Lecture 9
Biomolecule Characterization

Question 1: Electromagnetic radiation and matter

a) UV/Vis spectrometry relies on electromagnetic radiation with energy that corresponds to:
- Transitions between vibrational levels of chemical bonds
- Interference with nuclear energy levels
- Transitions between electron energy levels
- Molecular diffusion states

b) Select a statement that is **false**:
- Differential scanning calorimetry is a spectroscopic method
- Visible light has higher energy compared to microwave radiation
- Radio-frequent radiation is used in resonance experiments
- UV absorbance increases with the content of conjugated double bonds

c) Circular dichroism is based on:
- Differential scattering of left- and right- polarized UV radiation
- Differential absorption of non-polarized UV radiation
- Differential absorption of left- and right- polarized UV radiation
- Dynamic motion (vibrations) of secondary structure element

d) Select a statement that is **true**:
- Chromophores display stronger UV/Vis absorbance compared to fluorophores
- Fluorophores have larger optically-active groups compared to chromophores
- Fluorophores emit radiation of lower energy compared to the absorbed radiation.
- Fluorophores emit radiation of lower wavelength compared to the absorbed radiation.

e) Stokes-Einstein equation is used in which method:
- Circular dichroism
- Differential Scanning Fluorimetry
- X-ray crystallography
- Dynamic light scattering

f) Select a statement that is **true**:
- The same chemical groups in proteins are used to measure UV absorbance and DSF.
- Unlike proteins, nucleic acids do not have an extinction coefficient at 280nm.
- The presence of double bonds in carbohydrates makes them detectable by UV.
- All amino-acids are detectable by circular dichroism.

Question 2: Select a biophysical method for your experiments

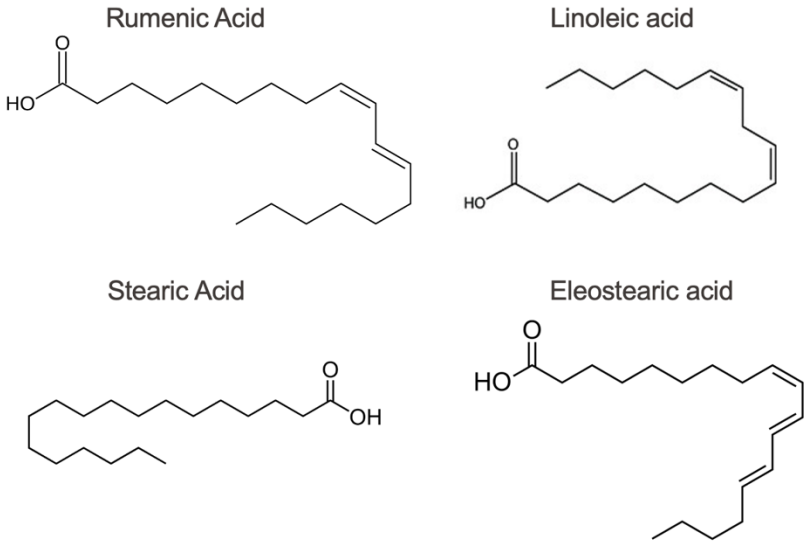
You produced a bacterial protein that has not been characterized previously. After performing extensive chromatographic purification step you wish to learn more about this protein. Select the biophysical methods from lectures 8 and 9 that you can use to measure the below-mentioned properties:

Inspect protein purity	SDS-PAGE
Determination of molecular radius in native state	DLS
Determine the content of secondary structure elements	CD, FT-IR
Determination of molecular weight in denatured state	SDS-PAGE
Measure concentration	UV Spectrometry
Measure protein stability to thermal denaturation	CD, DSF (or DSC), FT-IR
Determine atomic structures	NMR, XRC, cryoEM
Study protein aggregation	DLS (also cryoEM – Lecture 11)
Protein localization inside the cell	Fluorescent microscopy

The method options are:
- Dynamic Light Scattering (DLS)
- Sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE)
- Circular dichroism (CD)
- UV spectrometry/spectroscopy
- Differential Scanning Fluorimetry (DSF)
- Fluorescent microscopy
- Cryo-electron microscopy (cryoEM)
- Nuclear Magnetic Resonance (NMR)
- X-ray crystallography (XRC)
- Fourier-Transform Infrared Spectroscopy (FT-IR)

Question 3: UV absorbance

a) Below are structures of several fatty acids containing different content of double-bonds, resulting in different minimal energy difference between π and π* levels. Order the molecules based on the location of their maximum UV absorbance wavelength (λ_{max}), from highest to lowest.



b) Below are the sequences of 3 different short proteins (peptides). Based on the content of UV-active amino acids in each peptide, can you rank them based on the expected extinction coefficient at 280nm (from highest to lowest)?

Peptide M: MSTRKRVVNTSPLGNEDPKNYQMILAPRESNQGNKILSSGQEDPRTN
Peptide N: MTGSTTVNGTWADTIEGKQGRSDGDSYILSSASGSDRTYESNMPSL
Peptide O: MNGRGAGALMYRSDKDAKNGYLANVDAKDLVKKENGAASKLMIPNT

c) You wish to determine the concentration of Protein X in aqueous buffer. Its' molecular weight is 67kDa and the molar extinction coefficient at 280nm wavelength (ε_{280}) is 43'824 L/(mol*cm). Using a cuvette with a light path (l) of 1cm, you measured the absorbance at 280nm to be 3.5. Based on these values can you determine the molar concentration of this sample (mol/L). What is the mass concentration (g/L)?

Answer:

a) The greater the number of conjugated double bonds -> the higher the λ_{max}
- Eleostearic acid - λ_{max} = ~280nm (3 conjugated double bonds)
- Rumenic acid - λ_{max} = ~235nm (2 conjugated double bonds)
- Linoleic acid - λ_{max} = ~215nm (double bonds are separated by 2 single bonds, resulting in weak conjugation)
- Stearic acid - λ_{max} = ~200nm (no double bonds)

b) Given that F, H and C are absent from the polypeptide sequences, W and Y are the only amino acids that would determine the extinction coefficient. The greater the number of these amino-acids -> The higher the extinction coefficient. So, the correct order is:
- Peptide N: 1 W + 2 Y
- Peptide O: 2 Y
- Peptide M: 1 Y

c) First, we use Lambert-Beer's law to determine the molar concentration (C_M):
$A = \varepsilon_{280} * l * C$

$C_M = A / (\varepsilon_{280} * l) = 3.5 / (43'824 * 1)$ mol/L = $79.9 * 10^{-6}$ mol/L = **79.9µM**

Then we use this value together with the molecular weight information to calculate the mass concentration (C_g):

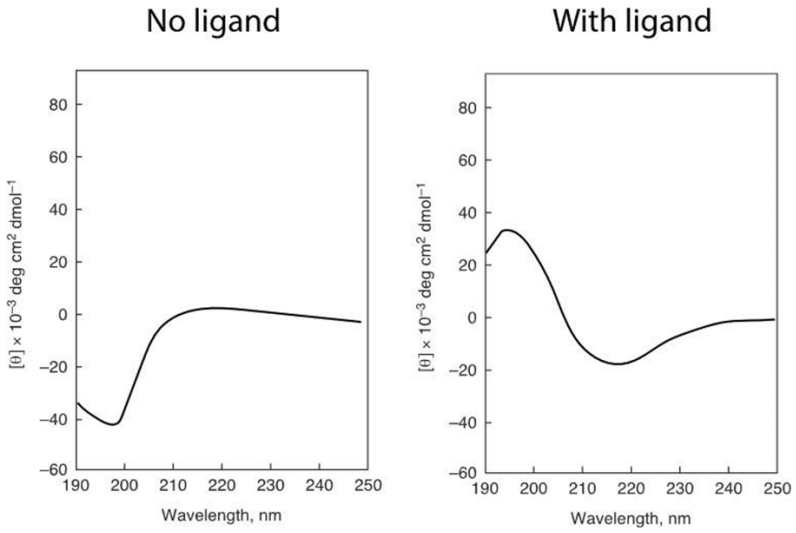
MW = 67kDa = 67'000 g/mol

$C_g = MW * C_M$ = 67'000 g/mol * $79.9*10^{-6}$ mol/L = **5.35 g/L**

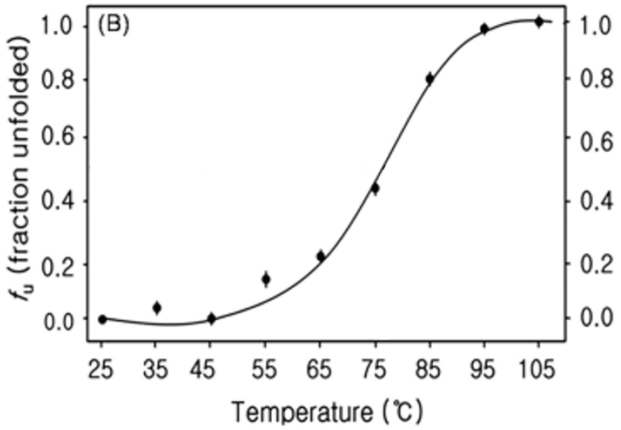
Question 4: Circular dichroism

You expressed and purified a bacterial carbohydrate-binding protein Z. This protein can detect galactose and switch on downstream signaling pathways to metabolize this sugar. You are interested in uncovering the molecular processes that regulate this signaling and discover a way to exploit it for therapeutic purposes.

a) You performed circular dichroism (CD) on protein Z in the presence and absence of the ligand (at 25˚C). Please describe what type of an effect the ligand has on protein? Based on comparison to the reference CD spectra shown on lecture slides can you elaborate on the specific details of this change?



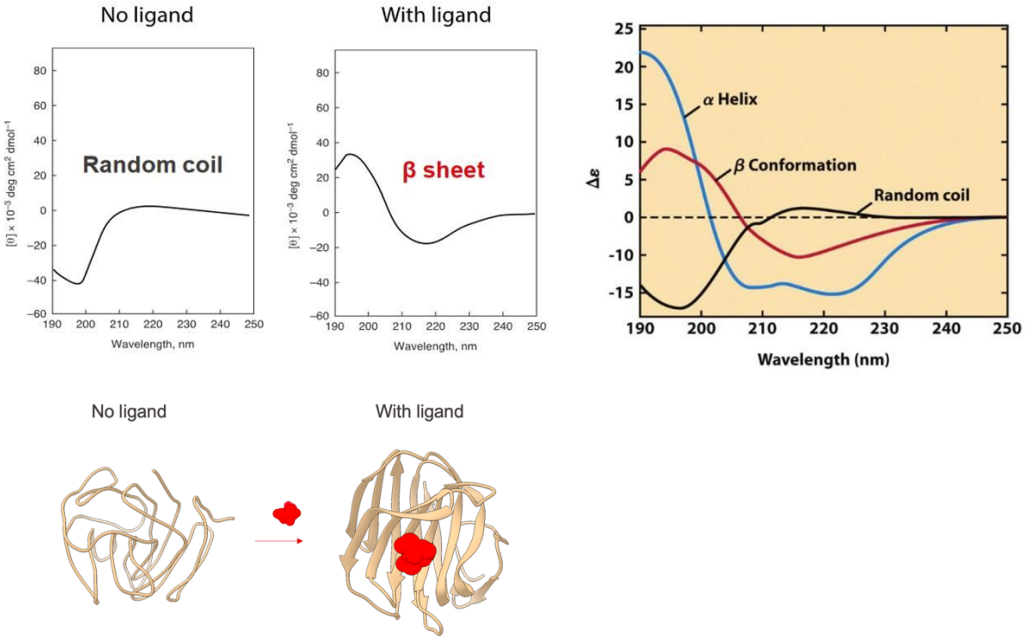
b) You wanted to measure the thermal stability of protein Z in the presence of the ligand using CD and you obtain the graph shown below. Can you estimate the Tm value from this plot? What would the raw CD curve (ellipticity vs wavelength) look like at temperatures >95˚C?



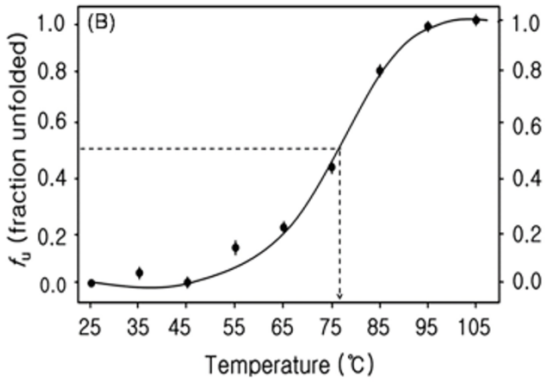
c) You are interested in developing a new antibiotic candidate that will inhibit this critical galactose detection mechanism and block downstream metabolism. You have a library of potential compounds to screen from. How would you setup the CD assay to determine which of these candidate compounds can block the activity of this protein?

Answer:

a) The ligand causes protein Z to take on a **structured** 3D state, converting from random coil (unstructured) architecture into a beta-sheet based architecture.

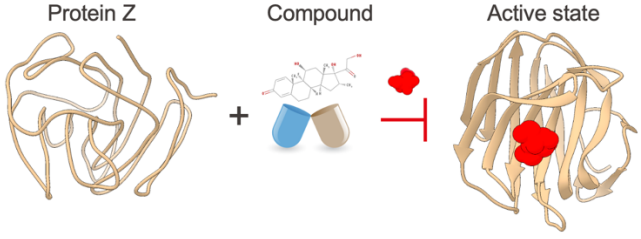


b) The Tm value is calculated as the temperature where 50% of the protein (0.5 fraction) is in the unfolded state. Therefore, Tm estimated based on this curve is ~76-77˚C.



At temperatures >95˚C most of protein Z will be in unfolded state and the raw CD spectra will show random-coil assembly, very similar to the one that the protein Z has in the ligand-free state shown in panel a) of this question.

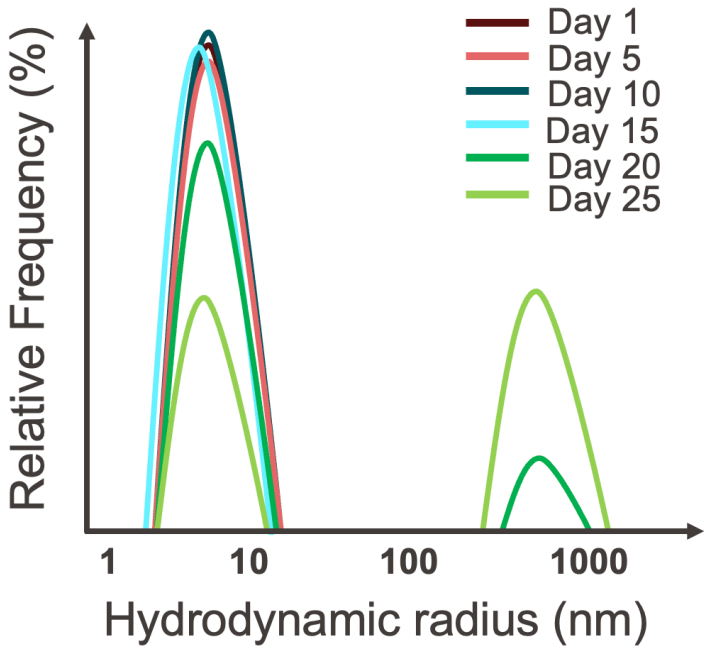
c) While CD is a relatively costly experiment in terms of sample requirements and throughput, it can certainly be used to evaluate inhibition by compounds. In this case you can screen by adding the candidate compounds with protein Z and then adding galactose to the mix. By recording the CD spectra (at 25˚C) you will be able to evaluate if the compound prevents transition from unfolded (random-coil) to the active beta-sheet state. If the compound really inhibits the process, the transition should not take place and all protein will stay in random coil state. Several other possibilities can also take place.



Question 5: Light Scattering

Dynamic and multi-angle light scattering methods are particularly useful in pharmaceutical industry for monitoring the quality and homogeneity of different therapeutic formulations.

a) Below you will see a DLS size-distribution curve of a monoclonal antibody cocktail after storage at room temperature for different periods of time. Can you infer what could be happening to the sample over time? What would be the max storage time under these conditions to assure the sample is still effective for therapeutic use?



b) You propose that one way to preserve the sample is to store it at lower temperatures (e.g., 4˚C). The viscosity of aqueous solutions is 2x higher under these conditions. How will this affect the diffusion coefficient of the monoclonal antibody? What about the hydrodynamic radius? Qualitative descriptions are sufficient.

c) Calculate the r_h for this monoclonal antibody in storage solution at 4 ˚C (T = 277 K). The diffusion coefficient (D) is $2.1*10^{-11}$ m^2/s and the viscosity (η) of the storage solution at this temperature is $1.6*10^{-3}$ Pa*s (= $J*s/m^3$). Assume that the Boltzmann constant (k_b) is $1.38*10^{-23}$ J/K, and π=3.14

Answers:

a) The monoclonal antibody sample is aggregating over time which is reflected by the formation of a second species with diameter between 500-1000nm. The last time point at which no aggregation is observed is at Day 15. Therefore, the max storage time without any visible signs of issues (at least by DLS) is 15 days.

b) Consider Stokes-Einstein equation:

$$D = \frac{k_B T}{6\pi \eta r_h}$$

Given that the radius of hydration is the property of a given protein dependent on its size, shape and density, the change in temperature will have minimal influence (if any) on this feature. It can still happen that temperature triggers a conformational change that would consequently change the r_h, but this is rare.

However, due to the two-fold increase in viscosity and a reduction in temperature, the net effect will be a decrease in the diffusion coefficient. This is relatively intuitive since (i) lower temperature will reduce the speed at which particles move in solution (less thermal energy), and (ii) increased viscosity will result in greater liquid resistance to particle movement. Altogether, both effects add to reduced diffusion.

c) If we solve Stokes-Einstein equation for r_h we get:

$$rh = \frac{\mathrm{kB} * \mathrm{T}}{6 * \pi * \eta * \mathrm{D}}$$

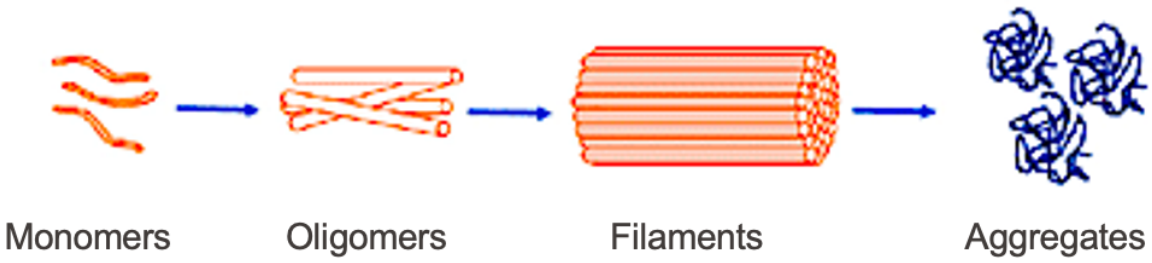
By plugging in the numbers, we get:

$r_h = (1.38*10^{-23}$ J/K * 277 K) / (6 * 3.14 * $1.6*10^{-3}$ Js/m3 * 2.1 *10^{-11} m^2/s)

rh = $382.3*10^{-23}$ / $63.3*10^{-14}$ m = **$6.0*10^{-9}$m = 6.0nm**

Question 6: Protein aggregation

Aggregation of certain cellular proteins (e.g., alpha-synuclein) is believed to be the primary causative factor of neurodegenerative disorders. You discovered one protein of only 80 amino-acids (MW = 9.5kDa) that resembles alpha-synuclein, and you suspect that it may undergo similar aggregation process upon triggering with chemical stimuli (e.g., ligand or low pH).



a) Which methods would you use to determine if the protein is structured or disordered before chemical induction?

b) After getting a better understanding of the protein in its monomeric form, you now wish to study the aggregation process. You can trigger aggregating by lowering the pH of the buffer which allows to control the start of the process. Which method can you apply to study the size of the oligomers and aggregates at different time points?

c) In alpha-synuclein the transition to fully aggregated state proceeds through several intermediates (e.g., oligomers, filaments) which form regular b-sheet-enriched assembly. The transition usually takes a few hours. Which method could you use to monitor this process over time?

d) You are suspicious that the monomers undergo peptide bond cleavage and fragmentation into smaller segments, in order for the aggregation to occur. Which approach would you use to check if the fragmentation is indeed occurring?

Answers:

a) You can use several different methods (including structural biology approaches), but CD or FT-IR would be the quickest/simplest ways to determine if the protein has any secondary structure elements (structured) or if it is all random coil (unstructured/disordered).

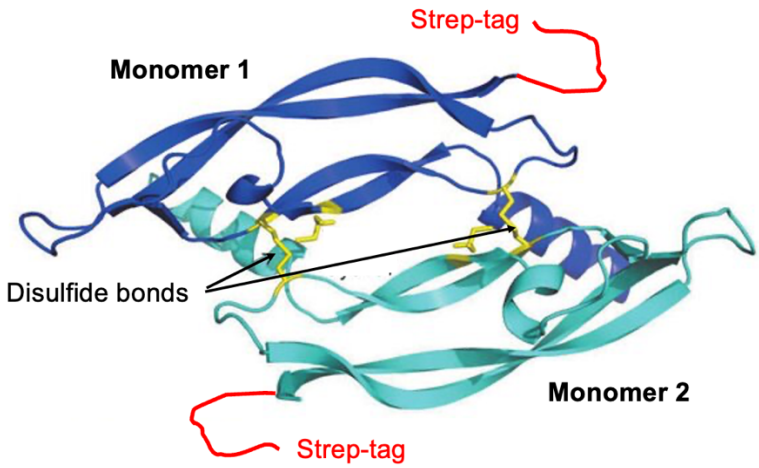
b) You can use dynamic light scattering (DLS) to determine the hydrodynamic radius of molecular clusters that are formed during the aggregation process.

c) Both CD and FT-IR can provide information on changes in secondary structure arrangement of molecules. In both cases, only seconds/minutes are needed to collect a full spectrum so you can closely monitor the reaction as it progresses.

d) SDS PAGE would be the best way to check for fragmentation. For this experiment to work, you need to denature the molecule and separate all the peptide components, which is done in SDS PAGE by default. Alternatively, mass spectrometry could also be used, but that method we didn't cover in class.

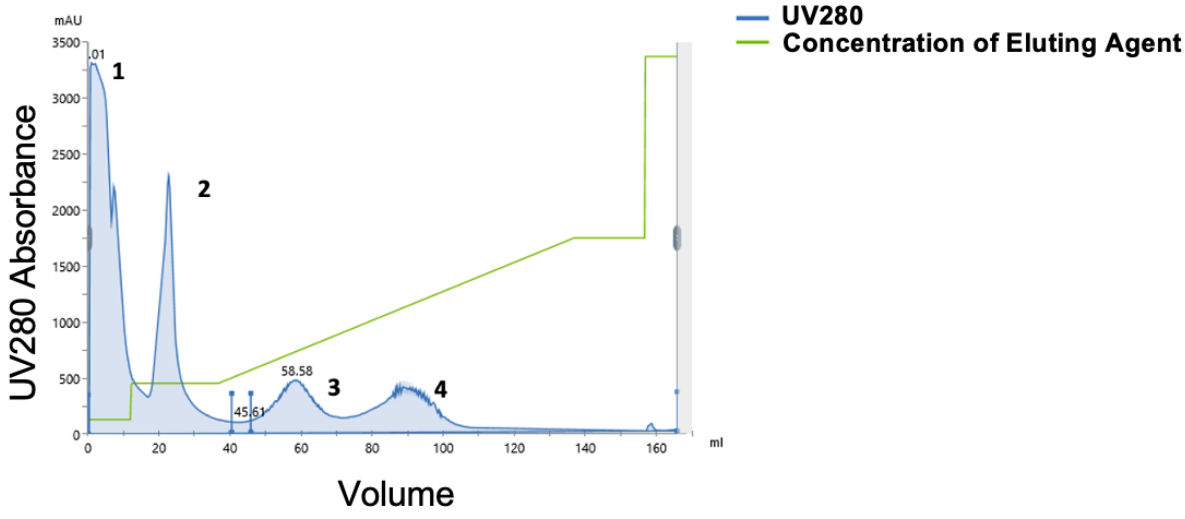
Question 7: Gel electrophoresis coupled with chromatography

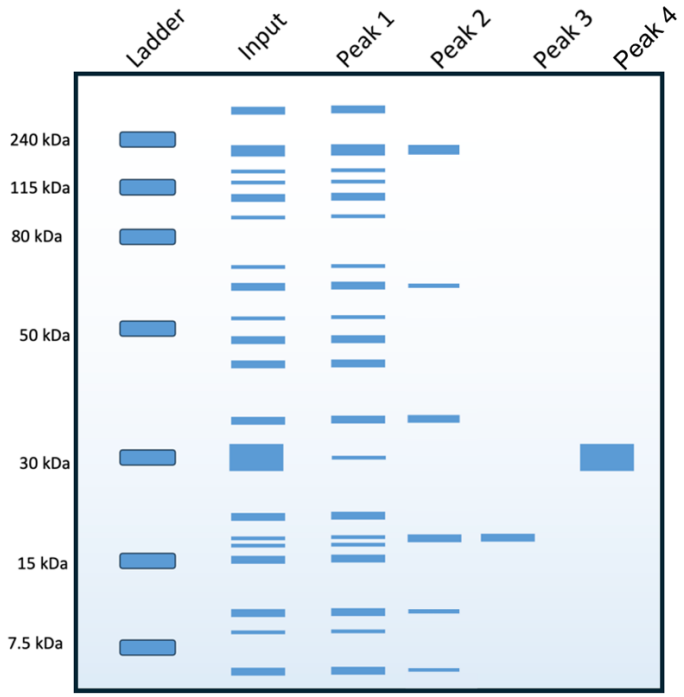
You are working with an enzyme with the predicted molecular weight of 30 kDa (based on amino-acid sequence). In nature this protein needs to dimerize in order to be in its' functional state. The dimeric state is further stabilized by 2 disulfide bonds formed between the monomers. The protein has been engineered with a C-terminal Strep-Tag and recombinantly expressed in *E. Coli.*



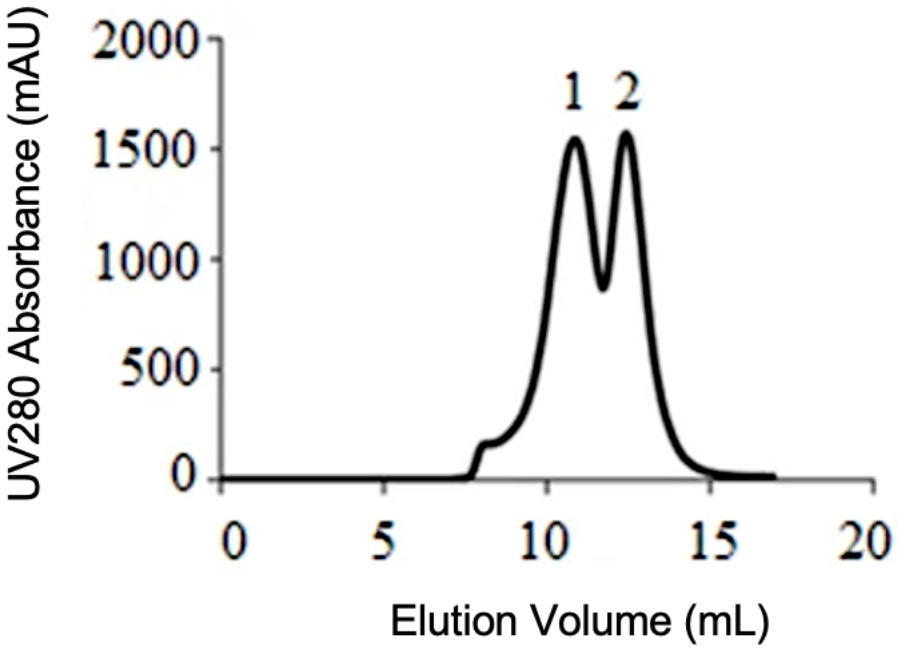
a) What type of affinity column would you use to purify this protein?

b) After purification you obtain the following chromatogram which shows more peaks than you anticipated. To determine which peak contains the pure protein you run small amounts on an SDS PAGE gel under reducing conditions (+ DTT). Which peak contains the pure protein of interest? Please explain.

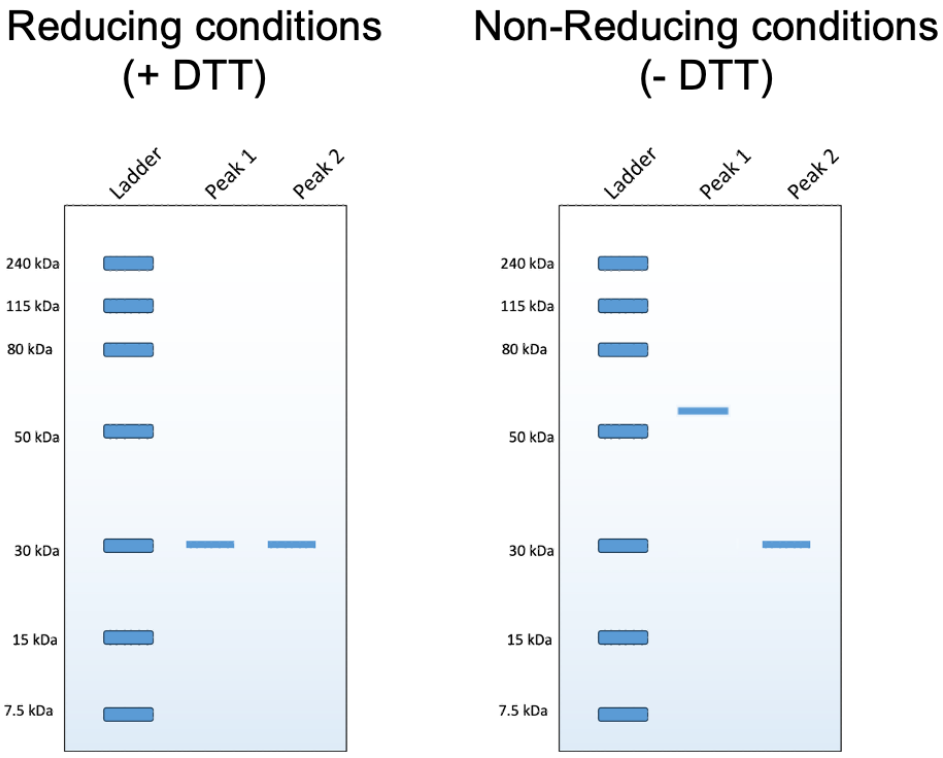




c) To separate different oligomeric states of the protein (monomer vs dimer) you perform size-exclusion chromatography. In the chromatogram you see two peaks and you get very excited because you assume that one of the peaks belongs to the dimeric active form of the enzyme. Which peak corresponds to the monomer and which peak to the dimer based on the elution volume?



d) You take a small amount of each peak and denature it under reducing (+DTT) and non-reducing (-DTT) conditions, followed by SDS PAGE analysis. The results are shown below. What can you say about the status of inter-monomeric disulfide bonds in Peak 1 and Peak 2? How do these results correlate to size-exclusion chromatography in terms of the oligomeric state?



Answers:

a) For the Strep-tag a streptavidin coated column is used. The protein is eluted with biotin.

b) Peak 4 elutes at the highest biotin concentration compared to all other peaks, and the corresponding protein runs on the SDS PAGE gel as a single clean band at molecular weight of 30kDa, which is expected for your target protein based on amino-acid sequence. The presence of DTT (reducing conditions) assures that all disulfide bridges will be reduced which is why there are no bands corresponding to dimers (~60kDa) or other oligomeric states of this protein whose existence would be dependent on disulfide bonds.
Altogether, Peak 4 appears to contain our protein of interest.

c) The dimer is in peak 1, the monomer in peak 2. The dimer elutes earlier as it is bigger and on a SEC column large proteins elute before small ones.

d) DTT is a reducing agent, and its presence during protein denaturation will disrupt disulfide bonds leading to the separation of dimers into monomers. Therefore, when DTT is present, your protein will run as a single band at ~30kDa regardless of the oligomeric state it had prior to denaturation. This is why in both cases, Peak 1 and Peak 2, we observe a single protein band corresponding to the monomeric protein.

When the reducing agent is not included, the disulfide bridges are preserved. In the case of Peak 1, we observe that the protein band is now located at 60kDa which corresponds to the dimeric state, held together by the disulfide bonds. In the case of Peak 2, the protein band is still located at 30kDa which indicates that this sample did not exist as a dimer prior to gel loading.

This is consistent with the size-exclusion chromatography in that Peak 1 appears to contain a dimeric form of your protein while Peak 2 is expected to be monomeric based on the elution volumes.