Lecture 8:

Biomolecule Production and Purification

Question 1: True of False

Indicate if the below statements are True or False, and provide explanation where applicable:

a) Chemical synthesis methods, such as solid-phase synthesis, generally yield peptides with natural post-translational modifications like glycosylation and disulfide bonds.

Answer: **False**. SPPS allows precise amino acid sequence control but lacks biological modification machinery. Such modifications usually require enzymatic or cell-based systems.

b) When purifying intracellular proteins that contain free cysteine residues, reducing agents such as DTT or β-mercaptoethanol are added to prevent oxidation and unwanted disulfide bond formation.

Answer: **True**. These agents maintain cysteines in their reduced state, preserving protein structure and activity.

c) Moderate salt concentrations in purification buffers can improve protein stability by shielding surface charges and reducing nonspecific aggregation, thereby improving solubility.

Answer: **True**. At moderate ionic strengths, salts (like NaCl) help stabilize proteins by minimizing electrostatic repulsion and preventing aggregation. However, high salt levels can cause disruption of electrostatic balance and protein precipitation ("salting-out").

d) Mammalian expression systems are often used instead of bacterial systems when proper protein folding and post-translational modifications are required.

Answer: **True**. Bacterial systems have a much simpler chaperone system and lacks the machinery needed for mammalian-like post-translational modifications.

e) In affinity chromatography, the target protein binds specifically to a ligand attached to the stationary phase, and is typically eluted by adding a competing molecule.

Answer: **True**. This is the basic principle of affinity chromatography although elution with salt and pH change (similar to ion-exchange) are also used in some cases.

f) In size-exclusion chromatography, smaller proteins elute before larger ones because they enter more of the pores in the resin beads.

Answer: **False**. Large proteins elute first.

g) In cation exchange chromatography, proteins with a net negative charge at the buffer pH will bind strongly to the column.

Answer: **False**: Cation exchangers bind positively charged proteins.

Question 2: Select an optimal production system for biomolecules

Many biomolecules can be obtained by **chemical synthesis, enzymatic synthesis, or cell-based expression**. For each of the examples below, select the most suitable production approach and briefly justify your choice (~2–3 sentences).

a) Carbohydrate - Branched oligosaccharide for vaccine application

Answer: Enzymatic synthesis (using glycosyltransferases) provides excellent regio- and stereoselectivity under mild solution conditions. Chemical synthesis is possible but complex for multiple linkage types. Cell-based synthesis is possible but hard to control since this is performed without a template and there is no option to strongly influence synthesis of a specific oligosaccharide.

b) Nucleic Acid - 25-mer silencing RNA Duplex

Answer: Chemical synthesis (solid-phase oligonucleotide synthesis) provides precise control over short sequences with high purity and scalability. Enzymatic or cell-based expression is possible, but unnecessary for such small molecules.

c) Protein - Antibody (multiple chains, ~1400 amino acids + glycosylation + disulfides)

Answer: Given the complexity of this protein, the only reasonable option is cell-based expression. Enzymatic synthesis is not practical due to the complex nature, size and post-translational modifications.

d) Lipid - Phosphatidylcholine (PC) with Controlled Fatty Acid Composition

Answer: Chemical synthesis or enzymatic (lipase-mediated) synthesis allows precise control over acyl chain composition. Cell-based production yields complex lipid mixtures with limited control (no ability to strongly influence fatty acid incorporation)

Question 3: Select an optimal expression system for your protein

What expression systems would you choose for the expression of the following proteins? Why or why not and explain:

a) 12 kDa single-domain protein of bacterial origin, residing in cytoplasm (example: KaiB oscillator protein)
Answer: Given the relatively small size, bacterial origin, and the lack of post-translational modifications (none mentioned in the question), **E. coli** would be the preferred cell system, although other systems can also be tested.

b) 280 kDa multi-domain protein of human origin, residing in cytoplasm (example: mTOR kinase)
Answer: In theory, this protein can be expressed in any system, but given the mammalian origin and multi-domain nature of the protein, preference should be given to **mammalian or insect cells** since they have sufficiently complex folding machinery (e.g., chaperones) to produce functional protein.

c) 15 kDa disulfide-linked chemokine of mouse origin, secreted outside the cell (example: Interleukin 11)
Answer: Given the small size, need for secretion, and disulfide bond formation, **yeast** would represent an ideal system for the production of this protein. E. coli cannot be used due to the lack of optimal secretory machinery and issues with formation of correct disulfide bond network. Insect cells and mammalian cells can also be applied for the production of this protein, given the biochemical properties and mouse origin.

d) 249 kDa multidomain, N-glycosylated, transmembrane protein of human origin (example: integrin)
Answer: Given the relatively large size, transmembrane localization and need for glycosylation, this protein would be ideally suited for **mammalian cells**. Insect cells can also be considered for this expression, although N-linked glycosylation is slightly different compared to mammalian cells, which could be problematic for some proteins (very rare). Yeast system may work in theory, but the multidomain nature of this protein would present a considerable challenge for these relatively simple cells. Still, many complex transmembrane proteins are produced in yeast. E. coli does not have the glycosylation machinery so it cannot be applied here.

e) 194 MDa virus of bacterial origin (example: T4 bacteriophage)
Answer: Viruses are most commonly produced in cells that match the origin species since they have all the necessary cellular machinery to correctly produce and assemble the virus. So in this case you should select **bacterial cells (e.g., E. coli)**. Similarly, mammalian or insect viruses should be produced in mammalian or insect cells from tissues matching those at which the viruses replicate (e.g., lung cell lines for respiratory viruses).

Question 4: Cell culture contamination

When working with mammalian cells (such as CHO, HEK), the work is always conducted in sterile environments (biological safety cabinets) to avoid any contamination with bacteria or yeast.

a) Why is bacterial contamination such a big problem when growing mammalian cells? Discuss as many potential issues as you can think of.

Answer:
Bacterial contamination is a major problem because bacteria grow much faster than mammalian cells, quickly consuming nutrients and producing toxic byproducts. These toxins and waste products can kill or stress the mammalian cells, while bacteria can also physically disrupt cell layers. Even a small contamination can invalidate experiments and ruin entire cultures.

b) Imagine that you were growing CHO (Chinese Hamster Ovary) cells in a 2 L bioreactor. The cell density is 1 x 10^6 cells/ml and you just provided them with fresh media for optimal growth. Accidentally, you introduce 1 bacterial cell (E coli) into this cell culture. Assuming the doubling rates of 24 hours for CHO cells and 0.5h for bacterial cells under these conditions, estimate the length of time that will take bacteria to match the total concentration of CHO cells in this flask?

Answer:
Here, we need to use the growth equation from class:

$$N_{(t)} = N_0 \times 2^{t/d}$$

CHO and bacterial cells will grow at different speeds:

- For CHO cells: $N_{(t)}\ (CHO) = N_0\ (CHO) \times 2^{t\,/\,d(CHO)}$
- For bacterial cells: $N_{(t)}\ (E\ coli) = N_0\ (E\ coli) \times 2^{t\,/\,d(E\ coli)}$

We are looking for the time (t) when:

$$N_{(t)}\ (CHO) = N_{(t)}\ (E\ coli)$$

Which gives:

$$1\ \text{cell} \times 2^{t\,/\,0.5} = 2000\ \text{mL} \times 10^6\ \text{cells/mL} \times 2^{t\,/\,24}$$

Apply base 2 logarithms on both sides (ignore “cell” since it is not really a physical unit):

$$t\,/\,0.5 = \log_2 (2\text{x}10^9) + t\,/\,24$$

And solve for t:

$$t = 30.897\,/\,1.9583 = \mathbf{15.77\ hours}$$

It will take less than one day before E coli matches (and ultimately overgrows) CHO cells.

Question 5: Select optimal columns for biomolecule purification

What columns or chromatography techniques would you need to complete the following biomolecule purification experiments:

a. Separate two proteins based on differing molecular weights
Answer: Size-exclusion (or gel filtration) column (e.g., Superose-based resin)

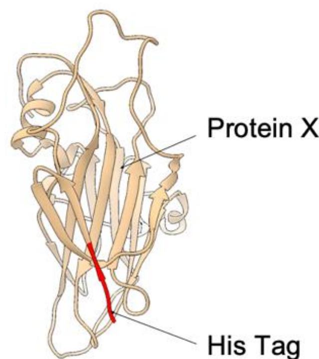
b. Separate a mixture of lipids based on differing hydrophobic properties
Answer: Hydrophobic interaction chromatography (e.g., silica column)

c. Separate a mixture of DNA molecules that have different lengths and resulting negative charge (e.g., 20, 40 and 60 nucleotide base pairs)
Answer: Anion-exchange chromatography (e.g., DEAE matrix), or size-exclusion

d. Purify a strep-tagged protein from other cellular biomolecules
Answer: Affinity chromatography using streptavidin-functionalized resins

Question 6: Protein X purification

You would like to purify a human intracellular protein (Protein X) which has a molecular weight of 40 kDa and is associated to several diseases. Based on sequence analysis and AlphaFold structure prediction, you conclude that it consists of a single domain and is not expected to undergo any post-translational modifications. Further, the protein does not have any cysteine residues that can form disulfide bridges. You added a poly-histidine tag to the C-terminus of the protein to facilitate purification.



a. Which expression system would you first test for this protein and why?

Answer: In theory, all expression systems we mentioned in class can be used to produce this protein. However, given the low molecular weight, single-domain nature and absence of predicted post-translational modifications, the first choice should be **E. coli** as it is the simplest, cheapest and most versatile system.

b. Following the expression in the host chosen previously, you continue with affinity chromatography in order to purify the protein. Describe the specific type of affinity purification you'll employ and identify which part of the protein is required for this method. Additionally, explain the working principle of this technique and specify the type of column you'll be using.

Answer: His-tag Affinity chromatography purification. For this method a column containing Ni-NTA groups conjugated to a resin will bind to groups enriched in histidine residues (i.e., his-tag). The proteins that do not contain a his-tag will wash through the column while the protein of interest, which contains a his-tag, will stick to the column. It will be eluted in a next step with imidazole which out-competes and displaces the his-tags bound on the resin.

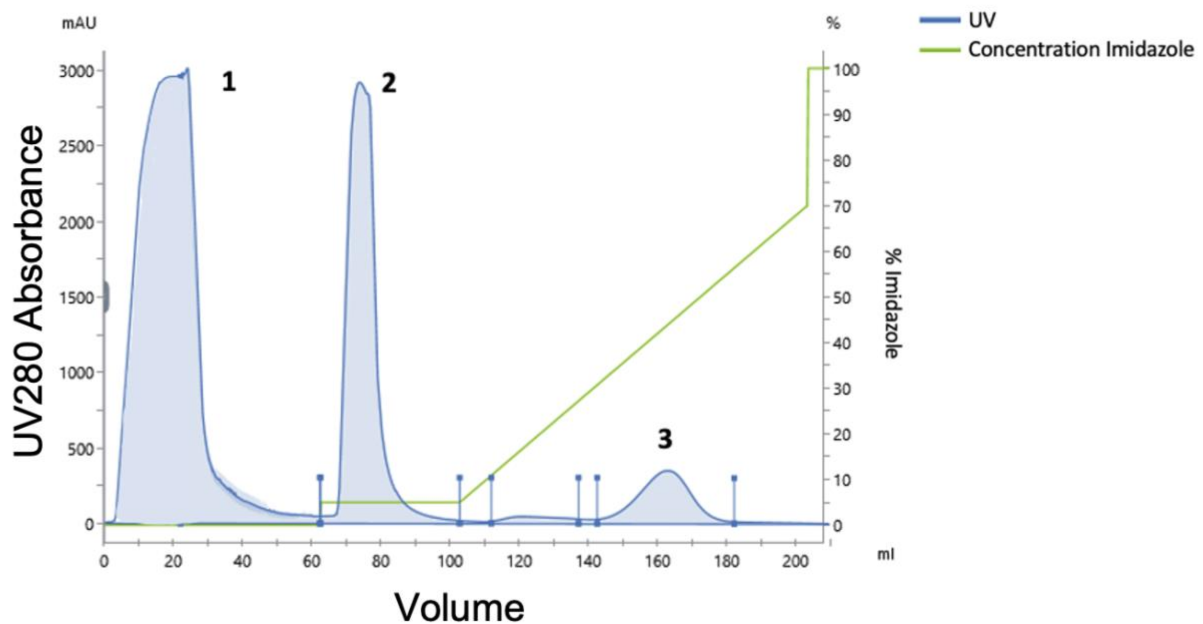
c. Select the minimal set of necessary buffer components that should be included in the Elution Buffer for the his-tag purification step?

Answer:

Reagent	Yes	No	Why?
20mM Tris-HCl	X		Buffer capacity to maintain the pH
0.1% Triton-X		X	This is a detergent. Not needed for this.
150mM NaCl	X		Salt to mimic physiological conditions.
5mM DTT		X	Reducing agent. Not needed in this case due to the lack of disulfides in the protein

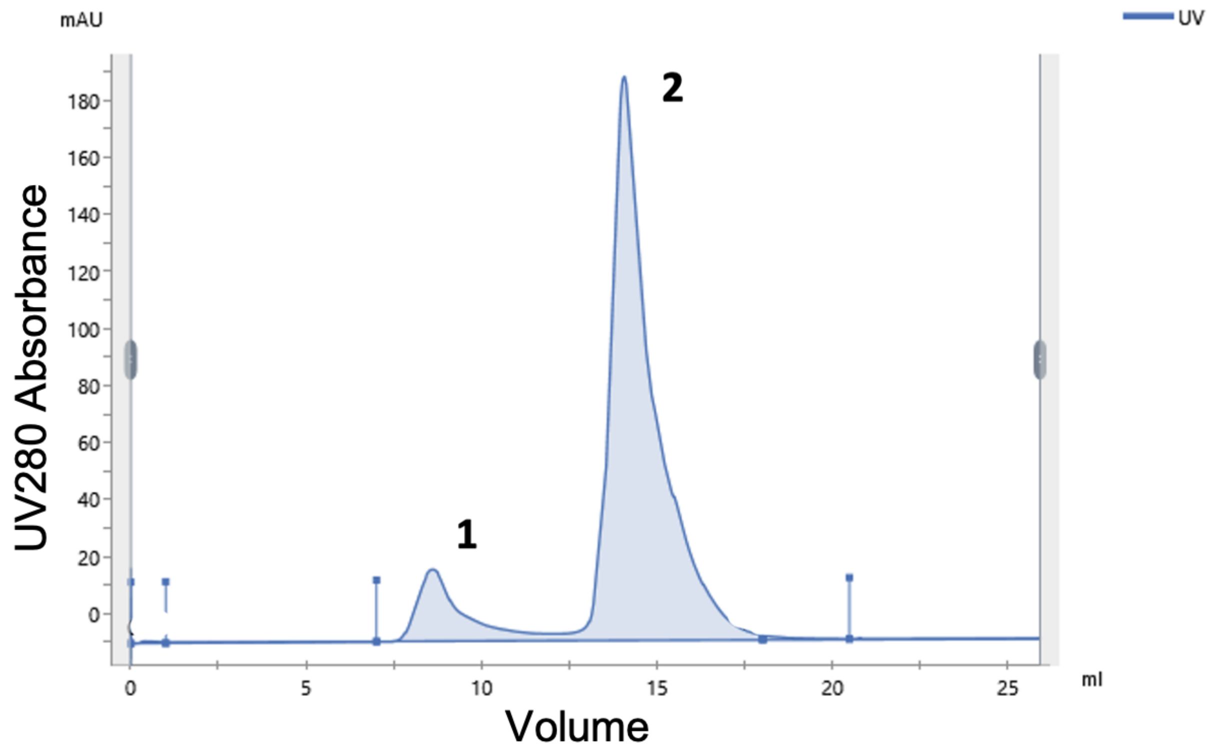
10mM Ribose		X	Carbohydrates can stabilize certain proteins. There is not enough information to justify its use here.
20% Glycerol		X	Organic solvents can help to prevent protein aggregation. Insufficient information to justify its use here.
0.1mM Imidazole		X	Imidazole is the eluting agent, but it needs to be at a much higher concentration to have effect.
500mM Imidazole	X		Needed to elute the protein from the Ni-NTA resin

d) For the His-Purification you obtain the following chromatogram. Label the peaks 1, 2 and 3 and explain what proteins are in each peak. In blue you can follow the UV absorbance of the protein at 280 nm and in green the concentration of imidazole.



Answer:
Peak 1: Protein flow through → Cellular proteins that do not bind to the His-column.
Peak 2: Cellular proteins that bind the nickel-column with low affinity probably because they contain some exposed histidine residues. They are eluted with a low concentration of imidazole (5% of the 500mM imidazole stock). This represents a wash step.
Peak 3: Protein of interest eluted using 30-60% of the 500mM imidazole stock solution. This represents an elution step.

e) You suspect that your protein of interest may be partially aggregated after the affinity purification. This could lead to problems with downstream structural and functional analyses. So you decide to separate the monomeric and aggregated forms of the protein using Size-Exclusion-Chromatography. How does size-exclusion chromatography work? Looking at the chromatogram, where is the aggregated and where is the soluble protein?



Answer: Size exclusion chromatography is used to separate proteins by size. The resin of the column consists of spherical beads which contain pores of a specific size distribution. Smaller proteins diffuse into the pores and therefore the passage through the column takes longer (higher elution volume). Larger proteins diffuse less into the pores and therefore elute faster (lower elution volume). Consequently, the proteins are separated by size with the large proteins eluting first and the small one last.

Aggregated proteins have a larger radius and a higher molecular weight and therefore they travel faster through the SEC column. The aggregated protein is therefore in the peak 1 and the folded protein of interest in peak 2.

Question 7: Isoelectric points and Ion-exchange chromatography

You are analyzing a mixture of two synthetic peptides produced by solid-phase synthesis:

Peptide A: **WDGDAKRVTSA**
Peptide B: **SEDDVKKKAGL**

Assume that the peptide is under standard aqueous conditions and the following pKa values apply:

Group	pKa
C-term (α–COO^-)	2.0
N-term (α–NH_3^+)	9.0
Aspartate side chain	4.0
Glutamate side chain	4.1
Lysine side chain	10.8
Arginine side chain	12.5

a) Define what charged groups will exist in each peptide at pH 7.

Answer:

Rule of thumb: side chains with pKa ≪ pH are deprotonated; with pKa ≫ pH are protonated. From the text we have **pH = 7**.

Peptide A (WDGDAKRVTSA)
- N-terminus (pKa 9.0): protonated → +1
- C-terminus (pKa 2.0): deprotonated → −1
- Asp (pKa 4.0) ×2: deprotonated → −1 each
- Lys (10.8): protonated → +1
- Arg (12.5): protonated → +1

So at pH 7 the charged groups are: N-term(+1), C-term(−1), Asp(−1, −1), Lys(+1), Arg(+1).

Peptide B (SEDDVKKKAGL)
- N-terminus (9.0): +1
- C-terminus (2.0): −1
- Glu (4.1): −1
- Asp (4.0) ×2: −1 each
- Lys ×3: +1 each

So charged groups: N-term(+1), C-term(−1), Glu(−1), Asp(−1, −1), Lys(+1, +1, +1).

b) Based on the analysis of charged groups in a) and their charge status you conclude that both peptides will carry net neutral charge at approximately pH 7. For each peptide, determine the bracket pKa values flanking pH 7? Use the bracket pKa values to calculate the exact pI value for each peptide?

Answer:

Peptide A: Ordered of relevant pKa values with bracket pKa values flanking pH 7 (and resulting in net charge = 0) shown in yellow:

2.0 (C-term) → 4.0 (D) → **4.0 (D) → 9.0 (N-term)** → 10.8 (Lys) → 12.5 (Arg).

The corresponding pI value is: **pI (A) = (4.0 + 9.0) / 2 = 6.50**

Peptide B: Ordered of relevant pKa values with bracket pKa values flanking pH 7 (and resulting in net charge = 0) shown in yellow:

2.0 (C-term) → 4.0 (D) → 4.0 (D) → **4.1 (E) → 9.0 (N-term)** → 10.8 ×3 (K)

The corresponding pI value is: **pI (B) = (4.1 + 9.0) / 2 = 6.55**

So both peptides have very similar pI value and will be at net neutral charge at pH 6.50 and pH 6.55 respectively.

c) What would be the net charge of each peptide at pH 3.0? Hint: Use lecture 5 to help you determine what happens with each charged groups at different pH.

Answer:

Rule of thumb: side chains with pKa ≪ pH are deprotonated; with pKa ≫ pH are protonated.

For acidic groups (e.g., R-COOH):

- Protonated state = COOH → Net charge = 0
- De-protonated state = COO^- → Net charge = -1

For basic groups (e.g., $R\text{-}NH_2$):

- Protonated state = NH_3^+ → Net charge = +1
- De-protonated state = NH_2 → Net charge = 0

Peptide A (WDGDAKRVTSA) – at pH 3.0:

- N-term (pKa 9.0) → **+1**
- C-term (pKa 2.0) → **-1**
- D (pKa 4.0) ×2 → **0** each
- K (pKa 10.8) → **+1**
- R (pKa 12.5) → **+1**

Sum ≈ +1 −1 +0 +0 +1 +1 **≈ +2** (more precisely ≈ +2.1 using fractional ionizations).

Peptide B (SEDDVKKKAGL) – at pH 3.0:

- N-term (pKa 9.0) → **+1**
- C-term (pKa 2.0) → **−1**
- E (pKa 4.1) → pH 3 < 4.1 → mostly **0**
- D (pKa 4.0) ×2 → **0** for both
- K (pKa 10.8) ×3 → **+3** for all three

Sum ≈ +1 −1 +0 +0 +3 = **≈ +3** (≈ +3.1 with fractional ionization).

Answer (c): Peptide A ≈ **+2**; Peptide B ≈ **+3** at pH 3.0.

d) If you wanted to immobilize the peptides to an ion-exchange column at pH 3.0, what would be the prefered type of column?

Answer:
At pH 3.0 both peptides are positively charged → they will bind to a cation-exchange column.

e) You proceeded with peptide immobilization at pH 3.0 using the selected column. Now both peptides are bound and you are trying to select an elution strategy that would allow to separate them. The decision is between using either (i) pH or (ii) salt gradient. Which approach would allow you to separate these peptides into distinct peaks? In both cases, discuss the order at which the proteins will elute.

Answer:

(i) **pH gradient elution:** peptides elute at their pI value. Because the two peptides have very similar pI values (≈6.50 vs 6.55), a pH gradient would make them elute at essentially the same pH → co-elution. So this would not allow to separate the two peptides.

(ii) **Salt gradient elution:** Increasing ionic strength (e.g., NaCl gradient) competes with the positively charged peptides for resin binding and elutes them according to binding strength / charge density. Since at pH 3 the peptides have different net charges (+2 vs +3) and different charge densities, they will elute at different salt concentrations (Peptide B, being more positively charged, will elute later with higher salt).

Question 8: Separating proteins by charge

You are working with a Protein Y with a molecular weight of 50 kDa that has a pI of 5.5. After the initial affinity purification you realize that the protein is contaminated with other protein species, so you decide to increase the purity of your protein by separating it based on charge, using ion-exchange chromatography.

Protein Y Contaminant

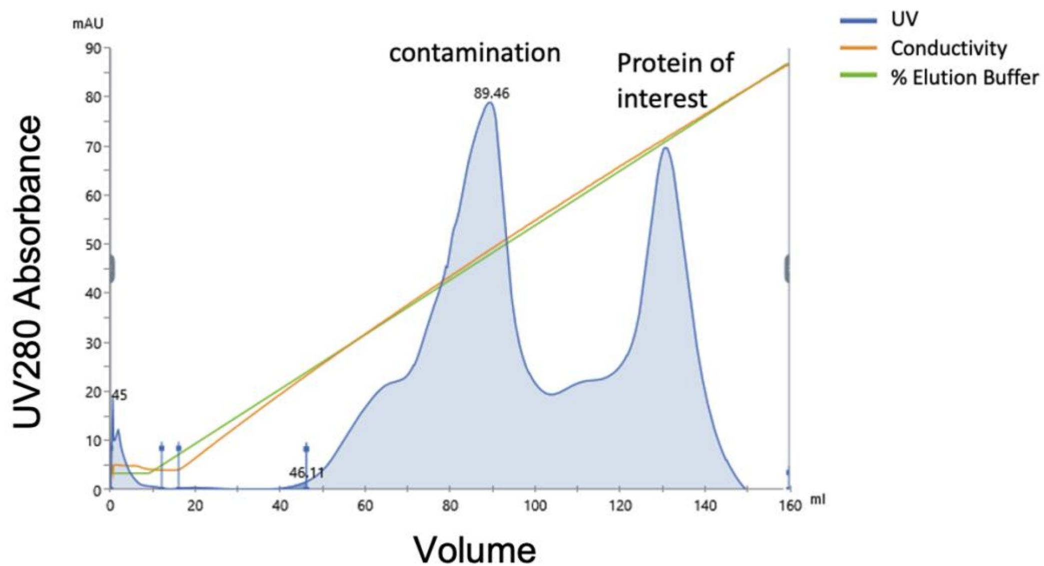
\+ -

Coulombic potential

a) Using a buffer of ph=7.5 what is the net charge of the protein? What subtype of column would you use to bind it?
Answer:
The pH of the buffer is higher than the pI of the protein, therefore it is negatively charged. Hence, you should use an anion exchange column to bind the protein as the resin of the column is positively charged and binds negatively charged proteins.

b) What are the different ways that you can use elute proteins from this type of column?
Answer:
Proteins can be eluted from an ion-exchange column either by:
(1) Gradually changing the pH of elution buffer – The protein will elute once the pH of the buffer drops below the pI value of the protein, which is 5.5.
(2) Gradually increasing salt (i.e., NaCl) concentration in the elution buffer – The proteins with elute depending on the ionic interactions formed with the resin. The stronger the ionic interaction -> the higher the concentration of salt needed to displace the protein

c) You elute the protein from the column using a linear gradient of NaCl and get the following chromatogram. In the chromatogram you follow the protein by looking at its absorption at 280 nm (UV). By increasing the relative percentage of the "elution buffer" (green) you slowly increase the salt (NaCl) concentration loaded onto the column which triggers the elution of the protein. You can follow the increase in salt concentration by looking at the increase in conductivity (orange). You observe that the contaminating proteins elute before your protein of interest. Why is that the case? What does the position of the peak depend on when eluting with salt?



Answer:

The location of the peak will depend on the relative strength of interactions formed between each protein species and the anion-exchange resin, which will be a function of the content of positively and negatively charged groups in the protein at the buffer pH (in this case 7.5), the relative sizes of each protein species (i.e., total number of amino-acids) and even their 3D structures. The stronger the ionic interaction -> the higher the concentration of salt need to displace the protein.

In this case the contaminating proteins form weaker interactions with the resin compared to the protein of interest, which is why they elute at lower salt concentrations (lower % of elution buffer).