PROTEIN EXPRESSION AND PURIFICATION

PROTEIN EXPRESSION AND PURIFICATION

Why do we decide to purify a protein?
What do we known about the protein?

What is the most abundant and cheap source?

- **-organism**
- **- tissue**
- **-subcellular localization**

- **- how much protein do we need?**
- **-how pure**
- **-is easy to purify the protein from natural sources?**
- **Abundance?**
- **-Stability, Molecular weight, Isoelectric point, Function-Activity, Isoforms, contaminant proteins**
- **- scheme of purification published**

Native source

-the gene is not available

-naturally abundant in the source

-the expression in recombinant system is complex like during purification of multiple complexes of proteins

Recombinant protein

- low abundance
- hard to purify from natural source

-genetic analysis; protein structural-function analysis; analysis of a domain

To obtain a recombinant protein

Obtain the cDNA clone

↓

Decide on the expression system and purification scheme

↓

Optimize the expression

↓

Purify the protein

↓

Protein characterization and quality control

To obtain a recombinant protein,

- **buy the clone (http://www.ncbi.nlm.nih.gov/clone)**
- **designed primers**
- **PCR amplification of the cDNA sequence**
- **select the corresponding vector**
- **Insert the cDNA amplified into the selected vector**

Subcloning for overexpression

Prokaryote systems: fast, cheap, high throughput
-most common Escherichia coli

Eukaryotic systems: expensive, laborious, high fidelity, natural post-tranlational processing

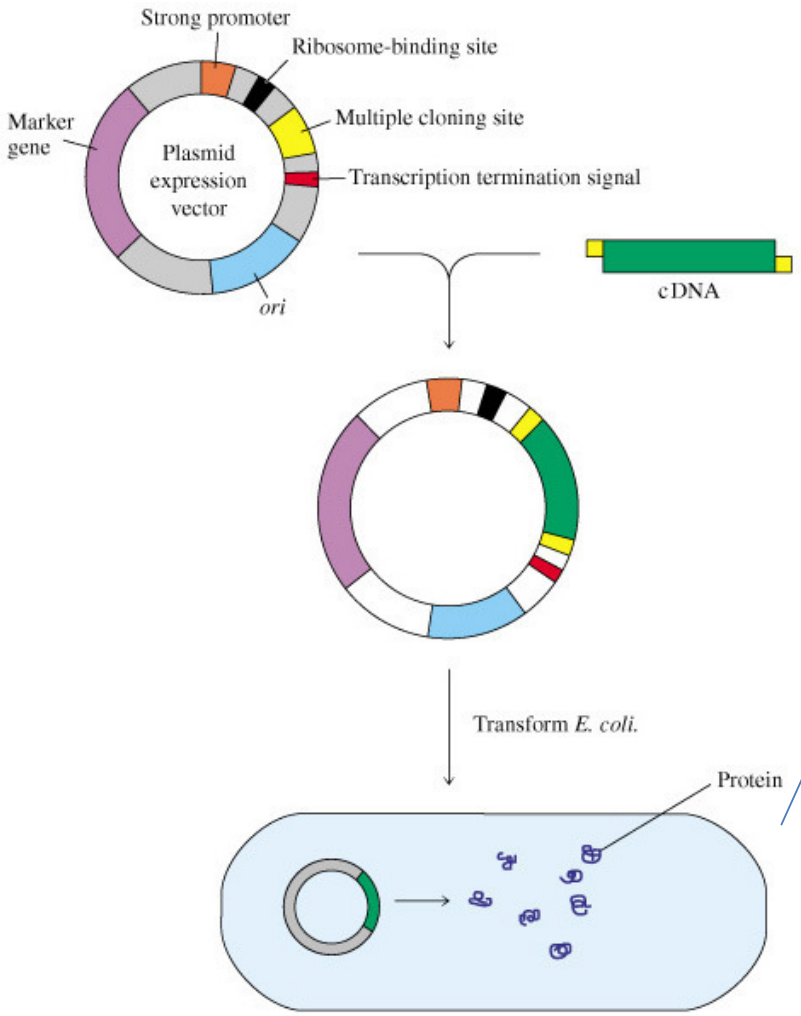
-yeast: yield 15g/L, slow growth, secreted protein, postranslational modifications

-insect cells: secretory pathway, high level of expression, glycosilation patterns, disulfide bonds, closer resemblance to mammalian cells

-mammalian cells high fidelity for postranslational modifications, expensive, low yield (<1 mg/ml)

Expression system	E.coli	Insect cells	Yeast cells	Mammalian cells
Proteolytic cleavage	+/-	+/-	+/-	+
Glycosylation	-	+	+	+
Secretion	+/-	+/-	+	+
Folding	+/-	+/-	+/-	+
Phosphorylation	-	+	+	+
Acetylation	-	+	+	+
Amidation	-	+	-	+
Percent yield	5-30%	1-30%		< 1%

Prokaryotic cells



Antibiotic to select cell transformed

↓

Induction of protein expression. Addition of inducer: IPTG, lactose

↓

Protein expression

↓

Centrifugation to collect the cells

↓

Cell lyses

↓

Protein purification

Insect cells, Baculovirus expression: Flow chart

- pFastBac donor plasmid

↓ clone gene of interest

- pFastBac recombinant

↓ transform in MAX efficiencyDH10Bac cells (containing bacmid and helper)

- E.coli colonies with recombinant Bacmid

↓ restreak

- Verified E.oli colonies with recombinant Bacmid

↓ Growth overnight culture and isolate recombinant bacmid DNA

- Recombinant Bacmid DNA

↓ Transfect insect cells using Cellfecting Reagent

- P1 Recombinant Baculovirus stock (>106 pfu/ml)

↓ Infect insect cells to amplify virus

- P2 recombinant Baculovirus stock (>107 pfu/ml)

↓ Titer and infect insect cells

Protein expression

↓

Cell lyses/ Media

↓

Protein purification

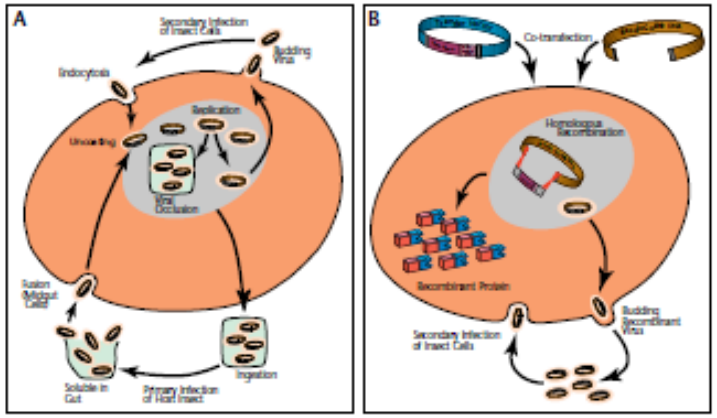
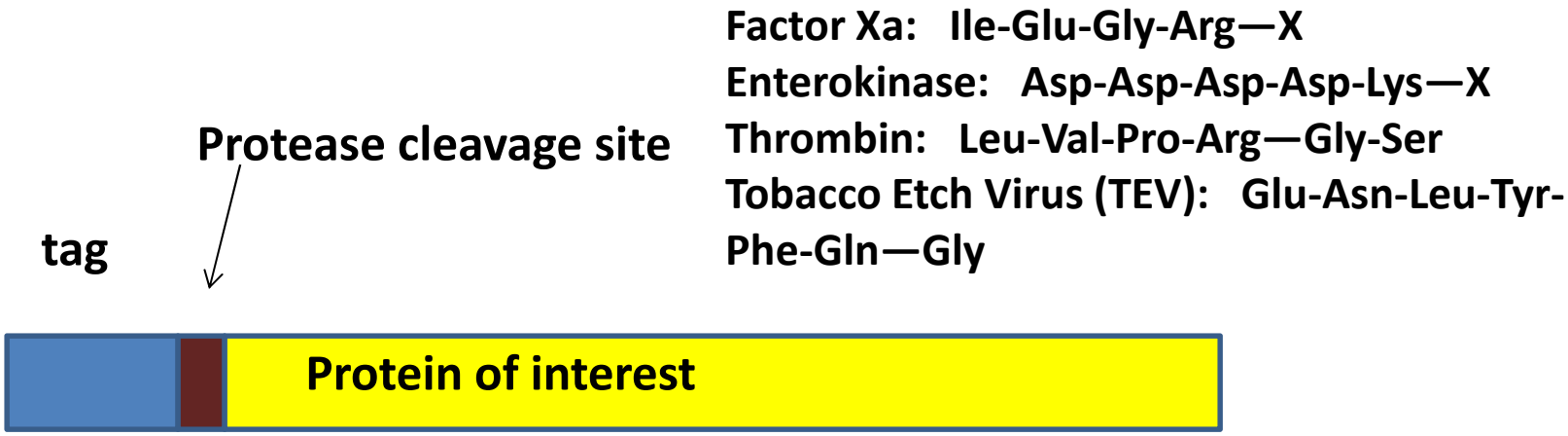


Figure 1. The Baculovirus life cycle *in vivo* and *in vitro*. A) *In vivo*. Two distinct viral populations are formed in infected insect cells, occluded and budded virions. Occluded virions are protected from desiccation in the environment, allowing primary infection in susceptible larva. Once ingested, the occlusion body is solubilized in the gut, releasing virions which fuse with midgut cells. The virion nucleocapsid migrates through the cytoplasm to the nucleus. The core is uncoated from the capsid structure in the nucleus. Here replication takes place. Secondary infection is mediated by the budded form of the virus entering adjacent cells via adsorptive endocytosis. B) *In vitro*. The Baculovirus genome is too large to directly insert foreign genes easily. Hence, the foreign gene is cloned into a transfer vector that contains flanking sequences which are homologous (5' and 3' to your insert) to the Baculovirus genome. BaculoGold™ DNA and the transfer vector containing your cloned gene are co-transfected into *Sf9* insect cells. Recombination takes place within the insect cells between the homologous regions in the transfer vector and the BaculoGold™ DNA. Recombinant virus produces recombinant protein and also infects additional insect cells thereby resulting in additional recombinant virus.

Recombinant protein purification

Recombinant proteins are typically expressed as a fusion with an "affinity tag"

Tag Protein of interest



Affinity tag		Size	Affinity resin
Maltose Binding Protein	(MBP)	40 kDa	Amylose
Glutathione S-transferase	(GST)	26 kDa	Glutathione
Poly-His	(His6)	<1 kDa	Ni2+

Protein purification for a typical soluble protein

1. **Homogenization** → prepare cell-free extract

Adjust conditions of buffer pH, salt, temperature

-Presence of proteases

-Adsorption to surface, denaturation air-water interface

-storage -80C, -196C (liquid nitrogen)

-stability

2. Centrifugation → remove membranes, nuclei, large organelles

To follow the protein through the purification process

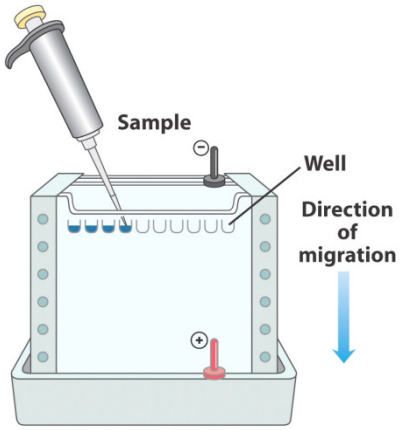
→ Protein content Ab280/Bradford/Lowry

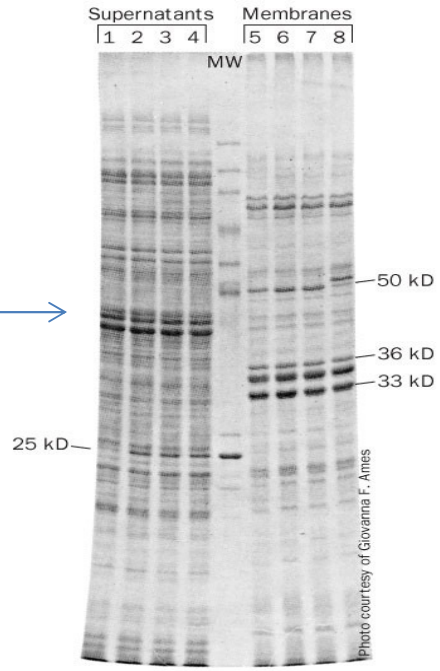
→ Activity coupled enzymatic reaction,

Immunoassays (RIA, ELISA), Western blot

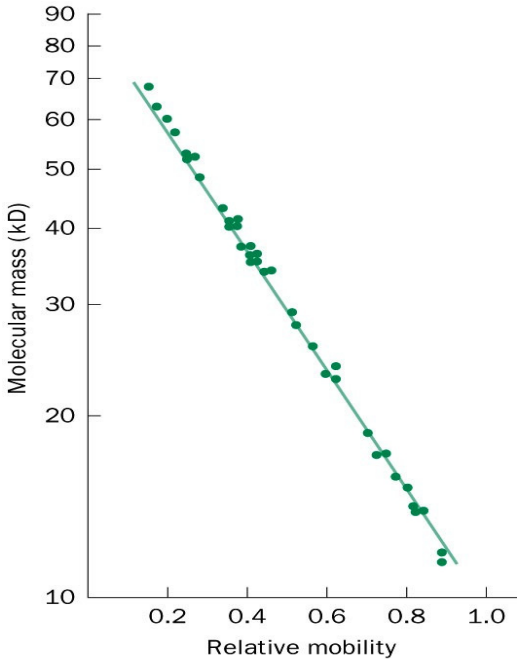
→ Electrophoresis SDS-PAGE

To follow the purification steps through Electrophoresis
SDS-PAGE (sodium dodecil sulphate polyacrylamide gel electrophoresis (ID, 2D IEF MW –pI)
Some methods are used to purify protein





Coomassie blue



Logarithmic relationship between the molecular mass of a protein and its relative electrophoretic mobility in SDS-PAGE.

-Molecular weight, Isoelectric point, Function-Activity

Protein characteristic	Purification Procedure
Solubility (pH, salt, temperature, solvent)	Salting out
Ion Charge	Ion exchange chromatography
Polarity	Hydrophobic interaction chromatography (HIC) Reverse phase chromatography (RPC)
Binding specificity (ligands, Ab, substrates)	Affinity chromatography
Molecular size	Gel filtration chromatography

3. Ammonium sulfate precipitation

4. Column Chromatography

Separation by precipitation

➢ **Salting-out separates proteins by their solubility**

Low salt concentration → increases the protein solubility→ **salting-in**

Higher salt concentration

↑hydrophobic interactions protein precipitation → **salting-out**

• Competition between the added salt ions and other dissolved solutes for molecules of solvents

•Depending of the hydrophobic protein composition protein precipitate at different salt concentrations

•Salt → multiple charged anions → ammonium sulfate, phosphate, citrates

• **Ammonium Sulfate** (cost/solubility)

High solubility that varies very little with the temperature (~4 M , 0°C, 100% solution)

Stabilize most of the proteins, and most protein precipitate 20-80%

Reduce lipid content of the sample

• The precipitates can be redissolved in small volume → concentration

•Protein precipitated contains salt → redissolved protein against low salt buffer → dialysis

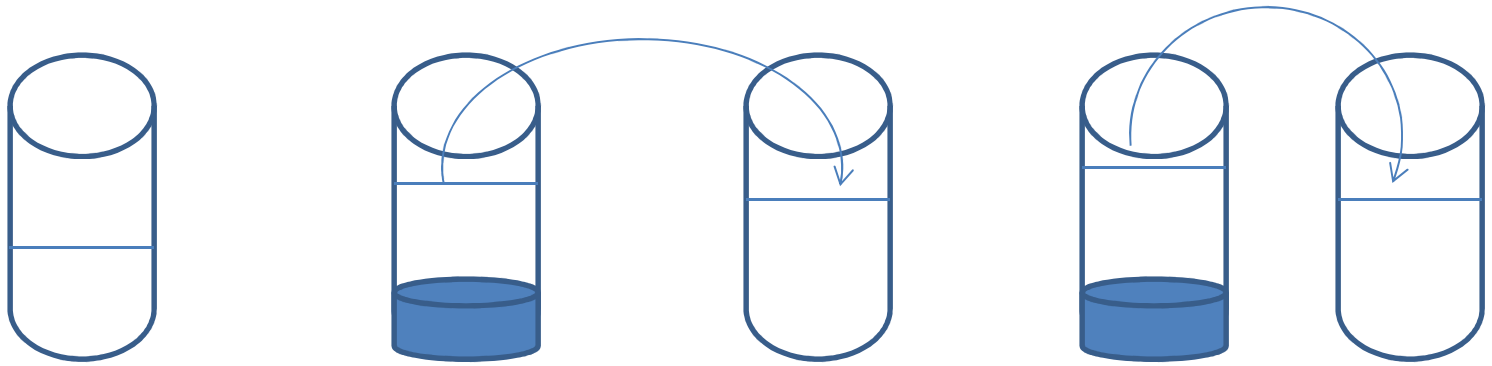
Salt fractionation

Add salt to 20% saturation ↓

Centrifuge and remove supernatant

Add salt to 40% saturation ↓

Centrifugate and remove supernatant

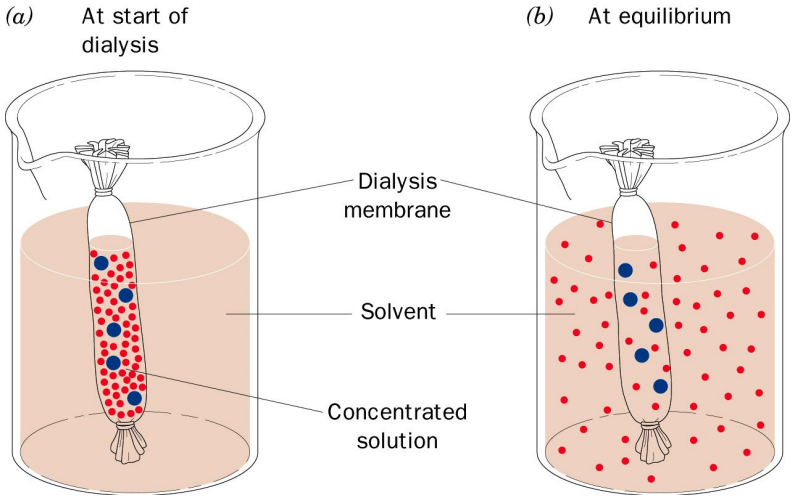


Cell free extract

Ammonium sulfate $(NH_4)_2SO_4$

$2NH^+_4\ SO4^-_2$

Dialysis lowers salt concentration in a protein solution and separates small and large molecules.



Dialysis protocol for decreasing salt concentration from 1M <1mM

Dialysis against 5 L of water
→swell to 110 ml, at equilibrium = 20 mM

Change dialysate, further 5 L of water
→no further swelling, at equilibrium = 0.4 mM

A single change would be sufficient even without complete equilibirum

Column chromatography

After the initial fractionation steps we move to **column chromatography**.

The mixture of substances (proteins) to be fractionated is dissolved in a liquid or gaseous fluid called the **mobile phase**.

This solution is passed through a column consisting of a porous solid matrix called the **stationary phase**. These are sometimes called **resins** when used in liquid chromatography.

The stationary phase has certain physical and chemical characteristics that allow it to interact in various ways with different proteins.

Common types of chromatographic stationary phases

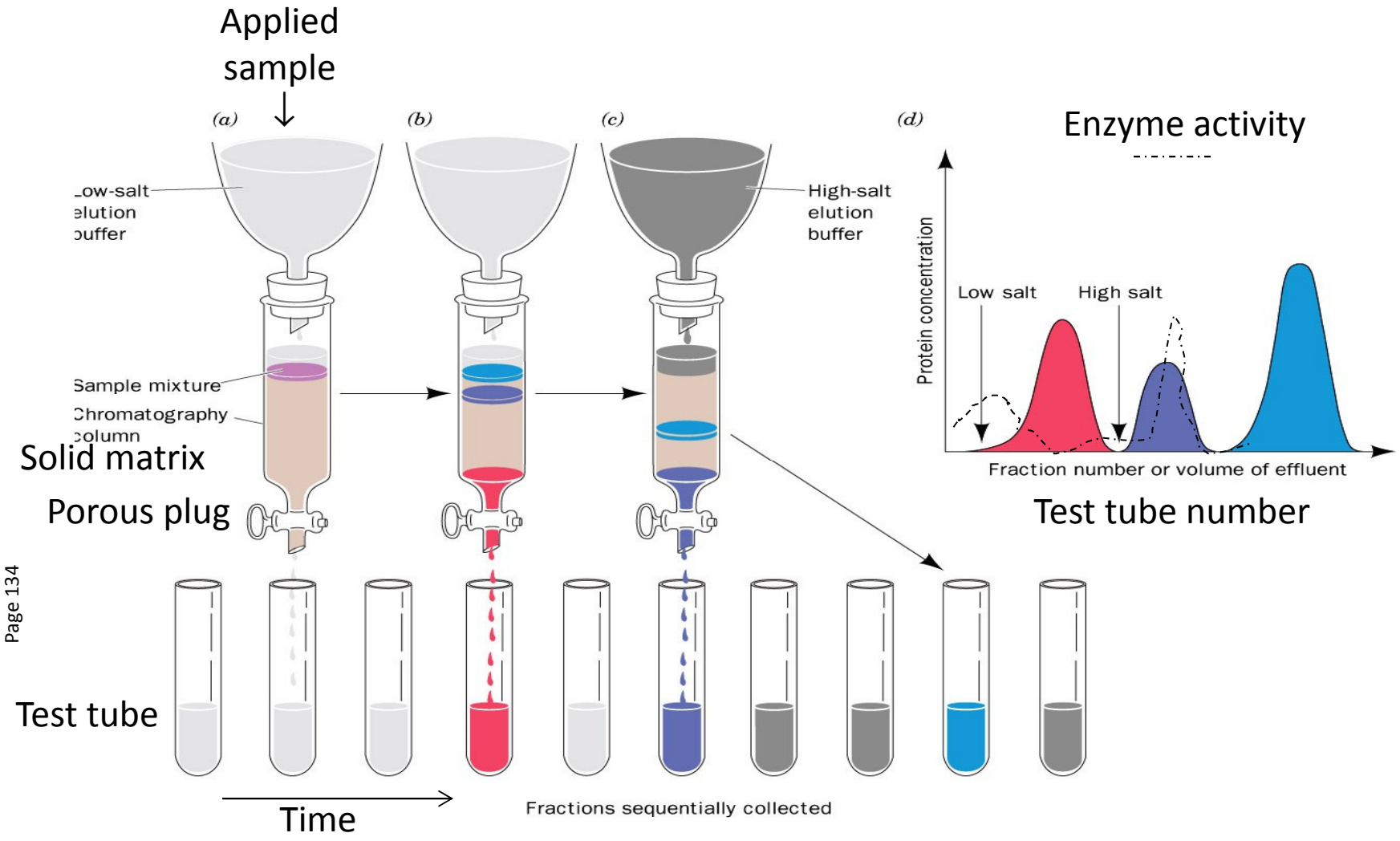
Ion exchange Anion exchange (DEAE), Cation exchange (CM)

Hydrophobic

Size exclusion Gel filtration

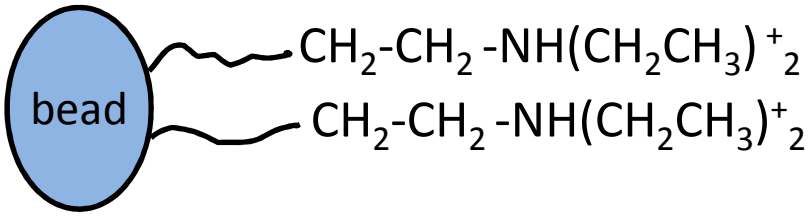
Specific Affinity

General Chromatography protocol

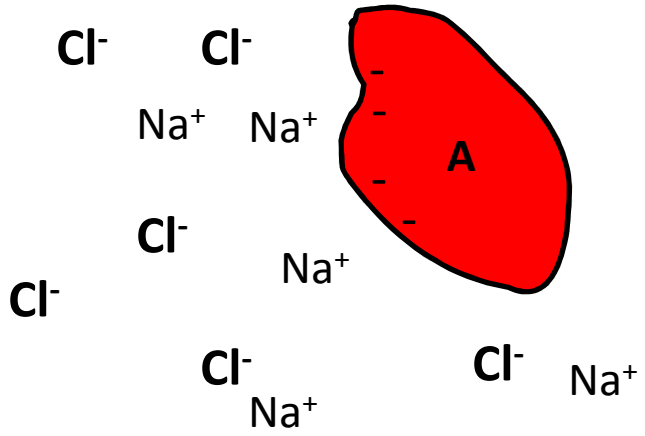


Ion exchange chromatography

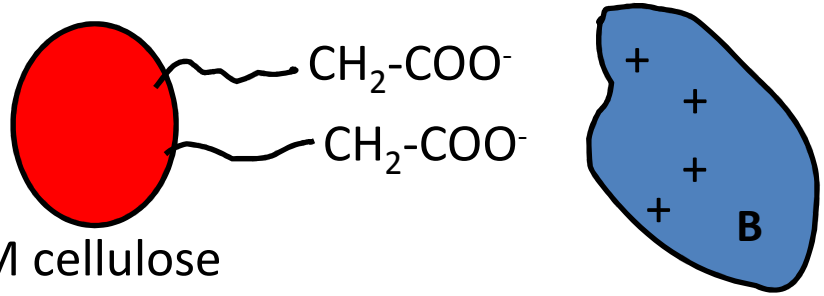
Ion exchange resins contain **charged groups** → *acidic* → interact with negatively charged proteins and are called **Anion** exchangers.



DEAE cellulose
diethylaminoethyl
anion exchanger



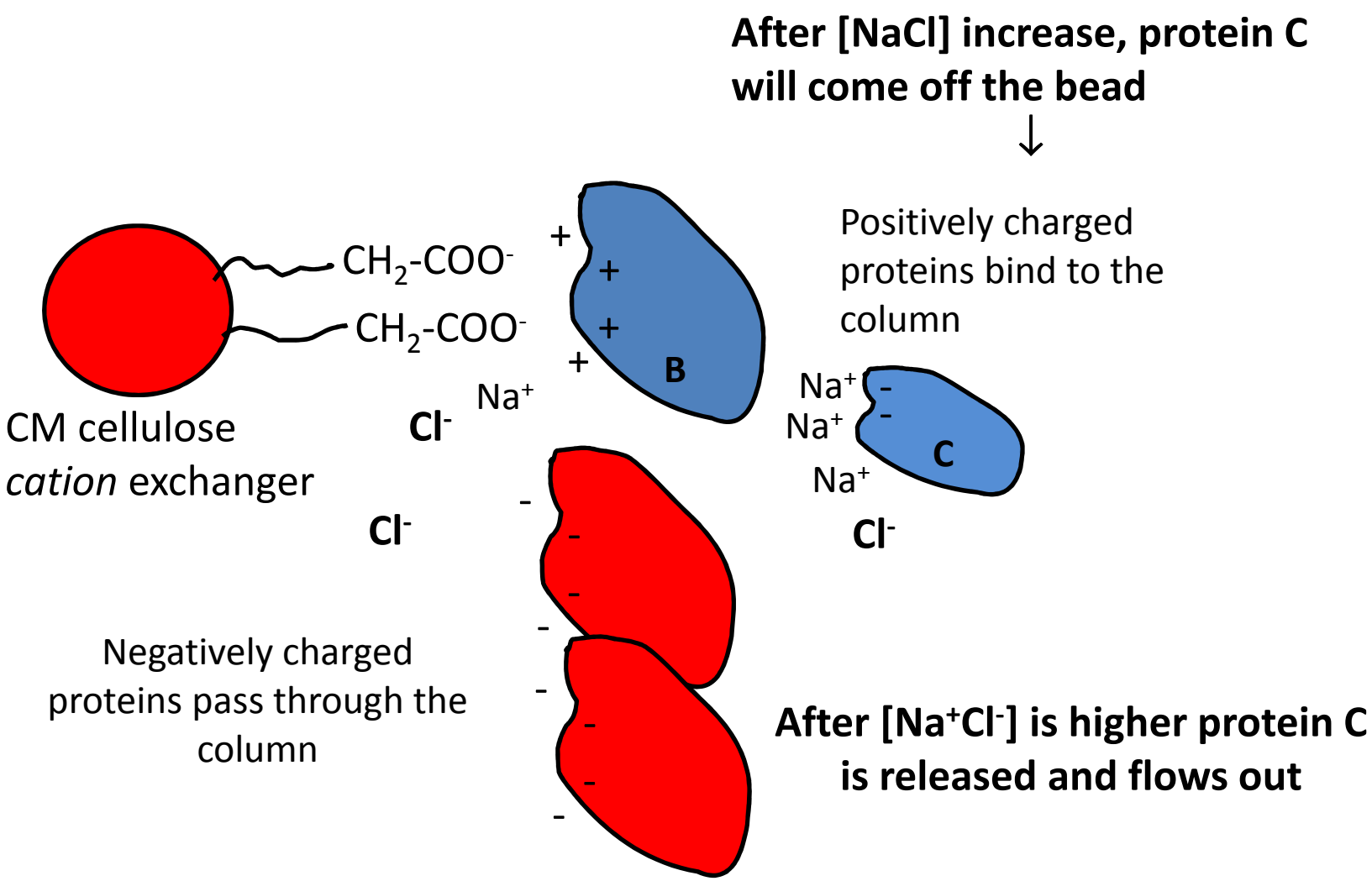
Ion exchange resins contain **charged groups** → *basic* → interact with positively charged proteins and are called **Cation** exchangers



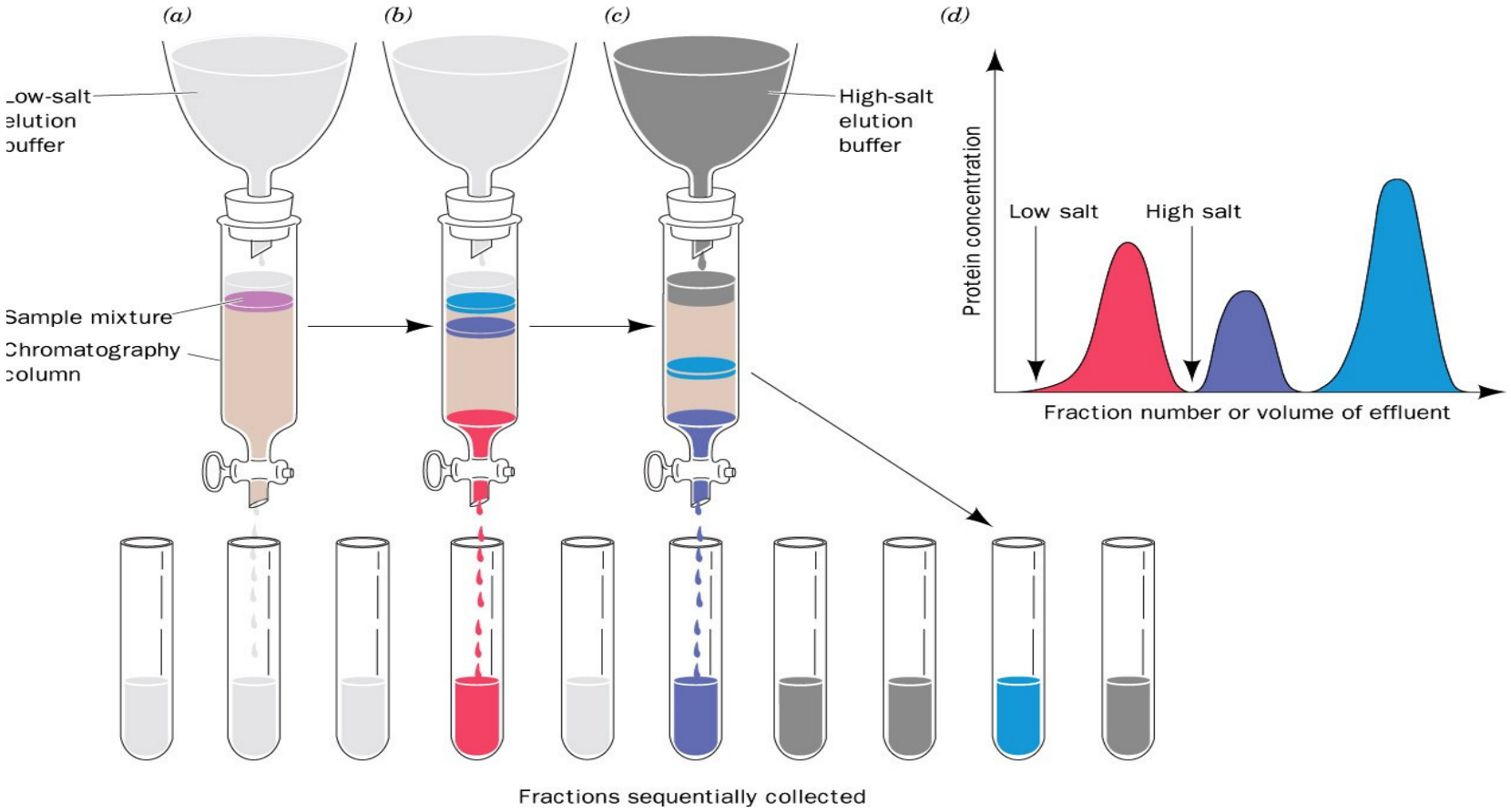
CM cellulose
carboxymehty
cation exchanger

Ion exchange chromatography

For protein binding, the pH is fixed (usually near neutral) under low salt conditions. Example cation exchange column…



Ion exchange chromatography using stepwise elution.



Some Biochemically Useful Ion Exchangers

Name[a]	Type	Ionizable group	Remarks
DEAE-cellulose	Weakly basic	Diethylaminoethyl $-CH_2CH_2N(C_2H_5)_2$	Used to separate acidic and neutral proteins
CM-cellulose	Weakly acidic	Carboxymethyl $-CH_2COOH$	Used to separate basic and neutral proteins
P-cellulose	Strongly and weakly acidic	Phosphate $-OPO_3H_2$	Dibasic; binds basic proteins strongly
Bio-Rex 70	Weakly acidic, polystyrene-based	Carboxylic acid $-COOH$	Used to separate basic proteins and amines
DEAE-Sephadex	Weakly basic cross-linked dextran gel	Diethylaminoethyl $-CH_2CH_2N(C_2H_5)_2$	Combined chromatography and gel filtration of acidic and neutral proteins
SP-Sepharose	Strongly acidic cross-linked agarose gel	Methyl sulfonate $-CH_2SO_3H$	Combined chromatography and gel filtration of basic proteins
CM Bio-Gel A	Weakly acidic cross-linked agarose gel	Carboxymethyl $-CH_2COOH$	Combined chromatography and gel filtration of basic and neutral proteins

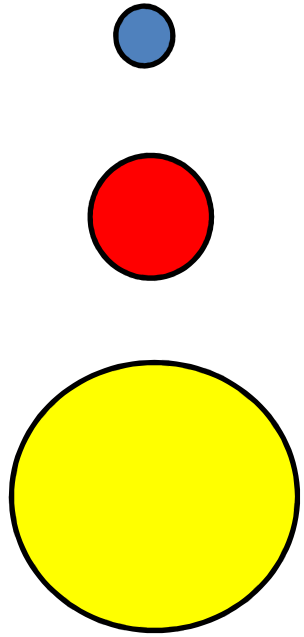
[a]Sephadex and Sepharose gels are manufactured by Amersham Pharmacia Biotech, Piscataway, New Jersey; Bio-Rex resins and Bio-Gels are manufactured by BioRad Laboratories, Hercules, California.

Mono-Q fast, high performance anion exchange separation
Mono -S fast, high performance cation exchange separation

Gel filtration chromatography

Mix of proteins of different with different molecular weight
Porous polymer beads

How does it work? If we assume proteins are spherical…



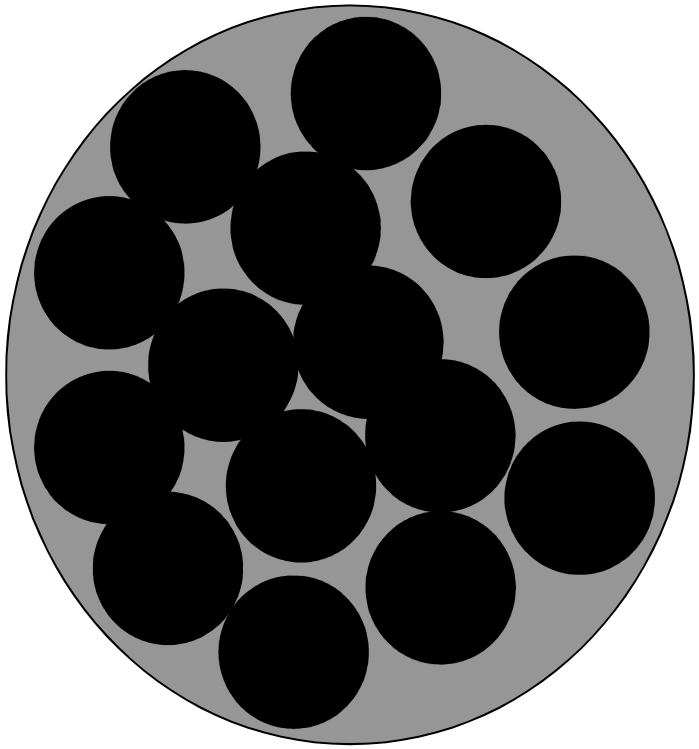
size

Molecular mass (daltons)

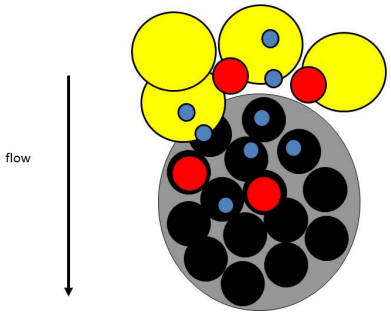
10,000

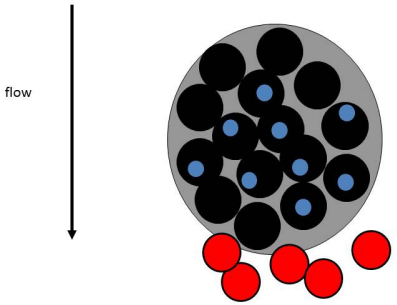
30,000

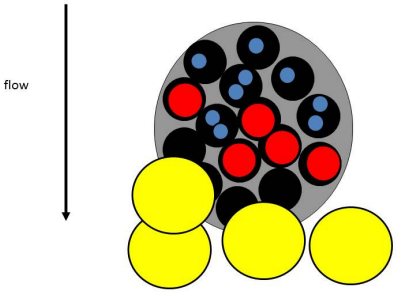
100,000

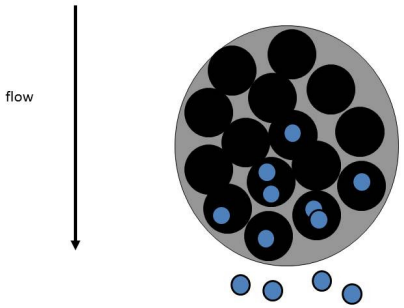


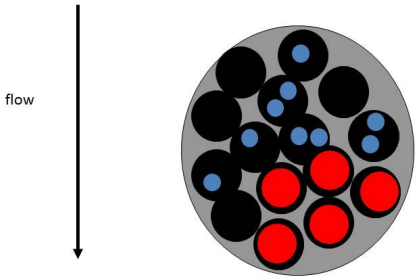
Gel filtration chromatography











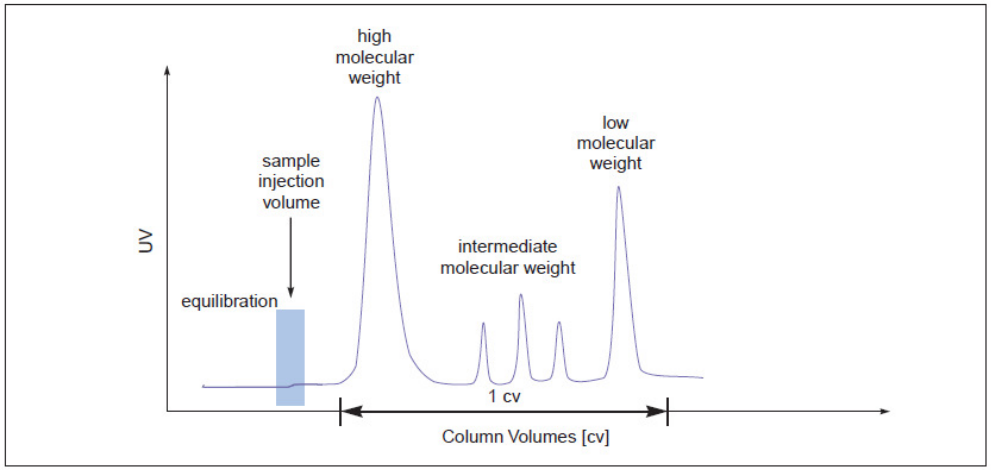


Fig. 41. Typical GF elution.

Gel filtration chromatography

- The molecular mass of the smallest molecule unable to penetrate the pores of the gel is at the **exclusion limit.**
- The exclusion limit is a function of molecular shape, since elongated molecules are less likely to penetrate a gel pore than other shapes.
- Behavior of the molecule on the gel can be quantitatively characterized.

Total bed volume of the column

$$V_t = V_x + V_0$$

V_x = volume occupied by gel beads

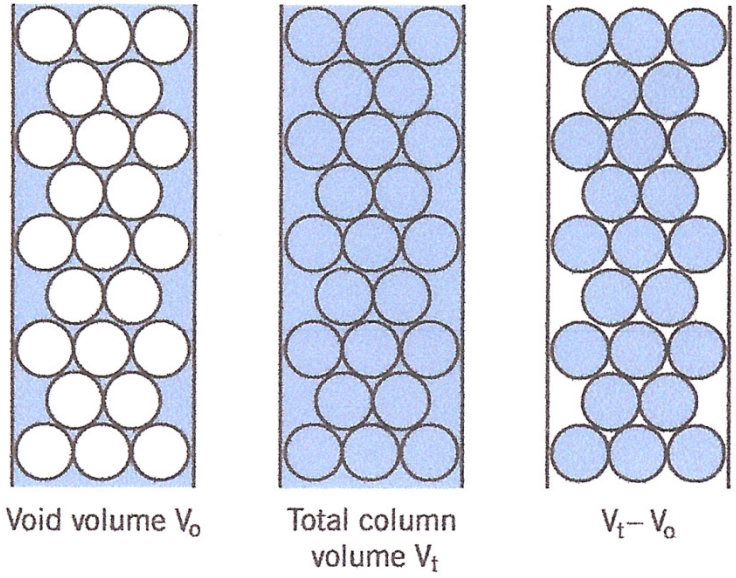
V_0 = volume of solvent space surrounding gel; Typically 35%

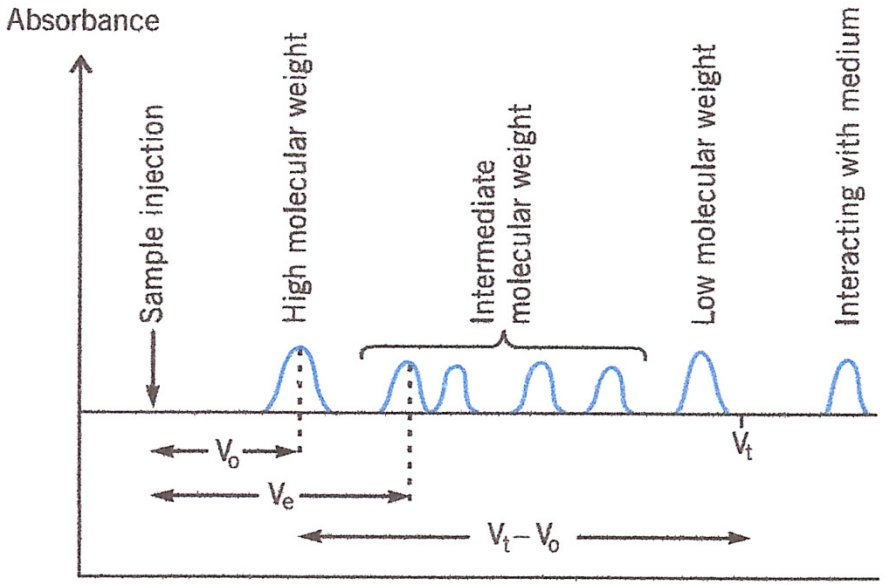
Gel filtration chromatography

- **Elution volume** (V_e) is the volume of a solvent required to elute a given solute from the column after it has first contacted the gel**.**

- **Relative elution volume** (V_e/V_0) is the behavior of a particular solute on a given gel that is independent of the size of the column.

- This effectually means that molecules with molecular masses ranging below the exclusion limit of a gel will elute from a gel in the order of their molecular masses with the largest eluting first.

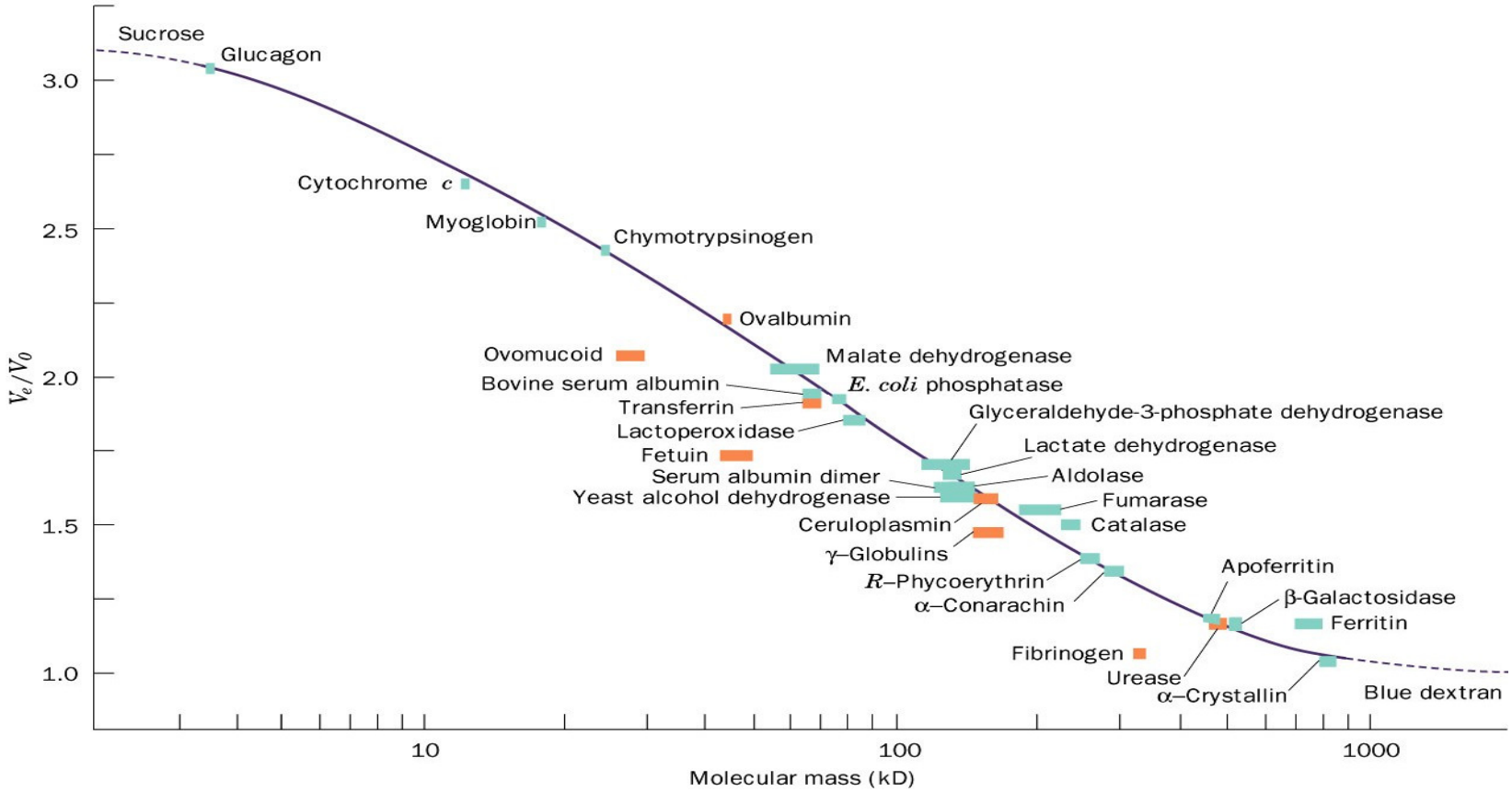




Ve = elution volume of the protein
Vo = void volume. Blue dextran 2000
Vt = total bed volume

$$Kav = \frac{Ve - Vo}{Vt - Vo}$$

Molecular mass determination by gel filtration chromatography



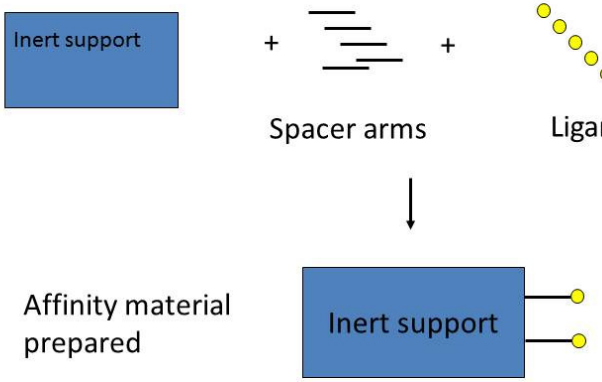
Some Commonly Used Gel Filtration Materials

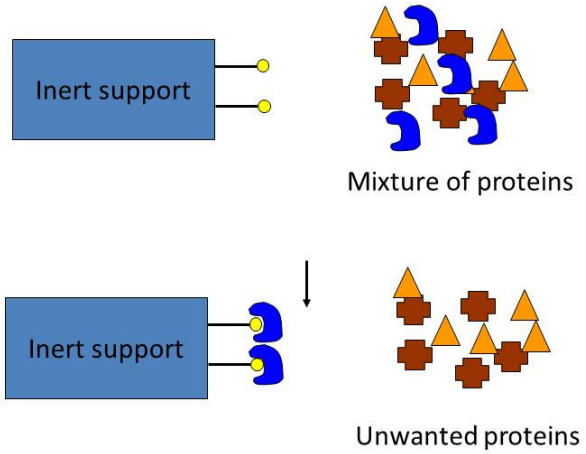
Name[a]	Type	Fractionation Range (kD)
Sephadex G-10	Dextran	0.05–0.7
Sephadex G-25	Dextran	1–5
Sephadex G-50	Dextran	1–30
Sephadex G-100	Dextran	4–150
Sephadex G-200	Dextran	5–600
Bio-Gel P-2	Polyacrylamide	0.1–1.8
Bio-Gel P-6	Polyacrylamide	1–6
Bio-Gel P-10	Polyacrylamide	1.5–20
Bio-Gel P-30	Polyacrylamide	2.4–40
Bio-Gel P-100	Polyacrylamide	5–100
Bio-Gel P-300	Polyacrylamide	60–400
Sepharose 6B	Agarose	10–4,000
Sepharose 4B	Agarose	60–20,000
Sepharose 2B	Agarose	70–40,000
Bio-Gel A-5	Agarose	10–5000
Bio-Gel A-50	Agarose	100–50,000
Bio-Gel A-150	Agarose	1000–150,000

[a]Sephadex and Sepharose gels are products of Amersham Pharmacia Biotech; Bio-Gel gels are manufactured by BioRad Laboratories.

Affinity chromatography

- How does it work?
- **Ligand** - a molecule that specifically binds to the protein of interest.





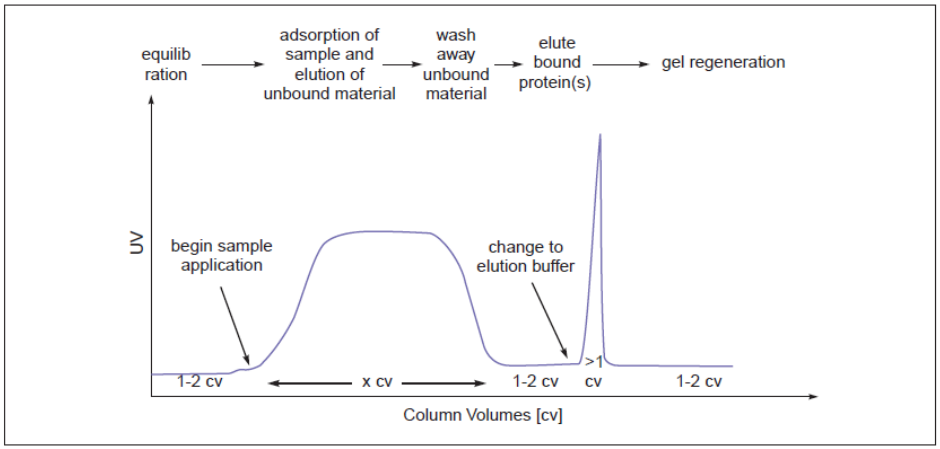
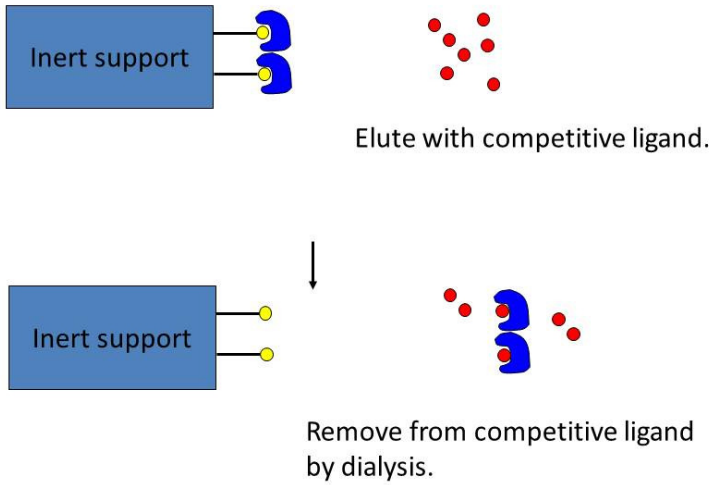


Fig. 40. Typical affinity separation.



Immunoaffinity chromatography

- **Monoclonal antibodies can be attached to the column material.**
- **The column only binds the protein against which the antibody has been raised.**
- **10,000-fold purification in a single step!**
- **Disadvantges**
  - **Difficult to produce monoclonal antibodies (expensive $$!)**
  - **Harsh conditions to elute the bound protein**

▪ **When a recombinant protein is expressed the specific tag added is used for the later purification.** **GST** **GSH-Sepharose**

6-His tag **Metal chelate affinity chromatography Zn^{2+}, Ni^{2+}**

Example: Recombinant protein **GST-~-Protein X**

Matrix **GSH-Sepharose**

Elution **free GSH, NaCl**

Protese cleavage **GSH-Sepharose + GST-~-Protein X**

↓

GSH-Sepharose – GST + Protein X

GST tag	(His)6 tag
Can be used in any expression system	Can be used in any expression system
High yields of pure product	High yields of pure product
Selection of purification products available for any scale	Selection of purification products available for any scale
Site-specific proteases enable cleavage of tag if required	Site-specific proteases enable cleavage of tag if required
pGEX6P PreScission™ protease vectors enable cleavage.	Small tag may not need to be removed. Purification in a single step. The fusion partner can be used directly as an antigen in antibody production
GST tag easily detected using an enzyme assay or an immunoassay	(His)6 tag easily detected using an immunoassay
Simple purification. Very mild elution conditions	Simple purification, but elution conditions are not as minimize risk of damage to functionality and mild as for GST fusion proteins. Neutral pH but imidazole may cause precipitation. Desalting to remove imidazole may be necessary

GST tag	(His)6 tag
GST tag can help stabilize folding of recombinant proteins	(His)6 - dihydrofolate reductase tag stabilizes small peptides during expression
Fusion proteins form dimers	Small tag is less likely to interfere with structure and function of fusion partner
	Mass determination by mass spectrometry not always accurate for some (His)6 fusion proteins*

Affinity chromatography

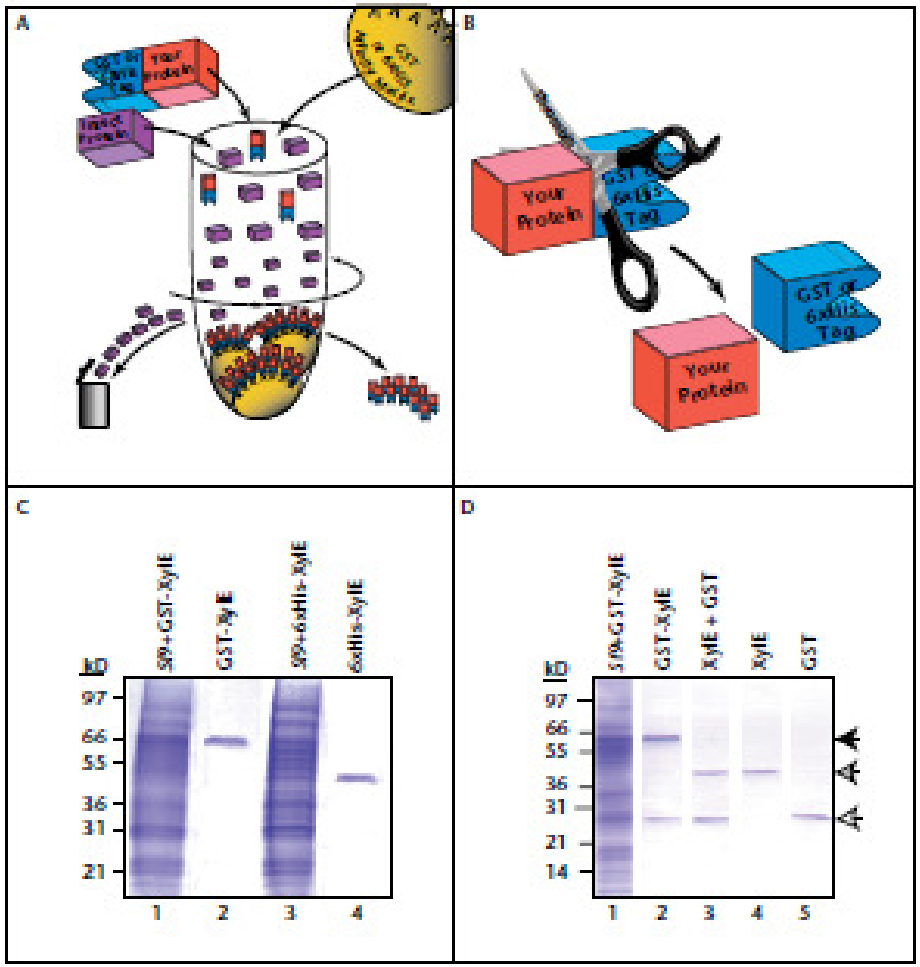


Figure 13. Expression, purification and cleavage of fusion proteins. A) Single-step protein purifi-

Other chromatographic methods

- **(HIC) Hydrophobic interaction chromatography (HIC)**- the stationary phase is hydrophillic (agarose gel) with substituted hydrophobic groups. Interactions with column are relatively weak and can be used for the separation of native proteins (not denatured), so proteins are separated based on surface hydrophobicity.

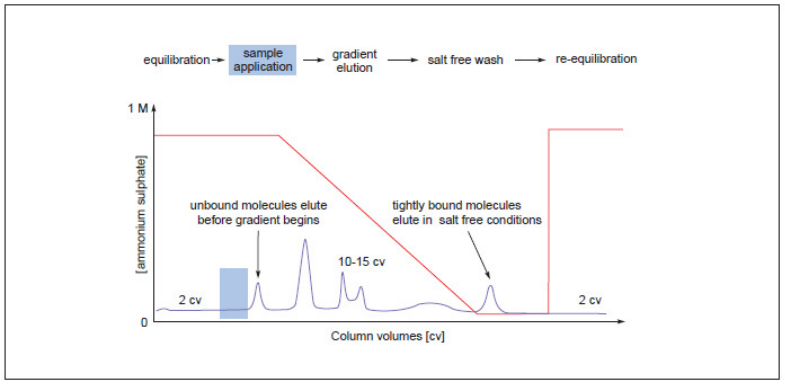


Fig. 36. Typical HIC gradient elution

- **Reverse-phase chromatography (RPC)**- separates nonpolar substances including denatured proteins.
- Stationary phase is non-polar and the mobile phase is a more polar liquid. Used to separate lipids but can also be used for proteins.
- Solvent must be highly non-polar usually high concentration of organic solvent (acetonitrile) so it denatures proteins so that the hydrophobic cores can interact with the matrix.

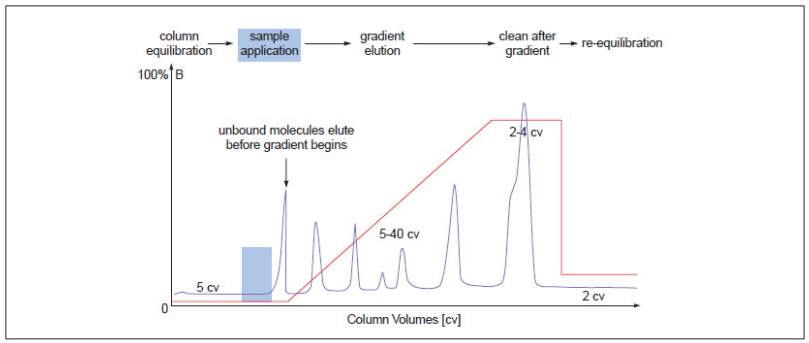


Fig. 43. Typical RPC gradient elution.

Other chromatographic methods

- **High performance liquid chromatography (HPLC)**- may be based on adsorption, ion exchange, size exclusion, HIC or RPC but is improved because of the noncompressible matrix.
- Can be made of silica and withstand very high pressures (up to 5000 psi) so flow rates can be very high.

- **Advantages of HPLC**
  - **High resolution**
  - **Fast**
  - **High sensitivity**
  - **Can be easily automated**

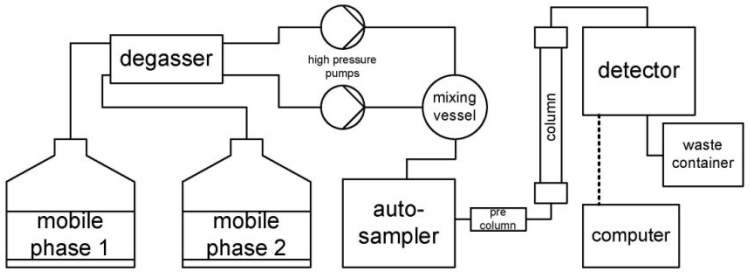


Fig.2 Basic flow chart of a HPLC system setup

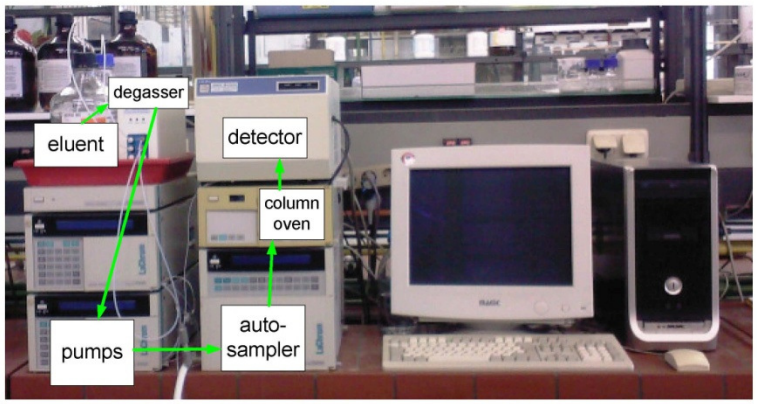


Fig.3 Picture of laboratory HPLC system setup

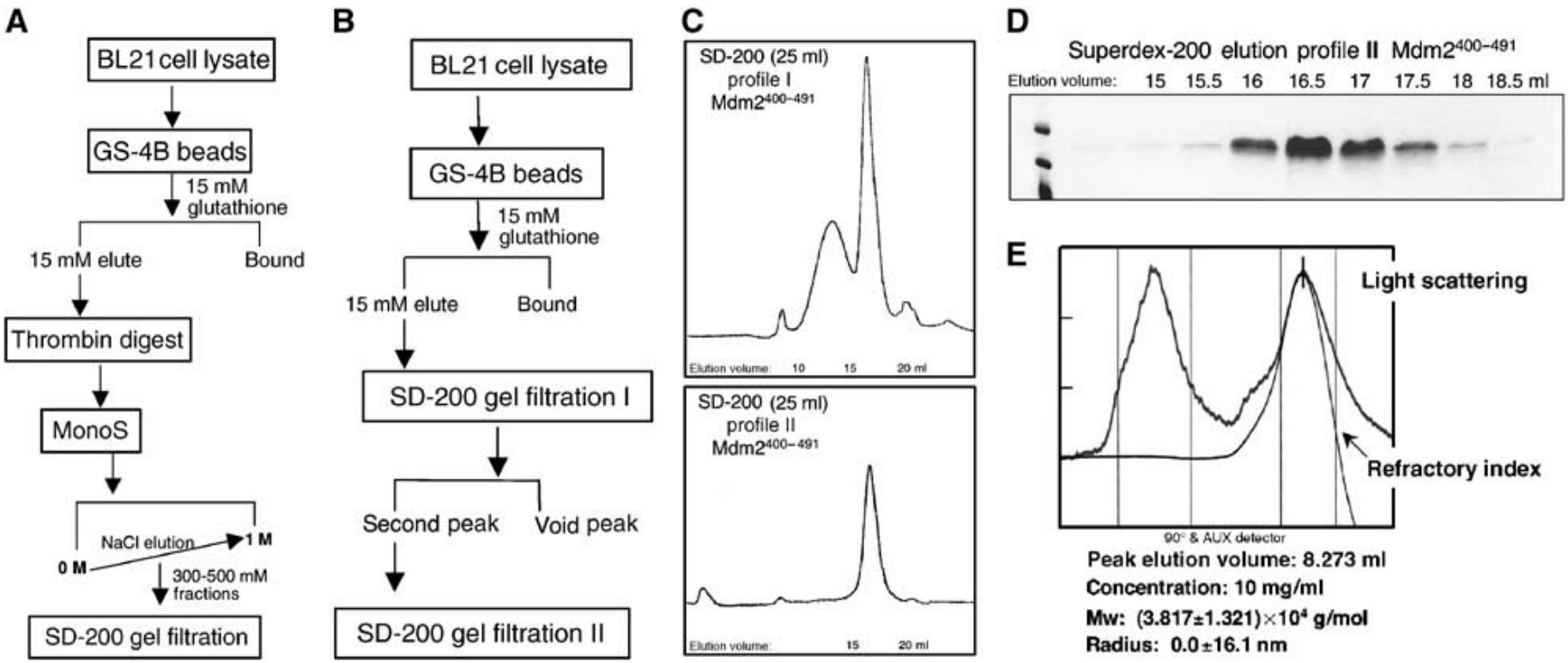
Purification of Rat Liver Glucokinase.

Stage	Specific Activity (nkat · g^{-1})[a]	Yield (%)	Fold[b] Purification
Scheme A: A "traditional" chromatographic procedure			
1. Liver supernatant	0.17	100	1
2. $(NH_4)_2SO_4$ precipitate	*c*	*c*	*c*
3. DEAE-Sephadex chromatography by stepwise elution with KCl	4.9	52	29
4. DEAE-Sephadex chromatography by linear gradient elution with KCl	23	45	140
5. DEAE-cellulose chromatography by linear gradient elution with KCl	44	33	260
6. Concentration by stepwise KCl elution from DEAE-Sephadex	80	15	480
7. Bio-Gel P-225 chromatography	130	15	780
Scheme B: An affinity chromatography procedure			
1. Liver supernatant	0.092	100	1
2. DEAE-cellulose chromatography by stepwise elution with KCl	20.1	104	220
3. Affinity chromatography[d]	**420**	**83**	**4500**

[a]A **katal** (abbreviation **kat**) is the amount of enzyme that catalyzes the transformation of 1 mol of substrate per second under standard conditions. One nanokatal (nkat) is 10^{-9} kat.
[b]Calculated from specific activity; the first step is arbitrarily assigned unity.
[c]The activity could not be accurately measured at this stage because of uncertainty in correcting for contamination by other enzymes.
[d]The affinity chromatography material was made by linking glucosamine (an inhibitor of glucokinase) through a 6-aminohexanoyl spacer arm to NCBr-activated agarose.

Source: Cornish-Bowden, A., *Fundamentals of Enzyme Kinetics, p.* 48, Butterworth (1979), as adapted from Parry, M.J. and Walker, D.G., *Biochem. J.* **99,** 266 (1966) for Scheme A and from Holroyde, M.J., Allen, B.M., Storer, A.C., Warsey, A.S., Chesher, J.M.E., Trayer, I.P., Cornish-Bowden, A., and Walker, D.G., *Biochem. J.* **153,** 363 (1976) for Scheme B.

Scheme of purification



Poyurovsky MV et al (2007). EMBO J. 25, 90-101

Figure 1 Purification of the RING domain of Mdm2. (A, B) Purification of Mdm2 protein. Flow-chart representations of the purification schemes for the RING domain of Mdm2 with (B) and without (A) a GST tag. (C) Mdm2 can be purified to homogeneity by size-exclusion chromatography. Mdm2 RING domain sample was prepared as outlined in (A). Upper panel: Superdex-200 gel-filtration absorbance profiles of Mdm2400491 RING (#I); lower panel: peak fractions of around 16 ml were collected and subjected to a second round of gel filtration (#II). (D) Purified Mdm2 RING domain is a single species. Fractions eluted from the second Superdex-200 gel-filtration column (#II) were subjected to 12% SDS–PAGE gel and stained with Coomassie blue. (E) Analytical gel-filtration and static light-scattering analysis of the Mdm2 RING monomer fraction. The GST-Mdm2 peak fractions from the second gel filtration (#II) were injected onto an analytical gel-filtration column at 200 mM concentration and the effluent was monitored by refractive index (bottom trace, arrow) and 901 static light-scattering (top trace) detectors. Calculations from the Debye plot estimate a molecular mass of 36.1 kDa for the protein peak of the elution profile.

References

-Protein purification, Principles and Practice. Scope R.K (1982). Springer-Verlag. New York Heidelberg Berlin

-Baculovirus Expression Vector System Manual. Introduction Manual. 6th Edition. URL://www.pharmigen .com

-Methods in Molecular Biology. Vol 235: E.coli Plasmid vectors. Edited by Casali N. and preston A.. Hmana Press Inc. Totowa, NJ

-Biochemistry. Voet, Voet.

- Handbooks from Amersham Biosciences