"Encoded proteins carry out most biological functions, and to understand how cells work, one must study what proteins are present, how they interact with each other and what they do".

PROTEOMICS

- What proteins are expressed and when?
- What amounts?
- How are they modified -phosphorylation, glycosylation, prenylation etc?
- How do they interact with each other?
- How do levels/types change during differentiation/disease etc?

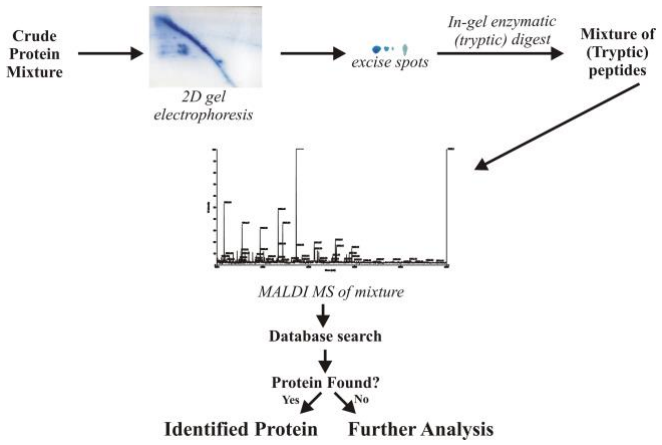
Two-dimensional gel electrophoresis (2DGE)
A widely used separation technology for proteomics
Start complex mixture of proteins, extract all and then separate them.

Why Gels?

- Quick
- Relatively Inexpensive
- Good resolving technique for complex mixtures
- Works with crude samples
- Moderately tolerant to salts and detergents
- Good visual representation of entire sample
- Easy comparison
- Relatively reproducible

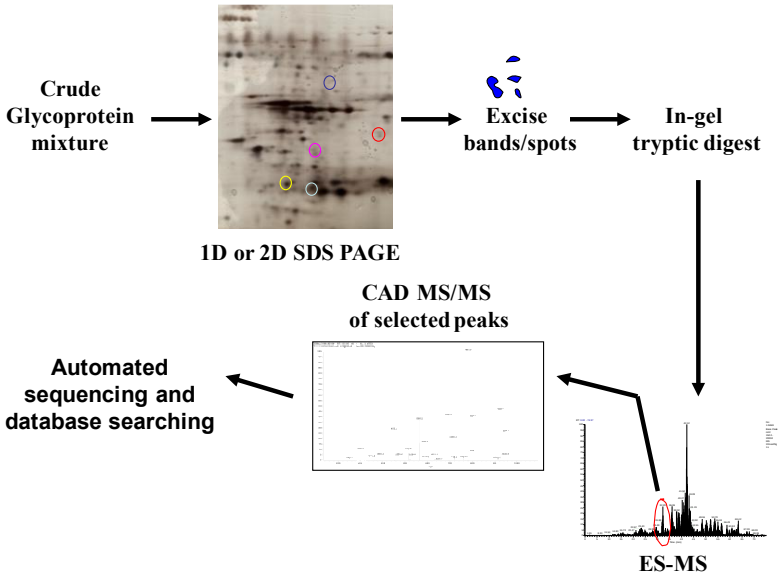
Detection in 2DGE
Coomassie blue, silver stain, radiolabelling, fluorescent stains…
MASS SPECTROMETRY - rigorous identification
2D gel you get info about charge and size but that is it (limited information)

PROTEOMICS: "Mapping" or "Fingerprinting" Strategy
To really answer those questions, get a better understanding of what those proteins are--> need to identify proteins → mass spec
Trypsin digest proteins into smaller tryptic peptides. Then run those on mass spec (typically MALDI or electrospray). Get a series of peptide molecular ions. Match theoretical tryptic peptides (from genome reference) to the observed peptides.
Limitations:
Need to have genome sequence

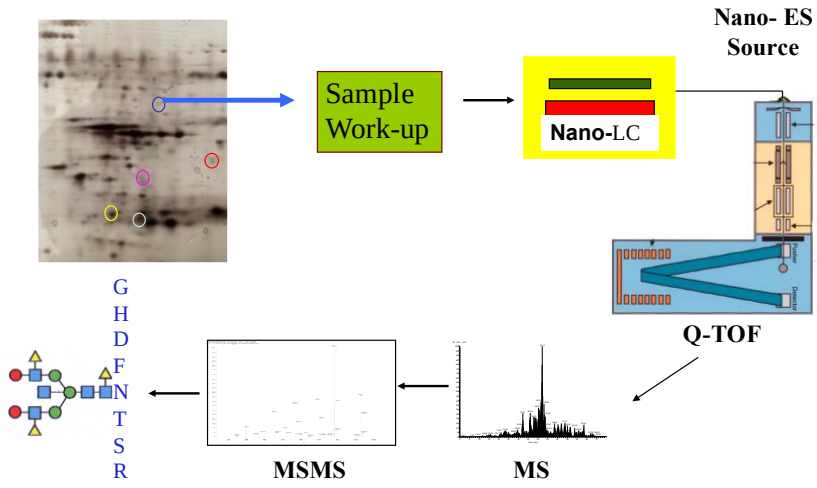


Get amino acid sequence of peptides → give a lot of info (bioinformatics searching).

PROTEOMICS: Basic sequencing strategy; high throughput difficult

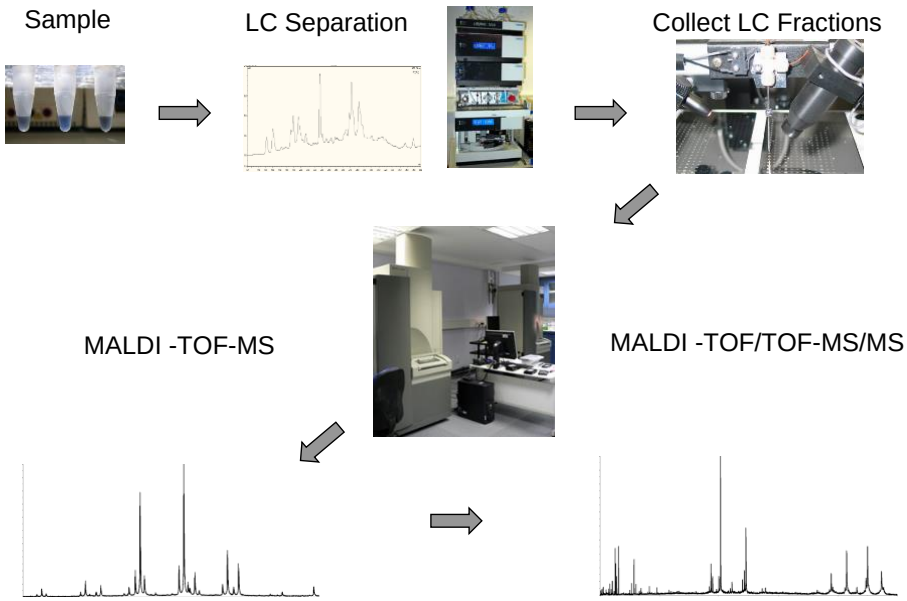


PROTEOMICS: greater automation; example of typical workflow using NanoLC-ES-MS/MS; PTMs can also be identified but MS/MS data might require manual interpretation



With electrospray → big advantage is that you can link that to liquid chromatography as well and therefore you can utilize resolving power of liquid chromatography (does not matter how complex initial sample is → separate it an LC room)

Peptide analysis by MALDI-TOF/TOF



It Is not just electrospray to use chromatography. Also with MALDI you can but not directly.

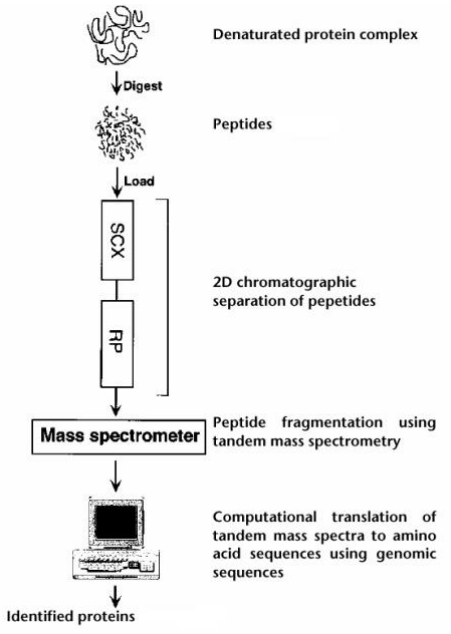
Shotgun Proteomics

- 2D-SDS-PAGE biased against-
- Low abundance proteins
- Integral membrane proteins
- Proteins with extremes of pI
- Proteins with extremes of MW
- Shotgun proteomics attempts to analyze all proteins in system

SDS page is not a very good technique with extremes of physical and chemical properties (not really large or small proteins/very acidic/very basic/ not good for membrane proteins etc..) . Here not separate at the protein level but at the peptide level.

MudPIT

- Multidimensional Protein Identification Technology
- Proteins extracted and digested
- 2D-online-LC separation
- ES-MS and MS/MS of resolved peptides
- Proteins identified by comparison to databases



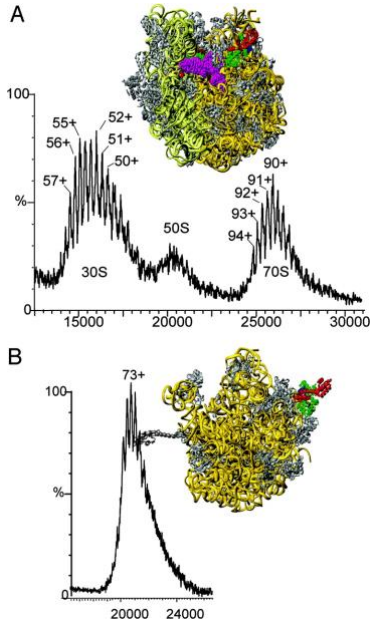
- Applied to yeast *Saccharomyces cerevisiae*
- 2D-PAGE identified 279 proteins
- MudPIT identified 1,484 proteins
- Including proteins of pI 3.82 and 12.55,

MW 559 kDa and membrane associated Proteins
“Micro-MudPIT” 4984 proteins were detected, 2/3 predicated reading frames of yeast genome

The more peptide mapped to protein (more percentage coverage of protein), the more confident that the protein is present. Remember to be critical about confidence level of data.
This is bottom-up proteomics

Mass spectra of 70S and 50S subunits of ribosomes from Thermus thermophilus from A ES-Q-ToF mass spectrometer

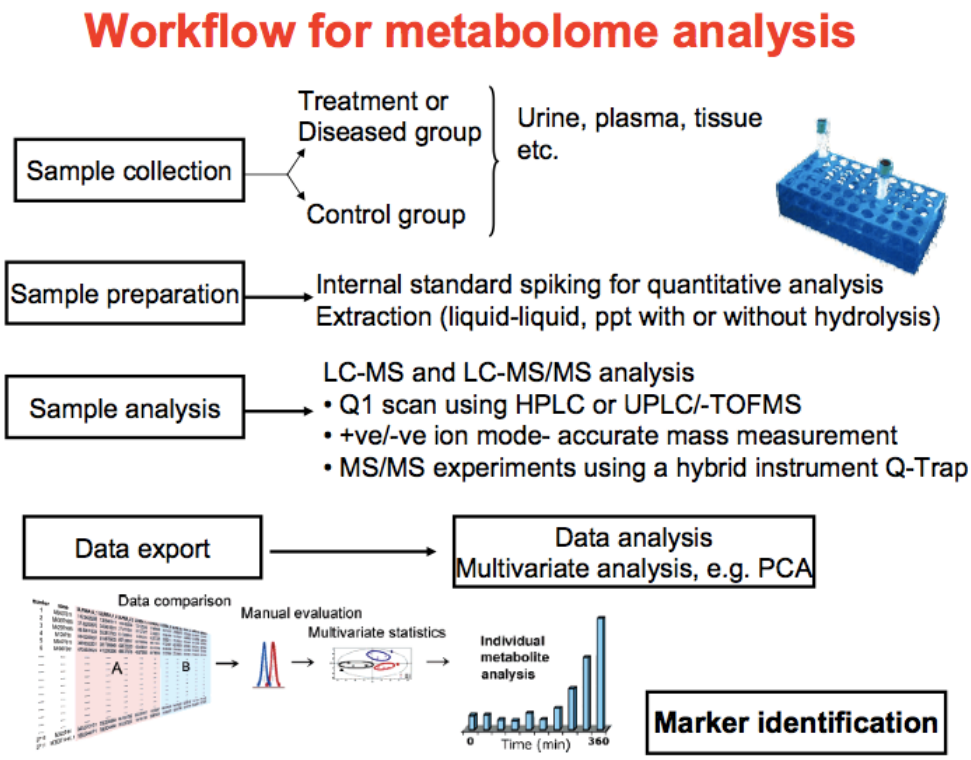
Top-down proteomics:
Don’t look at the peptides but analyse intact proteins. It is possible to look at very large molecules but machinery is more specialized.



Mass spectra of 70S and 50S subunits of ribosomes from Thermus thermophilus. (A) The MS of the 70S was recorded with a 100-V offset on the collision cell and reveals well resolved charge states for the 30S and 70S subunits. The 50S subunit was separated chromatographically, and a spectrum of the isolated subunit B was recorded. The structures of the 70S and 50S particles were produced from the Protein Data Bank coordinates 1GIX and 1GIY, respectively by using the program ribbons (34). The rRNA is shown in gold and yellow for the 50S and 30S subunits, respectively with proteins shown in gray with the exception of those released in mass spectra (see Fig. 2), L12 (red/green) and tRNA pink, red, and green in the A, P, and E sites, respectively.

What is Metabolomics?

Look at low molecular weight metabolites produced in cells (products of different several biochemical reactions occurring in the cell, products of several enzymes etc…). Therefore they are hugely information rich (a snapshot of what is occurring tin the cell at a certain time). Look at the chance in metabolism (cancer)



Very high resolution/mass accuracy can be important for metabolomics

FT-MS can be very useful for metabolomics because biological tissues and fluids have many thousands of low m/z metabolites that often have very similar m/z values (although lower performance is fine if suitable chromatographic methods are combined with MS)

Metabolite identification remains one of greatest technical challenges in metabolomics

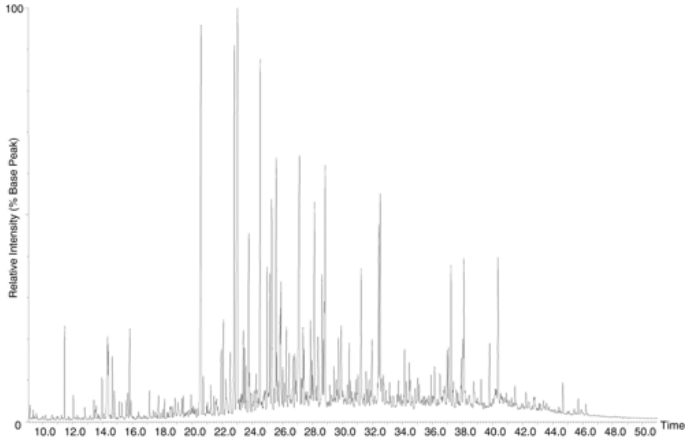
Number of possible empirical formulae ($C_cH_hN_nO_oP_pS_s$) that can be assigned to a peak at 400 Da.

Mass accuracy of spectrometer	No. of formulae
3 ppm (good MS)	23
1 ppm (excellent MS)	7
0.1 ppm (fantasy MS?)	1

Kind & Fiehn, BMC Bioinformatics 7:234 (2006)

GC-MS is also widely used for metabolomics

Quadrupole mass spec using electron impact is cheaper but still allows us to get good identification.

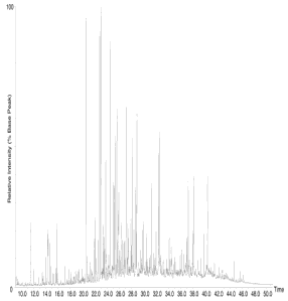


Total ion chromatogram of a human urine (1 mL) extract after silylation and GC-MS analysis. Resulted in 1,582 components detected.

MS-Imaging

Microscopy is very good to get spatial context. Can combine microscopy-mass spec.

MS-Data

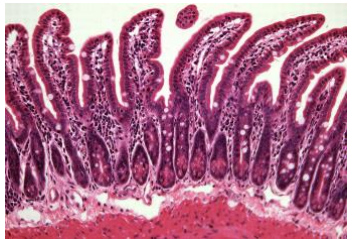


High sensitivity

Detailed molecular characterization

Lacks spatial context

Microscopy-Data



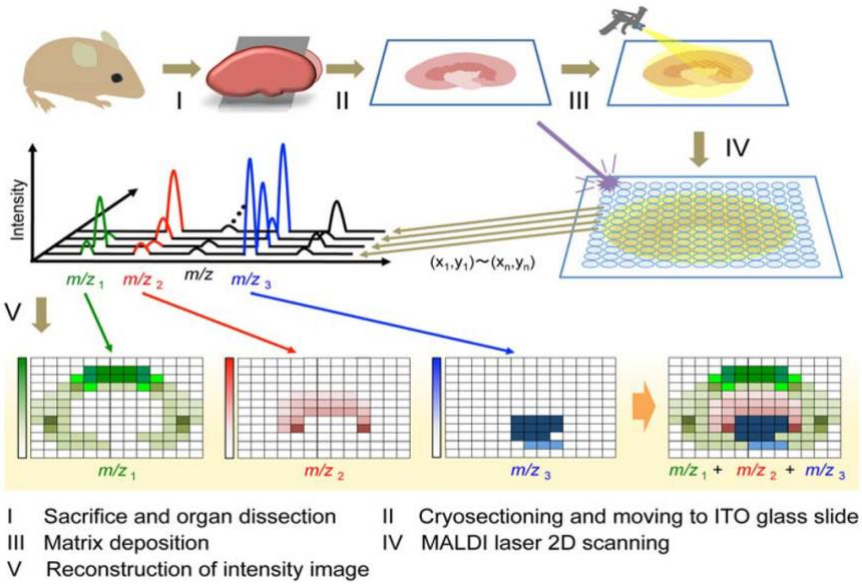
High spatial context

Immunomicroscopy can give some molecular context

Sensitivity ?

Can lack detailed molecular characterization

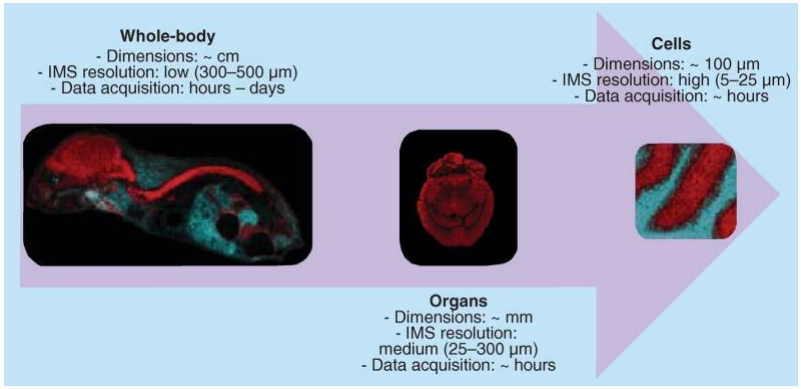
MALDI Imaging Mass Spectrometry



I Sacrifice and organ dissection
II Cryosectioning and moving to ITO glass slide
III Matrix deposition
IV MALDI laser 2D scanning
V Reconstruction of intensity image

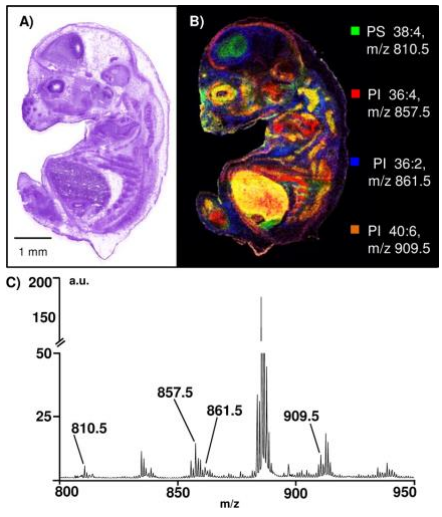
Combine mass spec with imaging techniques. Get molecular representations of molecules present in a slide. Superimpose maldi grid with image. Get several spots and physically related them to a specific area of tissue section.

MALDI Imaging Mass Spectrometry

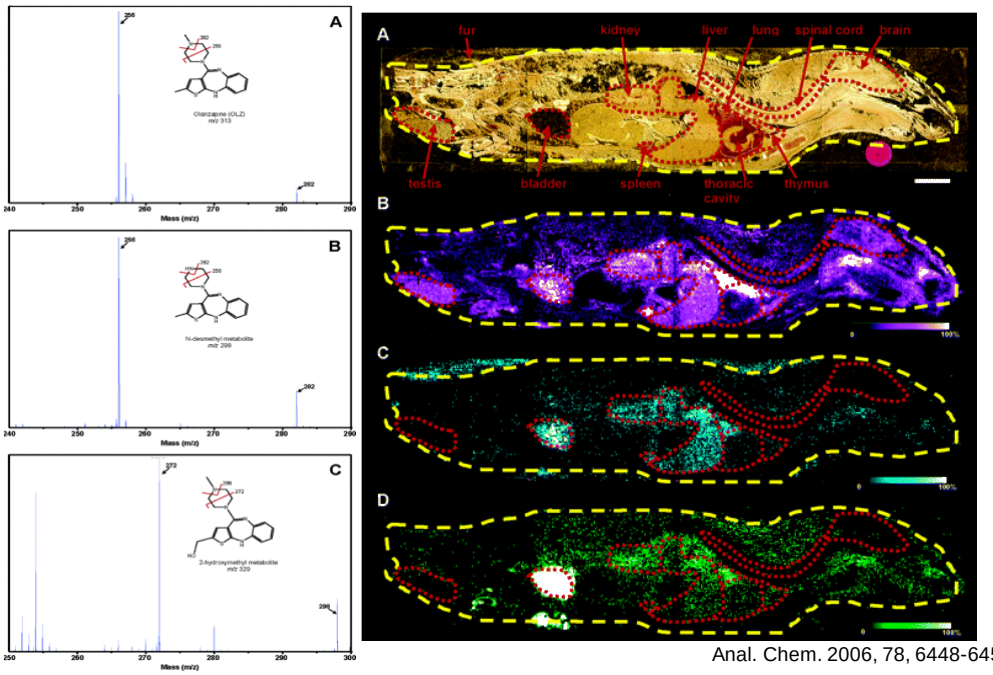


MALDI-IMS of lipids from mouse embryo

Combined ion image from four lipids from an imaging experiment on a mouse embryo at embryonic day 13.5. IMS data were collected at a 20 µm spatial resolution in negative ion mode.



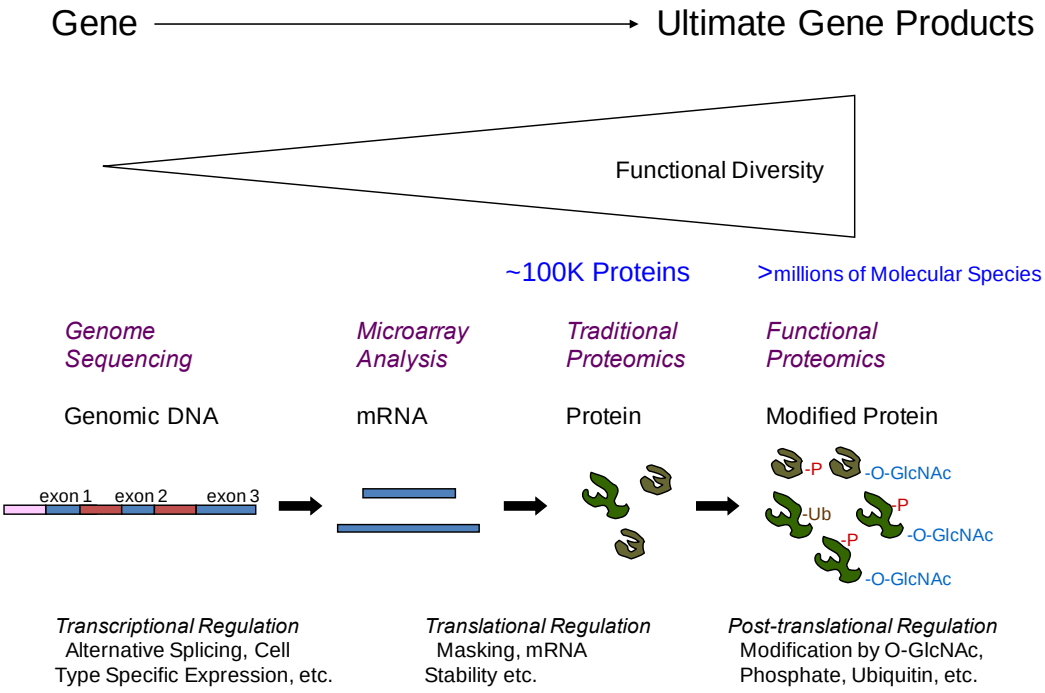
MALDI-IMS of Olanzapine and its metabolites 2h postdose



Anal. Chem. 2006, 78, 6448-6456

Get detailed used in pharma industry for pharmacokinetics. Treat schizophrenia bipolarism and other mental disorders.

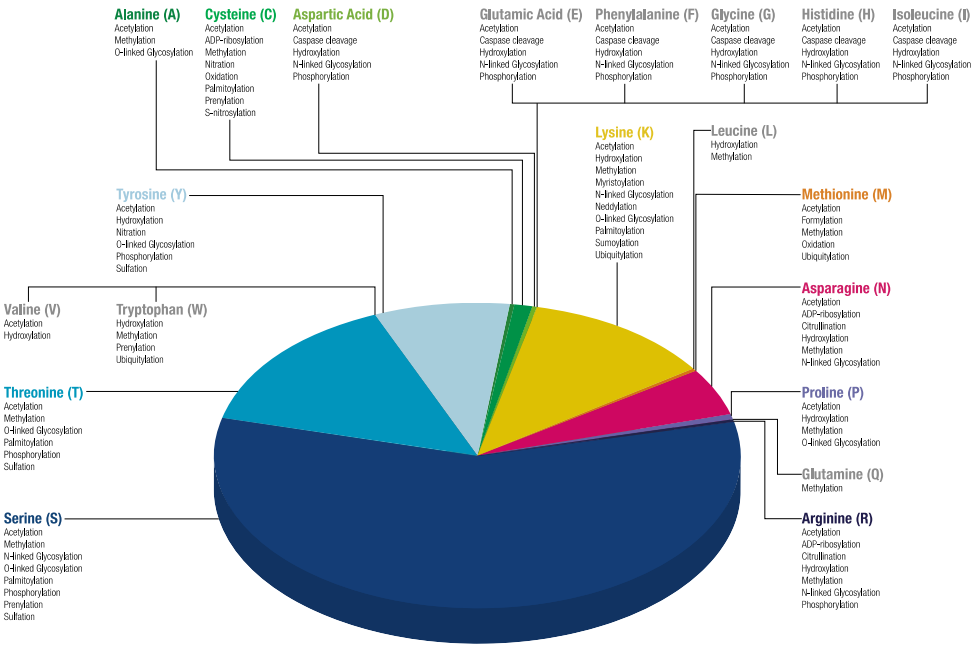
Genomics Does Not Explain Biology:
Biological complexity is not linearly related to the number of genes



Post-translational modifications

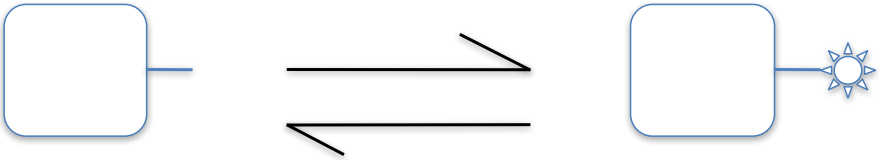
- PTMs are the chemical modification of a protein after its translation.
- They have profound effects on protein function by altering their activity state, localization, turnover, and interactions with other proteins.
- Over 200 different types of PTM, every amino acid can be modified.
- The majority of proteins are modified?

Distribution of Amino acid PTMs

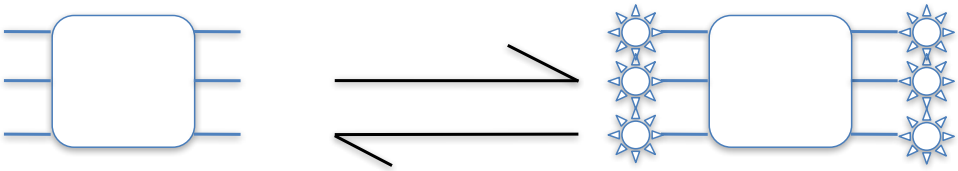


Mod-Forms of Proteins

Modifications allow for a lot more dynamic functional response. Human cells have limited amount of genes, if cell wants to adapt to changing environments → can change gene expression but that might not be enough. By keeping same protein scaffold but changing charge and other modifications → change functionality of proteins in a quick way.
PTM allow for simple on/off switch to gradual switch.



Number of Mod-Forms $2^1=2$



Number of Mod-Forms $2^6=64$

Serine/arginine repetitive matrix protein 2
Involved in pre-mRNA splicing
300 kDa, 2,752 AA
Over 300 phosphorylation sites
It is a heavily phosphorylated proteins. Not possible to have all the modification at the same time. It depends on the physiological state (proteins not a single entities but as the dynamic product of PTMs).

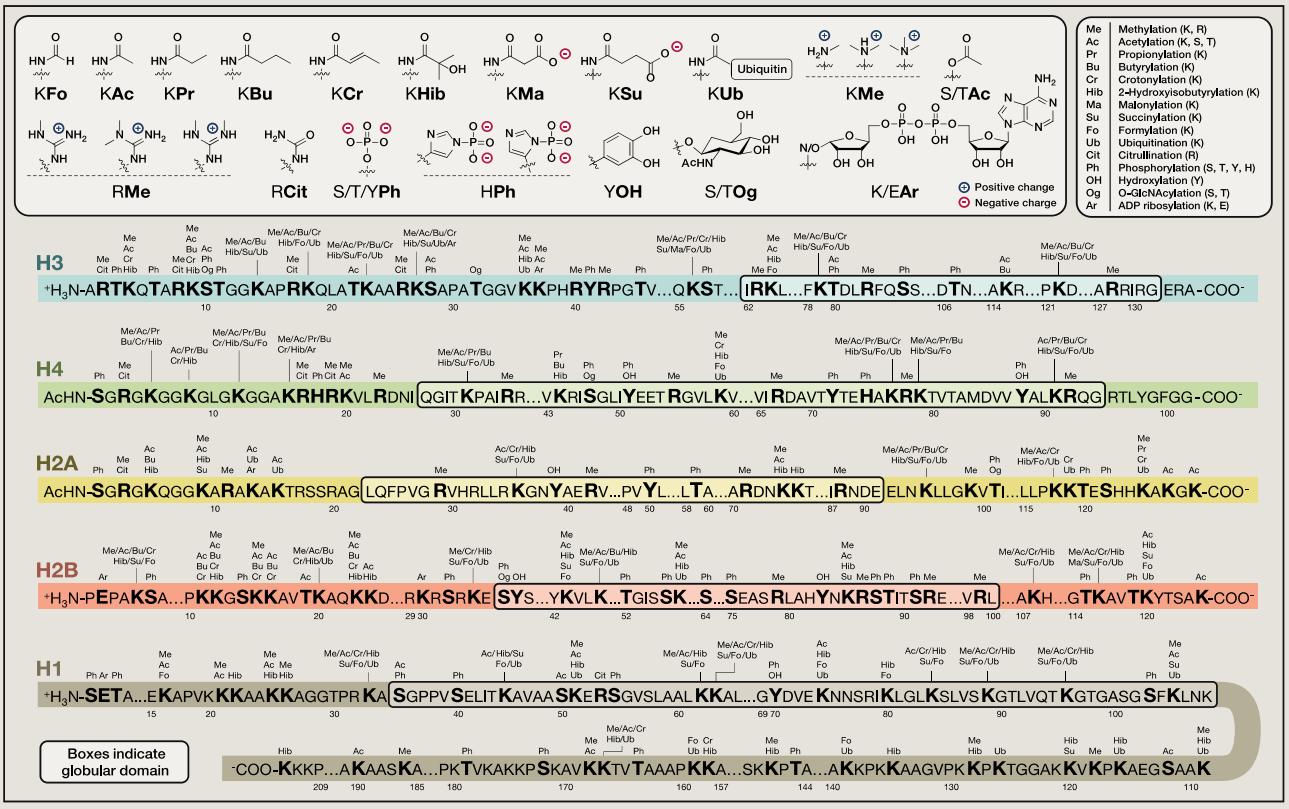
Histone code
Histones are some of the heaviest modificated protein known. There are hotspots.
Thousands of different modification forms of histones.

458 Cell 159, October 9, 2014 ©2014 Elsevier Inc. DOI http://dx.doi.org/10.1016/j.cell.2014.09.037 See online version for for legend and references.

SnapShot: Histone Modifications

He Huang,[1] Benjamin R. Sabari,[2] Benjamin A. Garcia,[3] C. David Allis,[2] and Yingming Zhao[1]
[1]Ben May Department of Cancer Research, The University of Chicago, Chicago, IL 60637, USA
[2]Laboratory of Chromatin Biology and Epigenetics, The Rockefeller University, New York, NY 10021, USA
[3]Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104, USA

Cell



Protein Phosphorylation-Kinases

- Human genome contains > 500 protein kinases, aprox 2% of all human genes
- Implicated in signal transduction, enzyme activity and regulation, transcription, cell cycle, apoptosis, differentiation etc……..
- Many kinase loci implicated in cancer and other major diseases

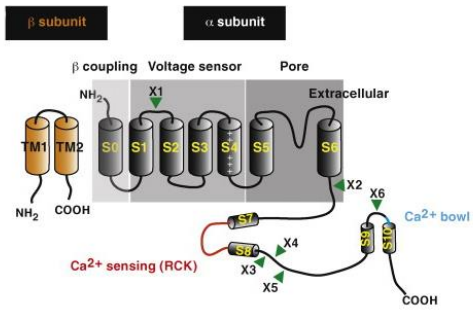
Protein Phosphorylation-Phosphatase

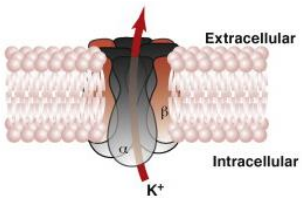
Human genome contains around 250 genes with a phosphatase domain

- Implicated in signal transduction, enzyme activity and regulation, transcription, cell cycle, apoptosis, differentiation etc……..
- >40% implicated in cancer

BK Channels

Affected by alcohol!!!
Expressed in neurons, sitting in the membrane. When opens allows calcium ions to go out from neuron (regulate release of neuropeptides).
When alcohol is assumed →alcohol affect functioning these channels. Lock the channel in an active form → open → lots of Ca exit from neurons. Constantly open → lose a lot of ions – reduction in neuronal excitement → shut down function of neurons → depression in central nervous system → loss of moto neuron and coordination. → then increase need to





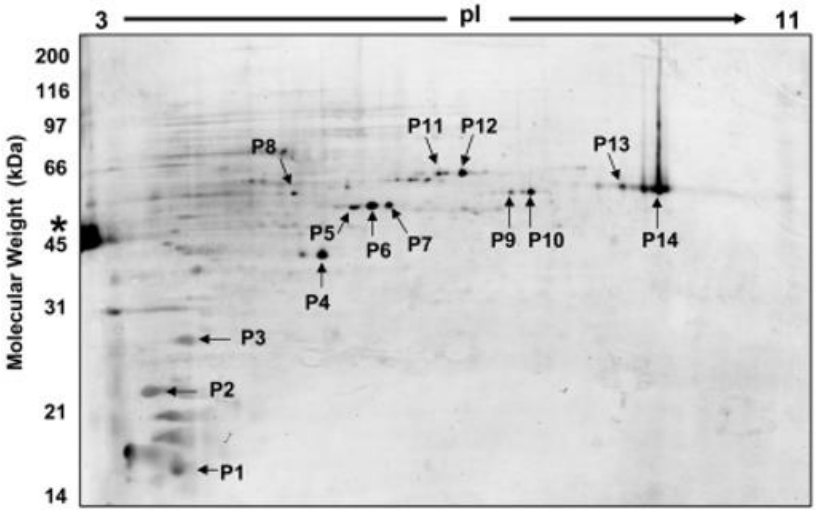
- Large Conductance Voltage- and Ca^{2+}-gated K^+ (BK) channel
- Involved in neuronal firing, neuropeptide release
- Pore forming a-subunits and modulatory b-subunits
- 4 a-subunits assemble to make a functional channel

- Alcohol activates, increasing K^+ current
- Causes reduced neuronal excitability and neuropeptide release
- Leads to depression of central neurons, motor incoordination and diuresis

- Phosphorylation of T107 by CaMKII kinase switches channel alcohol response from activation to inhibition
- Acts as a molecular dimmer switch
- Could mediate tolerance to alcohol
- Phosphorylation of b-subunits also has regulatory effects

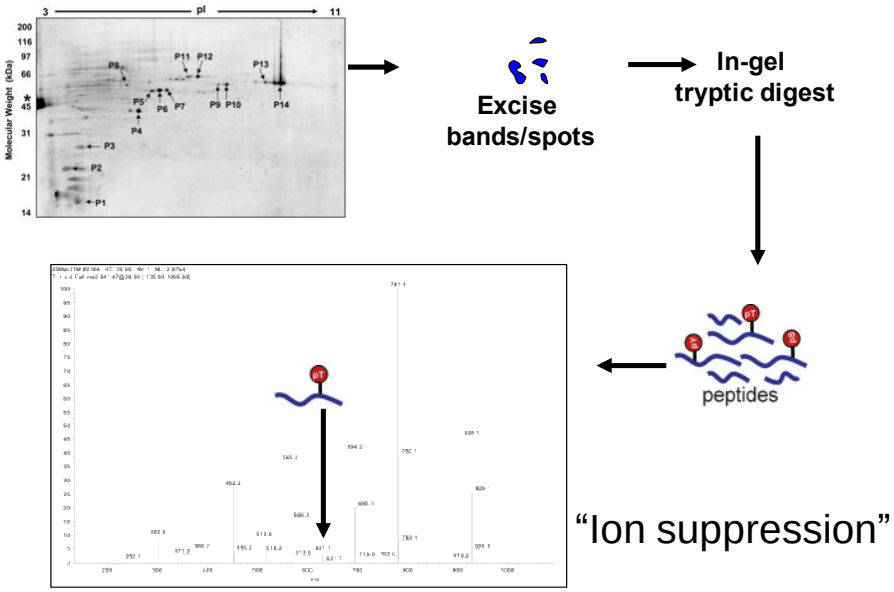
T107 is phosphorylation site in s0/s1 loop. Phosphorylation event changes the structure of the loop (encourage formation of alpha helix in the loop) –> change in protein structure induces change in activity of the ion channel. Ion channels doesn't stay permanently open → return to normal neurological state. Probably an evolutionary response to mammals such as cow who eat a lot of fermentative alcohol stuff.
In humans → alcohol abuse is dependency (tolerance). The more phosphorylation you have, the more alcohol you need to drink to feel the same effect.
Phosphorylation could be adaptation event. Look at the inhibitors of this molecule (stop alcohol adaptation).

Phosphoprotein Analysis



- Proteins separated by 1-D or 2-D SDS-PAGE
- Detection by phosphospecific stain, Pro-Q Diamond
- ^{32}P or ^{33}P labeling and autoradiography
- Western blotting with phospho-T, S, Y antibodies

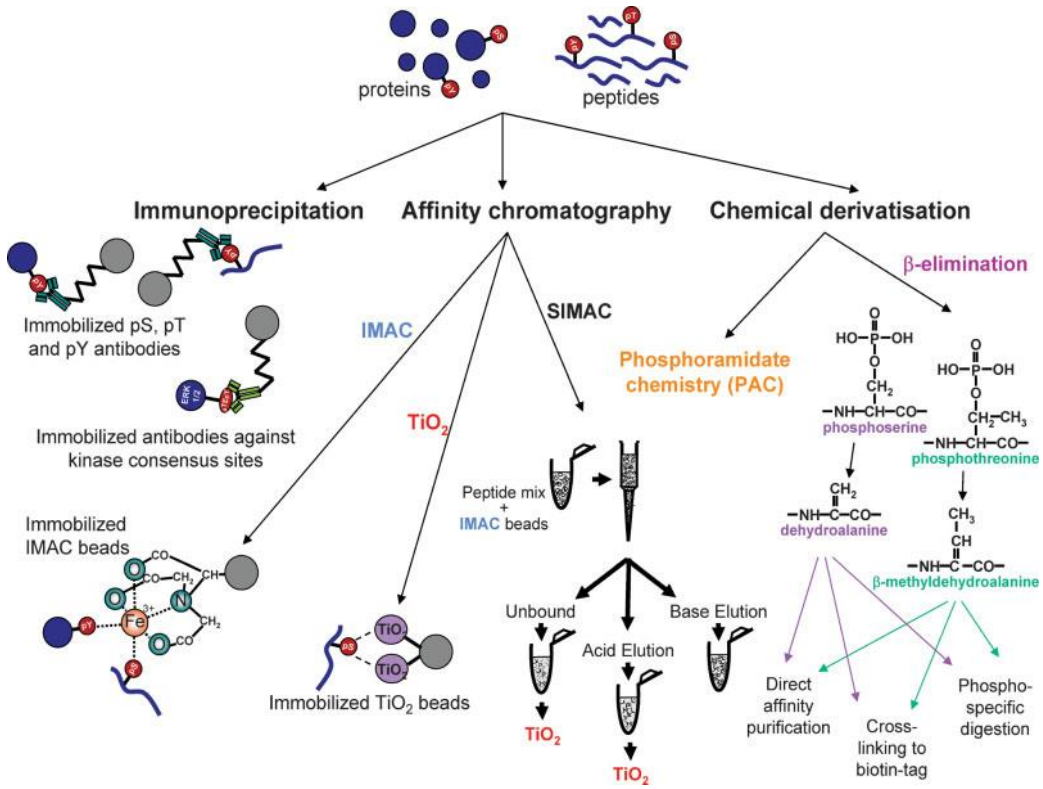
Phosphoproteomics



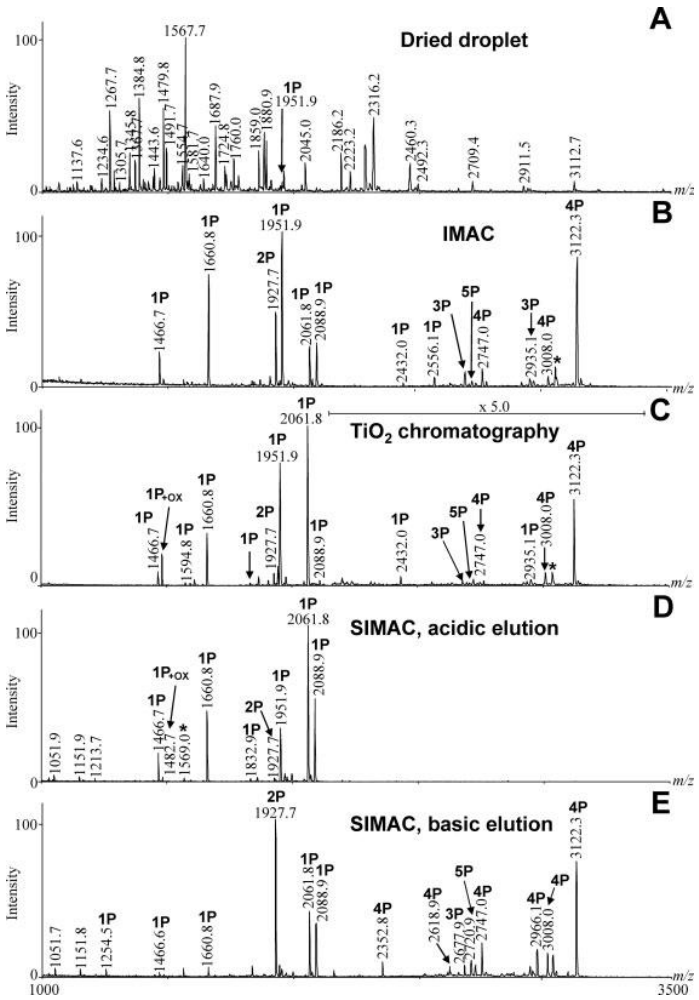
"Ion suppression"

Phosphoproteomics-enrichment

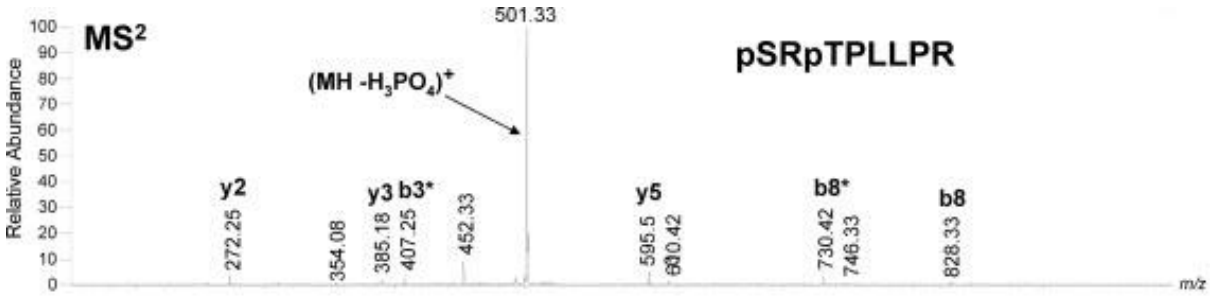
IMAC is charge-charge interaction, could get nonspecific binding. Can try to minimize interaction to lower pH. There is an improved resin → Titanium dioxide beads (form coordination bonds with phosphate groups, which are more specific than charge charge). For complex mixtures you can to combination → SIMAC (titanium + ion based charge-charge).



MALDI-MS analysis of 500 fmol Of peptide mixture obtained From Tryptic digestion of 12 proteins.



Phosphoproteomics-MS/MS
In a lot of situations, too much NG → phosphate groups can fall off ->> cannot see where they are attached to peptide. Careful on how much energy you use to fragment.

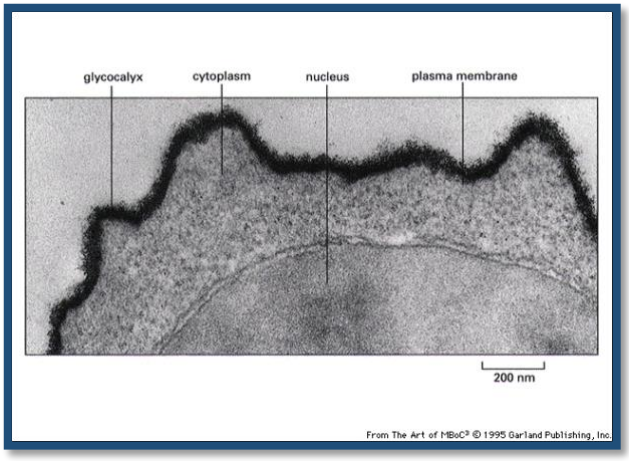


- P-Ser & P-Thr labile in CID, P-Tyr more stable
- Major ions observed -98 due to loss of H_3PO_4
- Alternative fragmentation methods

<u>Glucose Metabolism</u>

Sugars not just as energy but as recognition molecules.
Cells recognize molecules expressed on the cell surface.

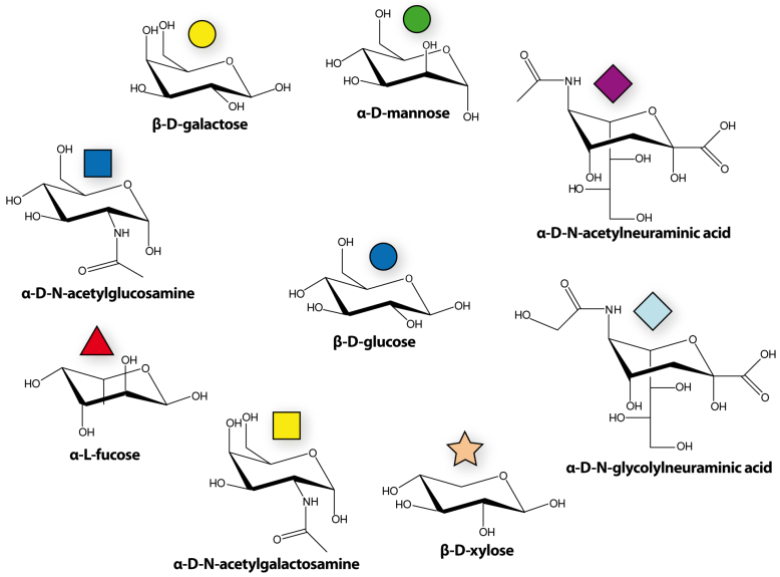
All cells have a sugar “coat” called the glycocalyx



EM of human white blood cell.
Outer surface has dense amorphic layer of sugar on the surface→ glycocalyx.
Not just mammalian blood cells; every cell in nature has some form of sugar, polysaccharides etc… on its Surface

Glycan-lectin recognition is key to cell-cell communication
Many fundamentals biological processes are driven by interaction between two cells and their specific recognition. These are often mediated by specific sugar molecules and specific sugar-binding protein (lectins) on the other cell. Lots of pathogen have hijacked sugar recognition process.

Glycoproteins carry a limited repertoire of sugars

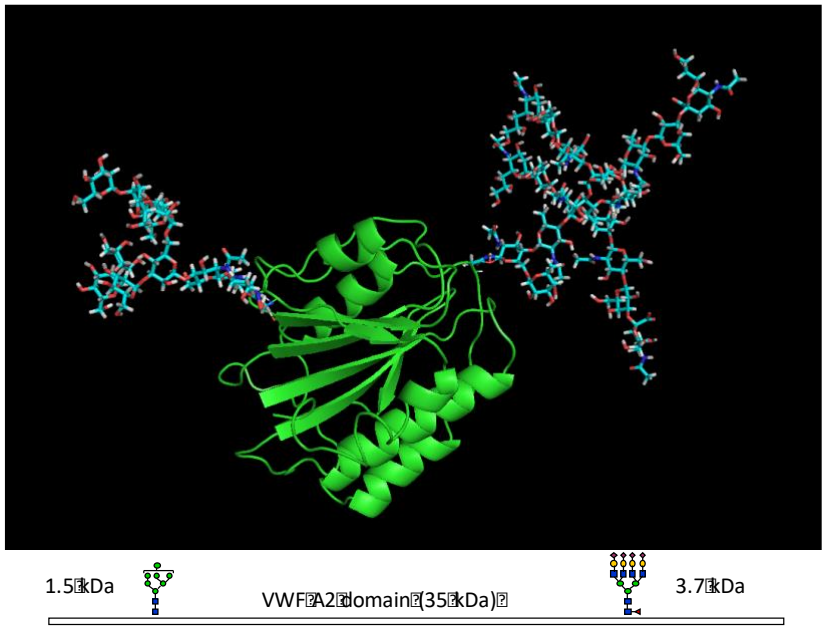


Prokaryotes have more ”exotic” monosaccharides.
Mammalians have quite limited repertoire.
Hexose family (mannose, glucose , galactose) → structural isomers of each other (change orientation of hydroxyl group) → can have profound effect on recognition of the molecules (recognition is dictated by 3D shape of sugar molecule)
Humans do not express n-glycolylneuramynicacid (only one hydroxylatioin difference), other apes can (one of the few biochemical differences between humans and apes, as we share 99. something of DNA)
Humans cannot oxidise NeuAc to NeuGc: partial gene deletion; inactive hydroxylase

<u>Glycosidic bond formation</u>
Lactose is made of galactose + glucose
Peptide bond is always the same, glycosidic bond varies.
Possible to link to multiple sites in sugar rings. (position 4, 2, and 3 etc…), potential to build branched structures
Possible to have alpha and beta glycosidic bond
Can get a lot more complexity with sugar residues than you would do with amino acid or nucleotides
Glycosidic bond formation is a very energetically unfavorable reaction
In cells, this reaction is coupled with breakage of high energy phosphate bond + they use catalytic power of enzymes (glycosyltransferases)

<u>Glycoproteins</u>
- Best characterized of all the glycoconjugates
- High proportion of secreted and membrane bound proteins are glycosylated
- Diverse functions
- Most glycoproteins are either at the cell surface or they are secreted (hydrophilic environment)

Glycans have a large hydrodynamic volume



Sugar that gets added to protein has large hydrodynamic volume
Glycan on the right is almost 10 times smaller than protein core, however it occupies a considerable a lot more space compared to the protein (hydrodynamic volume) + glycans are a lot more mobile than core.

Protein is mobile but glycan has a lot more freedom of more movement; in that movement it is potentially sampling a lot more space compared to the protein.

Protein glycosylation

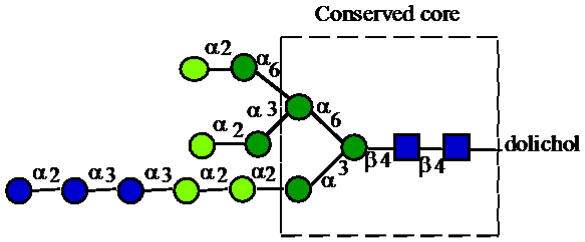
- Two forms of protein glycosylation
- N-glycosylation, sugar linked to amide nitrogen in the side chain of asparagine
- O-glycosylation, sugar linked to oxygen in the side chain of serine or threonine

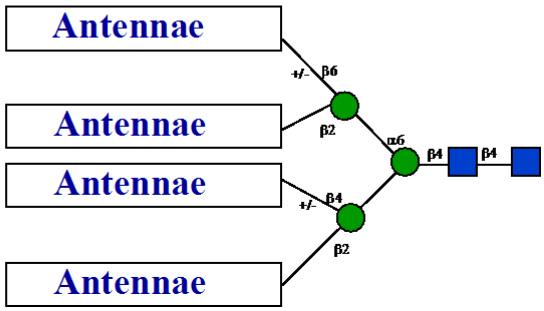
Protein N-Glycosylation

- Attached to Asn in the consensus sequon …Asn-X-Ser/Thr…where X is any AA except Pro
- Initiated in ER by *en bloc* transfer of a pre-formed lipid-anchored conserved glycan
- There has to be the consensus sequence on the protein in order to be glycosylated.
- N-linked is actually a co-translational modification --> do that as it is synthesized by the ribosome (before it is fully folded)

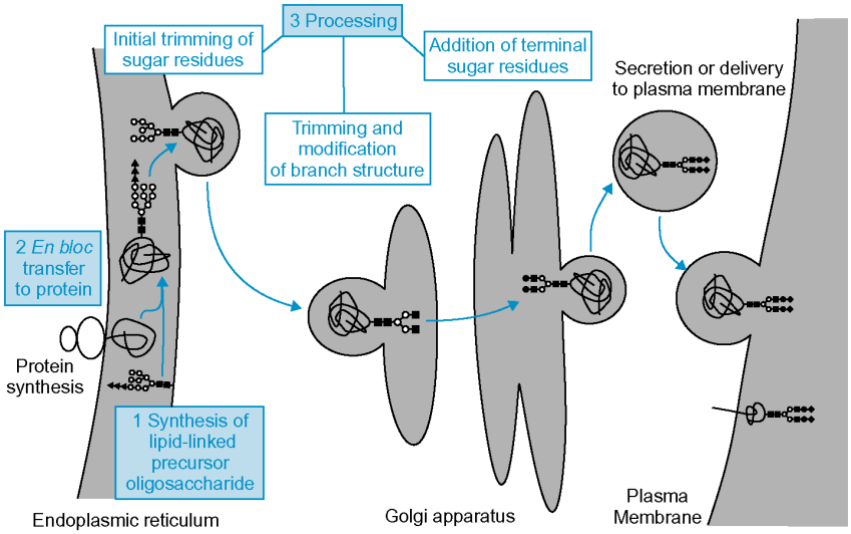
N-glycan structure

Start with precursor (dolichol type lipid) → 3 glucose, 9 mannose
N-glycans have 2 main domains to structures: all have conserved core residues (3 mannose, 2 NAG and then variable number of antennas) → constant core regions + variable set of antennas



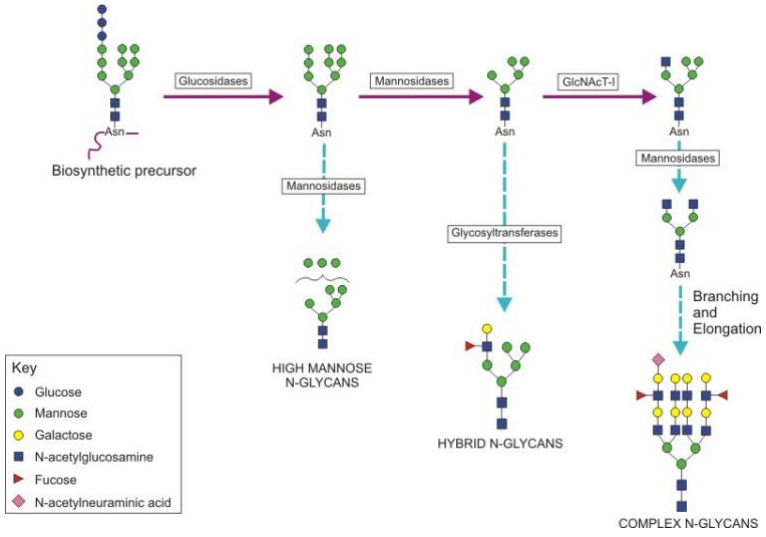


N-glycan biosynthesis



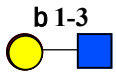
Eukaryotic N-glycan Biosynthesis

On some glycoproteins, start to remove mannose residues
Hybrid N glucans have some characteristics of high mannose and some of complex glycans.

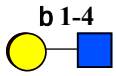


When building complex glycans the first thing is to build the antennas 8by addition of galactose to NAG)

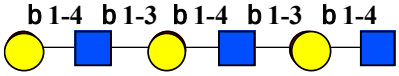
Antennae building blocks



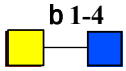
Type 1



Type 2 lacNAc

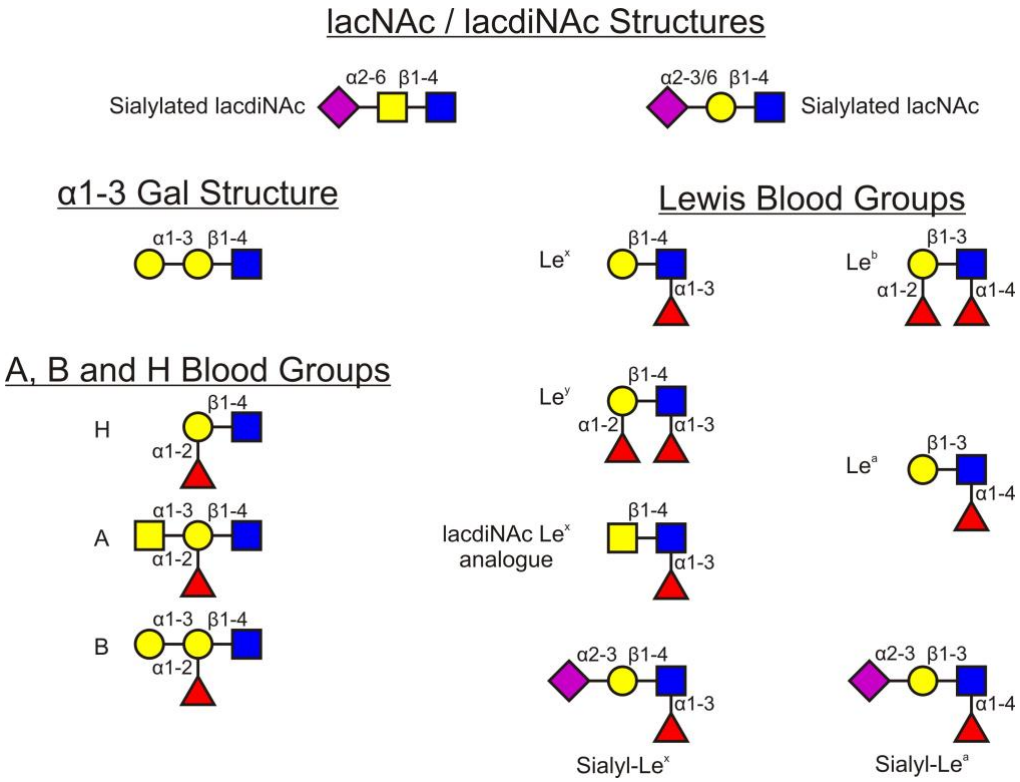


Polylac(Nac)



LacdiNAc

All of the glycosidic bonds on antennas are beta bonds. One of the consequences of having beta → tend to form long, straight rigid structures



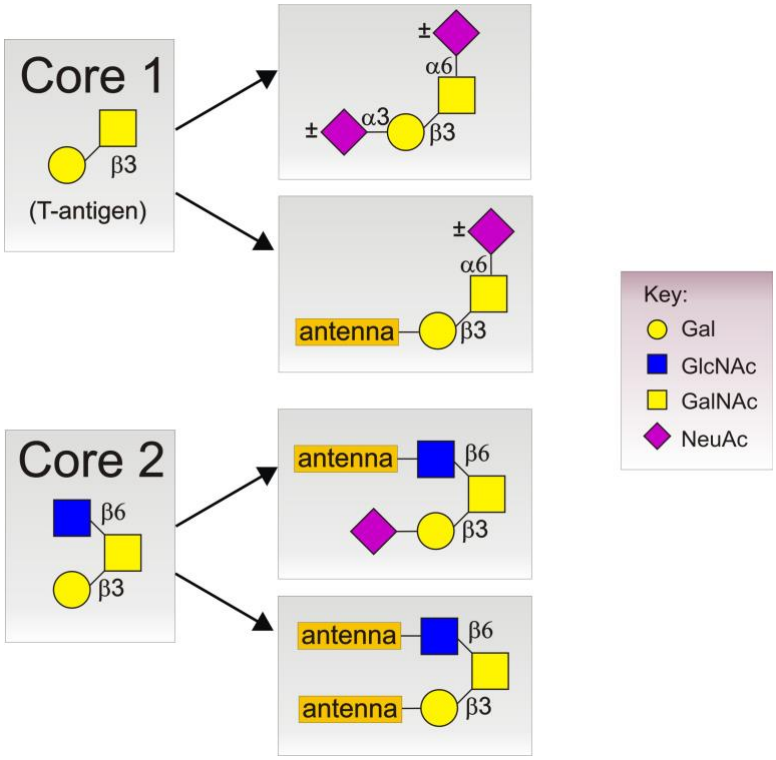
Capping sugars tend to be alpha linked → at an angle/kink (no long sequence but a kink)

O-Glycosylation

- Occurs on Ser and Thr
- No consensus sequence but some “rules” eg nearby proline, tandem repeats of Ser/Thr
- Initiated in Golgi by addition of a single sugar - usually GalNAc in mammals

O-glycans are classified by core structures

- At least 8 cores known
- Cores 1 & 2 very common in glycoproteins in general
- Cores 1-4 are the most common on mucins

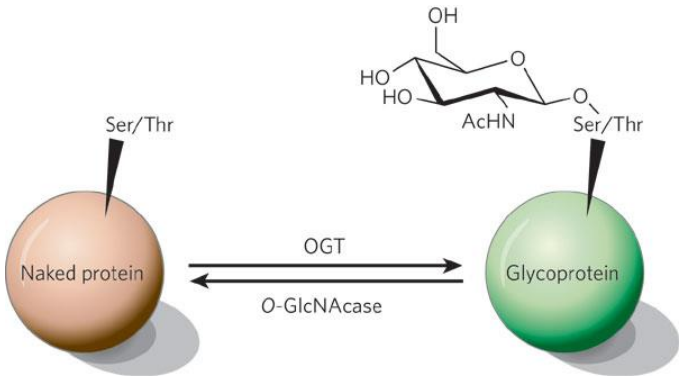


Sequential addition of monosaccharides residues (not addition of large pre-formed complex precursor). Most common type is core 1 (initial addition of GalNac); core 2 8addition of NAG in beta 1-6, then add antenna onto that.

Mucins

- Mucins are cell surface and excreted glycoproteins
- Protect mucus membranes by keeping them hydrated, acting as lubricants and prevent invasion by micro-organisms
- Heavy O-glycosylation occurs as a result of multiple S/T tandem repeats

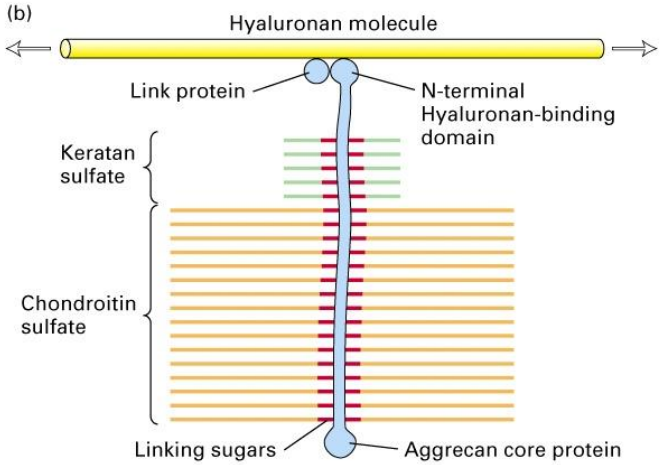
O-GlcNAc



- ♦ NOT elongated to more complex structures.
- ♦ Localized to the cytoplasm and nucleus.
- ♦ Present in all higher eukaryotes studied.
- ♦ As abundant as phosphorylation;
- ♦ O-GlcNAc proteins are also Phosphoproteins
- ♦ O-GlcNAc and Phosphorylation are often reciprocal.
- ♦ Highly dynamic modification - a regulatory role.

Proteoglycans

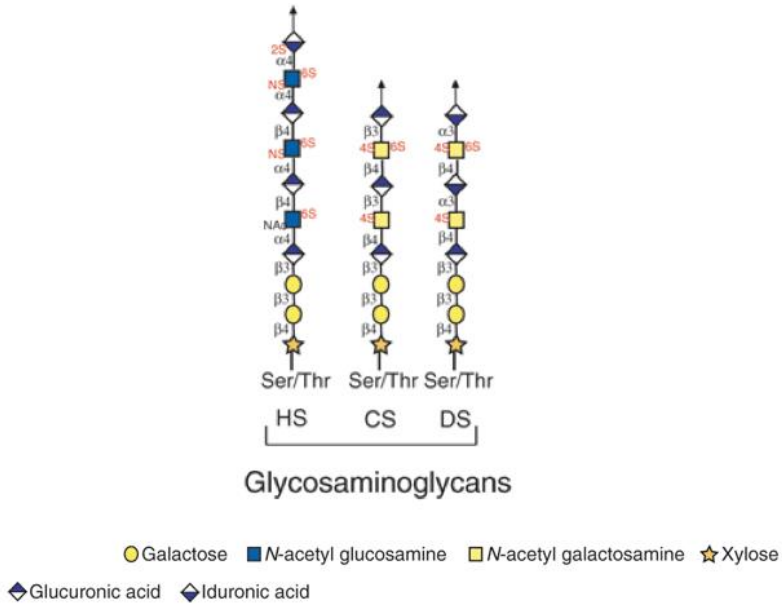
Are macromolecular structures, central protein cores with O- glycans attached to it.



- Glycoproteins found in the extracellular matrix and structural tissues such as cartilage
- Form massive macromolecular aggregates
- Highly hydrated, act as shock absorbers
- Also cell surface, specific recognition events

Proteoglycan O-glycan core structures

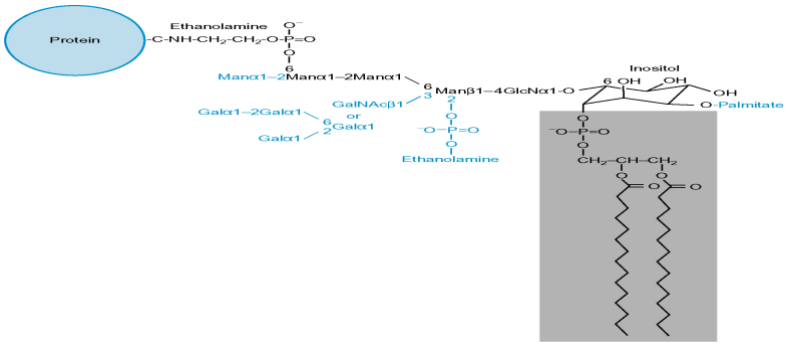
Very long linear chains of sulphated disaccharide units



HS=Heparin sulphate
CS=Chondroitin sulphate
DS=Dermatan sulphate

Glycosylphosphatidylinosotol (GPI)-anchors

Hybrid glycoprotein-glycolipid. Mechanism to anchor proteins to plasma membrane.



- Protein anchored to membrane via a diacylglycerol glycolipid
- Heterogeneity in structure between species
- Diverse array of protein use GPI-anchors to orientate them selves to the plasma membrane

Factors affecting glycosylation

- Protein sequence
- Sugar metabolism
- Expression of glycosyltransferases
- Competition between glycosyltransferases
- Physiological status
- Cell and tissue specific glycosylation

Glycosylation tend to be in hydrophilic outer area of the protein.
Different cells will express different glycosyltransferases and there is competition with the (can act on the same precursor)
Changes in glycosylation associated with stressing cells. Different disease and developmental states are associated with different glycosylations

Why Glycosylate ?

It is a huge energy investments so why bother? A lot of functionality

- Solubility
- Stability
- Conformation
- Organizational and barrier functions
- Cell-Cell and Cell-Matrix recognition

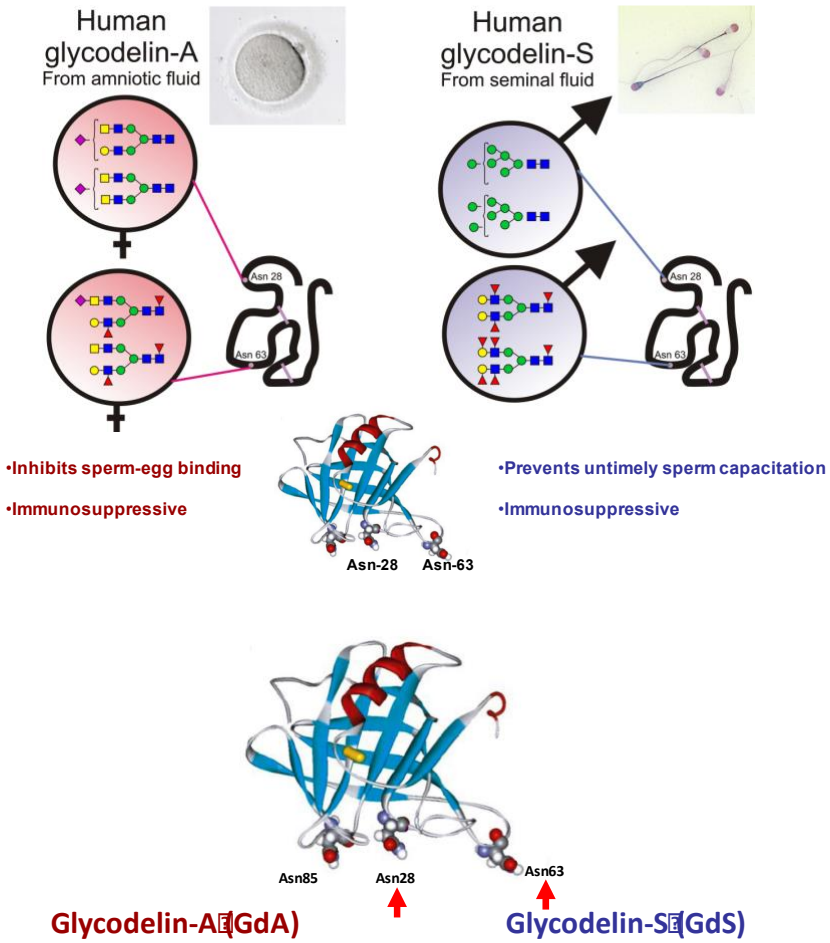
Definition of Glycoforms

Glycoproteins are mixtures of glycoforms i.e. same polypeptide but different glycans
The populations of sugars attached to an individual protein will depend on the cell type in which the protein is expressed and on the physiological status of the cell and may be developmentally regulated
Different glycoforms can potentially have different biological functions

Sperm-egg recognition involves glycan recognition

⇒ **Glycodelin**

24 kDa, 3 potential sites of N glycosylation. Described in female reproductive system. When pregnant → massively upregulated.
It is a potent immunosuppressant (knock down a lot of immune cells, t cell, macrophages etc…) -→ developing foetus is in knocked down envinonment (so immune system of women do not kill developing foetus).
Natural contraceptive, only expressed when it is a non fertile menstrual cycle. Regulating fertile periods in female reproductive tract.
Exactly the same gene was founfd in the male (found using antibodies targeted to female , which bound also to males).
Why does male ejaculant contain a contraceptive molecule ? Indeed actually it is not a contraceptive but it encourages sperm-egg binding (but is exaclty the same gene product, same sequence, same folding)
The answer is that different function is given by gender specific glycosylation.



Sources:
Endometrium, amniotic fluid, pregnancy serum

Activities:
- Inhibits sperm-egg binding
- Immuno-suppressive

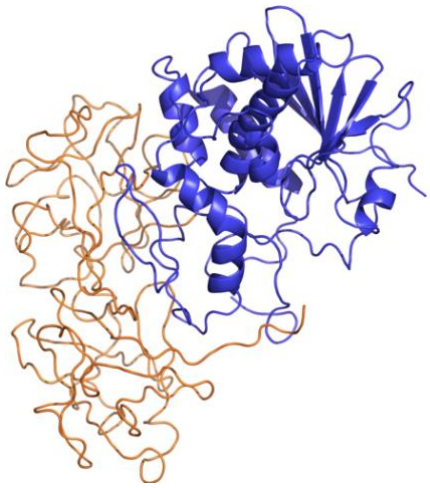
Sources:
Seminal vesicles, seminal plasma

Activities:
- Enhances sperm-egg binding
- Immuno-suppressive

Sugar Recognition

Lectins are proteins of non-immunoglobulin nature capable of specific recognition and reversible binding to carbohydrate moieties of complex carbohydrates without altering the covalent structure of any of the recognized glycosyl ligands.
(Kocourek and Horejsi, 1983, Clinical Biochem. 3, 3-6)

Ricin: a lectin and a toxin



A-chain is a toxin that inhibits protein synthesis
B-chain is a Gal-binding Lectin

Carbohydrate Recognition Domains (CRDs) of Lectins

- Usually found in shallow indentations on surfaces of lectins
- glycan binding modes include chelation with divalent cations especially Ca^{2+}, hydrogen-bonding with sugar OH and amide groups, van der Waals' interactions with sugar hydrophobic faces (eg with galactose) and ionic interactions (eg with carboxylate of sialic acids)
- The glycan ligand is usually one to four residues in size
- Binding is low affinity but high specificity but multivalency can result in tight binding despite low affinity

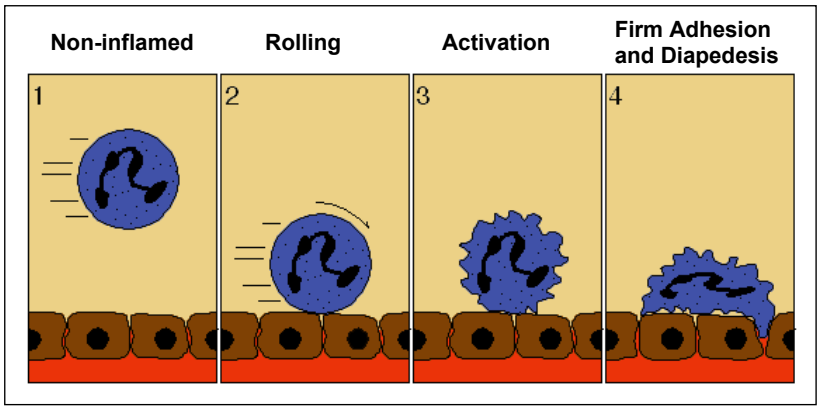
Classification of important animal lectins

Family	Defining feature	Ligand
C-type	Conserved CRD	Various
I-type (Siglec)	Ig-like CRD	Sialic acid
Galectins (S-type)	Conserved CRD	b-galactosides
P-type	"Conserved" CRD	Man-6-P

C-type lectins and the selectins

Professor Drickamer identified the amino acids that define the CRD of C-type lectins
Selectins are members of the C-type lectin family; they are very important for leukocyte trafficking
Various leukocytes such as macrophages and neutrophils need to be recruited to sites of acute/chronic inflammation or infection; Inducible process, only occurs on activated endothelium
Other leukocytes such as lymphocytes must traffic from the blood circulation into the lymphatic circulation

Leukocyte trafficking

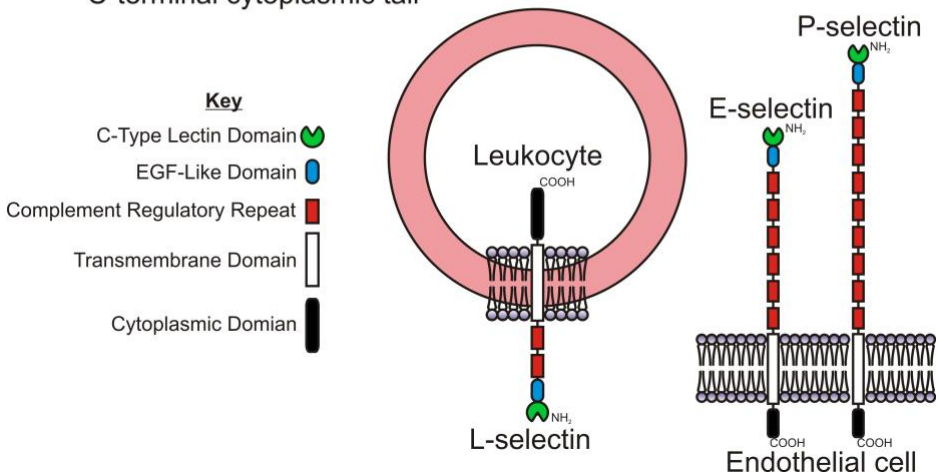


How were the selectins discovered?

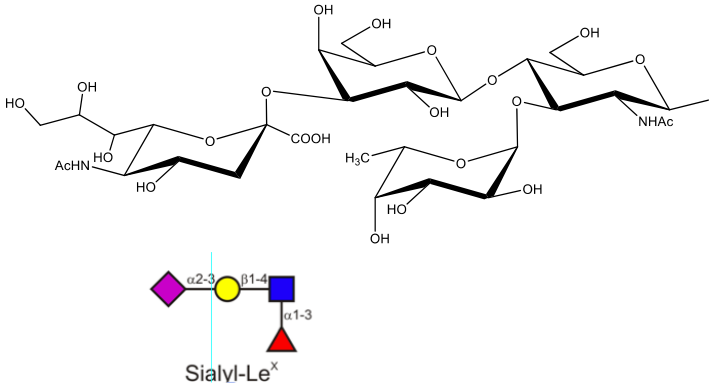
- Antibodies raised against activated endothelium and leukocytes which inhibited cell rolling
- The 3 proteins they bound to were immunopurified
- All 3 nearly simultaneously cloned in 1989
- Sequence identified N-terminal CRD's therefore lectins, named selectins

Structure of the selectins

All type 1 membrane proteins with N-terminal CRD, followed by EGF-like domain, variable numbers of complement-regulator repeats, transmembrane domain and short C-terminal cytoplasmic tail



In the early 1980's mass spectrometric analyses at Imperial College showed that leukocytes have SLe^x on their surface
(collaboration with Minoru Fukuda, La Jolla, USA)

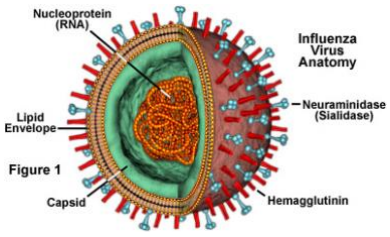


This glycan sequence was found to be the ligand for the selectins when they were cloned in 1989

Human diseases associated with selectins

- Leukocyte adhesion deficiency II- complex symptoms include recurrent infection; linked to Fuc metabolism, can't make SLeX
- Many diseases associated with chronic/acute inflammation
- metastatic cancers likely exploit the selectin pathway; SLeX is highly expressed on many cancer cells
- As yet no drug on the market but some recent promising progress

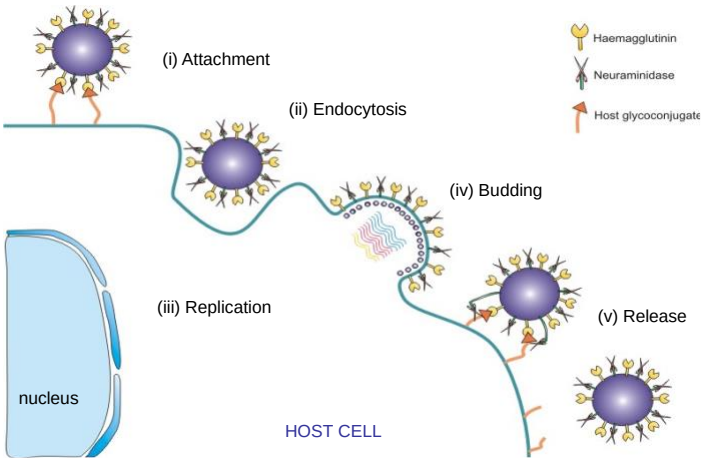
Influenza viruses



- RNA virus
- 3 types A, B, C
- A most virulent
- 16 serotypes of HA
- 9 serotypes of NA

Sugar-lectin interaction that is deleterious in the host by pathogen (whereas in immune system is beneficial this interaction)
Influenza viruses are able to infect a lot of different species (thought to be in aquatic birds). Mostly we infected by A and B (A is most common).
On surface there are homochromatic hemagglutinin and homotetramer neuraminidase.

Influenza infection



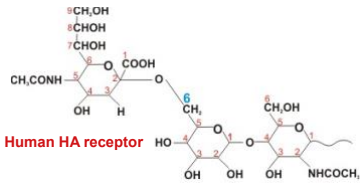
Both of the two proteins are involved in sugar recognition events. Influenza virus initiates infection process by binding by host glycoproteins and glycolipids on surface of host cells (in humans in epithelial cells that line in respiratory tract). Hemagglutinin so is a sialic acid specific lectin. HA has second important role → start endocytotic process, mediates fusion of viral cell membrane with host and internalization of viral particle.
Virus needs also to be able to release form cell. NA is an enzyme that cleaves sialic acid.
NA also is probably involved in cutting through mucin barrier in order to infect infection process.

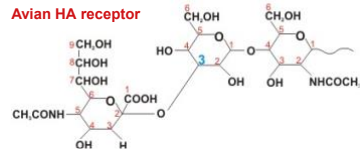
Human Pandemics

Every year there is a seasonal influenza. Big problem occurs every 10-20-30 years → influenza virus is produced that is able to infect that is able to infect millions of people (not just thousands).

- 1918 Spanish Flu H1N1
- 1957 Asian Flu H2N2
- 1967 Hong Kong Flu H3N2
- 2009 Swine Flu H1N1

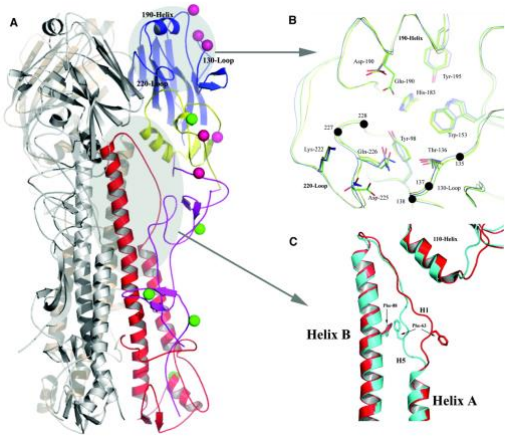
Haemagglutinin receptors



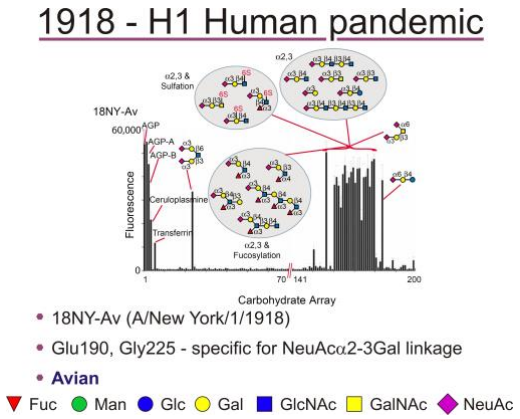


How does influenza virus adapt and start to jump species and infect ultimelty humans ? Major barrier to trans-species transmission is to do with HA molecule. HA binds to sialic acid but as more lectins is more specific than that.
In avian species HA binds sialic acid with 2-3 apla glycosidic bond.
Human adapted influenza viruses are able to bind to alpha 2-6 syalated glycans . Is this ability of HA to change specificity from alpha 2-3 to alpha 2-6 (human respiratory tract).

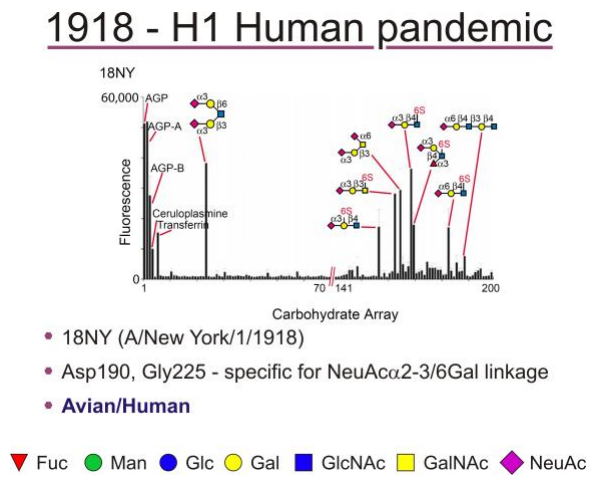
1918 H1 Human Pandemics



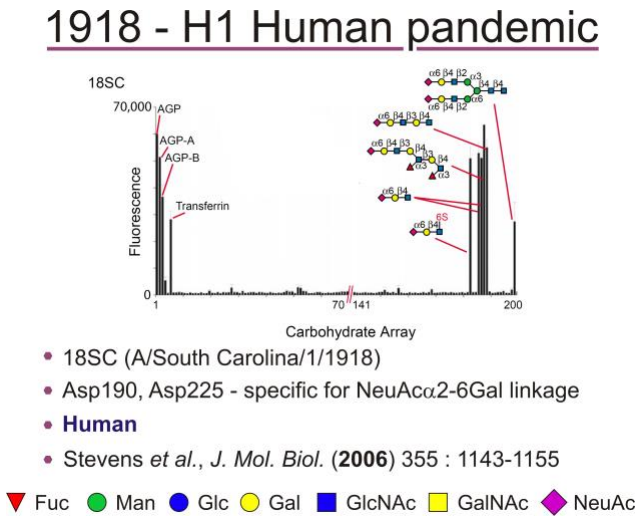
Able to identify key residues that change in avian and human adapted HA.
Glycan array to assess specificity of HA. Whole range of glycan structure arrayed on a surface (2-3 or 2-6 syalic acid) and then you see which glycan the HA bind to, then antibody to His tag and then secondary antibody and then fluorescent signal.



When ran avian virus → only alpha 2-3 structure (pure avian specificity).



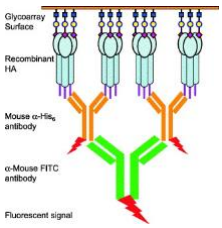
When looked at adapted virus → mixed specificity (human/avian); able to bind 2-3- and 2-6. It is only one single amino acid change. Asp at 190



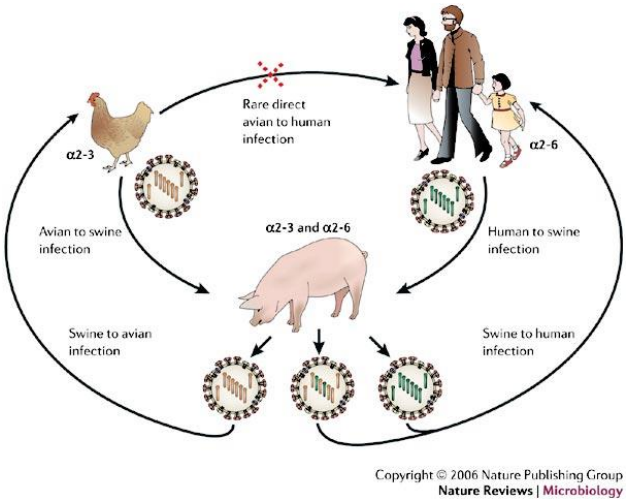
Exclusively human specific. Has a second mutation in Asp 225. It only takes mutations in 2 amino acids in HA to change specificity from 2-3 to 2-6.

Glycan Array

Human Pandemic H1N1



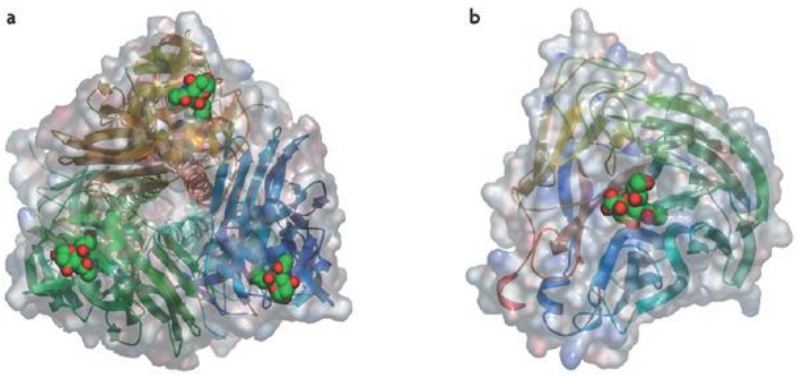
it is not that easy however to jump from bird to humans. Virus goes to a series of adaptation as they pass through different species (very rare to go from bird directly to human).



Pig express both 2-3 and 2-6 sialic acid. Virus jump from bird to pig, goes a period of adaptation and then human specificity.

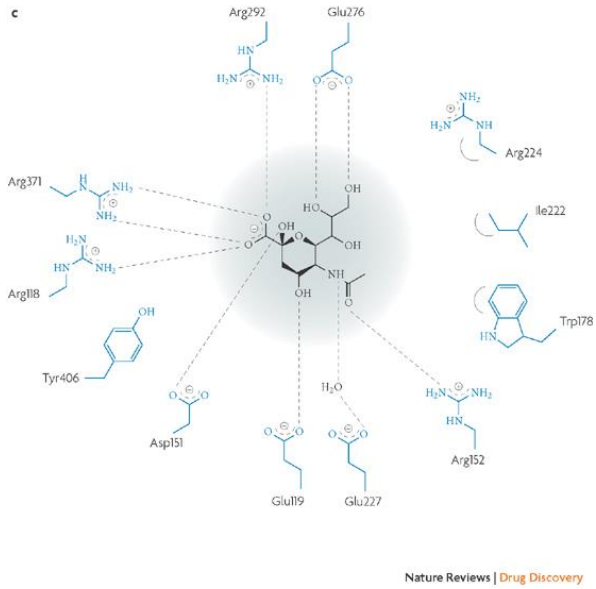
20 to 27% of global population got infected with H1N1. However, it wasn't that virulent. In fact, fatality rate has been estimated to be 0.02%. It could have been a lot worse.
Since 1997, only 860 confirmed laboratory cases of H5N1 avian Hong Kong influenza (not a lot), however the fertility rate is 53%. This particular strain produces an excessive immune response in the people infected. A lot of selectin mediated white cell trafficking to the lung. Your own immune response start to break down membrane barrier of the lung → impregnation with fluid from the body (patients drown in their own fluids).

Tamiflu and Relenza (neuraminidase inhibitors)

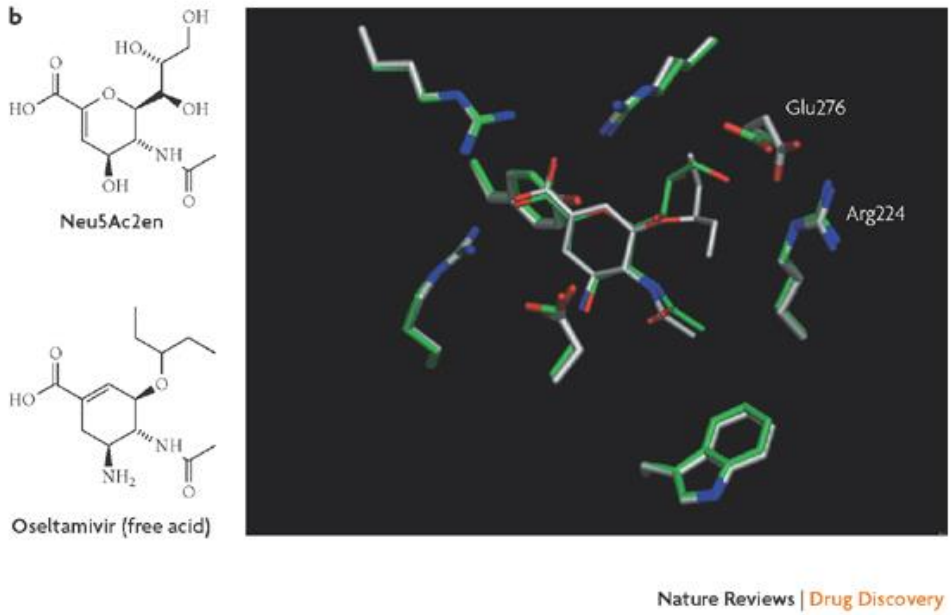


Crystal structures of HA and NA. HA is a lectin, it is not a good target to produce drugs against (lectins have shallow binding sites and weak interactions in those

sites). NA on the other site is an enzyme with deep active site with strong interactions with sialic acids. NA has been targeted for anti-influenza drugs.



8 key amino acid residues tend to be invariant tin NA (ionic interactions that allow to recognize sialic acid in active sites. Start to do rational based aleration of sualic acid to generate family of potential inhibitor molecules.



Oseltamivir is the most potent inhibitor. Glycerol side chain is replaced by more hydrophobic acyl chain without hydroxyl residues (no more H-bond with Glu276). Drug molecule is the white molecule, natural sialic acid is green and around are side chains. Normal interaction there is h bond with OH groups and Glu 276. When there are hydrophobic interaction → no more h bond → conformational change → glu interacts with Arg 224 → inhibited (virus no more able to release itself and no penetration of mucin barrier.

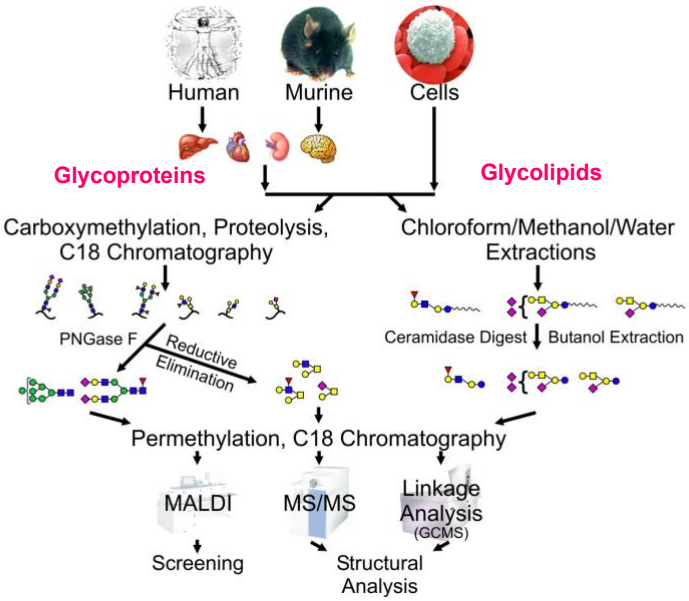
Tamiflu developed by Roche and has this strategy. Relenza by GSK (no oral administered, but inhaler and patients do not like inhaling). They are not perfect, you need to take them early in the infection process. Starting to be resistance strain of influenza mutating NA.

Glycomics

- Determining the glycan repertoire in cell, tissues organs etc as a first step to defining functions. Prior knowledge of glycan biosynthetic pathways is essential.

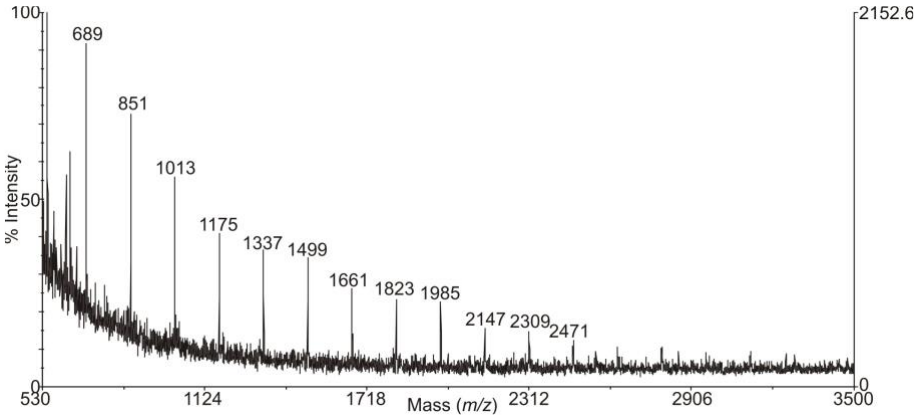
Mass spec is the most powerful technique to find glycan structures. Obkective is to find all the glycan structures in a particular system.

Glycomics Screening Strategy

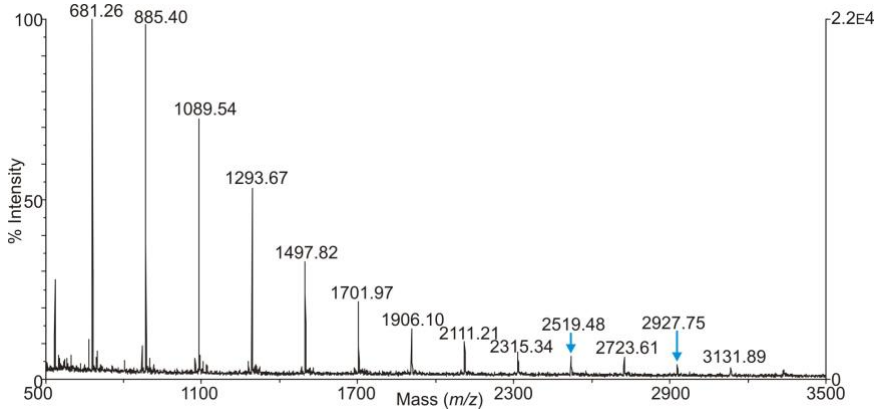


Start from complex biological systems. Extract glycoproteins using detergents, get population of ending glycans using PNGaseF, which cleaves glycan ending of the glycoprotein. DO a chemical reductive pemethylation process (gives O glycans). Using Ceramidase can get glycan ending of glycolipids.
Gycans are very hydrophilic molecules (do not tend to work well on mass spec). In order to get hugh sensitivity data we need derivatization → permethylation (replace hydroxyls with methyls).

Why permethylate? – Neutral saccharide, native



MALDI mass spectra of hexose polymer. If take the same amount of glycan but this time permethylated :

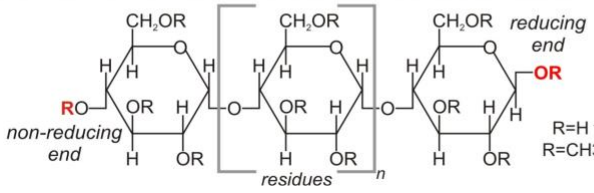


A lot more less background signal (5%, before it was 40%). Signal to noiuse is better (nice clear peaks). Quality of data is improved by permethylation

Glycan residues

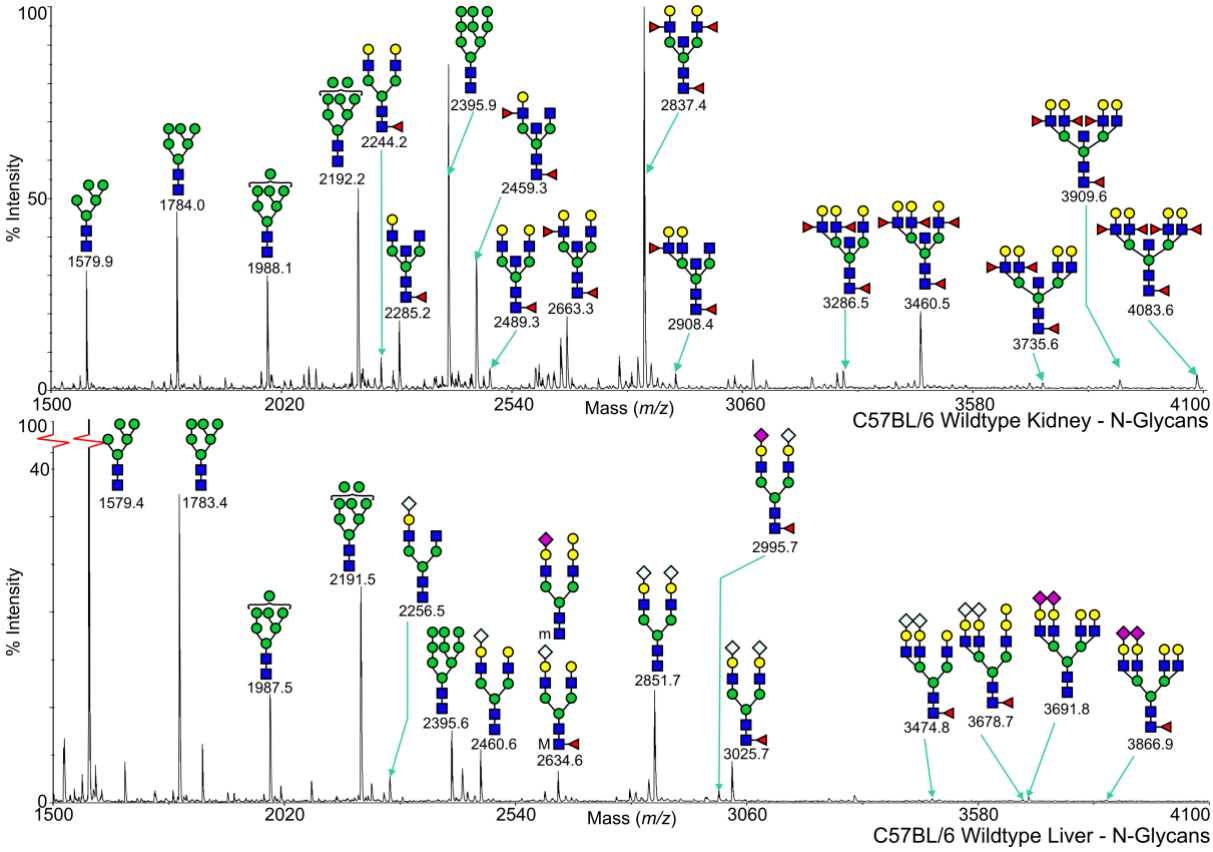
Monosaccharide	Abbreviation	Symbol	Permethylated residue mass (accurate)	Average No. Methylation events
Deoxyhexose				
Fucose	Fuc	▼	174.0892	2
Hexose				
Mannose	Man	●	204.0998	3
Galactose	Gal	○	204.0998	3
Glucose	Glc	●	204.0998	3
***N*-acetylhexosamine**				
N-acetylgalactosamine	GalNAc	□	245.1263	3
N-acetylglucosamine	GlcNAc	■	245.1263	3
Sialic acid				
N-acetylneuraminic acid	NeuAc	◆	361.1737	5
N-glycolylneuraminic acid	NeuGc	◇	391.1842	6

Non-reducing and reducing ends

Glycan Structure	Permethylated mass of ends
	R+OR = 46 R=H for underivatised glycans R=CH3 for permethyl derivatives

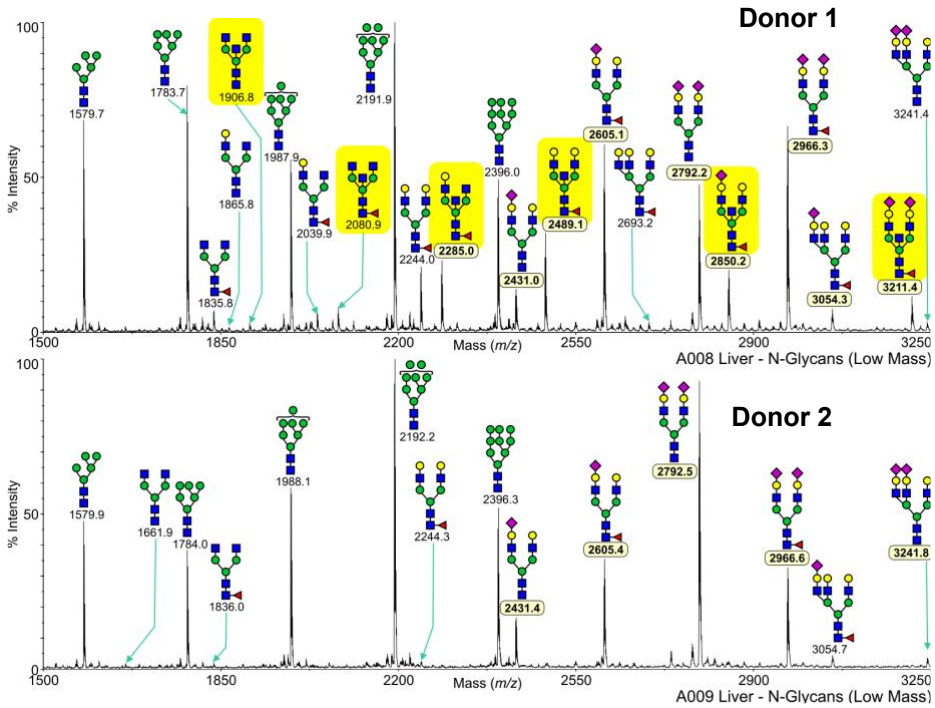
Need to be able to understand what happens to glycans in the methylation reaction. Deoxyhexose has 2 potential sites for methylation. A hexosugar has three potential sites for methylation etc… We can calculate the residue mass after permethylation. Murine kidney (upper) and liver (lower) N-glycans. Use residue mass as starting point in calcuòaiton. When looking at peptides → specific N terminus and O temniuns and need to take into account functional groups at these ends. It is the same for carbohydrates → have anon reducing and reducing end. Non reducing: add methyl group. Reducing: add an O-methyl group. We tend to use MALDI and weso we need charge on the glycan. Gycans tend to be M+Na.

LDI mapping showing organ specific glycomes



Compelx glycans tend to be decorated with fucose residues. Compare them with mouse liver sample: mostly biantennary compelx glycans, tend to be sialeted (not fucosilated). Organ specific glycosylation profiles; kidney has different function from liver, different glycosylation patterns.

MALDI mapping of human liver N-glycome illustrating potential for diagnostics: Bisected structures are indicative of liver disease



Donor 2: liver cancer patient. Look at N-glycan profiles (have changes in glycans profile in cancer patients. → increase in truncated complex glycans, bisecting GlcNAc). Also get increase in fucosylation events in the core.
Use glycomics methods to start look at differences of glycosylation in diseases.

Potentially exploit this as early biomarker for cancer.

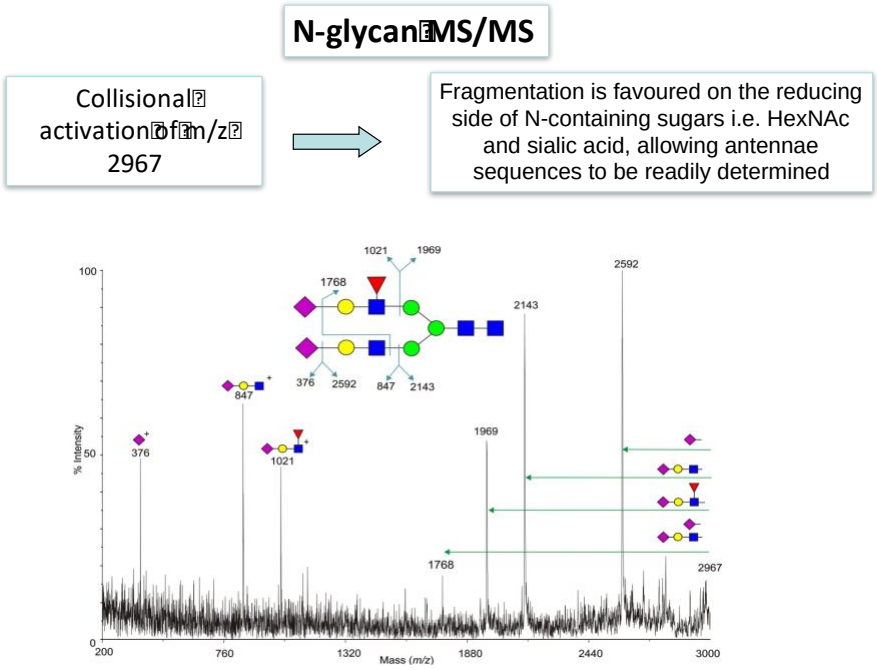
The problem is that mass spec measures m/z ration. Not able to differentiate between mannose, glucose and other structural isomers. However, we know the N-glycan biosynthetic pathways, we know which saccharides must be present. Biosynthetically, the only combination of monosaccharides that can give a peak of 5 hexoses and 2 glNAc is a Man5-2GluNac peak. They follow a sequential pattern. Start from core, then antenna, then decorate the fucose etc… By bringing knowledge of biosynthetic pathway, able to have a lot more of structural details.

Glycomics:MS/MS Sequencing

How do we differentiate the positions of biosynthetic possibilities? Use 2D Mass spec

A variety of suitable technologies is available eg ES-Q-TOF, ES-ion trap, MALDI-Q-TOF etc

MALDI-TOF/TOF is potentially the most powerful for automated high-throughput sequencing

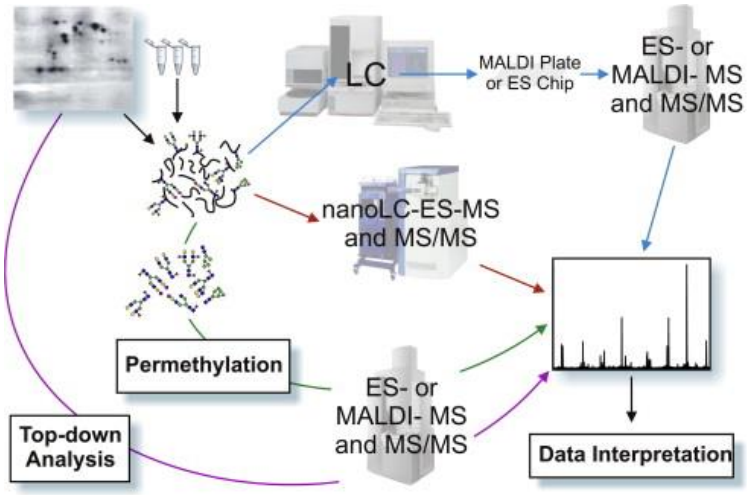


Able to say that fucose residue is in the antenn a and not in the core. 2D analysis give heightened level of definition.

Glycoproteomics

The problem with glycomics is that we are releasing glycans before the analysis. We are actually throwing a way the protein context of that glycan, don't know that glycoprotein, what sites on what glycoprotein those glycans are.

- Glycoproteomics-Defining the glycosylation status of individual proteins and individual sites of glycosylation (analyse glycan and protein component).
-
- More challenging and time consuming than glycomics



In glycoproteomics experiment, purify glycoproteins, proteolytic digestion to get glycopeptides, analyse them on mass spec. Complex nature of mixture, use LC seoaration (on line LC into ES mass spec or offline into MALDI). Then look at glycopeptides molecular ions and then fragment them and get both peptides and glycans fragment ions. Have to be able to interpret peptide and glycan fragment ions to work out structure e of glycan and peptides and the sites of glycosylation.

CID-MALDI-TOF/TOF MS/MS spectrum of the glycopeptide molecular ion at m/z 2555.2.