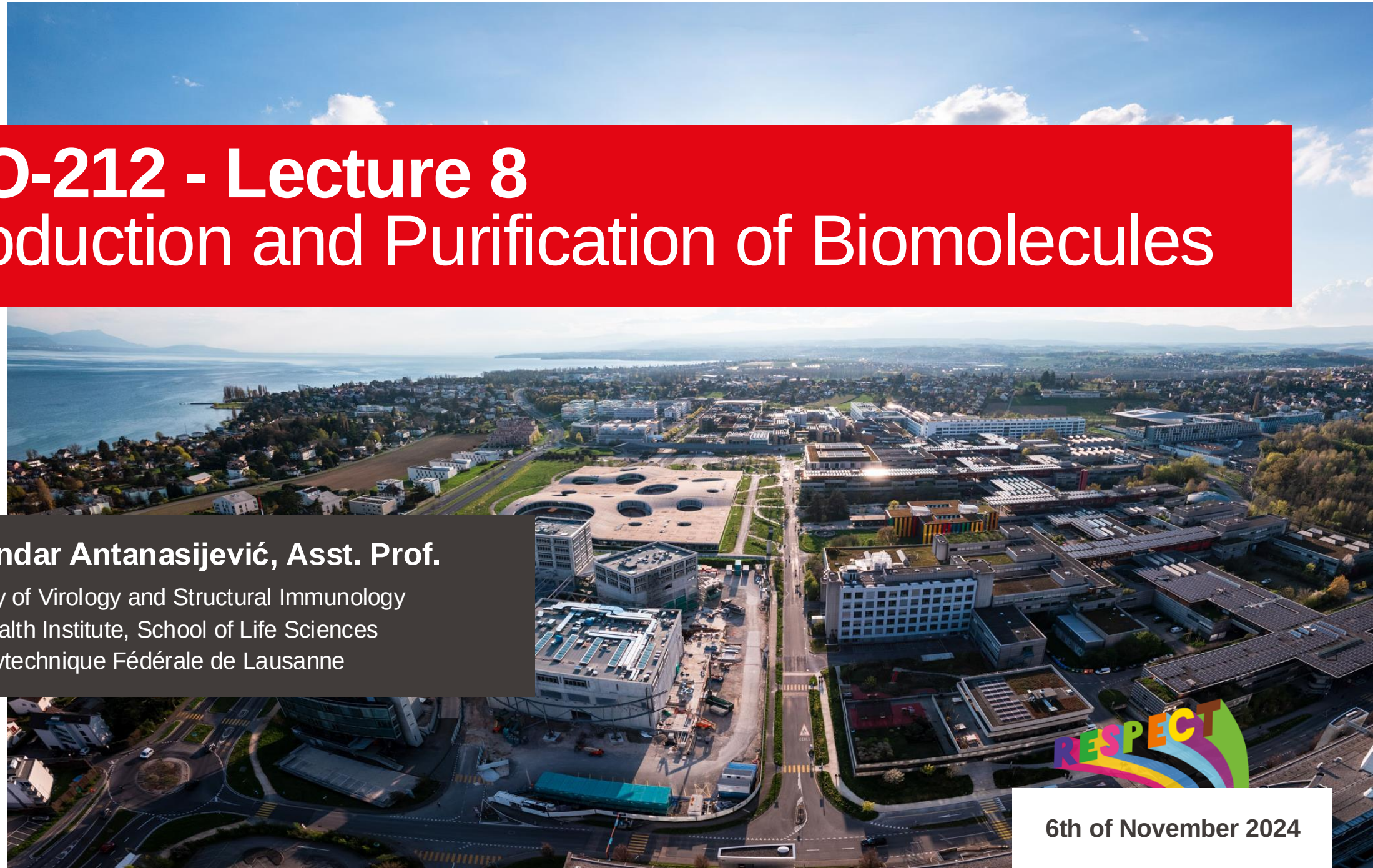
BIO-212 - Lecture 8
Production and Purification of Biomolecules



Aleksandar Antanasijević, Asst. Prof.

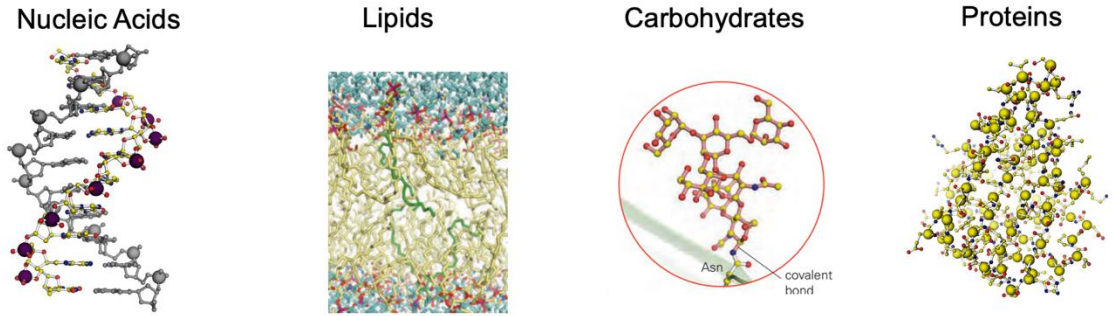
Laboratory of Virology and Structural Immunology
Global Health Institute, School of Life Sciences
École Polytechnique Fédérale de Lausanne

RESPECT

6th of November 2024

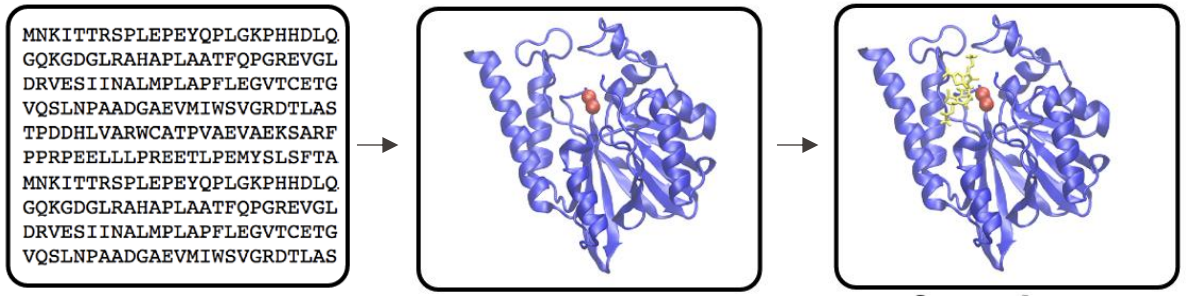
Quick Summary of the course so far

• Primary macromolecular components of cells



- Polymers composed of smaller building blocks

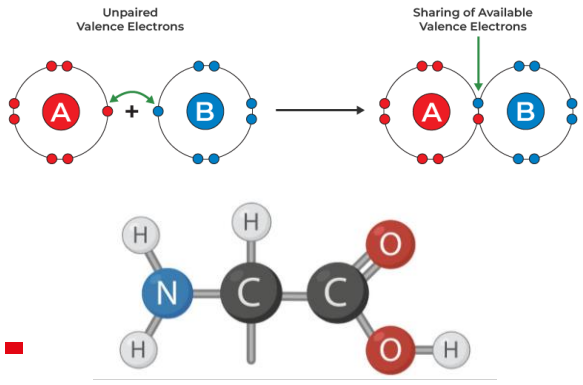
• Sequence -> Structure -> Function



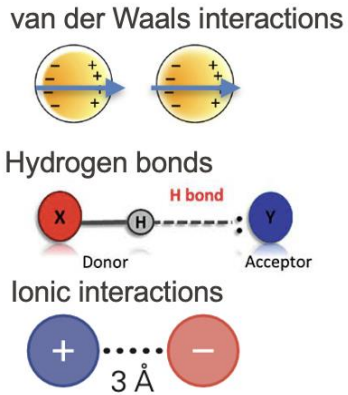
- 3D arrangment of atoms driven by thermodynamics

• Covalent and non-covalent interactions

Covalent

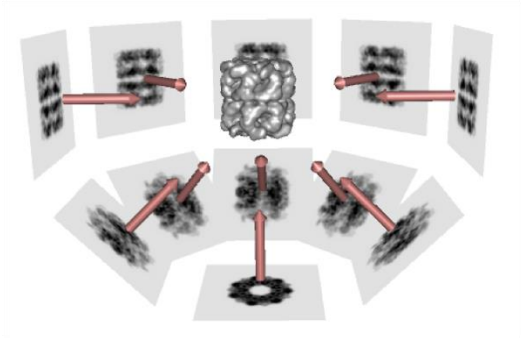


Non-covalent



• Structure determination and prediction

Experimental methods



- cryoEM, NMR, X-ray

Computational prediction

- AlphaFold2

Structural biology research on EPFL campus

• State-of-the-art facilities for cryoEM, NMR and X-ray

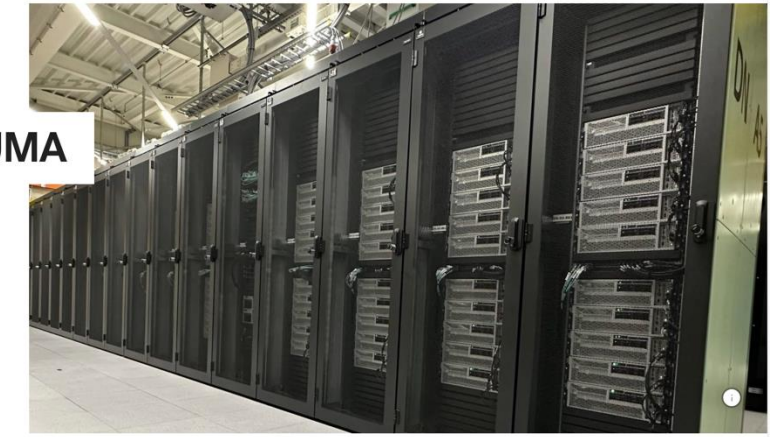
- Dubochet Center for Imaging
- Interdisciplinary Center for EM
- PTPSP Core Facility
- NMR Facility (ISIC)

• High-performance computing infrastructure

- SCITAS

KUMA

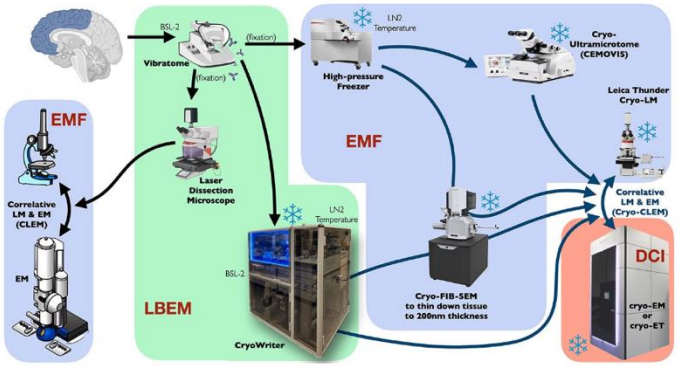
- 104 GPU nodes
- >480 GPUs
- 384 GB of DDR5 RAM
- Bandwidth: 2.4 TB/s



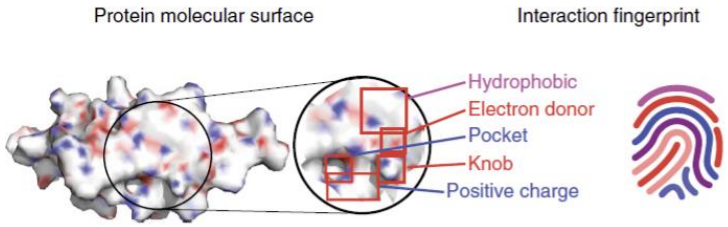
Available since September, 2024

• >10 research labs focusing on structural biology, protein structure prediction and protein engineering

Henning Stahlberg
Ulrich Lorenz
Andrea Ablasser
Nicolas Thoma
Suliana Manley
Pierre Gonczy



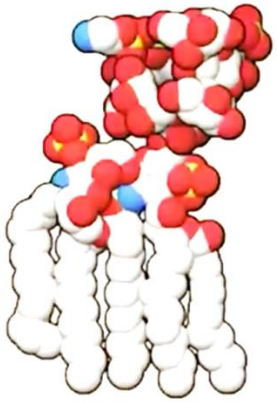
Bruno Correia
Patrick Barth
Aleksandra Radenovic
Matteo Dal Peraro
Aleksandar Antanasijevic



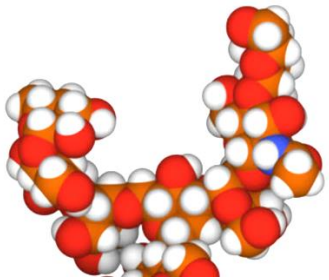
Production and purification of biomolecules

- Why is it important to generate purified biomolecule samples?

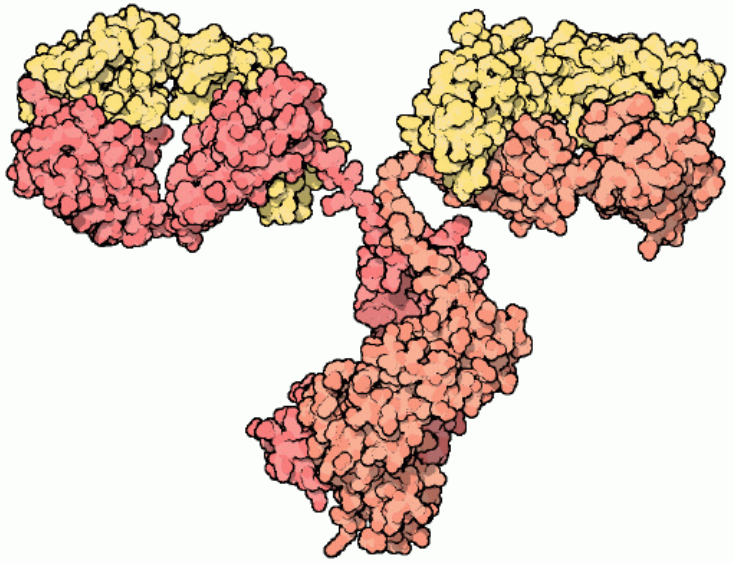
- To study their function
- To study how they fold (kinetics)
- To identify which other biomolecules they interact with
- To determine their sequence/composition
- To determine their structure (e.g., by X-ray crystallography)
- To apply for drug screening
- For therapeutic applications



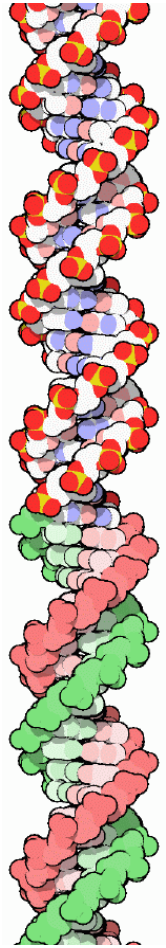
Lipid



Carbohydrate



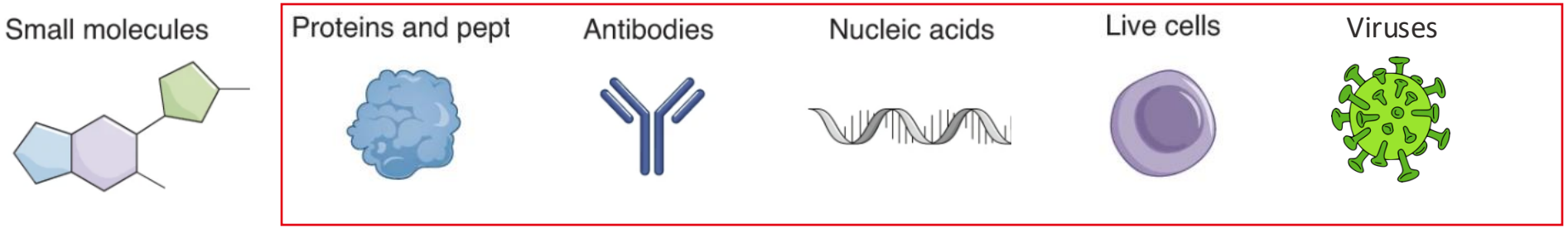
Protein



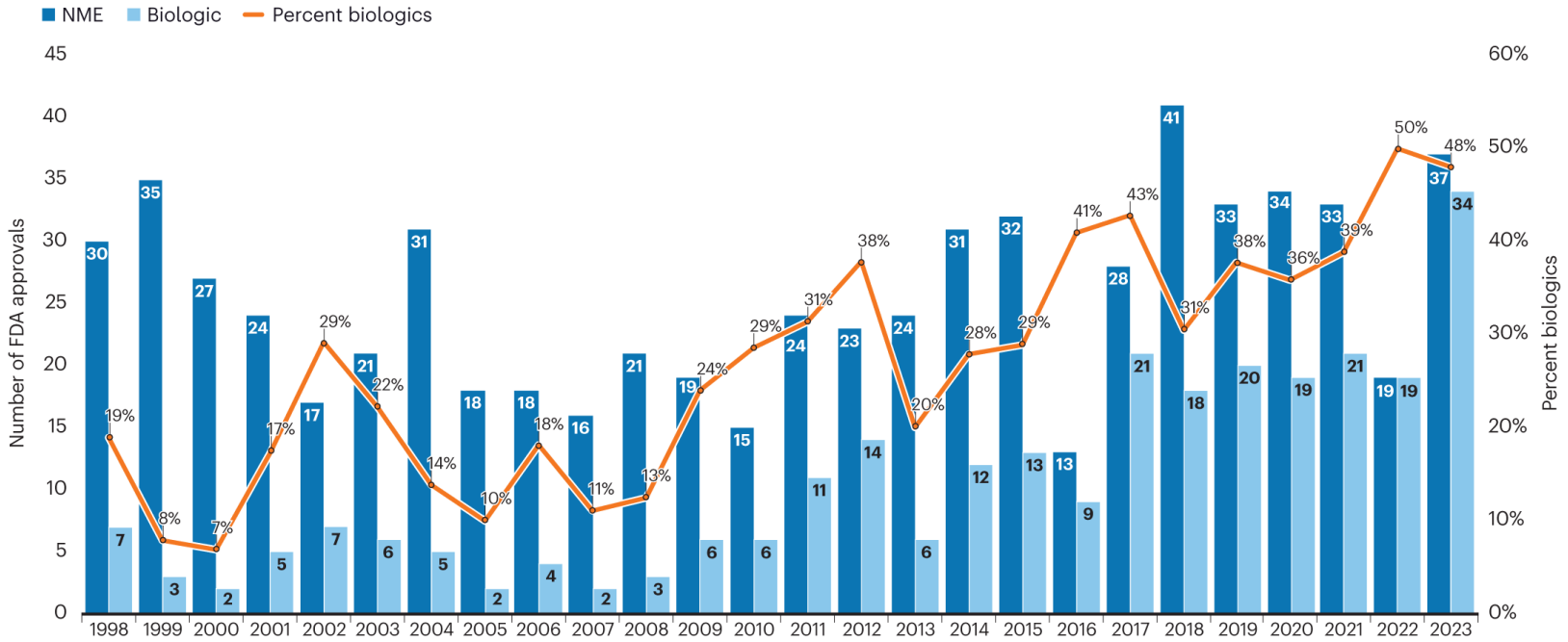
Nucleic Acids

Biomolecules as therapeutics and vaccines

- Newly approved drugs by the US Food and Drug Administration (FDA)



Biologics



Source: Senior et al., Nature Biotechnology, 2024

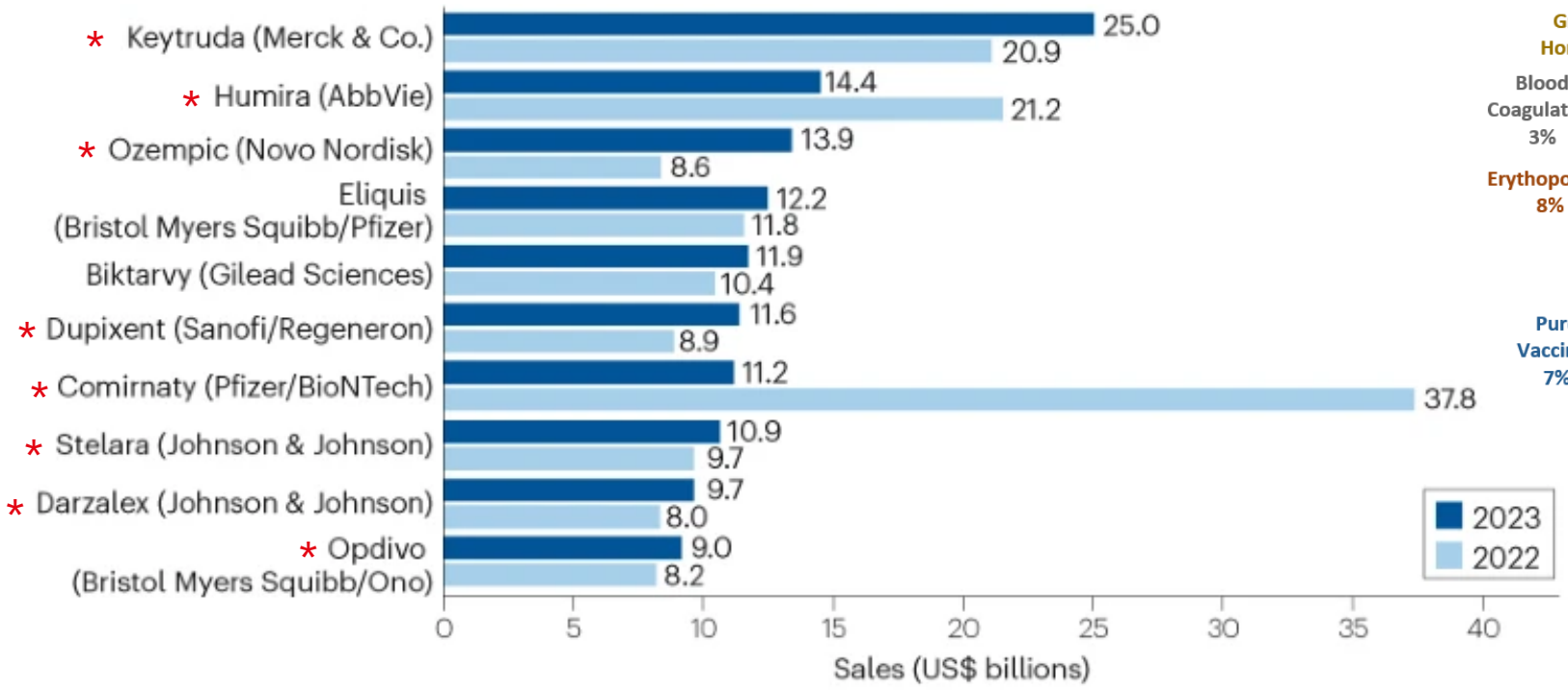
- 2023 was a record year in terms of the # of new approved drugs
- 48% of newly approved drugs in 2023 are biologics
- All vaccines are essentially of biomolecular origin

Biologics are important drugs

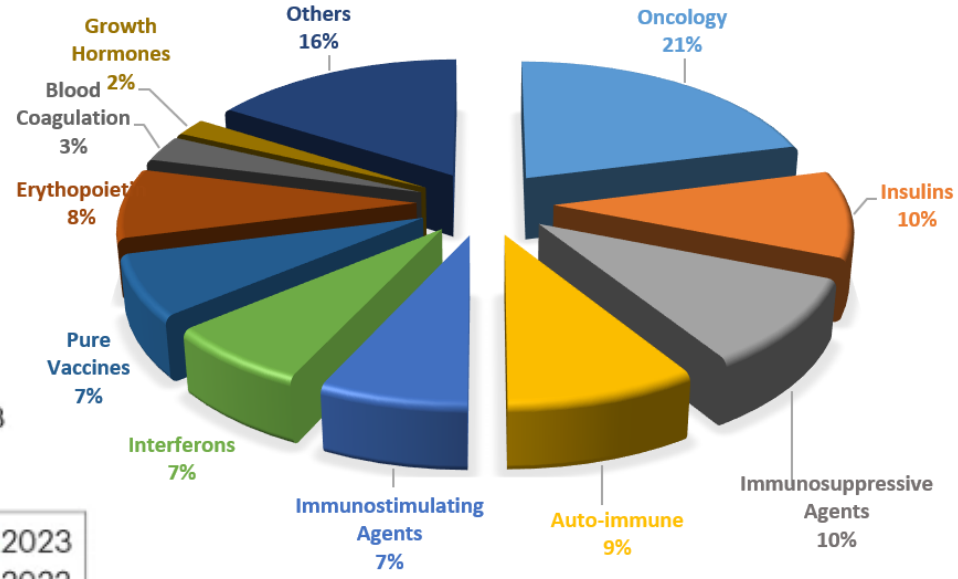
- Proteins and mRNA are a major class of biologics

* Biologics

Highest-selling drugs in 2022/2023



Applications of protein therapeutics



- Most protein therapeutics are based on monoclonal antibodies (more info during BIO-213)
- Advances in protein structure prediction and design are expected to lead to more biologics

Biomolecule Production

↓

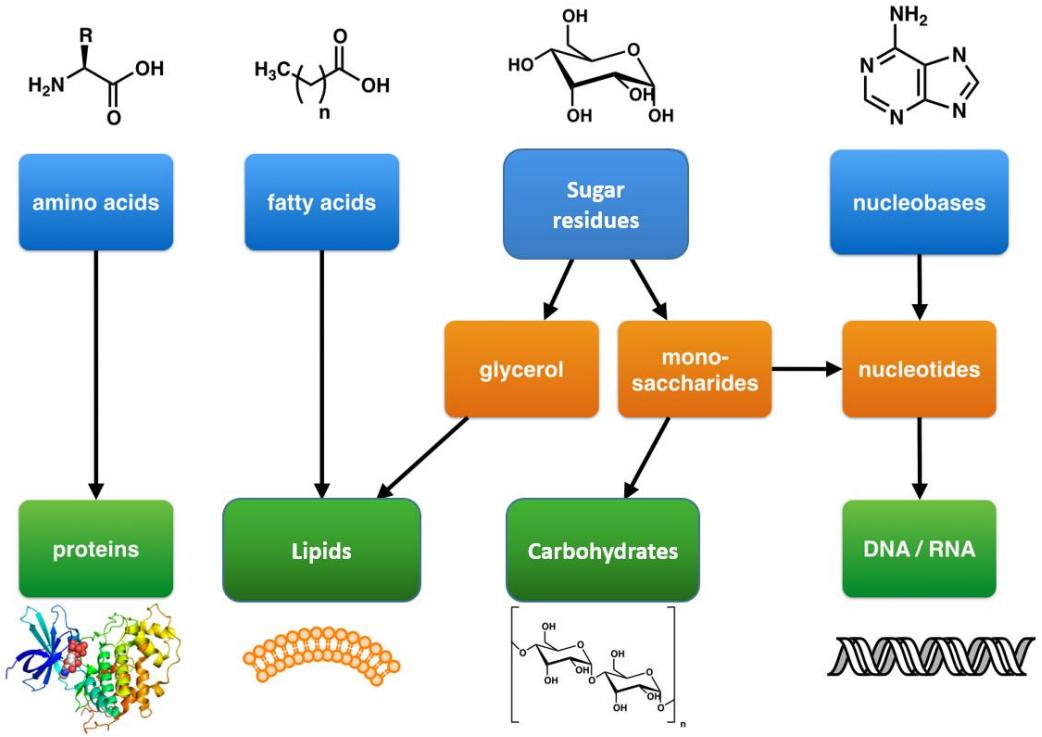
Biomolecule Purification

↓

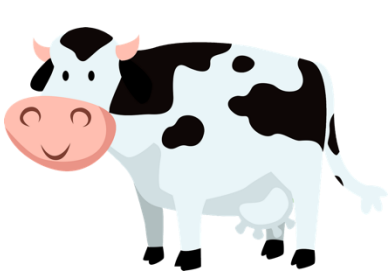
Measuring Quantity and Purity

Biomolecules are produced from building blocks

• **Biological macromolecules are produced from the corresponding building blocks** such as amino-acids, nucleotides, monosaccharides, fatty acids, which assemble into long linear or branched chains.



• First approaches for obtaining high purity biomolecules were based on **purification from biological samples** (e.g., animal tissues) enriched in the protein of interest.
• To date there are many biologic therapeutics that are still generated using these approaches (e.g., estrogen, heparin, antivenoms).



• Today's technology allows to generate target molecules without the need for natural source. There are 3 primary approaches to produce macromolecules:

- **Chemical synthesis**
- **Enzymatic synthesis**
- **Cell-based synthesis**

Chemical synthesis approaches

• These approaches rely on chemical reactions to fuse building blocks in the correct order (e.g., amino-acid sequence) and using the correct chemical linkage (e.g., peptide bond)

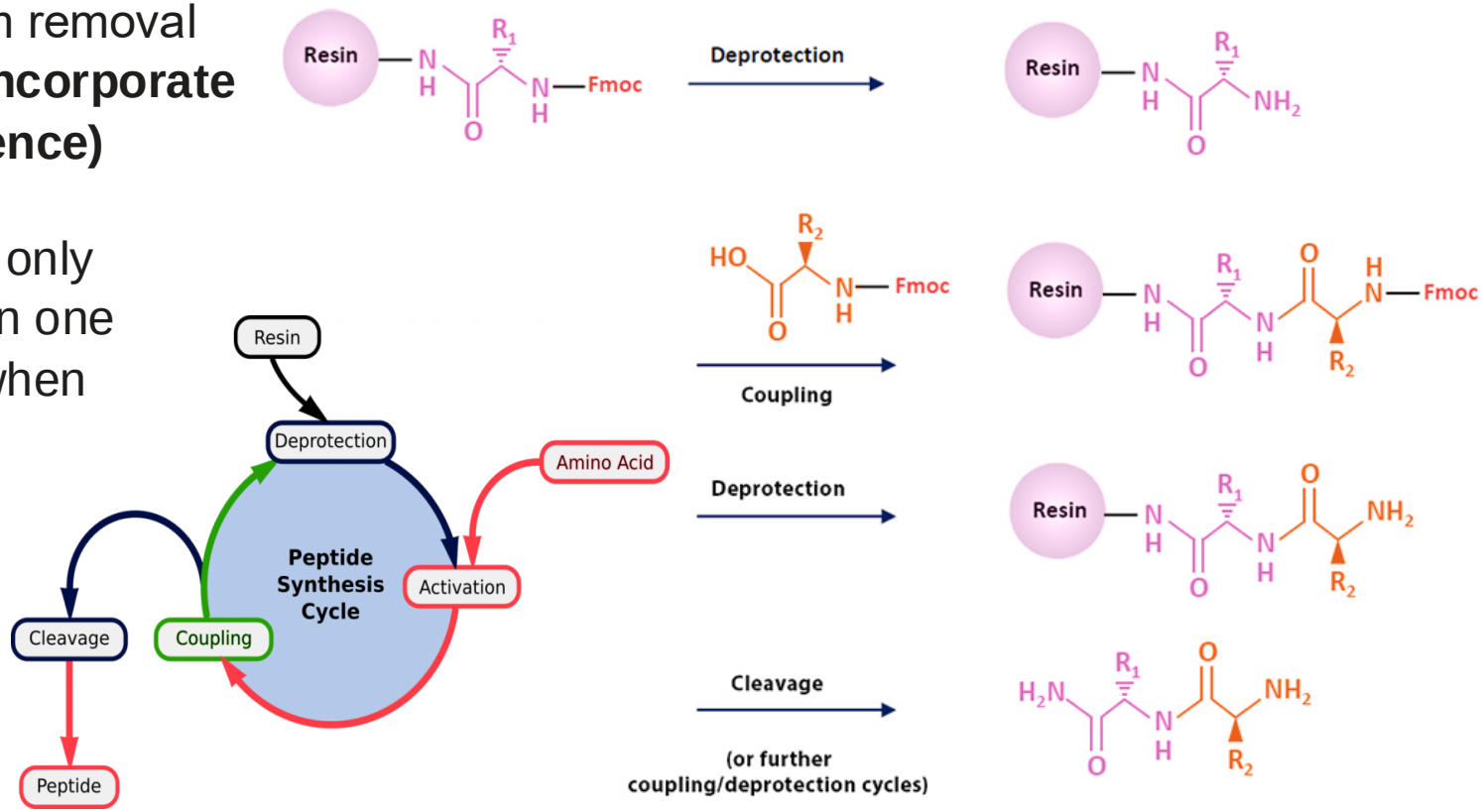
• Most current approaches utilize **solid-phase synthesis** where one terminus of the biomolecule (e.g., C-terminal amino acid) is conjugated to solid support (e.g., resin)

• This allows to use centrifugation to perform removal and replacement of chemicals, allowing to **incorporate building blocks in a desired order (sequence)**

• Protective groups (e.g., Fmoc) assure that only a single amino-acid is coupled to the chain in one step. The group is removed (deprotection) when the next building block is ready to be added.

• After adding the last building block of the desired sequence, the product is cleaved from the resin and ready for use.

<u>Example: Peptide synthesis from amino-acids</u>



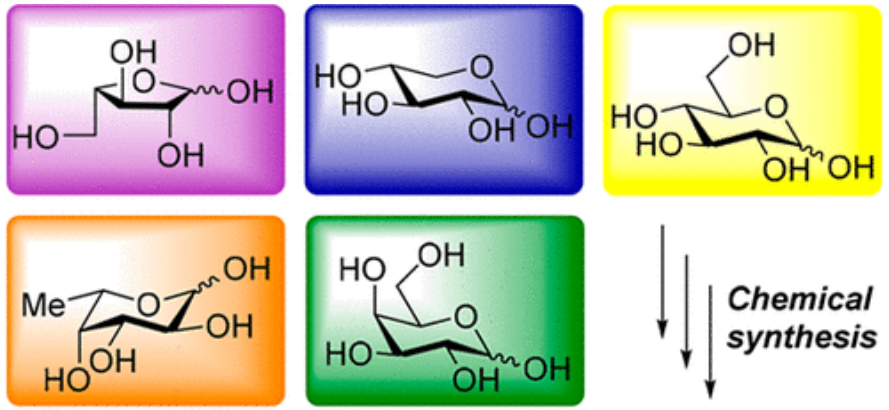
Chemical synthesis approaches

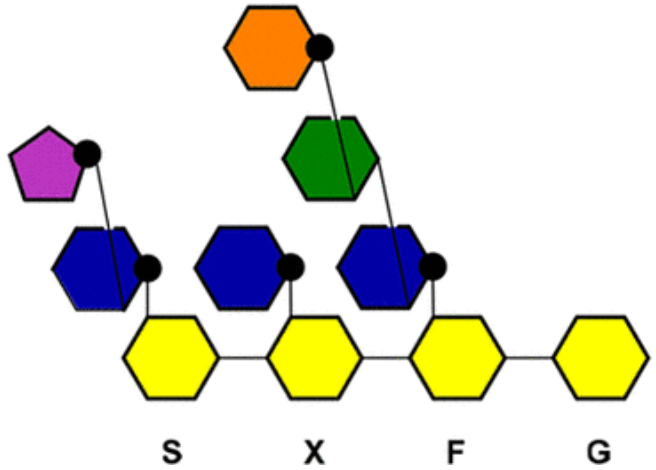
- Each biomolecule family has an equivalent set of chemical procedures and reagents that allow to perform synthesis of desired products.
- There are commercial kits that can be purchased, and many companies provide this service

- Advantages:
  - Simple and well-controlled (good in industry setting)
  - High-yield chemical reactions
  - Allows to produce molecules that are toxic to cells
  - Great for oligonucleotides and short peptides

- Disadvantages:
  - Requires specialized reagents (e.g., protected amino-acids)
  - Chemistry is not always simple (e.g., carbohydrates)
  - **Limited to shorter sequences (e.g., 20-100 residues)**
  - More reaction byproducts with more steps
  - Difficult to produce properly folded protein domains

Carbohydrates have more complex linkages





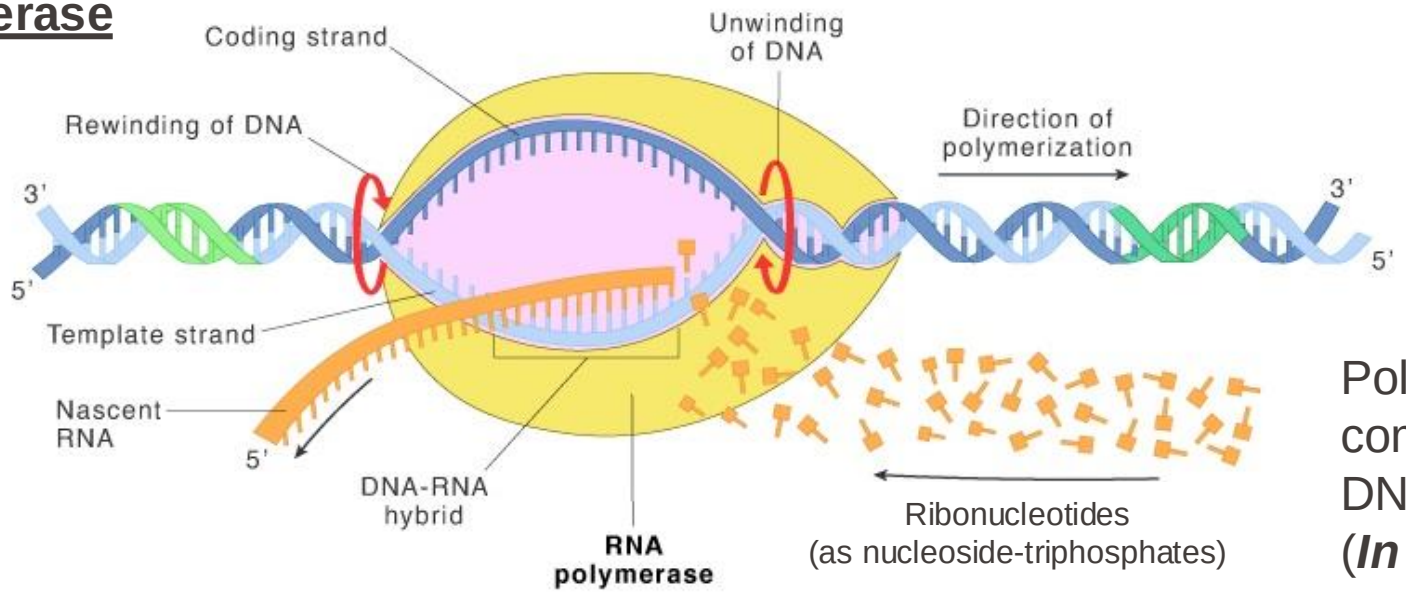
Enzymatic approaches for biomolecule synthesis

• These approaches are based on ***in vitro*** **application of enzymes** that naturally catalyze the formation of the corresponding chemical linkages, such as:

- DNA polymerase -> Phosphodiester bonds between deoxyribonucleotides
- RNA polymerase -> Phosphodiester bonds between ribonucleotides
- Ribosome -> Peptide bonds
- Carbohydrate synthases -> Glycosidic bonds

• The reactions require the enzyme, corresponding building blocks, buffering components, and in some cases (e.g., protein and nucleic acids) a template nucleic acid sequence to be copied or translated.

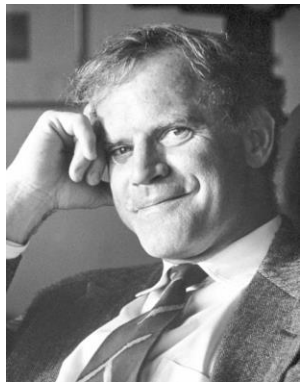
RNA polymerase



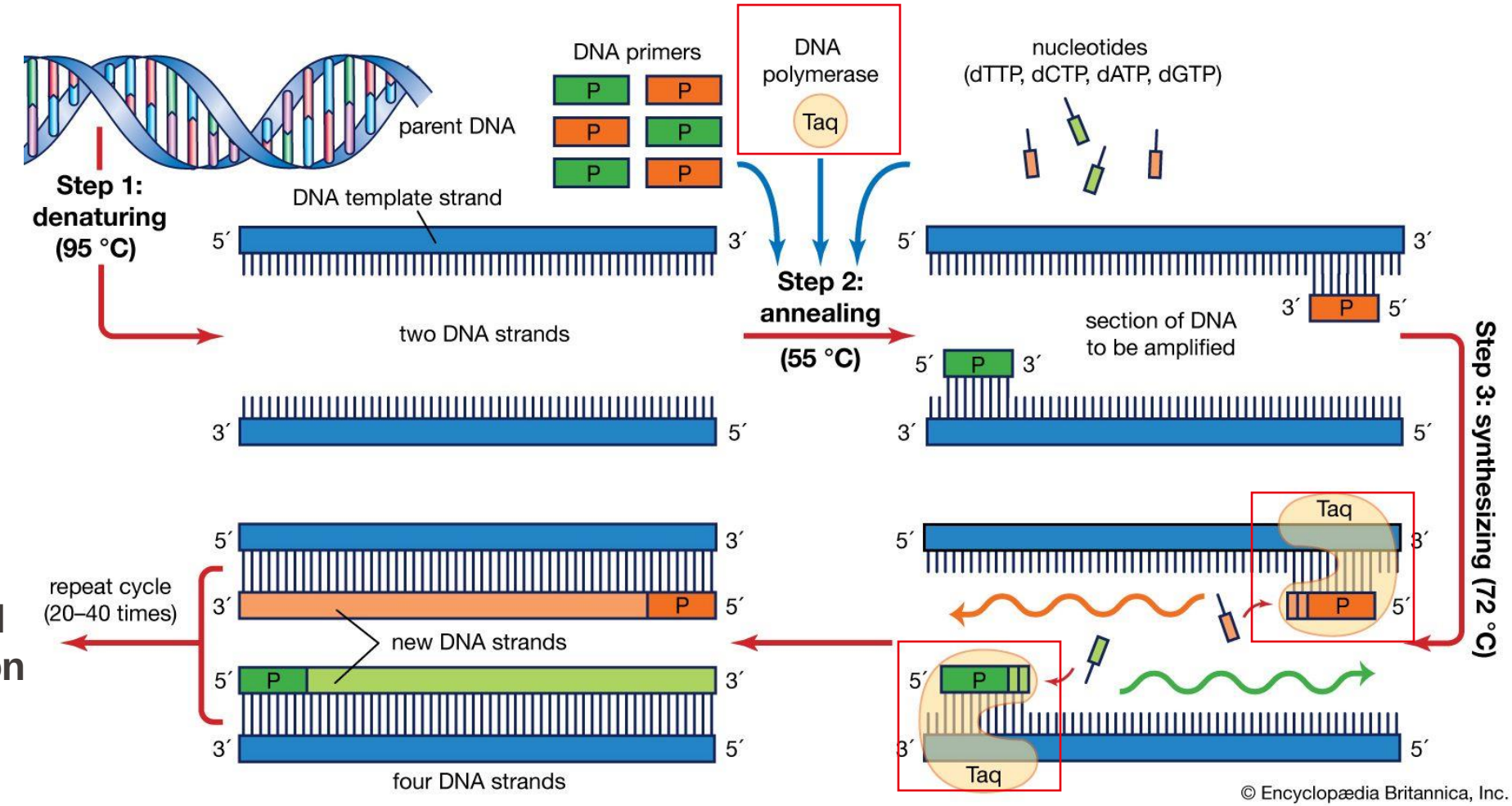
Polymerase builds RNA strand complementary to the tempate DNA strand using nucleotides (***In vitro*** **transcription**)

Polymerase chain reaction (PCR)

- Enzymatic system for DNA production based on template (parent) DNA
- Watson-Crick pairing is used to generate copies of complementary DNA strands



Kary Banks Mullis
(1944-2019)

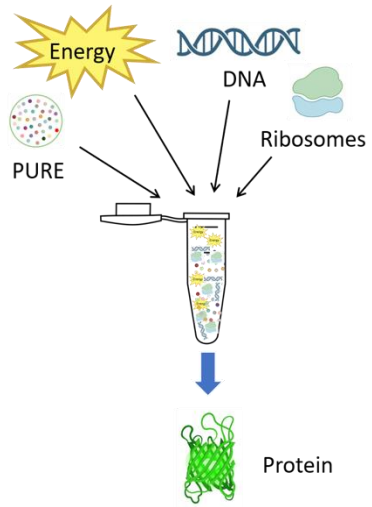


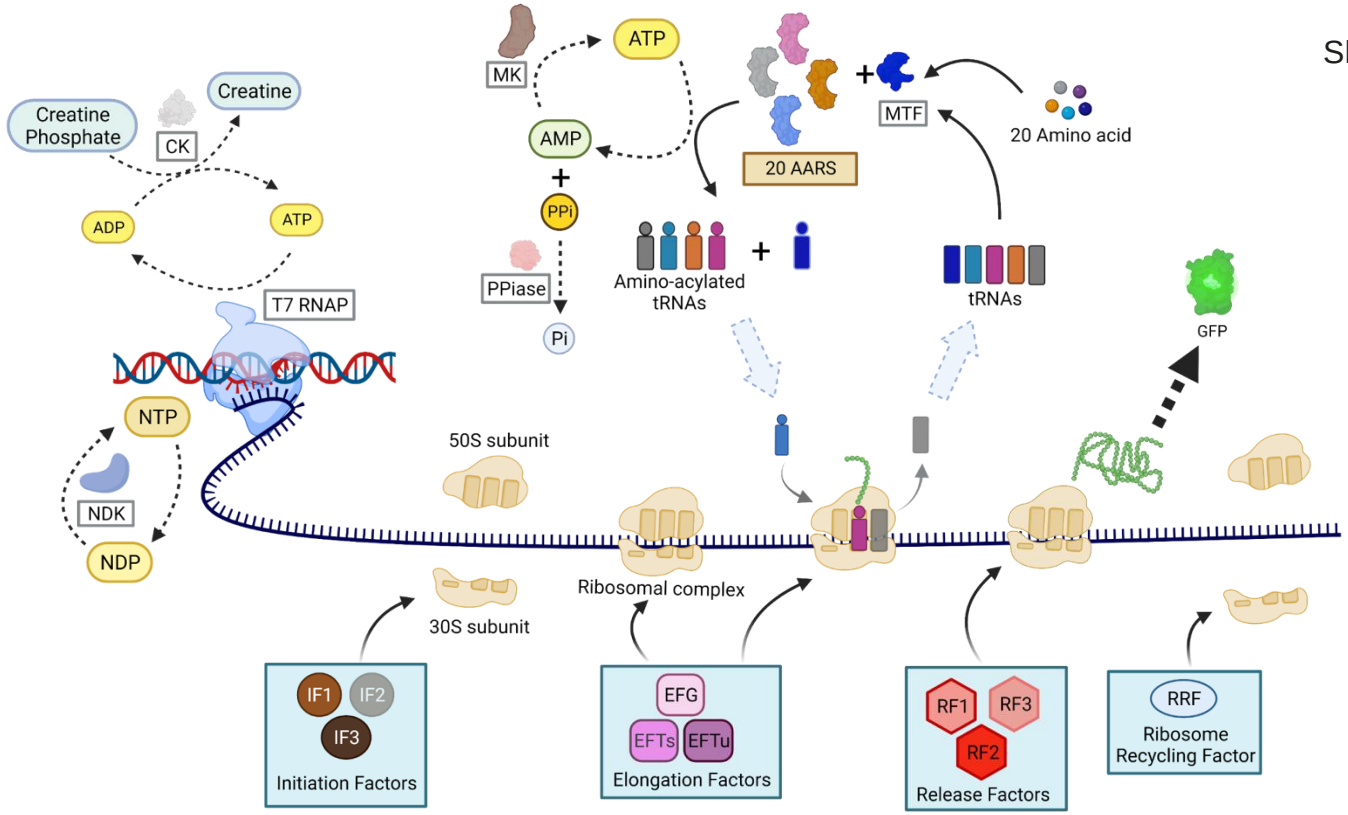
Exponential amplification

~10^9 copies

PURE Technology for Protein Synthesis

- PURE = Protein synthesis Using Recombinant Elements
- The mix comprises all the building blocks and enzymes needed for **transcription and translation**





Shimizu Y et al. 2001

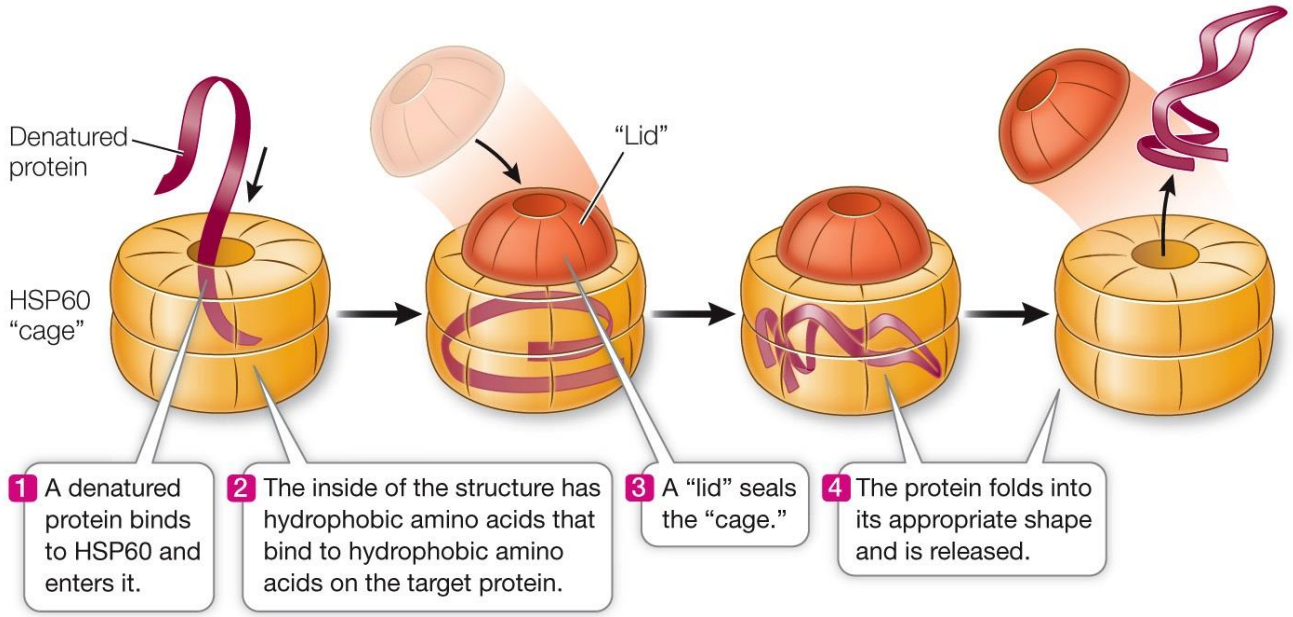
DNA -> RNA -> Protein

- Very useful for production of **toxic proteins and ^{13}C and ^{15}N isotope labeling for NMR**

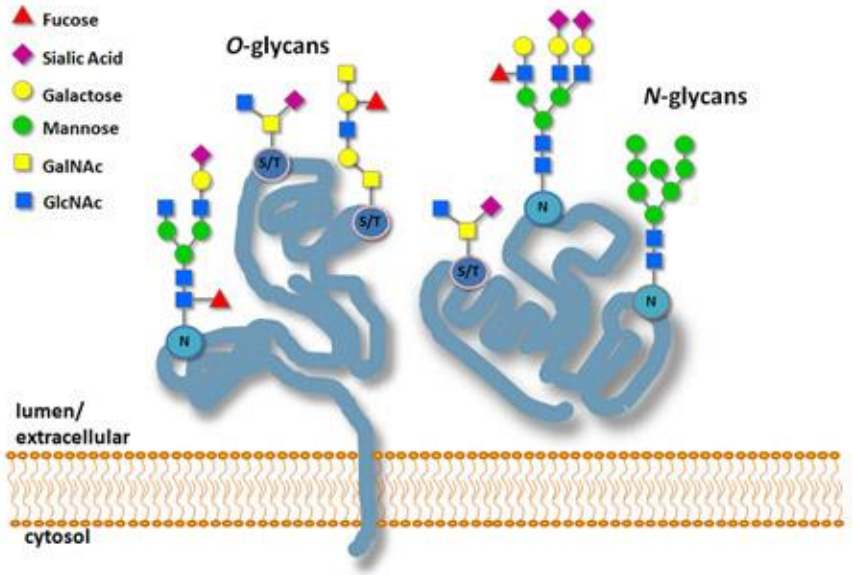
Pros and Cons of enzymatic biosynthesis

• Enzymatic biosynthesis approaches are versatile (i.e., many enzymes exist and can be applied towards different purpose) and significantly faster compared to chemical methods (e.g., DNA polymerase copies up to 700 base pairs per second)

• However, application to protein synthesis is relatively limited due to the lack of cellular proteins that assist with protein folding (**chaperones**) and post-translational modifications (**e.g., glycosylation**)



Chaperones help newly synthesized proteins to fold into their functional states



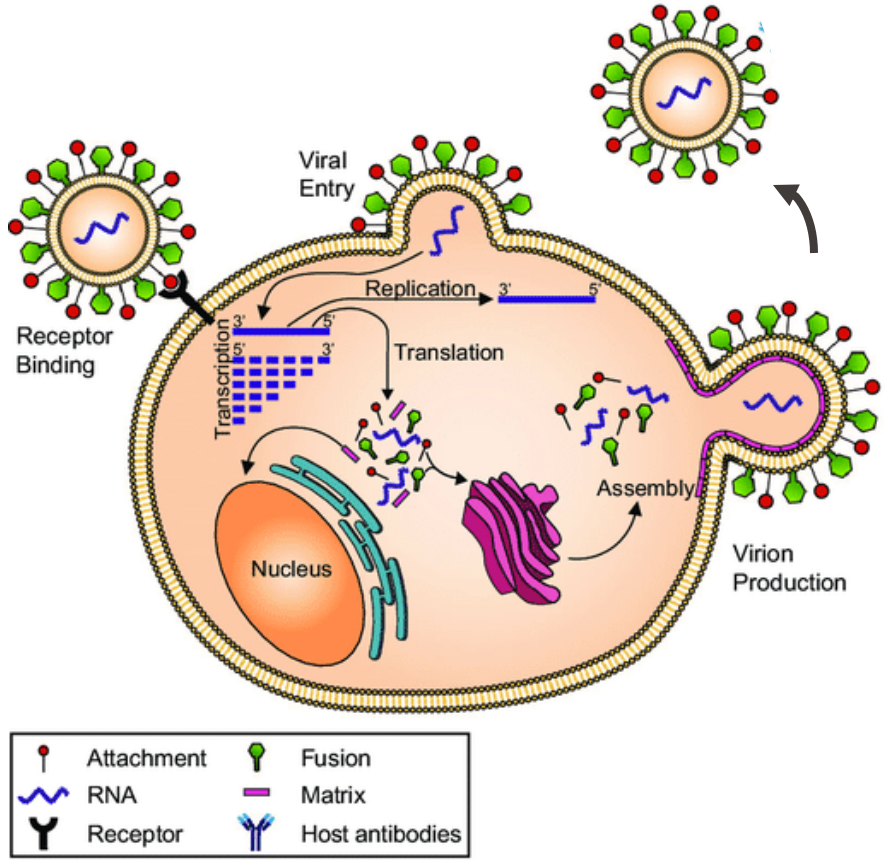
Glycosylation is attachment of oligosaccharides onto protein

Cell-based approaches for biomolecule synthesis

• Biomolecule synthesis is performed using living cells, typically transformed/engineered to produce **high quantity** of the biomolecule of interest.

• The biggest benefit comes from using (hijacking) existing cell machinery for synthesis, folding, PTMs and other modifications necessary to produce functional biomolecule.

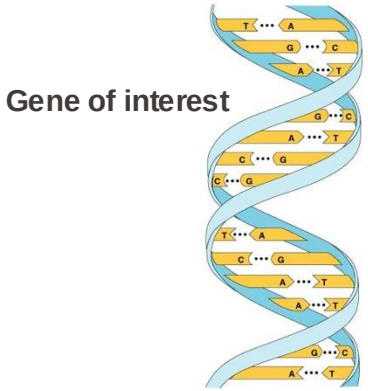
• A few examples:
  - DNA production at high quantities in bacterial cells
  - Synthesis of complex carbohydrates with specific branching points in mammalian cells
  - Production of proteins based on DNA sequence
  - Synthesis of modified, non-native lipids
  - Virus and vaccine manufacturing



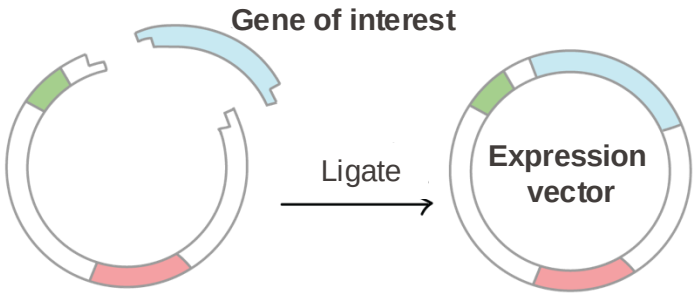
Virus = Nucleic Acids + (Glyco)proteins + Lipids

Recombinant production of proteins

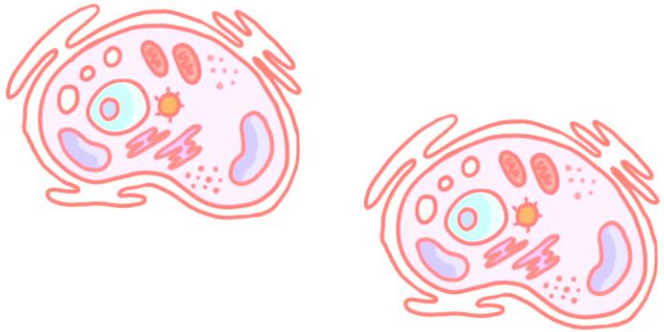
1) DNA encoding the protein of interest



2) Incorporate into an expression vector



3) Transduce the vector into host cells



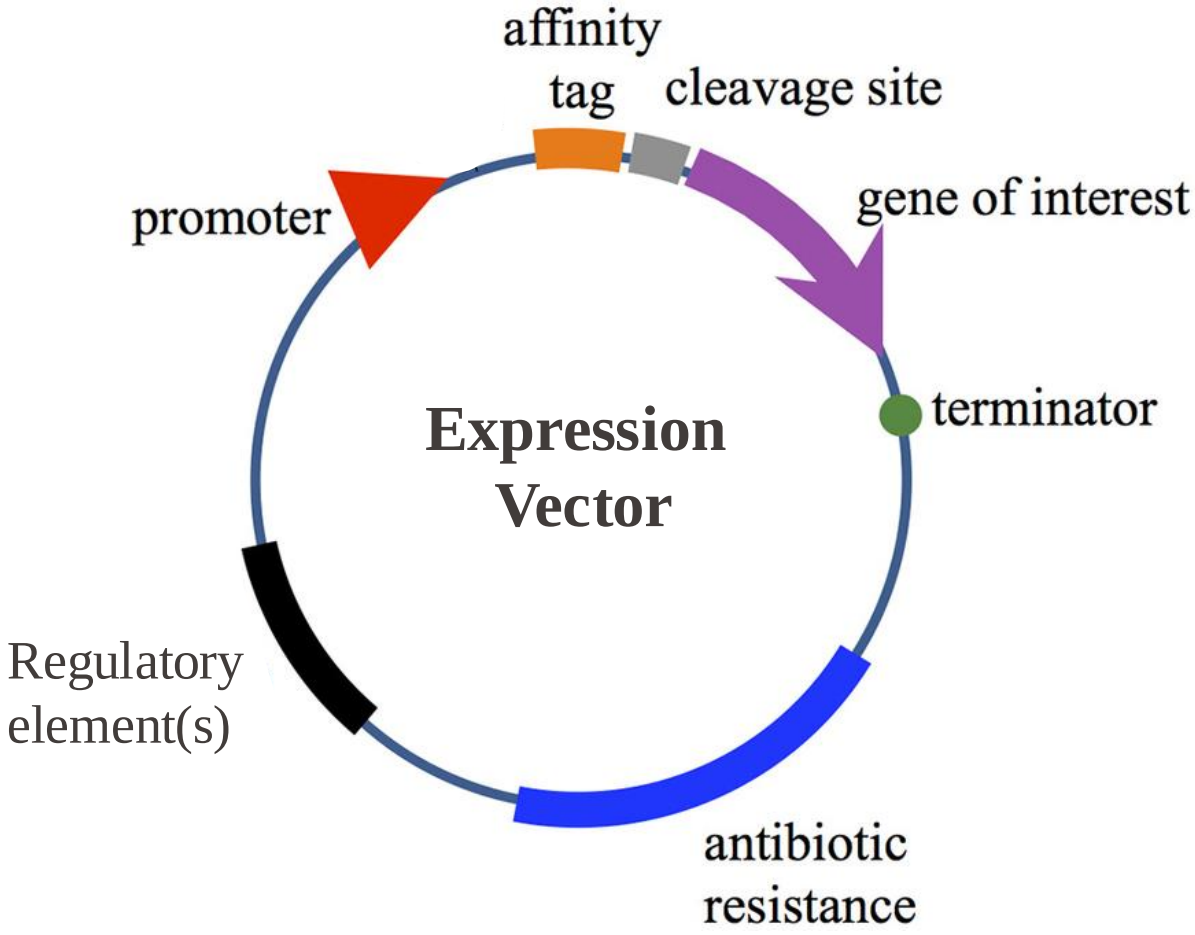
4) Cell growth in a suitable environment

Expression vector components

• Expression vector is based on circular DNA that has the necessary genetic elements (**promoters**) to assure robust transcription of the gene of interest, thereby creating lots of mRNA for translation.

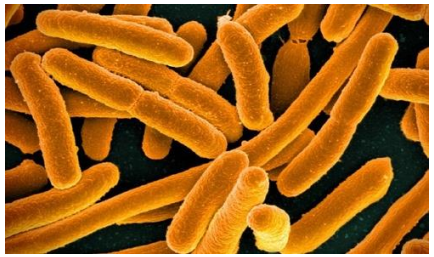
• Additional elements include **antibiotic resistance genes** (e.g., AmpR) for selection of cells that have taken up the vector and **regulatory elements** to (e.g., Lac) to switch on/off the transcription.

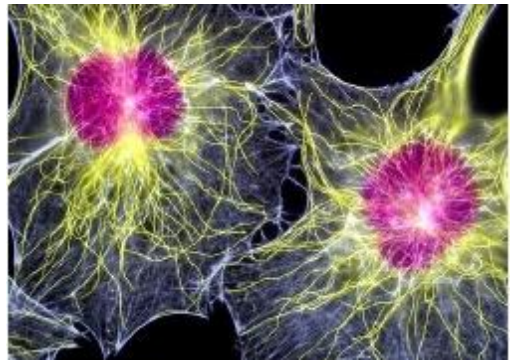
• Additional sequence can be added to the gene of interest to encode a short peptide extension on the protein that can be used for purification (i.e., **affinity tags**)



Recombinant Protein Production Hosts

- Microorganisms
  - ***E. coli***
  - ***Pichia pastoris* (Yeast)**
  - ***Saccharomyces cerevisiae* (Yeast)**
- Cultivated animal cells
  - **Mammalian cells (e.g., HeLa, CHO)**
  - **Insect cells (e.g., SF9)**
- Less common
  - **Transgenic animals**
  - **Cultivated plant cells**



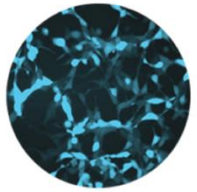


• Cell lines have typically been engineered to produce high quantity of nucleic acids, proteins, carbohydrates or any other biomolecule of interest.

When to use which cell line?

- ***E. coli*:** Small to medium, single domain, no post-translational modifications (PTMs)
- ***Yeast*:** Secreted (extracellular) proteins, antibodies, simpler multi-domain proteins
- **Insect cells:** Larger, more complex multi-domain proteins, may contain some PTMs
- **Mammalian cells:** Larger, more complex multi-domain and membrane proteins with PTMs



	Bacterial Expression System	Yeast Expression System	Insect Expression System	Mammalian Expression System
Speed	★★★★	★★★	★★	★
Yield	★★★	★★★★	★★	★
PTM (relative to human)	✗	★★	★★★	★★★★
Cost	★★★★	★★★	★★	★
Application	Small to medium, Simple single domain proteins	Secreted Protein, Disulfide-bonded protein, Glycosylated protein	Larger, more complex multi-domain proteins	Complex multi-domain and membrane proteins

- Always consider *E. coli* first since it is the cheapest, fastest and most resilient method.
- However, proteins often need to be produced in cells that resemble their origin tissue

Cells use the expression vector to produce protein

• Following the transduction with the expression vector each cell line is placed in a suitable **medium rich with nutrients** to assure optimal growth and translation of protein of interest

• The environment is also adjusted to **maintain temperature, humidity and atmospheric composition**

Usually ~mL - L volumes

• *E. coli* divide (double) every 20-30mins and reach max confluency in less than 24 hours

• Mammalian cells double every ~18-32 hours depending on the cell line (need longer time)

Cell bioreactors in pharmaceutical industry

- Industrial production of biomolecules requires scales that are in the 10^2-10^4L range

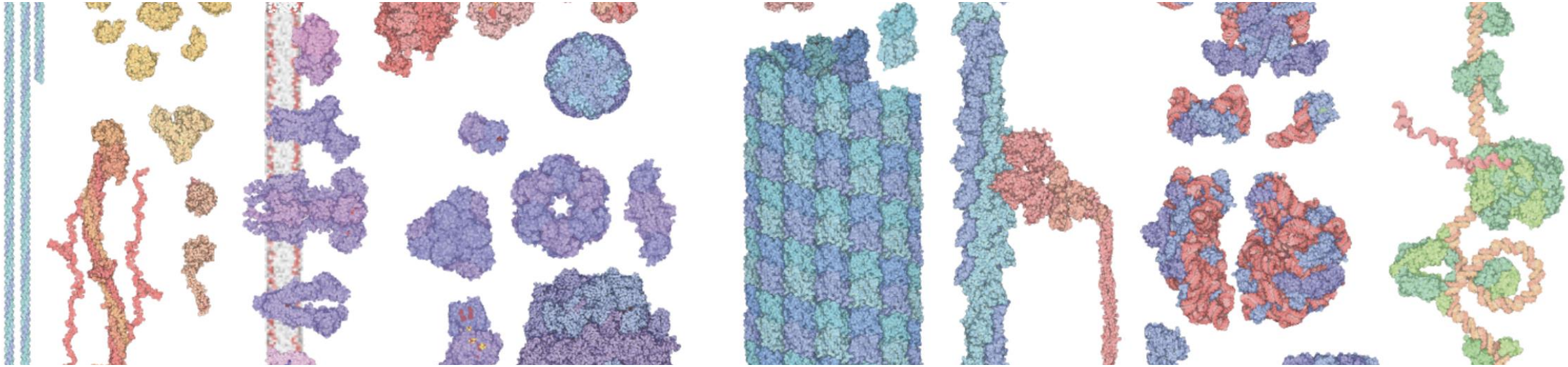
Roche

Sanofi

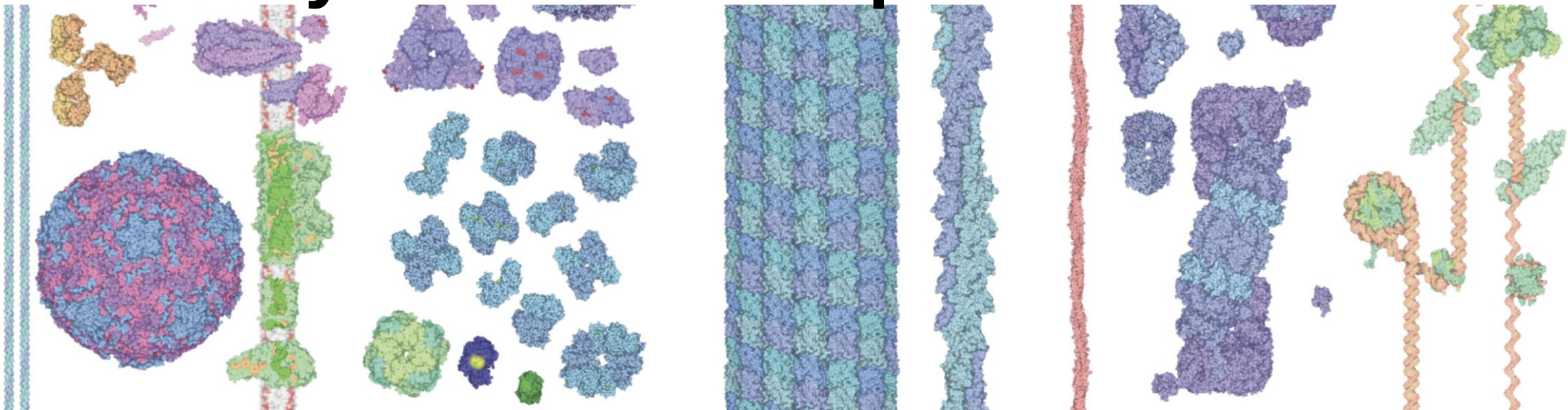
Lonza

- Additional safety measures need to be implemented if making vaccines using live pathogens

Problem: Cells have many biomolecular components



How do you isolate the protein of interest?



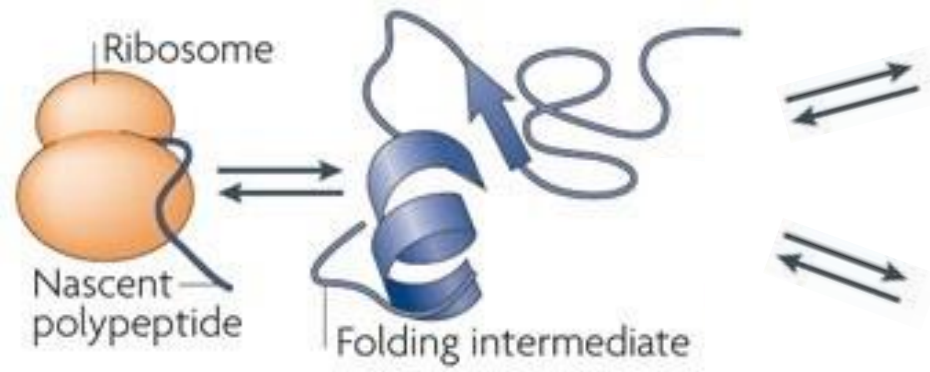
Biomolecule Production

Biomolecule Purification

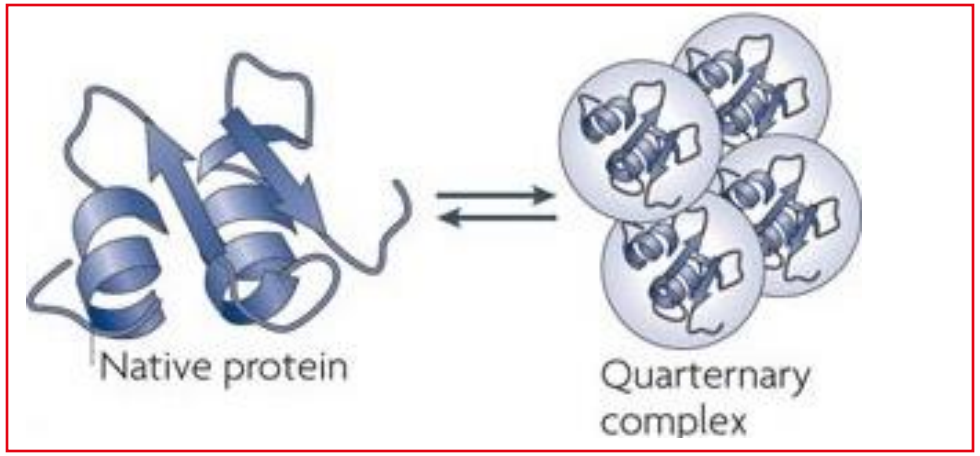
Measuring Quantity and Purity

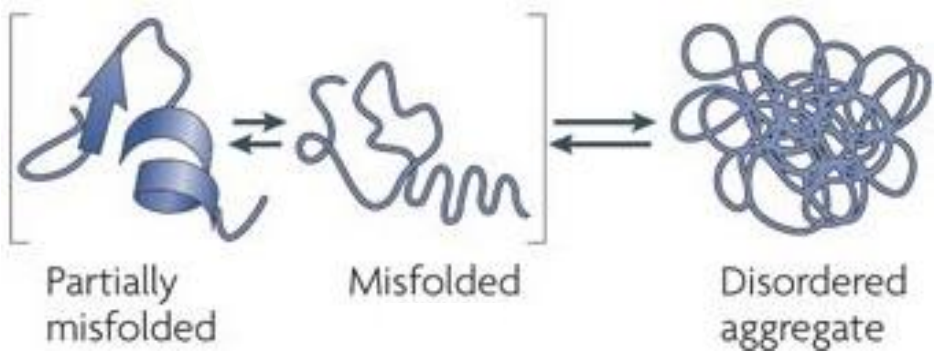
What are the goals of biomolecule purification?

- Specific applications and needs can vary, but generally speaking the goals are:
  - Obtaining ~mg quantities of the specific molecule
  - It needs to be pure of the remaining cell material
  - Natively-folded (=in active state)
  - Native oligomeric state (monomer, dimer, or other)
  - Stable at high concentrations (no aggregation)
  - Stable over time (days or even weeks)



Preferred outcome





- Protein expression can lead to a mix of folded and missfolded molecular species

Typical steps in protein production and purification

Design expression vector

↓

Express protein in cell culture

↓

Collect cells by centrifugation
(or medium if protein is secreted)

↓

Cell lysis to release proteins
(not needed if secreted)

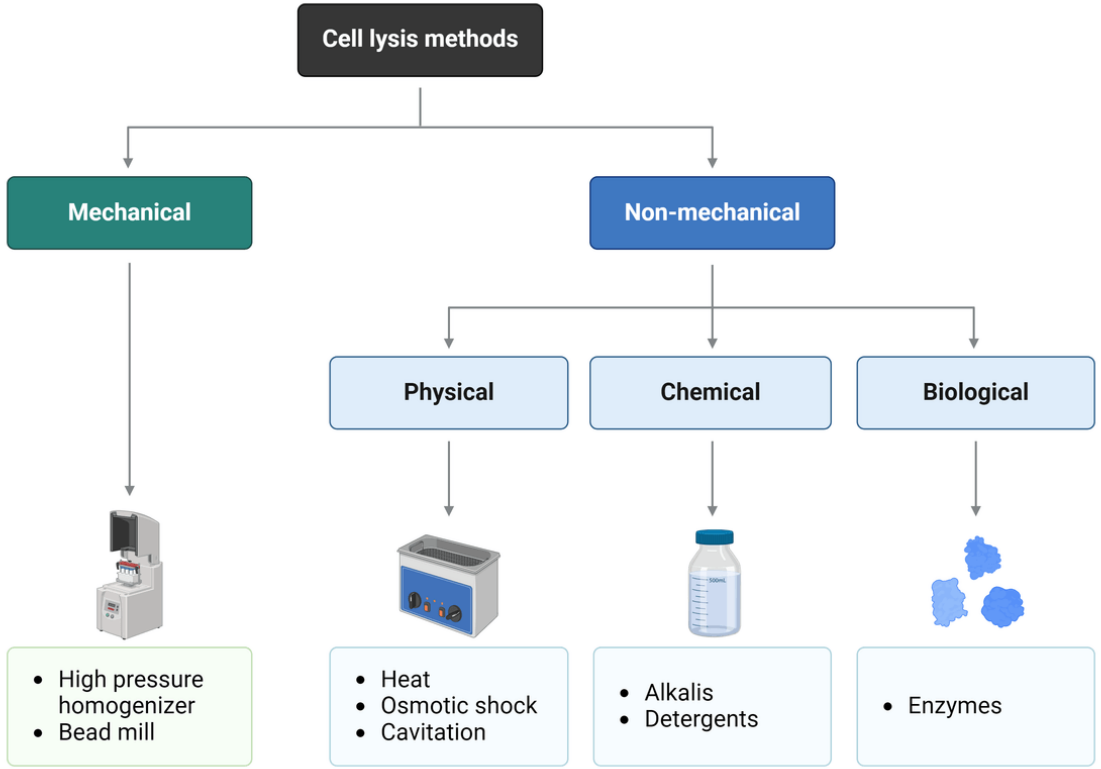
↓

Purify protein using
liquid chromatography

↓

Assess quantity and purity
(UV absorbance, SDS PAGE)

- If the protein is expressed intracellularly (cytoplasm or specific organelle) the cells need to be lysed to release the proteins
- Cell lysis can be performed using several methods:



- If the protein is expressed extracellularly (into the medium) then the cells can be discarded, and medium should be collected

Buffers - Liquid medium for biomolecule handling

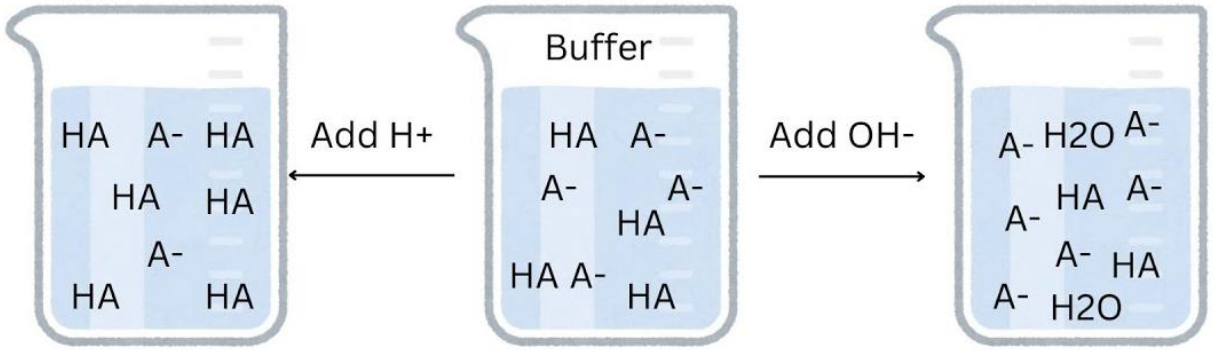
• A buffer is a solution that can resist pH change upon the addition of an acidic or basic components. It is able to neutralize small amounts of added acid or base, thus **maintaining the pH of the solution**.

Acid Base

$$HA \rightleftarrows A^- + H^+$$

$$pH = pK_a + \log \frac{[A^-]}{[HA]}$$

Buffer components bind H^+ and OH^- and maintain the pH



• Typically based on weak acids and bases with pKa value in the 3-10 range (e.g., phosphate, tris-HCl, MES, carbonate, acetate)

• **The buffering capacity is highest in pH ranges that are within 1 unit from the pKa**

• Other buffer components: salt (e.g., NaCl), detergents (e.g., Triton-X), organic solvents (e.g., glycerol), reducing agents (e.g., dithiothreitol-DTT)

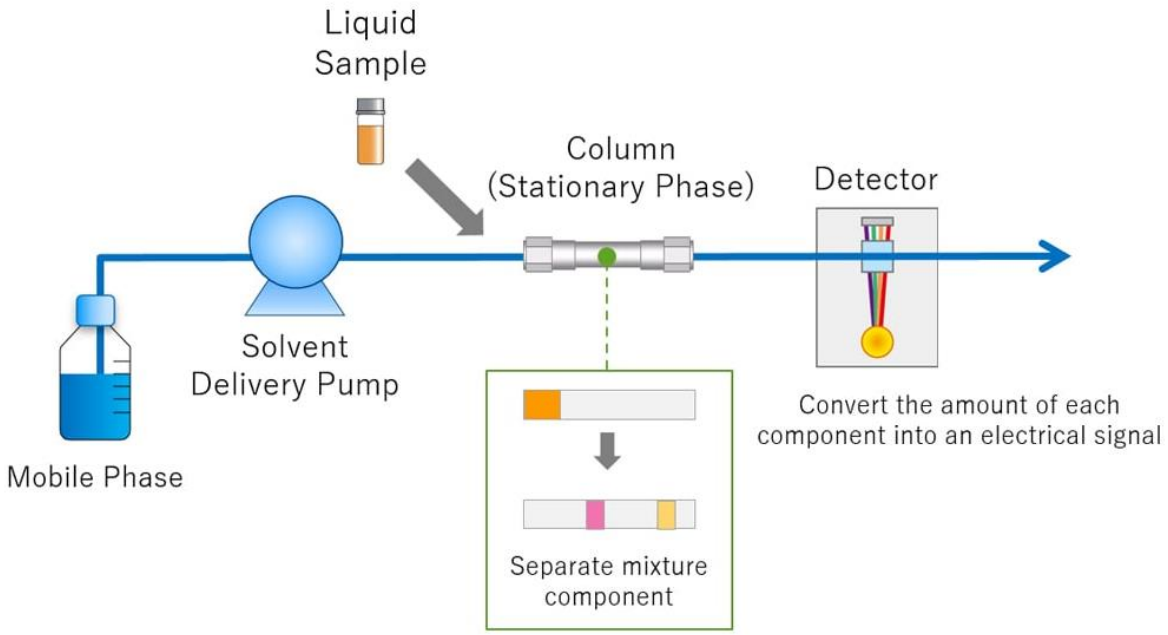
Phosphate-buffered saline (PBS): **10mM Phosphate (pH 7.4) + 137mM NaCl + 2.7mM KCl**

Liquid chromatography

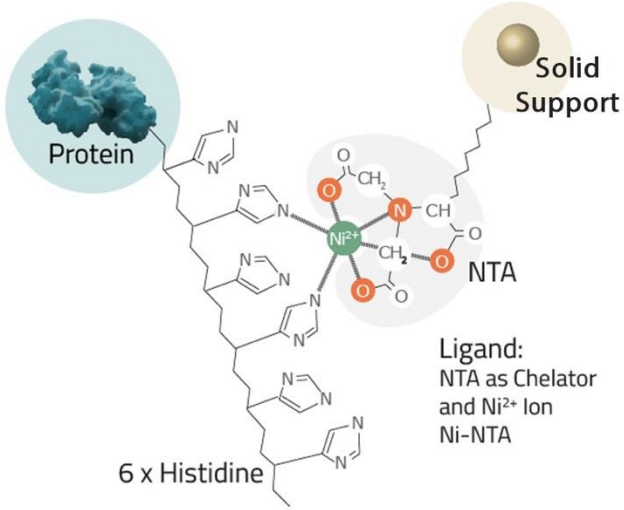
- Liquid chromatography is a technique used to separate a mixture of (biological) samples into its individual parts.

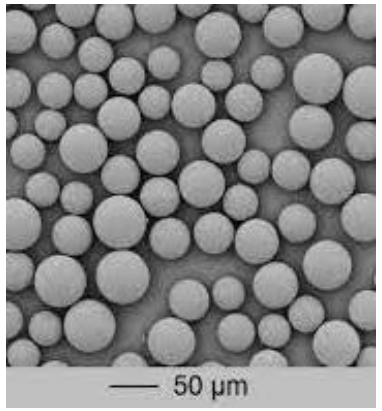
- This separation occurs based on the differential interactions of each sample with the **mobile** and **stationary** phases.

- Stationary phase consists of resin (matrix) which is packed into columns
  - Matrix is composed of highly polymerized small gel beads (some 10 µm in diameter)
  - Common materials are polysaccharides or synthetic polymers
  - Matrix is chemically derivatized with different groups (charged, polar, hydrophobic, antibody)



Ni-NTA matrix (binds poly-histidine tag)

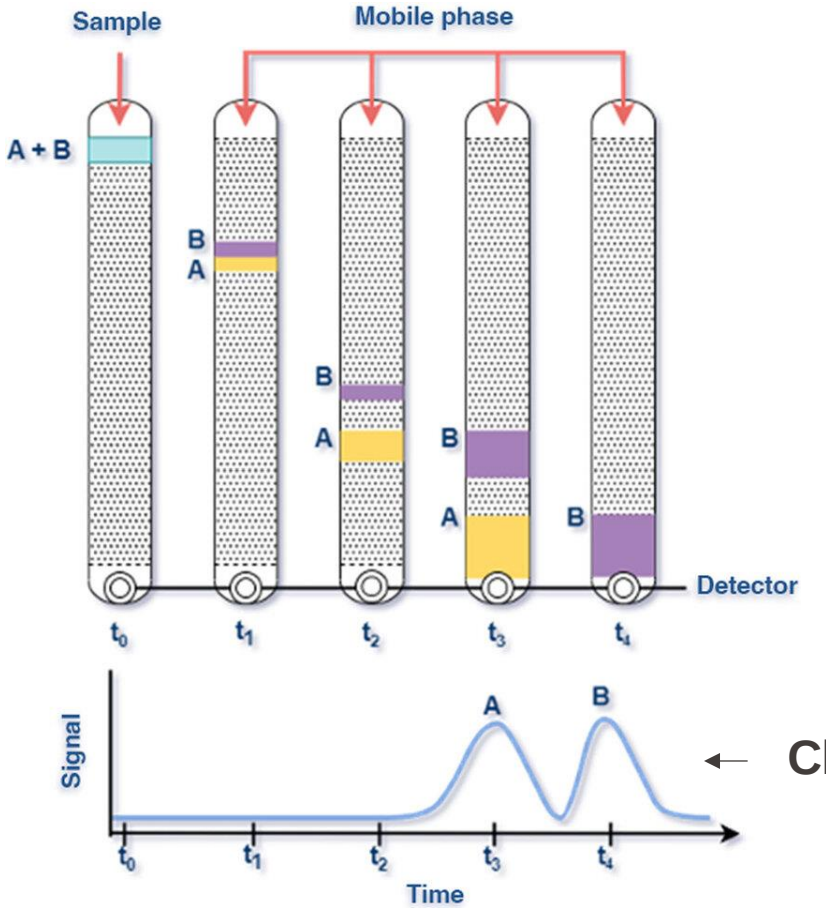




Matrix beads

Liquid chromatography

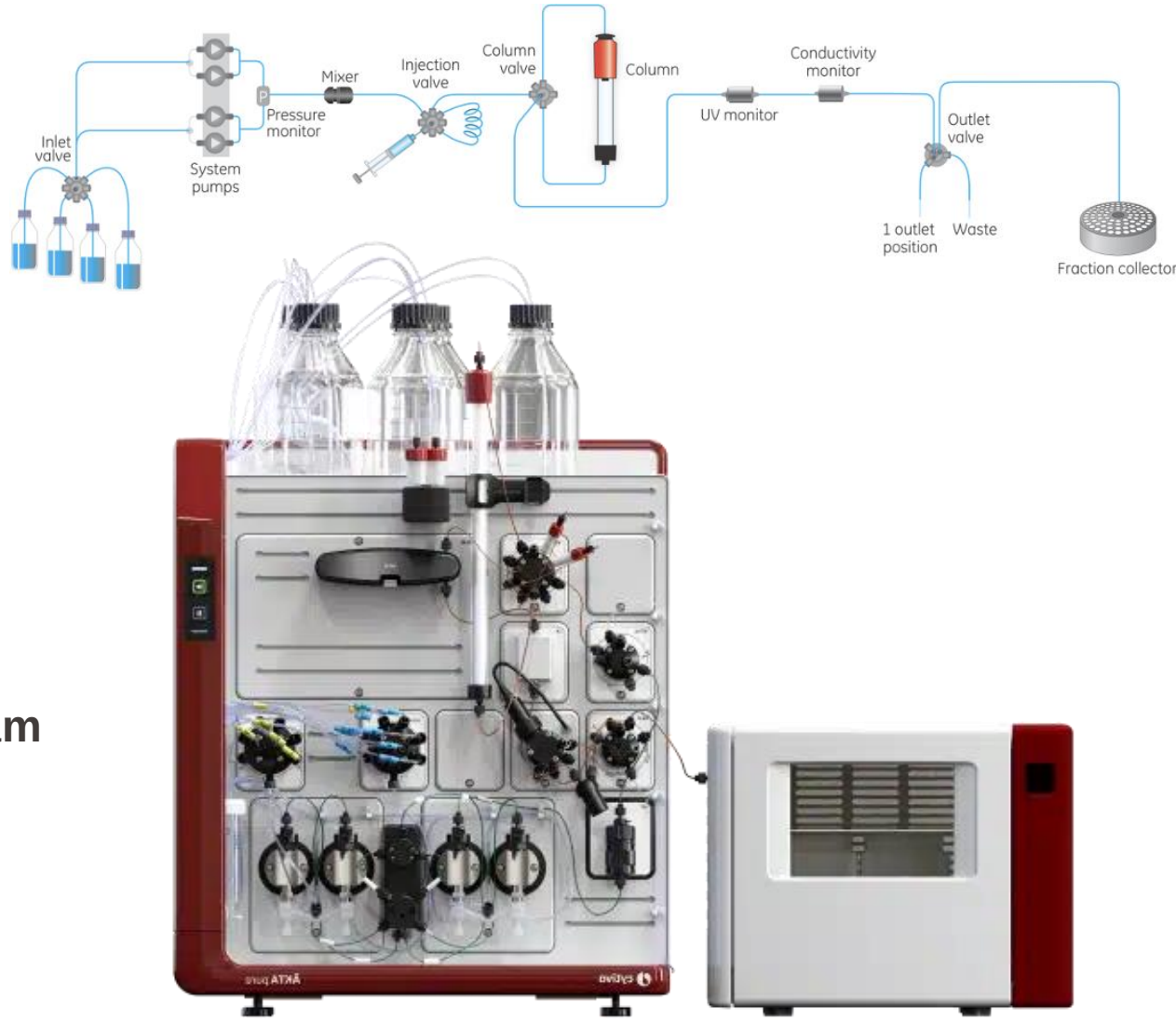
Separation principle



← **Chromatogram**

Samples A and B elute at different times

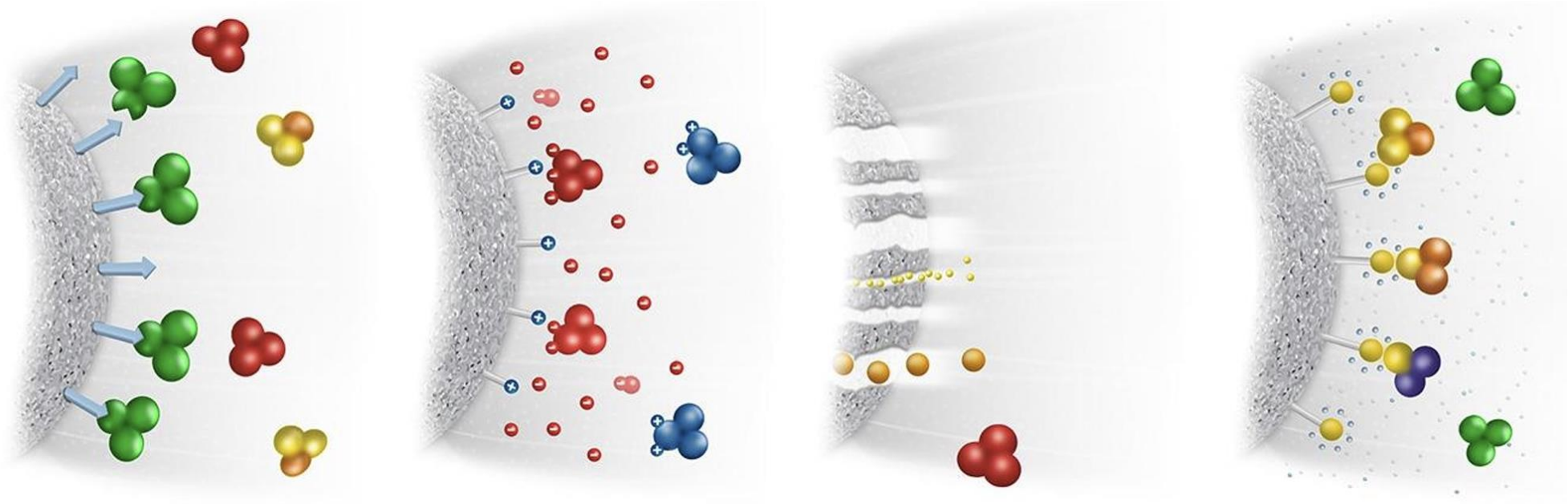
Fast Protein Liquid Chromatography (FPLC) System



The main types of chromatographic methods

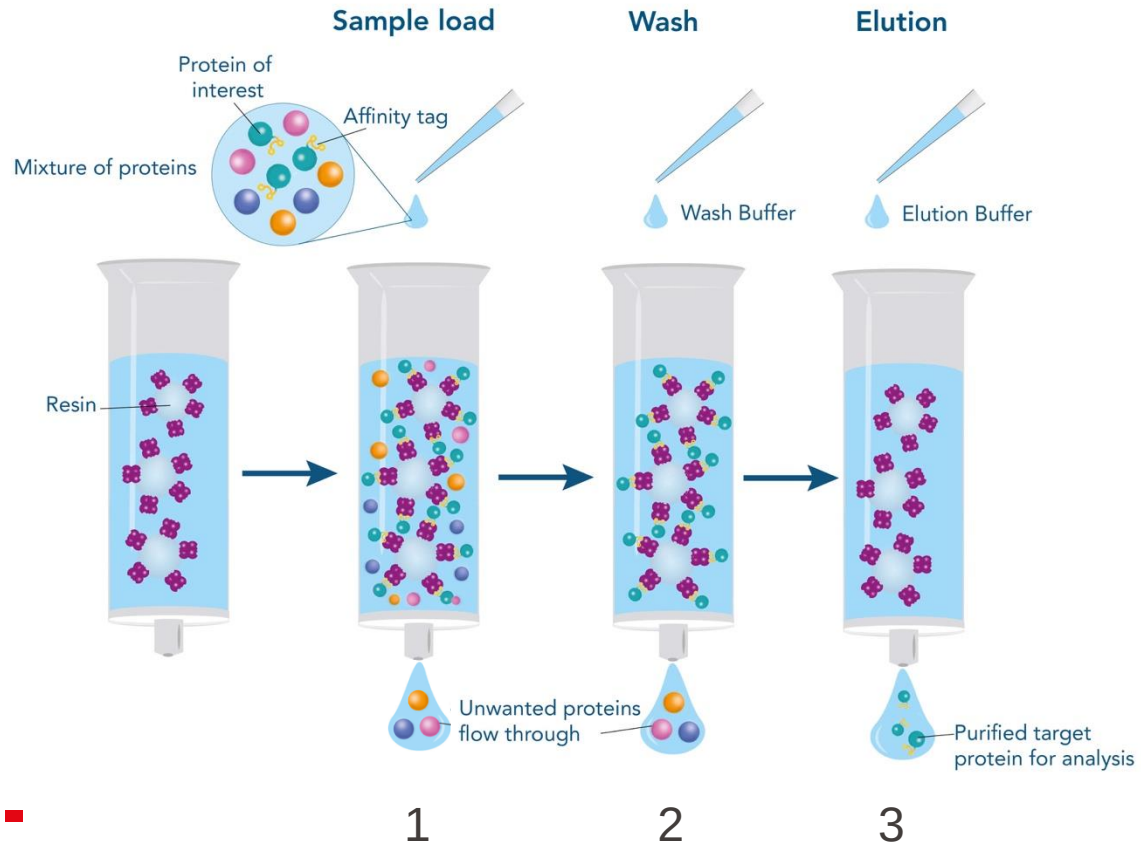
Method	Biochemical principle	Applications
Ion-exchange chromatography (IEX)	Different charges of biomolecules	Protein, DNA, charged lipids
Size-exclusion chromatography (SEC)	Different sizes of biomolecules	Protein, DNA, lipids, sugars
Affinity chromatography (AC)	Binding to a chemical group	Mainly proteins
Hydrophobic interaction chromatography (HIC)	Differing hydrophobic properties	Lipids, sugars, proteins

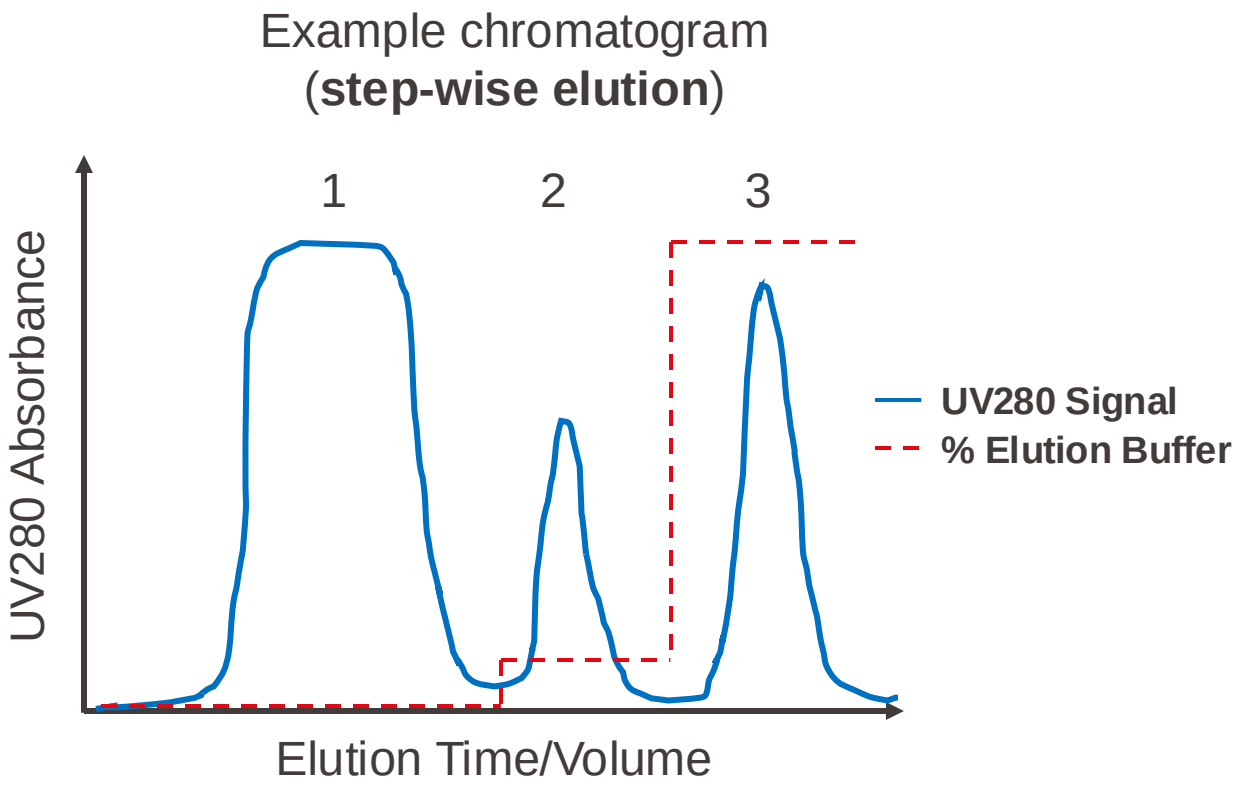
AC **IEX** **SEC** **HIC**



Affinity chromatography

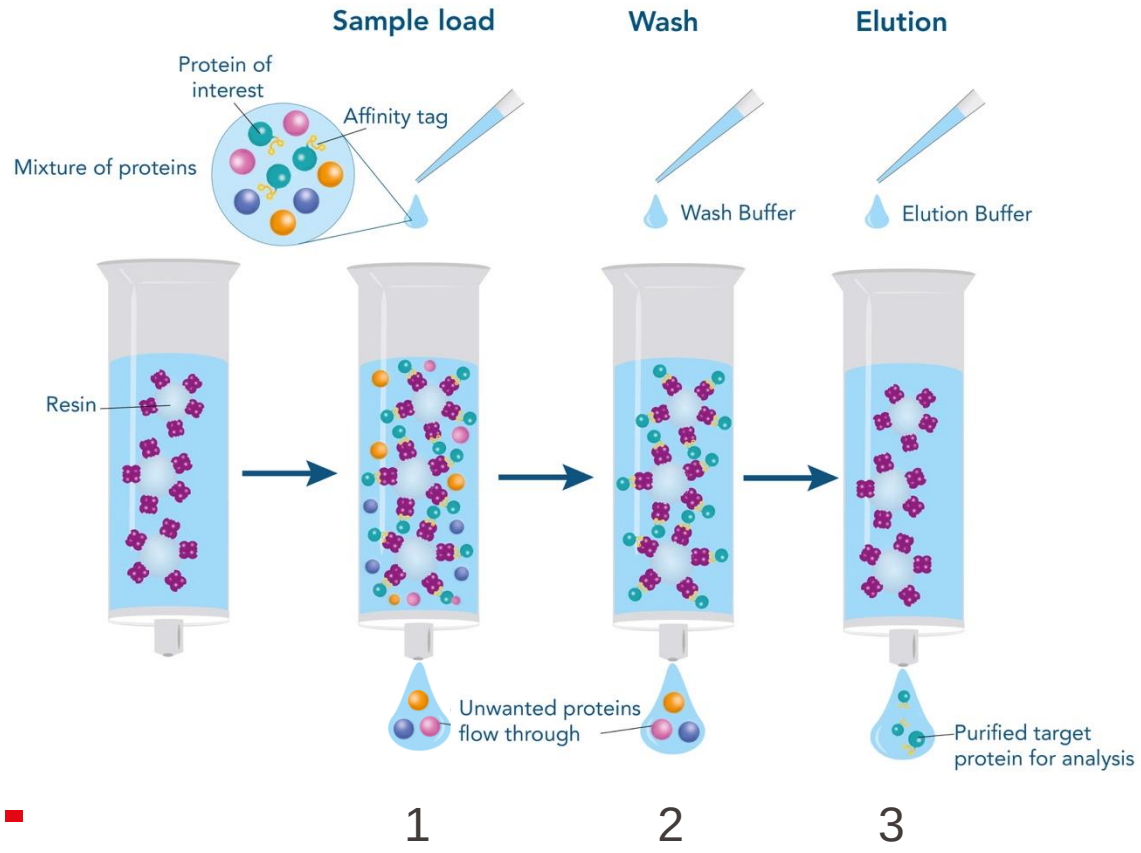
- Most commonly it is the first method for purification of target biomolecule from cell lysate.
- The stationary phase (matrix) is based on immobilized chemical or protein groups that selectively bind the protein of interest (via a tag), while they display minimal affinity to any other protein

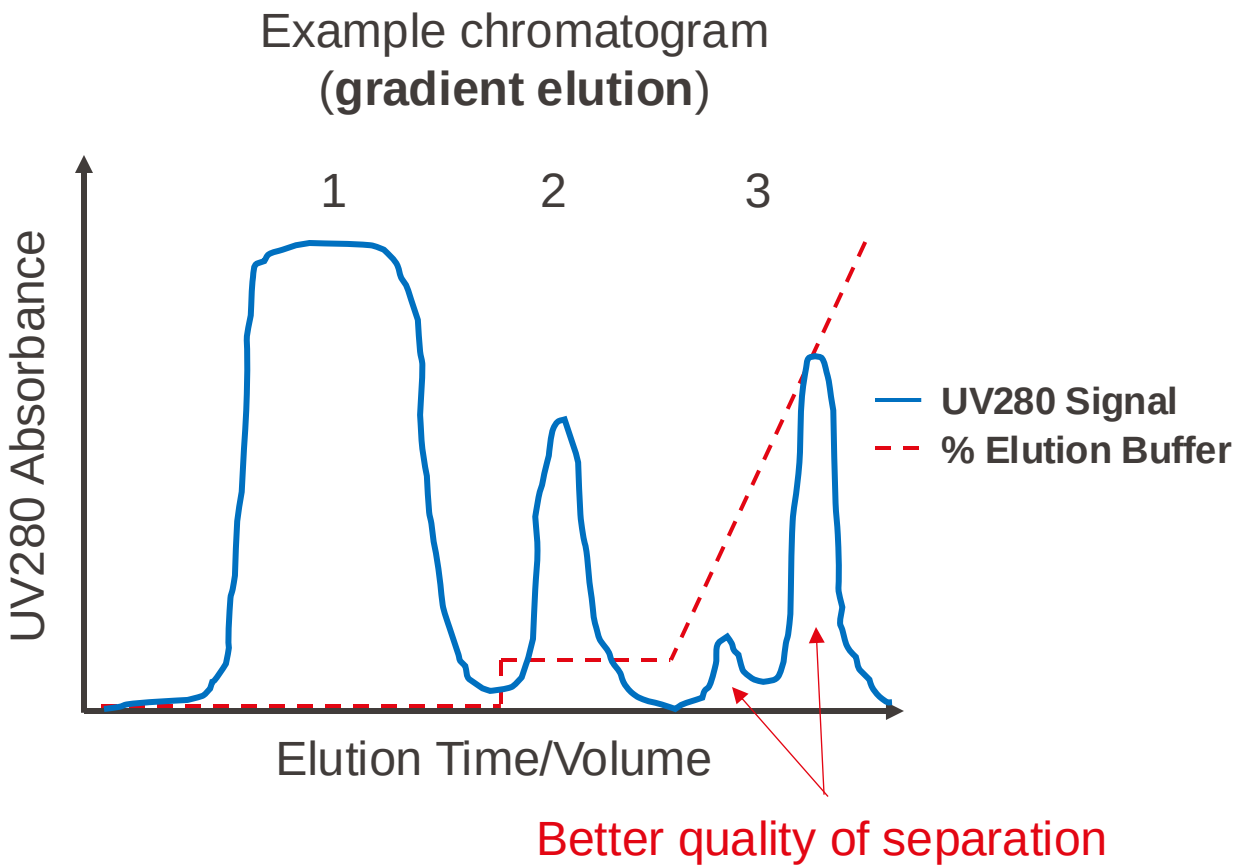




Affinity chromatography

- Most commonly it is the first method for purification of target biomolecule from cell lysate.
- The stationary phase (matrix) is based on immobilized chemical or protein groups that selectively bind the protein of interest (via a tag), while they display minimal affinity to any other protein

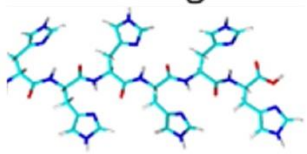


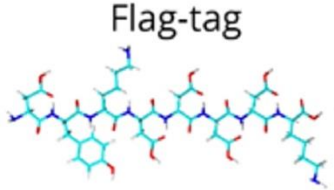


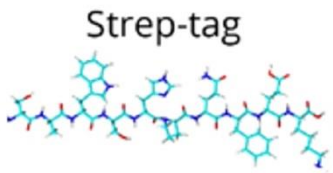
Affinity chromatography - Some common tags

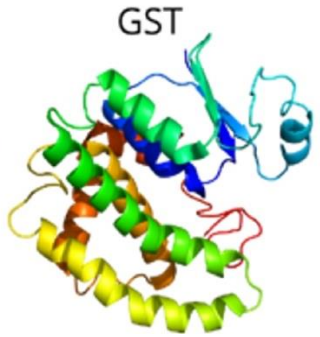
- Protein tags are added during the expression vector design and are fused in frame with the protein gene so that they can be coded into the protein sequence
- The location of the tag is typically at the N- or C-terminus of the protein
- Some common choices and their properties:

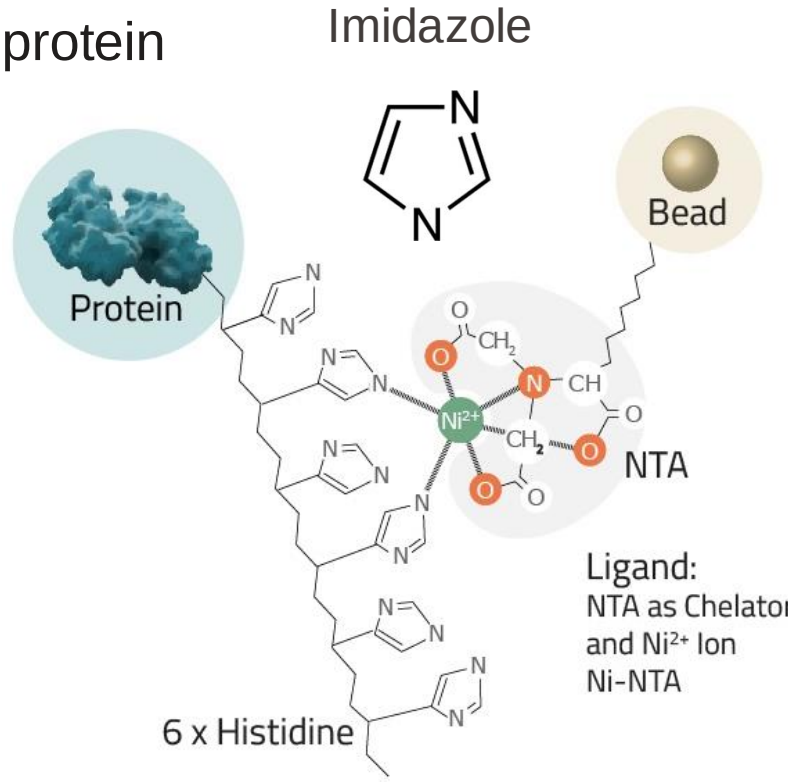
Tag	Sequence	Resin	Elution agent
His-tag	HHHHHH	Ni-NTA	500mM Imidazole
Strep-tag	WSHPQFEK	Streptavidin	10mM Biotin
Flag-tag	DYKDDDDKG	Anti-Flag Ab	0.1mM Flag peptide
GST	Whole domain	Glutathione	20mM Glutathione











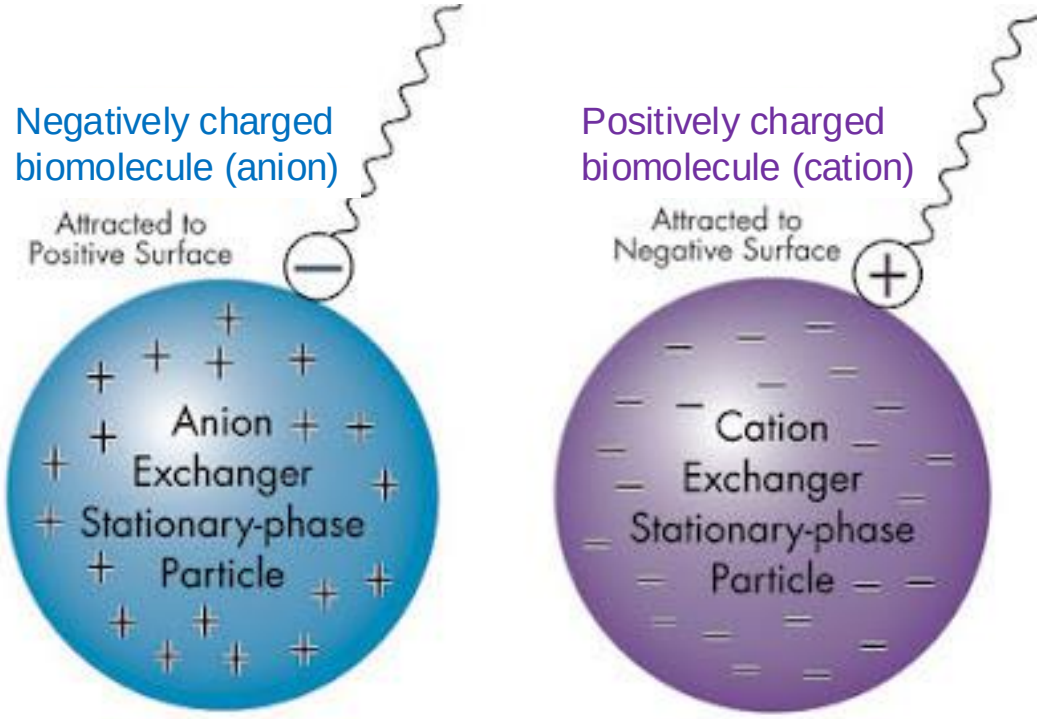
Typical Elution Buffer composition:

- 10-100mM Buffering component (e.g., Tris-HCl)
- 10-1000mM Salt (e.g., NaCl)
- Elution agent (from the table)
- Other components if/when needed

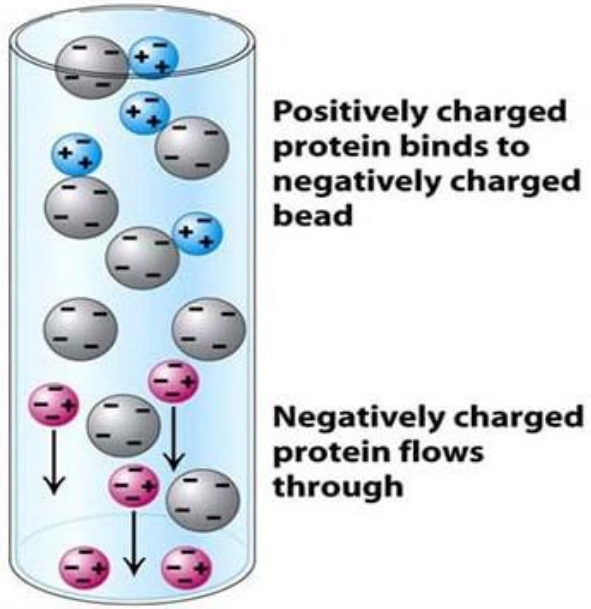
Ion Exchange Chromatography

- Ion exchange chromatography is a **separation method based on different properties of charges and different amounts of charge** in target biomolecules

- The stationary phase (resins) comprise chemical groups that are **positively or negatively charged**



Principle of work

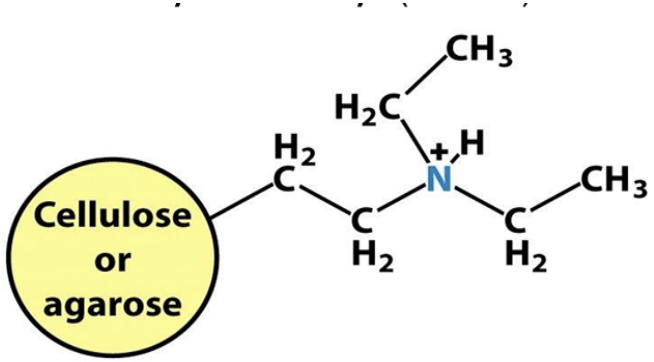


Ion Exchange Chromatography

- Ion exchange chromatography is a **separation method based on different properties of charges and different amounts of charge** in target biomolecules

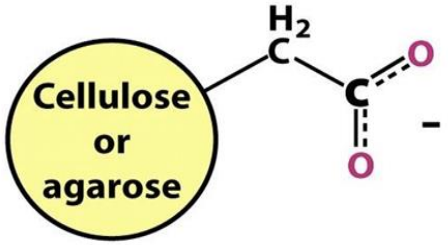
- The stationary phase (resins) comprise chemical groups that are **positively or negatively charged**

Anion exchanger



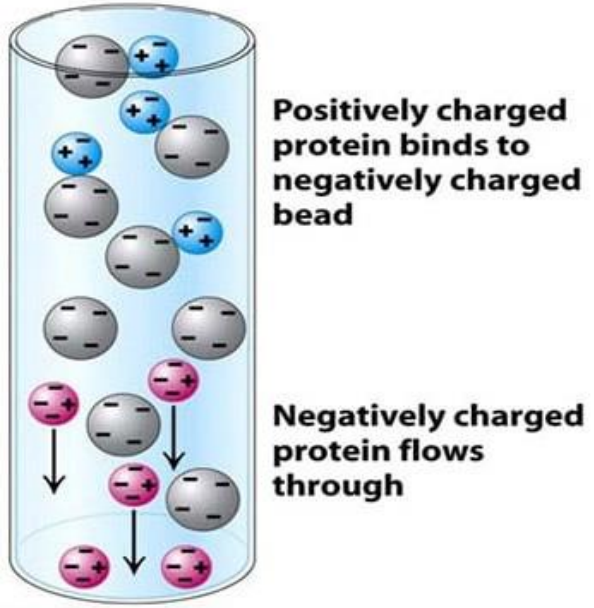
Diethylaminoethyl (DEAE) group (protonated form)

Cation exchanger



Carboxymethyl (CM) group (ionized form)

Principle of work



Which ion exchange resin would you use to immobilize or purify DNA?

Ion Exchange Chromatography

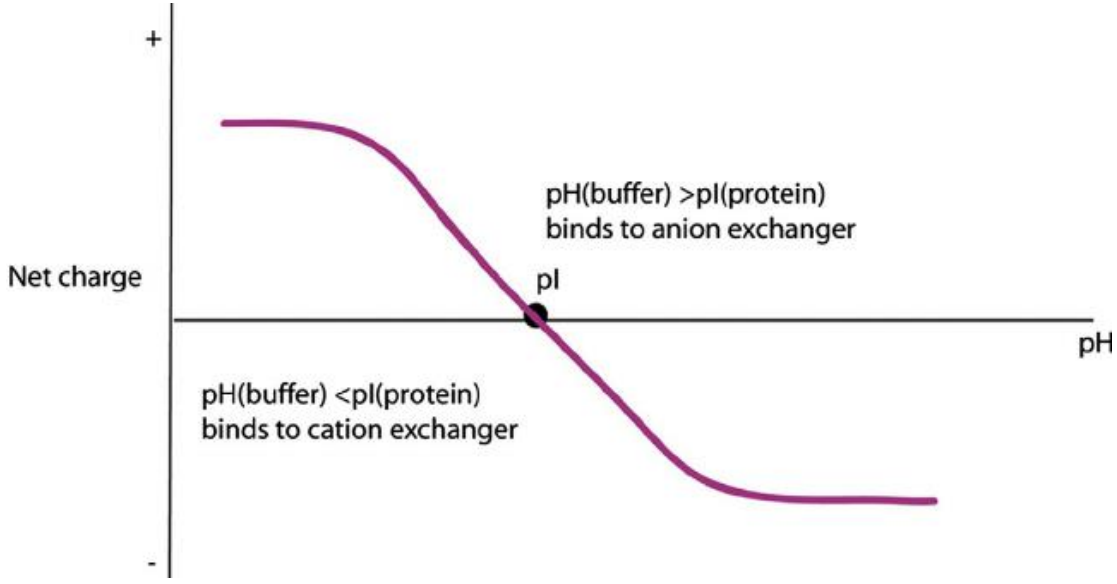
- The separation of protein species is achieved by differing net charge at a given pH, which is based on the protonation state of the charged groups (e.g., **Lys, Arg, Asp, Glu, N- and C-termini**)
- **Isoelectric point (pI)** of a protein corresponds to the pH at which the protein has no net charge. It is calculated as an average of the pKa values of the charged groups.
- Each protein has a different amino-acid composition and **different net charge at neutral pH** which means that different proteins will interact weaker or stronger with the same ion exchange resin

$$pI = \frac{pKa_1 + pKa_2 + \ldots. + pKa_n}{n}$$

pH > pI : Protein **negatively** charged

pH < pI : Protein **positively** charged

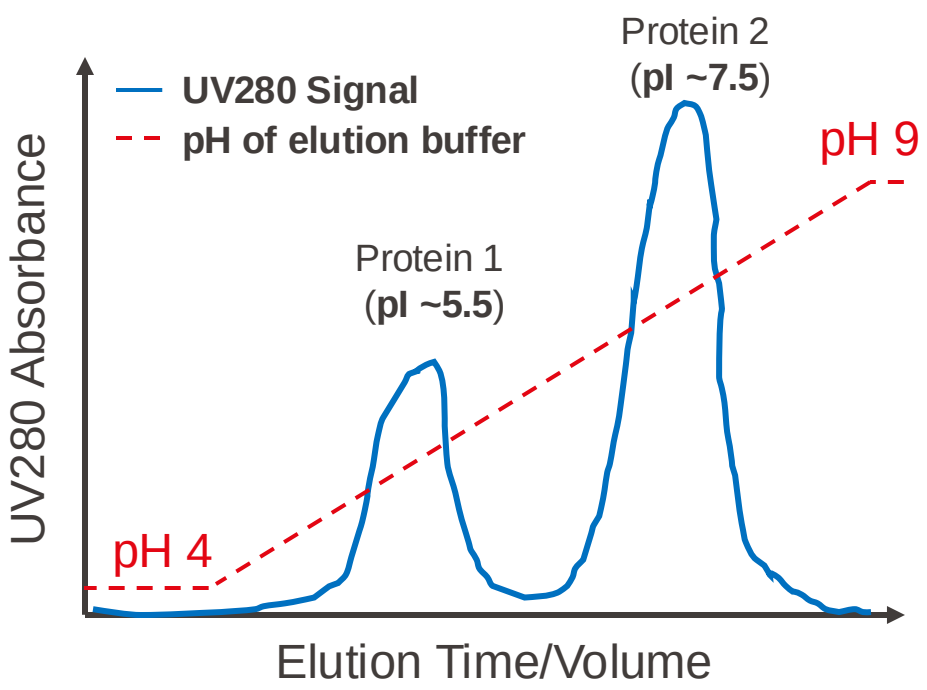
Selecting the optimal column for your protein



Ion Exchange Chromatography : Elution

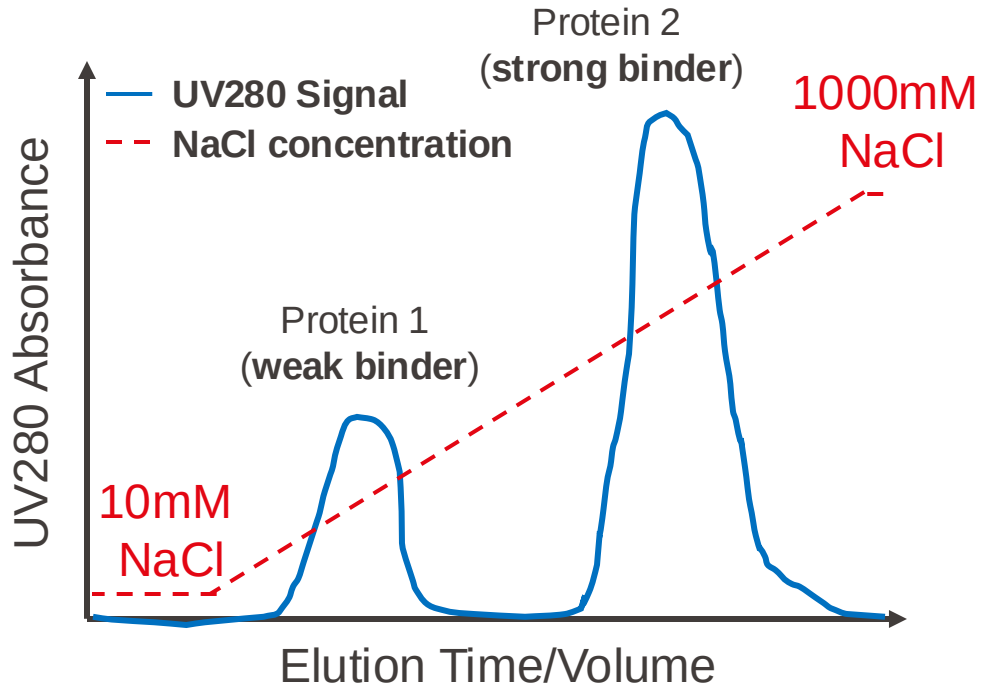
- After the protein is bound to the appropriate cation- or anion-exchange resin, the elution can be performed either by **(1) gradually changing the pH** or **(2) by changing the salt concentration**

pH gradient elution



• Each protein elutes when pH reaches their pI value due to charge reversal

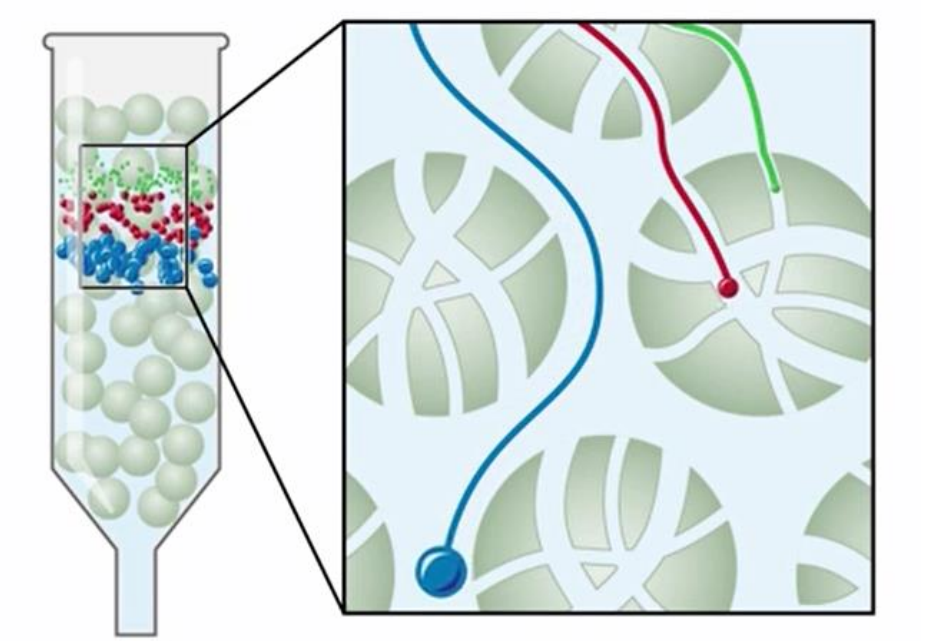
Salt gradient elution

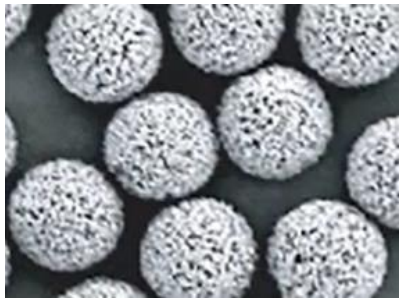


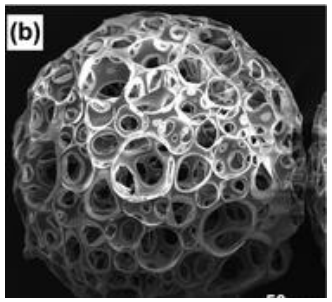
• Each protein elutes when the salt reaches sufficient concentration to outcompete it on the resin

Size-exclusion (gel filtration) chromatography

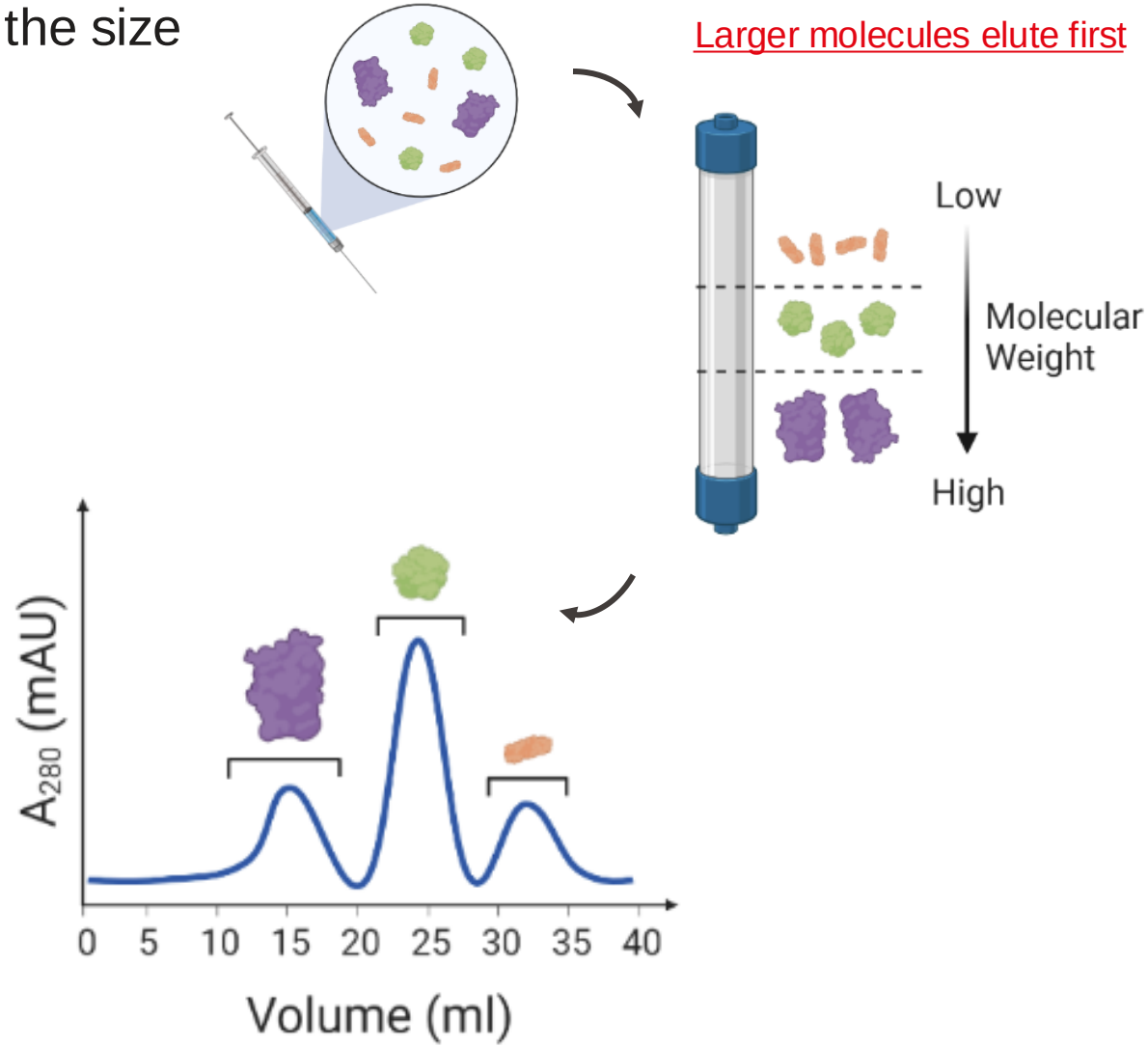
- Protein purification based on **different sizes (molecular weights)**
- Columns are based on porous resins through which the molecules migrate and get retained differently depending on the size





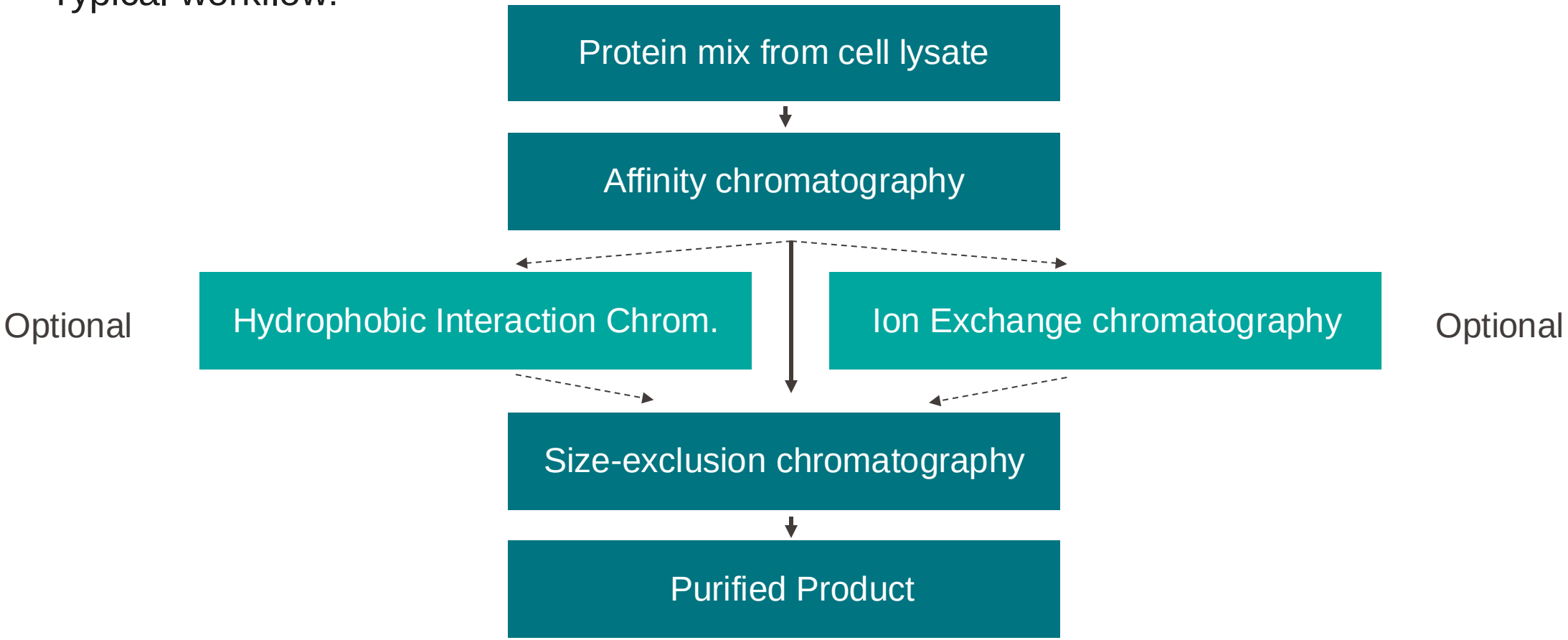


Larger molecules elute first



Size-exclusion (gel filtration) chromatography

- Size exclusion chromatography is typically applied at the end of the purification experiment for sample polishing (removal of last impurities) as well as for separation of different oligomeric forms (e.g., monomer, dimer, aggregate) of the same protein.
- Typical workflow:



Biomolecule Production

Biomolecule Purification

Measuring Quantity and Purity

Confirming the properties of the molecule of interest

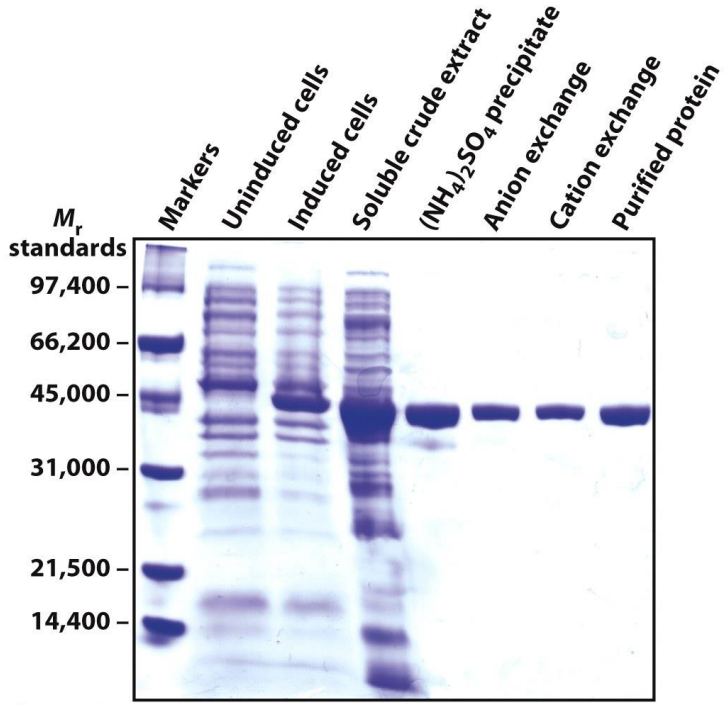
- Following the purification it is essential to confirm the **identity of the biomolecule, correct size (MW), functionality and to measure the concentration/quantity (yield)**.

- In some cases, this is also performed during purification to keep track of the progress (see example on the right containing different samples during purification - each band corresponds to 1 protein).

- Some basic analyses include:

Assay	Property
Gel electrophoresis	Molecular Weight, Purity
UV absorbance	Concentration, quantity
Western blot	Recognition/binding by antibody
ELISA	Recognition/binding by antibody
Enzymatic assays	Catalytic activity

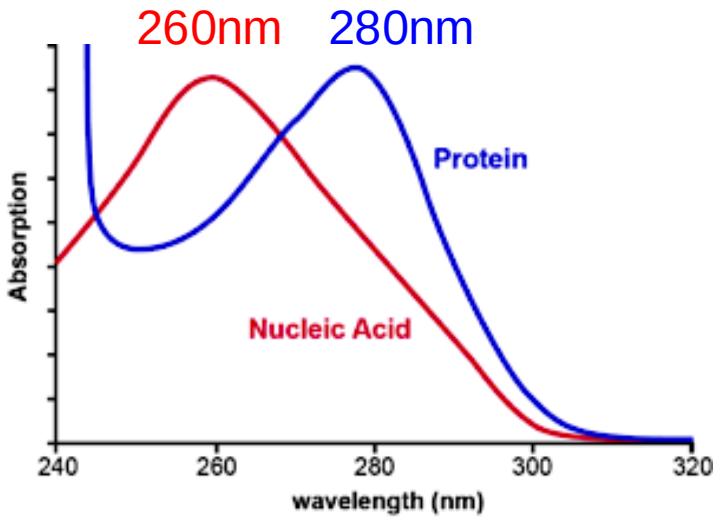
- These analyses are typically done with only a small portion of the sample (ng-µg quantities)



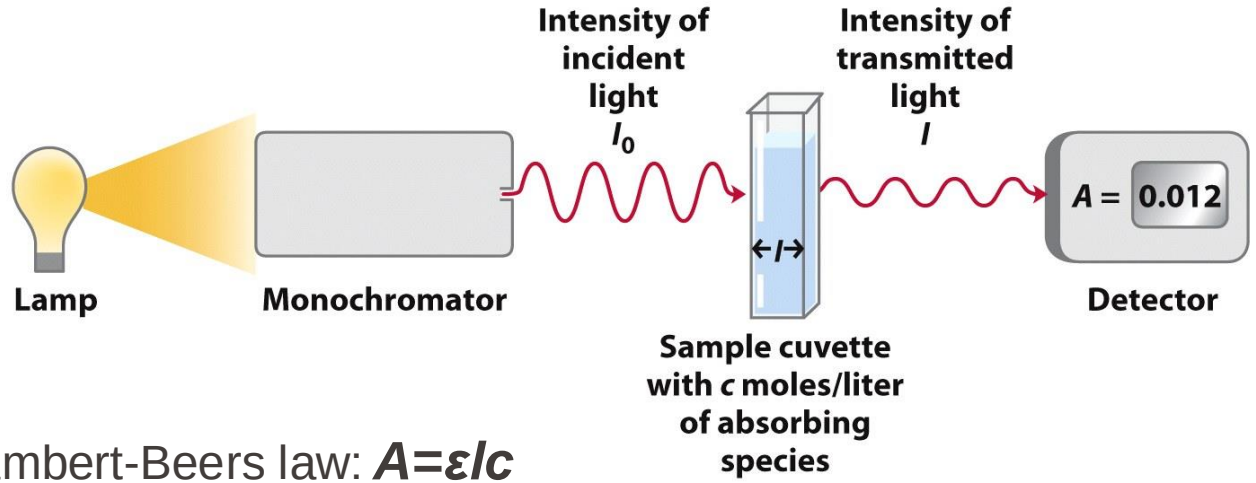
Gel electrophoresis result:

Biomolecule quantification methods

- Nucleic acids and proteins absorb UV light with maximum absorbance at 260nm and 280nm, respectively
- The concentration is calculated based on 260/280nm absorbance, using the **Lambert-Beer equation** below



Absorbance originates from aromatic chemical groups (e.g., Trp, Tyr, bases)



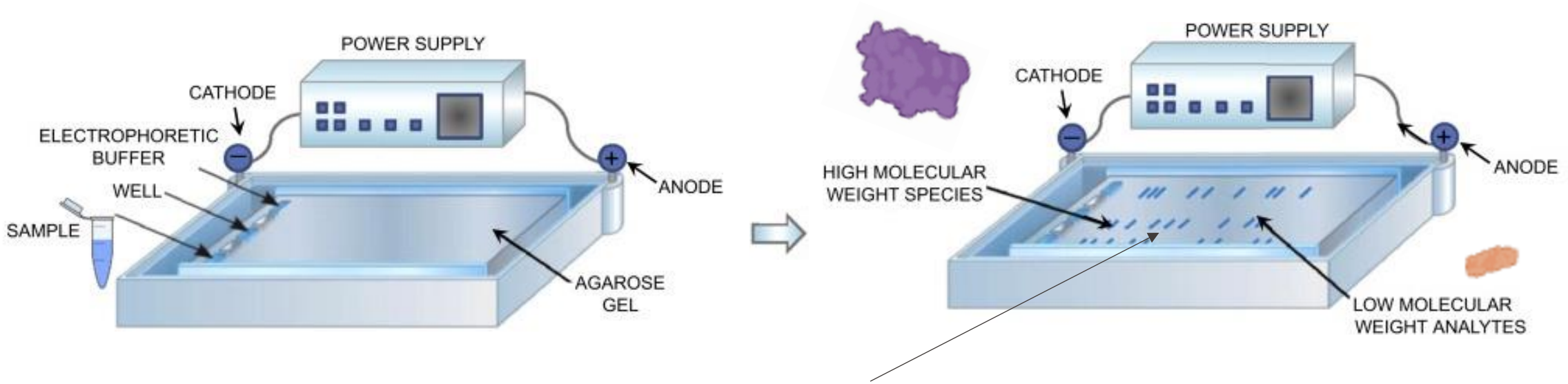
Lambert-Beers law: $A=\varepsilon lc$

A=absorbance,
*ε=extinction coefficient in 1/M*cm,* ← Depends on the sequence
l=path length in cm,
c=concentration in M

- Due to the great chemical group diversity in carbohydrate and lipid molecules, their UV absorbance properties can vary making the quantification more challenging (but still possible)
- Alternative methods for quantification are based on chemical titrations, refraction-index measurements, antibody binding, biochemical kits etc.

Gel electrophoresis methods

- **Gel electrophoresis** methods are used to evaluate biomolecule purity and molecular weight
- Biomolecule-containing sample is pulled through a gel matrix using electric field. Depending on the size the biomolecule will **migrate slower (high MW) or faster (low MW)**

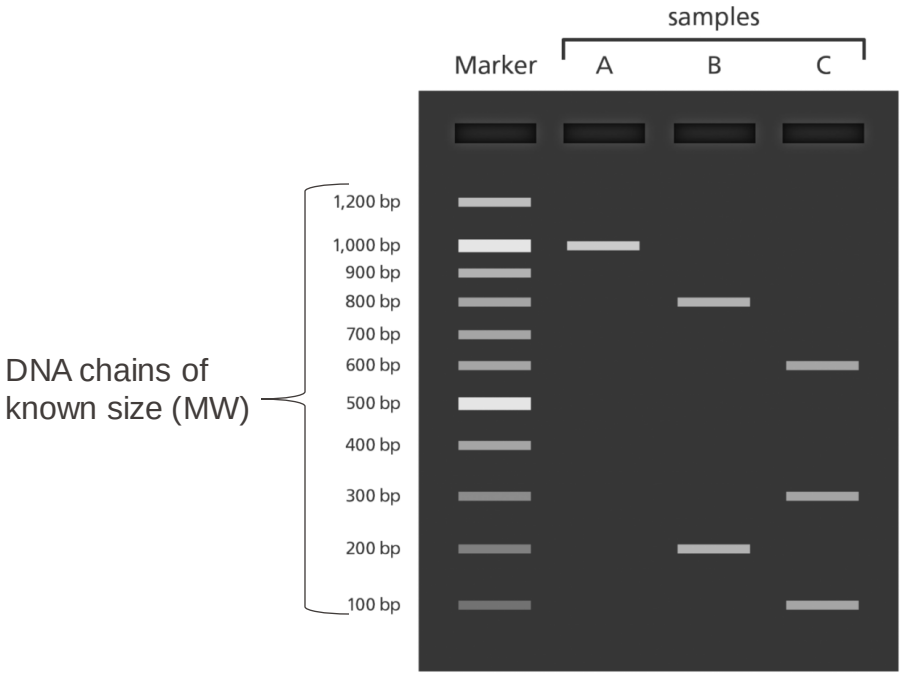


- Each blue band corresponds to one biomolecule species of the given MW
- Molecular weight markers are run in parallel to approximate the molecular weight

Nucleic acid and protein gels

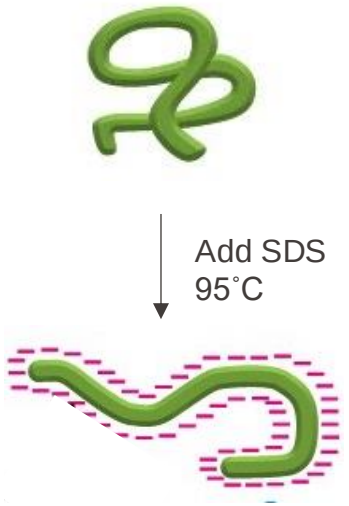
Nucleic acid gels:

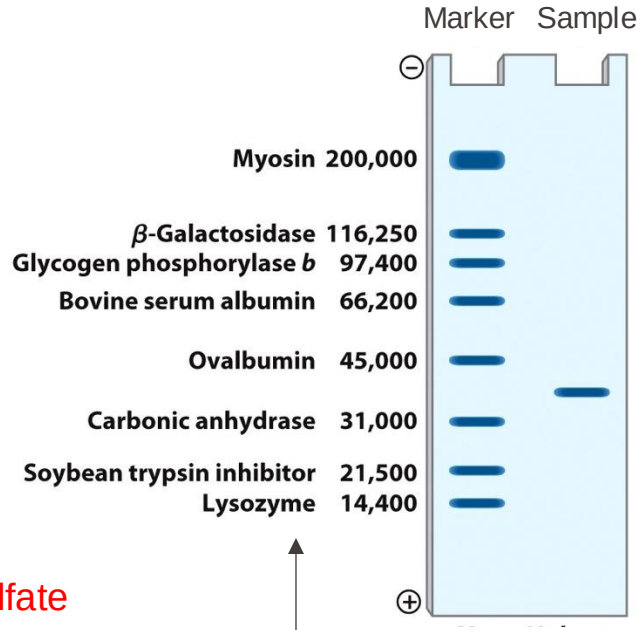
- Agarose-based gels
- Native charge from the backbone
- No need for thermal denaturation



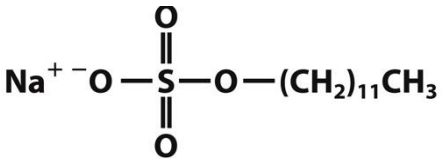
Protein gels (SDS PAGE):

- Polyacrylamide-based gels
- External charge using ionic detergent
- Thermally denatured (linearized) proteins



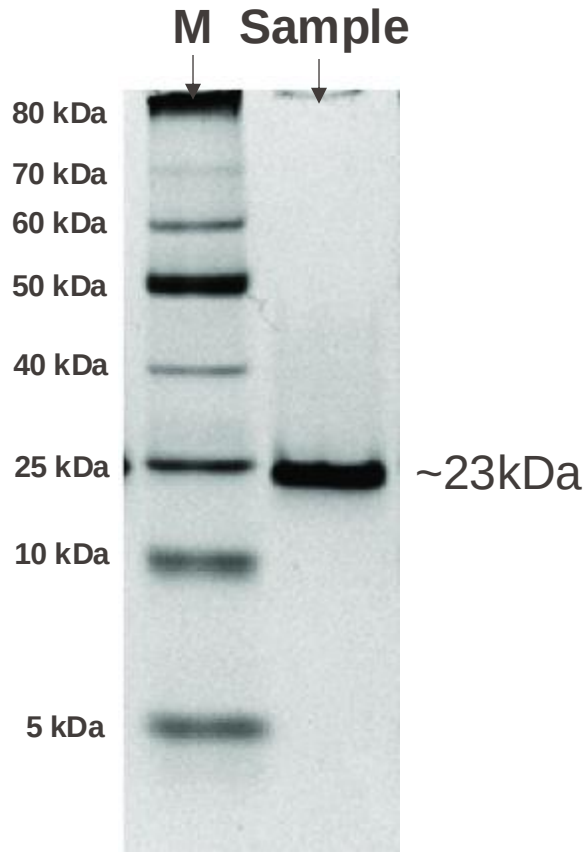


SDS = Sodium Dodecyl-Sulfate

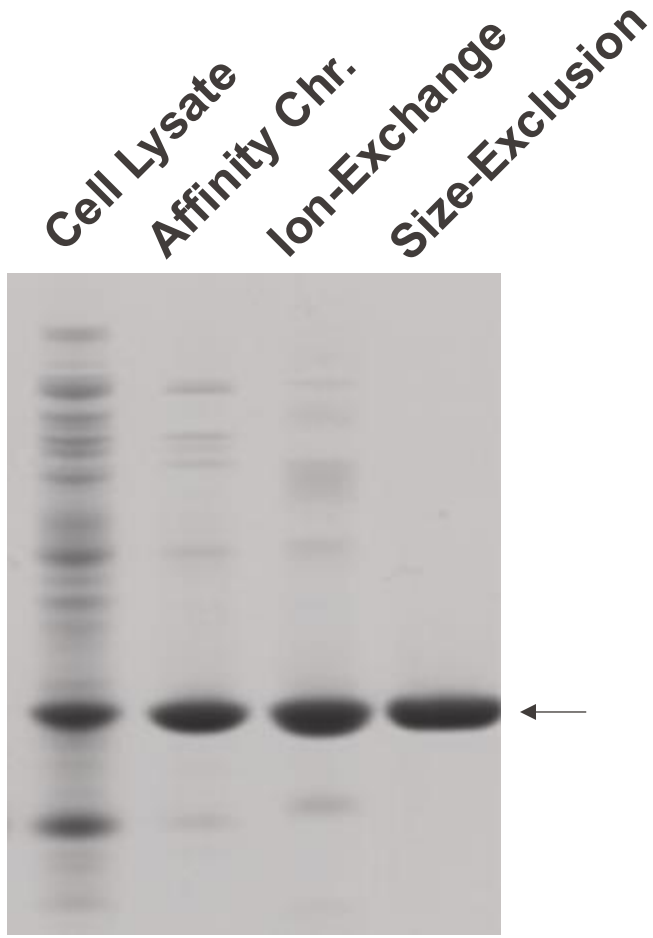


Important applications of gel electrophoresis

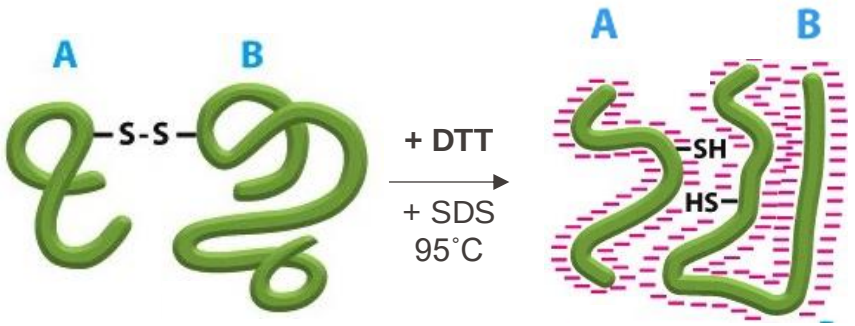
Determine molecular weight

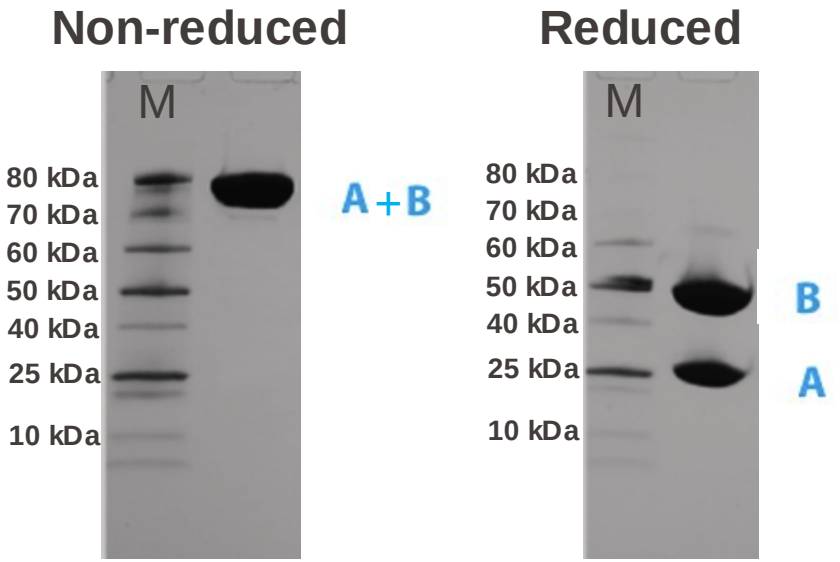


Evaluate sample purity



Study disulfides in protein complexes





DTT = Dithiothreitol (reducing agent)

Summary

- There are conceptually 3 different ways to synthetize biomolecules: chemical, enzymatic and cell-based
- Cell-based methods are the most versatile, particularly for protein production, but there are several steps to follow starting with the generation of expression vectors
- Selection of cell system for protein expression is done by taking into consideration the origin, size and complexity of the target protein, as well as the costs and time of using different systems
- Biomolecule purification is performed using liquid chromatography methods such as ion-exchange, size-exclusion and affinity chromatography
- For rapid evaluation of the yield and purity for biomolecules the standard methods are UV spectrometry and SDS PAGE analysis.