Example Synthesis Questions

These questions are extracted/modified from last years’ exam(s) to give you an idea of what

Synthesis questions might look like. The actual exam questions may be more conceptual. Don’t overfit to these questions. We will not be releasing the answers to these questions.

1. A colleague has sequenced a new protein sequence and wants to learn more about its potential biological role. They heard that you learned a lot about amino acid sequence analysis in BIOMEDIN 214, and so decide to consult you on this project. They want to learn as much as they can about this protein’s likely function based off the sequence information, and whatever can be derived from that, and ask you how to go about that. How do you reply? Suggest two questions that they could try to answer from this sequence, and briefly explain how they would go about answering those questions. One paragraph on each will suffice.
2. Person X is a medical researcher trying to figure out what is functionally happening in the cell such that some of their clinic’s patients respond to xyz chemotherapy and others do not. They have a bunch of tumor biopsies (samples) pre-treatment and has information on whether or not the person responded to treatment. Because you a world-famous bioinformatician, they approach you to get your advice on how to get started.
   1. You will be using gene expression data. Name at least 2 methods that researchers use to obtain this information
   2. What are the two gene expression groups that you are comparing?
   3. Person X says that they want to take the top genes from Group 1 and the top genes from Group 2 and functionally compare the 2 groups. Explain any problems with this approach. Name an alternative approach and explain in 1-3 sentences.
3. A medical application of HMMs (total 20pts) Creutzfeldt-Jakob **disease** (CJD) is a prion disease (prions are misfolded proteins) which involves fatal, degeneration of neural tissue. It is hypothesized that this disease is due to self-propagating conformations of proteins. This means that a misfolded protein (X\_sc) can cause disease by coming in contact to the normally folded version of that protein (X\_n), causing it to misfold (X\_n -> X\_sc). These misfolded proteins aggregate, leading to plaque build-up in the brain and thus rapid death.

What makes a protein prion-like is not well understood. Your goal is to informatically predict if a given protein can be a prion. For the first part, you will be building an HMM model to predict whether or not a protein can be a prion. Assume you are only given protein sequence information

* 1. What are the hidden states? What are your observations?
  2. How do you build your HMM model using this alignment? 1-3 sentences.
  3. Explain why using a position specific scoring matrix (PSSM) such as BLOSUM or PAM would not be helpful in this setting.