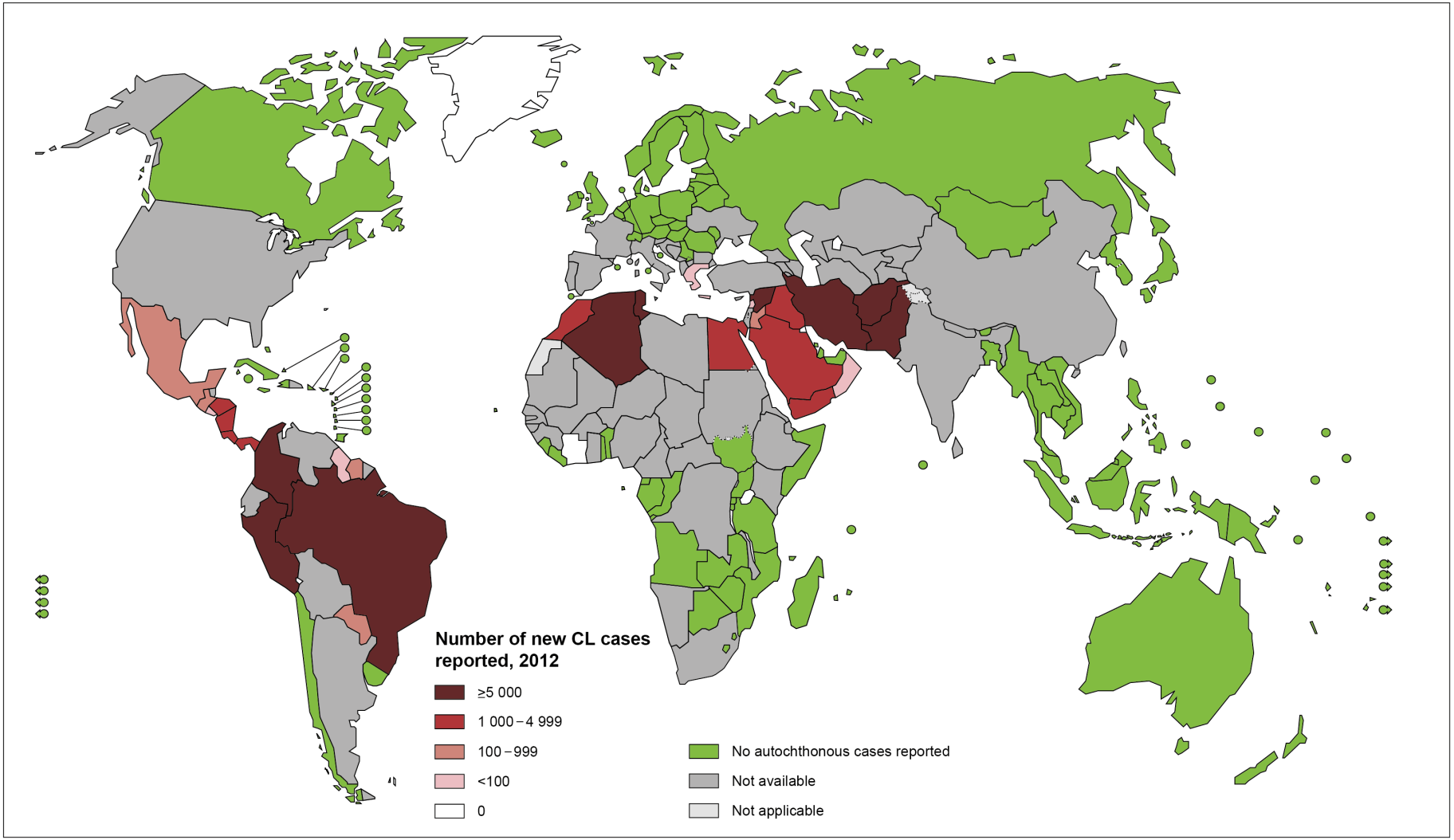
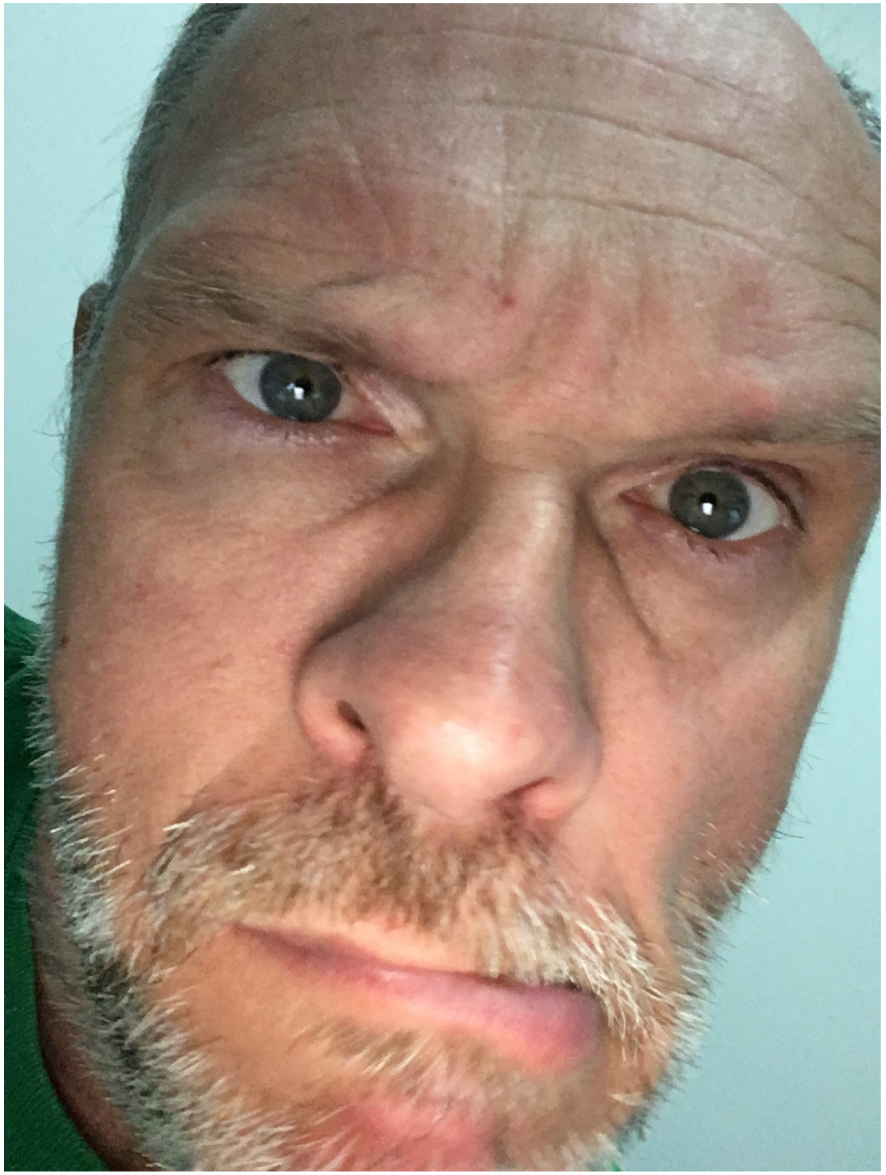
Status of endemicity of cutaneous leishmaniasis, worldwide, 2012



Data Source: World Health Organization
Map Production: Control of Neglected Tropical Diseases (NTD)
World Health Organization

World Health Organization

Leish MANIAC



University of Glasgow

- In Scotland's largest city
- Founded 1451
- ~25,000 students
- ~7,000 postgraduates

Glasgow Polyomics

- Data collection
- Data analysis
- Software development
- Omics training

Institute of Infection, Immunity & Inflammation

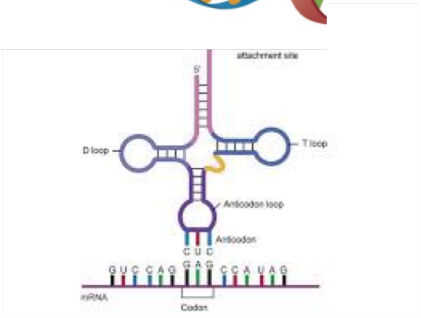
- Pathogen biology
- Immunology
- inflammation

Introduction to proteomics

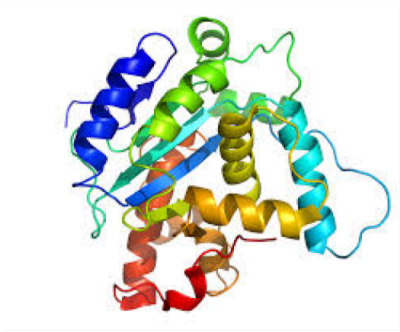
- Why study the (parasite) proteome?
  - Potential benefits
  - Challenges
- How can proteomes be characterized?
  - Protein separation
  - Protein characterization
  - Databases
- Some example applications in parasitology

The omics information cascade

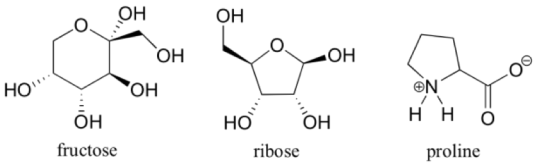
Genome



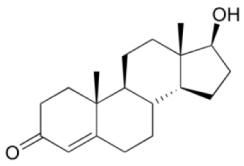
Transcriptome

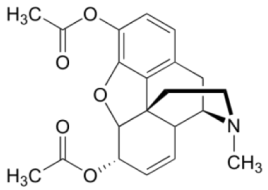


Proteome



Metabolome





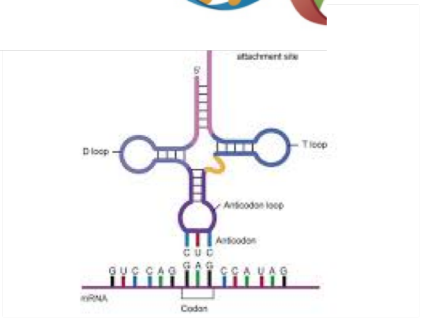
heroin

GENOTYPE

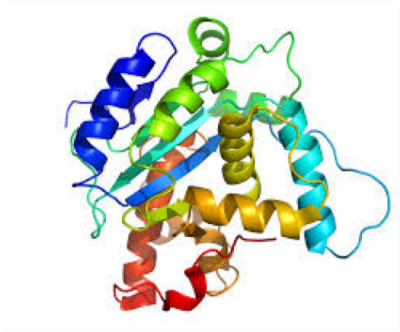
PHENOTYPE

Gene expression is regulated

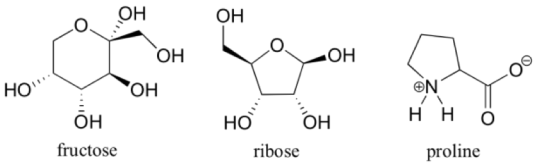
Genome



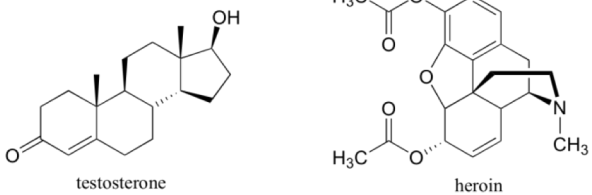
Transcriptome



Proteome



Metabolome

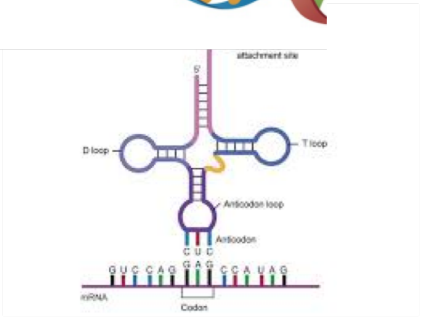


STATIC

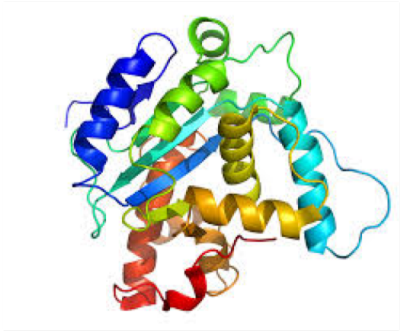
DYNAMIC

Everything contributes to the phenotype

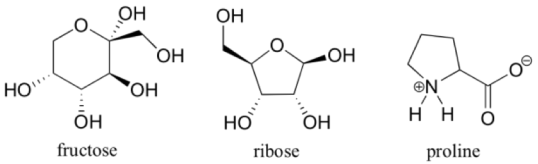
Genome



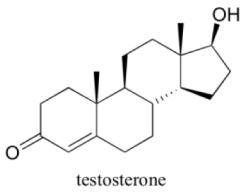
Transcriptome

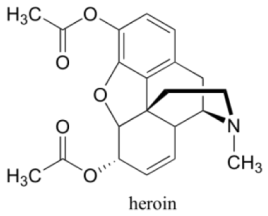


Proteome



Metabolome





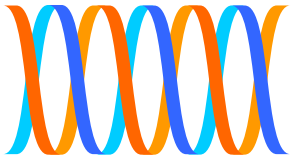
FEEDBACK

The omic information cascade is *orchestrated*

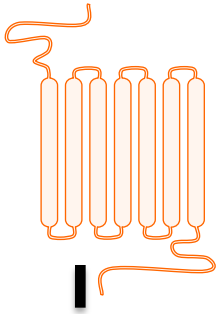
Score (DNA) → **Conductor (mRNA)** → **Musicians (proteins)** → **Instruments (metabolites)**

Proteins at the interface between information and activity

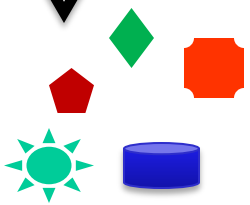
Genotype



Transcription, translation



Diverse enzymatic activities



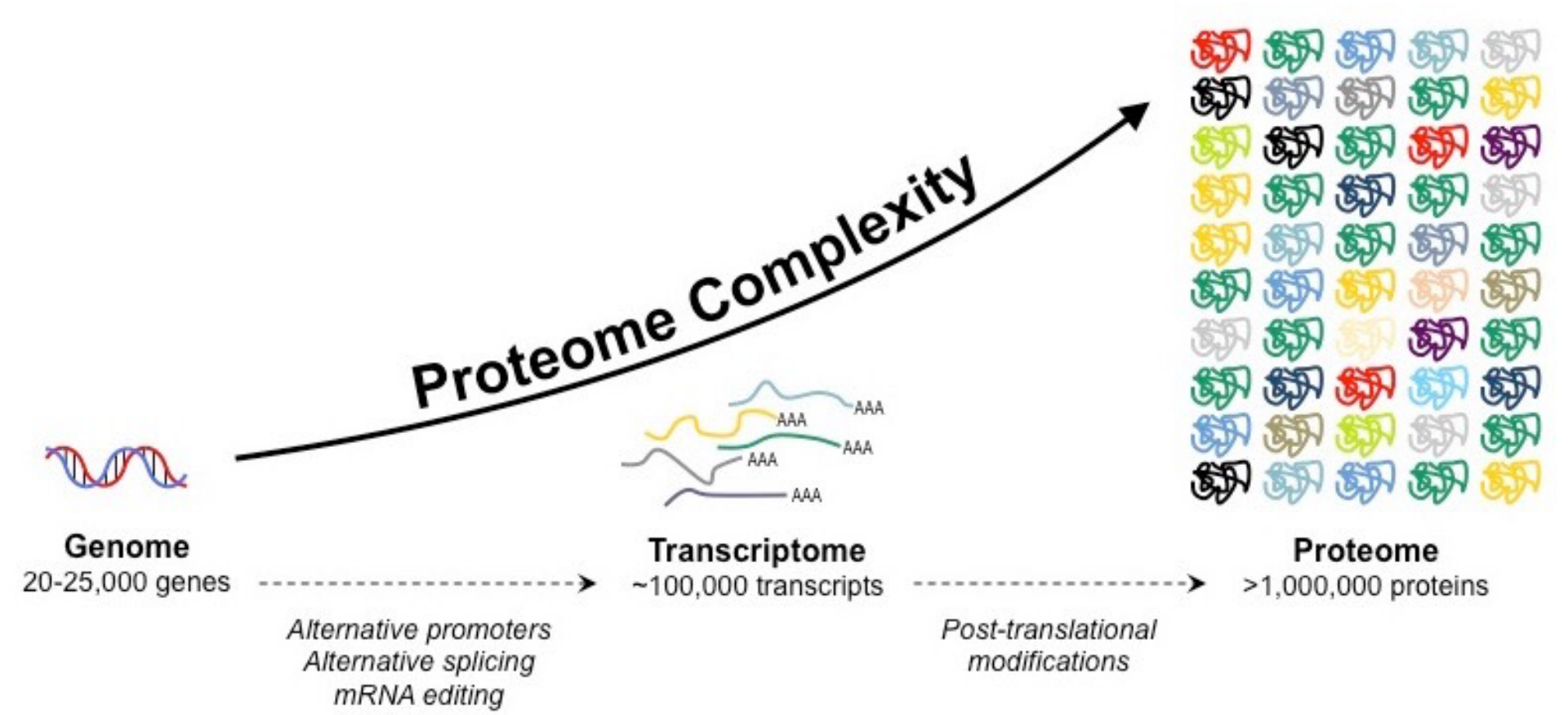
Phenotype

Proteome is related to **genome** by a defined code

Proteome can be related to **phenotype** *only by experiment*

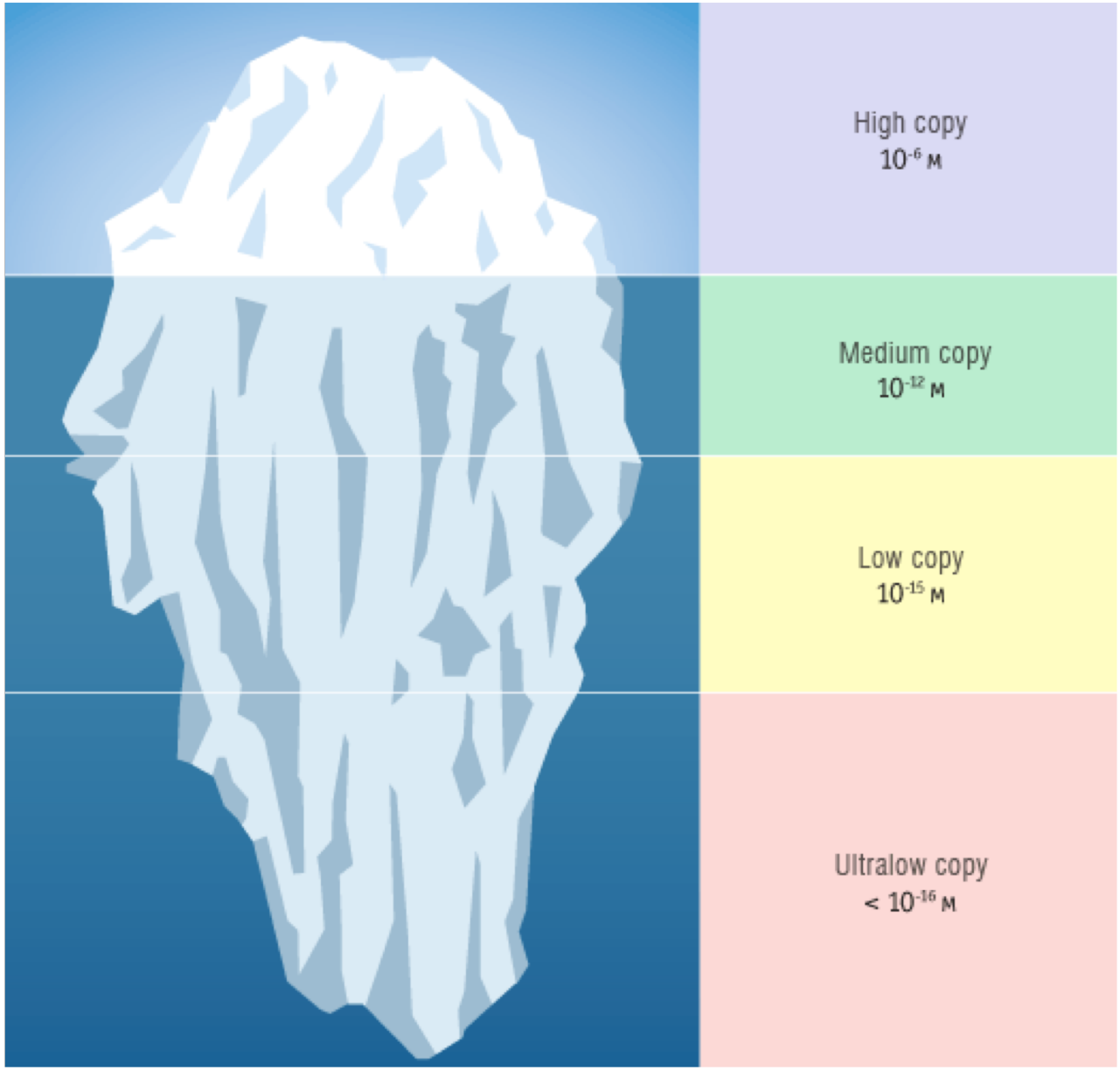
PROTEINS are effector molecules: Responsible for growth, reproduction, movement, communication……

Proteome is complex and dynamic

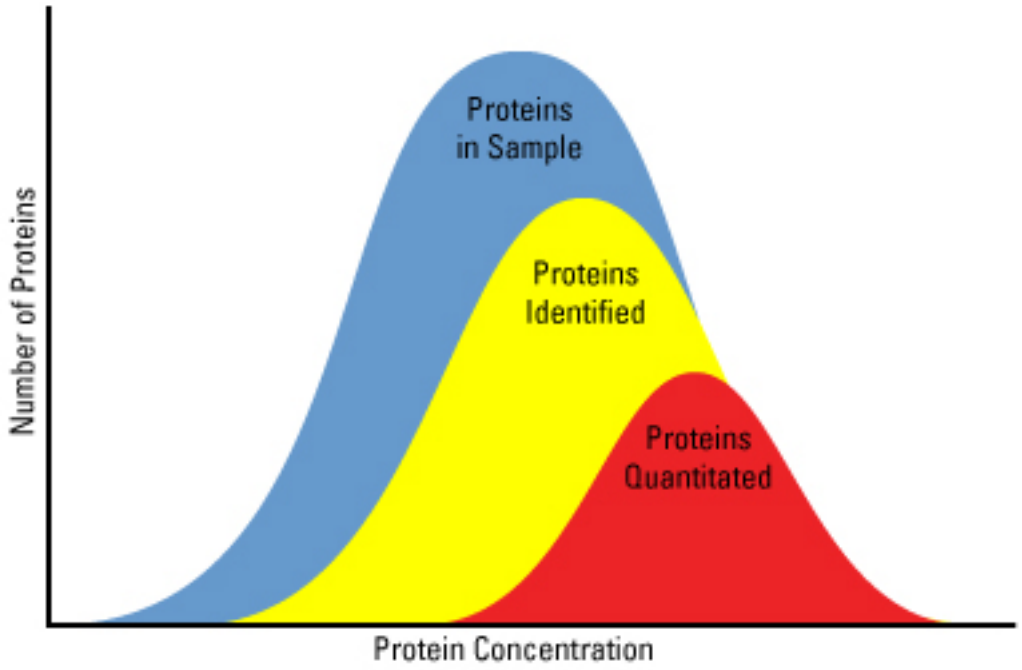


STATIC → **DYNAMIC**

Most proteins are rare



More abundant proteins are often more readily detected

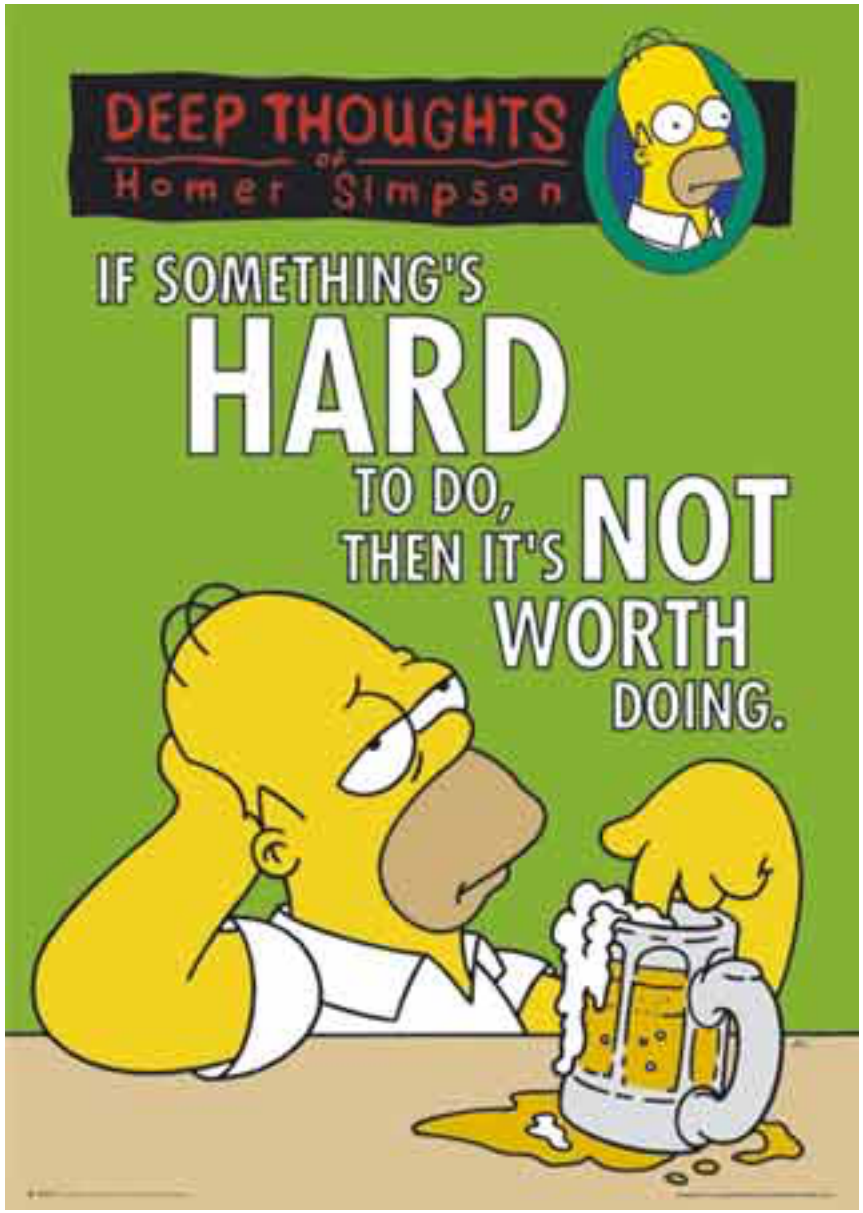


What is proteomics?

- In theory, proteomics is characterisation of **all** the proteins in a system
- In practice, proteomics is characterisation of **some** of the proteins in a system

Proteomic analysis:

- Partial, biased coverage is the norm
- Goal is to provide molecular basis for a phenotype



Proteomics is challenging!

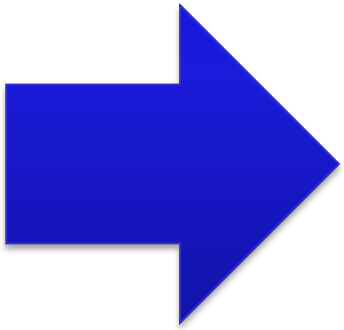
Gene expression can be studied at the RNA level – *why bother with proteomics?*

Why is it relevant to investigate parasite (or host) proteomes?

Proteins are the primary effector molecules

Parasite protein

- **Expression**
- Activity
- Localisation
- Interaction

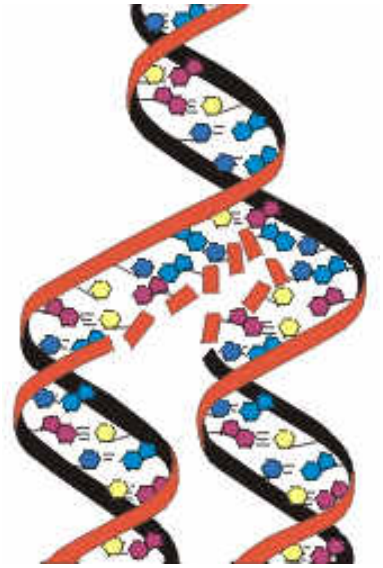


Parasite phenotype

- Virulence
- Differentiation
- Drug resistance
- Host tropism
- Immune evasion
- ……

We think that genes that are expressed have a function – proteomics can tell us about expression, but function is more difficult

One genotype encodes many phenotypes



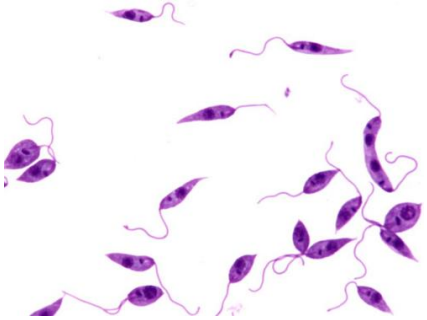
Genotype

Gene expression can be differentially regulated
- at transcription or downstream

RNA

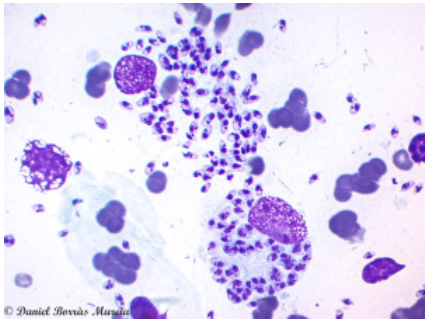
RNA

PROTEIN

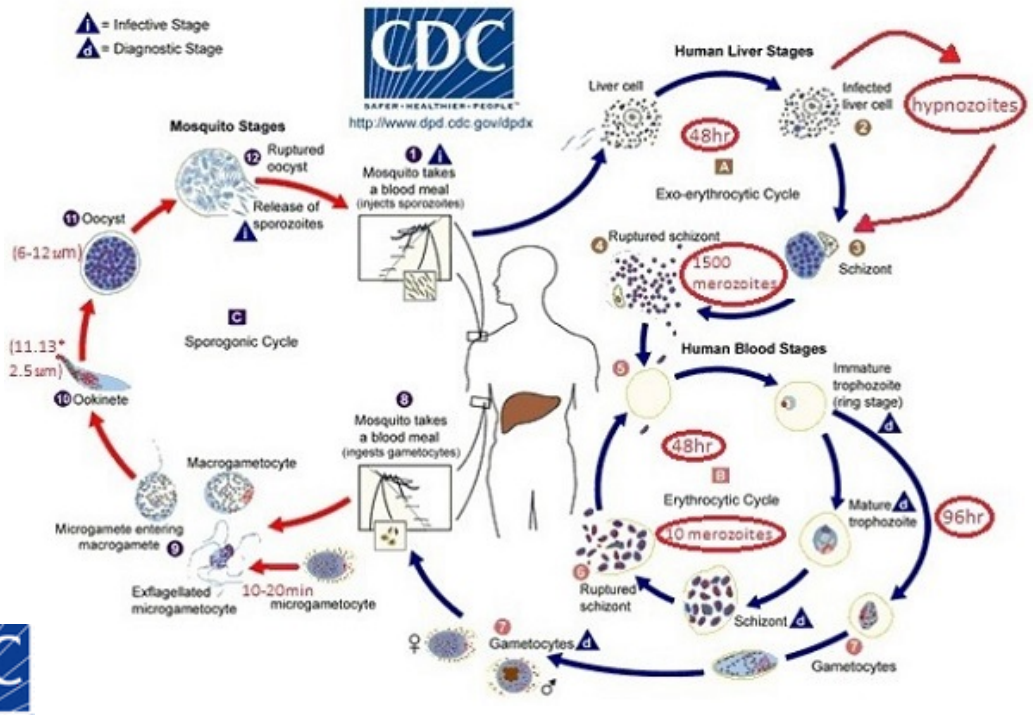


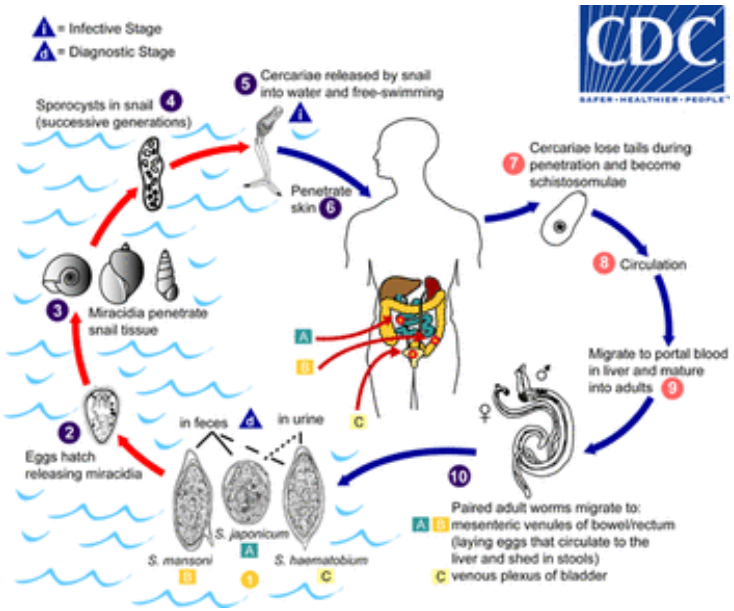
Phenotype

One gene can give rise to many products
- Through RNA processing or post-translational modification

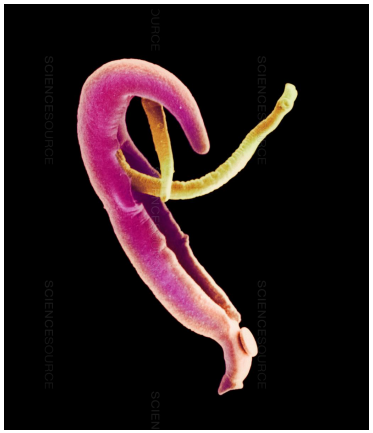


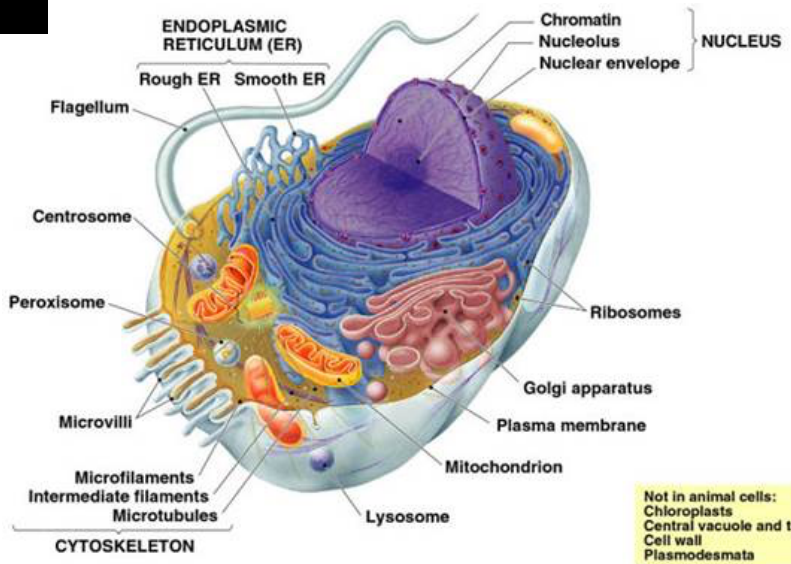
Complex developmental regulation of gene expression is a feature of parasite life cycles

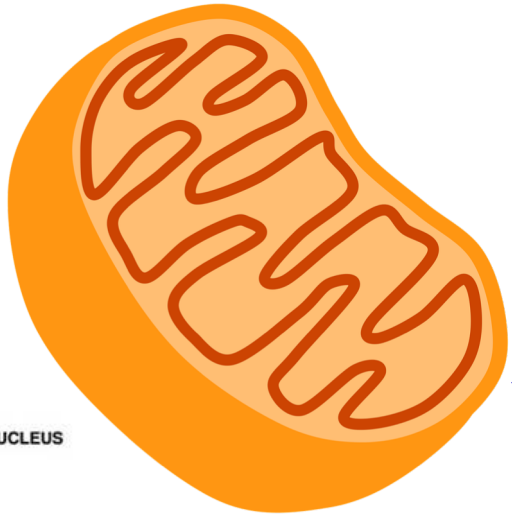


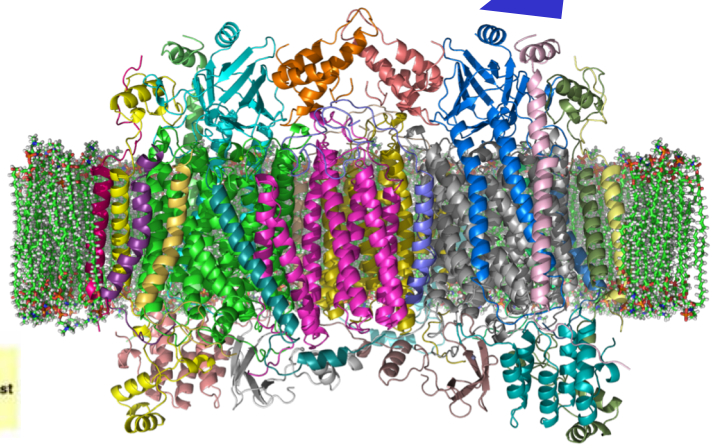


Most ”proteomic” analyses focus on a sub-proteome









Selection of an appropriate sub-proteome requires a hypothesis

How are proteomic analyses performed?

Biochemistry, mass spectrometry and database searching

Protein extraction

- Mechanical methods
  - Bead-beating, sonication, freeze-thaw, homogenizers, cavitation…….
- Detergents or chaotropes
  - Ionic detergents better than non-ionic (more powerful, easier to remove)
  - Chaotropes like urea, guanidinium disrupt hydrogen bonding and aid solubilisation
- A combination is often most effective
  - Avoid protein modification/degradation
  - Consider subsequent steps, especially MS

Protein selection

- **ALL** proteomic analyses are performed on SUB-PROTEOMES
  - Without fractionation, coverage is limited to abundant, soluble proteins
- Targeted analysis of a selected protein fraction
  - IP, affinity purification
  - Subcellular or biochemical fractionation
- Global analysis of a fractionated proteome

Top down or bottom up?

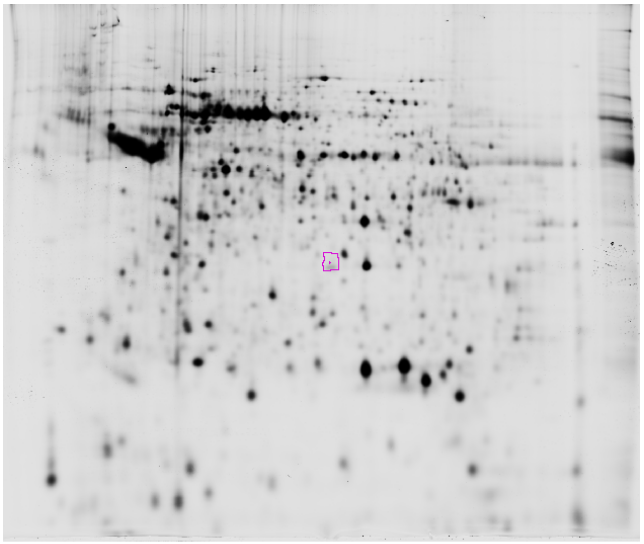
Bottom up

- Digest proteins
- Separate peptides
- MS to identify

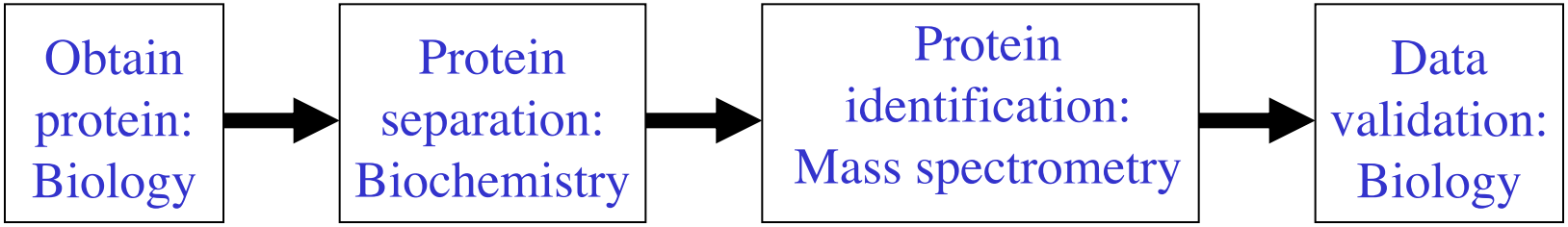
31. SmpA
33. Plp4
34. RcpA
35. TadD
37. LspB_2
39. HasR
40. TonB-dependent protein PM0803
41. TonB-dependent protein PM1428
42. HgbB
43. HgbB
44. Hypothetical protein PM1134
45. Opa
46. HlpB

Top down

- Separate proteins
- Select proteins of interest
- Digest and MS to identify



Proteomics workflow:



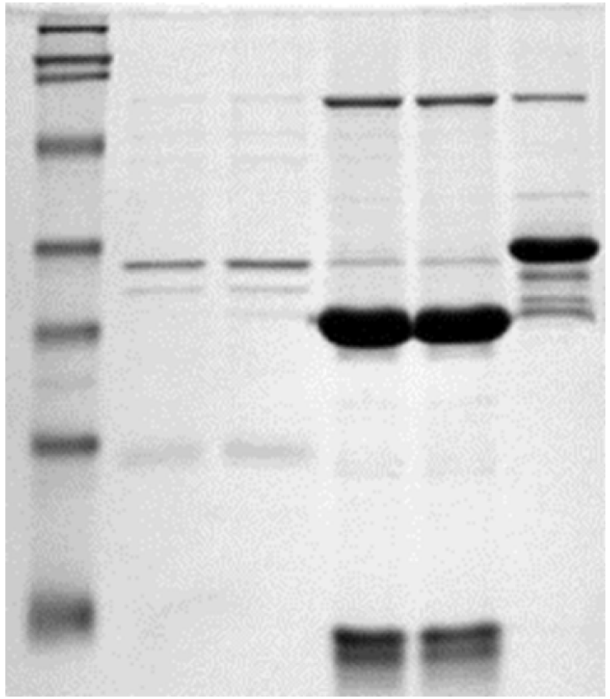
- Exploits a variety of approaches
- Enabled by information technology
- Limited by sensitivity threshold
- Challenged by complexity

The BIG question:

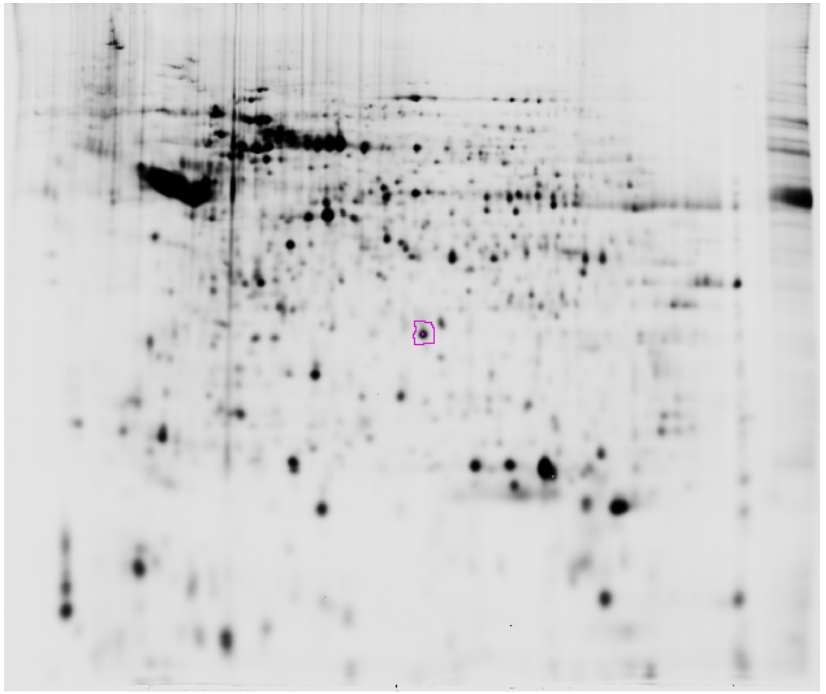
Which proteins are interesting (to you)?

Gel-based “top-down” proteomics

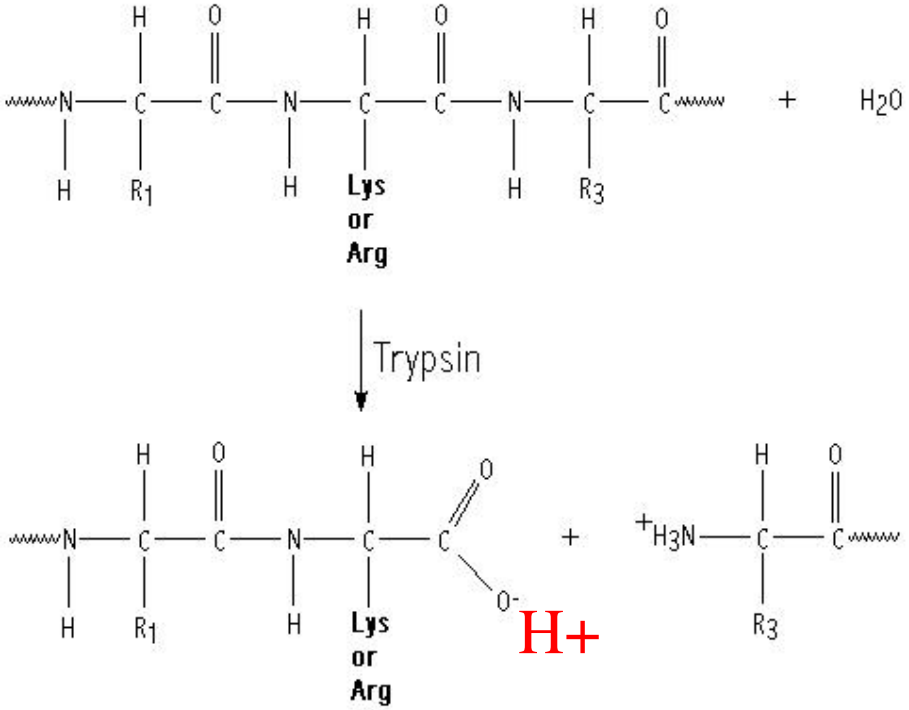
1-dimensional gel



2-dimensional gel



Trypsin cuts at lysine and arginine residues



Digest of a protein with trypsin gives rise to a series of peptides that is diagnostic for the protein and predictable from a genome database

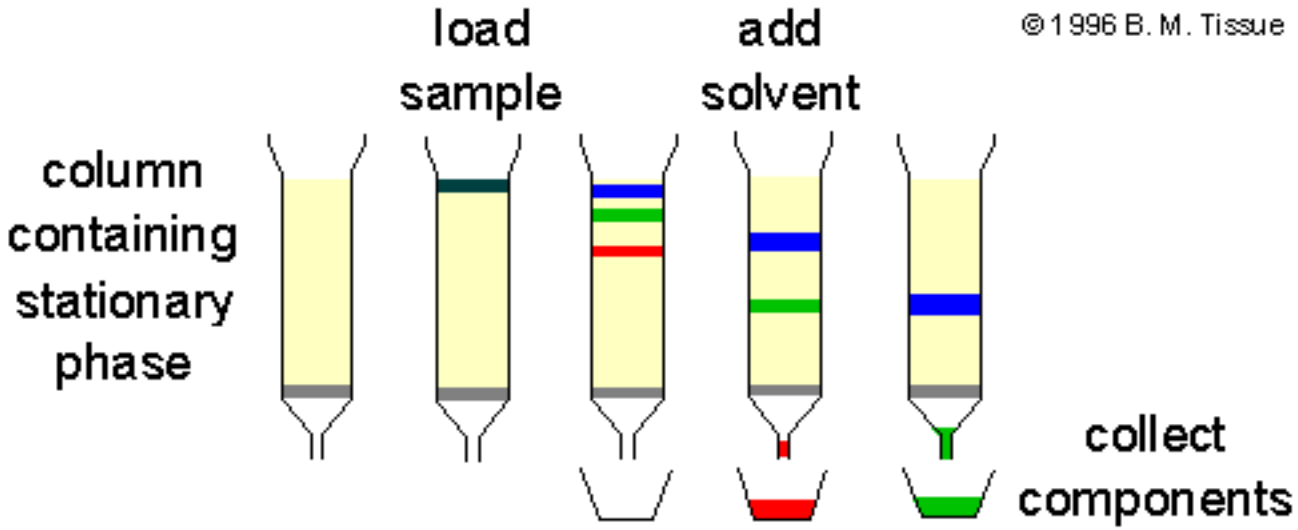
Accurate mass determination for some of these peptides can enable us to infer amino acid composition

Every protein will give a unique peptide mass fingerprint

mass	position	peptide sequence
1607.6788	1-15	MHATAETCETPSSSR
1269.6447	16-25	RPPNDRPDFR
634.3406	26-31	EGSTLK
538.2507	32-35	EFDK

Accurate mass of a peptide provides amino acid composition (but leucine and isoleucine cannot be distinguished by mass)

Liquid chromatography

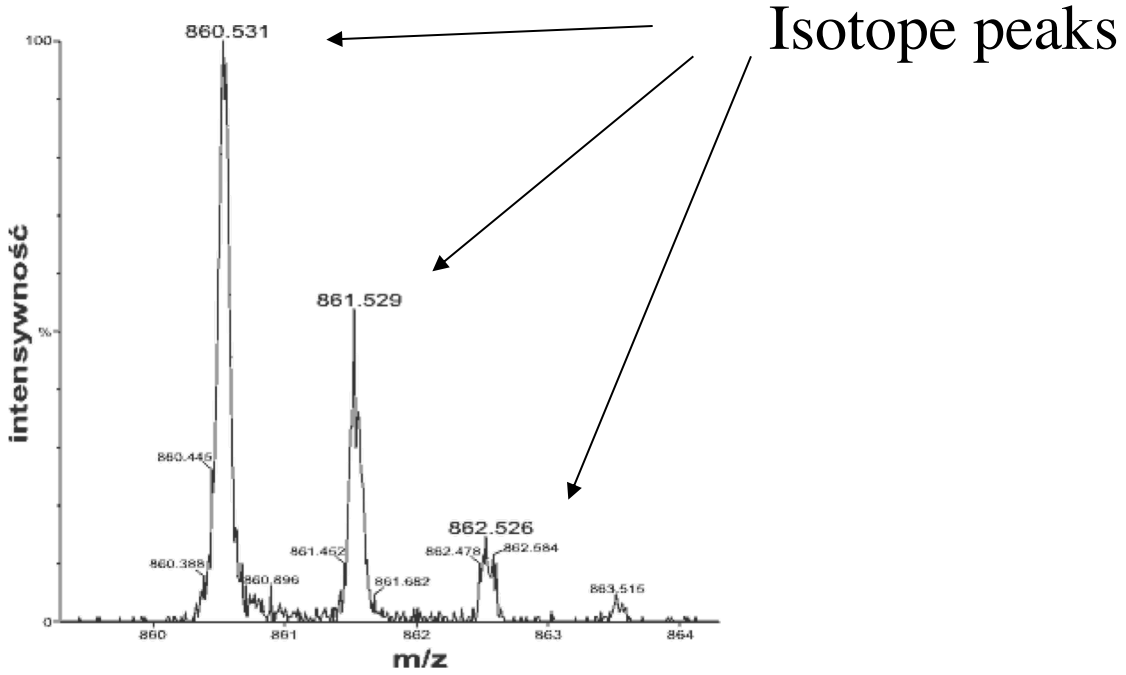


- In proteomics, always “bottom-up” (applied to peptides)
- Can be 2 dimensional for higher resolution
- Can be coupled directly to MS - automation

Nanoflow LC – MS

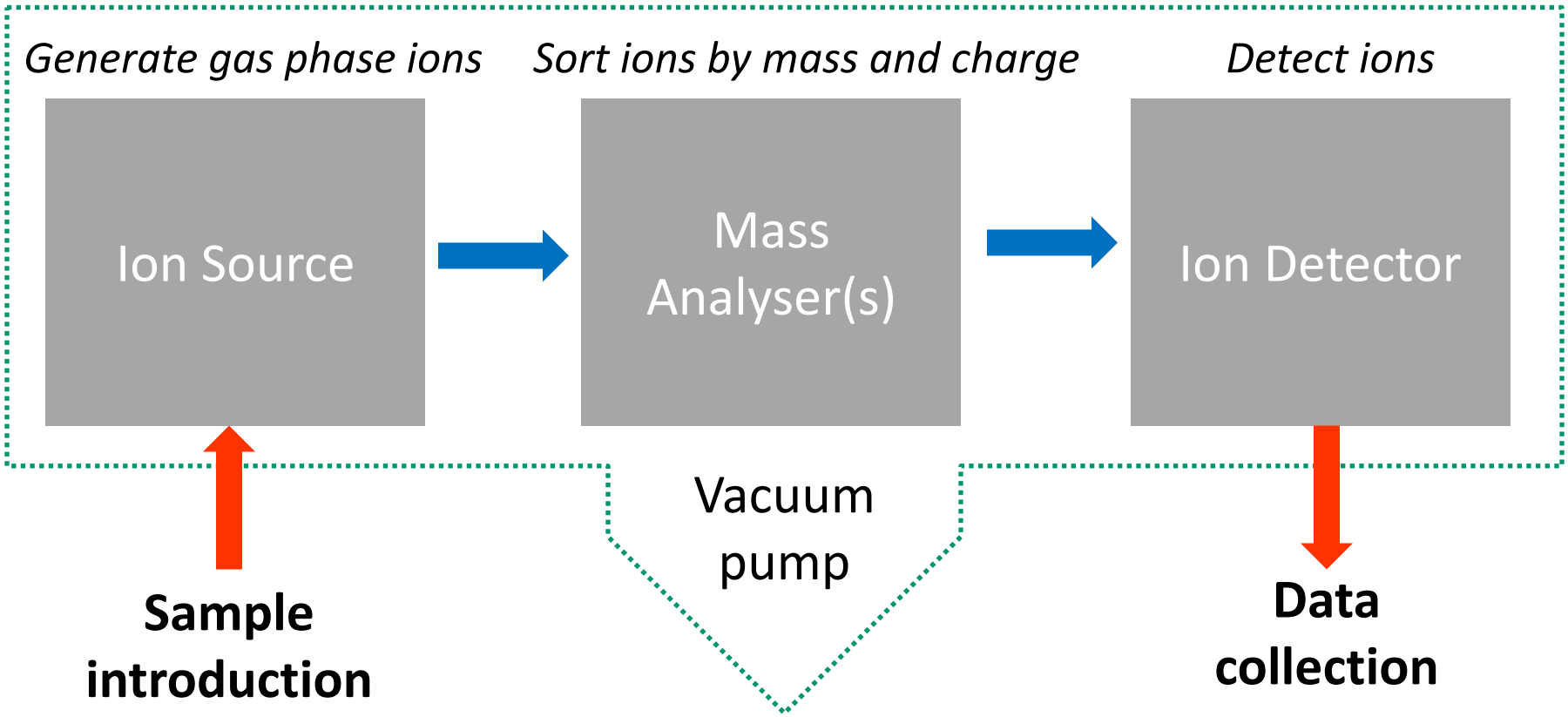
- Reversed phase chromatographic separation of peptides
- Online electrospray ionization of eluate
- MS and MS/MS data collected continuously
- Can be very automated and moderately high-throughput

A mass spectrometer can be a (very accurate) weighing machine



- For proteins, mass accuracy sufficient to infer processing etc
- For metabolites, mass accuracy sufficient to infer formula
- **For peptides, mass accuracy sufficient to infer aa composition**

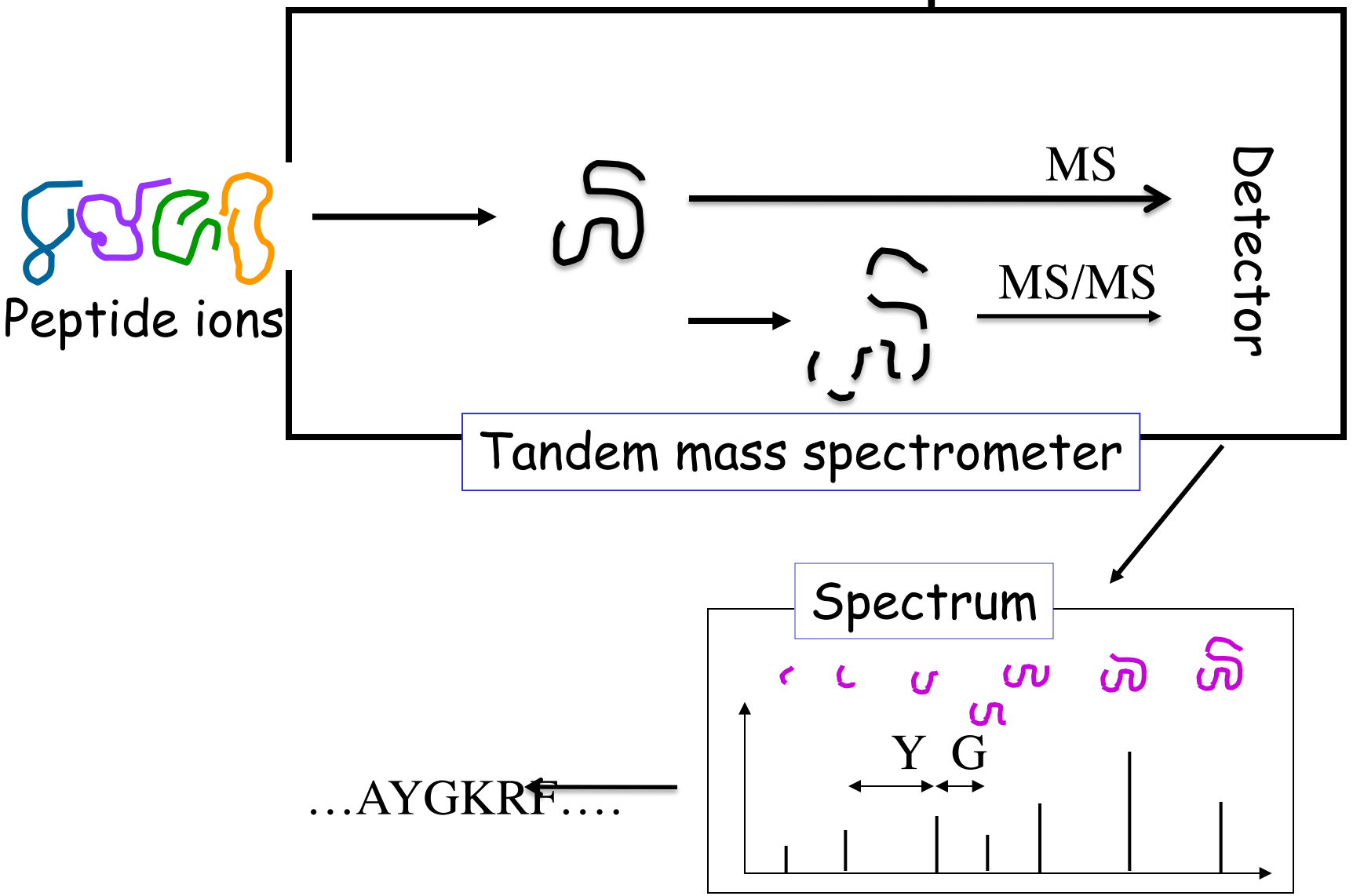
Mass Spectrometers – What's in the box?



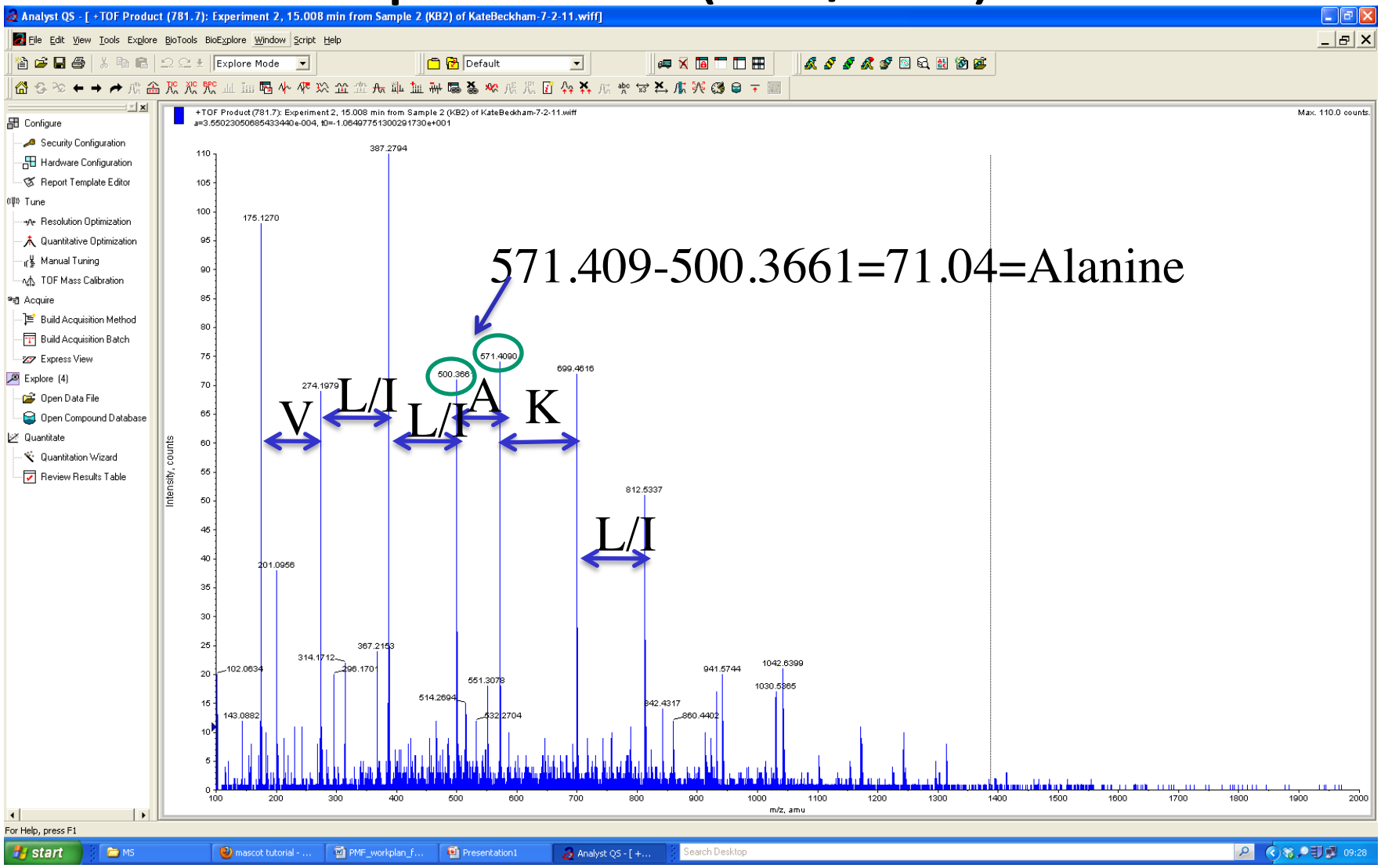
How complex is the analyte?
How can it be ionized?
Is further fragmentation required?
What mass accuracy is required?

What is the question?
How will data be calibrated?
How will data be stored?
How will data be analysed?

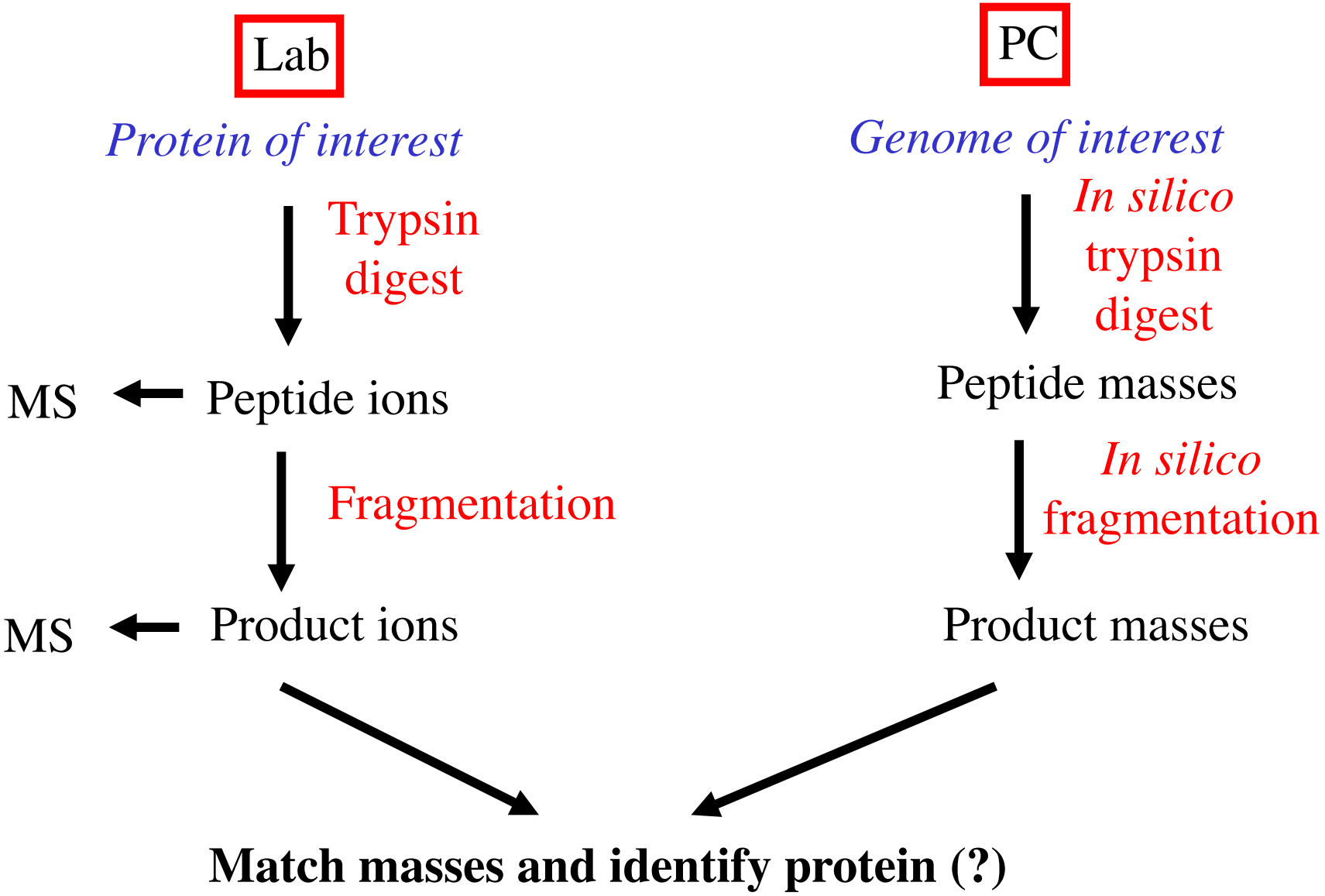
Tandem MS analysis of peptides can infer amino acid sequence



Annotated peptide fragmentation spectrum (MS/MS)

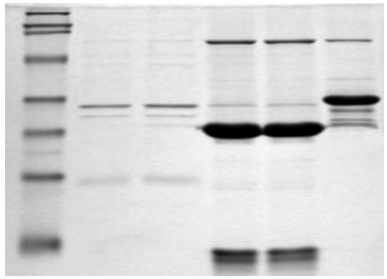


Protein identification using Mascot



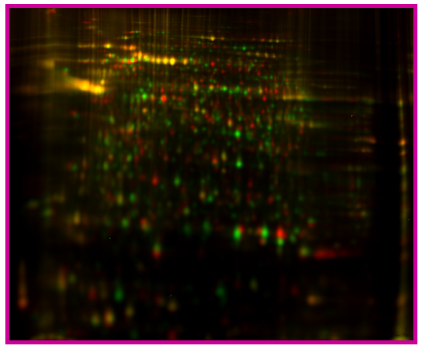
Standard proteomic workflows

GeLC-MS



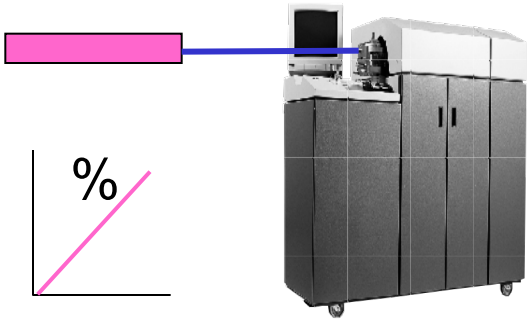
Robust, rapid,
Quantitative
Good for hydrophobic proteins
Low resolution

2-DE

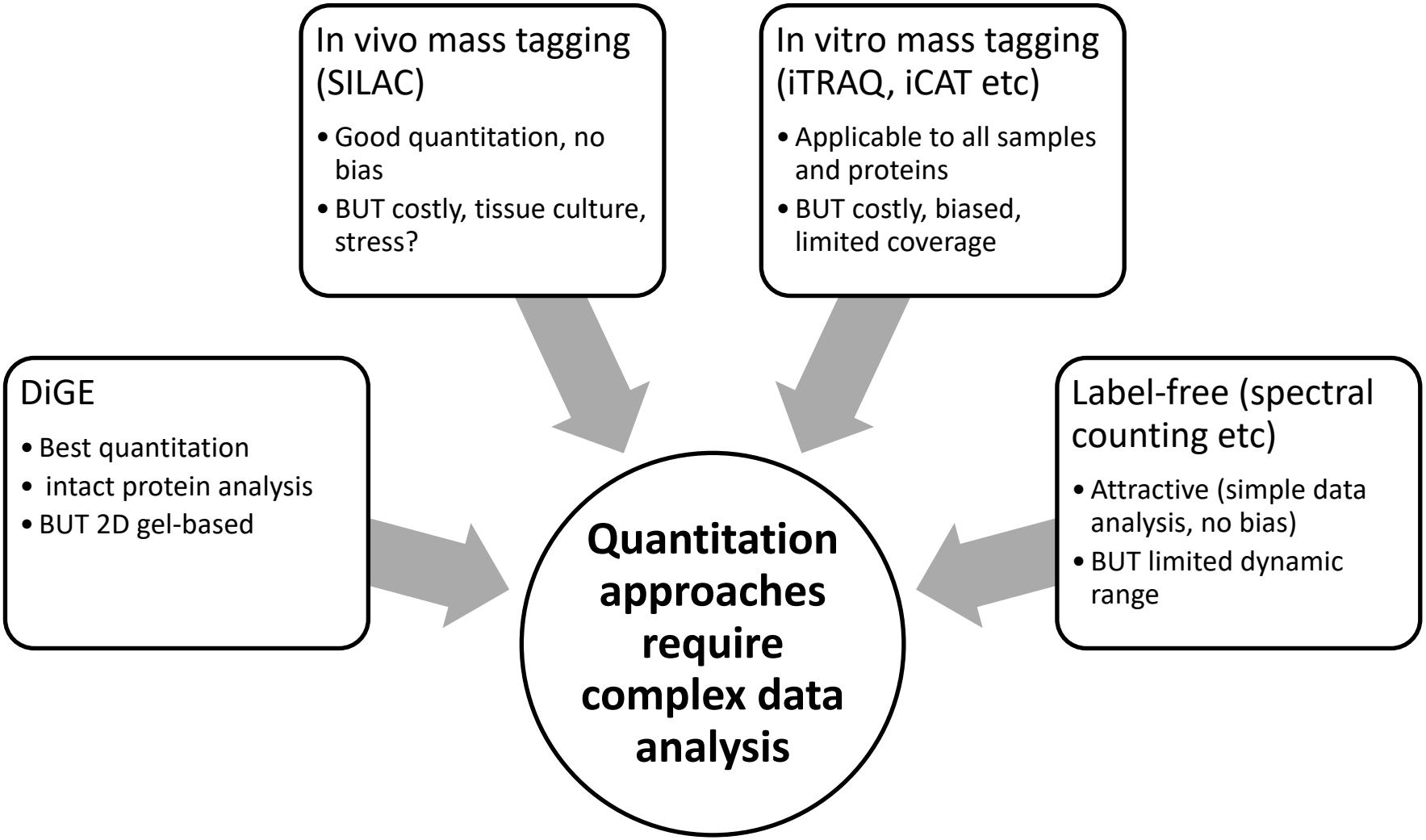


Best resolution
Best quantitation
Poor for hydrophobic proteins
Complex and labour intensive

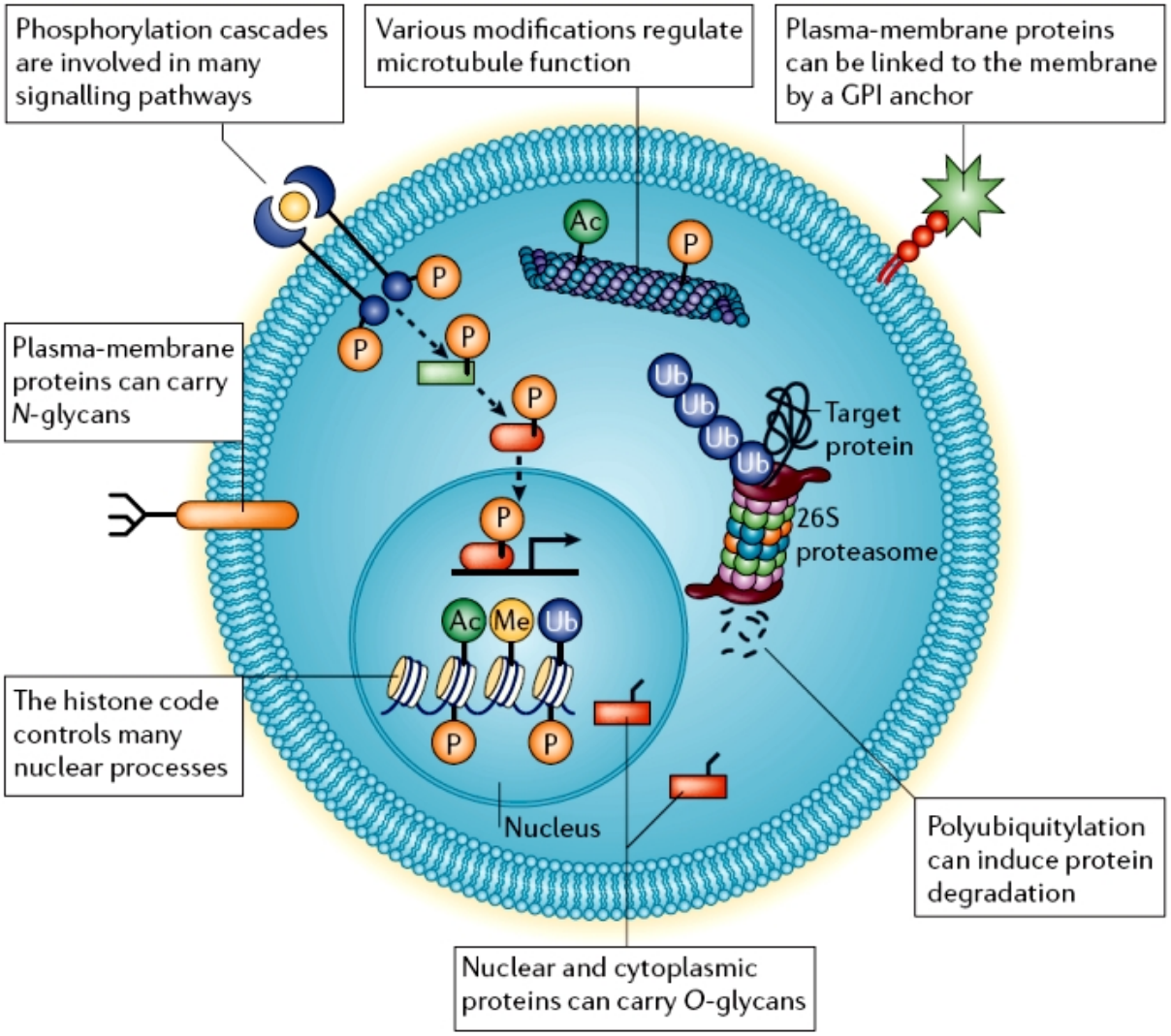
MuDPIT



Good proteomic coverage
Good resolution
Poorly quantitative
Challenging to reproduce


In vivo mass tagging (SILAC)
• Good quantitation, no bias
• BUT costly, tissue culture, stress?
In vitro mass tagging (iTRAQ, iCAT etc)
• Applicable to all samples and proteins
• BUT costly, biased, limited coverage
DiGE
• Best quantitation
• intact protein analysis
• BUT 2D gel-based
Label-free (spectral counting etc)
• Attractive (simple data analysis, no bias)
• BUT limited dynamic range
Quantitation approaches require complex data analysis

Characterisation of post-translational modifications



- PTMs can be:
  - diverse
  - substoichiometric
  - Unpredictable
  - Heterogeneous
  - Labile
  - Artifactual
  - *Important!*
- Post-translational modifications are very challenging to investigate.
- Proteomic workflows have been developed for many specific modifications

http://www.nature.com/scitable/nated/content/ne0000/ne0000/ne0000/ne0000/99009/10.1038_nrm1939-f1_full.jpg

Downstream of the proteome

- Metabolomics, lipidomics
  - Also enabled by mass spectrometry (and NMR)
  - No link back to genome, so identification is based on characterisation (generally by mass, often also by chemistry)
  - Can cut to the chase (metabolic markers)
- Omic technologies link to phenotypic analysis. Validation of omic results is critical and can be very informative

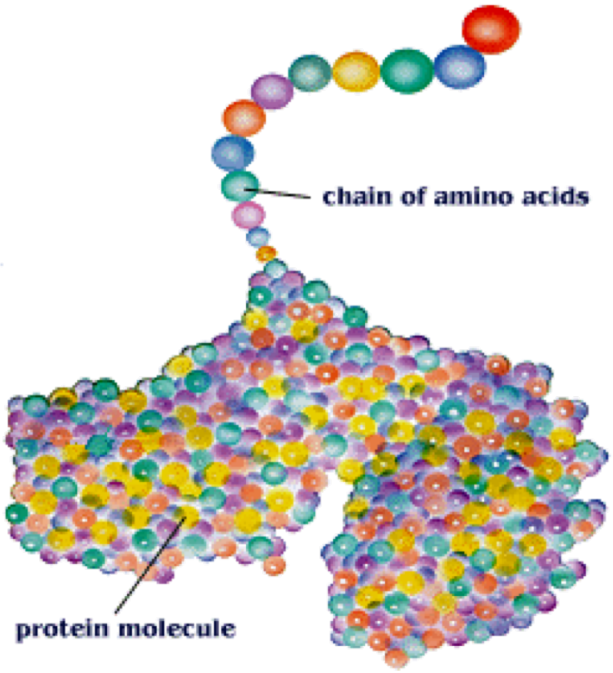
What is a metabolome?

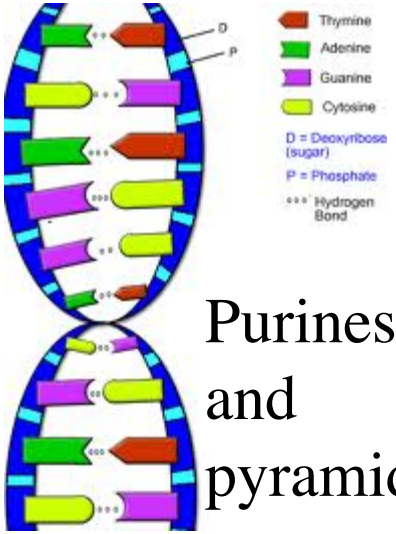
- The small molecule complement of a biological sample
- Which can be: tissue, cell culture, bacterial culture, secretion, etc.
- Products of anabolism, catabolism
- Secondary metabolites, primary metabolites

What is metabolomics?

- Analysis of the complete metabolome
- Tied to specific state
  - Time after drug treatment 0 – 24h
  - Healthy/Diseased
  - Wild-type/Mutant
  - Different diets
  - Stem cell/differentiated cell
  - Biofilm/planktonic bacteria

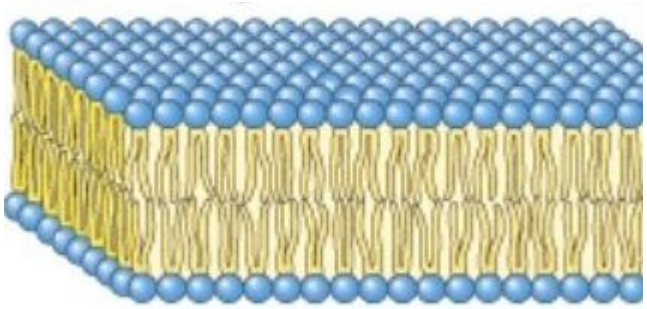
Metabolites



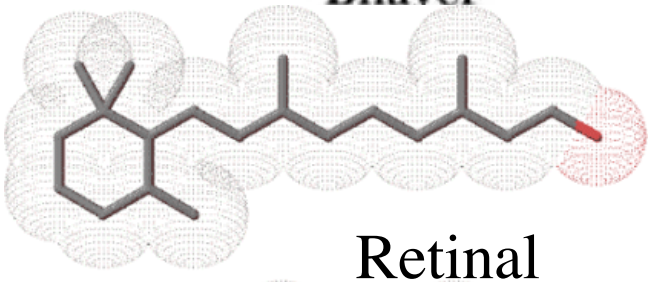


Purines and pyramidines

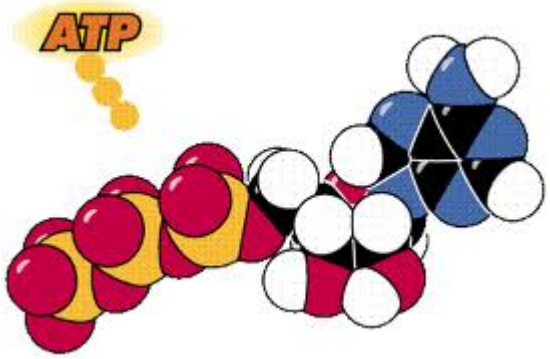
Fatty acids

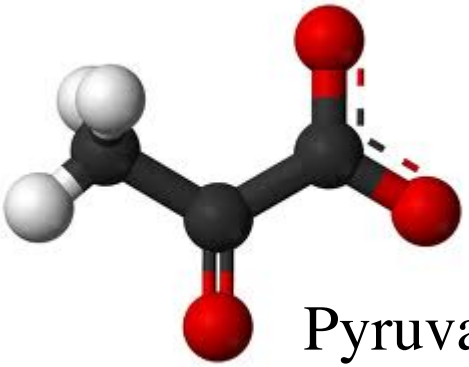


Bilaver

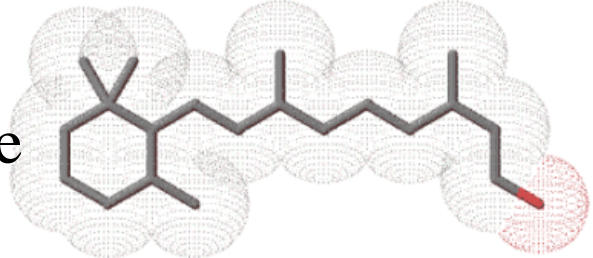


Retinal

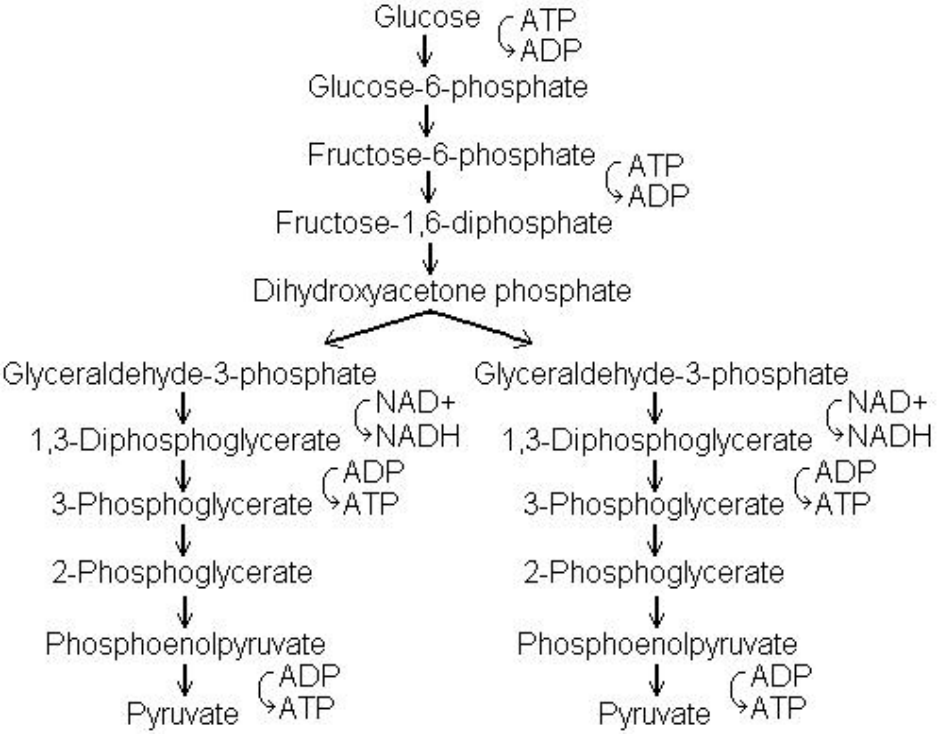


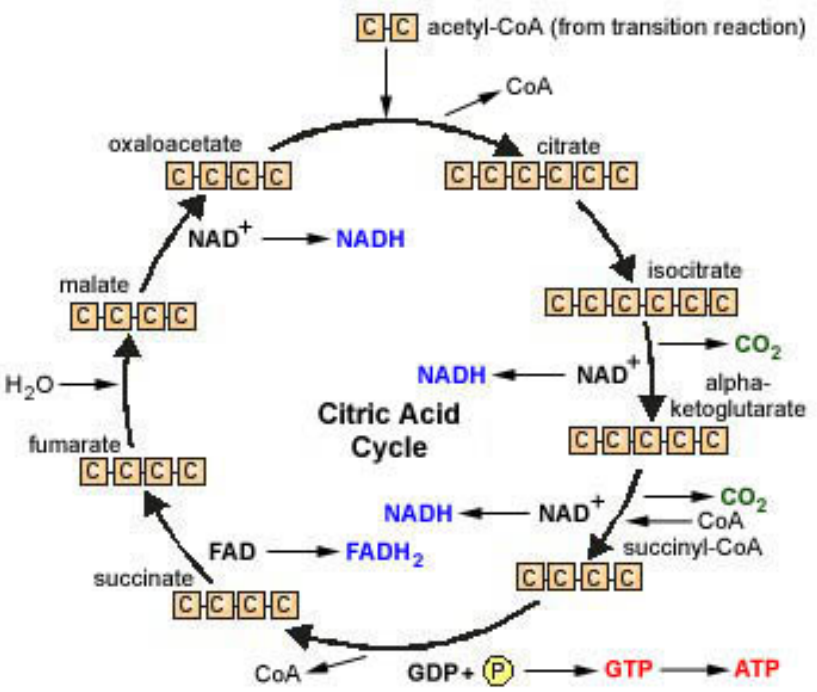


Pyruvate



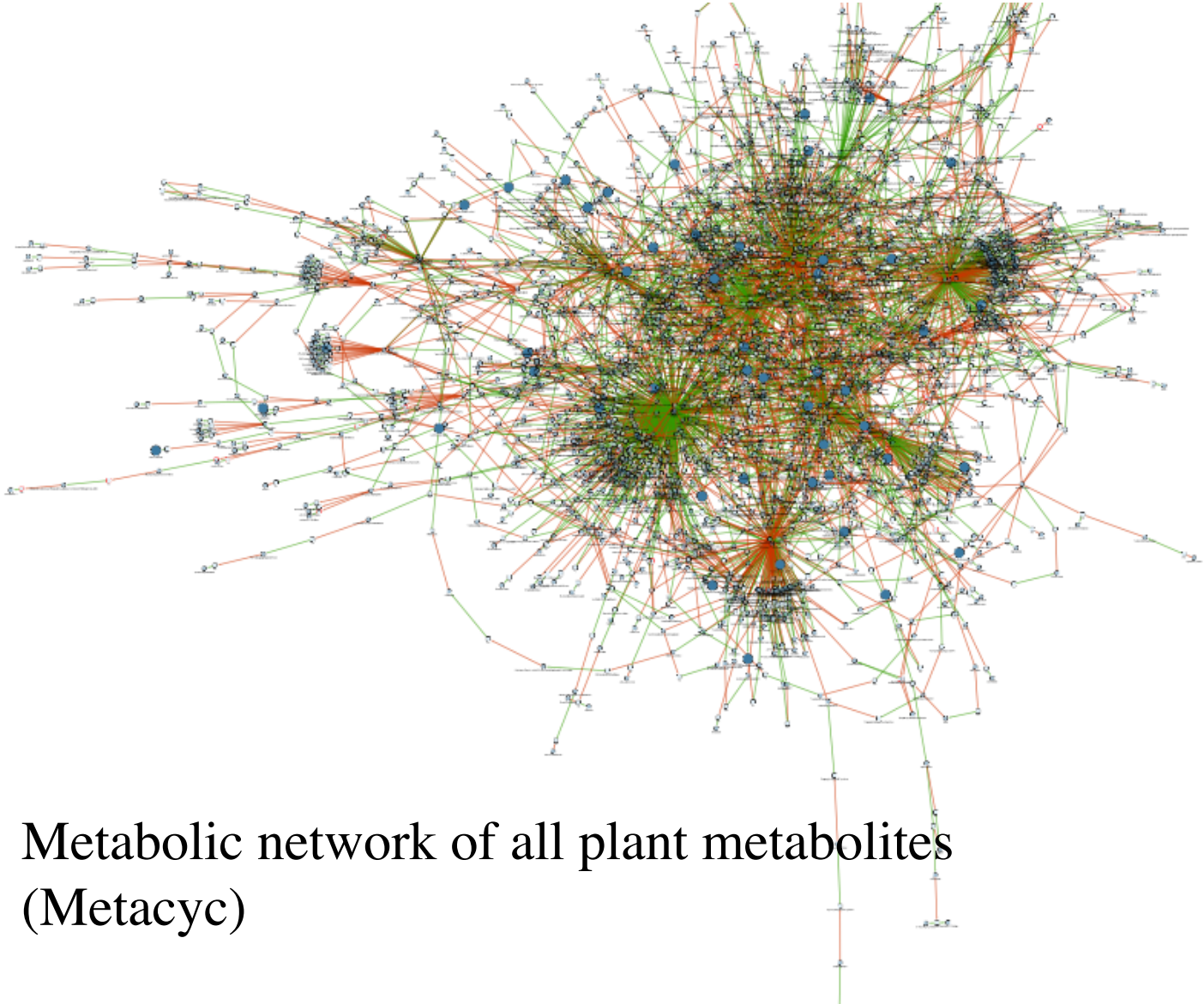
Metabolic Pathways





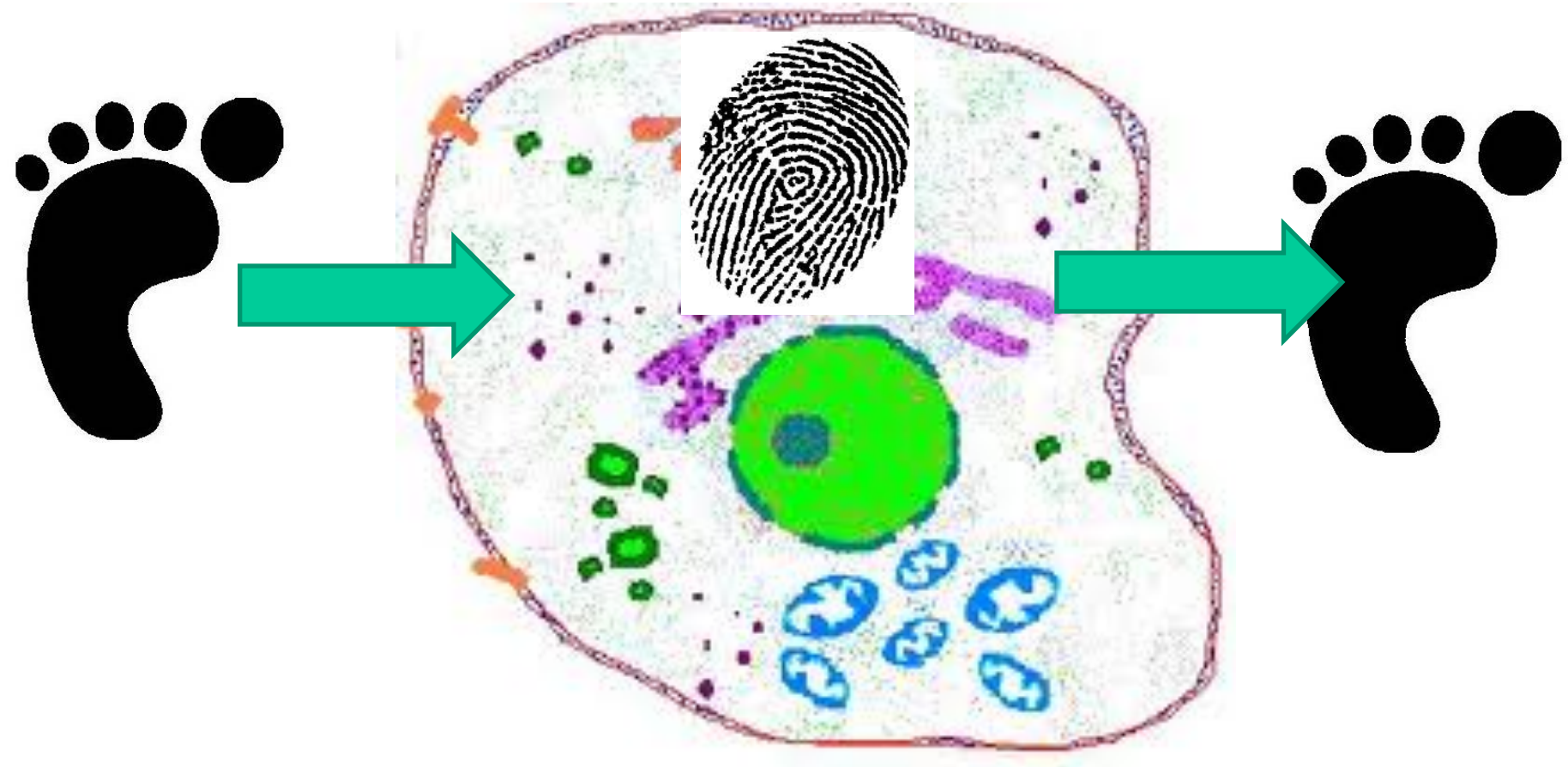
Metabolites are not in isolation

Metabolic networks



Metabolic network of all plant metabolites (Metacyc)

Metabolic Footprinting and Fingerprinting



How to do metabolomics

- Prepare samples
- Analyse by gas or liquid chromatography - mass spectrometry or nuclear magnetic resonance (and other minor methods)
- Collate data and perform quantitation
- Interpret in relation to biological knowledge