

## ECS129: Quiz 5 Answers

### AlphaFold and AlphaFold2

#### 1. AlphaFold and AlphaFold2 are two programs that have been designed to:

- a : Predict how proteins fold (The protein folding problem)
- b : Predict putative binding sites on proteins (for drug design)
- c : Predict the native structure of a protein (The protein structure prediction problem)
- d : Predict if two proteins interact, and, if they do, the structure of the complex

AlphaFold and AlphaFold2 were both designed to predict the native structure of a protein, and not how the protein folds.

#### 2. Which of these statements related to AlphaFold is most likely to be incorrect?

- a : AlphaFold needs information from large databases of protein sequences and structures
- b : AlphaFold performs best on multimeric proteins
- c : AlphaFold is likely to perform poorly on a protein that has a very small number of homologs
- d : AlphaFold does not predict the structures of RNA

We are looking for an incorrect statement. Many multimeric proteins are not fully globular, and AlphaFold does not perform well on such proteins, as there are no contacts per se that can help find the global fold.

#### 3. Protein sequence co-evolution gives information on

- a : The conformations of sidechains in a protein
- b : The presence of tryptophanes in a protein
- c : The stability of a protein
- d : 3D contacts between residues in a protein

The concept of co-evolution is to identify residues within a protein that may have co-evolved, i.e., when one of the residues is modified, they other “co-evolve” to adapt. Such co-evolution is expected to be a consequence of the fact that the two residues are next to each other, i.e. in contact.

#### 4. Two protein sequences differ by a single amino acid. You use AlphaFold to predict the structures of those sequences and find two structure models that are basically identical. You

- a : Trust those results as this is exactly what you expected
- b : Trust those results as AlphaFold has solved the protein structure prediction problem
- c : Do not fully trust those results as single mutations are known to sometimes alter the structure of a protein

d : Do not fully trust those results as AlphaFold predictions are usually poor

This one was a little tricky. It is true that two proteins with nearly identical sequences are expected to have similar structures. However, biology is not an exact science and there are examples of proteins in which a single mutation can lead to significant structural changes... this is why it is always good to be cautious.. and why c is the most likely answer!

**5. One of these statements is likely incorrect. Which one?**

a. AlphaFold relies heavily on multiple sequence alignment

b : AlphaFold predicts the full conformation of a protein, including the conformations of its sidechains

c : AlphaFold is most likely to fail on human proteins

d : AlphaFold does not consider information about ions or ligands when reading a file defining the structure of a protein

We are looking for an incorrect statement. It is true that AlphaFold relies on multiple sequence alignments (to predict contacts), that it predicts the full structure of a protein and not just its mainchain, and that it works on the protein alone, and not its environment. So the most likely (incorrect) statement is c.

**6. To assess the quality of a protein structure prediction, we compute either the RMSD or the TM score between the model and the actual experimental structure of the protein. What can you say about those scores?**

a : The higher the RMSD, the better the prediction

b : The higher the TM score, the better the prediction

c : A reasonable value for a good TM score is 10

We know that RMSD is akin to a distance, and therefore is small when the structures are similar. It cannot be answer a. The TM score is normalized to be between 0 and 1, so it cannot be answer c. It is true that TM was designed such that high scores indicate similarity, so answer b is valid.