

HGPI Initiative
HUPO Meeting : Long Beach, CA
October 29, 2006

**Bioinformatics Approach to
Functional Glycomics**

James C. Paulson - The Scripps Research Institute

HGPI 2006



CONSORTIUM FOR FUNCTIONAL GLYCOMICS

funded by National Institute of General Medical Sciences

Consortium for Functional Glycomics

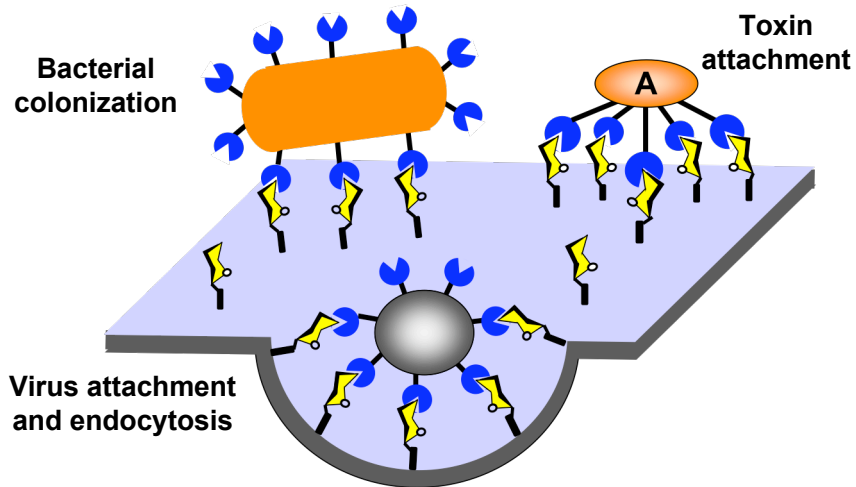
Goal:

***Define paradigms by which protein-
carbohydrate interactions mediate cell
communication***

<http://www.functionalglycomics.org>

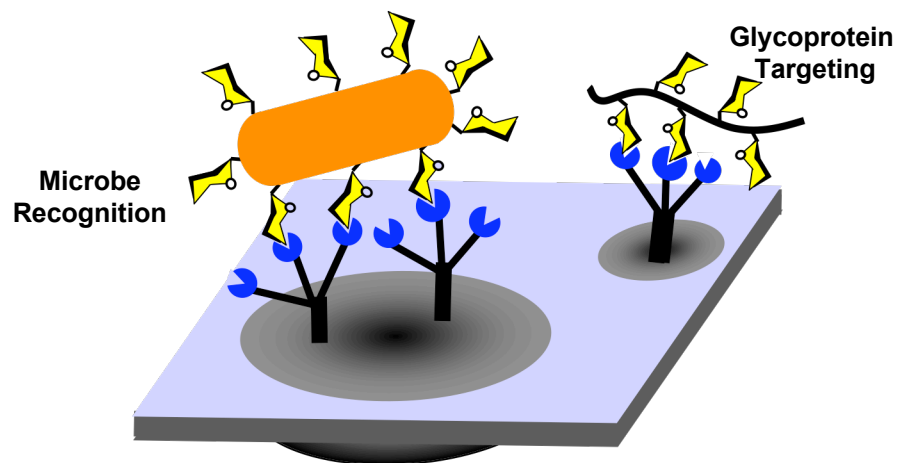
HGPI 2006

Attachment of microbial pathogens to host cell glycans



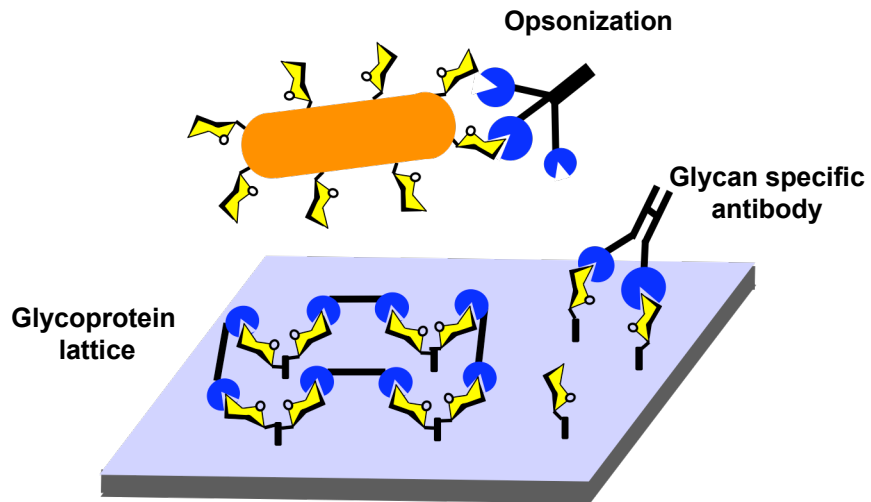
HGPI 2006

Cell surface GBPs binding pathogen and glycoprotein glycans



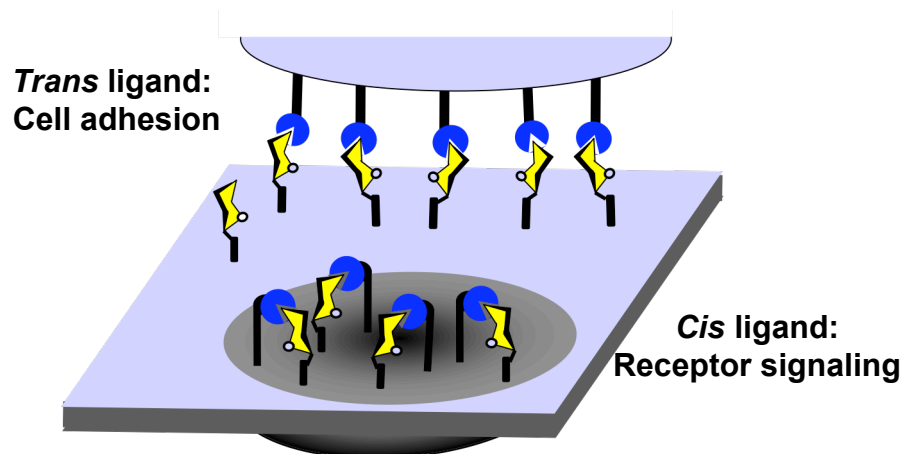
HGPI 2006

Soluble GBPs binding cell surface and pathogen glycans

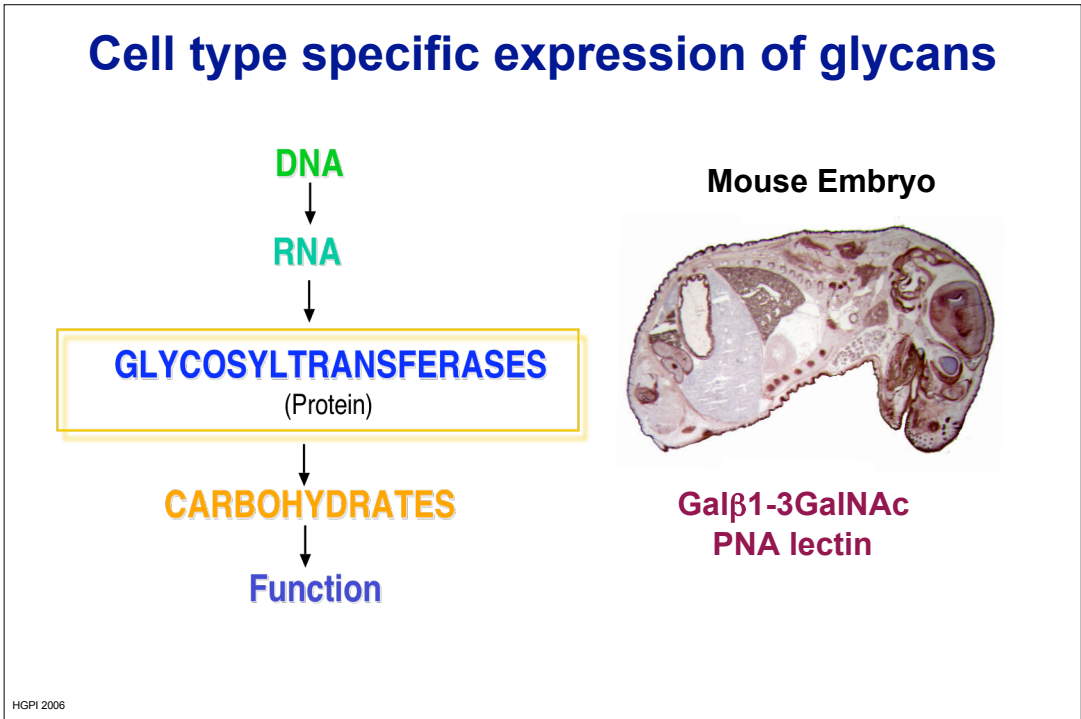
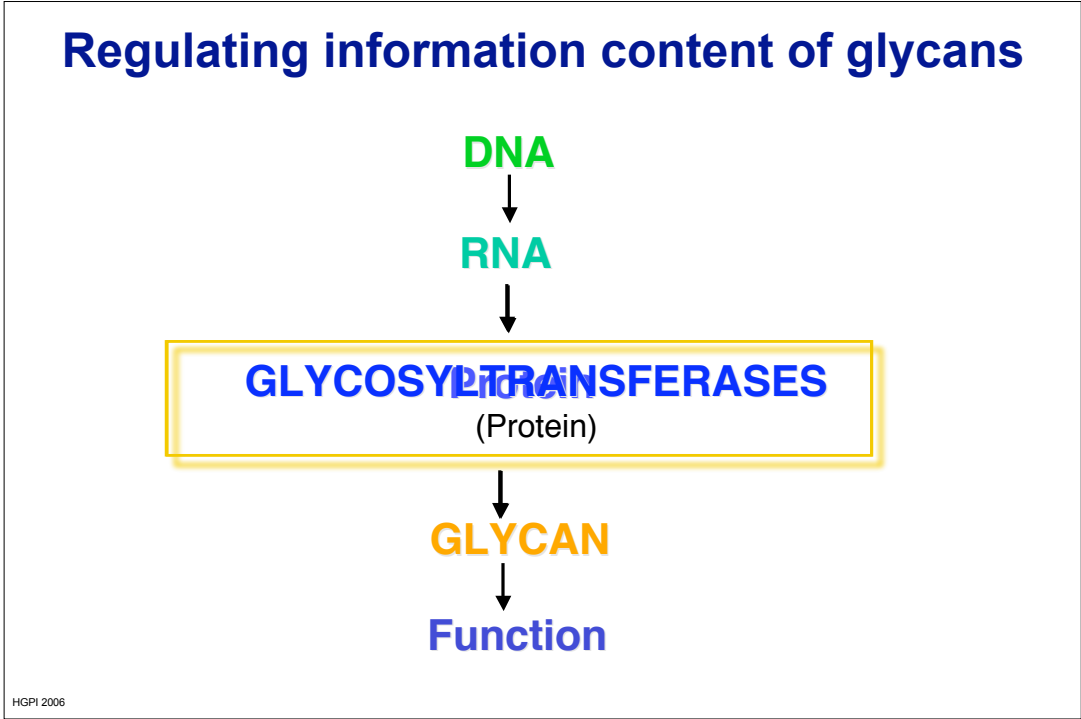


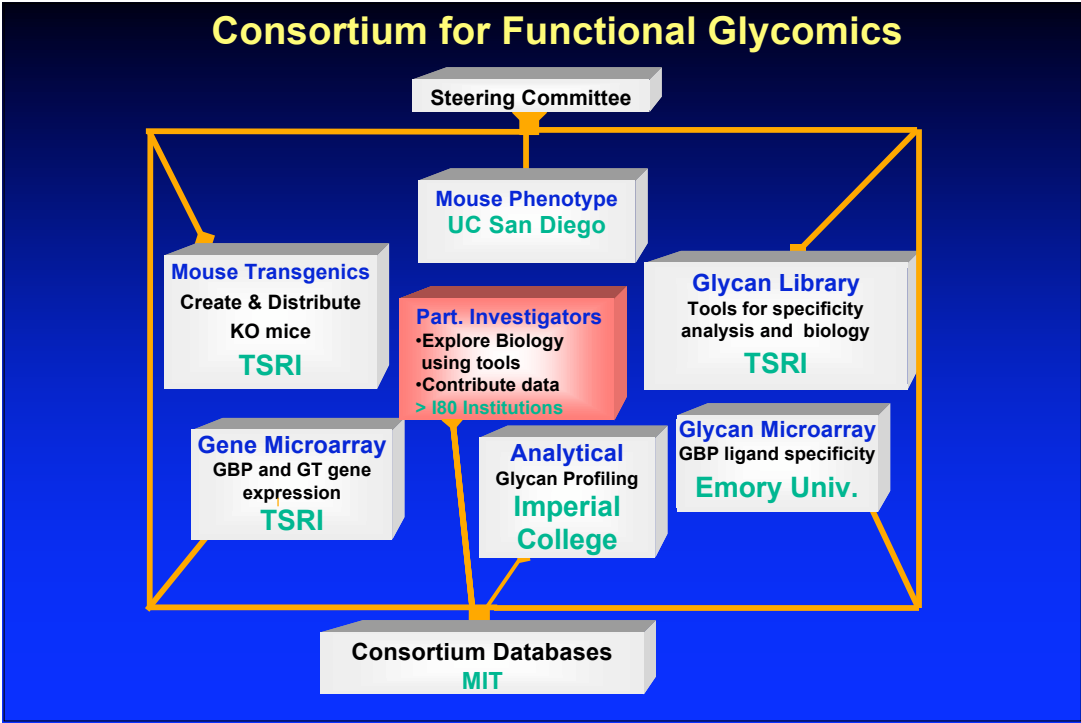
HGPI 2006

Cell surface GBPs binding cell surface glycans



HGPI 2006





Consortium for Functional Glycomics : Glycobiology Resources Offered

Carbohydrate compounds



Over 230 carbohydrate compounds available with azido or biotin functional groups

Glycan array screening



Specificity of your glycan-binding protein on microarray with ~300 glycoprotein and glycolipid glycans

Glyco-gene microarray



Analyze your mRNA on a custom gene microarray including 2,000 human and mouse glyco genes

Mouse knockout strains



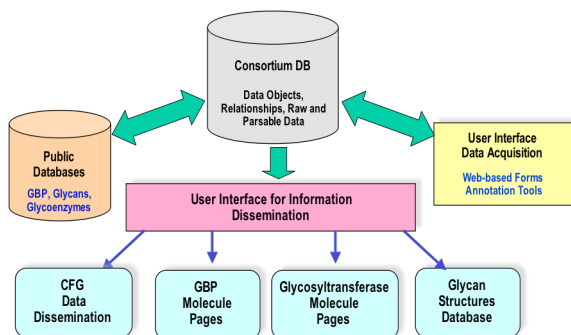
Novel knockout mouse strains with genes ablated for glycan-binding proteins

<http://www.functionalglycomics.org>

HGPI 2006

Bioinformatics

Bioinformatics
Rahul Raman
Ram Sasisekharan
MIT



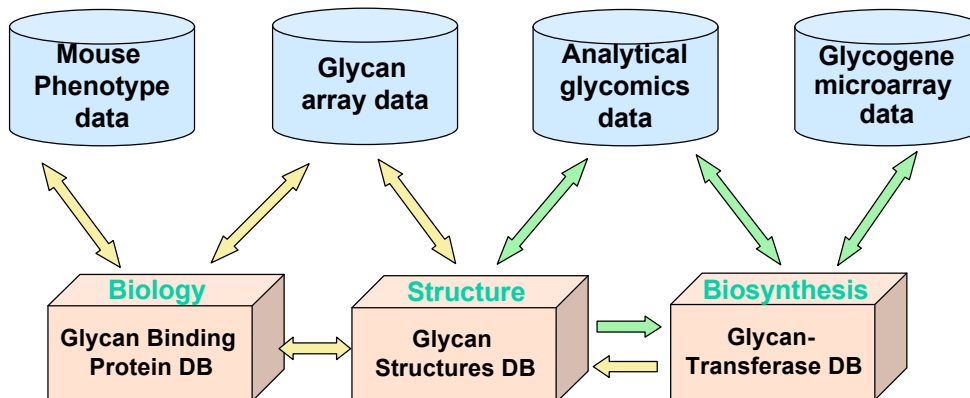
- Implement vision for CFG databases developed with the SC, Scientific Cores and Participating Investigators.
- Develop interfaces for acquisition and dissemination of Consortium data.
- Develop specialty databases integrated with CFG data as resources for PIs and the scientific community.

HGPI 2006

Integration of interfaces for CFG data and specialty databases

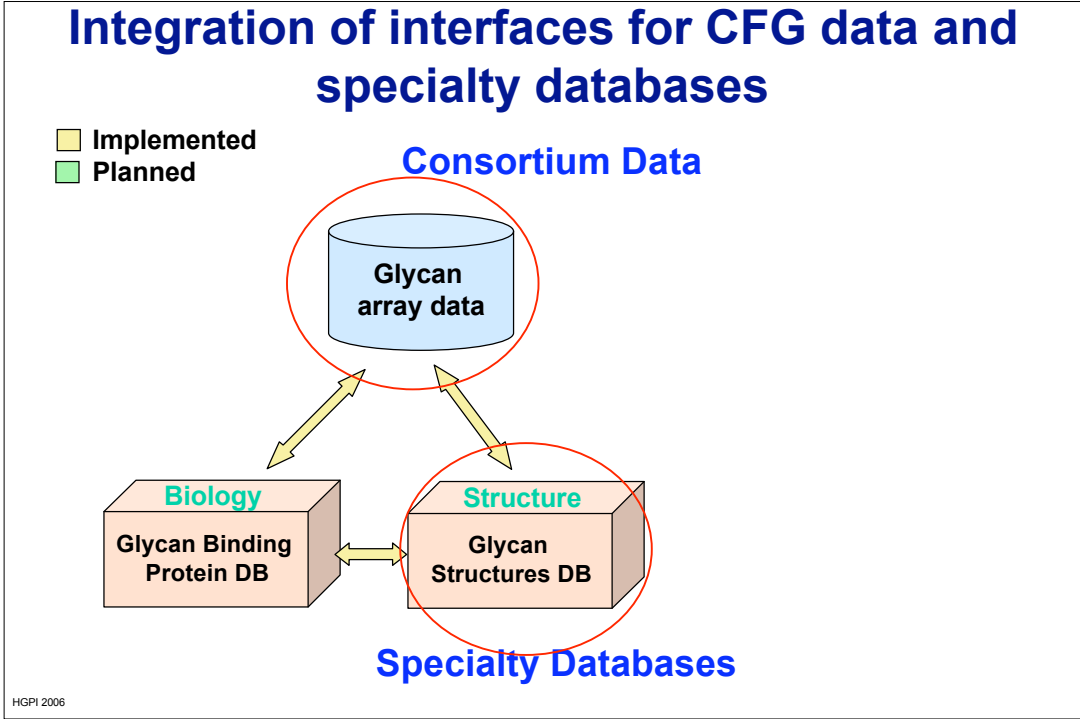
■ Implemented
 ■ Planned

Consortium Data



Specialty Databases

HGPI 2006



CFG Glycan Microarray

Synthetic Glycans + **NHS-activated glass surface**

OC[C@H]1O[C@H](CO)O[C@H](CO)O[C@H]1O + CCCC(=O)N(CCC1OC(CO)CO1)C(=O)O

Standard microarray robotics

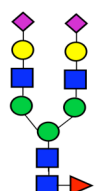
Blixt et al. (2004)
PNAS 101, 17033

Year	Number
Year 1	~10
Year 2	~20
Year 3	~50
Year 4	~150

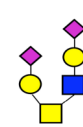
<http://functionalglycomics.org>

Stable amide bond

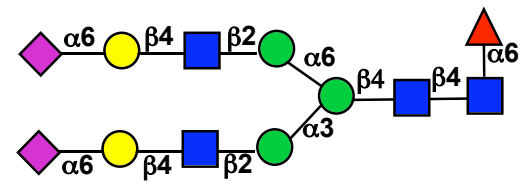
Symbol Nomenclature



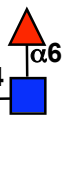
N-Linked



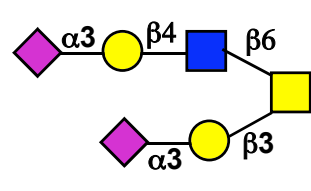
O-Linked



NeuAc α 2-6Gal β 1-4GlcNAc β 1-2Man α 1
 NeuAc α 2-6Gal β 1-4GlcNAc β 1-2Man α 1



Fuc α 1
 Man β 1-4GlcNAc β 1-4GlcNAc-Asn



NeuAc α 2-3Gal β 1-4GlcNAc β 1
 NeuAc α 2-3Gal β 1
 GalNAc α -Thr

Modified IUPAC nomenclature

Key: ◆ NeuAc ● Gal ● Glc ● Man ▲ Fuc ■ GalNAc ■ GlcNAc

HGPI 2006

Consortium for Functional Glycomics - Netscape

http://www.functionalglycomics.org/static/consortium/

CONSORTIUM FOR FUNCTIONAL GLYCOMICS

funded by National Institute of General Medical Sciences

CONSORTIUM HOME DATABASES SITE MAP CONTACT US

Consortium

People

Resources

Consortium Data

CFG databases

Policies

Join

Member Login

Partners

Links

What is the CFG?

The Consortium for Functional Glycomics (CFG) is a large research initiative funded by **NIGMS** to understand the role of carbohydrate-protein interactions at the cell surface in cell-cell communication. The **Scientific Cores** are producing a variety of resources and services for **distribution to investigators** for use in performing experiments that address the Specific Aims of the Consortium. Members of the Consortium (Participating Investigators) are able to access data posted in the **members' database** prior to its being posted in the **public database**. Click [here](#) to see an overview of the Consortium's strategic plan.

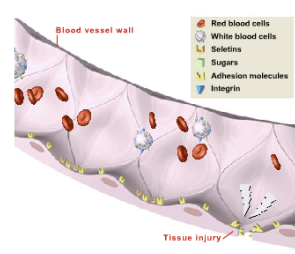


Illustration by John MacNeil
Animation by Matthew Beauchamp

[Back](#) [Next](#) [Animation Credit](#)

In The News

The Consortium Quarterly ([Full List](#))

- Vol. 4 No. 3, March 2006

Research Articles ([Full List](#))

- Stevens J et al. *Science*. 2006;312:404-410.
- McGrail SP et al. *Glycobiology*. 2006;16:422-430.
- Stevens J et al. *J Mol Biol*. 2006;335:1143-1155.
- Comelli EM et al. *Glycobiology*. 2006;16:117-131.
- Desplats PA et al. *J Neurochem*. 2006;96:743-757.
- Ohtsubo K et al. *Cell*. 2005;123:1307-1321.
- Xia B et al. *Nat Methods*. 2005;2:845-850.
- Kawano S et al. *Bioinformatics*. 2005;21:3976-3982.
- Zalonsk et al. *Nat Immunol*. 2005;6:810-818.
- Coombs et al. *J Biol Chem*. 2005;280:22993-22999.

Review Articles ([Full List](#))

- Paulson JC et al. *Nat Chem Biol*. 2006;2:238-248.
- Raman et al. *Glycobiology*. 2006;16:82R-90R.
- van Die I and Cummings RD. *Chem Immunol Allergy*. 2006;90:91-112. (Reprinted with permission from S. Karger AG, Basel.)
- Lehman MA. *Crit Rev Biochem Mol Biol*. 2006;41:51-75.
- Crocker PC. *Curr Opin Pharmacol*. 2005;5:431-437. (Reprinted with

Consortium for Functional Glycomics - Netscape

http://www.functionalglycomics.org/glycomics/common/jsp/firstpage.jsp

CONSORTIUM FOR FUNCTIONAL GLYCOMICS
Funded by National Institute of General Medical Sciences

CONSORTIUM HOME DATABASES SITE MAP CONTACT US [Feedback](#)

Consortium data

- Glycan Profiling
- Gene Microarray
- Mouse Phenotyping
- Glycan Array**

GBP database
Glycan database
GT database
Resources

Consortium Member
LOGIN

Click icon to login to Central Database

Databases

Part of the overall objective of the **Bioinformatics Core (Core B)** is presenting consortium information to investigators and potential contributors and disseminating results generated to the public and scientific community. For this purpose Core B is developing several complex relational databases and interfaces to facilitate linking appropriate data (from within the consortium as well as from other related databases worldwide) and engineering them for best usability. These databases and tools should prove a major resource for both the Consortium and the larger scientific community.

© 2002-2006 Consortium for Functional Glycomics. All rights reserved.
This website is best viewed with Internet Explorer 5.0+ or Netscape 6.0+

Consortium for Functional Glycomics - Netscape

http://www.functionalglycomics.org/glycomics/publicdata/selectedScreens.jsp?sortBy=investigatorName&sortOrder=asc

CONSORTIUM FOR FUNCTIONAL GLYCOMICS
Funded by National Institute of General Medical Sciences

CONSORTIUM HOME DATABASES SITE MAP CONTACT US [Feedback](#)

Readout for this primary screen
Search Readout
Bar Chart For 100 MicroMolar Concentration

Core H Bar Chart

Glycan Being Browsed

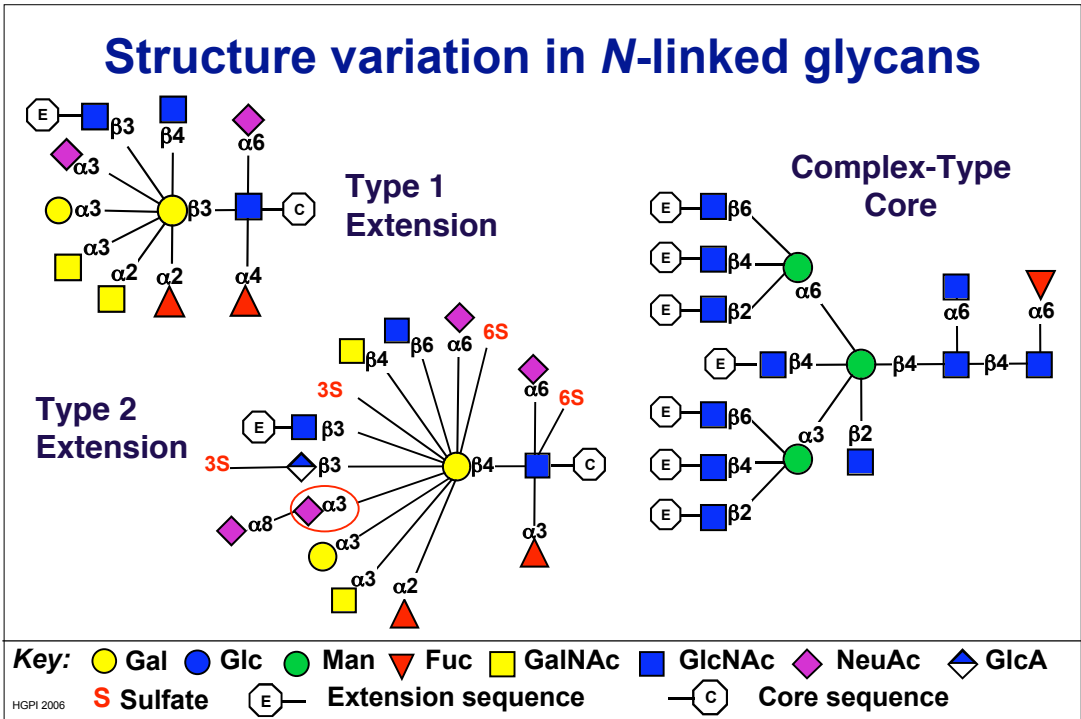
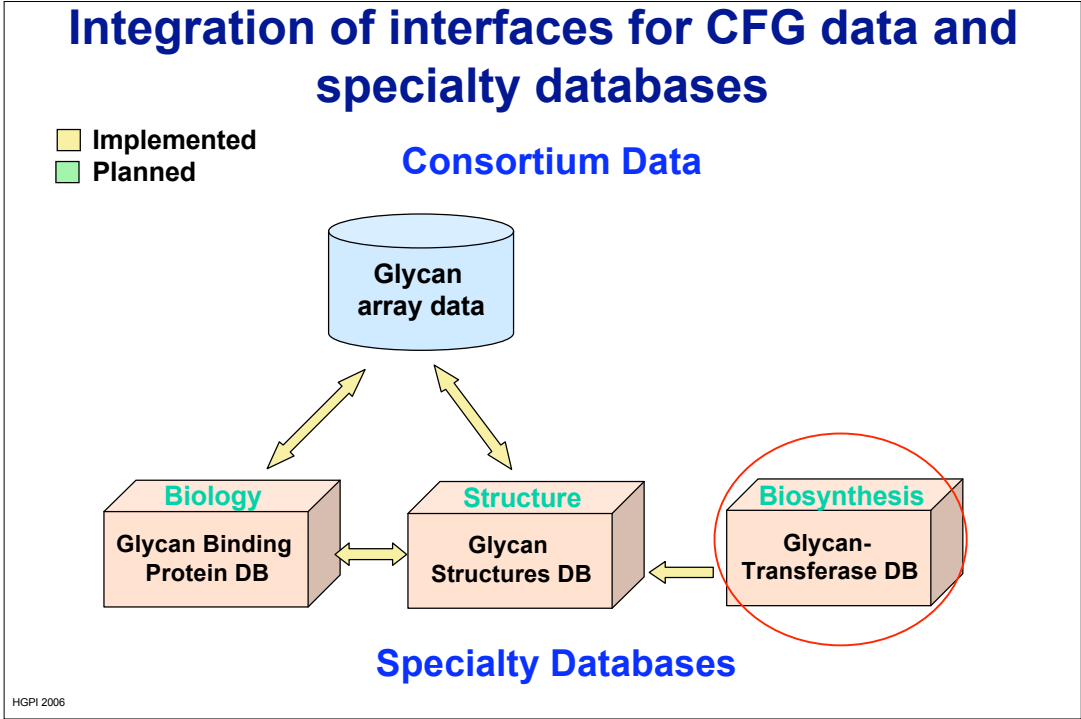
Cartoon Representation
IUPAC 2D Representation
IUPAC Code
Linear Code
Sub Structure Search Interface
General Information

Glycan : carbSynthe_0751_D000

Glycan Family: Synthetic
Sub. Family: Synthetic
Last Updated: 04/20/2005

Date Received [mm/dd/yyyy]: 11/01/2005

Serum from Sheep Vaccinated with <i>H. contortus</i>	Other	Other	Irma Van Die	Sheep sera inoculated with <i>H. contortus</i> (3-49)	Printed Array ver. 2 May 17, 06
Serum from Sheep Vaccinated with <i>H. contortus</i>	Other	Other	Irma Van Die	Sheep sera inoculated with <i>H. contortus</i> (4-49)	Printed Array ver. 2 May 17, 06



Consortium for Functional Glycomics - Netscape

http://glycomics3.mit.edu/glycomics/molecule/jsp/glycoEnzyme/gtdb.jsp?attachingNumber=54&baseNumber=49&type=Type+2&sequenceType=Extension&title=Te...

CONSORTIUM FOR FUNCTIONAL GLYCOMICS
funded by National Institute of General Medical Sciences

CONSORTIUM HOME DATABASES SITE MAP CONTACT US Feedback

General Reference Genome Proteome Activity

ST3Gal III (Human)

Putative involvement in biosynthesis of the linkages at the non-reducing end

Extension Term/Extension: Type 1

Glycosphingolipid Term/Extension: Lacto(L)

Extension Term/Extension: Type 2

Link to KEGG pathways KEGG entry for 2.4.99.6

© 2002-2006 Consortium for Functional Glycomics. All rights reserved.
This website is best viewed with Internet Explorer 5.0+ or Netscape 6.0+

Netscape

http://glycomics3.mit.edu/glycomics/molecule/jsp/carbohydrate/searchResults.jsp?sortBy=moWt&sortOrder=asc

CONSORTIUM FOR FUNCTIONAL GLYCOMICS
funded by National Institute of General Medical Sciences

CONSORTIUM HOME DATABASES SITE MAP CONTACT US

Home Search By: Sub Structure | Mol.Wt. | Composition | Linear Nomenclature | Multiple Criteria

Primary Screen GlycanScreen primscreen_GLYCAN_v3_81_03312005 - Netscape

Glycan Carbohydrate carbSynthe_0136_D000 - Netscape

Search Interface

Core H Bar Chart

Glycan Being Browsed

Bar Chart Without PAA Ligands

Binding GBP Information

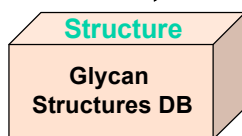
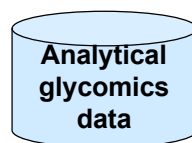
674.595	NeuAc2-3Galb1-3GlcNAc#SP	HEX1HEXNAc1NEUAc1	Synthetic	Synthetic
---------	--------------------------	-------------------	-----------	-----------

© 2006 Consortium for Functional Glycomics. All rights reserved.
This website is best viewed with Internet Explorer 5.0+ or Netscape 6.0+

Integration of interfaces for CFG data and specialty databases

- Implemented
- Planned

Consortium Data



Specialty Databases

HGPI 2006

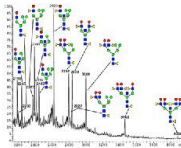
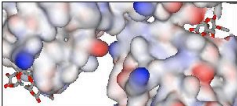

Address: <http://www.functionalglycomics.org/glycomics/molecule.jsp/carbohydrate/carbMoleculeHome.jsp>

CONSORTIUM FOR FUNCTIONAL GLYCOMICS
Funded by National Institute of General Medical Sciences

CONSORTIUM HOME DATABASES SITE MAP CONTACT US

Home Search By: Sub Structure | Mol.Wt. | Composition | Linear Nomenclature | Multiple Criteria

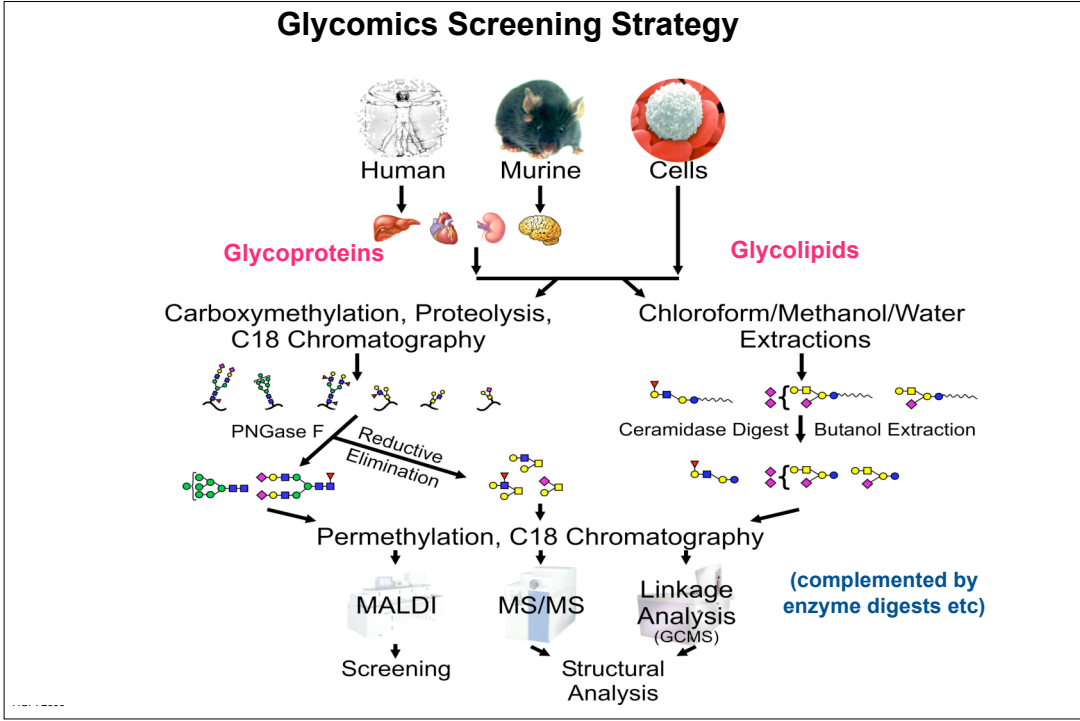
Glycan Database

<p style="text-align: center;">Updates</p> <ul style="list-style-type: none"> First cut version of glycan structures database Contains nearly 7500 entries Each entry contains structural and chemical information as well as related references Different search interfaces are provided via the menu above The database will be regularly updated with newly synthesized or discovered glycans 	<p style="text-align: center;">Search for glycans</p> <ul style="list-style-type: none"> Sub-structure Molecular weight Composition Linear nomenclature Use multiple search criteria
<p style="text-align: center;">Source of glycan structures</p> <ul style="list-style-type: none"> <i>N</i>- and <i>O</i>-linked glycans from CarbBank Glycominds Ltd. seed database <i>N</i>- and <i>O</i>-linked glycans identified in tissues and cells analyzed by the Analytical Glycotechnology Core (C) Glycans elaborated on the glycan array Glycans synthesized by the Carbohydrate Synthesis Core (D) and available as a resource 	<p style="text-align: center;">Glycan nomenclature</p> <ul style="list-style-type: none"> Glycans are displayed in several formats for ease of use. The Consortium nomenclature for representing glycans can be found here.

This page has been visited **42900** times.

© 2002-2005 Consortium for Functional Glycomics. All rights reserved.



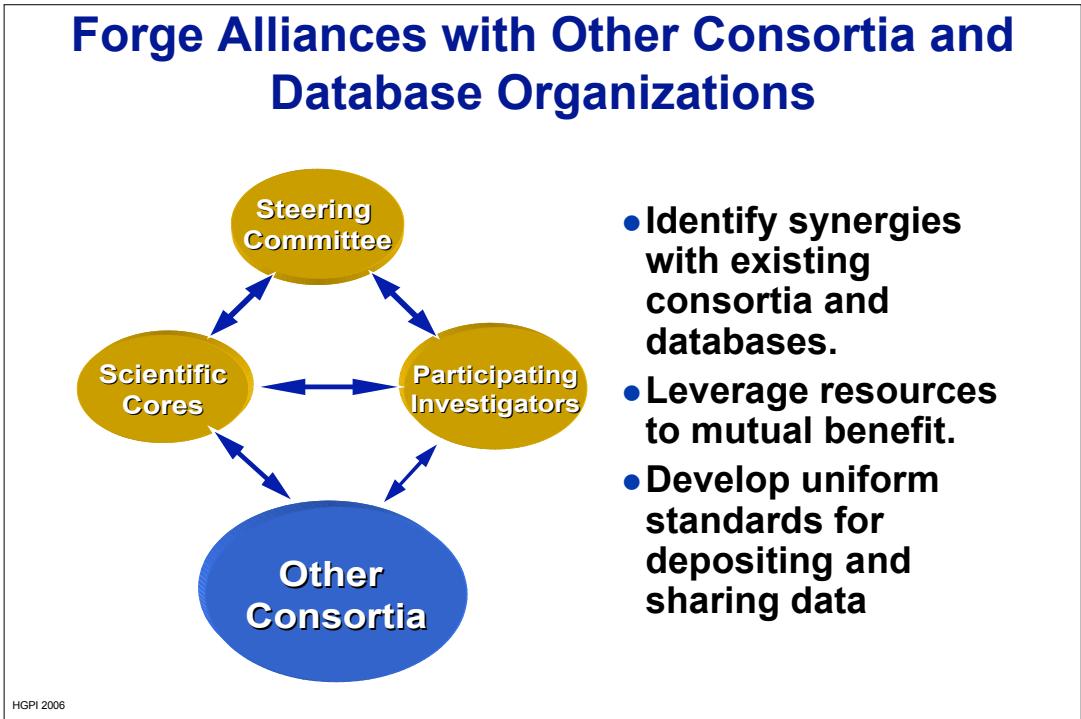
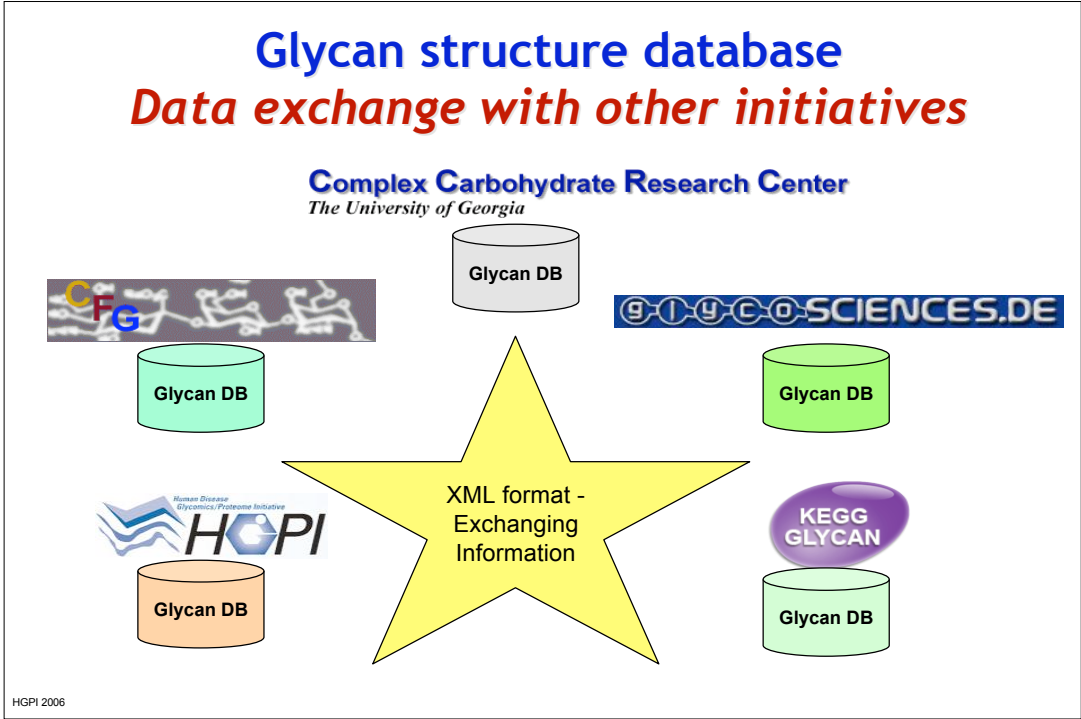
Glycan Structures Database

Integration with Glycan Analysis Data


Glycan : carbNlink_40124_D000

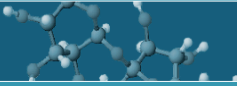
Cartoon Representation		Biological Sources			
IUPAC 2D Representation		Taxonomy Name	Organ	Tissue Type	Cell Type
			[Lewis lung cancer] Urine, Placenta [Cancer, Tamm - Horsfall, Invasive Mole]		
		Homo sapiens (Human)			Myelomonocyte cells
<p>IUPAC Code: NeuAc a2-3 Gal b1-4 GlcNAc b1-2 Man a1-3 [NeuAc a2-3 Gal b1-4 GlcNAc b1-2 Man a1-6) Man b1-4 GlcNAc b1-4 (Fuc a1-6) GlcNAc</p>		Sus scrofa (Pig)			
<p>Linear Code: NNA3a4GNB2Ma3(NNA3a4GNB2Ma5)Mb4GNB4(Fa5)GN</p>		Cercopithecus aethiops (Green Monkey)		[HSV -1]	Kidney cells
<p>Sub Structure Search Interface Load this structure for sub structure search</p>		Cricetus griseus (Chinese Hamster)			CHO cells
<p>General Information</p>		Human	Colon	Maldi MS only	
Glycan Family:	N-linked				
Sub. Family:	Complex				
Last Updated:	08/23/2004				
Oligosaccharide Molecular Wt.:	2370.15				
Calculated Oligosaccharide Molecular Wt.:	2370.15				
Per Methylated MW.:	2967				
Composition:	dHex ₁ NeuAc ₂ HexNAc ₄ Hex ₅				

HGPI 2006



[Jump to main content](#) [Jump to navigation](#)





[Register for email alerts](#)

Home

FG Update

Consortium for Functional Glycomics (CFG)

CFG Resources

CFG Data

CFG Databases

Links

- Database Resources
- Glycomics Related Websites
- NIGMS Glue Grants
- About us
- Sitemap

Toolbox

- E-alert signup
- RSS

NPG resources

- Nature
- Nature Methods
- Nature Chemical Biology
- Nature Immunology
- Nature Microbiology
- Nature Medicine
- Nature EMBO Journal
- Signaling Gateway

DATABASE RESOURCES

Carbohydrate databases

- KEGG-Glycan KEGG
- GLYCOSCIENCES.de DKFZ-Heidelberg
- CFG-Glycan Database
- GlycoSuite Proteome Systems

Glycan topology description

- GLYDE Glycan data exchange standard
- LINUCS Linear encoding of sugars
- CaboM Carbohydrate sequence markup Language
- KeqDraw
- LiGraph Graphical representation

Glycan MS tools

- GlycoMod Glycan composition from molbreak
- Glycofragment Masses from glycan fragments
- GlycoSearchMS MS-spectra comparison
- GlycosidIO Glycofragment mass fingerprinting

Glycan NMR tools

- CASPER 1H,13C-NMR estimation
- SugarBase 1H,13C-NMR search
- NMR-Search 1H,13C-NMR search

Glycan 3D modelling

- GLYCOSCIENCES modelling toolbox
- GLYCAM Generation of 3D structures

Glycans in PDBs

- PDB - repository DB of all 3D bio-structures
- dbPTM: DB of protein post-translational modification
- Glycoconjugate DB PDB glycan structures
- PDB2LINUCS Glycoproteins in PDB
- DOUGAL Database of glycoprotein structures

Glycan Processing Pathways & Enzymes

- Glycosylation pathways carbohydrate metabolism
- GGDB human glycoenes
- KEGG orthology glycosyltransferases
- CAZY carbohydrate active enzymes
- KEGG pathway carbohydrate metabolism
- GPI anchor biosynthesis
- SphingOMAP pathways of sphingolipid biosynthesis

Glycan Binding Proteins

- Lectines 3D structure of lectins
- CFG-GBP-DB Glycan binding protein
- GlycoEpitope DB Antibodies for glyco epitopes
- CFG- Microarray Data Glycogene chip

Prediction of glycosylation sites

- NetGlyc N-glycosylation
- NetOGlyc O-glycosylation
- YnOYans Glyco- phosphorylation
- big-PIPredictor GPI-anchor prediction
- DGPI GPI-anchor prediction
- Structural Assessment of Glycosylation (SAGS)

Nomenclature

- IUPAC Nomenclature of Carbohydrates

Acknowledgments

Steering Committee Scientific Cores Advisory Committee

Paul Crocker
Richard Cummings
Anne Dell
Kurt Drickamer
Pamela Marino
Jamey Marth
James Paulson
Steve Rosen
Ram Sasisekharan
Ron Schnaar
Pamela Stanley

Ola Blixt
Steve Head
Stuart Haslam
Bo Ma
Sally Orr
Rahul Raman
David Smith

Participating Investigators

~300

Elizabeth Neufeld
Ellis Reinherz
Nathan Sharon
Sponsors:
NIGMS
Neose Technologies
Kyowa Hakko
Otsuka Pharmaceuticals
NRC Ottawa