AST, Jan 2016

1. In the problem of protein sequence optimisation, how large is the potential search space, as a function of protein size (N_{res}) and the number of types of amino acid (N_{type}) ?

2. What might make the search problem easier than considering all possible sequences ?

3. You have a function which tells you how well a protein sequence fits to a structure. It acts as if it were an energy function $E(S)$ of the sequence S so a more negative energy is more favourable. Write in pseudocode a Monte Carlo method for simulating a walk over possible sequences at constant temperature.

4. You have a quasi-energy energy function $E(S)$ of the protein-sequence S. It contains two terms, $E^1(S)$ and $E^2(S)$ which correspond to the interactions of an amino acid with its fixed environment (E^1) and pairwise interactions (E^2). Describe a method with pseudocode to find the best possible score a residue of type a could have at a position i in the sequence. This is not the same as asking for the complete dead end elimination method.

5. Why is it not adequate to use an energy or free energy function as the score for protein sequence optimisation ?

6. Explain why protein sequence optimisation is a discrete optimisation problem. Give an example of a continuous optimisation problem.

7. If protein sequence design is a discrete optimization problem, what optimisation methods can I not use ?

8. I develop a method for finding the best energy sequence for a protein structure. Why may it not be important to find the single best sequence ?

9. I have successfully described an optimisation method (maybe Monte Carlo) which lets me find a good sequence for a given protein structure. How would you generalise the method to handle arbitrary rotamers ?

10. Describe in words and a diagram why the methods used in the program "psi-blast" are able to find more remote homologues than those used in "blast".

11. Explain why a good amino acid substitution matrix is more important for comparing remote protein homologues than for similar proteins.

12. You have to build a model for a protein sequence. When would you prefer to use a sequence alignment ? When would you consider using a protein sequence to structure alignment ?

13. Describe the entries you would put in a scoring matrix in order to find the alignment of a protein sequence to some structure. What kind of scoring scheme would one use and how does it differ from a sequence-sequence scoring scheme ?

14. Difficult: You have a sequence to structure alignment.

sequence	A C - E F A C D E F
structure	1 2 3 4 5 6 7 - 8 9

What are the geometric implications of the gap above place 3 and the gap below the D ?

15. I want to align a protein sequence to a protein structure. I have a perfect energy function which always returns the correct energy for any atoms interacting with each other. Why can I not use a simple scoring scheme as in sequence alignments ?

16. Describe a method which allows one to align a sequence into a protein structure.