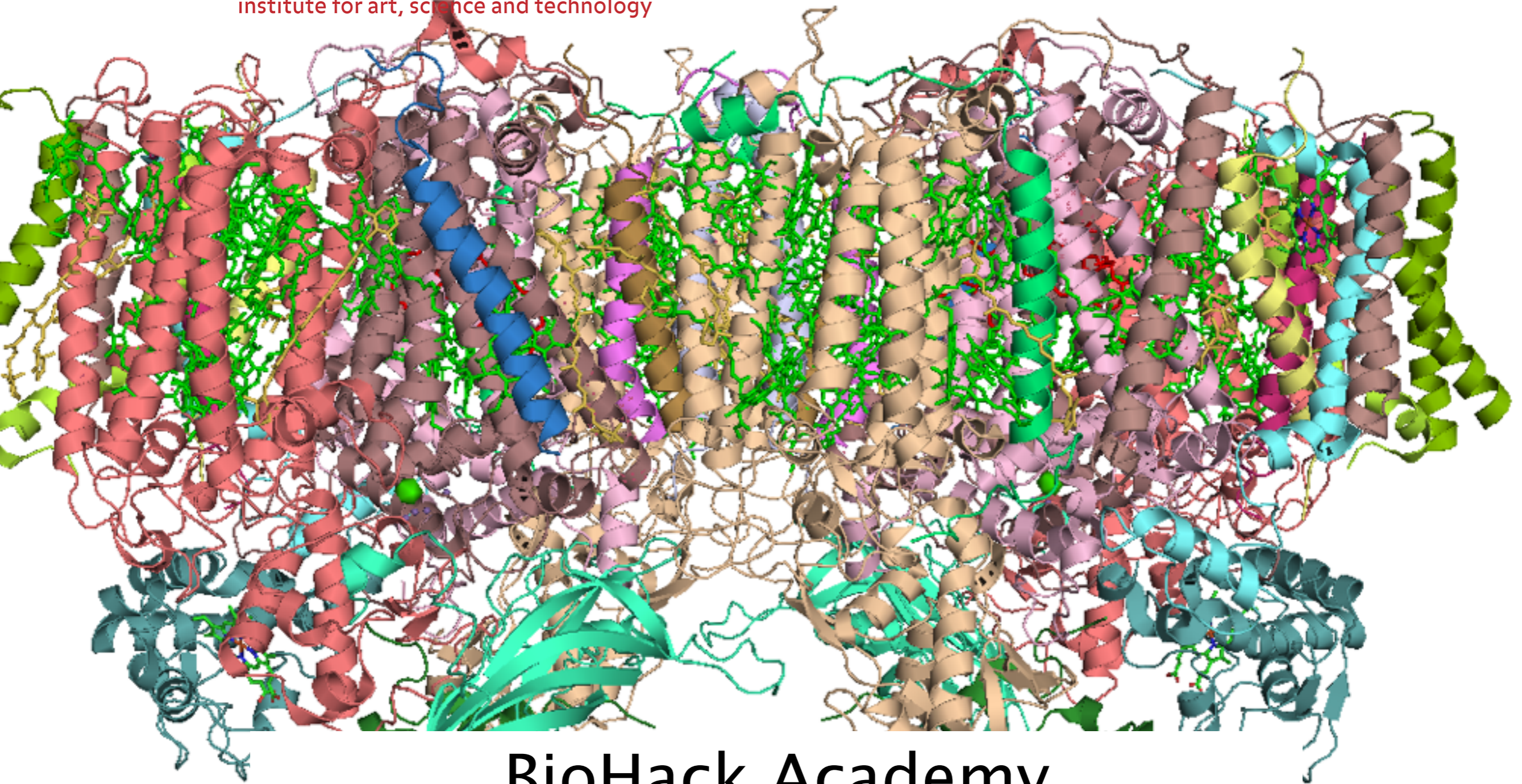




**waag society**

*institute for art, science and technology*



**BioHack Academy**  
**Bio Informatics**



# The “Omics”

- “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





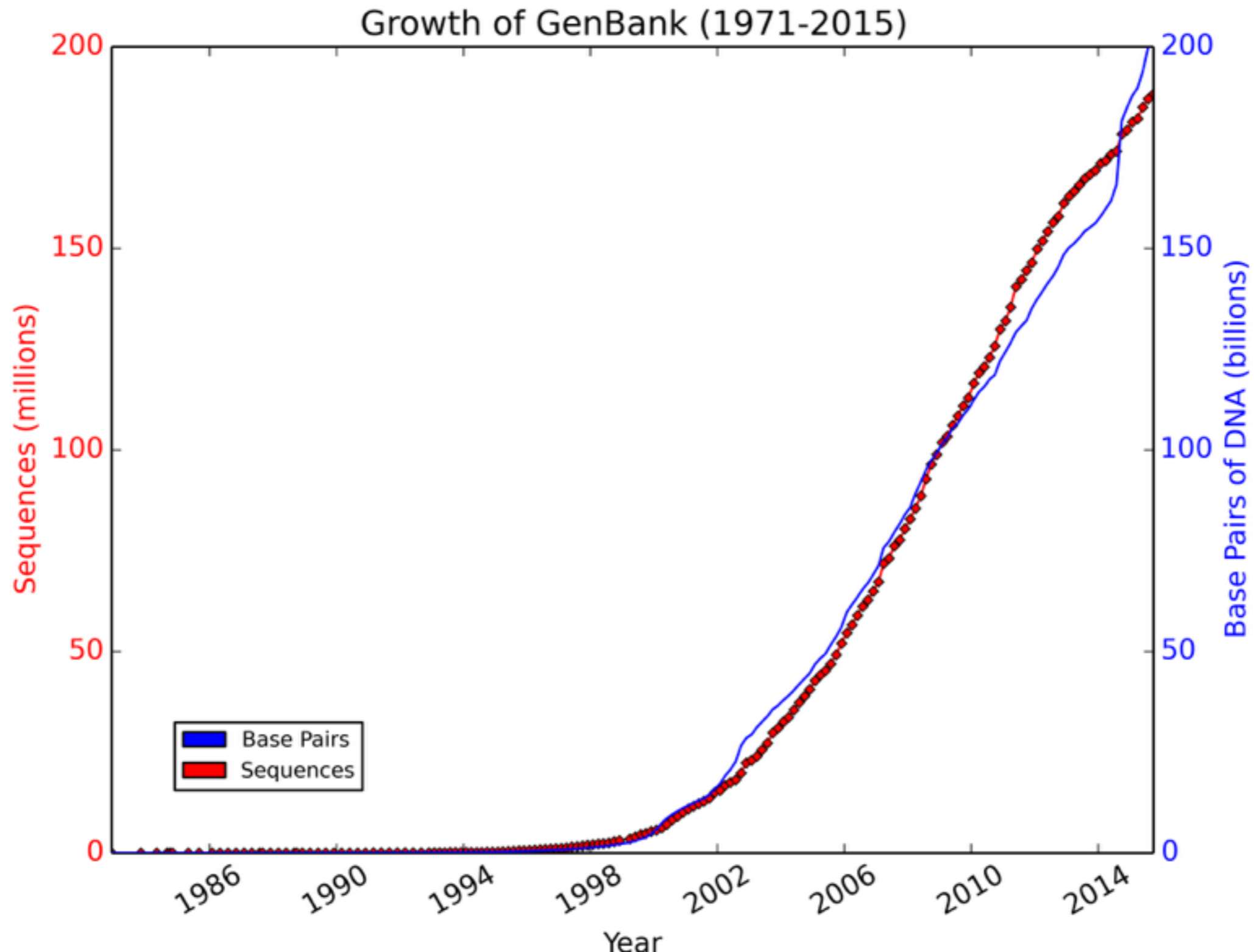
# Genomics

- Functional genomics
- Metagenomics
- Personal Genomics
- Epigenomics





# DNA database GenBank





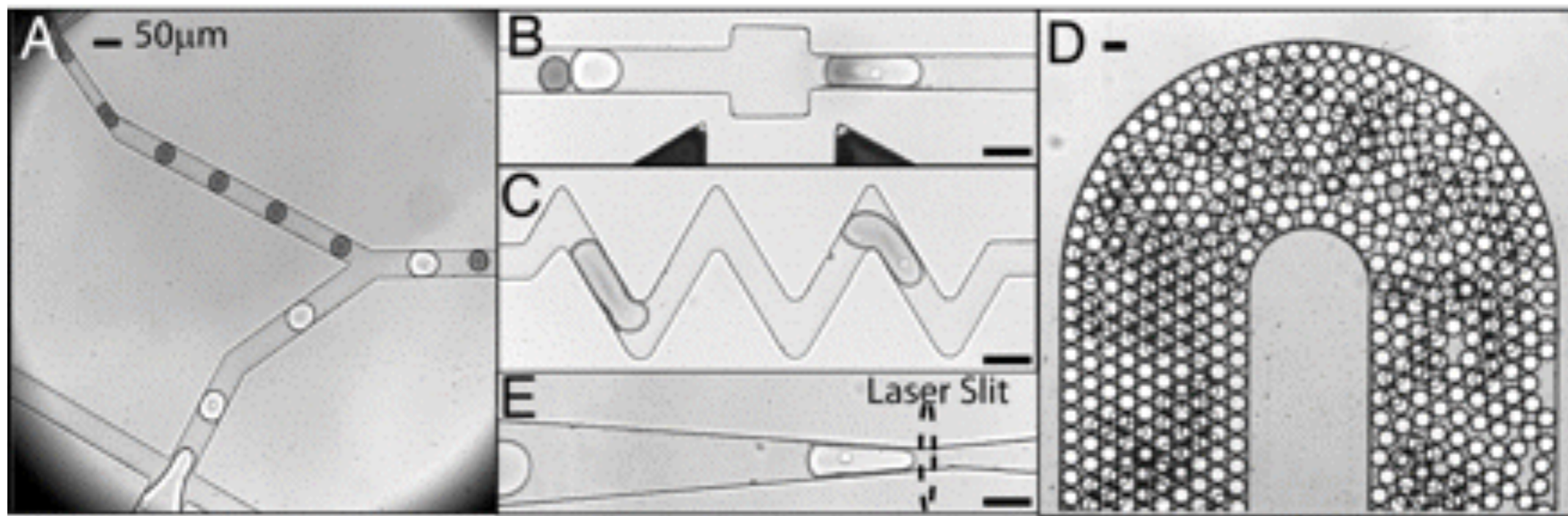
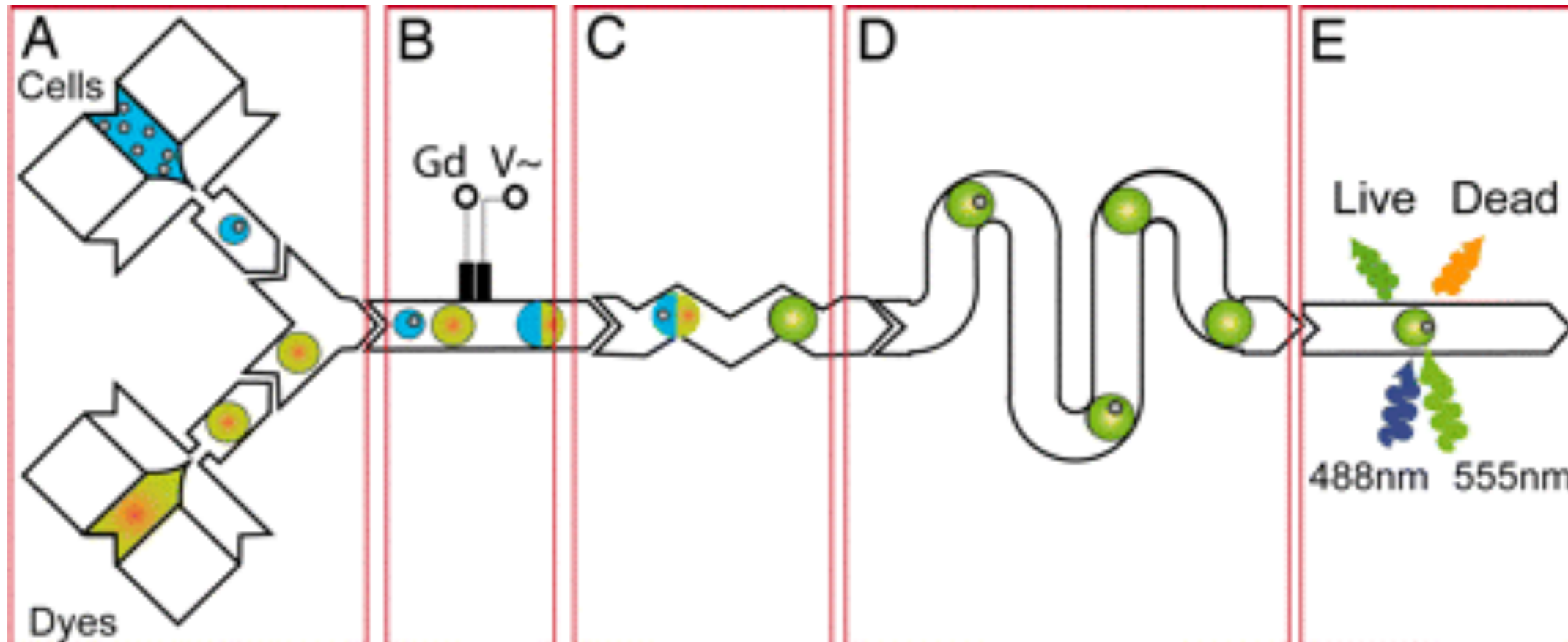
# Drivers

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



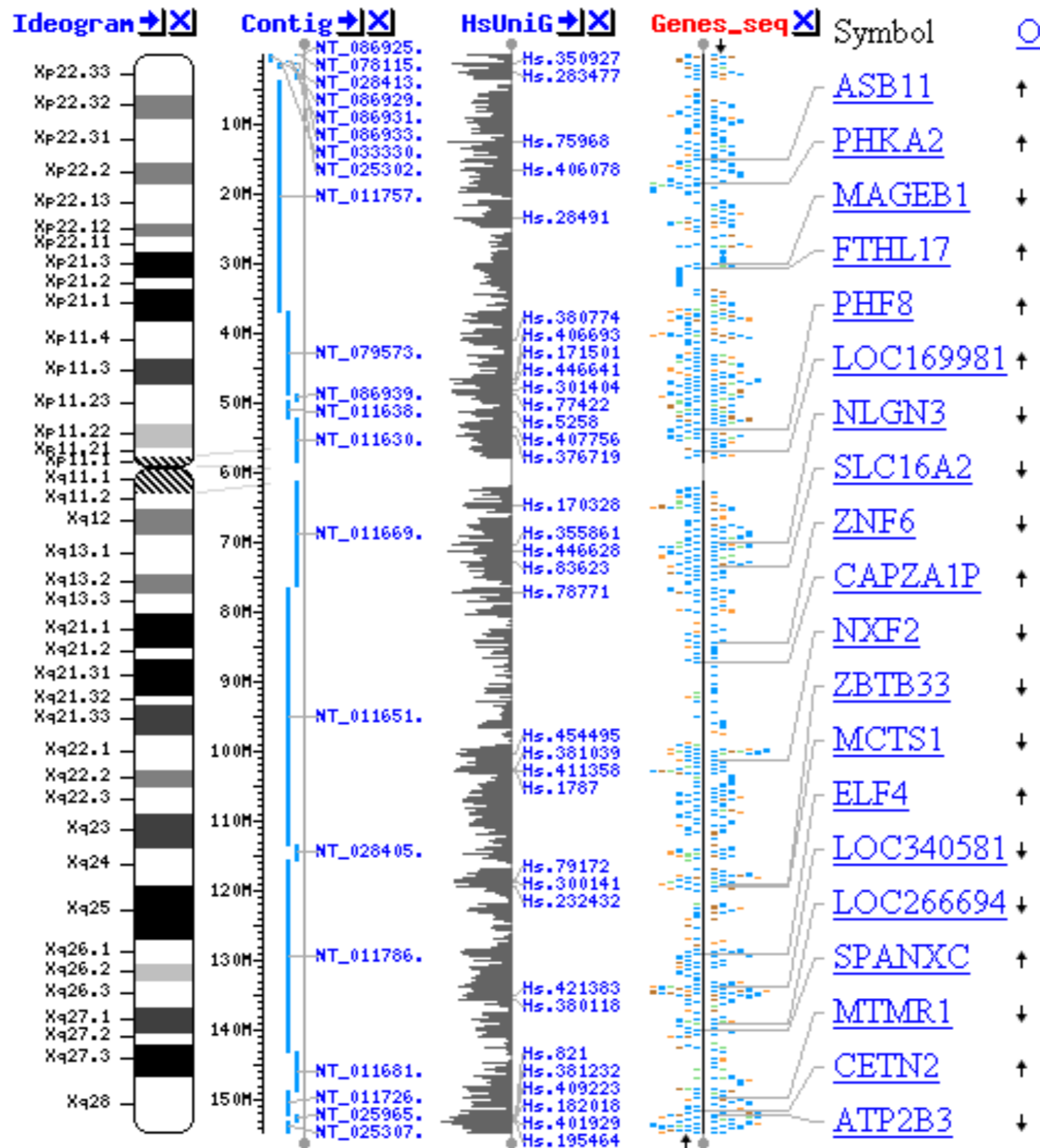
# High Throughput Screening

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





# Gene annotations







# BLAST: Basic Local Alignment Search Tool

<http://blast.ncbi.nlm.nih.gov/Blast.cgi>

**BLAST®** Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

NCBI/ BLAST/ blastn suite **Standard Nucleotide BLAST**

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

### Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)  [Clear](#)

Query subrange  From  To

Or, upload file  no file selected

Job Title   
Enter a descriptive title for your BLAST search

Align two or more sequences

### Choose Search Set

Database  Human genomic + transcript  Mouse genomic + transcript  Others (nr etc.):

Organism Optional   Exclude   
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude Optional  Models (XM/XP)  Uncultured/environmental sample sequences

Limit to Optional  Sequences from type material

Entrez Query Optional  [YouTube](#) [Create custom database](#)  
Enter an Entrez query to limit search

### Program Selection

Optimize for  Highly similar sequences (megablast)  
 More dissimilar sequences (discontiguous megablast)  
 Somewhat similar sequences (blastn)  
Choose a BLAST algorithm

Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)  
 Show results in a new window



# Sequence Alignment

```
14 SIKLWPPSQTRLLVERMANNLST..PSIFTRK..YGSLSKEEARENAKQIEEVACSTANQ.....HYEKEPDGDGGSASVQLYAKECSKLILEVLK 101
13 SIKLWPPSESTRIMLVDRMTNNLST..ESIFSRK..YRLLGKQEAHENAKTIEELCFALADE.....HFREEPDGDGSSAVQLYAKETSKMMLEVLK 100
23 VFKLWPPSQGTREAVRQKMAKLSS..ACFESQS..FARIELADAQE HARAIIEEVAFGAAQE.....ADSGGDKTGSASVVMVYAKHASKLMLETLR 109
13 SVKLWPPGQSTRMLMVERMTKNFIT..PSFISRK..YGLLSKEEAEDAKKIEEVAFAAANQ.....HYEKQPDGDGSSAVQIYAKESRRLMLEVLK 100
30 SFSIWPPPTQRTRDAVVRRLVDTLGG..DTILCKR..YGAVPAADAEPARGIEAEAFDAAA..SGEAAATASVEEGIKALQLYSKEVSRRLDFVK 120
44 SLSIWPPSQRTRDAVVRRLVQTLVA..PSILSKR..YGAVPEAEAGRAAAVEAEAYAAVTES..SSAAAAPASVEDGIEVLQAYSKEVSRRLLELAK 135
56 SFSIWPPPTQRTRDAIISRLIETLST..TSVLSKR..YGTIPKEEAASEARRIEEEAFSGAST.....VASSEKDGLEVLQLYSKEISKRMLETVK 141
29 SFAVWPPTRRTRDAVVRRLVAVLSGDTTALRKRYRYGAVPAADAERAARAVEAQAFDAASA...SSSSSSSVEDGIETLQLYSREVSNRLLAFVR 121
13 SIKLWPPSESTRMLMVERMTDNLSS..VSFFSRK..YGLLSKEEAENAKRIEETAFLAAND.....HEAKEPNLDDSSVVQFYAREASKLMLEALK 100
57 SLRIWPPTQKTRDAVLNRLIETLST..ESILSKR..YGTLSKSDATTVAKLIEEEAYGVASN.....AVSSDDDGKILELYSKEISKRMLESVK 142
25 NYSIWPPKQRTRDAVKNRLIETLST..PSVLTKR..YGTMSADEASAAAIQIEDEAFSVANA.....SSSTSDNVNVTILEVYSKEISKRMLETVK 110
28 SFKIWPPTQRTREAVVRRLVETLTS..QSVLSKR..YGVIPPEEDATSAARIIEEEAFSVASV..ASAASTGGRPEDEWIEVLHIYSQEIXQRVVESAK 119
25 SFSIWPPPTQRTRDAVINRLIESLST..PSILSKR..YGTLPQDEASETARLIEEEAFAAAGS.....TASDADDGIEILQVYSKEISKRMIDTVK 110
14 SVKMWPPSKSTRMLMVERMTKNITT..PSIFSRK..YGLLSVEEAEQDAKRIEDLAFATANK.....HFQNEPDGDGTSASVHVYAKESSKLMLDVIK 101
13 SIKLWPPSLPTRKALIERITNNFSS..KTIFTEK..YGS�TKDQATENAKRIEDIAFSTANQ.....QFEREPDGDGGSASVQLYAKECSKLILEVLK 100
48 SLSIWPPPTQRTRDAVITRLIETLSS..PSVLSKR..YGTISHDEAESARRIEDEAFGVANT.....ATSAEDDGLEILQLYSKEISRRMLDTVK 133
```



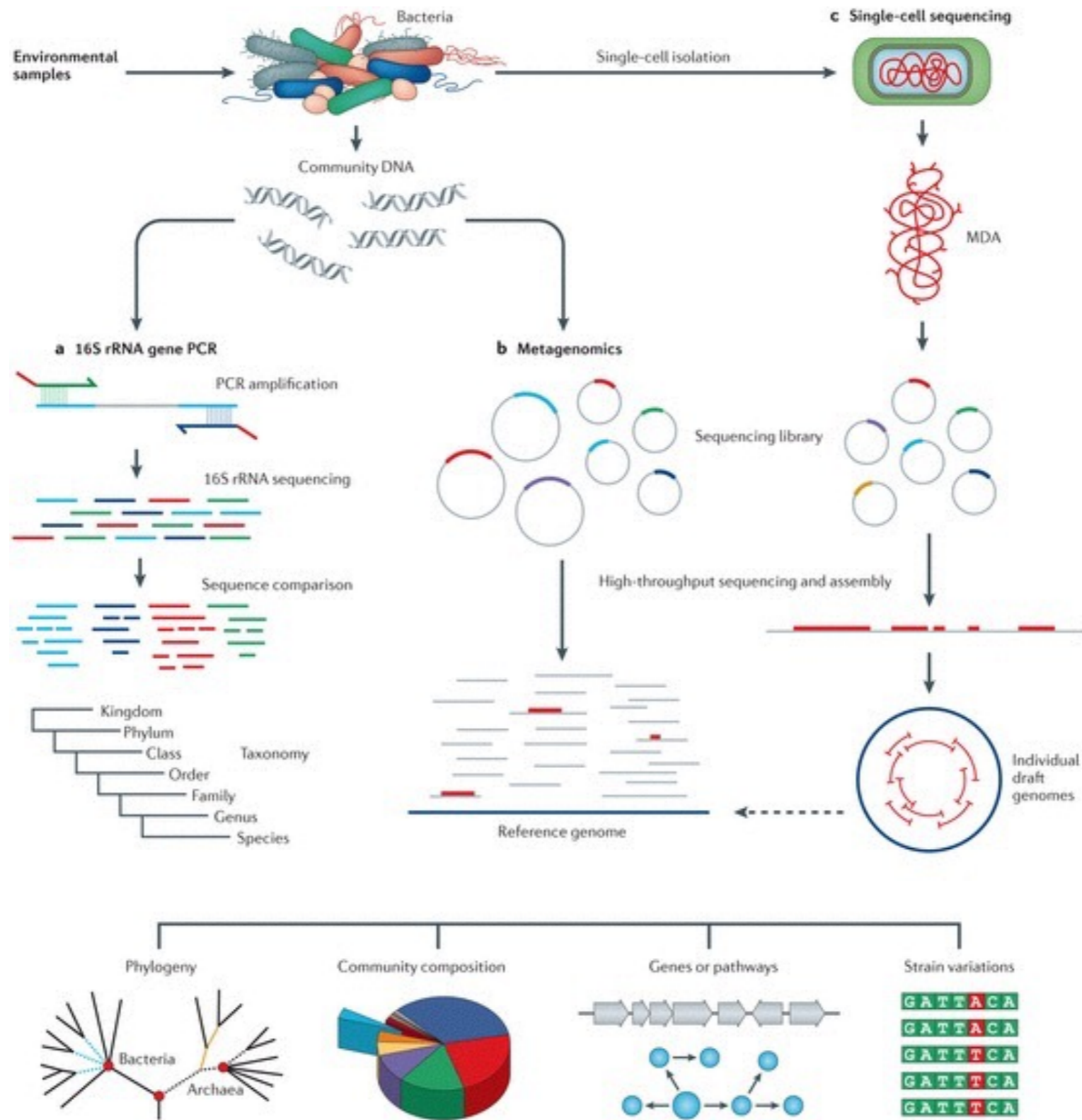
# Scoring Matrix BLOSUM

(BLOcks SUbstitution Matrix)

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



# Environmental DNA analysis



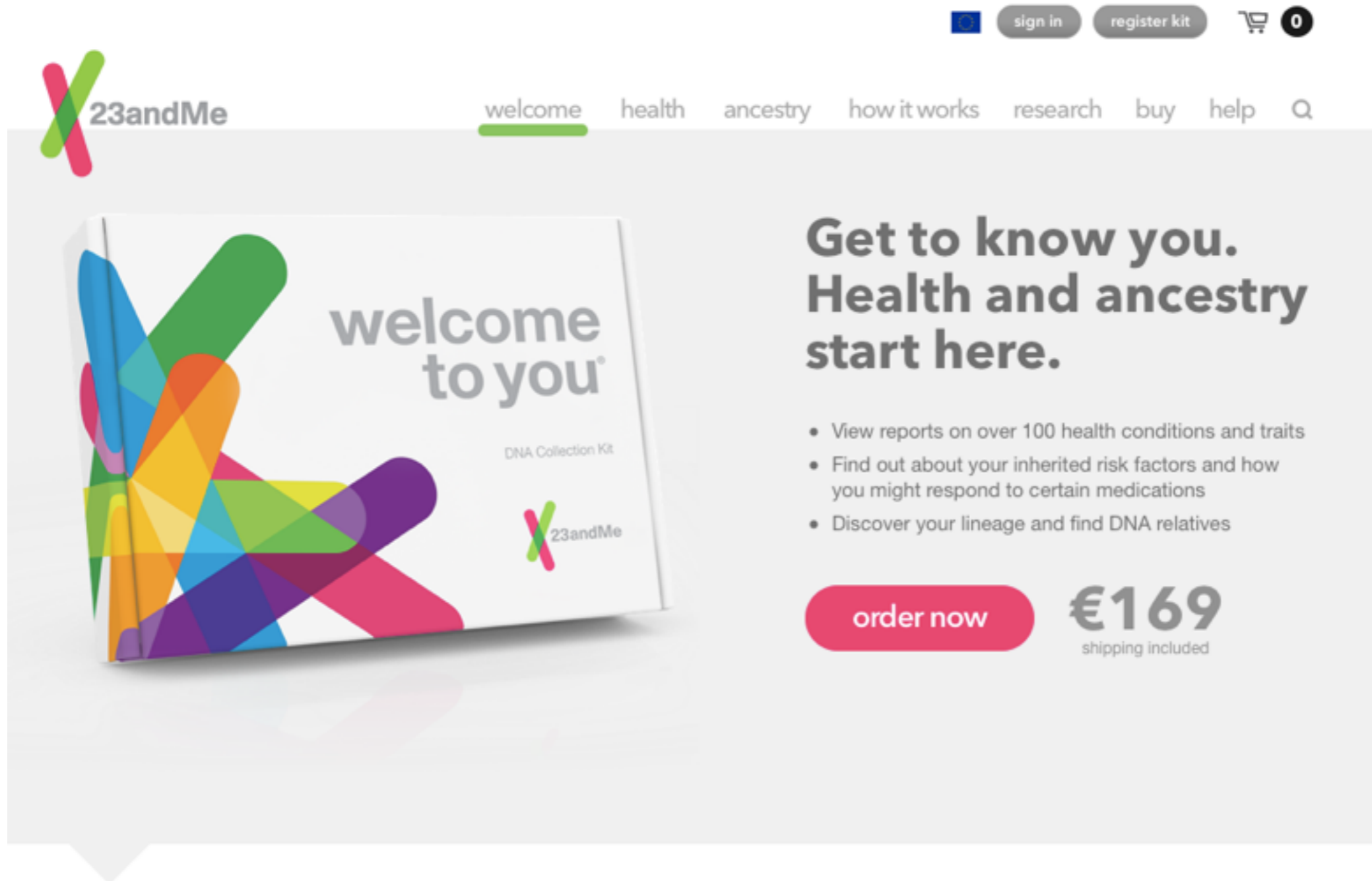


# Challenges

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



# Bio Informatics for the public



The screenshot shows the 23andMe website homepage. At the top right, there are links for 'sign in' and 'register kit', along with a shopping cart icon showing 0 items. The navigation menu includes 'welcome', 'health', 'ancestry', 'how it works', 'research', 'buy', 'help', and a search icon. The main content area features a large image of a 'welcome to you' DNA Collection Kit box on the left. To the right of the box, the text reads 'Get to know you. Health and ancestry start here.' Below this, there is a list of three bullet points: 'View reports on over 100 health conditions and traits', 'Find out about your inherited risk factors and how you might respond to certain medications', and 'Discover your lineage and find DNA relatives'. At the bottom right of this section, there is a red 'order now' button and a price tag of '€169 shipping included'.

## 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 Expression Analysis

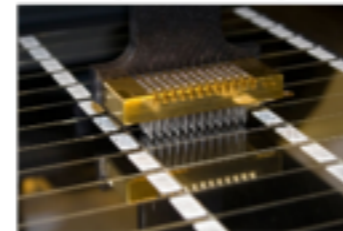
DNA extraction and digestion



Test = Tumoral DNA  
labeled with **Cy5**

Reference = Normal DNA  
labeled with **Cy3**

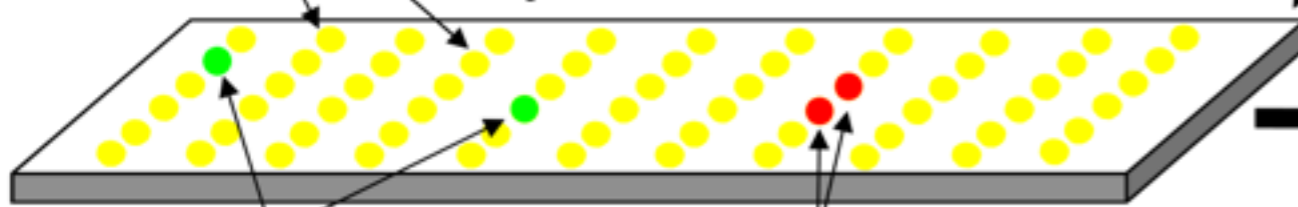
Spotting



combine equal  
amounts of DNA

Probes

Hybridisation



Lack of tumoral DNA  
Deletion

Excess of tumoral DNA  
Amplification

Scanning

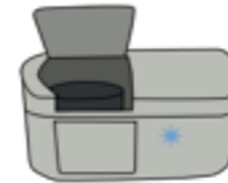
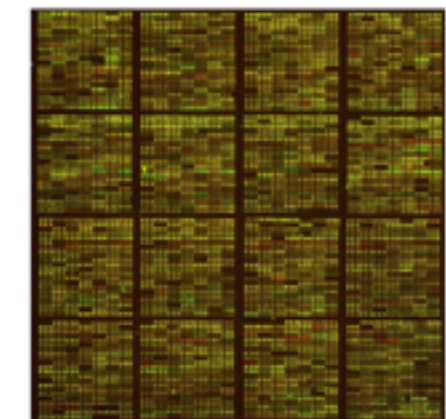
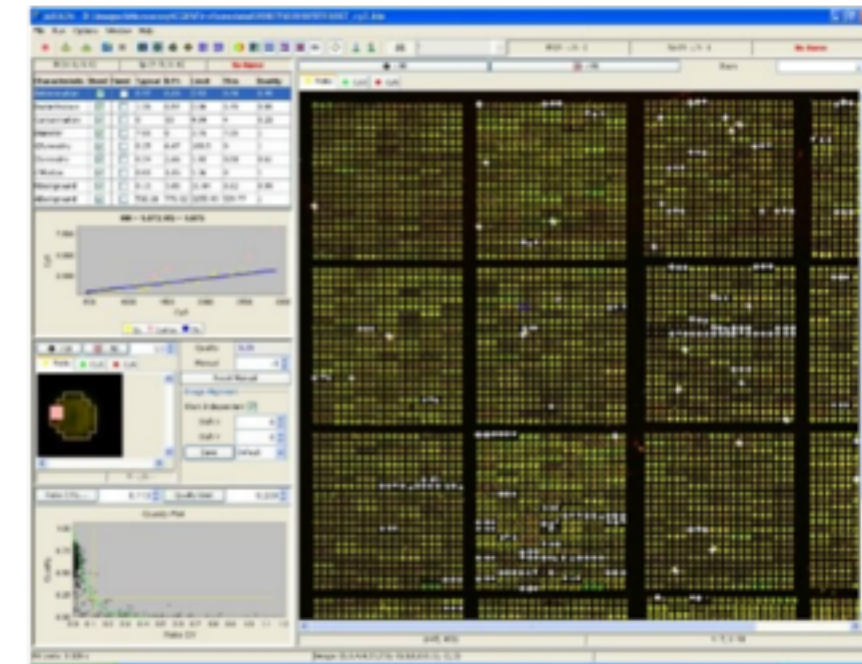


Image analysis software



Image



**waag society**

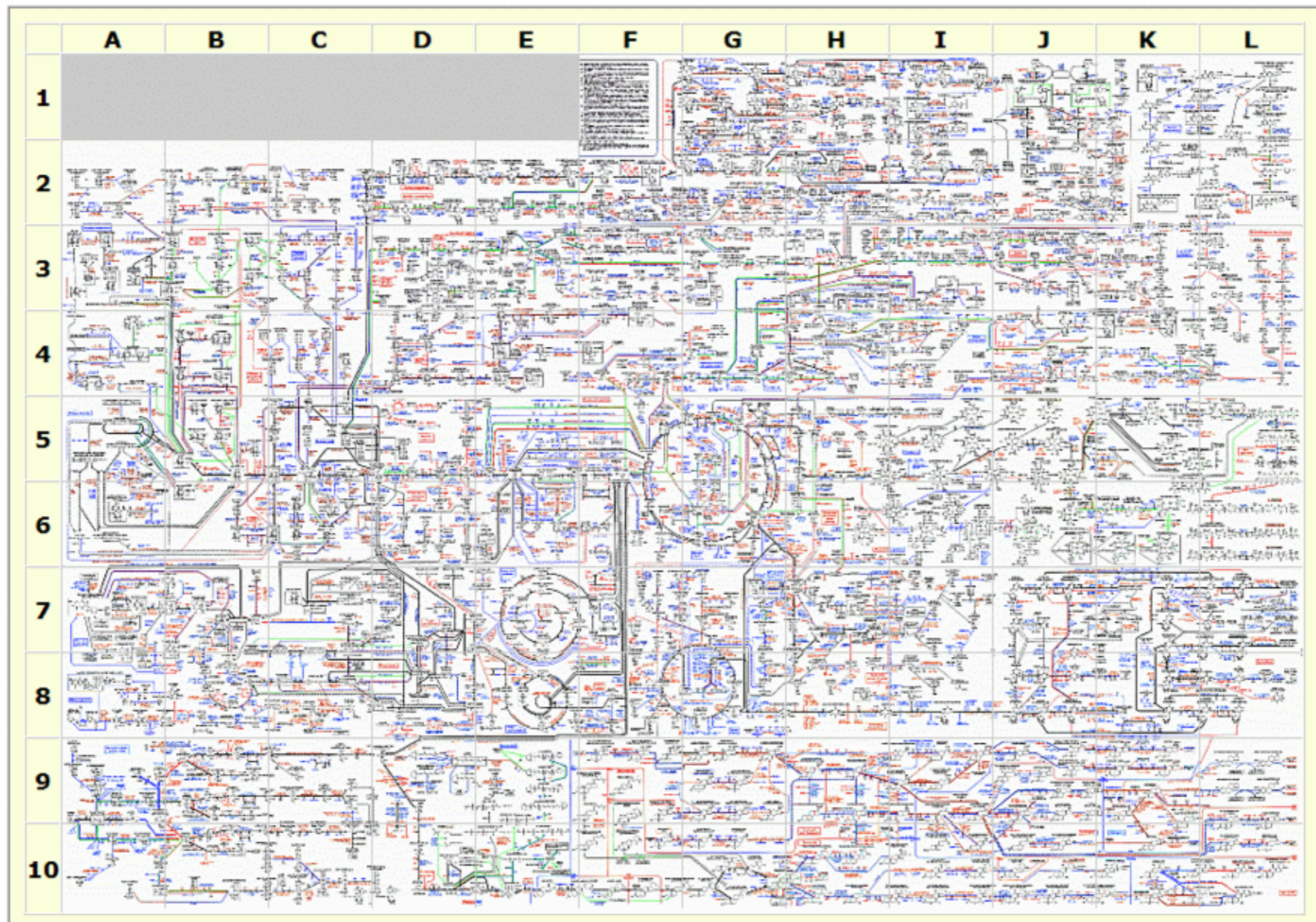
institute for art, science and technology

# Proteins



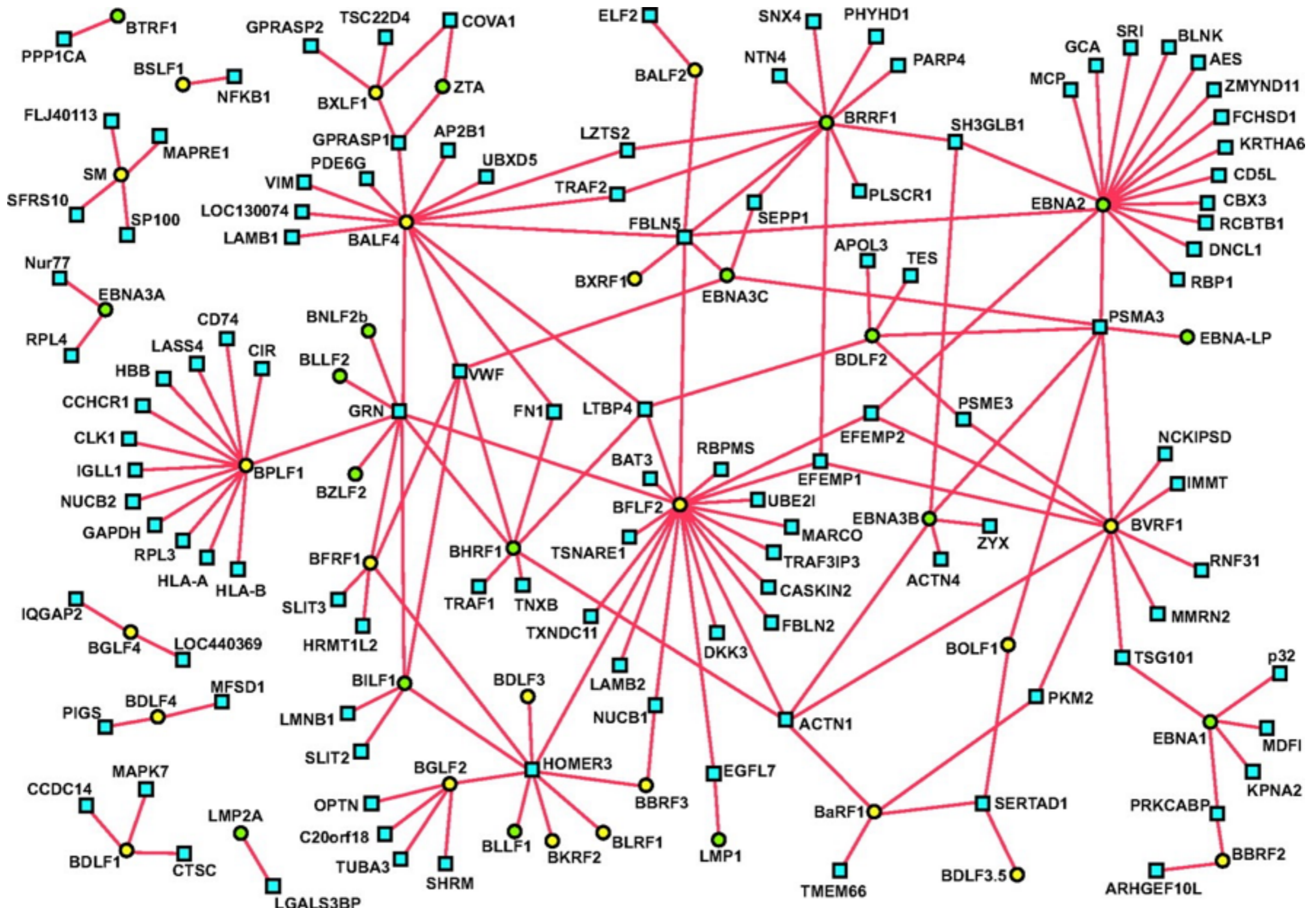


# Biochemical Pathways of the Cell





# Protein interaction mapping: MS

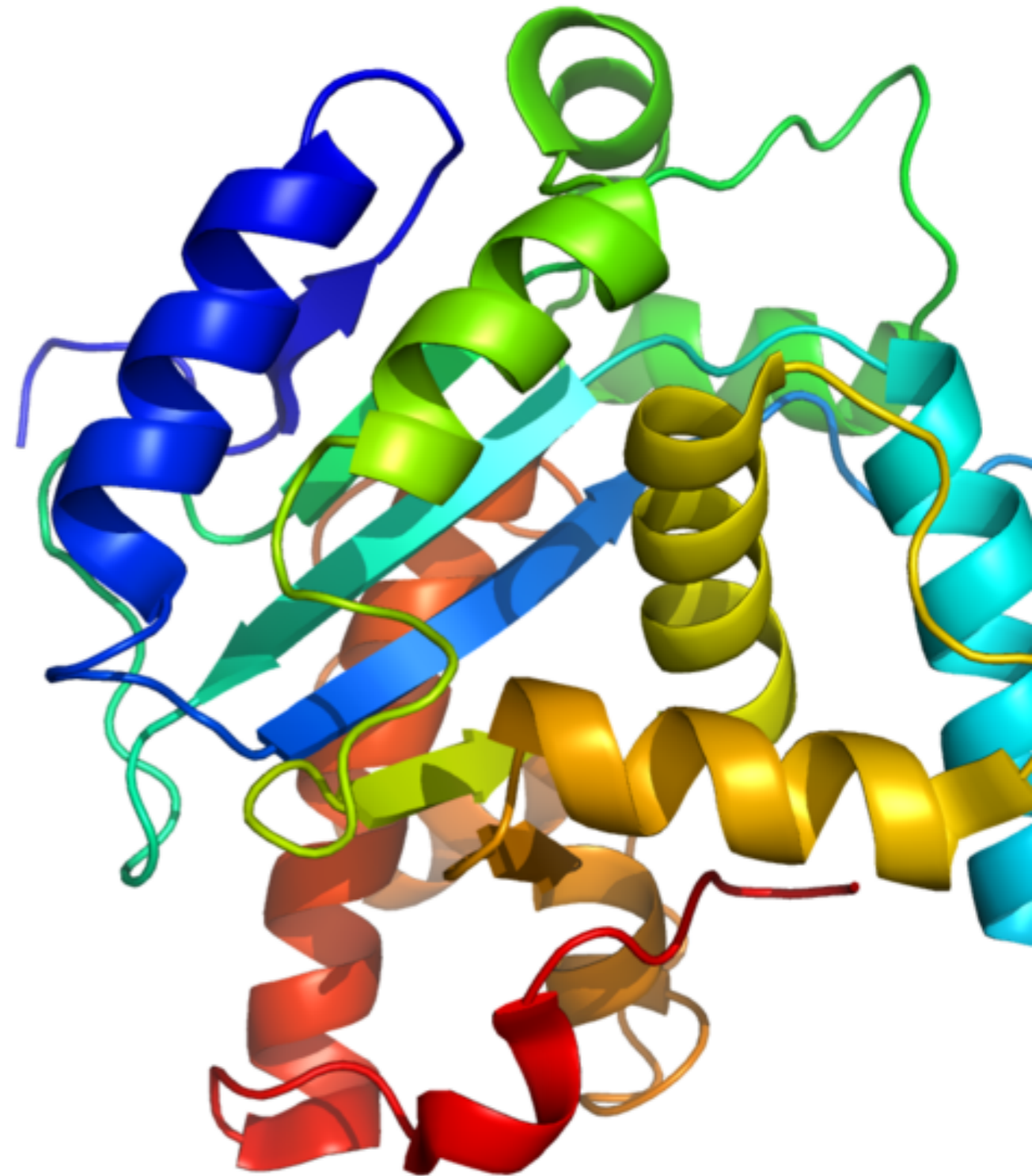




# What is this?

Simon Eugstar - CC-BY-SA 3.0

Debstart - CC-BY-SA 3.0





# NMR Machines

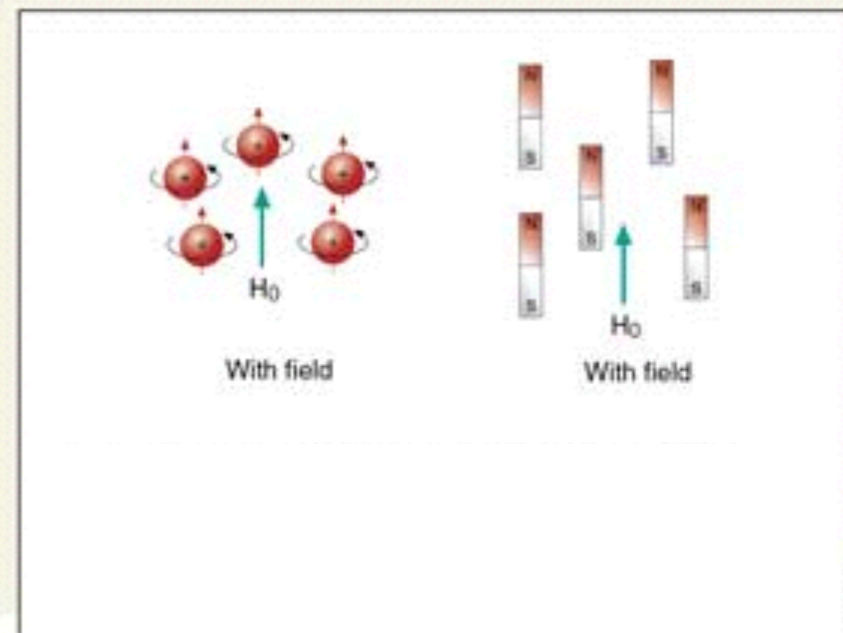
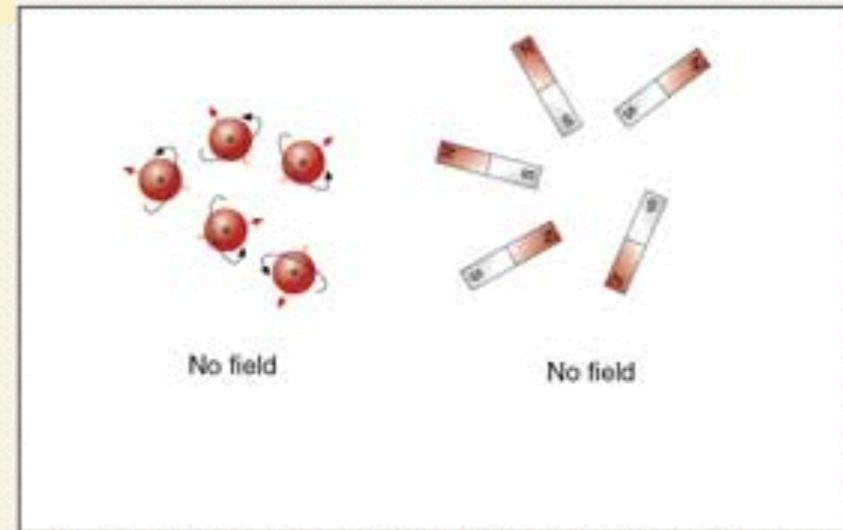
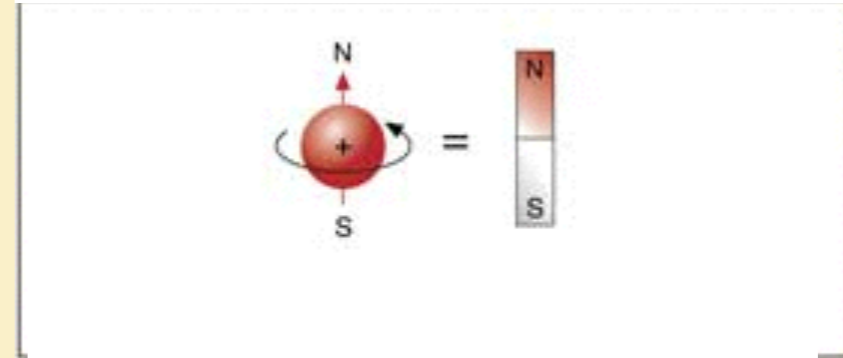
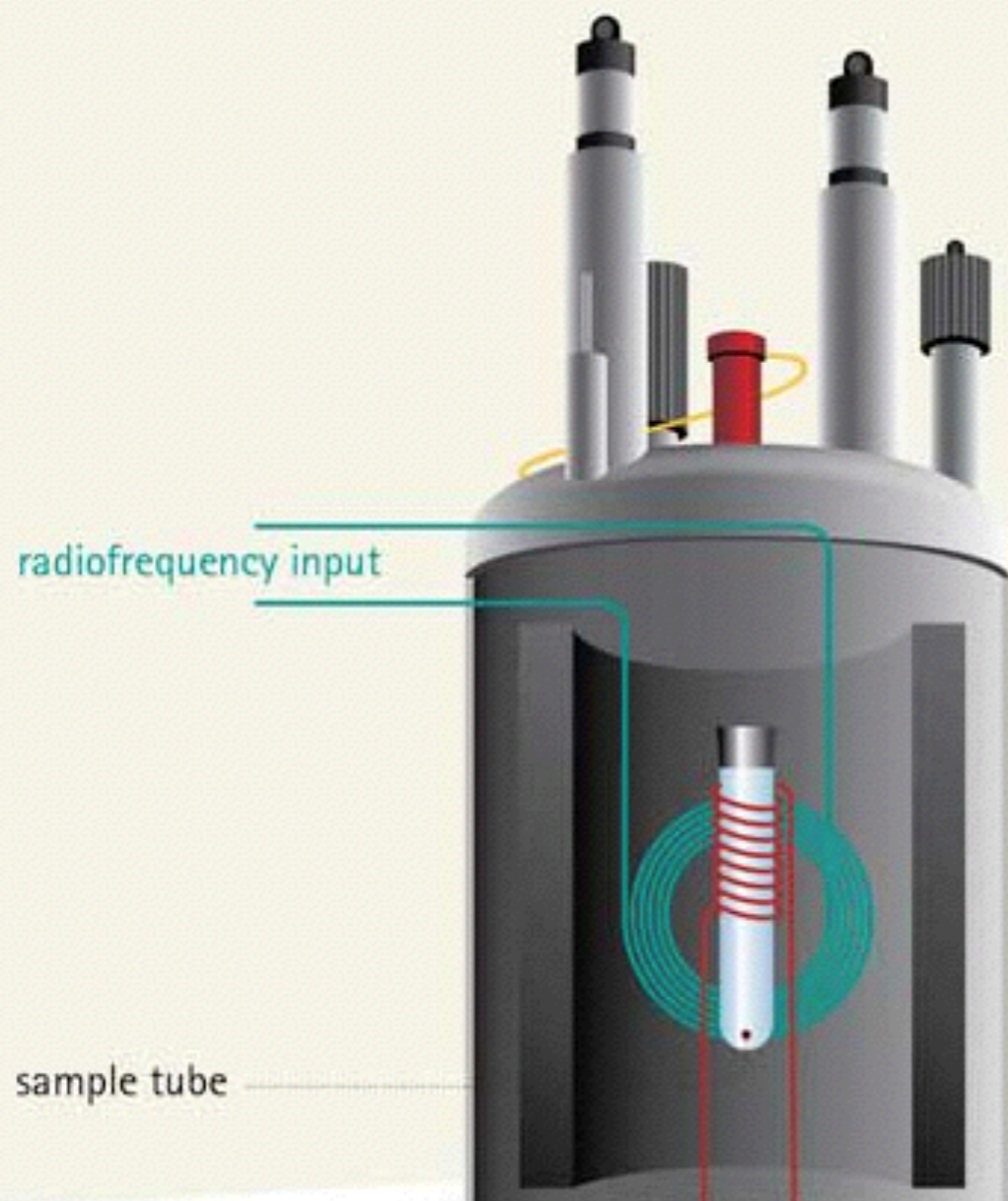
MartinSaunders

Public Domain





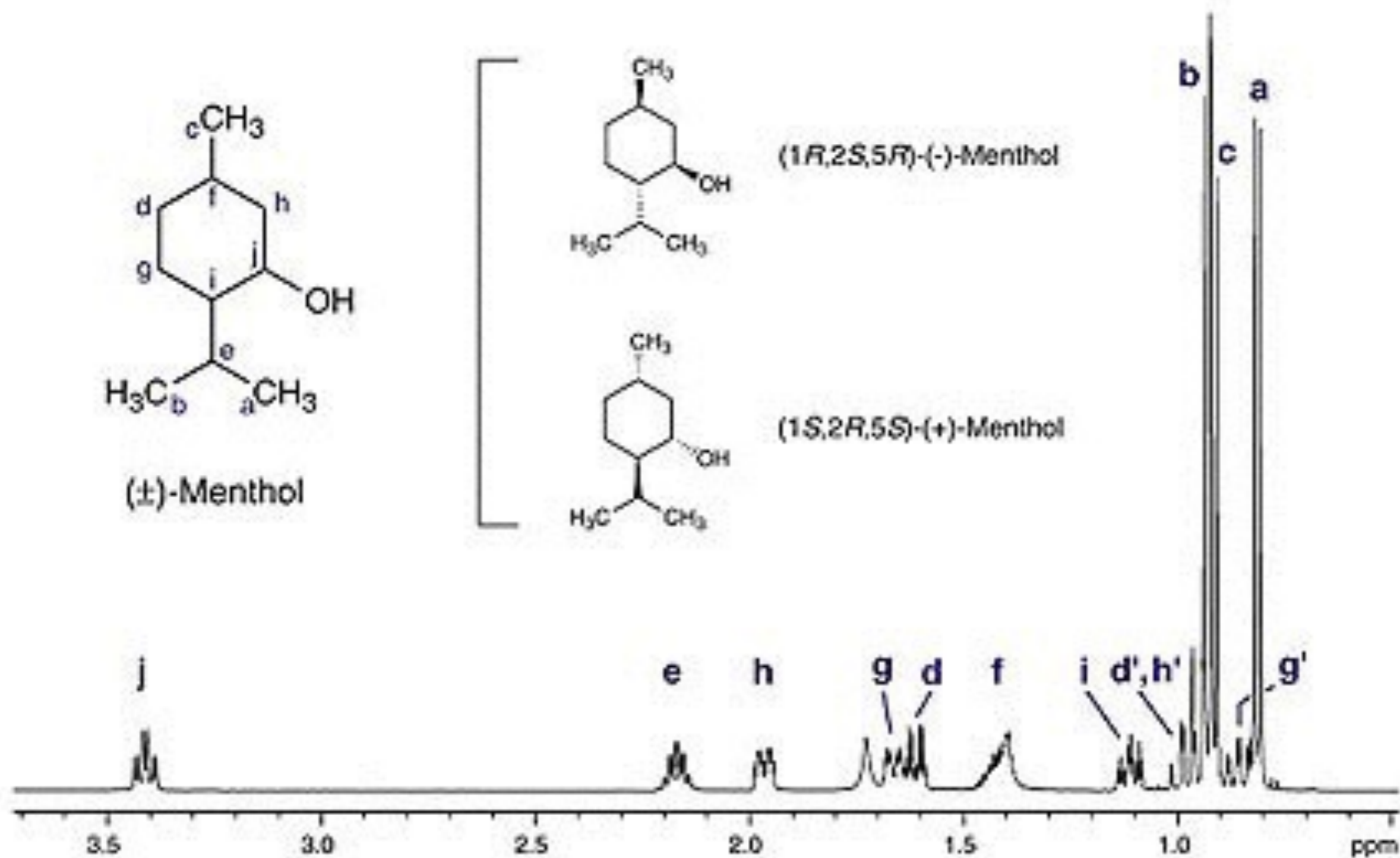
# NMR principles





# NMR Spectrum

## 1D PROTON SPECTRUM

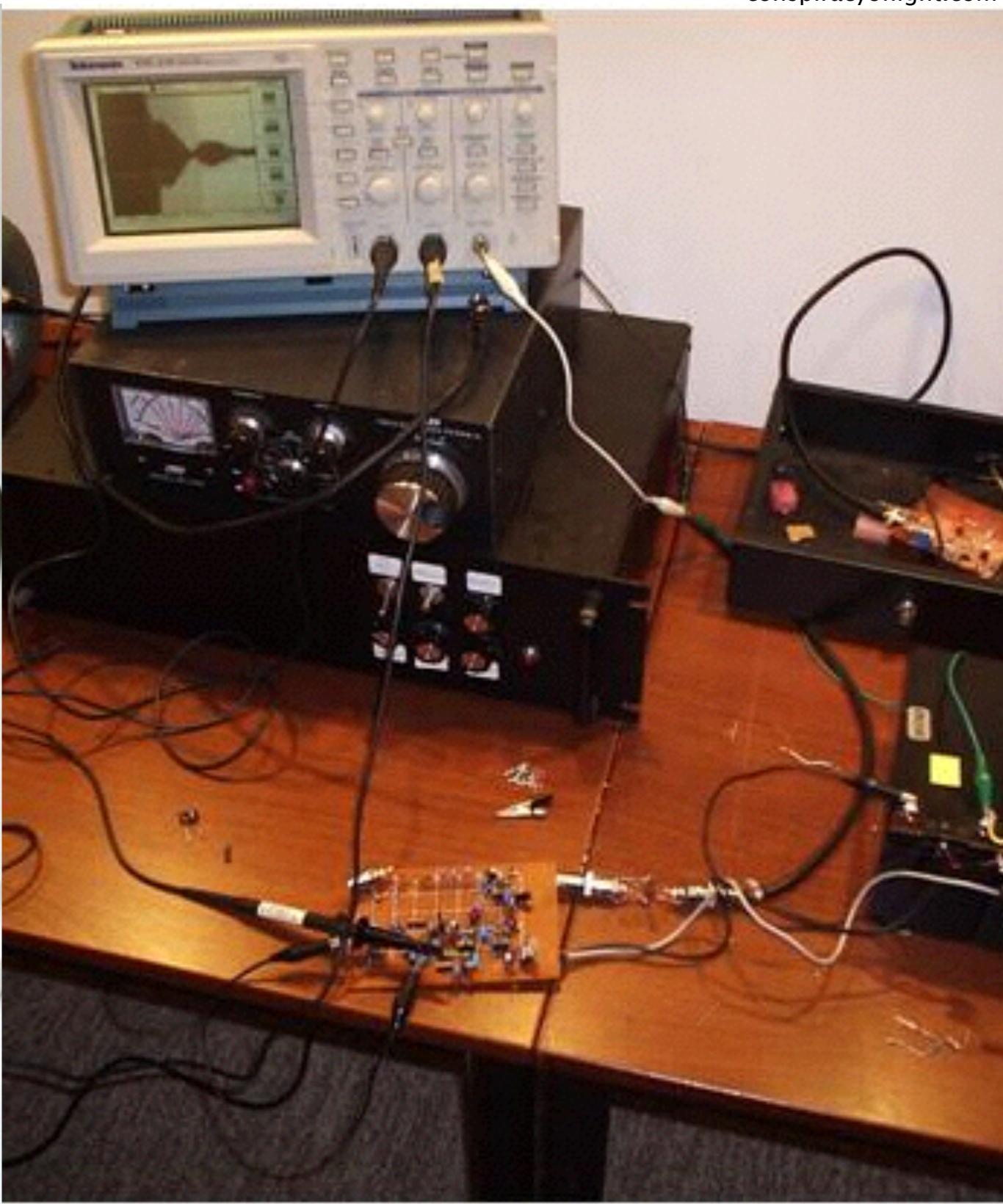
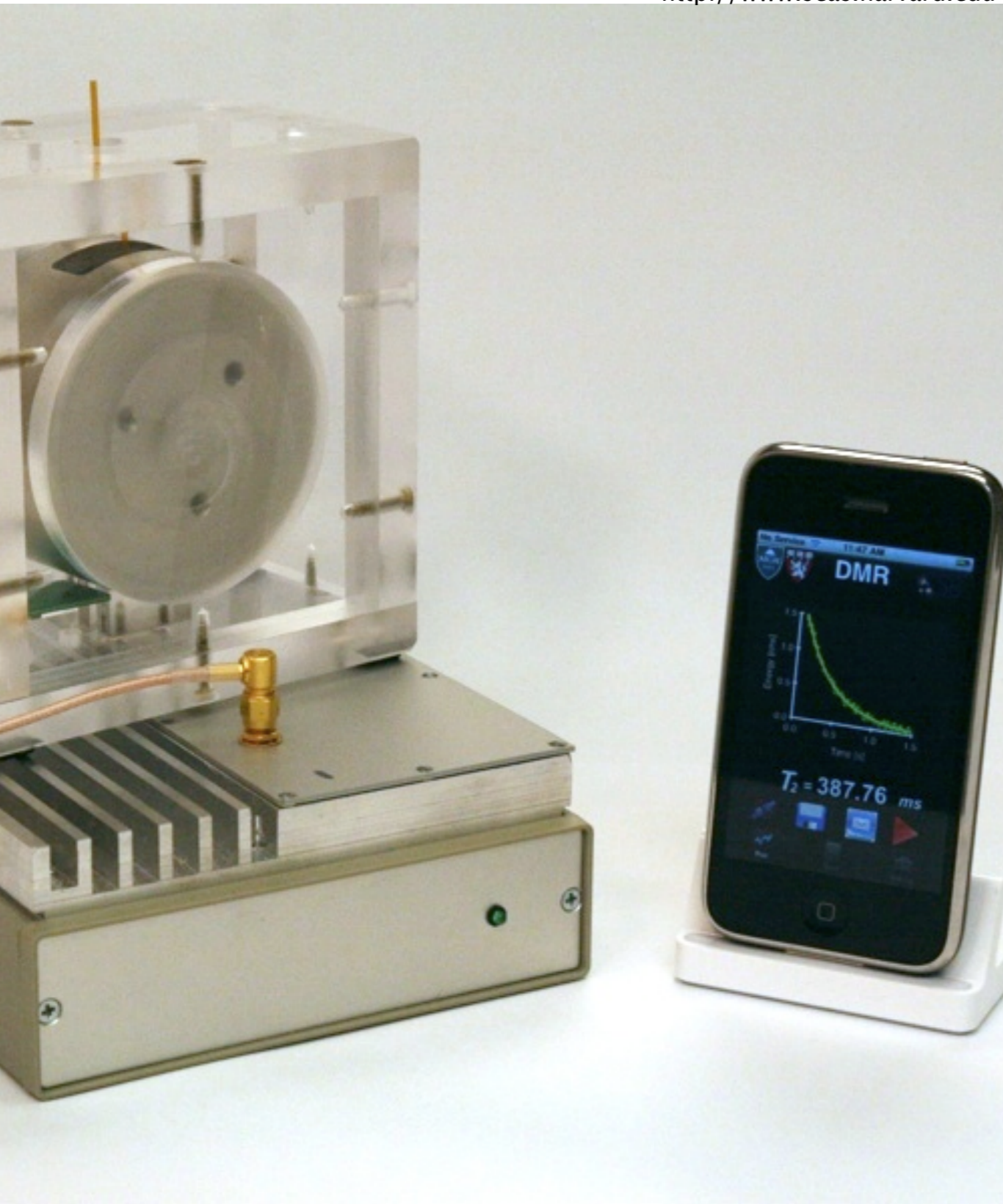




# DIY NMR?

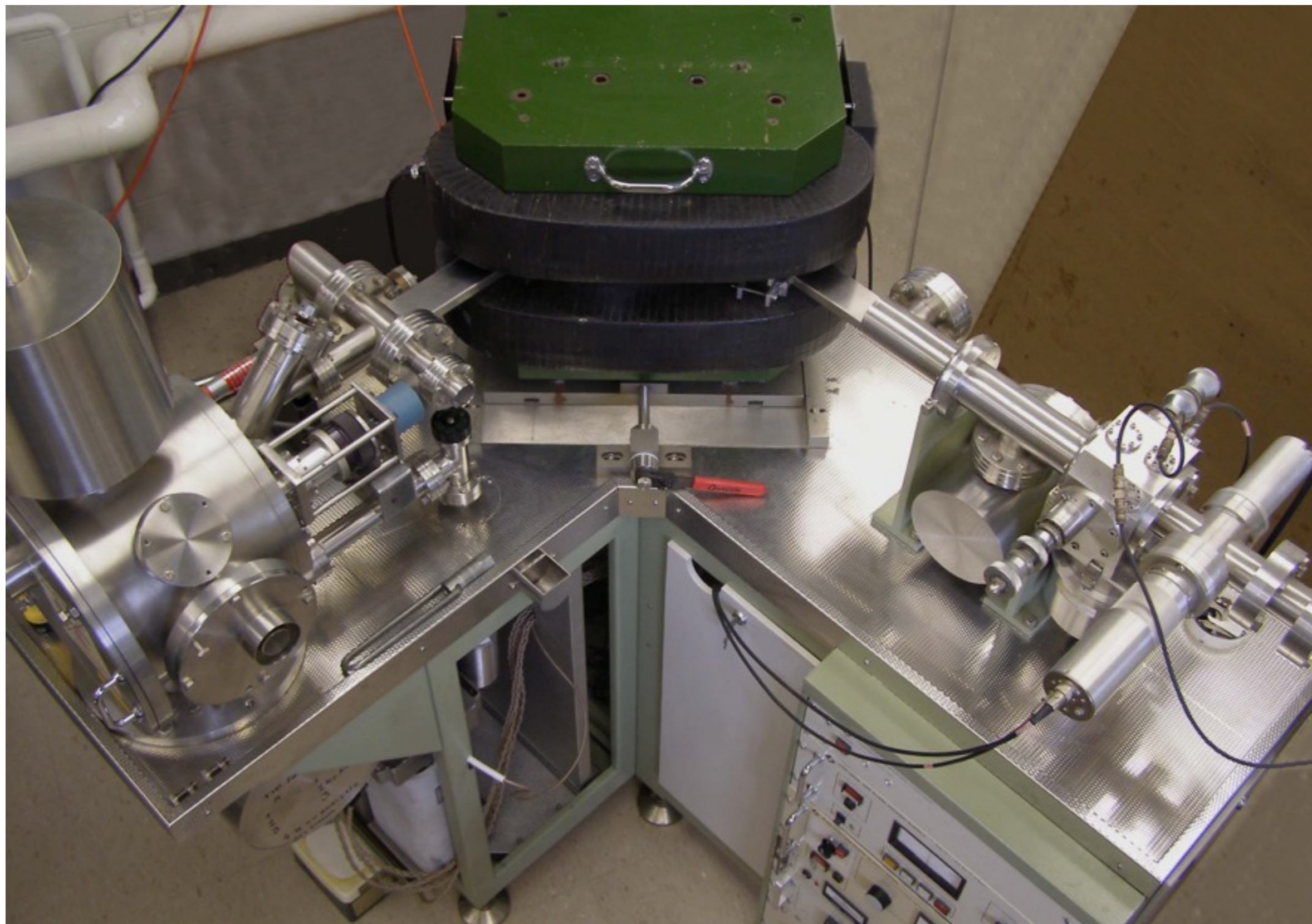
<http://www.seas.harvard.edu>

[conspiracyoflight.com](http://conspiracyoflight.com)





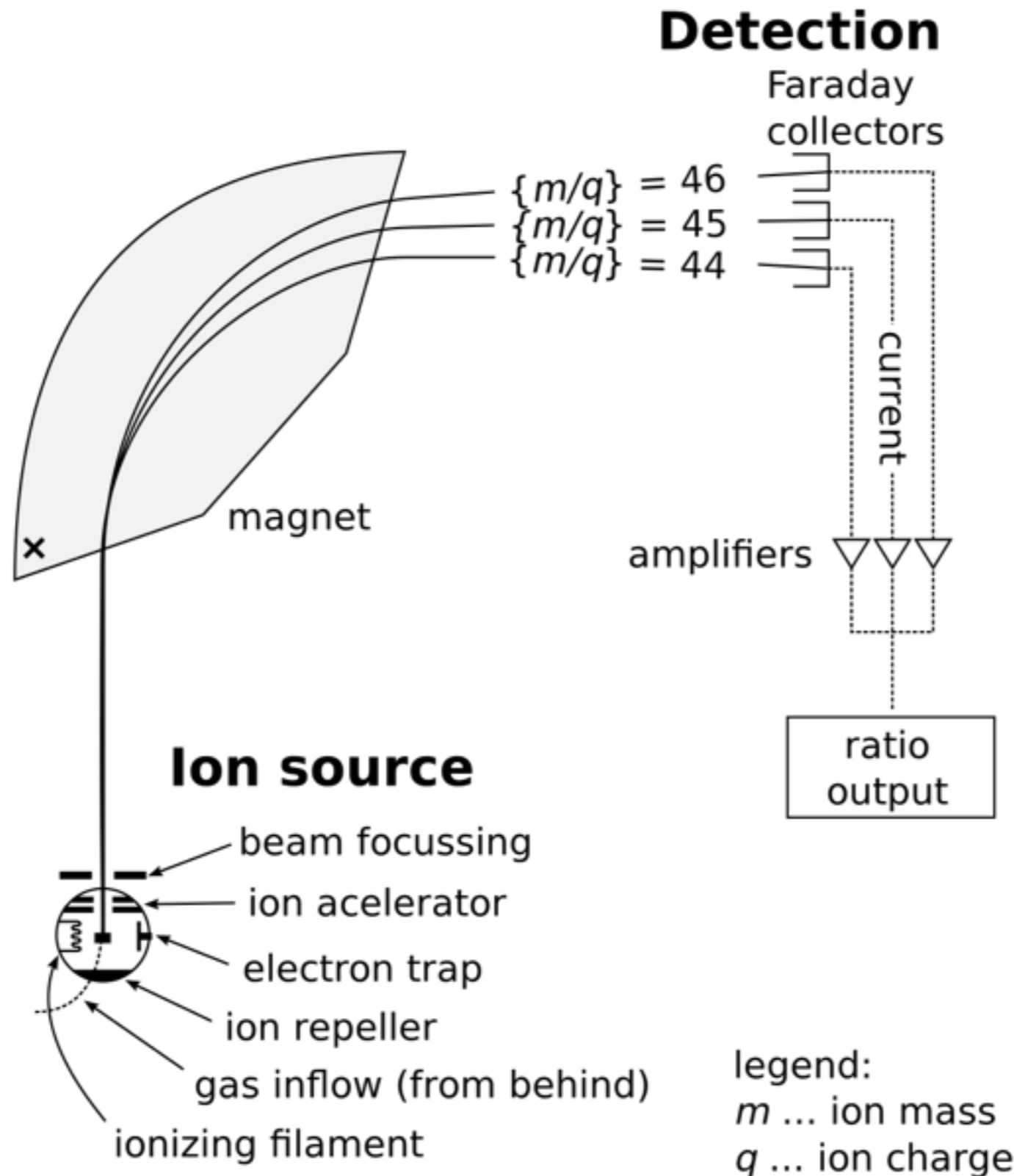
# Mass Spectrometer





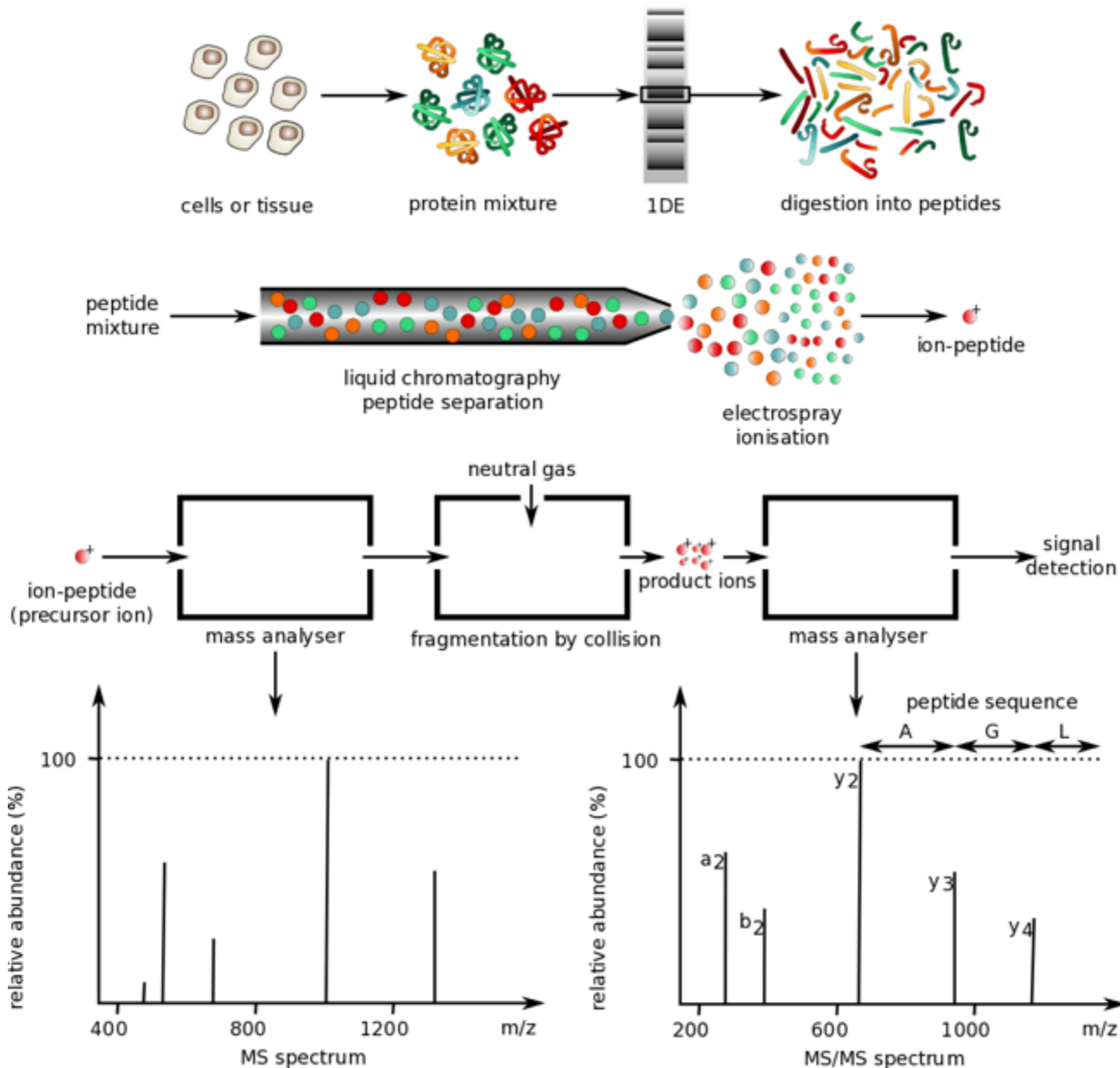


# Simple Mass Spectrometry



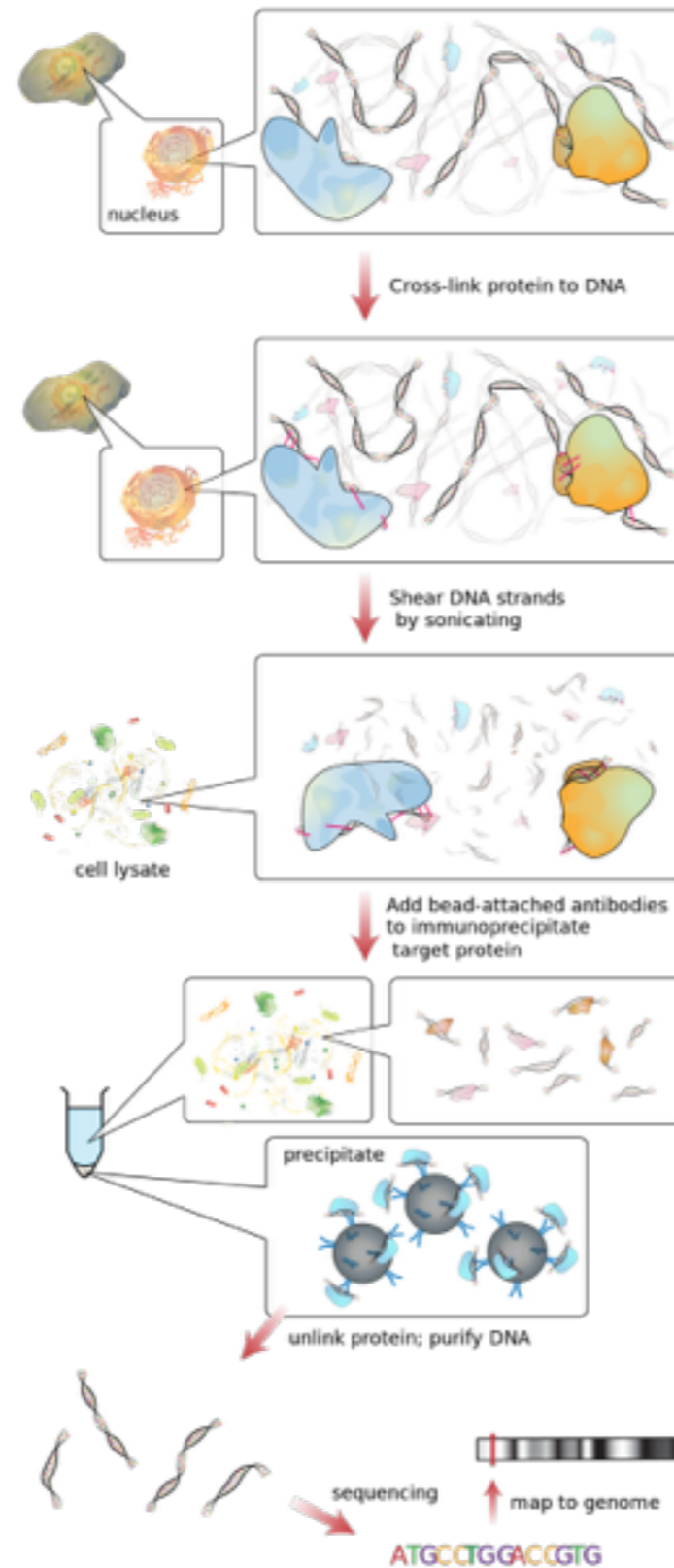


# Mass Spectrometry





# Chromatin Immunoprecipitation ChIP





# Visualisation PyMol

The screenshot displays the PyMOL Molecular Graphics System interface. The top window, titled "The PyMOL Molecular Graphics System", contains a command-line interface with the following text:

```
parser: no matching commands.  
PyMOL>map_double  
Parsing-Error: missing required argument: name  
PyMOL>map_double map  
PyMOL>rebuild  
ObjectMesh: updating "map".  
PyMOL>ray  
Ray: render time: 18.13 sec. = 198.6 frames/hour (61.42 sec. accum.).  
PyMOL>
```

The bottom window, titled "PyMOL Viewer", shows a 3D molecular model. The model consists of a green and blue ribbon structure representing the protein backbone, with red spheres representing atoms. The model is overlaid on a white wireframe mesh representing the electron density map. The interface includes a menu bar (File, Edit, Build, Movie, Display, Setting, Scene, Mouse, Wizard, Plugin) and a toolbar with buttons for Reset, Zoom, Orient, Draw, Ray, Unpick, Deselect, Rock, GetView, Stop, Play, MClear, Rebuild, and Abort. A legend on the right side of the viewer window lists the following objects and their corresponding colors:

Object	A	S	H	L	C
all	Blue	Red	Green	Yellow	White
icll	Blue	Red	Green	Yellow	White
icll_fofc	Blue	Red	Green	Yellow	White
map	Blue	Red	Green	Yellow	White

At the bottom right of the viewer window, there is a "Mouse Mode 3-Button Viewing" section with the following controls:

```
Buttons L M R Wheel  
& Keys Rota Move MovZ Slab  
ShFt +Box -Box Clip MovS  
Ctrl +/- PkAt Pk1 MvSZ  
CtSh Sele Orig Clip MovZ  
SnglClk +/- Cent Menu  
DbClk Menu - PkAt  
Selecting Residues  
State 1/ 1
```



# Issues

- Ethical
  - Who owns bio data?
  - Who decides what to use data for?
  - Is de-personalized bio information possible?
- Legal
  - Genetic discrimination
- Imagine:
  - You are immune to Zika virus. Are you entitled to royalties on the vaccine derived from your blood?



**some  
rights  
reserved**