# BMB/Bi/Ch 173 – Winter 2018

## Homework Set 8.1 (100 Points) – Assigned 2-27-18, due 3-6-18 by 10:30 a.m.

TA: Rachael Kuintzle. Office hours: SFL 220, Friday 3/2 4:00-5:00pm and SFL 229, Monday 3/5 4:00-5:30pm.

1. **NMR Basics (30 points)**
   a. (4 points) What makes a nucleus NMR active?
      i. An NMR-active nuclide will have a non-zero spin, and therefore a **non-zero magnetic moment**. So it has a magnetic dipole moment that can interact with the applied magnetic field. (Accept for partial credit: nucleus must have an odd number of protons and/or neutrons.)
      ii. A nucleus with spin=0 is not NMR active
      iii. A nucleus with spin=1/2 is NMR active
      *iv.* A nucleus with spin ≥ 1 is a quadrupolar nucleus and behaves differently than a spin=1/2 nucleus. For more information see *https://en.wikipedia.org/wiki/Nuclear_quadrupole_resonance*
      *v.*

Regularity of Nuclear Spin (I) for Different Nuclei:

| Atomic Number | Mass Number | Nuclear Spin (I) |
|---|---|---|
| odd or even | odd | 1/2, 3/2, ... 9/2 |
| even | even | 0 |
| odd | even | 1, 2, ... 6 |

   b. (5 points) What happens to an NMR active atom in an applied magnetic field?
      A moving charged particle creates a magnetic field. This is the magnetic moment of the nucleus. When this atom is placed in an external magnetic field, its magnetic dipole will align to the external field and precess at the Larmor frequency.

   c. (5 points) Why is a radiofrequency pulse used in an NMR experiment?
      i. After allowing the sample to align to the external magnetic field, a radiofrequency pulse is applied to disrupt the alignment. This induces "spin flips", forcing the sample into a higher energy level. Gradually the sample will relax back to alignment with the constant, external magnetic field. As it relaxes, it gives off the FID in the time domain.
      ii. The external magnetic field is along the z axis and the RF pulse pushes the net magnetic moment of the precessing nuclei into the XY plane where the signal can be read out by the detector. Gradually the relaxation brings the magnetic moments back into the Z direction.

   d. (2 points) What is the free induction decay (FID)?
      The FID is the current generated in the receiver coil by the magnetic moment. After taking a Fourier transform, the FID is in the frequency domain and yields the spectrum. A series of FID are collected and used to create the spectrum.

   *e.* (2 points) Why does the FID oscillate up and down?
      Alternating current

   *f.* (2 points) Why does the FID decay?
      The sample starts to relax back to equilibrium after the radiofrequency pulse.

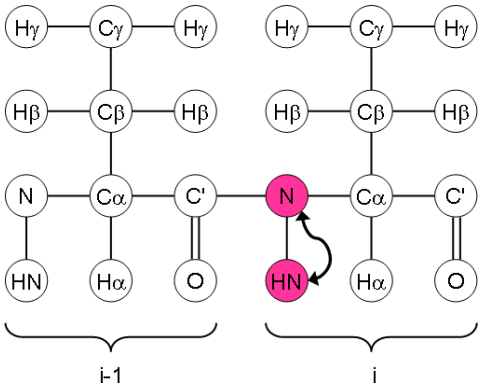
   *g.* (5 points) How is a Fourier transform used in NMR? What do different frequencies correspond to?
      i. Do a Fourier transform on the FID to figure out what frequencies are there.
      ii. Different nuclei have different precession frequencies. The FID is sum of the sine waves representing the different frequencies in the signal.

h. (5 points) Define "chemical shift" in your own words. Incorporate the concept of "shielding" and describe how it is related to the magnetic field perceived by an NMR-active nucleus.

Chemical shift is the resonance frequency of a nucleus in a magnetic field, relative to a standard, and it depends on the local magnetic environment of the nucleus in the molecule and solvent. Chemical shift is proportional to the strength of the magnetic field felt by an NMR active nucleus. When a hydrogen is bonded to an electronegative atom, that atom will draw electron density away from the protein, "de-shielding" it. A nucleus with more electron density around it (such as a methyl group hydrogen) is more shielded, and feels a weaker magnetic field than a de-shielded hydrogen (such as a hydroxyl group hydrogen) with little electron density around it.

## 2) $^{15}$N-HSQC (Heteronuclear Single Quantum Correlation) (28 Points)

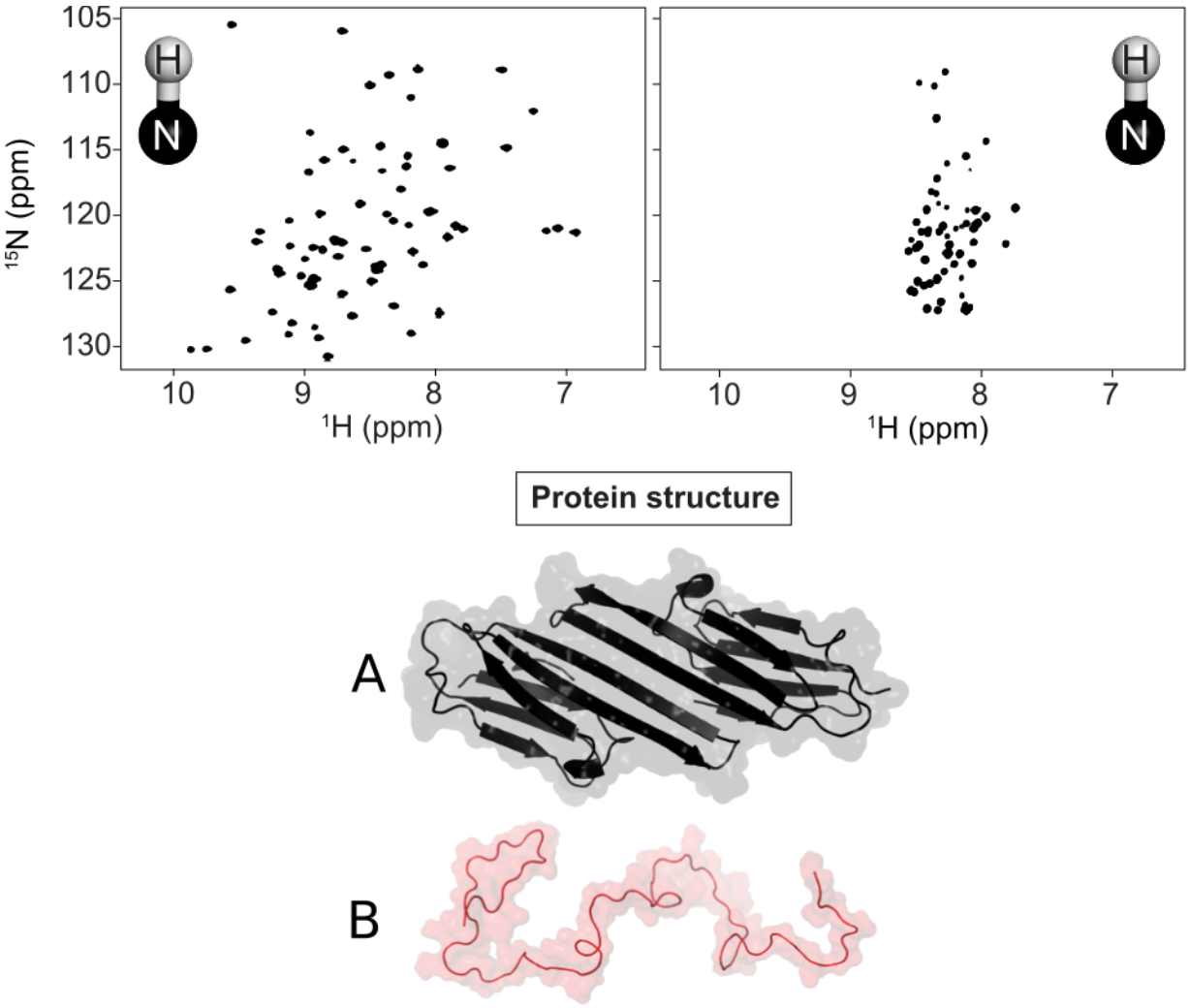
A. (5 points) $^{15}$N-HSQC is a type of 2D NMR experiment that measures $^{1}$H-$^{15}$N one-bond coupling. The second dimension, which arises from transfer of magnetization from the proton to the NMR-active nitrogen isotope, spreads peaks out and makes them easier to assign. It requires only a small amount of $^{15}$N-labeled material (less than 2 mg for a 20 kDa protein) and only ~30 minutes of spectrometer time. It yields a signal for almost every amino acid in a protein. For these reasons, a $^{15}$N-HSQC spectrum is often the first NMR spectrum recorded on a new protein under investigation. An example of a nuclei pair which would give rise to an $^{15}$N-HSQC peak is highlighted on the dipeptide below:




Which amino acid(s) doesn't (don't) have $^{15}$N-HSQC signals? Explain why.

Proline doesn't have a $^{15}$N-HSQC signal because it doesn't have an amide proton (unless it's at the N-terminus of the protein).

B. (10 points) Below are two $^{15}$N-HSQC spectra of the same protein. Label each spectrum with the letter corresponding to the protein conformation it represents. **Explain your reasoning.**




First spectrum: A. Second spectrum: B. The well-folded protein will have greater proton chemical shift dispersion, because these nuclei are experiencing more diverse shielding environments based on their proximity to other atoms in the protein. The unfolded protein has poor chemical shift dispersion because most protons are exposed to solvent and therefore experience very similar environments. When a protein is disordered, the corresponding $^{1}$H resonances will cluster between 8 and 8.5 ppm.

C. (3 points) In NMR, T2 relaxation time depends on the "tumbling rate," or rotational diffusion rate, of a molecule in solution. Quicker T2 relaxation times yield broad NMR peaks and poor signal-to-noise ratios. One of the fundamental limitations of NMR is the size of the molecule under study, in part due to the fact that larger molecules tumble more slowly and yield broad peaks which overlap too much. Assuming you already know the size of your protein, what is one thing you can learn about your protein sample from the broadness of your $^{15}$N-HSQC peaks?
Provides information on the aggregation state of the protein population. (If aggregated, peaks will be very broad). You could reason that it might also give info on whether the protein forms dimers or oligomers.

D. It is estimated that 40% of proteins have disordered regions of 40 residues or more (https://www.ncbi.nlm.nih.gov/pubmed/12368089).

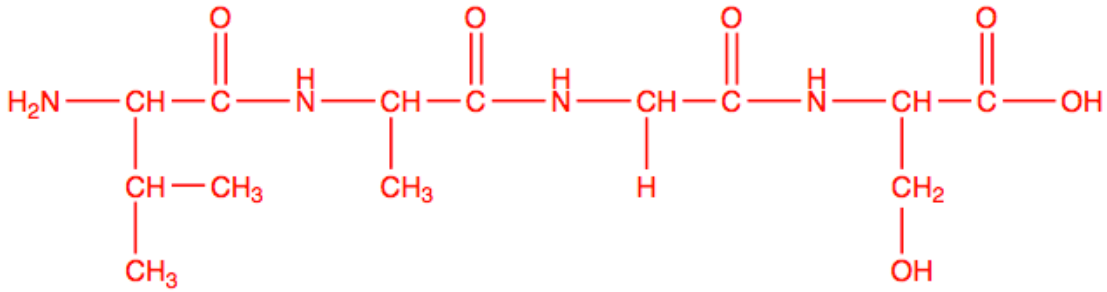
a. (2 points) Why is it difficult to assign resonances for disordered regions?
The protons are mostly exposed to solvent and therefore have very similar chemical shifts (their peaks are too close together to be distinguished).

b. (5 points) One approach to solve the problem above is to tag the protein with a paramagnetic tag. For example, a DOTA-M8 ytterbium tag (Yb-M8) (DOTA=1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid) could be attached to a cysteine in the protein. Briefly, how does this enable ambiguous signals to be assigned to specific N-H groups in an HSQC spectrum?
Accepted answer: The tag will change the chemical shift of nearby nuclei.
Long answer: It yields large chemical shift changes in neighboring residues (up to 26 residues away) in the $^{1}H^{15}N$ spectra. The signals that shift the most are closest to the Yb tag.
http://onlinelibrary.wiley.com/doi/10.1002/anie.201607261/full

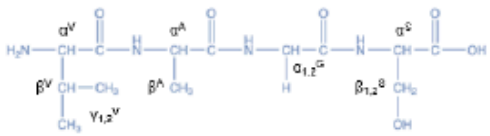
c. (3 points) Why is it important to choose a tag that will not spontaneously isomerize?
Because it could double your peaks, making it harder to assign them.

**3) Predicting 2D COSY data (20 points)**

A. (30 points) You have the following four peptide sequence: Val-Ala-Gly-Ser

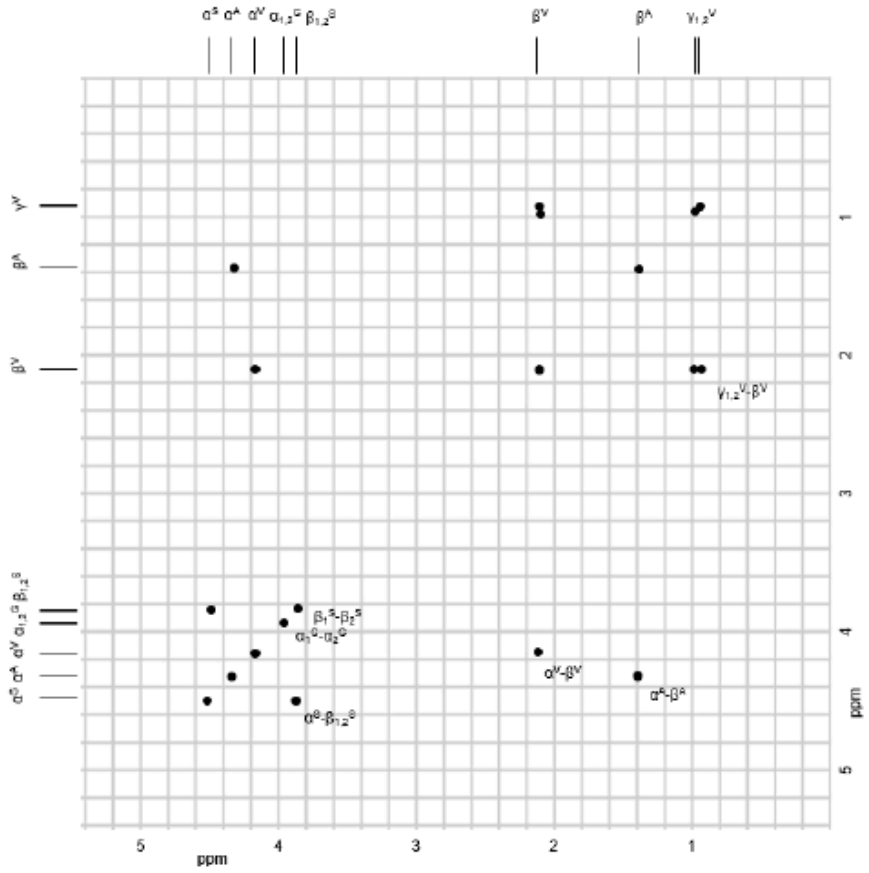


Sketch the 2D-COSY spectrum, if it was in the linear form. Use the table provided at the end of the assignment as a guide. Assume you are using a deuterated solvent; therefore, amine and hydroxyl protons will likely be replaced by deuterium and will not be observed.



Assume amine, amide, and hydroxyl protons solvent-exchange with deuterated solvent (NMR invisible).

| Residue | NH | $C_\alpha H$ | $C_\beta H$ | Others |
|---|---|---|---|---|
| Gly | 8.39 | 3.97 | 1.39 | 0.97, 0.94 |
| Ala | 8.25 | 4.35 | 2.13 | $C_\gamma H_3$ |
| Val | 8.44 | 4.18 | 3.88, 3.88 | |
| Ser | 8.38 | 4.50 | | |




**4) Other NMR experiments (10 points)**

A. (5 points) What type of magnetization transfer discussed in class can give you information about secondary structures present in your protein? **How** can you learn about your protein's secondary (and tertiary) structure from a NOESY experiment?
**Through-space NOE couplings or J-coupling** (which gives dihedral angles). Crosspeaks in a NOESY spectrum represent protons coupled through space via NOE (<5 Å). Using this info, you can figure out distance constraints and compare them to known values characteristic of different secondary and tertiary structures.

B. (5 points) Another technique you could use to investigate the tertiary structure of your protein is hydrogen-deuterium exchange (https://en.wikipedia.org/wiki/Hydrogen%E2%80%93deuterium_exchange). How does this experiment give information about tertiary structure?
Put your protein in a deuterated solvent. Take several NMR spectra over time and watch how fast different proton signals disappear as they are replaced by deuterium. You can learn from this how "exchangeable" the protons are; for example, buried residues are not readily accessible to solvent and will take much longer to exchange.

**5) Comparing NMR and X-ray crystallography (12 points)**

A. (4 points) We can determine 3D molecular structure with both X-ray crystallography and NMR, but the process required to build a model from the data is quite different for each method. Briefly (in 1-2

sentences) describe the process of model-building for each method. (For X-ray crystallography, assume we already have phases.)

- With X-ray crystallography you have an electron density map and we know that the 3D structure of our protein should fit into this map, we just need to find out where each residue is located within the map.
- With NMR, we assign peaks in different NMR experiments to specific nuclei to identify residues and obtain constraints for secondary structure and long range tertiary contacts. We then iterate through a variety of possible structures until we find an ensemble of structures that satisfy the constraints and end up with an ensemble of structures that satisfy constraints.

B. (4 points) Describe 2 technical limitations of NMR and X-ray crystallography (2 for each technique).

- NMR
  - requires small proteins (10-20 kDa ideally, up to ~50 kDa is possible)
  - proteins must be stable in solution for long periods of time
  - large amounts of pure protein are needed
  - need access to strong instrument
- X-ray crystallography
  - requires protein to be in crystals (often a non-trivial problem)
  - it can be difficult to get phases
  - need access to beamline

C. (4 points) Describe a protein structure that would be better to solve with NMR, and one that would be better to solve with X-ray crystallography. Explain your reasoning.
Many possible answers.

## $^1$H Chemical Shifts for the 20 Common Amino Acids (in ppm)

| Residue | NH | $C_\alpha H$ | $C_\beta H$ | Others | |
|---|---|---|---|---|---|
| Gly | 8.39 | 3.97 | | | |
| Ala | 8.25 | 4.35 | 1.39 | | |
| Val | 8.44 | 4.18 | 2.13 | $C_\gamma H_3$ | 0.97, 0.94 |
| Ile | 8.19 | 4.23 | 1.90 | $C_\gamma H_2$ | 1.48, 1.19 |
| | | | | $C_\gamma H_3$ | 0.95 |
| | | | | $C_\delta H_3$ | 0.89 |
| Leu | 8.42 | 4.38 | 1.65, 1.65 | $C_\gamma H$ | 1.64 |
| | | | | $C_\delta H_3$ | 0.94, 0.90 |
| Pro(*trans*) | | 4.44 | 2.28, 2.02 | $C_\gamma H_2$ | 2.03, 2.03 |
| | | | | $C_\delta H_2$ | 3.68, 3.65 |
| Ser | 8.38 | 4.50 | 3.88, 3.88 | | |
| Thr | 8.24 | 4.35 | 4.22 | $C_\gamma H_3$ | 1.23 |
| Met | 8.42 | 4.52 | 2.15, 2.01 | $C_\gamma H_2$ | 2.64, 2.64 |
| | | | | $C_\epsilon H_3$ | 2.13 |
| Cys | 8.31 | 4.69 | 3.28, 2.96 | | |
| Asp | 8.41 | 4.76 | 2.84, 2.75 | | |
| Asn | 8.75 | 4.75 | 2.83, 2.75 | $N_\gamma H_2$ | 7.59, 6.91 |
| Glu | 8.37 | 4.29 | 2.09, 1.97 | $C_\gamma H_2$ | 2.31, 2.28 |
| Gln | 8.41 | 4.37 | 2.13, 2.01 | $C_\gamma H_2$ | 2.38, 2.38 |
| | | | | $N_\delta H_2$ | 6.87, 7.59 |
| Lys | 8.41 | 4.36 | 1.85, 1.76 | $C_\gamma H_2$ | 1.45, 1.45 |
| | | | | $C_\delta H_2$ | 1.70, 1.70 |
| | | | | $C_\epsilon H_2$ | 3.02, 3.02 |
| | | | | $N_\epsilon H_3$ | 7.52 |
| Arg | 8.27 | 4.38 | 1.89, 1.79 | $C_\gamma H_2$ | 1.70, 1.70 |
| | | | | $C_\delta H_2$ | 3.32, 3.32 |
| | | | | $N_\delta H$ | 7.17, 6.62 |
| His | 8.41 | 4.63 | 3.26, 3.20 | $C_2 H$ | 8.12 |
| | | | | $C_4 H$ | 7.14 |
| Phe | 8.23 | 4.66 | 3.22, 2.99 | $C_2 H, C_6 H$ | 7.30 |
| | | | | $C_3 H, C_5 H$ | 7.39 |
| | | | | $C_4 H$ | 7.34 |
| Tyr | 8.18 | 4.60 | 3.13, 2.92 | $C_2 H, C_6 H$ | 7.15 |
| | | | | $C_3 H, C_5 H$ | 6.86 |
| Trp | 8.09 | 4.70 | 3.32, 3.19 | $C_2 H$ | 7.24 |
| | | | | $C_4 H$ | 7.65 |
| | | | | $C_5 H$ | 7.17 |
| | | | | $C_6 H$ | 7.24 |
| | | | | $C_7 H$ | 7.50 |
| | | | | NH | 10.22 |

Source: Adapted from data in K. Wüthrich (1986).