Question 1:

You are a scientist studying protein structure, function and interactions. You have access to X-ray crystallography, NMR and cryo-EM to solve some interesting problems from the world of protein biochemistry. Which of the 3 methods would you choose for each problem and why?

a) You are studying a protein receptor on cell surface (MW=150 kDa) as a potential drug target. Through cycles of drug discovery, you identified that ligand X binds to your cell receptor and exerts the desired functional effect (i.e., inhibiton of downstream signaling). Which method would you use to study the molecular details of ligand-receptor interactions?

We want to identify how a ligand interacts with the binding pocket of the protein. We can apply both X-ray crystallography and Cryo-EM on the the receptor-ligand complex. We cannot use NMR due to the size of the protein (150kDa).

b) In the pandemic time, scientists are intrigued by how the spike protein looks like on the virus particle itself in its native state. You decide to utilize the available facilities and expertise to answer that question. How would you solve the structure of the full COVID- 19 virus, knowing that it's molecular weight is approximately ~40MDa and the virus is pleomorphic (i.e., individual virions vary in size and shape)?

Cryo-EM is the best suited method for large particles like viruses, particualrly if they are pleomorphic. In the case of highly homogenous and symmetrical viral particles (e.g., poliovirus, dengue) the structural analysis can be performed using X-ray crystallography, but not for pleomorphic virions. NMR cannot be used for direct structural analysis due to the high molecular weight of the viral particle.

c) You have a protein with molecular weigh of 20 kDa, which interacts with ligand Y through a known mechanism and a well-stablished binding site (pocket). You would like to introduce point mutations in the pocket that disrupt the binding to this molecule. What structural methods could you use to assess how these mutations influence the ligand binding pocket?

Given the size of the protein and focus on protein-ligand interactions NMR is a suitable method, but you can also use X-ray crystallography. NMR may be a bit more versatile as it does not require the protein to be crystallized and can provide further information on protein stability and kinetics of binding (equilibrium constant) to the ligand by simply titrating the defined amounts of the ligand into the NMR tube with the protein. cryoEM cannot be used due the the relatively small size of the protein.

d) You are trying to solve a structure of a protein (MW ~200kDa) which aggregates in solution at concentrations above 0.5 mg/ml. What method would be the most applicable in this scenario?

We cannot use X-ray crystallography since the protein aggregates at relatively low concentrations. We need to have the protein as an ordered crystal, not a disordered aggregate. NMR is also not applicable due to the high MW. Cryo-EM will not be limited by the MW or low concentrations and is therefore the most appropriate method for this problem.

Question 2:

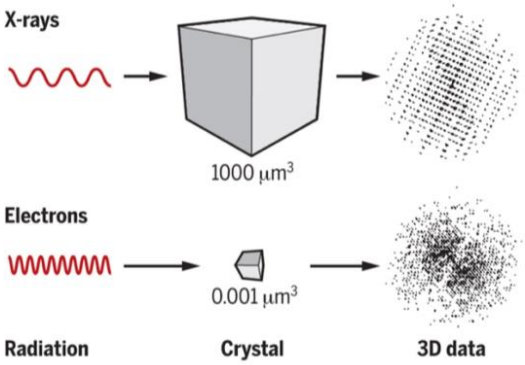
X-ray diffraction is one of the main methods for determination of macromolecular structures from crystals. However, there are some potential alternatives within the electromagnetic spectrum and within the subatomic particle space. Let's explore their applicability to this problem.

a) Microwave radiation – Can it be used in crystal diffraction experiments and allow to achieve atomic resolution? Explain.

b) Gamma radiation – Gamma (rays) have smaller wavelength compared to X-rays, and therefore a much lower diffraction limit. In theory, the maximum achievable resolution when using these rays would be superior. However they are not used for determination of biomolecular structures. Why do you think that is? Also, practically speaking, where could you perform such experiments?

c) Electrons - Can we use accelerated electrons instead of X-rays for crystal diffraction experiments? Do you think that crystal size would be an important factor for feasibility? Explain.

Answers:
a) Per Abbe's law the maximum achievable resolution when using microwave radiation in microscopy or diffraction experiments would be in the milimeter range.
b) Indeed this is theoretically feasible. However, the problem with gamma rays is that they are high frequency which means that they have high energy. Therefore, their scattering by biomolecular material would be even weaker compared to X-rays and the rate at which they would destroy the sample would be much faster. The most notable man-made sources of gamma rays are nuclear fission events, so the experiment would have to be performed in a nuclear research facility or a reactor.
c) This is possible and it is actually used by many scientists in the world. Just like in X-ray crystallography, electron diffraction involves the scattering of electrons by the biological sample in a crystalline form. The elastically scattered electrons can be used to generate a diffraction pattern, from which the structure can be calculated. For this to work, the sample needs to be crystalline in nature. The crystalline form of the protein amplifies the diffraction signal, as scattered electrons interfere with each other, increasing the signal's intensity. However, electrons interact with matter much better than X-rays and there is a major issue with inelastic and dynamical scattering (dynamical = multiple elastic scattering events while electron travels through the crystal). Therefore, the crystals used for electron diffraction need to be much smaller in size. Hence the name microcrystal electron diffraction (microED) for this method.



Question 3:

Multiple choice questions of Cryo-Electron Microscopy:

1. During EM, a vacuum is created inside the _____
    a) room of operation
    b) specimen
    c) microscope
    d) objective lens system

Answer: c) Microscope. Vacuum is needed to prevent the electrons form interacting with air molecules

2. Which of the following component of EM collects the beam of electrons from the source and directs it towards the sample?
    a) ocular lens
    b) condenser lens
    c) stage
    d) projector lens

Answer b) condenser lens collects electrons and assures that the beam is parallel to the optical axis of the instrument and directed to the sample.

3. Image formation in electron microscope is based on _____
    a) column dimensions
    b) objective aperture size
    c) differential electron scattering
    d) specimen size

Answer: c) Differential scattering of electrons from protein compared to surrounding ice is what generates contrast.

4. The biological materials have little intrinsic capability to scatter electrons compared to water/ice which is why it is essential to have _____
    a) low voltage in the microscope
    b) highly focused beam from the condenser lens
    c) thin ice layer surrounding the biomolecule
    d) very steady stage with minimal drift

Answer c) The density of protein is slightly higher compared to water, which is why thin ice surrounding the sample is necessary to have contrast.

5. _____ is the capacity to distinguish between two adjacent atoms or groups.
    a) Magnification
    b) Resolution
    c) Ionization
    d) Division

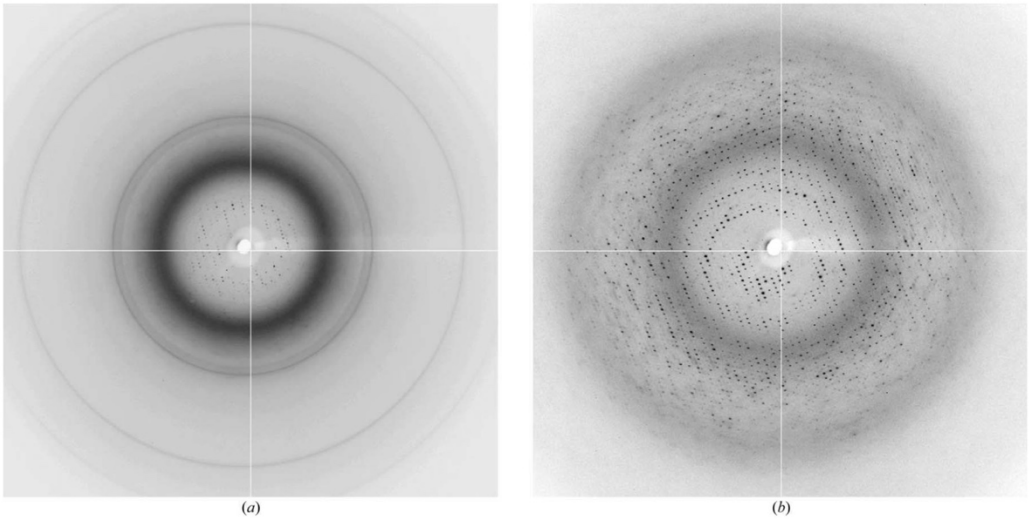
Answer: b) Resolution

6. How is resolution calculated in cryo-EM?
    a) Fourier Shell Correlation of half-sets
    b) Contrast Transfer Function
    c) R-factors
    d) Model-to-map agreement

Answer: a) The curve based on Fourier Shell Correlation of 2 half-sets is used to define the resolution in cryoEM experiments.

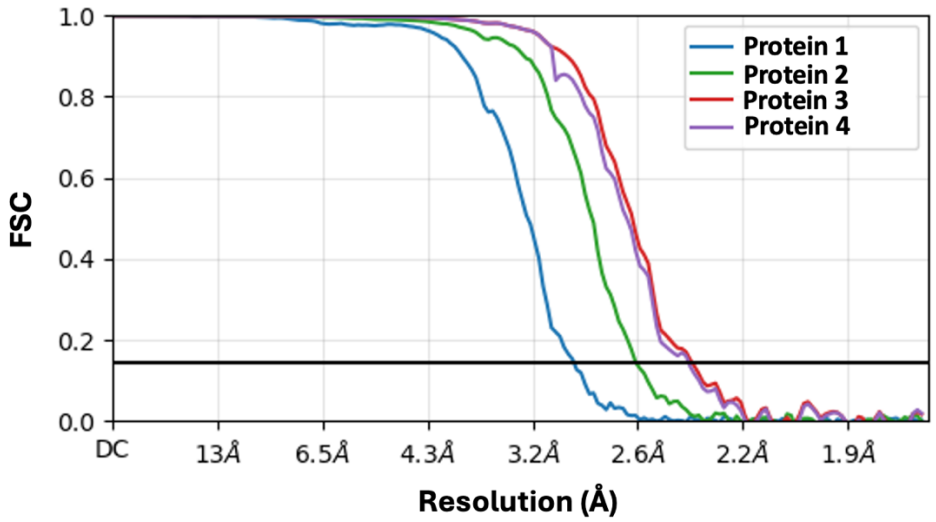
Question 4:

1) Below 2 pictures of X-ray diffraction pattern for the same protein. Which one has higher resolution? Why?



Answer:
b) has higher resolution because there are more spots that are further from the center.

2) Below is the overlay of 4 FSC curves obtained by electron mciroscopy analysis of different proteins (1-4). The gold-standard line is shown in black. From these plots, estimate the resolution of each protein reconstruction and rank the maps from highest to lowest quality (in terms of resolution)?



Answer:
Looking at the intersection of the gold-standard line and each FSC curve we come up with these numbers:

Protein 1 ~3.0Å
Protein 2 ~2.6Å
Protein 3 ~2.4Å (slightly higher resolution than Protein 4)
Protein 4 ~2.4Å

In terms of quality they are ranked form highest to lowest resolution:

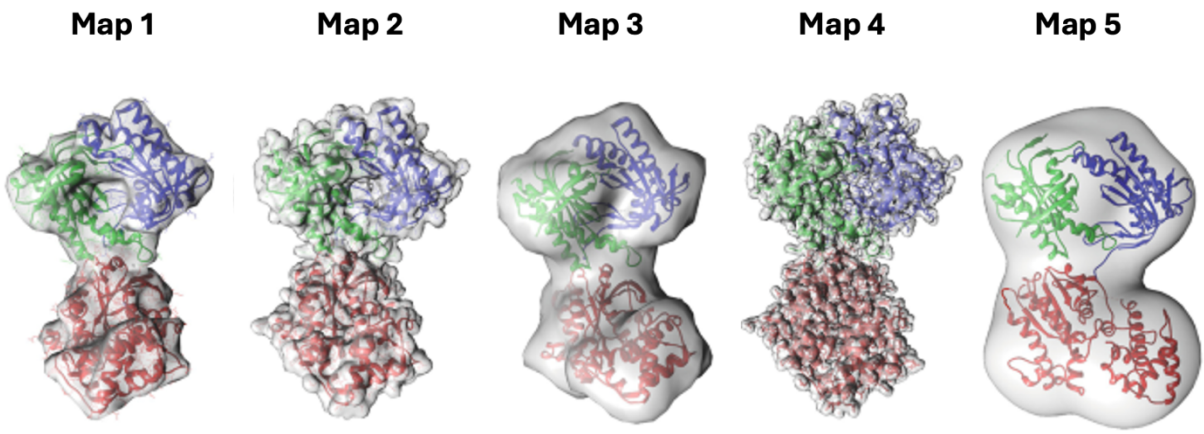
Protein 3 - Protein 4 - Protein 2 - Protein 1

Note that “higher resolution” means lower number because the amount of detail is greater.

Question 5:

Below are 5 maps of the same protein obtained by cryo-EM. The maps are presented as transparent gray surface and the corresponding model of the protein is shown with different domains colored in green, blue and red. The maps are all at different resolutions.

a) Can you rank the maps from highest to lowest resolution? Optional: Try to estimate the resolutions by comparing the level of detail in each map to the underlying protein models.
b) Which of the map(s) have sufficient quality to reliably place amino-acids into density?
c) Which of the map(s) only offer information on the level of protein domain but do not have sufficient resolution to accurately build amino-acids or secondary structure elements?



Answers:
a) The ranking is as follows:

Map 4: Res=3Å
Map 2: Res=5Å
Map 1: Res=10Å
Map 3: Res=15Å
Map 5: Res=20Å

The resolutions are provided just to illustrate how the numbers correlate to visual appearance of cryo-EM maps. This is hard to estimate visually, but it is good to have some idea of what different resolution values mean in terms of map quality and interpretability.

b) If you look carefully the two highest resolution maps (4 and 2) you will see the small side chains densities extending from the secondary structure elements (e.g., helices). Therefore, in these two cases you could use the map density to place the side chains and build models. Practically speaking, this would be much easier and more reliable in Map 4 since the resolution is higher. Higher resolution in this case means that the side-chain densities are better defined which increases the placement accuracy.

c) The resolution of Map 3 and Map 5 is too low to offer any interpretation on the atomic or amino-acid level. In fact, the density in these maps prohibits the analysis on the level of secondary structure because one cannot readily discern how different a-helices, b-strands or loops correlate to their locations in the density.

Question 6:

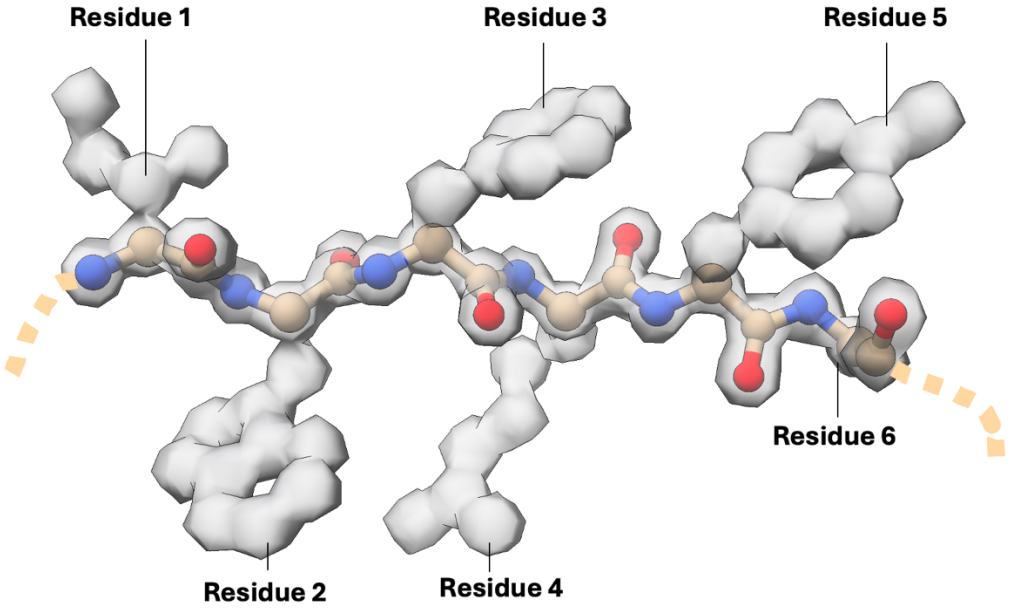
Jacques Dubochet, a Swiss scientist from Aigle, received the Nobel Prize in 2017 "for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution". His contribution was the development of a method to rapidly freeze liquid protein sample in liquid ethane.

a) Why does the protein sample need to be frozen for cryo-EM?
b) Why does the protein sample need to be frozen rapidly in liquid ethane?
c) It is common practice to wear masks while freezing grids and to dry all instruments that come into contact with the liquid ethane and nitrogen. Why?

Answers:
a) The protein sample needs to be frozen to preserve the sample from radiation damage and to protect it from dehydration and evaporation in the vacuum of the electron microscope.
b) Rapid freezing is required to obtain vitrified water without ice crystals. Only in vitrified water particles/proteins are clearly visible and can be imaged.
c) Both procedures are performed to minimize the water contamination in the liquid ethane and nitrogen which can in turn lead to ice contamination on the grid. Areas with large ice contaminations cannot be imaged.

Question 7:

You have performed structural analysis of a new protein using cryoEM and recovered an excellent quality map that is at 1.5Å There are no previously published structures of related proteins and you have to build the atomic model from beginning. You arrive to this section of the map that consists of 6 connected amino-acids and you need to figure out which part of the protein it corresponds to. The polypeptide backbone is already included but the amino-acid identities (side-chains) are not known.

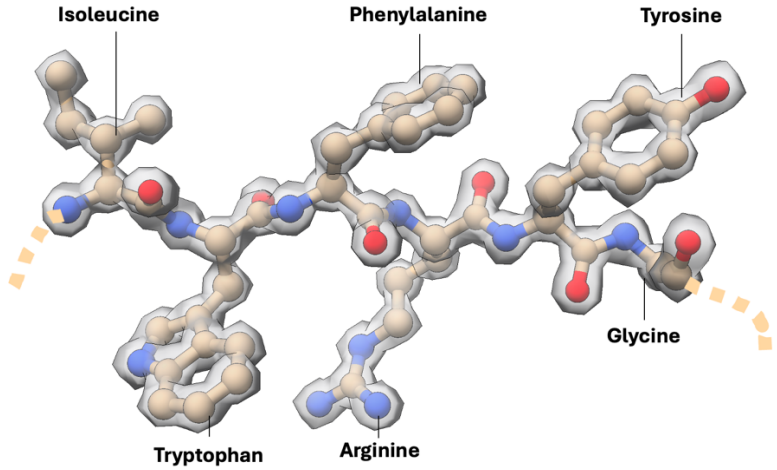


The sequence of the entire domain that also includes this region is:

----AFGRNTSWFRSSGIWFRYGTKEDGYYNPCLIHFNYRGSNTIWSRYGW----

Use your knowledge of amino-acid side-chains to infer which 6 consecutive amino-acids are shown in the map, and identify which section of the provided protein sequence it corresponds to. Note that at these resolutions you primarily see densities for heavy atoms (C, N, O) but not for hydrogens.

Answers:

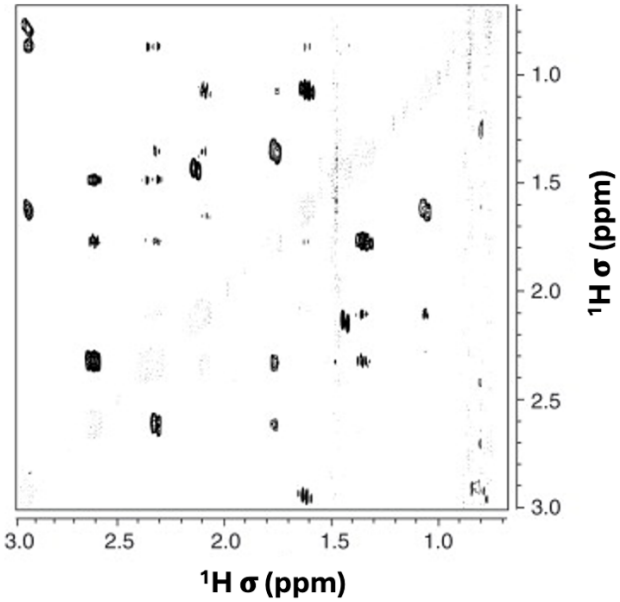


Based on the map density the inferred sequence of this region is “IWFRYG” corresponding to the red section of the protein sequence below:

----AFGRNTSWFRSSGIWFRYGTKEDGYYNPCLIHFNYRGSNTFRSRYGW----

Question 8:

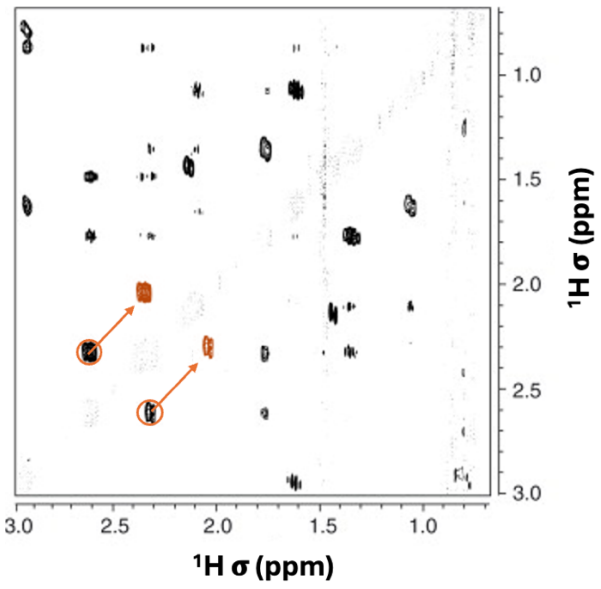
You are interested in resolving a structure of a small Calcium-binding protein by NMR. You collected a ^{1}H-^{1}H NOESY spectra of the protein in the absence of calcium and it looks like this:



a) What do the peaks in the spectra correspond to? What do the peak intensities tell you about the corresponding ^{1}H atom pairs?

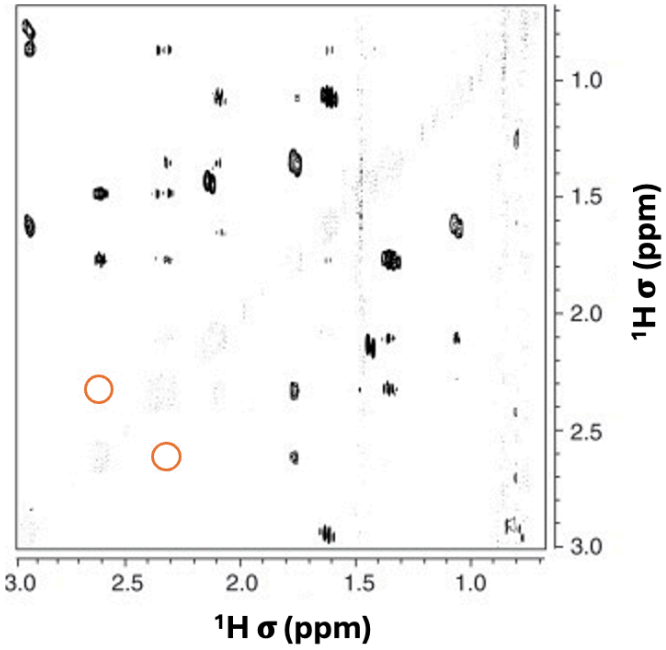
^{1}H-^{1}H NOESY experiments result in a set of peaks corresponding to pairs of hydrogen atoms that are in close proximity (<5Å) in 3D space. Each hydrogen atom in the pair will have a unique chemical shift (**σ**) which will be shown on X and Y axis for the first and the second hydrogen, respectively. The peak intensities depend on relative proximity of the two hydrogen atoms and will change dramatically with even small changes in distance (NOE signal ~$1/r^{6}$). The lower the intensity -> The greater the distance between hydrogens

b) Scenario 1: You added calcium to the protein and that resulted in shift of several peaks in the ^{1}H-^{1}H NOESY spectra (see below). What do you think is the explanation? Why did these ^{1}H exhibit a change in chemical shifts and others did not?



Chemical shits (**σ**) will change if there is a change of local chemical environment. In this case this is triggered by the addition of calcium which means that either (1) these hydrogen atoms (or the amino-acids they are a part of) are directly involved in the interaction with calcium ions, or (2) that calcium triggers some type of conformational change in the protein and indirectly changes the local chemical environment of these hydrogen atoms. The hydrogens that did not exhibit any change in chemical shifts are distal to calcium binding site and they do not experience any major changes in 3D assembly of the residues surounding them.

b) Scenario 2: You added calcium to the protein and that resulted in loss of several peaks in the ^{1}H-^{1}H NOESY spectra (see orange circles below). What do you think is the explanation? What is this reduction of peak intensity telling you about the corresponding ^{1}H pairs?



Peak intensities will change in NOESY spectra if the distances between the corresponding hydrogen pairs change. Given the sharp dependance on distance (NOE signal ~$1/r^6$), the hydrogen pairs that are more than ~5Å apart will not be detectable in these experiments. Therefore, the disappearance of peaks in NOESY indicates that the corresponding atoms shifted together with the underlying amino-acids. In this case, this is triggered by the addition of calcium, which leads to 2 possible explanations similar to what was described above (direct or indirect effect of calcium binding).

Question 9:

Below is a small part of a PDB file from two different structures (A and B). Can you identify what is different in terms of atom composition? Considering the atom composition, which structural biology method could have these structures originated from?

Structure A:

```
ATOM   2270  N   THR A 364      37.053  61.679  55.135  1.00 45.18           N
ATOM   2271  CA  THR A 364      36.644  63.002  55.575  1.00 41.40           C
ATOM   2272  C   THR A 364      35.133  63.162  55.498  1.00 42.51           C
ATOM   2273  O   THR A 364      34.526  63.741  56.395  1.00 41.87           O
ATOM   2274  CB  THR A 364      37.351  64.100  54.751  1.00 39.60           C
ATOM   2275  OG1 THR A 364      38.772  63.954  54.915  1.00 36.65           O
ATOM   2276  CG2 THR A 364      36.957  65.486  55.238  1.00 38.36           C
ATOM   2277  N   GLN A 365      34.539  62.643  54.433  1.00 40.74           N
ATOM   2278  CA  GLN A 365      33.089  62.723  54.269  1.00 40.85           C
ATOM   2279  C   GLN A 365      32.396  61.991  55.413  1.00 42.74           C
ATOM   2280  O   GLN A 365      31.415  62.482  55.961  1.00 44.33           O
ATOM   2281  CB  GLN A 365      32.673  62.090  52.945  1.00 42.48           C
ATOM   2282  CG  GLN A 365      33.070  62.906  51.716  1.00 43.48           C
ATOM   2283  CD  GLN A 365      32.670  62.179  50.442  1.00 41.36           C
ATOM   2284  OE1 GLN A 365      31.508  62.163  50.065  1.00 43.05           O
ATOM   2285  NE2 GLN A 365      33.623  61.540  49.808  1.00 42.06           N
ATOM   2286  N   ILE A 366      32.898  60.810  55.761  1.00 46.58           N
ATOM   2287  CA  ILE A 366      32.307  60.043  56.861  1.00 46.20           C
ATOM   2288  C   ILE A 366      32.432  60.816  58.169  1.00 49.94           C
ATOM   2289  O   ILE A 366      31.492  60.866  58.955  1.00 50.63           O
ATOM   2290  CB  ILE A 366      33.008  58.673  56.996  1.00 43.57           C
ATOM   2291  CG1 ILE A 366      32.766  57.878  55.730  1.00 42.97           C
ATOM   2292  CG2 ILE A 366      32.495  57.932  58.231  1.00 42.86           C
ATOM   2293  CD1 ILE A 366      33.601  56.636  55.618  1.00 42.38           C
ATOM   2294  N   SER A 367      33.582  61.441  58.413  1.00 51.38           N
ATOM   2295  CA  SER A 367      33.704  62.192  59.647  1.00 53.17           C
ATOM   2296  C   SER A 367      32.759  63.380  59.676  1.00 53.23           C
ATOM   2297  O   SER A 367      32.302  63.772  60.740  1.00 54.16           O
ATOM   2298  CB  SER A 367      35.167  62.637  59.903  1.00 54.34           C
ATOM   2299  OG  SER A 367      35.575  63.694  59.063  1.00 56.93           O
ATOM   2300  N   SER A 368      32.457  63.962  58.512  1.00 50.86           N
ATOM   2301  CA  SER A 368      31.548  65.107  58.465  1.00 48.18           C
ATOM   2302  C   SER A 368      30.132  64.650  58.878  1.00 45.09           C
ATOM   2303  O   SER A 368      29.438  65.325  59.646  1.00 47.04           O
ATOM   2304  CB  SER A 368      31.536  65.712  57.052  1.00 49.16           C
ATOM   2305  OG  SER A 368      30.680  66.848  57.009  1.00 52.42           O
ATOM   2306  N   ALA A 369      29.710  63.503  58.361  1.00 44.98           N
ATOM   2307  CA  ALA A 369      28.404  62.945  58.712  1.00 44.29           C
ATOM   2308  C   ALA A 369      28.348  62.684  60.208  1.00 44.40           C
ATOM   2309  O   ALA A 369      27.358  63.001  60.862  1.00 43.34           O
ATOM   2310  CB  ALA A 369      28.162  61.625  57.956  1.00 44.57           C
```

Structure B:

```
ATOM   1007  CA  LEU A  97      -8.354   1.914  -3.884  1.00  0.22           C
ATOM   1008  C   LEU A  97      -7.637   2.570  -5.058  1.00  0.21           C
ATOM   1009  O   LEU A  97      -7.259   3.734  -4.992  1.00  0.25           O
ATOM   1010  CB  LEU A  97      -7.319   1.264  -2.966  1.00  0.25           C
ATOM   1011  CG  LEU A  97      -6.284   2.248  -2.408  1.00  0.38           C
ATOM   1012  CD1 LEU A  97      -6.926   3.157  -1.381  1.00  0.52           C
ATOM   1013  CD2 LEU A  97      -5.097   1.524  -1.799  1.00  0.62           C
ATOM   1014  H   LEU A  97      -9.265   0.002  -4.103  1.00  0.23           H
ATOM   1015  HA  LEU A  97      -8.885   2.676  -3.331  1.00  0.24           H
ATOM   1016 1HB  LEU A  97      -7.837   0.802  -2.137  1.00  0.28           H
ATOM   1017 2HB  LEU A  97      -6.796   0.500  -3.520  1.00  0.47           H
ATOM   1018  HG  LEU A  97      -5.921   2.867  -3.215  1.00  0.46           H
ATOM   1019 1HD1 LEU A  97      -6.208   3.895  -1.056  1.00  1.11           H
ATOM   1020 2HD1 LEU A  97      -7.778   3.652  -1.820  1.00  1.18           H
ATOM   1021 3HD1 LEU A  97      -7.247   2.570  -0.533  1.00  1.02           H
ATOM   1022 1HD2 LEU A  97      -5.412   0.988  -0.913  1.00  1.03           H
ATOM   1023 2HD2 LEU A  97      -4.690   0.829  -2.517  1.00  1.29           H
ATOM   1024 3HD2 LEU A  97      -4.340   2.248  -1.529  1.00  1.25           H
ATOM   1025  N   GLU A  98      -7.467   1.814  -6.136  1.00  0.18           N
ATOM   1026  CA  GLU A  98      -6.790   2.319  -7.326  1.00  0.19           C
ATOM   1027  C   GLU A  98      -7.507   3.553  -7.874  1.00  0.17           C
ATOM   1028  O   GLU A  98      -6.926   4.635  -7.956  1.00  0.20           O
ATOM   1029  CB  GLU A  98      -6.735   1.227  -8.394  1.00  0.21           C
ATOM   1030  CG  GLU A  98      -5.724   1.491  -9.495  1.00  0.27           C
ATOM   1031  CD  GLU A  98      -6.131   2.632 -10.405  1.00  0.88           C
ATOM   1032  OE1 GLU A  98      -5.528   3.721 -10.296  1.00  0.96           O
ATOM   1033  OE2 GLU A  98      -7.051   2.438 -11.226  1.00  1.87           O
ATOM   1034  H   GLU A  98      -7.838   0.907  -6.144  1.00  0.18           H
ATOM   1035  HA  GLU A  98      -5.783   2.591  -7.050  1.00  0.22           H
ATOM   1036 1HB  GLU A  98      -6.481   0.290  -7.921  1.00  0.27           H
ATOM   1037 2HB  GLU A  98      -7.711   1.136  -8.848  1.00  0.31           H
ATOM   1038 1HG  GLU A  98      -4.773   1.731  -9.045  1.00  0.85           H
ATOM   1039 2HG  GLU A  98      -5.625   0.595 -10.088  1.00  1.00           H
```

Answer:
If you look at the last column you will notice that Structure A does not have hydrogen atoms while structure B does. This could be caused by the methods used for structure determination or the achieved resolution. Hydrogens cannot be readily visualized by X-ray crystallography or cryoEM unless the resolution is <1.5Å.

Structure A could have been resolved by X-ray crystallography or cryoEM but not NMR. NMR structure determination is based on hydrogens and they would have been present in the PDB file.

Structure B could have been solved by either of the three methods, but in the case of X-ray crystallography and cryoEM this would require a very high resolution to be achieved. This structure actually came from NMR.