeyer, B. Hazes, and I. Lasters (1992) The Dead-End Elimination Theorem and Its Use in Protein Side-Chain Positioning, *Nature* **356**, 539-542.

Doekel, S.,, M-F. Coeffet-Le Gal, J-Q. Gu, M. Chu, R. H. Baltz, and P. Brian (2008) Non-ribosomal Peptide Synthetase Module Fusions to Produce Derivatives of Daptomycin in Streptomyces roseosporus, *Microbiology* **154**, 2872-2880.

Dubos, R. J. (1939) Studies on a Bactericidal Agent Extracted from a Soil Bacillus, I. Preparation of the Agent. Its Activity *in Vitro*, *J. Exp. Med.*, **70**, 1-10.

Dubos, R.J. (1941) Bacteriostatic and Bactericidal Agents Obtained from Saprophytic Microorganisms, *J. Pediatrics*, **19** 588-595.

Emmerich, R., and O. Loew (1899) (Title unknown, Paper unavailable) *Zentr. Bakt.* **26**, 237. This is reference #12 from (my ref. B.1-12): R.J. Dubos (1941) Bacteriostatic and Bactericidal Agents Obtained from Saprophytic Microorganisms, *J. Pediatrics*, **19** 588-595.

Eppelmann, K.,, T. Stachelhaus, and M. A. Marahiel (2002) Exploitation of the Selectivity-Conferring Code of the Nonribosomal Peptide Synthetases for the

Rational Design of Novel Peptide Antibiotics, *Biochemistry* **41**, 9718-9726.

Fischbach, M. A., and C. T. Walsh (2006) Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms, *Chem. Rev.* **106**, 3468-3496.

Fleming, A. (1922) On a Remarkable Bacteriolytic Element Found in Tissues and Secretions, *Proc. Royal Society B*, **93**, 306-317.

Fleming, A. (1929) On the Antibacterial Action of Cultures of a Penicillium, With Special Reference to Their Use in the Isolation of *B. Influenzae*, *Br. J. Exper. Pathol*., **10**, 226-236.

Fleming, A. (1945a) Penicillin*, Nobel Lecture, Prize in Physiology or Medicine*

Fleming, A. (1945b) Banquet Speech, *Nobel Prize in Physiology or Medicine*

Fleming, A., and V.D. Allison (1922) Further Observations on a Bacteriolytic Element Found in Tissues and Secretions, *Proc. Royal Society B*, **94**, 142-151.

Folena-Wasserman, G.,, R. D. Sitrin, F. Chapin, and K. M. Snader (1987) Affinity Chromatography of Glycopeptide Antibiotics, *Journal of Chromatography*, **392**, 225-238.

Franklin, R. E., and R. G. Gosling (1953) Molecular Structure of Nucleic Acids: Molecular Configuration in Sodium Thymonucleate, *Nature* **171**, 740-741.

Fu, X.,, C. Albermann, J. Jiang, J. Liao, C. Zhang, and J. S. Thorson (2003) Antibiotic optimization via *in vitro* glycorandomization, *Nature Biotechnology*, **21**, 1467-1469.

Gainza, P.,, K. E. Roberts, I. Georgiev, R. H. Lilien, D. A. Keedy, C-Y. Chen, F. Reza, A. C. Anderson, D. C. Richardson, J. S. Richardson, and B. R. Donald (2013) OSPREY: Protein Design with Ensembles, Flexibility, and Provable Algorithms, *Methods in Enzymology* **523**, 87-107.

Georgiev, I.,, R. H. Lilien, and B. R. Donald (2006) Improved Pruning Algorithms and Divide-and-Conquer Strategies for Dead-End Elimination, with Application to Protein Design, *Bioinformatics* **22**, e174-e183.

Georgiev, I.,, and B. R. Donald (2007) Dead-End Elimination with Backbone Flexibility, *Bioinformatics* **23**, i185-i194.

Georgiev, I.,, D. Keedy, J. S. Richardson, D. C. Richardson, and B. R. Donald (2008a) Algorithm for Backrub Motions in Protein Design, *Bioinformatics* **24**, i196-i204.

Georgiev, I.,, R. H. Lilien, and B. R. Donald (2008b) The Minimized Dead-End Elimination Criterion and Its Application to Protein Redesign in a Hybrid Scoring and Search Algorithm for Computing Partition Functions Over Molecular Ensembles, *J. Comput. Chem.* **29**, 1527-1542.

Gloster, T. M., and D. J. Vocadlo (2012) Developing Inhibitors of Glycan Processing Enzymes as Tools for Enabling Glycobiology, *Nature Chemical Biology*, **8**, 683-694.

Gold, H. S., and R. C. Moellering (1996) Antimicrobial-Drug Resistance, *New Engl. J. Med.* **335**, 1445-1453.

Gomez, L., and R. Villalonga (2000) Functional Stabilization of Invertase by Covalent Modification With Pectin, *Biotechnology Letters* **22**, 1191-1195.

Griffith, B. R.,, C. Krepel, X. Fu, S. Blanchard, A. Ahmed, C. E. Edmiston, and J. S. Thorson (2007) Model for Antibiotic Optimization via Neoglycosylation: Synthesis of Liponeoglycopeptides Active against VRE, *J. Am. Chem. Soc.*, **129**, 8150-8155.

Groves, P.,, M. S. Searle, J. P. Mackay, and D. H. Williams (1994) The Structure of an Asymmetric Dimer Relevant to the Mode of Action of the Gglycopeptide Antibiotics, *Structure*, **2**, 747-754.

Harduin-Lepers, A.,, R. Mollicone, P. Delannoy, and R. Oriol (2005) The animal sialyltransferases and sialyltransferase-related genes: a phylogenetic approach, *Glycobiology*, **15**, 805-817.

Harris, C. M.,, H. Kopecka, and T. M. Harris (1983) Vancomycin: Structure and Transformation to CDP-I, *J. Am. Chem. Soc.* **105**, 6915-6922.

Hegde, S. S., and J. W. Janc (2014) Efficacy of Telavancin, a Lipoglycopeptide Antibiotic, in Experimental Models of Gram-Positive Infection, *Expert Rev. Anti Infect. Ther.*, **12**, 1463-1475.

Higgins, D. L.,, R. Chang, D. V. Debabov, J. Leung, T. Wu, K. M. Krause, E. Sandvik, J. M. Hubbard, K. Kaniga, D. E. Schmidt Jr., Q. Gao, R. T. Cass, D. E. Karr, B. M. Benton, and P. P. Humphrey (2005) Telavancin, a Multifunctional

Lipoglycopeptide, Disrupts both Cell Wall Synthesis and Cell Membrane Integrity in Methicillin-Resistant *Staphylococcus aureus*, *Antimicrob. Agents Chemother.*, **49**, 1127-1134.

Hosoguchi, K.,, T. Maeda, J-i. Furukawa, Y. Shinohara, J. Hinou, M. Sekiguchi, J. Togame, H. Takemoto, H. Kondo, and S-I. Nishimura (2010) An Efficient Approach to the Discovery of Potent Inhibitors against Galactosyltransferases, *J. Med. Chem.*, **53**, 5607-5619.

Hotchkiss, R. D., and R. J. Dubos (1940a) Fractionation of the Bactericidal Agent From Cultures of a Soil Bacillus, *J. Biol. Chem.*, **132**, 791-792.

Hotchkiss, R. D., and R. J. Dubos (1940b) Chemical Properties of Bactericidal Substances Isolated From Cultures of a Soil Bacillus, *J. Biol. Chem.*, **132**, 793-794.

Hubbard, B. K., and C. T. Walsh (2003) Vancomycin Assembly: Nature's Way, *Angew. Chem. Int. Ed.*, **42**, 730-765.

Jankovic, M. (2011) Glycans as Biomarkers: Status and Perspectives, *J. Med. Biochem.*, **30**, 213-223.

Jarvis, L. M. (2015) The Year in New Drugs: FDA's Drug Approvals in 2014 Were Remarkable for Both Their Quantity and Quality, *C&EN*, **93**, 11-16.

Johal, A. R.,, R. J. Blackler, J. A. Alfaro, B. Schuman, S. Borisova, and S. V. Evans (2014) pH-induced conformational changes in human ABO(H) blood group glycosyltlransferases confirm the importance of electrostatic interactions in the formation of the semi-closed state, *Glycobiology*, **24**, 237-246.

Johnson, L. N., and D. C. Phillips (1965) Structure of Some Crystalline Lysozyme-Inhibitor Complexes Determined by X-ray Analysis at 6 Angstrom Resolution, *Nature*, **206**, 761-763.

Jung, H-M.,, J-Y. Kim, H-J. Moon, D-K. Oh, and J-K. Lee (2007) Optimization of culture conditions and scale-up to pilot and plant scales for vancomycin production by *Amycolatopsis orientalis*, *Appl. Microbiol. Biotechnol*, **77**, 789-795.

Kahne, D.,, C. Leimkuhler, W. Lu, and C. Walsh (2005) Glycopeptide and Lipoglycopeptide Antibiotics, *Chem. Rev.*, **105**, 425-448.

Karaoui, L. R.,, R. El-Lababidi, and E. B. Chahine (2013) Oritavancin: An investigational lipoglycopeptide antibiotic, *Am. J. Health-Syst. Pharm.*, **70**, 23-33.

Kendrew, J. C.,, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips (1958) A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis, *Nature* **181**, 662-666.

Kendrew, J. C.,, R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips, and V. C. Shore (1960) Structure of Myoglobin - A Three-Dimensional Fourier Synthesis at 2 Angstrom Resolution, *Nature*, **185**, 422-427.

Kominato, Y.,, P. D. McNeill, M. Yamamoto, M. Russell, S. Hakomori and F. Yamamoto (1992) Animal Histo-blood Group ABO Genes, *Biochem. Biophys. Res. Commun.*, **189**, 154-164.

Kopp, F., and M. A. Marahiel (2007) Macrocyclization Strategies in Polyketide and Nonribosomal Peptide Biosynthesis, *Nat. Prod. Rep.* **24**, 735-749.

Kruger, R. G.,, W. Lu, M. Oberthur, J. Tao, D. Kahne, and C. T. Walsh (2005) Tailoring of Glycopeptide Scaffolds by the Acyltransferases from the Teicoplanin and A-40,926 Biosynthetic Operons, *Chemistry & Biology*, **12**, 131-140.

Kuroki, R.,, L. H. Weaver and B. W. Matthews (1995) Structure-based design of a lysozyme with altered catalyltic activity, *Nature Structural Biology*, **2**, 1007-1011.

Laferte, S.,, N. W. C. Chan, K. Sujino, T. L. Lowary, and M. M. Palcic (2000) Intracellular inhibition of blood group A glycosyltransferase, *Eur. J. Biochem.*, **267**, 4840-4849.

Lairson, L. L.,, B. Henrissat, G. J. Davies, and S. G. Withers (2008) Glycosyltransferases: Structures, Functions, and Mechanisms, *Annu. Rev. Biochem.*, **77**, 521-555.

Lal, R.,, S. Lal, E. Grund and R. Eichenlaub (1991) Construction of a Hybrid Plasmid Capable of Replication in *Amycolatopsis mediterranei*, *Appl. Environ. Microbiol.*, **57**, 665-671.

Laudano, J. B. (2011) Ceftaroline fosamil: A new broad-spectrum cephalosporin, *J.*

Antimicrob. Chemother., **66**, iii11-iii18.

Leach, A. R., and A. P. Lemon (1998) Exploring the Conformational Space of Protein Side Chains Using Dead-End Elimination and the A* Algorithm, *Proteins* **33**, 227-239.

Leadbetter, M. R.,, S. M. Adams, B. Bazzini, P. R. Fatheree, D. E. Karr, K. M. Krause, B. M. T. Lam, M. S. Linsell, M. B. Nodwell, J. L. Pace, K. Quast, J-P. Shaw, E. Soriano, S. G. Trapp, J. D. Villena, T. X. Wu, B. G. Christensen, and J. K. Judice (2004) Hydrophobic Vancomycin Derivatives with Improved ADME Properties: Discovery of Telavancin (TD-6424), *J. Antibiotics*, **57**, 326-336.

Lee, K-Y.,, H. G. Kim, M. R. Hwang, J. I. Chae, J. M. Yang, Y. C. Lee, Y. K. Choo, J. I. Lee, S-S. Lee, and S-I. Do (2002) The Hexapeptide Inhibitor of Galβ1,3GalNAc-specific α2,3-Sialyltransferase as a Generic Inhibitor of Sialyltransferases, *J. Biol. Chem.*, **277**, 49341-49351.

Levine, D. P. (2006) Vancomycin: A History, *Clinical Infectious Diseases*, **42**, s5-s12.

Lightman, A. (2005) Antibiotics, *The Discoveries: Great Breakthroughs in 20th-Century Science*, 253-277, Vintage Books, New York.

Lilien, R. H.,, B. W. Stevens, A. C. Anderson, and B. R. Donald (2005) A Novel Ensemble-Based Scoring and Search Algorithm for Protein Redesign and Its Application to Modify the Substrate Specificity of the Gramicidin Synthetase A Phenylalanine Adenylation Enzyme, *J. Comp. Biol.* **12**, 740-761.

Linne, U.,, D. B. Stein, H. D. Mootz, and M. A. Marahiel (2003) Systematic and Quantitative Analysis of Protein-Protein Recognition Between Nonribosomal Peptide Synthetases Investigated in the Tyrocidine Biosynthetic Template, *Biochemistry* **42**, 5114-5124.

Losey, H. C..,, M. W. Peczuh, Z. Chen, U. S. Eggert, S. D. Dong, I. Pelczer, D. Kahne, and C. T. Walsh (2001) Tandem Action of Glycosyltransferases in the Maturation of Vancomycin and Teicoplanin Aglycones: Novel Glycopeptides, *Biochemistry*, **40**, 4745-4755.

Losey, H. C.,, J. Jiang, J. B. Biggins, M. Oberthur, X-Y. Ye, S. D. Dong, D. Kahne, J. S. Thorson, and C. T. Walsh (2002) Incorporation of Glucose Analogs by GtfE and GtfD from the Vancomycin Biosynthetic Pathway to Generate Variant Glycopeptide, *Chem. Biol.*, **9**, 1305-1314.

Lowary, T., and O. Hindsgaul (1993) Recognition of synthetic deoxy and deoxyfluoro analogs of the acceptor α-L-Fuc*p*-(1→2)-β-D-Gal*p*-OR by the blood-group A and B gene-specified glycosyltransferases, *Carbohydr. Res.*, **249**, 163-195.

Lowary, T., and O. Hindsgaul (1994) Recognition of synthetic *O*-methyl, epimeric, and amino analogues of the acceptor α-L-Fuc*p*-(1→2)-β-D-Gal*p*-OR by the blood-group A and B gene-specified glycosyltransferases, *Carbohydr. Res.*, **251**, 33-67.

Lowe, J. B. (1993) Chapter 7, The Blood Group-Specific Human Glycosyltransferases, *Bailliere's Clinical Haematology*, **6**, 465-492.

Madon, J., and R. Hutter (1991) Transformation System for *Amycolatopsis* (*Nocardia*) *mediterranei*: Direct Transformation of Mycelium with Plasmid DNA, *Journal of Bacteriology*, **173**, 6325-6331.

Marahiel, M. A.,, T. Stachelhaus, and H. D. Mootz (1997) Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis, *Chem. Rev.* **97**, 2651-2673.

Marcus, S. L.,, R. Polakowski, N. O. L. Seto, E. Leinala, S. Borisova, A. Blancher, F. Roubinet, S. V. Evans, and M. M. Palcic (2003) A Single Point Mutation Reverses the Donor Specificity of Human Blood Group B-synthesizing Galactosyltransferase, *J. Biol. Chem.*, **278**, 12403-12405.

Marshall, J. J. (1976) Preparation and Characterization of a Dextran-Amylase Conjugate, *Carbohydrate Research* **49**, 389-398.

Matsushima, P.,, M. A. McHenney, and R. H. Baltz (1987) Efficient Transformation of *Amycolatopsis orientalis* (*Nocardia orientalis*) Protoplasts by *Streptomyces* Plasmids, *Journal of Bacteriology*, **169**, 2298-2300.

Matthews, B. W., and S. J. Remington (1974) The Three Dimensional Structure of the Lysozyme from Bacteriophage T4, *Proc. Nat. Acad. Sci. USA*, **71**, 4178-4182.

McGuire, J. M.,, R. N. Wolfe, and D. W. Ziegler (1955-1956) Vancomycin, a new antibiotic. II. *In vitro* antibacterial studies, *Antibiotics Annual*, **3**, 612-618. (unavailable through Duke) This is reference #4 from: D. P. Levine (2006) Vancomycin: A History, *Clinical Infectious Diseases*, **42**, s5-s12.

McIntyre, J. T.,, A. T. Bull, and A. W. Bunch (1996) Vancomycin Production in Batch and Continuous Culture, *Biotechnology and Bioengineering*, **49**, 412-420.

Mistry, P. K.,, E. Lukina, H. B. Turkia, D. Amato, H. Baris, M. Dasouki, M. Ghosn, A. Mehta, S. Packman, G. Pastores, M. Petakov, S. Assouline, M. Balwani, S. Danda, E. Hadjiev, A. Ortega, S. Shankar, M. H. Solano, L. Ross, J. Angell, and J. Peterschmitt (2015) Effect of Oral Eliglustat on Splenomegaly in Patients With Gaucher Disease Type 1, *J. Am. Med. Assoc.*, **313**, 695-706.

Mulichak, A. M.,, H. C. Losey, C. T. Walsh, and R. M. Garavito (2001) Structure of the UDP-Glucosyltransferase GtfB That Modifies the Heptapeptide Aglycone in the Biosynthesis of Vancomycin Group Antibiotics, *Structure*, **9**, 547-557.

Mulichak, A. M.,, H. C. Losey, W. Lu, Z. Wawrzak, C. T. Walsh, and R. M. Garavito (2003) Structure of the TDP-*epi*-vancosaminyltransferase GtfA from the Chloroeremomycin Biosynthetic Pathway, *Proc. Natl. Acad. Sci. USA*, **100**, 9238-9243.

Mulichak, A. M.,, W. Lu, H. C. Losey, C. T. Walsh, and R. M. Garavito (2004) Crystal Structure of Vancosaminyltransferase GtfD from the Vancomycin Biosynthetic Pathway: Interactions with Acceptor and Nucleotide Ligands, *Biochemistry*, **43**, 5170-5180.

Muller, B.,, C. Schaub, and R. R. Schmidt (1998) Efficient Sialyltransferase Inhibitors Based on Transition-State Analogues of the Sialyl Donor, *Angew. Chem. Int. Ed.*, **37**, 2893-2897.

Nguyen, H. P.,, N. O. L. Seto, Y. Cai, E. K. Leinala, S. N. Borisova, M. M. Palcic, and S. V. Evans (2003) The Influence of an Intramolecular Hydrogen Bond in Differential Recognition of Inhibitory Acceptor Analogs by Human ABO(H) Blood Group A and B Glycosyltransferases, *J. Biol. Chem.*, **278**, 49191-49195.

Nitanai, Y.,, T. Kikuchi, K. Kakoi, S. Hanamaki, I. Fujisawa, and K. Aoki (2009) Crystal Structures of the Complexes Between Vancomycin and Cell-Wall Precursor Analogs, *J. Mol. Biol.* **385**, 1422-1432.

Oberthur, M.,, C. Leimkuhler, R. G. Kruger, W. Lu, C. T. Walsh, and D. Kahne (2005) A Systematic Investigation of the Synthetic Utility of Glycopeptide Glycosyltlransferases, *J. Am. Chem. Soc.*, **127**, 10747-10752.

Otten, H. (1986) Domagk and the Development of the Sulphonamides, *J. Antimicrobial Chemotherapy*, **17**, 689-690.

Park, S-H.,, H-Y. Park, J. K. Sohng, H. C. Lee, K. Liou, Y. J. Yoon, and B-G. Kim (2009a) Expanding Substrate Specificty of GT-B Fold Glycosyltlransferase via Domain Swapping and High-Throughput Screening, *Biotechnology and Bioengineering*, **102**, 988-994.

Park, S-H.,, H-Y. Park, B-K. Cho, Y-H. Yang, J. K. Sohng, H. C. Lee, K. Liou, and B-G. Kim (2009b) Reconstitution of Antibiotics Glycosylation by Domain Exchanged Chimeric Glycosyltransferase, *J. Molecular Catalysis B: Enzymatic*, **60**, 29-35.

Patenaude, S. I.,, N. O. L. Seto, S. N. Borisova, A. Szpacenko, S. L. Marcus, M. M. Palcic, and S. V. Evans (2002) The structural basis for specificity in human ABO(H) blood group biosynthesis, *Nat. Struct. Biol.*, **9**, 685-690.

Paulson, J. C.,, J. I. Rearick, and R. L. Hill (1977) Enzymatic Properties of β-D-Galactoside α2-6 Sialytransferase from Bovine Colostrum, *J. Biol. Chem.*, **252**, 2363-2371.

Pearce, C. M., and D. H. Williams (1995) Complete Assignment of the ^{13}C NMR Spectrum of Vancomycin, *J. Chem. Soc. Perkins Trans.* II 153-158.

Pearce, C. M.,, U. Gerhard, and D. H. Williams (1995) Ligands Which Bind Weakly to Vancomycin: Studies by ^{13}C NMR, *J. Chem. Soc. Perkins Trans.* II 159-162.

Pelzer, S.,, W. Reichert, M. Huppert, D. Heckmann, and W. Wohlleben (1997) Cloning and analysis of a peptide synthetase gene of the balhimycin producer *Amycolatopsis mediterranei* DSM5908 and development of a gene disruption/replacement system, *Journal of Biotechnology*, **56**, 115-128.

Pennisi, E. (2003) Reaching Their Goal Early, Sequencing Labs Celebrate, *Science* **300**, 409.

Peracaula, R.,, G. Tabares, A. Lopez-Ferrer, R. Brossmer, C. de Bolos, and R. de Llorens (2005) Role of Sialyltransferases Involved in the Biosynthesis of Lewis Antigens in Human Pancreatic Tumour Cells, *Glycoconjugate Journal*, **22**, 135-144.

Perez-Garay, M.,, B. Arteta, E. Llop, L. Cobler, L. Pages, R. Ortiz, M. J. Ferri, C. de Bolos, J. Figueras, R. de Llorens, F. Vidal-Vanaclocha, and R. Peracaula (2013) α2,3-Sialyltransferase ST3 Gal IV Promotes Migration and Metastasis in Pancreatic Adenocarcinoma Cells and Tends to Be Highly Expressed in Pancreatic Adenocarcinoma Tissues, *Int. J. Biochem. & Cell Biol.*, **45**, 1748-

1757.

Persson, M.,, J. A. Letts, B. Hosseini-Maaf, S. N. Borisova, M. M. Palcic, S. V. Evans and M. L. Olsson (2007) Structural Effects of Naturally Occurring Human Blood Group B Galactosyltransferase Mutations Adjacent to the DXD Motif, *J. Biol. Chem.*, **282**, 9564-9570.

Perutz, M. F.,, M. G. Rossmann, A. F. Cullis, H. Muirhead, G. Will and A. C. T. North (1960) Structure of Haemoglobin – A Three-Dimensional Fourier Synthesis at 5.5 Angstrom Resolution, Obtained by X-Ray Analysis, *Nature*, **185**, 416-422.

Pesnot, T.,, R. Jorgensen, M. M. Palcic, and G. K. Wagner (2010) Structural and mechanistic basis for a new mode of glycosyltransferase inhibition, *Nat. Chem. Biol.*, **6**, 321-323.

Raju, T. N. K. (1999) The Nobel Chronicles. 1939: Gerhard Domagk (1895-1964), *The Lancet*, **353**, 681.

Ramakrishnan, B.,, E. Boeggeman, V. Ramasamy, and P. K. Qasba (2004) Structure and catalytic cycle of β-1,4-galactosyltransferase, *Curr. Opin. Struct. Biol.*, **14**, 593-600.

Rao, A. V. R.,, M. K. Gurjar, K. L. Reddy, and S. Rao (1995) Studies Directed toward the Synthesis of Vancomycin and Related Cyclic Peptides, *Chemical Reviews*, **95**, 2135-2167.

Rausch, C.,, T. Weber, O. Kohlbacher, W. Wohlleben, and D. H. Hudson (2005) Specificty Prediction of Adenylation Domains in Nonribosomal Peptide Synthetases (NRPS) Using Transductive Support Vector Machines (TSVMs), *Nucl. Acids Res.* **33**, 5799-5808.

Richardson, J. S., and D. C. Richardson (2014) Biophysical Highlights from 54 Years of Macromolecular Crystallography, *Biophysical Journal*, **106**, 510-525.

Riedel, S. (2005) Edward Jenner and The History of Smallpox and Vaccination, *Baylor Univ. Med. Center Proceedings* **18**, 21-25.

Rillahan, C. D.,, A. Antonopoulos, C. T. Lefort, R. Sonon, P. Azadi, K. Ley, A. Dell, S. M. Haslam, and J. C. Paulson (2012) Global metabolic inhibitors of sialyl- and fucosyltransferases remodel the glycome, *Nat. Chem. Biol.*, **8**, 661-668.

Rolinson, G. N.,, F. R. Batchelor, S. Stevens, and J. C. Wood (1960) Bacteriological

Studies on a New Penicillin – BR1241, *The Lancet*, **2 (originally, now 276)**, 564-567.

Rolinson, G. N., and A. M. Geddes (2007) The 50th anniversary of the discovery of 6-amino penicillanic acid (6-APA), *International Journal of Antimicrobial Agents*, **29**, 3-8.

Roper, D. I.,, T. Huyton, A. Vagin, and G. Dodson (2000) The Molecular Basis of Vancomycin Resistance in Clinically Relevant Enterococci: Crystal Structure of D-Alanyl-D-Lactate Ligase (VanA), *Proc. Natl. Acad. Sci.* **97**, 8921-8925.

Sainz-Polo, M. A.,, M. Ramirez-Escudero, A. Lafraya, B. Gonzalez, J. Marin-Navarro, J. Polaina, J. Sanz-Aparicio (2013) Three-Dimensional Structure of *Saccharomyces* Invertase: Role of a Non-Catalytic Domain in Oligomerization and Substrate Specificity, *J. Biological Chemistry*. **288**, 9755-9766.

Salas, J. A., and C. Mendez (2007) Engineering the glycosylation of natural products in actinomycetes, *TRENDS in Microbiology*, **15**, 219-232.

Schaefer, K.,, J. Albers, N. Sindhuwinata, T. Peters, and B. Meyer (2012) A New Concept for Glycosyltransferase Inhibitors: Nonionic Mimics of the Nucleotide Donor of the Human Blood Group B Galactosyltransferase, *ChemBioChem*, **13**, 443-450.

Schaefer, K.,, N. Sindhuwinata, T. Hackl, M. P. Kotzler, F. C. Niemeyer, M. M. Palcic, T. Peters, and B. Meyer (2013) A Nonionic Inhibitor with High Specificity for the UDP-Gal Donor Binding Site of Human Blood Group B Galactosyltransferase: Design, Synthesis, and Characterization, *J. Med. Chem.*, **56**, 2150-2154.

Schafer, M.,, T. R. Schneider, and G. M. Sheldrick (1996) Crystal Structure of Vancomycin, *Structure*, **4**, 1509-1515.

Schatz, A., and S. Waksman (1944) Effect of streptomycin upon mycobacterium tuberculosis and related organisms, *Proc. Soc. Exptl. Biol. Med.*, **57**, 244-248. (unavailable through Duke) This is reference #3 from: R. Zetterstrom (2007) Selman A. Waksman (1888-1973) Nobel Prize in 1952 for the discovery of streptomycin, the first antibiotic effective against tuberculosis, *Acta Paediatrica*, **96**, 317-319.

Schaub, C.,, B. Muller, and R. R. Schmidt (1998) New sialyltransferase inhibitors based on CMP-quinic acid: Development of a new sialyltransferase assay, *Glycoconjugate Journal*, **15**, 345-354.

Schwarz, F.,, Y-Y. Fan, M. Schubert, and M. Aebi (2011) Cytoplasmic *N*-Glycosyltransferase of *Actinobacillus pleuropneumoniae* Is an Inverting Enzyme and Recognizes the NX(S/T) Consensus Sequence, *J. Biol. Chem.* **286**, 35267-35274

Seto, N. O. L.,, M. M. Palcic, C. A. Compton, H. Li, D. R. Bundle, and S. A. Narang (1997) Sequential Interchange of Four Amino Acids from Blood Group B to Blood Group A Glycosyltransferase Boosts Catalytic Activity and Progressively Modifies Substrate Recognition in Human Recombinant Enzymes, *J. Biol. Chem.*, **272**, 14133-14138.

Seto, N. O. L.,, C. A. Compton, S. V. Evans, D. R. Bundle, S. A. Narang, and M. M. Palcic (1999) Donor Substrate Specificity of Recombinant Human Blood Group A, B, and Hybrid A/B Glycosyltlransferases Expressed in *Escherichia coli*, *Eur. J. Biochem.*, **259**, 770-775.

Shayman, J. A. (2010) Eliglustat Tartrate: Glucosylceramide Synthase Inhibitor Treatment of Type 1 Gaucher Disease, *Drugs Future,* **35**, 613-620.

Sheehan, J. C.,, K. R. Henery-Logan, and D. A. Johnson (1953) The Synthesis of Substituted Penicillins and Simpler Structural Analogs. VII. The Cyclization of a Penicilloate Derivative to Methyl Phthalimidopenicillanate, *J. Amer. Chem. Soc.*, **75**, 3292-3293.

Sheehan, J. C., and K. R. Henery-Logan (1957) The total synthesis of Penicillin V, *J. Am Chem Soc*, **79**, 1262-1263.

Sieber, S. A., and M. A. Marahiel (2005) Molecular Mechanisms Underlying Nonribosomal Peptide Synthesis: Approaches to New Antibiotics, *Chem. Rev.* **105**, 715-738.

Sola, R. J., and K. Griebenow (2006) Chemical Glycosylation: New Insights on the Interrelation Between Protein Structural Mobility, Thermodynamic Stability, and Catalysis, *FEBS Letters* **580**, 1685-1690.

Stevens, B. W.,, T. M. Joska, and A. C. Anderson (2006a) Progress Toward Re-Engineering Non-Ribosomal Peptide Synthetase Proteins: A Potential New Source of Pharmacological Agents, *Drug Development Research* **66**, 9-18.

Stevens, B. W.,, R. Lilien, I. Georgiev, B. R. Donald, and A. C. Anderson (2006b)

Redesigning the PheA Domain of Gramicidin Synthetase Leads to a New Understanding of the Enzyme's Mechanism and Selectivity, *Biochemistry* **45**, 15495-15504.

Takaya, K.,, N. Nagahori, M. Kurogochi, T. Furuike, N. Miura, K. Monde, Y. C. Lee, and S-I. Nishimura (2005) Rational Design, Synthesis, and Characterization of Novel Inhibitors for Human β 1,4-Galactosyltransferase, *J. Med. Chem.*, **48**, 6054-6065.

The College of Physicians of Philadelphia. *http://www.historyofvaccines.org/content/timelines/pasteur#EVT_100871* (2014) The History of Vaccines, a project of The College of Physicians of Philadelphia.

The Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome, *Nature* **486**, 207-214.

Truman, A. W.,, M. V. B. Dias, S. Wu, T. L. Blundell, F. Huang, and J. B. Spencer (2009) Chimeric Glycosyltransferases for the Generation of Hybrid Glycopeptides, *Chem. Biol.*, **16**, 676-685.

Tsugita, A.,, M. Inouye, E. Terzaghi and G. Streisinger (1968) Purification of Bacteriophage T4 Lysozyme, *J. Biol. Chem.*, **243**, 391-397.

van Wageningen, A. M. A.,, P. N. Kirkpatrick, D. H. Williams, B. R. Harris, J. K. Kershaw, N. J. Lennard, M. Jones, S. J. M. Jones, and P. J. Solenberg (1998) Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic, *Chemistry & Biology*, **5**, 155-162.

Varki, A.,, R. D. Cummings, J. D. Esko, et al., editors, Cold Spring Harbor (NY), CSHL Press (2009) Chapter 8. N-Glycans, *Essentials of Glycobiology*, **2nd edition,** NCBI Bookshelf. http://www.ncbi.nlm.nih.gov/books/NBK1908/

Varki, A.,, R. D. Cummings, J. D. Esko, et al., editors, Cold Spring Harbor (NY), CSHL Press (2009) Chapter 13. Structures Common to Different Glycans, *Essentials of Glycobiology*, **2nd edition,** NCBI Bookshelf. http://www.ncbi.nlm.nih.gov/books/NBK1908/

Varki, A.,, R. D. Cummings, J. D. Esko, et al., editors, Cold Spring Harbor (NY), CSHL Press (2009) Chapter 30. P-type Lectins, *Essentials of Glycobiology*, **2nd edition,** NCBI Bookshelf. http://www.ncbi.nlm.nih.gov/books/NBK1908/

Villalonga, M. L.,, P. Diez, A. Sanchez, M. Gamella, J. M. Pingarron, and R. Villalonga (2014) Neoglycoenzymes, *Chemical Reviews* **114**, 4868-4917.

Villalonga, R.,, M. Fernandez, A. Fragoso, R. Cao, P. DiPierro, L. Mariniello, and R. Porta (2003) Transglutaminase-Catalyzed Synthesis of Trypsin-Cyclodextrin Conjugates: Kinetics and Stability Properties, *Biotechnology and Bioengineering* **81**, 732-737.

Vrijbloed, J. W.,, J. Madon and L. Dijkhuizen (1995) Transformation of the Methylotrophic Actinomycete *Amycolatopsis methanolica* with Plasmid DNA: Stimulatory Effect of a pMEA300-Encoded Gene, *Plasmid*, **34**, 96-104.

Waksman, S.A. (1947) What is an Antibiotic or an Antibiotic Substance?, *Mycologia* **39**, 565-569.

Walsh, C. T.,, C. L. F. Meyers, and H. C. Losey (2003) Antibiotic Glycosyltransferases: Antibiotic Maturation and Prospects for Reprogramming, *Journal of Medicinal Chemistry*, **46**, 3425-3436.

Wang, S.,, J. A. Cuesta-Seijo, D. Lafont, M. M. Palcic, and S. Vidal (2013) Design of Glycosyltransferase Inhibitors: Pyridine as a Pyrophosphate Surrogate, *Chem. Eur. J.*, **19**, 15346-15357.

Watson, J. D., and F. H. C. Crick (1953a) Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid, *Nature* **171**, 737-738.

Watson, J. D., and F. H. C. Crick (1953b) Genetical Implications of the Structure of Deoxyribonucleic Acid, *Nature* **171**, 964-967.

Wilkins, M. H. F.,, A. R. Stokes, and H. R. Wilson (1953) Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids, *Nature* **171**, 738-740.

Williams, G. J.,, C. Zhang, and J. S. Thorson (2007) Expanding the Promiscuity of a Natural-Product Glycosyltransferase by Directed Evolution, *Nat. Chem. Biol.*, **3**, 657-662.

Williams, G. J.,, R. D. Goff, C. Zhang, and J. S. Thorson (2008) Optimizing Glycosyltransferase Specificty via “Hot Spot” Saturation Mutagenesis Presents a Catalyst for Novobiocin Glycorandomization, *Chemistry & Biology*, **15**, 393-401.

Wu, C-Y.,, C-C. Hsu, S-T. Chen, and Y-C. Tsai (2001) Soyasaponin 1, a Potent and Specific Sialytransferase Inhibitor, *Biochem. Biophys. Res. Commun.*, **284**, 466-469.

Xie, J.,, J. G. Pierce, R. C. James, A. Okano, and D. L. Boger (2011) A Redesigned Vancomycin Engineered for Dual D-Ala-D-Ala and D-Ala-D-Lac Binding Exhibits Potent Antimicrobial Activity Against Vancomycin-Resistant Bacteria, *J. Am. Chem. Soc.*, **133**, 13946-13949.

Xie, J.,, A. Okano, J. G. Pierce, R. C. James, S. Stamm, C. M. Crane, and D. L. Boger (2012) Total Synthesis of [Ψ[C(=S)NH]Tpg^4]Vancomycin Aglycon, [Ψ[C(=NH)NH]Tpg^4]Vancomycin Aglycon, and Related Key Compounds: Reengineering Vancomycin for Dual D-Ala-D-Ala and D-Ala-D-Lac Binding, *J. Am. Chem. Soc.*, **134**, 1284-1297.

Yamamoto, F.,, H. Clausen, T. White, J. Marken, and S. Hakomori (1990) Molecular genetic basis of the histo-blood group ABO system, *Nature*, **345**, 229-233.

Yamamoto, F., and S. Hakomori (1990) Sugar-nucleotide Donor Specificity of Histo-blood Group A and B Transferases Is Based on Amino Acid Substitutions, *J. Biol. Chem.*, **265**, 19257-19262.

Yamamoto, F., and P. D. McNeill (1996) Amino Acid Residue at Codon 268 Determines Both Activity and Nucleotide-Sugar Donor Substrate Specificity of Human Histo-blood Group A and B Transferases, *J. Biol Chem.*, **271**, 10515-10520.

Yamamoto, M.,, X-H. Lin, Y. Kominato, Y. Hata, R. Noda, N. Saitou, and F. Yamamoto (2001) Murine Equivalent of the Human Histo-blood Group ABO Gene Is a *cis*-AB Gene and Encodes a Glycosyltlransferase with Both A and B Transferase Activity, *J. Biol. Chem.*, **276**, 13701-13708.

Yarlagadda, V.,, P. Akkapeddi, G. B. Manjunath, and J. Haldar (2014) Membrane Active Vancomycin Analogues: A Strategy to Combat Bacterial Resistance, *J. Med. Chem.*, **57**, 4558-4568.

Yim, G.,, M. N. Thaker, K. Koteva, and G. Wright (2014) Glycopeptide Antibiotic Biosynthesis, *J. Antibiotics*, **67**, 31-41.

Zeng, J.,, J. Boyles, C. Tripathy, L. Wang, A. Yan, P. Zhou, and B. R. Donald (2009) High-Resolution Protein Structure Determination Starting With a Global Fold Calculated From Exact Solutions to the RDC Equations, *J. Biomol. NMR* **45**, 265-281.

Zeng, X.,, S. Wang, K. Jing, Z. Zhang, and Y. Lu (2013) Use of biodiesel-derived crude glycerol for vancomycin production by *Amycolatopsis orientalis* XMU-VS01, *Eng. Life Sci.*, **13**, 109-116.

Zhanel, G. G.,, D. Calic, F. Schweizer, S. Zelenitsky, H. Adam, P. R. S. Lagace-Wiens, E. Rubinstein, A. S. Gin, D. Hoban, and J. A. Karlowsky (2010) New Lipoglycopeptides: A Comparative Review of Dalbavancin, Oritavancin, and Telavancin, *Drugs*, **70**, 859-886.

Zhou, P., and G. Wagner (2010) Overcoming the Solubility Limit With Solubility-Enhancement Tags: Successful Applications in Biomolecular NMR Studies, *J. Biomol. NMR* **46**, 23-31.

Biography

John Samuel MacMaster

Place and Date of Birth: Chicago, Illinois; 2/16/1956

Education:

2015	Ph.D., Duke University, Durham, NC Department of Computer Science, Advisor: Bruce R. Donald, Ph.D.
2000	M.S., University of Utah, Salt Lake City, UT Department of Human Genetics, Advisor: Mario R. Capecchi, Ph.D.
1996	B.S., Massachusetts Institute of Technology, Cambridge, MA Department of Biology Department of Chemistry

Publications:

G.D. Prestwich, Y. Luo, M.R. Ziebell, K.P. Vercruysse, K.R. Kirker, J.S. MacMaster. Chemically-modified hyaluranon: New biomaterials and probes for cell biology. New Frontiers in Medical Science: Redefining Hyaluranon, 181-194, 2000.

M. Bogyo, S. Shin, J.S. McMaster, and H.L. Ploegh. Substrate binding and sequence preference of the proteasome revealed by active site-directed affinity probes. Chemistry & Biology, **5**, 307-320, 1998.

R. Glas, M. Bogyo, J.S. McMaster, M. Gaczynska, and H.L. Ploegh. A proteolytic system that compensates for loss of proteasome function. Nature, **392**, 618-622, 1998.

M. Bogyo, J.S. McMaster, M. Gaczynska, D. Tortorella, A.L. Goldberg, and H.L. Ploegh. Covalent modification of the active site threonine of proteasomal β subunits and the *E. coli* homolog HslV by a new class of inhibitors. Proc. Nat. Acad. Sci. USA, **94**, 6629-6634, 1997.

P. Pereira, M. Zijlstra, J.S. McMaster, J.M. Loring, R. Jaenisch, and S. Tonegawa. Blockade of transgenic γδ T cell development in β2-microglobulin deficient mice. EMBO Journal, **11**, 25-31, 1992.

K.W. Wegmann, J.S. McMaster, and W.R. Green. Mechanisms of nonresponsiveness to AKR/Gross leukemia virus in AKR.H-2^b:Fv-1^b mice: An analysis of precursor CTL frequencies in young versus moderately aged mice. J. Immunology, **146**, 2469-2477, 1991.

K.L. Fitzgerald, J.S. McMaster, and N.M. Ponzio. Natural killer cell activity in reticulum-cell sarcomas of SJL/J mice. III. Characterization of the effector cells within RCS that mediate NK lysis. International J. Cancer, **28**, 635-645, 1981.