Proteomics

What is proteomics?

- Studying the protein content of a particular organism, organ or cell type
- We have to resolve all the proteins in a particular system
- This can number over 40 000 proteins
- Proteomics is part of functional genomics

Why use Proteomics ?

- Allows us to compare control and experimental samples to look for differences in the protein profile
- Looks directly at the protein level rather than trying to make inferences from genetic information
- We can eventually use the information gathered to isolate the genes involved

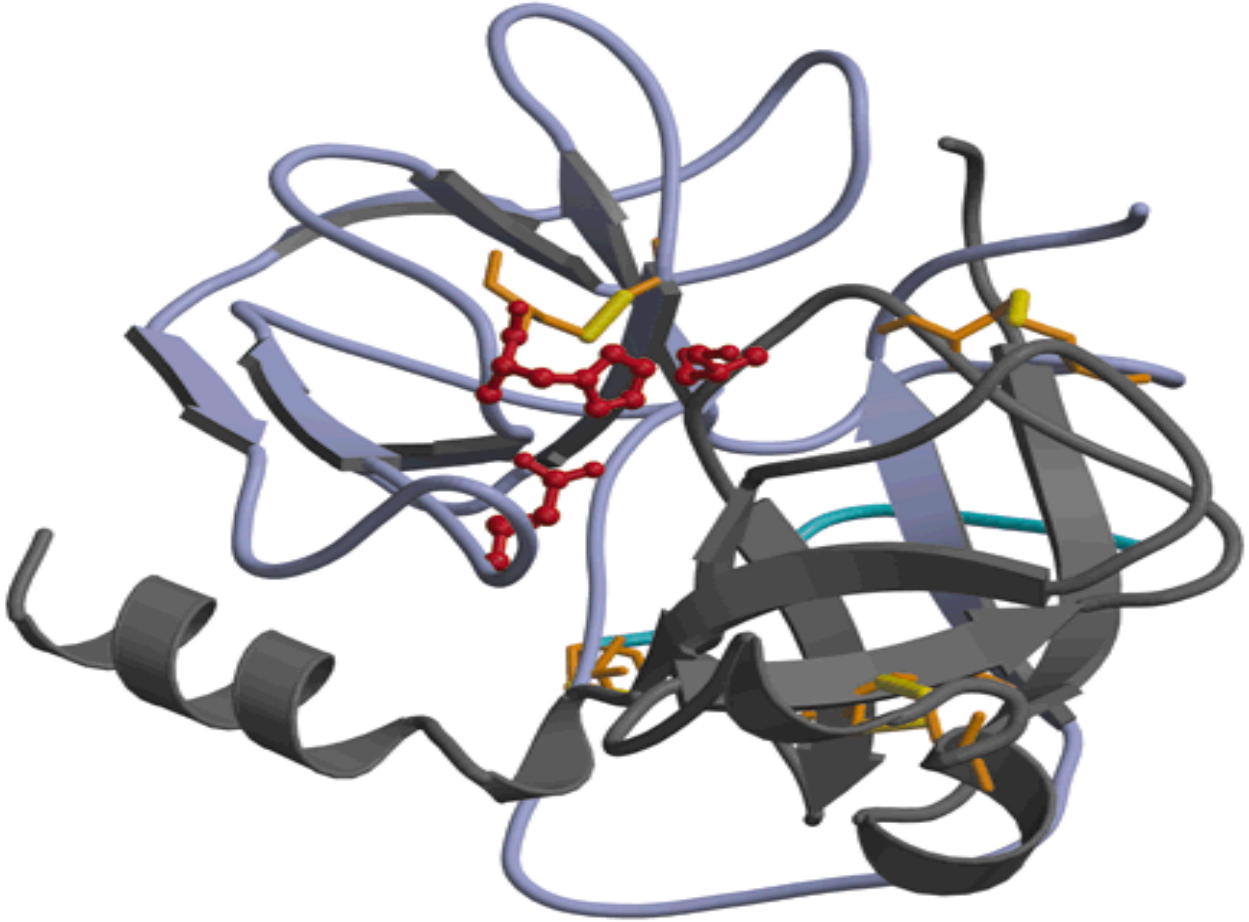
Genomics can be limited.

- From a protein function aspect there maybe post-translational modifications. These modifications may not be apparent from gene sequence
  - Phosphorylation
  - Glycosylation
  - Gernylation (C20) Membrane assoc.
  - Palmitoylation (C16) Membrane assoc.

Proteomics

- Divided up into two distinct areas
  - 2-D gel electrophoresis for the characterisation of protein profiles
  - Mass spectroscopy for the accurate identification of isolated proteins, including post-translational modifications.

Proteins are complicated macro-molecules



We can be reductionist and look at amino acids….

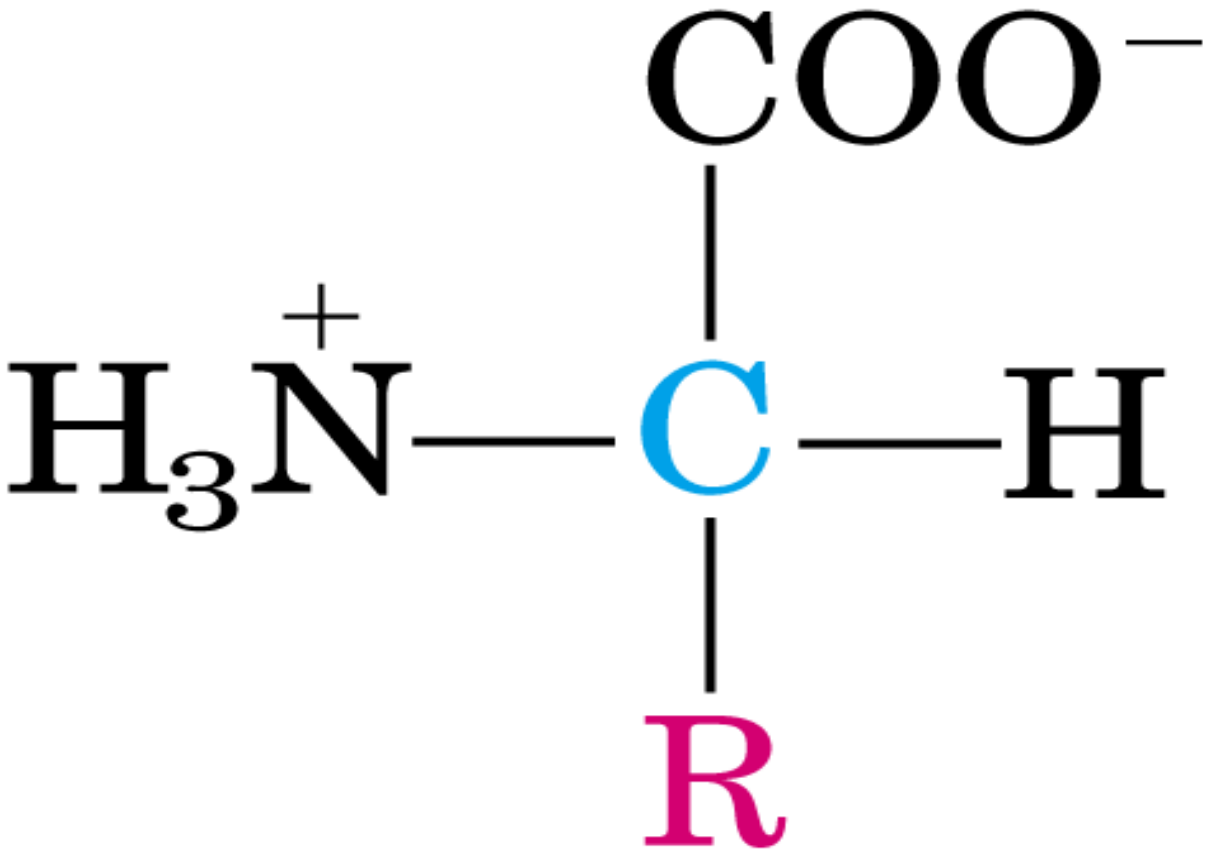


table 5–1

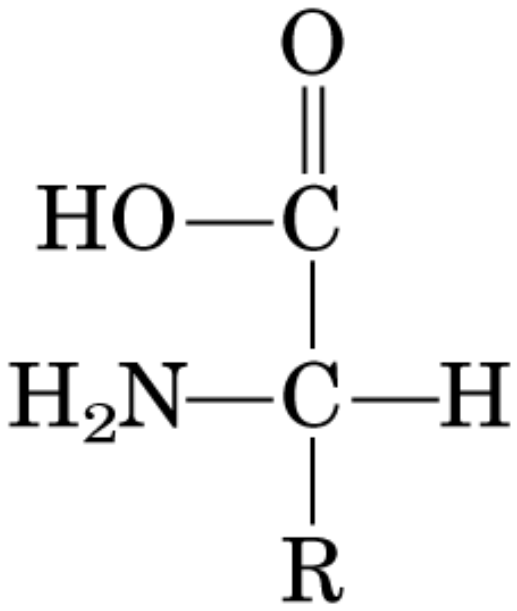
Properties and Conventions Associated with the Standard Amino Acids

Amino acid	Abbreviated names		M_r	pK_a values: pK_1 (—COOH)	pK_2 ($—NH_3^+$)	pK_R (R group)	pI	Hydropathy index*	Occurrence in proteins (%)†
Nonpolar, aliphatic R groups									
Glycine	Gly	G	75	2.34	9.60		5.97	−0.4	7.2
Alanine	Ala	A	89	2.34	9.69		6.01	1.8	7.8
Valine	Val	V	117	2.32	9.62		5.97	4.2	6.6
Leucine	Leu	L	131	2.36	9.60		5.98	3.8	9.1
Isoleucine	Ile	I	131	2.36	9.68		6.02	4.5	5.3
Methionine	Met	M	149	2.28	9.21		5.74	1.9	2.3
Aromatic R groups									
Phenylalanine	Phe	F	165	1.83	9.13		5.48	2.8	3.9
Tyrosine	Tyr	Y	181	2.20	9.11	10.07	5.66	−1.3	3.2
Tryptophan	Trp	W	204	2.38	9.39		5.89	−0.9	1.4
Polar, uncharged R groups									
Serine	Ser	S	105	2.21	9.15		5.68	−0.8	6.8
Proline	Pro	P	115	1.99	10.96		6.48	1.6	5.2
Threonine	Thr	T	119	2.11	9.62		5.87	−0.7	5.9
Cysteine	Cys	C	121	1.96	10.28	8.18	5.07	2.5	1.9
Asparagine	Asn	N	132	2.02	8.80		5.41	−3.5	4.3
Glutamine	Gln	Q	146	2.17	9.13		5.65	−3.5	4.2
Positively charged R groups									
Lysine	Lys	K	146	2.18	8.95	10.53	9.74	−3.9	5.9
Histidine	His	H	155	1.82	9.17	6.00	7.59	−3.2	2.3
Arginine	Arg	R	174	2.17	9.04	12.48	10.76	−4.5	5.1
Negatively charged R groups									
Aspartate	Asp	D	133	1.88	9.60	3.65	2.77	−3.5	5.3
Glutamate	Glu	E	147	2.19	9.67	4.25	3.22	−3.5	6.3

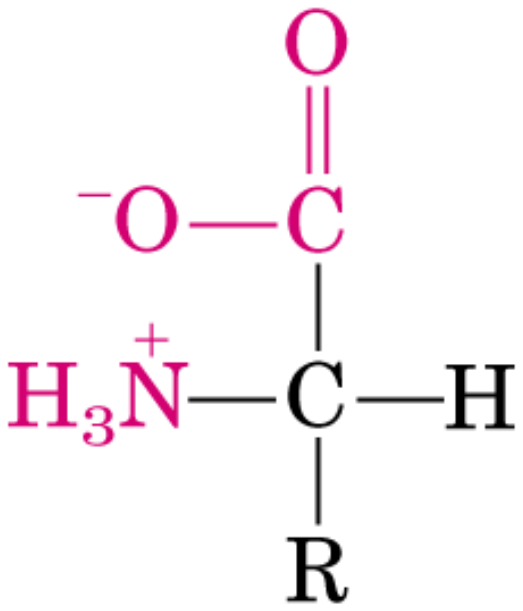
*A scale combining hydrophobicity and hydrophilicity of R groups; it can be used to measure the tendency of an amino acid to seek an aqueous environment (− values) or a hydrophobic environment (+ values). See Chapter 12. From Kyte, J. & Doolittle, R.F. (1982) *J. Mol. Biol.* **157**, 105 – 132.

†Average occurrence in over 1150 proteins. From Doolittle, R.F. (1989) Redundancies in protein sequences. In *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G.D., ed) Plenum Press, NY, pp. 599–623.

In effect all proteins are zwitterions

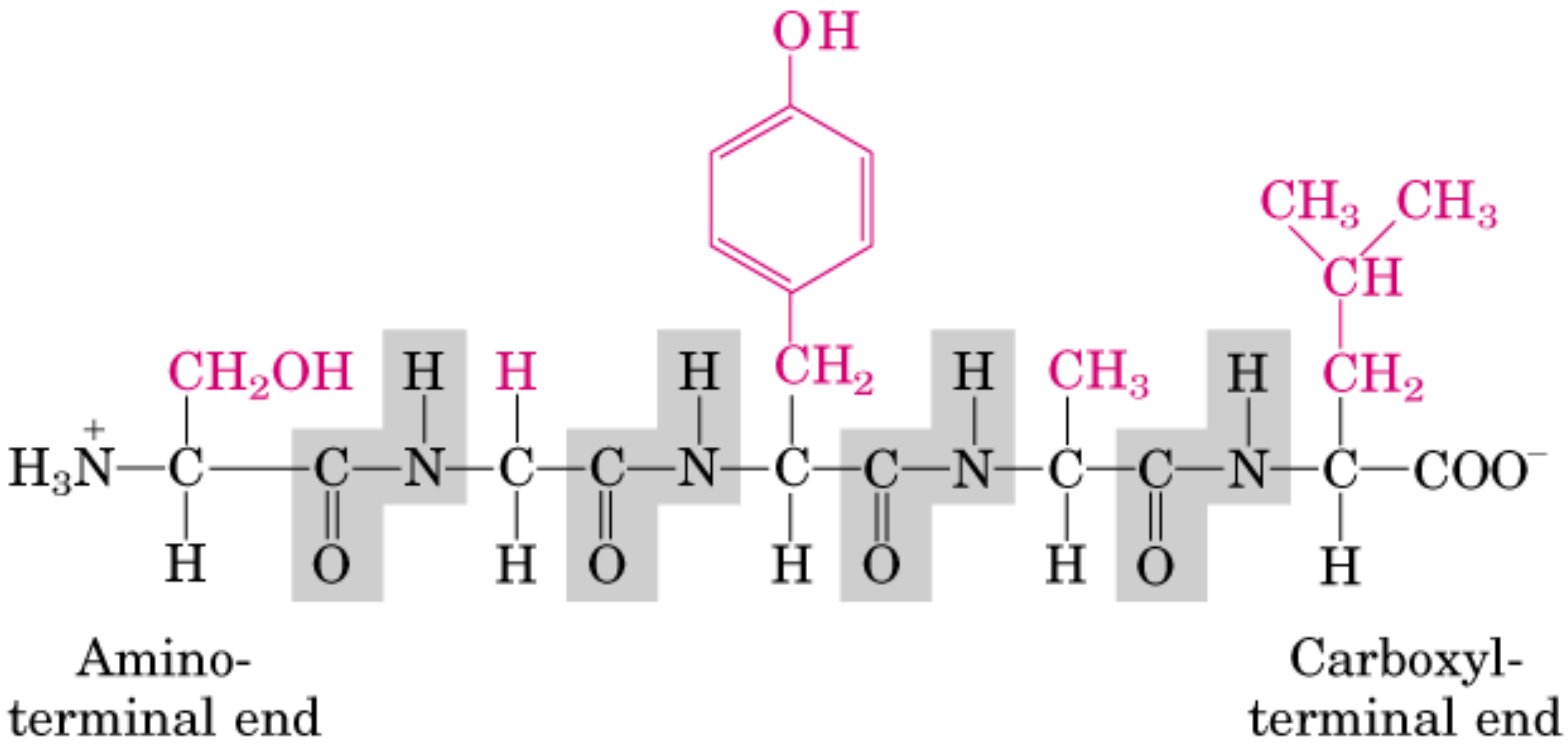


Nonionic form



Zwitterionic form

Proteins are sequential molecules with defined ends



Proteins vary greatly in size

table 5–2

Molecular Data on Some Proteins

	Molecular weight	Number of residues	Number of polypeptide chains
Cytochrome *c* (human)	13,000	104	1
Ribonuclease A (bovine pancreas)	13,700	124	1
Lysozyme (egg white)	13,930	129	1
Myoglobin (equine heart)	16,890	153	1
Chymotrypsin (bovine pancreas)	21,600	241	3
Chymotrypsinogen (bovine)	22,000	245	1
Hemoglobin (human)	64,500	574	4
Serum albumin (human)	68,500	609	1
Hexokinase (yeast)	102,000	972	2
RNA polymerase (*E. coli*)	450,000	4,158	5
Apolipoprotein B (human)	513,000	4,536	1
Glutamine synthetase (*E. coli*)	619,000	5,628	12
Titin (human)	2,993,000	26,926	1

Amino acid composition defines the physical nature of a protein

- Each amino acid has a particular pI
- These pI's combine to give a protein an over all pI.
- As proteins are zwitterions when they are maintained at their pI they have no charge.

table 5–3

Amino Acid Composition of Two Proteins*

Amino acid	Number of residues per molecule of protein	
	Bovine cytochrome *c*	Bovine chymotrypsinogen
Ala	6	22
Arg	2	4
Asn	5	15
Asp	3	8
Cys	2	10
Gln	3	10
Glu	9	5
Gly	14	23
His	3	2
Ile	6	10
Leu	6	19
Lys	18	14
Met	2	2
Phe	4	6
Pro	4	9
Ser	1	28
Thr	8	23
Trp	1	8
Tyr	4	4
Val	3	23
Total	104	245

*Note that standard procedures for the acid hydrolysis of proteins convert Asn and Gln to Asp and Glu, respectively. In addition, Trp is destroyed. Special procedures must be employed to determine the amounts of these amino acids.

Protein separation for proteomic investigations

1st dimension

Separation by charge

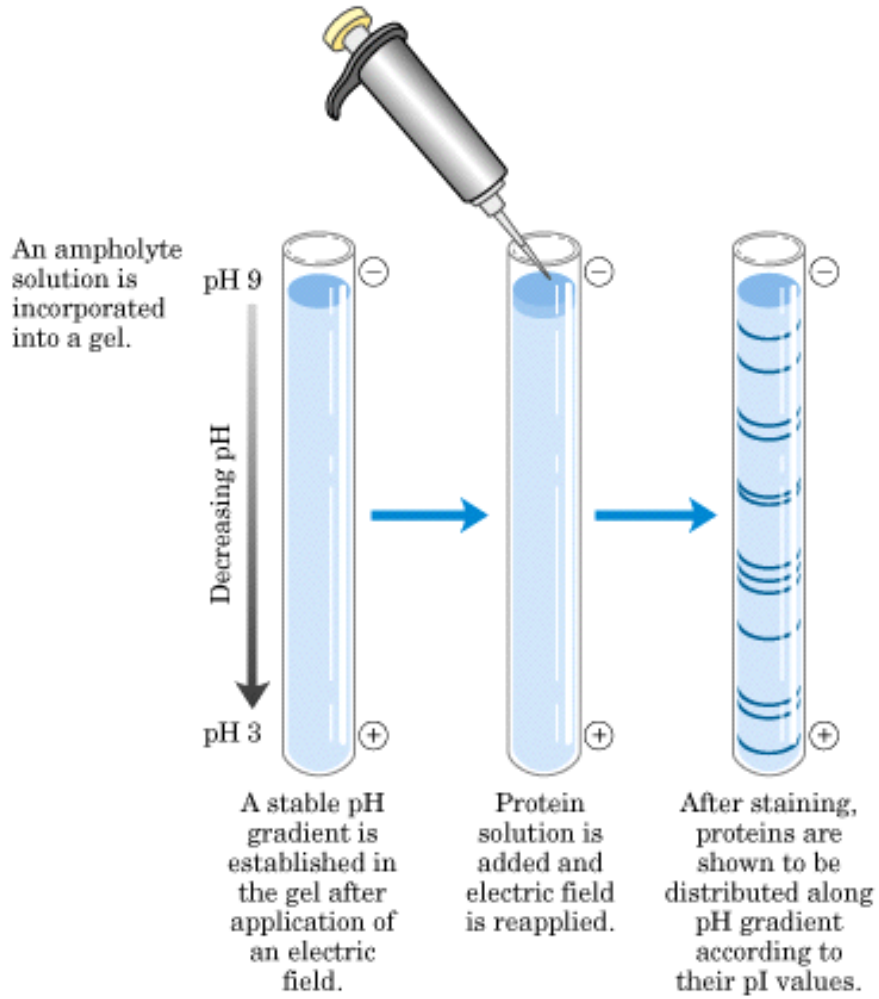
Proteins vary in isoelectric point

table 5–6

The Isoelectric Points of Some Proteins

Protein	pI
Pepsin	~1.0
Egg albumin	4.6
Serum albumin	4.9
Urease	5.0
β-Lactoglobulin	5.2
Hemoglobin	6.8
Myoglobin	7.0
Chymotrypsinogen	9.5
Cytochrome *c*	10.7
Lysozyme	11.0

Isoelectric focusing.
Separation by charge (pI)



Protein separation for proteomic investigations

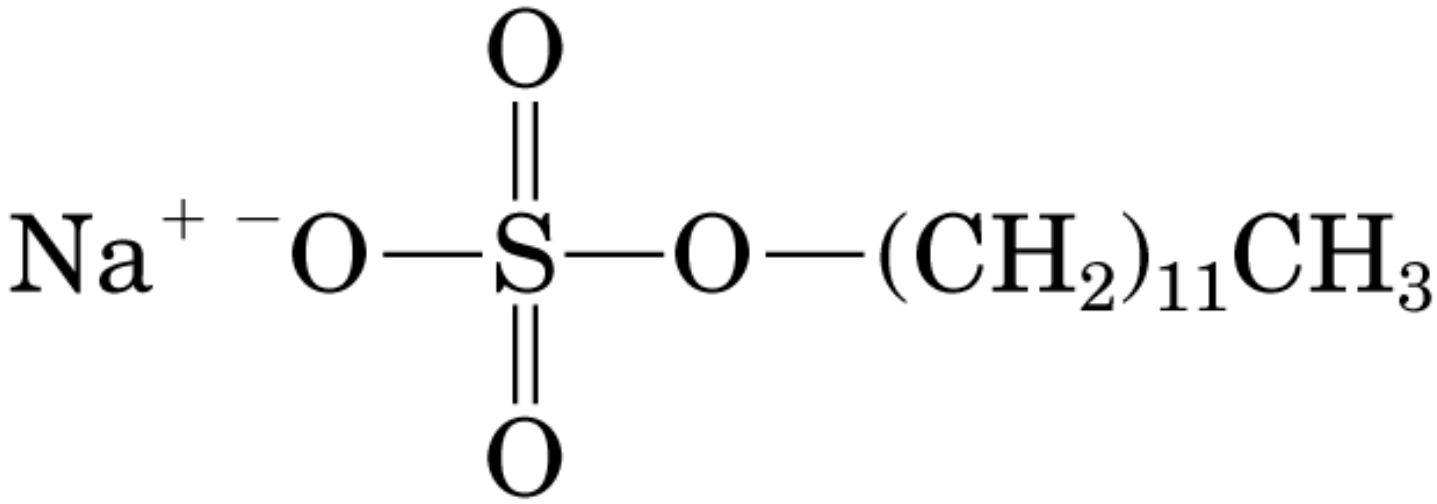
2nd dimension

Separation by size

Sodium dodecyl sulphate polyacrylamide electrophoresis (SDS-PAGE)

- System relies on two chemicals
  - Acrylamide
    - Forms a gel-like matrix of known composition
    - The “denser” a gel is made, the slower proteins will move.
  - Sodium dodecyl sulphate
    - A detergent which binds to the protein giving all proteins a gross negative charge.
    - All proteins now migrate towards a +ve charge.

SDS gives all proteins a net negative charge



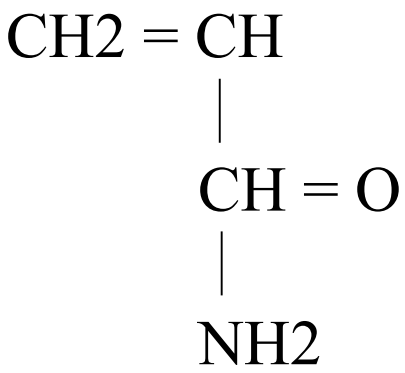
Sodium dodecyl sulfate
(SDS)

Acrylamide reaction

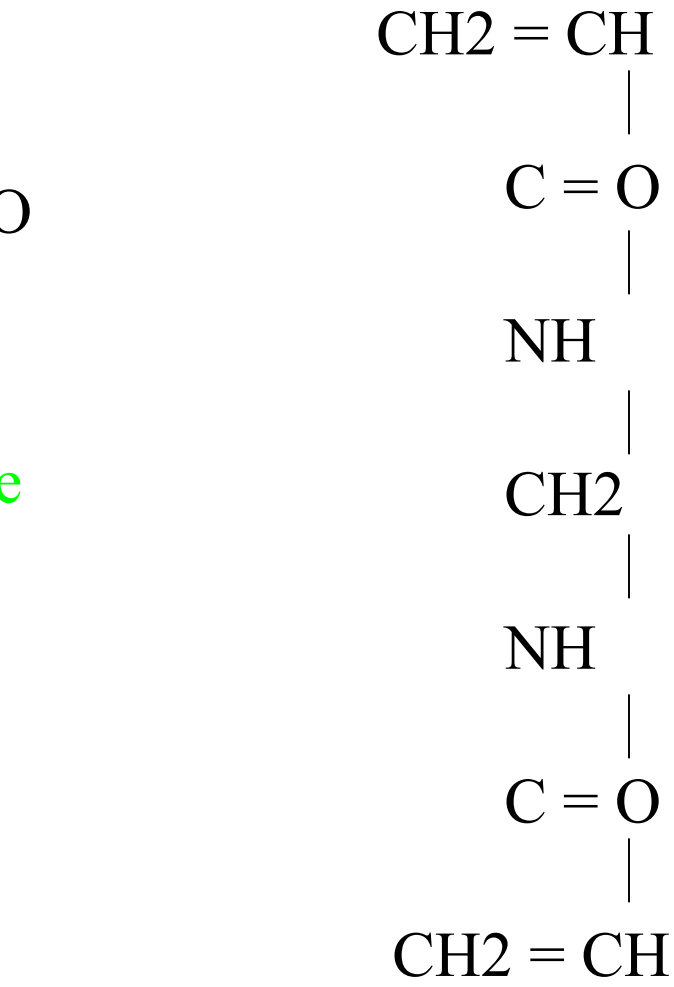
Acrylamide gels are formed by co-polymerisation of acrylamide monomer , $CH_2=CH-CO-NH_2$, and a crosslinking co-monomer, N,N'-methylenebisacrylamide, $CH_2=CH-CO-NH-CH_2-NH-CO-CH=CH_2$, bisacrylanide

The mechanism of gel formation is vinyl addition polymerisation and is catalysed by free radical-generation system composed of ammonium persulphate (initiator) and accelerator tetramethylethylenediamine (TEMED).

Components of acrylamide

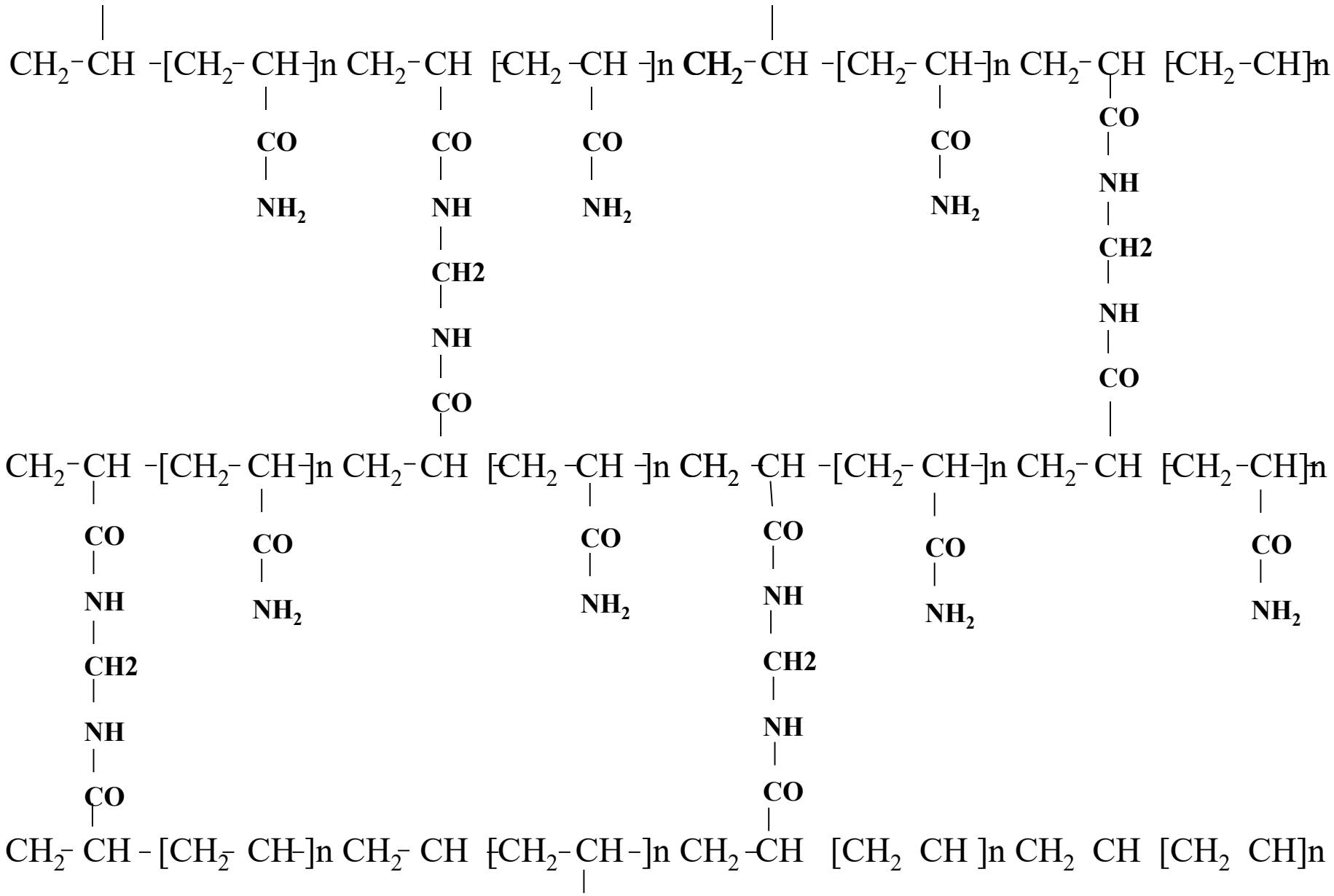


Acrylamide

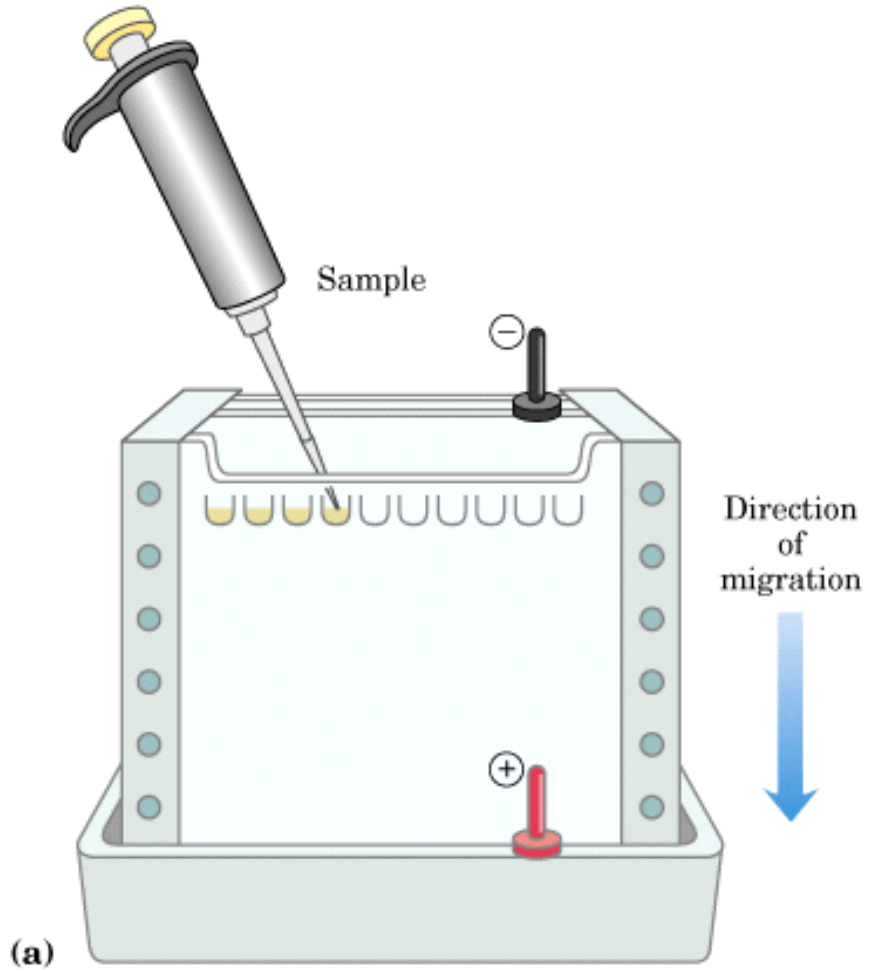


N,N’-methylene bisacrylamide

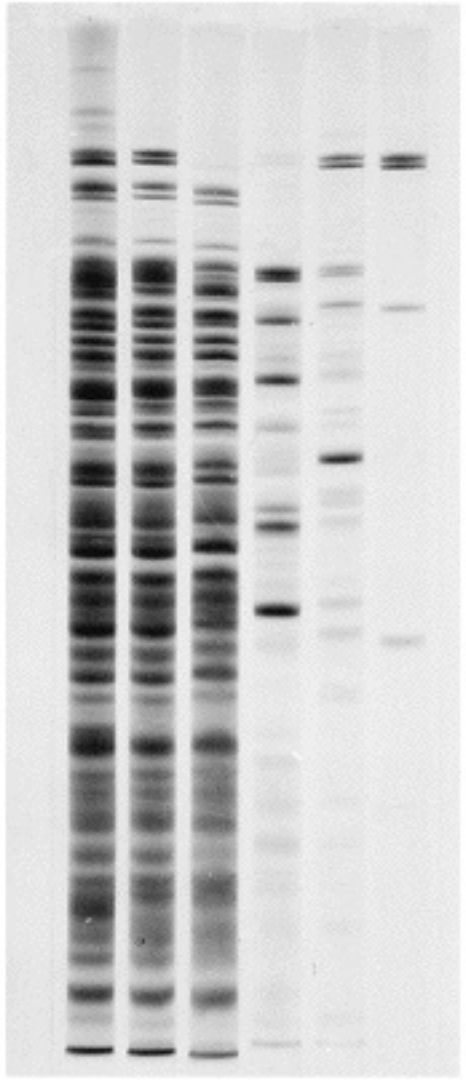
Acrylamide polymerisation



When run on a gel all proteins will move dependant on molecular weight alone

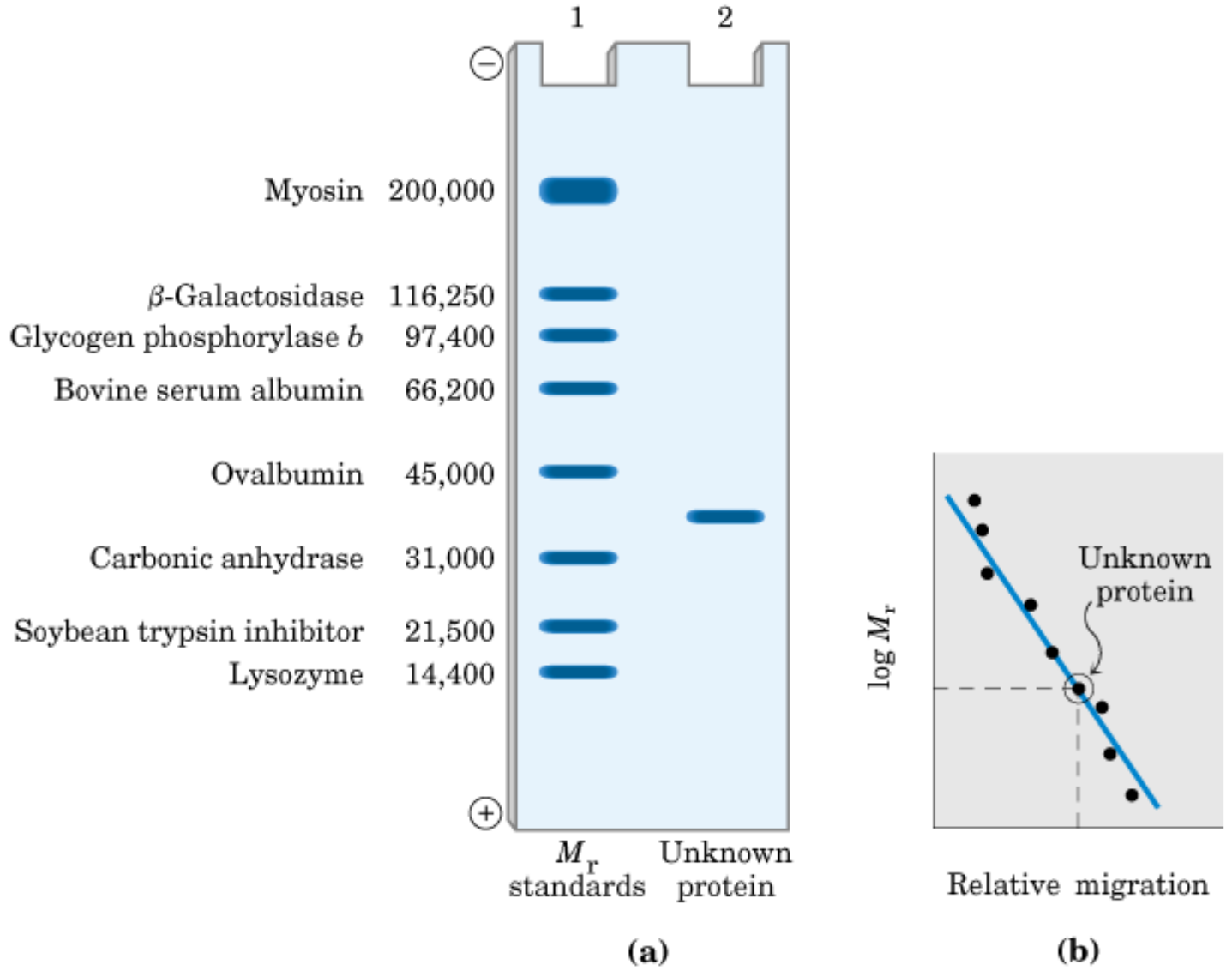


Example of protein separation on PA gel

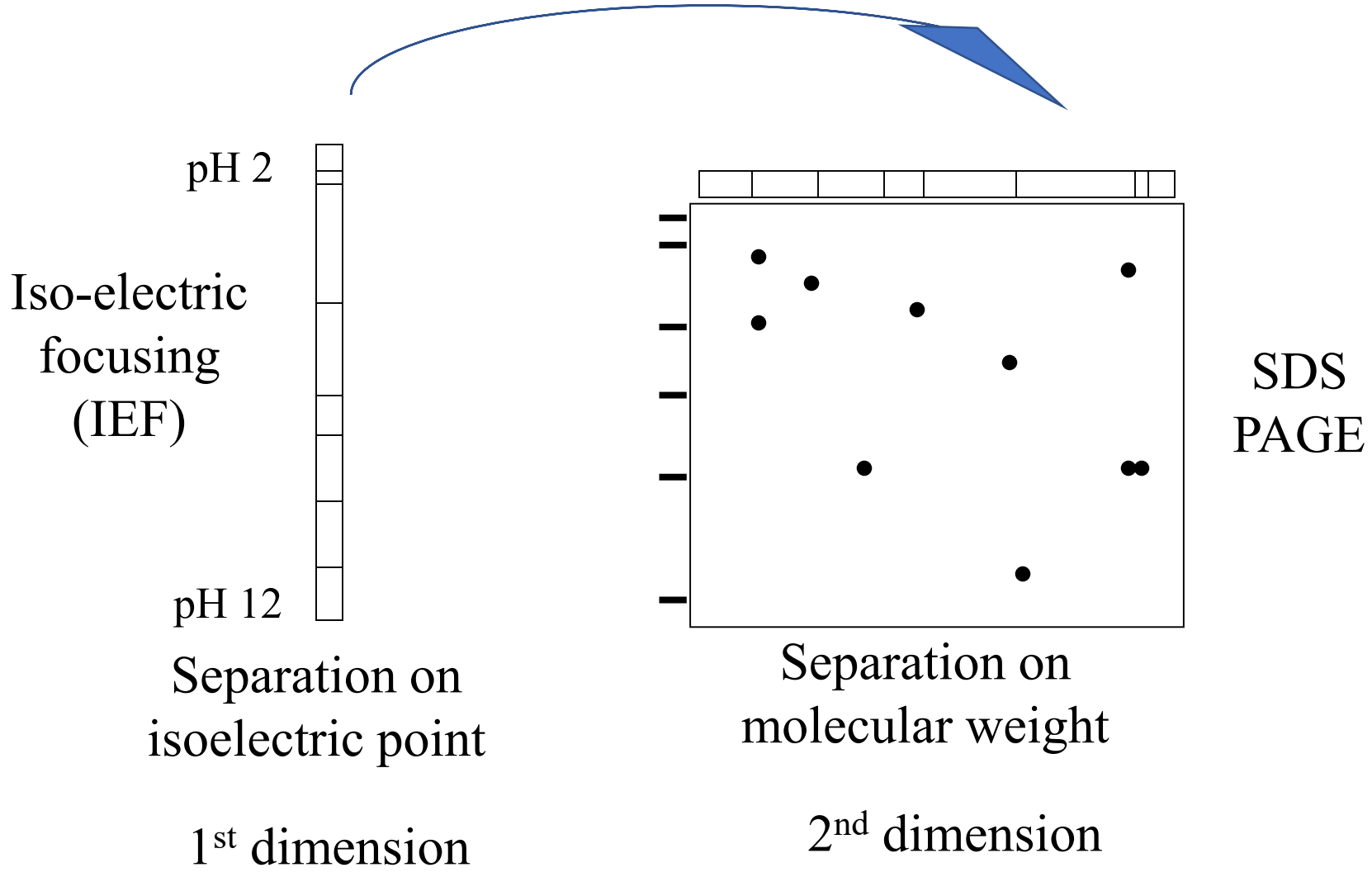


(b)

Sizing of proteins

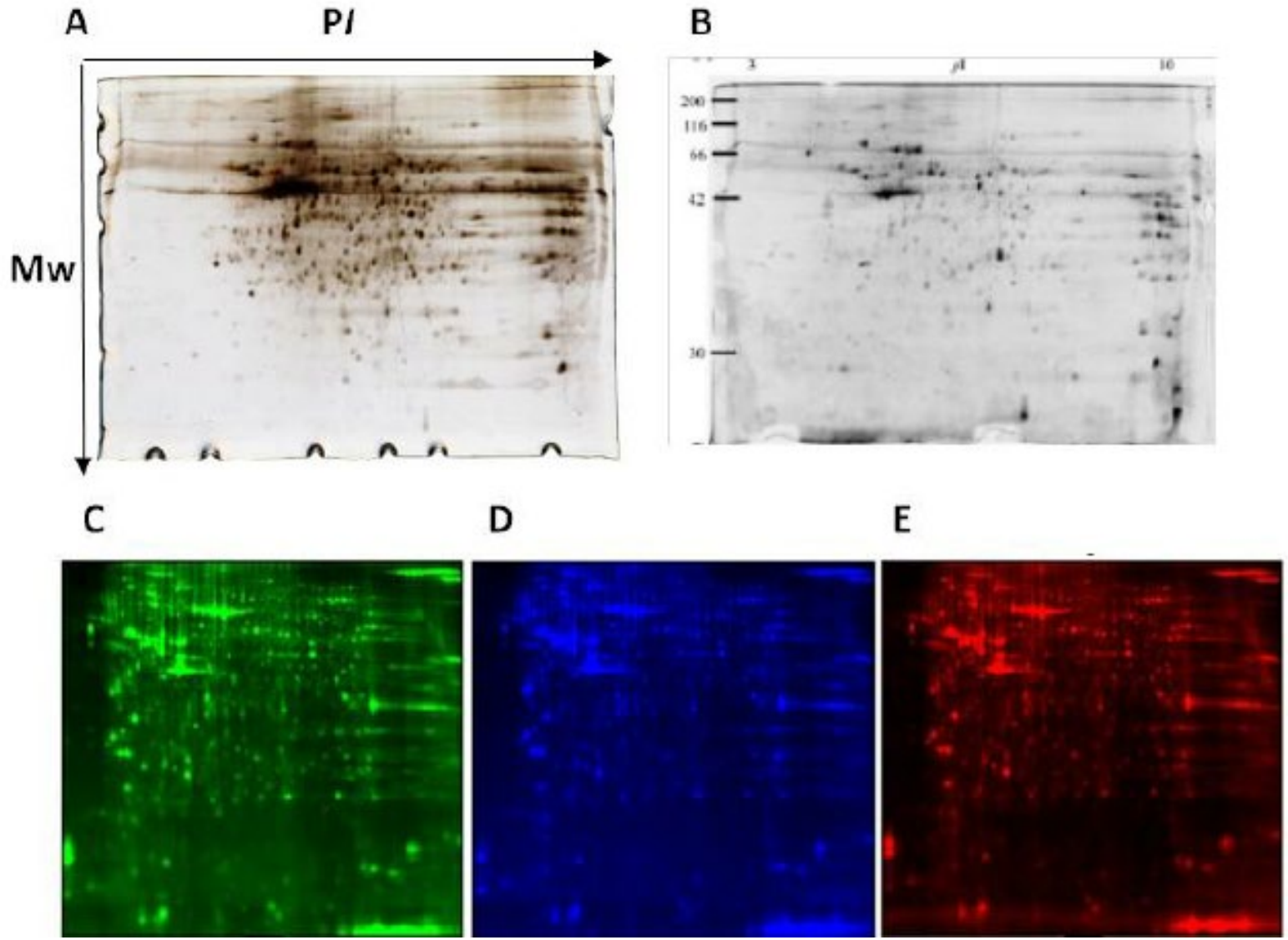


2-D gel electrophoresis

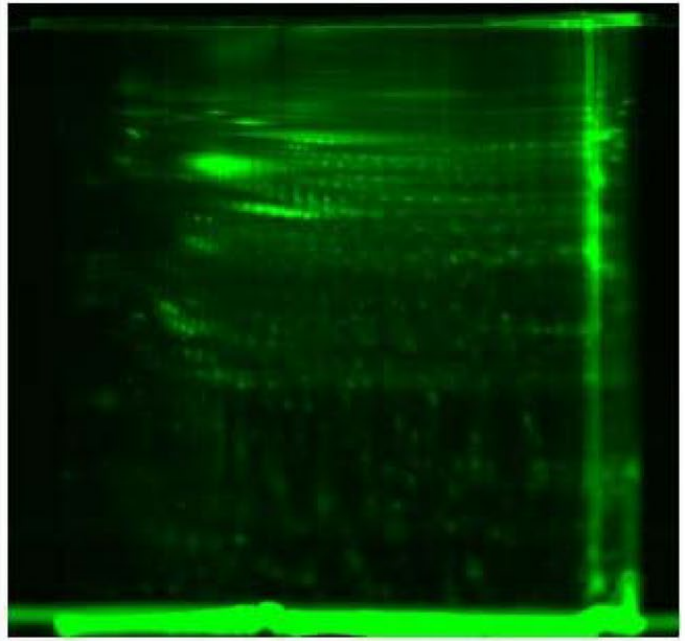


Visualisation of proteins

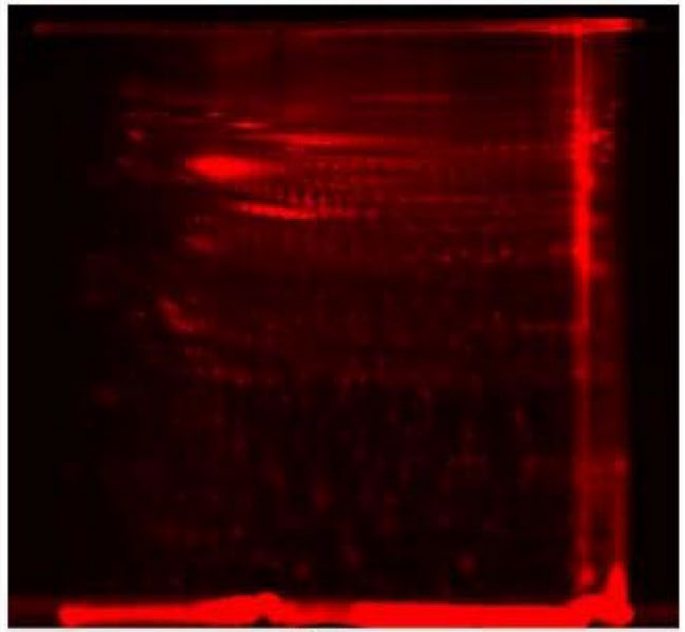
- Coomassie stain
  - Quick (1 hour) and cheap (20p/gel)
  - Not very sensitive (30ng/spot)
- Fluorescent stains (SYPRO Orange)
  - Slower (3 hours)
  - Sensitive (5ng/spot)
- Silver stain
  - Slower (2 hours)
  - Very sensitive (0.5ng/spot)



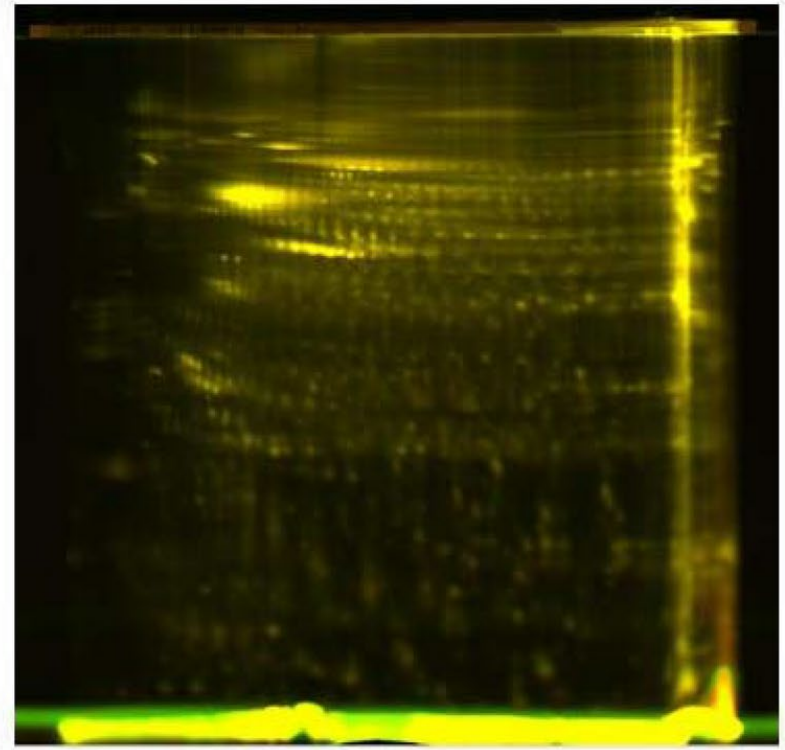
From Gel Electrophoresis - Principles and Basics



Cy3

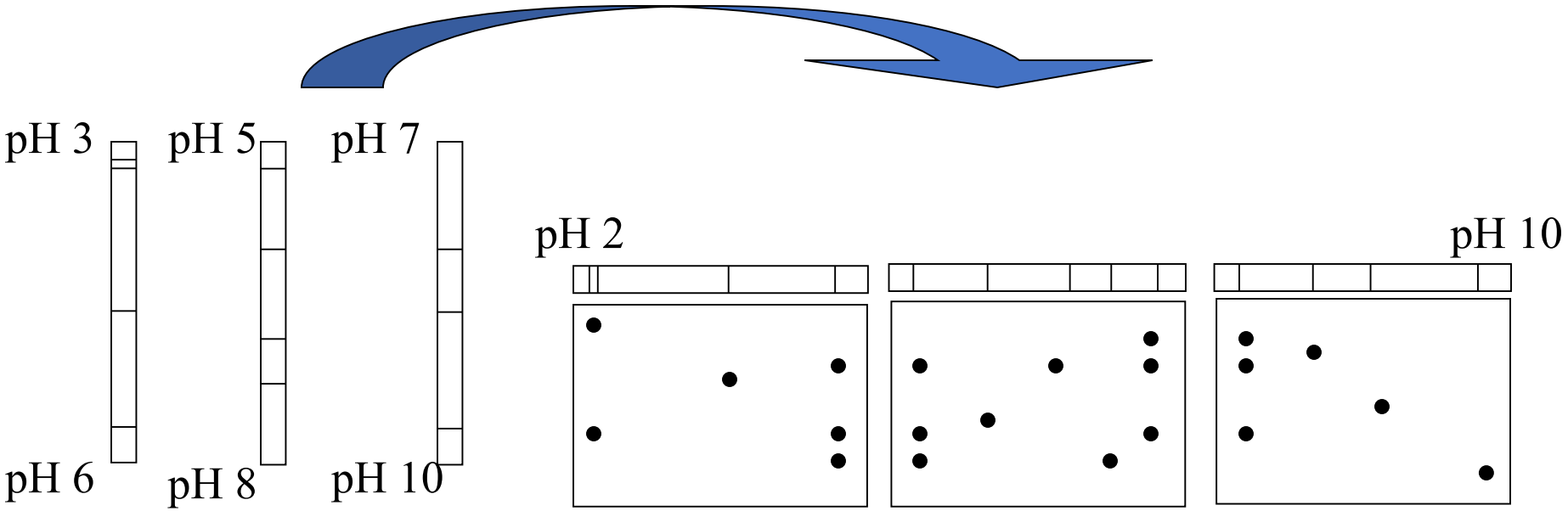


Cy5



overlap

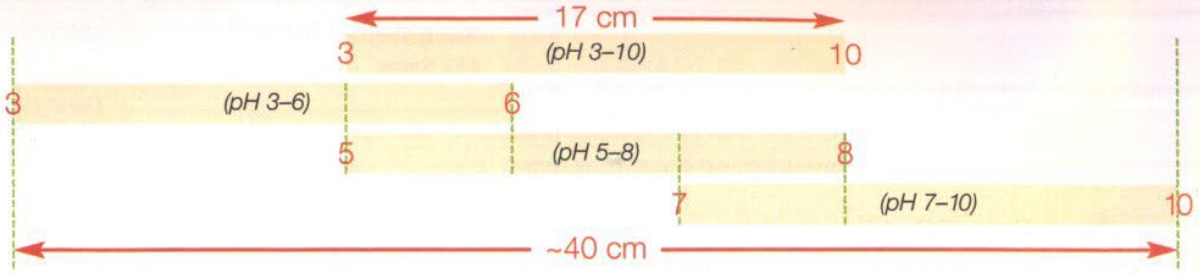
A range of IEF gels can be used

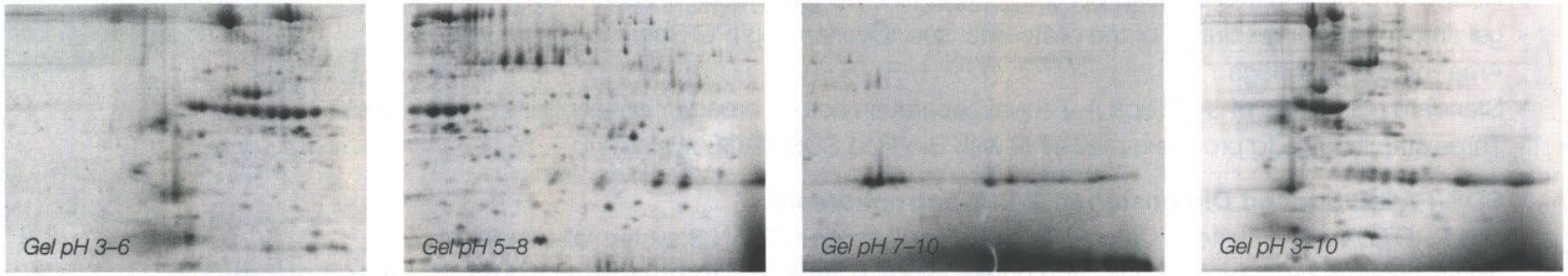


- We now have a pH range from 2 to 10 over three gels.

Gel arrays can increase separation

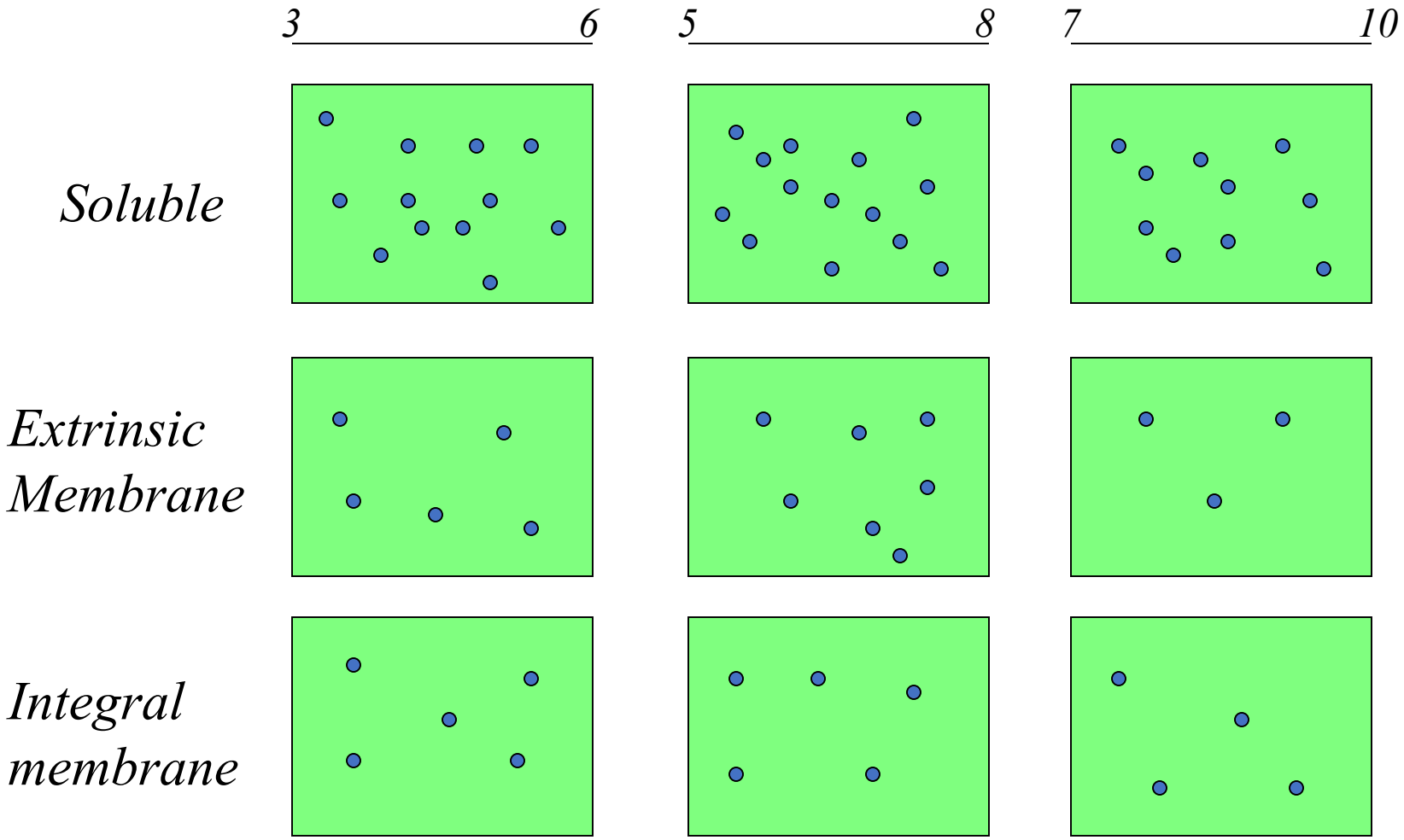
IPG Strip Length	pH Range	Combined Overlapping Gel Area
17 cm	3–10	~40 cm
11 cm	3–10	~26 cm
7 cm	3–10	~17 cm



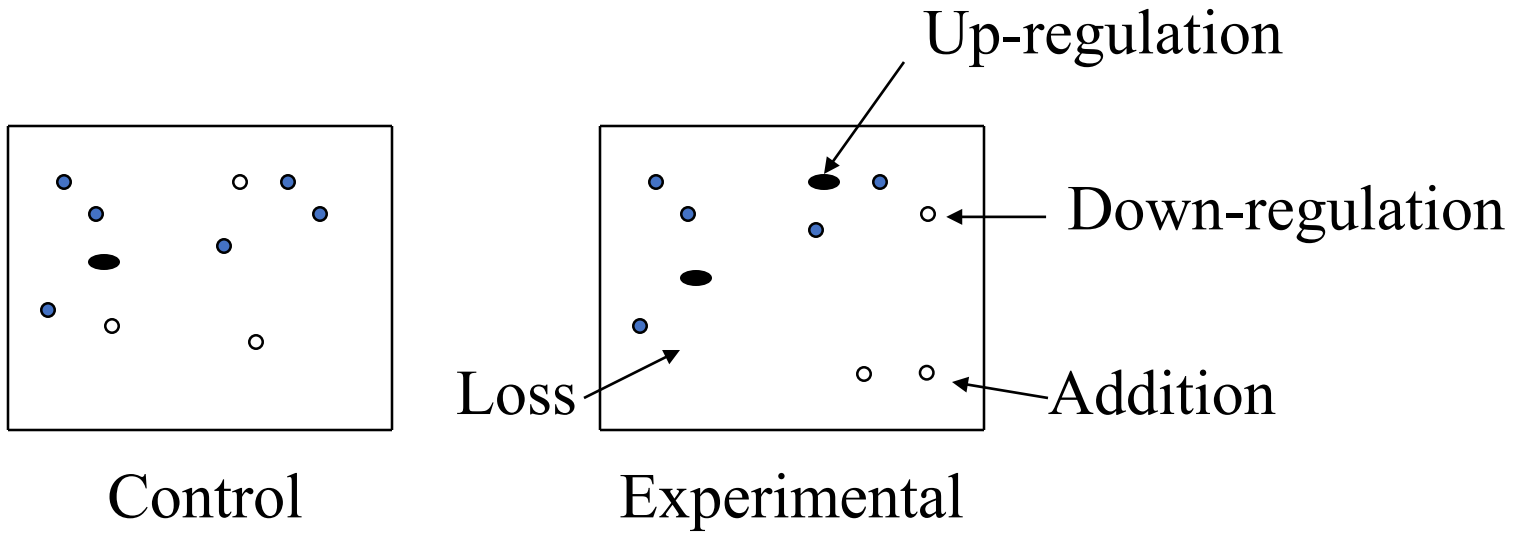


Soy storage proteins separated by the ReadyStrip narrow overlapping IPG strips, courtesy of Dr. Jim Malone, Monsanto, Inc.

We can also carry out differential extractions to increase resolution.



We can compare samples from control and experimental systems



We are now able to isolate proteins whose expression patterns are altered by a particular treatment.

What to do with these spots?

- Image analysis and associated software can allow differences to be noted.
- This information can then be fed to a robot spot cutter.
- The stained protein spots are isolated and placed in 96 well plates.
- The spots are digested with trypsin.
- Peptide mass maps can allow the identification of proteins.

We can fragment proteins at specific amino acid residues

table 5–7

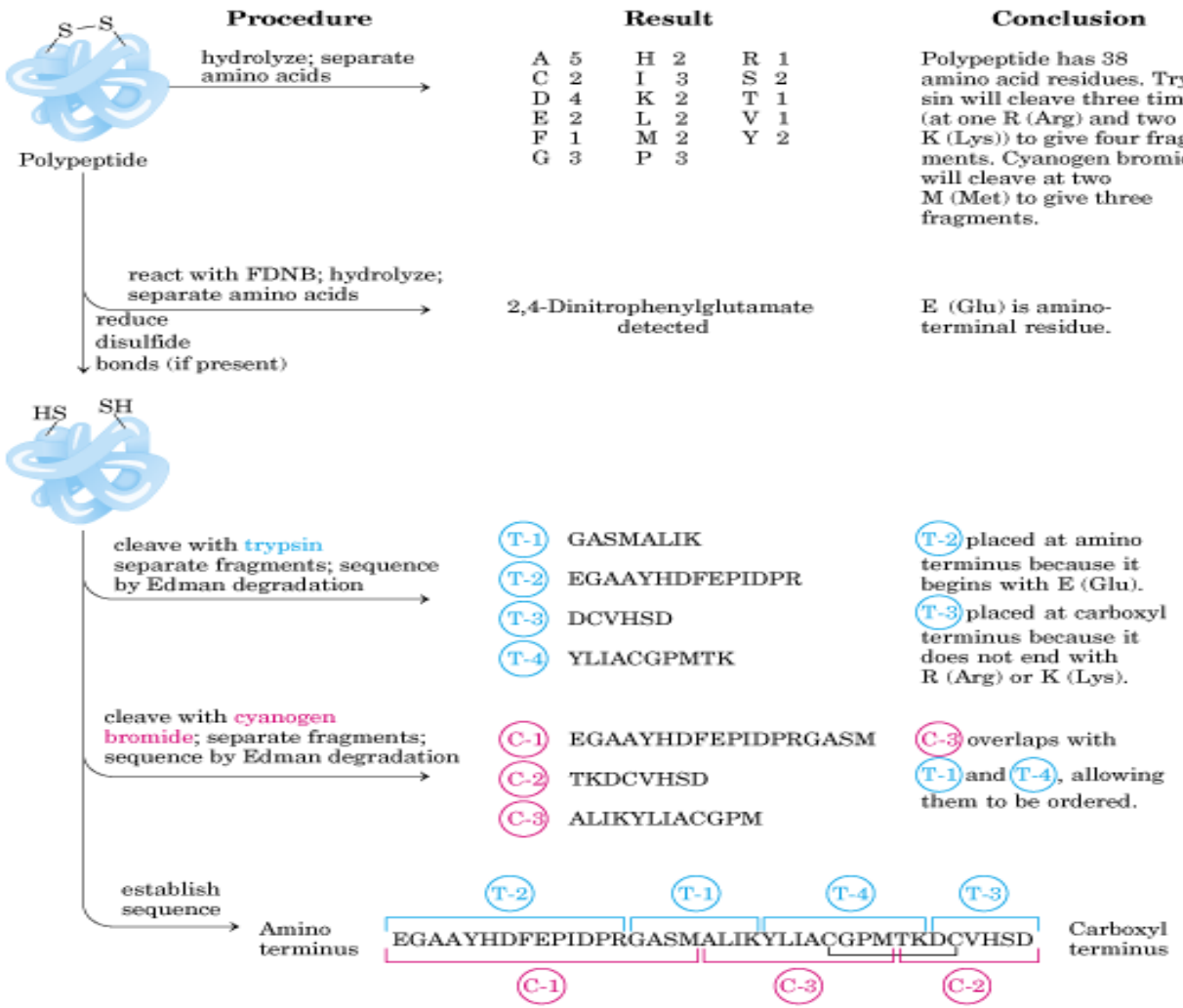
The Specificity of Some Common Methods for Fragmenting Polypeptide Chains

Treatment*	Cleavage points†
Trypsin	Lys, Arg (C)
Submaxillarus protease	Arg (C)
Chymotrypsin	Phe, Trp, Tyr (C)
Staphylococcus aureus V8 protease	Asp, Glu (C)
Asp-*N*-protease	Asp, Glu (N)
Pepsin	Phe, Trp, Tyr (N)
Endoproteinase Lys C	Lys (C)
Cyanogen bromide	Met (C)

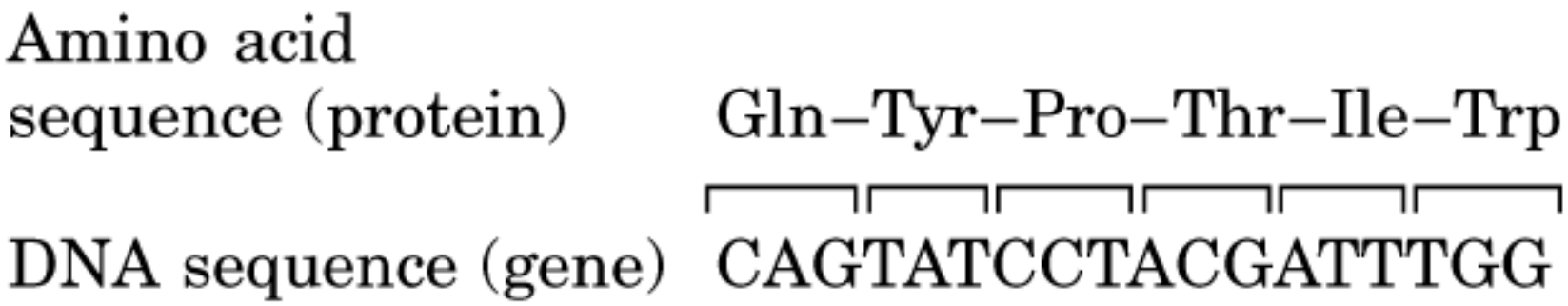
*All except cyanogen bromide are proteases. All are available from commercial sources.

†Residues furnishing the primary recognition point for the protease or reagent; peptide bond cleavage occurs on either the carbonyl (C) or the amino (N) side of the indicated amino acid residues.

Cleavage of proteins can allow sequence determination



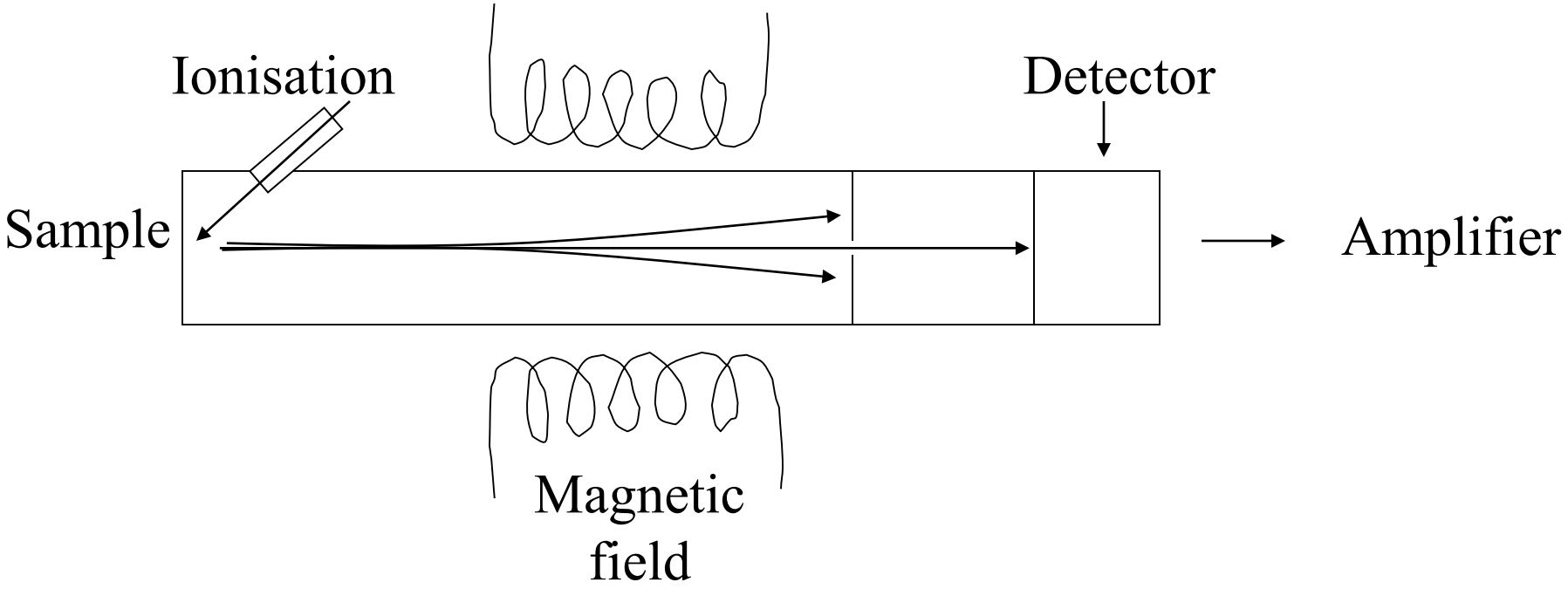
Once the amino acid sequence is known the gene can be isolated



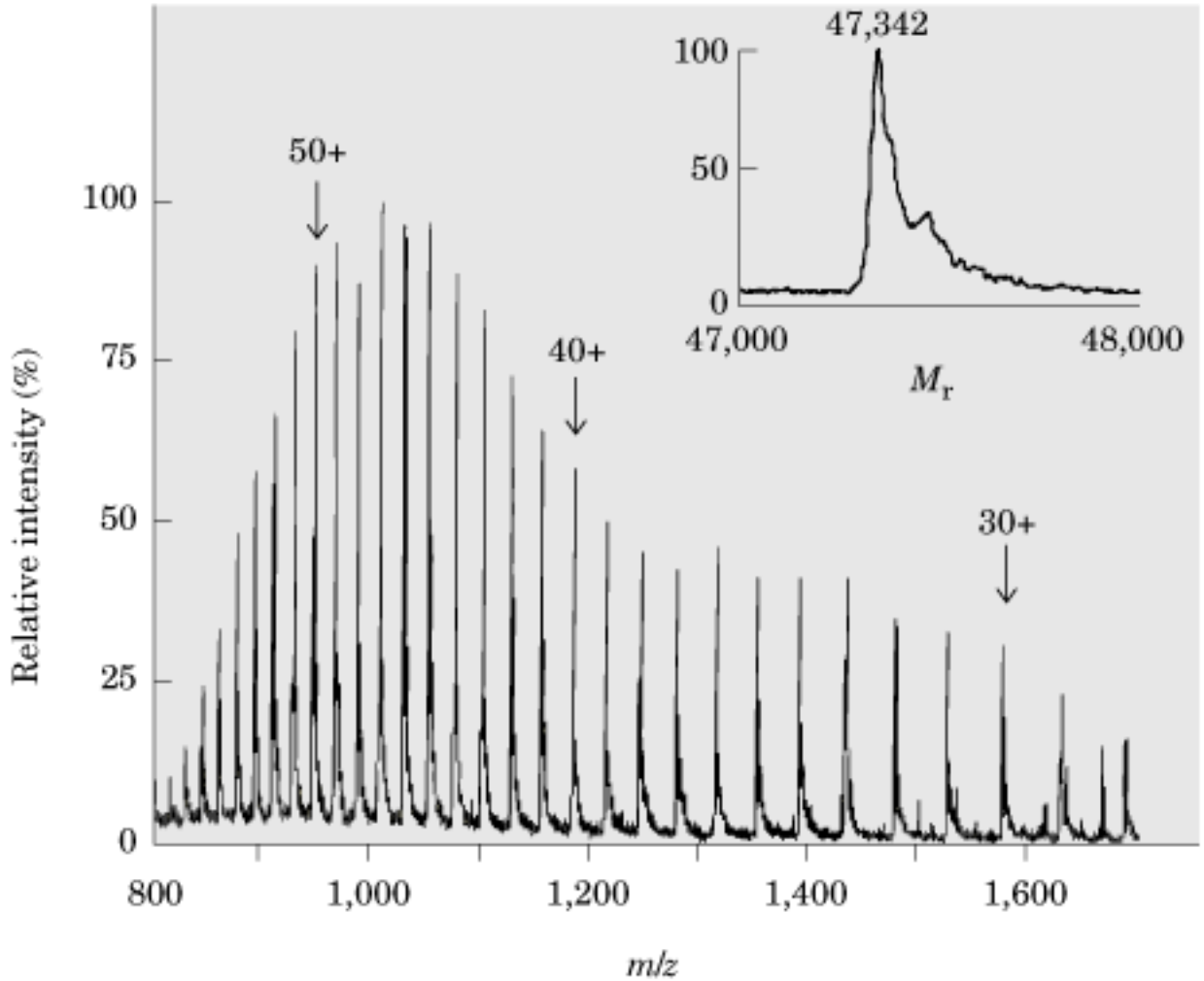
How Does MALDI-TOF Work?

- **The sample is mixed with the matrix and an aliquot is deposited onto the target plate.**
- **The target plate spot where the sample is located is irradiated by a pulse of light from the laser.**
- **Ions are desorbed into the gas phase.**
- **Ions are accelerated to a high kinetic energy by the plates of the ion source. These ions enter the flight tube. The unit of time that it takes the ion to move from the source to the detector is called the TIME-OF-FLIGHT.**
- **Ions strike the detector which converts them to electrical signal.**

Mass spectroscopy



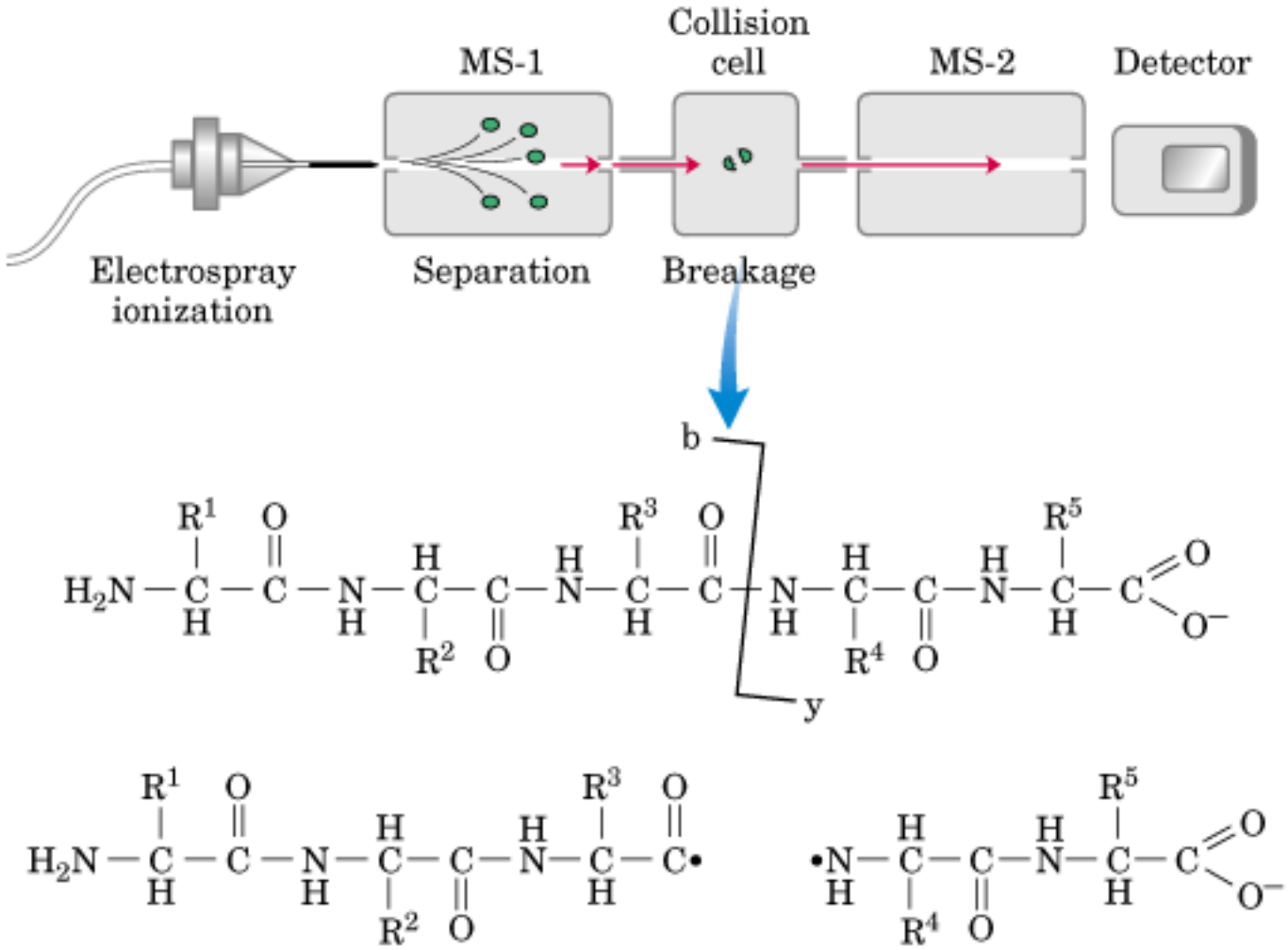
Varying the magnetic field allows different ions to pass through to the detector.


Relative intensity (%)
100
75
50
25
0
50+
40+
30+
800
1,000
1,200
1,400
1,600
m/z
47,342
100
50
0
47,000
48,000
M_r

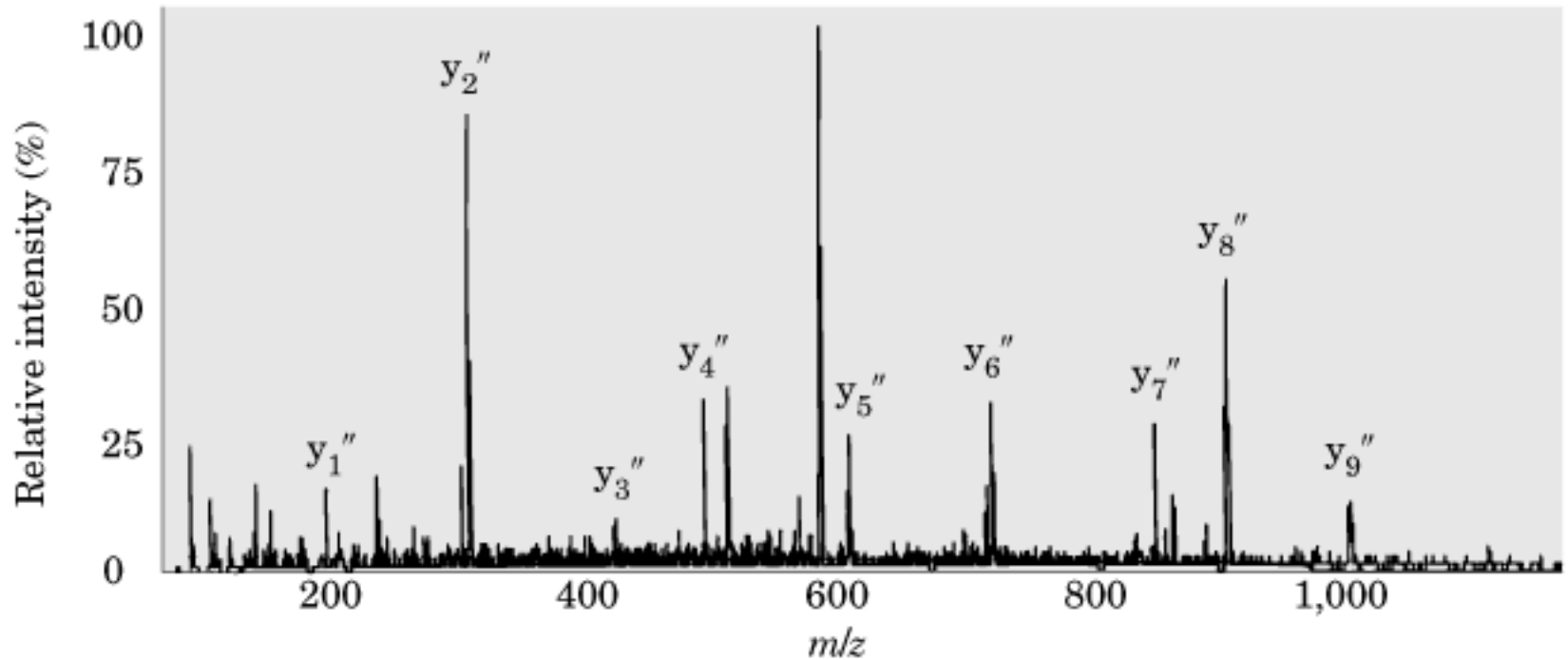
(b)

How does electrospray MS-MS work?

- **A mixture of peptides (in our case the tryptic digested spot) are sprayed through a fine needle (1μm) into the mass spec.**
- **Individual peptides are isolated during the first step and then fragmented during the second step.**
- **Fragments are obtained from the N- or C-terminal and are designated ‘b’ or ‘y’ ions.**
- **It is then a case of using software to look at ion profiles and reconstruct the protein sequence.**


Collision
cell
MS-1
MS-2
Detector
Electrospray
ionization
Separation
Breakage
b
y

(a)


100
75
50
25
0
Relative intensity (%)
y2″
y1″
y3″
y4″
y5″
y6″
y7″
y8″
y9″
200
400
600
800
1,000
m/z

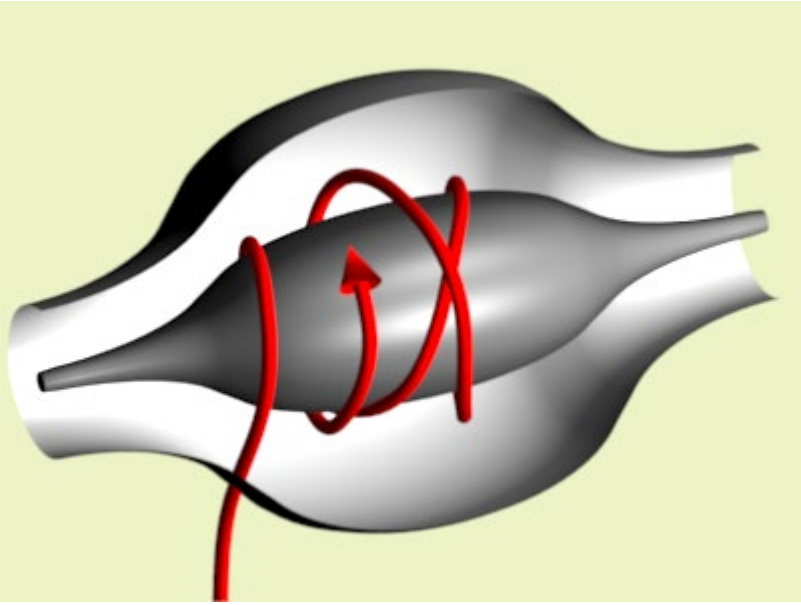
(b)

How do these two systems work together?

- We can now isolate proteins from a 2-D gel and identify the protein using mass spectroscopy.
- This type of process can be carried out rapidly allowing the development of high through-put systems.
- We can identify up to 200 spots a day from peptide mass maps and 20 from MS-MS.

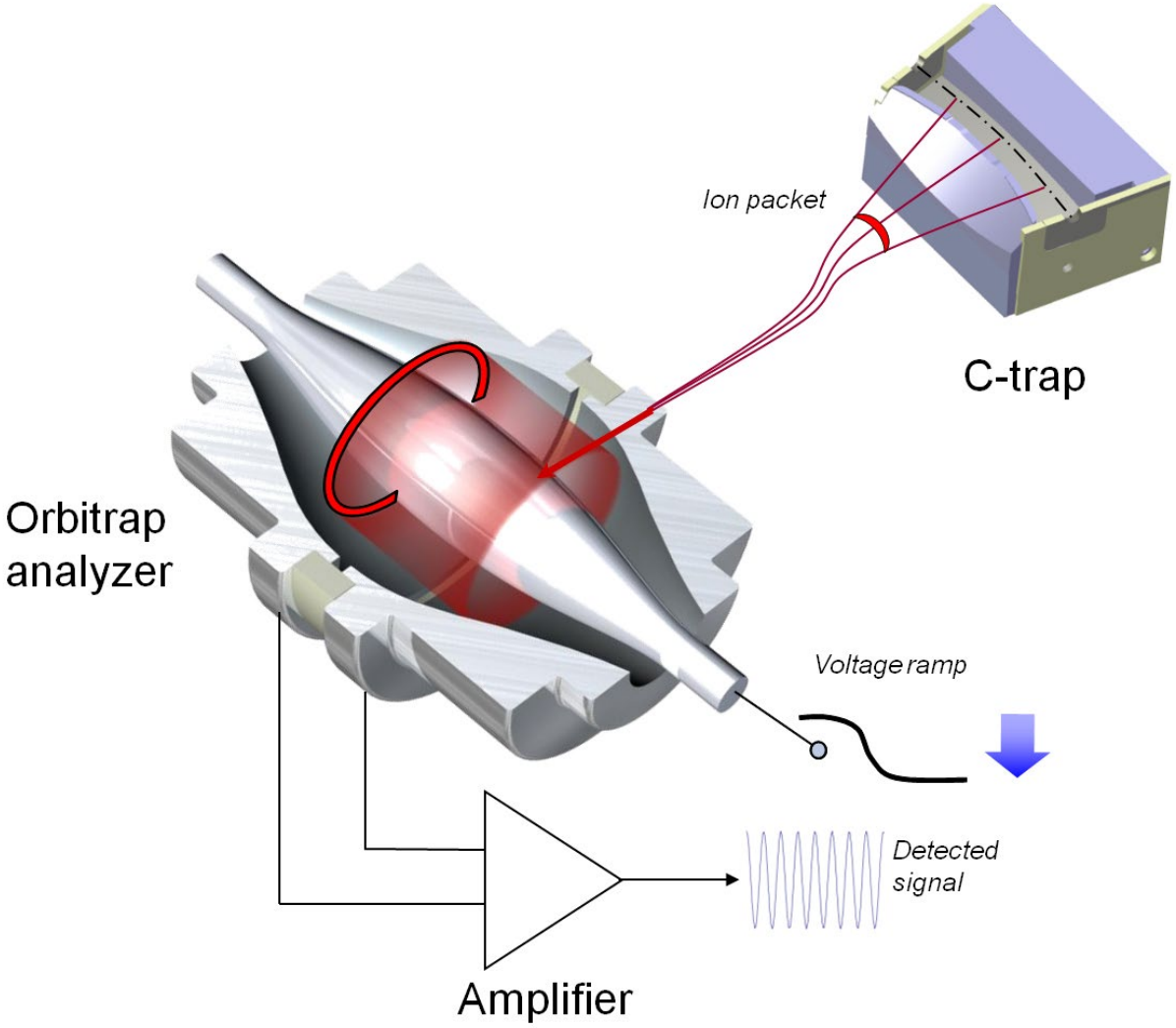
Orbitrap

Ion trap mass analyzer consisting of an outer barrel-like electrode and a coaxial inner spindle-like electrode that traps ions in an orbital motion around the spindle.

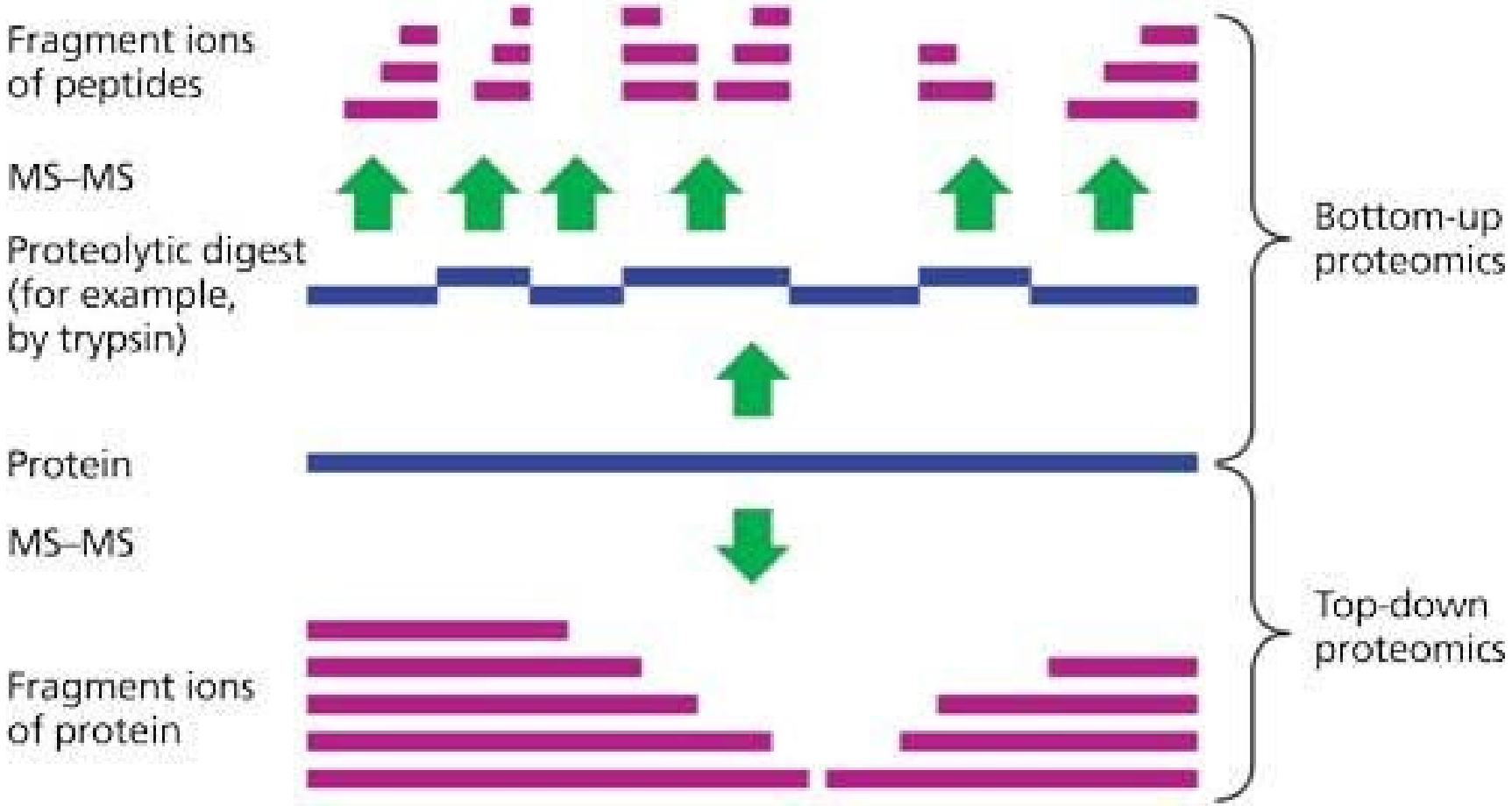


Orbitrap

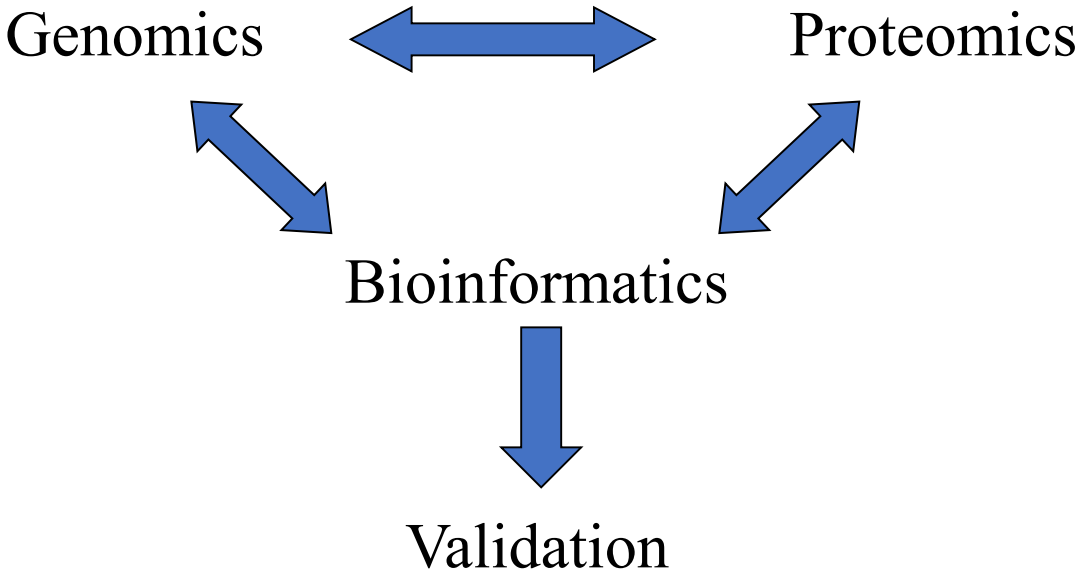
The image of trapped ions is detected and converted to a mass spectrum by first using the Fourier transform of time domain of the harmonic to create a frequency signal which is converted to mass



Top-down and bottom-up sequencing for identifying a specific protein or proteins of interest



Summary slide



https://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=PMF

https://www.matrixscience.com/training/pmf1_q.html