A TALE OF TWO TRANSFERS:

POMGNT2 SPECIFICITY IN PROTEIN O-MANNOSYLATION

AND PROBLEM SOLVING IN UNDERGRADUATE BIOCHEMISTRY

by

STEPHANIE MARIE HALMO

(Under the Direction of Paula P. Lemons and Lance Wells)

ABSTRACT

Composed of two seemingly disparate parts, this dissertation ultimately seeks to extend our understanding of biochemistry. The first part investigates a glycosyltransferase enzyme in the *O*-mannosylation pathway responsible for glycosylating the cell surface glycoprotein, α-dystroglycan. α-Dystroglycan is extensively modified with *O*-mannose glycans that are critical to basement membrane assembly via interactions with laminin and other proteins. Disruption of the *O*-mannosylation pathway involved in functional glycosylation of α-dystroglycan gives rise to congenital muscular dystrophies. With recent advancements, the entire functional glycan structure on α-dystroglycan and the enzymes responsible for its biosynthesis are known. POMGNT2 catalyzes the first step toward the functional glycan structure on α-dystroglycan. Yet, how specificity at each step in the biosynthesis of the elaborate glycan structure is achieved remains unknown. Studies undertaken in the first part of this dissertation aim to elucidate the substrate selectivity of POMGNT2 for two sites on α-dystroglycan. *In vitro*, POMGNT2 displays significant primary amino acid selectivity near the site of *O*-mannosylation. Further studies show that addition of primary amino acid determinants for POMGNT2 activity can lead to functional glycosylation

elsewhere on α-dystroglycan. The second part of this dissertation investigates undergraduate student thinking and targeted instruction about noncovalent interactions. Noncovalent interactions represent a core concept in undergraduate biochemistry courses. Yet, how undergraduate biochemistry students build conceptual understanding and problem-solving skill pertaining to noncovalent interactions remains largely unknown. In light of rejuvenated effort to improve undergraduate biochemistry education, the studies undertaken in the second part of this dissertation aim to characterize student thinking about noncovalent interactions in order to develop targeted instruction for this concept. Interviews with experts and students as they solved a protein structure-function problem revealed several student difficulties including challenges with explaining electrostatic principles of noncovalent interactions. Using the identified student difficulties, various types of targeted instruction deriving from distinct conceptual frameworks were developed and compared. This study suggests more than unscaffolded guidance is necessary for successful near transfer of knowledge about noncovalent interactions. Together, this body of work is an interdisciplinary research effort comprising both glycobiology and discipline-based education research in biochemistry.

INDEX WORDS: *O*-mannosylation, alpha-dystroglycan, POMGNT2, congenital muscular dystrophy, secondary dystroglycanopathies, STEM education, biochemistry undergraduate education, problem solving, noncovalent interactions, active learning, productive failure, guided inquiry

A TALE OF TWO TRANSFERS:

POMGNT2 SPECIFICITY IN PROTEIN O-MANNOSYLATION

AND PROBLEM SOLVING IN UNDERGRADUATE BIOCHEMISTRY

by

STEPHANIE MARIE HALMO

B.S., University of Georgia, 2012

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

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2020

A TALE OF TWO TRANSFERS:

POMGNT2 SPECIFICITY IN PROTEIN O-MANNOSYLATION

AND PROBLEM SOLVING IN UNDERGRADUATE BIOCHEMISTRY

by

STEPHANIE MARIE HALMO

Major Professors:	Paula P. Lemons
	Lance Wells
Committee:	Kelley W. Moremen
	Julie Dangremond Stanton

Electronic Version Approved:

Ron Walcott
Interim Dean of the Graduate School
The University of Georgia
August 2020

DEDICATION

For my Dad,

Paul Michael William Halmo.

ACKNOWLEDGEMENTS

I would like to thank my mentors, Paula Lemons and Lance Wells, for their tutelage, understanding, and patience on my journey to becoming an independent scientist. I would also like to thank the members of my committee, Julie Stanton and Kelley Moremen. Collectively, these individuals took on the task of supporting me in my pursuit of a rather unique degree and I am grateful for their efforts and insight.

In addition, I am indebted to the incredibly supportive and intellectually stimulating environments found within the Biology Education Research Group and the Complex Carbohydrate Research Center at UGA. I would like to acknowledge my lab mates and colleagues who have contributed to my development and made it fun, especially Cheryl Sensibaugh, Patricia Zagallo, Sophia Jeong, Gretchen King, Lisa Limeri, Michelle Ziadie, Kush Bhatia, Trevor Tuma, Peter Reinhart, Sasha Stogniy, Grace Snuggs, Vanessa Alele, Angie Stockton, Rebecca McCurley, Michelle Dookwah-Smith, Rob Bridger, Osman Sheikh, Stephanie Stalnaker, Sneha Patel, Danish Singh, Jessica Ho, David Live, Jeremy Praissman, Trevor Adams, Johnathan Mayfield, Hannah Stephen, and Chelsea Desbiens.

I would like to thank my entire family for being so supportive and understanding over the past five years. Most importantly, I want to thank my mom, Karen Halmo, my sister, Laura Halmo, and my partner, Hunt Gentry, for their unconditional love and unwavering support. This accomplishment is as much theirs as it is mine.

I would like to thank my friends for their constant encouragement throughout graduate school. To the six and the boysch – thanks for always lifting my spirits. To Amanda Shaver and

the rest of the Modern Pin-Ups – you've made Athens a home and been my safe harbor during turbulent times. I'm a better person because I know you.

I would like to thank my therapist, Tara Weiszer, for listening and giving me the extra encouragement and accountability that I needed during this past year.

Finally, I would like to acknowledge all my former teachers and students who helped me get to this point. I'm grateful to the universe for intertwining our paths. I would not be who I am today without each of you. Thank you.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION - PART 1

Summary and Scope of Dissertation

This chapter outlines the structure of this two-part dissertation and provides relevant background information for part one. The present chapter includes brief sections on the biological roles of glycans and types of protein glycosylation with a specific focus on *O*-mannosylation of α-dystroglycan (α-DG). The following chapters in part one of this dissertation consist of two previously published papers as well as a chapter detailing current progress towards another research article. Chapter 2 is a review article on recent advancements in *O*-mannosylation (Sheikh, M.O., Halmo, S.M., et al. 2017). Chapter 3 consists of a research article detailing the kinetics of two enzymes in the *O*-mannosylation pathway with synthetic glycopeptides derived from the α-DG sequence (Halmo, S.M., Singh, D., et al. 2017). Chapter 4 reports our progress towards another research article aimed at testing the sufficiency of the POMGNT2 motif for functional glycosylation elsewhere on α-DG.

The second part of this dissertation begins with an introduction to relevant background information in Chapter 5. Chapter 5 includes brief sections on the purpose of discipline-based education research and the movement towards evidence-based pedagogies in college STEM courses, particularly in the context of undergraduate biochemistry education. The following chapters in part two of this dissertation consist of two previously published papers. Chapter 6 is a qualitative research article describing how students think and solve a biochemistry problem involving the physical basis of non-covalent interactions. Chapter 7 is an experimental research

article building on Chapter 6 that investigates the comparative impacts of different pedagogies, varying in the nature and timing of guidance, on student learning of the physical basis of non-covalent interactions. The final chapter of this dissertation summarizes the major findings of both parts, discusses future experiments, and joins this body of work in broader context and significance.

Glycans and their Biological Roles

There are four major classes of macromolecules in biology: DNA/RNA, proteins, lipids and carbohydrates. From cellulose in plants to chitin in arthropod exoskeletons, the carbohydrate or glycan class of macromolecules are ubiquitous and abundant in all cell types and taxa of life. While glycans span taxa and species, the focus of this introduction is on glycans in eukaryotic organisms, and even more specifically on mammalian protein glycosylation. Glycans play many biological roles (Varki, A. 2017). Two biological roles of particular relevance to this dissertation are structural function and recognition events at the cell surface. All eukaryotic cells are covered in a dense layer of carbohydrates called the glycocalyx. One important component of the glycocalyx with key roles in extracellular matrix organization and cell-matrix interactions are glycoproteins.

Protein Glycosylation and Nomenclature

Glycoproteins are proteins that are post-translationally modified with glycans. Glycosylation is the most abundant post-translational modification of proteins and occurs across all domains of life. It's estimated that up to 50% of eukaryotic proteins are modified by some glycan structure (Apweiler, R., Hermjakob, H., et al. 1999). Unlike nucleotide and polypeptide production, glycan synthesis is not template driven. Given the non-template directed production of glycans, heterogeneity results. For glycoproteins, this heterogeneity occurs at two levels: site

occupancy and structural diversity at a given site. This key characteristic of glycosylation has made studying this core group of macromolecules challenging, but not impossible. Glycan structures on proteins can be annotated in multiple ways. Chair conformations provide the most stereochemical information, but a symbol nomenclature was developed to allow for convenient and standardized annotation (Neelamegham, S., Aoki-Kinoshita, K., et al. 2019, Varki, A., Cummings, R.D., et al. 2015). This simplified symbol nomenclature for glycans will be used throughout this dissertation.

Types of Protein Glycosylation

Eukaryotic protein glycosylation varies in glycan content, glycan-protein linkage, and the site of glycosylation in the cell. While nuclear and cytoplasmic protein glycosylation via the modification of proteins with *O*-GlcNAc is an important and rich area of study (Hart, G.W. 2019), the focus of this dissertation is on glycoproteins that traverse the secretory pathway to reach the cell surface.

Several different types of protein glycosylation occur in the secretory pathway via enzymes residing in the endoplasmic reticulum and the Golgi. The secretory pathway can be viewed as an assembly line for glycoprotein production. From this view, the workhorses of the assembly line are the ER and Golgi resident enzymes or glycosyltransferases. These glycosyltransferases move sugars from sugar nucleotide or lipid-linked oligosaccharide donors to glycoprotein substrates as they traffic through the secretory pathway. The localization of the membrane bound glycosyltransferases to specific subsites within the secretory pathway dictates the order of events in the biosynthesis of the different types of protein glycosylation.

The glycans on glycoproteins can broadly be classified into two predominant subgroups according to their glycan-peptide linkage: *N*-glycans and *O*-glycans. *N*-glycans are linked to proteins via the nitrogen of the amide group on an asparagine residue, whereas *O*-glycans are

linked to proteins via the oxygen of the hydroxyl group on serine or threonine residues. For *N*-glycans, the reducing end terminal sugar is N-acetylglucosamine. In contrast, for *O*-glycans, the reducing end terminal sugar can be a variety of sugar residues: GalNAc, GlcNAc, fucuse, glucose, xylose, galactose and mannose (Joshi, H.J., Narimatsu, Y., et al. 2018).

N-linked Glycans. The most well-studied and well-known type of secretory and cell-surface protein glycosylation is the class of *N*-linked glycans (Aebi, M. 2013). *N*-linked glycans are large branched structures with a conserved chitobiose core. These glycans get their class name from being linked to asparagine residues in a conserved Asn-X(but not Pro)-Ser/Thr sequon in proteins. *N*-linked glycans are synthesized in a step-by-step process initiated in the ER and continued in the Golgi. A $Man_5GlcNAc_2$-P-P-dolichol structure is first built on the cytoplasmic face of the ER. This structure is then flipped into the lumen of the ER where it is extended to a $Glc_3Man_9GlcNAc_2$-P-P-dolichol before being transferred co-translationally to asparagine residues on nascent polypeptide chains by the enzyme oligosaccharyltransferase. Processing of the *N*-linked glycan on proteins then begins in the ER where glucosidases remove the three glucose residues and a mannosidase removes a mannose. These initial processing steps regulate glycoprotein folding mediated by ER chaperones. For most glycoproteins, additional mannose residues are trimmed in the *cis*-Golgi to achieve a $Man_5GlcNAc_2$ structure. Further processing in the medial-Golgi results in the biosynthesis of hybrid and complex *N*-glycan structures. Further additions such as a fucosylated core, GlcNAc branching, Gal additions, poly-LacNAc and LacdiNAc extensions, as well as capping by sialic acid yield a diverse array of possible mature *N*-glycan structures on glycoproteins (Stanley, P., Taniguchi, N., et al. 2015).

O-linked Glycans. As stated previously, *O*-glycans are linked to proteins via the oxygen of the hydroxyl group on serine or threonine residues. For *O*-glycans, the reducing end terminal sugar

can be a variety of sugar residues (Joshi, H.J., Narimatsu, Y., et al. 2018). One common *O*-linked reducing-end sugar is N-acetylgalactosamine (GalNAc) (Brockhausen, I. and Stanley, P. 2015). Mucins, a type of glycoprotein, contain the greatest number of *O*-GalNAc initiated structures. The simplest *O*-GalNAc initiated structure is the Tn antigen. This single *O*-linked GalNAc on serine or threonine residues can be extended by a β-1,3 linked galactose residue to produce the Core 1 structure. There are six major core structures that can be further extended to give rise to linear or branched chains that vary in length and can carry blood group epitopes (Tran, D.T. and Ten Hagen, K.G. 2013). Another common *O*-linked reducing-end sugar is xylose. Glycosaminoglycans (GAGs) are linked to their protein cores through xylose (Lindahl, U., Couchman, J., et al. 2015). All GAGs share a common tetra-saccharide core composed of a xylose *O*-linked to serine that is capped by two galactoses and glucuronic acid. This common tetra-saccharide core is then elaborated with a linear polymer of repeating disaccharide units of amino sugars (GlcNAc or GalNAc) and uronic acids (glucuronic acid or iduronic acid). This linear polymer can then be modified to result in various GAG structures. Some common GAGs are heparin, heparan sulfate, and chondroitin sulfate. Two other *O*-linked reducing-end sugars are fucose and mannose (Haltiwanger, R.S., Wells, L., et al. 2015). *O*-fucose initiated glycans are small glycans composed of four or fewer sugars attached to serine residues. Glycoproteins known to be modified with *O*-fucose, like Notch, play important roles in signaling during development (Takeuchi, H. and Haltiwanger, R.S. 2014). *O*-mannose glycans are so named because they all initiate with a mannose that is *O*-linked to serine or threonine residues. Glycoproteins modified by *O*-mannose, like α-DG, are essential for proper brain, eye and skeletal muscle function, and are central to this dissertation.

O-mannosylation

Originally discovered in yeast 50 years ago, the post-translational modification of proteins by *O*-linked mannose is conserved from yeast to humans. Until 2017, the homologous family of protein *O*-mannosyl transferases (PMT1-7 in yeast and POMT1/POMT2 in humans) were the only known enzymes involved in *O*-Man synthesis in eukaryotes (Larsen, I.S.B., Narimatsu, Y., et al. 2019, Larsen, I.S.B., Narimatsu, Y., et al. 2017b). Today, we know there are several distinct *O*-mannosylation pathways in eukaryotes. There are three main pathways in metazoa: 1) the TMTC1-4 pathway in metazoans and protists that leads to *O*-mannosylation of proteins in the cadherin superfamily, 2) an uncharacterized pathway in metazoans that leads to *O*-mannosylation of plexins, and 3) the PMT1-7 and POMT1/POMT2 pathway conserved across eukaryotes (except in nematodes and plants) that leads to *O*-mannosylation of α-DG and some other proteins. Each pathway will be discussed in turn, with particular emphasis on the POMT1/POMT2 pathway.

O-mannosylation of cadherins by TMTC1-4. Two groups independently discovered *O*-man structures on cadherins in 2013 (Vester-Christensen, M.B., Halim, A., et al. 2013, Winterhalter, P.R., Lommel, M., et al. 2013). The first group probed the *O*-man glycoproteome using their SimpleCell strategy where nuclease-mediated gene editing of a human breast cancer cell line was employed to reduce the structural heterogeneity of *O*-man glycans. In this cell line they found members of the cadherin family to be *O*-mannosylated (Vester-Christensen, M.B., Halim, A., et al. 2013). In tandem, another group probed the *O*-man glycoproteome in tissues using a combination of glycosidase treatment and LC-MS analyses. Using this orthogonal method to the SimpleCell strategy, the second group identified T-cadherin (CDH13) from rabbit skeletal muscle as being modified by a single mannose residue (Winterhalter, P.R., Lommel, M., et al. 2013). Shortly after this discovery, the same group also showed that affinity purified E-cadherin was also

modified by a single *O*-mannose residue and suggested that *O*-mannosylation of E-cadherin is essential for cell-cell adhesion (Lommel, M., Winterhalter, P.R., et al. 2013).

The finding of single mannose residues on cadherins that were not further elaborated was surprising and warranted further investigation into their biosynthesis. Knocking out the POMT1/POMT2 enzymes indicated that these *O*-mannosylation enzymes were not required for the recently discovered *O*-mannosylation on cadherins (Larsen, I.S.B., Narimatsu, Y., et al. 2017b). The search for the enzymes responsible for the *O*-mannosylation found on cadherins identified a family of four genes called transmembrane and tetratricopeptide repeat containing proteins (TMTC1-4) as putative *O*-man transferases. Knocking out all four genes resulted in the loss of O-mannosylation on cadherins but not on other known *O*-mannosylated proteins, like α-DG, suggesting these enzymes serve as strong candidates for *O*-man transferases on cadherins (Larsen, I.S.B., Narimatsu, Y., et al. 2017a). TMTC1 and TMTC2 have been identified as ER localized enzymes (Sunryd, J.C., Cheon, B., et al. 2014), and more recently, the overexpression of TMTC3 rescued the *O*-mannosylation of E-cadherin and cell adherence in a genetically engineered cell line lacking all for TMTC genes (Graham, J.B., Sunryd, J.C., et al. 2020). Interestingly, biallelic mutations in TMTC3 result in cobblestone lissencephaly, which also occurs in dystroglycanopathies caused by mutations in classical *O*-mannosylation related enzymes (Jerber, J., Zaki, M.S., et al. 2016).

O-mannosylation of IPT domains. Aside from cadherins, Vester-Christensen, M.B., Halim, A., et al. (2013) also identified plexins and two structurally homologous receptor tyrosine kinases, hepatocyte growth factor receptor (c-Met) and macrophage stimulating 1 receptor (RON), as being *O*-mannosylated using their SimpleCell strategy. These authors noted a similar structural feature among the *O*-mannosylated proteins they discovered: the sites of *O*-mannosylation they identified

on plexins, c-Met and RON were all located on Ig-like, plexin, and transcription factor domains. Interestingly, *O*-mannosylation sites in these domains, like on cadherin, are not further extended. Additionally, the *O*-mannosylation of IPT domains is not POMT1/POMT2 or TMTC1-4 dependent (Larsen, I.S.B., Narimatsu, Y., et al. 2017a, Larsen, I.S.B., Narimatsu, Y., et al. 2017b). Therefore, the enzyme(s) that *O*-mannosylate IPT domains are currently unknown.

O-mannosylation of α-DG by POMT1/POMT2. *O*-mannosylation of α-DG begins in the endoplasmic reticulum. Two protein *O*-mannosyltransferase enzymes, POMT1 and POMT2, form a complex in the endoplasmic reticulum where they transfer mannose from dolichol-phosphate-mannose to serine or threonine residues on proteins in an α stereochemical linkage (Manya, H., Chiba, A., et al. 2004). Protein substrates that are modified by POMT1/2-dependent *O*-mannosylation are provided in Table 1.1. Similar to *O*-GalNAc initiated glycans, *O*-mannosylated glycans can be categorized into different core structures: M0, M1, M2, and M3 (Yoshida-Moriguchi, T., Willer, T., et al. 2013). Core M0 structures consist of a single unextended mannose (see *O-mannosylation of cadherins by TMTC1-4*) and POMT1/POMT2-initiated core M0 structures are currently not thought to exist *in vivo*. The initial *O*-man residue catalyzed by POMT1/POMT2 activity is further modified throughout the secretory pathway with N-acetylglucosamine, galactose, sialic acid, phosphate, N-acetylgalactosamine, xylose and glucuronic acid to generate a variety of different *O*-man initiated glycan structures (Praissman, J.L. and Wells, L. 2014).

Most often, the initial mannose is extended by the addition of an N-acetylglucosamine in a β-1,2-linkage by the *cis*-Golgi-resident enzyme POMGNT1 (Yoshida, A., Kobayashi, K., et al. 2001). This POMGNT1 extended mannose structure is known as the core M1 *O*-man glycan. The core M1 structure can be branched by the addition of a β-1,6-linked N-acetylglucosamine via the

Table 1.1. *O*-mannosylated Proteins and POMT1/2 Dependency

UniProt Accession	Protein Name	POMT1/2 Dependency	Reference
P62249	40S ribosomal protein S16	POMT1/POMT2 dependent	Larsen, I.S.B., Narimatsu, Y., et al. (2017a), Larsen, I.S.B., Narimatsu, Y., et al. (2017b), Vester-Christensen, M.B., Halim, A., et al. (2013)
Q9UHI8	A disintegrin and metalloproteinase with thrombospondin motifs 1		
P35611	A-adducin		
Q13618	Cullin-3		
Q14118	Dystroglycan		
Q7Z6Z7	E3 ubiquitin-protein ligase HUWE1		
P35658	Nuclear pore complex protein Nup214		
Q15293	Reticulocalbin-1		
Q9UQ35	Serine/arginine repetitive matrix protein 2		
Q9HCN8	Stromal cell-derived factor 2-like protein 1		
Q9UBS9	SUN domain-containing ossification factor		
P10599	Thioredoxin		
Q86YP4	Transcriptional repressor p66-α		
Q9ULS5	Transmembrane and coiled-coil domains protein 3		
Q9BQE3	Tubulin α-1C chain		
Q13162	Peroxiredoxin-4		
Q9Y6A1	Protein O-mannosyl-transferase 1		
Q9UKY4	Protein O-mannosyl-transferase 2		
Q9HCM3	UPF0606 protein KIAA1549		
P23471	Receptor-type tyrosine-protein phosphatase zeta/phosphacan	Likely POMT1/POMT2 dependent	Bartels, M.F., Winterhalter, P.R., et al. (2016), Trinidad, J.C., Schoepfer, R., et al. (2013)
O14594	Neurocan	Unknown	Bartels, M.F., Winterhalter, P.R., et al. (2016), Pacharra, S., Hanisch, F.G., et al. (2013)
Q9Y4C0	Neurexin 3 α	Unknown	Bartels, M.F., Winterhalter, P.R., et al. (2016)
Q9HDB5	Neurexin 3 β		
Q86UX2	Inter-α-trypsin inhibitor heavy chain family member 5		
P30101	Protein disulfide isomerase A3		
Q8TDW7	Fat3		
P25063	CD24	Unknown	Bleckmann, C., Geyer, H., et al. (2009)
Q92752	Tenascin-R	Unknown	Wing, D.R., Rademacher, T.W., et al. (1992)
P16112	Aggrecan	Unknown	Pacharra, S., Hanisch, F.G., et al. (2013)
P13611	Versican		
Q96GW7	Brevican		
O94856	Neurofascin 186	Unknown	Pacharra, S., Hanisch, F.G., et al. (2012), Pacharra, S., Hanisch, F.G., et al. (2013)

GNT-VB enzyme, yielding the core M2 structure (Praissman, J.L. and Wells, L. 2014). Collectively, core m1 and core M2 glycans account for the largest fraction of *O*-mannose glycans released from mouse brain proteins and 20% of total brain protein *O*-glycans (Stalnaker, S.H., Aoki, K., et al. 2011). However, the function of the core M1 and core M2 extended structures remains largely unknown.

The most well studied *O*-man glycan core structure is core M3. Core M3 structures are defined by the addition of a β-1,4-linked N-acetylglucosamine to the initial mannose residue (Yagi, H., Nakagawa, N., et al. 2013). This addition is catalyzed by the ER-resident enzyme POMGNT2 (Ogawa, M., Nakamura, N., et al. 2013). The core M3 structure is further elaborated by a suite of enzymes (see Chapter 2) that ultimately results in the synthesis of a chain of disaccharide units of xylose and glucuronic acid known as matriglycan. Matriglycan is responsible for binding to extracellular matrix proteins and promoting the entry of certain arenaviruses into cells (Yoshida-Moriguchi, T. and Campbell, K.P. 2015). Core M3 glycans make up a small portion of the *O*-glycome (Stalnaker, S.H., Aoki, K., et al. 2011), but serve important functional roles. Mutations in genes that encode the enzymes responsible for building and elaborating the functional core M3 structure result in a subset of congenital muscular dystrophies known as secondary dystroglycanopathies (Di Costanzo, S., Balasubramanian, A., et al. 2014, Endo, Y., Dong, M., et al. 2015, Godfrey, C., Foley, A.R., et al. 2011, Manzini, M.C., Tambunan, Dimira E., et al. 2012, Stevens, E., Carss, K.J., et al. 2013). These conditions range in spectrum from the less severe Limb-Girdle muscular dystrophy to the more severe Fukuyama congenital muscular dystrophy, Muscle-Eye-Brain disease, and Walker-Warburg syndrome (Falsaperla, R., Pratico, A.D., et al. 2016). These disease states are characterized by hypotonia and muscle tissue degeneration. In the

more severe cases, cobblestone lissencephaly and other brain and eye abnormalities are present at birth.

α-DG

α-DG is the most extensively characterized glycoprotein modified with *O*-mannose glycans. Initially discovered as a 156 kDa dystrophin-associated glycoprotein in the dystrophin glycoprotein complex in skeletal muscle (Campbell, K.P. and Kahl, S.D. 1989), α-DG's primary sequence was first reported via complementary DNA cloning in 1992 (Ibraghimov-Beskrovnaya, O., Ervasti, J.M., et al. 1992). Based solely on the primary sequence, α-DG is predicted to be a 74 kDa core protein. Due to heavy glycosylation, α-DG appears as a broad smear in SDS-PAGE and ranges in size in a tissue-specific manner (Ibraghimov-Beskrovnaya, O., Ervasti, J.M., et al. 1992). Based on electron microscopy studies, α-DG is described and often depicted in schematics as a dumbbell shaped protein (Brancaccio, A., Schulthess, T., et al. 1995). The rod-shaped central mucin-like domain (amino acids 316-485) is heavily glycosylated. On either side of the rod-shaped mucin-like domain are two globular domains (the N-terminus and C-terminus). In its mature form, α-DG lacks the globular N-terminal domain due to Furin cleavage (Kanagawa, M., Saito, F., et al. 2004, Singh, J., Itahana, Y., et al. 2004). The N-terminal domain of α-DG is secreted and trace amounts are found in serum (Saito, F., Saito-Arai, Y., et al. 2008).

Research into α-DG's biological roles in disease and development were a focus of research in the 1990s (Durbeej, M., Henry, M.D., et al. 1998). As a result of this body of work, α-DG was identified as a laminin receptor (Ervasti, J.M. and Campbell, K.P. 1993, Ibraghimov-Beskrovnaya, O., Ervasti, J.M., et al. 1992) involved in basement membrane assembly (Henry, M.D. and Campbell, K.P. 1998) with implications for congenital muscular dystrophies, a receptor for lymphocytic choriomeningitis virus and Lassa fever virus (Cao, W., Henry, M.D., et al. 1998), and

a Schwann cell receptor for *Mycobacterium leprae* (Rambukkana, A., Yamada, H., et al. 1998). This early work into α-DG revealed that it was heavily glycosylated in a tissue-specific manner (Ibraghimov-Beskrovnaya, O., Ervasti, J.M., et al. 1992) and could be stained by Alcian Blue (Ervasti, J.M. and Campbell, K.P. 1993). Work also showed α-DG's interaction with laminin was calcium-dependent and inhibited by heparin (Ervasti, J.M. and Campbell, K.P. 1993). This suggested that the post-translational glycan modifications on α-DG may include polyanionic glycosaminoglycans. Treatment with glycosaminoglycan-specific removing enzymes, the N-glycan removing enzyme PNGaseF, and a typical *O*-GalNAc removing *O*-glycosidase had no effect on the laminin binding ability of α-DG. However, treatment with periodic acid or trifluoromethanesulfonic acid resulted in the loss of laminin binding (Ervasti, J.M. and Campbell, K.P. 1993), suggesting a novel glycan moiety exists on α-DG that is essential for laminin-binding.

O-mannose initiated structures were first discovered on α-DG in 1997 (Chiba, A., Matsumura, K., et al. 1997). The structure, Siaα2-3Galβ1-4GlcNAcβ1-2Manα-Ser/Thr, was initially identified on bovine peripheral nerve α-DG. This structure was also found on rabbit skeletal muscle α-DG alongside core 1 *O*-GalNAc glycans (Sasaki, T., Yamada, H., et al. 1998). Additionally, an *O*-mannose glycan with a Lewis X epitope was found on sheep brain α-DG (Smalheiser, N.R., Haslam, S.M., et al. 1998). These reports confirmed the existence of *O*-mannose linked glycans in mammalian brain proteoglycan fractions that were reported earlier (Finne, J., Krusius, T., et al. 1979, Krusius, T., Finne, J., et al. 1986, Yuen, C.T., Chai, W., et al. 1997) and further identified *O*-mannose initiated glycan structures on a specific protein, α-DG. The discovery of *O*-mannose initiated glycan structures on the same protein in various tissues across species, suggested biological importance, and launched the study of the *O*-mannosylation pathway in mammals and the search for the enzymes involved in the pathway (Endo, T. 1999).

For the first decade of the twenty-first century, research into *O*-mannosylation of α-DG focused on the core M1 extended tetra-saccharide identified by Chiba, A., Matsumura, K., et al. (1997) and related structures. The discovery of LARGE's dual glycosyltransferase activities (Inamori, K., Yoshida-Moriguchi, T., et al. 2012), a set of enzymes (POMGNT2, B3GALNT2, and POMK) essential for functional α-DG (Yoshida-Moriguchi, T., Willer, T., et al. 2013), and a haploid gene trap insertion experiment that identified a host of positive regulators of functional α-DG glycosylation (Jae, L.T., Raaben, M., et al. 2013) represent a turning point in the field and a shift to the concerted study of core M3 structures on α-DG.

With positive regulators of functional α-DG identified via a haploid screen (Jae, L.T., Raaben, M., et al. 2013), recent research efforts into understanding *O*-mannosylation on α-DG have focused on fully elucidating the functional glycan structure responsible for laminin binding. Chapter 2 provides a review of these recent advancements in mammalian *O*-mannosylation including a summary of the elucidation of the full structure of the functional glycan with emphasis on the functions of key enzymes involved in the pathway, namely FKRP, FKTN, TMEM5, B4GAT1 and ISPD (Sheikh, M.O., Halmo, S.M., et al. 2017). With the pathway to the full functional glycan structure on α-DG defined, a few outstanding questions in the field remain.

POMGNT2

Human α-DG has at least twenty-five *O*-mannosylation sites (Harrison, R., Hitchen, P.G., et al. 2012, Nilsson, J., Nilsson, J., et al. 2010, Stalnaker, S.H., Hashmi, S., et al. 2010). Core M1 structures are far more abundant in the mucin-like domain of α-DG than core M3 structures. Site mapping studies have identified only two positions, Thr317 and Thr379, on α-DG with core M3 structures, although some evidence suggests 319 and 381 may also be sites of M3 modification (Hara, Y., Kanagawa, M., et al. 2011, Yagi, H., Nakagawa, N., et al. 2013). Given that there are

only two sites on αDG, Thr317 and Thr379, that carry the functional glycan responsible for binding to laminin-globular domains on extracellular matrix proteins, one critical gap in the literature is the manner in which protein and site specificity is determined for core M3 glycan biosynthesis. Of particular importance to addressing this question, and a central player in the first part of this dissertation, is the enzyme Protein *O*-linked mannose β-1,4-N-acetylglucosaminyltransferase 2 (POMGNT2). POMGNT2 catalyzes the first committed step towards the functional matriglycan structure on α-DG by transferring a β-1,4-linked N-acetylglucosamine to a mannose on Ser/Thr residues of α-DG (Yagi, H., Nakagawa, N., et al. 2013).

The first part of this dissertation seeks to understand what biochemical determinants result in β1,4-GlcNAc modification by POMGNT2 at a limited set of sites. Paradoxically, from a spatial-temporal perspective, *O*-man-modified α-DG encounters POMGNT2 in the ER before POMGNT1 in the *cis*-Golgi but is preferentially modified by POMGNT1. This led us to hypothesize that POMGNT2 must demonstrate substrate selectivity beyond simply an *O*-man-modified amino acid. Chapter 3 takes an *in vitro* approach using various synthetic *O*-mannosylated peptides derived from site mapping efforts on α-DG to demonstrate that POMGNT2 possesses significant primary amino acid selectivity near the sites of *O*-mannosylation. Additionally, Chapter 3 defines a POMGNT2 acceptor motif, conserved among 59 vertebrate species, in α-DG that when engineered into synthetic glycopeptides is sufficient to convert the glycopeptides into POMGNT2 substrates (Halmo, S.M., Singh, D., et al. 2017). Chapter 4 extends the work of Chapter 3 by moving *in cellulo* to demonstrate that the introduction of the POMGNT2 acceptor motif is sufficient for functional matriglycan extension at novel sites on α-DG. These findings contribute to the current working hypothesis that POMGNT2 functions as a gatekeeper enzyme to prevent the vast majority

of *O*-mannosylated sites on α-DG and other proteins from being modified with functional matriglycan.

References

Aebi M. 2013. N-linked protein glycosylation in the ER. *Biochimica et biophysica acta*, 1833:2430-2437.

Apweiler R, Hermjakob H, Sharon N. 1999. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1473:4-8.

Bartels MF, Winterhalter PR, Yu J, Liu Y, Lommel M, Mohrlen F, Hu H, Feizi T, Westerlind U, Ruppert T, *et al.* 2016. Protein O-Mannosylation in the Murine Brain: Occurrence of Mono-O-Mannosyl Glycans and Identification of New Substrates. *PloS one*, 11:e0166119.

Bleckmann C, Geyer H, Lieberoth A, Splittstoesser F, Liu Y, Feizi T, Schachner M, Kleene R, Reinhold V, Geyer R. 2009. O-glycosylation pattern of CD24 from mouse brain. *Biological chemistry*, 390:627-645.

Brancaccio A, Schulthess T, Gesemann M, Engel J. 1995. Electron microscopic evidence for a mucin-like region in chick muscle α-dystroglycan. *FEBS letters*, 368:139-142.

Brockhausen I, Stanley P. 2015. O-GalNAc Glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH, Prestegard JH, et al. editors. *Essentials of Glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press Copyright 2015-2017 by The Consortium of Glycobiology Editors, La Jolla, California. All rights reserved. p. 113-123.

Campbell KP, Kahl SD. 1989. Association of dystrophin and an integral membrane glycoprotein. *Nature*, 338:259-262.

Cao W, Henry MD, Borrow P, Yamada H, Elder JH, Ravkov EV, Nichol ST, Compans RW, Campbell KP, Oldstone MB. 1998. Identification of alpha-dystroglycan as a receptor for lymphocytic choriomeningitis virus and Lassa fever virus. *Science*, 282:2079-2081.

Chiba A, Matsumura K, Yamada H, Inazu T, Shimizu T, Kusunoki S, Kanazawa I, Kobata A, Endo T. 1997. Structures of sialylated O-linked oligosaccharides of bovine peripheral nerve alpha-dystroglycan. The role of a novel O-mannosyl-type oligosaccharide in the binding of alpha-dystroglycan with laminin. *J Biol Chem*, 272:2156-2162.

Di Costanzo S, Balasubramanian A, Pond HL, Rozkalne A, Pantaleoni C, Saredi S, Gupta VA, Sunu CM, Yu TW, Kang PB, *et al.* 2014. POMK mutations disrupt muscle development leading to a spectrum of neuromuscular presentations. *Human molecular genetics*, 23:5781-5792.

Durbeej M, Henry MD, Campbell KP. 1998. Dystroglycan in development and disease. *Current opinion in cell biology*, 10:594-601.

Endo T. 1999. O-mannosyl glycans in mammals. *Biochimica et biophysica acta*, 1473:237-246.

Endo Y, Dong M, Noguchi S, Ogawa M, Hayashi YK, Kuru S, Sugiyama K, Nagai S, Ozasa S, Nonaka I, *et al.* 2015. Milder forms of muscular dystrophy associated with POMGNT2 mutations. *Neurology. Genetics*, 1:e33.

Ervasti JM, Campbell KP. 1993. A role for the dystrophin-glycoprotein complex as a transmembrane linker between laminin and actin. *J Cell Biol*, 122:809-823.

Falsaperla R, Pratico AD, Ruggieri M, Parano E, Rizzo R, Corsello G, Vitaliti G, Pavone P. 2016. Congenital muscular dystrophy: from muscle to brain. *Italian journal of pediatrics*, 42:78.

Finne J, Krusius T, Margolis RK, Margolis RU. 1979. Novel mannitol-containing oligosaccharides obtained by mild alkaline borohydride treatment of a chondroitin sulfate proteoglycan from brain. *J Biol Chem*, 254:10295-10300.

Godfrey C, Foley AR, Clement E, Muntoni F. 2011. Dystroglycanopathies: coming into focus. *Current Opinion in Genetics & Development*, 21:278-285.

Graham JB, Sunryd JC, Mathavan K, Weir E, Larsen ISB, Halim A, Clausen H, Cousin H, Alfandari D, Hebert DN. 2020. Endoplasmic reticulum transmembrane protein TMTC3 contributes to O-mannosylation of E-cadherin, cellular adherence, and embryonic gastrulation. *Mol Biol Cell*, 31:167-183.

Halmo SM, Singh D, Patel S, Wang S, Edlin M, Boons GJ, Moremen KW, Live D, Wells L. 2017. Protein O-Linked Mannose beta-1,4-N-Acetylglucosaminyl-transferase 2 (POMGNT2) Is a Gatekeeper Enzyme for Functional Glycosylation of alpha-Dystroglycan. *J Biol Chem*, 292:2101-2109.

Haltiwanger RS, Wells L, Freeze HH, Stanley P. 2015. Other Classes of Eukaryotic Glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer

NH, Prestegard JH, et al. editors. *Essentials of Glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press
Copyright 2015-2017 by The Consortium of Glycobiology Editors, La Jolla, California. All rights reserved. p. 151-160.

Hara Y, Kanagawa M, Kunz S, Yoshida-Moriguchi T, Satz JS, Kobayashi YM, Zhu Z, Burden SJ, Oldstone MBA, Campbell KP. 2011. Like-acetylglucosaminyltransferase (LARGE)-dependent modification of dystroglycan at Thr-317/319 is required for laminin binding and arenavirus infection. *Proceedings of the National Academy of Sciences*, 108:17426-17431.

Harrison R, Hitchen PG, Panico M, Morris HR, Mekhaiel D, Pleass RJ, Dell A, Hewitt JE, Haslam SM. 2012. Glycoproteomic characterization of recombinant mouse alpha-dystroglycan. *Glycobiology*, 22:662-675.

Hart GW. 2019. Nutrient regulation of signaling and transcription. *J Biol Chem*, 294:2211-2231.

Henry MD, Campbell KP. 1998. A role for dystroglycan in basement membrane assembly. *Cell*, 95:859-870.

Ibraghimov-Beskrovnaya O, Ervasti JM, Leveille CJ, Slaughter CA, Sernett SW, Campbell KP. 1992. Primary structure of dystrophin-associated glycoproteins linking dystrophin to the extracellular matrix. *Nature*, 355:696-702.

Inamori K, Yoshida-Moriguchi T, Hara Y, Anderson ME, Yu L, Campbell KP. 2012. Dystroglycan function requires xylosyl- and glucuronyltransferase activities of LARGE. *Science*, 335:93-96.

Jae LT, Raaben M, Riemersma M, van Beusekom E, Blomen VA, Velds A, Kerkhoven RM, Carette JE, Topaloglu H, Meinecke P, *et al.* 2013. Deciphering the Glycosylome of Dystroglycanopathies Using Haploid Screens for Lassa Virus Entry. *Science*, 340:479-483.

Jerber J, Zaki MS, Al-Aama JY, Rosti RO, Ben-Omran T, Dikoglu E, Silhavy JL, Caglar C, Musaev D, Albrecht B, *et al.* 2016. Biallelic Mutations in TMTC3, Encoding a Transmembrane and TPR-Containing Protein, Lead to Cobblestone Lissencephaly. *American journal of human genetics*, 99:1181-1189.

Joshi HJ, Narimatsu Y, Schjoldager KT, Tytgat HLP, Aebi M, Clausen H, Halim A. 2018. SnapShot: O-Glycosylation Pathways across Kingdoms. *Cell*, 172:632-632.e632.

Kanagawa M, Saito F, Kunz S, Yoshida-Moriguchi T, Barresi R, Kobayashi YM, Muschler J, Dumanski JP, Michele DE, Oldstone MBA, *et al.* 2004. Molecular Recognition by LARGE Is Essential for Expression of Functional Dystroglycan. *Cell*, 117:953-964.

Krusius T, Finne J, Margolis RK, Margolis RU. 1986. Identification of an O-glycosidic mannose-linked sialylated tetrasaccharide and keratan sulfate oligosaccharides in the chondroitin sulfate proteoglycan of brain. *J Biol Chem*, 261:8237-8242.

Larsen ISB, Narimatsu Y, Clausen H, Joshi HJ, Halim A. 2019. Multiple distinct O-Mannosylation pathways in eukaryotes. *Curr Opin Struct Biol*, 56:171-178.

Larsen ISB, Narimatsu Y, Joshi HJ, Siukstaite L, Harrison OJ, Brasch J, Goodman KM, Hansen L, Shapiro L, Honig B, *et al.* 2017a. Discovery of an O-mannosylation pathway selectively serving cadherins and protocadherins. *Proc Natl Acad Sci U S A*, 114:11163-11168.

Larsen ISB, Narimatsu Y, Joshi HJ, Yang Z, Harrison OJ, Brasch J, Shapiro L, Honig B, Vakhrushev SY, Clausen H, *et al.* 2017b. Mammalian O-mannosylation of cadherins and plexins is independent of protein O-mannosyltransferases 1 and 2. *J Biol Chem*, 292:11586-11598.

Lindahl U, Couchman J, Kimata K, Esko JD. 2015. Proteoglycans and Sulfated Glycosaminoglycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH, Prestegard JH, et al. editors. *Essentials of Glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press
 p. 207-221.

Lommel M, Winterhalter PR, Willer T, Dahlhoff M, Schneider MR, Bartels MF, Renner-Muller I, Ruppert T, Wolf E, Strahl S. 2013. Protein O-mannosylation is crucial for E-cadherin-mediated cell adhesion. *Proc Natl Acad Sci U S A*, 110:21024-21029.

Manya H, Chiba A, Yoshida A, Wang X, Chiba Y, Jigami Y, Margolis RU, Endo T. 2004. Demonstration of mammalian protein O-mannosyltransferase activity: Coexpression of POMT1 and POMT2 required for enzymatic activity. *Proceedings of the National Academy of Sciences of the United States of America*, 101:500-505.

Manzini MC, Tambunan Dimira E, Hill RS, Yu Tim W, Maynard Thomas M, Heinzen Erin L, Shianna Kevin V, Stevens Christine R, Partlow Jennifer N, Barry Brenda J, *et al.* 2012. Exome Sequencing and Functional Validation in Zebrafish Identify GTDC2 Mutations as a Cause of Walker-Warburg Syndrome. *The American Journal of Human Genetics*, 91:541-547.

Neelamegham S, Aoki-Kinoshita K, Bolton E, Frank M, Lisacek F, Lütteke T, O'Boyle N, Packer NH, Stanley P, Toukach P, *et al.* 2019. Updates to the Symbol Nomenclature for Glycans guidelines. *Glycobiology*, 29:620-624.

Nilsson J, Nilsson J, Larson G, Grahn A. 2010. Characterization of site-specific O-glycan structures within the mucin-like domain of α-dystroglycan from human skeletal muscle. *Glycobiology*, 20:1160-1169.

Ogawa M, Nakamura N, Nakayama Y, Kurosaka A, Manya H, Kanagawa M, Endo T, Furukawa K, Okajima T. 2013. GTDC2 modifies O-mannosylated α-dystroglycan in the endoplasmic reticulum to generate N-acetyl glucosamine epitopes reactive with CTD110.6 antibody. *Biochemical and Biophysical Research Communications*, 440:88-93.

Pacharra S, Hanisch FG, Breloy I. 2012. Neurofascin 186 is O-mannosylated within and outside of the mucin domain. *Journal of proteome research*, 11:3955-3964.

Pacharra S, Hanisch FG, Muhlenhoff M, Faissner A, Rauch U, Breloy I. 2013. The lecticans of mammalian brain perineural net are O-mannosylated. *Journal of proteome research*, 12:1764-1771.

Praissman JL, Wells L. 2014. Mammalian O-Mannosylation Pathway: Glycan Structures, Enzymes, and Protein Substrates. *Biochemistry*, 53:3066-3078.

Rambukkana A, Yamada H, Zanazzi G, Mathus T, Salzer JL, Yurchenco PD, Campbell KP, Fischetti VA. 1998. Role of alpha-dystroglycan as a Schwann cell receptor for Mycobacterium leprae. *Science*, 282:2076-2079.

Saito F, Saito-Arai Y, Nakamura A, Shimizu T, Matsumura K. 2008. Processing and secretion of the N-terminal domain of alpha-dystroglycan in cell culture media. *FEBS letters*, 582:439-444.

Sasaki T, Yamada H, Matsumura K, Shimizu T, Kobata A, Endo T. 1998. Detection of O-mannosyl glycans in rabbit skeletal muscle alpha-dystroglycan. *Biochimica et biophysica acta*, 1425:599-606.

Sheikh MO, Halmo SM, Wells L. 2017. Recent advancements in understanding mammalian O-mannosylation. *Glycobiology*, 27:806-819.

Singh J, Itahana Y, Knight-Krajewski S, Kanagawa M, Campbell KP, Bissell MJ, Muschler J. 2004. Proteolytic enzymes and altered glycosylation modulate dystroglycan function in carcinoma cells. *Cancer Res*, 64:6152-6159.

Smalheiser NR, Haslam SM, Sutton-Smith M, Morris HR, Dell A. 1998. Structural analysis of sequences O-linked to mannose reveals a novel Lewis X structure in cranin (dystroglycan) purified from sheep brain. *J Biol Chem*, 273:23698-23703.

Stalnaker SH, Aoki K, Lim JM, Porterfield M, Liu M, Satz JS, Buskirk S, Xiong Y, Zhang P, Campbell KP, *et al.* 2011. Glycomic analyses of mouse models of congenital muscular dystrophy. *J Biol Chem*, 286:21180-21190.

Stalnaker SH, Hashmi S, Lim J-M, Aoki K, Porterfield M, Gutierrez-Sanchez G, Wheeler J, Ervasti JM, Bergmann C, Tiemeyer M, *et al.* 2010. Site Mapping and Characterization of O-Glycan Structures on α-Dystroglycan Isolated from Rabbit Skeletal Muscle. *The Journal of Biological Chemistry*, 285:24882-24891.

Stanley P, Taniguchi N, Aebi M. 2015. N-Glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH, Prestegard JH, et al. editors. *Essentials of Glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press
Copyright 2015-2017 by The Consortium of Glycobiology Editors, La Jolla, California. All rights reserved. p. 99-111.

Stevens E, Carss KJ, Cirak S, Foley AR, Torelli S, Willer T, Tambunan DE, Yau S, Brodd L, Sewry CA, *et al.* 2013. Mutations in B3GALNT2 cause congenital muscular dystrophy and hypoglycosylation of α-dystroglycan. *American journal of human genetics*, 92:354-365.

Sunryd JC, Cheon B, Graham JB, Giorda KM, Fissore RA, Hebert DN. 2014. TMTC1 and TMTC2 are novel endoplasmic reticulum TPR-containing adapter proteins involved in calcium homeostasis. *Journal of Biological Chemistry*.

Takeuchi H, Haltiwanger RS. 2014. Significance of glycosylation in Notch signaling. *Biochem Biophys Res Commun*, 453:235-242.

Tran DT, Ten Hagen KG. 2013. Mucin-type O-glycosylation during development. *J Biol Chem*, 288:6921-6929.

Trinidad JC, Schoepfer R, Burlingame AL, Medzihradszky KF. 2013. N- and O-glycosylation in the murine synaptosome. *Molecular & cellular proteomics : MCP*, 12:3474-3488.

Varki A. 2017. Biological roles of glycans. *Glycobiology*, 27:3-49.

Varki A, Cummings RD, Aebi M, Packer NH, Seeberger PH, Esko JD, Stanley P, Hart G, Darvill A, Kinoshita T, *et al.* 2015. Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology*, 25:1323-1324.

Vester-Christensen MB, Halim A, Joshi HJ, Steentoft C, Bennett EP, Levery SB, Vakhrushev SY, Clausen H. 2013. Mining the O-mannose glycoproteome reveals cadherins as major O-mannosylated glycoproteins. *Proceedings of the National Academy of Sciences*, 110:21018-21023.

Wing DR, Rademacher TW, Schmitz B, Schachner M, Dwek RA. 1992. Comparative glycosylation in neural adhesion molecules. *Biochem Soc Trans*, 20:386-390.

Winterhalter PR, Lommel M, Ruppert T, Strahl S. 2013. O-glycosylation of the non-canonical T-cadherin from rabbit skeletal muscle by single mannose residues. *FEBS letters*, 587:3715-3721.

Yagi H, Nakagawa N, Saito T, Kiyonari H, Abe T, Toda T, Wu S-W, Khoo K-H, Oka S, Kato K. 2013. AGO61-dependent GlcNAc modification primes the formation of functional glycans on α-dystroglycan. *Scientific Reports*, 3:3288.

Yoshida A, Kobayashi K, Manya H, Taniguchi K, Kano H, Mizuno M, Inazu T, Mitsuhashi H, Takahashi S, Takeuchi M, *et al.* 2001. Muscular dystrophy and neuronal migration disorder caused by mutations in a glycosyltransferase, POMGnT1. *Developmental cell*, 1:717-724.

Yoshida-Moriguchi T, Campbell KP. 2015. Matriglycan: a novel polysaccharide that links dystroglycan to the basement membrane. *Glycobiology*, 25:702-713.

Yoshida-Moriguchi T, Willer T, Anderson ME, Venzke D, Whyte T, Muntoni F, Lee H, Nelson SF, Yu L, Campbell KP. 2013. SGK196 Is a Glycosylation-Specific O-Mannose Kinase Required for Dystroglycan Function. *Science*, 341:896-899.

Yuen CT, Chai W, Loveless RW, Lawson AM, Margolis RU, Feizi T. 1997. Brain contains HNK-1 immunoreactive O-glycans of the sulfoglucuronyl lactosamine series that terminate in 2-linked or 2,6-linked hexose (mannose). *J Biol Chem*, 272:8924-8931.

CHAPTER 2

RECENT ADVANCEMENTS IN UNDERSTANDING MAMMALIAN *O*-MANNOSYLATION[1]

[1] Halmo, S.M.[*], Sheikh, M.O.[*], and Wells, L. 2017. *Glycobiology*. 27(9): 806–819.
[*]These authors contributed equally to this work.
Reprinted here with permission of the publisher.

Abstract

The post-translational glycosylation of select proteins by *O*-linked mannose (*O*-mannose or *O*-man) is a conserved modification from yeast to humans and has been shown to be necessary for proper development and growth. The most well studied *O*-mannosylated mammalian protein is α-dystroglycan (α-DG). Hypoglycosylation of α-DG results in varying severities of congenital muscular dystrophies, cancer progression and metastasis, and inhibited entry and infection of certain arenaviruses. Defects in the gene products responsible for post-translational modification of α-DG, primarily glycosyltransferases, are the basis for these diseases. The multitude of clinical phenotypes resulting from defective *O*-mannosylation highlights the biomedical significance of this unique modification. Elucidation of the various *O*-mannose biosynthetic pathways is imperative to understanding a broad range of human diseases and for the development of novel therapeutics. In this review, we will focus on recent discoveries delineating the various enzymes, structures and functions associated with *O*-mannose-initiated glycoproteins. Additionally, we discuss current gaps in our knowledge of mammalian *O*-mannosylation, discuss the evolution of this pathway, and illustrate the utility and limitations of model systems to study functions of *O*-mannosylation.

Introduction

Oxygen-linked alpha-mannose (hereinafter referred to as either *O*-mannose or *O*-man) covalently attached to serine or threonine residues was discovered nearly 50 years ago in yeast (Sentandreu and Northcote 1968) and nearly 40 years ago in rat brain (Finne et al. 1979). This post-translational modification is conserved from fungi to humans and plays a role in a wide variety of human diseases, such as congenital muscular dystrophy (CMD), cancer metastasis and viral entry (Dobson et al. 2013; Panin and Wells 2014). Almost 20 years ago, it was discovered that the causative genes for CMD were primarily glycosyltransferase (GT) genes that specifically modify the most well studied *O*-mannosylated protein, α-dystroglycan (α-DG) (Michele et al. 2002; Moore et al. 2002; Muntoni et al. 2002; Yoshida et al. 2001). The human *DAG1* gene (dystrophin-associated glycoprotein 1) encodes the 895-residue dystroglycan precursor protein, which is post-translationally cleaved into the peripheral membrane subunit (α-DG, residues 1–653) and the transmembrane subunit (β-DG, residues 654–895) (Holt et al. 2000). α-DG is post-translationally processed in the secretory pathway by furin to generate the mature α-DG (313–653). In addition to being modified by *N*-linked glycans (Ervasti and Campbell 1991), α-DG is highly *O*-glycosylated (serine- or threonine-linked) within its mucin-like domain (residues 316–485) which includes mucin-type, *O*-GalNAc (*N*-acetylgalactosamine)-initiated structures as well as *O*-mannose-initiated structures (Barresi and Campbell 2006; Chiba et al. 1997; Praissman and Wells 2014; Sasaki et al. 1998; Smalheiser et al. 1998; Wells 2013). The best-characterized function of α-DG is its role in the dystrophin-glycoprotein complex (DGC), in which α-DG contributes a glycan-dependent link between the actin cytoskeleton and extracellular matrix (ECM) (Ervasti and Campbell 1993) (Figure 2.1 A).

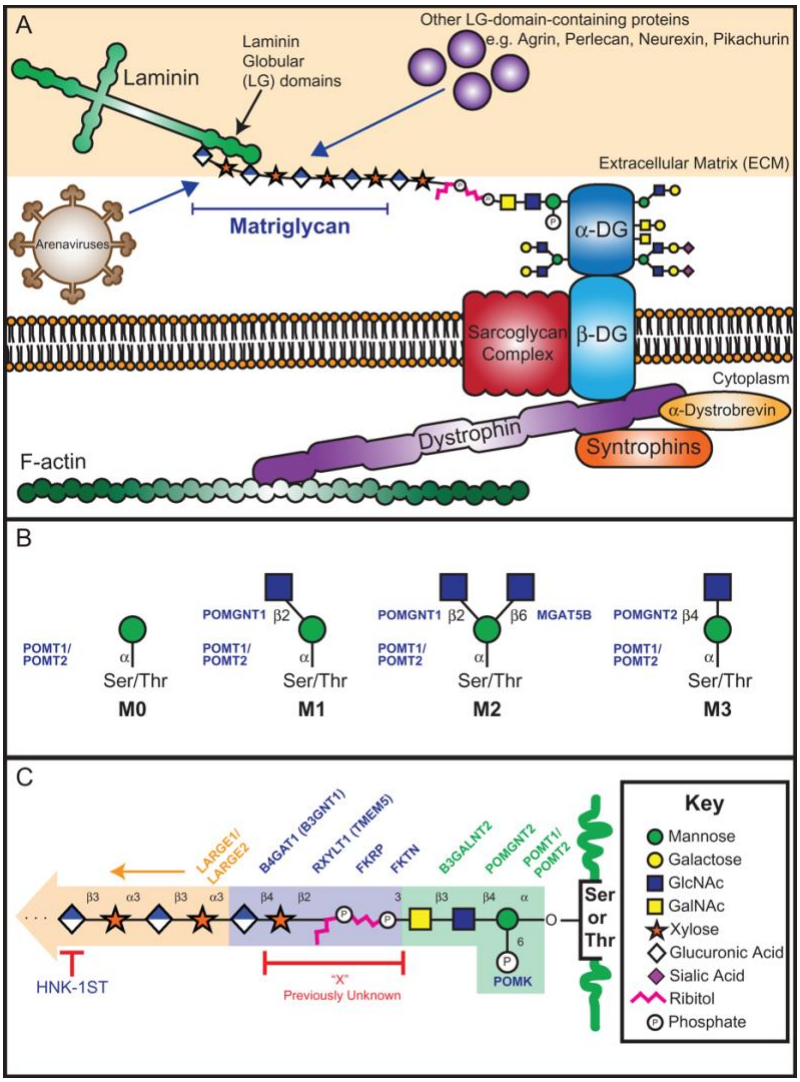


Figure 2.1. The DGC and core *O*-mannose structures.
(**A**) The DGC consists of the cytosolic dystrophin protein which connects the actin cytoskeleton to other intracellular and extracellular proteins. The heavily glycosylated peripheral membrane protein α-DG participates in the DGC and mediates interactions with ECM proteins through the GAG-like repeating disaccharide, matriglycan. In addition to binding to LG-domain-containing proteins, matriglycan has been implicated in interactions with certain arenaviruses which utilize α-DG as their primary cell-surface receptor.
(**B**) Four classifications of core *O*-mannose structures are presented here. Enzymes that catalyze the sugar transfer are indicated.
(**C**) Summary of the fully elaborated core M3 functional glycan. Previously unknown region of this structure ("X") is indicated in red. The phospho-trisaccharide, as well as the enzymes that catalyze the indicated sugar or phosphate transfer for this region, are highlighted in green. The linker and priming region as well as the linker and priming enzymes are indicated in blue, while matriglycan is highlighted in orange. Glycan symbol representation is adapted from Varki et al. (2015).

The biosynthesis of *O*-mannose-initiated glycans begins on the cytosolic face of the endoplasmic reticulum (ER) with the generation of the lipid-linked mannosyl donor molecule dolichol-phosphate mannose (Dol-P-Man or DPM). The enzymatic activity of the DPM synthase complex (DPMS, comprised of DPM1, DPM2 and DPM3) catalyzes the transfer of mannose from guanidine diphosphate mannose (GDP-Man) to dolichol-phosphate (Maeda et al. 2000). In addition to *O*-mannosylation, various mannosyltransferases involved in asparagine-linked (*N*-linked) glycosylation (Aebi 2013), tryptophan-linked *C*-mannosylation (Doucey et al. 1998) and glycosylphosphatidylinositol-anchor biosynthesis (Kang et al. 2005; Orlean 1990) are dependent on Dol-P-Man as a donor substrate.

Four major classifications of core *O*-mannose-initiated glycans are proposed in Figure 2.1 B, where the last three have been extensively reviewed by Praissman and Wells (2014) and Yoshida-Moriguchi et al. (2013). Hereinafter, we will refer to core *O*-mannose substructures consistent with the nomenclature proposed by Praissman and Wells (2014). *O*-mannose glycan biosynthesis is initiated in the ER by the Protein *O*-mannosyltransferase 1 (POMT1) and Protein *O*-mannosyltransferase 2 (POMT2) enzyme complex which catalyzes the transfer of mannose from Dol-P-Man to the hydroxyl oxygen of serine or threonine side chains (Manya et al. 2004). We refer to a serine or threonine residue linked with a single α-mannose as core M0. Core M0 can be modified with β1,2-linked *N*-Acetylglucosamine (GlcNAc) by Protein *O*-Linked Mannose *N*-Acetyl-glucosaminyltransferase 1 (POMGNT1), resulting in the core M1 structure. The core M1 structure can be modified with β1,6-linked GlcNAc by the activity of Mannosyl (α1,6)-Glycoprotein β1,6-*N*-Acetyl-Glucosaminyltransferase (MGAT5B), yielding the core M2 structure. Core M3 glycans are generated by the addition of β1,4-linked GlcNAc to

core M0 that is catalyzed by Protein *O*-Linked Mannose *N*-Acetyl-glucosaminyltransferase 2 (POMGNT2).

While M1 and M2 cores can be further elaborated, most of the recent research focus has been on the M3 core that can become functionally modified to bind laminin globular (LG)-domain-containing proteins. The core M3 structure is extended with β1,3-linked GalNAc by β-1,3-*N*-Acetylgalactosaminyltransferase 2 (B3GALNT2). After assembly of this trisaccharide, the six-position of mannose is subject to phosphorylation by Protein *O*-Mannose Kinase (POMK, formerly SGK196) in the ER, in which this reaction product will be referred to as the "phospho-trisaccharide" throughout this review (Yoshida-Moriguchi et al. 2013). Once formed, the phospho-trisaccharide is further elaborated by the activities of the recently discovered phospho-ribitol transferases and the *priming enzymes*, which will be discussed later. Finally, the synthesis of matriglycan, a glycosaminoglycan (GAG)-like repeating disaccharide of xylose (Xyl) and glucuronic acid (GlcA) ([-Xyl-α3-GlcA-β3-]$_n$), is polymerized on the primed structure by the activities of the bifunctional GTs LARGE1 (formerly LARGE) and/or its paralog LARGE2 (formerly GYLTL1B) in the Golgi apparatus (Ashikov et al. 2013; Inamori et al. 2012, 2013) (Figure 2.1 C). Matriglycan is the functional epitope that binds the LG-domain-containing proteins in the ECM. This linkage appears to be critical for the structural integrity of skeletal muscle and proper brain development (Yoshida-Moriguchi and Campbell 2015).

A disruption of this link between α-DG and the ECM results in a subset of CMDs. When mutations in specific genes result in hypoglycosylation or loss of α-DG, these diseases are referred to as dystroglycanopathies. The evolving nomenclature of dystroglycanopathies is chiefly classified by the origin of the genetic defect, where *primary* dystroglycanopathies result from mutation in the *DAG1* gene (Geis et al. 2013; Hara, Balci-Hayta et al. 2011),

while *secondary* dystroglycanopathies result from defects in any of the genes encoding enzymes directly modifying of α-DG (Beedle et al. 2012; Brockington et al. 2010) (Table 2.1). Secondary dystroglycanopathies include a wide range of muscular dystrophic phenotypes including the most severe case of Walker–Warburg syndrome (WWS) to the less severe limb-girdle muscular dystrophies (LGMDs). Recently, *tertiary* dystroglycanopathies have been described as originating from defects in genes required for the proper biosynthesis of donor substrate molecules, such as Dol-P-Man or CDP-ribitol, used by the α-DG modifying enzymes and will be discussed further in this review (Figure 2.2) (Barone et al. 2012; Lefeber et al. 2009, 2011; Riemersma, Froese et al. 2015; Yang et al. 2013).

An astonishing amount of progress has been made by multiple laboratories over the past 4 years in order to elucidate the central link between the core M3 phospho-trisaccharide and matriglycan. Until 2016; the linker region (which has been frequently named "X" in the literature) was unknown (Figure 2.1 C). The identification of previously unknown activities of multiple causal gene products including B4GAT1 (formerly B3GNT1), TMEM5, ISPD, FKTN, and FKRP further assisted in the complete assignment of the unknown linker region. Structural determination of both POMGNT1 and POMK and mechanistic studies on POMGNT1 and POMGNT2 provided insight into substrate specificity and regulation. This review focuses on recently published work, expands on the current gaps in the field, and highlights future directions.

Phosphodiester linkages connecting matriglycan to α-DG

In 2010, it was inferred that the functional glycan, matriglycan was connected to core M3 substructures by an unknown moiety known as "X" (Figure 2.1 C), presumably via a phosphate group linked to position 6 of α-mannose of the core M3 phospho-trisaccharide [GalNAc-β1,3-GlcNAc-β1,4(6-phospho)-Man] (Yoshida-Moriguchi et al. 2010). These conclusions were made

Table 2.1. Proteins involved with functional glycosylation of α-DG.

Gene Product Name (Former Name)	UniProt ID	Function	Core Glycan	Subcellular Localization	Dystroglycanopathy classification based on mutation
DAG1	Q14118	extracellular glycoprotein that acts as a receptor for LG-domain containing ECM proteins		Plasma membrane/ Extracellular space	Primary
POMT1	Q9Y6A1	transfers α-mannose from DPM to serine or threonine residues in a complex with POMT2	M0, M1, M2, M3	ER	Secondary
POMT2	Q9UKY4	transfers α-mannose from DPM to serine or threonine residues in a complex with POMT1	M0, M1, M2, M3	ER	Secondary
POMGNT1	Q8WZA1	transfers β1,2-GlcNAc to *O*-mannose on serine or threonine residues	M1, M2	Golgi	Secondary
POMGNT2 (AGO61, GTDC2)	Q8NAT1	transfers β1,4-GlcNAc to *O*-mannose on serine or threonine residues	M3	ER	Secondary
MGAT5B (GNT-VB, GNT-IX)	Q3V5L5	transfers β1,6-GlcNAc to *O*-mannose on serine or threonine residues	M2	Golgi	*Not Observed*
B3GALNT2	Q8NCR0	transfers β1,3-GalNAc to core M3	M3	ER	Secondary
POMK (SGK196)	Q9H5K3	Carbohydrate kinase that phosphorylates 6 position of *O*-mannose	M3	ER	Secondary
FKTN (FCMD)	O75072	transfers a ribitol-phosphate to the core M3 trisaccharide at the 3 position of GalNAc in a phosphodiester linkage	M3	Golgi	Secondary
FKRP	Q9H9S5	transfers a ribitol-phosphate to the ribitol-phosphate product of FKTN in a phosphodiester linkage	M3	Golgi	Secondary
B2XYLT1 (TMEM5)	Q9Y2B1	transfers β1,2-xylose to ribitol	M3	Golgi	Secondary
B4GAT1 (B3GNT1)	O43505	transfers a β1,4-glucuronic acid to an underlying xylose	M3	Golgi	Secondary
LARGE1 (LARGE)	O95461	polymerizes an α1,3-xylose-β1,3-glucuronic acid repeat	M3	Golgi	Secondary
LARGE2 (GYLTL1B)	Q8N3Y3	polymerizes an α1,3-xylose-β1,3-glucuronic acid repeat	M3	Golgi	*Not Observed*
HNK-1ST (CHST10)	O43529	presumably sulfates terminal glucuronic acid of matriglycan	M3	Golgi	*Not Observed*
DPM1	O60762	transfers mannose from GDP-mannose to dolichol monophosphate to form dolichol phosphate mannose (Dol-P-Man)	M1, M2, M3	ER	Tertiary
DPM2	O94777	regulatory subunit of the DPM synthase complex	M1, M2, M3	ER	Tertiary
DPM3	Q9P2X0	tethers catalytic subunit DPM1 to the ER	M1, M2, M3	ER	Tertiary
DOLK	Q9UPQ8	phosphorylates dolichol to produce dolichol-phosphate	M1, M2, M3	ER	Tertiary
CRPPA (ISPD)	A4D126	catalyzes the transfer of a pyrophosphate group from CTP to synthesize CDP-ribitol	M3	Cytosolic	Tertiary
HK	P19367 P52790 P52789	phosphorylates mannose to produce mannose-6-phosphate	M1, M2, M3	Mitochondrial/ Cytosolic	*Not Observed*
MPI	P34949	catalyzes the isomerization of mannose-6-phosphate and fructose-6-phosphate	M1, M2, M3	Cytosolic	Tertiary[1,2]
PMM2	O15305	catalyzes the isomerization of mannose-6-phosphate to mannose-1-phosphate	M1, M2, M3	Cytosolic	Tertiary[1,2]
GMPPB	Q9Y5P6	synthesizes GDP-mannose from GTP and mannose-1-phosphate	M1, M2, M3	Cytosolic	Tertiary[1]

[1] Indicates CDG
[2] Indicates *Putative* Dystroglycanopathy

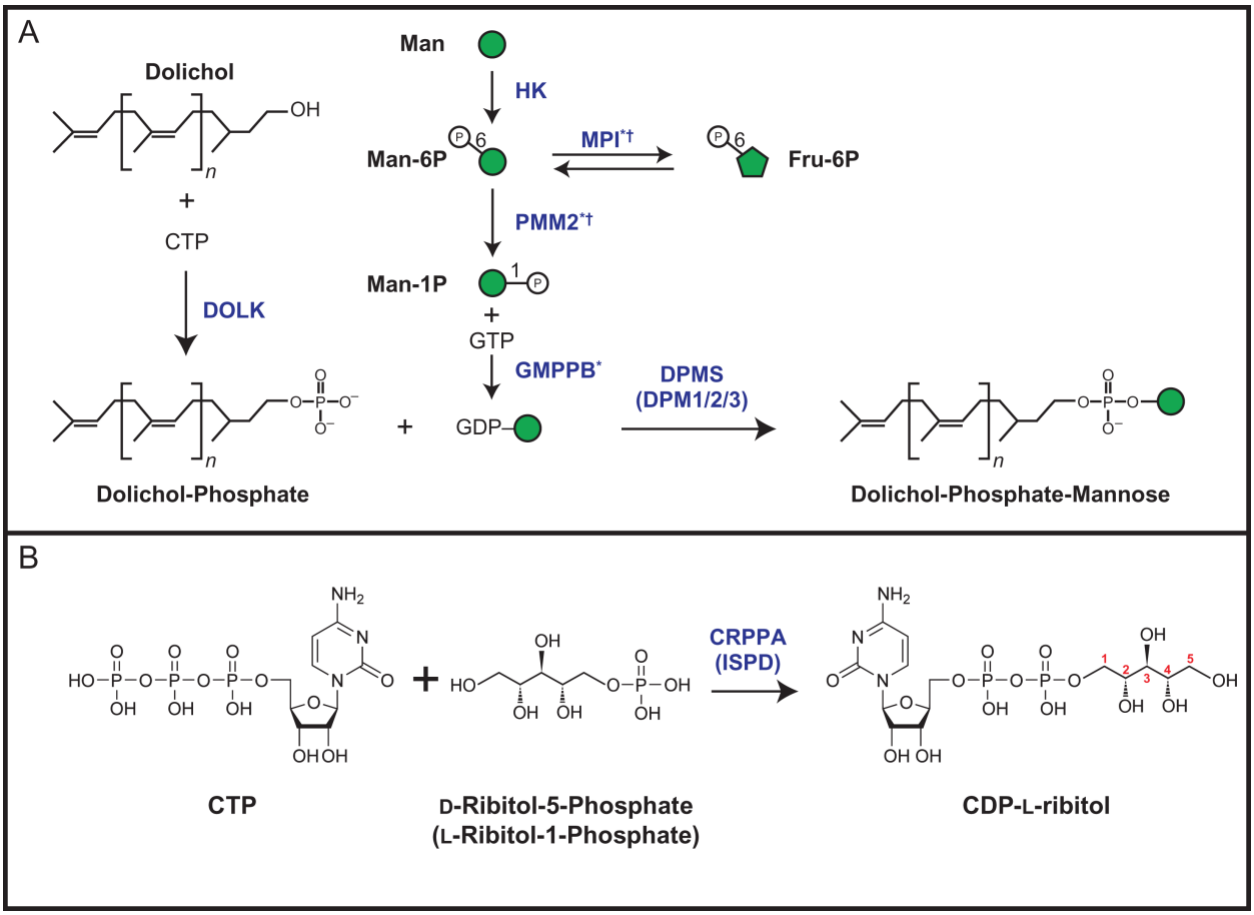


Figure 2.2. Donor synthesis of Dol-P-Man and CDP-ribitol.
(**A**) Reaction scheme for the synthesis of Dol-P-Man. Dolichol is phosphorylated by the CTP-mediated kinase activity of Dolichol kinase (DOLK) to form Dolichol-Phosphate (Dol-P). To generate GDP-Man, Mannose (Man) is phosphorylated by Hexokinase (HK) to yield Mannose-6-Phosphate (Man-6P) and can undergo isomerization by the activity of Mannose-6-Phosphate Isomerase (MPI) for conversion to Fructose-6-Phosphate (Fru-6P). Phosphomannomutase 2 (PMM2) converts Man-6P to Man-1P. Man-1P and GTP is converted to GDP-Man by the activity of GDP-Mannose Pyrophosphorylase B (GMPPB). Dolichol-Phosphate-Mannose (Dol-P-Man) is synthesized from Dol-P and GDP-Man by the activities of the DPM synthase (DPMS or DPM1/2/3) complex. A superscript *indicates that the gene has been implicated in CDG (congenital disorders of glycosylation) and †indicates a *putative* dystroglycanopathy [(Belaya et al. 2015; Luo et al. 2017; Schollen et al. 2000; Sparks et al. 1993), see Table 2.1].
(**B**) Reaction scheme for the synthesis of CDP-L-ribitol from CTP and ribitol-5-phosphate by the activity of CRPPA (ISPD). Carbon numbering assignments for ribitol in CDP-L-ribitol, based on IUPAC rules, are shown in red.

from experiments demonstrating that reactivity with an α-DG glyco-specific antibody, IIH6, was abolished upon treatment of α-DG with aqueous hydrofluoric acid (HFaq) which chemically cleaves phosphodiester bonds (Ilg et al. 1996). Interestingly, phosphate analysis of native α-DG purified from rabbit skeletal muscle revealed 4–5 moles of phosphate per mole of protein (Yoshida-Moriguchi et al. 2010), which could either be attributed to multiple sites of core M3-type modification or additional phosphates within the functional glycan. In 2016, our group (Praissman et al. 2016) and others (Kanagawa et al. 2016) identified ribitol-phosphate (RboP)-containing species within the functional glycan on α-DG by mass spectrometry. Kanagawa et al. demonstrated that the RboP incorporation occurred in tandem and was linked to carbon-3 of GalNAc, instead of the previously hypothesized mannose-6-phosphate of the core M3 glycan (Kanagawa et al. 2016) (Figure 2.1 C). Interestingly, the Kato and Khoo laboratories detected a novel, lower abundance modification of a single glycerol-phosphate (GroP) moiety linked to the core M3 phospho-trisaccharide on truncated, recombinant α-DG (Yagi et al. 2016). In this study, no tandem GroP or GroP-RboP modifications were detected, nor was the GroP modification extended with Xyl-GlcA repeats, which suggests that GroP might serve as a molecular brake for core M3 functional glycan synthesis under certain physiological conditions. The discovery of tandem ribitol-phosphate moieties within the α-DG functional glycan identifies a gap in the symbol representation (Varki et al. 2015). Therefore, we suggest to represent the linear ribitol (reduced ribose) as a pink zigzag line containing five points, as shown in Figure 2.1.

CDP-l-ribitol - A Novel "Nucleotide-Sugar Alcohol" In Mammals

Having determined that RboP is a part of the functional M3 glycan brought up the question of the donor for the FKTN and FKRP enzymes. One hint came from an effort to identify host factors required for Lassa virus (LASV) cellular entry, which is known to utilize α-DG as its

primary cellular receptor (Cao et al. 1998; Oldstone and Campbell 2011). Using a gene trap–insertion screen in the near-haploid HAP1 human cell line (Carette et al. 2011) the dystroglycan gene *DAG1*, in addition to other genes known to be causal for CMD and involved with α-DG modification were identified (Table 2.1) (Jae et al. 2013). Among these genes, *ISPD* (isoprenoid synthase domain-containing) was of unknown function and not predicted to have any GT domains. *ISPD* was of particular interest due to reports of WWS patients harboring mutations in this gene (Roscioli et al. 2012; Willer et al. 2012). Near the end of 2015, four independent research groups demonstrated that recombinant ISPD was able to utilize CTP and D-ribitol-5-phosphate (as well as ribose-5-phosphate or ribulose-5-phosphate) to generate CDP-L-ribitol (or CDP-ribose or CDP-ribulose, respectively, Figure 2.2 B) (Gerin et al. 2016; Kanagawa et al. 2016; Praissman et al. 2016; Riemersma, Froese et al. 2015). Note that by IUPAC convention as well as stereospecific numbering nomenclature D-ribitol-5-phosphate is the preferred name but identical to L-ribitol-1-phosphate and only CDP-L-ribitol (alternatively referred to as CDP 5-ester of D-ribitol) exists for this polyol nucleotide where the carbon in ribitol nearest the phosphate is the 1 carbon (Korte et al. 1976). While all groups converged on demonstrating ISPD's enzymatic activity and requirement for α-DG glycosylation, three of these studies provided unique contributions to understanding the function of ISPD. Riemersma et al. solved the 2.4 Å crystal structure of human ISPD (residues 43–451) and mapped disease causing mutations to the N-terminal cytidyltransferase domain (Riemersma, Froese et al. 2015), while the Bommer laboratory successfully detected a CDP-pentitol, likely CDP-L-ribitol (CDP-Rbo), in rat muscle and mouse myotubes (Gerin et al. 2016). Furthermore, Kanagawa et al. demonstrated that supplementation of CDP-Rbo to cells deficient in ISPD can restore functional glycosylation of α-DG, and that further investigation of CDP-Rbo

supplementation therapy using animal models should be considered as a potential therapeutic (Kanagawa et al. 2016).

Interestingly, all groups have defined the enzymatic activity of ISPD in similar, yet alternative ways. For instance, Kanagawa et al. use the broadest classification of ISPD as a CDP-ribitol synthase (i.e. an enzyme that catalyzes the linking together of two molecules). Riemersma et al. classify ISPD as a cytidyltransferase, where a transferase is an enzyme that catalyzes the transfer of a particular moiety from one molecule to another, consistent with the classification used in defining homologous enzymes, such as TarI, in bacterial systems (Baur et al. 2009). However, following mammalian enzyme nomenclature, ISPD is most appropriately defined as a pyrophosphorylase (i.e. an enzyme that catalyzes the transfer of a *pyrophosphate* group from one molecule to another). Praissman et al. and Gerin et al. define ISPD as a CDP-ribitol pyrophosphorylase (Gerin et al. 2016; Praissman et al. 2016). This terminology is more specific than synthase and is most consistent with the naming schemes of other mammalian sugar-nucleotide biosynthetic enzymes, like GDP-Mannose Pyrophosphorylase B (GMPPB). Accordingly, we propose to rename ISPD to CDP-L-ribitol (ribose, ribulose) pyrophosphorylase A, or **CRPPA**.

It is puzzling why this cytidine-containing nucleotide-sugar was not previously identified in mammals. However, it may be due to the coelution of CDP-glucose during sugar-nucleotide analysis by HPLC (Gerin et al. 2016). Outstanding areas of research include identification of the pentose reductase and the location of this activity in the CDP-Rbo biosynthetic pathway in mammals (i.e. does the reductase act on ribose-5-phosphate and then CRPPA acts or does CRPPA convert ribose-5-phosphate to CDP-ribose and then is it reduced to CDP-ribitol?). Upon identification of a pentose reductase involved in this pathway, it would be intriguing to determine

if any mutations in the reductase are causal for aberrant α-DG glycosylation as multiple dystroglycanopathies are still of unknown genetic etiology.

Another gene identified in the gene trap–insertion screen (Carette et al. 2011) was S*LC35A1* (Solute Carrier Family 35 Member A1), a CMP-sialic acid transporter (Patnaik and Stanley 2006). *SLC35A1* mutations are causal for CMD and result in defective α-DG glycosylation (Riemersma, Sandrock et al. 2015). While it has yet to be demonstrated, it is enticing to hypothesize that SLC35A1 is also a CDP-Rbo transporter as *SLC35A1*-deficient HAP1 cells lack the functional glycan as detected by IIH6 staining, independent of sialic acid (Riemersma, Sandrock et al. 2015).

Fukutin and FKRP - enzymes that utilize CDP-l-ribitol

LARGE1/2-mediated matriglycan synthesis on core M3 glycans requires extension of the phospho-trisaccharide by several recently characterized phosphoglycosyltransferases (Figure 2.1 C). The genes *FKTN* (Fukutin) and *FKRP* (Fukutin related protein) were initially predicted to encode phosphoryl-ligand transferases based on sequence analysis (Aravind and Koonin 1999) and were established to be *medial*-Golgi-resident proteins (Esapa et al. 2002; Lynch et al. 2012). Mutations in *FKTN* and *FKRP* result in α-DG hypoglycosylation (Brockington et al. 2001; Kobayashi et al. 1998) and are causative for Fukuyama-type CMD (FCMD), LGMD and WWS (Taniguchi-Ikeda et al. 2016). Recently, recombinantly expressed FKTN, lacking the transmembrane domain, was shown to transfer RboP from CDP-Rbo to a phospho-trisaccharide peptide or purified fragments of rabbit α-DG, and once this transfer has occurred, FKRP can then transfer the second RboP group (Gerin et al. 2016; Kanagawa et al. 2016). Nuclear magnetic resonance (NMR) analyses of the reaction products revealed that FKTN transfers alditol-1-P to the C3 position of GalNAc through a phosphodiester linkage and FKRP transfers the second RboP to

the C5 (note: we refer to this as C5 using IUPAC nomenclature) position of the underlying RboP by a phosphodiester linkage (Kanagawa et al. 2016) (Figure 2.1 C). Thus, this tandem RboP addition mediated by the sequential phosphoglycosyltransferase activities of FKTN and FKRP, respectively, occurs in the Golgi after addition of the phosphate to the 6-position of the M3 trisaccharide by POMK in the ER (Yoshida-Moriguchi et al. 2013).

Despite understanding of the enzymatic activities of FKTN and FKRP, further studies are warranted in order to establish their unique donor and acceptor substrate specificities. Comparative structural analyses of FKTN and FKRP will aid in understanding the acceptor substrate requirements and processive nature of this enzyme pair. It is unclear if GTs that extend the phospho-trisaccharide can recognize the distal 6-phosphate on the core M3 mannose (transferred by POMK) as a part of their functional specificity. In regard to the GroP modification mentioned above (Yagi et al. 2016), it is unknown if human ISPD can synthesize CDP-glycerol using CTP and glycerol-3-phosphate, however, the homologous bacterial IspD enzyme has been shown to catalyze such a reaction (Majumdar et al. 2009). Regardless of the source of CDP-glycerol, it has yet to be determined if FKTN or FKRP can transfer GroP from CDP-glycerol. Given that only GroP modifications on the phospho-trisaccharide were identified and that no tandem GroP or GroP-RboP modifications were detected, one hypothesis is that FKTN (and not FKRP) might be able to transfer the initial GroP from CDP-glycerol and that the stringency of FKRP's acceptor substrate specificity prohibits priming.

The priming enzymes

Collectively, we refer to B4GAT1 and TMEM5 as the *priming enzymes*. The β1,4-glucuronyltransferase B4GAT1 (formerly B3GNT1) was identified in 2014 as the enzyme responsible for adding β1,4-linked GlcA to an underlying β-linked Xyl which serves as a primer

for extension by LARGE1 (Praissman et al. 2014; Willer et al. 2014). Among other α-DG-related gene products of unknown function, the Wells and Campbell groups showed that recombinant TMEM5 can hydrolyze UDP-Xyl in the absence of any acceptor substrate and could transfer ^{14}C-labeled Xyl to a truncated α-DG construct (α-DG-Fc340) that was expressed in a *TMEM5*-deficient patient cell line, demonstrating that TMEM5 is a xylosyltransferase (Praissman et al. 2016). Subsequently, Manya et al. established that TMEM5 transfers Xyl in a β-linkage to the ribitol moiety of the FKRP reaction product based on NMR studies, and this TMEM5 reaction product is an acceptor substrate for B4GAT1 (Manya et al. 2016). The anomeric configuration of Xyl, as predicted by Praissman et al. (2014) and Willer et al. (2014), is in complete agreement with the substrate specificity of B4GAT1, which can complete priming of the β-linked Xyl via transfer of β1,4-linked GlcA and allow for LARGE1/2 to synthesize matriglycan (Figure 2.1 C).

Manya et al. identified the Xyl-linkage position of the TMEM5 reaction product using NMR spectroscopy and refer to TMEM5 as a ribitol β-*1,4* Xylosyltransferase (Manya et al. 2016). This carbon numbering is based on the fact that D-ribitol-5-phosphate is derived from D-ribose-5-phosphate and CDP-ribitol is synthesized from CTP and D-ribitol-5-phosphate. The IUPAC name, however, for CDP-ribitol is {[[(2 R,3 S,4 R,5 R)-5-(4-amino-2-oxopyrimidin-1-yl)-3,4-dihydroxyoxolan-2-yl]methoxy-hydroxyphosphoryl] [(2 R,3 S,4 S)-2,3,4,5-tetrahyroxypentyl] hydrogen phosphate} that is more commonly referred to as CDP-L-ribitol or the CDP 5-ester of D-ribitol. Manya et al. maintained the carbohydrate nomenclature and thus referred to the carbon on ribitol as the 4-carbon that TMEM5 acts on. However, in CDP-L-ribitol, carbon-1 of ribitol is directly linked to the β-phosphate of the nucleotide phosphate (Figure 2.2 B). FKTN and FKRP thus transfer alditol-1-phosphate moieties, in nomenclature agreement with the teichoic acid synthesis enzymes in bacteria (Figure 2.1 C) (Korte et al. 1976). Therefore, based on the NMR

analyses by Manya et al. and following IUPAC nomenclature, we refer to TMEM5 as a ribitol β-*1,2* Xylosyltransferase (Figure 2.1 C). We propose renaming TMEM5 to **RXYLT1** consistent with its defined activity and the naming convention used for the CAZy (Carbohydrate-Active enZYmes Database) GT8 family.

Functional relevance of core M1 and M2 structures on α-DG and other proteins

While progress has been made in elucidating the link between the core M3 *O*-man glycan structure and matriglycan (Gerin et al. 2016; Kanagawa et al. 2016; Praissman et al. 2014, 2016; Willer et al. 2014), little is known about the functional relevance of core M1 and M2 structures. Core M1 structures are formed by the POMGNT1-dependent extension of the initial mannose residue with β1,2-linked GlcNAc (Figure 2.1 B) and can be further extended to form the classical tetrasaccharide (Neu5Ac-α2,3-Gal-β1,4-GlcNAc-β1,2-Man), a Lewis X epitope (Gal-β1,4-(Fuc-α1,3)-GlcNAc-β1,2-Man) or a Human Natural Killer-1 epitope (HNK-1; 3S-GlcA-β1,3-Gal-β1, 4-GlcNAc-β1,2-Man), among others (Praissman and Wells 2014) using GTs and modifying enzymes that are involved in multiple glycan pathways. Core M1 structures account for over 15% of brain *O*-glycans (Stalnaker et al. 2011) and are far more abundant in the mucin-like domain of α-DG than core M3 structures (Harrison et al. 2012; Nilsson et al. 2010; Stalnaker et al. 2010). Mutations in POMGNT1 leading to loss of core M1 glycans are causal for various forms of CMDs (Falsaperla et al. 2016), and core M1 structures are necessary for functional glycosylation of α-DG (Liu et al. 2006). One hypothesis is that these structures serve as a scaf-fold for the core M3 GTs (Kuwabara et al. 2016; Xiong et al. 2006). Evidence that overexpressed POMGNT1 co-precipitates with overexpressed FKTN and forms a complex (Xiong et al. 2006) suggests one potential model to be tested for the role of POMGNT1 in the generation of functional M3 glycan structures. Further support for this model comes from a POMGNT1 crystal structure

showing a lectin-like stem domain capable of binding the enzyme's reaction product (Figure 2.1 B) (Kuwabara et al. 2016). Thus, POMGNT1's binding to M1 sites may assist in the recruitment of core M3 GTs through protein–protein associations to facilitate elaboration of nearby core M3 sites.

Core M1 structures also serve as a precursor for core M2 structures (Figure 2.1 B). Multiple core M2 structures exist, including HNK-1 and Lewis X epitope-containing structures (Praissman and Wells 2014; Stalnaker et al. 2010). Core M2 structures account for ~5% of brain *O*-glycans (Stalnaker et al. 2011). Altering core M2 levels changes integrin-dependent cell adhesion and migration in vitro (Abbott et al. 2006, 2008), however, a mouse model lacking core M2 shows no neuronal development problems and does not impede functional glycosylation of α-DG (Kanekiyo et al. 2013; Lee et al. 2012). While clear biological roles of the core M1 and M2 glycan structures are not fully understood, their functional relevance may become more apparent as research into α-DG and other *O*-mannosylated proteins continues.

Expanding the *O*-mannosylated proteome beyond α-DG

Hypoglycosylation of α-DG that results from mutations in GTs explains the CMD phenotypes observed in skeletal muscle, but it does not fully explain the associated spectrum of neurological phenotypes seen in patients. While hypoglycosylation of α-DG and loss-of-function mutations in dystrophin (see Figure 2.1) both disrupt the DGC and lead to muscle disease (CMDs and Duchenne's muscular dystrophy, respectively), critical neurological complications are only observed in the severe forms of CMD (Falsaperla et al. 2016; Yiu and Kornberg 2015). In line with this observation, similar levels of *O*-man initiated structures were found to be present in the brain of *DAG1* knockout mice compared to WT mice, suggesting that there must be *O*-mannosylated proteins other than α-DG (Stalnaker et al. 2011). This begged the question: Other

than α-DG, are there additional proteins that are functionally modified by *O*-mannose-initiated glycans?

Recently, two groups (Lommel et al. 2013; Vester-Christensen et al. 2013) identified E-cadherin as an *O*-mannosylated protein in mammals (Baenziger 2013). Cadherins are a class of cell-surface membrane glycoproteins that have multiple extracellular cadherin (ECs) domains. Clausen et al. identified *O*-mannose sites in EC2-5 of classical Types 1 and 2 cadherins, EC2-3 of the clustered protocadherins, and plexins, as well as all known *O*-mannose sites on α-DG (Vester-Christensen et al. 2013). Additionally, Strahl et al. demonstrated that *O*-mannosylation is essential for E-cadherin mediated cell adhesion in mouse embryos (Lommel et al. 2013). In order to simplify the *O*-glycoproteome, Clausen's group used their "SimpleCell" breast cancer line that contains a genetic inactivation of POMGNT1. Lectin-weak affinity chromatography (LWAC) followed by identification using mass spectrometry was used to elucidate the *O*-mannose glycoproteome. While the data strongly suggest that cadherin/plexin-derived peptides are *O*-glycosylated, the lectin (Concanavalin A; ConA) used for LWAC enrichment of α-mannose also has known affinity for other hexose sugars, including α-glucose (Goldstein and Poretz 2012; Goldstein et al. 1973). Hexoses are indistinguishable in mass spectrometric analysis of glycopeptides, therefore, experiments to eliminate alternative possibilities, such as *O*-glucosyl modification of cadherins/plexins, should be performed. For example, it is imperative to determine if these presumed *O*-mannosyl modified proteins are sensitive to α-mannosidase treatment. Further efforts should be put forth to understand how those putative mannose residues on cadherins/plexins either remain unextended as core M0, or are extended into core M1, M2 or M3 structures in biologically pertinent cell lines and tissues, such as muscle and brain.

Two other groups (Dwyer et al. 2012; Yaji et al. 2015) found that receptor protein tyrosine phosphatase ζ (RPTPζ)/phosphacan is also *O*-mannosylated in mouse brain. More specifically, Dwyer et al. found that RPTPζ/phosphacan is *O*-mannosylated and hypoglycosylated in brains of POMGNT1-knockout mice, a model of Muscle-Eye-Brain disease (Dwyer et al. 2012). In this study, they also demonstrate that RPTPζ/phosphacan is not modified by LARGE1 in the mouse brain, suggesting that there are only core M1 and core M2 structures present on this protein. In fact, Morise et al. found that RPTPζ/phosphacan in mouse brain has the *O*-mannose-linked HNK-1 glycan epitope that is an elaboration of the core M1/M2 structure (Morise et al. 2014). In a similar vein, the Lewis X epitope was found to be mainly expressed on RPTPζ/phosphacan in the developing mouse brain (Yaji et al. 2015). Since the Lewis X epitope almost disappeared in POMGNT1-knockout mouse brains, it was suggested that the *O*-man glycan is responsible for presenting the Lewis X epitope as well. Taken together, the abnormal glycosylation of RPTPζ/phosphacan in POMGNT1-knockout mice brains may contribute to the spectrum of neurological phenotypes seen in mutant-POMGNT1, POMT1 and POMT2 CMDs (Dwyer et al. 2012). A summary of currently known and putative *O*-mannosylated proteins is presented in Supplemental Table S2.1 (Abbott et al. 2008; Bartels et al. 2016; Bleckmann et al. 2009; Pacharra et al. 2012, 2013; Vester-Christensen et al. 2013; Winterhalter et al. 2013).

Evolutionary perspectives of *O*-mannosylation and model organisms

O-mannosylation of α-DG is evolutionarily conserved, and has been extensively studied in mammals (Yoshida-Moriguchi and Campbell 2015). α-DG has also been studied in the powerful model organism *Drosophila melanogaster* (Nakamura et al. 2010). All subunits of the DGC are present in *Drosophila*, however, only one isoform from *DAG1* splicing has a glycosylated mucin-like domain similar to human α-DG. Unlike human α-DG, all *DAG1* isoforms in flies are a single

polypeptide and are not cleaved into α and β subunits (Sciandra et al. 2015). It is unclear if α-DG in flies plays a similar biological role to α-DG in humans. Interestingly, in *Drosophila* there are no identified elaborated *O*-man structures (Aoki et al. 2008) nor obvious homologs for most of the enzymes needed to make and elaborate the core M3 glycan [(Grewal et al. 2005), Figure 2.3]. However, there are homologs for *POMT1/2*, *rotated* (*rt*) and *twisted* (*tw*) (Ichimiya et al. 2004; Nakamura et al. 2010), indicating there are *O*-mannosylated proteins present in this organism that is relatively distant from mammals on the evolutionary tree. Phylogenetic analysis of human *LARGE1*, in addition to other genes involved with functional glycosylation of α-DG, indicates orthologues are present in most metazoans, especially vertebrates. Divergence begins in insects followed by complete absence of most of the genes in fungi and lower organisms. This suggests higher organisms have evolved a uniquely multifaceted functional glycosylation, presumably to accommodate increasing complexity in tissue structure and function (Figure 2.3) (Sadreyev et al. 2015). However, due to the evolutionary conservation of *O*-mannosyl modification of proteins, reduced heterogeneity of structures outside of the mammalian clade, and the lower costs associated with research, using model organisms, like *Drosophila*, presents unique advantages for studying the first step of protein *O*-mannosylation.

The zebrafish (*Danio rerio*) has also emerged as a powerful vertebrate model organism to study dystroglycanopathies and other neuromuscular disorders (Pappalardo et al. 2013; Steffen et al. 2007). Genetic manipulation of zebrafish embryos is fast, effective and inexpensive, and all of the human dystroglycanopathy-related genes have been identified in zebrafish (Moore et al. 2008; Wood and Currie 2014). While homozygous dystroglycan mutations in mice are embryonic lethal (Williamson et al. 1997), morpholino-mediated knockdown of dystroglycan results in viable zebrafish, however, with disorganized muscle and disruption of the DGC (Parsons et al. 2002).

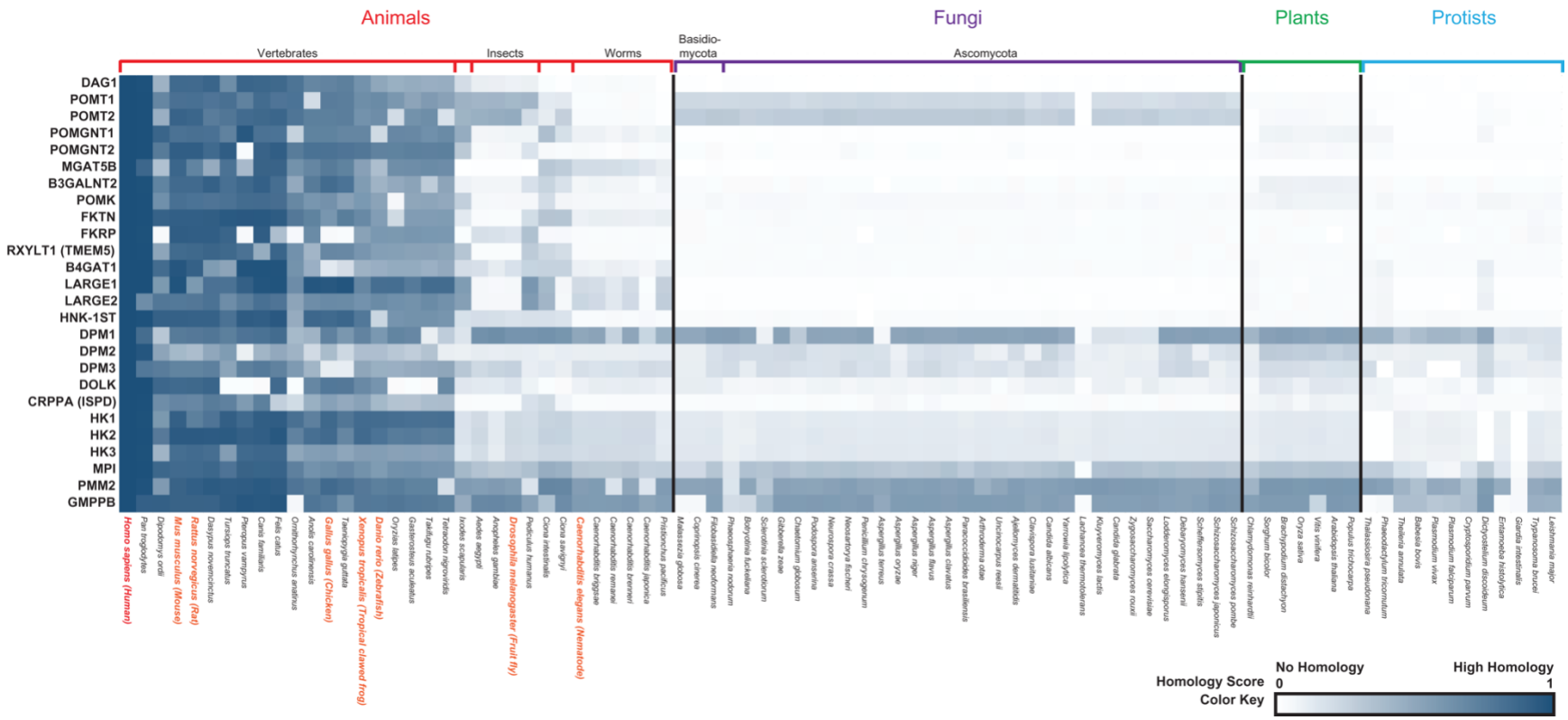


Figure 2.3. Phylogenetic profile of genes involved with functional glycosylation of α-DG.
A phylogenetic profile heatmap of human genes (left) involved with functional glycosylation of α-DG (from Table 2.1) was generated using *PhyloGene* (Sadreyev et al. 2015). Species analyzed are indicated at the bottom, and the heatmap is categorized according to indicated taxa (top and vertical black lines) and subtaxa (additional colored vertical line separators). Model organisms that have been described in this review or used in dystroglycan studies (not mentioned in this review) are labeled in orange, and *Homo sapiens* is labeled in red. Protein sequence similarity values range from 0 (white, no homology) to 1 (blue, high homology relative to *H. sapiens*).

Recent zebrafish studies knocking down genes implicated in secondary dystroglycanopathies, such as those described in Table 2.1, have recapitulated the muscle, eye and brain phenotypes typically observed in CMD patients, such as hydrocephaly, reduced eye size, impaired muscle development and reduced α-DG glycosylation (Avsar-Ban et al. 2010; Buysse et al. 2013; Di Costanzo et al. 2014; Manzini et al. 2012; Praissman et al. 2016; Stevens et al. 2013). Interestingly, while FKTN and FKRP knockdown in zebrafish results in hypoglycosylated α-DG and reduced laminin binding (Kawahara et al. 2010; Thornhill et al. 2008), other studies suggest that FKTN and FKRP may also play roles in protein secretion, the unfolded protein response, and angiogenesis (Lin et al. 2011; Wood et al. 2011). For tertiary dystroglycanopathies (Table 2.1 and Figure 2.2), morpholino-mediated knockdown of the zebrafish DPMS complex and ISPD resulted in anticipated hypoglycosylation of α-DG and a dystrophic muscle phenotype (Marchese et al. 2016; Roscioli et al. 2012).

Other aspects of *O*-mannosylation

The loss of α-DG expression and hypoglycosylation has also been documented in many types of epithelial and neuronal cancers, with aberrant glycosylation of α-DG being implicated in cancer progression and metastasis (de Bernabe et al. 2009; Martin et al. 2007; Muschler et al. 2002; Sgambato et al. 2003; Shen et al. 2012). It has been shown that maturation of Core M3-type glycans on α-DG is critical for binding of lymphocytic choriomeningitis virus (LCMV) and LASV, in addition to other arenaviruses, Mobala and Oliveros (Kunz, Rojek, Kanagawa et al. 2005). Furthermore, LCMV and LASV compete with laminin for binding α-DG glycans (Kunz, Rojek, Perez et al. 2005; Kunz et al. 2001). Ubiquitous expression of α-DG in various tissues and the competitive binding nature of LCMV and LASV with LG-domain-containing proteins indicate

these viruses have evolved to infect a broad range of cell types and the potential disruption of cell-ECM homeostasis likely contributes to pathogenesis.

Further insights into the mechanism of ECM receptor binding to the functional glycan on α-DG have come out of the Hohenester and Campbell laboratories with a high-resolution (1.4 Å) crystal structure of matriglycan bound to the LG4 and LG5 domains of laminin-α2 (Briggs et al. 2016). Structural analysis of this complex show the coordination of a single Ca^{2+} ion by a single [-GlcA-β3-Xyl-α3-] repeat, providing a snapshot of this high affinity Ca^{2+}-dependent protein-carbohydrate interaction. Precise regulation of the control of matriglycan chain length is currently unknown, however expression levels of human natural killer-1 sulfotransferase (HNK-1ST) and LARGE1 have been implicated in the regulatory mechanism (Nakagawa et al. 2012, 2013). Overexpression of LARGE1 results in increased levels of α-DG glycosylation (Barresi et al. 2004; Patnaik and Stanley 2005), whereas sulfation of the functional glycan, likely on a non-reducing end GlcA residue, by HNK-1ST reduces levels of LARGE-mediated α-DG glycosylation (Nakagawa et al. 2012, 2013) (Figure 2.1 C). While the sulfation has been determined to be within the functional moiety on core M3 glycans (Nakagawa et al. 2013), the exact site of HNK-1ST-mediated sulfation has yet to be formally demonstrated. Although the precise range of matriglycan GlcA-Xyl repeats is unknown, the large shifts in electrophoretic mobility of α-DG in SDS-PAGE (>60 kDa which is approximately >200 repeats) suggests that the LARGE1-dependent modification might serve as a long scaf-fold potentially allowing multiple LG-domain-containing ECM proteins to bind and tether to the cell-surface. However, LG-domain-containing ligand proteins have been shown to competitively bind to α-DG (Kanagawa et al. 2004). Thus, it remains to be determined whether a single matriglycan chain can bind multiple LG-domain-containing proteins concurrently. Complete structural determination of the fully elaborated core M3 glycan

built in vitro by chemoenzymatic synthesis using recombinant enzymes is necessary to define the comprehensive three dimensional structure and will allow for direct testing of the impact of repeat length and binding to other proteins of interest, in addition to being used as a tool to identify other proteins that may participate in the DGC.

Using Chinese Hamster ovary cell mutants that lacks both *O*-mannose and complex *N*-glycans, work from the Stanley laboratory suggests that when overexpressed in these systems, LARGE1 can modify substrates other than *O*-mannose, such as *N*-glycans and mucin-type *O*-glycans on α-DG (Aguilan et al. 2009; Patnaik and Stanley 2005). Additionally, LARGE1 overexpression in neural stems cells deficient in POMT2 or both POMT2 and α-DG resulted in the reporting of IIH6-reactive proteins and laminin-binding epitopes that were PNGaseF-sensitive (which removes *N*-glycans), suggesting that LARGE1 could potentially modify *N*-glycans on proteins other than α-DG (Zhang and Hu 2012; Zhang et al. 2011). However, it was recently shown that LARGE2, and not LARGE1, is capable of modifying proteoglycans, such as Glypican-4, presumably by extending the non-reducing end GlcA on a common core GAG tetrasaccharide linker, GlcAβ1-3-Galβ1-3-Galβ1-4Xylβ-, with matriglycan when overexpressed in dystroglycan-deficient mouse embryonic stem cells (Inamori et al. 2016). While LARGE1 and LARGE2 catalyze the same bifunctional GT reaction (Inamori et al. 2012, 2013), LARGE1 is highly expressed in skeletal muscle, heart and brain, whereas LARGE2 is highly expressed in kidneys and testis (Grewal et al. 2005; Peyrard et al. 1999). This differential tissue expression, different pH optima for GT activities (Inamori et al. 2013), and different substrate specificities (Inamori et al. 2016) suggests that LARGE1 and LARGE2 may not have overlapping functions. Thus, in a tissue-specific manner, LARGE1- and/or LARGE2-mediated hyperglycosylation of non-native structures (other than *O*-mannose, such as proteoglycans) may serve as a compensatory

mechanism, and potential therapeutic approach, for dystroglycanopathies where there are no *O*-mannose structures present.

Recent progress has been made in understanding the structural aspects of enzymes in the *O*-mannosylation pathway. Specifically, elucidation of the POMGNT1 structure revealed a carbohydrate-binding stem domain and further support for the promiscuity of this enzyme towards various glycopeptides substrates (Kuwabara et al. 2016). Additionally, POMK is an unusual pseudokinase that lacks certain primary sequence elements thought to be required for kinase activity (Yoshida-Moriguchi and Campbell 2015), however, a crystal structure of zebrafish POMK reveals the mechanism in which the enzyme recognizes the GalNAc-β3-GlcNAc-β4-Man trisaccharide acceptor substrate and catalyzes the phosphate transfer from ATP (Zhu et al. 2016).

Given that the POMGNTs dictate the core structure to be synthesized, an understanding of substrate specificity for these enzymes needs to be elucidated. Our group recently investigated the substrate specificities of POMGNT1 and POMGNT2 and found POMGNT1 to be promiscuous while an identified amino acid motif, R-X-R-X-X-I-X-X-T(*O*-Man)-P-T, that appears to only be present and conserved in vertebrate α-DGs is the preferred substrate for POMGNT2 (Halmo et al. 2017). Thus, POMGNT2 appears to be the "gatekeeper" enzyme for the generation of functional M3 glycans. Elucidation of POMGNT2 structure, to complement POMGNT1's, will greatly facilitate our understanding of what sites of *O*-man can be elaborated into full-length M3 glycans capable of binding LG-domain-containing ECM proteins.

Concluding remarks

Although major accomplishments have been made in understanding the uniquely complex functional modification of α-DG, there are still many unknowns in the field. These questions include: 1. What are the mechanisms behind the unique specificities of the pathway enzymes?, 2.

What are the origins and significance of the GroP modification?, 3. What is the functional relevance of core M1 and M2 glycans?, 4. What is the role of the furin-cleaved N-terminus of α-DG? and 5. Are newly identified proteins that contain *O*-mannose further extended and functional?. Many of these studies could take advantage of the aforementioned model organisms established for the study of CMDs. Now that the vast majority of enzyme functions involved with core M3 elaboration have been identified, structural analyses can be undertaken to establish catalytic mechanisms, identify how CMD patient mutations affect enzymatic functions, and propose testable mechanisms to ameliorate these defects. It should also be noted that there are still CMD and cobblestone lissencephaly patients with unknown genetic etiology or known etiology where the functional link to *O*-mannosylation has not been established such as mutations in TMTC3 (Jerber et al. 2016). In conclusion, while it would seem inconceivable, from an evolutionary point of view, that such an elaborate modification that requires multiple gene products and at least one novel donor substrate is targeted to only a single cell-surface protein, α-DG, on two sites (T317 and T379), no other proteins have yet to be identified with M3 glycans (Hara, Kanagawa et al. 2011; Yagi et al. 2013). With the aid of the research findings described here in the last few years and the advancements in the tools used to study these biological molecules and processes (Porterfield et al. 2014; Steentoft et al. 2011; Sun et al. 2016; Zhao et al. 2013), continued research will help us understand the evolutionary perspective of *O*-mannosylation and achieve the ultimate goal of discovering novel therapeutics relevant to the associated human diseases.

Acknowledgements

We thank all members of the Wells Laboratory for helpful discussions. This work was supported by the National Institutes of Health [R01GM111939 to L.W., P41GM103490 to L.W. (Senior Investigator), P01GM107012 to L.W. (Senior Investigator), T32GM107004 to S.M.H.].

References

Abbott KL, Matthews RT, Pierce M. 2008. Receptor tyrosine phosphatase beta (RPTPbeta) activity and signaling are attenuated by glycosylation and subsequent cell surface galectin-1 binding. *J Biol Chem*, 283:33026-33035.

Abbott KL, Troupe K, Lee I, Pierce M. 2006. Integrin-dependent neuroblastoma cell adhesion and migration on laminin is regulated by expression levels of two enzymes in the O-mannosyl-linked glycosylation pathway, PomGnT1 and GnT-Vb. *Experimental cell research*, 312:2837-2850.

Aebi M. 2013. N-linked protein glycosylation in the ER. *Biochim Biophys Acta*, 1833:2430-2437.

Aguilan JT, Sundaram S, Nieves E, Stanley P. 2009. Mutational and functional analysis of Large in a novel CHO glycosylation mutant. *Glycobiology*, 19:971-986.

Aoki K, Porterfield M, Lee SS, Dong B, Nguyen K, McGlamry KH, Tiemeyer M. 2008. The diversity of O-linked glycans expressed during Drosophila melanogaster development reflects stage- and tissue-specific requirements for cell signaling. *J Biol Chem*, 283:30385-30400.

Aravind L, Koonin EV. 1999. The fukutin protein family - predicted enzymes modifying cell-surface molecules. *Current Biology*, 9:R836-R837.

Ashikov A, Buettner FF, Tiemann B, Gerardy-Schahn R, Bakker H. 2013. LARGE2 generates the same xylose- and glucuronic acid-containing glycan structures as LARGE. *Glycobiology*. 23:303–309.

Avsar-Ban E, Ishikawa H, Manya H, Watanabe M, Akiyama S, Miyake H, Endo T, Tamaru Y. 2010. Protein O-mannosylation is necessary for normal embryonic development in zebrafish. *Glycobiology*, 20:1089-1102.

Baenziger JU. 2013. O-mannosylation of cadherins. *Proc Natl Acad Sci U S A*, 110:20858-20859.

Barone R, Aiello C, Race V, Morava E, Foulquier F, Riemersma M, Passarelli C, Concolino D, Carella M, Santorelli F, *et al.* 2012. DPM2-CDG: A muscular dystrophy-dystroglycanopathy syndrome with severe epilepsy. *Ann Neurol*, 72:550-558.

Barresi R, Campbell KP. 2006. Dystroglycan: from biosynthesis to pathogenesis of human disease. *Journal of cell science*, 119:199-207.

Barresi R, Michele DE, Kanagawa M, Harper HA, Dovico SA, Satz JS, Moore SA, Zhang WL, Schachter H, Dumanski JP, *et al.* 2004. LARGE can functionally bypass alpha-dystroglycan glycosylation defects in distinct congenital muscular dystrophies. *Nat Med*, 10:696-703.

Bartels MF, Winterhalter PR, Yu J, Liu Y, Lommel M, Mohrlen F, Hu H, Feizi T, Westerlind U, Ruppert T, *et al.* 2016. Protein O-Mannosylation in the Murine Brain: Occurrence of Mono-O-Mannosyl Glycans and Identification of New Substrates. *PLoS One*, 11:e0166119.

Baur S, Marles-Wright J, Buckenmaier S, Lewis RJ, Vollmer W. 2009. Synthesis of CDP-activated ribitol for teichoic acid precursors in Streptococcus pneumoniae. *J Bacteriol*, 191:1200-1210.

Beedle AM, Turner AJ, Saito Y, Lueck JD, Foltz SJ, Fortunato MJ, Nienaber PM, Campbell KP. 2012. Mouse fukutin deletion impairs dystroglycan processing and recapitulates muscular dystrophy. *J Clin Invest*, 122:3330-3342.

Belaya K, Rodriguez Cruz PM, Liu WW, Maxwell S, McGowan S, Farrugia ME, Petty R, Walls TJ, Sedghi M, Basiri K *et al*. 2015. Mutations in GMPPB cause congenital myasthenic syndrome and bridge myasthenic disorders with dystroglycanopathies. *Brain*. 138:2493–2504.

Bleckmann C, Geyer H, Lieberoth A, Splittstoesser F, Liu Y, Feizi T, Schachner M, Kleene R, Reinhold V, Geyer R. 2009. O-glycosylation pattern of CD24 from mouse brain. *Biol Chem*, 390:627-645.

Briggs DC, Yoshida-Moriguchi T, Zheng T, Venzke D, Anderson ME, Strazzulli A, Moracci M, Yu L, Hohenester E, Campbell KP. 2016. Structural basis of laminin binding to the LARGE glycans on dystroglycan. *Nat Chem Biol*.

Brockington M, Blake DJ, Prandini P, Brown SC, Torelli S, Benson MA, Ponting CP, Estournet B, Romero NB, Mercuri E, *et al.* 2001. Mutations in the fukutin-related protein gene (FKRP) cause

a form of congenital muscular dystrophy with secondary laminin alpha 2 deficiency and abnormal glycosylation of alpha-dystroglycan. *Am J Hum Genet*, 69:1198-1209.

Brockington M, Torelli S, Sharp PS, Liu K, Cirak S, Brown SC, Wells DJ, Muntoni F. 2010. Transgenic overexpression of LARGE induces alpha-dystroglycan hyperglycosylation in skeletal and cardiac muscle. *PLoS One*, 5:e14434.

Buysse K, Riemersma M, Powell G, van Reeuwijk J, Chitayat D, Roscioli T, Kamsteeg EJ, van den Elzen C, van Beusekom E, Blaser S, *et al.* 2013. Missense mutations in beta-1,3-N-acetylglucosaminyltransferase 1 (B3GNT1) cause Walker-Warburg syndrome. *Hum Mol Genet*, 22:1746-1754.

Cao W, Henry MD, Borrow P, Yamada H, Elder JH, Ravkov EV, Nichol ST, Compans RW, Campbell KP, Oldstone MB. 1998. Identification of α-Dystroglycan as a Receptor for Lymphocytic Choriomeningitis Virus and Lassa Fever Virus. *Science*, 282:2079-2081.

Carette JE, Raaben M, Wong AC, Herbert AS, Obernosterer G, Mulherkar N, Kuehne AI, Kranzusch PJ, Griffin AM, Ruthel G, *et al.* 2011. Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature*, 477:340-343.

Chiba A, Matsumura K, Yamada H, Inazu T, Shimizu T, Kusunoki S, Kanazawa I, Kobata A, Endo T. 1997. Structures of sialylated O-linked oligosaccharides of bovine peripheral nerve alpha-dystroglycan The role of a novel O-mannosyl-type oligosaccharide in the binding of alpha-dystroglycan with laminin. *Journal of Biological Chemistry*, 272:2156-2162.

de Bernabe DB, Inamori K, Yoshida-Moriguchi T, Weydert CJ, Harper HA, Willer T, Henry MD, Campbell KP. 2009. Loss of alpha-dystroglycan laminin binding in epithelium-derived cancers is caused by silencing of LARGE. *J Biol Chem*, 284:11279-11284.

Di Costanzo S, Balasubramanian A, Pond HL, Rozkalne A, Pantaleoni C, Saredi S, Gupta VA, Sunu CM, Yu TW, Kang PB, *et al.* 2014. POMK mutations disrupt muscle development leading to a spectrum of neuromuscular presentations. *Human Molecular Genetics*, 23:5781-5792.

Dobson CM, Hempel SJ, Stalnaker SH, Stuart R, Wells L. 2013. O-Mannosylation and human disease. *Cellular and molecular life sciences : CMLS*, 70:2849-2857.

Doucey MA, Hess D, Cacan R, Hofsteenge J. 1998. Protein C-mannosylation is enzyme-catalysed and uses dolichyl-phosphate-mannose as a precursor. *Mol Biol Cell*, 9:291-300.

Dwyer CA, Baker E, Hu H, Matthews RT. 2012. RPTPzeta/phosphacan is abnormally glycosylated in a model of muscle-eye-brain disease lacking functional POMGnT1. *Neuroscience*, 220:47-61.

Ervasti JM, Campbell KP. 1991. Membrane organization of the dystrophin-glycoprotein complex. *Cell*, 66:1121-1131.

Ervasti JM, Campbell KP. 1993. A role for the dystrophin-glycoprotein complex as a transmembrane linker between laminin and actin. *J Cell Biol*, 122:809-823.

Esapa CT, Benson MA, Schroder JE, Martin-Rendon E, Brockington M, Brown SC, Muntoni F, Kroger S, Blake DJ. 2002. Functional requirements for fukutin-related protein in the Golgi apparatus. *Human Molecular Genetics*, 11:3319-3331.

Falsaperla R, Pratico AD, Ruggieri M, Parano E, Rizzo R, Corsello G, Vitaliti G, Pavone P. 2016. Congenital muscular dystrophy: from muscle to brain. *Ital J Pediatr*, 42:78.

Finne J, Krusius T, Margolis RK, Margolis RU. 1979. Novel mannitol-containing oligosaccharides obtained by mild alkaline borohydride treatment of a chondroitin sulfate proteoglycan from brain. *J Biol Chem*, 254:10295-10300.

Geis T, Marquard K, Rodl T, Reihle C, Schirmer S, von Kalle T, Bornemann A, Hehr U, Blankenburg M. 2013. Homozygous dystroglycan mutation associated with a novel muscle-eye-brain disease-like phenotype with multicystic leucodystrophy. *Neurogenetics*, 14:205-213.

Gerin I, Ury B, Breloy I, Bouchet-Seraphin C, Bolsee J, Halbout M, Graff J, Vertommen D, Muccioli GG, Seta N, *et al.* 2016. ISPD produces CDP-ribitol used by FKTN and FKRP to transfer ribitol phosphate onto alpha-dystroglycan. *Nature communications*, 7.

Goldstein IJ, Poretz RD. 2012. Isolation, physicochemical characterization, and carbohydrate-binding specificity of lectins. In: Liener IE, Sharon N, Goldstein IJ editors. *The Lectins Properties, Functions and Applications in Biology and Medicine*. p. 33–247.

Goldstein IJ, Reichert CM, Misaki A, Gorin PA. 1973. An "extension" of the carbohydrate binding specificity of concanavalin A. *Biochim Biophys Acta*, 317:500-504.

Grewal PK, McLaughlan JM, Moore CJ, Browning CA, Hewitt JE. 2005. Characterization of the LARGE family of putative glycosyltransferases associated with dystroglycanopathies. *Glycobiology*, 15:912-923.

Halmo SM, Singh D, Patel S, Wang S, Edlin M, Boons GJ, Moremen KW, Live D, Wells L. 2016. Protein O-linked mannose beta-1,4-N-acetylglucosaminyltransferase 2 (POMGNT2) is a gatekeeper enzyme for functional glycosylation of alpha-dystroglycan. *J Biol Chem*.

Hara Y, Balci-Hayta B, Yoshida-Moriguchi T, Kanagawa M, Beltran-Valero de Bernabe D, Gundesli H, Willer T, Satz JS, Crawford RW, Burden SJ, *et al.* 2011a. A dystroglycan mutation associated with limb-girdle muscular dystrophy. *N Engl J Med*, 364:939-946.

Hara Y, Kanagawa M, Kunz S, Yoshida-Moriguchi T, Satz JS, Kobayashi YM, Zhu Z, Burden SJ, Oldstone MB, Campbell KP. 2011b. Like-acetylglucosaminyltransferase (LARGE)-dependent modification of dystroglycan at Thr-317/319 is required for laminin binding and arenavirus infection. *Proc Natl Acad Sci U S A*, 108:17426-17431.

Harrison R, Hitchen PG, Panico M, Morris HR, Mekhaiel D, Pleass RJ, Dell A, Hewitt JE, Haslam SM. 2012. Glycoproteomic characterization of recombinant mouse alpha-dystroglycan. *Glycobiology*, 22:662-675.

Holt KH, Crosbie RH, Venzke DP, Campbell KP. 2000. Biosynthesis of dystroglycan: processing of a precursor propeptide. *FEBS Lett*, 468:79-83.

Ichimiya T, Manya H, Ohmae Y, Yoshida H, Takahashi K, Ueda R, Endo T, Nishihara S. 2004. The twisted abdomen phenotype of Drosophila POMT1 and POMT2 mutants coincides with their heterophilic protein O-mannosyltransferase activity. *J Biol Chem*. 279:42638–42647.

Ilg T, Stierhof YD, Craik D, Simpson R, Handman E, Bacic A. 1996. Purification and structural characterization of a filamentous, mucin-like proteophosphoglycan secreted by Leishmania parasites. *Journal of Biological Chemistry*, 271:21583-21596.

Inamori K, Hara Y, Willer T, Anderson ME, Zhu Z, Yoshida-Moriguchi T, Campbell KP. 2013. Xylosyl- and glucuronyltransferase functions of LARGE in alpha-dystroglycan modification are conserved in LARGE2. *Glycobiology*, 23:295-302.

Inamori K, Yoshida-Moriguchi T, Hara Y, Anderson ME, Yu L, Campbell KP. 2012. Dystroglycan function requires xylosyl- and glucuronyltransferase activities of LARGE. *Science*, 335:93-96.

Inamori KI, Beedle AM, de Bernabe DB, Wright ME, Campbell KP. 2016. LARGE2-dependent glycosylation confers laminin-binding ability on proteoglycans. *Glycobiology*, 26:1284-1296.

Jae LT, Raaben M, Riemersma M, van Beusekom E, Blomen VA, Velds A, Kerkhoven RM, Carette JE, Topaloglu H, Meinecke P, *et al.* 2013. Deciphering the Glycosylome of Dystroglycanopathies Using Haploid Screens for Lassa Virus Entry. *Science*, 340:479-483.

Jerber J, Zaki MS, Al-Aama JY, Rosti RO, Ben-Omran T, Dikoglu E, Silhavy JL, Caglar C, Musaev D, Albrecht B *et al*. 2016. Biallelic Mutations in TMTC3, Encoding a Transmembrane and TPR-Containing Protein, Lead to Cobblestone Lissencephaly. *Am J Hum Genet*. 99:1181–1189.

Kanagawa M, Kobayashi K, Tajiri M, Manya H, Kuga A, Yamaguchi Y, Akasaka-Manya K, Furukawa J, Mizuno M, Kawakami H, *et al.* 2016. Identification of a Post-translational Modification with Ribitol-Phosphate and Its Defect in Muscular Dystrophy. *Cell Rep*, 14:2209-2223.

Kanagawa M, Saito F, Kunz S, Yoshida-Moriguchi T, Barresi R, Kobayashi YM, Muschler J, Dumanski JP, Michele DE, Oldstone MB *et al*. 2004. Molecular recognition by LARGE is essential for expression of functional dystroglycan. *Cell*. 117:953–964.

Kanekiyo K , Inamori K, Kitazume S, Sato K, Maeda J, Higuchi M, Kizuka Y, Korekane H, Matsuo I, Honke K *et al*. 2013. Loss of branched O-mannosyl glycans in astrocytes accelerates remyelination. *J Neurosci*. 33:10037–10047.

Kang JY, Hong Y, Ashida H, Shishioh N, Murakami Y, Morita YS, Maeda Y, Kinoshita T. 2005. PIG-V involved in transferring the second mannose in glycosylphosphatidylinositol. *J Biol Chem*, 280:9489-9497.

Kawahara G, Guyon JR, Nakamura Y, Kunkel LM. 2010. Zebrafish models for human FKRP muscular dystrophies. *Hum Mol Genet*, 19:623-633.

Kobayashi K, Nakahori Y, Miyake M, Matsumura K, Kondo-Iida E, Nomura Y, Segawa M, Yoshioka M, Saito K, Osawa K, *et al.* 1998. An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. *Nature*, 394:388-392.

Korte F , Goto M, Anno K. 1976. *Nucleic acids, proteins, and carbohydrates. 3.4:2.2 Carbohydrates; Cell Wall Substances: Teichoic Acids.* New York: Academic Press. 231, p. viii.

Kunz S, Rojek JM, Kanagawa M, Spiropoulou CF, Barresi R, Campbell KP, Oldstone MB. 2005a. Posttranslational modification of α-dystroglycan, the cellular receptor for arenaviruses, by the glycosyltransferase LARGE is critical for virus binding. *J Virol*, 79:14282-14296.

Kunz S, Rojek JM, Perez M, Spiropoulou CF, Oldstone MB. 2005b. Characterization of the interaction of lassa fever virus with its cellular receptor α-dystroglycan. *J Virol*, 79:5979-5987.

Kunz S, Sevilla N, McGavern DB, Campbell KP, Oldstone MB. 2001. Molecular analysis of the interaction of LCMV with its cellular receptor α-dystroglycan. *J Cell Biol*, 155:301-310.

Kuwabara N, Manya H, Yamada T, Tateno H, Kanagawa M, Kobayashi K, Akasaka-Manya K, Hirose Y, Mizuno M, Ikeguchi M, *et al.* 2016. Carbohydrate-binding domain of the POMGnT1 stem region modulates O-mannosylation sites of alpha-dystroglycan. *Proc Natl Acad Sci U S A*, 113:9280-9285.

Lee JK, Matthews RT, Lim JM, Swanier K, Wells L, Pierce JM. 2012. Developmental expression of the neuron-specific N-acetylglucosaminyltransferase Vb (GnT-Vb/IX) and identification of its in vivo glycan products in comparison with those of its paralog, GnT-V. *J Biol Chem*, 287:28526-28536.

Lefeber DJ, de Brouwer APM, Morava E, Riemersma M, Schuurs-Hoeijmakers JHM, Absmanner B, Verrijp K, van den Akker WMR, Huijben K, Steenbergen G, *et al.* 2011. Autosomal Recessive Dilated Cardiomyopathy due to DOLK Mutations Results from Abnormal Dystroglycan O-Mannosylation. *Plos Genet*, 7.

Lefeber DJ, Schonberger J, Morava E, Guillard M, Huyben KM, Verriip K, Grafakou O, Evangelioi A, Preijers FW, Manta P, *et al.* 2009. Deficiency of Dol-P-Man Synthase Subunit

DPM3 Bridges the Congenital Disorders of Glycosylation with the Dystroglycanopathies. *Am J Hum Genet*, 85:76-86.

Lin YY, White RJ, Torelli S, Cirak S, Muntoni F, Stemple DL. 2011. Zebrafish Fukutin family proteins link the unfolded protein response with dystroglycanopathies. *Hum Mol Genet*, 20:1763-1775.

Liu J, Ball SL, Yang Y, Mei P, Zhang L, Shi H, Kaminski HJ, Lemmon VP, Hu H. 2006. A genetic model for muscle-eye-brain disease in mice lacking protein O-mannose 1,2-N-acetylglucosaminyltransferase (POMGnT1). *Mech Dev*, 123:228-240.

Lommel M, Winterhalter PR, Willer T, Dahlhoff M, Schneider MR, Bartels MF, Renner-Muller I, Ruppert T, Wolf E, Strahl S. 2013. Protein O-mannosylation is crucial for E-cadherin-mediated cell adhesion. *Proc Natl Acad Sci U S A*, 110:21024-21029.

Luo S , Cai S, Maxwell S, Yue D, Zhu W, Qiao K, Zhu Z, Zhou L, Xi J, Lu J *et al*. 2017. Novel mutations in the C-terminal region of GMPPB causing limb-girdle muscular dystrophy overlapping with congenital myasthenic syndrome. *Neuromuscul Disord*. 27:557–564.

Lynch TA, Lam LT, Man NT, Kobayashi K, Toda T, Morris GE. 2012. Detection of the dystroglycanopathy protein, fukutin, using a new panel of site-specific monoclonal antibodies. *Biochem Bioph Res Co*, 424:354-357.

Maeda Y, Tanaka S, Hino J, Kangawa K, Kinoshita T. 2000. Human dolichol-phosphate-mannose synthase consists of three subunits, DPM1, DPM2 and DPM3. *Embo J*, 19:2475-2482.

Majumdar A, Shah MH, Bitok JK, Hassis-LeBeau ME, Meyers CLF. 2009. Probing phosphorylation by non-mammalian isoprenoid biosynthetic enzymes using H-1-P-31-P-31 correlation NMR spectroscopy. *Mol Biosyst*, 5:935-944.

Manya H , Chiba A, Yoshida A, Wang X, Chiba Y, Jigami Y, Margolis RU, Endo T. 2004. Demonstration of mammalian protein O-mannosyltransferase activity: coexpression of POMT1 and POMT2 required for enzymatic activity. *Proc Natl Acad Sci USA*. 101:500–505.

Manya H, Yamaguchi Y, Kanagawa M, Kobayashi K, Tajiri M, Akasaka-Manya K, Kawakami H, Mizuno M, Wada Y, Toda T, *et al.* 2016. The Muscular Dystrophy Gene TMEM5 Encodes a

Ribitol beta1,4-Xylosyltransferase Required for the Functional Glycosylation of Dystroglycan. *J Biol Chem*, 291:24618-24627.

Manzini MC, Tambunan DE, Hill RS, Yu TW, Maynard TM, Heinzen EL, Shianna KV, Stevens CR, Partlow JN, Barry BJ, *et al.* 2012. Exome Sequencing and Functional Validation in Zebrafish Identify GTDC2 Mutations as a Cause of Walker-Warburg Syndrome. *Am J Hum Genet*, 91:541-547.

Marchese M, Pappalardo A, Baldacci J, Verri T, Doccini S, Cassandrini D, Bruno C, Fiorillo C, Garcia-Gil M, Bertini E, *et al.* 2016. Dolichol-phosphate mannose synthase depletion in zebrafish leads to dystrophic muscle with hypoglycosylated alpha-dystroglycan. *Biochem Biophys Res Commun*, 477:137-143.

Martin LT, Glass M, Dosunmu E, Martin PT. 2007. Altered expression of natively glycosylated alpha dystroglycan in pediatric solid tumors. *Hum Pathol*, 38:1657-1668.

Michele DE, Barresi R, Kanagawa M, Saito F, Cohn RD, Satz JS, Dollar J, Nishino I, Kelley RI, Somer H, *et al.* 2002. Post-translational disruption of dystroglycan-ligand interactions in congenital muscular dystrophies. *Nature*, 418:417-422.

Moore CJ, Goh HT, Hewitt JE. 2008. Genes required for functional glycosylation of dystroglycan are conserved in zebrafish. *Genomics*, 92:159-167.

Moore SA, Saito F, Chen J, Michele DE, Henry MD, Messing A, Cohn RD, Ross-Barta SE, Westra S, Williamson RA, *et al.* 2002. Deletion of brain dystroglycan recapitulates aspects of congenital muscular dystrophy. *Nature*, 418:422-425.

Morise J, Kizuka Y, Yabuno K, Tonoyama Y, Hashii N, Kawasaki N, Manya H, Miyagoe-Suzuki Y, Takeda S, Endo T, *et al.* 2014. Structural and biochemical characterization of O-mannose-linked human natural killer-1 glycan expressed on phosphacan in developing mouse brains. *Glycobiology*, 24:314-324.

Muntoni F, Brockington M, Blake DJ, Torelli S, Brown SC. 2002. Defective glycosylation in muscular dystrophy. *Lancet*, 360:1419-1421.

Muschler J, Levy D, Boudreau R, Henry M, Campbell K, Bissell MJ. 2002. A role for dystroglycan in epithelial polarization: loss of function in breast tumor cells. *Cancer Res*, 62:7102-7109.

Nakagawa N, Manya H, Toda T, Endo T, Oka S. 2012. Human natural killer-1 sulfotransferase (HNK-1ST)-induced sulfate transfer regulates laminin-binding glycans on alpha-dystroglycan. *J Biol Chem*, 287:30823-30832.

Nakagawa N, Takematsu H, Oka S. 2013. HNK-1 sulfotransferase-dependent sulfation regulating laminin-binding glycans occurs in the post-phosphoryl moiety on alpha-dystroglycan. *Glycobiology*, 23:1066-1074.

Nakamura N, Stalnaker SH, Lyalin D, Lavrova O, Wells L, Panin VM. 2010. Drosophila Dystroglycan is a target of O-mannosyltransferase activity of two protein O-mannosyltransferases, Rotated Abdomen and Twisted. *Glycobiology*, 20:381-394.

Nilsson J, Nilsson J, Larson G, Grahn A. 2010. Characterization of site-specific O-glycan structures within the mucin-like domain of α-dystroglycan from human skeletal muscle. *Glycobiology*, 20:1160-1169.

Oldstone MB, Campbell KP. 2011. Decoding arenavirus pathogenesis: essential roles for alpha-dystroglycan-virus interactions and the immune response. *Virology*, 411:170-179.

Orlean P. 1990. Dolichol phosphate mannose synthase is required in vivo for glycosyl phosphatidylinositol membrane anchoring, O mannosylation, and N glycosylation of protein in Saccharomyces cerevisiae. *Mol Cell Biol*, 10:5796-5805.

Pacharra S, Hanisch FG, Breloy I. 2012. Neurofascin 186 is O-mannosylated within and outside of the mucin domain. *J Proteome Res*, 11:3955-3964.

Pacharra S, Hanisch FG, Muhlenhoff M, Faissner A, Rauch U, Breloy I. 2013. The lecticans of mammalian brain perineural net are O-mannosylated. *J Proteome Res*, 12:1764-1771.

Panin VM, Wells L. 2014. Protein O-mannosylation in metazoan organisms. *Curr Protoc Protein Sci*, 75:Unit 12 12.

Pappalardo A, Pitto L, Fiorillo C, Alice Donati M, Bruno C, Santorelli FM. 2013. Neuromuscular disorders in zebrafish: state of the art and future perspectives. *Neuromolecular Med*, 15:405-419.

Parsons MJ, Campos I, Hirst EM, Stemple DL. 2002. Removal of dystroglycan causes severe muscular dystrophy in zebrafish embryos. *Development*, 129:3505-3512.

Patnaik SK, Stanley P. 2005. Mouse large can modify complex N- and mucin O-glycans on alpha-dystroglycan to induce laminin binding. *J Biol Chem*, 280:20851-20859.

Patnaik SK, Stanley P. 2006. Lectin-resistant CHO glycosylation mutants. *Methods Enzymol*, 416:159-182.

Peyrard M, Seroussi E, Sandberg-Nordqvist AC, Xie YG, Han FY, Fransson I, Collins J, Dunham I, Kost-Alimova M, Imreh S, *et al.* 1999. The human LARGE gene from 22q12.3-q13.1 is a new, distinct member of the glycosyltransferase gene family. *Proc Natl Acad Sci U S A*, 96:598-603.

Porterfield M, Zhao P, Han HY, Cunningham J, Aoki K, Von Hoff DD, Demeure MJ, Pierce JM, Tiemeyer M, Wells L. 2014. Discrimination between Adenocarcinoma and Normal Pancreatic Ductal Fluid by Proteomic and Glycomic Analysis. *J Proteome Res*, 13:395-407.

Praissman JL, Live DH, Wang S, Ramiah A, Chinoy ZS, Boons GJ, Moremen KW, Wells L. 2014a. B4GAT1 is the priming enzyme for the LARGE-dependent functional glycosylation of alpha-dystroglycan. *eLife*, 3.

Praissman JL, Wells L. 2014b. Mammalian O-mannosylation pathway: glycan structures, enzymes, and protein substrates. *Biochemistry-Us*, 53:3066-3078.

Praissman JL, Willer T, Sheikh MO, Toi A, Chitayat D, Lin YY, Lee H, Stalnaker SH, Wang S, Prabhakar PK, *et al.* 2016. The functional O-mannose glycan on α-dystroglycan contains a phospho-ribitol primed for matriglycan addition. *eLife*, 5.

Riemersma M, Froese DS, van Tol W, Engelke Udo F, Kopec J, van Scherpenzeel M, Ashikov A, Krojer T, von Delft F, Tessari M, *et al.* 2015a. Human ISPD Is a Cytidyltransferase Required for Dystroglycan O-Mannosylation. *Chem Biol*, 22:1643-1652.

Riemersma M, Sandrock J, Boltje TJ, Bull C, Heise T, Ashikov A, Adema GJ, van Bokhoven H, Lefeber DJ. 2015b. Disease mutations in CMP-sialic acid transporter SLC35A1 result in abnormal alpha-dystroglycan O-mannosylation, independent from sialic acid. *Hum Mol Genet*, 24:2241-2246.

Roscioli T, Kamsteeg EJ, Buysse K, Maystadt I, van Reeuwijk J, van den Elzen C, van Beusekom E, Riemersma M, Pfundt R, Vissers LE, *et al.* 2012. Mutations in ISPD cause Walker-Warburg syndrome and defective glycosylation of alpha-dystroglycan. *Nat Genet*, 44:581-585.

Sadreyev IR, Ji F, Cohen E, Ruvkun G, Tabach Y. 2015. PhyloGene server for identification and visualization of co-evolving proteins using normalized phylogenetic profiles. *Nucleic acids research*, 43:W154-159.

Sasaki T, Yamada H, Matsumura K, Shimizu T, Kobata A, Endo T. 1998. Detection of O-mannosyl glycans in rabbit skeletal muscle alpha-dystroglycan. *Bba-Gen Subjects*, 1425:599-606.

Schollen E , Dorland L, de Koning TJ, Van Diggelen OP, Huijmans JG, Marquardt T, Babovic-Vuksanovic D, Patterson M, Imtiaz F, Winchester B *et al*. 2000. Genomic organization of the human phosphomannose isomerase (MPI) gene and mutation analysis in patients with congenital disorders of glycosylation type Ib (CDG-Ib). *Hum Mutat*. 16:247–252.

Sciandra F, Bigotti MG, Giardina B, Bozzi M, Brancaccio A. 2015. Genetic Engineering of Dystroglycan in Animal Models of Muscular Dystrophy. *BioMed research international*, 2015:635792.

Sentandreu R, Northcote DH. 1968. The structure of a glycopeptide isolated from the yeast cell wall. *Biochem J*, 109:419-432.

Sgambato A, Migaldi M, Montanari M, Camerini A, Brancaccio A, Rossi G, Cangiano R, Losasso C, Capelli G, Trentini GP, *et al.* 2003. Dystroglycan expression is frequently reduced in human breast and colon cancers and is associated with tumor progression. *Am J Pathol*, 162:849-860.

Shen JG, Xu CY, Li X, Dong MJ, Jiang ZN, Wang J, Wang LB. 2012. Dystroglycan is associated with tumor progression and patient survival in gastric cancer. *Pathol Oncol Res*, 18:79-84.

Smalheiser NR, Haslam SM, Sutton-Smith M, Morris HR, Dell A. 1998. Structural analysis of sequences O-linked to mannose reveals a novel Lewis X structure in cranin (Dystroglycan) purified from sheep brain. *Journal of Biological Chemistry*, 273:23698-23703.

Sparks SE , Krasnewich DM. 1993. PMM2-CDG (CDG-Ia). In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJH, Bird TD, Ledbetter N, Mefford HC, Smith RJH *et al.* editors. *GeneReviews(R)*. Seattle (WA): University of Washington, Seattle.

Stalnaker SH, Aoki K, Lim JM, Porterfield M, Liu M, Satz JS, Buskirk S, Xiong Y, Zhang P, Campbell KP, *et al.* 2011. Glycomic analyses of mouse models of congenital muscular dystrophy. *J Biol Chem*, 286:21180-21190.

Stalnaker SH, Hashmi S, Lim J-M, Aoki K, Porterfield M, Gutierrez-Sanchez G, Wheeler J, Ervasti JM, Bergmann C, Tiemeyer M, *et al.* 2010. Site Mapping and Characterization of O-Glycan Structures on α-Dystroglycan Isolated from Rabbit Skeletal Muscle. *The Journal of Biological Chemistry*, 285:24882-24891.

Steentoft C, Vakhrushev SY, Vester-Christensen MB, Schjoldager KTBG, Kong Y, Bennett EP, Mandel U, Wandall H, Levery SB, Clausen H. 2011. Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat Methods*, 8:977-982.

Steffen LS, Guyon JR, Vogel ED, Beltre R, Pusack TJ, Zhou Y, Zon LI, Kunkel LM. 2007. Zebrafish orthologs of human muscular dystrophy genes. *BMC genomics*, 8:79.

Stevens E, Carss KJ, Cirak S, Foley AR, Torelli S, Willer T, Tambunan DE, Yau S, Brodd L, Sewry CA, *et al.* 2013. Mutations in B3GALNT2 cause congenital muscular dystrophy and hypoglycosylation of alpha-dystroglycan. *Am J Hum Genet*, 92:354-365.

Sun T, Yu SH, Zhao P, Meng L, Moremen KW, Wells L, Steet R, Boons GJ. 2016. One-Step Selective Exoenzymatic Labeling (SEEL) Strategy for the Biotinylation and Identification of Glycoproteins of Living Cells. *J Am Chem Soc*, 138:11575-11582.

Taniguchi-Ikeda M, Morioka I, Iijima K, Toda T. 2016. Mechanistic aspects of the formation of alpha-dystroglycan and therapeutic research for the treatment of alpha-dystroglycanopathy: A review. *Mol Aspects Med*, 51:115-124.

Thornhill P, Bassett D, Lochmuller H, Bushby K, Straub V. 2008. Developmental defects in a zebrafish model for muscular dystrophies associated with the loss of fukutin-related protein (FKRP). *Brain*, 131:1551-1561.

Varki A, Cummings RD, Aebi M, Packer NH, Seeberger PH, Esko JD, Stanley P, Hart G, Darvill A, Kinoshita T, *et al.* 2015. Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology*, 25:1323-1324.

Vester-Christensen MB, Halim A, Joshi HJ, Steentoft C, Bennett EP, Levery SB, Vakhrushev SY, Clausen H. 2013. Mining the O-mannose glycoproteome reveals cadherins as major O-mannosylated glycoproteins. *Proceedings of the National Academy of Sciences*, 110:21018-21023.

Wells L. 2013. The o-mannosylation pathway: glycosyltransferases and proteins implicated in congenital muscular dystrophy. *J Biol Chem*, 288:6930-6935.

Willer T, Inamori K, Venzke D, Harvey C, Morgensen G, Hara Y, Beltran Valero de Bernabe D, Yu L, Wright KM, Campbell KP. 2014. The glucuronyltransferase B4GAT1 is required for initiation of LARGE-mediated alpha-dystroglycan functional glycosylation. *eLife*, 3.

Willer T, Lee H, Lommel M, Yoshida-Moriguchi T, de Bernabe DB, Venzke D, Cirak S, Schachter H, Vajsar J, Voit T, *et al.* 2012. ISPD loss-of-function mutations disrupt dystroglycan O-mannosylation and cause Walker-Warburg syndrome. *Nat Genet*, 44:575-580.

Williamson RA, Henry MD, Daniels KJ, Hrstka RF, Lee JC, Sunada Y, IbraghimovBeskrovnaya O, Campbell KP. 1997. Dystroglycan is essential for early embryonic development: Disruption of Reichert's membrane in Dag1-null mice. *Human Molecular Genetics*, 6:831-841.

Winterhalter PR, Lommel M, Ruppert T, Strahl S. 2013. O-glycosylation of the non-canonical T-cadherin from rabbit skeletal muscle by single mannose residues. *FEBS Lett*, 587:3715-3721.

Wood AJ, Currie PD. 2014. Analysing regenerative potential in zebrafish models of congenital muscular dystrophy. *Int J Biochem Cell Biol*, 56:30-37.

Wood AJ, Muller JS, Jepson CD, Laval SH, Lochmuller H, Bushby K, Barresi R, Straub V. 2011. Abnormal vascular development in zebrafish models for fukutin and FKRP deficiency. *Hum Mol Genet*, 20:4879-4890.

Xiong H, Kobayashi K, Tachikawa M, Manya H, Takeda S, Chiyonobu T, Fujikake N, Wang F, Nishimoto A, Morris GE, *et al.* 2006. Molecular interaction between fukutin and POMGnT1 in the glycosylation pathway of alpha-dystroglycan. *Biochem Biophys Res Commun*, 350:935-941.

Yagi H, Kuo CW, Obayashi T, Ninagawa S, Khoo KH, Kato K. 2016. Direct Mapping of Additional Modifications on Phosphorylated O-glycans of alpha-Dystroglycan by Mass Spectrometry Analysis in Conjunction with Knocking Out of Causative Genes for Dystroglycanopathy. *Mol Cell Proteomics*, 15:3424-3434.

Yagi H, Nakagawa N, Saito T, Kiyonari H, Abe T, Toda T, Wu SW, Khoo KH, Oka S, Kato K. 2013. AGO61-dependent GlcNAc modification primes the formation of functional glycans on alpha-dystroglycan. *Sci Rep*, 3:3288.

Yaji S, Manya H, Nakagawa N, Takematsu H, Endo T, Kannagi R, Yoshihara T, Asano M, Oka S. 2015. Major glycan structure underlying expression of the Lewis X epitope in the developing brain is O-mannose-linked glycans on phosphacan/RPTPbeta. *Glycobiology*, 25:376-385.

Yang AC, Ng BG, Moore SA, Rush J, Waechter CJ, Raymond KM, Willer T, Campbell KP, Freeze HH, Mehta L. 2013. Congenital disorder of glycosylation due to DPM1 mutations presenting with dystroglycanopathy-type congenital muscular dystrophy. *Mol Genet Metab*, 110:345-351.

Yiu EM, Kornberg AJ. 2015. Duchenne muscular dystrophy. *J Paediatr Child Health*, 51:759-764.

Yoshida-Moriguchi T, Campbell KP. 2015. Matriglycan: a novel polysaccharide that links dystroglycan to the basement membrane. *Glycobiology*, 25:702-713.

Yoshida-Moriguchi T, Willer T, Anderson ME, Venzke D, Whyte T, Muntoni F, Lee H, Nelson SF, Yu LP, Campbell KP. 2013. SGK196 Is a Glycosylation-Specific O-Mannose Kinase Required for Dystroglycan Function. *Science*, 341:896-899.

Yoshida-Moriguchi T, Yu LP, Stalnaker SH, Davis S, Kunz S, Madson M, Oldstone MBA, Schachter H, Wells L, Campbell KP. 2010. O-Mannosyl Phosphorylation of Alpha-Dystroglycan Is Required for Laminin Binding. *Science*, 327:88-92.

Yoshida A, Kobayashi K, Manya H, Taniguchi K, Kano H, Mizuno M, Inazu T, Mitsuhashi H, Takahashi S, Takeuchi M, *et al.* 2001. Muscular dystrophy and neuronal migration disorder caused by mutations in a glycosyltransferase, POMGnT1. *Dev Cell*, 1:717-724.

Zhang P, Hu H. 2012. Differential glycosylation of alpha-dystroglycan and proteins other than alpha-dystroglycan by like-glycosyltransferase. *Glycobiology*, 22:235-247.

Zhang Z, Zhang P, Hu H. 2011. LARGE expression augments the glycosylation of glycoproteins in addition to alpha-dystroglycan conferring laminin binding. *PLoS One,* 6:e19080.

Zhao P, Stalnaker SH, Wells L. 2013. Approaches for site mapping and quantification of O-linked glycopeptides. *Methods in molecular biology*, 951:229-244.

Zhu Q, Venzke D, Walimbe AS, Anderson ME, Fu Q, Kinch LN, Wang W, Chen X, Grishin NV, Huang N, *et al.* 2016. Structure of protein O-mannose kinase reveals a unique active site architecture. *Elife*, 5.

Supplemental Material Table S2.1. Currently known O-mannosylated proteins.

Protein Family	UniProt ID	Gene	Function	Origin	Site-Mapping Reference(s)
Dystroglycan	Q14118	*DAG1*	extracellular glycoprotein that acts as a receptor for LG-domain containing ECM proteins	MDA-MB-231 cells	Stalnaker et al. 2010; Yoshida-Moriguchi et al. 2010; Nilsson et al. 2010; Harrison et al. 2012
Cadherins	P12830	*CDH1*	calcium-dependent cell adhesion proteins	Recombinantly expressed in MDCK cells	Lommel et al. 2013
	Q9UEJ6	*ME5*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9NYQ6	*CELSR1*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9HCU4	*CELSR2*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	P55287	*CDH11*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	P55290	*CDH13*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9H159	*CDH19*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	P19022	*CDH2*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9H251	*CDH23*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	P22223	*CDH3*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	P55283	*CDH4*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	P55285	*CDH6*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	P55286	*CDH8*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	P55288	*CDH11*		Mouse brain	Bartels et al. 2016
	Q9WTR5	*CDH13*		Mouse brain	Bartels et al. 2016
	G1TL92	*G1TL92_RABIT*	uncharacterized calcium ion binding protein	Rabbit skeletal muscle	Winterhalter et al. 2013
Desmocollin-2	Q02487	*DSC2*	component of intercellular desmosome junctions	MDA-MB-231 cells	Vester-Christensen et al. 2013
Desmoglein-2	Q14126	*DSG2*	component of intercellular desmosome junctions	MDA-MB-231 cells	Vester-Christensen et al. 2013
Hepatocyte growth factor receptor	P08581	*MET*	receptor tyrosine kinase that transduces signals from the extracellular matrix into the cytoplasm by binding to hepatocyte growth factor/HGF ligand	MDA-MB-231 cells	Vester-Christensen et al. 2013
inter-alpha-trypsin inhibitor 5	Q8BJD1	*ITIH5*	protease inhibitor important in stabilizing the extracellular matrix	Mouse brain	Bartels et al. 2016
Intercellular adhesion molecule 1	P05362	*ICAM1*	ligand for the leukocyte adhesion protein LFA-1 and receptor for rhinovirus A-B capsid proteins	MDA-MB-231 cells	Vester-Christensen et al. 2013
Lecticans	P55066	*NCAN*	proteoglycans of the perineuronal net that interact with hyaluronic acid and tenascin-R	Mouse brain	Pacharra et al. 2013; Bartels et al. 2016
	F1N2Y8	*NCAN*		Calf brain	Pacharra et al. 2013
	P81282	*VCAN*		Calf brain	Pacharra et al. 2013
Macrophage-stimulating protein receptor	Q04912	*MST1R*	receptor tyrosine kinase that transduces signals from the extracellular matrix into the cytoplasm by binding to MST1 ligand	MDA-MB-231 cells	Vester-Christensen et al. 2013
Neurofascin 186	Q810U3	*NFASC*	cell adhesion, ankyrin-binding protein which may be involved in neurite extension, axonal guidance, synaptogenesis, myelination and neuron-glial cell interactions	Mouse brain and HEK293 EBNA cells	Pacharra et al. 2012
Neurexin 3	Q6P9K9	*NRX3*	cell adhesion protein involved in synaptic plasticity	Mouse brain	Bartels et al. 2016
Plexins	Q9UIW2	*PLXNA1*	receptors for semaphorin	MDA-MB-231 cells	Vester-Christensen et al. 2013
	O75051	*PLXNA2*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	O43157	*PLXNB1*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	O15031	*PLXNB2*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9ULL4	*PLXNB3*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9Y4D7	*PLXND1*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	P70206	*PLXNA1*		Mouse brain	Bartels et al. 2016
	B2RXS4	*PLXNB2*		Mouse brain	Bartels et al. 2016

Protein disulfide-isomerase A3	P27773	*PDIA3*	catalyzes the rearrangement of -S-S- bonds in proteins	Mouse brain	Bartels et al. 2016
	P30101	*PDIA3*		MDA-MB-231 cells	Vester-Christensen et al. 2013
Protein disulfide-isomerase	P07237	*P4HB*		MDA-MB-231 cells	Vester-Christensen et al. 2013
Protocadherins	Q9Y5I4	*PCDHAC2*	cell-adhesion proteins	MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9UN67	*PCDHB10*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9Y5F2	*PCDHB11*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9Y5E9	*PCDHB14*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9Y5E7	*PCDHB2*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9UN66	*PCDHB8*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9Y5E1	*PCDHB9*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q14517	*FAT1*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q8BNA6	*FAT3*		Mouse brain	Bartels et al. 2016
	Q9QXA3	*FAT1*		Mouse brain	Bartels et al. 2016
	Q6V0I7	*FAT4*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9Y5H3	*PCDHGA10*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9Y5G9	*PCDHGA4*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9Y5G3	*PCDHGB1*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9Y5G2	*PCDHGB2*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9UN70	*PCDHGC3*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9Y5F6	*PCDHGC5*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q08174	*PCDH1*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9P2E7	*PCDH10*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9NPG4	*PCDH12*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q96JQ0	*DCHS1*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	O14917	*PCDH17*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	O14917	*PCDH17*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q9HCL0	*PCDH18*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	O95206	*PCDH8*		MDA-MB-231 cells	Vester-Christensen et al. 2013
	Q91XX5	*PCDHGB5*		Mouse brain	Bartels et al. 2016
	Q91Y09	*PCDHAC2*		Mouse brain	Bartels et al. 2016
	F8VPK8	*PCDH9*		Mouse brain	Bartels et al. 2016
	E9PXQ7	*PCDH10*		Mouse brain	Bartels et al. 2016
	Q91XY7	*PCDHGA12*		Mouse brain	Bartels et al. 2016
	Q9HC56	*PCDH9*		MDA-MB-231 cells	Vester-Christensen et al. 2013
Receptor-type tyrosine-protein phosphatase beta	P23467	*PTPRB*	blood vessel remodeling and angiogenesis	Human bone marrow neuroblast (SH-SY5Y cells)	Abbott et al. 2008
Receptor-type tyrosine-protein phosphatase zeta	B9EKR1	*PRPTZ1*	required for normal differentiation of the precursor cells into mature, fully myelinating oligodendrocytes	Mouse brain	Dwyer et al. 2012, Morise et al. 2014, Yaji et al. 2015, Bartels et al. 2016
Sinal transducer CD24	P24807	*CD24*	early thymocyte development	Mouse brain	Bleckmann et al. 2009
SUN domain-containing ossification factor	Q9UBS9	*SUCO*	required for bone modeling during late embryogenesis	MDA-MB-231 cells	Vester-Christensen et al. 2013
TGF-beta receptor type-1	P36897	*TGFBR1*	transmembrane serine/threonine kinase forming a complex with TGFBR2 that transduces signals from the cell surface to the cytoplasm by binding TGF-beta cytokines (TGFB1, TGFB2 and TGFB3)	MDA-MB-231 cells	Vester-Christensen et al. 2013
UPF0606 protein KIAA1549	Q9HCM3	*KIAA1549*	unknown	MDA-MB-231 cells	Vester-Christensen et al. 2013

CHAPTER 3

PROTEIN *O*-LINKED MANNOSE β-1,4-*N*-ACETYLGLUCOSAMINYL-TRANSFERASE 2 (POMGNT2) IS A GATEKEEPER ENZYME FOR FUNCTIONAL GLYCOSYLATION OF α-DYSTROGLYCAN[1]

[1]Halmo, S.M., Singh, D., Patel, S., Wang, S., Edlin, M., Boons, G.J., Moremen, K.W., Live, D. and Wells, L. 2017. *J Biol Chem*. 292(6): 2101–2109.

Abstract

Disruption of the *O*-mannosylation pathway involved in functional glycosylation of α-dystroglycan gives rise to congenital muscular dystrophies. Protein *O*-linked mannose β-1,4-*N*-acetylglucosaminyltransferase 2 (POMGNT2) catalyzes the first step toward the functional matriglycan structure on α-dystroglycan that is responsible for binding extracellular matrix proteins and certain arenaviruses. Alternatively, protein *O*-linked mannose β-1,2-*N*-acetylglucosaminyltransferase 1 (POMGNT1) catalyzes the first step toward other various glycan structures present on α-dystroglycan of unknown function. Here, we demonstrate that POMGNT1 is promiscuous for *O*-mannosylated peptides, whereas POMGNT2 displays significant primary amino acid selectivity near the site of *O*-mannosylation. We define a POMGNT2 acceptor motif, conserved among 59 vertebrate species, in α-dystroglycan that when engineered into a POMGNT1-only site is sufficient to convert the *O*-mannosylated peptide to a substrate for POMGNT2. Additionally, an acceptor glycopeptide is a less efficient substrate for POMGNT2 when two of the conserved amino acids are replaced. These findings begin to define the selectivity of POMGNT2 and suggest that this enzyme functions as a gatekeeper enzyme to prevent the vast majority of *O*-mannosylated sites on proteins from becoming modified with glycan structures functional for binding laminin globular domain-containing proteins.

Introduction

Congenital muscular dystrophy (CMD; Live, D., Wells, L., et al. 2013) describes a family of genetic, degenerative diseases characterized by contractures, myopathy, and in some cases, central nervous system abnormalities. Many CMDs are caused by defects in the formation of a functional dystrophin glycoprotein complex (DGC) that links the actin cytoskeleton to the extracellular matrix (ECM). α-Dystroglycan (α-DG), encoded by the *DAG1* gene, provides the physical link to LG-domain containing proteins in the ECM (Ervasti, J.M. and Campbell, K.P. 1993), however, there are only a few known mutations in the DAG1 coding sequence that lead to CMD (Yoshida-Moriguchi, T. and Campbell, K.P. 2015). A subset of CMDs, termed secondary dystroglycanopathies, is caused by mutations in genes encoding enzymes responsible for glycosylating α-DG in its mucin-like domain (residues 313-489). These secondary dystroglycanopathies range in severity from mild Limb-Girdle muscular dystrophy to the more severe Walker-Warburg syndrome (Godfrey, C., Foley, A.R., et al. 2011, Live, D., Wells, L., et al. 2013, Voglmeir, J., Kaloo, S., et al. 2011). The causal genes for secondary dystroglycanopathies have been identified as encoding enzymes in the pathway associated with the biosynthesis of the *O*-mannosyl (*O*-Man) glycans (Jae, L.T., Raaben, M., et al. 2013, Wells, L. 2013).

The *O*-mannosylation pathway begins in the endoplasmic reticulum (ER) where a complex of POMT1 and POMT2 catalyze the transfer of mannose from dolicholphosphomannose to serine and threonine residues in an α-linkage to α-DG (Manya, H., Chiba, A., et al. 2004), and presumably a handful of other proteins (Vester-Christensen, M.B., Halim, A., et al. 2013). Bifurcation of the pathway then occurs by the addition of an N-acetylglucosamine in either a β2 or a β4 linkage (Figure 3.1). Two enzymes, POMGNT1 and POMGNT2, mediate these additions, respectively. In most cases on α-DG, a β-1,2 linked GlcNAc residue can be added to the initial mannose residue

by POMGNT1 in the *cis*-Golgi (Takahashi, S., Sasaki, T., et al. 2001). This core M1 structure can be branched by another GlcNAc addition to give rise to the core M2 glycan structure (Praissman, J.L. and Wells, L. 2014). Much more rarely on α-DG, POMGNT2 will add a β-1,4 linked GlcNAc to the initial mannose residue in the ER, leading to the formation of the core M3 glycan structure (Figure 3.1).

After POMGNT2 mediated β-1,4 GlcNAc addition, the glycan is subject to further extension with a β-1,3 linked N-acetyl galactosamine (GalNAc) by B3GALNT2 and phosphorylation of the reducing-end mannose at the 6-position by POMK to give rise to the phosphotrisaccharide core M3 glycan structure while still in the ER (Ogawa, M., Nakamura, N., et al. 2013, Yagi, H., Nakagawa, N., et al. 2013, Yoshida-Moriguchi, T., Willer, T., et al. 2013). From here, it has been recently demonstrated that Fukutin (FKTN) and Fukutin-related protein (FKRP) appear to be responsible for extending the core M3 phosphotrisaccharide in the Golgi by addition of two ribitol-phosphate units in phosphodiester linkages (Kanagawa, M., Kobayashi, K., et al. 2016). TMEM5 then apparently adds a xylose (Xyl) to the distal ribitol that is followed by B4GAT1-catalyzed addition of glucuronic acid (GlcA) in a β-1,4 linkage to the xylose (Praissman, J.L., Live, D.H., et al. 2014, Praissman, J.L., Willer, T., et al. 2016). This primer permits LARGE1 to catalyze the addition of a repeating disaccharide (α-1,3 linked Xyl – β-1,3 linked GlcA) that is the functional component, termed matriglycan, responsible for binding to laminin globular (LG) domains of ECM proteins (Hara, Y., Kanagawa, M., et al. 2011, Kanagawa, M., Saito, F., et al. 2004, Yoshida-Moriguchi, T. and Campbell, K.P. 2015).

Human α-DG has at least 25 *O*-mannosylation sites (Harrison, R., Hitchen, P.G., et al. 2012, Nilsson, J., Nilsson, J., et al. 2010, Stalnaker, S.H., Hashmi, S., et al. 2010). The majority of the *O*-mannosylation sites on α-DG are populated by core M1 and M2 glycan structures via the

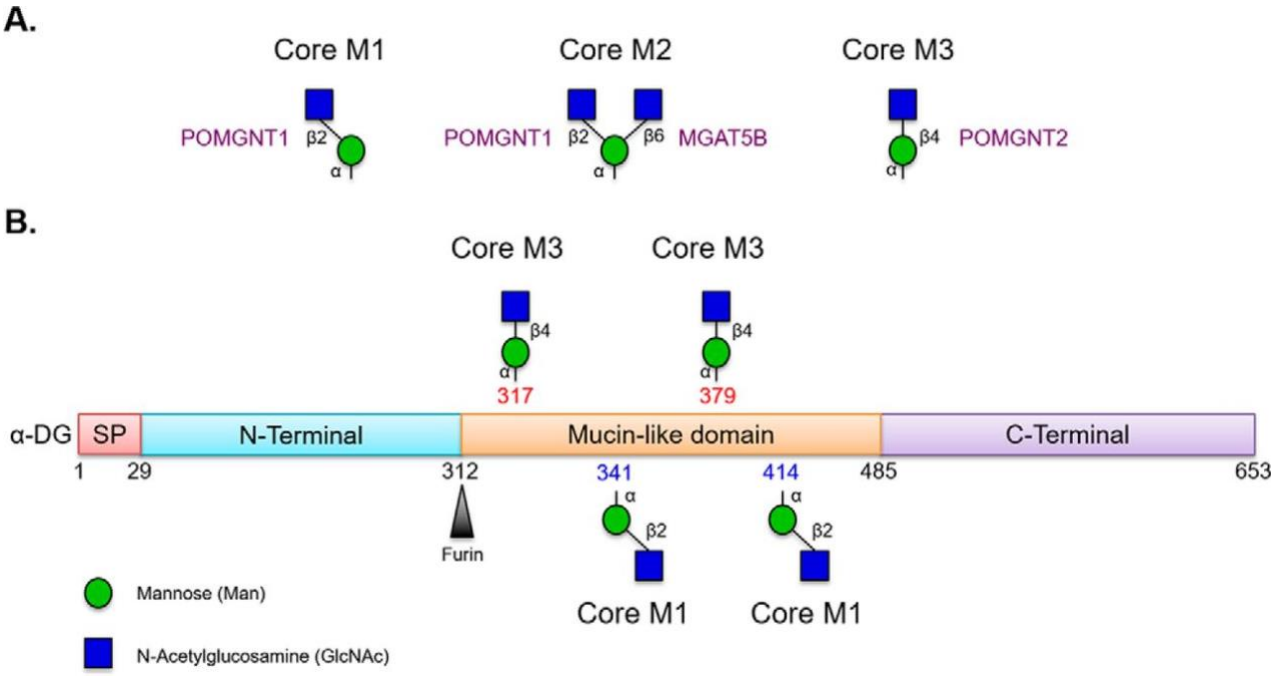


Figure 3.1. Core *O*-Man structures on α-Dystroglycan
(**A**) POMGNT1 is responsible for generating the M1 core glycan structure that can be branched by MGAT5B to generate the M2 core, while POMGNT2 is responsible for generating the M3 core glycan structure.
(**B**) Schematic of known *O*-mannosylated sites on α-dystroglycan addressed in this study. Thr317 and Thr379 are elaborated with the M3 core glycan structure, while Thr341 and Thr414 are elaborated with M1 core glycan structures that can be further elaborated to core M2 glycan structures. Glycan symbols follow guidelines outlined in Varki, A., Cummings, R.D., et al. (2015).

action of POMGNT1 (M1) followed by MGAT5B (M2) (Harrison, R., Hitchen, P.G., et al. 2012, Lee, J.K., Matthews, R.T., et al. 2012, Nilsson, J., Nilsson, J., et al. 2010, Stalnaker, S.H., Hashmi, S., et al. 2010). Site-mapping studies have identified only two positions, Thr317 and Thr379, on α-DG with M3 core structures, though some evidence suggests 319 and 381 may also be sites of M3 modification (Figure 3.1) (Hara, Y., Kanagawa, M., et al. 2011, Nilsson, J., Nilsson, J., et al. 2010, Stalnaker, S.H., Hashmi, S., et al. 2010, Yagi, H., Nakagawa, N., et al. 2013). Paradoxically from a spatial-temporal perspective, *O*-Man modified α-DG encounters POMGNT2 in the ER before POMGNT1 in the *cis*-Golgi yet is preferentially modified by POMGNT1. This led us to hypothesize that POMGNT2 must demonstrate substrate selectivity beyond simply an *O*-Man modified amino acid.

Here, we explore the specificity of POMGNT2, and compare it with POMGNT1. We synthesized multiple *O*-mannosylated peptides derived from known M1 and M3 modified sites of α-DG and tested their ability to be acceptor substrates for the two enzymes. POMGNT2 displays selectivity based on the primary amino acid sequence in proximity to the site of *O*-mannosylation while POMGNT1 is promiscuous. We identified a sequence motif, highly conserved in vertebrates, in α-DG that appears to modulate POMGNT2 substrate specificity *in vitro*. We demonstrated sufficiency of the extended motif by engineering the sequence into a typical M1 *O*-mannosylated peptide that resulted in it being a POMGNT2 acceptor. We also demonstrate that replacement of conserved amino acids compromises an M3 peptide for extension by POMGNT2. Intriguingly, a conservative degenerate sequence based on our identified motif is present in several human membrane/secreted proteins.

Results

Acceptor Selectivity of POMGNT1 and POMGNT2 using Synthetic α-DG Glycopeptides.

In order to identify primary amino acid determinants of POMGNT2 selectivity, we used solid-phase peptide synthesis to generate synthetic glycopeptides whose sequences are those from known *O*-mannosylated regions of human α-DG (Table 3.1). Direct physical evidence for core M3 extension at position 379 in α-DG has previously been shown (Yoshida-Moriguchi, T., Yu, L., et al. 2010), while the threonine at position 341 in α-DG has been demonstrated as a POMT1/POMT2 acceptor that does not carry an M3 core (Manya, H., Suzuki, T., et al. 2007). We selected these two sites (379 and 341) because we predicted their extensions differ in core glycan structure while their immediate primary amino acid sequences share a similar Thr(-*O*-Man)-Pro-Thr (TPT) motif. The synthetic glycopeptides were designed to be 21 amino acids in length with the mannosylated threonine as the central residue (residue 11) to evaluate nearby C-terminal and N-terminal amino acid determinants (Table 3.1).

To establish if the synthetic glycopeptides were substrates for POMGNT1 and POMGNT2, we performed overnight radioactive transfer assays. Recombinant human POMGNT1 catalyzed GlcNAc transfer to both the Man341 and Man379 synthetic glycopeptides (Figure 3.2 A,B). To confirm the composition of POMGNT1 reaction products, parallel transfer assays using non-radiolabeled UDP-GlcNAc were performed using the Man341 and Man379 glycopeptides as acceptor substrates, and the reaction products were analyzed by mass spectrometry (MS) (Figure 3.2 C,D). The observed peaks at 859.121 and 899.504 *m/z* in the full FTMS correspond to the addition of a HexNAc residue [+203] to the Man341 and Man379 glycopeptides, respectively. Thus, POMGNT1 will extend the mannose in synthetic glycopeptides at positions 341 and 379 of

Table 3.1. Comparison of POMGNT1 and POMGNT2 kinetics with various M1 and M3 synthetic glycopeptide acceptors

Asterisk in acceptor sequence indicates the mannosylated threonine residue. Kinetic parameters of Man341 and Man414 with POMGNT2 were not measurable as indicated by the dashed lines.

	Acceptor	Acceptor Sequence	Core Glycan Structure	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)
POMGNT1	Man317	PKRVRRQIHAT*PTPVTAIGPP	M3	1.2 ± 0.1	13 ± 0.5	10 x 10_3
	Man379	TIRTRGAIIQT*PTLGPIQPTR	M3	2.6 ± 0.4	32 ± 3.3	12 x 10_3
	Man379-ETP	TIETPGAIIQT*PTLGPIQPTR	Modified M3	0.9 ± 0.1	11 ± 0.4	12 x 10_3
	Man341	IQEPPSRIVPT*PTSPAIAPPT	M1	0.1 ± 0.04	1.0 ± 0.1	10 x 10_3
	Man341-RPR	IQRPRSRIVPT*PTSPAIAPPT	Modified M1	3.2 ± 0.7	13 ± 1.5	4.0 x 10_3
	Man414	YVEPT*AV	M1	11 ± 3.0	16 ± 3.3	1.5 x 10_3
POMGNT2	Man317	PKRVRRQIHAT*PTPVTAIGPP	M3	2.2 ± 0.2	12 ± 0.5	5.5 x 10_3
	Man379	TIRTRGAIIQT*PTLGPIQPTR	M3	0.8 ± .04	16 ± 1.4	20 x 10_3
	Man379-ETP	TIETPGAIIQT*PTLGPIQPTR	Modified M3	2.9 ± 0.7	11 ± 1.4	3.8 x 10_3
	Man341	IQEPPSRIVPT*PTSPAIAPPT	M1	--	--	--
	Man341-RPR	IQRPRSRIVPT*PTSPAIAPPT	Modified M1	6.0 ± 1.1	3.3 ± 1.0	0.6 x 10_3
	Man414	YVEPT*AV	M1	--	--	--
	ShortMan379	GAIIQT*PTLGPIQPTR	Modified M3	0.8 ± 0.3	10 ± 1.4	12 x 10_3

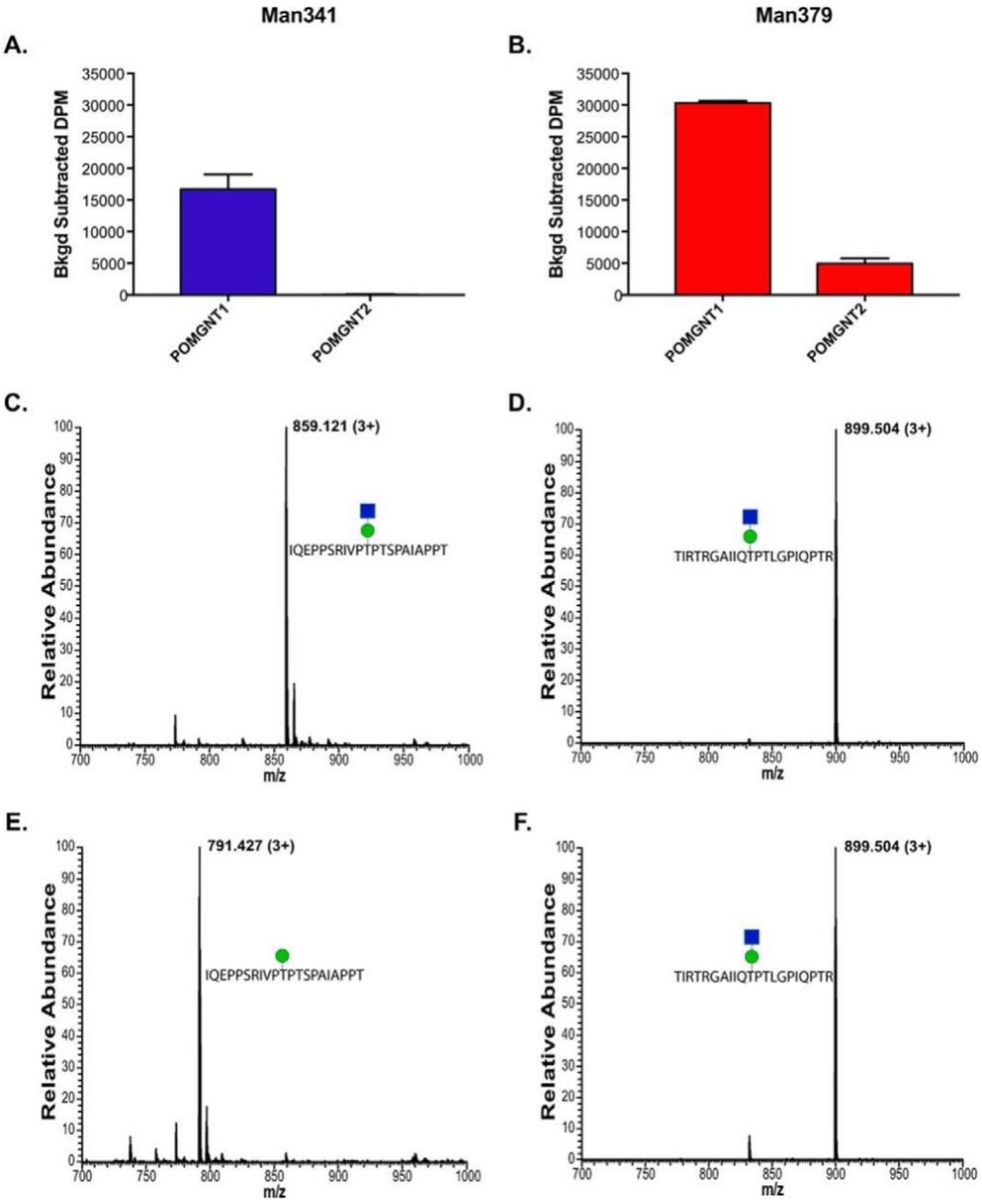


Figure 3.2. Unlike POMGNT1, POMGNT2 exhibits acceptor selectivity
(**A-B**) Radioactive assay of POMGNT1 and POMGNT2 activity with (**A**) Man341, an M1 acceptor and (**B**) Man379, an M3 acceptor. Transfer is measured in background corrected disintegrations per minute (DPM). Error bars represent standard error from the mean of 3 replicates.
(**C-D**) FTMS spectra verifying (**C**) POMGNT1 extended Man341 (1.16 ppm mass accuracy) and (**D**) POMGNT1 extended Man379 (2.53 ppm mass accuracy). Green circle represents a mannose and the blue square represents an N-acetylglucosamine Varki, A., Cummings, R.D., et al. (2015).
(**E-F**) FTMS spectra verifying (**E**) POMGNT2 Man341 product (1.11 ppm mass accuracy) and (**F**) POMGNT2 extended Man379 (1.11 ppm mass accuracy). Green circle represents a mannose and the blue square represents an N-acetylglucosamine Varki, A., Cummings, R.D., et al. (2015).

the α-DG sequence, *in vitro*. These results clearly demonstrate that POMGNT1 exhibits minimal acceptor selectivity between core M1 and M3 sites on these two α-DG-derived glycopeptides.

In comparison, POMGNT2 showed preferential acceptor selectivity for the known M3 site in the α-DG sequence. Radiolabel transfer assays showed no detectable transfer of the sugar to the acceptor Man341 glycopeptide by POMGNT2 (Figure 3.2 A) but transfer of GlcNAc to the Man379 synthetic glycopeptide (Figure 3.2 B). Parallel non-radioactive transfer assays followed by MS analysis identified the composition of the POMGNT2 reaction products and verified the transfer results observed in the radioactive assays (Figure 3.2 E,F). The predominant peak at 791.427 *m/z* in the full FTMS corresponds to the unmodified Man341 glycopeptide with a single mannose (Figure 3.2 E). In contrast to the results seen in Figure 3.2 E, the observed peak at 899.504 *m/z* in the full FTMS in Figure 3.2 F corresponds to the addition of a HexNAc residue [+203] to the Man379 glycopeptide. These results suggest that POMGNT2 preferentially modifies specific sites on α-DG intended for core M3 glycan elaboration.

Kinetic Parameters of POMGNT1 and POMGNT2 with Synthetic α-DG Glycopeptides. To further characterize the substrate specificities of POMGNT1 and POMGNT2, additional core M1 and core M3 synthetic glycopeptides were generated. Man414 and Man317 are known *O*-mannosylated regions of α-DG (Hara, Y., Kanagawa, M., et al. 2011, Stalnaker, S.H., Hashmi, S., et al. 2010). Evidence for core M3 extension at position 317 in α-DG has been shown previously (Hara, Y., Kanagawa, M., et al. 2011, Yagi, H., Nakagawa, N., et al. 2013). Man317 is 21 amino acids in length, contains the TPT motif, and has the mannosylated threonine as the central residue (residue 11), similar to Man379 and Man341 (Table 3.1). Man414 is only 7 amino acids in length, and lacks the TPT motif (Table 3.1). However, the kinetics of Man414 with POMGNT1 has previously been studied (Mo, K.-F., Fang, T., et al. 2011) and the homologous residue in rabbit

(*Oryctolagus cuniculus)* has been site-mapped with mannose (Stalnaker, S.H., Hashmi, S., et al. 2010) making it a useful predicted core M1 glycopeptide for this study.

Glycosyltransferase reaction kinetics for POMGNT1 with the four synthesized glycopeptides were investigated by UDP-Glo™ assays. The α-DG sequences in the four synthetic glycopeptides were all utilized by POMGNT1 as acceptors (Figure 3.3 A-D). Inspection of the K_m values derived from nonlinear regression analyses of the experimentally obtained values reveals that the affinity of POMGNT1 for synthetic acceptor glycopeptides containing a TPT motif (Man317, Man379, Man341) is greater than the affinity for the Short Man414 synthetic glycopeptide lacking a TPT motif (Table 3.1). POMGNT1 has the fastest turnover (k_{cat}) with the Man341 synthetic glycopeptide, but catalytic efficiency (k_{cat}/K_m) is an order of magnitude greater for the core M3 synthetic glycopeptides, Man317 and Man379 (Table 3.1).

To validate the acceptor selectivity of POMGNT2, we also performed UDP-Glo™ assays with the four synthesized glycopeptides to investigate glycosyltransferase reaction kinetics. Transfer of GlcNAc to Man341 and Man414 by POMGNT2 (Figure 3.4 A,C) was below the level of detection, while Man379 and Man317 (Figure 3.4 B,D) are clearly acceptor substrates for POMGNT2 activity. The measured K_m, k_{cat}, and k_{cat}/ K_m for Man317 and Man379 synthetic glycopeptides with POMGNT2 are similar (Table 3.1). These data are consistent with the results obtained in our initial transfer assays and support the proposal that the features of the primary amino acid sequence in the region of the TPT sequence are determinants of POMGNT2 selectivity.

A Primary Amino Acid Motif in α-DG is Favorable for POMGNT2 Activity. It was previously suggested that all *O*-mannosylated sites on α-DG have a conserved Thr-Pro-Thr (TPT) motif at the mannosylated threonine (Hara, Y., Kanagawa, M., et al. 2011, Manya, H., Suzuki, T., et al. 2007) though site mapping studies have demonstrated that only a subset of mapped sites

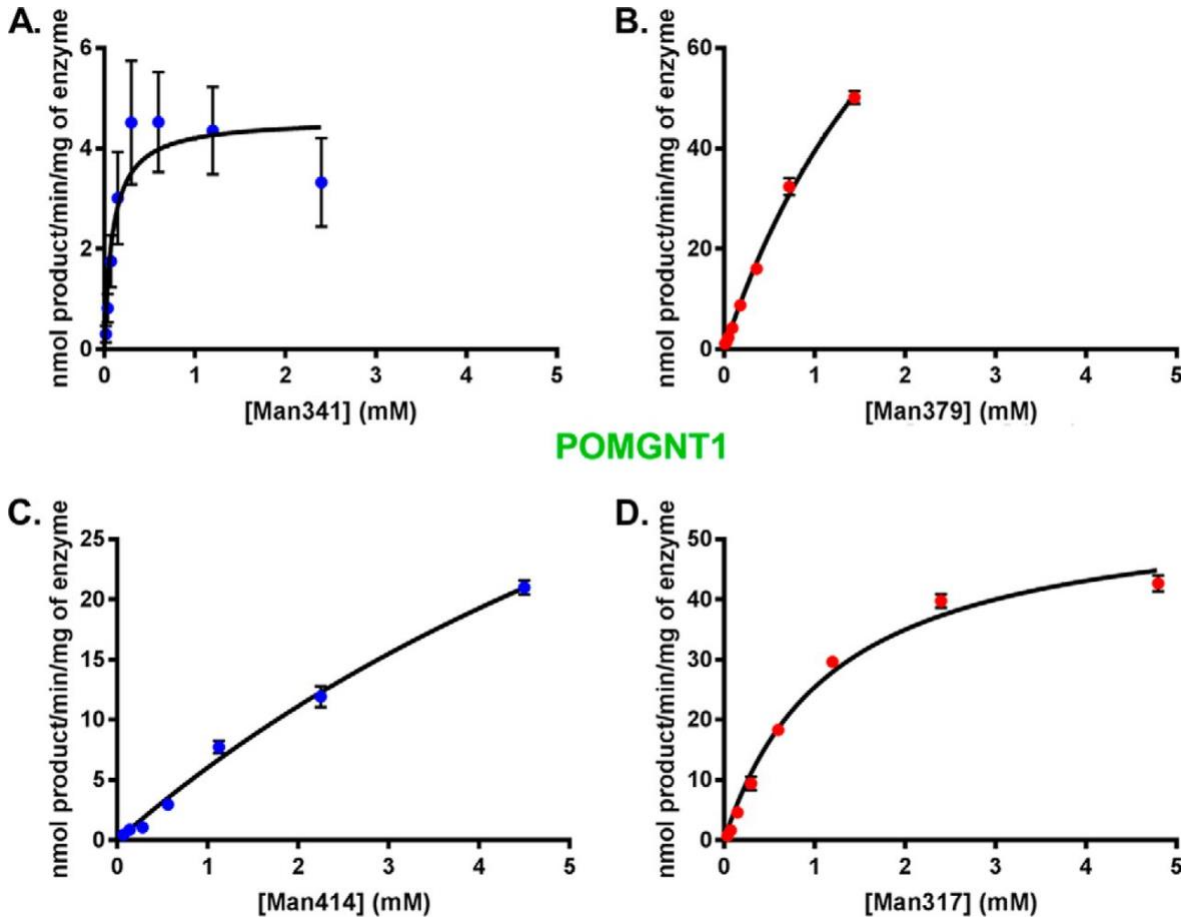


Figure 3.3. POMGNT1 transfers to both M1 and M3 acceptors
POMGNT1 kinetics with (**A**) Man341, (**B**) Man379, (**C**) Man414, or (**D**) Man317 acceptor glycopeptide measured by UDP-Glo assay. Error bars represent standard error from the mean from three experiments. See Table 3.1 for a list of kinetic parameters.

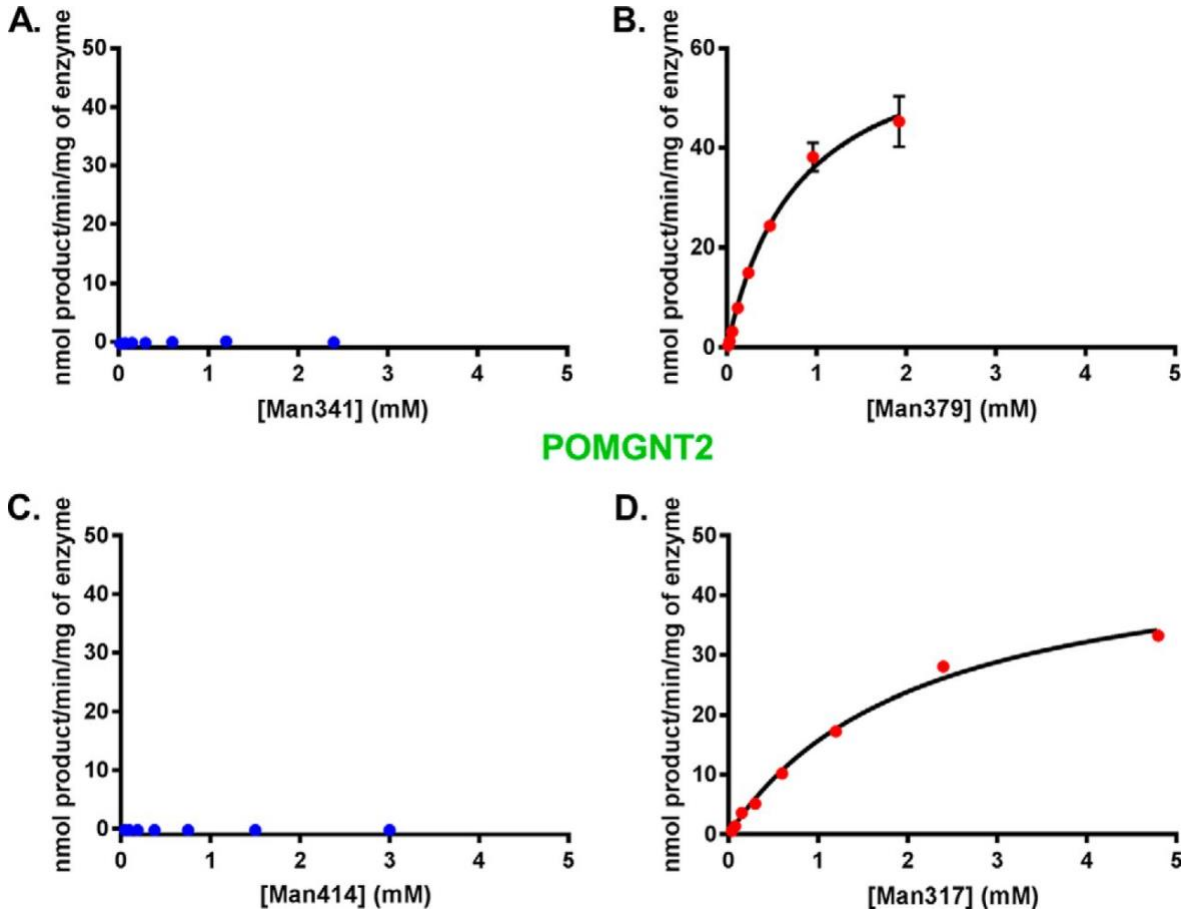


Figure 3.4. POMGNT2 only transfers to M3 acceptors
POMGNT2 kinetics with (**A**) Man341, (**B**) Man379, (**C**) Man414, or (**D**) Man317 acceptor glycopeptide measured by UDP-Glo assay. Error bars represent standard error from the mean from three experiments. See Table 3.1 for a list of kinetic parameters.

follow this pattern (Harrison, R., Hitchen, P.G., et al. 2012, Nilsson, J., Nilsson, J., et al. 2010, Stalnaker, S.H., Hashmi, S., et al. 2010). Indeed, the primary amino acid sequences around mapped *O*-mannose sites on α-DG (excluding sites Thr317 and Thr379) are heterogeneous (Figure 3.5 A). To identify primary sequence elements that govern the observed preferences on POMGNT2 acceptor substrate selectivity, α-DG amino acid sequences surrounding sites Thr317 and Thr379 from fifty-nine vertebrate species with orthologues of human DAG1, POMGNT1 or POMGNT2, and FKTN or B4GAT1 were aligned using WebLogo (Crooks, G.E., Hon, G., et al. 2004) (Figure 3.5 B). The previously identified TPT motif was evident in our alignment, but other conserved amino acids were observed that are not present around Thr341. Interestingly, our alignment of 317/379 M3 sites across species demonstrated that arginines at -6 and -8 and an Ile at -3 were conserved in addition to the P at +1 and the T at +2. Thus, the R-X-R-X-X-I-X-X-T-P-T motif is a proposed conserved sequence for M3 extension (Figure 3.5 B).

Since this motif is only present at known core M3 sites and not at core M1 sites, we hypothesized that the R-X-R portion of the primary amino acid sequence motif of α-DG might confer extension by POMGNT2. To test this, we synthesized a modified version of the Man341 peptide that already contains the TPT sequence and an Ile at -3, that we refer to as Man341-RPR. In this glycopeptide, we replaced the two divergent amino acids at -6 and -8 with arginines to introduce the conserved R-X-R motif. Glycosyltransferase reaction kinetics of POMGNT2 with this modified glycopeptide was investigated by UDP-Glo™ assay. In contrast to the undetectable reaction with Man341, POMGNT2 transferred GlcNAc to Man341-RPR (Figure 3.6 A, Table 3.1). Thus, we successfully converted a core M1 non-acceptor peptide into a core M3 acceptor for POMGNT2 *in vitro* by the addition of our identified motif.

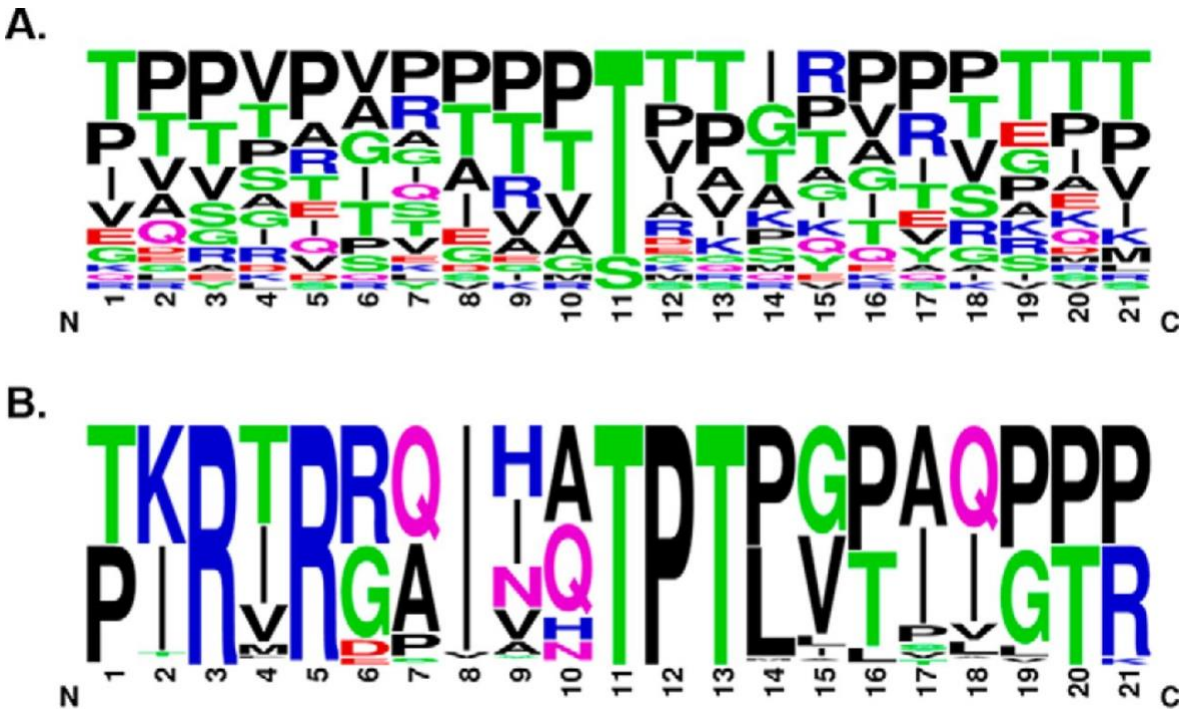


Figure 3.5. A Conserved Consensus Sequence for POMGNT2 Activity
(**A**) Sequence alignment of 21-mer sequences of α-DG centered on known *O*-mannose sites from human and rabbit (Nilsson, J., Nilsson, J., et al. 2010, Stalnaker, S.H., Hashmi, S., et al. 2010) excluding Thr317 and Thr379. Logo made using Berkeley's WebLogo program.
(**B**) Sequence alignment of 21-mer sequences of α-DG centered on human sites Thr317 and Thr379 from all Ensembl vertebrata with orthologues of DAG1, POMGNT1 or 2 and FKTN or B4GAT1 (total of 59 species, see Experimental Methods for a complete list). Logo made using Berkeley's WebLogo program.

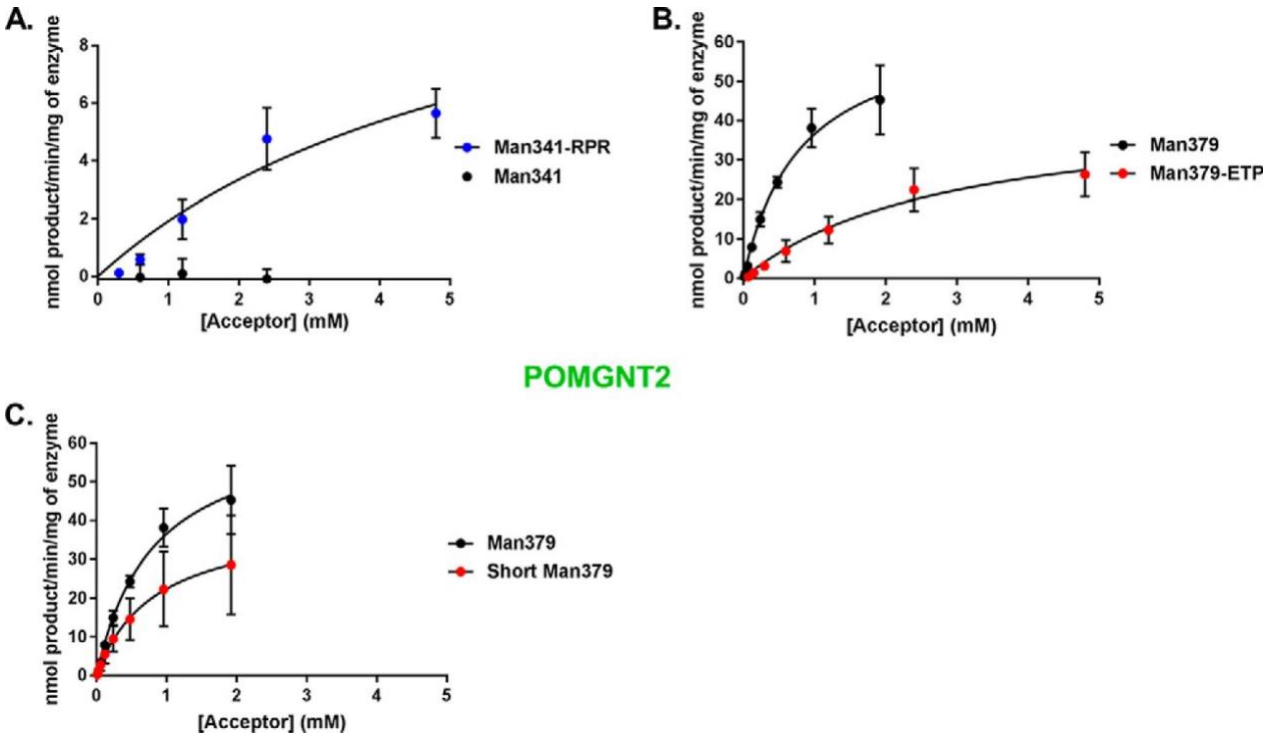


Figure 3.6. A Primary Amino Acid Motif in α-DG is Permissible for POMGNT2 Activity
POMGNT2 kinetics with (**A**) Man341-RPR and Man341, (**B**) Man379-ETP and Man379, and (**C**) ShortMan379 and Man379 acceptor glycopeptide measured by UDP-Glo assay. Error bars represent standard error from the mean from three experiments. See Table 3.1 for a list of kinetic parameters.

To further test that the R-X-R portion of the primary amino acid sequence motif of α-DG is important for glycan extension by POMGNT2, we synthesized a new glycopeptide based on the sequence at a known core M3 site (379) but with the two N-terminal arginine residues altered to the divergent residues of a core M1 acceptor (Man341). We have designated this modified core M3 glycopeptide as Man379-ETP. Glycosyltransferase reaction kinetics of POMGNT2 with this modified glycopeptide were investigated by UDP-Glo™ assays. In comparison to the kinetics of POMGNT2 with Man379, POMGNT2 has a lower affinity and a greater than five-fold reduction in catalytic efficiency for Man379-ETP (Figure 3.6 B, Table 3.1). The replacement of our identified R-X-R motif in a core M3 acceptor with divergent residues reduced but did not eliminate POMGNT2 activity.

Lastly, to test the necessity of the R-X-R portion of the primary amino acid sequence motif of α-DG for POMGNT2 activity, we synthesized a truncated version of the Man379 glycopeptide we refer to as ShortMan379. This N terminus of this glycopeptide begins immediately after the R-X-R motif. Glycosyltransferase reaction kinetics of POMGNT2 with this truncated glycopeptide was investigated by UDP-Glo™ assay. POMGNT2 utilized ShortMan379 as an acceptor substrate with a similar affinity to Man379 but with a less than one fold reduction in catalytic efficiency (Figure 3. 6 C, Table 3.1). Thus, the R-X-R portion of the motif is sufficient but not necessary for POMGNT2 activity.

Discussion

While POMGNT2 is poised to modify α-DG in the ER before it encounters POMGNT1 in the *cis*-Golgi, only two M3 sites have been identified on α-DG. Thus, it seems likely that POMGNT2 demonstrates acceptor substrate preferences beyond simply an *O*-Man modified residue. We tested this hypothesis regarding specificity by examining the impact of local primary

amino acid sequence around *O*-Man sites on synthetic peptides as acceptor substrates for POMGNT1 and POMGNT2.

Employing a set of *O*-Man glycopeptide substrates, we have shown that POMGNT2 has a preference for acceptors with mannosylated residues at positions Thr317 and Thr379 while POMGNT1 has no significant acceptor substrate preferences among the various synthetic glycopeptides tested (Figure 3.2-3.4 and Table 3.1). Analysis of the sites that are POMGNT2-dependent demonstrate a R-X-R-X-X-I-X-X-T-P-T motif that is conserved among vertebrata α-DG (Figure 3.5 B). We also observed that this sequence is not found in any of the mapped sites from other *O*-mannosylated proteins (Vester-Christensen, M.B., Halim, A., et al. 2013) consistent with α-DG being the only demonstrated protein to contain M3 glycans (Figure 3. 5 A). We found that replacement of a divergent sequence on a POMGNT1 acceptor that was only missing the conserved R-X-R motif converted it to a POMGNT2 acceptor, demonstrating that replacing the two amino acids was sufficient to confer activity (Figure 3.6 A). Likewise, replacement of the arginines in the Man379 peptide with amino acids found in the M1 peptide of Man341 reduced the efficiency of POMGNT2 to catalyze the addition of GlcNAc to the *O*-Man peptide more than five-fold (Figure 3.6 B). However, while the addition of the R-X-R motif to a core M1 acceptor is sufficient to make it a substrate for POMGNT2, the complete removal of the R-X-R motif on a core M3 acceptor does not abolish POMGNT2 activity (Figure 3.6 C). Taken together, this suggests that the R-X-R motif allows for extension by POMGNT2 at core M3 sites, but is not essential for a short synthetic *O*-Man peptide. These results support a case of sufficiency in the absence of necessity which deviates from the normal necessary and sufficient or necessary but not sufficient arguments. We would rationalize that when there is sequence upstream of the site of action, as that actually found in the full-length α-DG protein, that non-basic amino acids replacing

the R-X-R portion of the motif generate steric or electrostatic clashes that prevent proper binding of the substrate protein.

Interestingly for Man317, the identified R-X-R motif is upstream of the known Furin cleavage site. However, as POMGNT2 is an ER-resident glycosyltransferase and Furin is located in the Golgi, POMGNT2 acts first and thus has the capability to interact with residues upstream of the Furin cleavage site and this may at least partially explain the requirement for the N-terminus for synthesis of functionally glycosylated mature α-DG (Kanagawa, M., Saito, F., et al. 2004). Our current model, based on the data presented here, is that POMGNT2 selectivity determines which sites on α-DG become modified with the core M3 glycan structure. In turn, only the core M3 glycan structure can be extended by B3GALNT2, phosphorylated by POMK and further elaborated to become the functional matriglycan for α-DG (Yoshida-Moriguchi, T. and Campbell, K.P. 2015, Yoshida-Moriguchi, T., Willer, T., et al. 2013). Functional glycosylation of α-DG, and, in particular, matriglycan synthesis stemming from the POMGNT2-dependent core M3 glycan structure is required for binding to extracellular matrix (ECM) proteins with laminin globular domains and maintaining overall ECM integrity (Praissman, J.L. and Wells, L. 2014). Thus, POMGNT2 acts a *gatekeeper enzyme* for functional glycosylation of α-DG.

The strict R-X-R-X-X-I-X-X-T-P-T motif that we have presented here is not present on any other secreted or membrane associated protein in humans except for α-DG at T317/319 and T379/381 (Yoshida-Moriguchi, T. and Campbell, K.P. 2015, Yoshida-Moriguchi, T., Willer, T., et al. 2013). Relaxing the sequence constraints to allow for conservative replacements generates a motif of R/K-X-R/K-X-X-I/L/V-X-X-T/S-P-T/S. This motif is found on a handful of membrane/secreted human proteins including SRPX, CLEC18C, FREM2, MANBA, SEMA3E, SPACA7, and TMEM182. However, if we examine conservation of the motif in these proteins

across 59 vertebrate species, as we did for α-DG, we see poor conservation (data not shown). This lends further support to the working model that only α-DG contains sequences that are substrates for POMGNT2 that go on to become functionally glycosylated with matriglycan (Yoshida-Moriguchi, T. and Campbell, K.P. 2015).

We have identified and partially characterized a primary amino acid sequence motif governing acceptor specificity for POMGNT2 towards *O*-mannosylated substrates. Additional studies are required to fully characterize the functional roles of individual amino acids in this motif. Structural analyses of POMGNT2 in complex with various acceptor substrates would greatly assist in defining the molecular details of the POMGNT2 *gatekeeping* mechanism that we have established here. Furthermore, future *in vivo* studies testing the role of the R-X-R-X-X-I-X-X-T-P-T motif in POMGNT2 acceptor selectivity will be invaluable to complement our *in vitro* findings presented here.

Experimental Procedures

Cell Culture and Protein Purification. The catalytic domains of human POMGNT1 (amino acid residues 60–660, UniProt Q8WZA1) and POMGNT2 (amino acid residues 25–580, UniProt Q8NAT1) were expressed as soluble, secreted fusion proteins by transient transfection of HEK293 suspension cultures (Meng, L., Forouhar, F., et al. 2013). The coding regions were amplified from Mammalian Gene Collection (Gerhard, D.S., Wagner, L., et al. 2004) clones using primers that appended a tobacco etch virus (TEV) protease cleavage site (Phan, J., Zdanov, A., et al. 2002) to the NH_2-terminal end of the coding region and attL1 and attL2 Gateway adaptor sites to the 5′ and 3′ terminal ends of the amplimer products. The amplimers were recombined via BP clonase reaction into the pDONR221 vector and the DNA sequences were confirmed. The pDONR221 clones were then recombined via LR clonase reaction into a custom Gateway adapted version of

the pGEn2 mammalian expression vector (Barb, A.W., Meng, L., et al. 2012, Meng, L., Forouhar, F., et al. 2013) to assemble a recombinant coding region comprised of a 25 amino acid NH_2-terminal signal sequence from the *T. cruzi* lysosomal α-mannosidase (Vandersall-Nairn, A.S., Merkle, R.K., et al. 1998) followed by an 8xHis tag, 17 amino acid AviTag (Beckett, D., Kovaleva, E., et al. 1999), 'superfolder' GFP (Pedelacq, J.D., Cabantous, S., et al. 2006), the nine amino acid sequence encoded by attB1 recombination site, followed by the TEV protease cleavage site and the respective glycosyltransferase catalytic domain coding region. Suspension culture HEK293f cells (Life Technologies, Grand Island, NY) were transfected as previously described (Meng, L., Forouhar, F., et al. 2013) and the culture supernatant was subjected to Ni-NTA superflow chromatography (Qiagen, Valencia, CA). Enzyme preparations eluted with 300 mM imidazole were concentrated to ~1 mg/ml using an ultrafiltration pressure cell membrane (Millipore, Billerica, MA) with a 10 kDa molecular weight cutoff.

Glycopeptide Synthesis. The glycopeptide synthesis here extends earlier work describing synthesis of *O*-Man-Ser and -Thr peptide synthesis building blocks, as well as *O*-Man glycopeptides (Liu, M., Borgert, A., et al. 2008). The glycopeptides were prepared as C-terminal carboxamides and acetylated at the N-terminus to emulate the situation in the native protein. For this work all couplings except those for glycosylated residues were carried out on an automated microwave-assisted solid-phase peptide synthesizer (Liberty CEM Microwave Synthesizer) using standard protocols in the instrument software, on Rink amide resin (~0.5 meq/gm, Novabiochem) via an N_R-Fmoc-based approach with DMF as the primary solvent. 20% 4-methyl piperidine in DMF was used for Fmoc removal. 2-(1H-benzotriazole-1-yl)-oxy-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU)/1-hydroxybenzotriazole (HOBt) in the presence of N,N-diisopropylethylamine (DIPEA) were employed as the coupling reagents for standard amino acids.

For the coupling of the glycosylated amino acid, Fmoc-Thr(α-D-Man(Ac)$_4$)-OH (Sussex Research), the peptide resin was removed from the synthesizer and coupling performed manually using a CEM Discover microwave apparatus. 2-(7-aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU)/ 1-hydroxy-7-azabenzotriazole (HOAt) in the presence of DIPEA were the activating reagents. Typically two couplings at ~1.5 fold excess of glycosylated amino acid to the resin loading were done for this amino acid to conserve reagent. Upon completion of the manual coupling reaction, as determined by Matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS), glycopeptide resins were returned to the automatic synthesizer to complete assembly. After final N-deprotection, the glycopeptides were manually N-acetylated by treatment with DMF/Acetic Anhydride/DIPEA 85/10/5 v/v for ~30 min, and the O-acetyl protection on the mannosyl residues were subsequently removed by two successive treatments with Hydrazine/MeOH 70/20 v/v for an hour each. Glycopeptides were then cleaved from the resin as C-terminal carboxamides, with simultaneous removal of remaining amino acid side chain protection through treatment with TFA/TIPS/H_2O 95/2.5/2.5 for ~4 hrs. The resin was filtered off and the TFA solution concentrated on a rotary evaporator to a few mL. The remaining concentrate was added drop wise to cold ether from which the crude glycopeptides precipitated. After centrifugation and removal of the ether supernatant, the glycopeptides were redissolved and purified via HPLC over an Ultra II 250x10.0mm 5µm C_{18} column (RESTEK) with a 0.1% TFA in water/0.1% TFA in Acetonitrile solvent gradient. Purity was verified by analytical HPLC and MALDI-TOF MS (see Supplementary Material). Yields were in the range of 30 -50%.

Radiolabel Transfer Assays. The radiometric assays were carried out in reactions containing 100 mM MES (pH 6.5), 10 mM $MnCl_2$, 2 mM UDP-GlcNAc mixed with 10nCi ^{3}H-

UDP-GlcNAc, and 1mM glycopeptide acceptor. Reactions were incubated for 21 hrs at 37°C, then quenched by addition of 5 µl 1% TFA and boiled at 100°C for 5 min. Reaction products were purified by reverse phase separation using C_{18} SepPak micro spin columns (The Nest Group) by loading and washing with 0.1% formic acid and elution with 80% acetonitrile with 0.1% formic acid. Disintegrations per minute (DPMs) were counted using a liquid scintillation counter (Beckman) to determine the amount of ^{3}H-GlcNAc incorporated into the glycopeptides. The data presented represent the average of at least 3 independent experiments.

Mass Spectrometry. Cold glycosyltransferase reactions used for analysis by mass spectrometry were carried out identical to the radioactive transfer assays but without radioactive UDP-GlcNAc. After reverse phase separation, the product was vacuumed to dryness and resuspended in 100 µl of 0.1% formic acid. Samples were filtered using a 0.2 µm nanosep microcentrifuge filter (Pall Life Sciences) and transferred to an autosampler vial with glass insert (Thermo Scientific). The samples were run on a Thermo Scientific™ Orbitrap Fusion™ Lumos™ mass spectrometer. Full Fourier transform MS Spectra were analyzed using Xcalibur Qual Browser software, and MS/MS scans were analyzed using Byonic™ Version 2.6.46 (Protein Metrics Inc.), using a precursor mass tolerance of 10 ppm and a fragmentation mass tolerance of 0.3 Daltons followed by manual interpretation.

UDP-Glo™ Glycosyltransferase Assays. UDP-Glo™ Glycosyltransferase Assays (Promega) were performed using 50 mM Tris-HCl (pH 7.5), 5 mM $MnCl_2$, 100 µM UDP-GlcNAc, 40 ng of enzyme, and varying amounts of glycopeptide acceptor-substrates at 37°C for 2 hrs in a white, flat bottom, 384-well plate. After the glycosyltransferase reaction, an equal volume of UDP Detection Reagent was added to simultaneously convert the UDP product to ATP and generate light in a luciferase reaction. The light generated was detected using a GloMax-Multi+

luminometer (Promega). Luminescence was correlated to UDP concentration by using a UDP standard curve. Kinetic parameters were extracted from the data after fitting to the Michaelis-Menten equation using the non-linear regression fit in GraphPad Prism Version 7.1. The data presented represent the average of at least 3 independent experiments.

WebLogo Consensus Sequence Alignment. All vertebrate species available on the Ensembl genome browser (release 85) (Yates, A., Akanni, W., et al. 2016) with orthologues of human DAG1, POMGNT1 or POMGNT2, and FKTN or B4GAT1 (*X. tropicalis, L. chalumnae, A. carolinensis, C. hoffmanni, T. truncatus, P. sinensis, M. gallopavo, G. gallus, A. platyrhynchos, T. guttata, F. albicollis, D. ordii, E. telfairi, O. princeps, L. africana, P. capensis, M. lucifugus, C. porcellus, R. norvegicus, M. musculus, E. europaeus, D. novemcinctus, O. garnettii, I. tridecemlineatus, B. taurus, O. aries, O. cuniculus, M. putorius furo, C. lupus familiaris, F. catus, A. melanoleuca, T. syrichta, T. belangeri, P. vampyrus, S. scrofa, E. caballus, C. jacchus, P. anubis, M. mulatta, C. sabaeus, N. leucogenys, G. gorilla gorilla, P. troglodytes, H. sapiens, O. anatinus, M. eugenii, M. domestica, S. harrisii, L. oculatus, D. rerio, A. mexicanus, T. nigroviridis, T. rubripes, O. latipes, X. maculatus, P. formosa, G. morhua, G. aculeatus,* and *O. niloticus*) were aligned to human DAG1 using Clustal Omega (Sievers, F., Wilm, A., et al. 2011). The ten amino acids upstream and downstream of the Threonine at position 317 and 379 in human DAG1 for all species were extracted from the alignment and used for analysis in Berkeley's WebLogo program (version 3) (Crooks, G.E., Hon, G., et al. 2004).

Acknowledgements

The authors would like to thank M. Osman Sheikh for assistance in the preparation of this manuscript, and Robert Bridger and Jeremy L. Praissman for technical assistance with mass spectrometry. The authors would like to thank all members of the Wells lab for thoughtful

discussions. This work was supported in part by grants from the National Institutes of Health (R01GM111939, DL and LW; P01GM107012, GJB, KM, LW; P41GM103490, KM and LW; P41GM103390, KM and DL; S10RR027155 for purchase of the peptide synthesizer, DL). SH is supported in part by a T32 NIH training grant (T32GM107004). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Barb AW, Meng L, Gao Z, Johnson RW, Moremen KW, Prestegard JH. 2012. NMR characterization of immunoglobulin G Fc glycan motion on enzymatic sialylation. *Biochemistry-Us*, 51:4618-4626.

Beckett D, Kovaleva E, Schatz PJ. 1999. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci*, 8:921-929.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome research*, 14:1188-1190.

Ervasti JM, Campbell KP. 1993. A role for the dystrophin-glycoprotein complex as a transmembrane linker between laminin and actin. *The Journal of Cell Biology*, 122:809-823.

Gerhard DS, Wagner L, Feingold EA, Shenmen CM, Grouse LH, Schuler G, Klein SL, Old S, Rasooly R, Good P, *et al.* 2004. The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res*, 14:2121-2127.

Godfrey C, Foley AR, Clement E, Muntoni F. 2011. Dystroglycanopathies: coming into focus. *Current Opinion in Genetics & Development*, 21:278-285.

Hara Y, Kanagawa M, Kunz S, Yoshida-Moriguchi T, Satz JS, Kobayashi YM, Zhu Z, Burden SJ, Oldstone MBA, Campbell KP. 2011. Like-acetylglucosaminyltransferase (LARGE)-dependent modification of dystroglycan at Thr-317/319 is required for laminin binding and arenavirus infection. *Proceedings of the National Academy of Sciences*, 108:17426-17431.

Harrison R, Hitchen PG, Panico M, Morris HR, Mekhaiel D, Pleass RJ, Dell A, Hewitt JE, Haslam SM. 2012. Glycoproteomic characterization of recombinant mouse alpha-dystroglycan. *Glycobiology*, 22:662-675.

Jae LT, Raaben M, Riemersma M, van Beusekom E, Blomen VA, Velds A, Kerkhoven RM, Carette JE, Topaloglu H, Meinecke P, *et al.* 2013. Deciphering the Glycosylome of Dystroglycanopathies Using Haploid Screens for Lassa Virus Entry. *Science*, 340:479-483.

Kanagawa M, Kobayashi K, Tajiri M, Manya H, Kuga A, Yamaguchi Y, Akasaka-Manya K, Furukawa J-i, Mizuno M, Kawakami H, *et al.* 2016. Identification of a Post-translational Modification with Ribitol-Phosphate and Its Defect in Muscular Dystrophy. *Cell Reports*, 14:2209–2223.

Kanagawa M, Saito F, Kunz S, Yoshida-Moriguchi T, Barresi R, Kobayashi YM, Muschler J, Dumanski JP, Michele DE, Oldstone MBA, *et al.* 2004. Molecular Recognition by LARGE Is Essential for Expression of Functional Dystroglycan. *Cell*, 117:953-964.

Lee JK, Matthews RT, Lim JM, Swanier K, Wells L, Pierce JM. 2012. Developmental expression of the neuron-specific N-acetylglucosaminyltransferase Vb (GnT-Vb/IX) and identification of its in vivo glycan products in comparison with those of its paralog, GnT-V. *J Biol Chem*, 287:28526-28536.

Liu M, Borgert A, Barany G, Live D. 2008. Conformational consequences of protein glycosylation: preparation of O-mannosyl serine and threonine building blocks, and their incorporation into glycopeptide sequences derived from alpha-dystroglycan. *Biopolymers*, 90:358-368.

Live D, Wells L, Boons G-J. 2013. Dissecting the Molecular Basis of the Role of the O-Mannosylation Pathway in Disease: α-Dystroglycan and Forms of Muscular Dystrophy. *ChemBioChem*, 14:2392-2402.

Manya H, Chiba A, Yoshida A, Wang X, Chiba Y, Jigami Y, Margolis RU, Endo T. 2004. Demonstration of mammalian protein O-mannosyltransferase activity: Coexpression of POMT1 and POMT2 required for enzymatic activity. *Proceedings of the National Academy of Sciences of the United States of America*, 101:500-505.

Manya H, Suzuki T, Akasaka-Manya K, Ishida H-K, Mizuno M, Suzuki Y, Inazu T, Dohmae N, Endo T. 2007. Regulation of Mammalian Protein O-Mannosylation: PREFERENTIAL AMINO ACID SEQUENCE FOR O-MANNOSE MODIFICATION. *Journal of Biological Chemistry*, 282:20200-20206.

Meng L, Forouhar F, Thieker D, Gao Z, Ramiah A, Moniz H, Xiang Y, Seetharaman J, Milaninia S, Su M, *et al.* 2013. Enzymatic basis for N-glycan sialylation: structure of rat alpha2,6-sialyltransferase (ST6GAL1) reveals conserved and unique features for glycan sialylation. *J Biol Chem*, 288:34680-34698.

Mo K-F, Fang T, Stalnaker SH, Kirby PS, Liu M, Wells L, Pierce M, Live DH, Boons G-J. 2011. Synthetic, Structural, and Biosynthetic Studies of an Unusual Phospho-Glycopeptide Derived from α-Dystroglycan. *Journal of the American Chemical Society*, 133:14418-14430.

Nilsson J, Nilsson J, Larson G, Grahn A. 2010. Characterization of site-specific O-glycan structures within the mucin-like domain of α-dystroglycan from human skeletal muscle. *Glycobiology*, 20:1160-1169.

Ogawa M, Nakamura N, Nakayama Y, Kurosaka A, Manya H, Kanagawa M, Endo T, Furukawa K, Okajima T. 2013. GTDC2 modifies O-mannosylated α-dystroglycan in the endoplasmic reticulum to generate N-acetyl glucosamine epitopes reactive with CTD110.6 antibody. *Biochemical and Biophysical Research Communications*, 440:88-93.

Pedelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS. 2006. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol*, 24:79-88.

Phan J, Zdanov A, Evdokimov AG, Tropea JE, Peters HK, 3rd, Kapust RB, Li M, Wlodawer A, Waugh DS. 2002. Structural basis for the substrate specificity of tobacco etch virus protease. *The Journal of biological chemistry*, 277:50564-50572.

Praissman JL, Live DH, Wang S, Ramiah A, Chinoy ZS, Boons G-J, Moremen KW, Wells L. 2014. B4GAT1 is the priming enzyme for the LARGE-dependent functional glycosylation of α-dystroglycan. *eLife*, 3.

Praissman JL, Wells L. 2014. Mammalian O-Mannosylation Pathway: Glycan Structures, Enzymes, and Protein Substrates. *Biochemistry*, 53:3066-3078.

Praissman JL, Willer T, Sheikh MO, Toi A, Chitayat D, Lin YY, Lee H, Stalnaker SH, Wang S, Prabhakar PK, *et al.* 2016. The functional O-mannose glycan on alpha-dystroglycan contains a phospho-ribitol primed for matriglycan addition. *Elife*, 5.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, *et al.* 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7.

Stalnaker SH, Hashmi S, Lim J-M, Aoki K, Porterfield M, Gutierrez-Sanchez G, Wheeler J, Ervasti JM, Bergmann C, Tiemeyer M, *et al.* 2010. Site Mapping and Characterization of O-Glycan Structures on α-Dystroglycan Isolated from Rabbit Skeletal Muscle. *The Journal of Biological Chemistry*, 285:24882-24891.

Takahashi S, Sasaki T, Manya H, Chiba Y, Yoshida A, Mizuno M, Ishida H-K, Ito F, Inazu T, Kotani N, *et al.* 2001. A new β-1,2-N-acetylglucosaminyltransferase that may play a role in the biosynthesis of mammalian O-mannosyl glycans. *Glycobiology*, 11:37-45.

Vandersall-Nairn AS, Merkle RK, O'Brien K, Oeltmann TN, Moremen KW. 1998. Cloning, expression, purification, and characterization of the acid a-mannosidase from *Trypanosoma cruzi*. *Glycobiology*, 8:1183-1194.

Varki A, Cummings RD, Aebi M, Packer NH, Seeberger PH, Esko JD, Stanley P, Hart G, Darvill A, Kinoshita T, *et al.* 2015. Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology*, 25:1323-1324.

Vester-Christensen MB, Halim A, Joshi HJ, Steentoft C, Bennett EP, Levery SB, Vakhrushev SY, Clausen H. 2013. Mining the O-mannose glycoproteome reveals cadherins as major O-mannosylated glycoproteins. *Proceedings of the National Academy of Sciences*, 110:21018-21023.

Voglmeir J, Kaloo S, Laurent N, Meloni Marco M, Bohlmann L, Wilson Iain BH, Flitsch Sabine L. 2011. Biochemical correlation of activity of the α-dystroglycan-modifying glycosyltransferase POMGnT1 with mutations in muscle-eye-brain disease. *Biochemical Journal*, 436:447-455.

Wells L. 2013. The O-Mannosylation Pathway: Glycosyltransferases and Proteins Implicated in Congenital Muscular Dystrophy. *The Journal of Biological Chemistry*, 288:6930-6935.

Yagi H, Nakagawa N, Saito T, Kiyonari H, Abe T, Toda T, Wu S-W, Khoo K-H, Oka S, Kato K. 2013. AGO61-dependent GlcNAc modification primes the formation of functional glycans on α-dystroglycan. *Scientific Reports*, 3:3288.

Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, Cummins C, Clapham P, Fitzgerald S, Gil L, *et al.* 2016. Ensembl 2016. *Nucleic Acids Research*, 44:D710-D716.

Yoshida-Moriguchi T, Campbell KP. 2015. Matriglycan: a novel polysaccharide that links dystroglycan to the basement membrane. *Glycobiology*, 25:702-713.

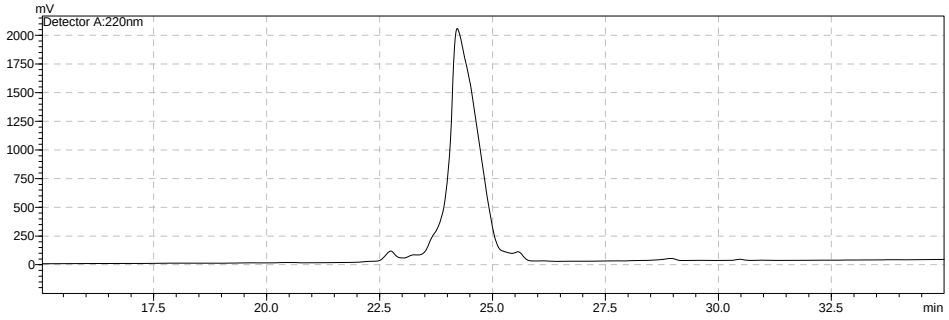
Yoshida-Moriguchi T, Willer T, Anderson ME, Venzke D, Whyte T, Muntoni F, Lee H, Nelson SF, Yu L, Campbell KP. 2013. SGK196 Is a Glycosylation-Specific O-Mannose Kinase Required for Dystroglycan Function. *Science*, 341:896-899.

Yoshida-Moriguchi T, Yu L, Stalnaker SH, Davis S, Kunz S, Madson M, Oldstone MB, Schachter H, Wells L, Campbell KP. 2010. O-mannosyl phosphorylation of alpha-dystroglycan is required for laminin binding. *Science*, 327:88-92.

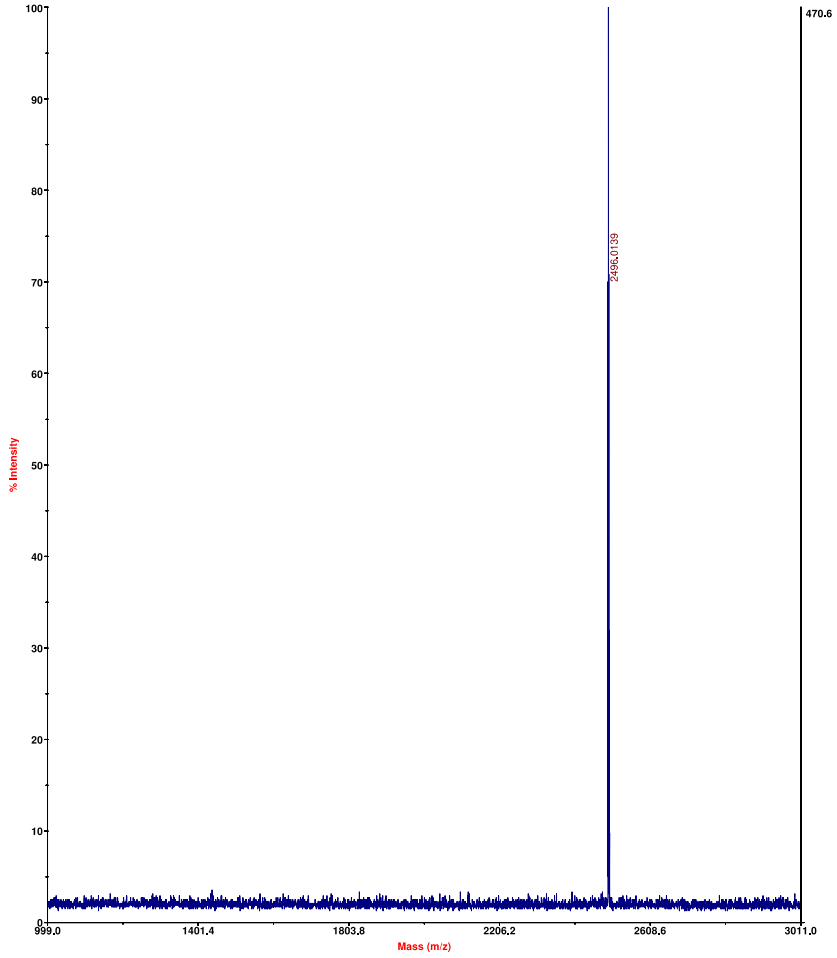
Supplemental Material Figure S3.1. (**A**) HPLC chromatogram (**B**) and MALDI-TOF spectra of purified glycopeptides Man 317, Man 379, Man 379-ETP, ShortMan379, Man341, Man414, Man341- RPR.

Man317

A.

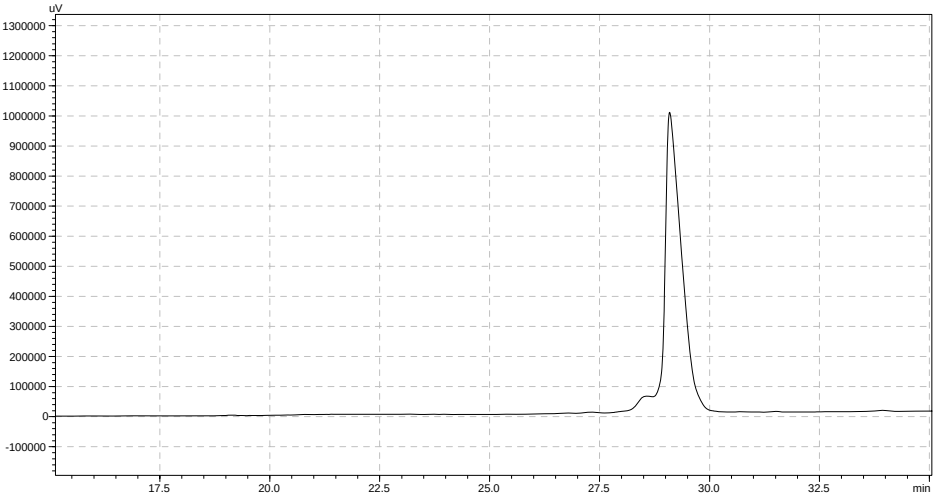


B.

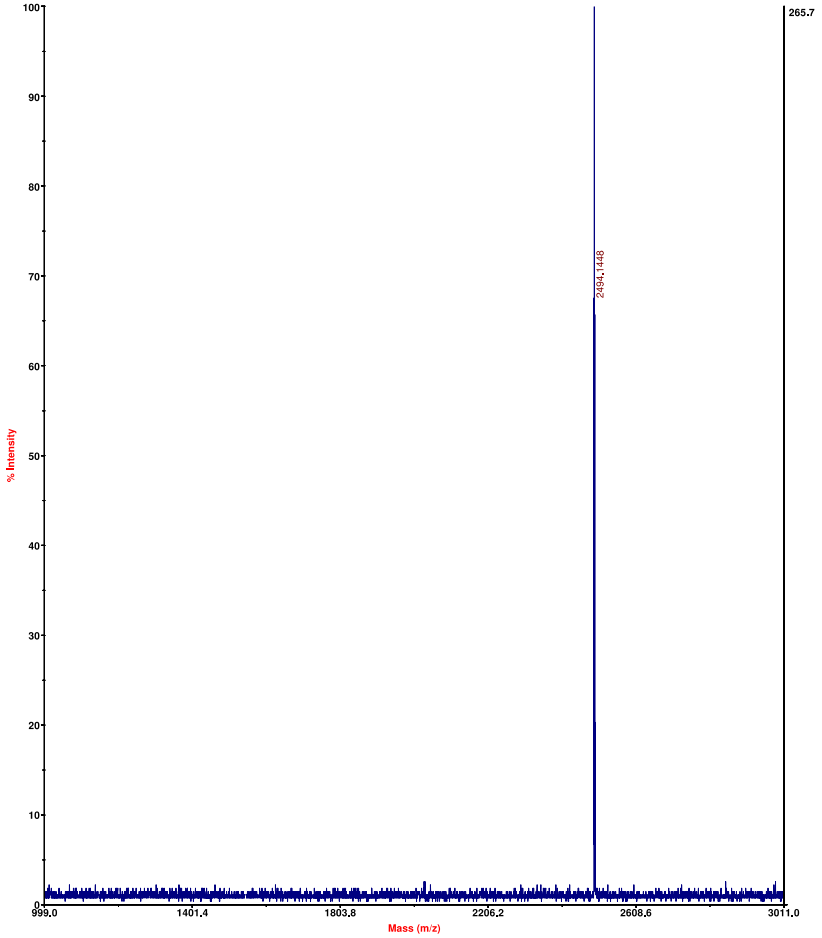


Man379

A.

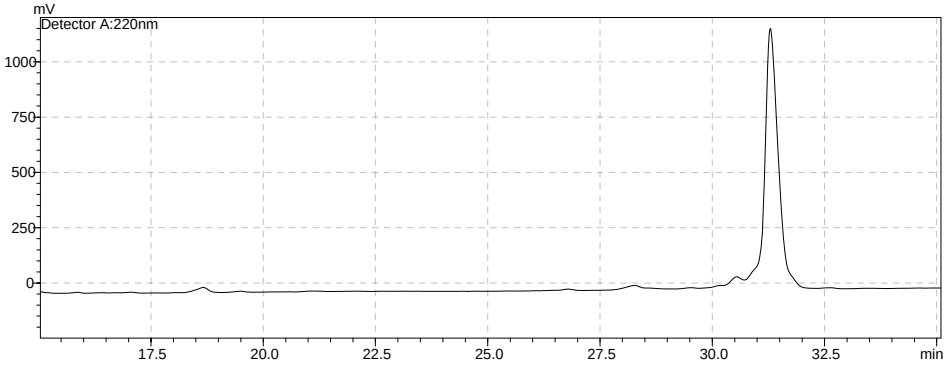

uV
1300000
1200000
1100000
1000000
900000
800000
700000
600000
500000
400000
300000
200000
100000
0
-100000
17.5
20.0
22.5
25.0
27.5
30.0
32.5
min

B.

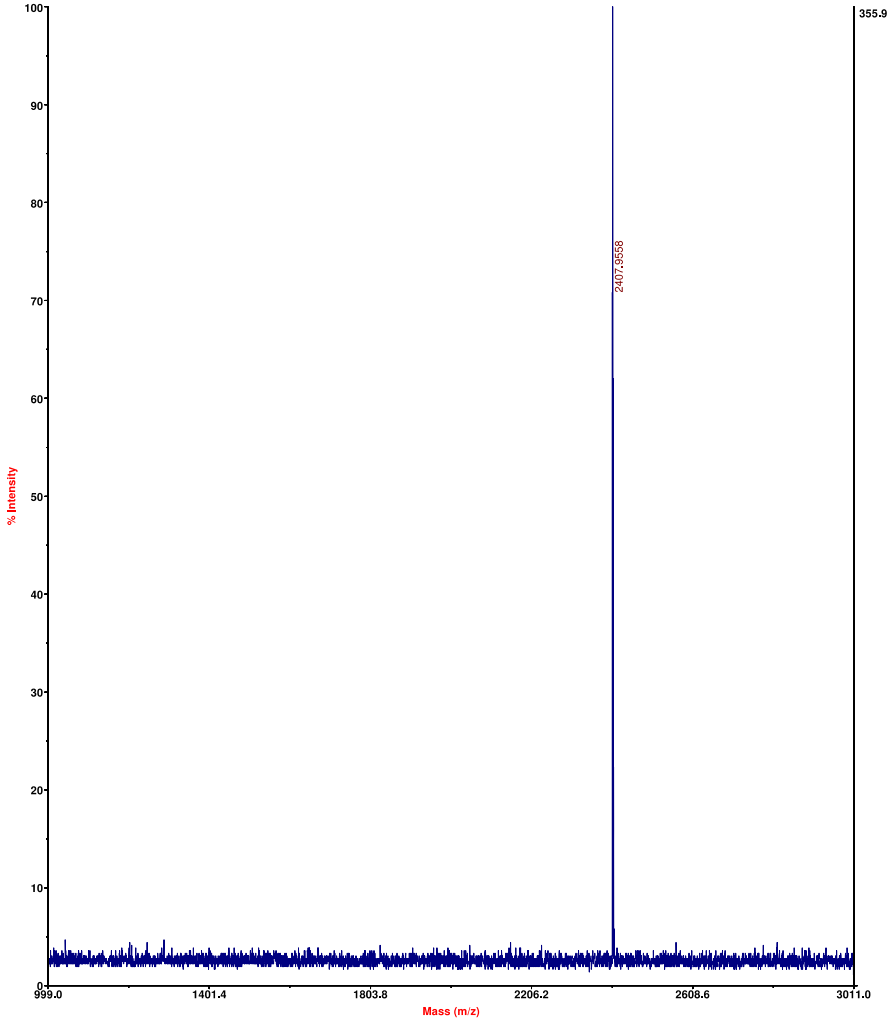

100
90
80
70
60
50
40
30
20
10
0
% Intensity
2494.1448
265.7
999.0
1401.4
1803.8
2206.2
2608.6
3011.0
Mass (m/z)

Man379-ETP

A.

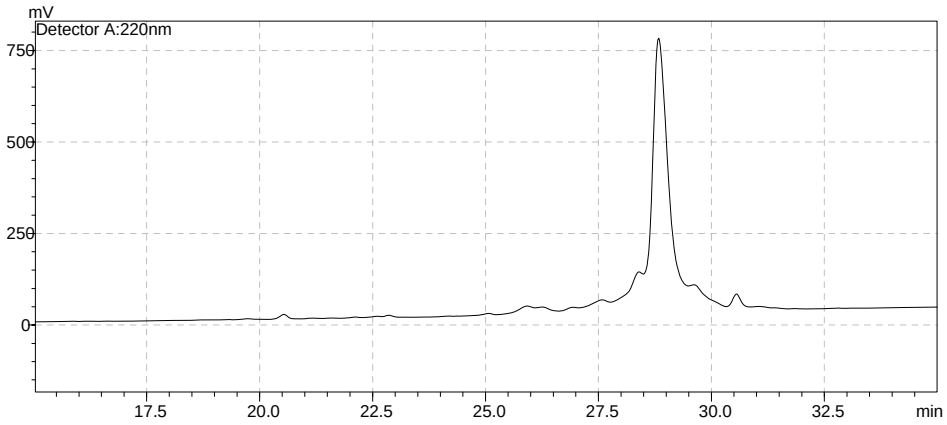


B.

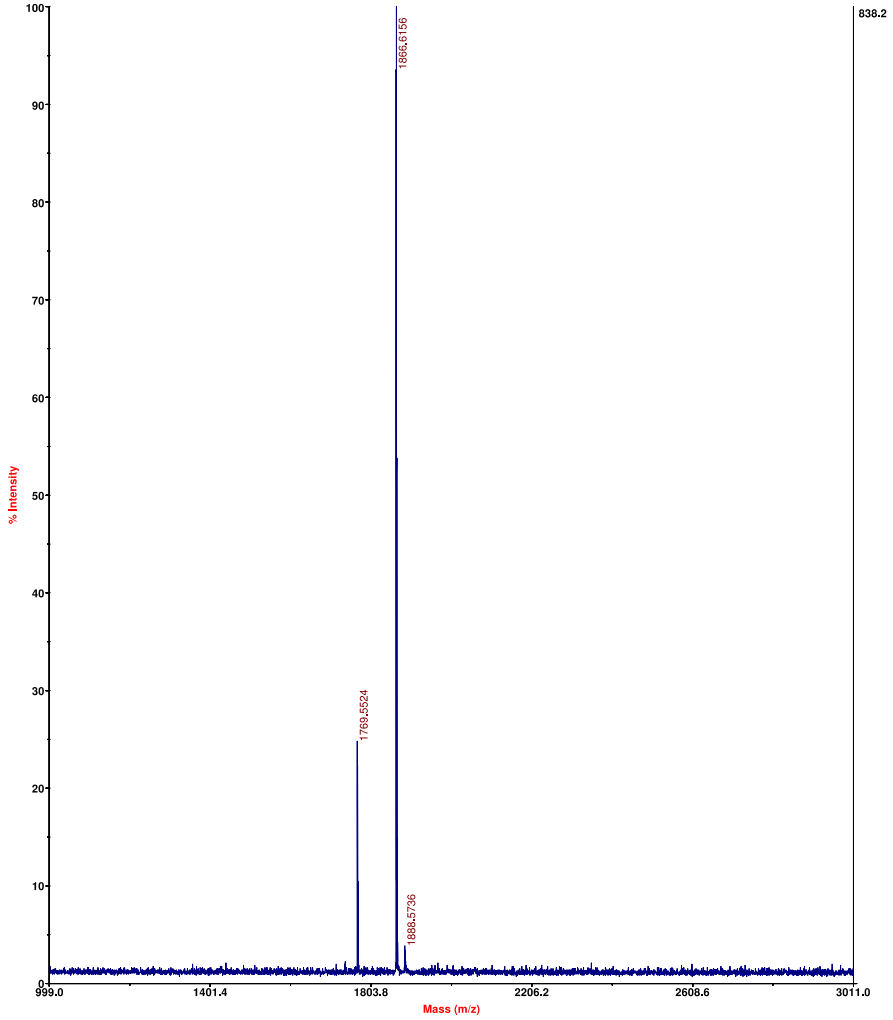


Short Man379

A.

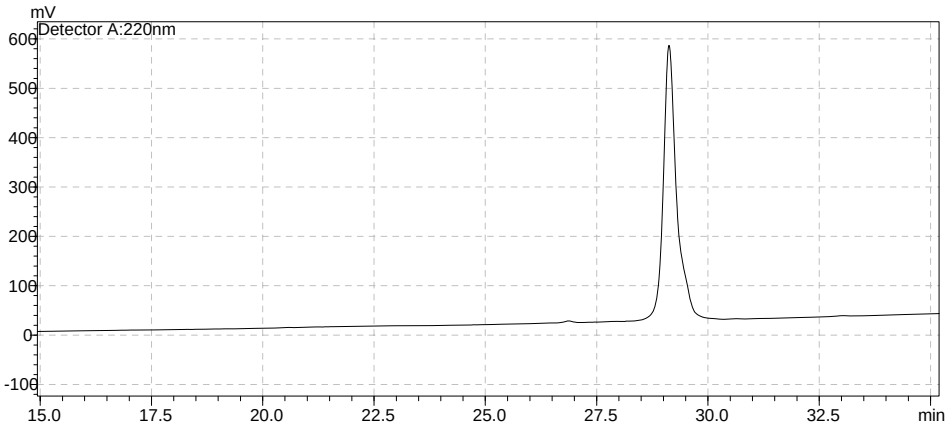


B.

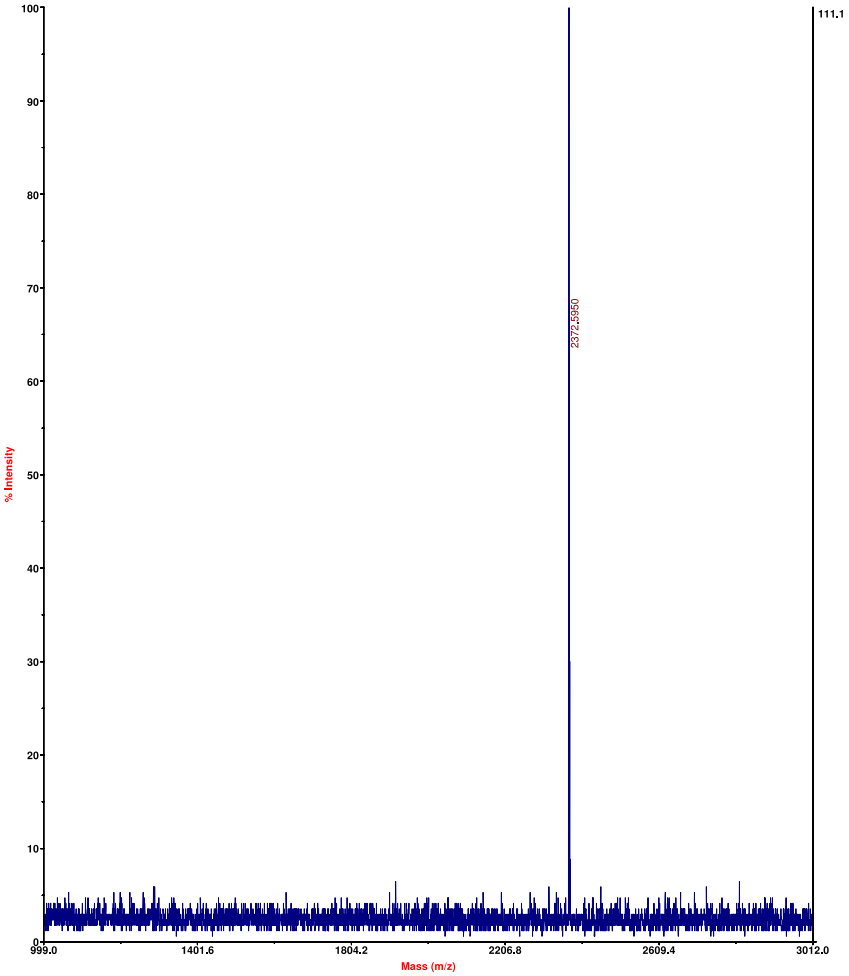


Man341

A.

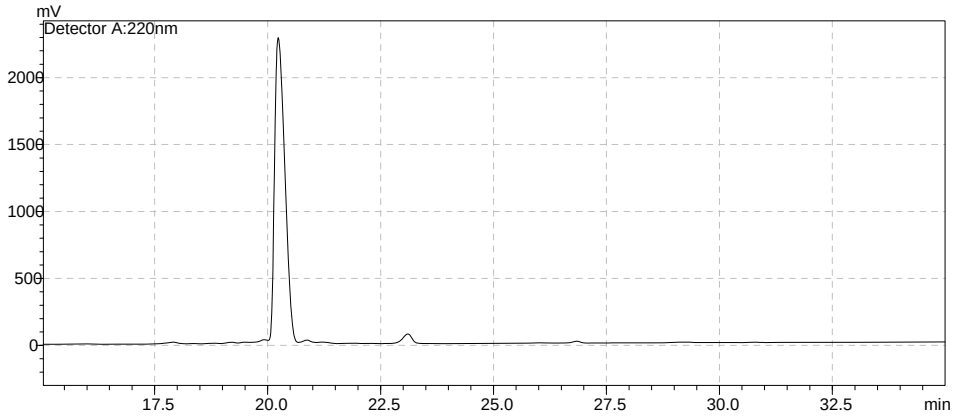

mV
Detector A:220nm
600
500
400
300
200
100
0
-100
15.0
17.5
20.0
22.5
25.0
27.5
30.0
32.5
min

B.

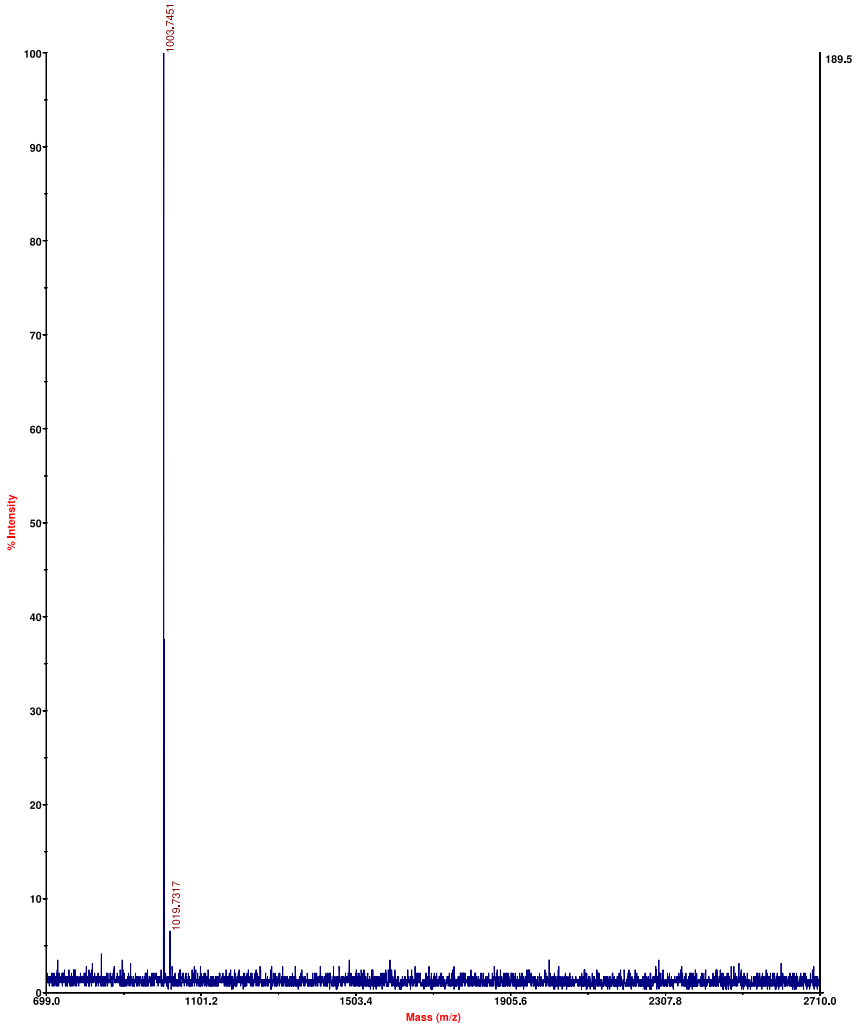

100
90
80
70
60
50
40
30
20
10
0
% Intensity
2372.5950
111.1
999.0
1401.6
1804.2
2206.8
2609.4
3012.0
Mass (m/z)

Man414

A.

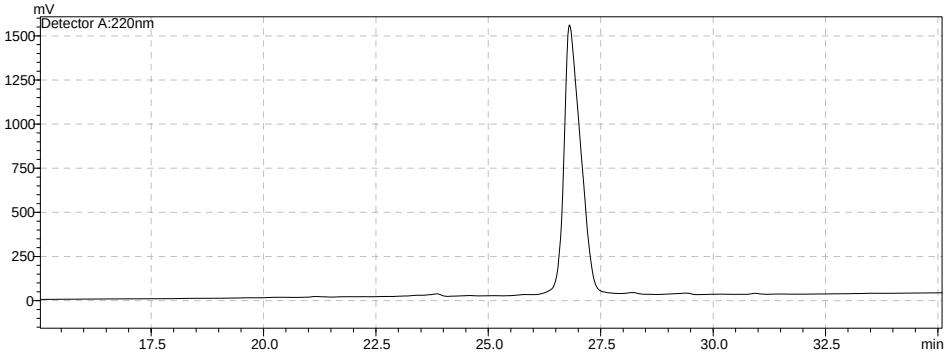

mV
Detector A:220nm
2000
1500
1000
500
0
17.5
20.0
22.5
25.0
27.5
30.0
32.5
min

B.

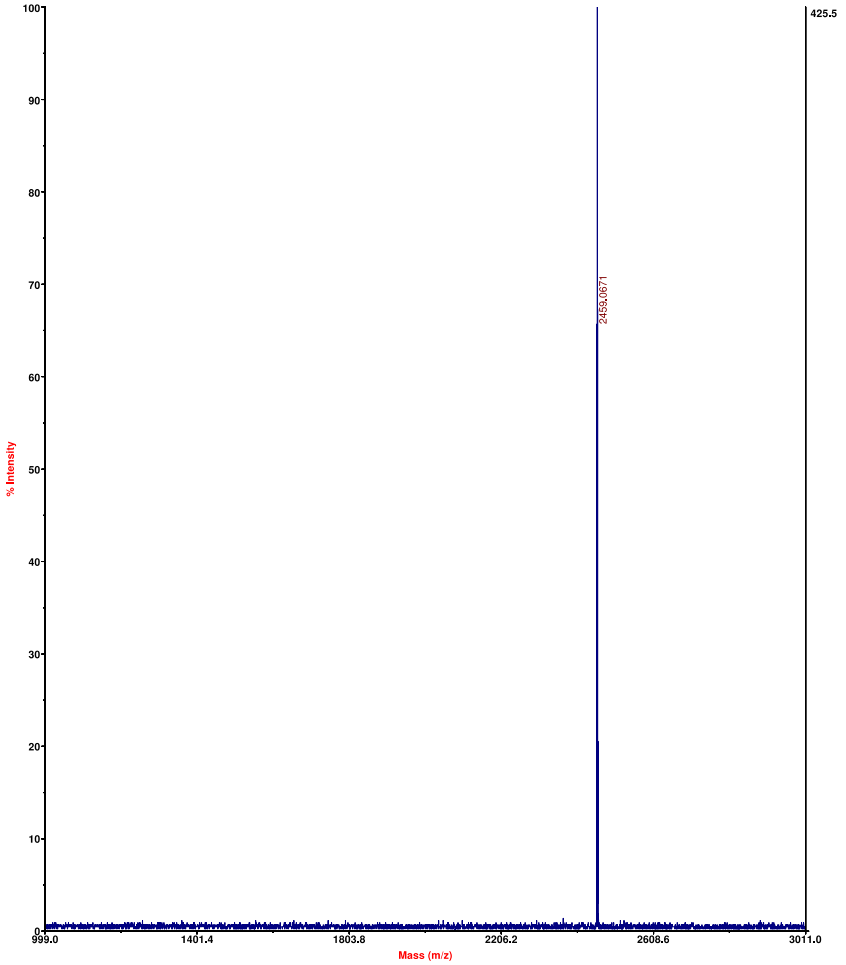

1003.7451
1019.7317
189.5
% Intensity
100
90
80
70
60
50
40
30
20
10
0
699.0
1101.2
1503.4
1905.6
2307.8
2710.0
Mass (m/z)

Man341-RPR

A.



B.



CHAPTER 4

POMGNT2 SUBSTRATE SELECTIVITY

Introduction

Dystroglycan is a transmembrane protein, encoded by a single gene (*DAG1*), and is cleaved into two subunits, α-dystroglycan (α-DG) and β-dystroglycan, via posttranslational processing. β-dystroglycan serves as the transmembrane subunit and binds dystrophin within the cell, which in turn, binds to the actin cytoskeleton. α-DG, the extracellular subunit, is anchored to the plasma membrane via its interaction with β-dystroglycan. α-DG consists of three domains: an N-terminal domain, a mucin-like domain and a C-terminal domain. α-DG is highly glycosylated in its mucin-like domain by *O*-GalNAc and *O*-mannose initiated glycans. The glycans on α-DG, specifically the glycosaminoglycan (GAG)-like polymer known as matriglycan, binds to laminin-globular domain-containing proteins in the extracellular matrix, such as laminin, agrin, neurexin, and perlecan (Hohenester, E. 2019, Yoshida-Moriguchi, T. and Campbell, K.P. 2015). Matriglycan, the functional glycan on α-DG, is a sugar chain synthesized by the bifunctional glycosyltransferase and Golgi-resident enzyme LARGE1 and consists of repeating disaccharide units of xylose and glucuronic acid [(-GlcA-β3-Xyl-α3-)$_n$] (Yoshida-Moriguchi, T. and Campbell, K.P. 2015). The laminin-G-like domain 4 of laminin-α2 interacts directly with a single disaccharide repeat of matriglycan and this binding is chelated by a calcium ion (Briggs, D.C., Yoshida-Moriguchi, T., et al. 2016). In essence, matriglycan on α-DG serves as a physical link between the extracellular matrix and the inside of the cell. Proper glycosylation of α-DG is essential for the formation and function of the dystrophin-glycoprotein complex at the cell surface and basement membrane

integrity in skeletal and brain tissues (Henry, M.D. and Campbell, K.P. 1998, Michele, D.E., Barresi, R., et al. 2002). Perturbation of α-DG glycosylation results in disease subtypes of congenital muscular dystrophy known as dystroglycanopathies, which are characterized by hypoglycosylated forms of α-DG (Godfrey, C., Foley, A.R., et al. 2011, Wells, L. 2013).

LARGE, the enzyme responsible for building matriglycan on α-DG, binds to the N-terminal domain of α-DG during α-DG's posttranslational maturation. Disruption of this interaction leads to loss of functional α-DG glycosylation (Kanagawa, M., Saito, F., et al. 2004), which is in line with the fact that *Large*^{myd} mice that harbor a mutation in the *LARGE* gene and patients with mutations in the *LARGE* gene have α-DG with altered laminin-binding capabilities (Durbeej, M. and Campbell, K.P. 2002, Grewal, P.K., Holzfeind, P.J., et al. 2001). Ultimately, the N-terminal domain of α-DG is cleaved by furin protease and released from the mature form of α-DG.

While the specificity of LARGE may be partially determined by its ability to interact with the N-terminal domain of α-DG, further specificity is gained by the unique glycan structure that is required as an acceptor for initial LARGE activity. The matriglycan polymer is linked to α-DG via a highly unusual *O*-mannose initiated heptasaccharide linker. The complete chemical structure of this unique glycan was deciphered recently (Sheikh, M.O., Halmo, S.M., et al. 2017). The linker between α-DG and matriglycan begins with the core M3 phosphotrisaccharide (Yoshida-Moriguchi, T., Yu, L., et al. 2010). The core M3 phosphotrisaccharide is then further extended by a novel tandem phospho-ribitol component and primed for LARGE extension by the addition of a xylose and glucuronic acid (Gerin, I., Ury, B., et al. 2016, Kanagawa, M., Kobayashi, K., et al. 2016, Praissman, J.L., Live, D.H., et al. 2014, Praissman, J.L., Willer, T., et al. 2016). A suite of enzymes is responsible for building this unique heptasaccharide glycan structure, including

POMT1/POMT2, POMGNT2, B3GALNT2, POMK, FKTN, FKRP, RXYLT1, and B4GAT1. Adding to the specificity of this glycan structure, only a few known sites on α-DG are modified with the functional glycan. Threonines at sites 317 and 319 in α-DG are demonstrated sites of functional glycosylation (Hara, Y., Kanagawa, M., et al. 2011, Yagi, H., Nakagawa, N., et al. 2013), and threonine 379 has been shown to possess the core M3 phosphotrisaccharide as well (Yoshida-Moriguchi, T., Yu, L., et al. 2010).

Given the the required coordination of several enzymes necessary to build the functional glycan on α-DG, and only a few known sites on α-DG that serve as scaffolds for its synthesis, it is hypothesized that specificity for the final glycan structure exists at every step in the pathway. Several threonine and serine residues in α-DG are *O*-mannosylated by the enzyme complex POMT1/POMT2 (Nilsson, J., Nilsson, J., et al. 2010, Stalnaker, S.H., Hashmi, S., et al. 2010, Vester-Christensen, M.B., Halim, A., et al. 2013). However, only a handful of these *O*-mannose sites are extended by the enzyme POMGNT2 (Hara, Y., Kanagawa, M., et al. 2011, Yagi, H., Nakagawa, N., et al. 2013, Yoshida-Moriguchi, T., Yu, L., et al. 2010). Previous work has shown that POMGNT2 demonstrates substrate selectivity *in vitro* for the primary amino acid sequence of α-DG beyond simply an *O*-mannose modified threonine residue (Halmo, S.M., Singh, D., et al. 2017). Additionally, a POMGNT2 acceptor motif in α-DG, conserved among 59 vertebrate species, has been proposed (Halmo, S.M., Singh, D., et al. 2017). Here, we attempt to probe the necessity and sufficiency of the R-X-R-X-X-(L/I/V)-X-X-T-P-T POMGNT2 acceptor motif in α-DG by testing glycosylation status and laminin-binding of various α-DG constructs. Using biochemical analyses, we demonstrate that the POMGNT2 minimal acceptor motif is sufficient for functional matriglycan extension elsewhere on α-DG and that this glycosylation can be enhanced by exogenous LARGE1.

Materials and Methods

Constructs. DGFc5, a fusion construct encoding amino acids 1-653 of rabbit α-DG (Uniprot accession number: Q28685) and an Fc tag was obtained from the Baum lab at UCLA. The portion of this construct encoding amino acids 1-653 of rabbit α-DG (Uniprot accession number: Q28685) and additional elements including a 3X FLAG tag were introduced by restriction cloning into the pGEc2-DEST vector to generate the WT α-DG construct. The 4TA, RXR, and Ndel α-DG constructs were designed and ordered from GeneArt (Thermo Fischer Scientific) and introduced into the pGEc2-DEST vector via gateway cloning. Cloning of each construct was verified by sequencing analysis. Full-length LARGE1, a fusion construct encoding the entire human LARGE1 enzyme (Uniprot accession number: O95461) and a C-terminal 3X HA tag in an Ad5 vector with a CMV promoter and an internal ribosome entry site for mCherry was designed and ordered from Genescript via the University of Iowa Viral Vector Core.

Protein Expression and Purification. The WT, 4TA, RXR and Ndel α-DG constructs were expressed as soluble, secreted fusion proteins (C-terminal 3X FLAG tag, TEV-protease cleavage site, 8X His-tag, AviTag and 'superfolder' GFP) by transient co-transfection of HEK293F suspension cultures alongside full-length LARGE1 (Moremen, K.W., Ramiah, A., et al. 2018). Suspension culture HEK293F cells (Life Technologies, Grand Island, NY) were transfected as previously described (Meng, L., Forouhar, F., et al. 2013), and the cell culture media was subjected to Ni-NTA chromatography (Millipore Sigma, St. Louis, MO). Protein preparations eluted with 300 mM imidazole were concentrated to ~1 mg/mL using an Amicon centrifugal concentrator (Millipore Sigma) with a 30 kDa molecular weight cutoff and buffer exchanged into 25 mM HEPES and 200 mM NaCl, pH 7.4. Recombinant expression of a soluble, secreted version of green

fluorescent protein (GFP)-LARGE1 was expressed and purified as previously described (Praissman, J.L., Live, D.H., et al. 2014).

Immunoblotting. Following SDS-PAGE, proteins were transferred to PVDF-FL (Millipore), and probed with antibodies as follows: The anti-glyco α-DG mAb IIH6C4 (1:2,000 Dilution, Millipore) was detected by secondary antibody goat anti-mouse IgM IR680RD (1:10,000, Li-Cor). The anti-α-DG core primary antibody (AF6868, 1:1,000 dilution) was detected by secondary antibody donkey anti-goat IgG IR800CW (1:10,000, Li-Cor). All immunoblots were imaged using a Li-Cor Odyssey scanner.

Laminin Overlay Assay. Laminin overlay assays were performed as previously described (Michele, D.E., Barresi, R., et al. 2002). Natural mouse laminin from the Engelbreth-Holm-Swarm (EHS) sarcoma was purchased from Invitrogen (Catalog number: 23017-015). The rabbit polyclonal anti-laminin antibody (Sigma L9393, 1:1000 dilution) was detected by secondary antibody donkey anti-rabbit IgG IR800CW (1:5000, Li-Cor). All immunoblots were imaged using a Li-Cor Odyssey scanner.

Exogenous LARGEylation Assay. LARGE1 reactions were performed in 0.1 mM MES pH 6.5, 5 mM MgCl, 5 mM MnCl, 2 mM UDP-GlcA and UDP-Xyl, with 0.1 µg/µl of α-DG and 0.01 µg/µl µM GFP-LARGE1. The reaction was allowed to incubate at 37◦C for 16 hrs. Reaction products were subjected to treatment with PNGaseF at 37◦C for 7 hrs and then analyzed by immunoblotting and laminin overlay assays.

Results

POMGNT2 minimal acceptor motif is sufficient for functional matriglycan extension elsewhere on α-DG. In order to test the hypothesized POMGNT2 acceptor motif, we generated α-DG constructs based on the previously published DGFc5 construct (Kanagawa, M., Saito, F., et

al. 2004, Kunz, S., Sevilla, N., et al. 2001) which spans the entire rabbit α-DG protein from amino acids 1-653 (Figure 4.1). The mucin domain of the WT construct is extensively *O*-mannosylated. Previous work has shown direct physical evidence for core M3 extension at threonines 317 and 379 (Hara, Y., Kanagawa, M., et al. 2011, Yoshida-Moriguchi, T., Yu, L., et al. 2010). However, the threonine at position 341, while a demonstrated POMT1/POMT2 acceptor, does not appear to be extended into a core M3 structure (Manya, H., Suzuki, T., et al. 2007). Despite the differences in core *O*-man structures at sites 317 and 379 compared to site 341, all three sites share a similar primary amino acid Thr(-*O*-Man)-Pro-Thr sequence motif. Given the presence of the hypothesized POMGNT2 acceptor motif [RXRXX(L/I/V)XXTPT] surrounding sites 317 and 379 (Halmo, S.M., Singh, D., et al. 2017) and only the partial POMGNT2 acceptor motif [(L/I/V)XXTPT] surrounding site 341, site 341 was selected as a target for probing the POMGNT2 acceptor motif hypothesis. To generate an α-DG construct that lacked any core M3 extended structures, the threonines at site 317, 319, 379, and 381 in the WT construct were mutated to alanine. This construct, called 4TA α-DG, should serve as a negative control for POMGNT2 and core M3 extension (Figure 4.1). An additional α-DG construct, called RXR α-DG, was also generated (Figure 4.1). Similar to 4TA α-DG, RXR α-DG lacks the core M3 extended structures at sites 317 and 379. In addition, the sequence upstream of threonine 341 in the RXR α-DG construct was mutated to contain the complete POMGNT2 acceptor motif via mutation of both a glutamate and a proline to arginine. The RXR α-DG construct was strategically designed to determine if the POMGNT2 minimal acceptor motif is sufficient for functional matriglycan extension elsewhere on a α-DG. An N-terminal deletion construct of α-DG was also generated by deleting amino acids 30-311 from the WT sequence (Figure 4.1). Similar to previous investigations of DGFc5 constructs and given that the N-terminal domain of α-DG is necessary for LARGE activity *in vivo*, the Ndel

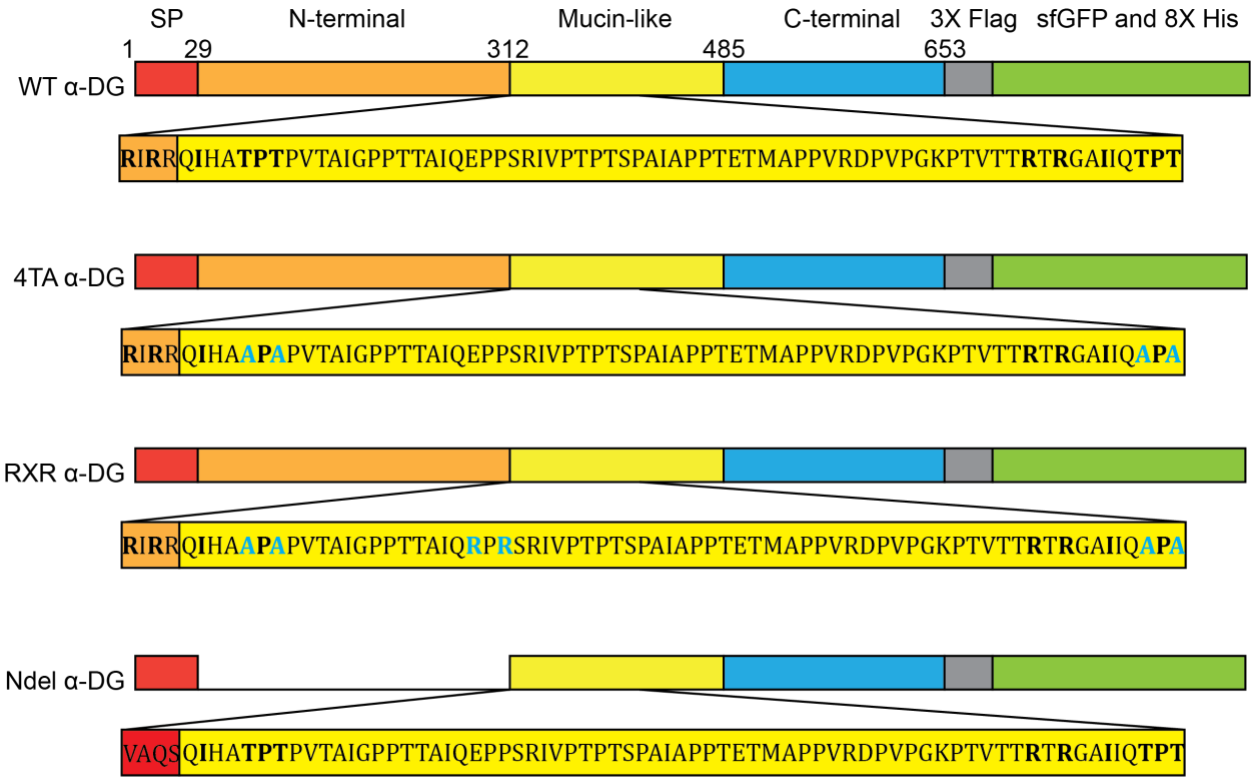


Figure 4.1. α-DG Constructs

Schematic of α-DG protein constructs used in this study. α-DG is composed of a signal peptide (SP) shown in red, an N-terminal domain shown in orange, a mucin-like domain shown in yellow, and a C-terminal domain shown in blue. Laminin-binding epitopes are located within the first 69 amino acids of the mucin-like domain in α-DG (yellow sequence). In the WT sequence, Thr317/319 and Thr379/381 are elaborated with the core M3 glycan structure, while Thr341 is elaborated with core M1 glycan structures. In the 4TA sequence, Thr317/319 and Thr379/381 have been mutated to alanine residues. In the RXR sequence, Thr317/319 and Thr379/381 have been mutated to alanine residues and the sequence upstream of Thr341 is modified to incorporate the hypothesized POMGNT2 motif. The Ndel sequence is identical to the WT sequence, except the N-terminal domain has been removed.

α-DG construct was designed to serve as an additional negative control for core M3 extension (Kanagawa, M., Saito, F., et al. 2004). In contrast to the previously used DGFc constructs which utilized the heavy-chain constant moiety of human IgG1 (Fc) for purification, all constructs utilized in this study were constructed with a C-terminal fusion tag encoding a superfolder GFP and 8X His tag to enable purification of the recombinant proteins (Figure 4.1).

To establish the glycosylation status of the α-DG proteins depicted in Figure 4.1, the constructs were co-transfected with full length LARGE1 into HEK 293 suspension cultures. The secreted α-DG proteins were harvested and purified from culture supernatant, and probed using standard immunoblotting assays. Antibodies used in immunoblotting include the monoclonal antibody IIH6, which recognizes matriglycan or the functional glycan present on glycosylated α-DG, and a core α-DG antibody that recognizes the α-DG protein directly, which serves as a protein loading indicator. In addition, the ligand binding ability of the α-DG constructs was probed via laminin overlay (L/O) assays (Michele, D.E., Barresi, R., et al. 2002). When equally loaded, only WT α-DG was found to react with IIH6 and to bind laminin (Figure 4.2 A). Both 4TA α-DG and RXR α-DG, while expressed, did not react with IIH6 or bind to laminin when loaded at concentrations equal to WT α-DG. Consistent with previous reports and what is known about α-DG, the N-terminal deletion of α-DG (Ndel α-DG) was also expressed but did not react with IIH6 or bind laminin (Figure 4.2 A) (Kanagawa, M., Saito, F., et al. 2004). Interestingly, when five times as much 4TA and RXR α-DG protein were loaded compared to Ndel and WT α-DG, RXR α-DG showed reactivity to IIH6 and could bind laminin while 4TA α-DG still showed no reactivity or binding. This result indicates that combined point mutations of threonines 317, 319, 379, and 381 to alanines abolishes IIH6 reactivity and laminin binding capabilities of α-DG (see 4TA α-DG lanes in Figure 4.2), and that addition of two arginines upstream of site 341 rescues IIH6 reactivity

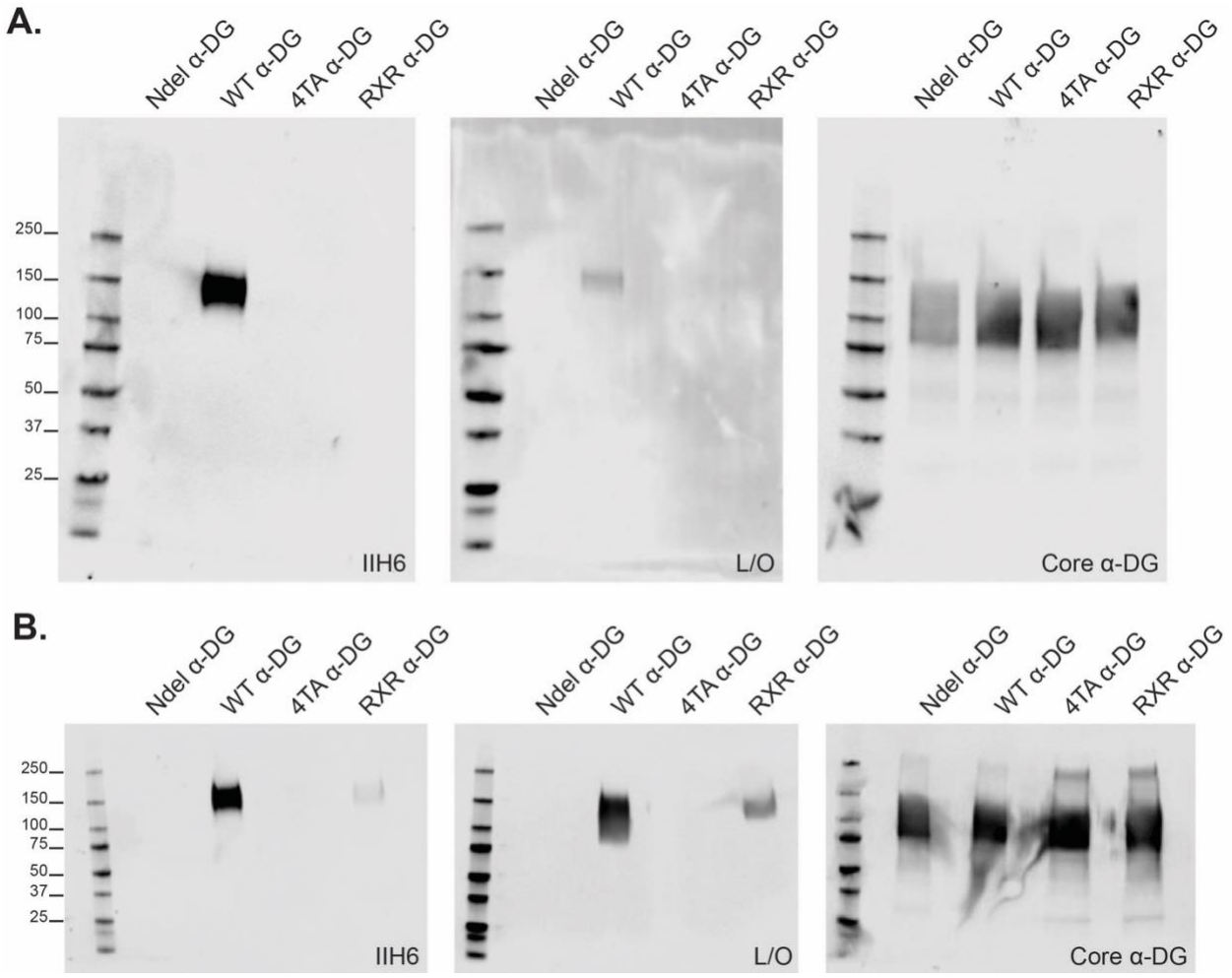


Figure 4.2. POMGNT2 minimal acceptor motif is sufficient for functional matriglycan extension elsewhere on a α-DG

(**A**) Western blots of α-DG constructs that are equally loaded (1 µg per lane). The constructs shown in Figure 4.1 were expressed in HEK293 F suspension cells, and α-DGsfGFP proteins were purified from the culture medium and analyzed by Western blotting with antibodies against the core α-DG protein (Core α-DG) and the glycosylated form of α-DG (IIH6) and by laminin overlay (L/O). Apparent molecular masses (kDa) are indicated.

(**B**) Western blots of α-DG constructs with five times as much 4TA and RXR loaded (10 µg per lane) compared to NdeI and WT (2 µg per lane). The constructs shown in Figure 4.1 were expressed in HEK293 F suspension cells, and α-DGsfGFP proteins were purified from the culture medium and analyzed by Western blotting with antibodies against the core α-DG protein (Core α-DG) and the glycosylated form of α-DG (IIH6) and by laminin overlay (L/O). Apparent molecular masses (kDa) are indicated.

and the laminin binding capabilities of α-DG (see RXR α-DG lanes in Figure 4.2), albeit not to the level of WT α-DG. In essence, engineering in the POMGNT2 motif around a known core M1/M2 site in the α-DG sequence turned this site (341) into a core M3 site, enabling functional glycosylation elsewhere on α-DG.

Exogenous LARGEylation enhances functional matriglycan extension elsewhere on a α-DG. The reduced levels of RXR α-DG IIH6 reactivity and laminin binding compared to WT α-DG (Figure 4.2), begged the question if this new site on α-DG was an acceptable substrate for LARGE activity. We hypothesized that perhaps while containing the necessary primary amino acid determinants for POMGNT2 activity, site 341 in the RXR α-DG construct might be obstructed from access to LARGE *in vivo* due to three-dimensional structural constraints. In order to test this idea, we employed *in vitro* LARGEylation of our α-DG constructs by combining our α-DG proteins with recombinant LARGE1 protein and LARGE1's sugar nucleotide donors, UDP-Xyl and UDP-GlcA. LARGE1 has been shown to extend N-glycans (Aguilan, J.T., Sundaram, S., et al. 2009, Patnaik, S.K. and Stanley, P. 2005), so PNGaseF was included in these reactions to remove potentially LARGEylated N-glycans, which were not the focus of this experiment. As shown in Figure 4.3, exogenous LARGEylation enhances the IIH6 reactivity and laminin binding capability of RXR and WT α-DG, but not that of 4TA α-DG. Exogeneous LARGEylation of Ndel α-DG leads to the appearance of IIH6 reactivity and laminin binding (Figure 4.3). This is not surprising given that the N-terminal domain of α-DG is likely only required to recruit LARGE1 *in vivo*, but not *in vitro*. In sum, exogenous LARGEylation enhances functional matriglycan extension at canonical sites and elsewhere on a α-DG.

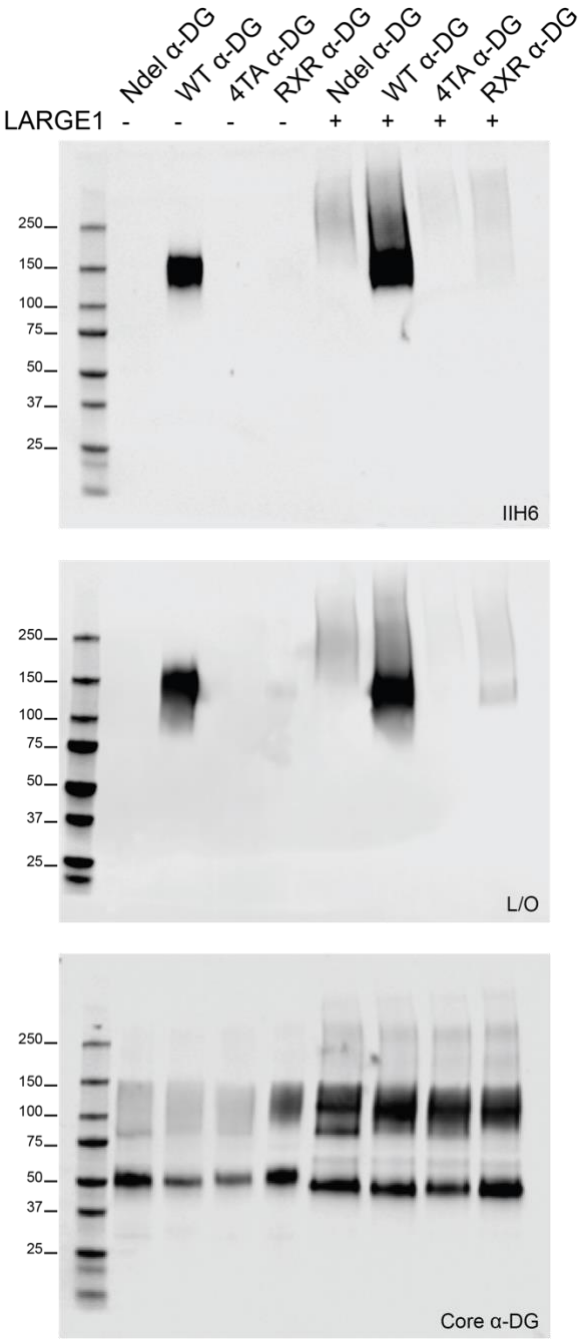


Figure 4.3. Exogenous LARGEylation enhances functional matriglycan extension elsewhere on a α-DG

Western blots of exogenously LARGEylated α-DG constructs that are equally loaded (1 µg per lane). The constructs shown in Figure 4.1 were expressed in HEK293 F suspension cells, and α-DGsfGFP proteins were purified from the culture medium. The α-DGsfGFP proteins were incubated with LARGE1 enzyme and sugar donors (UDP-Xyl and UDP-GlcA), then treated with PNGaseF and analyzed by Western blotting with antibodies against the core α-DG protein (Core α-DG) and the glycosylated form of α-DG (IIH6) and by laminin overlay (L/O). Apparent molecular masses (kDa) are indicated.

Discussion

LARGE-dependent modification of a unique *O*-mannose linked heptasaccharide on α-DG is essential for its ability to bind laminin at the cell surface (Hara, Y., Kanagawa, M., et al. 2011, Yoshida-Moriguchi, T., Yu, L., et al. 2010). However, before LARGE can synthesize matriglycan, a set of enzymes must add and extend certain *O*-mannosylated threonine residues in α-DG (Sheikh, M.O., Halmo, S.M., et al. 2017). The first enzyme in the pathway towards functional *O*-mannose glycan elongation is POMGNT2, which displays primary amino acid selectivity towards α-DG (Halmo, S.M., Singh, D., et al. 2017). Based on the known sites of POMGNT2 extension on α-DG, a POMGNT2 acceptor motif was proposed (Halmo, S.M., Singh, D., et al. 2017). Here, we demonstrate that the core M3 structure, which serves as a scaffold for LARGE extension, can be moved elsewhere on α-DG by engineering in the POMGNT2 motif. Additionally, we show that this functional matriglycan extension elsewhere on α-DG can be enhanced by exogenous LARGE activity.

Although we demonstrate that the core M3 glycan structure can be moved elsewhere on α-DG by introducing the proposed POMGNT2 acceptor motif, we note that extension at this new site, site 341 in the RXR α-DG construct, is not as efficient as extension at the natural sites 317 and 379 in the WT construct. The difference between the engineered site and natural sites is evident by the reduced IIH6 reactivity and laminin binding of the RXR α-DG construct compared to WT in immunoblots and the need to load five times as much RXR α-DG protein to detect reactivity (Figure 4.2). One possible explanation for this finding includes the fact that site 341 may not have structural access to the LARGE enzyme as α-DG transverses through the secretory pathway whereas sites 317 and 379 may reside in a portion of the mucin domain of α-DG that has easy access to LARGE as it interacts with the N-terminal domain. Additionally, given the recent

structural insights into POMGNT1's lectin domain and the resulting hypothesis that POMGNT1 may act as a scaffold for core M3 enzymes (Kuwabara, N., Manya, H., et al. 2016), the *O*-mannosylation status of surrounding serines and threonines near the sites in question may play a role. Notably, nearby sites 328 and 329 to site 317/319 have been shown to be *O*-mannosylated, and nearby sites 367, 369, 372, and 378 to sites 379/381 have be shown to possess *O*-mannose glycans (Nilsson, J., Nilsson, J., et al. 2010, Vester-Christensen, M.B., Halim, A., et al. 2013, Yoshida-Moriguchi, T., Yu, L., et al. 2010). Nearby *O*-mannosylation to site 341 includes sites 343, 344, and 351 (Vester-Christensen, M.B., Halim, A., et al. 2013). The idea that a glycosyltransferase, like POMGNT2, may recognize glycan structures distal to the catalytic site does not seem too far-fetched given that these enzymes often contain lectin-like stem domains. This line of thought is especially intriguing considering the recent structural findings of FKRP, which indicate that the stem domain of one FKRP subunit, the second RboP transferase responsible for building the heptasacchraide on which LARGE is built, recognizes the phosphate group on the distal mannose (Kuwabara, N., Imae, R., et al. 2020).

We demonstrated that functional matriglycan extension elsewhere on α-DG can be enhanced by exogenous LARGE activity. However, we direct the reader to consider an important caveat. Namely, while we eliminated the possibility for N-glycans to be extended by LARGE via the inclusion of PNGaseF in the reactions, we did not exclude the possibility of *O*-GalNac linked glycans from being LARGEylated (Patnaik, S.K. and Stanley, P. 2005). The exogenous LARGEylation assays may have non-specifically LARGEylated *O*-GalNAc glycans leading to the increased signal seen in Figure 4.3 that we attribute to the LARGEylation of *O*-mannose linked glycans. Therefore, direct evidence of the functional glycan, or at least the core M3 phosphotrisaccharide present at site 341, via LC-MS analysis is needed. While we speculate that

the primed and extended core M3 structure is likely present at site 341 in the RXR α-DG construct given that it is IIH6 reactive, can bind laminin, and can be further LARGEylated, direct evidence of the glycan structure would verify our results.

The main finding that the core M3 glycan structure can be moved elsewhere on α-DG by introducing the proposed POMGNT2 acceptor motif raises a few questions that will need to be addressed in future studies. First, given that site 367 in α-DG has been shown to be *O*-mannosylated (Nilsson, J., Nilsson, J., et al. 2010) and the fact that surrounding amino acids include partial pieces of the POMGNT2 acceptor motif (RXXXXXXXTXT in RDPVPGKPTVT), engineering in the POMGNT2 acceptor motif near site 367 in the 4TA α-DG background could recapitulate the findings presented here. Second, a clear next question to address is if the core M3 structure can be introduced onto other proteins by simply engineering in the POMGNT2 acceptor motif. A few intriguing candidate proteins exist, including the highly *O*-mannosylated multipass transmembrane protein of unknown function, KIAA1549 (Uniprot accession number: Q9HCM3) (Larsen, I.S.B., Narimatsu, Y., et al. 2019), and the *O*-mannosylated single-pass type I membrane protein tyrosine phosphatase, phosphacan (Uniprot accession number: P23471). Both proteins are known to be *O*-mannosylated and are highly expressed in the brain (de Bruijn, S.E., Verbakel, S.K., et al. 2018, Dwyer, C.A., Baker, E., et al. 2012, Dwyer, C.A., Katoh, T., et al. 2015, Larsen, I.S.B., Narimatsu, Y., et al. 2017, Vester-Christensen, M.B., Halim, A., et al. 2013), which is one of the affected tissues in dystroglycanopathies. Moving functional matriglycan that is currently only known to occur on α-DG onto other proteins like KIAA1549 and phosphacan is a clear next step. We hypothesize that functional matriglycan extension on these proteins is unlikely because these proteins lack the N-terminal domain of α-DG which is required for LARGE activity (Kanagawa, M., Saito, F., et al. 2004). However, the absence of the N-terminal domain of α-DG

does not exclude the possibility of POMGNT2 activity and perhaps the synthesis of the core M3 phosphotrisaccharide on these candidate proteins.

References

Aguilan JT, Sundaram S, Nieves E, Stanley P. 2009. Mutational and functional analysis of Large in a novel CHO glycosylation mutant. *Glycobiology*, 19:971-986.

Briggs DC, Yoshida-Moriguchi T, Zheng T, Venzke D, Anderson ME, Strazzulli A, Moracci M, Yu L, Hohenester E, Campbell KP. 2016. Structural basis of laminin binding to the LARGE glycans on dystroglycan. *Nature chemical biology*, 12:810-814.

de Bruijn SE, Verbakel SK, de Vrieze E, Kremer H, Cremers FPM, Hoyng CB, van den Born LI, Roosing S. 2018. Homozygous variants in KIAA1549, encoding a ciliary protein, are associated with autosomal recessive retinitis pigmentosa. *Journal of medical genetics*, 55:705-712.

Durbeej M, Campbell KP. 2002. Muscular dystrophies involving the dystrophin-glycoprotein complex: an overview of current mouse models. *Curr Opin Genet Dev*, 12:349-361.

Dwyer CA, Baker E, Hu H, Matthews RT. 2012. RPTPzeta/phosphacan is abnormally glycosylated in a model of muscle-eye-brain disease lacking functional POMGnT1. *Neuroscience*, 220:47-61.

Dwyer CA, Katoh T, Tiemeyer M, Matthews RT. 2015. Neurons and glia modify receptor protein-tyrosine phosphatase zeta (RPTPzeta)/phosphacan with cell-specific O-mannosyl glycans in the developing brain. *J Biol Chem*, 290:10256-10273.

Gerin I, Ury B, Breloy I, Bouchet-Seraphin C, Bolsee J, Halbout M, Graff J, Vertommen D, Muccioli GG, Seta N, *et al.* 2016. ISPD produces CDP-ribitol used by FKTN and FKRP to transfer ribitol phosphate onto alpha-dystroglycan. *Nature communications*, 7:11534.

Godfrey C, Foley AR, Clement E, Muntoni F. 2011. Dystroglycanopathies: coming into focus. *Current Opinion in Genetics & Development*, 21:278-285.

Grewal PK, Holzfeind PJ, Bittner RE, Hewitt JE. 2001. Mutant glycosyltransferase and altered glycosylation of α-DG in the myodystrophy mouse. *Nature Genetics*, 28:151-154.

Halmo SM, Singh D, Patel S, Wang S, Edlin M, Boons GJ, Moremen KW, Live D, Wells L. 2017. Protein O-Linked Mannose beta-1,4-N-Acetylglucosaminyl-transferase 2 (POMGNT2) Is a

Gatekeeper Enzyme for Functional Glycosylation of alpha-Dystroglycan. *J Biol Chem,* 292:2101-2109.

Hara Y, Kanagawa M, Kunz S, Yoshida-Moriguchi T, Satz JS, Kobayashi YM, Zhu Z, Burden SJ, Oldstone MBA, Campbell KP. 2011. Like-acetylglucosaminyltransferase (LARGE)-dependent modification of dystroglycan at Thr-317/319 is required for laminin binding and arenavirus infection. *Proceedings of the National Academy of Sciences*, 108:17426-17431.

Henry MD, Campbell KP. 1998. A role for dystroglycan in basement membrane assembly. *Cell*, 95:859-870.

Hohenester E. 2019. Laminin G-like domains: dystroglycan-specific lectins. *Current Opinion in Structural Biology,* 56:56-63.

Kanagawa M, Kobayashi K, Tajiri M, Manya H, Kuga A, Yamaguchi Y, Akasaka-Manya K, Furukawa J-i, Mizuno M, Kawakami H, *et al.* 2016. Identification of a Post-translational Modification with Ribitol-Phosphate and Its Defect in Muscular Dystrophy. *Cell Reports*, 14:2209–2223.

Kanagawa M, Saito F, Kunz S, Yoshida-Moriguchi T, Barresi R, Kobayashi YM, Muschler J, Dumanski JP, Michele DE, Oldstone MBA, *et al.* 2004. Molecular Recognition by LARGE Is Essential for Expression of Functional Dystroglycan. *Cell*, 117:953-964.

Kunz S, Sevilla N, McGavern DB, Campbell KP, Oldstone MB. 2001. Molecular analysis of the interaction of LCMV with its cellular receptor [alpha]-dystroglycan. *J Cell Biol,* 155:301-310.

Kuwabara N, Imae R, Manya H, Tanaka T, Mizuno M, Tsumoto H, Kanagawa M, Kobayashi K, Toda T, Senda T, *et al.* 2020. Crystal structures of fukutin-related protein (FKRP), a ribitol-phosphate transferase related to muscular dystrophy. *Nature communications,* 11:303.

Kuwabara N, Manya H, Yamada T, Tateno H, Kanagawa M, Kobayashi K, Akasaka-Manya K, Hirose Y, Mizuno M, Ikeguchi M, *et al.* 2016. Carbohydrate-binding domain of the POMGnT1 stem region modulates O-mannosylation sites of α-DG. *Proceedings of the National Academy of Sciences*, 113:9280-9285.

Larsen ISB, Narimatsu Y, Clausen H, Joshi HJ, Halim A. 2019. Multiple distinct O-Mannosylation pathways in eukaryotes. *Curr Opin Struct Biol,* 56:171-178.

Larsen ISB, Narimatsu Y, Joshi HJ, Yang Z, Harrison OJ, Brasch J, Shapiro L, Honig B, Vakhrushev SY, Clausen H, *et al.* 2017. Mammalian O-mannosylation of cadherins and plexins is independent of protein O-mannosyltransferases 1 and 2. *J Biol Chem*, 292:11586-11598.

Manya H, Suzuki T, Akasaka-Manya K, Ishida H-K, Mizuno M, Suzuki Y, Inazu T, Dohmae N, Endo T. 2007. Regulation of Mammalian Protein O-Mannosylation: PREFERENTIAL AMINO ACID SEQUENCE FOR O-MANNOSE MODIFICATION. *Journal of Biological Chemistry*, 282:20200-20206.

Meng L, Forouhar F, Thieker D, Gao Z, Ramiah A, Moniz H, Xiang Y, Seetharaman J, Milaninia S, Su M, *et al.* 2013. Enzymatic basis for N-glycan sialylation: structure of rat alpha2,6-sialyltransferase (ST6GAL1) reveals conserved and unique features for glycan sialylation. *J Biol Chem*, 288:34680-34698.

Michele DE, Barresi R, Kanagawa M, Saito F, Cohn RD, Satz JS, Dollar J, Nishino I, Kelley RI, Somer H, *et al.* 2002. Post-translational disruption of dystroglycan-ligand interactions in congenital muscular dystrophies. *Nature*, 418:417-422.

Moremen KW, Ramiah A, Stuart M, Steel J, Meng L, Forouhar F, Moniz HA, Gahlay G, Gao Z, Chapla D, *et al.* 2018. Expression system for structural and functional studies of human glycosylation enzymes. *Nature chemical biology*, 14:156-162.

Nilsson J, Nilsson J, Larson G, Grahn A. 2010. Characterization of site-specific O-glycan structures within the mucin-like domain of α-DG from human skeletal muscle. *Glycobiology*, 20:1160-1169.

Patnaik SK, Stanley P. 2005. Mouse large can modify complex N- and mucin O-glycans on alpha-dystroglycan to induce laminin binding. *J Biol Chem*, 280:20851-20859.

Praissman JL, Live DH, Wang S, Ramiah A, Chinoy ZS, Boons G-J, Moremen KW, Wells L. 2014. B4GAT1 is the priming enzyme for the LARGE-dependent functional glycosylation of α-DG. *eLife*, 3.

Praissman JL, Willer T, Sheikh MO, Toi A, Chitayat D, Lin YY, Lee H, Stalnaker SH, Wang S, Prabhakar PK, *et al.* 2016. The functional O-mannose glycan on alpha-dystroglycan contains a phospho-ribitol primed for matriglycan addition. *Elife*, 5.

Sheikh MO, Halmo SM, Wells L. 2017. Recent advancements in understanding mammalian O-mannosylation. *Glycobiology*, 27:806-819.

Stalnaker SH, Hashmi S, Lim J-M, Aoki K, Porterfield M, Gutierrez-Sanchez G, Wheeler J, Ervasti JM, Bergmann C, Tiemeyer M, *et al.* 2010. Site Mapping and Characterization of O-Glycan Structures on α-DG Isolated from Rabbit Skeletal Muscle. *The Journal of Biological Chemistry*, 285:24882-24891.

Vester-Christensen MB, Halim A, Joshi HJ, Steentoft C, Bennett EP, Levery SB, Vakhrushev SY, Clausen H. 2013. Mining the O-mannose glycoproteome reveals cadherins as major O-mannosylated glycoproteins. *Proceedings of the National Academy of Sciences*, 110:21018-21023.

Wells L. 2013. The O-Mannosylation Pathway: Glycosyltransferases and Proteins Implicated in Congenital Muscular Dystrophy. *The Journal of Biological Chemistry*, 288:6930-6935.

Yagi H, Nakagawa N, Saito T, Kiyonari H, Abe T, Toda T, Wu S-W, Khoo K-H, Oka S, Kato K. 2013. AGO61-dependent GlcNAc modification primes the formation of functional glycans on α-DG. *Scientific Reports*, 3:3288.

Yoshida-Moriguchi T, Campbell KP. 2015. Matriglycan: a novel polysaccharide that links dystroglycan to the basement membrane. *Glycobiology*, 25:702-713.

Yoshida-Moriguchi T, Yu L, Stalnaker SH, Davis S, Kunz S, Madson M, Oldstone MB, Schachter H, Wells L, Campbell KP. 2010. O-mannosyl phosphorylation of alpha-dystroglycan is required for laminin binding. *Science*, 327:88-92.

CHAPTER 5

INTRODUCTION - PART 2

This chapter describes the purpose of the studies within the second half of this dissertation, outlines relevant theoretical and conceptual frameworks, and provides an overview of the remaining chapters.

Purpose of Studies

There is a sincere need and rejuvenated effort to improve undergraduate science education and the disciplines it encompasses (AAAS, 2010). From addressing climate change to resolving racial and social health disparities, complex challenges face our nation and world today. These complex problems impact every human life, and will likely require an array of interdisciplinary scientific collaborators, policy-makers, and citizens to solve. Whether in a voting booth, at the bench, or in the hospital, every person needs a basic understanding of science and scientific principles in order to actively contribute to solving the challenging problems of the world. In undergraduate science education, this means all students need to develop a level of scientific literacy. In addition to a scientifically literate population, undergraduate science education must also prepare and supply the world with a diverse scientific workforce. A diverse workforce ensures a diversity of solutions to our complex problems. However, science suffers from a loss of contributors to its talent pool with attrition from science, technology, engineering, and mathematics (STEM) occurring at every progression, from high school to college, college to professional school, and professional school to career (Barr, Gonzalez, & Wanat, 2008; Chen, 2013). This STEM attrition problem disproportionately affects individuals from diverse and

underrepresented backgrounds (National Research Council, 2011). Undergraduate science education represents a key point in student development. This time point can be leveraged to create a more scientifically literate and diverse scientific workforce. Improvements to undergraduate science education are necessary to better prepare all students for the unique challenges of the 21st century. AAAS (2010) specifically calls for undergraduate science education to use evidence-based active learning strategies, integrate and organize factual knowledge into a larger conceptual context for students, focus on the processes, nature, and limits of science, and heed critical insights from cognitive and learning sciences.

Discipline-based education research (DBER) answers the call for reformed undergraduate science education by investigating learning within specific sub-disciplines of STEM. DBER is an "interdisciplinary research enterprise that combines the expertise of scientists with methods and theories that explain learning" (National Research Council, 2012). This differs from and complements research on general learning and cognition by looking at learning through a lens that prioritizes the discipline's perspective, knowledge, and practices (National Research Council, 2012). DBER is grounded in expert knowledge of the discipline and the associated challenges with learning and teaching within that discipline. Due to this fact, some DBER scholars have dual training in a natural science discipline and in education research. DBER scholars bring unique capabilities to this research context including deep disciplinary knowledge, understanding of the nature of science, and social science expertise. For me, this means I have sought out formal training in biochemistry and education research. DBER scholars are called to this work to understand how people learn the concepts, practices, and ways of thinking in science, understand the nature and development of expertise within science, identify and measure instructional approaches that

advance students toward learning objectives while making science education broad and inclusive, and translate research findings into classroom practice (National Research Council, 2012).

DBER has generated evidence-based pedagogies for improving undergraduate science education. Active learning, which requires students to engage cognitively and meaningfully with learning materials, improves performance in STEM courses (Freeman et al., 2014). Additionally, compared to lecture, active learning reduces the achievement gap for students (Haak, HilleRisLambers, Pitre, & Freeman, 2011) and increases actual learning despite students feeling as though they learned less (Deslauriers, McCarty, Miller, Callaghan, & Kestin, 2019). Although, active learning does not always work as designed (Andrews, Leonard, Colgrove, & Kalinowski, 2011). In practice, types of active learning range from including clicker questions in lecture-based courses to completely revamping courses based on principles like case-based learning (Stains et al., 2018). Several types of active learning have been shown to be effective, including worked examples plus practice (Paas & van Gog, 2006), productive failure (Kapur, 2008), and guided inquiry (Hmelo-Silver, Duncan, & Chinn, 2007). However, direct comparisons between these effective types of active learning are limited. Even though evidence abounds for the above pedagogies in math, chemistry, and biology, there is still a gap in our understanding of how learning can be best supported in the discipline of biochemistry.

Biochemistry is one subdiscipline of the life sciences that integrates the disciplines of biology and chemistry to study the chemical processes that occur within and relating to living organisms. The perspective from which biochemistry operates includes the principles of evolution and homeostasis. The practices that define biochemistry include the process of science, collaboration, and accessing, assessing and communicating science. The knowledge that defines biochemistry includes foundational concepts such as energy, information storage, and structure

and function (ASBMB, 2020). One core concept related to structure and function in biochemistry, and the focus of the second half of this dissertation, is the physical basis of noncovalent interactions (Loertscher, Green, Lewis, Lin, & Minderhout, 2014).

The physical basis of noncovalent interactions encompasses molecular forces which are electrostatic in nature. Interactions occur because of the electrostatic properties of molecules which can involve full, partial, and/or temporary charges. The unifying principle for this core concept is Coulomb's Law, which explains that a molecular force, whether within a molecule or between molecules, depends on the magnitude (full, partial, temporary), sign (positive or negative), and distance between the charges involved. Biochemists have categorized noncovalent interactions by both the ways in which charges arise and the terminology used to describe the resulting charges. Three main categories are: ion pairing, hydrogen bonds, and van der Waals interactions. Once the concept of the physical basis of interactions is understood, similarities between different interactions become clear, and a student recognizes that although noncovalent interactions are given different names, they are all based on the same electrostatic principles of attraction due to opposite charge (Loertscher et al., 2014). Deep understanding of noncovalent interactions is essential for students to integrate structure and function relationships in biochemistry. Students are introduced to noncovalent interactions earlier in their undergraduate careers. Intramolecular and intermolecular forces are covered in general chemistry and introductory biology courses also teach noncovalent interactions to some extent. While the names of noncovalent interactions may be familiar to students, it is unclear if they learned about the physical basis of noncovalent interactions as a concept. Therefore, it is critical to understand how students entering and leaving biochemistry are thinking about noncovalent interactions.

Several student difficulties with the concept of noncovalent interactions have been identified. For example, middle school students find ionic bonding difficult to explain because they view bonds as entities rather than interactions between charged particles (Boo & Watson, 2001). Xu and colleagues used the Instrument of Foundational Concepts for Biochemistry (IFCB) (Villafañe, Bailey, Loertscher, Minderhout, & Lewis, 2011) to uncover incorrect ideas that biochemistry students have about hydrogen bonding, such as all hydrogens are capable of participating in a hydrogen bond regardless of their covalent bond participation (Xu, Lewis, Loertscher, Minderhout, & Tienson, 2017). In a study on chemistry student understanding of intermolecular forces, 55% of students incorrectly and unambiguously represented London dispersion forces as an interaction or bond within a single molecule (Cooper, Williams, & Underwood, 2015). Another study on students' mechanistic reasoning of London dispersion forces showed that immediately after instruction a majority of students explain interactions between helium atoms as electrostatic in nature. However, most students provided limited detail as to the mechanism by which these electrostatic interactions arise between neutral atoms (Becker, Noyes, & Cooper, 2016). Additional work indicates that students enrolled in a transformed general chemistry curriculum have a more scientifically correct and coherent understanding of noncovalent interactions compared to students in a traditional general chemistry course, and that this difference persists through organic chemistry (Williams, Underwood, Klymkowsky, & Cooper, 2015). Taken together, this body of work indicates that from middle school through undergraduate science courses, students hold a mixture of ideas about noncovalent interactions, including non-canonical ideas. These mixed, incomplete ways of thinking likely account for students' difficulties explaining the underlying causes and electrostatic mechanisms of noncovalent interactions.

Previous work also indicates that instruction can target these mixed ideas and lead to deep and meaningful learning (Williams et al., 2015).

Given the importance of the concept of the physical basis of interactions for understanding macromolecular structure and function in biochemistry, the purpose of the studies presented in the second half of this dissertation are two-fold. First, this body of work aims to characterize the knowledge and problem-solving steps of students and experts as they think through noncovalent interactions in a protein structure-function problem unique to the discipline of biochemistry. Second, using the insights into student difficulties with a protein structure-function problem, this work aims to create instruction that will promote deep learning of the physical basis of noncovalent interactions in biochemistry.

Theoretical and Conceptual Frameworks

This body of work draws upon several theoretical and conceptual frameworks to investigate biochemistry-specific learning.

Constructivism and How People Learn. The goal of education shifted from the early part of the twentieth century to now (National Research Council, 2000). The former goal of education was knowledge acquisition in reading, writing, and arithmetic. Because of this view, most curricula emphasized memory with assessments testing students' ability to remember discrete facts learned in textbooks. State-of-the-art education research no longer supports this transmissionist view. Instead, it is widely accepted that the goal of education is to help students develop knowledge and skills to think and problem solve productively between and across subject areas. Given this constructivist view, curricula should focus on deep understanding rather than memorization with assessments testing students' ability to apply and transfer usable knowledge to various contexts (National Research Council, 2000).

The work described within this dissertation uses the lens of constructivism as the main guiding theoretical framework. Constructivist learning theory posits that knowledge is constructed in the mind of the learner, and that existing knowledge is used to build new knowledge (Bodner, Klobuchar, & Geelan, 2001). Extensions of this theory also acknowledge that learning happens socially and in context, and that learning is facilitated by more advanced guides and scaffolds (i.e., instructors, advanced peers, and instructional materials) (National Academies of Sciences & Medicine, 2018; Wood, Bruner, & Ross, 1976). Two key assumptions with implications for learning and research emerge from constructivist theory. First, prior knowledge matters. Students enter education with a range of prior knowledge, skills, and beliefs that influence how they interact and engage with the learning environment. This impacts their ability to construct new knowledge. Therefore, instructors must draw out and work with students' preexisting understandings (Vosniadou & Brewer, 1989). In other words, the instructor must actively inquire into what the student is thinking. Revealing student thinking helps an instructor recognize what makes mastery challenging, build on student ideas, challenge those ideas, and monitor students changing conceptions throughout instruction. The assumption that all knowledge is constructed from previous knowledge should not be misconstrued to mean that instructors should never directly tell students anything. Knowledge construction can occur even when listening to a lecture. Second, developing expertise for retrieval and application requires a deep foundation of knowledge that is organized and connected. Therefore, instructors must teach core concepts in depth and use assessments that test deep understanding rather than surface knowledge (National Research Council, 2000). In addition to these broad constructivist assumptions, my research drew upon several conceptual frameworks to further define concepts and beliefs that support and guide the research.

Knowledge and Domain-Specificity. A key guiding belief for the work presented here is that optimal learning is knowledge and context dependent, or domain-specific. Knowledge can be broadly defined as a concept, misconception, principle, fact or skill. This knowledge can be domain-general or domain-specific. Domain-general knowledge can be applied across domains, whereas domain-specific knowledge is applicable to a specific field of study like biochemistry (Alexander, Schallert, & Hare, 1991). Importantly, domain-specific knowledge is more than simply knowing discrete facts or pieces of knowledge, but knowing how and when to utilize and retrieve well-connected pieces of knowledge (Alexander, 1992). An individual's knowledge base is composed of both domain-general and domain-specific knowledge. Some research focuses solely on the construction of domain-general knowledge or strategies. However, attempts to teach general skills or strategies in the absence of a specific context or in the absence of domain-specific knowledge does not support problem solving or transfer (National Research Council, 2000). In contrast, some research focuses solely on the construction of domain-specific knowledge because the researchers argue that domain-general knowledge is innate and only domain-specific knowledge matters in educational settings (Tricot & Sweller, 2014). The research presented here assumes that learning and knowledge construction rely on both domain-general and domain-specific knowledge (Alexander & Judy, 1988). The recognition of the domain-specificity of knowledge calls for research in particular domains and not just research on learning in general.

Knowledge-in-Pieces. An important framework for thinking about the construction of domain-specific knowledge in undergraduate education is the knowledge-in-pieces framework (diSessa, 2018). This framework models knowledge as loosely aggregated or in pieces. Students come to instruction with disconnected and haphazardly organized pieces of knowledge. The goal of instruction is then for students to build increasingly sophisticated networks of knowledge by

assembling their knowledge pieces into conceptual understanding. Knowledge-in-pieces offers a dynamic perspective on cognition with patterns of thinking emerging from how knowledge pieces are activated (Gouvea & Simon, 2018). In this model, all knowledge pieces can be productive stepping stones towards more sophisticated, domain-specific knowledge webs. The trajectory of this research is to identify knowledge pieces and design instructional interventions that encourage students to build and integrate their knowledge pieces around core concepts.

Problem Solving and Expertise. One source of information on domain-specific knowledge and successful problem solving comes from the study of expertise (National Research Council, 2000). Broadly speaking, problem solving is the application of knowledge to a task. Problem solving is required whenever a goal state exists and the path from the initial state to the goal state is uncertain and not possible by direct action or memory (Newell & Simon, 1972; Novick & Bassok, 2005). Determining the tip on a restaurant bill, figuring out a route to a destination given certain road closures, playing a game of chess, and developing a cure for a disease are all examples of where problem solving is needed. Experts think effectively about problems in a given domain or discipline, and their expertise is often marked by several thousand hours of experience and practice in said domain (Chase & Simon, 1973). Novices, on the other hand, are less experienced with solving problems in a given domain.

Research comparing expert and novice problem-solving in specific domains provides several substantiated claims. First, experts are efficient problem solvers compared to novices because they have a rich, well-organized, and connected knowledge base. Expert knowledge is organized around core concepts or big ideas that guide an expert's thinking and problem solving in a domain (National Research Council, 2000). This enables automatic and fluent retrieval of knowledge that is relevant to a particular problem. In contrast, novice knowledge can be rich, but

is often loosely-connected or in pieces which makes retrieval less fluent and automatic (diSessa, 2018). Second, experts recognize features and patterns in problems that go unnoticed by novices (National Research Council, 2000). For example, experts classify common classroom problems based on deep features and underlying concepts, while novices classify problems based on surface features (Chi, Feltovich, & Glaser, 1981; Nehm & Ridgway, 2011; Smith, 1992). Third, compared to novices, experts recognize the limits of their current knowledge and devise ways to obtain the knowledge they need to solve a problem. Given these expert-novice differences, experts may forget what is easy and what is difficult for novices learning in a domain. Moving students on the novice-expert continuum will involve helping students reckon with the discrepancies between their non-canonical and canonical ideas as well as assist them to organize and connect new and prior knowledge. Novice students might benefit from models of how experts approach problem solving.

Most of the problems used in this body of work and that students encounter in classrooms are well-defined. A well-defined problem is a task with clearly defined initial and goal states and a limited number of possible and correct solutions. In contrast, an ill-defined problem is a task where one or more of the problem components are not clearly defined and the possible solutions are limitless (Chi & Glaser, 1985). Most of the important problems that face society today are ill-defined. These are the types of problems today's students will have to solve after they graduate, yet most of schooling deals with learning how to solve well-defined problems. Therefore, for successful and meaningful learning to occur, transfer of knowledge learned in school to new contexts is necessary.

Transfer. Transfer, between school and workplace, from one course to another, and from one problem to another, is defined in the classical cognitive perspective as the ability to apply knowledge learned in one context to another context (National Research Council, 2000). In

contrast to measuring retention through rote memorization, measuring meaningful learning and understanding via transfer is aligned with the constructivist view (Mayer, 2002). However, transfer as a concept is critiqued by more sociocultural-centered learning theories like situated cognition that posit that knowledge is strongly embedded in the context in which it is learned (Lobato, 2006, 2012). These other learning theories take the position that the concept of transfer represents a passive application of knowledge and ignores the social context of learning (Kohn, Underwood, & Cooper, 2018; Lobato, 2006). This dissertation assumes the classical cognitive perspective of transfer, and by doing so, relies on models of expert performance and ignores the contribution of the environment (Lobato, 2012).

Transfer can be classified as near or far depending on whether the two contexts are similar or dissimilar, respectively. For example, transfer from one school task to another highly similar school task can be called near transfer, and transfer from school domains to a non-school setting can be called far transfer (National Research Council, 2000). In this example, the setting of the task (school vs. non-school) is the distinguishing factor. In other instances, near and far transfer have been distinguished by the surface form of the task itself. In this body of work, near transfer tasks are highly similar to tasks used in instruction both in structure and surface form, whereas far transfer tasks rely on principles learned during instruction but require different solution structures than the tasks used during instruction.

Research has defined conditions for effective transfer. First, initial learning is necessary for transfer (National Research Council, 2000). Initial learning that supports transfer focuses on meaningful learning rather than rote learning (Mayer, 2002) and attends to motivation and relevance. Instructional materials designed to help students notice deep underlying structure can be used for initial conceptual learning (Gentner, Loewenstein, & Thompson, 2003; Schwartz &

Bransford, 1998; Schwartz, Chase, Oppezzo, & Chin, 2011). Second, transfer tasks must share cognitive elements and deep abstract structure for transfer to occur (Gick & Holyoak, 1980).

Cognitive Architecture and Processing. Much of the research on problem-solving transfer is based in information processing theory (Newell & Simon, 1972). Information processing theory focuses on the domain-general cognitive processes used to reach a problem solution such as brainstorming and means-end analysis which involves working backwards from the goal state to the initial problem state (Chi & Glaser, 1985). However, information processing theory does not consider the knowledge base of the learner, a key principle of the constructivist learning theory. To effectively support meaningful learning from a constructivist lens, it is prudent to consider cognitive architecture and processing that attends to prior knowledge.

Two aspects of cognitive architecture that consider the knowledge base of the learner are long-term memory and working memory (Sweller, Van Merrienboer, & Paas, 1998). Long-term memory consists of stored knowledge or information. Knowledge stored in long-term memory is limitless and can include large, complex webs of interacting knowledge. Contents of long-term memory filter through working memory. Working memory is often equated with consciousness and is where information is processed and knowledge webs are constructed (Sweller, 2016). Working memory is limited in that it is capable of holding about seven pieces of information at a time (Miller, 1956).

Cognitive load theory is an instructional design theory that emphasizes working memory constraints as determinants of effective instruction. From this view, instruction should aim to limit extraneous cognitive load on the learner's working memory to make efficient use of working memory capacity (Kalyuga & Singh, 2016; Paas, Renkl, & Sweller, 2003). In addition, cognitive load theory notes that there is an optimal, germane level of cognitive load or amount of working

memory resources that must be devoted to the mental effort of processing knowledge for long-term memory storage. Germane cognitive load enhances learning (Sweller et al., 1998).

Extending this idea of germane cognitive load is the concept of desirable difficulty in learning. The concept of desirable difficulty embodies the idea that instruction should aim to create an appropriate level of challenge for the learner (Bjork, 1994), and stems from research on how to optimize learning in real-world settings. Specifically, research on verbal and motor learning revealed two key findings that led to the concept of desirable difficulty (Schmidt & Bjork, 1992). First, manipulations that maximize performance during training can be detrimental for long-term transfer. Second, manipulations that slow performance during training can improve transfer performance or the ability to generalize training to new contexts (Kapur, 2016). From a cognitive perspective, desirable difficulty conditions that slow performance during training actually prepare the learner for processing activities required during testing or transfer scenarios. Examples of desirable difficulty conditions include delaying feedback during initial learning and variable task practice (Schmidt & Bjork, 1992). In essence, the proper level of difficulty is needed to optimize learning and maintain motivation. Tasks that are too easy will become boring and tasks that are too challenging will cause frustration (National Research Council, 2000).

The work presented in the remaining chapters relies on the assumptions and beliefs of constructivist learning theory and the associated concepts of domain-specific knowledge, knowledge-in-pieces, expertise, problem solving, transfer, and cognitive architecture to investigate biochemistry-specific learning and instruction.

Overview of Remaining Chapters

The second half of this dissertation is composed of two manuscripts that support problem solving in undergraduate biochemistry courses. The first manuscript, Chapter 6, is a study

describing how students think and solve a structure-function biochemistry problem (Halmo et al., 2018). Using think-aloud interviews and a qualitative research methodology, this manuscript characterizes the knowledge and problem-solving steps that six experts and thirteen undergraduate students use to solve and reason through a biochemistry problem involving the physical basis of noncovalent interactions. From this work, several student difficulties with the problem are described and discussed. Implications for biochemistry educators and researchers are proposed to support the learning and problem-solving application of the physical basis of noncovalent interactions.

The second manuscript, Chapter 7, is an empirical study investigating the comparative impacts of different pedagogies—varying in the nature and timing of guidance—on student learning of the physical basis of noncovalent interactions (Halmo et al., in press). Using an experimental design and quantitative methodology, this manuscript compares various pedagogies and their respective impact on basic knowledge performance, near-transfer problem solving, and far-transfer problem solving, while accounting for differing levels in prior knowledge. The various pedagogies selected for comparison are worked examples plus practice, productive failure and two forms of guided inquiry termed scaffolded and unscaffolded guidance. This works shows 1) the four pedagogies do not differentially impact basic knowledge performance, 2) worked examples plus practice, productive failure, and scaffolded guidance lead to greater near-transfer performance compared to unscaffolded guidance, and 3) productive failure and scaffolded guidance do not differentially impact far-transfer performance. Based on these findings, recommendations for future research and biochemistry instruction on the physical basis of noncovalent interactions are discussed.

The final chapter of this dissertation summarizes the major findings and future directions for both parts of this dissertation, in turn. It concludes by joining the two bodies of work together in their broader context and significance.

References

AAAS. (2010). *Vision and Change: A Call to Action*. Retrieved from Washington, DC: www.visionandchange.org/VC_report.pdf

Alexander, P. A. (1992). Domain knowledge: Evolving themes and emerging concerns. *Educational Psychologist, 27*(1), 33-51.

Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of educational research, 58*(4), 375-404.

Alexander, P. A., Schallert, D. L., & Hare, V. C. (1991). Coming to Terms: How Researchers in Learning and Literacy Talk About Knowledge. *Review of educational research, 61*(3), 315-343. doi:doi:10.3102/00346543061003315

Andrews, T. M., Leonard, M. J., Colgrove, C. A., & Kalinowski, S. T. (2011). Active learning not associated with student learning in a random sample of college biology courses. *CBE Life Sci Educ, 10*(4), 394-405. doi:10.1187/cbe.11-07-0061

ASBMB. (2020). Core Concepts of Macromolecular Structure and Function. Retrieved from http://www.asbmb.org/education/teachingstrategies/foundationalconcepts/MacromolecularStructureFunction/

Barr, D. A., Gonzalez, M. E., & Wanat, S. F. (2008). The leaky pipeline: Factors associated with early decline in interest in premedical studies among underrepresented minority undergraduate students. *Acad Med, 83*(5), 503-511.

Becker, N., Noyes, K., & Cooper, M. (2016). Characterizing Students' Mechanistic Reasoning about London Dispersion Forces. *Journal of Chemical Education, 93*(10), 1713-1724. doi:10.1021/acs.jchemed.6b00298

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In *Metacognition: Knowing about knowing.* (pp. 185-205). Cambridge, MA, US: The MIT Press.

Bodner, G., Klobuchar, M., & Geelan, D. (2001). The Many Forms of Constructivism. *Journal of Chemical Education, 78*(8), 1107. doi:10.1021/ed078p1107.4

Boo, H. K., & Watson, J. (2001). Progression in high school students'(aged 16–18) conceptualizations about chemical reactions in solution. *Science education, 85*(5), 568-585.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*(1), 55-81.

Chen, X. (2013). *STEM attrition: College students' paths into and out of STEM fields*. Retrieved from Washington, D.C.:

Chi, M., & Glaser, R. (1985). Problem solving ability. R. Sternberg (Ed.), Human abilities: An information-processing approach (pp. 227-257). In: San Francisco: WH Freeman & Co.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science, 5*, 121-152.

Cooper, M. M., Williams, L. C., & Underwood, S. M. (2015). Student Understanding of Intermolecular Forces: A Multimodal Study. *Journal of Chemical Education, 92*(8), 1288-1298. doi:10.1021/acs.jchemed.5b00169

Council, N. R. (2012). *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*. Washington, DC: The National Academies Press.

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences, 116*(39), 19251-19257. doi:10.1073/pnas.1821936116

diSessa, A. A. (2018). *A Friendly Introduction to "Knowledge in Pieces": Modeling Types of Knowledge and Their Roles in Learning*.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proc Natl Acad Sci U S A, 111*(23), 8410-8415. doi:10.1073/pnas.1319030111

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and Transfer: A General Role for Analogical Reasoning. *Journal of Educational Psychology, 95*(2), 393-408.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving.

Gouvea, J. S., & Simon, M. R. (2018). Challenging Cognitive Construals: A Dynamic Alternative to Stable Misconceptions. *CBE—Life Sciences Education, 17*(2), ar34. doi:10.1187/cbe.17-10-0214

Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science, 332*(6034), 1213-1216. doi:10.1126/science.1204820

Halmo, S. M., Sensibaugh, C. A., Bhatia, K. S., Howell, A., Ferryanto, E. P., Choe, B., . . . Lemons, P. P. (2018). Student difficulties during structure–function problem solving. *Biochemistry and Molecular Biology Education, 46*(5), 453-463. doi:doi:10.1002/bmb.21166

Halmo, S. M., Sensibaugh, C. A., Reinhart, P., Stogniy, O., Fiorella, L., & Lemons, P. P. (in press). Advancing the Guidance Debate: Lessons from Educational Psychology and Implications for Biochemistry Learning. *CBE Life Sci Educ*.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99-107. doi:10.1080/00461520701263368

Kalyuga, S., & Singh, A.-M. (2016). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review, 28*(4), 831-852.

Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26*(3), 379-424.

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist, 51*(2), 289-299.

Kohn, K. P., Underwood, S. M., & Cooper, M. M. (2018). Energy Connections and Misconnections across Chemistry and Biology. *CBE-Life Sciences Education, 17*(1). doi:10.1187/cbe.17-08-0169

Lobato, J. (2006). Alternative perspectives on the transfer of learning: History, issues, and challenges for future research. *The Journal of the Learning Sciences, 15*(4), 431-449.

Lobato, J. (2012). The Actor-Oriented Transfer Perspective and Its Contributions to Educational Research and Practice. *Educational Psychologist, 47*(3), 232-247. doi:10.1080/00461520.2012.693353

Loertscher, J., Green, D., Lewis, J. E., Lin, S., & Minderhout, V. (2014). Identification of threshold concepts for biochemistry. *CBE-Life Sciences Education, 13*(3), 516-528.

Mayer, R. E. (2002). Rote Versus Meaningful Learning. *Theory Into Practice, 41*(4), 226-232. doi:10.1207/s15430421tip4104_4

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81.

National Academies of Sciences, E., & Medicine. (2018). *How People Learn II: Learners, Contexts, and Cultures*. Washington, DC: The National Academies Press.

National Research Council. (2000). *How people learn: Brain, mind, experience, and school (Expanded edition)*. Washington, D.C.: National Academies Press.

National Research Council. (2011). *Expanding Underrepresented Minority Participation: America's Science and Technology Talent at the Crossroads*. Retrieved from Committee on Underrepresented Groups and the Expansion of the Science and Engineering Workforce Pipeline, Committee on Science, Engineering, and Public Policy. Washington, D.C.:

National Research Council. (2012). *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering* (S. R. Singer, N. R. Nielsen, & H. A. Schweingruber Eds.). Washington, DC: The National Academies Press.

Nehm, R. H., & Ridgway, J. (2011). What Do Experts and Novices "See" in Evolutionary Problems? *Evolution: Education and Outreach, 4*(4), 666-679. doi:10.1007/s12052-011-0369-7

Newell, A., & Simon, H. A. (1972). *Human Problem Solving* (Vol. 104): Prentice-Hall Englewood Cliffs, NJ.

Novick, L. R., & Bassok, M. (2005). Problem solving. In K. Holyoak & B. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 321--349): Cambridge University Press.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1-4.

Paas, F., & van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning and Instruction, 16*(2), 87-91. doi:https://doi.org/10.1016/j.learninstruc.2006.02.004

Schmidt, R. A., & Bjork, R. A. (1992). New Conceptualizations of Practice: Common Principles in Three Paradigms Suggest New Concepts for Training. *Psychological Science, 3*(4), 207-218. doi:10.1111/j.1467-9280.1992.tb00029.x

Schwartz, D. L., & Bransford, J. D. (1998). A Time For Telling. *Cognition and Instruction, 16*(4), 475-5223. doi:10.1207/s1532690xci1604_4

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology, 103*(4), 759-775. doi:10.1037/a0025140

Smith, M. U. (1992). Expertise and the organization of knowledge: Unexpected differences among genetic counselors, faculty, and students on problem categorization tasks. *Journal of Research in Science Teaching, 29*(2), 179-205.

Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., . . . Young, A. M. (2018). Anatomy of STEM teaching in North American universities. *Science, 359*(6383), 1468-1470. doi:10.1126/science.aap8892

Sweller, J. (2016). Working memory, long-term memory, and instructional design. *Journal of Applied Research in Memory and Cognition, 5*(4), 360-367.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296.

Tricot, A., & Sweller, J. (2014). Domain-Specific Knowledge and Why Teaching Generic Skills Does Not Work. *Educational Psychology Review, 26*(2), 265-283. doi:10.1007/s10648-013-9243-1

Villafañe, S. M., Bailey, C. P., Loertscher, J., Minderhout, V., & Lewis, J. E. (2011). Development and analysis of an instrument to assess student understanding of foundational concepts before biochemistry coursework. *Biochemistry and Molecular Biology Education, 39*(2), 102-109.

Vosniadou, S., & Brewer, W. F. (1989). The concept of the earth's shape: A study of conceptual change in childhood. *Center for the Study of Reading Technical Report; no. 467*.

Williams, L. C., Underwood, S. M., Klymkowsky, M. W., & Cooper, M. M. (2015). Are noncovalent interactions an Achilles heel in chemistry education? A comparison of instructional approaches. *Journal of Chemical Education, 92*(12), 1979-1987.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of child psychology and psychiatry, 17*(2), 89-100.

Xu, X., Lewis, J. E., Loertscher, J., Minderhout, V., & Tienson, H. L. (2017). Small Changes: Using Assessment to Direct Instructional Practices in Large-Enrollment Biochemistry Courses. *CBE Life Sci Educ, 16*(1). doi:10.1187/cbe.16-06-0191

CHAPTER 6

STUDENT DIFFICULTIES DURING STRUCTURE-FUNCTION PROBLEM SOLVING[1]

[1]Halmo, S.M., Sensibaugh, C.A, Bhatia, K.S., Howell, A., Ferryanto, E.P., Choe, B., Kehoe, K.P., Watson, M. and Lemons, P.P. 2018. *Biochem Mol Bio Educ*. 46: 453–463.

Abstract

Protein structure-function is a key concept in biochemistry. We used the perspective of domain-specific problem solving to investigate students' solutions to a well-defined protein structure-function problem. We conducted think-aloud interviews with thirteen undergraduate students and performed qualitative analysis to examine the differences in the domain-general and domain-specific knowledge among correct and incorrect solutions. Our work revealed that students used domain-general and domain-specific knowledge in their problem solving. We also identified difficulties for students with the amino acid backbone, amino acid categorization, and causal mechanisms of noncovalent interactions. Using the identified difficulties, we make recommendations for the design of instructional materials targeted to improve protein structure-function problem solving in the biochemistry classroom.

Introduction

Structure and function, which refers to the phenomenon that the shape of a structure determines many of its properties (National Research Council 2011), is widely recognized as a core concept in biochemistry (Loertscher, J., Green, D., et al. 2014, Rowland, S.L., Smith, C.A., et al. 2011, Tansey, J.T., Baird, T., Jr., et al. 2013, Wright, A., Provost, J., et al. 2013). Additionally, structure and function is one of five core concepts for biological literacy (American Association for the Advancement of Science 2011) and one of seven cross-cutting concepts in the Next Generation Science Standards (Laverty, J., Underwood, S., et al. 2016, National Research Council 2014).

Students must integrate knowledge to solve problems about biochemical structure and function (Anderson, T.R. and Schonborn, K.J. 2008). This knowledge concerns two threshold concepts for biochemistry, the *physical basis of interactions* and *thermodynamics of macromolecular structure formation* (Loertscher, J., Green, D., et al. 2014). The *physical basis of interactions* consists of knowledge that full, partial, and/or momentary charges are the causal mechanistic basis for electrostatic properties, and that those properties make noncovalent interactions possible (Becker, N., Noyes, K., et al. 2016, Cooper, M.M., Corley, L.M., et al. 2013, Loertscher, J., Green, D., et al. 2014). Undergraduates often can state the definitions of noncovalent interactions but cannot expand on the definitions with explanations (Loertscher, J., Green, D., et al. 2014). Additionally, undergraduates lack causal mechanistic reasoning to explain the attraction between two nonpolar atoms (Becker, N., Noyes, K., et al. 2016, Cooper, M.M., Corley, L.M., et al. 2013). The *thermodynamics of macromolecular structure formation* includes knowledge that biochemical phenomena occur in aqueous environments and that entropy and enthalpy both contribute to structure (Loertscher, J., Green, D., et al. 2014).

When students integrate these pieces of knowledge, they should be able to accomplish structure-function-related learning goals set by ASBMB, such as determining the impact of chemical modification on protein structure, evaluating the contributions to specificity of non-covalent interactions in a ligand-macromolecule complex, or predicting the effects of mutation on the structure and activity of a protein (ASBMB). To assess student growth in structure-function-related learning goals, educators can use well-defined problems. Well-defined problems require a decision-making process where the path to a solution is uncertain (Martinez, M.E. 1998). In contrast to ill-defined problems, which have myriad solutions, well-defined problems have a small number of possible solutions. In order to investigate well-defined problem solving about protein structure-function, it is useful to consider findings from problem-solving research.

Problem-solving research shows that the most important factor in problem solving is one's knowledge base (Chase, W.G. and Simon, H.A. 1973), which includes: (1) domain-specific knowledge, or knowledge about a specific field of study (e.g., biochemistry) (Alexander, P.A., Schallert, D.L., et al. 1991); (2) domain-general knowledge, or knowledge that can be applied across domains (Chase, W.G. and Simon, H.A. 1973). For example, knowing that alanine is a nonpolar amino acid is biochemistry-specific knowledge, while knowing visual representations should be used during problem solving is domain-general knowledge. Experts in a particular domain (e.g., biochemistry) possess a well-organized knowledge base (Chase, W.G. and Simon, H.A. 1973) that consists of schemas, which are mental constructs connecting related ideas into manageable chunks (Marshall, S.P. 1995). Schemas allow experts to recognize problems as belonging to a previously learned category and specify the steps that are appropriate for solving the problem (Chase, W.G. and Simon, H.A. 1973, Renkl, A. and Atkinson, R.K. 2003, Sweller, J. and Cooper, G.A. 1985). In contrast, novice students struggle with problem solving because they

have yet to build relevant schemas. When they solve problems, they tend to focus on superficial features and access knowledge in pieces that are not well connected (diSessa, A.A. 2008, Nehm, R.H. and Ridgway, J. 2011, Renkl, A. and Atkinson, R.K. 2003, Smith, M.U. 1991, Smith, M.U. 1992). Researchers recognize that while domain-specific knowledge is an important factor in problem solving, domain-general processes can be helpful too, particularly for novices who are actively building schemas (Chi, M. and Glaser, R. 1985, Newell, A. and Simon, H.A. 1972, Pressley, M., Borkowski, J.G., et al. 1987, Smith, M.U. 1991). Domain-general processes like brainstorming (Halpern, D.F. 1997, Runco, M.A. and Chand, I. 1995), and working backwards, which is beginning with the problem goal and working in reverse toward the initial problem state (Chi, M. and Glaser, R. 1985, Newell, A., Shaw, J.C., et al. 1958), can enable a problem solver to organize the problem-solving process when domain-specific knowledge is lacking. Taken together, these findings form the basis of a conceptual framework known as domain-specific problem solving.

Domain-specific problem solving has been used to identify student difficulties from multiple science disciplines. For example, undergraduate physics students struggled with the principle of vector addition of forces in textbook physics problems (Larkin, J.H. 1981). In chemistry, undergraduates were unable to apply the laws of thermodynamics to typical chemical equilibrium problems (Camacho, M. and Good, R. 1989). Additionally, the most difficult aspect for undergraduate genetics students was applying probability concepts to classical genetics problems (Smith, M.U. and Good, R. 1984). Each of these studies compared novice problem solving with that of more expert (e.g., advanced students or PhDs) participants.

To date, domain-specific problem solving has not been emphasized in biochemistry education research. Instead, research on student difficulties primarily has used the framework of

conceptual and reasoning difficulties. This body of work has revealed students' problems with visual representations (e.g., antibody-antigen interactions) (Schönborn, K.J. and Anderson, T.R. 2009, Schönborn, K.J., Anderson, T.R., et al. 2002), drawing and explaining the role of hydrogen bonds in protein secondary structure (Harle, M. and Towns, M.H. 2013), the use of analogies (Orgill, M. and Bodner, G. 2007), enzyme-substrate interactions (Linenberger, K.J. and Bretz, S.L. 2014), and others (Grayson, D.J., Anderson, T.R., et al. 2001, Miller, K. and Kim, T. 2017).

Here we took the approach of domain-specific problem solving because it enables identification of students' biochemistry-specific difficulties as well as difficulties that may be more generally applied across science. It also encourages comparisons among novices at different stages and experts. The National Research Council has called for this type of research (National Research Council 2012). Specifically, we investigated the research question: What are the domain-specific and domain-general difficulties for students when solving a protein structure-function problem?

Methods

Participants & Context. Students were recruited in fall 2014 from an introductory biology course and an introductory biochemistry course at a large public research university in southeastern United States. The introductory biology course is a prerequisite for upper-level biology courses and consists of mostly first and second year students. The introductory biochemistry course consists of mostly third year students because introductory biology and a semester of organic chemistry are prerequisites to enroll. The introductory biology course is the first in a two-semester introductory biology sequence for life sciences majors and focuses on biomolecules, cells, molecular biology, genetics, and metabolism. The introductory biochemistry course is a one-semester course and focuses on proteins, enzymes, carbohydrate and lipid metabolism, and molecular biology. Both the introductory biology and introductory biochemistry courses are large-

enrollment courses with approximately 300 and 200 students per section respectively at the time of recruitment. We refer to participants from the introductory biology course as beginning students and participants from the introductory biochemistry course as advanced students. Both beginning and advanced students were recruited to provide a range of relevant knowledge among participants.

Students were recruited to volunteer for the study through an in-class announcement during the tenth week of the fifteen-week semester. We sought twenty participants, and twenty students agreed to participate. However, after the first five interviews we shortened the problem set (see *Supplemental Material*), and two students did not show up for the interview. In total, thirteen students participated, including six females and seven males. The thirteen students were pursuing the following majors: agricultural engineering (n=1), biology (n=4), biology with psychology (n=1), biochemistry (n=2), economics (n=1), health promotion (n=2), nutrition science (n=1), and pharmaceutical sciences (n=1). Students earned the following final grades in introductory biology or biochemistry: A (n=6), B (n=6), and C (n=1). A monetary incentive of $20 was given to students who participated in the study.

For an expert comparison to the novice participants, we recruited experts (n=8) at a summer 2015 national biochemistry education workshop, including PhDs in biology (n=2), biochemistry (n=3), and chemistry (n=3). They did not receive an incentive for participation. Expert participants included four females and four males.

This study was approved by the UGA IRB under exempt status (STUDY00000660).

Data Collection. Problem Development. The problem we designed, hereafter referred to as the Protein X problem (Figure 6.1), prompts critical elements of biochemistry problem solving. In particular, it prompts students to make predictions about protein structure and function based upon

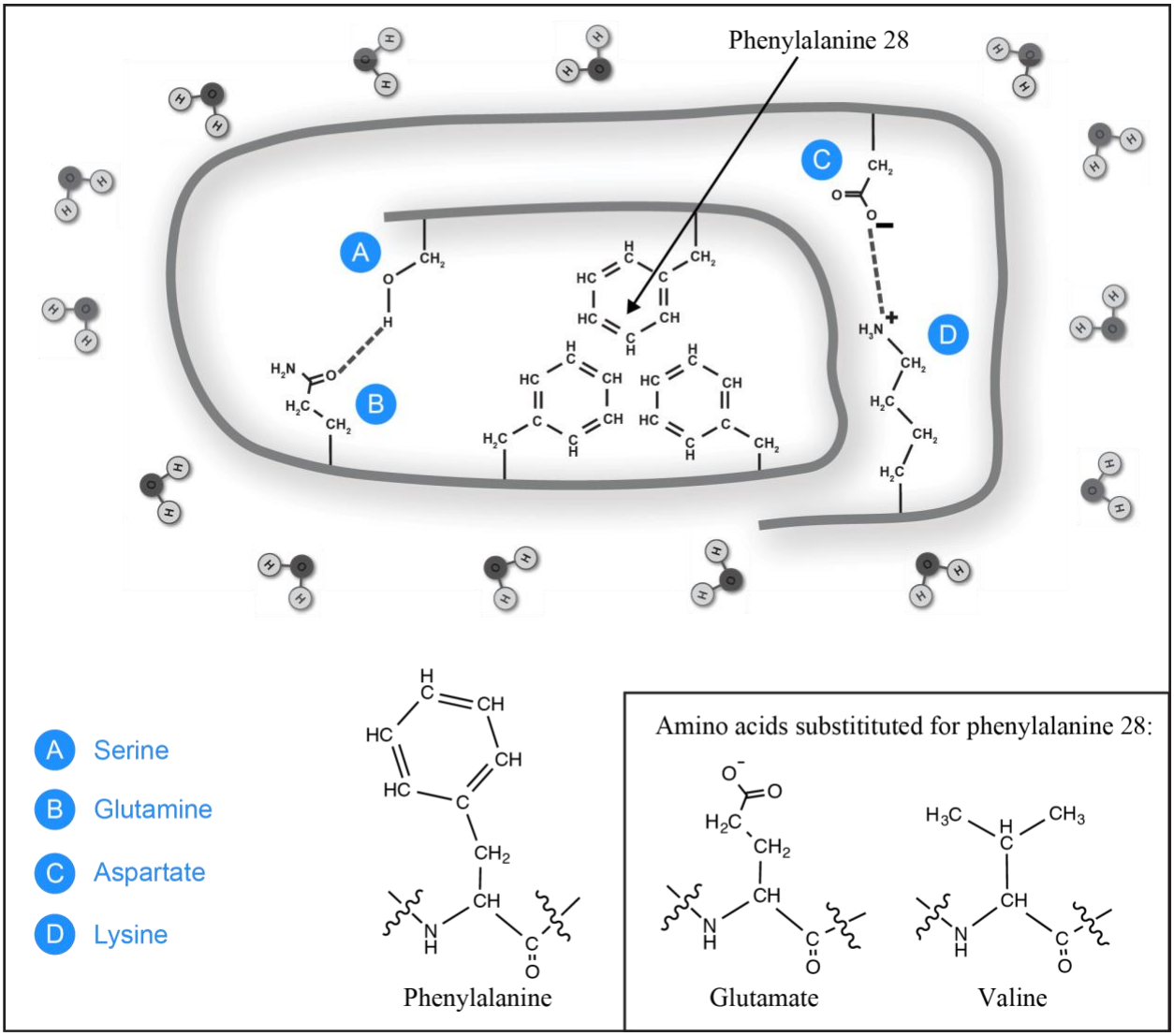


Figure 6.1. Protein X Problem

This problem was developed to assess student knowledge about protein structure-function. The version provided to participants in this study did not include the amino acid names or labels seen in blue. The prompt was as follows: *Protein X is a protein that functions in the cytoplasm of the cell. A model of normal protein X is shown below. The gray line represents the amino acid sequence folded into its tertiary structure; some but not all of the amino acid side chains are shown. Normal protein X has the amino acid phenylalanine at position 28 as labeled. Researchers studied variants of protein X with phenylalanine 28 substituted for other amino acids. For each of the amino acid substitutions shown in the box, predict its impact on the following: A. the folding of protein X. Justify your prediction. B. the function of protein X. Justify your prediction.*

their knowledge of noncovalent interactions and thermodynamics of macromolecular structure formation (Loertscher, J., Green, D., et al. 2014). We designed the Protein X problem to be solvable by beginning and advanced students. In the design of the Protein X problem, we developed four preliminary versions and collected feedback via discussions and think-aloud interviews with fifteen beginning or advanced students and 36 experts who have PhDs in biology, biochemistry, or chemistry (see *Supplemental Material*). For the expert comparison, we used the Protein X problem in think-aloud interviews with experts. Seven out of eight solved the Protein X problem correctly, and the only expert who was incorrect missed only one of four predictions. Additionally, during Protein X problem solving, experts used their knowledge of noncovalent interactions and thermodynamics of macromolecular structure, as we intended.

Think-aloud Interviews. We collected data on participants' problem-solving knowledge using a think-aloud interview protocol, which reveals cognitive processes that are not revealed in written answers alone and limits unnecessary interruption by the interviewer (Ericsson, K.A. and Simon, H.A. 1980). We modeled our protocol from other studies (Cooper, M.M., Corley, L.M., et al. 2013, Keys, C.W. 2000). In order to familiarize participants with the think-aloud approach, at the start of each interview we described the think-aloud protocol to participants and allowed them to practice using a problem requiring interpretation of a simple line graph. Then we asked participants to think aloud while solving the Protein X problem. We told interviewees to take as much time as they needed to respond. As in a typical think-aloud interview, the interviewer regularly prompted participants to "keep telling me what you're thinking." Additionally, as in the study by Cooper and colleagues (Cooper, M.M., Corley, L.M., et al. 2013), in order to reveal additional participant thinking, the interviewers occasionally prompted the participants with statements such as, "Can you tell us more about that?" or "What do you mean by [scientific term]?"

We did not intentionally provide information during prompting. We only asked participants to clarify their meaning. We video recorded all interviews so we could see the parts of the problem participants were pointing to while they were talking. Based on the video recordings, we were able to clarify what participants meant by "this" or "that" or "here," an approach taken by others (J., L.K. and Lowery, B.S. 2015). We directly transcribed the interviews into text format and imported them to MaxQDA software (Version 12, VERBI GmbH, Berlin, Germany). We assigned pseudonyms to participants. Additionally, we kept identifiers such as course level and class standing hidden until the end of data analysis in order to avoid bias.

Data Analysis. Problem Scoring. To have a correct solution to the Protein X problem, students correctly predicted the impact of both mutations on protein folding and function (Figure 6.1). Based on our biochemical knowledge and experts' solutions, a correct solution contained the following conclusions: glutamate mutant will fold differently, valine mutant will fold similarly, glutamate mutant will function differently, and valine mutant will function similarly (underlined text in Figure 6.2). If a student made a correct prediction but by the end changed his or her answer to an incorrect idea, we counted the solution as incorrect.

Qualitative Content Analysis. We conducted qualitative content analysis to determine the types of knowledge used by students and experts while solving the Protein X problem. First, a subset of authors developed a codebook that included content knowledge and procedures necessary for solving the problem. Second, the same set of authors individually applied these codes to the transcripts in a deductive manner and inductively created new codes as needed during the coding process. Coders met to discuss new codes and negotiate any disagreements. These steps of independent and group coding were iterative. At the end of coding, each coded transcript was reviewed and checked by all authors. Third, all authors (except Howell) organized the codes into

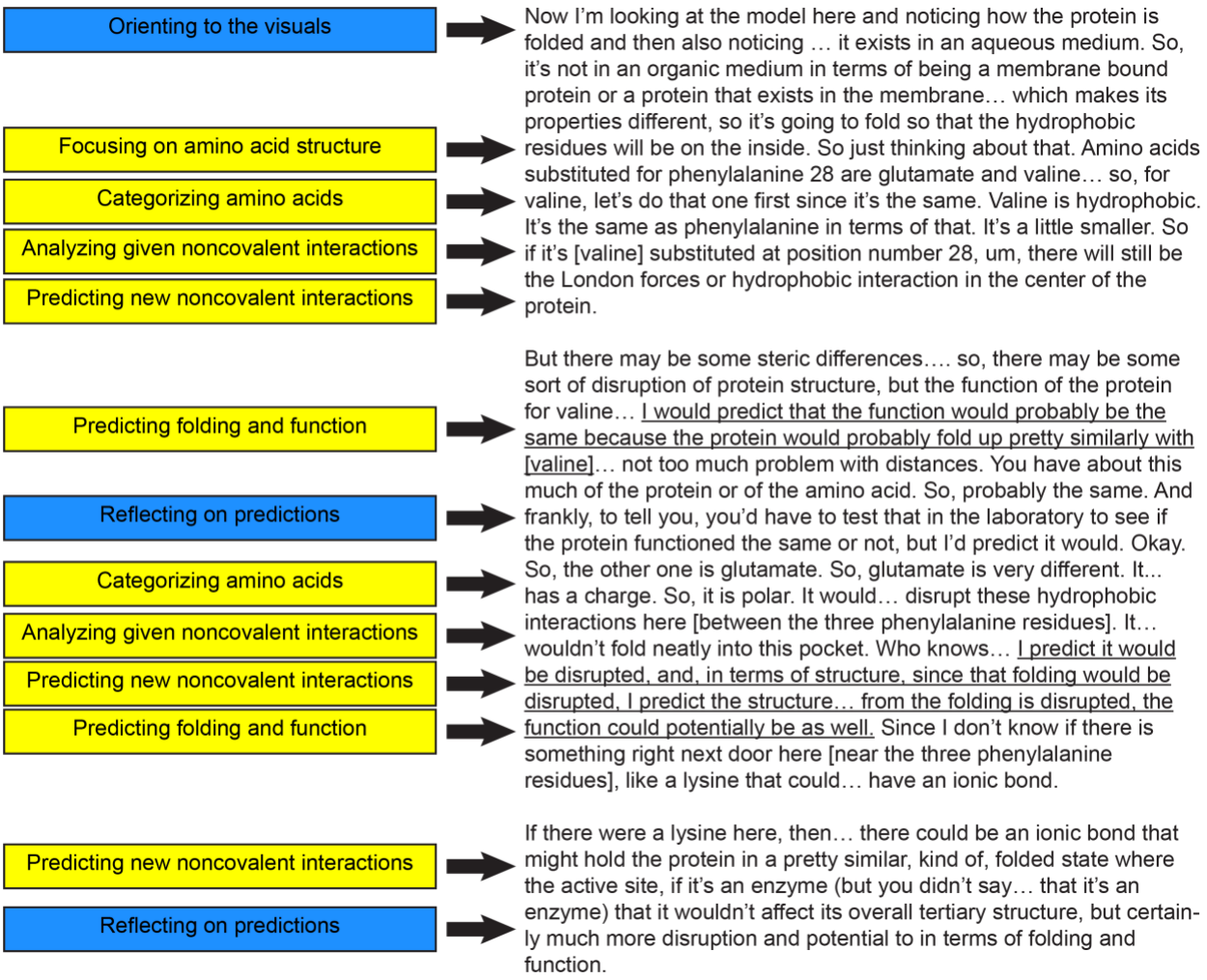


Figure 6.2. An Expert Solution to the Protein X Problem
Excerpt from expert Pamela's think-aloud interview. Correct predictions are underlined. Problem-solving steps are indicated to the left of the text with domain-general steps in blue and domain-specific steps in yellow.

themes (Saldana, J. 2015): domain-specific and domain-general knowledge (Alexander, P.A. and Judy, J.E. 1988). This process consisted of reasoning which codes represented parts of the interview where problem solvers were (1) stating facts, using a process or deciding when and where to apply knowledge tied to the domain of biochemistry (i.e., domain-specific) or (2) using a process or deciding when and where to apply knowledge that can be applied across domains (i.e., domain-general). For example, we reasoned that the code *identifying a hydrogen bond* is tied to the domain of biochemistry, so it is domain-specific. In contrast, we reasoned that the code *expressing tentativeness* represents a process that can be applied to all scientific problem-solving, not just biochemistry problem solving, so we categorized it as domain-general. Fourth and finally, a subset of authors assigned each code to a problem-solving step based on the typical process participants used to solve the problem. Thus, at the end of analysis, we had determined the domain-specific and domain-general problem-solving steps taken by each participant.

Results

Here, we describe our findings that address the research question: What are the domain-specific and domain-general difficulties for students when solving a protein structure-function problem? We provide evidence of: (1) domain-general and domain-specific knowledge when solving the Protein X problem; (2) the domain-specific and domain-general difficulties students encountered.

The Protein X Problem Elicits Domain-General and Domain-Specific Knowledge. We found during analysis that participants used the following steps, not always in this exact order, to solve the Protein-X problem:

1. orienting to the visuals
2. focusing on amino acid structure

3. categorizing amino acids
4. analyzing the noncovalent interactions indicated in normal Protein X
5. predicting potential new noncovalent interactions in Protein X variants
6. predicting the folding and function of each Protein X variant
7. reflecting on their predictions

Steps 1 and 7 are domain-general because they could be applied to many other types of problems in biochemistry as well as other domains. Steps 2-5 are domain-specific in that they pertain to biochemical structure-function problems. Step 6 is the solution step itself in response to the problem prompt, i.e. predicting folding and function for each amino acid substitution. Importantly, these steps emerged from qualitative content analysis; we did not predetermine them.

To illustrate these problem-solving steps in a concise format in context, we present an excerpt from an expert interview (Figure 6.2). As Figure 6.2 shows, this expert answered the problem correctly, predicting that the valine substitution would not impact protein structure or function and that the glutamate substitution would impact protein structure and function (Figure 6.2). This expert started and ended her solution with domain-general knowledge. She became familiar with the visuals at the beginning and reflected on her predictions at the end. In the middle, this expert used domain-specific knowledge, which provided the rationale for her predictions.

Seven students also solved the problem correctly. The remaining six did not. The correct solutions came from two beginning and five advanced students. The incorrect solutions came from three beginning and three advanced students. In the following two sections, we compare the domain-specific and domain-general problem-solving steps (Prevost, L.B. and Lemons, P.P. 2016) of correct and incorrect solutions in order to highlight student difficulties (Grayson, D.J., Anderson, T.R., et al. 2001). We illustrate our findings with quotes. To increase readability, we

use pseudonyms for students. The words in brackets within the quotes are based on video evidence, and ellipses indicate sections removed for clarity or brevity.

Domain-Specific Problem-Solving Steps of Correct and Incorrect Solutions. Focusing on Amino Acids (Step 2). In all correct and incorrect solutions, students explicitly "focused on the amino acid side chains*"* of phenylalanine, glutamate and/or valine (Table 6.1). However, two students that reached incorrect solutions also "focused on the amino acid backbone*"* (Table 6.1), thinking incorrectly that the backbones should be considered in the predictions about folding and function. For example, Robert spent several minutes working with the side chain structures of each amino acid in order to determine their polarity. During this part of the interview, he circled the amino acid side chains (Figure 6.3) and said the following:

> *This aspect [side chain] of the amino acid [valine] is nonpolar as well as this aspect [side chain] of the amino acid [phenylalanine]. So, I guess it doesn't really matter if it's a keto group or a benzo [sic] group. Um, but it does – it does matter the charge of the amino acid. And then this amino acid [glutamate] has a negative charge while these [valine and phenylalanine] don't, and they're [valine and phenylalanine] also nonpolar as well. So, basically, they are … similar, even though it – this [phenylalanine side chain] is a benzo [sic] group.*

However, after Robert worked with the side chains, he noticed that the amino acid backbones are identical to each other, and he circled them (Figure 6.3). This observation intrigued him and led him down an unproductive path:

> *But then, something that I find interesting is this part … each of these [side chains] are all attached to the same thing [amino acid backbone]. I just found that pretty interesting… if*

Table 6.1. Domain-Specific Problem-Solving Differences Between Correct and Incorrect Solutions

Domain-specific codes are presented in the general order of problem-solving steps used to solve the Protein X problem. Italicized codes indicate student difficulties. Code usage is depicted using shading and numbers. Saturated blue indicates all members of a group used a code. White indicates no members of a group used a code.

Domain-Specific Problem Solving Step	Code	Correct Solutions (n=7)	Incorrect Solutions (n=6)
Focusing on amino acids	Focusing on amino acid side chains	7	6
	Focusing on amino acid backbone	0	2
Categorizing amino acids	Categorizing at least one amino acid correctly	7	6
	Not recognizing the centrality of categorization	0	3
Analyzing given noncovalent interactions	Identifying hydrogen bond or ionic interaction	6	2
	Identifying interactions among Phe side chains	6	4
	Stating misconceptions about Phe interactions	4	4
Predicting new noncovalent interactions	Predicting nonpolars will interact with nonpolars	3	0
	Predicting Glu charge will need to be satisfied	7	4
	Predicting Glu will interact more strongly with Phe	0	4

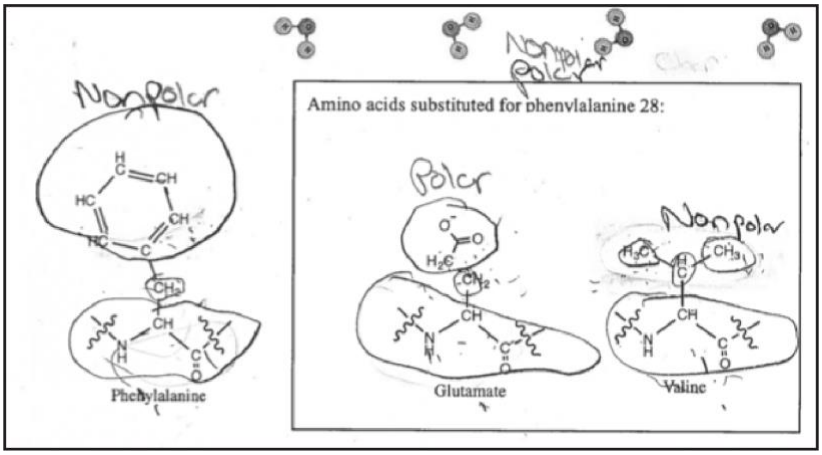


Figure 6.3. Student Depiction of Side Chains and Categorization of Amino Acids

As explained in the text, even though Robert initially categorized the amino acids based on the polarity of the amino acid side chains, he was ultimately distracted by the similarity of the amino acid backbones.

it's this part of the [amino acid backbone] that goes in the sequence, and if it [amino acid backbone] never changes, then the folding of the protein should never change.

Categorizing Amino Acids (Step 3). In all correct and incorrect solutions, students "categorized at least one amino acid correctly" (Table 6.1). Among correct solutions, students showed no trouble categorizing phenylalanine and valine as nonpolar/uncharged and glutamate as polar/charged, and they used this information in future problem-solving steps. In contrast, among incorrect solutions, half "did not recognize the centrality of categorization" (Table 6.1). That is, while these students did at least some categorization, they did not recognize that this was critical information for solving the problem. For example, in Nicole's incorrect solution she stated,

I can notice that this glutamate has a negative charge because of the oxygen and it looks like valine is kind of nonpolar, like it has no charges… I feel like if you substitute any type of amino acid or any type of protein with a different amino acid, it's automatically going to change the folding.

As illustrated by Nicole's quote, she accurately categorized glutamate and valine and recognized they are different, but she went on to solve the problem in an alternative way. She did not take amino acid categorization into account in her prediction of folding.

Analyzing Given Noncovalent Interactions (Step 4). Students with correct solutions readily "identified hydrogen bonding or ionic interaction" between serine and glutamine on the far left and aspartate and lysine on the far right, respectively (Table 6.1). Among incorrect solutions, only two students did this (Table 6.1). When it came to analyzing the interactions among the phenylalanine side chains, most of the students with correct as well as incorrect solutions "identified interactions between phenylalanine side chains," even if they did not name the interactions (Table 6.1). However, among both correct and incorrect solutions, students "stated

misconceptions about phenylalanine interactions" (Table 6.1). Daniel stated in his correct solution that nonpolar molecules do not participate in noncovalent interactions with each other. After examining the phenylalanine cluster, he said:

> *I think these are all phenylalanines, but they're not interacting with each other, so if this [phenylalanine 28] was changed to valine, the same thing would occur. These three would not be attracting.*

When prompted to explain why the phenylalanines are not interacting, Daniel responded with,

> *Because they're three nonpolar, so they don't have a charge. If they don't have a charge, they will not interact with each other. That's my reasoning, at least.*

As an additional example of the misconceptions we found, Emily in contrast to Daniel did draw potential interactions among the phenylalanines (Figure 6.4) however she then misidentified them:

> *I've got, like, three interactions there [three interactions drawn on phenylalanine cluster (Figure 6.4)] that I think would happen…*
>
> Interviewer: *Do you have a name for those interactions?*
>
> *I would call them hydrogen bonds… Because I think that the hydrogens would be the ones that were attracted to each other through hydrogen bonding.*

Predicting New Noncovalent Interactions (Step 5). Regarding the valine substitution, a few correct solutions "predicted that nonpolar amino acids will interact with other nonpolar amino acids" (Table 6.1) as Michael did in his correct solution,

> *If we were to substitute phenylalanine 28 with the valine, it could have the same… probably the same folding because this [valine] is nonpolar; these are nonpolar [phenylalanine side chains]. The nonpolar molecules would interact with each other and cause the van der*

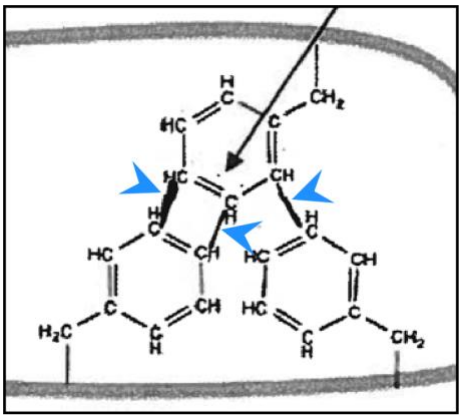


Figure 6.4. Student Depiction of Interactions Between Phenylalanine Side Chains
As described in the text, Emily depicted the interactions she described between phenylalanine residues (blue arrows) in the center of Protein X, which she later misidentified as hydrogen bonds.

Waal forces to interact.

In contrast, no incorrect solutions included this correct idea (Table 6.1).

Regarding the glutamate substitution, all correct solutions and four of the six incorrect solutions "predicted that glutamate's negative charge would need to be satisfied" (Table 6.1). That is, students knew that the full negative charge on glutamate would tend to be attracted to a full (or at least permanent partial) positive charge. For example, Tyler, whose solution was correct, stated:

The negative charge on the oxygen [of glutamate] will likely attract to positive charges elsewhere in… the amino acid sequence of the protein.

Interestingly, even though the Protein X problem depicts water molecules, only one correct solution discussed the idea that glutamate's charge could be satisfied through an interaction with water (see *Supplemental Material*). Finally, only among incorrect solutions did students incorrectly "predict that glutamate will interact more strongly with phenylalanine" (Table 6.1). As Ashley stated,

The 'O' negative would make it [glutamate] want to bond more to this [phenylalanine] anyway.

While it is true that an ion-induced dipole interaction is stronger than a van der Waals interaction, in the context of Protein X, a better prediction is that a glutamate substituted for phenylalanine 28 will be attracted to a positive charge somewhere else in the protein or in water.

Domain-General Problem-Solving Steps of Correct and Incorrect Solutions. We did not identify differences between correct and incorrect solutions for the domain-general step of orienting to the visuals (Step 1). However, we found important differences between correct and incorrect solutions for the domain-general step of reflecting on predictions (Step 7). First, we found that similar to the experts interviewed, students with correct solutions "expressed tentativeness

about folding/function" predictions (Table 6.2). Tentativeness shown by experts centered on the need for empirical data to support predictions, and they expressed this need confidently (see Figure 6.2*, "frankly, to tell you, you'd have to test that in the laboratory to see if the protein functioned the same or not.").* Students with correct solutions expressed a similar type of tentativeness, but they did so less emphatically. For example, Samantha said the following about her valine prediction,

> *Whereas valine, it might or might not have an effect on the function of protein X. Um 'cause it would have less effect on the folding than glutamate would, but I don't know how much of an effect it would have on the folding.*

Among incorrect solutions, only one student expressed tentativeness. In contrast to the prior examples, the student's tentativeness centered on whether she made a complete prediction:

> *I don't know. I'm kind of at a standstill, and I feel like I'm not sure where to go from there. But I know that the folding will differ, and I know that it'll change the function… because it's been substituted. So, at that point, I'm not really sure on… what to do. So, I'll probably just either go on to the next question or just maybe stare at the page and think of something random to pop in my head.*

The second important difference we found between correct and incorrect solutions was that no students with correct solutions changed their answers once they solved the problem. Yet among students with incorrect solutions, many flip-flopped their answers while solving. In two cases, the inconsistency contributed to solving the problem (i.e., "changing answer to a correct prediction") (Table 6.2). In all other cases, the students "contradicted themselves" (Table 6.2) by making conflicting statements, changing their answer to an incorrect solution, or second-guessing themselves without resolving the indecision. For example, Ashley contradicted herself by first

Table 6.2. Domain-General Problem-Solving Differences Between Correct and Incorrect Solutions

These domain-general codes occurred during the part of problem solving where participants reflected on predictions. Italicized code indicates a student difficulty. Code usage is depicted using shading and numbers. Saturated blue indicates all members of a group used a code. White indicates no members of a group used a code.

Domain-General Problem Solving Step	Code	Correct Solutions (n=7)	Incorrect Solutions (n=6)
Reflecting on predictions	Expressing tentativeness about folding/function	7	1
	Changing answer to a correct prediction	0	2
	Contradicting self	0	5

correctly stating that glutamate would have an effect on the function but then going on to state that glutamate would not change the function,

> *Trying to think if it [glutamate] would change the function, 'cause there's not much of a difference [in structure between glutamate and valine]. So, maybe it [glutamate] wouldn't change the function.*

Discussion

In this study, we identified domain-specific and domain-general problem-solving steps and associated difficulties for students solving a protein structure-function problem. Our findings lead to several suggestions for teaching and learning to improve student problem solving about protein structure-function.

Recommendations for Improving Domain-Specific Problem-Solving. Based on our findings, students would benefit from more instruction on protein structure-function problem solving. Specifically, students need instruction to facilitate learning that (1) amino acid side chains rather than the amino acid backbone drive tertiary structure interactions, (2) amino acid categories are not something only to be memorized, but rather provide the key to predicting the dynamics of protein structure-function, and (3) all noncovalent interactions, including those among nonpolar molecules, share the underlying causal mechanism of attraction due to charge.

To provide this instruction, problem-solving research suggests two potential approaches: worked examples (Renkl, A. 2014) and preparation for future learning (Kapur, M. 2008, Schwartz, D.L. and Martin, T. 2004). Worked examples give students detailed descriptions of the steps to solving a problem prior to asking them to solve problems independently. A strong body of evidence from domains like physics, mathematics, and statistics shows that worked examples promote better acquisition of problem-solving skill than problem solving alone (Paas, F., Renkl, A., et al. 2003,

Renkl, A. 2014, Renkl, A. and Atkinson, R.K. 2003). Another problem-solving approach is preparation-for-future-learning, which asks students to explore solutions through invention or generation prior to seeing worked examples (Kalyuga, S. and Singh, A.-M. 2016, Rittle-Johnson, B. and Star, J.R. 2007, Rittle-Johnson, B. and Star, J.R. 2009, Schwartz, D.L., Chase, C.C., et al. 2011, Star, J.R., Pollack, C., et al. 2014). Preparation-for-future-learning appears to be particularly advantageous in helping students transfer their knowledge to novel problems (Kapur, M. 2008, Kapur, M. 2011, Schwartz, D.L., Chase, C.C., et al. 2011, Schwartz, D.L. and Martin, T. 2004). In both approaches, students' should be prompted to explain their reasoning, because students learn the deep features of a problem type as they refine and expand their knowledge of a single problem (Chi, M.T.H., Bassok, M., et al. 1989).

Given both these evidence-based approaches from the literature, we envision instructional materials that explicitly draw students' attention to the difficulties we reported here, before or after problem exploration. For instance, worked examples could be designed that direct students to notice (1) amino acid side chains rather than the amino acid backbone are involved in tertiary structure interactions, (2) amino acid categorization is key to solving a protein structure-function problem, and (3) the causal mechanistic basis for noncovalent interactions. The worked examples may not need to include ionic interactions, because students appear to easily master this knowledge. Rather, worked examples should focus on hydrogen bonds and van der Waals interactions, as our data show that students' memorized definitions are not adequate for protein structure-function problem solving. This finding is consistent with the literature (Becker, N., Noyes, K., et al. 2016, Cooper, M.M., Corley, L.M., et al. 2013, Cooper, M.M., Williams, L.C., et al. 2015, Loertscher, J., Green, D., et al. 2014). Alternatively, preparation-for-future-learning activities could be designed that direct students' attention to the same problem features, yet give

them opportunities to activate their prior knowledge before being asked to solve a problem. Instructional materials also should attend to the visual cues provided in the problem to ensure that they align with best practices (Hegarty, M. 2011, Linenberger, K.J. and Bretz, S.L. 2014, Offerdahl, E.G., Arneson, J.B., et al. 2017, Schönborn, K.J. and Anderson, T.R. 2006). Future studies will investigate which approach is better for targeting the problem-solving difficulties we identified.

Recommendations for Improving Domain-General Problem-Solving. Our findings also illuminate ways to improve the teaching and learning of domain-general problem solving. A critical domain-general difficulty for students was that they contradicted themselves during problem solving. It appears that these students were unable to maintain knowledge of their early problem-solving steps during later phases of problem solving (Smith, M.U. 1991). We hypothesize that students who struggled with these foundational aspects of the problem do so because of a lack of well-developed schemas (Smith, M.U. 1991). One potential way to help students construct organized schemas is through conceptual curriculum design such as POGIL (Bailey, C.P., Minderhout, V., et al. 2012), Chemical Thinking (Sevian, H. and Talanquer, V. 2014, Talanquer, V. and Pollard, J. 2017), or CLUE Chemistry (Cooper, M. and Klymkowsky, M. 2013, Cooper, M.M. and Klymkowsky, M.W. 2013). Each of these approaches thoughtfully scaffolds disciplinary knowledge in a meaningful progression, which can contribute to dramatic improvements in conceptual understanding (Williams, L.C., Underwood, S.M., et al. 2015).

Additionally, we found that expert solutions to the Protein X problem include a degree of tentativeness surrounding protein folding and function predictions (Figure 6.2). If a goal of undergraduate biochemistry education is to develop critical scientific practices like explanation and argumentation, then generating tentativeness falls under this mandate. In order for

tentativeness to blend into the biochemistry curriculum, instructors must explicitly teach the kind of skepticism and need for empirical data for which scientists are known (Buck, Z.E., Lee, H.-S., et al. 2014). This scientific practice could be more explicitly incorporated into the Protein X problem by asking what data would be needed to test the protein folding and function predictions.

Limitations of Our Work. The methodology used to investigate our research question was not intended to generate quantitative data representing all undergraduate biology students; rather, it was intended to provide a rich description of problem solving within a manageable sample. Regardless, the study presented here is limited due to the small sample size. Our data show where students experienced difficulties in problem solving, but additional difficulties might be found in a larger sample. Also, our data suggest that beginning and advanced students have similar difficulties with protein structure-function problems. Yet our sample size may mask differences between beginning and advanced students. Thus, forthcoming studies should explore students' solutions from a larger sample. Finally, this work does not address the role of metacognition, motivation, or effort in problem solving nor consider the three-dimensional nature of protein structure-function relationships (Offerdahl, E.G., Arneson, J.B., et al. 2017, Schönborn, K.J. and Anderson, T.R. 2006), although these are important aspects to consider in future work.

Acknowledgments

This material is based on work supported by the National Science Foundation under grant DRL 1350345. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We thank the students and experts who participated in this study. We also thank the Biology Education Research Group (BERG) at the University of Georgia, and the reviewers who improved the quality of this work with critical feedback on the manuscript.

References

Alexander PA, Judy JE. 1988. The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research*, 58:375-404.

Alexander PA, Schallert DL, Hare VC. 1991. Coming to Terms: How Researchers in Learning and Literacy Talk About Knowledge. *Review of Educational Research*, 61:315-343.

American Association for the Advancement of Science. 2011. Vision and Change in Undergraduate Biology Education: A Call to Action.

Anderson TR, Schonborn KJ. 2008. Bridging the educational research-teaching practice gap: Conceptual understanding, part 1: The multifaceted nature of expert knowledge. *Biochem. Mol. Biol. Educ.*, 36:309-315.

ASBMB. Core Concepts of Macromolecular Structure and Function.

Bailey CP, Minderhout V, Loertscher J. 2012. Learning transferable skills in large lecture halls: Implementing a POGIL approach in biochemistry. *Biochem. Mol. Biol. Educ.*, 40:1-7.

Becker N, Noyes K, Cooper M. 2016. Characterizing Students' Mechanistic Reasoning about London Dispsersion Forces. *J. Chem. Educ.*, 93:1713-1724.

Buck ZE, Lee H-S, Flores J. 2014. I Am Sure There May Be a Planet There: Student articulation of uncertainty in argumentation tasks. *Intl. J. of Sci. Educ.*, 36:2391-2420.

Camacho M, Good R. 1989. Problem solving and chemical equilibrium: Successful versus unsuccessful performance. *J. Res. Sci. Teaching*, 26:251-272.

Chase WG, Simon HA. 1973. Perception in chess. *Cognitive Psychology*, 4:55-81.

Chi M, Glaser R. 1985. Problem solving ability. R. Sternberg (Ed.), Human abilities: An information-processing approach (pp. 227-257). San Francisco: WH Freeman & Co.

Chi MTH, Bassok M, Lewis MW, Reimann P, Glaser R. 1989. Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. *Cognitive Science*, 13:145-182.

Cooper M, Klymkowsky M. 2013. Chemistry, life, the universe, and everything: a new approach to general chemistry, and a model for curriculum reform. *J. Chem. Educ.*, 90:1116-1122.

Cooper MM, Corley LM, Underwood SM. 2013. An investigation of college chemistry students' understanding of structure–property relationships. *J. Res. Sci. Teaching*, 50:699-721.

Cooper MM, Klymkowsky MW. 2013. The trouble with chemical energy: why understanding bond energies requires an interdisciplinary systems approach. *CBE Life Sci. Educ.*, 12:306-312.

Cooper MM, Williams LC, Underwood SM. 2015. Student Understanding of Intermolecular Forces: A Multimodal Study. *J. Chem. Educ.*, 92:1288-1298.

diSessa AA. 2008. A bird's eye view of the "pieces" vs. "coherency" controversy (From the "pieces" side of the fence). p. 35-60.

Ericsson KA, Simon HA. 1980. Verbal reports as data. *Psychological review*, 87:215.

Grayson DJ, Anderson TR, Crossley LG. 2001. A four-level framework for identifying and classifying student conceptual and reasoning difficulties. *Intl. J. of Sci. Educ.*, 23:611-622.

Halpern DF. 1997. *Critical Thinking Across the Curriculum: A Brief Edition of Thought and Knowledge* Routledge:New York.

Harle M, Towns MH. 2013. Students' understanding of primary and secondary protein structure: Drawing secondary protein structure reveals student understanding better than simple recognition of structures. *Biochem. Mol. Biol. Educ.*, 41:369-376.

Hegarty M. 2011. The cognitive science of visual-spatial displays: implications for design. *Top Cogn Sci*, 3:446-474.

J. LK, Lowery BS. 2015. Biochemistry students' ideas about how an enzyme interacts with a substrate. *Biochem. Mol. Biol. Educ.*, 43:213-222.

Kalyuga S, Singh A-M. 2016. Rethinking the boundaries of cognitive load theory in complex learning. *Educ. Psychol. Rev.*, 28:831-852.

Kapur M. 2008. Productive failure. *Cognition and Instruction*, 26:379-424.

Kapur M. 2011. A further study of productive failure in mathematical problem solving: Unpacking the design components. *Instructional Science*, 39:561-579.

Keys CW. 2000. Investigating the Thinking Processes of Eighth Grade Writers during the Composition of a Scientific Laboratory Report. *J. Res. Sci. Teaching*, 37:676-690.

Larkin JH. 1981. Enriching formal knowledge: A model for learning to solve textbook physics problems. *Cognitive skills and their acquisition*:311-334.

Laverty J, Underwood S, Matz R, Posey L, Carmel J, Caballero M. 2016. Characterizing College Science Assessments: The Three-Dimensional Learning Assessment Protocol. *PLoS ONE*, 11.

Linenberger KJ, Bretz SL. 2014. Biochemistry students' ideas about shape and charge in enzyme–substrate interactions. *Biochem. Mol. Biol. Educ.*, 42:203-212.

Loertscher J, Green D, Lewis JE, Lin S, Minderhout V. 2014. Identification of threshold concepts for biochemistry. *CBE Life Sci. Educ.*, 13:516-528.

Marshall SP. 1995. *Schemas in Problem Solving* Cambridge University Press.

Martinez ME. 1998. What is problem solving? *Phi Delta Kappan*, 79:605-609.

Miller K, Kim T. 2017. Examining student heuristic usage in a hydrogen bonding assessment. *Biochem. Mol. Biol. Educ.*, 45:411-416.

National Research Council. 2011. A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Committee on a Conceptual Framework for New K-12 Science Education Standards, Board on Science Education, Division of Behavioral and Social Scienes and Education. Washington, D.C.

National Research Council. 2012. *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering* The National Academies Press:Washington, DC.

National Research Council. 2014. *Developing assessments for the Next Generation Science Standards* National Academies Press:Washington, D.C.

Nehm RH, Ridgway J. 2011. What Do Experts and Novices "See" in Evolutionary Problems? *Evolution: Education and Outreach*, 4:666-679.

Newell A, Shaw JC, Simon HA. 1958. Elements of a theory of human problem solving. *Psychological review*, 65:151.

Newell A, Simon HA. 1972. *Human Problem Solving* Prentice-Hall Englewood Cliffs, NJ.

Offerdahl EG, Arneson JB, Byrne N. 2017. Lighten the Load: Scaffolding Visual Literacy in Biochemistry and Molecular Biology. *CBE Life Sci. Educ.*, 16.

Orgill M, Bodner G. 2007. Locks and keys. *Biochem. Mol. Biol. Educ.*, 35:244-254.

Paas F, Renkl A, Sweller J. 2003. Cognitive load theory and instructional design: Recent developments. *Educ. Psychol.*, 38:1-4.

Pressley M, Borkowski JG, Schneider W. 1987. Cognitive strategies: Good strategy users coordinate metacognition and knowledge. *Annals of Child Development*: JAI Press Inc. p. 89-129.

Prevost LB, Lemons PP. 2016. Step by Step: Biology Undergraduates' Problem-Solving Procedures during Multiple-Choice Assessment. *CBE Life Sci. Educ.*, 15.

Renkl A. 2014. Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, 38:1-37.

Renkl A, Atkinson RK. 2003. Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educ. Psychol.*, 38:15-22.

Rittle-Johnson B, Star JR. 2007. Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, 99:561-574.

Rittle-Johnson B, Star JR. 2009. Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology*, 101:529-544.

Rowland SL, Smith CA, Gillam EM, Wright T. 2011. The concept lens diagram: a new mechanism for presenting biochemistry content in terms of "big ideas". *Biochem. Mol. Biol. Educ.*, 39:267-279.

Runco MA, Chand I. 1995. Cognition and creativity. *Educ. Psychol. Rev.*, 7:243-267.

Saldana J. 2015. *The Coding Manual for Qualitative Researchers* SAGE.

Schönborn KJ, Anderson TR. 2006. The importance of visual literacy in the education of biochemists. *Biochem. Mol. Biol. Educ.*, 34:94-102.

Schönborn KJ, Anderson TR. 2009. A model of factors determining students' ability to interpret external representations in biochemistry. *Intl. J. of Sci. Educ.*, 31:193-232.

Schönborn KJ, Anderson TR, Grayson DJ. 2002. Student difficulties with the interpretation of a textbook diagram of immunoglobulin G (IgG)*. *Biochem. Mol. Biol. Educ.*, 30:93-97.

Schwartz DL, Chase CC, Oppezzo MA, Chin DB. 2011. Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, 103:759-775.

Schwartz DL, Martin T. 2004. Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, 22:129-184.

Sevian H, Talanquer V. 2014. Rethinking chemistry: a learning progression on chemical thinking. *Chemistry Education Research and Practice*, 15:10-23.

Smith MU. 1991. A view from biology. *Toward a Unified Theory of Problem Solving: Views from the Content Domains*. New York: Routledge. p. 1-20.

Smith MU. 1992. Expertise and the organization of knowledge: Unexpected differences among genetic counselors, faculty, and students on problem categorization tasks. *J. Res. Sci. Teaching*, 29:179-205.

Smith MU, Good R. 1984. Problem Solving and Classical Genetics: Successful Versus Unsuccessful Performance. *J. Res. Sci. Teaching*, 21:895-912.

Star JR, Pollack C, Durkin K, Rittle-Johnson B, Lynch K, Newton K, Gogolen C. 2014. Learning from comparison in algebra. *Contemporary Educational Psychology*, 40:41-54.

Sweller J, Cooper GA. 1985. The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and instruction*, 2:59-89.

Talanquer V, Pollard J. 2017. Reforming a large foundational course: Successes and challenges. *J. Chem. Educ.*

Tansey JT, Baird T, Jr., Cox MM, Fox KM, Knight J, Sears D, Bell E. 2013. Foundational concepts and underlying theories for majors in "biochemistry and molecular biology". *Biochem. Mol. Biol. Educ.*, 41:289-296.

Williams LC, Underwood SM, Klymkowsky MW, Cooper MM. 2015. Are noncovalent interactions an Achilles heel in chemistry education? A comparison of instructional approaches. *J. Chem. Educ.*, 92:1979-1987.

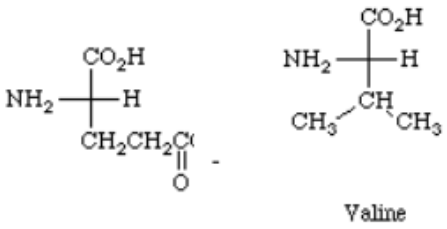
Wright A, Provost J, Roecklein-Canfield JA, Bell E. 2013. Essential concepts and underlying theories from physics, chemistry, and mathematics for "biochemistry and molecular biology" majors. *Biochem. Mol. Biol. Educ.*, 41:302-308.

Supplemental Material

1. Sickle cell anemia is characterized by an altered protein structure of hemoglobin in blood. This disease occurs when the amino acid, glutamate, at position 6 on the polypeptide chain is replaced with valine. Because of this mutation, red blood cells change from their normal disc shape to a sickled shape.

a. Given the structures of glutamate and valine (Figure 1 below), which tertiary structure interactions (hydrogen bonding, hydrophobic effect, van der Waals interactions, covalent bonding, or electrostatic interactions) do you think contribute the most to hemoglobin in normal cells? Explain.

b. How do these interactions change in the sickled cell?



Glutamate

Valine

Figure 1

c. Researchers wanted to test a variety of amino acids to find out which one would be less detrimental than valine to the hemoglobin structure. They tested three point mutations with three different amino acids (other than valine) substituted for glutamate at position 6. They observed the percentage of sickled cells in the blood with each point mutation. Given the data that is presented in Figure 2 below, which amino acid, 1, 2, or 3, do you think has a structure that is most similar to glutamate?

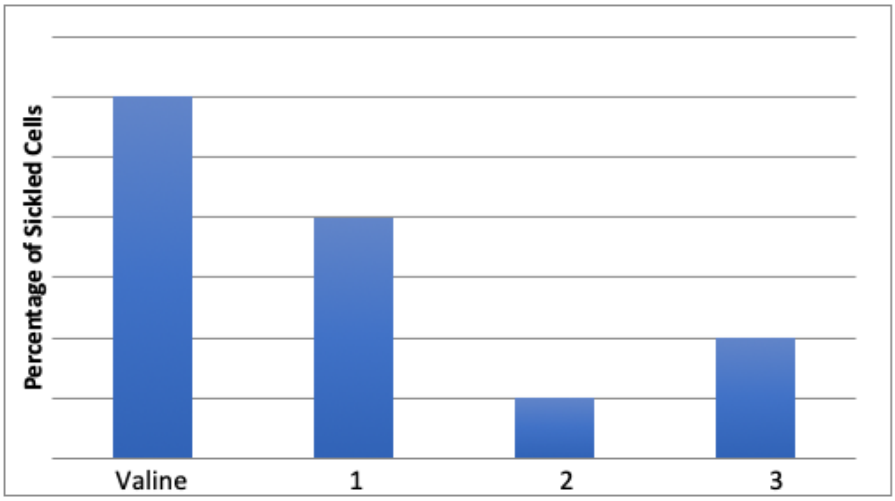
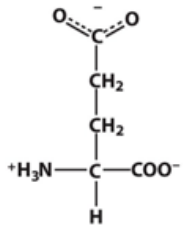


Figure 2

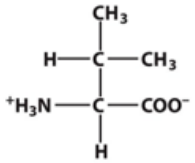
Supplemental Material Figure S6.1. Problem Development Version 1 used in think aloud interviews, n=10 students

1. Sickle cell anemia is characterized by an altered protein structure of hemoglobin in blood. This disease occurs when the amino acid glutamate at position six on the polypeptide chain is replaced with valine. Because of this mutation, red blood cells change from their normal disc shape to a sickled shape.

A. Tertiary structure interactions can be broken into covalent and non-covalent interactions. Non-covalent interactions include the following electrostatic interactions: hydrogen bonding, van der Waals interactions, and ion-ion interactions. Given the structures of glutamate and valine shown below, which tertiary structure interactions do you think contribute the most to hemoglobin in normal cells? Choose from covalent, hydrogen bonding, van der Waals interactions, or ion-ion interactions. Explain the rationale for your selection.



Glutamate

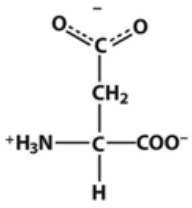
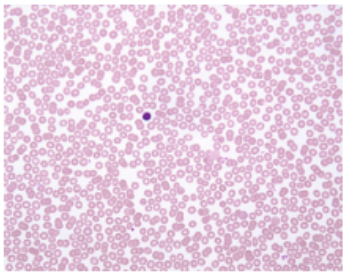
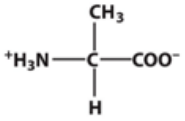
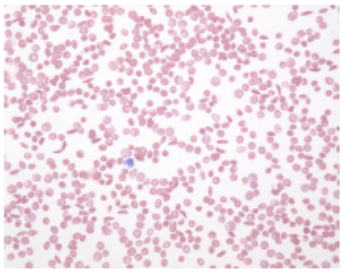


Valine

B. How do these tertiary structure interactions change in the sickled cell?

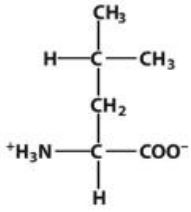
C. Researchers tested a variety of amino acid substitutions at position six of hemoglobin to find out if they also cause red blood cells to sickle. Their results are shown below. (1) Describe the results for both panels. (2) Based on your knowledge of amino acid structure and tertiary structure interactions, provide an explanation for these results.

Table 1. Results from sickle cell test performed on blood from individuals with two different amino acid substitutions and position six of hemoglobin.

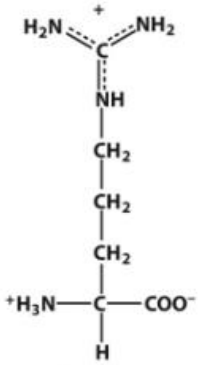
Glutamic acid at position six was substituted for:	Blood smear of red blood cells from individuals with this amino acid substitution
 Aspartate	
 Alanine	

Supplemental Material Figure S6.2. Problem Development Version 2 used in focus groups, n=22 experts. Rationale: Student interviews with Version 1 showed that the graph in Version 1 was too easy. It did not prompt for careful analysis. Thus, we modified part C, along with some of the language of other parts to create Version 2.

1. Imagine a protein that contains the amino acid leucine. If leucine was changed to arginine, would the shape of the protein change? The structure of leucine and arginine are shown below.

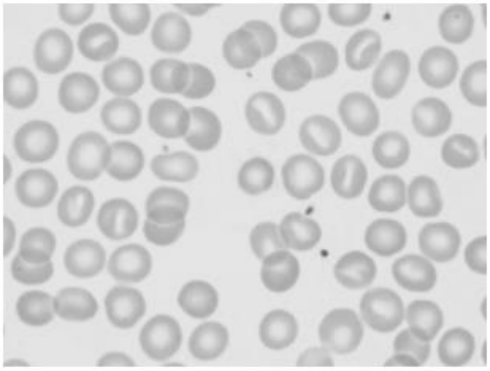


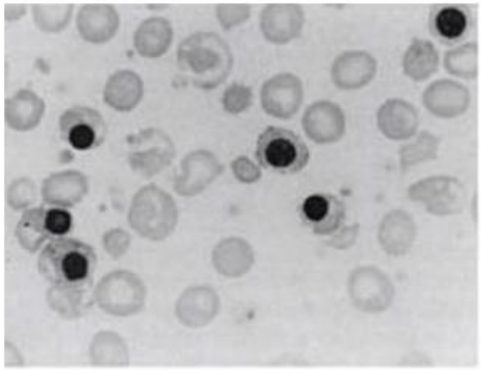
Leucine



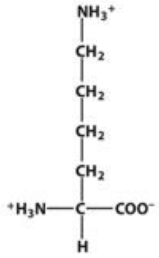
Arginine

2. Hemoglobin is a protein in red blood cells that carries oxygen. Changes in particular amino acids in hemoglobin can lead to disease. One type of disease that can happen is called inclusion body disease. Normal red blood cells and inclusion body red blood cells are shown below (normal on the left and inclusion body on the right).

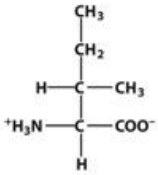




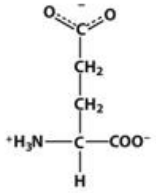
The amino acid that is critical in inclusion body disease is leucine 28. In some cases, but not all, if leucine 28 is changed to another amino acid, inclusion body disease results. Of the amino acids shown below, which one(s) do you predict would lead to inclusion body disease? Justify your prediction.



Lysine



Isoleucine



Glutamate

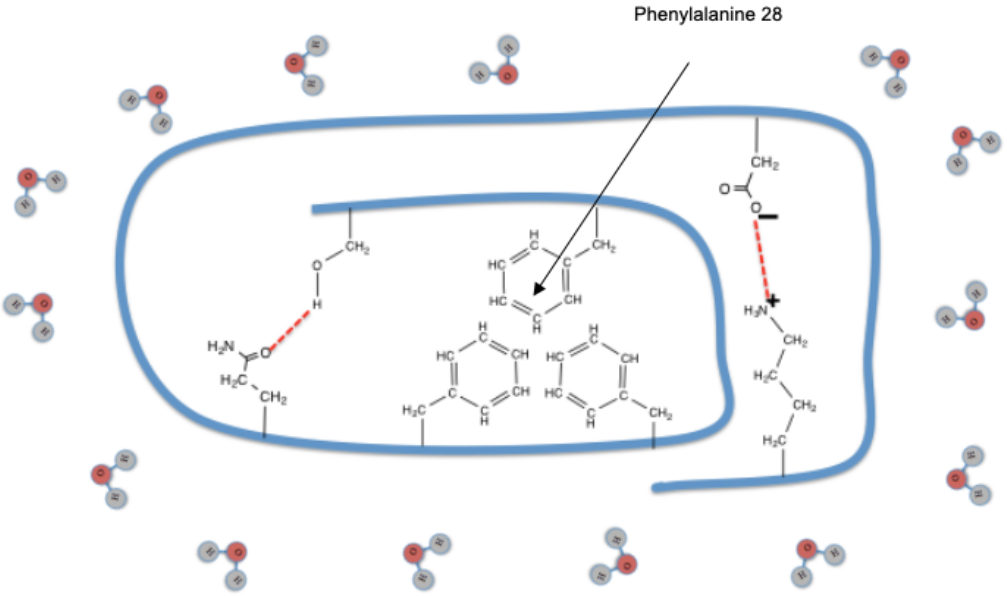
Supplemental Material Figure S6.3. Problem Development Version 3 used in two focus groups, n=6 experts. Rationale: Experts who reviewed Version 2 worried that students would be distracted by their prior knowledge of sickle cell anemia and focus on the wrong parts of the problem. Thus, we created Version 3 to focus on a different disease-causing mutation in hemoglobin, and we modified the rest of the question to match.

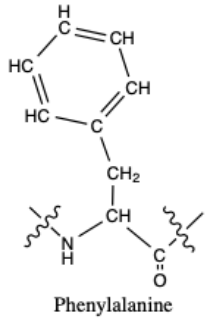
Protein X is a protein that functions in the cytoplasm of the cell. A model of normal protein X is shown below. The gray line represents the amino acid sequence folded into its tertiary structure; some but not all of the amino acid side chains are shown. Normal protein X has the amino acid phenylalanine at position 28 as labeled. Researchers studied variants of protein X with phenylalanine 28 substituted for other amino acids.

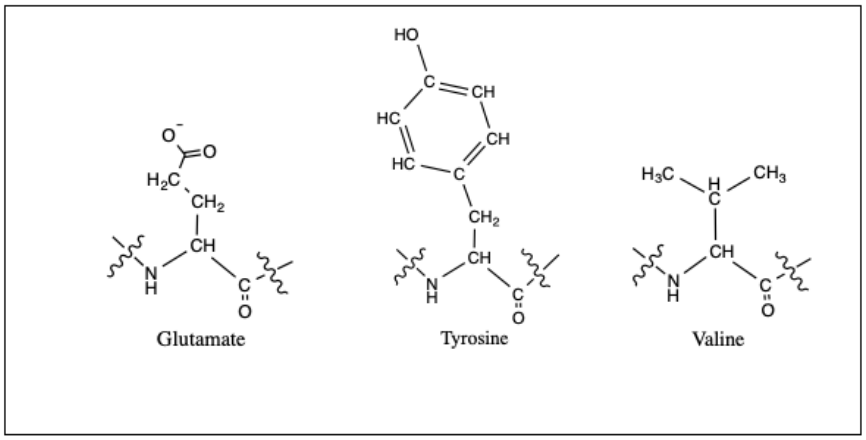
For each of the amino acid substitutions shown in the box, predict its impact on the following:

A. the folding of protein X? Justify your prediction.

B. the function of protein X. Justify your prediction.







Supplemental Material Figure S6.4. Problem Development Version 4 used in think-aloud interviews, n=5 students. Rationale: Upon expert review of Version 3, we realized the connection between mutations in hemoglobin and data on hemoglobin-related diseases is not clear enough for students. Also, we decided that we were more interested in students' ability to explain the role of non-covalent interactions and the entropy of water in protein folding and function. Thus, for Version 4, we removed all reference to hemoglobin and used the generic Protein X instead. We also removed the data provided in previous versions and focused instead on a model of protein folding.

There's a hydrogen bond over here [serine-glutamine interaction], and that's keeping this end [left part of protein X] together, and … I'm trying to think if… the attraction between the water molecules and this oxygen at the end of the … glutamate … is enough to overcome a hydrogen bond [between serine and glutamine]. And we're assuming that this [Protein X] can't rotate out like this [out of the page]. Then this [glutamine] would have to become unhitched, so this end [serine] could roll out like this [unwind into a straight line instead of look like a paper clip]. I think. I'm not really educated in this, but if it [serine] would come around and rotate around the other side, it would have to overcome this hydrogen bond [serine-glutamine interaction] to do it. Otherwise, it would just poof this end [phenylalanine cluster] out – mm, would it poof that end out or would it just stay inside? Because it's [glutamate's] not being repelled by these [adjacent phenylalanine side chains].

Supplemental Material Figure S6.5. As stated in *Predicting New Noncovalent Interactions (Step 5),* one student, Kevin, discussed the idea that glutamate's charge could be satisfied through an interaction with water. Kevin described this tentatively, even stating outright that he was not educated in the constraints of protein folding, but he showed sophisticated thinking regarding whether glutamate's attraction to water can overcome the hydrogen bond between serine and glutamine on the left side of Protein X.

CHAPTER 7

ADVANCING THE GUIDANCE DEBATE: LESSONS FROM EDUCATIONAL PSYCHOLOGY AND IMPLICATIONS FOR BICOHEMISTTRY LEARNING[1]

[1]Halmo, S.M.*, Sensibaugh, C.A., Reinhart, P., Stogniy, O., Fiorella, L. and Lemons, P.P. 2020. *Life Sciences Education*. in press.

Abstract

Research in science, technology, engineering, and mathematics education supports a shift from traditional lecturing to evidence-based instruction in college courses, yet it is unknown whether particular evidence-based pedagogies are more effective than others for learning outcomes like problem solving. Research supports three distinct pedagogies: worked examples plus practice, productive failure, and guided inquiry. These approaches vary in the nature and timing of guidance, all while engaging the learner in problem solving. Educational psychologists debate their relative effectiveness, but the approaches have not been directly compared. In this study, we investigated the impact of worked examples plus practice, productive failure, and two forms of guided inquiry (unscaffolded and scaffolded guidance) on student learning of a foundational concept in biochemistry. We compared all four pedagogies for basic knowledge performance and near-transfer problem solving, and productive failure and scaffolded guidance for far-transfer problem solving. We showed that 1) the four pedagogies did not differentially impact basic knowledge performance, 2) worked examples plus practice, productive failure, and scaffolded guidance led to greater near-transfer performance compared to unscaffolded guidance, and 3) productive failure and scaffolded guidance did not differentially impact far-transfer performance. These findings offer insights for researchers and college instructors.

Introduction

Trailblazing work over the last twenty years supports a shift from traditional lecturing to evidence-based pedagogies in college science, technology, engineering and mathematics (STEM) courses (Deslauriers, McCarty, Miller, Callaghan, & Kestin, 2019; Freeman et al., 2014; Haak, HilleRisLambers, Pitre, & Freeman, 2011; Knight & Wood, 2005). For example, discipline-based education research (DBER) has shown that active learning improves performance and reduces the achievement gap for STEM students compared to lecture (Freeman et al., 2014; Freeman, Haak, & Wenderoth, 2011; Haak et al., 2011). Since prominent studies like these, DBER has focused increasingly on second-generation instructional research, using findings from educational psychology to inform instructional design and testing these designs for certain topics and student populations (Eddy & Hogan, 2014; Freeman et al., 2014). As more STEM instructors join the movement towards evidence-based pedagogy, one enduring question remains: what type of instruction is optimal for student learning?

In order to optimize student learning in biology, instruction should be aligned to desired learning outcomes. Biology lessons almost always teach basic knowledge, including key terminology, the use of terms in context, and interpretation of common visual representations. Many instructors also aim for students to build procedural and conceptual knowledge, which enables them to explain how facts and terms connect and facilitates principle-based reasoning (Loibl, Roll, & Rummel, 2017; Rittle-Johnson & Schneider, 2015). This type of learning can be assessed using problems that resemble those used during instruction, referred to as near-transfer problems (McDaniel et al., 2018). Finally, some ambitious instructors aim for students to adapt learned concepts to new situations or different types of problems (Loibl et al., 2017). This type of learning can be assessed using problems that appear foreign or different from all previous practice,

referred to as far-transfer problems (Loibl et al., 2017). Indeed, since *Vision and Change*, the ability to solve both near- and far-transfer problems has been viewed as a key learning outcome for biology education and, thus, the focus of instructional design (American Association for the Advancement of Science, 2011).

While biology educators agree that instruction should focus on transfer, determining the most effective type of instruction for enhancing transfer is a topic of ongoing debate, particularly in educational psychology (Hmelo-Silver, Duncan, & Chinn, 2007; Kapur, 2016; Kirschner, Sweller, & Clark, 2006; Sweller, Kirschner, & Clark, 2007). At the heart of this debate is the nature and timing of the guidance provided during instruction (Lazonder & Harmsen, 2016; Mayer, 2004; Schwartz & Bransford, 1998). We define guidance broadly as any form of assistance offered during the learning process that aims to either provoke or provide information concerning the process or content involved (adapted from Lazonder & Harmsen, 2016). Regarding the nature of guidance, researchers debate whether guidance should be highly explicit such as providing explanations or less explicit such as providing prompts (Lazonder & Harmsen, 2016). The timing of guidance is also debated. Some argue that novice learners should be explicitly told concepts and procedures before solving problems independently (Glogger-Frey, Fleischer, Grüny, Kappich, & Renkl, 2015; Hsu, Kalyuga, & Sweller, 2015; Paas, Renkl, & Sweller, 2003; Renkl, 2014; Sweller, 2016; Sweller, Van Merrienboer, & Paas, 1998). Others argue that learners should explore problems on their own before being given explicit instructions (Kapur, 2008, 2011; Kapur & Bielaczyc, 2012; Kapur & Rummel, 2012; Schwartz, Chase, Oppezzo, & Chin, 2011; Schwartz & Martin, 2004; Weaver, Chastain, DeCaro, & DeCaro, 2018). Furthermore, some argue that guidance that fades away as knowledge and skills are built should be provided throughout the learning event (Hmelo-Silver et al., 2007). From this debate, three evidence-based pedagogies

emerge: worked examples plus practice, productive failure, and guided inquiry. All three pedagogies engage learners in problem solving and share the ultimate goal of enhancing student learning. Notably, none of these pedagogies involve unguided problem-solving practice (Mayer, 2004). Yet, the nature and timing of guidance recommended by each approach varies based on the theories in which they are situated (Figure 7.1). Likewise, each pedagogy is hypothesized to target different levels of transfer (Kapur, 2016).

Distinct Pedagogies and their Theoretical Underpinnings. Worked Examples Plus Practice. In worked examples plus practice (Figure 7.1), students receive explicit step-by-step explanations on how to solve a problem, usually through an expert solution, and then practice implementing these solution procedures through independent problem solving. According to cognitive load theory, worked examples reduce the amount of cognitive load or mental effort invested in working memory during learning (Paas et al., 2003; Sweller et al., 1998). Cognitive load theory suggests that when students study worked examples, they can focus their limited working memory on constructing the knowledge needed to solve the problem rather than using cognitive resources to search the problem space for a solution (Kirschner et al., 2006; Sweller, 2016). Studies in support of cognitive load theory demonstrate that students who learn using worked examples plus practice perform better on subsequent problem-solving tests than students who only solve practice problems on their own without guidance (Renkl, 2014; Sweller & Cooper, 1985). Some cognitive load theorists use these findings to argue that less-guided pedagogies, such as guided inquiry, are not ideal for learners, particularly learners with limited prior knowledge, because of their high demands on working memory (Kirschner et al., 2006). Yet others have argued that research on worked examples plus practice relies on weak controls (i.e., minimal guidance) and results in narrow learning outcomes (e.g., near transfer) (Kapur, 2016), and that high levels of

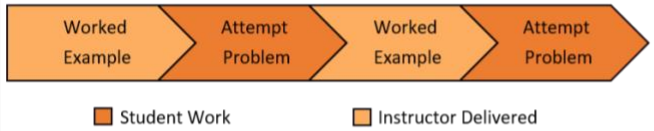
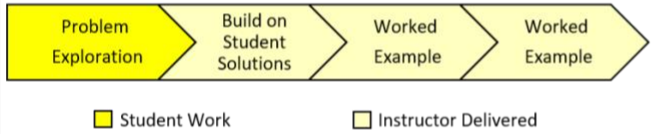
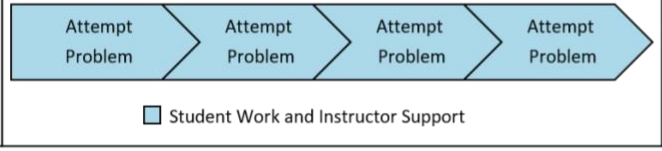
	Instruction	Nature of Guidance	Timing of Guidance	Targeted Learning Outcome
Worked Examples Plus Practice		Explicit explanations	Early, before problem solving	Near transfer
Productive Failure		Explicit explanations	Later, after problem exploration	Far transfer
Guided Inquiry		Scaffolded, prompts and explicit explanations	Distributed, during problem solving	Far transfer

Figure 7.1. Three Evidence-Based Pedagogies. Worked examples plus practice, productive failure, and guided inquiry are three evidence-based pedagogies that vary in the nature of guidance, the timing of guidance, and their targeted learning outcome

cognitive load directed towards exploring problems can benefit the development of deeper levels of conceptual understanding and transfer (Kapur, 2016; Schwartz et al., 2011).

Productive Failure. In productive failure (Figure 7.1), students explore problems and generate possible solutions on their own prior to receiving explicit guidance (e.g., explanations) (Kapur, 2008). The productive failure approach stems from research on desirable difficulties. Learning tasks that contain desirable difficulties require more effort and make learning more challenging in the short-term but more durable in the long-term (Bjork, 1994; Schmidt & Bjork, 1992). The demands on cognitive load are useful because presenting students with challenging problems first, before guidance, prepares them for future learning (Schwartz et al., 2011; Schwartz & Martin, 2004). Studies show that students who learn from productive failure outperform students who receive instruction in the form of lecture followed by problem-solving practice (Kapur, 2011; Kapur & Bielaczyc, 2012; Steenhof, Woods, Van Gerven, & Mylopoulos, 2019; Weaver et al., 2018). The benefits of productive failure include gains in conceptual knowledge and far transfer (Kapur, 2016; Loibl et al., 2017; Schwartz et al., 2011). Proponents of productive failure hypothesize that it is advantageous, especially for far transfer, because it helps learners activate their prior knowledge, recognize their own knowledge gaps, and focus on the underlying structure of problems prior to explicit instruction (Kapur, 2016; Loibl et al., 2017). In contrast, the worked examples plus practice approach risks that students will merely learn to apply provided procedures to practice problems without the deep conceptual understanding needed for transfer (Schwartz et al., 2011). However, some have argued that research on productive failure also suffers from inappropriate control conditions (Glogger-Frey et al., 2015) and has primarily been tested across a limited range of topics in mathematics (Loibl et al., 2017).

Guided Inquiry. While worked examples plus practice and productive failure are well-defined approaches, guided inquiry is more ill-defined and suffers from imprecision in terminology. For example, depending on their specific implementation, one could categorize inquiry-based learning (Prince & Felder, 2006), problem-based learning (Dochy, Segers, Van den Bossche, & Gijbels, 2003; Hmelo-Silver, 2004), case-based learning (Herreid, 2007), peer-led guided inquiry (Lewis & Lewis, 2005, 2008), and Process Oriented Guided Inquiry Learning (POGIL https://pogil.org/; Bailey, Minderhout, & Loertscher, 2012; Farrell, Moog, & Spencer, 1999) as types of guided inquiry instruction. We acknowledge this variance in implementation and the fact that there are other structures used by instructors, such as hybrids of these techniques, that may be effective (Eberlein et al., 2008). However, for the purpose of this paper we define guided inquiry as an approach where students actively engage in solving problems to learn critical concepts and practices and are guided throughout the process (Hmelo-Silver, 2004). Guidance through this process ranges in level of explicitness based on the learner's prior knowledge, but broadly consists of hints, prompts, questions or even direct explanation from an instructor or learning assistant (Hmelo-Silver et al., 2007; Lazonder & Harmsen, 2016). Additionally, we consider guided inquiry instruction to have the following characteristics: 1) students working together in small groups, 2) the instructor and learning assistants acting as facilitators of learning rather than as proprietors of knowledge, and 3) scaffolds or instructional supports that fade away as knowledge is built (van Merriënboer & Kirschner, 2007).

As defined, guided inquiry stems from social constructivism theory, which recognizes knowledge is built by the learner and is impacted by cooperative social interactions (Bodner, Klobuchar, & Geelan, 2001; Eberlein et al., 2008). While proponents of worked examples plus practice criticize guided inquiry for ignoring the limitations of human working memory (Kirschner

et al., 2006), guided inquiry proponents argue that scaffolded guidance effectively manages students' cognitive load (Schmidt, Loyens, Van Gog, & Paas, 2007). Because guided inquiry suffers from imprecision in terminology, it is challenging to characterize the evidence base for this approach. In the K-12 literature, guided inquiry has been shown to improve student learning outcomes compared to unguided inquiry (Lazonder & Harmsen, 2016). In higher education, students in a POGIL-style chemistry course scored as high as or higher on the final exams than students who had taken a more traditional lecture-based course from the same instructor (Farrell et al., 1999). Additionally, students in a peer-led guided inquiry style chemistry course experienced improved performance on the ACS Exam compared to students in more traditional lecture-based courses (Lewis & Lewis, 2008). Compared with traditional lecture, case-based learning in an introductory biology course improved exam performance, including performance on questions requiring application and analysis (Chaplin, 2009). Problem-based learning has also been shown to improve retention, application, and skill development compared to more traditional teaching methods (Dochy et al., 2003; Prince, 2004). Overall, the literature substantiates that guided inquiry-related approaches can improve student learning. However, as evident from the above examples, the literature is limited due to the use of comparison groups that provide no guidance for problem solving or lecture only with no time for problem-solving practice.

Controversies and a Need for Comparison Studies. Worked examples plus practice, productive failure, and guided inquiry all have been shown to enhance student learning, yet they have not been directly compared. This is an important deficit in the literature. First, researchers have recently hypothesized unique advantages of each pedagogy for serving different learning outcomes (Kalyuga & Singh, 2016; Kapur, 2016). For instance, worked examples plus practice may be best for learning specific procedures and near transfer, while productive failure and guided

inquiry may be best for promoting far transfer. Second, considering guided inquiry is a common instructional approach in biology and chemistry, it is of interest to the DBER community to compare guided inquiry to the other pedagogies. Third, research studies for all three pedagogies are limited due to weak comparison groups (e.g., no guidance, no time for problem solving practice, weak forms of direct instruction, lecture only). Stronger comparisons among pedagogies involving guidance and problem-solving practice are more intriguing to researchers and educators. Lastly, the majority of research studies highlighted above focus on domains in mathematics, such as algebra and statistics. The context-specific boundaries of this research base should be expanded to include domains like biology that rely heavily on conceptual knowledge. In this paper, we address these gaps in the literature with an investigation that directly compares worked examples plus practice, productive failure and two forms of guided inquiry in the context of biochemistry.

Chosen Context of Biochemistry. We purposefully chose the context of biochemistry for this comparison. Introductory biochemistry courses play an important role in STEM undergraduate curriculum because they 1) are required for many STEM majors, 2) include content that is critical for health professional entrance exams, and 3) integrate the disciplines of biology and chemistry. Biochemists agree on a set of core concepts that define the discipline (ASBMB, 2020; Loertscher, Green, Lewis, Lin, & Minderhout, 2014). One particularly challenging concept for students is the physical basis of non-covalent interactions (PBI). PBI builds on students' general chemistry and introductory biology content knowledge to bring together the idea that non-covalent interactions occur due to the electrostatic properties of biological molecules (Cooper, Williams, & Underwood, 2015). Students with expertise of this concept recognize that although interactions are given different names (i.e., ionic interactions, hydrogen bonds, van der Waals forces), they are all based on the same electrostatic principle of attraction due to opposite charge (Loertscher et al., 2014).

PBI is so central to biochemistry that once students deeply understand it, their view of the discipline is transformed (Loertscher et al., 2014). PBI is a content area ripe for instructional design because of known student difficulties with causal mechanisms of how non-covalent interactions arise (Becker, Noyes, & Cooper, 2016; Halmo et al., 2018). Although problems that deal with PBI are difficult and challenging for undergraduate students, incoming biochemistry students have prior knowledge in biology and chemistry that could be activated to help them solve these problem types.

Cross-Disciplinary Research Question. Researchers in both DBER and educational psychology assert that direct comparisons of distinct evidence-based pedagogies could resolve outstanding questions in each field and enable the optimization of student learning of persistently troublesome biology concepts. By drawing upon the strengths and shared goals of educational psychology and DBER (McDaniel et al., 2017), we aim to test general learning mechanisms within a specific disciplinary context that is persistently troublesome for students and, thus, advance both research and practice. Specifically, we address the following research question: What are the comparative impacts on student learning of PBI for methods of instruction that vary in the nature and timing of guidance, namely, worked examples plus practice, productive failure, and two forms of guided inquiry?

Methods

Study Design. We compared the impacts on student learning of four instructional approaches: worked examples plus practice, productive failure, and two forms of guided inquiry (unscaffolded and scaffolded guidance). To do this, we recruited students from the two prerequisite courses for introductory biochemistry (introductory biology and organic chemistry I) in spring 2018 and spring 2019. Due to logistical constraints, our study utilized an unbalanced incomplete

block design, with semester of data collection (spring 2018 and spring 2019) serving as the block effect (Figure 7.2). Our study is unbalanced because there are unequal sample sizes for each treatment across blocks. Our study is incomplete because we did not test all treatments in each block: productive failure was tested in spring 2018 and spring 2019, worked examples plus practice and unscaffolded guidance were tested in spring 2018 only, and scaffolded guidance was tested in spring 2019 only. Students who agreed to participate and completed a basic knowledge pretest were randomly assigned to one of the conditions tested in each block. Each condition involved a 35-45-minute lesson about PBI. After instruction, participants completed an assessment of basic knowledge and transfer. We describe the participants, data collection, and data analysis in detail in the following sections.

Participants. Data collection for this study took place over the course of two semesters (spring 2018 and spring 2019) at a research-intensive university in the southeastern United States. Spring 2018 participants were enrolled in one of three sections of an introductory biology course taught by one professor. Spring 2019 participants were enrolled in either an introductory biology course taught by one professor or an introductory organic chemistry I course taught by two different professors. The researchers purposely chose to recruit students from introductory biology and introductory organic chemistry I courses at the end of the spring semester because these courses are required prerequisites for introductory biochemistry, and students at the end of these courses reflect the incoming biochemistry student population.

In spring 2018, the researchers announced the study to 416 introductory biology students through in-class announcements. The instructor also allowed us to contact all 416 students through email. One hundred and fifty-four of the 416 contacted students agreed to participate and completed the basic knowledge test (described below). We excluded data for 31 of these students

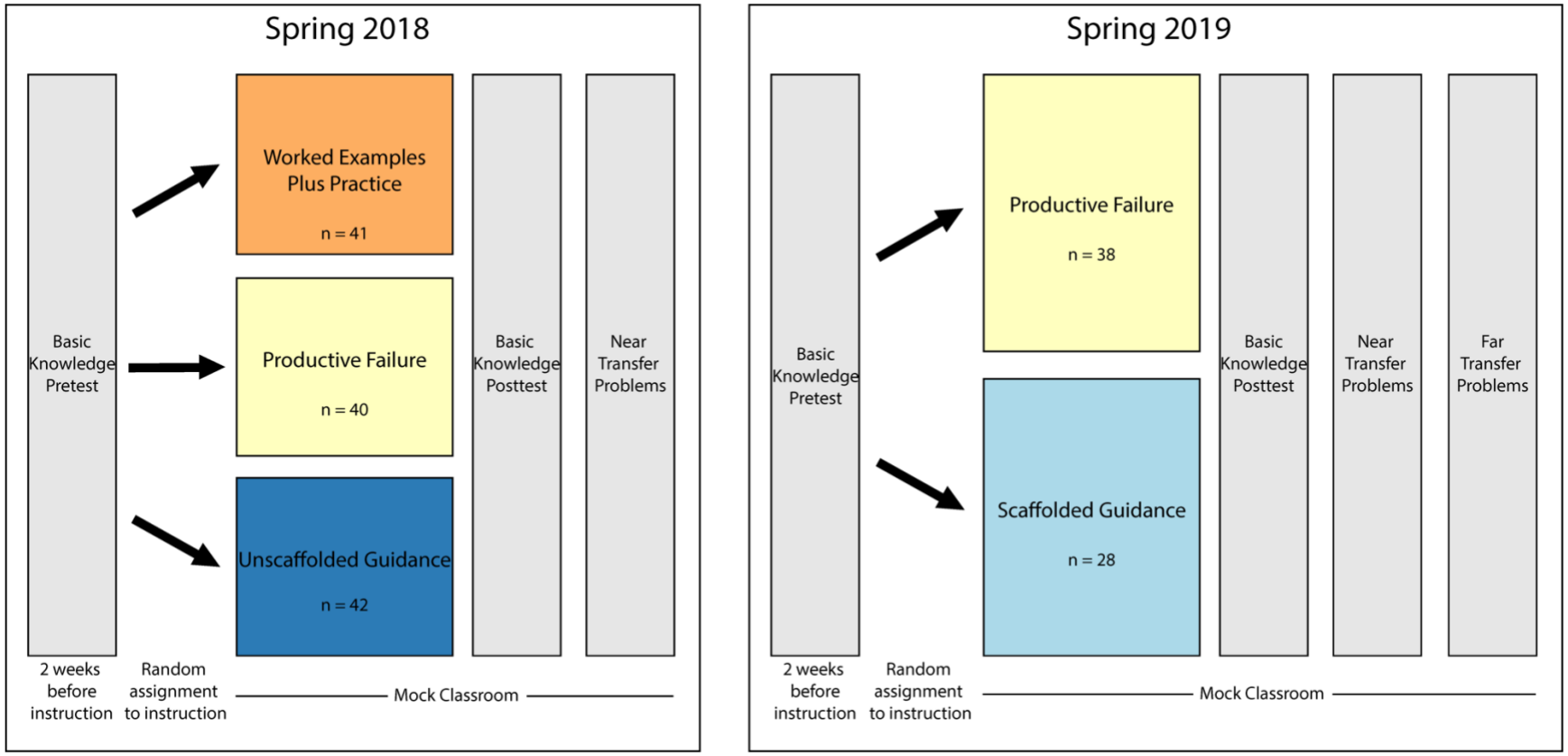


Figure 7.2. Study Design. Study design for the comparison of impacts on student learning of four instructional approaches: worked examples plus practice, productive failure, unscaffolded guidance, and scaffolded guidance.

because they did not complete the entire study. One hundred twenty-three (30%) of the 416 contacted students completed the entire study. We randomly assigned these participants to one of three instructional conditions: 41 participants to worked examples plus practice, 40 participants to productive failure, and 42 participants to unscaffolded guidance. The 123 participants who completed the entire study received ten points of extra credit towards their final introductory biology course grade (2.5% of the total possible points) as an incentive.

In spring 2019, the researchers used in-class announcements to announce the study to 931 introductory biology and introductory organic chemistry I students. The instructors allowed us to follow up via email with students who gave us their names and email addresses. Two hundred and twenty-seven students provided their names and email addresses. Ninety-five of the 227 contacted students agreed to participate and completed the basic knowledge test. We excluded data for 29 students from analyses because they did not complete the entire study. Sixty-six (29%) of the 227 contacted students completed the entire study. We randomly assigned these participants to one of two instructional conditions: 38 participants to productive failure, and 28 participants to scaffolded guidance. The 66 participants who completed the entire study received $25 cash as an incentive.

The UGA institutional review board approved this study under exempt status (STUDY00000660 and PROJECT000000090). Demographic information of the participants can be found in Supplementary Material Table S7.1.

Data Collection. Development of Instructional Materials. The authors developed all instructional materials. We designed them to help students achieve learning objectives pertaining to PBI and focused, in particular, on known student difficulties (e.g., Halmo et al., 2018). The materials are intended for use in introductory biochemistry courses. Three experts who are biochemistry instructors and discipline-based education researchers provided feedback on the

worked examples plus practice, productive failure, and unscaffolded guidance materials. Three experts on the method of guided inquiry, who are also discipline-based education researchers, provided feedback on the scaffolded guidance instructional materials. We pilot tested the productive failure and scaffolded guidance materials in two focus groups. Four introductory biology students participated in the productive failure focus group in spring 2018, and six introductory biology students and five introductory biochemistry students participated in the scaffolded guidance focus group in spring 2019. We revised the materials based on expert feedback and pilot testing. We provide the finalized lesson materials used in this study, including handouts, instructor slides, and notes in Appendix A.

Instructional Conditions. To compare the impact of worked examples plus practice, productive failure, unscaffolded guidance and scaffolded guidance, we randomly assigned participants to one of the conditions tested in each block. Each lesson lasted 35-45 minutes and took place in a SCALE-UP classroom (Beichner & Saul, 2003). SCALE-UP classrooms have several round tables with nine seats per table and are designed to facilitate student-instructor and student-student interactions. One of the authors taught all three lessons in Spring 2018 while a different instructor taught both lessons in Spring 2019. We randomly assigned participants in each session to seats in the SCALE-UP classroom. The materials used and type of instruction experienced by participants differed depending on the instructional condition:

Worked examples plus practice condition: The instructor reviewed the learning objectives, introduced participants to a problem, and presented a worked example solution to the problem (i.e., an explicit explanation). She then gave participants time to practice a similar problem independently for several minutes, and then asked them to compare their solutions with the two people sitting closest to them. The instructor did not assist participants during independent problem

solving or group sharing. In total, participants went through two rounds of this worked example-problem practice pairing.

Productive failure condition: The instructor reviewed the learning objectives and introduced participants to the same four problems that the worked examples plus practice participants practiced. However, the instructor provided no solution. Instead, she asked students to explore the problems with the participants at their table by comparing and contrasting the problems and generating as many possible solutions as they could (i.e., prompts). During this exploration, the instructor and two peer learning assistants in 2018 (one in 2019) walked around the room and noticed student work. They did not comment on the correctness of students' ideas or direct them to a solution. Instead, they repeatedly asked students to explain what they were doing and pushed them to expand their thinking to all four problems. The instructor and peer learning assistants quickly conferred on the common ideas students were expressing. Then after problem exploration, the instructor commented and built upon students' ideas (Loibl et al., 2017). For example, the instructor and peer learning assistants noticed that students frequently compared the differences among having C, H, N, or O atoms in an amino acid R group. The instructor pointed this out to students and said, "That's a good problem-solving step. The way to think about differences among atoms for this problem is to consider their differences in electronegativity and what this means in terms of full/partial and permanent/temporary charges." After building on students' solutions, the instructor presented two back-to-back worked examples (the same ones from the worked examples plus practice condition).

Unscaffolded guidance condition: In this condition, the instructor and two peer learning assistants provided guidance, but the instructional materials were not scaffolded (i.e., did not have supports that faded away as knowledge was built). We chose this approach to implement a guided

inquiry condition that was comparable to worked examples plus practice and productive failure in the number of problems covered and in overall session length. Thus, after the instructor reviewed the learning objectives, she gave participants the same four problems that the worked examples plus practice and productive failure participants received. However, she did not provide any solutions in the form of worked examples, and participants had the entire class period to work on the four problems with people at their table. During this work time, the instructor and two peer learning assistants circulated the room and addressed participants' questions. The instructor/peer learning assistants aimed to provide support based on a students' prior knowledge. When interacting with students they first diagnosed the students' prior knowledge (e.g., by looking at participants' work or asking them to explain their thinking). If possible, the instructors/peer learning assistants provided simple prompts (e.g., "Look at this aspect of the problem."), but if participants' prior knowledge was more limited, they provided explicit guidance (e.g., explanations of a concept). For example, if a student was stuck on hydrogen bonds, the instructor would ask probing questions like, "What is a hydrogen bond?" Next, the instructor might ask leading questions such as, "Do hydrogen bonds involve charges? If so, where do they come from?" At that point, if it was clear that participants had the knowledge they needed to proceed, the instructor would leave them to work. However, if it was clear that participants were unfamiliar with key ideas, the instructor would provide an explicit explanation. Anytime an explicit explanation was provided, participants were encouraged to use that explanation to help their work on subsequent problems.

Scaffolded guidance condition: In this condition, the instructor and four peer learning assistants provided guidance, and the instructional materials were scaffolded (i.e., had supports that faded away as knowledge was built). We chose this approach to implement a guided inquiry

condition that was comparable to worked examples plus practice and productive failure in session length, yet presented students with problems that progressed from simple to complex. The instructional materials included a total of 24 problems. These 24 problems encompassed the same four problems that participants in other conditions saw, but the problems were strategically broken down into component parts that built on one another. By the end of the problem set, students were solving a problem without any support. To start the session, the instructor reviewed the learning objectives. Then she gave participants the entire instruction time to work on the problem set with the participants at their table. During this work time, the instructor and four peer learning assistants circulated the room and addressed participants' questions. They followed the same principles of interaction as the unscaffolded guidance condition.

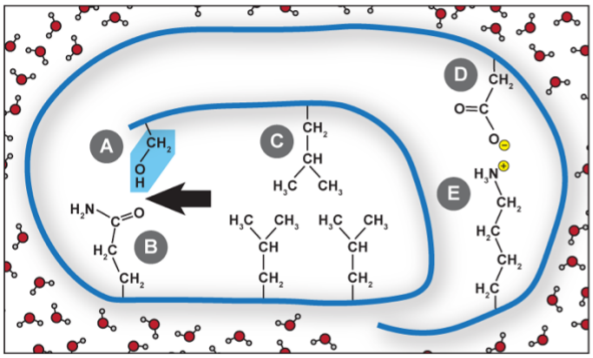
Assessments of Student Learning. We used three assessments of student learning, which are described below. We administered all assessments through the Qualtrics (SAP, Walldorf, Baden-Württemberg, Germany) online survey platform. We provide all assessment items used in this study in Appendix A.

Basic Knowledge Test. The basic knowledge test was developed as part of a separate longitudinal study on student learning (same institution, N=913). The test consists of nineteen multiple-choice and multiple true-false items and addresses key terminology, the use of terms in context, and interpretation of common visual representations associated with PBI. We present five of the nineteen items in Figure 7.3A and the full nineteen-item test in Appendix A. A key for this test is available from the corresponding author upon request. We scrutinized the psychometric properties of the test, including dimensionality and reliability (unpublished results). In so doing, we used an item response theory model. Item response theory is a probabilistic approach where a correct response to an item is defined as a function of person (i.e., ability) and item parameters

based on unidimensionality and local independence assumptions (Embretson & Reise, 2000). We used a 2-parameter logistic model. In the 2-parameter logistic (2-PL) model, the correct response to an item is defined as a function of the student's ability, and the item's difficulty level and discrimination power. The BILOG software was used to estimate person and item parameters (Zimowski, Muraki, Mislevy, & Bock, 1996). Results showed that the empirical reliability was acceptable at a value of 0.67 (Du Toit, 2003; Kline, 2000). We used the obtained item parameters from this 2-PL model to estimate students' ability in the current study. For the current study, participants completed the basic knowledge test before and after instruction (Figure 7.2). Participants took the pretest on their own time. Participants took the posttest immediately after instruction in the same classroom where they received instruction. The pre- and posttest were identical except they referred to proteins that differ in appearance (i.e., Protein Z for the pretest and Protein X for the posttest).

Near-Transfer Problems. The three near-transfer problems used in this study are based on the Protein X problem published in Halmo et al. (2018) and were further revised based on interviews with one introductory biology student and two introductory biochemistry students in spring 2016. These problems require students to make a prediction and explain their prediction (Figure 7.3B and Appendix A). As a reliability measure, we calculated Cronbach's alpha between the three problems to be 0.75, which indicated good internal consistency (Kline, 2000). The near-transfer problems resemble the problems used during instruction, but present proteins that differ in appearance and involve different amino acids. Participants completed the near-transfer problems immediately after instruction following the basic knowledge posttest (Figure 7.2). The near-transfer problems were presented to students in random order.

A. Protein X, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and gray). This environment has a pH of 7.4. The blue line represents the protein X backbone. Some, but not all, of the amino acid side chains are shown in chemical notation.



The amino acids shown are: (A) serine, (B) glutamine, (C) leucine, (D) aspartate, and (E) lysine.

The items below all relate to the most prominent non-covalent interaction occurring in the space pointed to by the arrow.

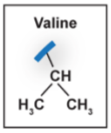
What is the name of this non-covalent interaction? Select one option.

- ○ hydrogen bond
- ○ ion pairing
- ○ van der Waals interaction

The charges involved in this non-covalent interaction are…

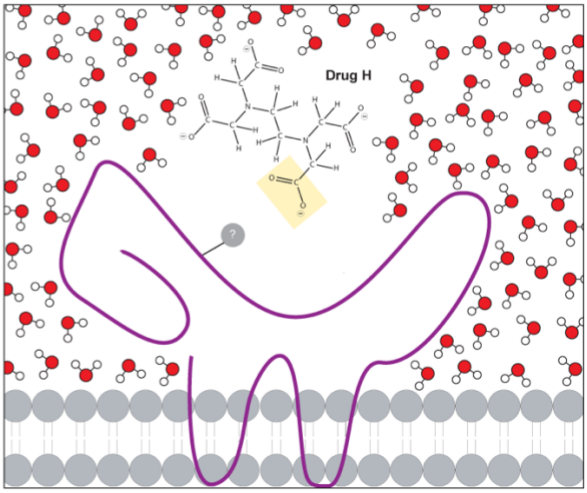
	True	False
partial	○	○
temporary	○	○
induced	○	○
due to differences in electronegativity	○	○

B. Sometimes, a mutation occurs that substitutes serine (blue highlight) with valine (below).

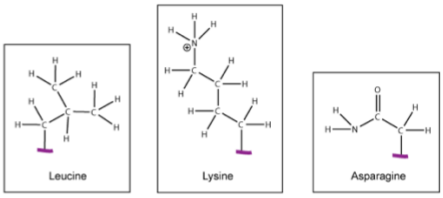


Do you predict that such a mutation would affect the non-covalent interaction pointed to by the arrow? Provide a scientific explanation to support your prediction.

C. Below is a model of Drug H and a protein with which it may interact. The protein is located on the cell surface situated within the cell membrane and surrounded by water molecules (red and white). The environment has a pH of 7.4. The purple line represents the protein backbone, and the section labeled with a question mark is a site for an amino acid side chain (R group).



Which amino acid below would interact non-covalently with the yellow highlighted section of Drug H?



Provide a scientific explanation describing how Drug H interacts non-covalently with the amino acid you selected. Be sure to describe how this interaction forms.

Figure 7.3. Three Measures of Student Learning of PBI. Selected items from the basic knowledge posttest (A) and examples of a near-transfer problem (B) and far-transfer problem (C) used in this study.

Far-Transfer Problems. The authors developed three far-transfer problems based on the previous work of Werth (2017). Six introductory biology students and five introductory biochemistry students in a spring 2019 focus group provided feedback on the far-transfer problems, and we revised the problems as needed. These problems require students to make a prediction and explain their prediction (Figure 7.3C and Appendix A). These far-transfer problems draw upon the same conceptual knowledge as the near-transfer problems, but the context differs from the problems provided during instruction, and, thus, requires a different solution structure. As for near transfer problems, we used Cronbach's alpha as a reliability measure. Cronbach's alpha (0.78) indicated good internal consistency (Kline, 2000). Due to the logistical constraints described above only spring 2019 participants (not spring 2018 participants) completed the far-transfer problems, and they did so immediately after instruction (Figure 7.2).

Data Analysis. We downloaded from Qualtrics (SAP, Walldorf, Baden-Württemberg, Germany) all participant responses to the assessment items. We collected data on the basic knowledge pre- and posttests and written responses to the near- and far-transfer problems.

Scoring of Basic Knowledge Test. As described in the previous section, we used the 2-PL IRT model to estimate item parameters with responses from 913 students. We estimated students' ability for pre- and posttest in this study using those item parameters. The ability parameter in the IRT model can be interpreted as a Z-score (M=0; SD=1). Hereafter, we refer to students' ability estimates as basic knowledge performance.

Analytical Coding of Near- and Far-Transfer Written Responses. Three of the authors (SMH, PR & OS) first read all written responses from both rounds of data collection. We developed an analytical codebook, informed by knowledge of published descriptions of student thinking about PBI (Becker et al., 2016; Cooper et al., 2015; Halmo et al., 2018; Loertscher et al.,

2014), to capture common ideas. The authors who served as coders and developers of the codebook were blind to condition during this phase of the research. We independently applied codes from the codebook to the written responses in a deductive manner and inductively created new codes as needed. The researchers then deductively applied these new codes to all written responses. The finalized codebook was applied to all written responses by two coders. After independent coding, we calculated intercoder reliability for all codes used in subsequent analyses (see below) using Cohen's Kappa (Gisev, Bell, & Chen, 2013). Cohen's Kappa coefficients ranged from 0.21 to 0.96 for near-transfer coding and from 0.45 to 0.93 for far-transfer coding. The overall average Cohen's Kappa for near-transfer and far-transfer coding was 0.74. and 0.76, respectively, which both indicated substantial agreement (Viera & Garrett, 2005). The researchers discussed all discrepancies in coding until complete agreement was reached. We provide the finalized analytical codebook for each problem in Appendix A.

After the written responses were coded to consensus, two of the authors (SMH & PPL) and a team of undergraduate researchers developed a rubric that enabled us to assign a score to each near- and far-transfer problem based on the analytical codebook. The researchers involved in this scoring process were blind to condition. The scoring rubric for each problem is available in Appendix A. The rubric captured the quality and correctness of predictions and the supporting evidence. The rubric also credited participants for attempting to support a prediction with evidence regardless of the quality or correctness of the prediction and evidence provided. More specifically, we awarded up to three points based on the quality and correctness of the prediction. We awarded up to three points based on the quality and correctness of the evidence. We awarded up to one point if both a prediction and evidence were provided, regardless of whether either was correct. We applied the scoring rubrics to all analytically coded written responses, resulting in a score from

zero to seven for every written response item. To calculate overall near-transfer and far-transfer performance, we summed the scores on the three problems for each participant and divided that sum by the highest possible score, generating an overall near-transfer and far-transfer score.

Statistical Analyses. We separately analyzed basic knowledge posttest performance, near-transfer performance, and far-transfer performance. We set our alpha level at 0.05 for these three independent statistical tests. We analyzed basic knowledge posttest performance using a type II sum of squares ANCOVA with semester of data collection serving as the block effect, basic knowledge pretest performance as a covariate, and instructional condition as the independent variable. We analyzed near-transfer performance using a type II sum of squares ANCOVA with instructional condition as the between-subject variable, semester of data collection serving as the block effect, and basic knowledge pretest performance as the covariate. We analyzed far-transfer performance using type II sum of squares ANCOVA, with instructional condition as the between-subject variable and basic knowledge pretest performance as the covariate. When instruction type had a significant effect on the omnibus test ($p<0.05$), Tukey's least-squares means post hoc test was used to perform multiple comparisons. We determined the estimates of adjusted group mean differences to be statistically significant if $p<0.05$. To calculate effect sizes for significant differences, we estimated Cohen's *d* by dividing the adjusted group mean difference by the square root of the MS_{Error} from the analysis of covariance. All statistical analyses were conducted in R (R Core Team, 2019).

Results

Prior knowledge, not method of instruction, predicts performance on the basic knowledge test. We first evaluated whether there was a difference on pretest performance among instructional conditions. ANOVA results indicated no significant differences on pretest performance among the

four conditions [$F(3, 184)= 0.19$, $p = 0.90$]. Previous work showed that worked examples plus practice, productive failure, and guided inquiry all enhance student learning. Thus, we did not expect differential advantages of different instructional methods on basic knowledge posttest performance. As anticipated, no instructional group outperformed the others on the basic knowledge posttest (Figure 7.4). The means for posttest performance (adjusted for pretest performance) were -0.48 for participants in the worked examples plus practice condition, -0.61 for participants in the productive failure condition, -0.48 for participants in the unscaffolded guidance condition, and -0.68 for participants in the scaffolded guidance condition (Table 7.1). The negative values do not indicate that no learning occurred (see Supplemental Material Table S7.2). The negative values only indicate that students in the current study performed below the population mean of the 913 students used for IRT analysis (see Methods). For basic knowledge posttest performance, we first tested to see if there were significant interaction effects between condition and pretest performance. The interaction effect was not significant [$F(3, 180) = 2.49$, $p = 0.06$], so we removed the interaction from the model and report the ANCOVA results without the interaction. Specifically, using a type II ANCOVA, we found no significant effect of instruction type on basic knowledge posttest performance [$F(3, 183) = 0.48$, $p = 0.69$]. We also found no significant effect of semester of data collection on basic knowledge posttest performance [$F(1, 183) = 1.12$, $p = 0.29$]. Additionally, when the data were filtered to look at participants in the productive failure condition only, the semester block (either spring 2018 or spring 2019) did not have a significant effect on posttest performance [$F(1, 76) = 0.52$, $p = 0.48$], suggesting no semester block differences. However, there was a significant effect of basic knowledge pretest performance on basic knowledge posttest performance [$F(1, 183) = 16.68$, $p < 0.001$]. In sum,

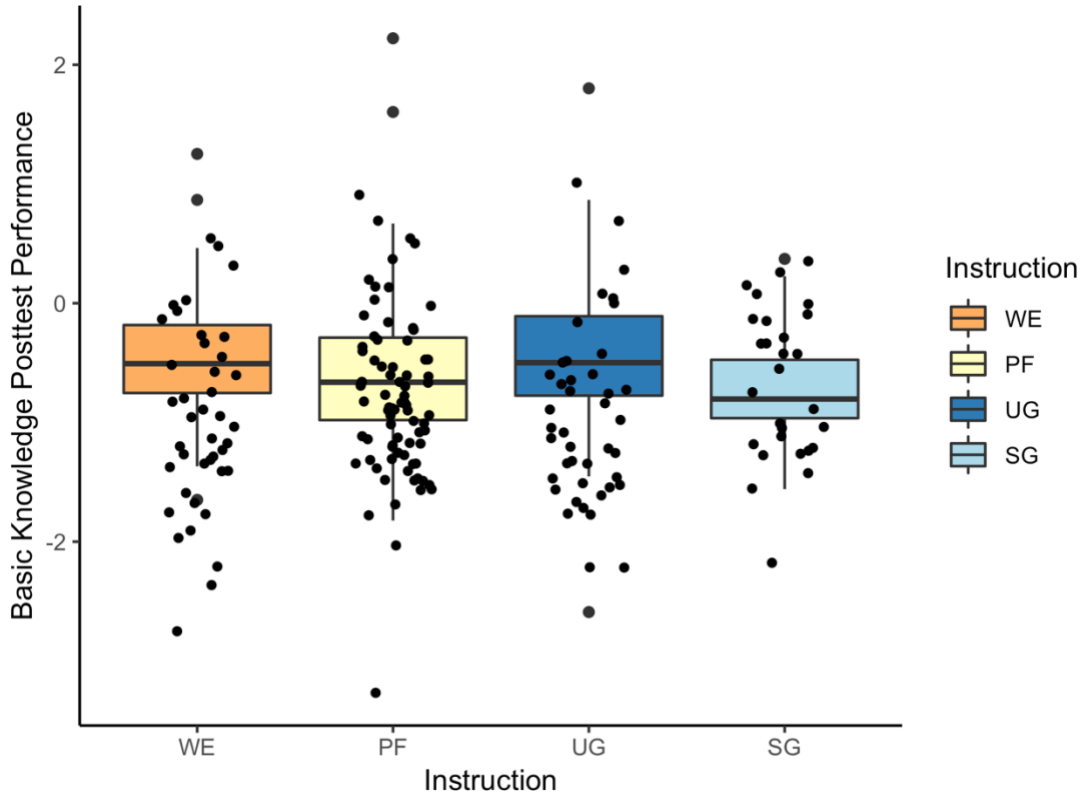


Figure 7.4. Basic knowledge posttest performance by instruction type. For each box and whisker plot, the black horizontal line represents the median basic knowledge posttest performance (unadjusted for pretest performance), the hinges represent the first and third quartiles, and the whiskers extend to the highest and lowest value that is within 1.5 times the interquartile range of the hinge. Dots represent individual participants and values can be interpreted as a Z-score (*M*=0; *SD*=1). Positive and negative values do not indicate learning and no learning, respectively. Rather, positive values indicate that a student performed above the population mean of the 913 students used for IRT analyses, while negative numbers indicate a student performed below the population mean.

Table 7.1. Basic knowledge posttest performance means and standard deviations adjusted for pretest performance

Instruction	Mean	Standard Deviation
Worked examples plus practice (n = 41)	-0.48	0.74
Productive failure (n = 78)	-0.61	0.60
Unscaffolded guidance (n = 42)	-0.48	0.74
Scaffolded guidance (n = 28)	-0.68	0.70

worked examples plus practice, productive failure, unscaffolded guidance, and scaffolded guidance did not differentially impact basic knowledge posttest performance (Figure 7.4).

More than unscaffolded guidance is needed for near-transfer problem solving. While we found no significant effect of instructional condition on basic knowledge posttest performance, we did find a significant effect of instructional condition on near-transfer problem solving (Figure 7.5). Specifically, worked examples plus practice, productive failure, and scaffolded guidance outperformed unscaffolded guidance on near-transfer problem solving. The mean near-transfer score was 0.63 for worked examples plus practice, 0.58 for productive failure, 0.44 for unscaffolded guidance, and 0.66 for scaffolded guidance participants (Table 7.2). For near transfer performance, we first tested to see if there were significant interaction effects between condition and pretest performance. The interaction effect was not significant [$F(3, 180) = 1.62$, $p = 0.19$], so we removed the interaction from the model and report the ANCOVA results without the interaction. Using a type II ANCOVA with semester of data collection as a block effect and basic knowledge pretest performance as a covariate, we found a significant effect of instruction type on near-transfer problem-solving performance [$F(3, 183) = 5.36$, $p = 0.001$]. Post hoc comparisons using the least-squares means Tukey adjusted test indicate that the mean near-transfer problem-solving performance for the unscaffolded guidance condition was significantly lower than the worked examples plus practice condition ($d = 0.72$), the productive failure condition ($d = 0.65$), and the scaffolded guidance condition ($d = 1.08$) (Table 7.3). Basic knowledge pretest performance significantly affected near-transfer problem-solving performance [$F(1, 183) = 4.21$, $p = 0.04$], whereas semester block did not significantly affect near-transfer problem-solving performance [$F(1, 183) = 1.43$, $p = 0.23$]. Additionally, when the near-transfer scores were filtered to look at

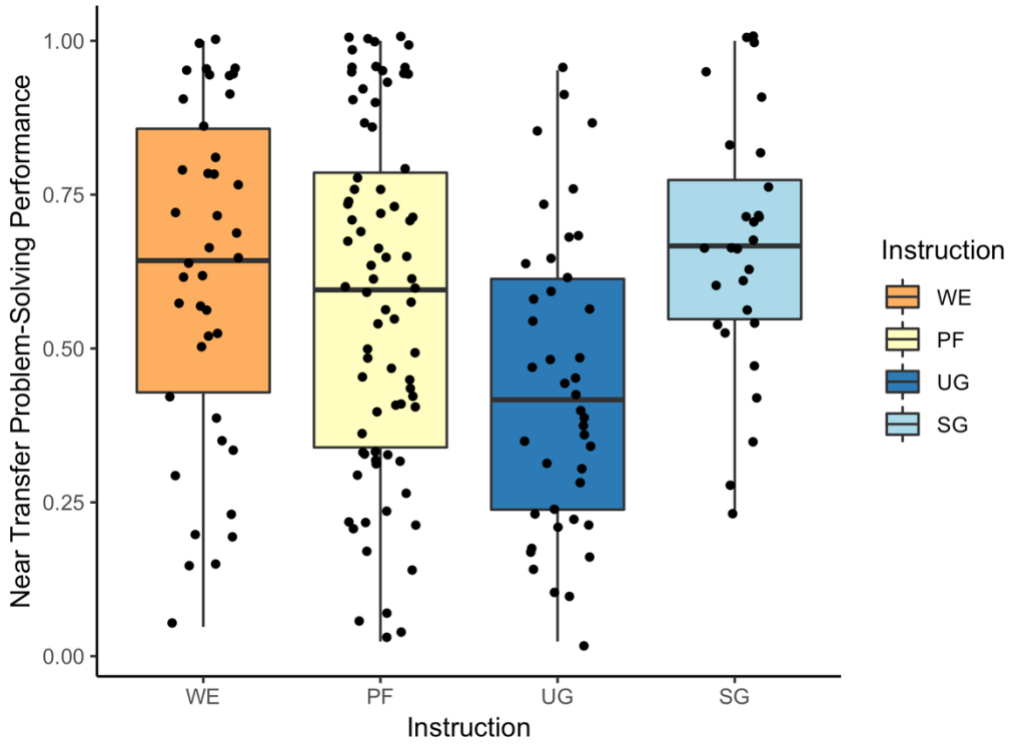


Figure 7.5. Near transfer problem-solving performance by instruction type. For each box and whisker plot, the black horizontal line represents the median near-transfer performance (unadjusted for pretest performance), the hinges represent the first and third quartiles, and the whiskers extend to the highest and lowest value that is within 1.5 times the interquartile range of the hinge. Dots represent individual participants.

Table 7.2. Near transfer performance unadjusted means and standard deviations

Instruction	Mean	Standard Deviation
Worked examples plus practice (n = 41)	0.63	0.28
Productive failure (n = 78)	0.58	0.28
Unscaffolded guidance (n = 42)	0.44	0.24
Scaffolded guidance (n = 28)	0.66	0.20

Table 7.3. Pairwise comparisons of near transfer performance means adjusted for pretest performance

Comparison	Adjusted Mean Difference	Standard Error	p-Value	Effect Size (Cohen's *d*)
Productive failure - scaffolded guidance	-0.11	0.06	0.32	
Productive failure - unscaffolded guidance	0.17	0.06	0.02*	0.65
Productive failure - worked examples plus practice	-0.02	0.06	0.99	
Scaffolded guidance - unscaffolded guidance	0.28	0.09	0.008*	1.08
Scaffolded guidance - worked examples plus practice	0.10	0.09	0.71	
Unscaffolded guidance - worked examples plus practice	-0.18	0.06	0.007*	0.72

participants in the productive failure condition only, the semester block (either spring 2018 or spring 2019) did not have a significant effect on near transfer performance [$F(1, 76) = 0.96$, $p = 0.33$], suggesting no semester block differences. This result suggests that more than unscaffolded guidance is needed to help learners solve problems similar to those seen in instruction.

Different types of failure do not differentially impact far-transfer problem solving. Given that more than unscaffolded guidance is needed for near-transfer problem solving, we investigated how the other forms of instruction impacted far-transfer problem solving. Only participants in Spring 2019 were asked to solve far-transfer problems, and we did not recruit enough participants in Spring 2019 to test three conditions. Therefore, we elected to test productive failure and scaffolded guidance, but not worked examples plus practice, because it was unknown if productive failure and guided inquiry differentially impact far-transfer problem solving (Kapur, 2016). In our experiment, scaffolded guidance and productive failure did not differentially impact far-transfer performance (Figure 7.6). The mean far-transfer score was 0.51 for participants in the productive failure condition and 0.58 in the scaffolded guidance condition (Table 7.4). For far transfer performance, we first tested to see if there were significant interaction effects between condition and pretest performance. The interaction effect was not significant [$F(1, 62) = 0.83$, $p = 0.37$], so we removed the interaction from the model and report the ANCOVA results without the interaction. Using a type II ANCOVA with basic knowledge pretest performance as a covariate, we found no significant effect of instruction type [$F(1, 63) = 0.96$, $p = 0.33$] or basic knowledge pretest performance [$F(1, 63) = 0.06$, $p = 0.81$] on far-transfer performance. Our data suggest that explicit guidance after problem solving (i.e., productive failure) and guidance distributed throughout problem solving in the form of scaffolded materials and instructor support (i.e., scaffolded guidance) do not differentially impact far-transfer problem solving.

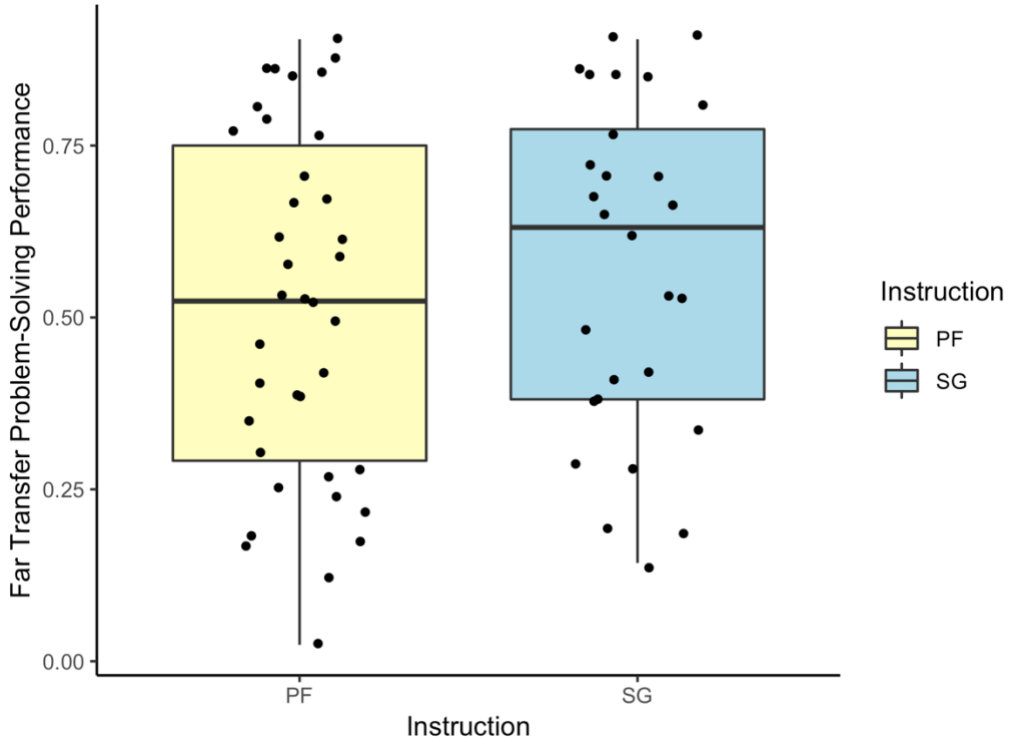


Figure 7.6. A comparison of far transfer problem-solving performance between productive failure and scaffolded guidance. For each box and whisker plot, the black horizontal line represents the median far-transfer performance (unadjusted for pretest performance), the hinges represent the first and third quartiles, and the whiskers extend to the highest and lowest value that is within 1.5 times the interquartile range of the hinge. Dots represent individual participants.

Table 7.4. Far transfer performance unadjusted means and standard deviations

Instruction	Mean	Standard Deviation
Productive failure (n = 38)	0.51	0.25
Scaffolded guidance (n = 28)	0.58	0.24

Limitations

Readers should consider the following limitations when evaluating our findings. First, the nature of the participant recruitment and data collection in this study led to an incomplete block design. While every instruction type was not represented in each block, there was overlap from spring 2018 to spring 2019 of the productive failure instructional condition. We accounted for block in our statistical analyses and found no significant differences. A second limitation related to the first is that low recruitment in spring 2019 necessitated a reduction in the number of treatments for that block. The research team prioritized the testing of productive failure and scaffolded guidance, which prevented us from investigating worked examples plus practice in our analysis of far transfer performance. Researchers hypothesize that worked examples plus practice may lead to lower performance on far transfer compared to productive failure or scaffolded guidance (Kapur, 2016). Future research should test this hypothesis. Third, we were limited by sample size. One hundred and eighty-nine (14%) of the 1,347 students recruited to participate in the study actually participated, suggesting the students who did participate may have differed from a typical student population. For example, our sample was disproportionately female (see Supplemental Material Table S7.1). In addition, the average posttest scores were below the mean for the instrument (see Table 7.1), suggesting the students in this sample differ from the students used in the 2-PL IRT model. One possible explanation for the below average basic knowledge scores is the fact that participants in this study took the test outside of class time whereas participants used to generate the 2-PL IRT model took the test during class time and therefore might have taken it more seriously. Also, we cannot rule out the possibility that the students who did participate may have had greater motivation for learning the material or for the extrinsic rewards offered as incentives. Therefore, future work should investigate similar research questions

in an authentic classroom setting. Additionally, given the sample size constraint, a sensitivity power analysis with alpha set to 0.05 and power set to 0.85 reveals a minimal detectable effect of 0.26 for the basic knowledge and near-transfer analyses and a minimal detectable effect of 0.37 for the far-transfer analysis, indicating that we could at minimum detect a medium (0.25-0.40) or large (>0.40) effect. Smaller differential effects may exist, but we did not have the statistical power in our dataset to detect them. Replicating the study with an increased sample size could resolve this issue. Fourth, to minimize participant burden and increase compliance, we administered assessments immediately following instruction. Therefore, we cannot answer the question of whether the results obtained persisted for longer periods of time. Fifth, our measures of basic knowledge, near transfer, and far transfer may not have been sensitive enough to detect differences that did exist. Finally, we investigated only one content area (PBI), so our claims are not generalizable to other content areas.

Discussion

Given the calls for second generation research on active learning (Freeman et al., 2014) and cross-disciplinary collaborations between biology education and educational psychology (McDaniel et al., 2017), we set out to determine the comparative impacts of worked examples plus practice, productive failure, and two forms of guided inquiry (i.e., unscaffolded and scaffolded guidance) on student learning of a challenging concept in biochemistry, PBI. Below we discuss our findings in the context of past research, explore their significance, and propose future directions.

The Nature and Timing of Guidance May Not Matter for Near Transfer. We show that multiple, but not all, instructional methods can achieve comparable learning gains. Prior research on worked examples plus practice, productive failure, and guided inquiry suffered from the use of

weak controls (e.g., minimal guidance or weaker forms of explicit instruction such as lecture only) (Glogger-Frey et al., 2015; Kapur, 2016). Pedagogies with some form of guidance unsurprisingly outperformed those with no guidance (Mayer, 2004), but how do guided pedagogies measure up to one another? Until now, the literature lacked a direct comparison of worked examples plus practice, productive failure, and guided inquiry (Figure 7.1). In this work, we conducted a head-to-head comparison to show that worked examples plus practice, productive failure, unscaffolded guidance, and scaffolded guidance led to comparable basic knowledge outcomes (Figure 7.4). We also show that worked examples plus practice, productive failure, and scaffolded guidance led to comparable near-transfer problem solving that is significantly better than unscaffolded guidance (Figure 7.5). Finally, we show that productive failure and scaffolded guidance produced comparable far-transfer problem solving (Figure 7.6). These novel findings shed light on the debate in educational psychology about the nature and timing of guidance (Hmelo-Silver et al., 2007; Kapur, 2016; Kirschner et al., 2006) and suggest that instructors have some flexibility in choosing among the tested approaches.

We did not detect differences among worked examples plus practice, productive failure, or scaffolded guidance on near-transfer performance. Likewise, for far transfer, we did not detect differences in learning for productive failure compared to scaffolded guidance. Taken together, these findings suggest that at least near transfer can be achieved whether the nature of guidance involves explicit explanations (i.e. worked examples plus practice and productive failure) or scaffolded instructional materials (i.e., scaffolded guidance). Transfer can also be achieved whether guidance is early, late, or distributed throughout problem solving, although future research should compare worked examples plus practice, productive failure, and scaffolded guidance for far-transfer problem solving (Kapur, 2016).

We stress one important caveat regarding the nature of guidance. If guided inquiry is used, scaffolded instructional materials seem to be important. Our scaffolded guidance materials sequenced problem solving into increasingly complex questions, while the unscaffolded guidance condition simply provided problems for participants to solve. Even though the unscaffolded guidance condition was designed to provide just-in-time support, perhaps this was not sufficient because only some students requested help, and there were not enough members of the instructional team to help students break down the problem into components pieces. The unscaffolded guidance condition involved two learning assistants while the scaffolded guidance condition involved four learning assistants. Unscaffolded guidance can occur unintentionally among instructors who aim to create active-learning environments. These instructors may give students problem sets, but not break them down into manageable chunks that lead students from simple to complex thinking. This could happen even if instructors stop lecturing and give students plenty of time to work in class. Along these lines, note that our implementation of unscaffolded guidance would be categorized as student-centered by the COPUS protocol (Lund et al., 2015; Smith, Jones, Gilbert, & Wieman, 2013), multiple-voice by the DART method (Owens et al., 2017), and high structure (Eddy & Hogan, 2014). Unscaffolded-guidance participants engaged in problem solving for nearly the entire class period, supported by an instructor and two learning assistants. Despite this design, near-transfer learning for the unscaffolded guidance condition was inferior to the three conditions receiving more guidance.

The general comparability of pedagogical approaches observed in the present study is somewhat surprising given the heated debates among their proponents. However, all the pedagogies tested have been shown to positively impact learning, so maybe their differences are not as important as previously suspected. Alternatively, the instructional approaches we

implemented may not differ enough from each other. We may have inadvertently omitted critical features from one or more of the methods that would have led to differential impacts on near- or far-transfer problem solving. For example, perhaps unscaffolded guidance could be successful if the number of learning assistants was increased. Another alternative explanation is that the dosage of instruction was not adequate. We might have detected differences if the instructional sessions lasted longer or if instruction spanned over multiple lessons, although a previous meta-analysis on instructional guidance indicates that the length of an instructional study does not impact its effect size (Lazonder & Harmsen, 2016). Finally, the limitations of our study may mask differential impacts (see Limitations).

Implications for Active Learning Instructors. Instructors new (and even somewhat new) to active learning frequently want to know whether one instructional method is better than another or whether there are things not to do. Our data provide much-needed guidance for these instructors. First, our data suggest that some variability in the nature and timing of guidance may be just fine for student learning. For example, instructors who struggle to see themselves going from straight lecture classrooms to guided inquiry (where most of the class period is spent in student work) may find productive failure as a potentially easier transition that appears to be equally effective. With productive failure, an instructor must carefully craft a challenging problem for the start of class and follow it by connecting students' solutions with the varieties of ways experts solve the problems (Kapur, 2016; Loibl et al., 2017). Crafting these problems may be challenging for a new active-learning instructor, but explicitly explaining the problem to students should feel familiar. Second, our data suggest that unscaffolded guidance should be avoided. Even though unscaffolded guidance looks and feels like active learning, it did not maximize near-transfer problem solving in our study. Getting students talking and working more in class, while a great first start, is not

sufficient to implement successful active learning. Instructors may experience less than optimal student outcomes if they only add clicker questions or challenging problems to their lessons. Instead, instructors should aim to create well-scaffolded lesson materials that break problems down into manageable pieces, sequence them carefully from start to finish of a lesson, and consider how they will introduce and follow-up questions. Lastly, and unsurprisingly, what students know coming into a classroom setting matters, and instructors should remember that eliciting students' prior knowledge is a worthwhile endeavor.

Potential Implications for Classroom Climate and Other Noncognitive Factors. Although our study did not investigate classroom climate or other noncognitive factors, future research should take these factors into consideration. First, different instructional methods may be better suited for building classroom equity. Classroom equity refers to promoting fairness in the classroom so that all students have the opportunity to participate, think, pose ideas, construct their knowledge, and feel welcomed into the intellectual discussion (Miller & Tanner, 2015; Tanner, 2013). In our study, only the scaffolded guidance condition provided all students with access to guidance from instructional materials along with opportunities to construct knowledge through interactions facilitated by the instructional team. Worked examples plus practice and productive failure provided all students with guidance through the problem-solving process. However, students had limited opportunities to pose ideas or to receive feedback from the instructor. On the other hand, the unscaffolded guidance condition provided all students with the chance to pose ideas and questions, construct knowledge, and participate in intellectual discussion. Yet the instructional materials themselves offered no guidance, only prompts to solve a challenging problem. These differences may impact students' perceptions of classroom equity. Second, various instructional methods may differentially interact with noncognitive aspects of student development. The amount

of struggle and level of challenge that students experience in worked examples plus practice, productive failure, and guided inquiry likely varies. Worked examples plus practice reduces challenge and provides explicit explanations and support, so little to no struggle is experienced. In contrast, productive failure and guided inquiry force students to struggle with the material and even to fail at solving problems correctly. Noncognitive aspects of student development that may interact with these instructional methods include, but are not limited to, motivation to learn, self-efficacy, and resilience (England, Brigati, Schussler, & Chen, 2019; Henry, Shorter, Charkoudian, Heemstra, & Corwin, 2019; Trujillo & Tanner, 2014). A potentially fruitful area of research would be to compare the impacts of each method on classroom equity and other noncognitive factors.

Conclusions

This work serves as a model of research that draws from theoretical and empirical work in educational psychology to inform classroom practice in biology, while refining context-specific boundaries for worked examples plus practice, productive failure, and guided inquiry approaches. Importantly, this work provides the first direct comparison of these approaches while simultaneously extending previous work on these pedagogies to the conceptual domain of biochemistry. The biochemistry lesson materials developed for this study target known student difficulties and help students craft explanations for near- and far-transfer problems. While this work advances both the fields of educational psychology and DBER, there is still more cross-disciplinary work to be done. Lessons developed here can be further improved by incorporating other known principles from educational psychology, like drawing to learn (Ainsworth, Prain, & Tytler, 2011; Fiorella & Zhang, 2018; Quillin & Thomas, 2015; Van Meter & Garner, 2005). Additionally, future work can address the question of whom these different pedagogies most benefit.

Acknowledgements

We would like to acknowledge the participants for their time and effort. We kindly thank Sarah Baas Robinson, Erin Dolan, Julie Dangremond Stanton, Regina Frey and Jennifer Loertscher for their helpful feedback during the development of our lesson materials. We thank Sarah Baas Robinson, Kush Bhatia, Jamie Pham, Justin Rubin, and Ave Fouriezos for their assistance with data collection. We also thank Vanessa Alele and Grace Snuggs for their assistance with coding and scoring. Special thanks to Dan Hall and Huimin Hu of the UGA Statistical Consulting Center and Hye-Jeong Choi for their advice with statistical analyses. We also express gratitude to our supportive colleagues in the Biology Education Research Group at UGA for their helpful feedback on earlier drafts of this manuscript. Paula Lemons would like to thank the UGA Study in a Second Discipline program for the fellowship that led to collaboration with Logan Fiorella. Stephanie Halmo would like to thank David Live for introducing her to Logan Fiorella's research and the ARCS Foundation for their financial support. This material is based upon work supported by the National Science Foundation under Grant No. DRL1350345.

References

Ainsworth, S., Prain, V., & Tytler, R. (2011). Drawing to Learn in Science. *Science, 333*(6046), 1096-1097. doi:10.1126/science.1204153

American Association for the Advancement of Science. (2011). *Vision and Change in Undergraduate Biology Education: A Call to Action*. Retrieved from http://www.visionandchange.org/

ASBMB. (2020). Core Concepts of Macromolecular Structure and Function. Retrieved from http://www.asbmb.org/education/teachingstrategies/foundationalconcepts/MacromolecularStructureFunction/

Bailey, C. P., Minderhout, V., & Loertscher, J. (2012). Learning transferable skills in large lecture halls: Implementing a POGIL approach in biochemistry. *Biochemistry and Molecular Biology Education, 40*(1), 1-7. doi:10.1002/bmb.20556

Becker, N., Noyes, K., & Cooper, M. (2016). Characterizing Students' Mechanistic Reasoning about London Dispsersion Forces. *Journal of Chemical Education, 93*, 1713-1724. doi:10.1021/acs.jchemed.6b00298

Beichner, R. J., & Saul, J. M. (2003). Introduction to the SCALE-UP (student-centered activities for large enrollment undergraduate programs) project. *Proceedings of the International School of Physics ‹‹Enrico Fermi››, Varenna, Italy*. Retrieved from https://projects.ncsu.edu/per/scaleup.html

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In *Metacognition: Knowing about knowing.* (pp. 185-205). Cambridge, MA, US: The MIT Press.

Bodner, G., Klobuchar, M., & Geelan, D. (2001). The Many Forms of Constructivism. *Journal of Chemical Education, 78*(8), 1107. doi:10.1021/ed078p1107.4

Chaplin, S. (2009). Assessment of the impact of case studies on student learning gains in an introductory biology course. *Journal of College Science Teaching, 39*(1), 72.

Cooper, M. M., Williams, L. C., & Underwood, S. M. (2015). Student Understanding of Intermolecular Forces: A Multimodal Study. *Journal of Chemical Education, 92*(8), 1288-1298. doi:10.1021/acs.jchemed.5b00169

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences, 116*(39), 19251-19257. doi:10.1073/pnas.1821936116

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: a meta-analysis. *Learning and Instruction, 13*(5), 533-568. doi:10.1016/s0959-4752(02)00025-7

Du Toit, M. (2003). *IRT from SSI: Bilog-MG, multilog, parscale, testfact*: Scientific Software International.

Eberlein, T., Kampmeier, J., Minderhout, V., Moog, R. S., Platt, T., Varma-Nelson, P., & White, H. B. (2008). Pedagogies of engagement in science. *Biochemistry and Molecular Biology Education, 36*(4), 262-273. doi:10.1002/bmb.20204

Eddy, S. L., & Hogan, K. A. (2014). Getting under the hood: how and for whom does increasing course structure work? *CBE Life Sci Educ, 13*(3), 453-468. doi:10.1187/cbe.14-03-0050

Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists.

England, B. J., Brigati, J. R., Schussler, E. E., & Chen, M. M. (2019). Student Anxiety and Perception of Difficulty Impact Performance and Persistence in Introductory Biology Courses. *CBE—Life Sciences Education, 18*(2), ar21. doi:10.1187/cbe.17-12-0284

Farrell, J. J., Moog, R. S., & Spencer, J. N. (1999). A Guided-Inquiry General Chemistry Course. *Journal of Chemical Education, 76*(4), 570. doi:10.1021/ed076p570

Fiorella, L., & Zhang, Q. (2018). Drawing Boundary Conditions for Learning by Drawing. *Educational Psychology Review, 30*(3), 1115-1137. doi:10.1007/s10648-018-9444-8

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410-8415. doi:10.1073/pnas.1319030111

Freeman, S., Haak, D., & Wenderoth, M. P. (2011). Increased course structure improves performance in introductory biology. *CBE Life Sci Educ, 10*(2), 175-186. doi:10.1187/cbe.10-08-0105

Gisev, N., Bell, J. S., & Chen, T. F. (2013). Interrater agreement and interrater reliability: key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy, 9*(3), 330-338.

Glogger-Frey, I., Fleischer, C., Grüny, L., Kappich, J., & Renkl, A. (2015). Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learning and Instruction, 39*, 72-87.

Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science, 332*(6034), 1213-1216. doi:10.1126/science.1204820

Halmo, S. M., Sensibaugh, C. A., Bhatia, K. S., Howell, A., Ferryanto, E. P., Choe, B., . . . Lemons, P. P. (2018). Student difficulties during structure–function problem solving. *Biochemistry and Molecular Biology Education, 46*(5), 453-463. doi:doi:10.1002/bmb.21166

Henry, M. A., Shorter, S., Charkoudian, L., Heemstra, J. M., & Corwin, L. A. (2019). FAIL Is Not a Four-Letter Word: A Theoretical Framework for Exploring Undergraduate Students' Approaches to Academic Challenge and Responses to Failure in STEM Learning Environments. *CBE—Life Sciences Education, 18*(1), ar11. doi:10.1187/cbe.18-06-0108

Herreid, C. F. (2007). Start with a story: The case study method of teaching college science: NSTA press.

Hmelo-Silver, C. E. (2004). Problem-Based Learning: What and How Do Students Learn? *Educational Psychology Review, 16*(3), 235-266. doi:10.1023/B:EDPR.0000034022.16470.f3

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99-107. doi:10.1080/00461520701263368

Hsu, C.-Y., Kalyuga, S., & Sweller, J. (2015). When should guidance be presented in physics instruction? *Archives of Scientific Psychology, 3*(1), 37.

Kalyuga, S., & Singh, A.-M. (2016). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review, 28*(4), 831-852.

Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26*(3), 379-424.

Kapur, M. (2011). A further study of productive failure in mathematical problem solving: Unpacking the design components. *Instructional Science, 39*(4), 561-579.

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist, 51*(2), 289-299.

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences, 21*(1), 45-83.

Kapur, M., & Rummel, N. (2012). Productive failure in learning from generation and invention activities. *Instructional Science, 40*(4), 645-650.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75-86.

Kline, P. (2000). *The handbook of psychological testing*: Psychology Press.

Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biol Educ, 4*(4), 298-310. doi:10.1187/05-06-0082

Lazonder, A. W., & Harmsen, R. (2016). Meta-Analysis of Inquiry-Based Learning:Effects of Guidance. *Review of Educational Research, 86*(3), 681-718. doi:10.3102/0034654315627366

Lewis, S. E., & Lewis, J. E. (2005). Departing from Lectures: An Evaluation of a Peer-Led Guided Inquiry Alternative. *Journal of Chemical Education, 82*(1), 135. doi:10.1021/ed082p135

Lewis, S. E., & Lewis, J. E. (2008). Seeking effectiveness and equity in a large college chemistry course: an HLM investigation of Peer-Led Guided Inquiry. *Journal of Research in Science Teaching, 45*(7), 794-811. doi:10.1002/tea.20254

Loertscher, J., Green, D., Lewis, J. E., Lin, S., & Minderhout, V. (2014). Identification of threshold concepts for biochemistry. *CBE-Life Sciences Education, 13*(3), 516-528.

Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review*, *29*(4), 693-715.

Lund, T. J., Pilarz, M., Velasco, J. B., Chakraverty, D., Rosploch, K., Undersander, M., & Stains, M. (2015). The best of both worlds: Building on the COPUS and RTOP observation protocols to easily and reliably measure various levels of reformed instructional practice. *CBE Life Sci Educ, 14*(2). doi:10.1187/cbe.14-10-0168

Mayer, R. E. (2004). Should There Be a Three-Strikes Rule Against Pure Discovery Learning? *American Psychologist, 59*(1), 14-19. doi:10.1037/0003-066X.59.1.14

McDaniel, M. A., Cahill, M. J., Frey, R. F., Rauch, M., Doele, J., Ruvolo, D., & Daschbach, M. M. (2018). Individual Differences in Learning Exemplars Versus Abstracting Rules: Associations with Exam Performance in College Science. *Journal of Applied Research in Memory and Cognition, 7*(2), 241-251. doi:https://doi.org/10.1016/j.jarmac.2017.11.004

McDaniel, M. A., Mestre, J. P., Frey, R. F., Gouravajhala, R., Hilborn, R. C., Miyatsu, T., & Yuan, H. (2017). Maximizing undergraduate STEM learning: Promoting research at the intersection of cognitive psychology and discipline-based education research. In.

Miller, S., & Tanner, K. D. (2015). A Portal into Biology Education: An Annotated List of Commonly Encountered Terms. *CBE—Life Sciences Education, 14*(2), fe2. doi:10.1187/cbe.15-03-0065

Owens, M. T., Seidel, S. B., Wong, M., Bejines, T. E., Lietz, S., Perez, J. R., . . . Tanner, K. D. (2017). Classroom sound can be used to classify teaching practices in college science courses. *Proceedings of the National Academy of Sciences, 114*(12), 3085. doi:10.1073/pnas.1618693114

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1-4.

Prince, M. (2004). Does Active Learning Work? A Review of the Research. *Journal of Engineering Education, 93*(3), 223-231. doi:10.1002/j.2168-9830.2004.tb00809.x

Prince, M. J., & Felder, R. M. (2006). Inductive Teaching and Learning Methods: Definitions, Comparisons, and Research Bases. *Journal of Engineering Education, 95*(2), 123-138. doi:10.1002/j.2168-9830.2006.tb00884.x

Quillin, K., & Thomas, S. (2015). Drawing-to-Learn: A Framework for Using Drawings to Promote Model-Based Reasoning in Biology. *CBE—Life Sciences Education, 14*(1), es2. doi:10.1187/cbe.14-08-0128

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science, 38*(1), 1-37.

Rittle-Johnson, B., & Schneider, M. (2015). Developing conceptual and procedural knowledge of mathematics. In *The Oxford handbook of numerical cognition.* (pp. 1118-1134). New York, NY, US: Oxford University Press.

Schmidt, H. G., Loyens, S. M. M., Van Gog, T., & Paas, F. (2007). Problem-Based Learning is Compatible with Human Cognitive Architecture: Commentary on Kirschner, Sweller, and. *Educational Psychologist, 42*(2), 91-97. doi:10.1080/00461520701263350

Schmidt, R. A., & Bjork, R. A. (1992). New Conceptualizations of Practice: Common Principles in Three Paradigms Suggest New Concepts for Training. *Psychological Science, 3*(4), 207-218. doi:10.1111/j.1467-9280.1992.tb00029.x

Schwartz, D. L., & Bransford, J. D. (1998). A Time For Telling. *Cognition and Instruction, 16*(4), 475-5223. doi:10.1207/s1532690xci1604_4

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology, 103*(4), 759-775. doi:10.1037/a0025140

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction, 22*(2), 129-184.

Smith, M. K., Jones, F. H. M., Gilbert, S. L., & Wieman, C. (2013). The classroom observation protocol for undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE-Life Sciences Education, 12*, 618-627.

Steenhof, N., Woods, N. N., Van Gerven, P. W. M., & Mylopoulos, M. (2019). Productive failure as an instructional approach to promote future learning. *Advances in Health Sciences Education, 24*(4), 739-749. doi:10.1007/s10459-019-09895-4

Sweller, J. (2016). Working memory, long-term memory, and instructional design. *Journal of Applied Research in Memory and Cognition, 5*(4), 360-367.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*(1), 59-89.

Sweller, J., Kirschner, P. A., & Clark, R. E. (2007). Why Minimally Guided Teaching Techniques Do Not Work: A Reply to Commentaries. *Educational Psychologist, 42*(2), 115-121. doi:10.1080/00461520701263426

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296.

Tanner, K. D. (2013). Structure matters: twenty-one teaching strategies to promote student engagement and cultivate classroom equity. *CBE life sciences education, 12*(3), 322-331. doi:10.1187/cbe.13-06-0115

Trujillo, G., & Tanner, K. D. (2014). Considering the Role of Affect in Learning: Monitoring Students' Self-Efficacy, Sense of Belonging, and Science Identity. *CBE—Life Sciences Education, 13*(1), 6-15. doi:10.1187/cbe.13-12-0241

van Merriënboer, J. J. G., & Kirschner, P. A. (2007). *Ten steps to complex learning: A systematic approach to four-component instructional design*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Van Meter, P., & Garner, J. (2005). The Promise and Practice of Learner-Generated Drawing: Literature Review and Synthesis. *Educational Psychology Review, 17*(4), 285-325. doi:10.1007/s10648-005-8136-3

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam med, 37*(5), 360-363.

Weaver, J. P., Chastain, R. J., DeCaro, D. A., & DeCaro, M. S. (2018). Reverse the routine: Problem solving before instruction improves conceptual knowledge in undergraduate physics. *Contemporary Educational Psychology, 52*, 36-47. doi:10.1016/j.cedpsych.2017.12.003

Werth, M. T. (2017). Serotonin in the Pocket: Non-covalent interactions and neurotransmitter binding. *CourseSource*. Retrieved from https://doi.org/10.24918/cs.2017.14

Zimowski, M. F., Muraki, E., Mislevy, R., & Bock, R. (1996). BILOG-MG: Multiple group IRT analysis and test maintenance for binary items. Chicago: Scientific Software International.

Supplemental Material

Supplemental Material Table S7.1. Demographic information of research participants

	Gender	Race/Ethnicity	Average GPA
Total (N=189)	81.5% Female (n=154) 18.5% Male (n=35)	23.3% Asian (n=44) 19.6% Black or African-American (n=37) 4.8% Hispanic or Latinx (n=9) 1.6% Not reported (n=3) 2.1% Two or more races (n=4) 48.7% White (n=92)	3.46
Worked examples plus practice (n=41)	75.6% Female (n=31) 24.4% Male (n=10)	9.8% Asian (n=4) 17.1% Black or African-American (n=7) 7.3% Hispanic or Latinx (n=3) 0% Not reported (n=0) 2.4% Two or more races (n=1) 63.4% White (n=26)	3.40
Productive failure (n=78)	88.5% Female (n=69) 11.5% Male (n=9)	25.6% Asian (n=20) 25.6% Black or African-American (n=20) 5.1% Hispanic or Latinx (n=4) 2.6% Not reported (n=2) 1.3% Two or more races (n=1) 39.7% White (n=31)	3.50
Unscaffolded guidance (n=42)	66.7% Female (n=28) 33.3% Male (n=14)	19% Asian (n=8) 11.9% Black or African-American (n=5) 4.8% Hispanic or Latinx (n=2) 0% Not reported (n=0) 4.8% Two or more races (n=2) 59.5% White (n=25)	3.41
Scaffolded guidance (n=28)	92.9% Female (n=26) 7.1% Male (n=2)	42.9% Asian (n=12) 17.9% Black or African-American (n=5) 0% Hispanic or Latinx (n=0) 3.6% Not reported (n=1) 0% Two or more races (n=0) 35.7% White (n=10)	3.53

Supplemental Material Table S7.2. Unadjusted basic knowledge pre- and posttest performance

	Pretest Mean (s.d.)	Posttest Mean (s.d.)
Worked examples plus practice (n=41)	-0.98 (0.77)	-0.44 (0.57)
Productive failure (n=78)	-0.81 (0.67)	-0.60 (0.66)
Unscaffolded guidance (n=42)	-0.95 (0.73)	-0.44 (0.69)
Scaffolded guidance (n=28)	-0.68 (0.63)	-0.70 (0.48)

CHAPTER 8

CONCLUSIONS

This chapter describes major contributions of the studies within each part of this dissertation and plans for future research. It closes by placing both parts in broader context and significance.

POMGNT2 Specificity in Protein *O*-mannosylation

Conclusions. The first half of this dissertation examines the biochemical determinants of POMGNT2 activity for functional core m3 *O*-mannosylation biosynthesis on α-dystroglycan. α-dystroglycan is extensively *O*-mannosylated (Stalnaker, S.H., Hashmi, S., et al. 2010, Vester-Christensen, M.B., Halim, A., et al. 2013). A unique glycan structure, termed matriglycan, is responsible for α-dystroglycan's interactions with laminin at the cell surface (Yoshida-Moriguchi, T. and Campbell, K.P. 2015). Matriglycan is synthesized on core m3 O-mannose glycan structures. Core m3 structures are found only on two sites, Thr317 and Thr379, on α-dystroglycan (Yoshida-Moriguchi, T., Willer, T., et al. 2013). POMGNT2 catalyzes the first committed step towards the functional matriglycan structure on α-dystroglycan by transferring a β-1,4-linked N-acetylglucosamine to a mannose on Thr residues of α-dystroglycan (Yagi, H., Nakagawa, N., et al. 2013). Alternatively, protein *O*-linked mannose β-1,2-*N*-acetylglucosaminyltransferase 1 (POMGNT1) catalyzes the first step toward other various glycan structures present on α-dystroglycan of unknown function (Sheikh, M.O., Halmo, S.M., et al. 2017). Chapter 3 investigates differential *in vitro* transfer of N-acetylglucosamine by POMGNT1 and POMGNT2 to synthetic glycopeptides (Halmo, S.M., Singh, D., et al. 2017). Using a combination of

luminescent and radioactive transfer assays verified by mass spectrometry, kinetic parameters of recombinant POMGNT1 and POMGNT2 were determined for various synthetic glycopeptides substrates based on the α-dystroglycan sequence. This works demonstrates that POMGNT1 is promiscuous for *O*-mannosylated peptides, whereas POMGNT2 displays significant primary amino acid selectivity near the site of *O*-mannosylation. Sequence alignment of α-dystroglycan sites that permit POMGNT2 extension defines a POMGNT2 acceptor motif in α-dystroglycan conserved among 59 vertebrate species. Further enzymatic assays reveal that when engineered into a POMGNT1-only site, the acceptor motif is sufficient to convert the *O*-mannosylated peptide to a substrate for POMGNT2. Conversely, when conserved amino acids in the motif are replaced by divergent amino acids, the acceptor glycopeptide is a less efficient substrate for POMGNT2. Chapter 4 expands on this work by moving *in cellulo* to demonstrate that the introduction of the minimal POMGNT2 acceptor motif is sufficient for functional matriglycan extension at novel sites on α-dystroglycan. This functional matriglycan extension is enhanced by addition of exogenous LARGE1. Taken together, these findings contribute to the current working hypothesis that POMGNT2 functions as a gatekeeper enzyme to prevent the vast majority of *O*-mannosylated sites on α-dystroglycan from being modified with functional matriglycan. This work helps to fill the gap in knowledge about how specificity is achieved in the first committed step towards core m3 *O*-mannosylation on α-dystroglycan. Chapters 3 and 4 lay important groundwork for further investigation into the biosynthesis of matriglycan on α-dystroglycan and core m3 structures on other *O*-mannosylated proteins.

Future Directions. With the complete pathway and identification of the enzymes involved in its biosynthesis known, the characterization of additional regulatory factors in classical *O*-mannosylation serves as a clear next step. The work presented here sets the stage for determining

how specificity is achieved at each step in the pathway. This will involve uncovering the mechanisms behind the unique specificities of the remaining pathway enzymes, including B3GALNT2, POMK, FKTN, FKRP, RXYLT1, B4GAT1, LARGE1, LARGE2 and HNK-1ST (Sheikh, M.O., Halmo, S.M., et al. 2017). Much of this work is already underway. The recombinant production of fusion-tagged enzymes in high yields as described in Chapter 3 will aid in these endeavors (Moremen, K.W., Ramiah, A., et al. 2018). Crystallographic studies of the pathway enzymes in complex with their glycan and glycopeptide acceptor substrates will further our understanding of these glycosyltransferases. For example, the recently elucidated POMGNT1, POMK, and FKRP crystal structures, provide glycan substrate recognition mechanisms (Kuwabara, N., Imae, R., et al. 2020, Kuwabara, N., Manya, H., et al. 2016, Zhu, Q., Venzke, D., et al. 2016). Collaborations have led to an apo crystal structure of POMGNT2, but the acceptor complex remains elusive. Combined, these efforts will generate more testable hypotheses for this unique glycan pathway and the resulting glycoprotein, α-dystroglycan, which is of functional significance in developmental and pathogenic diseases.

In regards to POMGNT2, the following could further our understanding of its specificity. Based on the findings presented in Chapter 4, assessing whether functional matriglycan can be moved onto other proteins, such as KIAA1549 and phosphacan, by engineering in the POMGNT2 acceptor motif serves as a clear next step. Functional matriglycan extension on these proteins is unlikely because they lack the N-terminal domain of α-DG which is required for LARGE activity (Kanagawa, M., Saito, F., et al. 2004). However, the absence of the N-terminal domain of α-DG does not exclude the possibility of POMGNT2 activity. Precise identification of primary amino acid determinants for POMGNT2 acceptors is needed and could be accomplished using single point mutation studies and random glycopeptide libraries. This information would assist in future

attempts to obtain an acceptor complex in crystallography experiments. Doubly mannosylated acceptor glycopeptides, where both Thr317 and Thr319 or Thr379 and Thr381 in the TPT portion of the acceptor motif, have been synthesized. Preliminary *in vitro* evidence suggests that both O-mannosylated threonines can be extended by POMGNT2. The fact that the two arginines and the two threonines in the minimal POMGNT2 acceptor motif are separated by one amino acid is intriguing, and leads one to speculate if this specific motif permits the enzyme to slide along the acceptor motif and act on both threonines. The identification of other POMGNT2 substrates aside from α-dystroglycan may expand our understanding of POMGNT2 specificity. Degenerate motif searches will identify candidate proteins known to traverse the secretory pathway, which can be cross referenced with the list of currently known POMT1/POMT2-dependent *O*-mannosylated proteins (see Table 1.1). Given that these candidate proteins will lack the N-terminal domain of α-dystroglycan that is required for LARGE activity (Kanagawa, M., Saito, F., et al. 2004), it is unlikely that probing for terminal IIH6 and laminin reactivity will prove fruitful. Therefore, it is proposed that these candidate proteins be recombinantly expressed in POMGNT1 knockout cells lines and then submitted to glycopeptide analysis. The search would need to include modifications ranging from a simple Hex-HexNAc modification that would correspond to POMGNT2-specific GlcNAc extended mannose all the way to the heptasaccharide glycan structure that results after B4GAT1 activity and prior to LARGE activity. In addition, these candidate proteins could be recombinantly expressed in POMGNT1 and POMGNT2 double knockout cell lines and then assessed for POMGNT2 extension using *in vitro* transfer assays and site-mapping methods of detection.

Other than α-dystroglycan, several proteins are known to be *O*-mannosylated (see Table 1.1). Characterization of the complete glycan structures present on other *O*-mannosylated proteins

and their functions are needed. Given the brain and eye abnormalities present in some patients with Congenital Muscular Dystrophy caused by mutations in *O*-mannosylation pathway enzymes (Godfrey, C., Foley, A.R., et al. 2011, Manzini, M.C., Tambunan, Dimira E., et al. 2012), *O*-mannosylated proteins highly expressed in these tissues are of particular interest to future studies. For example, KIAA1549 is extensively *O*-mannosylated in a POMT1/POMT2-dependent fashion and is highly expressed in the brain and retina (Larsen, I.S.B., Narimatsu, Y., et al. 2019). However, it is unknown if the *O*-mannose residues on KIAA1549 are further extended by POMGNT1 or POMGNT2, or what functional role these glycans may serve. Interestingly, KIAA1549 and POMGNT1 are both genes that harbor mutations associated with a disease of the eye called non-syndromic autosomal recessive retinitis pigmentosa (de Bruijn, S.E., Verbakel, S.K., et al. 2018, Wang, N.H., Chen, S.J., et al. 2016, Xu, M., Yamada, T., et al. 2016), suggesting their gene products may be involved in the same pathway. Inspection of the primary amino acid sequence of KIAA1549 reveals a portion of the POMGNT2 acceptor motif around *O*-mannosylated threonine 245. If not naturally extended by POMGNT2, this site could serve as a potential site for POMGNT2 motif engineering.

Problem Solving in Undergraduate Biochemistry

Conclusions. The second half of this dissertation examines student thinking and problem solving around a core concept in biochemistry and aims to contribute to the collective effort of improving undergraduate science education (AAAS 2010). The physical basis of noncovalent interactions deals with the fact that all noncovalent interactions result due to electrostatic properties of molecules, despite variation in the nature of these properties (Loertscher, J., Green, D., et al. 2014). Chapter 6 investigates student and expert thinking during structure-function problem solving that deals with noncovalent interactions (Halmo, S.M., Sensibaugh, C.A., et al. 2018). This

work revealed domain-general and domain-specific knowledge that participants brought to bear on the problem, and identified several student difficulties. Student difficulties included getting distracted by the amino acid backbone instead of focusing on amino acid side chains, categorizing amino acid side chains but not using the categorizations in future steps, and struggling to make predictions about noncovalent interactions that include electrostatic mechanisms for their formation. Chapter 7 aims to address these identified student difficulties through various instructional approaches (Halmo, S.M., Sensibaugh, C.A., et al. in press). Specifically, Chapter 7 investigates the comparative impacts of four forms of active learning instruction on student learning of the physical basis of noncovalent interactions. The forms of instruction tested vary in the nature and timing of guidance. Worked examples plus practice provided direct guidance through worked examples before problem practice. Productive failure provided direct guidance after a problem exploration phase. Scaffolded guidance provided guidance though instructional materials and instructors that faded away as knowledge was built. Unscaffolded guidance provided guidance only through instructors that faded away as knowledge was built. Our comparison indicated that the four pedagogies did not differentially impact basic knowledge performance. However, worked examples plus practice, productive failure, and scaffolded guidance led to greater near-transfer performance compared to unscaffolded guidance, suggesting that more than unscaffolded guidance is necessary for near-transfer problem solving. Additionally, we show that productive failure and scaffolded guidance did not differentially impact far-transfer performance. Taken together, these results provide biochemistry instructors with insight into student thinking around noncovalent interactions and instructional materials that can be employed in the classroom to address known difficulties. Chapters 6 and 7 lay important groundwork for further investigation

into the learning of noncovalent interactions in biochemistry and other important core concepts in the domain.

Future Directions. One clear next step is to assess the comparative impacts of worked examples plus practice, productive failure, and scaffolded guidance on far-transfer performance. This remains an important unanswered question in this line of work and in the field (Kapur, M. 2016). This work is currently underway with data collected in the ecologically valid environment of biochemistry classrooms. There is also a need to investigate the interaction of other variables with transfer performance, such as classroom equity, individual students' motivation to learn, and science self-efficacy (Glynn, S.M. and Koballa, T.R., Jr. 2006, Reinholz, D.L. and Shah, N. 2018).

The constructivist lens used throughout this work acknowledges sociocultural aspects of learning but does not prioritize these aspects in practice, especially when considering the classical cognitive perspective of transfer (Lobato, J. 2006). Namely, measuring performance based on expert models may underestimate the amount of transfer that actually occurs. In other words, performance may be an unreliable index of how much learning has occurred. Reconceptualizing transfer from a situated cognition perspective would require the use of methods rooted in a different theory. If this were undertaken, transfer could be measured by investigating if student solutions or reasoning in the transfer situation were influenced by learning experiences from instruction (Bransford, J.D. and Schwartz, D.L. 1999). Some evidence of this exists in our data by the presence of the tentative statement "further experimentation is needed to confirm this prediction" in student's transfer responses. This statement was explicitly presented in instruction but not accounted for in our scoring rubrics as it is not essential to the core concept of noncovalent interactions. Emphasizing the situated role and social aspects of cognition might mean moving away from the use of independent problem-solving performance as a measure of transfer entirely.

In this methodological reconceptualization, transfer might be best measured using group assessments where the use of resources is permitted. Interactions with peers and artifacts during the transfer task would become possible data sources. Methods of this nature would also be inclusive to cultural knowledge that may be absent from the current focus on written student responses that implicitly privileges schooling practices like grammar and sentence construction (National Academies of Sciences, E. and Medicine 2018).

The second half of this dissertation focused on a single core concept in biochemistry, and findings presented here cannot be generalizable to other topic areas. Similar lines of research can be pursued with other core concepts in biochemistry, such as metabolism, steady state, and thermodynamics (Loertscher, J., Green, D., et al. 2014).

Closing Remarks

This dissertation is unique because of its seemingly disparate parts. The research contained within splits across two fields: biochemistry and discipline-based education research. The first part deals with understanding the transfer of *N*-acetylglucosamine to certain *O*-mannosylation sites on α-dystroglycan and probing the specificity involved. The second part deals with understanding how transfer of the concept of noncovalent interactions can be best supported in biochemistry learning. Removing the researcher from this tale of two transfers obscures its connected purpose. Thus, I must interject myself in these concluding remarks.

My goal with this dissertation was to contribute to the ever-growing knowledge of biochemistry and how learning occurs in this field. This will serve me well in my future efforts as learner and teacher or biochemist and biochemistry instructor. Jointly pursuing training in both types of research has enabled me to build expertise in two fields. As I reflect on my body of work and my experiences, I prefer the model of *accomplished novice* proposed by the authors of How

People Learn (National Research Council 2000). Completing this dissertation led to skill development in many areas, but I recognize that my knowledge on these two subjects, while deep, is microscopic compared to all that is potentially knowable. There is still much more to learn, discover, and share.

References

AAAS. 2010. Vision and Change: A Call to Action. Washington, DC.

Bransford JD, Schwartz DL. 1999. Chapter 3: Rethinking transfer: A simple proposal with multiple implications. *Review of research in education*, 24:61-100.

de Bruijn SE, Verbakel SK, de Vrieze E, Kremer H, Cremers FPM, Hoyng CB, van den Born LI, Roosing S. 2018. Homozygous variants in KIAA1549, encoding a ciliary protein, are associated with autosomal recessive retinitis pigmentosa. *Journal of medical genetics*, 55:705-712.

Glynn SM, Koballa TR, Jr. 2006. Motivation to Learn in College Science. In: Mintzes JJ, Leonard WH editors. *Handbook of College Science Teaching*: NSTA Press.

Godfrey C, Foley AR, Clement E, Muntoni F. 2011. Dystroglycanopathies: coming into focus. *Current Opinion in Genetics & Development*, 21:278-285.

Halmo SM, Sensibaugh CA, Bhatia KS, Howell A, Ferryanto EP, Choe B, Kehoe K, Watson M, Lemons PP. 2018. Student difficulties during structure–function problem solving. *Biochemistry and Molecular Biology Education*, 46:453-463.

Halmo SM, Sensibaugh CA, Reinhart P, Stogniy O, Fiorella L, Lemons PP. in press. Advancing the Guidance Debate: Lessons from Educational Psychology and Implications for Biochemistry Learning. *CBE Life Sci Educ.*

Halmo SM, Singh D, Patel S, Wang S, Edlin M, Boons GJ, Moremen KW, Live D, Wells L. 2017. Protein O-Linked Mannose beta-1,4-N-Acetylglucosaminyl-transferase 2 (POMGNT2) Is a Gatekeeper Enzyme for Functional Glycosylation of alpha-Dystroglycan. *J Biol Chem*, 292:2101-2109.

Kanagawa M, Saito F, Kunz S, Yoshida-Moriguchi T, Barresi R, Kobayashi YM, Muschler J, Dumanski JP, Michele DE, Oldstone MBA, *et al.* 2004. Molecular Recognition by LARGE Is Essential for Expression of Functional Dystroglycan. *Cell*, 117:953-964.

Kapur M. 2016. Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist*, 51:289-299.

Kuwabara N, Imae R, Manya H, Tanaka T, Mizuno M, Tsumoto H, Kanagawa M, Kobayashi K, Toda T, Senda T, *et al.* 2020. Crystal structures of fukutin-related protein (FKRP), a ribitol-phosphate transferase related to muscular dystrophy. *Nature communications,* 11:303.

Kuwabara N, Manya H, Yamada T, Tateno H, Kanagawa M, Kobayashi K, Akasaka-Manya K, Hirose Y, Mizuno M, Ikeguchi M, *et al.* 2016. Carbohydrate-binding domain of the POMGnT1 stem region modulates O-mannosylation sites of α-dystroglycan. *Proceedings of the National Academy of Sciences,* 113:9280-9285.

Larsen ISB, Narimatsu Y, Clausen H, Joshi HJ, Halim A. 2019. Multiple distinct O-Mannosylation pathways in eukaryotes. *Curr Opin Struct Biol,* 56:171-178.

Lobato J. 2006. Alternative perspectives on the transfer of learning: History, issues, and challenges for future research. *The Journal of the Learning Sciences,* 15:431-449.

Loertscher J, Green D, Lewis JE, Lin S, Minderhout V. 2014. Identification of threshold concepts for biochemistry. *CBE-Life Sciences Education,* 13:516-528.

Manzini MC, Tambunan Dimira E, Hill RS, Yu Tim W, Maynard Thomas M, Heinzen Erin L, Shianna Kevin V, Stevens Christine R, Partlow Jennifer N, Barry Brenda J, *et al.* 2012. Exome Sequencing and Functional Validation in Zebrafish Identify GTDC2 Mutations as a Cause of Walker-Warburg Syndrome. *The American Journal of Human Genetics,* 91:541-547.

Moremen KW, Ramiah A, Stuart M, Steel J, Meng L, Forouhar F, Moniz HA, Gahlay G, Gao Z, Chapla D, *et al.* 2018. Expression system for structural and functional studies of human glycosylation enzymes. *Nature chemical biology,* 14:156-162.

National Academies of Sciences E, Medicine. 2018. *How People Learn II: Learners, Contexts, and Cultures* The National Academies Press:Washington, DC.

National Research Council. 2000. *How people learn: Brain, mind, experience, and school (Expanded edition)* National Academies Press:Washington, D.C.

Reinholz DL, Shah N. 2018. Equity analytics: A methodological approach for quantifying participation patterns in mathematics classroom discourse. *Journal for Research in Mathematics Education,* 49:140-177.

Sheikh MO, Halmo SM, Wells L. 2017. Recent advancements in understanding mammalian O-mannosylation. *Glycobiology,* 27:806-819.

Stalnaker SH, Hashmi S, Lim J-M, Aoki K, Porterfield M, Gutierrez-Sanchez G, Wheeler J, Ervasti JM, Bergmann C, Tiemeyer M, *et al.* 2010. Site Mapping and Characterization of O-Glycan Structures on α-Dystroglycan Isolated from Rabbit Skeletal Muscle. *The Journal of Biological Chemistry*, 285:24882-24891.

Vester-Christensen MB, Halim A, Joshi HJ, Steentoft C, Bennett EP, Levery SB, Vakhrushev SY, Clausen H. 2013. Mining the O-mannose glycoproteome reveals cadherins as major O-mannosylated glycoproteins. *Proceedings of the National Academy of Sciences*, 110:21018-21023.

Wang NH, Chen SJ, Yang CF, Chen HW, Chuang HP, Lu YH, Chen CH, Wu JY, Niu DM, Chen YT. 2016. Homozygosity Mapping and Whole-Genome Sequencing Links a Missense Mutation in POMGNT1 to Autosomal Recessive Retinitis Pigmentosa. *Investigative ophthalmology & visual science*, 57:3601-3609.

Xu M, Yamada T, Sun Z, Eblimit A, Lopez I, Wang F, Manya H, Xu S, Zhao L, Li Y, *et al.* 2016. Mutations in POMGNT1 cause non-syndromic retinitis pigmentosa. *Human molecular genetics*, 25:1479-1488.

Yagi H, Nakagawa N, Saito T, Kiyonari H, Abe T, Toda T, Wu S-W, Khoo K-H, Oka S, Kato K. 2013. AGO61-dependent GlcNAc modification primes the formation of functional glycans on α-dystroglycan. *Scientific Reports*, 3:3288.

Yoshida-Moriguchi T, Campbell KP. 2015. Matriglycan: a novel polysaccharide that links dystroglycan to the basement membrane. *Glycobiology*, 25:702-713.

Yoshida-Moriguchi T, Willer T, Anderson ME, Venzke D, Whyte T, Muntoni F, Lee H, Nelson SF, Yu L, Campbell KP. 2013. SGK196 Is a Glycosylation-Specific O-Mannose Kinase Required for Dystroglycan Function. *Science*, 341:896-899.

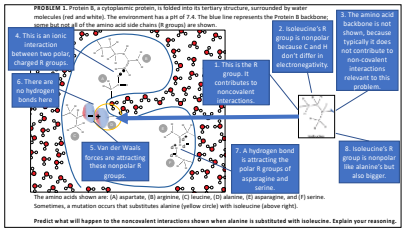
Zhu Q, Venzke D, Walimbe AS, Anderson ME, Fu Q, Kinch LN, Wang W, Chen X, Grishin NV, Huang N, *et al.* 2016. Structure of protein O-mannose kinase reveals a unique active site architecture. *eLife*, 5:e22238.

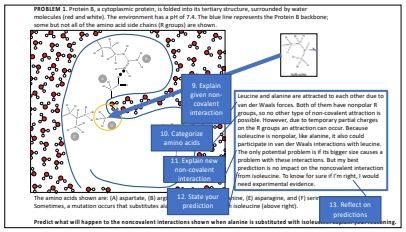
APPENDIX A

EXTENDED SUPPLEMENTAL MATERIAL FOR CHAPTER 7

Worked Examples Plus Practice.

Problem 1: Worked Example

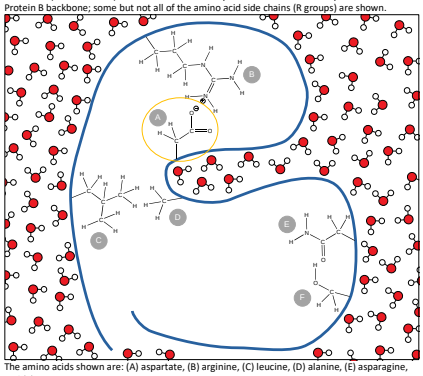




1

Problem 2:

Protein B, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and white). The environment has a pH of 7.4. The blue line represents the Protein B backbone; some but not all of the amino acid side chains (R groups) are shown.



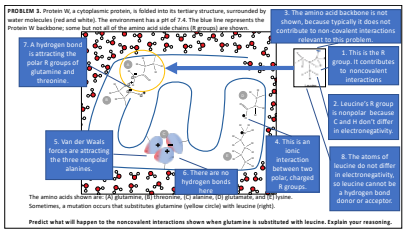
The amino acids shown are: (A) aspartate, (B) arginine, (C) leucine, (D) alanine, (E) asparagine, and (F) serine.

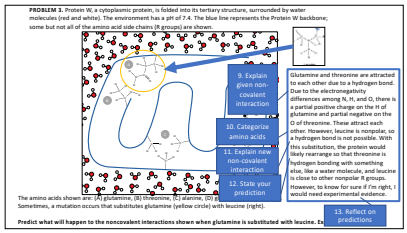
Sometimes, a mutation occurs that substitutes aspartate (yellow circle) with valine (below).

Predict what will happen to the noncovalent interactions shown when aspartate is substituted with valine. Explain your reasoning.

2

Problem 3: Worked Example

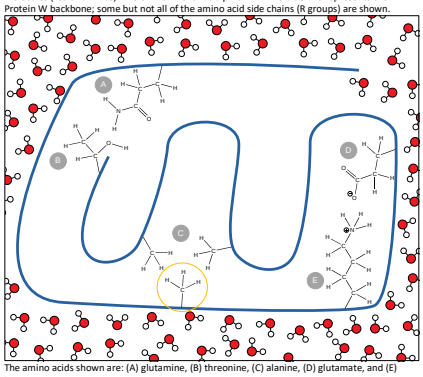




3

Problem 4:

Protein W, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and white). The environment has a pH of 7.4. The blue line represents the Protein W backbone; some but not all of the amino acid side chains (R groups) are shown.



The amino acids shown are: (A) glutamine, (B) threonine, (C) alanine, (D) glutamate, and (E) lysine.

Sometimes, a mutation occurs that substitutes alanine (yellow circle) with serine (below).

Predict what will happen to the noncovalent interactions shown when alanine is substituted with serine. Explain your reasoning.

4

Name: _____

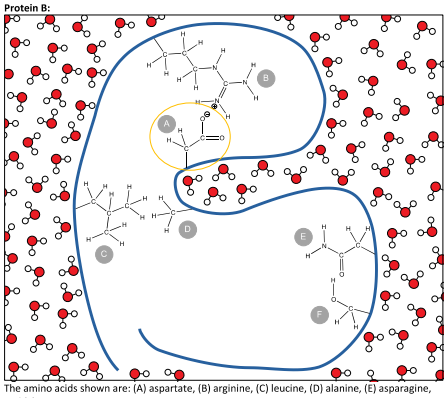
Problem Worksheet

Learning Objectives:

- Given a model of biological molecules, identify, categorize, and compare the important features.
- Recall that chemical composition affects the structure and interactions of macromolecules.
- Compare different types of noncovalent interactions occurring in biological molecules.
- Explain why and how it is that all noncovalent interactions are attractions of opposite charges.
- Predict the impact of changes in the building blocks of biological molecules on the structure of the macromolecules and formulate a scientific explanation for your prediction.
- Solve new problems that extend your thinking about the role of noncovalent interactions across biology.

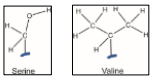
Protein B and Protein W, both cytoplasmic proteins, are folded into their tertiary structure, surrounded by water molecules (red and white). The environment has a pH of 7.4. The blue line represents the protein backbone; some but not all of the amino acid side chains (R groups) are shown for each protein. Sometimes, mutations occur that substitute aspartate (yellow circle) in Protein B and alanine (green circle) in Protein W with serine or valine (below).

Protein B:



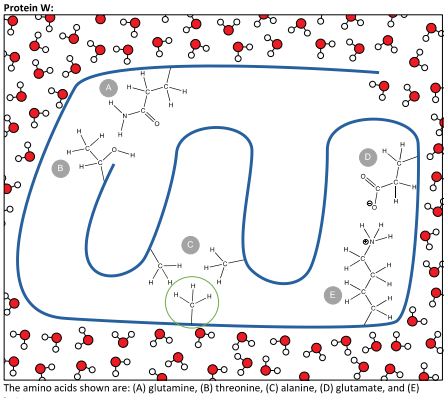
The amino acids shown are: (A) aspartate, (B) arginine, (C) leucine, (D) alanine, (E) asparagine, and (F) serine.

Substituted Amino Acids:



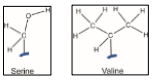
2

Protein W:



The amino acids shown are: (A) glutamine, (B) threonine, (C) alanine, (D) glutamate, and (E) lysine.

Substituted Amino Acids:



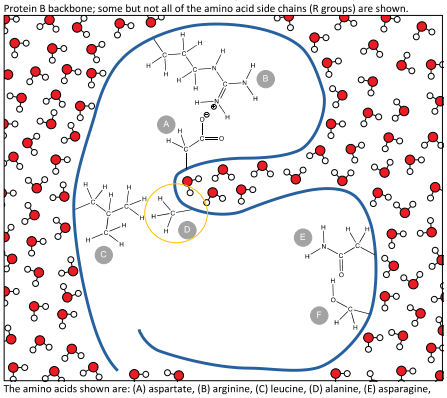
Predict the effect of each amino acid substitution on the existing noncovalent interactions. Compare and contrast the effects of the four amino acid substitutions. Explain your reasoning.

3

Unscaffolded Guidance.

Problem 1:

Protein B, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and white). The environment has a pH of 7.4. The blue line represents the Protein B backbone; some but not all of the amino acid side chains (R groups) are shown.



The amino acids shown are: (A) aspartate, (B) arginine, (C) leucine, (D) alanine, (E) asparagine, and (F) serine.

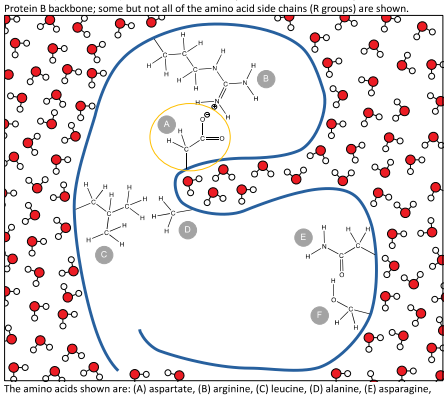
Sometimes, a mutation occurs that substitutes alanine (yellow circle) with isoleucine (below).

Predict what will happen to the noncovalent interactions shown when alanine is substituted with isoleucine. Explain your reasoning.

1

Problem 2:

Protein B, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and white). The environment has a pH of 7.4. The blue line represents the Protein B backbone; some but not all of the amino acid side chains (R groups) are shown.



The amino acids shown are: (A) aspartate, (B) arginine, (C) leucine, (D) alanine, (E) asparagine, and (F) serine.

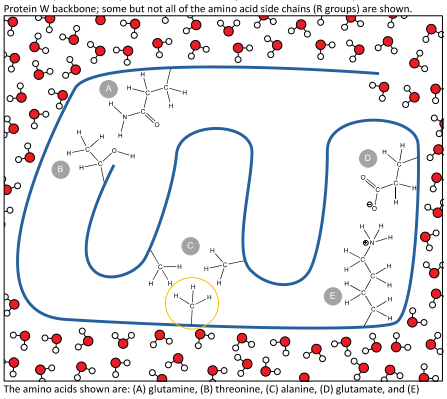
Sometimes, a mutation occurs that substitutes aspartate (yellow circle) with valine (below).

Predict what will happen to the noncovalent interactions shown when aspartate is substituted with valine. Explain your reasoning.

2

Problem 3:

Protein W, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and white). The environment has a pH of 7.4. The blue line represents the Protein W backbone; some but not all of the amino acid side chains (R groups) are shown.



The amino acids shown are: (A) glutamine, (B) threonine, (C) alanine, (D) glutamate, and (E) lysine.

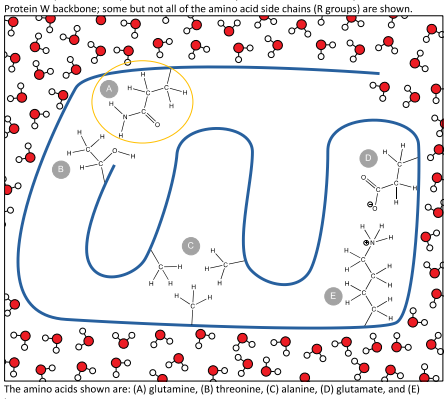
Sometimes, a mutation occurs that substitutes alanine (yellow circle) with serine (below).

Predict what will happen to the noncovalent interactions shown when alanine is substituted with serine. Explain your reasoning.

3

Problem 4:

Protein W, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and white). The environment has a pH of 7.4. The blue line represents the Protein W backbone; some but not all of the amino acid side chains (R groups) are shown.



The amino acids shown are: (A) glutamine, (B) threonine, (C) alanine, (D) glutamate, and (E) lysine.

Sometimes, a mutation occurs that substitutes glutamine (yellow circle) with leucine (below).

Predict what will happen to the noncovalent interactions shown when glutamine is substituted with leucine. Explain your reasoning.

4

Scaffolded Guidance.

Problem Worksheet

Learning Objectives:

- Given a model of biological molecules, identify, categorize, and compare the important features.
- Recall that chemical composition affects the structure and interactions of macromolecules.
- Compare how different types of noncovalent interactions occur in biological molecules.
- Explain why and how it is that all noncovalent interactions are attractions of opposite charges.
- Predict the impact of changes in the building blocks of biological molecules on the structure of the macromolecules and formulate a scientific explanation for your prediction.
- Solve new problems that extend your thinking about the role of noncovalent interactions across biology.

Model 1: General Structure of an Amino Acid

The building blocks of proteins are amino acids. There are at least 20 different amino acids found in proteins, each differing in their side chain or R group. Investigate the images below of one amino acid, alanine.

A) Alanine

B) Alanine

1. For each image, circle the amino acid R group and draw a box around the peptide backbone.

Part A: Categorizing Amino Acids Based on their R Groups

When looking at amino acid R groups, it can be helpful to categorize the R group as charged, polar, or nonpolar based on its chemical composition. Let's take another look at alanine.

2. What atoms make up the R group of alanine?

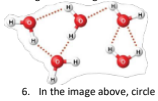
3. Compare the electronegativity of the atoms that make up alanine's R group.

4. Is alanine's R group charged? Does alanine's R group have a significant dipole? Explain your reasoning.

5. Using your answers to questions 1-3 questions, how would you classify alanine's R group (charged, polar, or nonpolar)? Explain your reasoning.

Model 2: Covalent Bonds vs. Noncovalent Interactions

Investigate the image below of water molecules:



6. In the image above, circle and label a covalent bond and a noncovalent interaction.
7. Using your knowledge from general chemistry, define a noncovalent interaction in your own words:

There are three main types of noncovalent interactions: ion pairing, hydrogen bonding, and van der Waals. These three types of noncovalent interactions vary based on the charges involved.

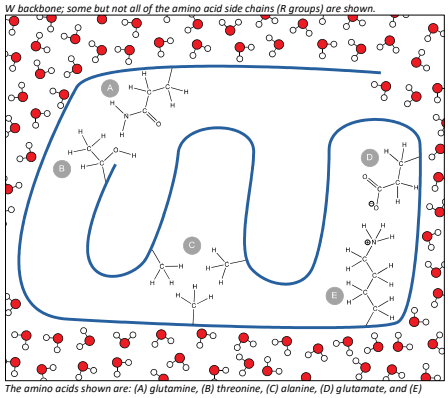
8. Complete the following chart to help you distinguish between the three types of noncovalent interactions:

	Ion Pairing	Hydrogen Bond	van der Waals
Full or partial charges involved?			
Permanent or temporary charges involved?			

Model 3. Noncovalent Interactions in the Context of a Folded Protein

Proteins fold in a way that maximizes the noncovalent interactions between amino acid R groups. Read the model legend in italics and investigate the model below:

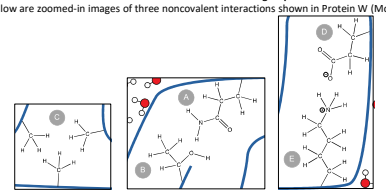
Protein W, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and white). The environment has a pH of 7.4. The blue line represents the Protein W backbone; some but not all of the amino acid side chains (R groups) are shown.



The amino acids shown are: (A) glutamine, (B) threonine, (C) alanine, (D) glutamate, and (E) lysine.

9. Circle a noncovalent interaction between amino acid R groups in the model above.

Model 4: Noncovalent Interactions Between Amino Acid R groups
Below are zoomed-in images of three noncovalent interactions shown in Protein W (Model 3).



Part A: Consider the interaction between the three R groups of amino acid C shown above.

10. Categorize the R group of C (refer to Model 1: Part A for help):

11. What type of noncovalent interaction is occurring among the three amino acids C? Explain how this interaction occurs.

Part B: Consider the interaction between the R group of amino acid A and the R group of amino acid B shown above.

12. Categorize the R group of A and the R group of B:

13. What type of noncovalent interaction is occurring between amino acid A and B? Explain how this interaction occurs.

Part C: Consider the interaction between the R groups of amino acid D and amino acid E shown above.

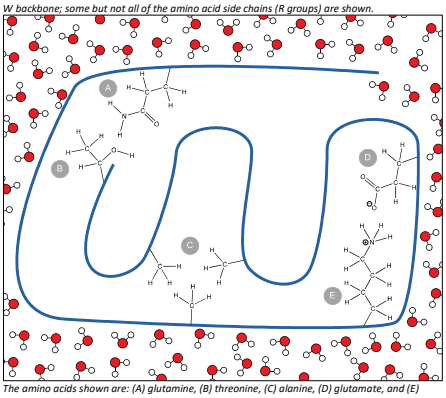
14. Categorize the R group of D and the R group of E:

15. What type of noncovalent interaction is occurring between amino acid D and E? Explain how this interaction occurs.

STOP: Before moving on, double-check your answers with a PLA.

Model 5. An Amino Acid Substitution
Let's return to Model 3 on page 3.

Protein W, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and white). The environment has a pH of 7.4. The blue line represents the Protein W backbone; some but not all of the amino acid side chains (R groups) are shown.



The amino acids shown are: (A) glutamine, (B) threonine, (C) alanine, (D) glutamate, and (E) lysine.

Sometimes, a mutation occurs that substitutes glutamine (A) in Protein W with leucine (below).

Critical Thinking Questions: Considering an Amino Acid Substitution

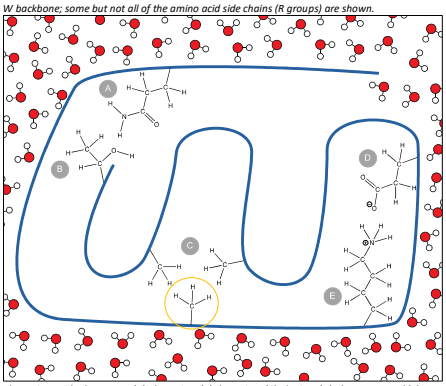
16. Predict what will happen to the noncovalent interactions shown in Protein W when glutamine is substituted with leucine (Model 5). Explain your reasoning and construct a scientific explanation by filling in the blanks below:

Glutamine is a _____ (polar or nonpolar) amino acid. Threonine is a _____ (polar or nonpolar) amino acid. In protein W, glutamine (A) and threonine (B) are attracted to each other through a _____ (type of noncovalent interaction). This interaction arises because there is a _____ (partial or full) positive charge on the H of glutamine and a partial _____ (positive or negative) charge on the O of threonine due to the electronegativity _____ (similarities or differences) among N, H, and O atoms. These _____ (permanent or temporary) charges on glutamine and threonine attract each other. Leucine is a _____ (polar or nonpolar) amino acid. If leucine replaced glutamine in the protein sequence, a hydrogen bond would _____ (still or no longer) form. With this substitution, the protein would likely rearrange so that threonine forms a _____ (type of noncovalent interaction) with something else, like a water molecule. This re-folding would also bring leucine closer to other _____ (polar or nonpolar) R groups, like _____ (name of amino acid in protein W). This is because leucine can interact noncovalently with these other amino acids through _____ (type of noncovalent interaction). Therefore, I predict that the leucine substitution will _____ (have or not have) an impact on the noncovalent interaction shown. In order to feel more confident about my prediction an experiment would need to be conducted to test my prediction.

NOTE: Reflect on the features of the explanation above. Use the structure of this explanation as a model for writing explanations for this type of problem.

Model 6: Another Amino Acid Substitution
Refer back to Protein W.

Protein W, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and white). The environment has a pH of 7.4. The blue line represents the Protein W backbone; some but not all of the amino acid side chains (R groups) are shown.



The amino acids shown are: (A) glutamine, (B) threonine, (C) alanine, (D) glutamate, and (E) lysine.

Sometimes, a mutation occurs that substitutes alanine (yellow circle) with serine (below).

Critical Thinking Questions: Considering an Amino Acid Substitution

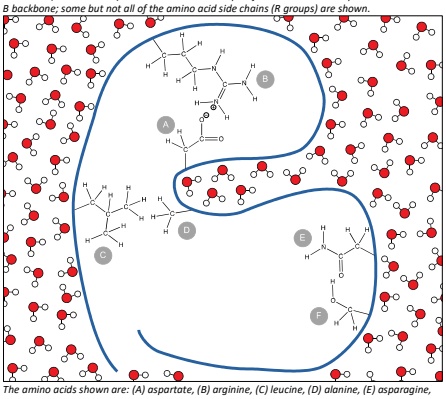
17. Predict what will happen to the noncovalent interactions shown when alanine is substituted with serine. Using complete sentences, explain your reasoning and construct a scientific explanation by completing the table below:

Categorize alanine (C):	
Categorize serine:	
Describe the given noncovalent interaction between the three alanines (C) in Protein W:	
When the substituted Protein W folds, would serine interact with the other alanines?	
State your prediction about the impact of the substitution on the given noncovalent interaction:	
Do you predict any new noncovalent interactions after the substitution? If so, describe them.	
Reflect on your prediction by thinking about what information you would need to feel more confident about your prediction:	

Model 7. More Noncovalent Interactions in the Context of a Different Folded Protein

Let's consider a new protein, Protein B. Read the model legend in italics:

Protein B, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and white). The environment has a pH of 7.4. The blue line represents the Protein B backbone; some but not all of the amino acid side chains (R groups) are shown.



The amino acids shown are: (A) aspartate, (B) arginine, (C) leucine, (D) alanine, (E) asparagine, and (F) serine.

18. On the model of Protein B above, circle and name the three noncovalent interactions that are occurring.

Sometimes, a mutation occurs that substitutes alanine (D) in Protein B with isoleucine (below).

Critical Thinking Questions:

Predict what will happen to the noncovalent interactions shown in Protein B when alanine is substituted with isoleucine (Model 7). Explain your reasoning and construct a scientific explanation by using the format below:

19. Categorize and compare alanine and isoleucine:

20. Describe the original noncovalent interaction that alanine (D) participates in and how it arises:

21. Will Protein B fold into the same 3D shape when alanine is substituted with isoleucine? Explain.

22. Will the original noncovalent interaction still form after substitution?

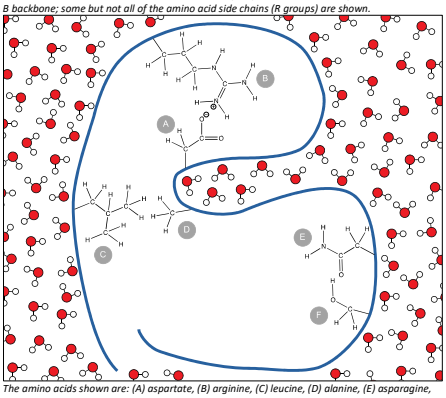
23. Explain any new non-covalent interactions that may occur after the substitution:

24. Reflect on your prediction by thinking about what type of evidence you would need to feel confident about your prediction:

Exercises:

1. Try to solve the following problem on your own.

Protein B, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and white). The environment has a pH of 7.4. The blue line represents the Protein B backbone; some but not all of the amino acid side chains (R groups) are shown.



The amino acids shown are: (A) aspartate, (B) arginine, (C) leucine, (D) alanine, (E) asparagine, and (F) serine.

Sometimes, a mutation occurs that substitutes aspartate (A) with valine (below).

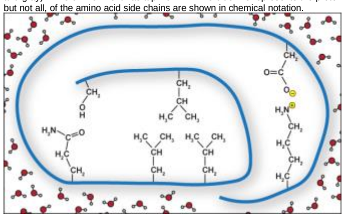
Predict what will happen to the noncovalent interactions shown when aspartate is substituted with valine. Explain your reasoning.

Assessments

Basic Knowledge Posttest

Basic Knowledge Post-Test

Protein X, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and gray). This environment has a pH of 7.4. The blue line represents the protein X backbone. Some, but not all, of the amino acid side chains are shown in chemical notation.



For questions that list multiple true/false statements:

- Decide for each statement whether it is more likely to be true, or more likely to be false.
- Multiple statements can be true.
- Multiple statements can be false.

What occupies the white area inside the protein, where no amino acid side chains are shown?

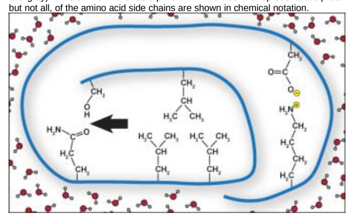
	True	False
a void of empty space (i.e., nothing at all)	O	O
chemical groups	O	O

The blue line represents …

	True	False
peptide bonds	O	O
amino acid side chains	O	O

Continue

Protein X, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and gray). This environment has a pH of 7.4. The blue line represents the protein X backbone. Some, but not all, of the amino acid side chains are shown in chemical notation.



The items below all relate to the most prominent non-covalent interaction occurring in the space pointed to by the arrow.

What is the name of this non-covalent interaction? Select one option.

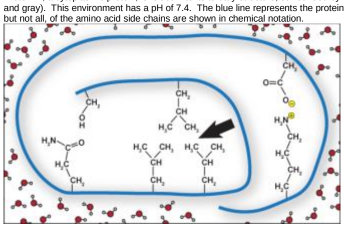
- O hydrogen bond
- O ion pairing
- O van der Waals interaction

The charges involved in this non-covalent interaction are …

	True	False
partial	O	O
temporary	O	O
induced	O	O
due to differences in electronegativity	O	O

Continue

Protein X, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and gray). This environment has a pH of 7.4. The blue line represents the protein X backbone. Some, but not all, of the amino acid side chains are shown in chemical notation.



The items below all relate to the most prominent non-covalent interaction occurring in the space pointed to by the arrow.

What is the name of this non-covalent interaction? Select one option.

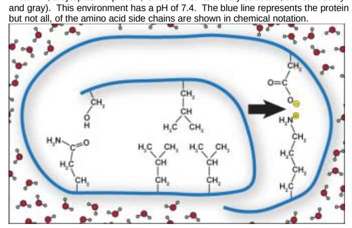
- O hydrogen bond
- O ion pairing
- O van der Waals interaction

The charges involved in this non-covalent interaction are …

	True	False
partial	O	O
temporary	O	O
induced	O	O
due to differences in electronegativity	O	O

Continue

Protein X, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and gray). This environment has a pH of 7.4. The blue line represents the protein X backbone. Some, but not all, of the amino acid side chains are shown in chemical notation.



The items below all relate to the most prominent non-covalent interaction occurring in the space pointed to by the arrow.

What is the name of this non-covalent interaction? Select one option.

- O hydrogen bond
- O ion pairing
- O van der Waals interaction

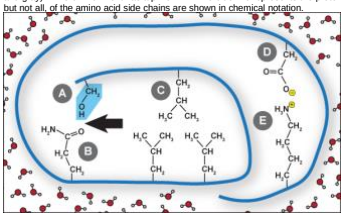
The charges involved in this non-covalent interaction are …

	True	False
partial	O	O
temporary	O	O
induced	O	O
due to differences in electronegativity	O	O

Near Transfer Problems.

Near Transfer Problems

Protein X, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and gray). This environment has a pH of 7.4. The blue line represents the protein X backbone. Some, but not all, of the amino acid side chains are shown in chemical notation.



The amino acids shown are: (A) serine, (B) glutamine, (C) leucine, (D) aspartate, and (E) lysine.

Sometimes, a mutation occurs that substitutes serine (blue highlight) with valine (below).

Do you predict that such a mutation would affect the non-covalent interaction pointed to by the arrow?

- ❍ Yes, the interaction would be affected.
- ❍ No, the interaction would not be affected.

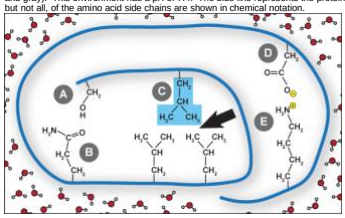
Provide a scientific explanation to support your prediction.

(The following prompt is shown only if "Yes" is selected above)

Predict any new non-covalent interactions that might occur with such a mutation, and provide scientific explanations of them.

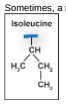
Continue

Protein X, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and gray). This environment has a pH of 7.4. The blue line represents the protein X backbone. Some, but not all, of the amino acid side chains are shown in chemical notation.



The amino acids shown are: (A) serine, (B) glutamine, (C) leucine, (D) aspartate, and (E) lysine.

Sometimes, a mutation occurs that substitutes leucine (blue highlight) with isoleucine (below).



Do you predict that such a mutation would affect the non-covalent interaction pointed to by the arrow?

- ❍ Yes, the interaction would be affected.
- ❍ No, the interaction would not be affected.

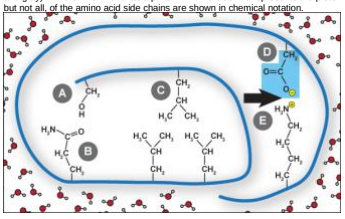
Provide a scientific explanation to support your prediction.

(The following prompt is shown only if "Yes" is selected above)

Predict any new non-covalent interactions that might occur with such a mutation, and provide scientific explanations of them.

Continue

Protein X, a cytoplasmic protein, is folded into its tertiary structure, surrounded by water molecules (red and gray). This environment has a pH of 7.4. The blue line represents the protein X backbone. Some, but not all, of the amino acid side chains are shown in chemical notation.



The amino acids shown are: (A) serine, (B) glutamine, (C) leucine, (D) aspartate, and (E) lysine.

Sometimes, a mutation occurs that substitutes aspartate (blue highlight) with alanine (below).

Do you predict that such a mutation would affect the non-covalent interaction pointed to by the arrow?

- ❍ Yes, the interaction would be affected.
- ❍ No, the interaction would not be affected.

Provide a scientific explanation to support your prediction.

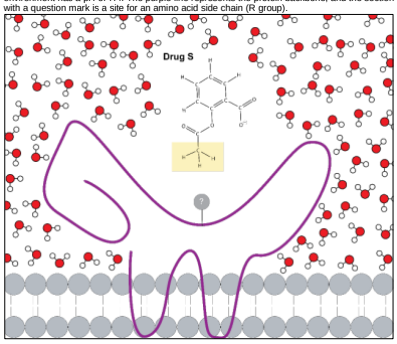
(The following prompt is shown only if "Yes" is selected above)

Predict any new non-covalent interactions that might occur with such a mutation, and provide scientific explanations of them.

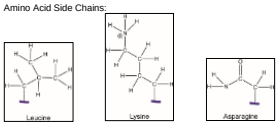
Far Transfer Problems.

Far Transfer Problems

Below is a model of Drug S and a protein with which it may interact. The protein is located on the cell surface situated within the cell membrane and surrounded by water molecules (red and white). The environment has a pH of 7.4. The purple line represents the protein backbone, and the section labeled with a question mark is a site for an amino acid side chain (R group).

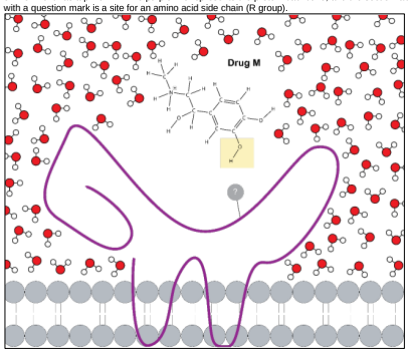


Amino Acid Side Chains:

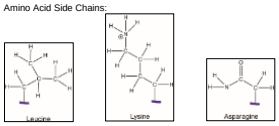


Which amino acid would interact non-covalently with the yellow highlighted section of Drug S? Provide a scientific explanation describing how Drug S interacts non-covalently with the amino acid you selected. Be sure to describe how this interaction forms.

Below is a model of Drug M and a protein with which it may interact. The protein is located on the cell surface situated within the cell membrane and surrounded by water molecules (red and white). The environment has a pH of 7.4. The purple line represents the protein backbone, and the section labeled with a question mark is a site for an amino acid side chain (R group).

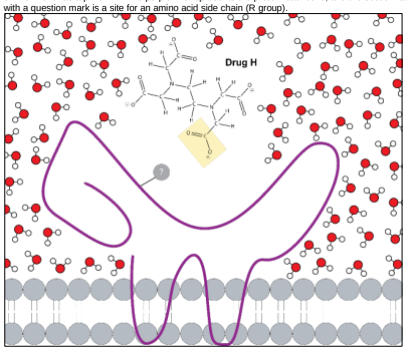


Amino Acid Side Chains:

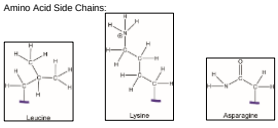


Which amino acid would interact non-covalently with the yellow highlighted section of Drug M? Provide a scientific explanation describing how Drug M interacts non-covalently with the amino acid you selected. Be sure to describe how this interaction forms.

Below is a model of Drug H and a protein with which it may interact. The protein is located on the cell surface situated within the cell membrane and surrounded by water molecules (red and white). The environment has a pH of 7.4. The purple line represents the protein backbone, and the section labeled with a question mark is a site for an amino acid side chain (R group).



Amino Acid Side Chains:



Which amino acid would interact non-covalently with the yellow highlighted section of Drug H? Provide a scientific explanation describing how Drug H interacts non-covalently with the amino acid you selected. Be sure to describe how this interaction forms.

Analytical Codebooks

Protein X – Valine Substitution Codebook

Code #	Code Description

1	**General comparison of amino acids**
	• Chemical properties are compared; there is some difference/similarity IN GENERAL
	  • Appearance, atoms, size, electronegativity or structure
	  ○ *"The hydrogen bonding that typically would occur between these two side chains would no longer be able to occur if serine was substituted for valine because valine does not have a hydrogen atom that can readily participate in this interaction."*
	  ○ *"The hydrogen attached to the oxygen in serine will be gone. The other hydrogens in valine are all attached to carbons."*
	  ○ *"This would happen because these 2 exhibit very similar chemical properties"*
	• Double-coding with codes 2-7 is okay IF specific differences are ALSO mentioned

2	**Serine categorized correctly** Level 1 – polar Level 2 – polar *and* partially charged/dipole *or* permanently charged Level 3 – polar *and* partially charged/dipole *and* permanently charged
3	**Glutamine categorized correctly** Level 1 – polar Level 2 – polar *and* partially charged/dipole *or* permanently charged Level 3 – polar *and* partially charged/dipole *and* permanently charged
4	**Valine categorized correctly** Level 1 – nonpolar/hydrophobic Level 2 – nonpolar *and* induced partial charge/dipole *or* temporarily charged Level 3 – nonpolar *and* induced partial charge/dipole *and* temporarily charged
5	**Serine categorized incorrectly** • Non-polar or hydrophobic • Positively charged (overall) • Charged (overall) • More/less polar • Do NOT use this for "uncharged"
6	**Glutamine categorized incorrectly** • Non-polar or hydrophobic • Positively charged (overall) • Charged (overall) • More/less polar • Do NOT use this for "uncharged"
7	**Valine categorized incorrectly** • Polar ○ *"the CH3 bonds are polar"* • More/less polar

8	**Electronegativity of specific atoms discussed** For Serine: • Hydrogen is less electronegative • Oxygen is more electronegative For Glutamine: • Carbon is less electronegative • Oxygen is more electronegative For Valine: • Neither carbon nor hydrogen are electronegative; no difference in electronegativities • Do NOT count references to chemical groups, i.e. *"stable CH3/methyl"* would not receive this code *"Serine allows for hydrogen bonding because there is a hydrogen on the molecule which is attached to an electronegative atom."* *"The new interaction would be hydrogen bonding with glutamine because the difference in electronegativity causes partial charges to occur which causes the hydrogen off the oxygen of the mutated R-group to be attracted to the oxygen of serine."* NOTE – can receive this code without amino acid category codes (2-7).
9	**Given interaction would no longer occur after substitution** Do NOT double-code simply because other codes about new interactions apply • Literal statements ○ *"Hydrogen bonding would no longer occur."* ○ *"The reaction (read, 'interaction') would no longer be able to take place."* ○ *"The bond is disrupted"* • Use code 15 for statements regarding the ABSENCE of any new interaction • "instead of" and "change to" statements do not qualify for this code
10	**Valine shape/size will affect the interaction**
11	**Valine and glutamine could interact – correct** NOT for hydrogen bonds (see code 14) • Van der Waals interactions • London dispersion forces • Dipole-induced dipole interaction • Methyl and carbonyl O (or N) could interact • Through the methyl/non-polar groups of valine NOTE: assume if they say "new interaction", then they mean between valine and glutamine
12	**Valine could interact with another residue** • Leucine, hydrophobic core, or a non-polar residue not shown NOTE: must be specific about what valine will interact with in order to receive this code
13	**Glutamine could interact with another residue or water**
14	**Valine and glutamine could interact through a hydrogen bond** Do NOT also apply code 11 • Hydrogen bond could still be intact ○ *" There are still hydrogens that can be bonded"* Through the methyl/non-polar groups of valine
15	**New interaction predicted incorrectly** • Incorrect predictions other than hydrogen bonding ○ *"An ionic bond may form because of the increased differences in electronegativity"* • No interaction would occur after substitution (lack or absence of an interaction) ○ *"There is no new non-covalent interaction"* ○ *"No interaction would occur because one is polar and one is non-polar"* ○ *"The amino acid cannot interact with anything."*
16	**Structure/folding/function affected or NOT affected by valine shape and/or size**
17	**Structure/folding/function affected or NOT affected by mutation and/or new interactions**
18	**Contradicting without a resolution** Do NOT use if contradiction is resolved

19	**Need for empirical evidence** • "more info" does not count, but "more scientific info" does count.
20	**Given interaction (hydrogen bond) identified correctly**
21	**Given interaction identified incorrectly** • Anything but a hydrogen bond.
22	**Mechanism for existing interaction described correctly** *"attraction due to opposite charges"* NOTE: must use the word attract to receive this code
23	**Mechanism for new interaction(s) described correctly** *"attraction due to opposite charges"* NOTE: must use the word attract to receive this code
24	**Mechanism for existing interaction described incorrectly** *"repulsion"* • Residues would repel each other ○ *"... there would be a great deal of repulsion between the two compounds because both are nonpolar."*
25	**Mechanism for new interaction(s) described incorrectly** *"repulsion"* • Residues would repel each other ○ *"... there would be a great deal of repulsion between the two compounds because both are nonpolar."*

Protein X – Isoleucine Substitution Codebook

Code #	Code Description

1	**General comparison of amino acids**
	• Chemical properties are compared; there is some difference/similarity IN GENERAL
	  • Appearance, atoms, size, electronegativity or structure
	  o *"Isoleucine is an isomer of leucine."*
	  o *"the molecules differ in orientation"*
	  • Polarity
	  o *"Leucine and isoleucine have the same polarity."*
	• Double-coding with codes 2-5 is okay IF specific similarities are ALSO mentioned

2	**Leucine categorized correctly** Level 1 – nonpolar/hydrophobic Level 2 – nonpolar *and* induced partial charge/dipole *or* temporarily charged Level 3 – nonpolar *and* induced partial charge/dipole *and* temporarily charged
3	**Isoleucine categorized correctly** Level 1 – nonpolar/hydrophobic Level 2 – nonpolar *and* induced partial charge/dipole *or* temporarily charged Level 3 – nonpolar *and* induced partial charge/dipole *and* temporarily charged
4	**Leucine categorized incorrectly**
5	**Isoleucine categorized incorrectly**
6	**Electronegativity of specific atoms discussed** • Neither carbon nor hydrogen are electronegative; no difference in electronegativities Do NOT count references to chemical groups, i.e. *"stable CH3/methyl"* would not receive this code. Must specify atoms. NOTE – can receive this code without amino acid category codes (2-5).
7	**Confuse hydrophobic amino acid with hydrophobic effect**

8	**Given interaction would no longer occur after substitution**
	• Do NOT double-code simply because other codes about new interactions apply
	• Literal statements
	  o *"Ion pairing would no longer occur."*
	  o *"The reaction (read, 'interaction') would no longer be able to take place."*
	  o *"The bond is disrupted"*
	• Use code 11 for statements regarding the ABSENCE of any new interaction
	• "instead of" and "change to" statements do not qualify for this code

9	**A new or different van der Waals interaction is formed** • Okay to say "the interaction" instead of van der Waals • Between isoleucine and a lower leucine (see Note) NOTE – Interaction: • Anything going on with the three residues • Do not differentiate between isoleucine with left leucine vs. isoleucine with right leucine • *"The interaction will remain the same"* counts for this code if they also got code 16

10	**Isoleucine shape/size will affect or NOT affect the interaction**
	• a general acknowledgement that size may have an effect qualifies for this code
	• Explanation of what causes a new van der Waals interaction (code 9)
	• Between isoleucine and a lower leucine (see Note 1)
	  o *"This mutation would affect the distance of the bottom left leucine from the isoleucine methyl group. This mutation might get rid of or severely weaken the bond between these two."*
	  o *" … the interaction with the molecule to the left will be affected due to the distance between the two (interacting) compounds."*

11	**New interaction predicted incorrectly** • Incorrect predictions other than LDF or VDW o *"An ionic bond may form because of the increased differences in electronegativity"* • Residues would repel each other o *"... there would be a great deal of repulsion between the two compounds because both are nonpolar."* • No interaction would occur after substitution o *"There is no non-covalent interaction"* • Don't code unintelligible answers that can't be interpreted as definitely incorrect
12	**Structure/folding/function affected or NOT affected by isoleucine shape and/or size**
13	**Structure/folding/function affected or NOT affected by mutation and/or new interactions** Capture both affected and not affected answers
14	**Contradicting without a resolution**
15	**Need for empirical evidence** • "more info" does not count, but "more scientific info" does count.
16	**Given interaction (van der Waals) identified correctly** • Induced dipole – induced dipole, London dispersion force, van der Waals
17	**Given interaction identified incorrectly** Anything but van der Waals, London dispersion force, or induced dipole - induced dipole
18	**Mechanism for existing interaction described correctly** *"attraction due to opposite charges"* NOTE: must use the word attract to receive this code
19	**Mechanism for new interaction(s) described correctly** *"attraction due to opposite charges"* NOTE: must use the word attract to receive this code
20	**Mechanism for existing interaction described incorrectly** *"repulsion"* • Residues would repel each other *"... there would be a great deal of repulsion between the two compounds because both are nonpolar."*
21	**Mechanism for new interaction(s) described incorrectly** *"repulsion"* • Residues would repel each other *"... there would be a great deal of repulsion between the two compounds because both are nonpolar."*

Protein X – Alanine Substitution Codebook

Code #	Code Description

1	**General comparison of amino acids**
	• Chemical properties are compared; there is some difference/similarity IN GENERAL
	  • Appearance, atoms, size, electronegativity, or structure
	  • Double-coding with codes 2-7 is okay IF specific differences are ALSO mentioned

2	**Aspartate categorized correctly** Level 1 – polar or charged Level 2 – polar or charged *and* fully charged *or* permanently charged Level 3 – polar or charged *and* fully charged *and* permanently charged
3	**Lysine categorized correctly** Level 1 – polar or charged Level 2 – polar or charged *and* fully charged *or* permanently charged Level 3 – polar or charged *and* fully charged *and* permanently charged
4	**Alanine categorized correctly** Level 1 – nonpolar/hydrophobic Level 2 – nonpolar *and* induced partial charge/dipole *or* temporarily charged Level 3 – nonpolar *and* induced partial charge/dipole *and* temporarily charged
5	**Aspartate categorized incorrectly** • Non-polar or hydrophobic • Uncharged
6	**Lysine categorized incorrectly** • Non-polar or hydrophobic • Uncharged
7	**Alanine categorized incorrectly** • Partially charged • Hydrogen is partially positively charged
8	**Electronegativity of specific atoms discussed** For Aspartate: • Carbon is less electronegative • Oxygen is more electronegative For Lysine: • Hydrogen is less electronegative • Nitrogen is more electronegative For Alanine: • Neither carbon nor hydrogen are electronegative; no difference in electronegativities • Do NOT count references to chemical groups, i.e. *"stable CH3/methyl"* would not receive this code NOTE – can receive this code without amino acid category codes (2-4).

9	**Given interaction would no longer occur after substitution**
	• Do NOT double-code simply because other codes about new interactions apply
	• Literal statements
	  ○ *"Ion pairing would no longer occur."*
	  ○ *"The reaction (read, 'interaction') would no longer be able to take place."*
	  ○ *"The reaction (read, 'interaction') would not form between the alanine and the lysine."*
	  ○ *"The bond is disrupted"*
	• Use code 15 for statements regarding the ABSENCE of any new interaction
	• "instead of" and "change to" statements do not qualify for this code

10	**Alanine shape and/or size will affect or NOT affect the interaction**

11	**Alanine and lysine could interact - correct** • NOT for ion pairing (see code 16) • Interaction MUST be named • Assume named interaction is between alanine and lysine unless otherwise specified • Ion-induced dipole interaction • Van der Waals interaction
12	**Alanine could interact with another residue** • Leucine, hydrophobic core, or a non-polar residue not shown • Van der Waals interaction • London dispersion forces
13	**Lysine could interact with another residue or water** • Note: It's okay for students to call this hydrogen bonding, ion-dipole interaction
14	**Ion pairing could still be intact or not affected** • Explanation of why intact or not affected, beyond answer for part a. • Do NOT also apply code 11 • Between alanine and lysine • Through the methyl/non-polar groups of alanine
15	**New interaction predicted incorrectly** • Incorrect predictions other than ion pairing (e.g., hydrogen bonding between alanine and lysine, repelling each other) • No interaction would occur after substitution o *"There is no new non-covalent interaction"* o *"… no bond will form because van der waals interactions are only bonded between neutral atoms. Since there is a positively charged nitrogen, there will not be any interaction taking place."*
16	**Structure/folding/function affected or NOT affected by alanine shape and/or size**
17	**Structure/folding/function affected or NOT affected by mutation and/or new interactions**
18	**Contradicting without a resolution** o *"I do not see any new non-covalent interactions that might occur. There maybe van Der waals forces interacting because of the nonpolar group of alanine."*
19	**Need for empirical evidence** • "more info" does not count, but "more scientific info" does count.
20	**Given interaction (ion pairing) identified correctly** • other acceptable terms: ionic interaction, ionic bonding, ionic bond, ion pair, ion-ion pair/interaction, ionic attraction
21	**Given interaction identified incorrectly** • Anything but ion pairing
22	**Mechanism for existing interaction described correctly** *"attraction due to opposite charges"* NOTE: must use the word attract to receive this code
23	**Mechanism for new interaction(s) described correctly** *"attraction due to opposite charges"* NOTE: must use the word attract to receive this code
24	**Mechanism for existing interaction described incorrectly** *"repulsion"* • Residues would repel each other *"… there would be a great deal of repulsion between the two compounds because both are nonpolar."*
25	**Mechanism for new interaction(s) described incorrectly** *"repulsion"* • Residues would repel each other *"… there would be a great deal of repulsion between the two compounds because both are nonpolar."*

Drug S Codebook

Code #	Code Description
1	**General comparison of amino acids/drug (chemical composition, polarity)**
2	**Electronegativity of specific atoms discussed** • O, N are more electronegative than C, H • C and H have no real difference in electronegativity
3	**Comparison of strength of intermolecular forces** • van der Waals, induced dipole-induced dipole < dipole-induced dipole < dipole-dipole, hydrogen bond < ion-induced dipole < ion-dipole < ion-ion pairing • Interactions being compared need to be clearly stated
4	**Need for empirical evidence**
5	**Contradicting self** • Without resolution
6	**Categorizing chosen amino acid correctly** For Leucine: nonpolar with induced partial temporary charges, For Asparagine: polar with partial permanent charges For Lysine: polar with full permanent charge • Level 1 – category • Level 2 – category AND magnitude OR permanency described • Level 3 – category AND magnitude AND permanency described
7	**Categorizing chosen amino acid incorrectly**
8	**Characterizing highlighted portion of drug correctly** Nonpolar methyl with induced partial temporary charges • Level 1 – category • Level 2 – category AND magnitude OR permanency described • Level 3 – category AND magnitude AND permanency described
9	**Characterizing highlighted portion of drug incorrectly**
10	**Predicting interaction correctly** van der Waals between Leucine and Drug Dipole-induced dipole between Asparagine and Drug (or vdw, but must name groups involved) Ion-induced dipole between Lysine and Drug (or vdw, but must name groups involved) • MUST name interaction
11	**Predicting interaction incorrectly**
12	**Describing how interaction will form correctly (mechanistic reasoning)** *"Attraction due to opposite charges"* van der Waals: attraction of oppositely signed induced temporary dipoles/charges on two nonpolar molecules Dipole-induced dipole: attraction between charge on a dipole and the opposite signed induced charge on a nonpolar molecule Ion-induced dipole: attraction between fully charged species and the opposite signed induced charge on a nonpolar molecule
13	**Describing how interaction will form incorrectly** • *"repulsion"*

Drug M Codebook

Code #	Code Description
1	**General comparison of amino acids/drug (chemical composition, polarity)**
2	**Electronegativity of specific atoms discussed** • O, N are more electronegative than C, H • C and H have no real difference in electronegativity
3	**Comparison of strength of intermolecular forces** • van der Waals, induced dipole-induced dipole < dipole-induced dipole < dipole-dipole, hydrogen bond < ion-induced dipole < ion-dipole < ion-ion pairing • Interactions being compared need to be clearly stated
4	**Need for empirical evidence**
5	**Contradicting self** • Without resolution
6	**Categorizing chosen amino acid correctly** For Leucine: nonpolar with induced partial temporary charges, For Asparagine: polar with partial permanent charges For Lysine: polar with full permanent charge • Level 1 – category • Level 2 – category AND magnitude OR permanency described • Level 3 – category AND magnitude AND permanency described
7	**Categorizing chosen amino acid incorrectly**
8	**Characterizing highlighted portion of drug correctly** Polar hydroxyl with partial permanent charges (O is partially negative, H is partially positive) • Level 1 – category • Level 2 – category AND magnitude OR permanency described • Level 3 – category AND magnitude AND permanency described
9	**Characterizing highlighted portion of drug incorrectly**
10	**Predicting interaction correctly** Dipole-induced dipole between drug and Leucine Dipole-dipole/hydrogen bond between drug and Asparagine Ion-dipole between Lysine and drug • MUST name interaction
11	**Predicting interaction incorrectly**
12	**Describing how interaction will form correctly (mechanistic reasoning)** *"attraction due to opposite charges"* Dipole-induced dipole: attraction between partially charged species and the oppositely signed induced temporary dipole/charge on a nonpolar molecule Dipole-dipole/hydrogen bond: attraction between two oppositely partially charged species Ion-dipole: attraction between fully charged ionic species and the oppositely signed partial charge on a polar molecule
13	**Describing how interaction will form incorrectly** *"repulsion"*

Drug H Codebook

Code #	Code Description
1	**General comparison of amino acids/drug (chemical composition, polarity)**
2	**Electronegativity of specific atoms discussed** • O, N are more electronegative than C, H • C and H have no real difference in electronegativity
3	**Comparison of strength of intermolecular forces** • van der Waals, induced dipole-induced dipole < dipole-induced dipole < dipole-dipole, hydrogen bond < ion-induced dipole < ion-dipole < ion-ion pairing • Interactions being compared need to be clearly stated
4	**Need for empirical evidence**
5	**Contradicting self** • Without resolution
6	**Categorizing chosen amino acid correctly** For Leucine: nonpolar with induced partial temporary charges, For Asparagine: polar with partial permanent charges For Lysine: polar with full permanent charge • Level 1 – category • Level 2 – category AND magnitude OR permanency described • Level 3 – category AND magnitude AND permanency described
7	**Categorizing chosen amino acid incorrectly**
8	**Characterizing highlighted portion of drug correctly** Polar carboxylate with full permanent negative charge on O • Level 1 – category • Level 2 – category AND magnitude OR permanency described • Level 3 – category AND magnitude AND permanency described
9	**Characterizing highlighted portion of drug incorrectly**
10	**Predicting interaction correctly** Ion-induced dipole between drug and Leucine Ion-dipole between drug and Asparagine Ion-ion between drug and Lysine • MUST name interaction
11	**Predicting interaction incorrectly**
12	**Describing how interaction will form correctly (mechanistic reasoning)** *"attraction due to opposite charges"* Ion-induced dipole: attraction between fully charged ionic species and the oppositely signed induced temporary dipole/charge on a nonpolar molecule Ion-dipole: attraction between fully charged ionic species and the oppositely signed partial charge on a polar molecule Ion-pairing: attraction between two oppositely fully charged species
13	**Describing how interaction will form incorrectly** *"repulsion"*

Scoring Rubrics

Protein X – Valine Substitution Rubric

Evidence (up to 3 pts)						**Prediction (up to 3 pts)**		**Linkage (up to 1 pt)**	
Categorize given amino acids	**pts**	**Categorize new amino acid**	**pts**	**Identify given interaction**	**pts**	**Predict new interaction**	**pts**	**Linking evidence and prediction**	**pts**
Ser and Gln both categorized correctly	1	Val categorized correctly	1	Hydrogen bond identified correctly	1	High quality prediction	3	At least one piece of evidence and a prediction made	1
Only Ser categorized	1	Different	0.5	Mixed ideas	0.5	Low quality prediction	2	Evidence or prediction missing	0
Only Gln categorized	1	Mixed ideas	0.5	Missing	0	Mixed ideas	1		
Mixed ideas	0.5	Missing	0	Interaction identified incorrectly	0	Missing	0		
Missing	0	Val categorized incorrectly	0			Unacceptable hydrogen bond prediction	0		
Ser or Gln categorized incorrectly	0					Other unacceptable prediction	0		

Protein X – Isoleucine Substitution Rubric

Evidence (up to 3 pts)						Prediction (up to 3 pts)		Linkage (up to 1 pt)	
Categorize given amino acids	**pts**	**Categorize new amino acid**	**pts**	**Identify given interaction**	**pts**	**Predict new interaction**	**pts**	**Linking evidence and prediction**	**pts**
Leu categorized correctly	1	Ile categorized correctly	1	van der Waals identified correctly	1	High quality prediction	3	At least one piece of evidence and a prediction made	1
Mixed ideas	0.5	Different	0.5	Mixed ideas	0.5	Low quality prediction	2	Evidence or prediction missing	0
Missing	0	Mixed ideas	0.5	Missing	0	Mixed ideas	1		
Leu categorized incorrectly	0	Missing	0	Interaction identified incorrectly	0	Missing	0		
		Ile categorized incorrectly	0			Unacceptable prediction	0		

Protein X – Alanine Substitution Rubric

Evidence (up to 3 pts)						**Prediction (up to 3 pts)**		**Linkage (up to 1 pt)**	
Categorize given amino acids	**pts**	**Categorize new amino acid**	**pts**	**Identify given interaction**	**pts**	**Predict new interaction**	**pts**	**Linking evidence and prediction**	**pts**
Asp and Lys both categorized correctly	1	Ala categorized correctly	1	Ion pairing identified correctly	1	High quality predictions	3	At least one piece of evidence and a prediction made	1
Only Asp categorized	1	Different	0.5	Mixed ideas	0.5	Low quality predictions	2	Evidence or prediction missing	0
Only Lys categorized	1	Mixed ideas	0.5	Missing	0	Mixed ideas	1		
Mixed ideas	0.5	Missing	0	Interaction identified incorrectly	0	Missing	0		
Missing	0	Ala categorized incorrectly	0			Unacceptable ion pairing prediction	0		
Asp or Lys categorized incorrectly	0					Other unacceptable prediction	0		

Drug S Rubric

Evidence (up to 3 pts)						Prediction (up to 3 pts)				Linkage (up to 1 pt)	
Categorize Drug S	**pts**	**Categorize chosen amino acid**	**pts**	**Mechanistic Reasoning**	**pts**	**Predict new interaction**	**pts**	**Chosen amino acid**	**pts**	**Linking evidence and prediction**	**pts**
Correctly (8)	1	Correctly (6)	1	Described how interaction will form correctly (12)	1	Predict interaction correctly (10)	2	Leu	1	At least one piece of evidence and a prediction made	1
Mixed (8 and 9)	0.5	General Comparison (1)	0.5	Mixed (12 and 13)	0.5	Mixed (10 and 11)	1	Lys	0	Evidence or prediction missing	0
Missing	0	Mixed (1 or 6 and 7)	0.5	Missing	0	Missing	0	Asn	0		
Incorrectly (9)	0	Missing	0	Incorrectly (13)	0	Incorrectly (11)	0				
		Incorrectly (7)	0								

Numbers in parentheses correspond to code numbers in codebook.

Drug M Rubric

Evidence (up to 3 pts)						Prediction (up to 3 pts)				Linkage (up to 1 pt)	
Categorize Drug S	**pts**	**Categorize chosen amino acid**	**pts**	**Mechanistic Reasoning**	**pts**	**Predict new interaction**	**pts**	**Chosen amino acid**	**pts**	**Linking evidence and prediction**	**pts**
Correctly (8)	1	Correctly (6)	1	Described how interaction will form correctly (12)	1	Predict interaction correctly (10)	2	Asn	1	At least one piece of evidence and a prediction made	1
Mixed (8 and 9)	0.5	General Comparison (1)	0.5	Mixed (12 and 13)	0.5	Mixed (10 and 11)	1	Lys	0	Evidence or prediction missing	0
Missing	0	Mixed (1 or 6 and 7)	0.5	Missing	0	Missing	0	Leu	0		
Incorrectly (9)	0	Missing	0	Incorrectly (13)	0	Incorrectly (11)	0				
		Incorrectly (7)	0								

Numbers in parentheses correspond to code numbers in codebook.

Drug H Rubric

Evidence (up to 3 pts)						Prediction (up to 3 pts)				Linkage (up to 1 pt)	
Categorize Drug S	**pts**	**Categorize chosen amino acid**	**pts**	**Mechanistic Reasoning**	**pts**	**Predict new interaction**	**pts**	**Chosen amino acid**	**pts**	**Linking evidence and prediction**	**pts**
Correctly (8)	1	Correctly (6)	1	Described how interaction will form correctly (12)	1	Predict interaction correctly (10)	2	Lys	1	At least one piece of evidence and a prediction made	1
Mixed (8 and 9)	0.5	General Comparison (1)	0.5	Mixed (12 and 13)	0.5	Mixed (10 and 11)	1	Leu	0	Evidence or prediction missing	0
Missing	0	Mixed (1 or 6 and 7)	0.5	Missing	0	Missing	0	Asn	0		
Incorrectly (9)	0	Missing	0	Incorrectly (13)	0	Incorrectly (11)	0				
		Incorrectly (7)	0								

Numbers in parentheses correspond to code numbers in codebook.