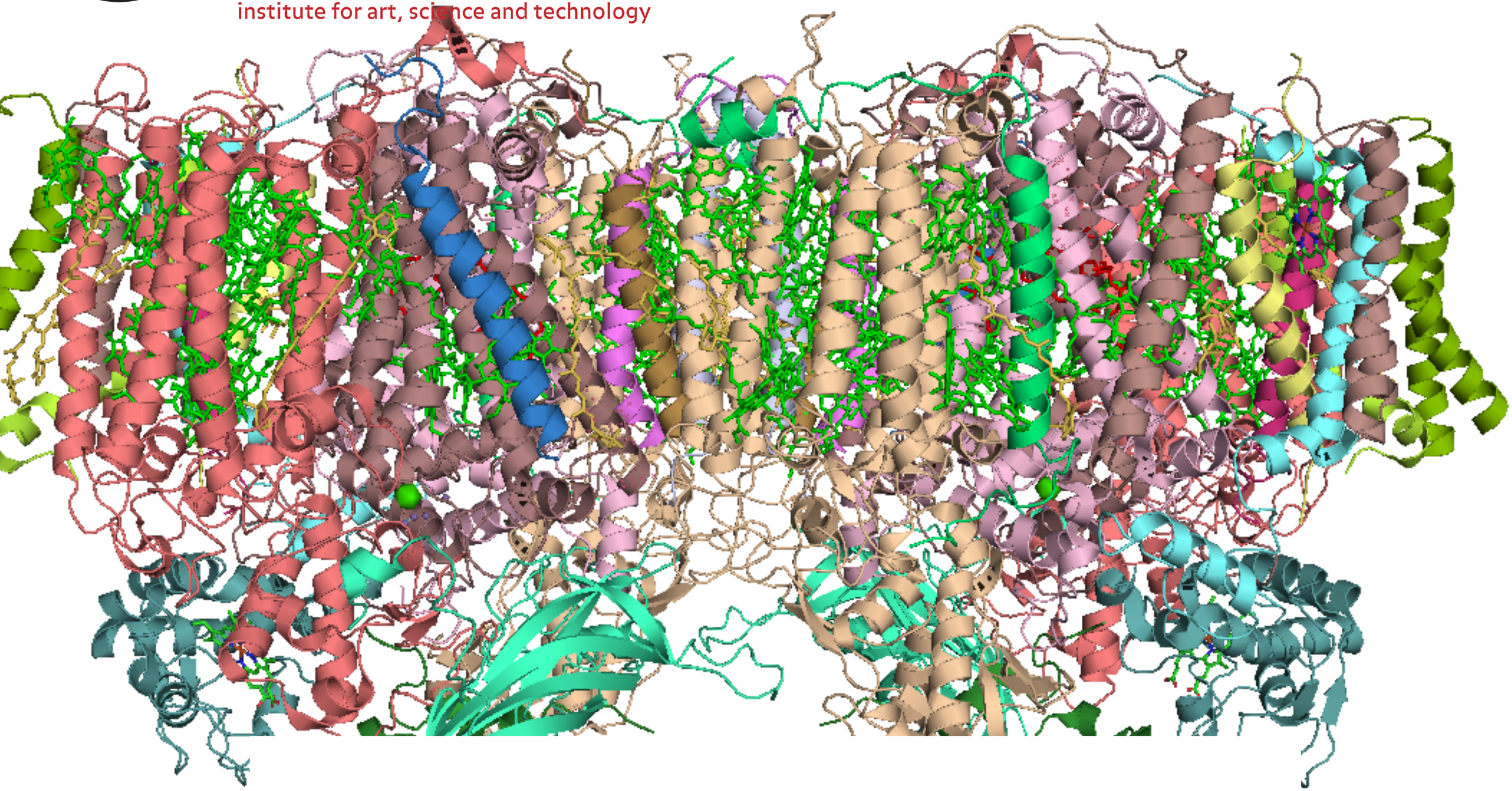
waag society

institute for art, science and technology



Bio Informatics

The information / “Omics” age

- “Genomics” DNA sequence analysis
- “Transcriptomics” DNA expression analysis
- “Proteomics” Protein (structure) prediction / analysis
- “Interactomics” Protein – Protein, DNA – Protein interaction
- “Metabolomics” Metabolism modeling

What is it used for

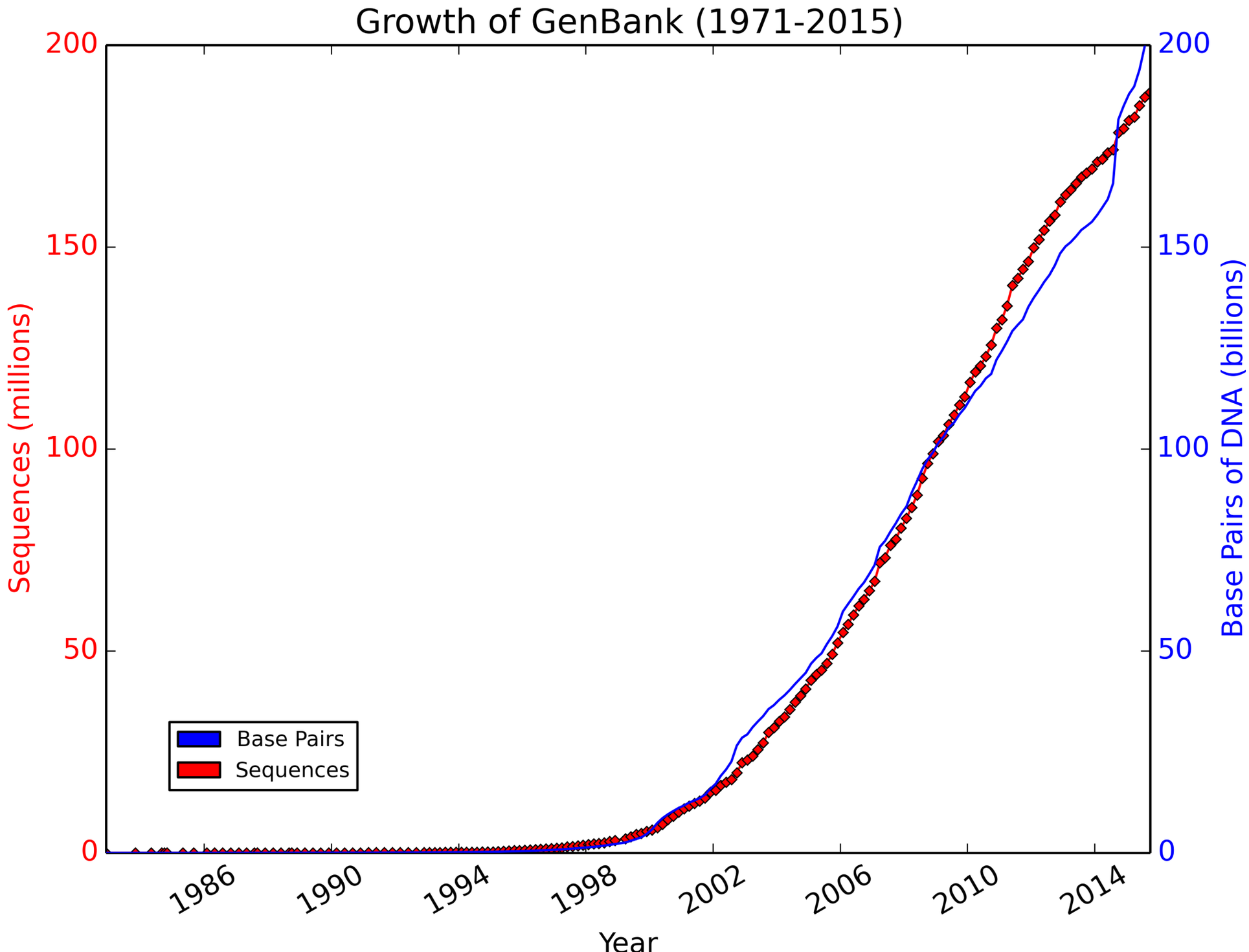
- Optimizing yield
- Predicting organisms behaviour
- Medical diagnostics
  - Personal medicine
- Drug discovery

OpenWorm

Genomics

- Functional genomics
- Metagenomics
- Personal Genomics
- Epigenomics

DNA database GenBank

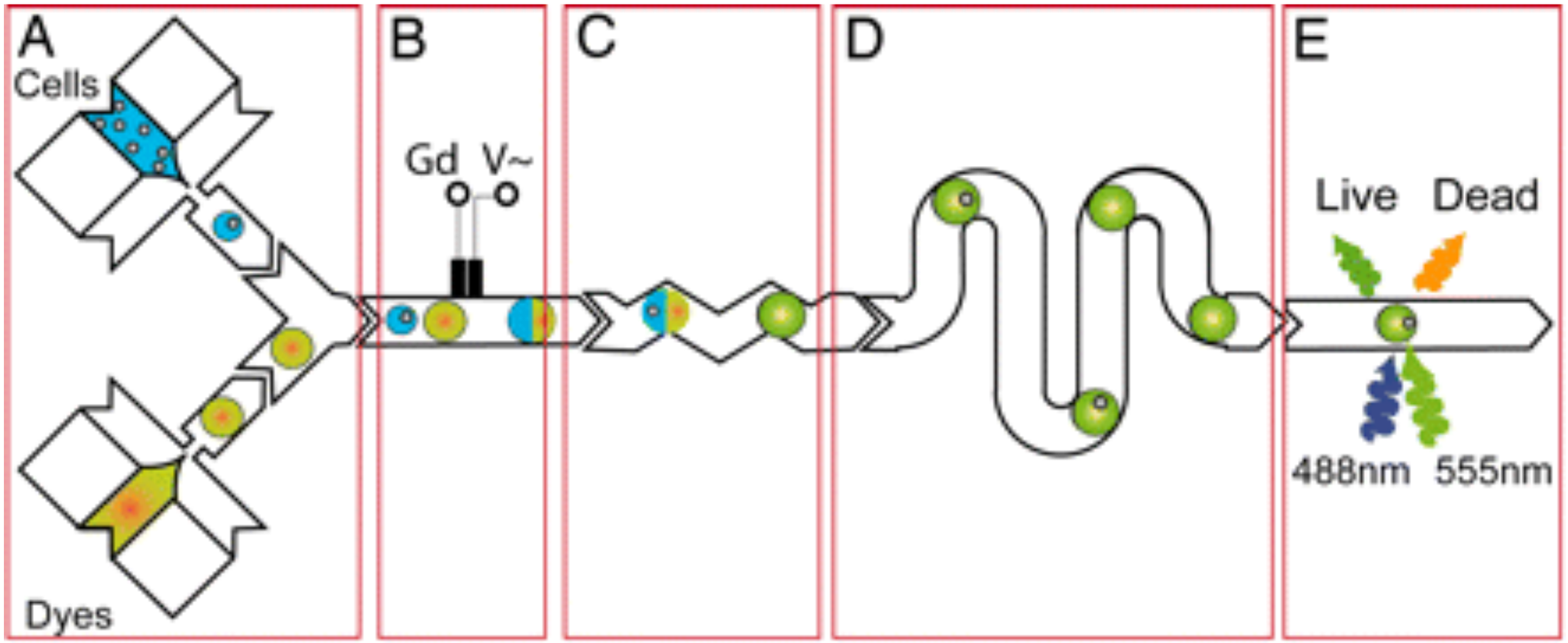


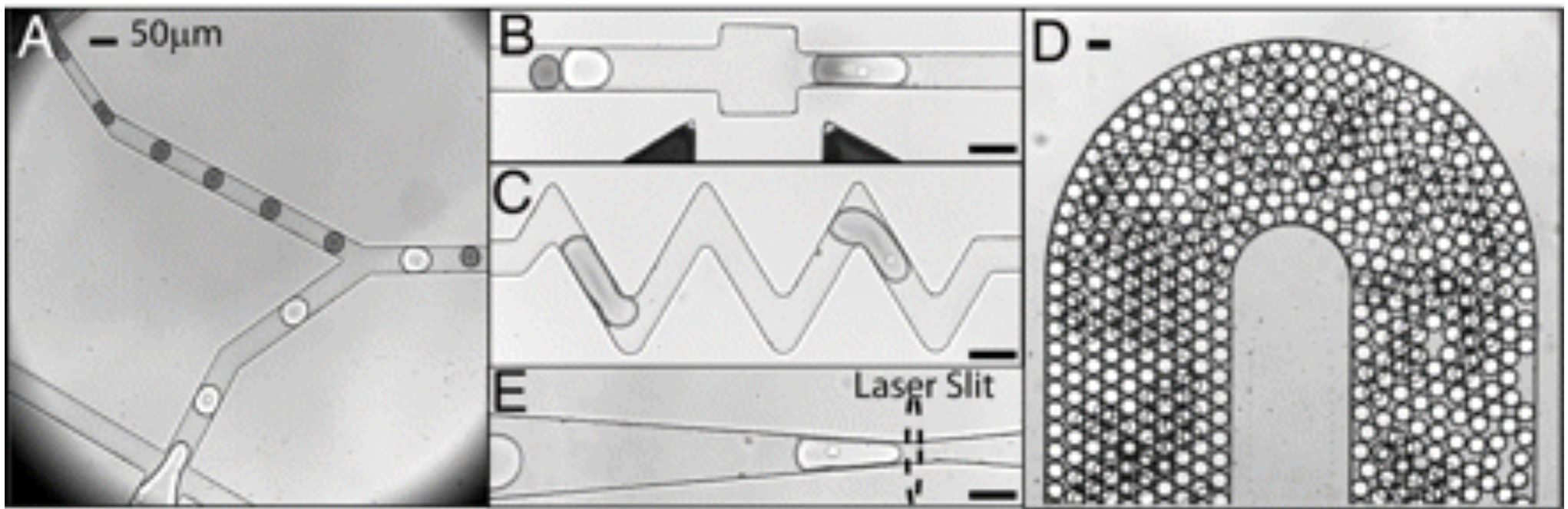
Drivers

- “High Throughput Research”
  - Robotics
  - Databases
  - Visualisation
- Public tools
- Open data

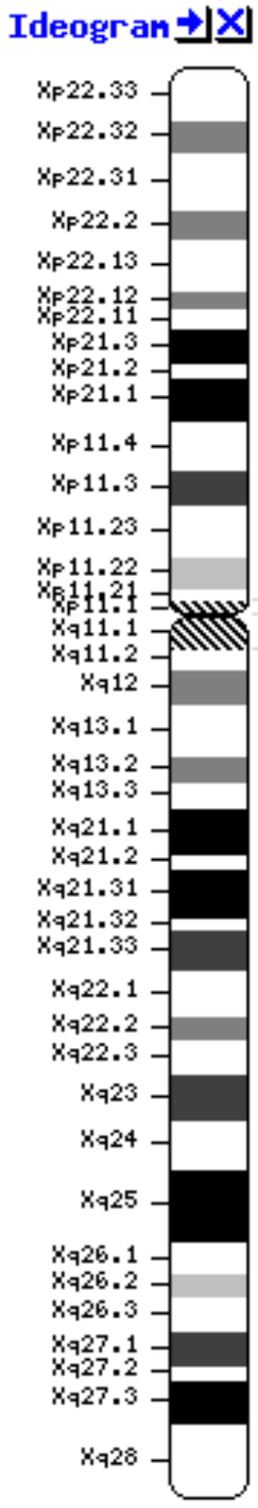
High Throughput Screening

Eric Brouzes et al. PNAS 2009;106:14195-14200

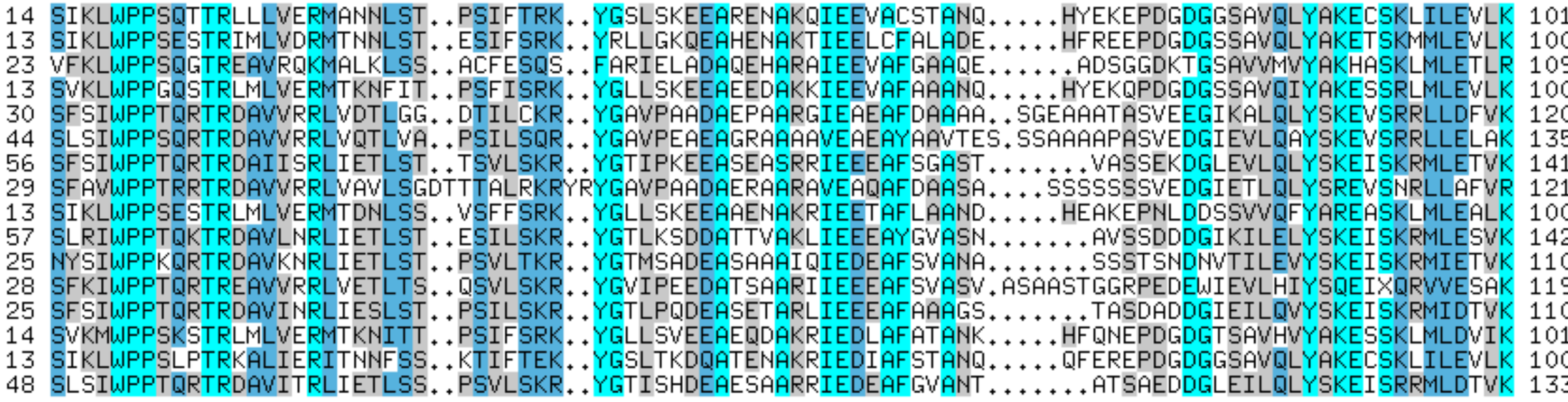




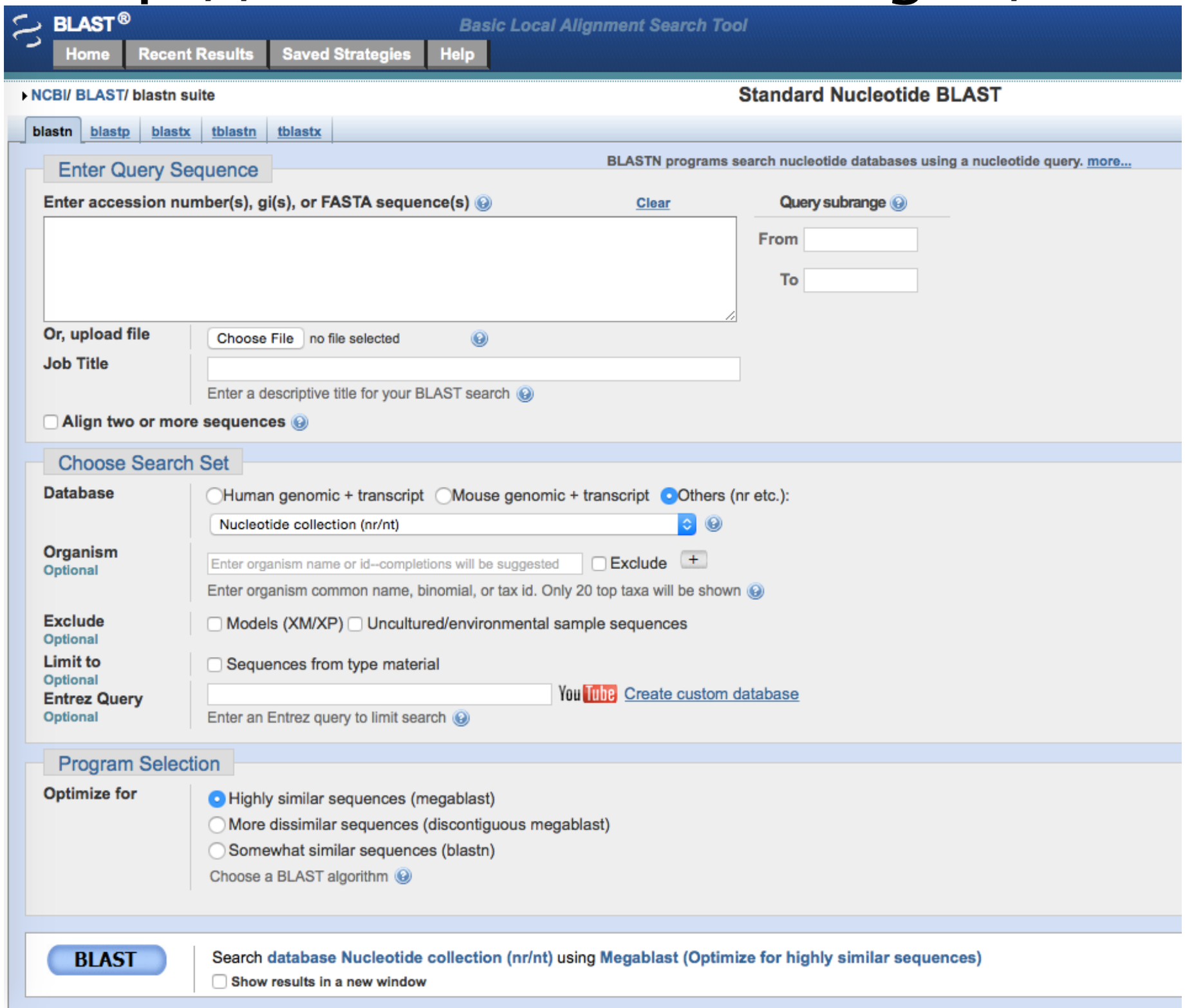
Gene annotations



Sequence Alignment



BLAST: Basic Local Alignment Search Tool
http://blast.ncbi.nlm.nih.gov/Blast.cgi

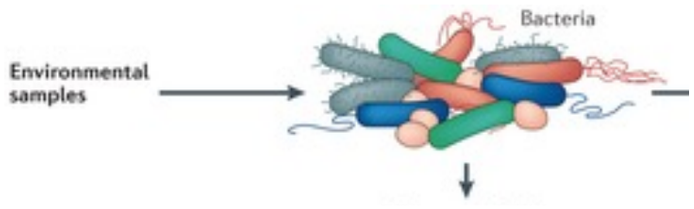


Scoring Matrix BLOSUM

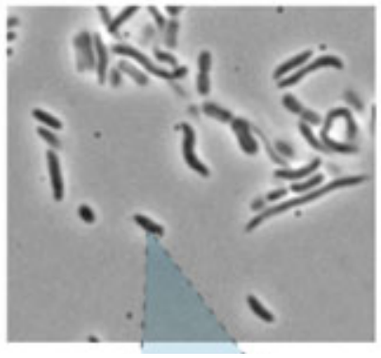
(**BLO**cks **SU**bstitution **M**atrix)

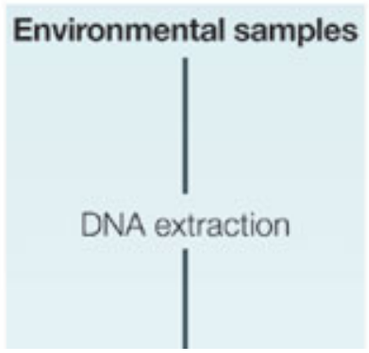
	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	4																			
Arg	-1	5																		
Asn	-2	0	6																	
Asp	-2	-2	1	6																
Cys	0	-3	-3	-3	9															
Gln	-1	1	0	0	-3	5														
Glu	-1	0	0	2	-4	2	5													
Gly	0	-2	0	-1	-3	-2	-2	6												
His	-2	0	1	-1	-3	0	0	-2	8											
Ile	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
Leu	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
Lys	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
Met	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
Phe	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
Pro	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
Ser	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
Thr	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
Trp	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Tyr	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
Val	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

Environmental DNA analysis

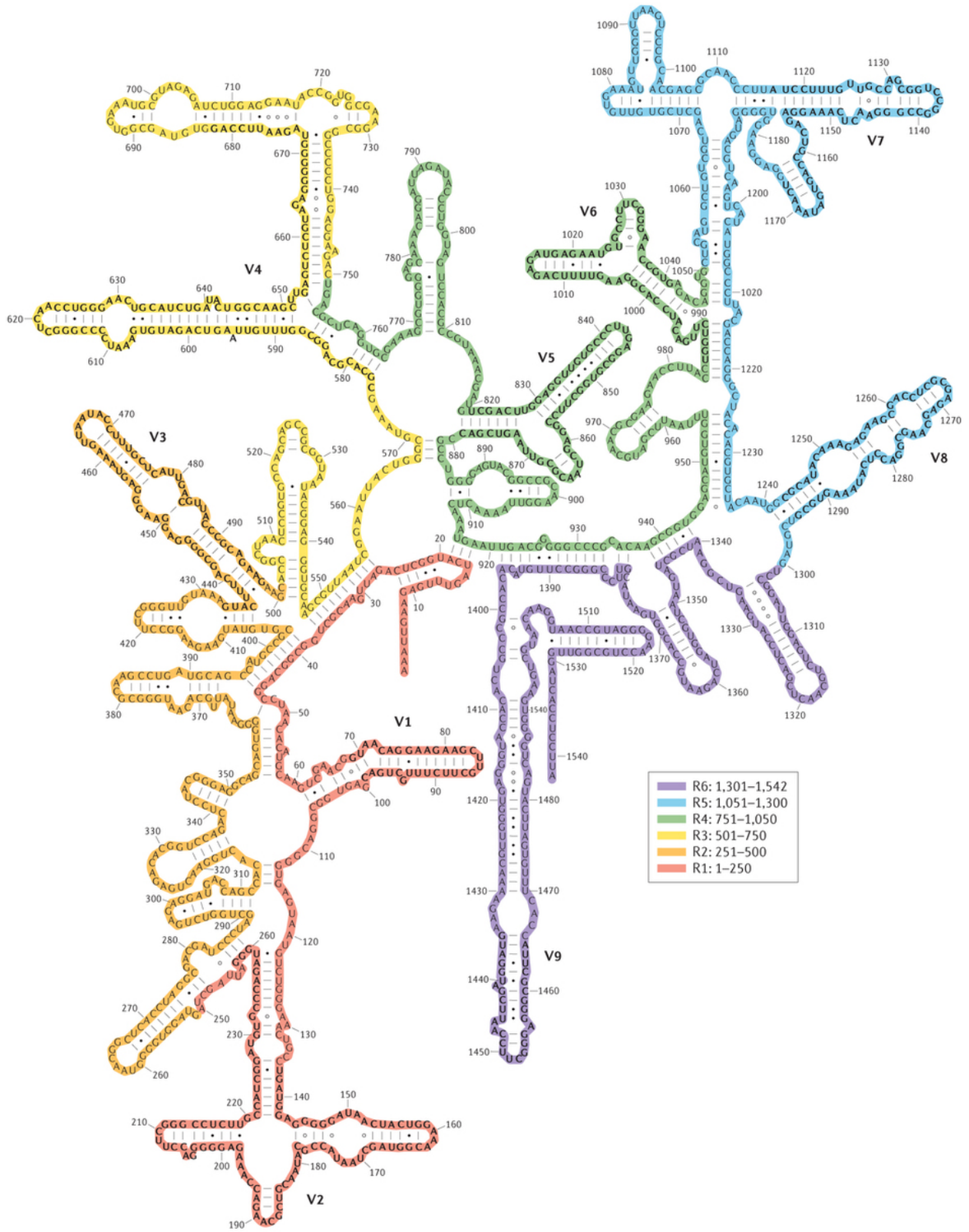


16S RNA





16S RNA molecule

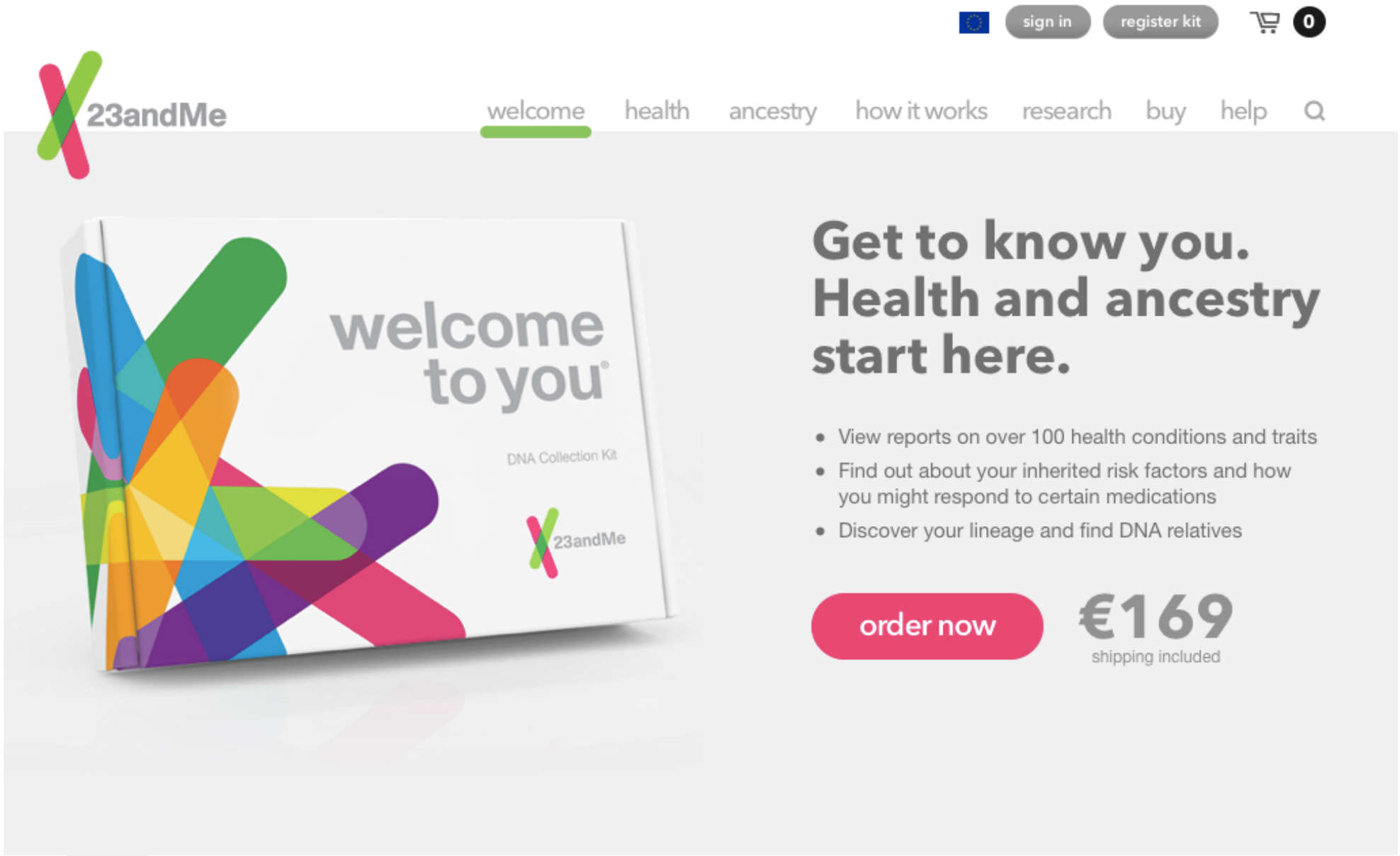


Nature Reviews | Microbiology

Challenges

- Cross linking data / Data mining
  - Relate Genomics to Transcriptomics, Proteomics
  - Relate to structure
  - Relate to disease

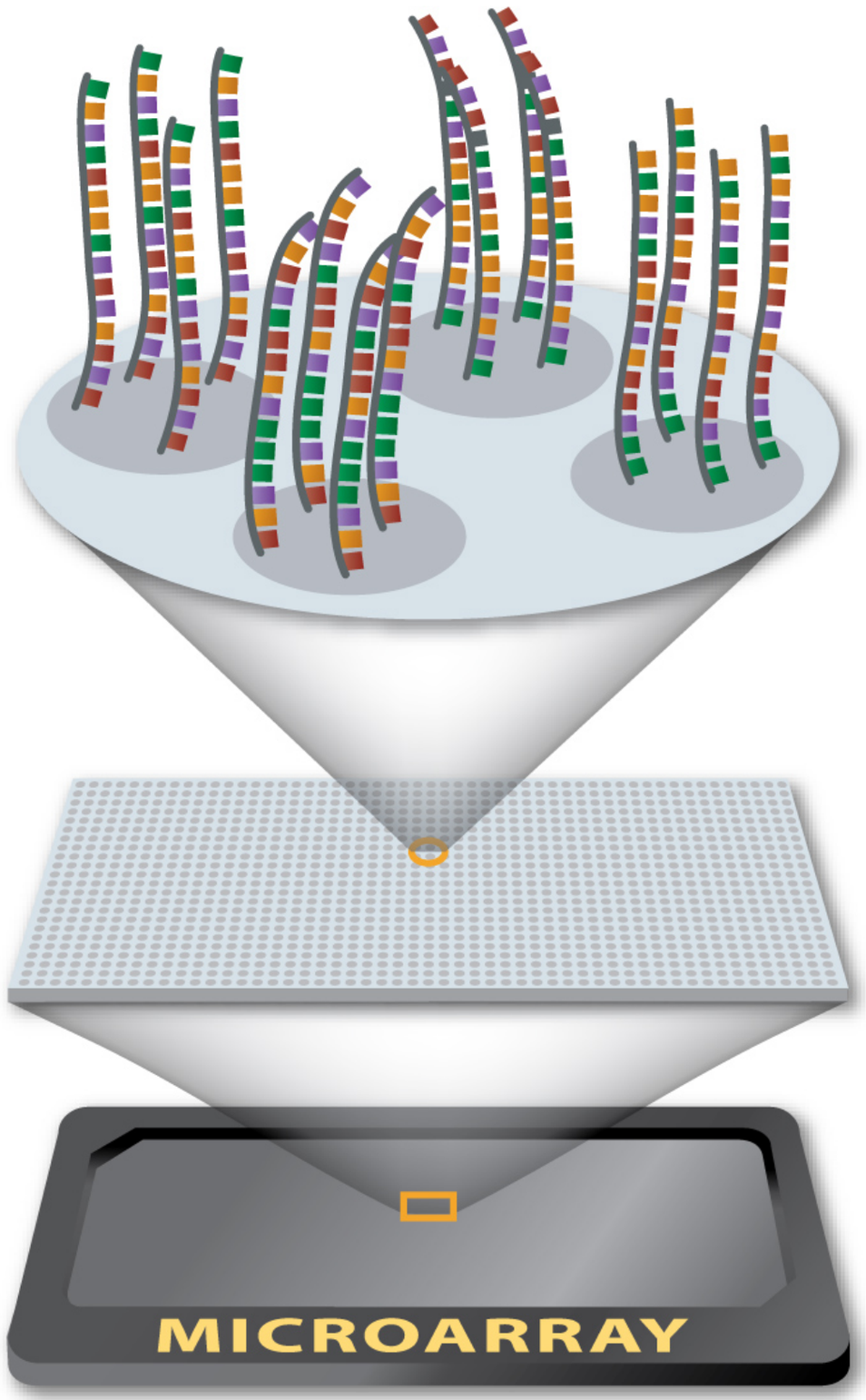
Bio Informatics for the public



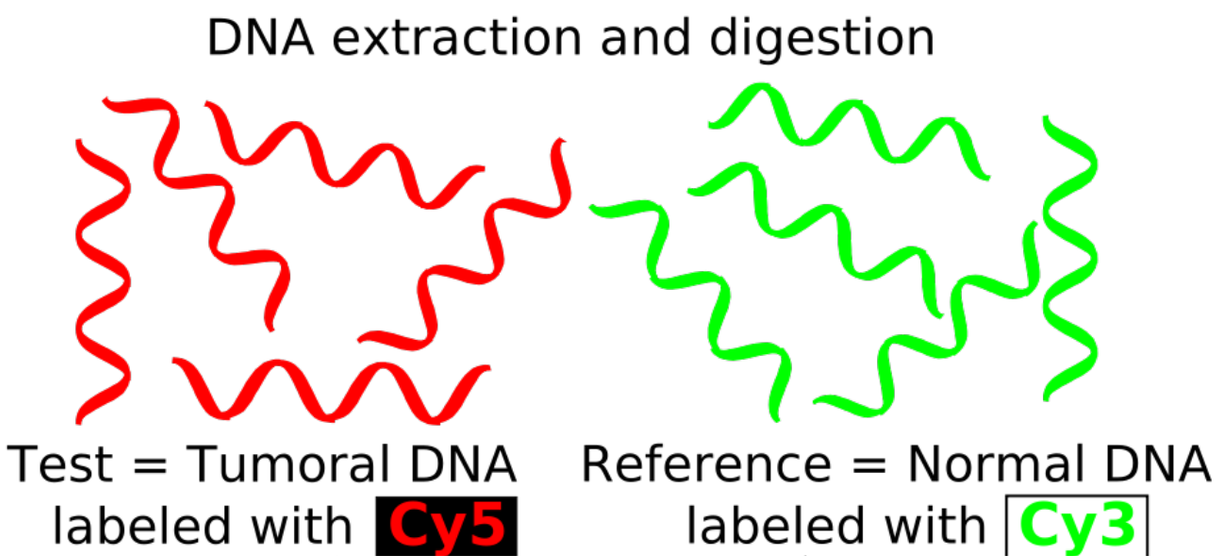
What your DNA says about you.

Find out how your genetics relate to things like abnormal blood clotting, cystic fibrosis or response to certain medications. You can also see if your body metabolises caffeine quickly or if you're likely lactose intolerant. We believe the more you know about

DNA Microarray



DNA Microarray Expression Analysis

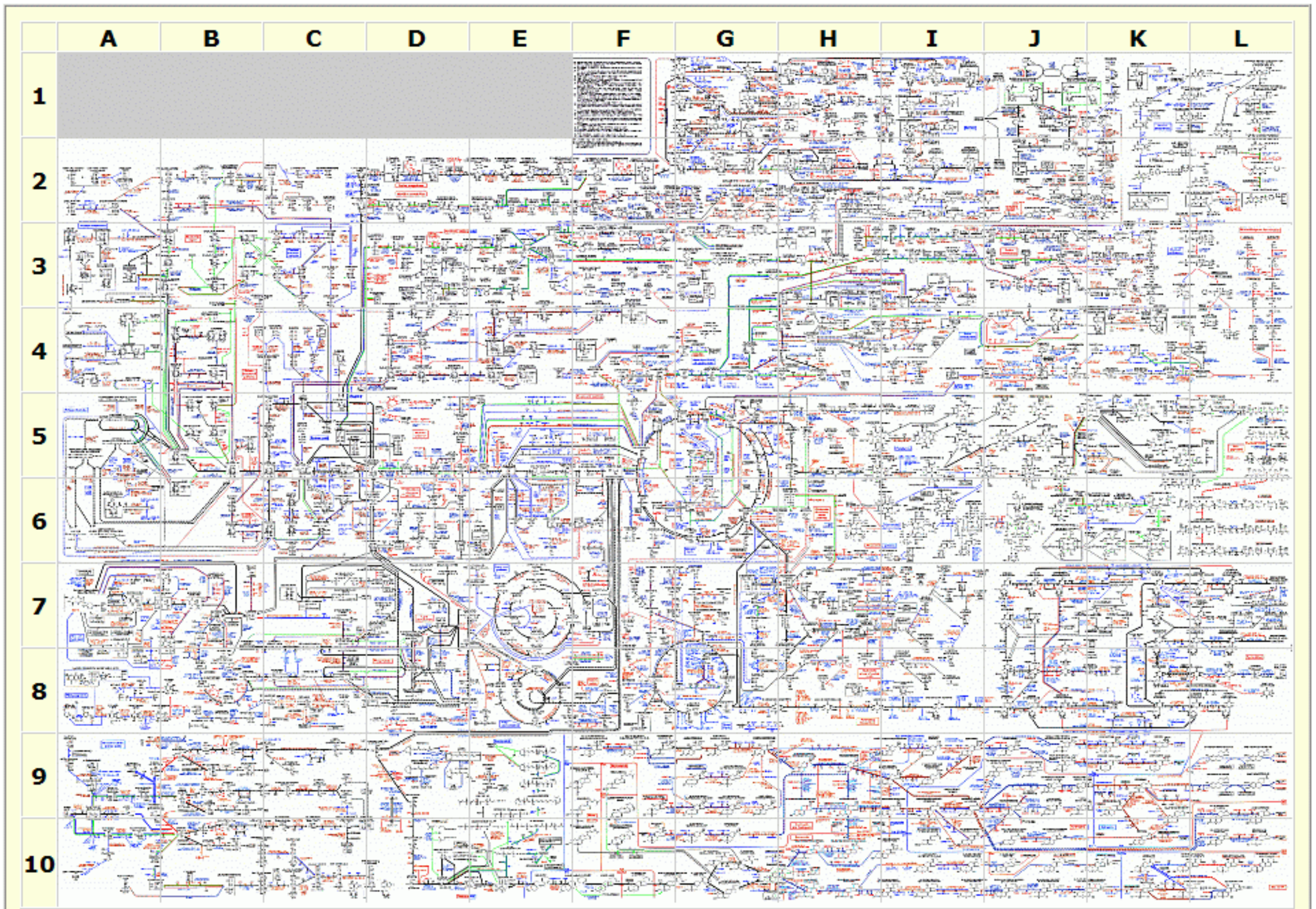


waag society

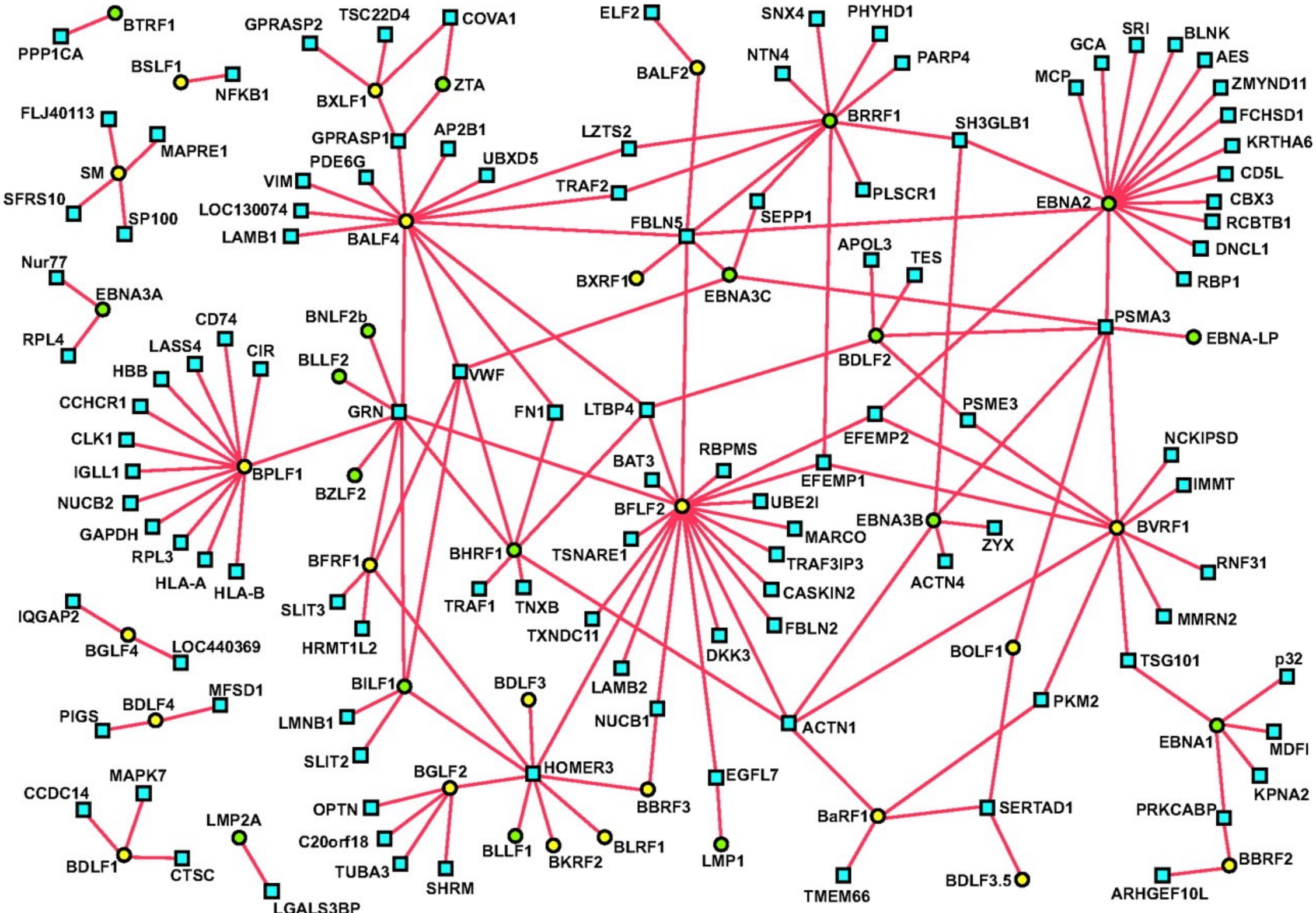
institute for art, science and technology

Proteins

Biochemical Pathways of the Cell



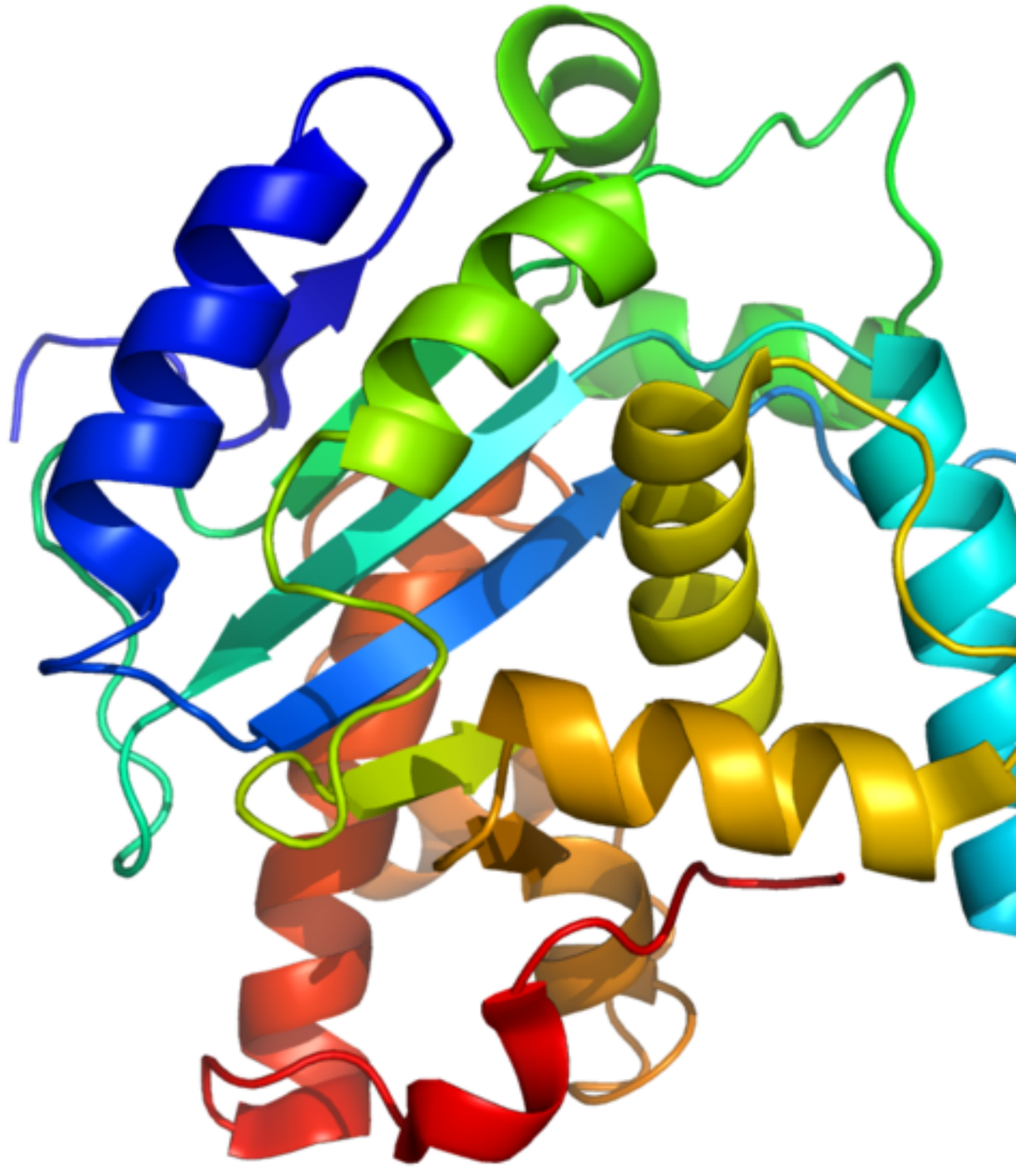
Protein interaction mapping: MS



What is this?

Simon Eugstar - CC-BY-SA 3.0

Debstart - CC-BY-SA 3.0

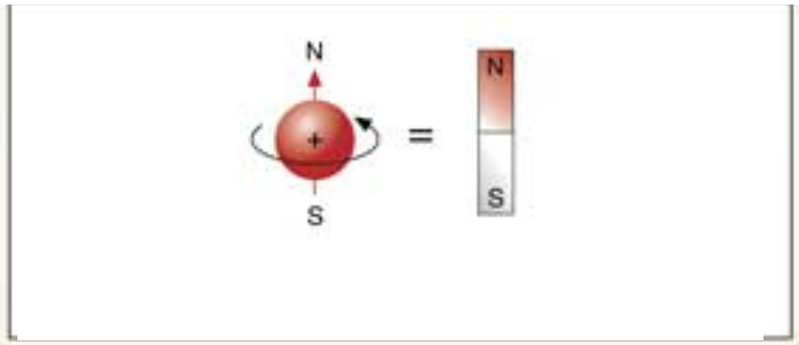


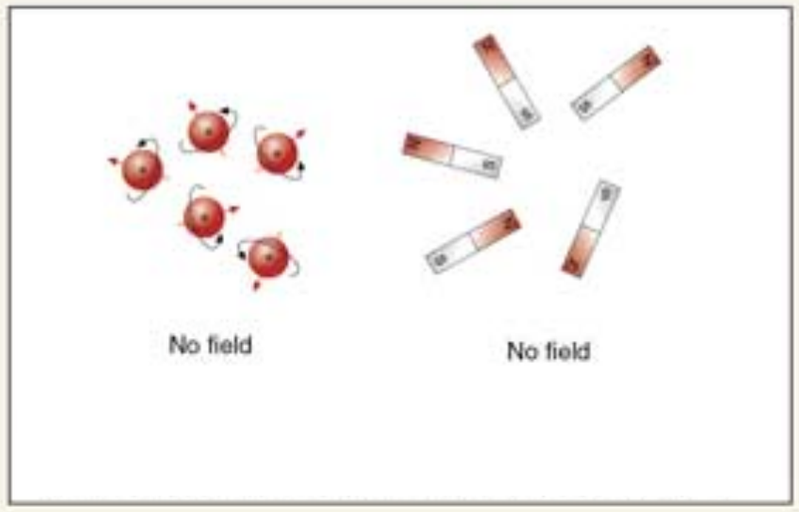
NMR Machines

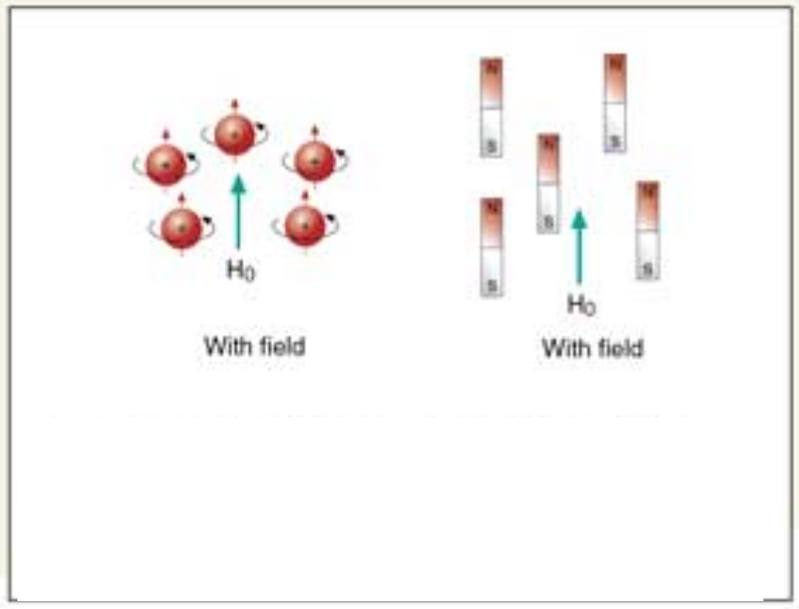
MartinSaunders

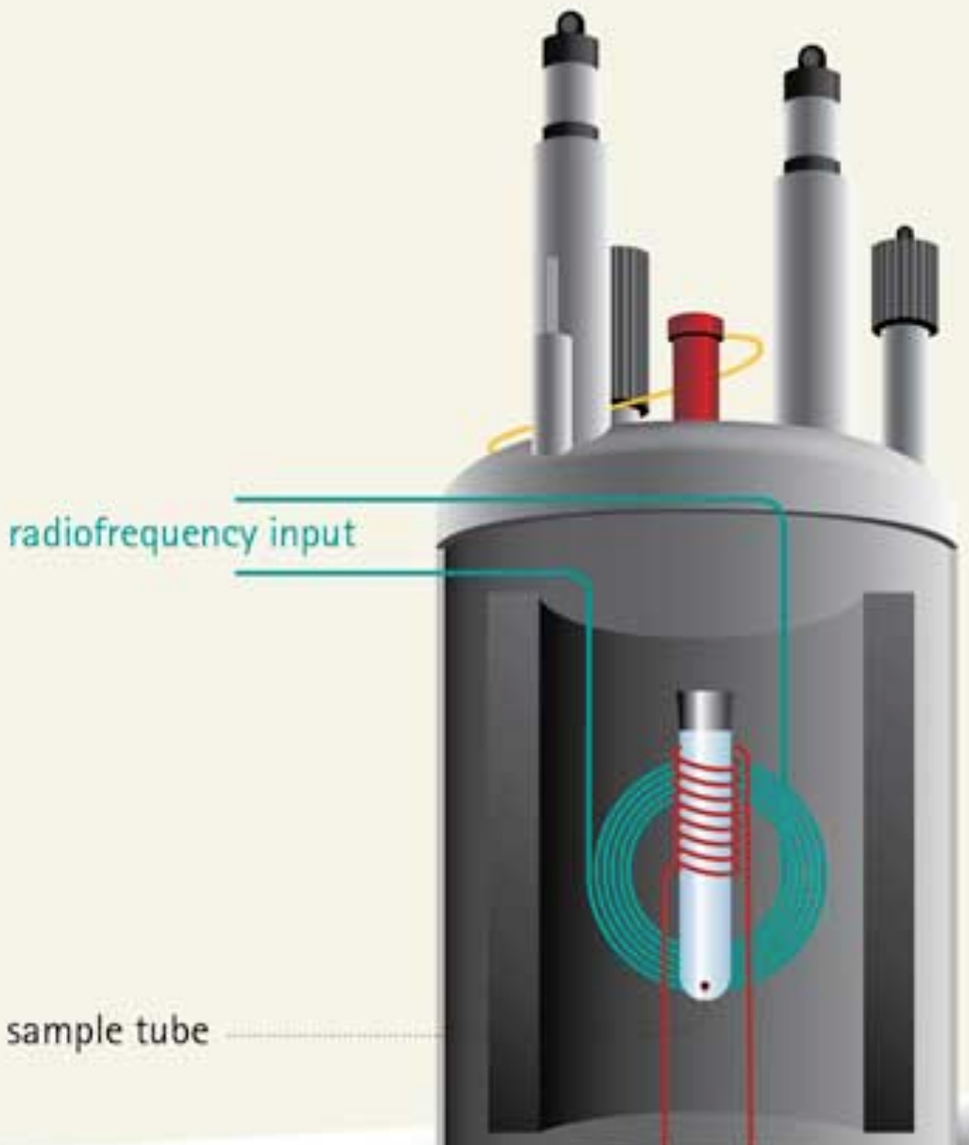
Public Domain

NMR principles

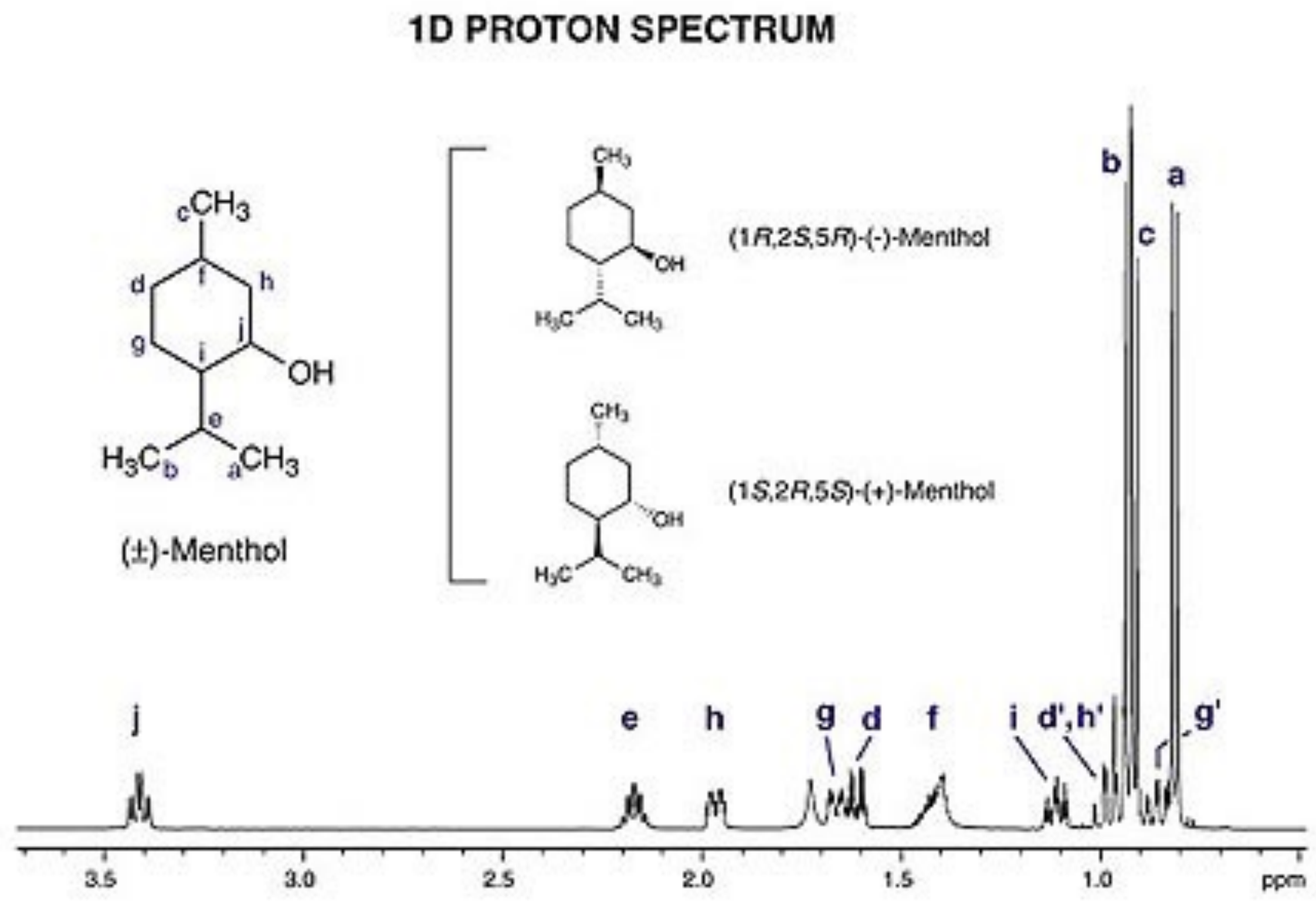






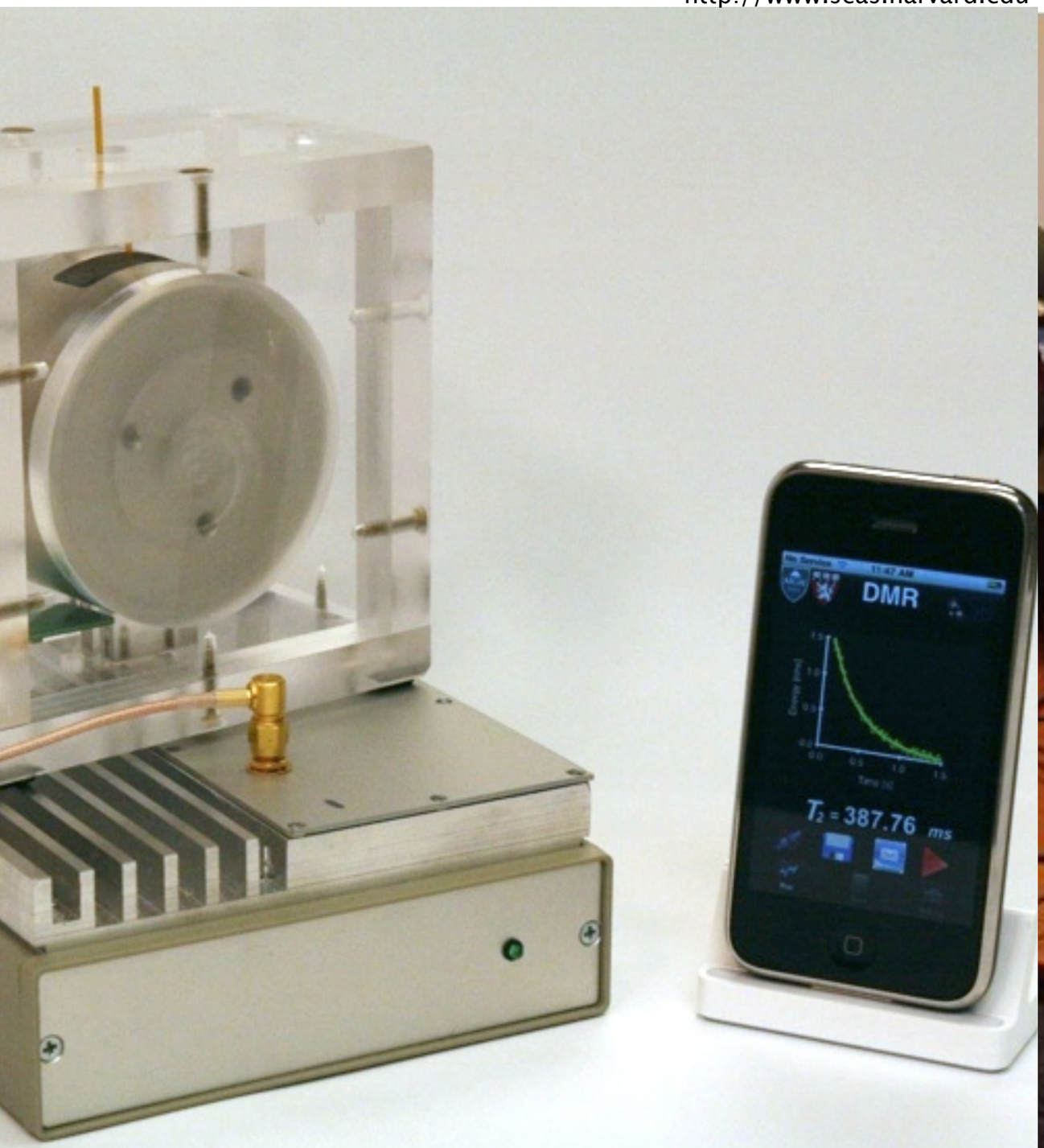


NMR Spectrum

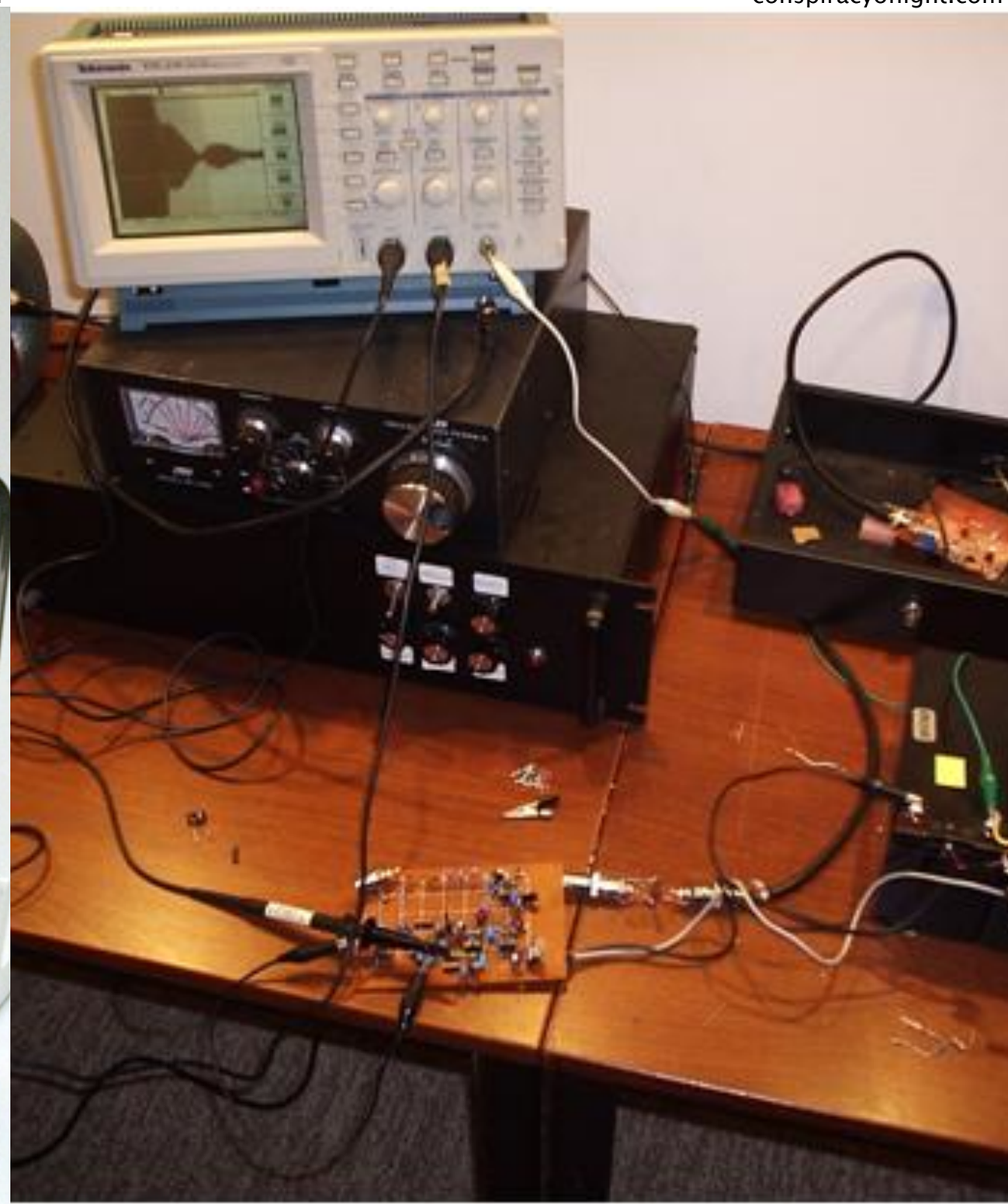


DIY NMR?

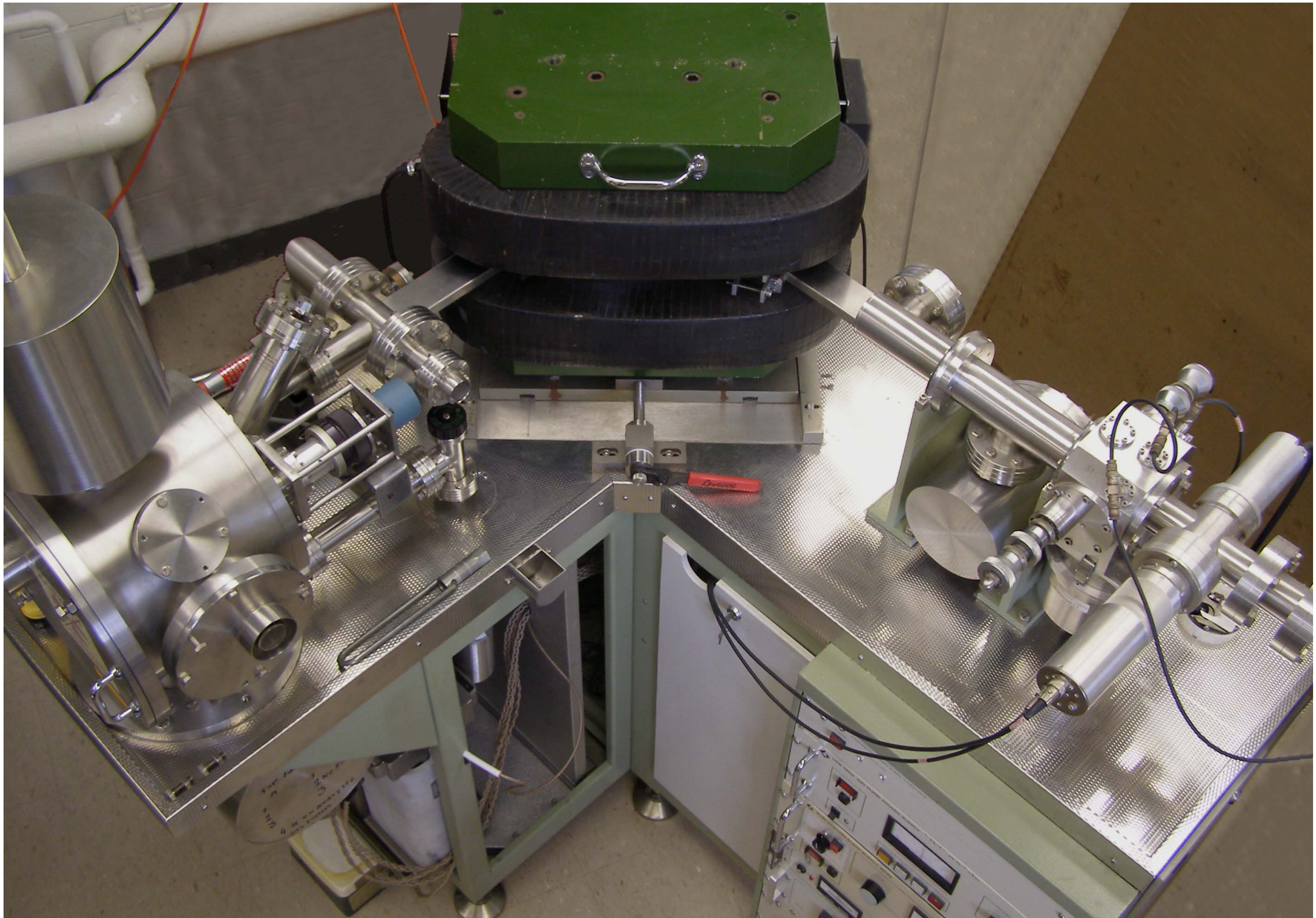
http://www.seas.harvard.edu



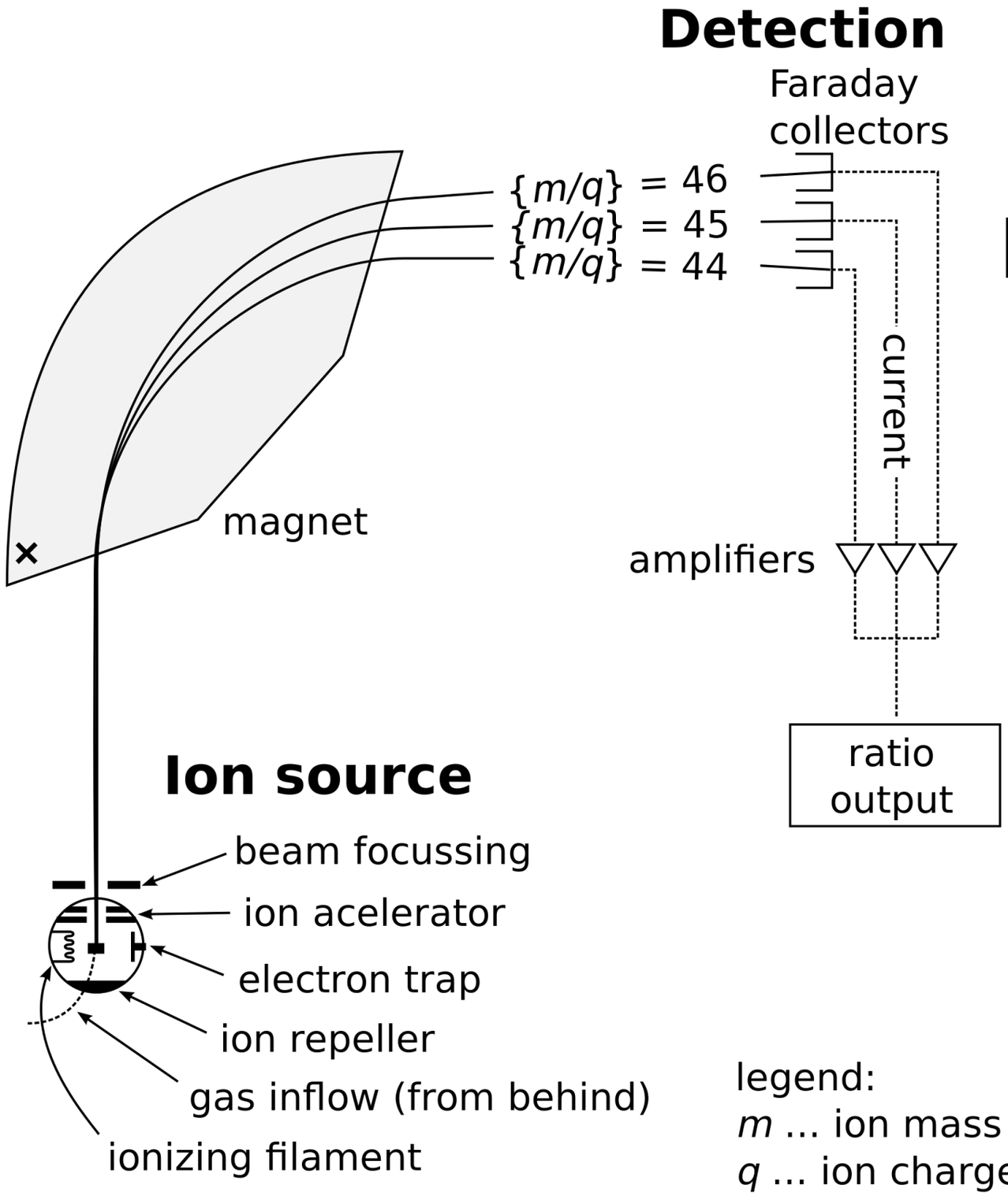
conspiracyoflight.com



Mass Spectrometer



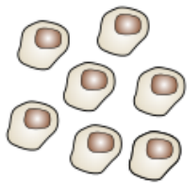
Simple Mass Spectrometry



High mass = high m/q

Big charge = low m/q

Mass Spectrometry



cells or tissue

MALDI

TOF

Tandem

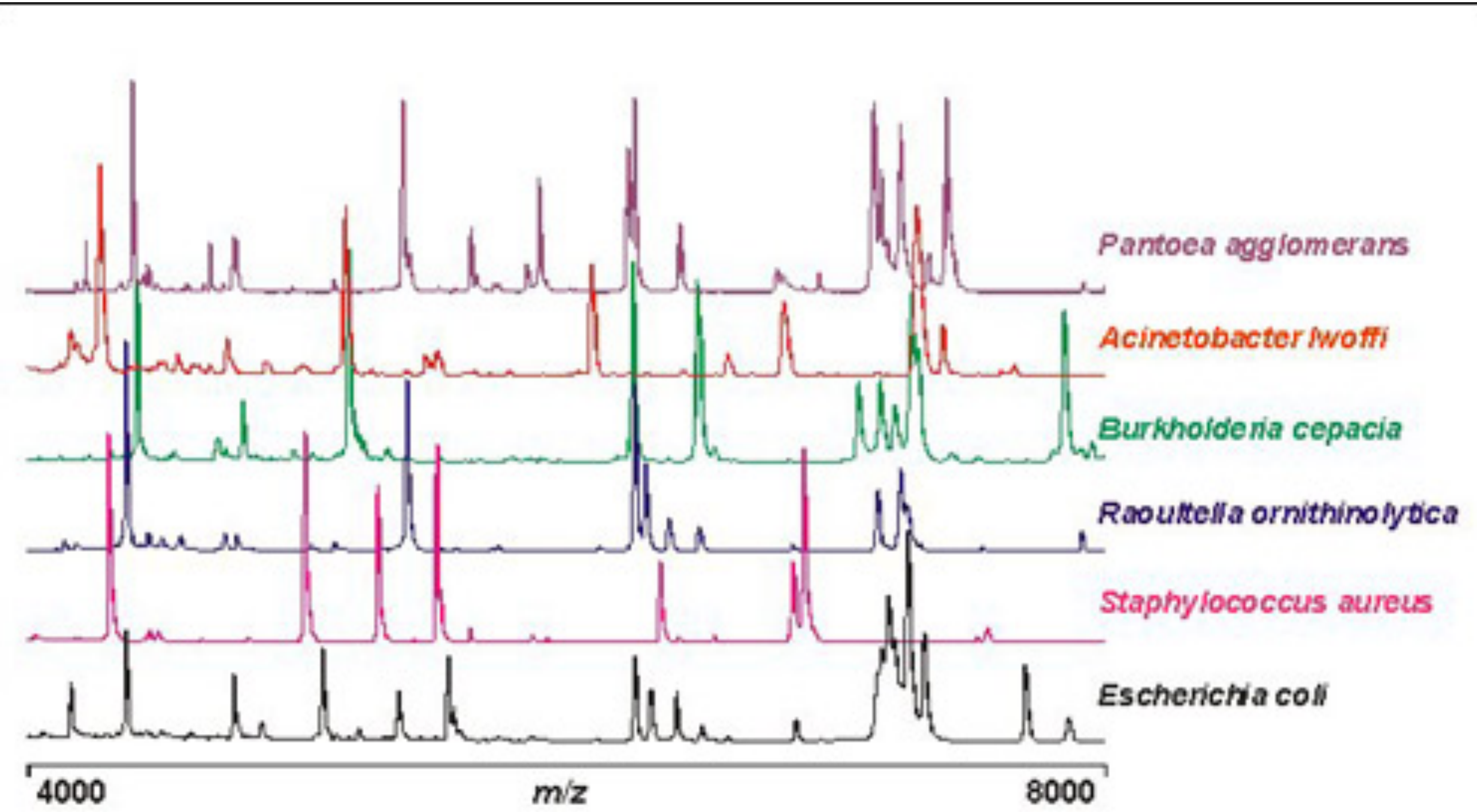
Procedure

Sampling the colony
± Extraction of intracellular protein in 70% formic acid and absolute ethanol

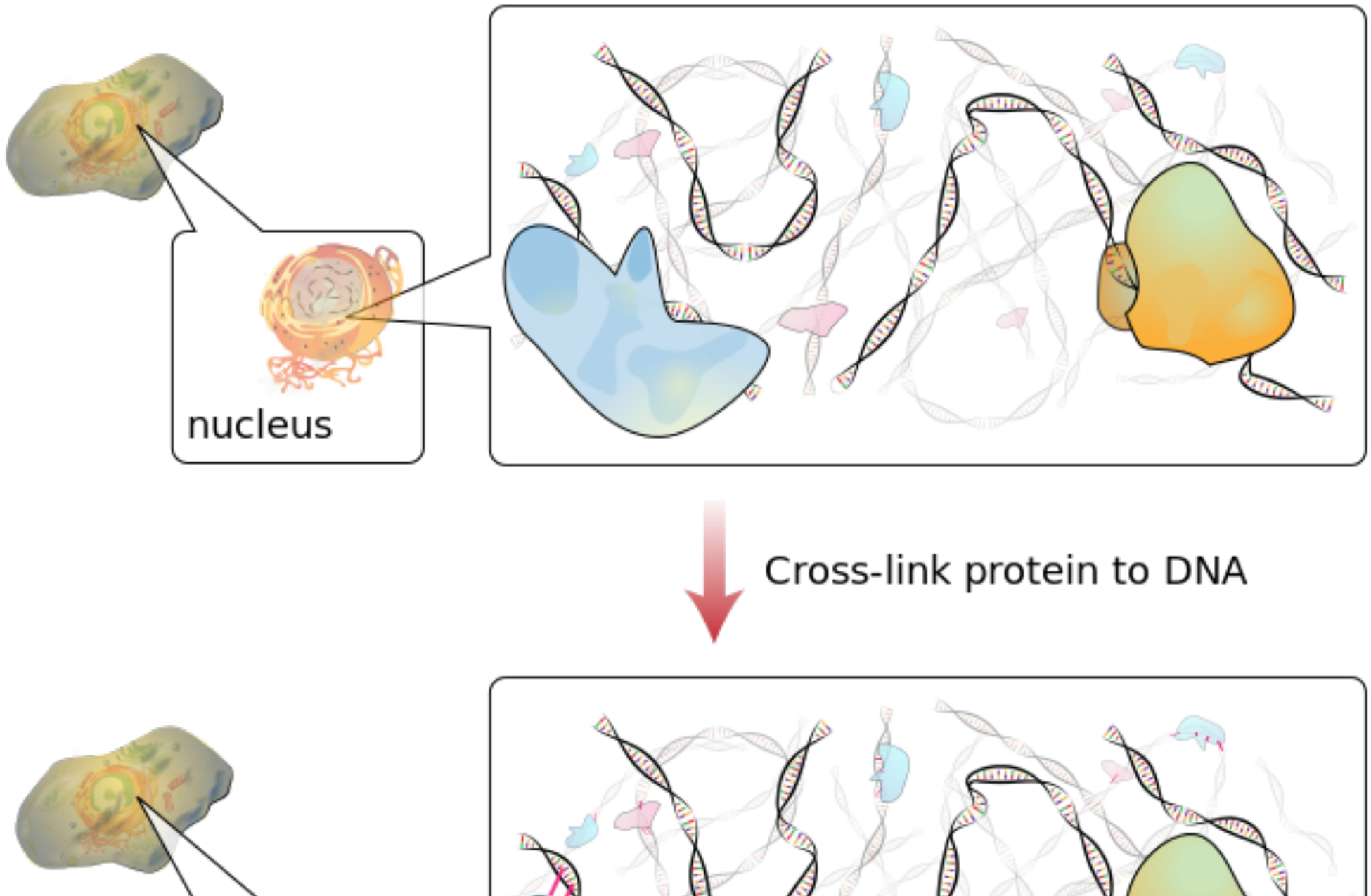
Spectrum and interpretation are obtained in less than 2 minutes per deposit

Introduction of the target into the apparatus

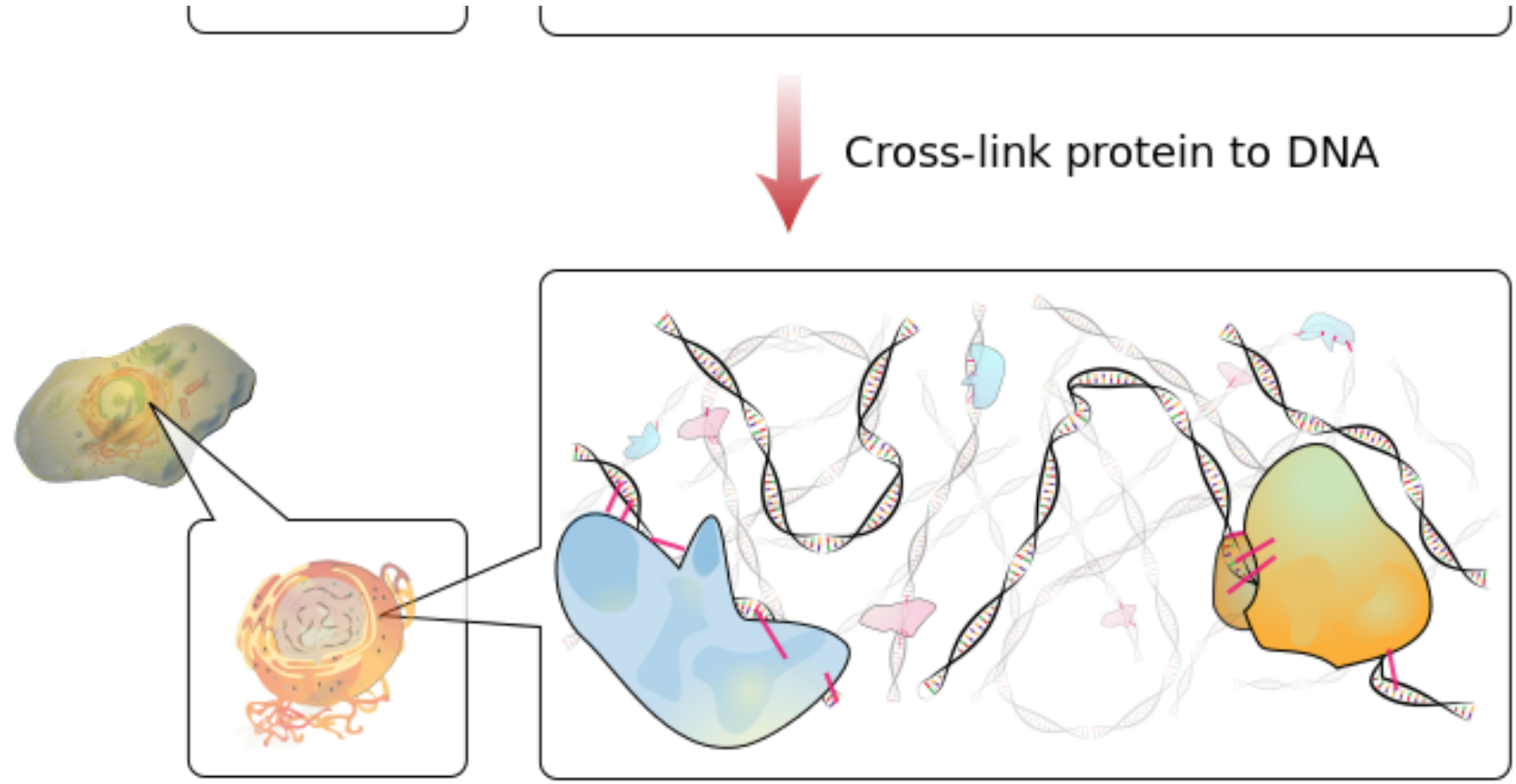
Bacterial profiles



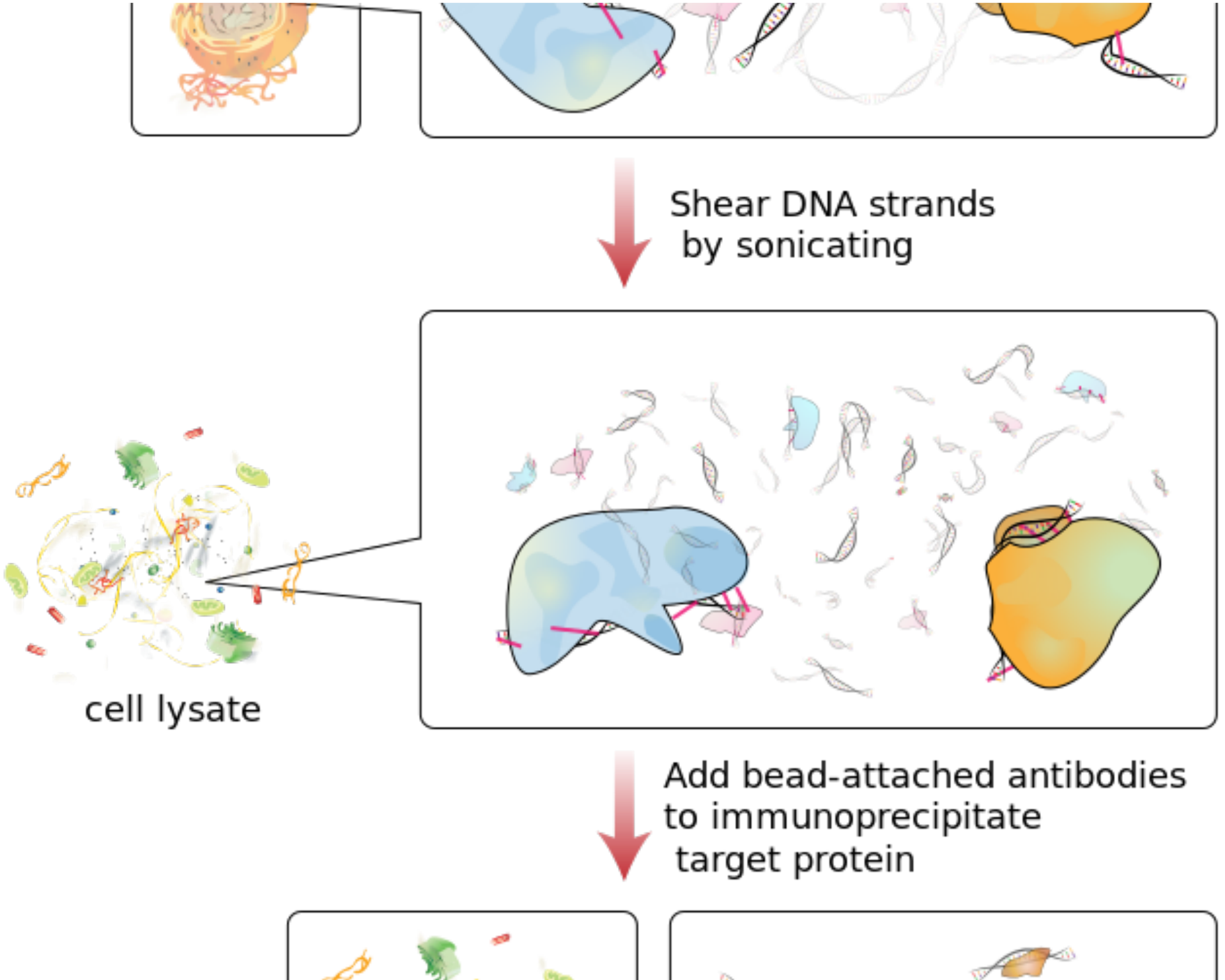
Chromatin ImmunoPrecipation ChIP



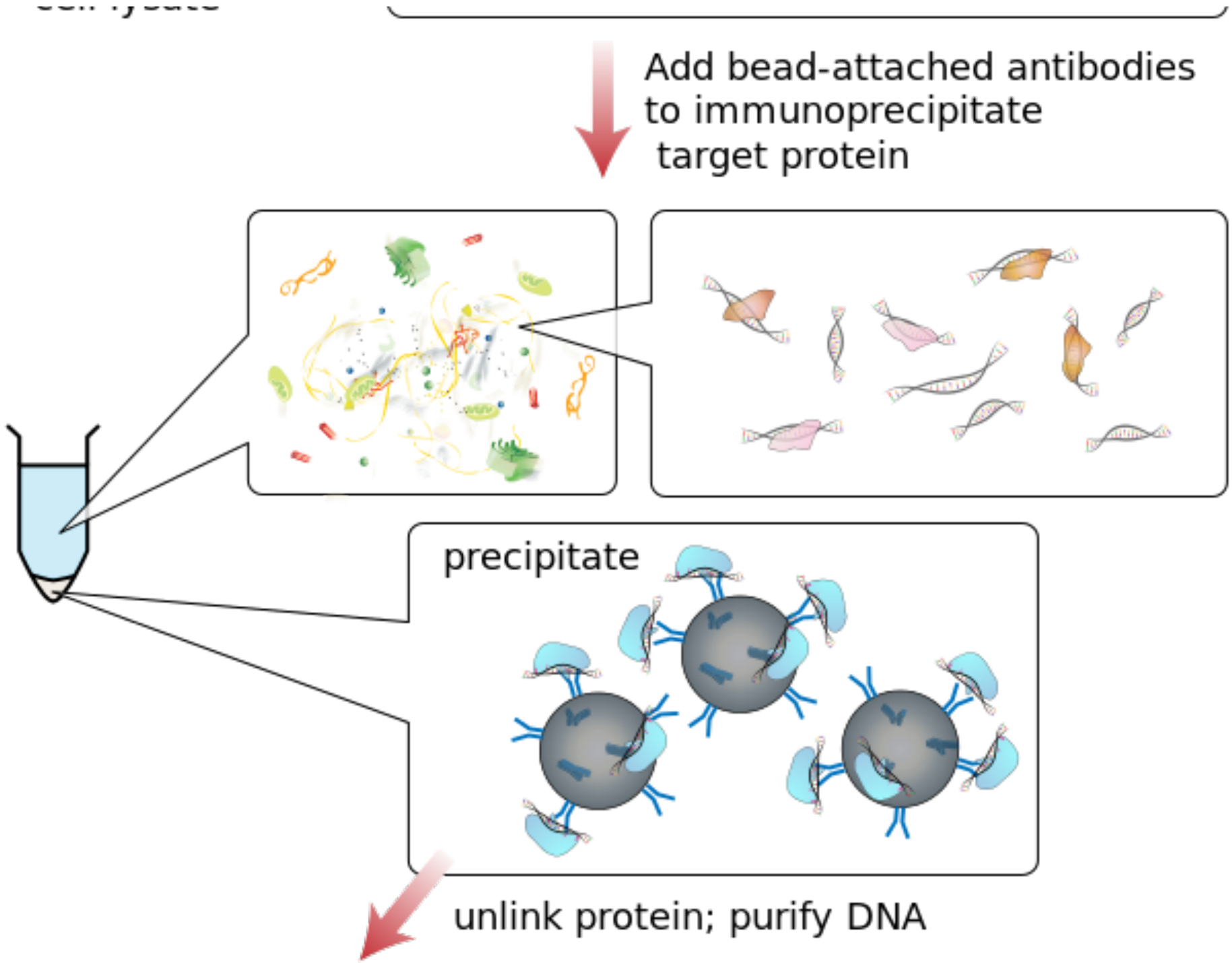
Chromatin ImmunoPrecipation ChIP



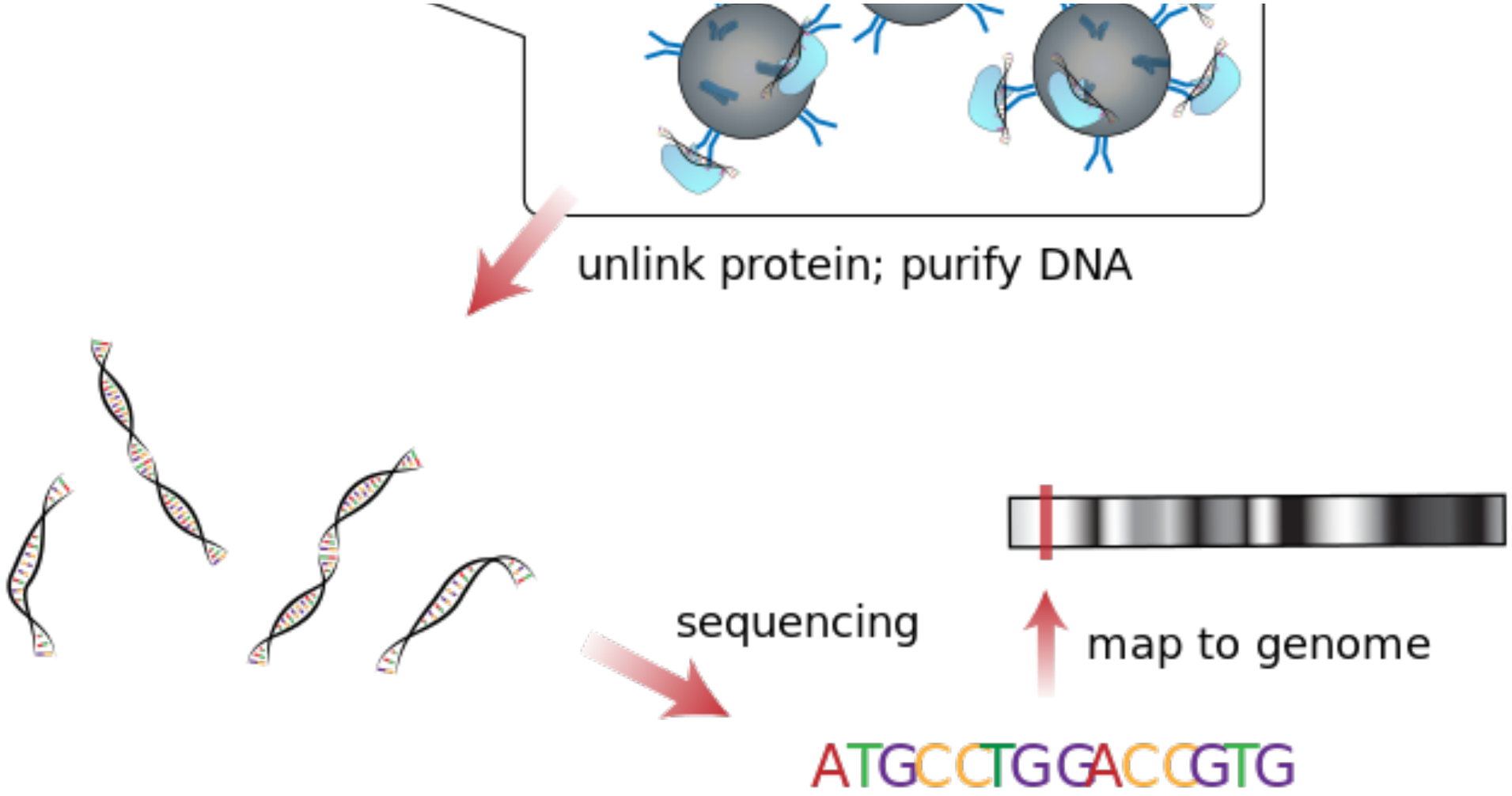
Chromatin ImmunoPrecipation ChIP



Chromatin ImmunoPrecipation ChIP



Chromatin ImmunoPrecipation ChIP



From analysis to synthesis

- Genomics: DNA sequence data
- Transcriptomics: Gene expression data
- Proteomics: Protein composition data

- Can we reverse this process and design our own bioproducts?

Bridging protein and dna data

20n is open sourcing its platform for synthetic biology

Over the last 4 years we have developed a better way to bioengineer organisms. We are now open sourcing our entire software stack, 20n/act. Find it at https://github.com/20n/act.

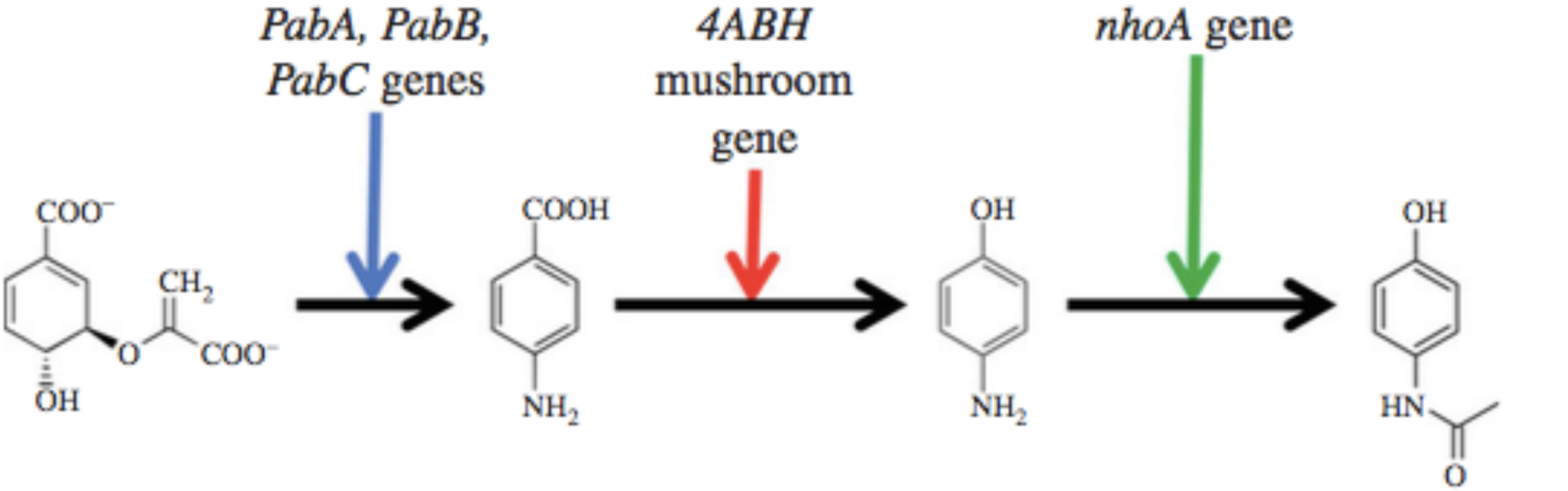
The stack will enumerate all bio-accessible chemicals, called *reachables* (20n/act/reachables). For each of those chemicals, it will design DNA blueprints. These DNA blueprints can bioengineering organisms with un-natural function. E.g., build organisms to make chemicals that were previously only sourced through petrochemistry.

To do that, the stack contains many modules built from scratch in-house. Some of them: mine raw biochemical data, integrate heterogenous sources, learn rules of biochemistry, automatically clean bad data, mine patents, mine plain text, bioinformatic identification of enzymes with desired function.

Once the suggested DNA is used to create new engineered cells those cells can be analyzed with LCMS for function. Our deep learning-based untargeted metabolomics stack processes the raw data and enumerates all side-effects of the changed genomic structure of the cell. Some would be expected, as the organism making the desired chemical, and some unexpected metabolic changes are highlighted.

We are also releasing a economic cost model for bioproduction. This economic cost model maps the desired the market price of the biological product to the "science needed" to get there. The "science needed" is measured in fermentation metrics, yield,

Predicted pathway



Issues

- Ethical
  - Who owns bio data?
  - Who decides what to use data for?
  - Is de-personalized bio information possible?

- Imagine:
  - You are one of the only persons immune to Zika virus. Are you entitled to royalties on the vaccine derived from your blood?

Visualisation PyMol