Glycans in Pathway Tools

by

Markus Krummenacker
SRI International
Q1 2013
What are Glycans?

- Macromolecules: Polysaccharides or Oligosaccharides
- Decoration of proteins or lipids
- Structural materials, cell walls: cellulose, chitin
- Storage polymers: glycogen, amylose
- Structures often contain 100s or 1000s of saccharide units
- Important biochemically, and for Biofuel production
Goals for Pathway Tools

• Glycan structures can be very big
• Showing big glycans in atomic detail results in overwhelming clutter
• Instead, icons with colors clarify what the building-blocks are and how they are connected
• Support editing and display of glycans
• For Biofuel production, the goal is to display summarized degradation pathways
Icons clarify glycan structure
Chosen Representation Standards

• Display:
  • CFG icons (Consortium for Functional Glycomics)
  • Different shapes and colors are combined to represent saccharide units
  • Standard chemical modifications have their own icons
Extended Carbohydrate Symbol Notation

Hexoses: All NAc's and amines are in the two position. All A's are in the 5 position.

- Glucose
- Galactose
- Mannose
- Idose
- Galactose
- Allose
- Lelose
- Allose

6-Deoxy Sugers: All NAc's and amines are in the two position.

- Fucose
- Quinovose
- Rhamnose
- Pentoses
- Ribose
- Arabinose
- Xylose
- Lycose

Ketoses
- Fructose
- Sorbose
- Polylactose
- Tagatose

Glutaric Acids:
- NAc
- NAc
- KDO

Amino Acids:
- Pseudaminic acid
- Legionaminic acid
- 8-Epi-legionaminic acid
- 4-Epi-legionaminic acid

Sugar is in L-configuration: (Draw to bottom-left of sugar)
Sugar is in D-configuration: (Draw to bottom-left of sugar)

All Sugars by default are in the pyranose form. If clarification or emphasis is needed, do as below:

- L: Example: indicates L-RhaP
- D: Example: indicates D-RhaP

Substituted functional groups are connected to the specified carbon:
- 4'-β-D-GlcpNAc (1-6)-α-D-GlcpNAc (1-4)-β-D-GlcpA (1-3)-α-D-GalpNAc (1-

Chosen Representation Standards

- File exchange:
  - GlycoCT XML
  - GlycoCT has 3 flavors: XML, condensed, compressed
  - Namespaces of all entities are controlled
  - Canonical numbering of residues and linkages
  - Allows a variety of chemical modifications of residues
  - Allows repetitive segments, structural ambiguities
GlycoCT technical details

1. configuration of anomeric C-atom, o for open chain
2. configuration and 3 letter code for stemtype. This section can be repeated to accommodate carbohydrates with more than 4 stereogenic centers. Optional.
3. superclass
4. numerical value ring begin
5. numerical value ring end
6. position of modifier multiple connections are separated by "|
7. type of modifier
   - d = deoxygenation
   - keto = carbonyl function
   - en, enx = double bond
   - a = acidic function
   - aldi = reduced C1-carbonyl
   - sp2 = outgoing linkage with double bond
   - sp = outgoing linkage with triple bond
   - geminal = 2 OH at one backbone carbon

Examples

- b-dglc-hex-1:5
  - β-D-Glucose, pyranose form
GlycanBuilder

- Open source JAVA applet, developed by the EUROCarbDB project
- Uses layout algorithms to draw glycans in standardized ways
- Pathway Tools communicates with GB by starting a local HTTP server, launching a Web browser, and displaying GB in the browser, while exchanging the glycan structure as GlycoCT XML (similar to the Marvin compound editor applet)
- Glycan icon structures are available for compounds under the Glycans class
Editing with GlycanBuilder

CPD-13416: GXGG xyloglucan oligosaccharide

m/z: 1045.5037 [MONO, perMe, Na, 0, freeEnd]
Pathway Tools technical details

• Compounds under the class Glycans can have both a traditional atomic structure and a icon structure
• Layout coordinates are sent by GlycanBuilder (special extension)
• Additional slots record icon structure:
  • GROUP-COORDS-2D: (186 82), (134 82), (82 82), (30 82), (82 30)
  • STRUCTURE-GROUPS: |B-DGLC-HEX-1:5|, |B-DGLC-HEX-1:5|, |B-DGLC-HEX-1:5|, |B-DGLC-HEX-1:5|, |A-DXYL-PEN-1:5|
  • STRUCTURE-LINKS: (1 2 О 4 D 1), (2 3 О 4 D 1), (3 4 О 4 D 1), (3 5 О 6 D 1)
Future Work

• Support for non-canonical NON residues. Extend GlycanBuilder to enter free-form text for them.
• Support polymer repeat units
• Implement Glycan (degradation) pathway displays that show at which linkages enzymes act
• Consistency checks between traditional compound structure versus icon structure, if both are present