

Functional annotation of sugar databases

Frederique Lisacek¹, Christine Hoogland¹, Elaine Mullen³,
Lynette Hirschman², Nicolle Packer⁴

¹ *Proteome Informatics Group, Swiss Institute of
Bioinformatics, Geneva, Switzerland*

² *The MITRE Corporation, Mass., USA*

³ *Biomolecular Frontiers Research Centre, Faculty of Science,
Macquarie University, Sydney, Australia*

GlycoSuiteDB

- N- and O-linked linear and branched glycan structures
 - linkage, anomeric configuration, mass and composition
- biological source information
 - native and recombinant sources
 - tissue/cell type, cell line, strain and disease state
 - protein name (UniProtKB)
 - site of attachment
- methods used
- literature references (PubMed)



you are here: ExPASy CH

The ExPASy (**Expert Protein Analysis System**) [prote](#) protein sequences and structures as well as 2-D PA

Databases

[UniProtKB](#), [PROSITE](#), [HAMAP](#), [SwissVar](#),
[ViralZone](#), [SWISS-MODEL Repository](#), [SWISS-2DPAGE](#), [World-2DPAGE Repository](#),
[MIAPEGelDB](#), [ENZYME](#), [GlycoSuiteDB](#),
[UniPathway](#)

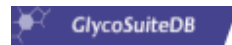
[\[details\]](#) [\[full list\]](#)

Education & services

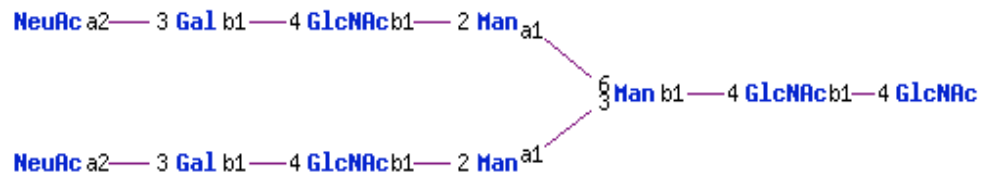
[Downloads](#), [Protein Spotlight](#),
[Protéines à la «Une»](#), [e-proxemis](#),
[Bioinformatics core facility for Proteomics](#)

[\[full list\]](#)

GlycoSuiteDB entry



entry: 1187-1149

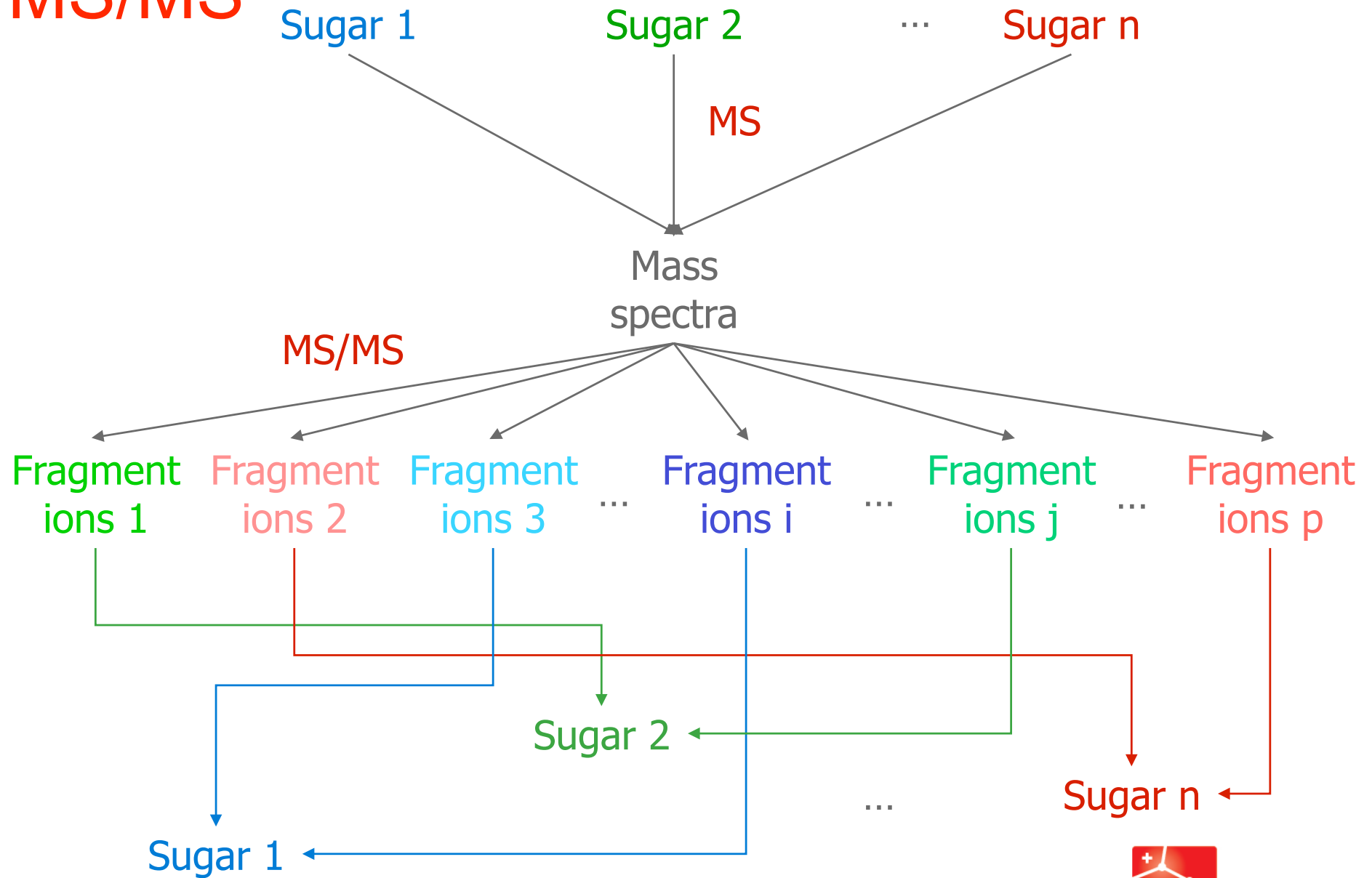


Species	<i>Homo sapiens</i> (HUMAN)
Class	MAMMALIA
Source	UROGENITAL SYSTEM, EXCRETION, URINE
Source notes	BIOLOGICALLY ACTIVE URINARY hCG PROVIDED BY A NUMBER OF INSTITUTES.
Attached to	CHORIOGONADOTROPIN BETA CHAIN (swiss-prot entry P01233); amino-acid ASN-33
Linkage	N-LINKED
Glycosylation sites	N-33, N-50, S-141, S-147, S-152 & S-158 [MORGAN ET AL. (1975) J. BIOL. CHEM. 250:5247-5258].
Identified by methods	PROTON NMR
References	Weisshaar (1991) Glycobiology 1: 393-404
Glycan structure	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)GlcNAc
Mass	2222.7830 Da (<i>monois</i>), 2224.0232 Da (<i>avg</i>), total residues: 11
Composition	Hex ₅ HexNAc ₄ NeuAc ₂
Release date	04-NOV-00 (<i>last updated</i> 04-DEC-00)

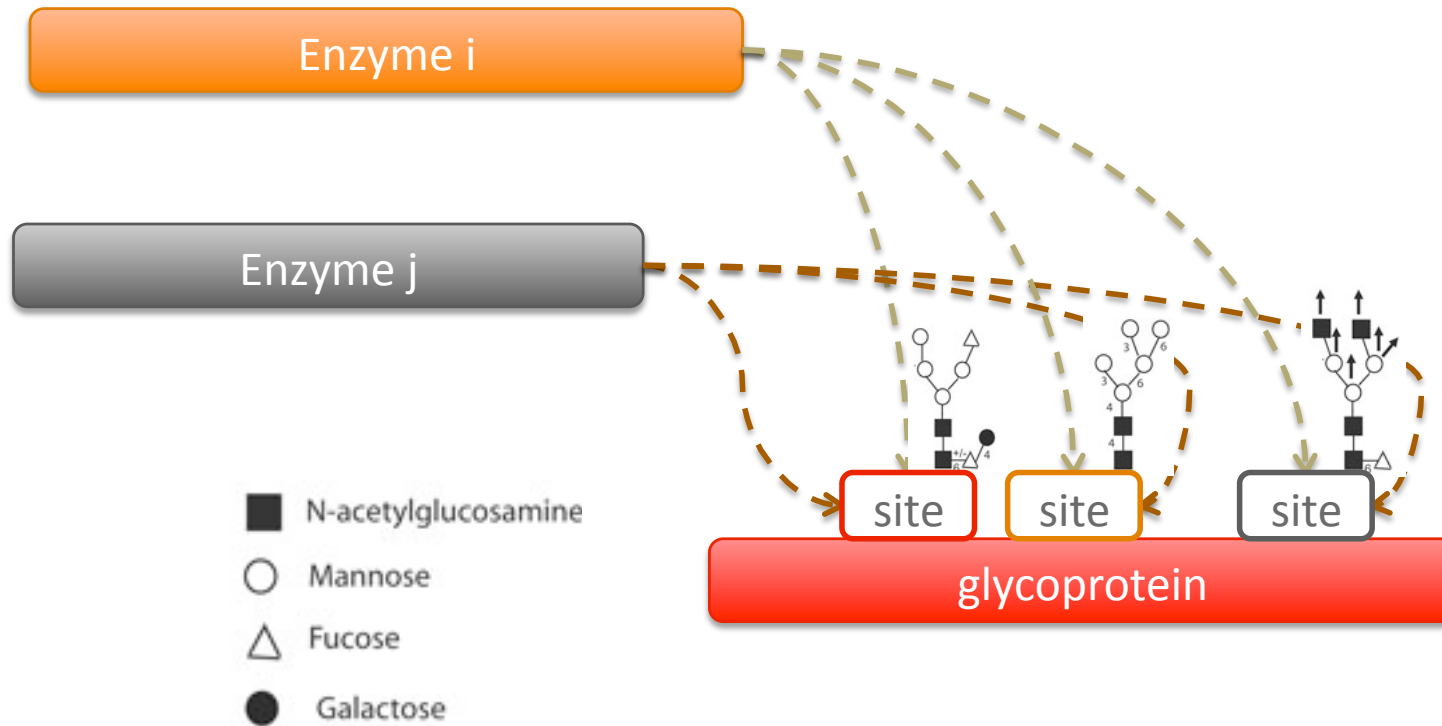
Basis of sugar annotation

- Methods for glycan structure determination: 57 listed in GlycoSuite DB, e.g., NMR, methylation, MS, HPLC, etc.
 - Scalable reliability
 - Potential for evidence tags
- Further insight with MS/MS
- Possible scenarios of synthesis:
 - Upon knowledge of enzymes (in genomes, in expression data)
 - Upon more or less established branching rules

MS/MS



Enzymatic constraints



Recent addition on ExPASy



SugarBindDB
Pathogen Sugar-Binding Database

Search Database

About us

Contact us

About This Database

The Pathogen Sugar-Binding Database was developed under MITRE internal funding, as part of a project to investigate use of floating glycoprotein films to capture pathogens. This database provides a list of known carbohydrate sequences to which pathogenic organisms specifically adhere. The data were compiled through an exhaustive search of literature published over the past 30 years by glycobiologists, microbiologists, and medical histologists.

The database allows users to search for bacteria, toxins, and viruses that bind to a particular sequence of sugars at the non-reducing terminus of an oligosaccharide. Alternatively, one can learn which glycans are ligands of adhesins (lectins) expressed by a particular species. Abstracts of primary literature are easily accessed, and a list of citations can be printed from search results.

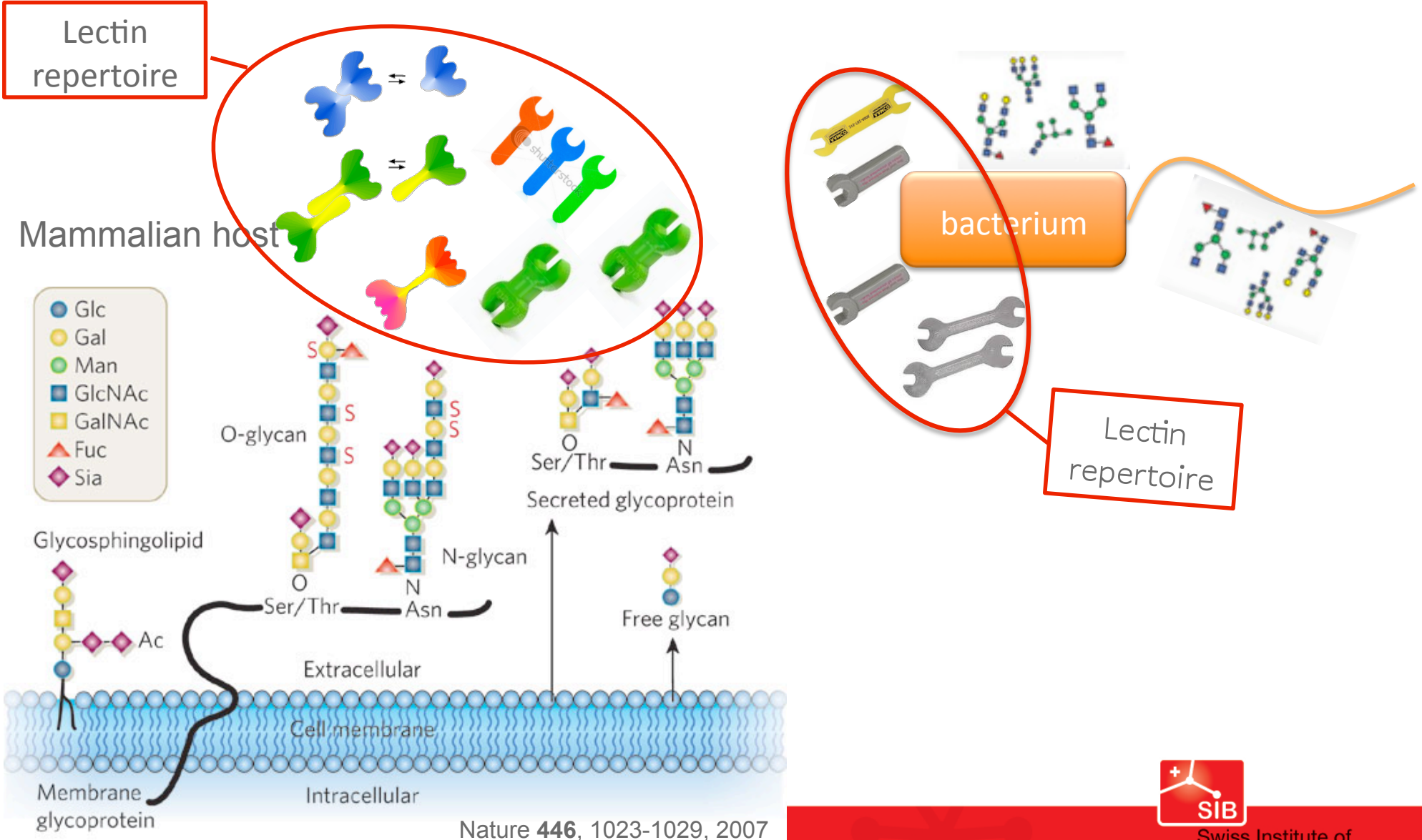
MITRE

Copyright © 2003-2010, The MITRE Corporation. All rights reserved.
MITRE is a registered trademark of The MITRE Corporation.
SugarBind™ and SugarBindDB™ are trademarks of The MITRE Corporation.
Material on this site may be copied and distributed as long as the
MITRE SugarBindDB is acknowledged.

20100611



Host-pathogen interactions via sugars



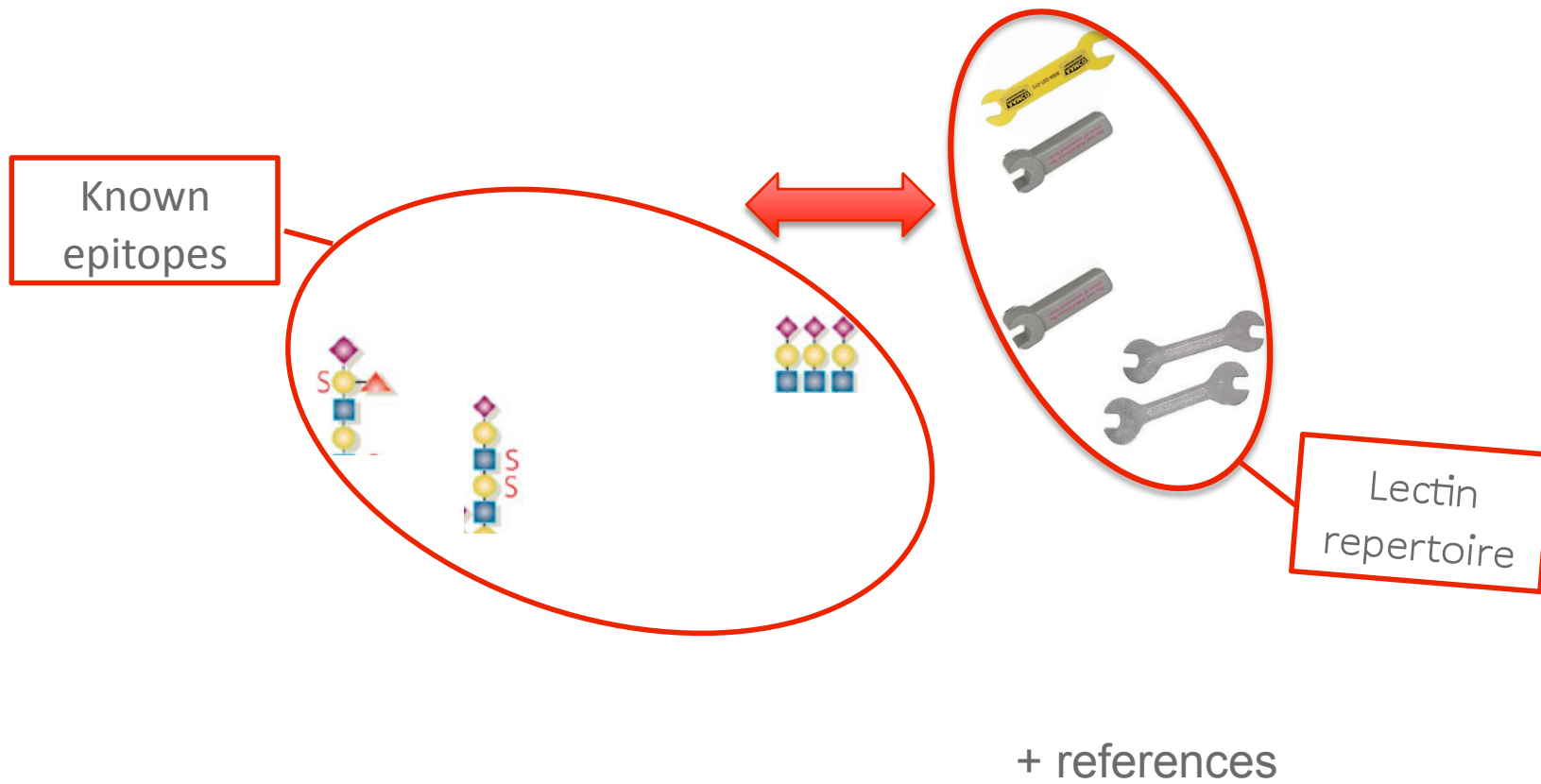
Nature 446, 1023-1029, 2007

Biocuration 2010














Swiss Institute of Bioinformatics

SugarBind



Example

Pathogen or Toxin	Fimbria / Pilus	Lectin / Adhesin	Gene Name	 Carbohydrate or Ligand	Pub Year		Citation
Pseudomonas aeruginosa		PA-IIL	lecB	Fuc	2004	<input type="checkbox"/>	Microbes and Infection, 2004 Feb; vol. 6 no. 2: 221-8 
Pseudomonas aeruginosa		PA-IIL	lecB	Fuc(a1-4)[Gal(b1-3)]GlcNAc(b1-3)Gal(b1-4)Glc	2004	<input type="checkbox"/>	Microbes and Infection, 2004 Feb; vol. 6 no. 2: 221-8 
Pseudomonas aeruginosa		PA-IIL	lecB	Fuc(a1-4) [Neu(a2-3)Gal(b1-3)]GlcNAc(b1-3)Gal(b1-4)Glc	2004	<input type="checkbox"/>	Microbes and Infection, 2004 Feb; vol. 6 no. 2: 221-8 
Pseudomonas aeruginosa		PA-IIL	lecB	Gal(b1-3)[Fuc(a1-4)]GlcNAc(b1-3)Gal(b1-4)Glc	2004	<input type="checkbox"/>	Microbes and Infection, 2004 Feb; vol. 6 no. 2: 221-8 
Pseudomonas aeruginosa		PA-IIL	lecB	NeuAc(a2-3)Gal(b1-3) [Fuc(a1-4)]GlcNAc(b1-3)Gal(b1-4)Glc	2004	<input type="checkbox"/>	Microbes and Infection, 2004 Feb; vol. 6 no. 2: 221-8 
Pseudomonas aeruginosa		PA-IL	lecA	Gal	2004	<input type="checkbox"/>	Microbes and Infection, 2004 Feb; vol. 6 no. 2: 221-8 
Pseudomonas aeruginosa		PA-IL	lecA	Gal(a1-3)Gal	2004	<input type="checkbox"/>	Microbes and Infection, 2004 Feb; vol. 6 no. 2: 221-8 
Pseudomonas aeruginosa		PA-IL	lecA	Gal(a1-4)Gal	2004	<input type="checkbox"/>	Microbes and Infection, 2004 Feb; vol. 6 no. 2: 221-8 
Pseudomonas aeruginosa		PA-IL	lecA	Gal(a1-6)Gal	2004	<input type="checkbox"/>	Microbes and Infection, 2004 Feb; vol. 6 no. 2: 221-8 

Open consortium

- ***N. Packer et al***, Biomolecular Frontiers Research Centre of Macquarie University, Sydney, Australia
- ***P. Rudd et al***, Dublin-Oxford Glycobiology Lab of the Conway Institute in Dublin, Ireland
- ***N. Karlsson et al***, Medical Biochemistry Dpt of University of Gothenburg in Sweden
- ***E. Mullen et al***, MITRE Corporation in Massachusetts, USA
- The more the merrier ...

Roadmap (1)

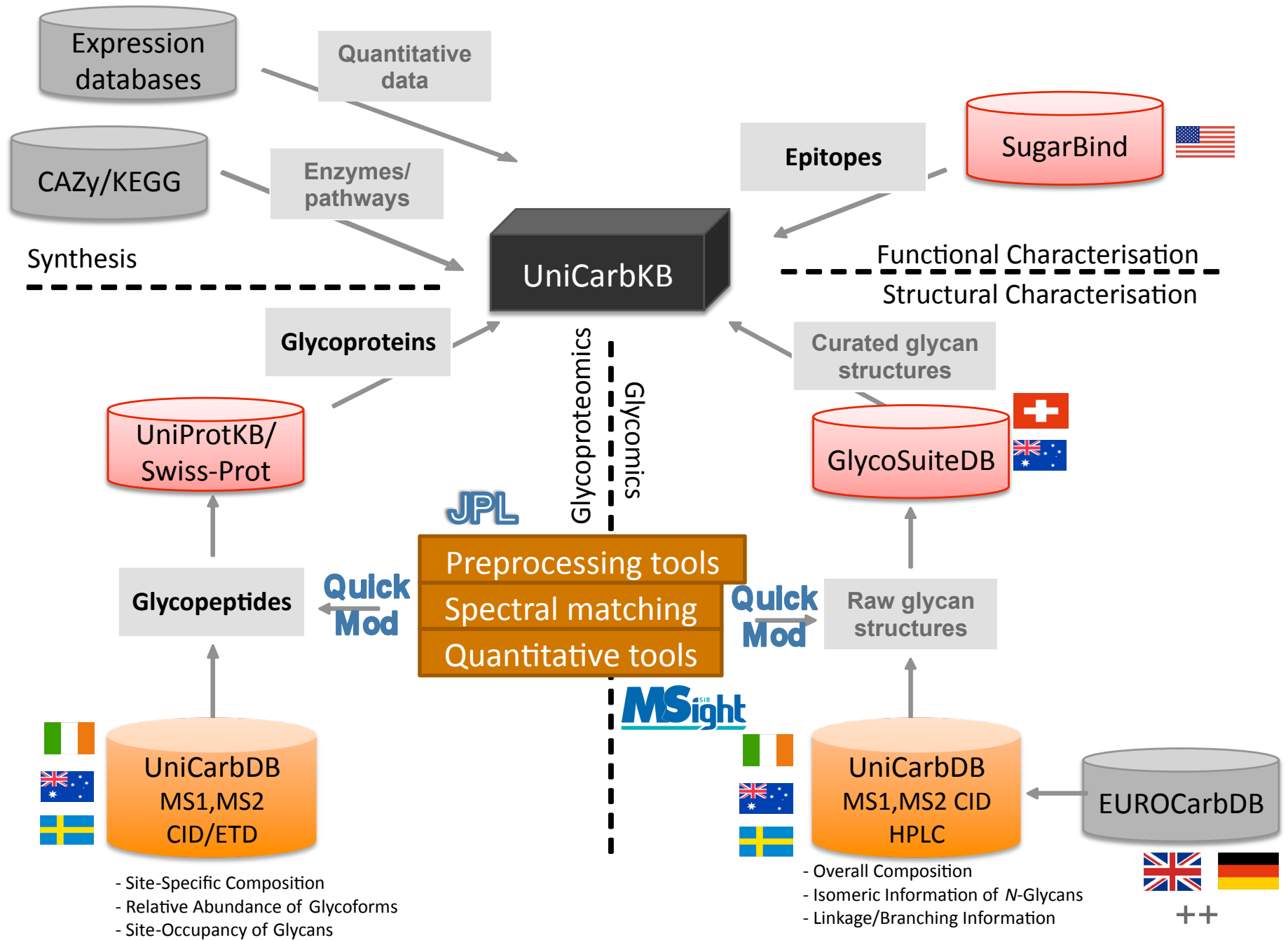
- Step 1: Define the basis of extensive linking of glycomics data (*standards* need refinement and reinforcement)
- Step 2: Link data production to structural determination (*manual curation* of MS1/MS2 data is necessary for the accurate determination of structural data)
- Step 3: Link sugar structures with protein attachment site (*existing databases* play central role for collecting data on structures released from specific proteins/tissues)

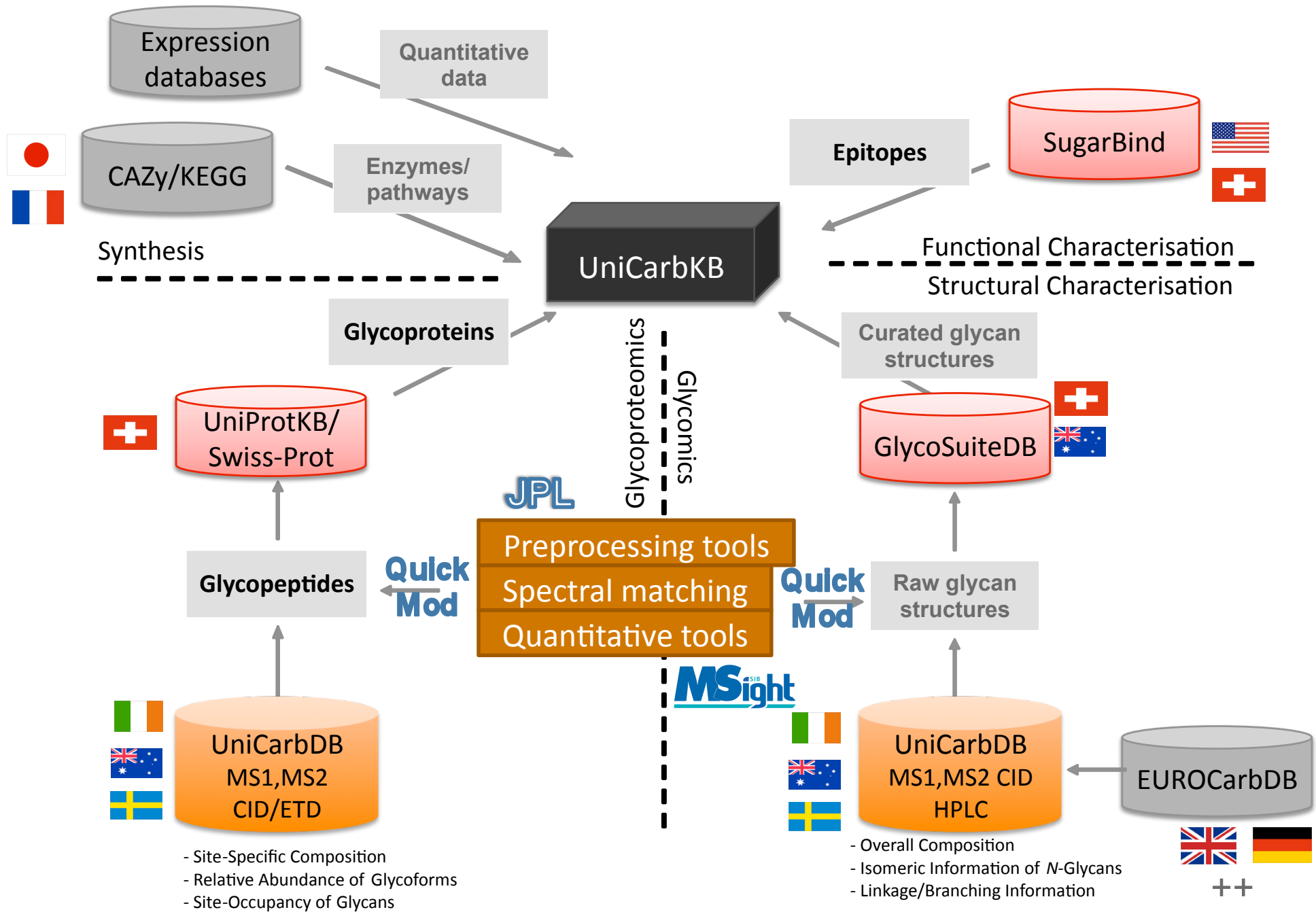
Roadmap (2)

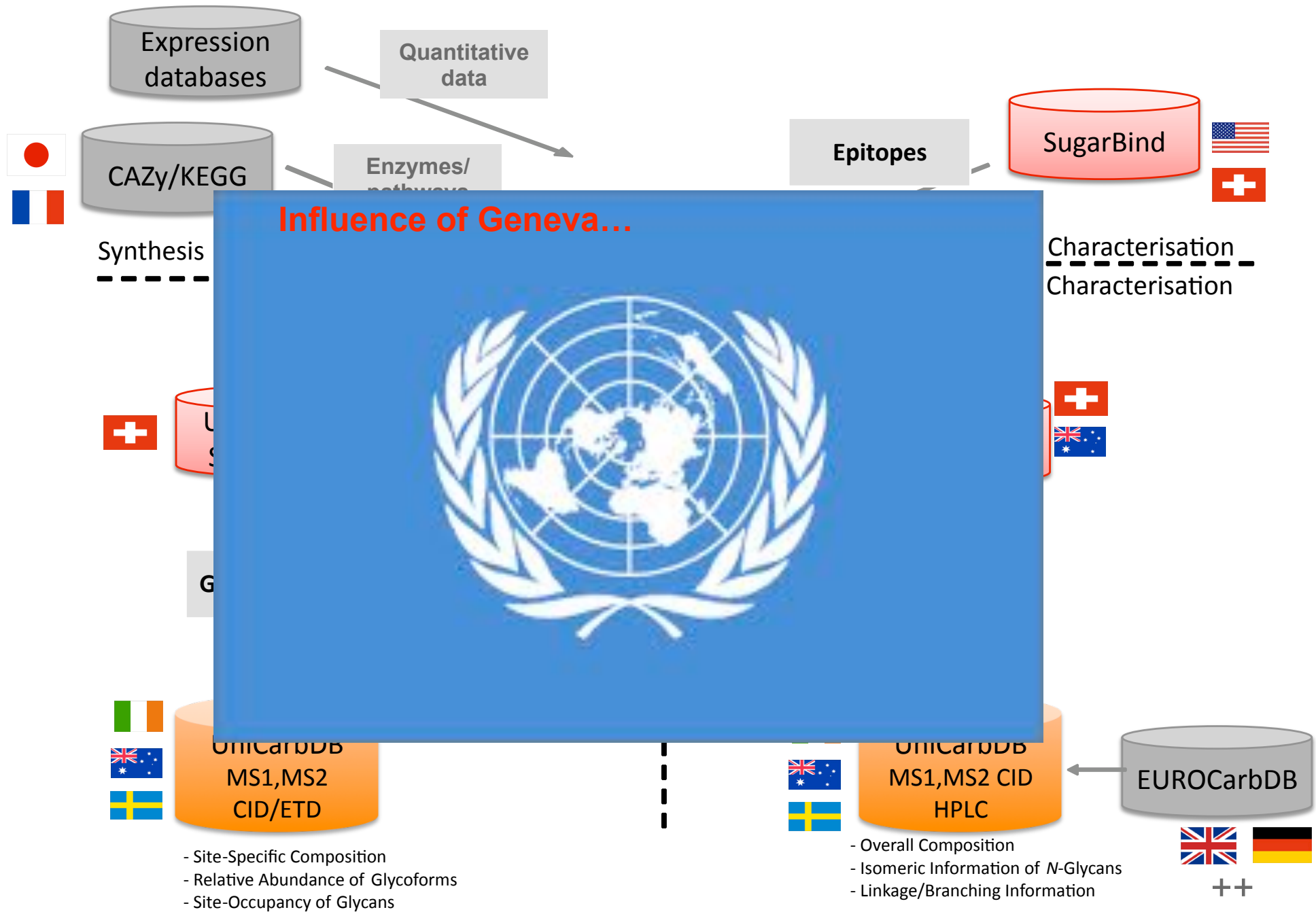
- Step 4: Link mass spectrometry data to the expression of glycosyltransferases or glycosidases (***reverse proteomics***)
- Step 5: Link glycosylation information to cellular function (understanding of ***functional role*** of sugars)

[strong human input, “manual” data curation]

In brief: knowledge of structure/function of glycosylation of proteins to be determined in much the same way as knowledge structure/function of proteins integrated into UniProtKB







Immediate concerns

- **Challenge 1:** Strengthen functional annotation in host-pathogen interactions by linking epitopes to full sugar structures on glycolipids and glycoproteins
- **Challenge 2:** Automate MS data processing for feeding MS and MS/MS into databases
- **Challenge 3:** Make sugar analysis easier using bioinformatic tools!

Acknowledgements

- SIB colleagues
- PIG members
- Consortium members



Thank you!

