Protein purification

Another critical step of protein engineering – large scale production and purification

5 L

50 L

100 L

different sizes of fermentors

large-scale chromatography

Protein purification - recombinant proteins

Protein over-expressed in a model system (*E. coli*, yeast, ..)

Usually plenty of desired protein *per* cell (> 40 % of the total cellular protein)

Recombinant proteins usually modified to make the purification easier (various tags)

Problems with **solubility** (aggregation into inclusion bodies) .... sometimes frustrating, vast number of protocols

Proteins have personality -> purification is a state of art

Biggest pitfall of recombinant protein purification – low solubility

Possible solutions (*E. coli*):

- Fusion Proteins –> using a protein tag
- Expression temperature -> less than 20°C overnight, 'Artic' cell line (10°C)
- Chaperone co-expression -> plasmid sets expressing GroEL, GroES, DnaJ,K..
- Low IPTG (or other expression inducers) conc., auto-inducible growth media
- Various 'magic' tricks ...

Cell disruption

Osmotic lysis: simple and gentle, place cells in a hypotonic solution, could not be enough, hypotonic buffer can disrupt protein interaction, lost of enzyme activity

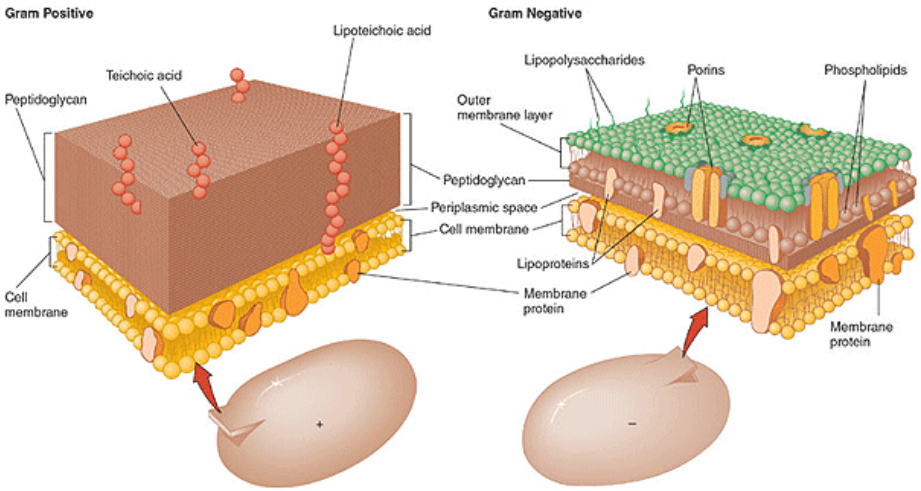
Mechanical: good if cells have a thick cell wall, crushing or grinding of cells, French press (pressure), sonication (sound), bead-beater (beads)

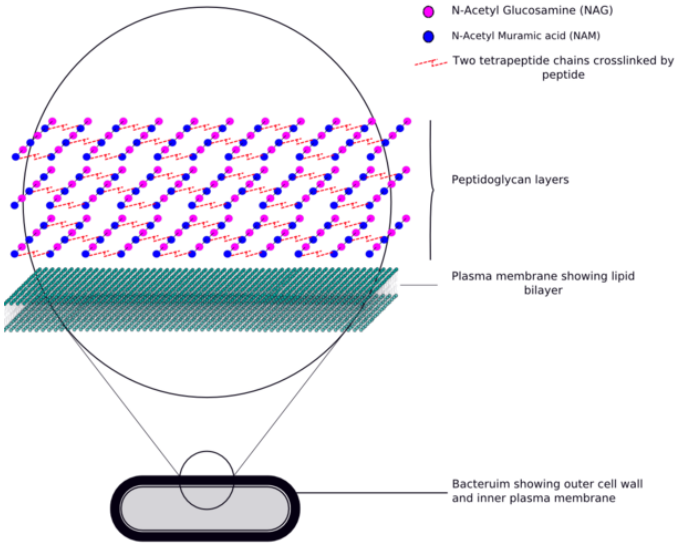
Detergents (Triton X100) + enzymes (lysozym, DNase I) – works well for both gram-positive and gram-negative bacteria (*E. coli*), often combined with sonication

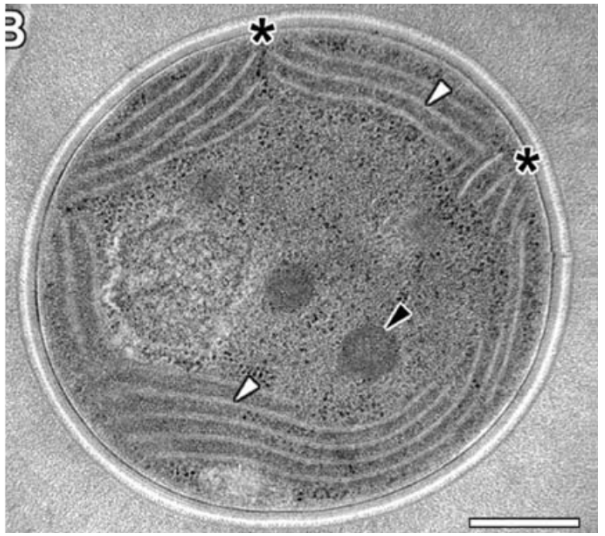
Cell disruption - Lysozyme

~15 KDa enzyme - catalyzes hydrolysis of 1,4-beta-linkages between N-acetylmuramic acid and N-acetyl-D-glucosamine residues in peptidoglycan

Discovered by accident in the A. Fleming's lab (1922)







Many bacterial species –
no chance, thick cell wall (glycoproteins, S-layer)

How to purify a desired protein from complex protein mix?

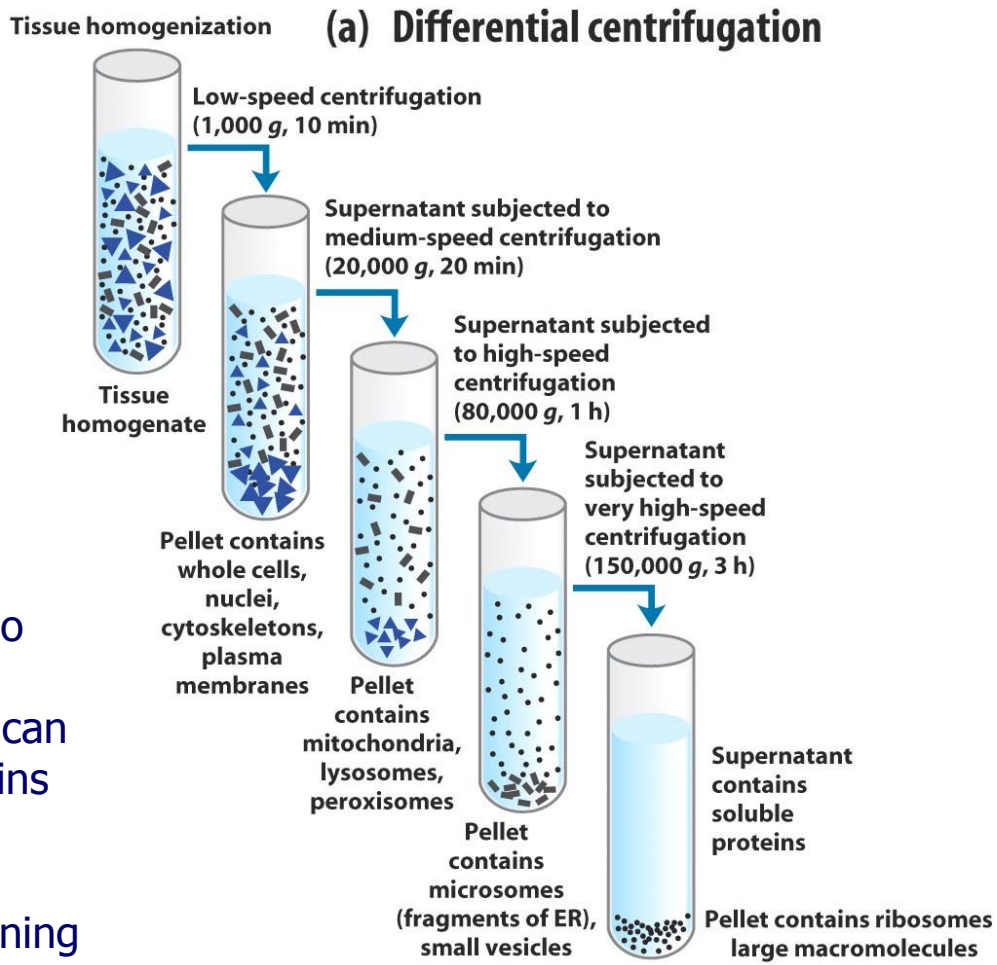
-> Sort proteins using their physical characteristics:

1. solubility
2. electrical charge
3. size + shape
4. affinity for other molecules
5. polarity

Separation of proteins using physical characteristics

Charge	1. Ion exchange chromatography 2. Electrophoresis 3. Isoelectric focusing
Size, shape:	1. Ultracentrifugation 2. Gel electrophoresis 3. Gel filtration (size exclusion) chromatography
Specific affinity to other molecules:	1. Affinity chromatography (+ tagging)
Polarity:	1. Phase separation 2. Reverse-phase chromatography 3. Hydrophobic/hydrophilic chromatography 4. Precipitation by salts/polymers

Differential centrifugation

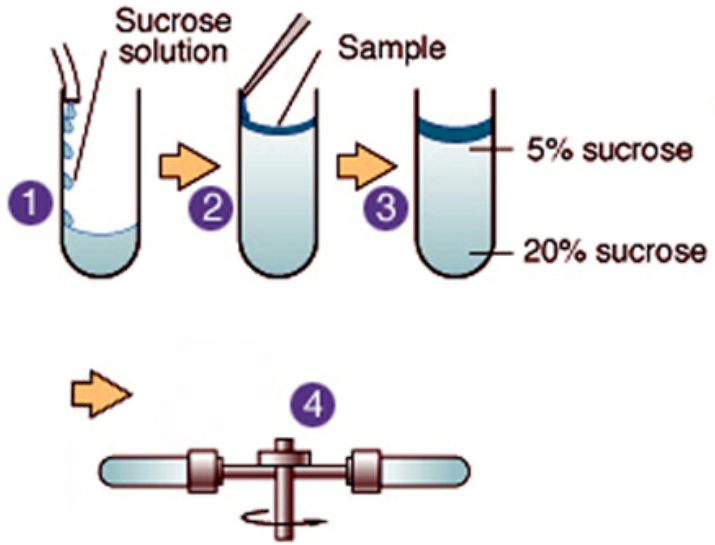


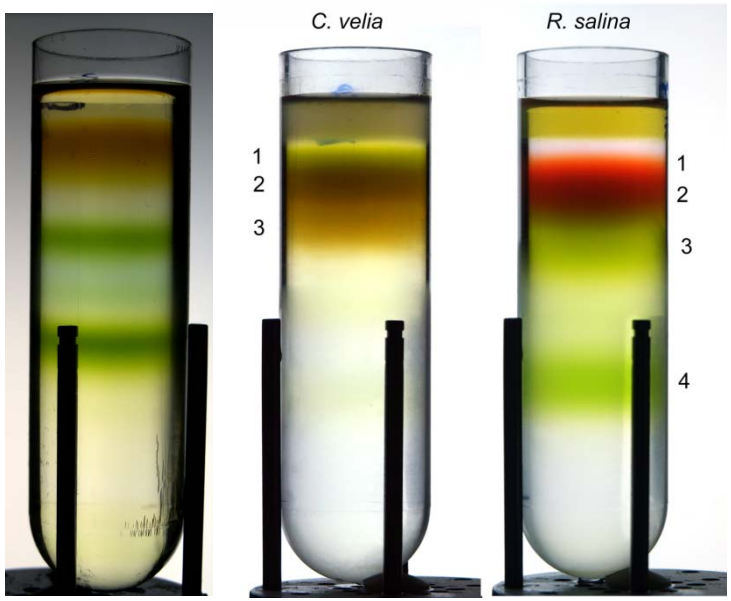
- A pre-purification step, useful to do
- Large complexes and membranes can be well separated from small proteins
- Result is a cell fractionation containing protein of interest

Centrifugation in sucrose / glycerol gradient

Sucrose (glycerol) density gradient is created by gently overlaying lower concentrations of sucrose on higher concentrations in a centrifuge tube

The sample is load on the top -> ultracentrifugation (~ 80 - 280 K x *g)*





Thylakoid membrane complexes from the cyanobacterium *Synechocystis* 6803 and two algae

Problem: takes long time (overnight..)

Precipitation of proteins by ammonium sulfate (salting out)

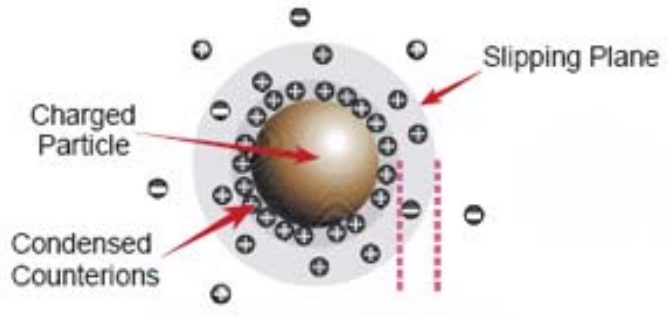
• The solubility of proteins varies according to the ionic strength of the solution, and hence according to the salt concentration. Protein precipitation at high salt conc. is called "salting out"

• Principe - the excess ions (not bound to the protein) compete with proteins for the solvent. The decrease in solvation allows the proteins to aggregate and precipitate

• Proteins differ markedly in their solubilities at high ionic strength, salting-out is a useful procedure to assist in the purification of a given protein

• Ammonium Sulfate $(NH_4)_2SO_4$ is commonly used as its solubility is so high that salt solutions with high ionic strength are allowed (PEG can be an alternative)

• Once the protein is precipitated (not denatured) can by separate by centrifugation and re-dissolved in buffer for further purification

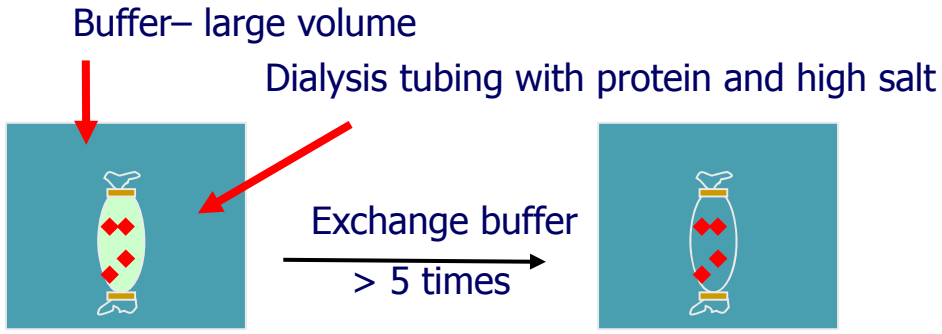


It does not work for hydrophobic proteins (membrane proteins)

Desalting, buffer exchange

Following e.g. salting-out step, the solution will contain a high concentration of salt that can be disruptive to subsequent chromatographic steps.

Salt can be removed by dialysis – dialysis tubing has pores with a specific molecular weight cut-off that allows smaller molecules (salt) to pass.



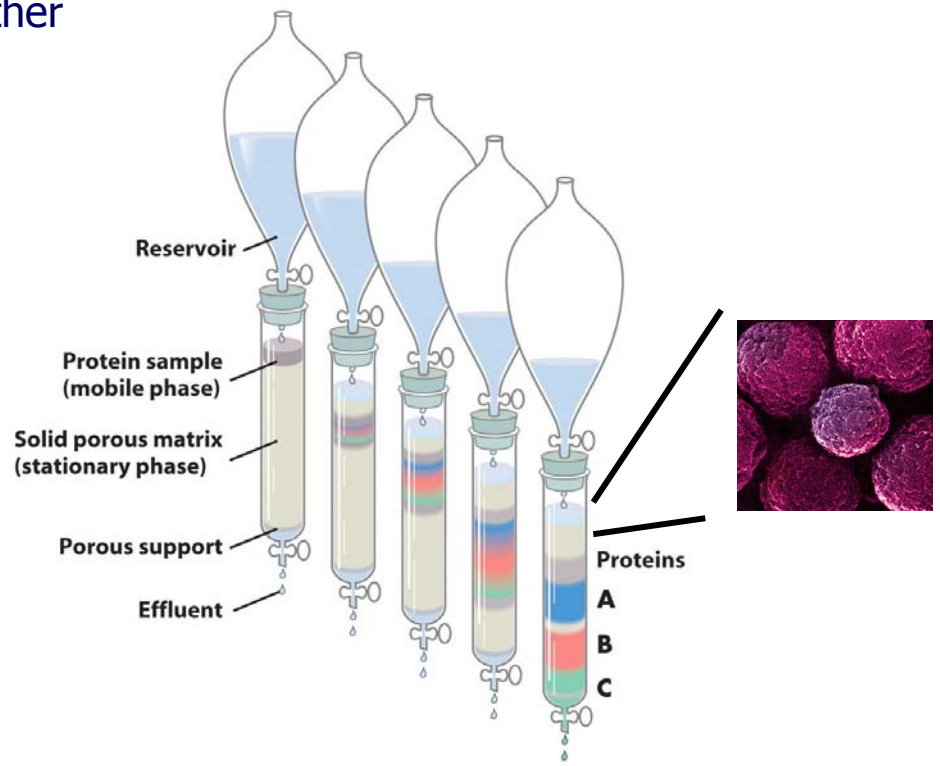
Slide-Lyzer (PIERCE)

Takes long time, alternative –> desalting columns (or another chromatography)

Chromatographic methods

How chromatography works:

- Immiscible stationary (usually small beads) and mobile phases (fluid - usually buffer)
- An arrangement where a mixture is deposited at one end of the stationary phase
- Flow of the mobile phase towards the other end of the stationary phase
- Different rates of partitioning for each component



Chromatography - terminology

Resin

- stationary phase (beads, gel with specific structure or functional groups)

Mobile Phase

- fluid (buffer) carrying the mixture of analytes

Elution

- The process of moving the solutes out of the chromatographic system
- fraction(s) with purified protein(s)

Eluent

- additional solvents used for elution

Residency

- time spent on column

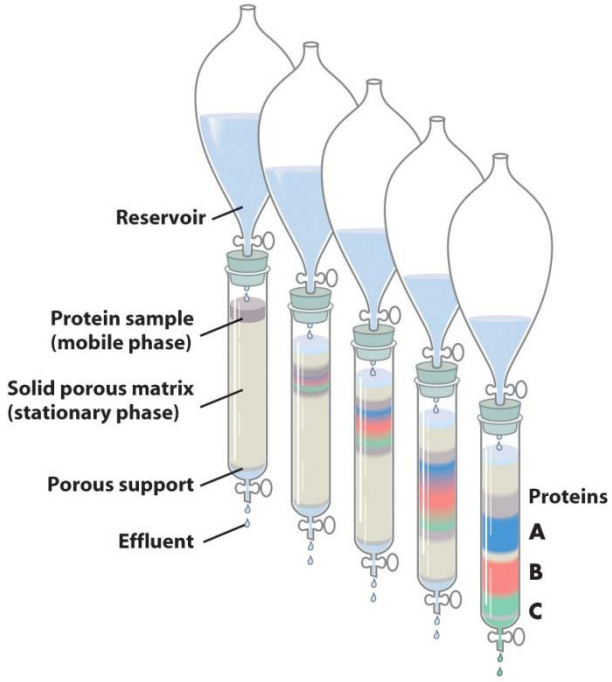


TABLE 3–5 A Purification Table for a Hypothetical Enzyme

Procedure or step	*Fraction volume (ml)*	*Total protein (mg)*	*Activity (units)*	*Specific activity (units/mg)*
1. Crude cellular extract	1,400	10,000	100,000	10
2. Precipitation with ammonium sulfate	280	3,000	96,000	32
3. Ion-exchange chromatography	90	400	80,000	200
4. Size-exclusion chromatography	80	100	60,000	600
5. Affinity chromatography	6	3	45,000	15,000

Note: All data represent the status of the sample *after* the designated procedure has been carried out. Activity and specific activity are defined on page 94.

Fractions are usually concentrated using concentrators (ultrafiltration) with well defined cut-off (3 KDa – 1 MDa) – sharper peaks, better purity.

There are two ways to run a column:

- Allow **gravity** to draw samples and buffers through the column resin
- Use **pumps** to push a sample and buffers through a column

a) Peristaltic pump – there are better ways..

b) Fast-Performance Liquid Chromatography (FPLC)

–> large volumes of sample can be injected
preparative methods, columns can be prepared in lab
Designed for biological buffers

c) High-Performance Liquid Chromatography (HPLC)

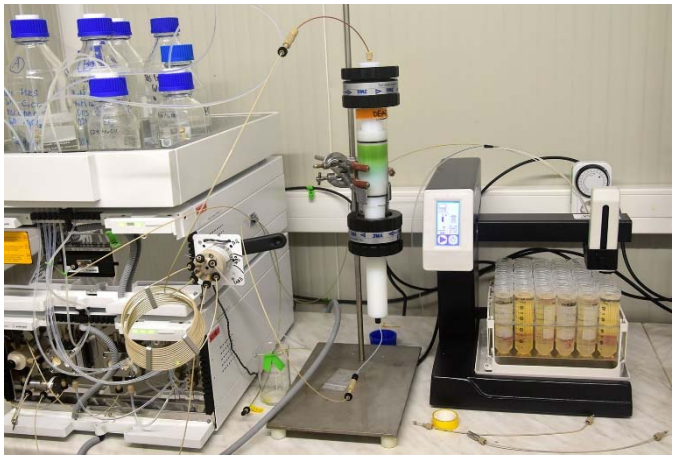
-> expensive columns, **analytical volumes** (10 – 500 μL), can be very accurate, problem with high salt conc.

d) Preparative HPLC

-> even more expensive columns, large volumes

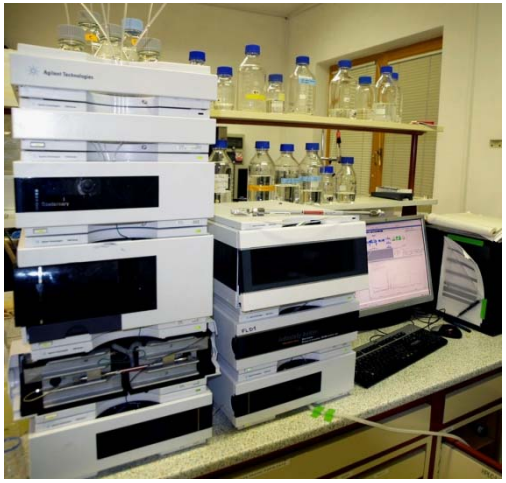
FPLC, HPLC, prepHPLC...

Current situation in the Laboratory of Photosythesis, Třeboň



FLPC + DAD detector

HPLC #1 – 1x DAD,
2x Fluorescence detector



HPLC #2 – 1x DAD, 1 x scintilation and
1 x light scatering detectors

prep HPLC – 1x 2-wavelengths spectrophotometer,
1 x Mass-Spec

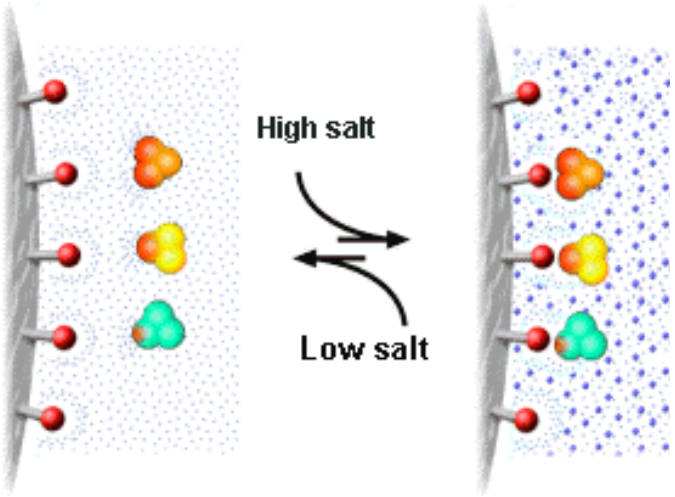
Hydrophobic Interaction Chromatography (HIC)

Although most hydrophobic amino acids are buried in the interior of proteins, many proteins have hydrophobic surfaces or patches which can be used for separation

HIC resin/gel is a hydrophilic polymer backbone (cross-linked dextran, agarose ...) with small non-polar groups (butyl, octyl, phenyl)

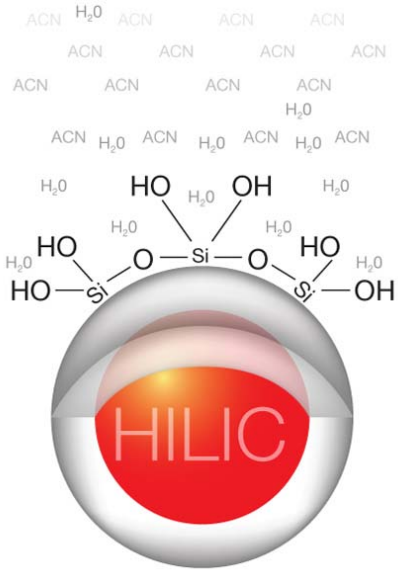
High salt concentrations often used to adsorb the sample to the HIC gel. This makes it especially well-suited as a capture step after ammonium sulfate precipitation

Proteins are eluted from HIC columns via a gradient of high salt to low salt concentrations (pH of the buffer increase/decrease binding)



Hydrophilic Interaction Chromatography (HILIC)

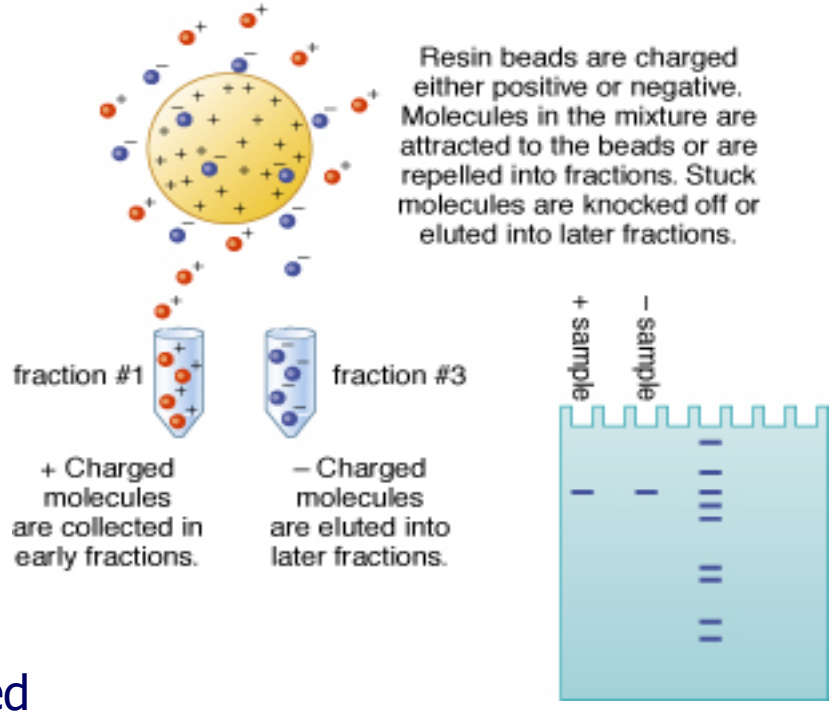
HILIC columns (HPLC, FPLC):



Better for small proteins or peptides (proteomics), bigger proteins have usually multiple binding sites

Ion exchange chromatography

Separation of ions and polar molecules (proteins) based on their charge



Resins are manufactured with ions attached

The ions present a certain degree of positive or negative charge, **depending on the buffer pH and salt concentration**

Ion exchange chromatography – two types o resin

The stationary phase (resin) surface displays ionic functional groups (R-X) that interact with analyte ions of opposite charge

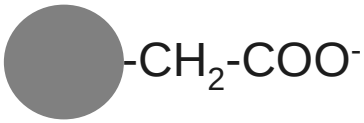
Cation exchange chromatography **retains positively charged proteins** because the stationary phase displays a negatively charged functional group. "S" exchangers.

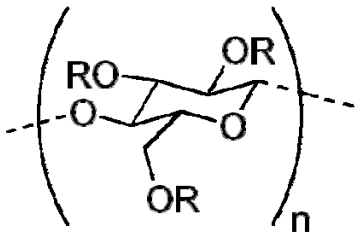
Anion exchange chromatography retains ... "Q" exchangers.

Cation / Anion exchangers

Textbook examples of resins: diethylaminoethylcellulose (DEAE-cellulose), which is positively charged, and carboxymethylcellulose (CM -cellulose), which is negatively charged.

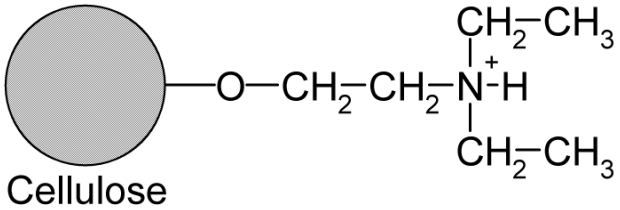
Cation exchanger - CM-cellulose





R = H or CH_2CO_2H

Anion exchanger - DEAE-cellulose



Rather weak binding, not always an disadvantage

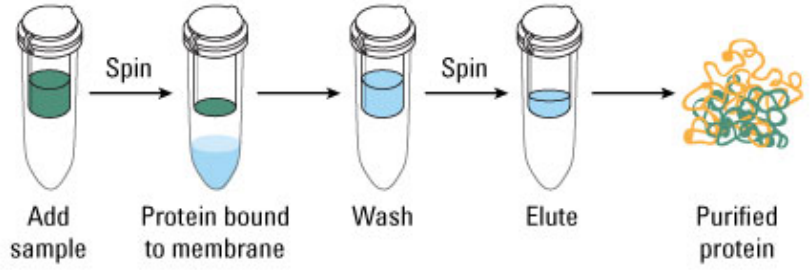
Strong Cation / Anion exchangers - new generation

$-CH_2-N^+(CH_3)_3$ as **Q** and $-CH_2-SO^{3-}$ as **S** exchanger functional groups

Monobeads columns (Mono Q/S): monodisperse 10 µm porous beads

Monolithic columns (Uno Q/S): In continuous-bed technology, monomers are polymerized directly in the column and aggregate into a dense homogeneous network. Expensive

Uno-Q-12 (~4500,- EUR)



Spin columns: best for the sample polishing, small volume of resin, very fast

Ion exchange chromatography – protocol design

Choosing an ion exchanger will depend on:

The pI value of the target protein (pI = pH where protein has zero net charge)
The pH of buffer used

Ionizable group	pKa	pH2	pH7	pH12
C-terminal (COOH)	4.0	oooooooo	----------	----------
Aspartate (COOH)	4.5	oooooooooo	----------	----------
Glutamate (COOH)	4.6	oooooooooo	----------	----------
Histidine (imidazole)	6.2	++++++++++++	oooooooooooooooooooo	
N-terminal (amino)	7.3	++++++++++++++	oooooooooooooooo	
Cysteine (SH)	9.3	oooooooooooooooooooo	----------	
Tyrosine (phenol)	10.1	oooooooooooooooooooooooo	----------	
Lysine (amino)	10.4	++++++++++++++++++++++++	ooooooo	
Arginine (guanido)	12.0	++++++++++++++++++++++++++++	o	

TABLE 3–6 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

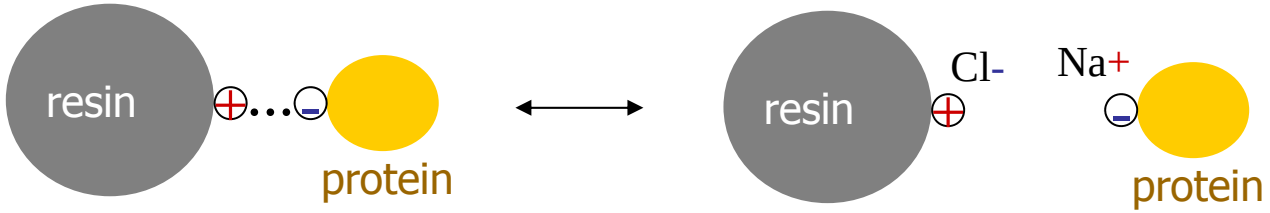
Ion exchange chromatography – protocol design

An (acidic) protein of a pI of ~5

At pH 7 (phosphate buffer, Tris buffer ...) - the protein will carry a net negative charge

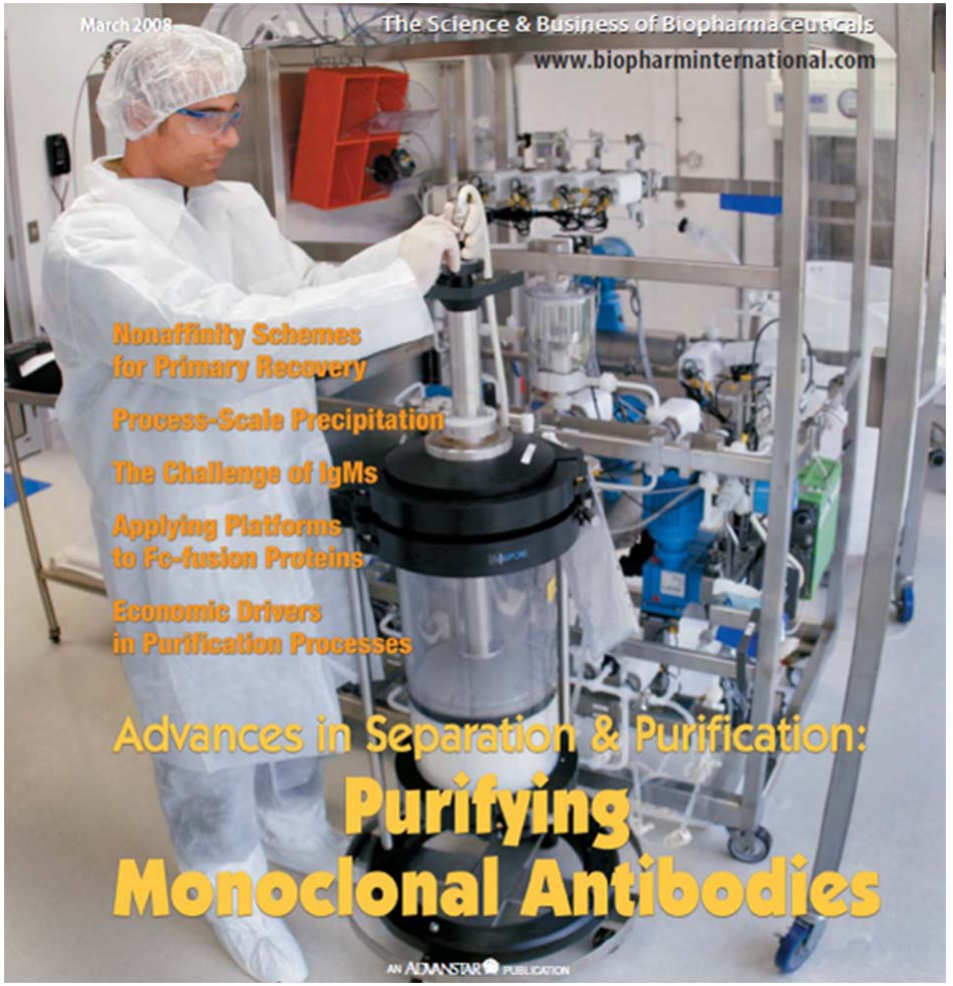
-> the protein will bind to an anion exchanger (Q columns)
-> will be repelled from a cation exchanger (S columns).

-> Apply the protein mixture containing the target protein
-> Remove neutral and basic proteins in flow-through
-> Elute the target protein with increasing salt (NaCl, KCl) or lower/higher pH

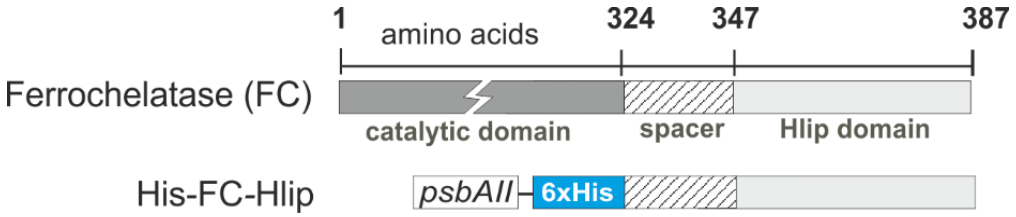


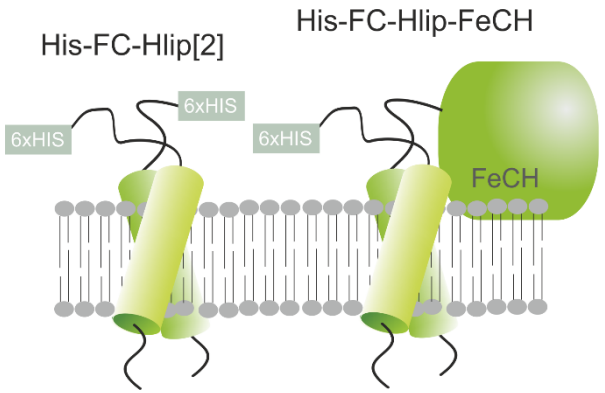
Ion-exchange focusation -> pH gradient e.g. sample in CHES buffer pH 9
-> bind to anion-exchange column -> eluted by increasing concentration
of MES buffer pH 6

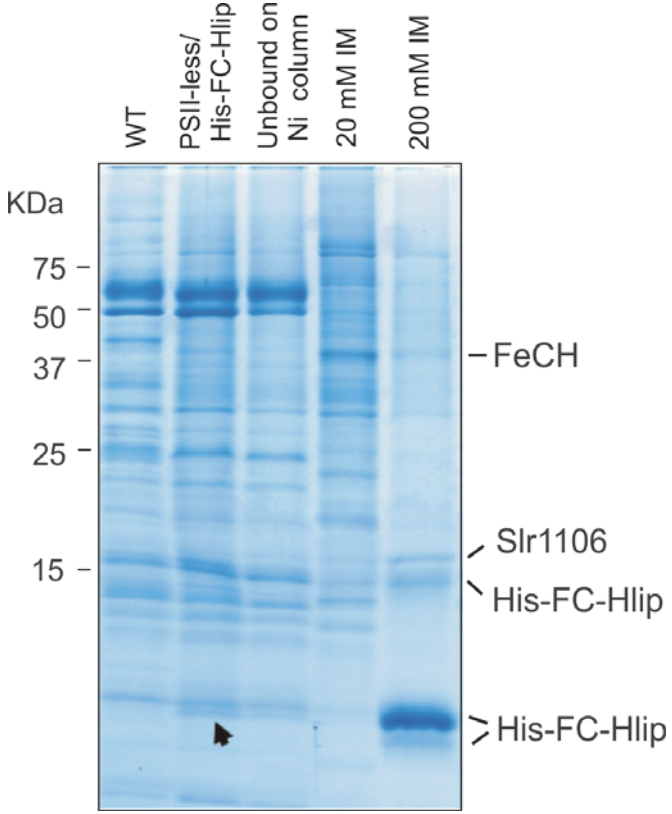
Ion exchange chromatography – workhorse for the industrial protein purification



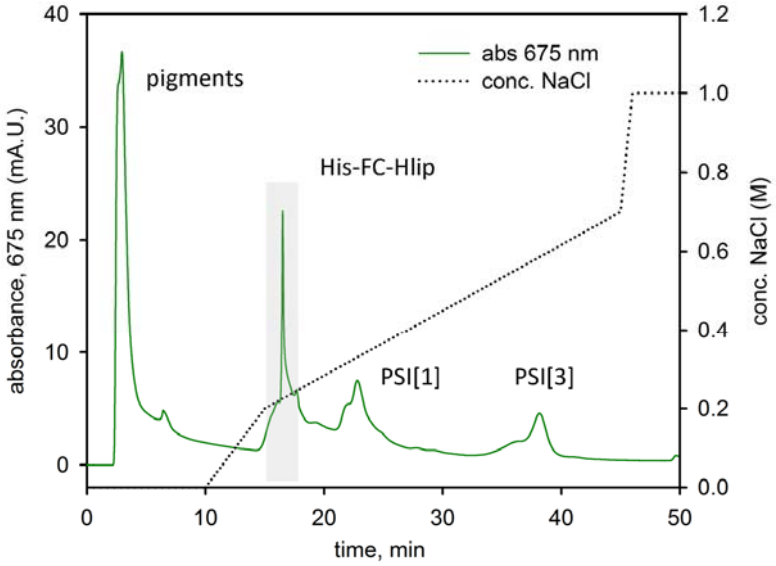
Analytical ion exchange chromatography – Ferrochelatase Hlip domain





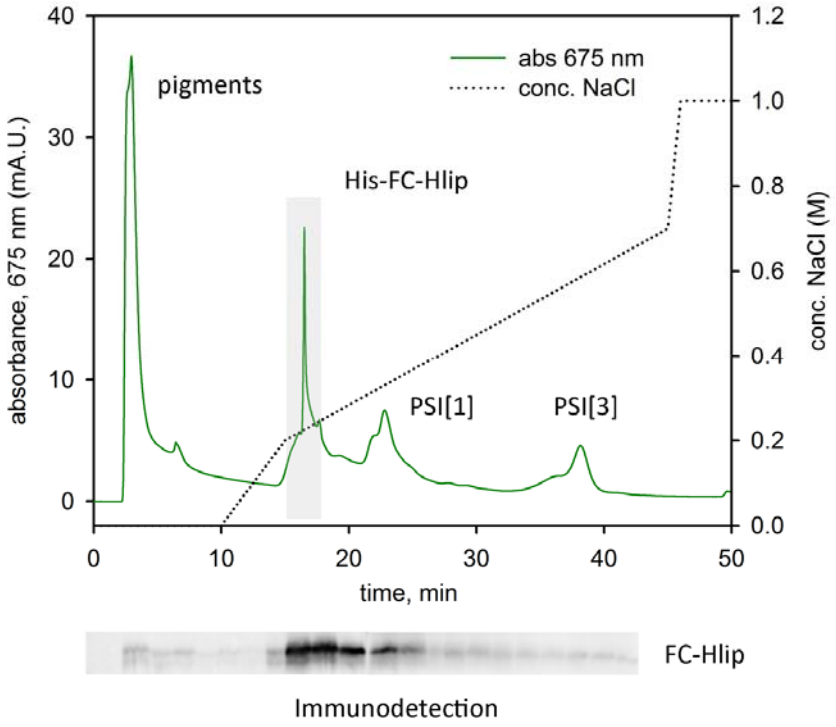


Ion-exchange chromatography (Mono-Q)



Immunodetection

Ion exchange chromatography – an example



MonoQ column (1.6 mL),

Buffer A: 20 mM Hepes pH 8.1 , 5 mM $MgCl_2$, 0.06% β-DDM, Buffer

Buffer B: A + 1M NaCl.

Conditions for HPLC (flow 0.8 mL/min):

0-10min - 0% B
15min - 20% B
45min - 65% B
47min - 100% B
53min - 100% B
54min - 0% B

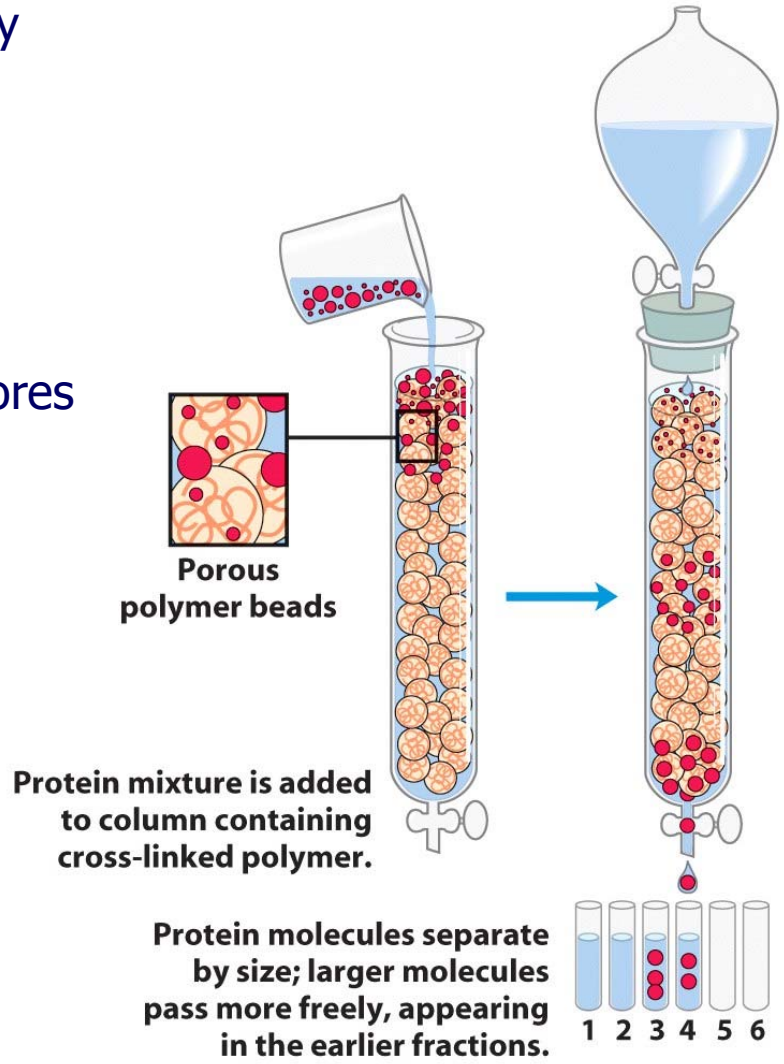
Size exclusion chromatography

Synonymum - gel filtration chromatography

Contains porous beads

Separates according to size and shape

Larger proteins excluded from the small pores

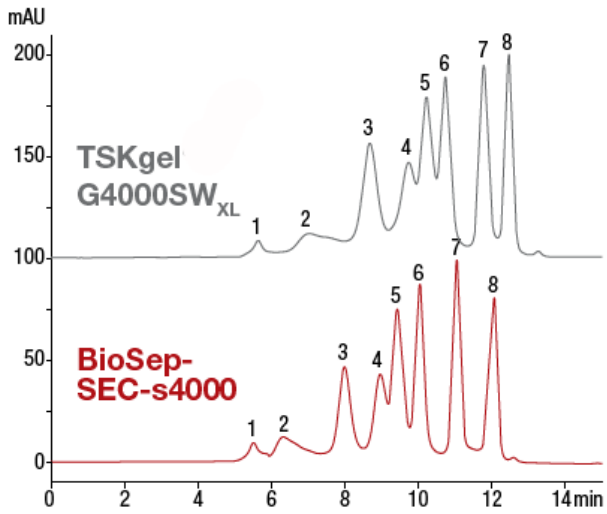


Size exclusion chromatography

Preparative SEC - Protein purification, bigger column (FPLC)

Analytical SEC – determination of native size of a protein/complex, quaternary structure determination, analysis of cofactor bound

Usually employed with HPLC -> SEC column for HPLC, small eluent volume (< 0.5 ml/min), low pressure, isocatric = no gradient, just buffer



Conditions for both columns:

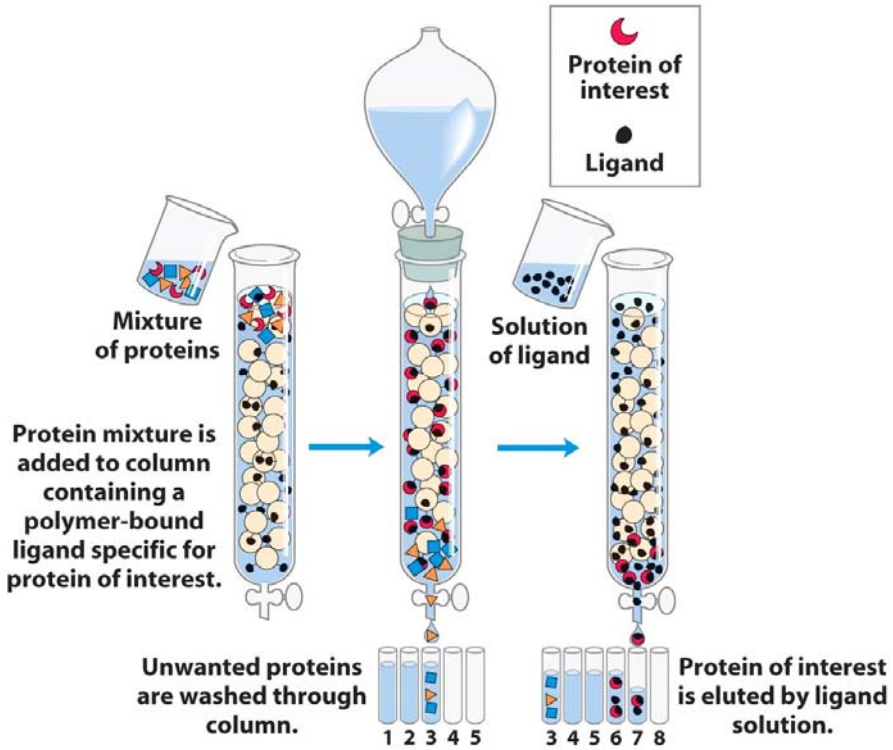
Columns: BioSep-SEC-s4000
TSKgel G4000SW$_{XL}$
Dimensions: 300 x 7.8 mm
Mobile Phase: 100 mM Sodium Phosphate pH 6.8
Flow Rate: 1 mL/min
Temperature: Ambient
Detection: UV @ 220 nm
Sample: 1. HMW impurity
2. IgM 900 kDa
3. Thyroglobulin 670 kDa
4. IgA 300 kDa
5. β-Amylase 200 kDa
6. BSA 66 kDa
7. Ribonuclease A 13.7 kDa
8. Uridine 244 Da

Affinity chromatography

Separates by specific interactions:

Intrinsic >

enzyme <-> substrate
receptor <-> hormone
antigen <-> antibody
metal binding protein <->Cu^{2+}, Ni^{2+}, Co^{2+}
....



Added (protein tagging)>

Specific protein domains can be fused
to proteins of interest at the gene level
to facilitate purification

Protein tagging

Once you have a gene cloned and can over-express the protein, you can alter the protein to improve the ease of purification or detection

You can fuse a tag to the N-or C- terminus of your protein

You can decide to remove the tag or not

Basic strategies:

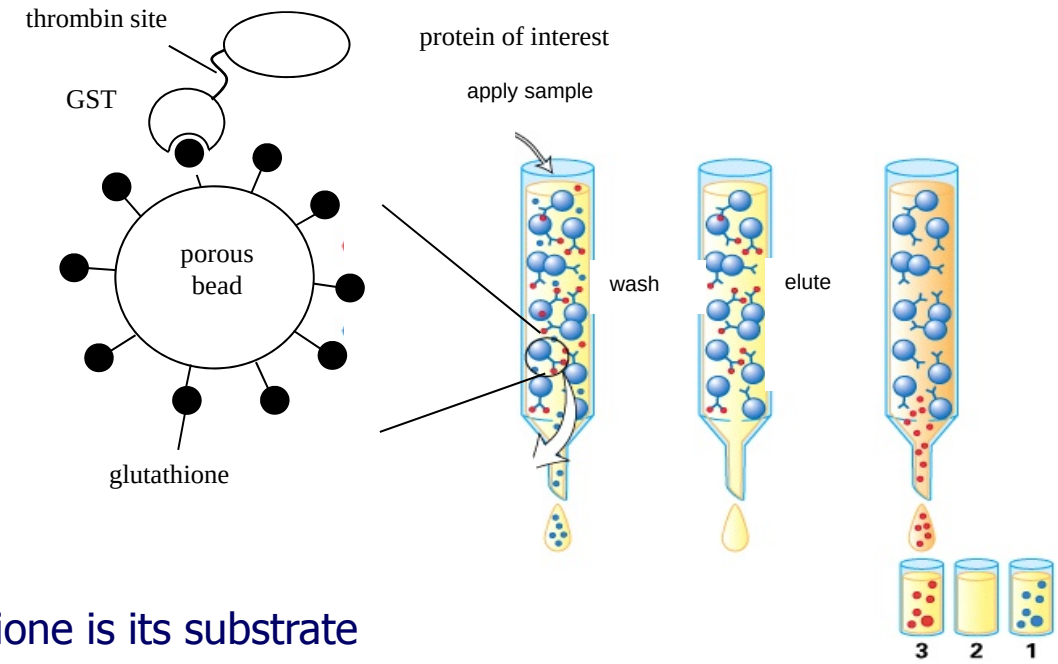
Add an affinity handle for affinity chromatography. By far the most used is the 6xHis-tag, other – GST, Strep, Flag, Myc, CBP ...)

Add sequence/protein that aids in detection (GFP, GST, 3xFlag ..)

Add protein that helps the protein stay soluble (GST, NusA, c-myc, MBP, CBP)

Add signal sequence that causes secretion into culture medium (OmpF)

Glutathion-S-transferase (GST) tag

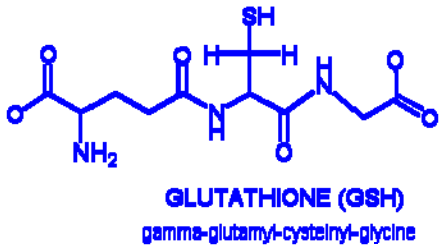


GST –> an enzyme (26 KDa), gluthatione is its substrate

GST-tagged proteins bind to gluthatione on beads

Non-specifically or weakly bound proteins washed off

GST-tagged proteins eluted with glutathione (competitor) or by cleaving off GST (thrombin, HRV protease)



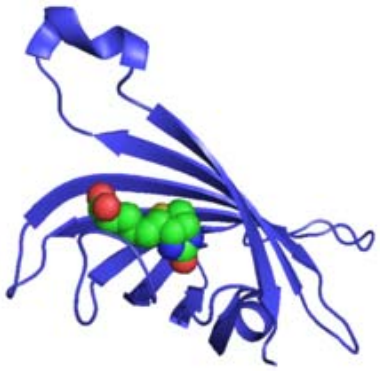
GLUTATHIONE (GSH)
gamma-glutamyl-cysteinyl-glycine

Strep(tavidin) – tag system

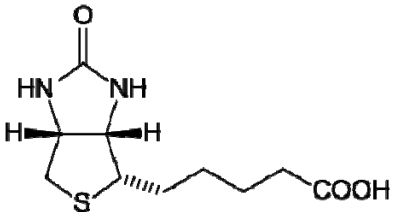
Streptavidin is a 60 KDa protein purified from the bacterium *Streptomyces avidinii*, has an extraordinarily high affinity for biotin

Strep II tag (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys) - peptide sequence exhibits intrinsic affinity towards Strep-Tactin, a specifically engineered streptavidin (provided by IBA company)

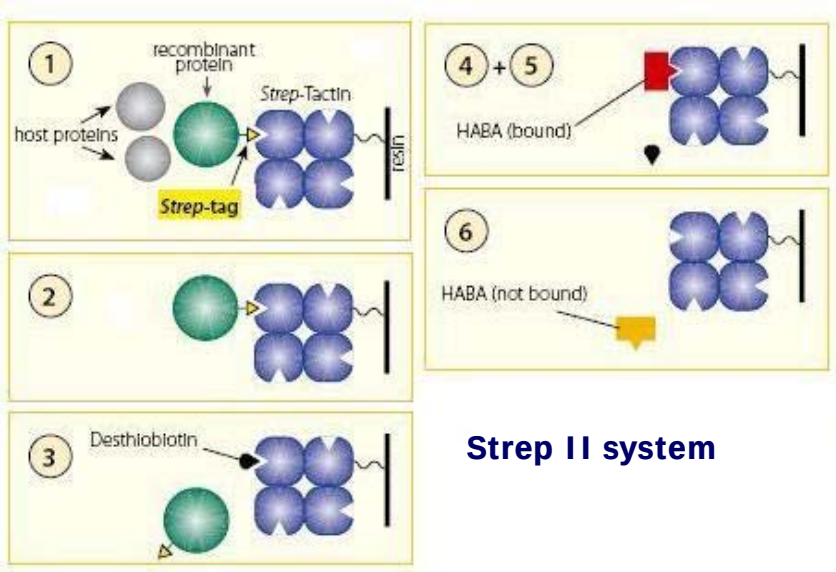
problems -> resin (IBA) binds unspecifically hydrophobic proteins



Streptavidin



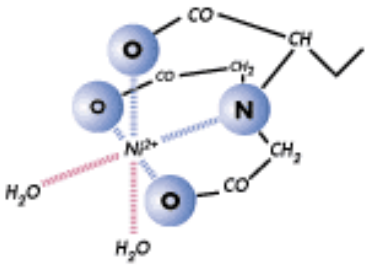
Biotin



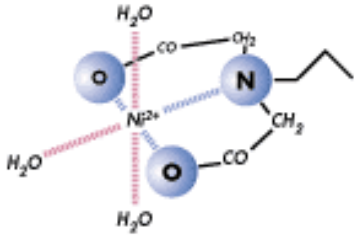
Strep II system

Poly-histidine tag (His-tag)

A fusion with chain of histidines (6-10) allow the protein to bind to columns containing chelated metal supports (usually Ni^{2+} or Co^{2+})

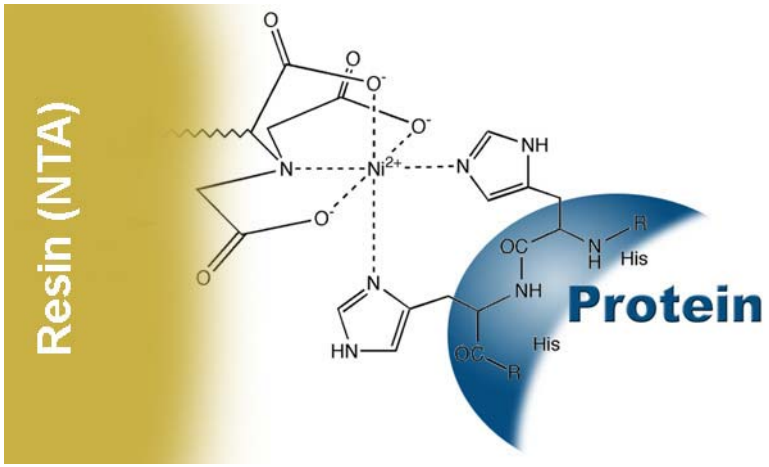


Ni-NTA resin



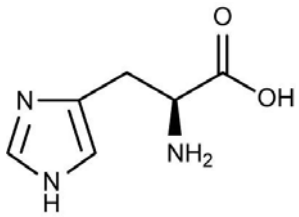
Most metal-chelating resins

NTA-agarose

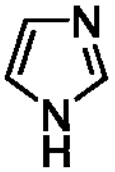


His-tag and metal affinity chromatography

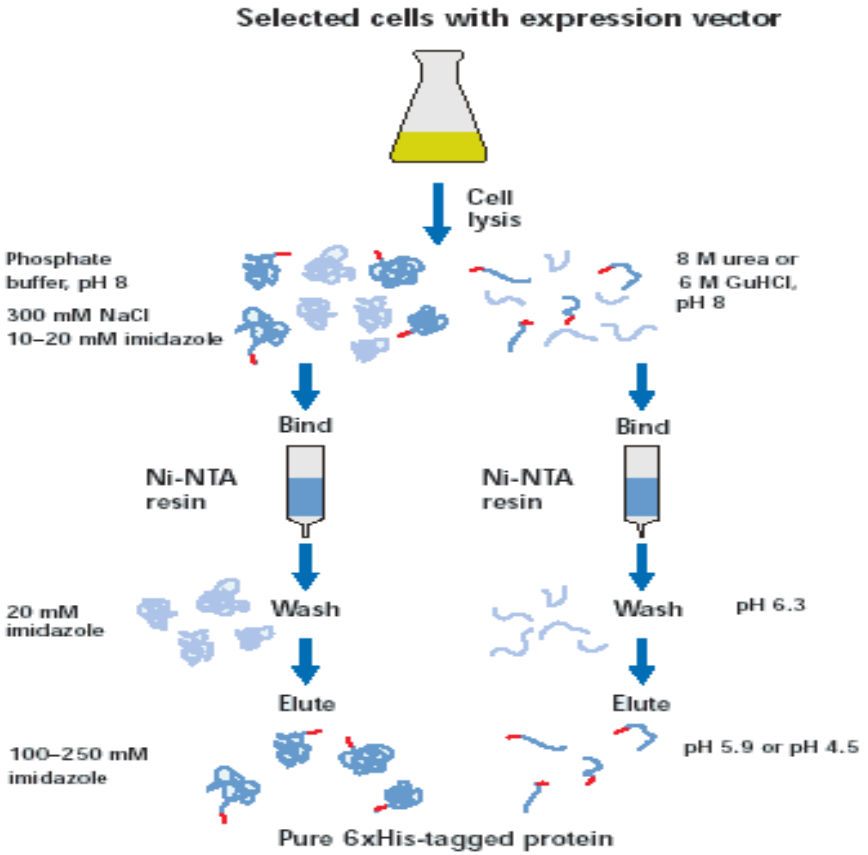
His-tagged protein bound to resin is washed and eluted by an increasing concentration of imidazole (0 – 200 mM)



Histidine



Imidazole



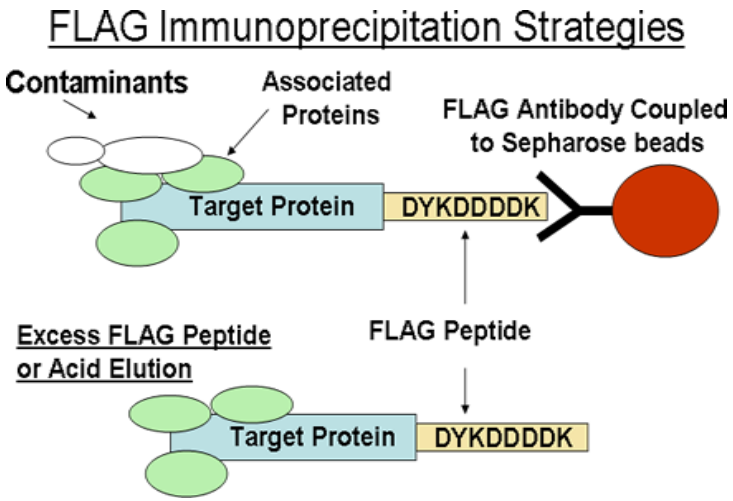
Immunoaffinity Chromatography – FLAG-tag

The FLAG-tag was the first example of a fully functional epitope tag published (and patented - 1988)

Epitope tag - short peptide sequence easily recognized by *tag*-specific antibodies

FLAG-tag is: DYKDDDDK (~1 KDa), 3xFLAG provides better results than 1xFLAG

Tagged protein purified using anti-FLAG resin, eluted by an excess of 3xFLAG peptide or low pH



MBP-tag - Maltose-binding protein tag, increases solubility

c-myc - a short segment of the human protooncogene myc (EQKLISEEDL)
HA - segment from human influenza hemagglutinin protein (YPYDVPDYA)

S-tag is derived from RNase A (KETAAAKFERQHMDS)

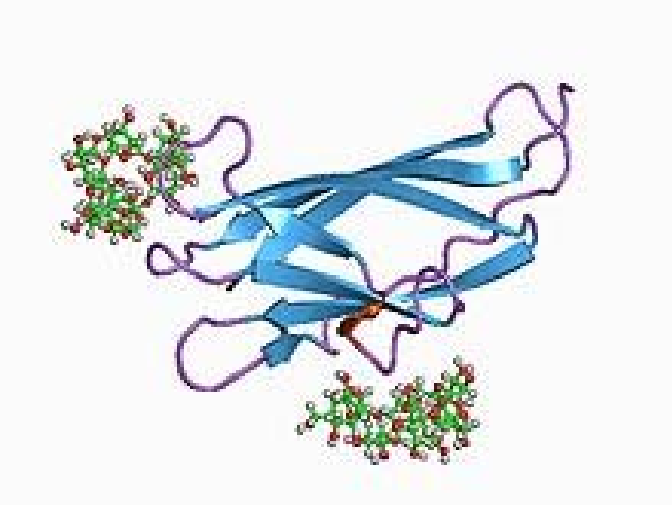
T7-Tag MASMTGGQQMG-K

Tag	Residues	Sequence	Size (kDa)
Poly-Arg	5–6 (usually 5)	RRRRR	0.80
Poly-His	2–10 (usually 6)	HHHHHH	0.84
FLAG	8	DYKDDDDK	1.01
Strep-tag II	8	WSHPQFEK	1.06
c-myc	11	EQKLISEEDL	1.20
S-	15	KETAAAKFERQHMDS	1.75
HAT-	19	KDHLIHNVHKEFHAHAHNK	2.31
3x FLAG	22	DYKDHDGDYKDHDIDYKDDDDK	2.73
Calmodulin-binding peptide	26	KRRWKKNFIAVSAANRFKKISSSGAL	2.96
Cellulose-binding domains	27–189	Domains	3.00–20.00
SBP	38	MDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP	4.03
Chitin-binding domain	51	TNPGVSAWQVNTAYTAGQLVTYNGKTYKCLQPHTSLAGWEPSNVPALWQLQ	5.59
Glutathione S-transferase	211	Protein	26.00
Maltose-binding protein	396	Protein	40.00

Cellulose-binding domain

CBD tag, 20 KDa

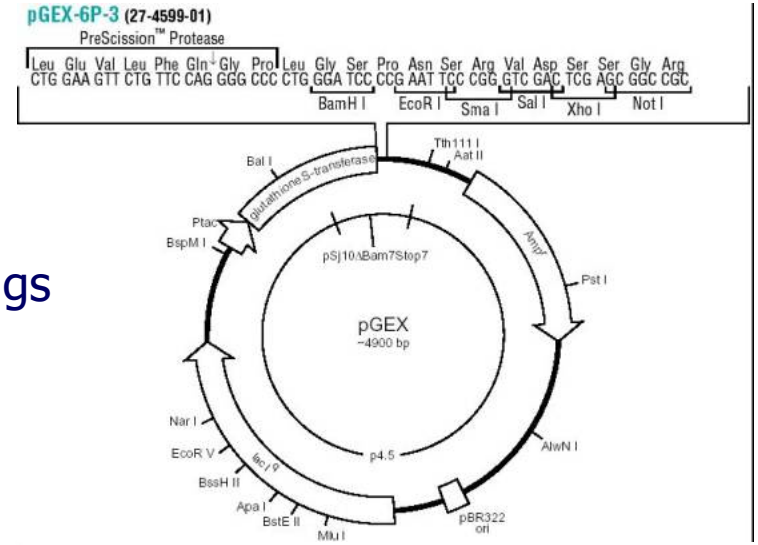
Affinity purification on a cellulose column, looks to be simple, effective and easily scalable. Eluted by glucose (up to 1M).



Cellulose resin (CBinD) provided by number of companies

Protease cleavage sites

Included in vectors to allow removal of tags



Enzyme	Recognition Site	Comments
Thrombin	LVPR/GS	Active, less specific
Factor Xa	IEGR/	Less specific
Enterokinase	DDDDK/	Very specific
PreScission™	LEVLFQ/GP	Very specific
TEV	ENLYFQ/G(S)	Very specific
TVMV	ETVRFQ/G(S)	Very specific
Ubq/SUMO Pro.	Ubq/SUMO fusion	

Protein tagging – commercial vectors / systems

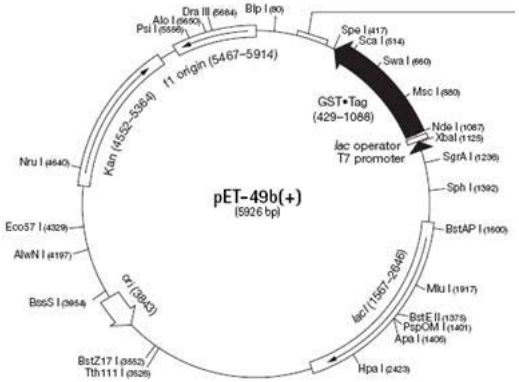
Number of commercial systems available providing all components for tagging (expression vectors), purification (resins, eluents), detection (antibodies).

Limited for model expression systems

Examples for *Escherichia coli*:

pET plasmids (Novagen) – pET9 (6xHis), pET17 (T7-tag), **pET49** (GST+His tag), pET50 (NusA) Strep II system (IBA), plasmid, tactin resin, desthibiothin for elution, anti-Strep antibody 3xFLAG system (Sigma), plasmids, anti-Flag resin, antibody

Various GST-Bind Purification Kits, His-Bind Purification Kits, Magnetic Streptavidin Bead, S-Tag, T7-Tag kits



Purification of chlorophyll synthase from the cyanobacterium *Synechocystis* 6803

4 L of cells

→ Beadbeating → Isolation of membranes

↙

Solubilization by 1% dodecyl-*b*-maltoside

↙

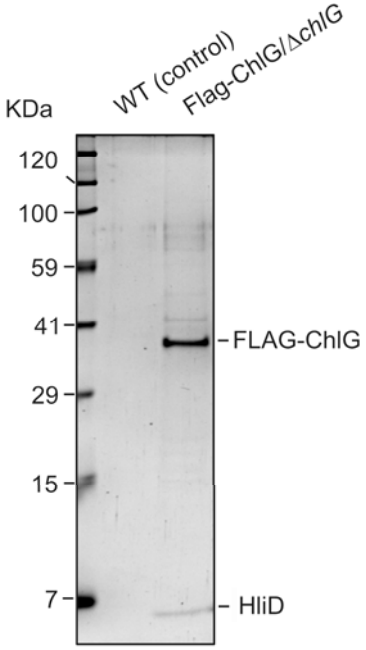
binding on the ~ 400 μL of anti-FLAG resin, intensive washing

↙

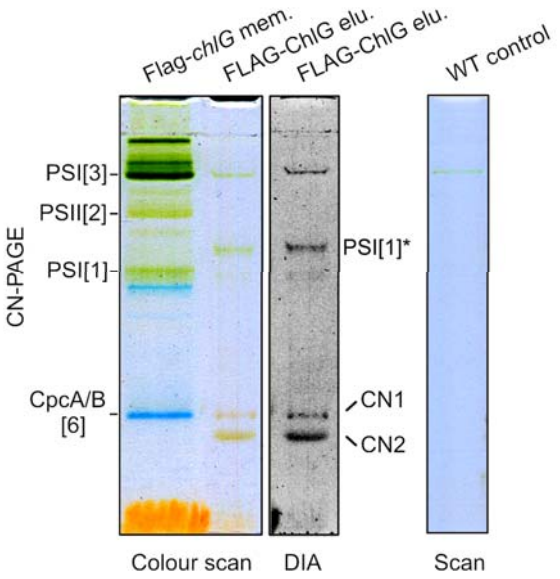
Elution by 100 μg of 3xFLAG peptide

↘

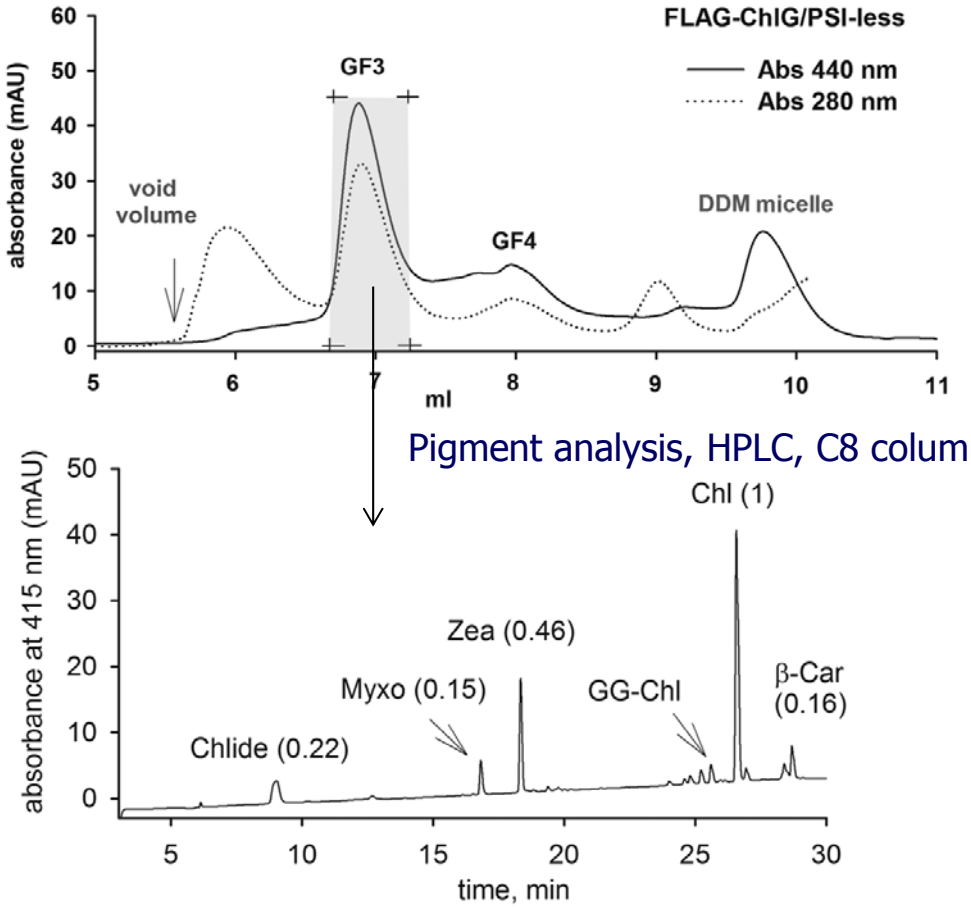
~500 μL of active pure enzymatic complex



Purification of chlorophyll synthase from the cyanobacterium *Synechocystis* 6803

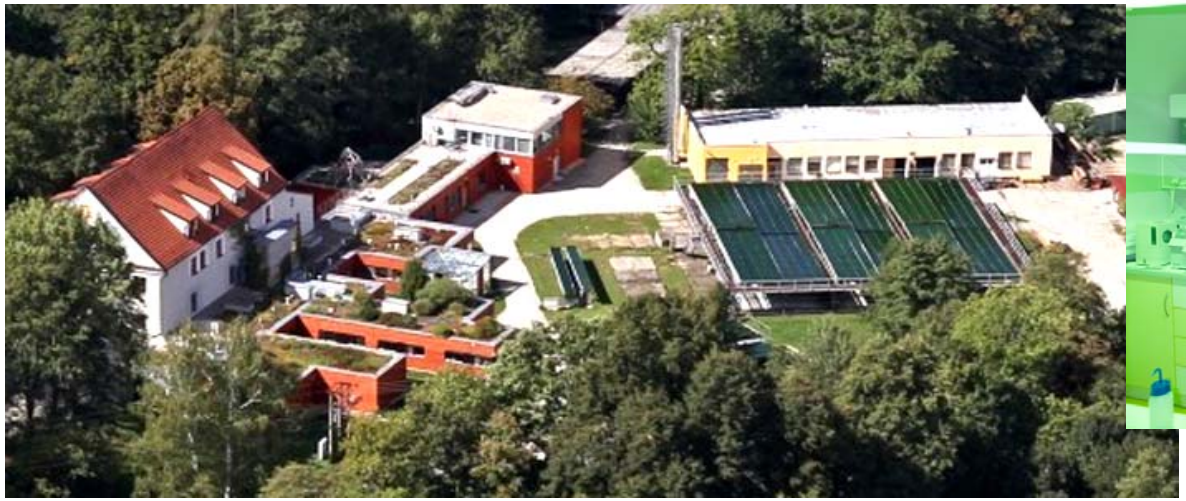


Clear-Native gel, the eluate contains pigmented complexes



Gel-filtration, HPLC Bio-Sec 3000 column, Phenomenex. Peak GF3 collected, pigments extracted and separated by C8 column – **20 mM MES buffer pH 6.5, 0.1M NaCl, 0.06% DDM, Flow 0.2 mL/min, 10°C**

Laboratory exercise, Jan 7-8, 8:30, Třeboň

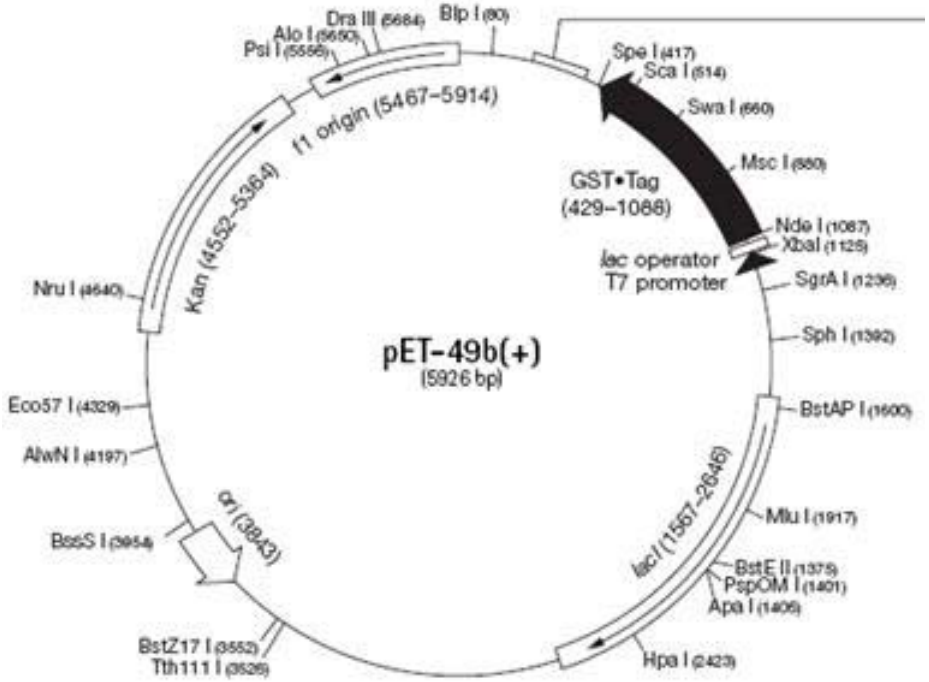


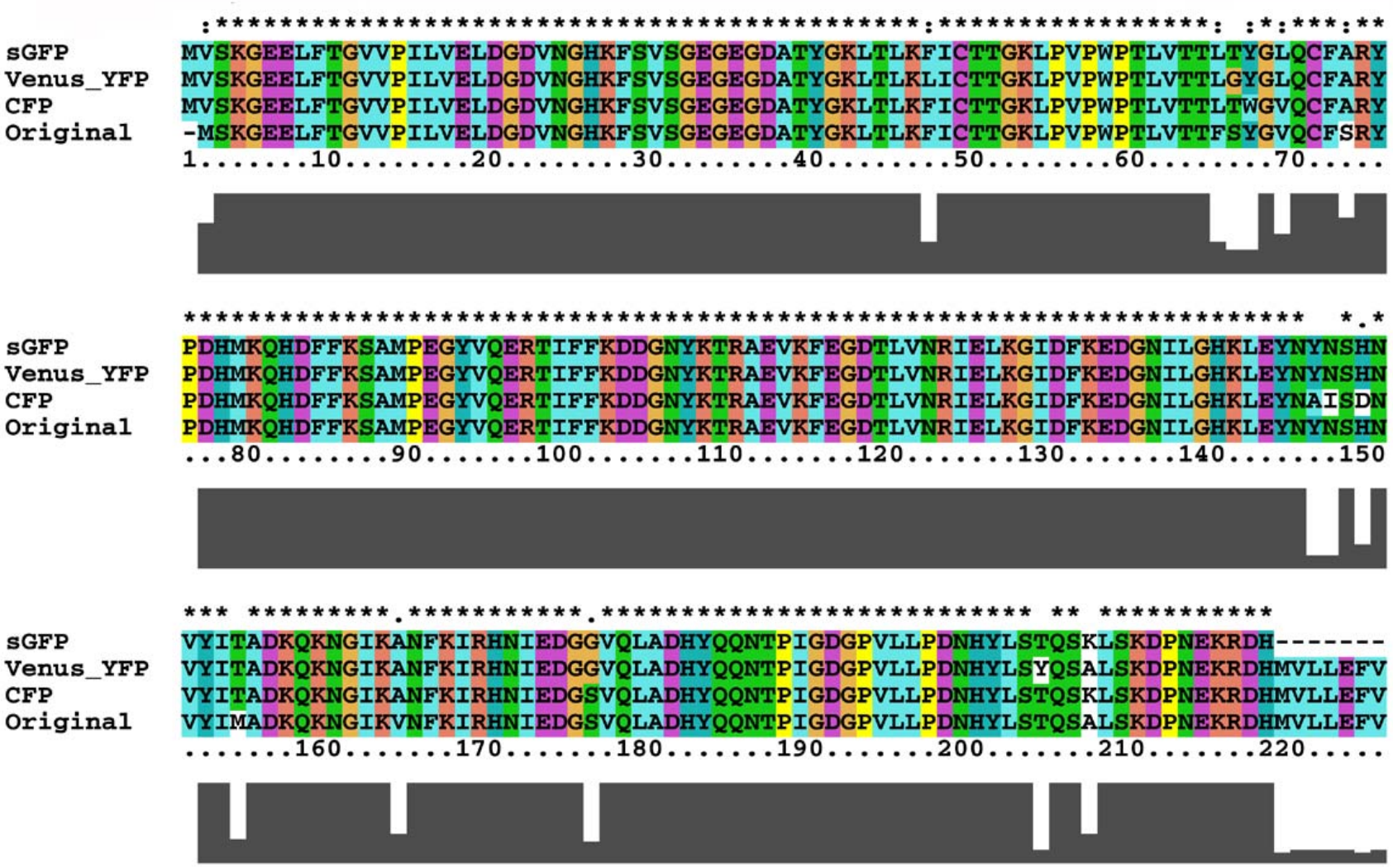
Day 1:

Cloning of genes coding for recombinant fluorescent proteins

Electroporation into BL21 (DE3) *E. coli* cells

sGFP, YFP, RFP and TFP cloned into pET9 vector, expression inducible by IPTG, 6xHis-tag, Kanamycin resistance




sGFP
Venus_YFP
CFP
Original
MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGLQCFARY
MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKLICTTGKLPVPWPTLVTTLGYGLQCFARY
MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTWGVQCFARY
-MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFSRY
1.....10.....20.....30.....40.....50.....60.....70.....
PDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHN
PDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHN
PDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNAISDN
PDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHN
...80.....90.....100.....110.....120.....130.....140.....150
VYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDH-------
VYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSYQSALSKDPNEKRDHMVLLEFV
VYITADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFV
VYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFV
.....160.....170.....180.....190.....200.....210.....220.....

Day 2:

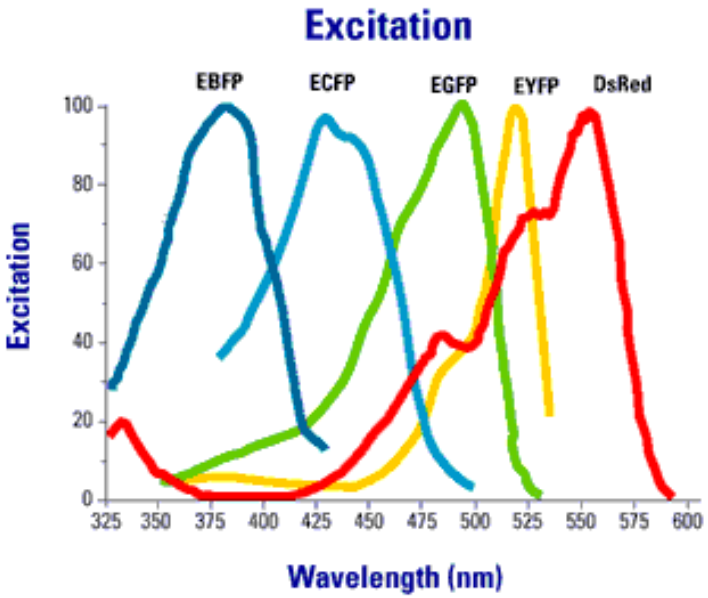
Purification of His-tagged recombinant fluorescent proteins

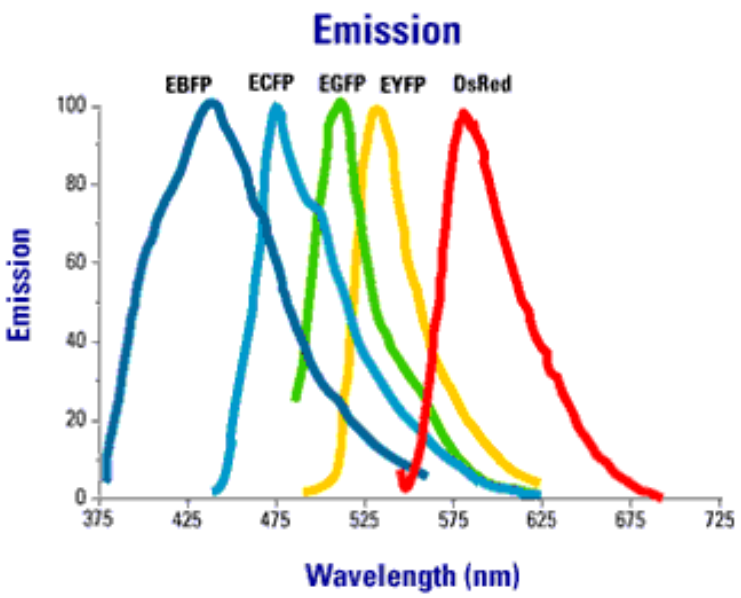
- Preparation of a nickel columns
- Purification, elution using imidazole

SDS-PAGE Preparation of samples

Day 2: Characterization of purified FPs

His-sGFP (super-folding GFP)
His-YFP (Venus)
His-RFP
TFP - blue





Day 2: Characterization of purified FPs

YFP (Venus)- Coreluan FP/TFP, a favored pair for FRET (Förster resonance energy transfer)

Are proteins Z and Y interacting *in vivo*?