

2007

GlycoO Ontology

Christopher Thomas
Wright State University - Main Campus

Follow this and additional works at: <https://corescholar.libraries.wright.edu/knoesis>



Part of the [Bioinformatics Commons](#), [Communication Technology and New Media Commons](#), [Databases and Information Systems Commons](#), [OS and Networks Commons](#), and the [Science and Technology Studies Commons](#)

Repository Citation

Thomas, C. (2007). GlycoO Ontology. .
<https://corescholar.libraries.wright.edu/knoesis/239>

This Presentation is brought to you for free and open access by the The Ohio Center of Excellence in Knowledge-Enabled Computing (Kno.e.sis) at CORE Scholar. It has been accepted for inclusion in Kno.e.sis Publications by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.

GLYCO

No.

Because Ontologies cannot give a full model of the world or even a single domain, the designer needs to make choices

Is the domain artificial or natural?

Are the modeled parts static or dynamic?

Are events or objects described?

to name a few possible choices.

Depending on the choices, the ontology will take a different form.

Glycomics, or glycobiology is a discipline of biology that deals with the structure and function of oligosaccharides (chains of sugars).

The term glycomics is derived from the chemical prefix for sweetness or a sugar, "glyco-", and was formed to follow the naming convention established by genomics (which deals with genes) and proteomics (which deals with proteins). The identity of the entirety of carbohydrates in an organism is thus collectively referred to as the glycome.

en.wikipedia.org/wiki/Glycomics

The human genome presumably encodes roughly 40,000 proteins

Not enough to account for all the work they're doing

The hypothesis is that complex sugars attach to proteins at so-called glycosylation sites and change the function of the protein.

Strong impact on stem-cell and cancer-cell research

Build a domain ontology that formalizes the predominant metaphor of Glycan structures and their biosynthetic pathways.

Show that an ontology can be more than a database with a more sophisticated schema and finer grained classification.

The computational use of the ontology for classification and analysis of glycans requires

- Modeling of glycan structures
- Modeling of dependencies among glycans and other biochemical entities, such as Enzymes, Peptides and Lipids

is a focused ontology for the description of glycomics
models the biosynthesis, metabolism, and biological
relevance of complex glycans
models complex carbohydrates as sets of simpler structures
that are connected with rich relationships
573 classes, 113 relationships
Published through the National Center for Biomedical
Ontology (NCBO)
URL: <http://lsdis.cs.uga.edu/projects/glycomics/glyco>

In order to easily analyze similarities between structures, the higher level structures, i.e. glycans should be modeled using lower level building blocks, i.e. carbohydrate residue instances.

Formal Representation in OWL of GlycoTree (Takahashi, Kato 2003)

Assumption: with a large body of background knowledge, learning and extraction techniques can be used to assert facts.

Asserted facts are compositions of individual building blocks

Because the building blocks are richly described, the extracted larger structures will be of high quality


- Multiple data sources used in populating the ontology
 - o KEGG - Kyoto Encyclopedia of Genes and Genomes
 - o SWEETDB
 - o CARBBANK Database
- Each data source has a different schema for storing data
- There is significant overlap of instances in the data sources
- Hence, entity disambiguation and a common representational format are needed

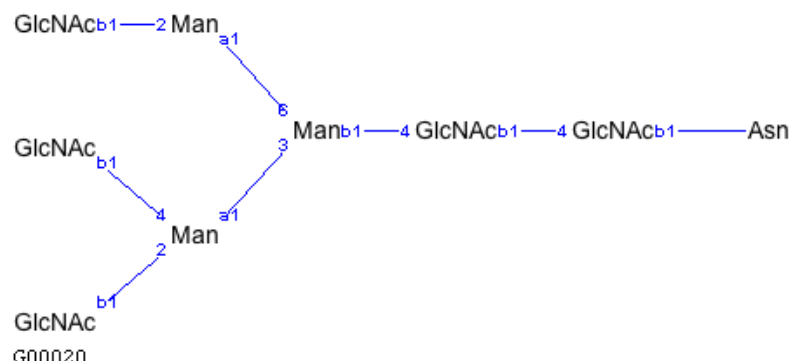
Democratic principle

Some sources can be wrong, but not all will be

More likely to have homogeneity in correct data than in erroneous data

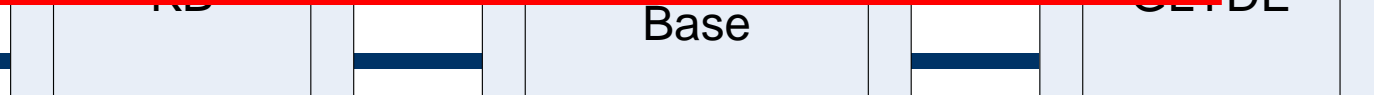
Ontology population workflow


GLYCAN: G00020
Help

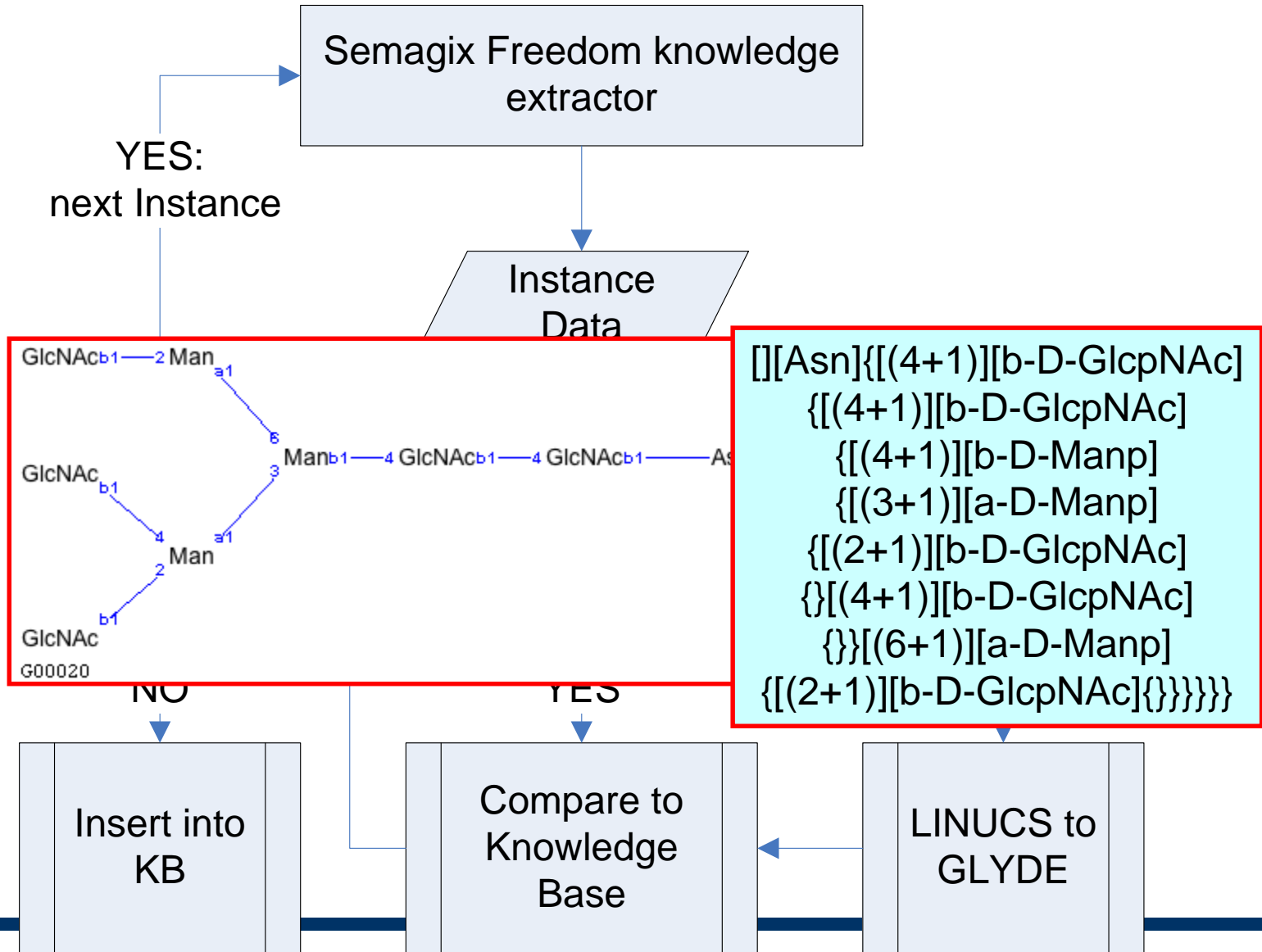
Entry	G00020 Glycan
Composition	(GlcNAc)5 (Man)3 (Asn)1
Mass	1502.4 (Asn)
Structure	 <p style="font-size: small; margin-top: 5px;">GlcNAc G00020</p> <div style="display: flex; gap: 10px;"> KCF file KCaM </div>
Class	Glycoprotein; N-Glycan
Reaction	R05987 R05991
Pathway	PATH: map00510 N-Glycan biosynthesis
Enzyme	2.4.1.145 2.4.1.155
Ortholog	KO: K00738 alpha-1,3-mannosylglycoprotein beta-1,4-N-acetylglucosaminyltransferase KO: K00744 alpha-1,3(6)-mannosylglycoprotein beta-1,6-N-acetyl-glucosaminyltransferase
Other DBs	CCSD: 4362 5058 7636 30028 33048 34937 35723 42024 43245
LinkDB	All DBs
KCF data	Show

C to
JCS

CS to
DE



Ontology population workflow



Semagix Freedom knowledge
extractor

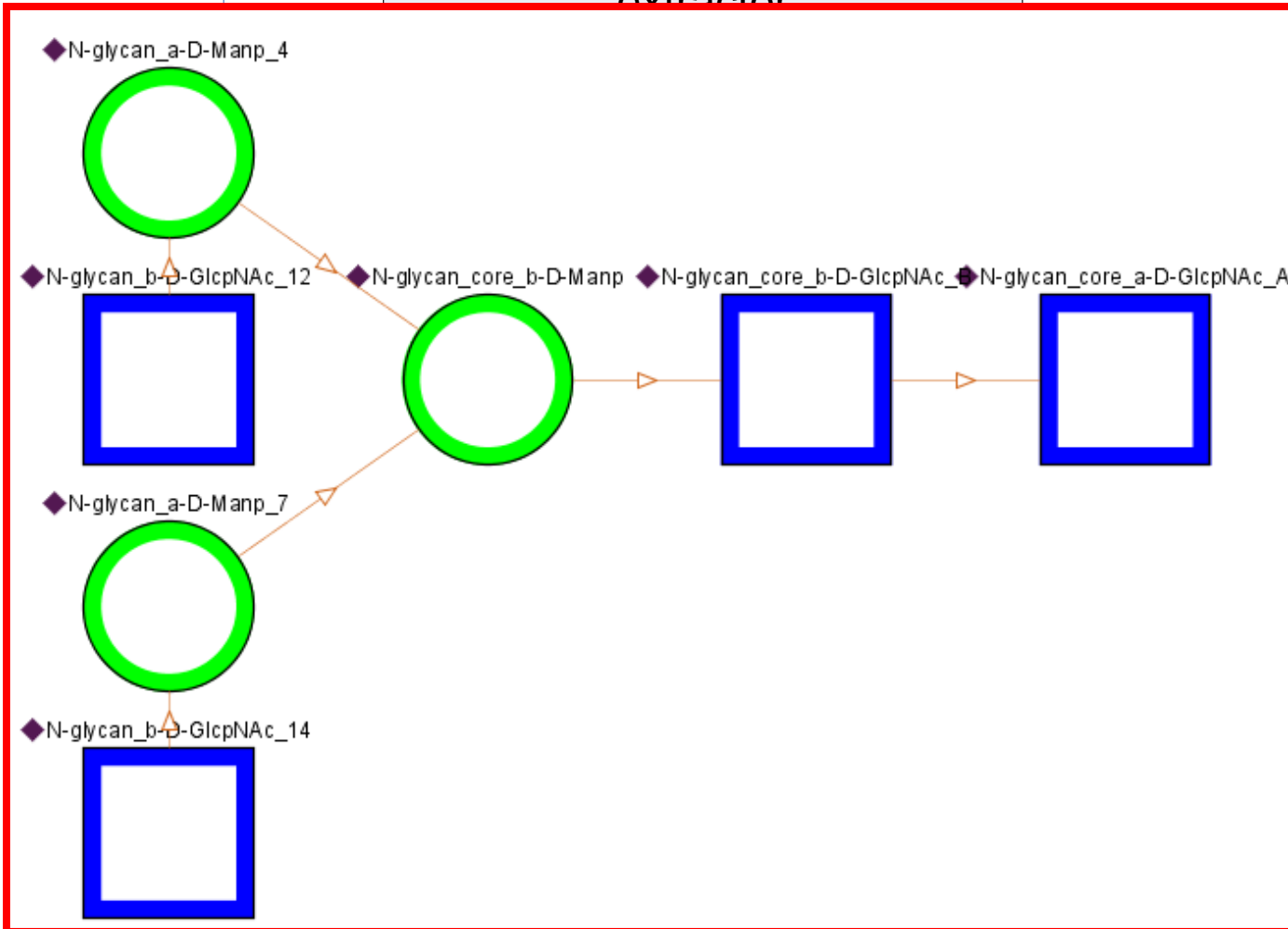
```
<Glycan>
  <aglycon name="Asn"/>
  <residue link="4" anomeric_carbon="1" anomer="b" chirality="D" monosaccharide="GlcNAc">
    <residue link="4" anomeric_carbon="1" anomer="b" chirality="D" monosaccharide="GlcNAc">
      <residue link="4" anomeric_carbon="1" anomer="b" chirality="D" monosaccharide="Man" >
        <residue link="3" anomeric_carbon="1" anomer="a" chirality="D" monosaccharide="Man" >
          <residue link="2" anomeric_carbon="1" anomer="b" chirality="D" monosaccharide="GlcNAc" >
            </residue>
          <residue link="4" anomeric_carbon="1" anomer="b" chirality="D" monosaccharide="GlcNAc" >
            </residue>
          </residue>
        <residue link="6" anomeric_carbon="1" anomer="a" chirality="D" monosaccharide="Man" >
          <residue link="2" anomeric_carbon="1" anomer="b" chirality="D" monosaccharide="GlcNAc">
            </residue>
          </residue>
        </residue>
      </residue>
    </residue>
  </residue>
</Glycan>
```

insert into
KB

Knowledge
Base

LINKS TO
GLYDE

Semagix Freedom knowledge extractor



Holds only, when the data in each source is independent
In the case of GlycO, the sources that were meant to assure quality were not diverse

- One original source (CarbBank) was copied by several Databases without curation
- Errors in the original propagated
- Errors in KEGG and CarbBank are the same

Cannot use these sources for comparison

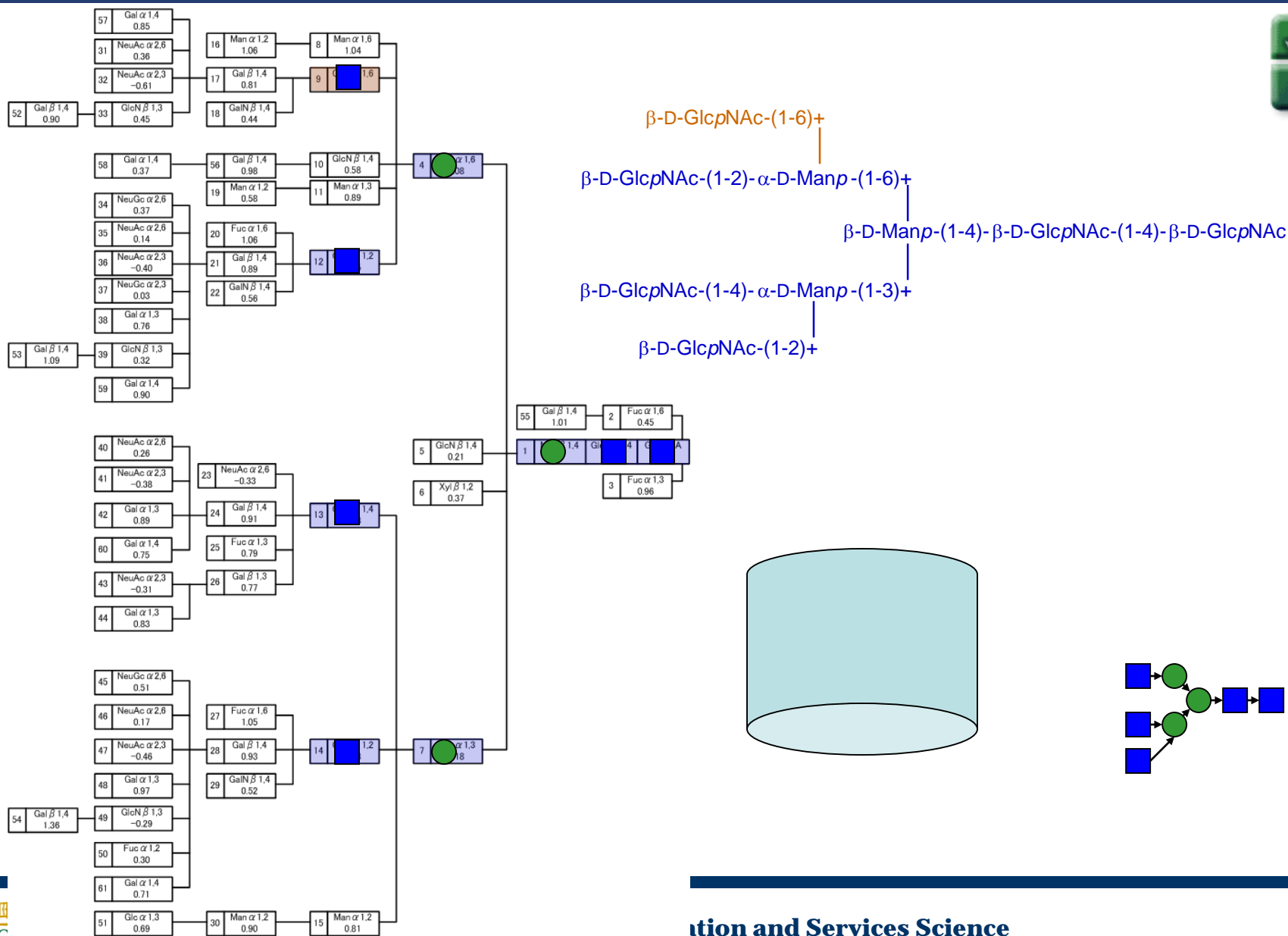
When structures (or data in general) follows well known patterns, unknown data can be compared to these patterns

In the case of GlycO, the number of possible patterns is limited by natural processes

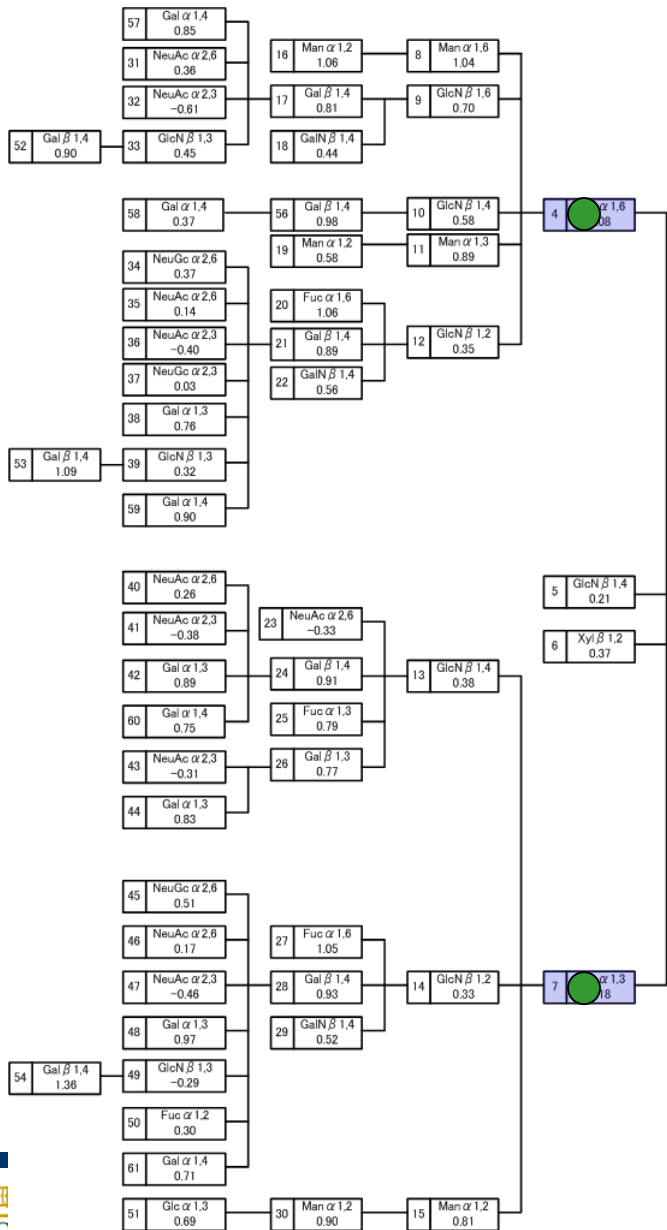
- Metabolic pathway steps
- Enzymes
- Monosaccharide residues

One identified pattern is the Glycotree

GlycoTree – A Canonical Representation of N-Glycans



Database Verification using Glyco

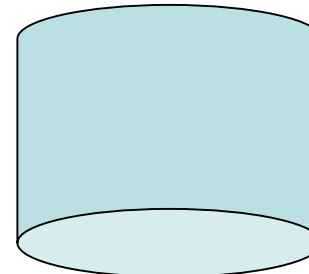


b-D-Manp-(1-6)+

a-D-Manp-(1-4)-b-D-GlcpNAc-(1-4)-D-GlcNAc

b-D-Manp-(1-3)+

a-D-Manp-(1-4) is not part of the identified canonical structure for N-Glycans, hence it is likely that the database entry is incorrect



Pathway Steps - Reaction

R06264

- has_acceptor_substrate:** lipid-linked N-glycan precursor molecule G10599
- has_product:** lipid-linked N-glycan precursor molecule G00008
- has_product:** dolichol phosphate
- has_glycosyl_donor_substrate:** dolichol-phosphate-glucose
- adds_glycosyl_residue:** N-glycan a-D-Glcp 71
- has_acceptor_residue:** N-glycan a-D-Glcp 70
- has_reactant:** dolichol-phosphate-glucose
- has_reactant:** lipid-linked N-glycan precursor molecule G10599

has_reactant

R06264

has_product

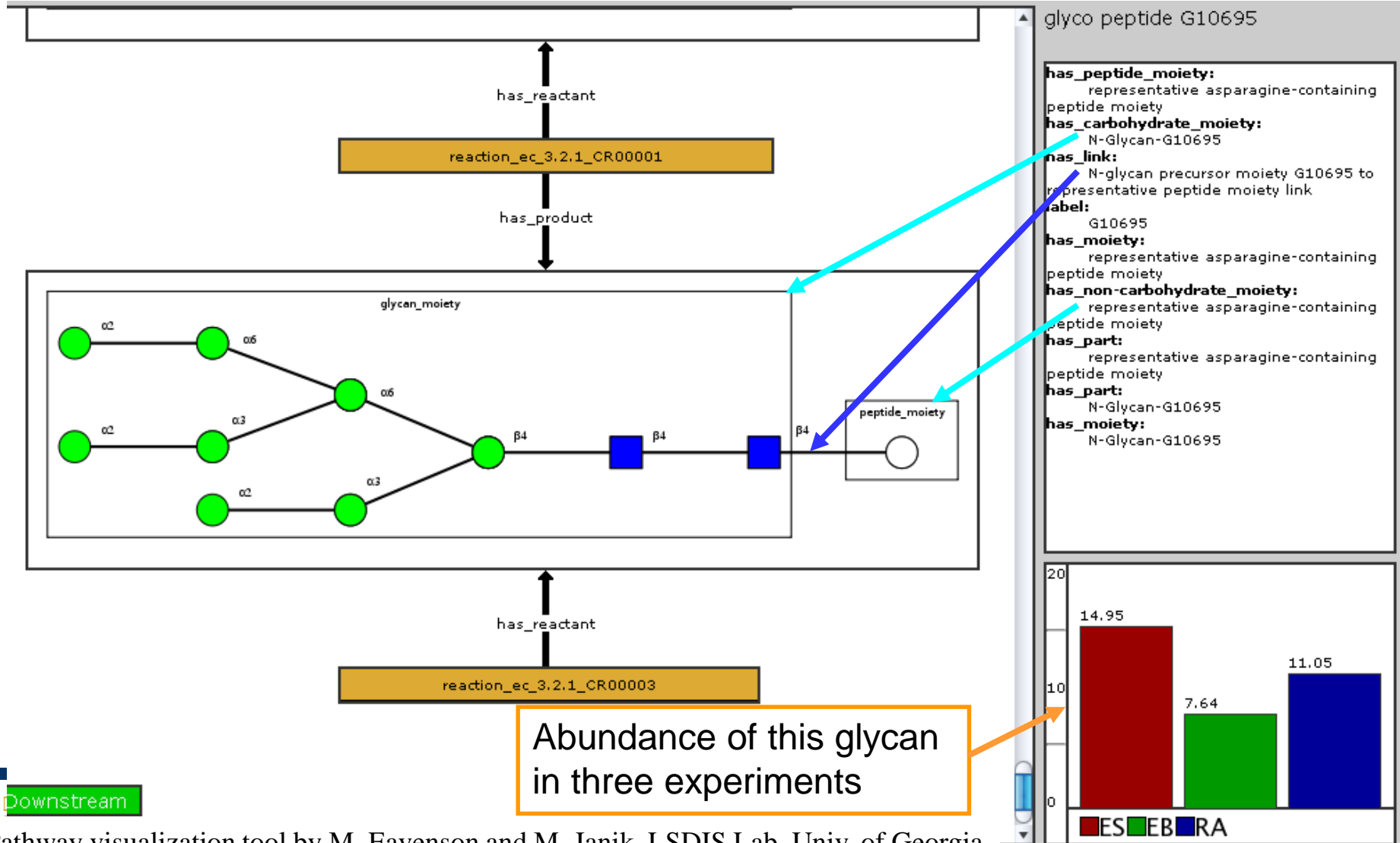
20

Evidence for this reaction from three experiments

0 .19 .09 .17

■ ES ■ EB ■ RA

Pathway Steps - Glycan



The diagram illustrates the synthesis of a glyco peptide (G10695) through two reaction steps. The central glycan moiety is composed of several sugar units (green circles) and peptide moieties (blue squares and white circle). The structure is defined by linkage types: $\alpha 2$, $\alpha 6$, $\alpha 3$, $\beta 4$, and $\beta 4$.

Reaction steps are shown as boxes:

- reaction_ec_3.2.1_CR00001**: has_reactant (top), has_product (bottom)
- reaction_ec_3.2.1_CR00003**: has_reactant (bottom)

The glyco peptide G10695 is detailed in the right panel:

- has_peptide_moiety:** representative asparagine-containing peptide moiety
- has_carbohydrate_moiety:** N-Glycan-G10695
- has_link:** N-glycan precursor moiety G10695 to representative peptide moiety link
- label:** G10695
- has_moiety:** representative asparagine-containing peptide moiety
- has_non-carbohydrate_moiety:** representative asparagine-containing peptide moiety
- has_part:** representative asparagine-containing peptide moiety
- has_part:** N-Glycan-G10695
- has_moiety:** N-Glycan-G10695

Abundance of this glycan in three experiments is shown in the bar chart:

Experiment	Abundance
ES	14.95
EB	7.64
RA	11.05

Downstream

Abundance of this glycan in three experiments

Pathway visualization tool by M. Eavenson and M. Janik, LSDIS Lab, Univ. of Georgia

