Glycosyltransferases and Glycoside Hydrolases

Kelley Moremen

Complex Carbohydrate Research Center
315 Riverbend Rd.
University of Georgia
Athens, GA 30602
706-542-1705
moremen@uga.edu


Major Classes of Vertebrate Glycan Structures

Oligosaccharide portion of a glycoprotein can occupy a significant amount of space, often more than the polypeptide.
Enormous diversity of glycan structures in animal systems

Huge number of glycan structures synthesized in a single cell type based on the array of enzymes expressed

Mouse ES cell glycans identified by MSn techniques

O-glycan structures  N-glycan structures

Two major carbohydrate-peptide linkages found in glycoproteins

O-linked (mucin-type, GalNAc) chain

Asn-linked sugar chain
O-linked oligosaccharides can be initiated by the addition of several different sugars to serine and threonine residues.

![Diagram showing different sugars and their associated proteins](image)

Other classes of carbohydrate-peptide linkages are also found in glycoproteins

<table>
<thead>
<tr>
<th>Type</th>
<th>Linkage</th>
<th>Enzyme</th>
<th>Consensus sequence</th>
<th>Domain</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-glycosylation</td>
<td>GalNAc-β-Asn</td>
<td>OST</td>
<td>N-X-(ST) (standard sequences)</td>
<td>Nascent polypeptides</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GalNAc-β-Thr</td>
<td>OGT</td>
<td>N-X-(ST) (non-standard sequences)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O-glycosylation</td>
<td>GalNAc-α-Thr</td>
<td>PGLA/Nts</td>
<td>Isoform specific†</td>
<td>Mucins</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GalNAc-β-Thr</td>
<td>OGT</td>
<td>No set consensus*</td>
<td>Cytoplasmic, nuclear proteins</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GlcNAc-β-Thr</td>
<td>EOGT</td>
<td>Unknown</td>
<td>EGF</td>
<td></td>
</tr>
<tr>
<td>Xyl-β-Ser</td>
<td>XYLTLXYLT2</td>
<td></td>
<td>α-4-linked sugars</td>
<td>Heparin, proteoglycan core proteins</td>
<td></td>
</tr>
<tr>
<td>Fuc-α-Thr</td>
<td>POFUT2</td>
<td></td>
<td>C-X-X-(S/T)</td>
<td>TSR</td>
<td>Thrombospondin, ADAMTS family</td>
</tr>
<tr>
<td>Glc-β-Ser</td>
<td>POGUT1</td>
<td></td>
<td>C-X-S-X-(A/P)-C</td>
<td>EGF</td>
<td>Notch, clotting factors</td>
</tr>
<tr>
<td>Man-α-Thr</td>
<td>POMT1, POMT2</td>
<td></td>
<td>I-X-P-T-P-X-T-X-P-X-X-P-T-X-(T/N)-X-X**</td>
<td>α-dystroglycan</td>
<td></td>
</tr>
<tr>
<td>Gal-β-Hyl</td>
<td>GLTLD1, GLTLD2</td>
<td></td>
<td>X-Hyl-Gly (collagen repeats)</td>
<td>Collagen, adiponectin</td>
<td></td>
</tr>
<tr>
<td>Glc-α-Tyr</td>
<td>GYG</td>
<td></td>
<td>Tyr194 of GYG†</td>
<td>GYG</td>
<td></td>
</tr>
<tr>
<td>C-mannosylation</td>
<td>Man-α-Trp</td>
<td>Unknown</td>
<td>W-X-W*</td>
<td>TSR</td>
<td>Thrombospondin, ADAMTS family</td>
</tr>
<tr>
<td>GlypHylation</td>
<td>Pri-O(β1,6)N-6-P-Man</td>
<td>Transaminase</td>
<td>No set consensus*</td>
<td>THY1, NCAM1</td>
<td></td>
</tr>
</tbody>
</table>

Most classes of glycan structures are created and modified within the ER and Golgi during transit to the cell surface

- Modification enzymes (GTs and GHs) are anchored to the membrane (usually via N-terminal TMD) with a catalytic domain that faces into the lumen
- Modify soluble and membrane associated glycan acceptors

Most proteins/lipids in the secretory pathway are glycosylated

- Glycans are extended in lumen of secretory pathway by glycosyltransferases
- Each enzyme has specificity for acceptor glycan, sugar donor, position of transfer
- Employ sugar-nucleotide donors for transfer one monosaccharide at a time
- Main role of the secretory pathway is post-translational modification

(From Darnell et al (1990) Molecular Cell Biology, 2nd ed.)
Complexity in the synthesis of glycan structures

- Anomeric configuration of glycosidic linkages: α or β
- Position of glycosidic linkages:
  - Hexose acceptors: C2, C3, C4, C6 (more for other sugars)
  - Large potential for branching
  - Peptide groups: Asn, Ser, Thr, Lys(OH), Trp
  - EthN: GPI anchors
- Sugar donor type
  - sugar-nucleotide
  - lipid(polyisoprene)-P-sugar (Dol-P-Man, Dol-P-Glc, Dol-P-P-OS)
- N(D/M)P-Sugar donors (9+1 in mammals, many more in other organisms):
  - UDP sugars: Glc, Gal, GlcNAc, GalNAc, Xyl, GlcA (IdoA)
  - GDP-sugars: Man, Fuc
  - CDP-ribitol
  - CMP-sugars: NeuAc
- Each donor has a single anomeric configuration

Biosynthesis and interconversion of monosaccharides

Essentials of Glycobiology
Second Edition
Complexity in the synthesis of glycan structures

- Sugar transfer to acceptors can result in product with:
  - *Identical anomeric configuration* (retention in configuration)
  - *Switch in anomeric configuration* (inversion in configuration)
- Sugar nucleotide linkages (in mammals):

<table>
<thead>
<tr>
<th>Donor</th>
<th>Inverted Product</th>
<th>Retaining Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>UDP-α-D-Glc</td>
<td>Glc-β-D-X</td>
<td>Glc-α-D-X</td>
</tr>
<tr>
<td>UDP-α-D-Gal</td>
<td>Gal-β-D-X</td>
<td>Gal-α-D-X</td>
</tr>
<tr>
<td>UDP-α-D-GalNAc</td>
<td>GalNAc-β-D-X</td>
<td>GalNAc-α-D-X</td>
</tr>
<tr>
<td>UDP-α-D-GlcNAc</td>
<td>GlcNAc-β-D-X</td>
<td>GlcNAc-α-D-X</td>
</tr>
<tr>
<td>UDP-α-D-GlcA</td>
<td>GlicA-β-D-X</td>
<td></td>
</tr>
<tr>
<td>UDP-α-D-Xyl</td>
<td>Xyl-β-D-X</td>
<td>Xyl-α-D-X</td>
</tr>
<tr>
<td>GDP-α-D-Man</td>
<td></td>
<td>Man-α-D-X</td>
</tr>
<tr>
<td>GDP-β-L-Fuc</td>
<td>Fuc-α-D-X</td>
<td></td>
</tr>
<tr>
<td>CMP-β-D-Neu5Ac</td>
<td>Neu5Ac-α-D-X</td>
<td></td>
</tr>
<tr>
<td>CDP-D-Ribitol</td>
<td></td>
<td>?Ribitol-1-PO₄-X</td>
</tr>
<tr>
<td>Dol-P-β-D-Man</td>
<td>Man-α-D-X</td>
<td>Man-β-D-X</td>
</tr>
<tr>
<td>Dol-P-β-D-Glc</td>
<td>Glc-α-D-X</td>
<td></td>
</tr>
</tbody>
</table>

Refresher discussion on sugar conformations

- Monosaccharides can interconvert between 6 membered (pyranose) and 5 membered (furanose) forms through a opened ring transition.
- Sugars can also have ring opening and interconvert between α and β configuration at the anomeric center (C1 hemiacetal carbon).
Glycosidic bond formation blocks interconversion of forms

Once a glycosidic linkage (acetal linkage) is formed from the C1OH, the ring can no longer open and the monosaccharide is locked into a specific ring form (furanose or pyranose).

The anomeric form (α or β) is also locked and can not interconvert.

Only the reducing terminal monosaccharide of an oligosaccharide (if there is one) can open to form an equilibrium mixture of the various forms.

Flexibility of closed sugar conformations is governed by steric hindrance

Six membered rings have three major and several minor configurations.

**Figure 11-11** A disaccharide is formed from two monosaccharides (here, two molecules of D-glucose) when an alcoholic —OH of one glucose molecule (right) condenses with the intramolecular hemiacetal of the other glucose molecule (left), with the elimination of H₂O and formation of a glycosidic bond. The reversal of this reaction is hydrolysis—attack by H₂O on the glycosidic bond. The maltose molecule retains a reducing hemiacetal at the C1 not involved in the glycosidic bond.

**Figure 10-6.** The conformations of the cyclohexane ring. (a) In the boat conformation, substituents at the “bow” and “stem” (red) are sterically crowded, whereas those along its sides (green) are eclipsed. (b) In the chair conformation, the substituents that extend parallel to the ring’s threefold rotation axis are designated axial [α] and those that extend roughly outward from this symmetry axis are designated equatorial [ε]. The equatorial substituents about the ring are staggered so that they alternately extend above and below the mean plane of the ring.

**Figure 10-7.** The two alternative chair conformations of β-o-glucopyranose. In the conformation on the left, which predominates, the relatively bulky OH and CH₂OH substituents all occupy equatorial positions, whereas in that on the right (drawn in ball-and-stick form in Fig. 10-5, right) they occupy the more crowded axial positions.

Energetically favored, all equatorial in glucose

Energetically unfavored, all axial in glucose (steric hindrance)
Cleavage of an acetal: non-enzymatic mechanisms

In the non-enzymatic cleavage of an acetal the progression is:

1) Protonation of a reactive oxygen
2) Cleavage of the C-O bond and formation of a resonance stabilized carbocation (called an oxonium ion or an oxocarbenium ion).
3) The oxocarbenium ion has double bond character and is planar in the O-C-H linkages.
4) Addition of water to the oxocarbenium ion yields the hemiacetal and regeneration of the acid catalyst.

Thus enzyme reaction should contain

1) Acid catalyst for proton donation
2) Nucleophile to attack the oxocarbenium ion

Bond formation or cleavage: transition state for both GTs and GHs

- The transition state (positively charged oxocarbenium ion) is a very high energy molecule
  - Positive charge is not just at C1 but shared between C1 and ring oxygen
  - Partial double bond character: C1 and ring oxygen are in same plane
  - Geometry changes from chair to half-chair/boat/skew boat
  - This stabilizes positive charge.
  - Need lots of energy to cause change in geometry of sugar
Glycosidases bind to a conformation near the transition state

Glycoside bond cleavage or formation

- Essentially just a group transfer reaction
Glycoside bond hydrolysis

- Leaving group is aglycone sugar, nucleophile is $\text{H}_2\text{O}$

Glycosyltransferase reaction

- Leaving group is NDP, nucleophile is sugar OH
Two different mechanisms of acid-base assisted glycoside bond hydrolysis

• How does an enzyme generate protons and hydroxyl ions?
• Two amino acids with carboxylic acid side-chains
  – Glutamate or aspartate
• Two mechanisms are as follows:

Acid-base assisted single displacement mechanism: inverting glycosidases

• Acid-base assisted single (direct) displacement
• The acid catalyst
  – Uncharged, hydrogen donated to the glycosidic oxygen.
• The catalytic base
  – Extracts a proton from water
  – Hydroxyl anion attacks C1 of the oxocarbenium ion transition state
• Direct displacement
  – $S_{N}2$ reaction, bimolecular, second order kinetics
Glucoamylase: classical model of an inverting mechanism

- Acid-base assisted double displacement mechanism
  - Glycosylation
    - Formation of a covalent glycosyl-enzyme intermediate (ester bond)
    - The aglycone sugar released from active site
  - Deglycosylation
    - The ester bond between the glycone sugar and the enzyme is hydrolyzed
    - Glycone sugar is released
- Two amino acids
  - One works as a general acid/base
  - The other as a nucleophile and leaving group
Mechanism of glycoside bond cleavage by GH20 β-hexosamindases: substrate-assisted catalysis

Figure 1. Proposed catalytic mechanism for family 20 β-hexosaminidase. The Hex B residues Glu355 (Hex A: Glu323) and Asp354 (Hex A: Asp332) are shown. Glu355 acts as a general acid–base residue, whereas Asp354 acts primarily to help in orienting the C2-acetamido group into a position for nucleophilic attack and subsequently to stabilize the positive charge on the oxazolinium ion intermediate. No attempt has been made to indicate the true positions of these residues. The hydroxyl groups and C5 have been removed from the pyranose ring of the substrate for clarity.


Generating a high-affinity, selective inhibitor: mimic the oxazolinium intermediate
Catalysis by Glycosyltransferases

- Inversion versus retention in anomeric configuration requires distinct enzymatic mechanisms

Glycosyltransferases: creating glycoside linkages

- Base-assisted direct displacement
  - The catalytic base
    - Extracts a proton from acceptor
    - Hydroxyl ion attacks C1 of the transition state
- Direct displacement
  - $S_N2$ reaction, bimolecular
  - Second order kinetics

NO!!!
Two (not very satisfying) models for retaining GT mechanisms

Needs to explain lack of nucleophile or any other catalytic residues in proximity of the active center!!!

Notch-modifying xylosyltransferase structures support an $S_{N1}$-like retaining mechanism

Examined the time-dependent transfer of Xyl from UDP-Xyl to Xyl-$\alpha$1,3-Glc-Notch EGF domain to form Xyl-$\alpha$1,3-Xyl-$\alpha$1,3-Glc-Notch by a retaining xyloside-$\alpha$1,3-xylosyltransferase

Three (+more?) fold types: GT-A, GT-B, GT-C
- Both GT-A and GT-B have two Rossmann domains
- GT-A tightly linked may look like a single $\beta$-sheet
- GT-B has two separate domains
- Requirement of nucleotide binding limits number of folds
- GT-C is multi-spanning TM protein, structure unknown

Glycosyltransferase structural folds

NDP-sugar donors
(or CMP-sugar for NeuAc)

Dol-P-sugar donors


Rossmann fold based on NAD$^+$ binding domain structures


Two domain structures:
- NAD$^+$ binding domain
- Substrate binding domain

Open twisted parallel $\beta$-sheet

Domain is commonly used in other nucleotide binding proteins

Surprisingly: the $\beta$-sheet scaffold conserved, not the nucleotide binding residues, multiple equally effective ways to bind NAD with the fold

Active site residues not conserved
Scaffold conserved

GTs are modular for binding of donor and acceptor

GT-A fold
- Looks like a single parallel β-sheet
- One side binds donor
- Other side binds acceptor

GT-B fold
- 1 Rossmann fold
- 6-8 β-strands
- Mixed and
  (+2-4 extra β-strands)
- 2 Rossmann folds
- 7+6 β-strands
- Usually all

GTs are modular for binding of donor and acceptor
GTs are modular for binding of donor and acceptor

**GT-A fold**
- Looks like a single parallel β-sheet
- One side binds donor
- Other side binds acceptor

**GT-B fold**
- 2 domains with separate Rossmann folds
- Can be independent with linker or can be tightly connected
- One domain binds donor
- Other domain and cleft binds acceptor
GTs are modular for binding of donor and acceptor

**GT-A fold**
- Looks like a single parallel β-sheet
- One side binds donor
- Other side binds acceptor

**GT-B fold**
- 2 domains with separate Rossmann folds
- Can be independent with linker or can be tightly connected
- One domain binds donor
- Other domain and cleft binds acceptor

**Correlation between inverting/retaining and GT-A/GT-B?**

- Inverting - clan I
  - 2, 7, 12, 13, 14, 16, 21, 25, 31, 40, 42, 43, 49, 82, 84
- Retaining - clan III
  - 6, 8, 15, 24, 27, 34, 44, 45, 55, 60, 62, 84, 78, 81

- Inverting - clan II
  - 1, 9, 10, 17, 19, 23, 26, 28, 30, 33, 41, 47, 56, 63, 80
- Retaining - clan IV
  - 3, 4, 5, 20, 32, 38, 72

**Correlation**


**NO CORRELATION** between fold and mechanism

**Synthesis of the lipid-linked precursor for Asn-linked glycosylation**

GT-C enzymes that use Dol-P-sugars


**Glycan precursor transferred to nascent polypeptides by the action of OST (oligosaccharyltransferase)**

Ost1p, Stt3p, Wbp1p, Ost3p, Ost6p, Swp1p, Ost2p, Ost5p, and Ost4p

Stt3p contains the catalytic activity
OST structure (PglB from *Campylobacter lari*) explains acceptor specificity

Protein/peptide binding

LLO binding/catalytic site

Acceptor

Asn

90°


OST (GT-C fold) mechanism inferred from crystal structure

Periplasmic domain

TM domain

El5

Groove

GT topologies, folds and mechanisms

Mechanism: Inv Inv Inv Ret Inv Ret Ret
Family: GT21 GT2 GT2 GT4 GT12 GT35 GT8

Cytosol

ER/Golgi

Lumen

Inv Inv Inv Ret Inv Inv Inv
Family: GT1 GT10, GT18, GT12, GT13, GT6, GT8 GT18, GT47, GT54, GT65GT25, GT61

GT-C

Two domain glycan co-polymerases

Matriglycan

Chondroitin Sulfate

Heparan Sulfate

GT49

GT8

LARGE

LARGE2

GT7

CHPF

CHPF2

CHSY1

CHSY3

GT64

GT74

Ext1

Ext2

ExtL1

ExtL3

GAG core protein

GAG core protein
Diversity of glycan structures reflected in complexity of glycan biosynthetic and catabolic genes

• CAZy database (www.cazy.org) assembles sequence families for:
  - Glycosyltransferases: 97 sequence families
  - Glycoside hydrolases: 133 sequence families
  - Polysaccharide lyases: 23 sequence families
  - Carbohydrate esterases: 16 sequence families
  - Carbohydrate binding modules: 70 sequence families
  
  (all as of July 2015)

• Families include all species, domains (bacteria, archaea, eukaryota, viruses)

• Database catalogs:
  - Sequences
  - Known activities (not very comprehensive)
  - Mechanisms (if known),
  - Structures (if any)
  - External links

Diversity of glycan structures reflected in complexity of glycan biosynthetic and catabolic genes

• CAZy database (www.cazy.org)
Mining the CAZy database for info on GTs

95 total CAZy GT families (GT46 and GT86 were deleted or merged)
355,418 total sequences from all 4 domains (many are species variants)
(structures determined for 150)

51 GT families bridge multiple domains of life
Mining the CAZy database for info on GTs

14 GT families have ONLY bacterial members

Mining the CAZy database for info on GTs

2 GT families have ONLY viral members (none are exclusively in Archaea)
26 GT families have ONLY eukaryotic members

73 GT families have eukaryotic members (77%) (Some are unique to eukaryotes, others span multiple domains of life)
30 GT families have eukaryotic (but not mammalian) members 
(mostly fungal or plant glycan biosynthetic enzymes)
### Mining the CAZy database for info on GTs

<table>
<thead>
<tr>
<th>Family</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammal</td>
<td>35</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Bacteria</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Archea</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>Viral</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mammal</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Bacteria</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Archea</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Viral</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### Estimates that <<<<1% of all GTs have been characterized (even for mammalian GTs <20% have been well characterized for substrate specificity)

### Mining the CAZy database for info on GTs

<table>
<thead>
<tr>
<th>Family</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacteria</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Archea</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Viral</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mammal</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Bacteria</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Archea</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Viral</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### Only 14 of 43 mammalian GT families have ≥1 structure for a mammalian member (23/199 = 11%) (fairly common for structures within a family to be quite distinct)
Case studies on GT families

- **Sialyltransferases**: BIG and DIVERSE family of GT-A inverting enzymes
- **GlcNAc transferases**: different CAZy families have similar fold but bind acceptors quite differently
- **GT2 enzymes**: can couple synthesis with extrusion across the membrane
- **β4-Galtransferase in mammary gland**: accessory proteins alters acceptor specificity
- **ABO glycosyltransferase**: polymorphisms control sugars that are transferred

Modular functions of GT-A fold enzymes in the CAZy database: sialyltransferases

<table>
<thead>
<tr>
<th>Family</th>
<th>Mammal</th>
<th>Bacteria</th>
<th>Archea</th>
<th>Viruses</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>27</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>35</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>43</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>50</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>57</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>64</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>71</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>78</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>16</td>
</tr>
</tbody>
</table>

8 CAZy GT families contain STs (mostly in bacteria)
Only 1 CAZy family in eukaryotes (GT29)
(20 mammalian GT29 STs)
Sialic acid linkages are generated by sialyltransferases (CAZY GT29)

- 20 mammalian family members, 4 subfamilies
- All employ CMP-sialic acid as donor
- Larger aglycone specificity for N-glycan, O-glycans, glycolipids
- Often have branch specificity even when terminal linkages are identical

Mammalian sialyltransferase subfamilies: CAZy GT29

- Similarity largely restricted to "sialylmotif" sequences

Modular conservation in GT-A fold: GT29 sialyltransferases

Mammalian sialyltransferase sequence alignments

Transmembrane domains


Structural similarity between rat ST6GAL1 and pig ST3GAL1

Structural similarity is mostly restricted to sialylmotif sequences
Comparison of ST6GAL1 and ST3GAL1 substrate interactions

- ST6GAL1
- CMP-NeuAc donor binding site
- Glycan acceptor binding site

- ST3GAL1
- CMP-NeuAc donor binding site
- Glycan acceptor binding site

Acceptor glycan position differs for ST6GAL1 and ST3GAL1

- ST6GAL1
  - CMP-NeuAc
  - 4'-GlcNAc

- ST3GAL1
  - CMP-NeuAc
  - 3'-GalNAc

- Similar positions of CMP-NeuAc donor, approximate position of acceptor sugar, OH nucleophile
- Rotation and reposition of sugar to provide correct acceptor OH as nucleophile
- Acceptor site residues are completely different to accommodate position and interactions with acceptor glycan
Proposed ST6GAL1 mechanism

Proposed ST6GAL1 mechanism

Proposed ST6GAL1 mechanism

**GlcNAc transferases: MGAT2 vs B3GNT2**

**MGAT2**

Requires
- Prior GlcNAc addition by MGAT1
- Prior α1,3 and α1,6-Man cleavage by GMII

Restrictions
- No prior addition to GlcNAc arm added by MGAT1 (e.g. β1,4-Gal)
- No bisecting GlcNAc

**MGAT2 complexes with UDP and acceptor**

With the additional modeling of the UDP-GlcNAc from the MGAT1 structure
MGAT2 structures demonstrate inverting GT mechanism with assistance of Mn\(^{2+}\) (and DXD motif)

- Asp319 deprotonates O2 hydroxyl of the \(\alpha_{1,6}\)-Man
- O2 attacks the C1 of GlcNAc in the UDP-GlcNAc donor (in-line nucleophilic substitution)
- Mn\(^{2+}\) stabilizes the forming charge on the UDP leaving group
- Product is a \(\beta_{1,2}\)GlcNAc linkage to the \(\alpha_{1,6}\)-Man
- Active site mutants generated for testing model for substrate recognition

B3GNT2 substrate specificity

Polylactosamine
B3GNT2 structure solved

B3GNT2 + UDP/MnCl₂ + Acceptor (LNnT)

MGAT2 and B3GNT2 have structural similarity, identical mechanism, but bind acceptors quite differently

Structural alignment of MGAT2 and B3GNT2
Glycan polymers synthesized at the cell surface

- Several glycan polymers are synthesized at the cell surface
- Employ cytosolic sugar-nucleotide donors
- Synthesize and extrude polymer through the PM

  - **Examples:**
    - Plants and bacteria: cellulose
    - Fungi: chitin
    - Mammals: hyaluronic acid

- Genes are restricted to GT2 family
- Multipass transmembrane proteins
- Model for polymer synthesis is the structure cellulose synthase (BcsA) from *Rhodobacter sphaeroides*

Structure of *Rhodobacter* cellulose synthase

Synthesis of lactose in mammary gland

- Lactose synthase is comprised of two proteins: B4GALT1 + α-lactalbumin (similar structural fold as lysozyme)
- In the absence of α-lactalbumin B4GALT1 has a very high Km for glucose (2 M) as an acceptor (but low Km for GlcNAc terminal glycans): synthesis of lactose (Gal-β1,4-Glc) is very inefficient
- α-Lactalbumin reduces the Km for glucose as an acceptor by 1000-fold: HUGE increase in lactose synthesis in mammary gland
- Expression of α-lactalbumin is induced in mammary tissues by prolactin (B4GALT1 expression is constitutively high)
- α-Lactalbumin binds to a B4GALT1-Mn^{2+}-UDP-Gal complex before binding Glc
- One of the first examples of a metal-dependent (Mn^{2+}) GTs
- Original proposal that D-X-D motif in GT sequences accounts for metal binding (NOT TRUE, correlation is good but only one Asp interacts with Mn^{2+})


Structure of B4GALT1

B4GALT1 with bound UDP-Gal (PDB 1TW1)

ABO/end
Why do you care???

Did you ever have a blood transfusion?
Did you ever have a paternity test?
Did you ever watch a police drama and see a suspect ruled in or out based on blood tests?

Blood groups are polymorphic antigen structures (usually carbohydrates) on cell surfaces that block the ability to transfer (transfuse) blood or tissues between incompatible individuals.

e.g. blood group A individuals recognize blood from group B individuals as “foreign” and mount an immune response to it.

<table>
<thead>
<tr>
<th>Blood Type</th>
<th>Antigen Present</th>
<th>Antibodies Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A antigen</td>
<td>Anti-B</td>
</tr>
<tr>
<td>B</td>
<td>B antigen</td>
<td>Anti-A</td>
</tr>
<tr>
<td>AB</td>
<td>A and B antigens</td>
<td>Anti-A and Anti-B</td>
</tr>
<tr>
<td>O</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

ABO blood groups: a natural mutagenesis experiment

Polymorphic ABO glycosyltransferase:
One gene, ABO locus (chromosome 9), three common alleles
- O allele = non-functional
- A allele (GTA) = GalNAc transferase
- B allele (GTB) = Gal transferase

Type-2 unit

How can variants of one transferase be able to transfer two different sugars?

ABO blood groups: a natural mutagenesis experiment

The difference in specificity is the substituent at C2 on the sugar-nucleotide donor
- N-acetyl for GTA
- -OH for GTB

Palic et al (2001)
Transfusion Med 11, 315-323
ABO blood groups: a natural mutagenesis experiment

GTA and GTB are identical in sequence except for 4 amino acids:

<table>
<thead>
<tr>
<th>Position</th>
<th>GTA residue</th>
<th>GTB residue</th>
</tr>
</thead>
<tbody>
<tr>
<td>176</td>
<td>Arg</td>
<td>Gly</td>
</tr>
<tr>
<td>235</td>
<td>Gly</td>
<td>Ser</td>
</tr>
<tr>
<td>266</td>
<td>Leu</td>
<td>Met</td>
</tr>
<tr>
<td>268</td>
<td>Gly</td>
<td>Ala</td>
</tr>
</tbody>
</table>

Table 1. Relative activity of hybrid A/B glycosyltransferases. A vs. B donor substrate specificity is dominated by the turnover number $k_{cat}$.

<table>
<thead>
<tr>
<th>Residue examined</th>
<th>$k_{cat}$ relative to GTA</th>
<th>$k_{cat}$ relative to GTB</th>
</tr>
</thead>
<tbody>
<tr>
<td>176, 235, 266, 288 (UDP-GalNAc)</td>
<td>Overall activity</td>
<td></td>
</tr>
<tr>
<td>A-BAA</td>
<td>1.0</td>
<td>0.003</td>
</tr>
<tr>
<td>B-BAA</td>
<td>11</td>
<td>0.006</td>
</tr>
<tr>
<td>ABA-B</td>
<td>4.9</td>
<td>0.005</td>
</tr>
<tr>
<td>AABA</td>
<td>1.2</td>
<td>0.042</td>
</tr>
<tr>
<td>AAB-B</td>
<td>1.6</td>
<td>0.15</td>
</tr>
<tr>
<td>BABA</td>
<td>2.5</td>
<td>0.07</td>
</tr>
<tr>
<td>BBA-B</td>
<td>2.0</td>
<td>0.3</td>
</tr>
<tr>
<td>BAB-B</td>
<td>1.1</td>
<td>0.5</td>
</tr>
<tr>
<td>R_{cat,GTB} / R_{cat,GTB}</td>
<td>0.06</td>
<td>1.0</td>
</tr>
</tbody>
</table>

*The $k_{cat}$ values for the hybrid A/B enzymes are expressed relative to the values for wild-type A and B enzymes with UDP-GalNAc and UDP-Gal donors, respectively.

Nature Struc. Biol. 9, 685-690

Transfus. Med 11, 315-323

a

UDP-GalNAc

b

Does not explain why GTA prefers GalNAc: Gal would fit just as well

Explains GTB preference for Gal: Steric hindrance of Met blocks access of -NAc
**ABO blood groups: a natural mutagenesis experiment**

What is “wild type” ABO sequence? Why retain ABO polymorphisms???

<table>
<thead>
<tr>
<th></th>
<th>O</th>
<th>A</th>
<th>B</th>
<th>AB</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of total</td>
<td>40.7</td>
<td>31.8</td>
<td>22</td>
<td>5.6</td>
</tr>
</tbody>
</table>

*Prevalence in the human population worldwide (prevalence varies significantly by region)*

http://www.blood.co.uk

**Epidemiology suggests:**

- Type A and B individuals less susceptible to Cholera infection than type O
- Type O may have higher incidence of peptic (stomach) ulcers, lower incidence of squamous cell carcinoma, basal cell carcinoma, pancreatic cancer
- Type A have higher incidence of stomach cancer, heart disease, pernicious anemia
- Other evolutionary forces may have driven the development of glycan diversity but may not be evident in the present population

**ABO blood groups: a natural mutagenesis experiment**

Most glycan diversity believed to result from **“Red Queen” effect**:

Lewis Carroll’s “Through the Looking Glass”

*the Red Queen states:*

**“It takes all the running you can do, to keep in the same place.”**

**Evolutionary theory:** Red Queen Principle states that continuing adaptation is needed in order for a species to maintain its relative fitness among the systems where it is co-evolving.

Explains the potential selective advantage at the level of individuals versus the constant evolutionary arms race between competing species. By mixing of the parents genes during sexual reproduction it allows a species to evolve quickly just to hold onto the ecological niche that it already occupies in the ecosystem.
Conclusions

• Extreme diversity of glycan structures in biological systems requires a similar diversity in GTs and GHs to accomplish their synthesis

• Enzymes leverage a limited number of protein folds to accomplish glycoside bond synthesis by employing one of two types of enzymatic mechanisms (inverting or retaining)

• Surprising diversity in glycosyltransferase topologies allow tethering of GT catalytic domains in unique subcellular contexts for appropriate glycan extension.

• Gene duplication and divergence (with additional potential for genetic polymorphisms within individual gene loci) led to diversification of glycan biosynthetic machinery

• All derived from a limited number of progenitor enzyme structural prototypes (GT-A, GT-B, GT-C, others?)