Bif401 Final term Feb, March 2017 solved Papers

Protein structure prediction:

There are three different strategies for structure prediction

1. Homology Modelling
2. Threading/Fold Recognition
3. Ab Initio Modelling.

Chou – Fasman method:

It is a technique for the prediction of secondary structures in proteins i.e. Alpha Helices, Beta Sheets and Turns is Chou – Fasman technique. The method is based on analyses of the relative frequencies of each amino acid in alpha helices, beta sheets, and turns based on known protein structures solved with X-ray crystallography. From these frequencies a set of probability parameters (in our handouts, it is propensity table) were derived

- **For the appearance** of each amino acid in each secondary structure type.
- **To predict the probability** that a given sequence of amino acids would form a helix, a beta strand, or a turn in a protein.

Chou-fasman algorithm (alpha helix):

- For Alpha Helices, 4 contiguous amino acids are required.
- Their Alpha-Helix propensity should be more than 1.0
- Once this propensity falls below 1.0, Alpha-Helix stops.

Chou- Fasman (Loop form after alpha helix and beta sheet) (10 marks):

After computing the propensity of alpha helices and beta sheets, we need to settle for loops.

- Let's see how can we find out the loops using Chou Fasman Algorithm.
- For any jth residue in sequence, we calculate f (Total) = f(j) f(j+1) f(j+2) f(j+3) (tetra peptide).

- If,
- f (Total) > 0.000075.
- The average value for P (turn) > 1.00 in the tetra peptide.
- The averages for the tetra peptide are such P(a-helix) < P(turn) > P(b-sheet),
- It is a Turn!

Advantage and disadvantage of Ab initio:

Advantages:

- Ab Initio methods can fold any target sequence using only physical atomic properties.
- Predictions are mostly accurate and correctly describe the natural folding process.

Disadvantages:

- Ab initio methods are the very difficult to design (energy function).
- These methods are slow due to the huge possibilities.

Bioinformatics role:

- An interdisciplinary field
- New but rapidly developing field
- Low requirement on infrastructure and research equipment
- Vast opportunities for scientific discovery

PAM matrix

- PAM means “Point Accepted Mutations”
- Point accepted mutation is a substitution of one amino acid by another such that the protein functions stays conserved.
- PAM unit is a time in which about 1% of amino acids in a sequence undergo accepted mutations
- PAM matrices are scoring matrices that are useful in computing sequence alignment scores.

CLUSTALW:

- Performs alignment in slow/accurate or fast/approximate
- Create multiple alignments, optimize existing alignments, profile analysis, create phylogenetic trees
- CLUSTALW can use multiple file formats including EMBL/SwissProt and Pearson (Fasta) etc

Function of Tandem Ms:

Tandem MS helps in measurement of mass to the fragments as well. This process provides another step in further scoring and ranking and Protein identification thus becomes easier.

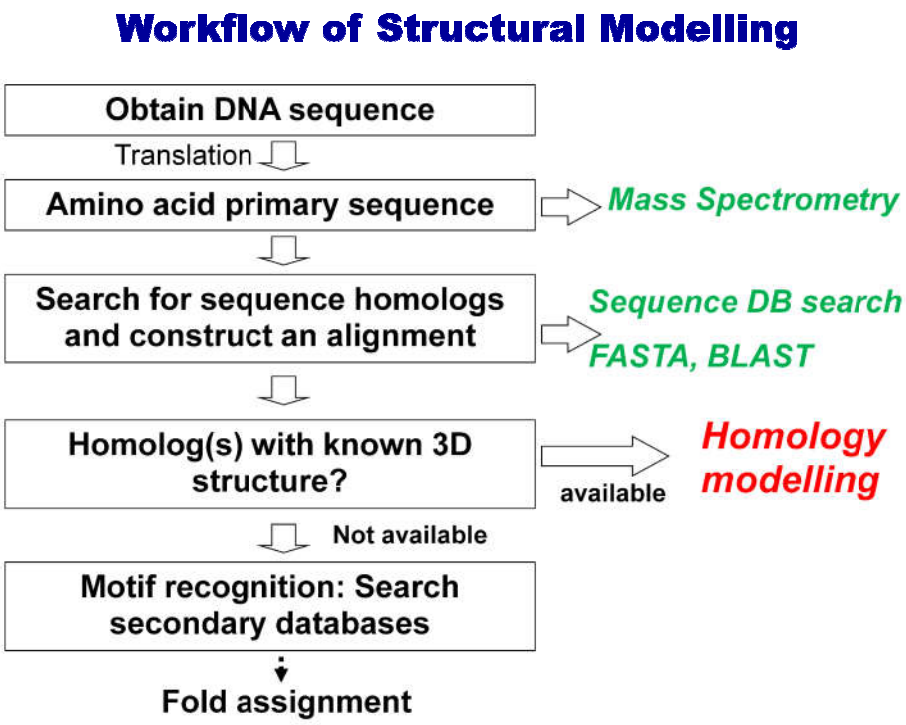
Steps present in flow chart of homology modeling: - (10 marks)

1. Template recognition and initial alignment.
2. Alignment correction.
3. Backbone generation.
4. Loop modeling.
5. Side-chain modeling.
6. Model optimization.
7. Model validation

How homology modeling is used for knowing the sequence of unknown protein sequence: (10 marks)

If another protein which has a similar sequence also has its structure known, the structure of an unknown protein can be predicted based on that similar protein. So, it is then possible to identify unknown protein structures by just examining the homologous protein sequences. Good sequence alignment and identity ensures that homology modelling will give accurate results

Thus, Homology modeling is used to predict structures of proteins having high sequence similarity with other proteins with known structures:



Silico fragment scoring:

- Count the matches between in silico and in vitro peaks.
- Give an equivalent score to the candidate protein.
- Weigh each of the aforementioned match by the mass error.
- Accumulate the score

How we can get scoring in BUP proteomics:

There are two approaches to perform BUP:

- Peptide Mass Fingerprinting.
- Shotgun Proteomics

ProSight pattern:

- Kelleher et al have developed an online Top Down Proteomics Search Engine i.e. Prosight PTM
- ProSight PTM is the state of the art in top down proteomics search.

- Using Prosight PTM, post-translational modifications can be accurately identified.

Top down proteomics:

Measures intact proteins followed by their peptides after fragmentation

Score in Protein Fragmentation:-

If we can:

- Measure the mass of fragments using MS.
- Calculate the Protein Fragmentation Techniques theoretical mass of the fragments.
  Then, we can award score on the basis of the similarity of experimental and theoretical mass.

Molecular biology:

It is the branch of biology that deals with the structure and function of the macromolecules (e.g. proteins and nucleic acids) essential to life. Molecular biology is the study of biology at a molecular level.

Protein Fragmentation techniques:

- Electron Capture Dissociation (ECD).
- Electron Transfer Dissociation (ETD).
- Collision Induced Dissociated (CID).

Types of secondary structures of RNA:

1. **Single stranded:**
   - 3'end may fold on to the 5' end.
2. **Helices:**
   - Double stranded RNA helix of stacked base pairs
3. **Hairpin loop:**
   - The loop of the hairpin must at least four bases long to avoid steric hindrance with base-pairing in the stem part of the structure

4. **Bulge Loops:**
    - Bulges, are formed when a double-stranded region cannot form base pairs perfectly
5. **Interior loop:**
    - Interior loops are formed by an asymmetric number of unpaired bases on each side of the loop.
6. **Junctions or intersections:**
    - Junctions include two or more double-stranded regions converging to form a closed structure.
    - The unpaired bases appear as a bulge.

Rooted trees:

- Each node with descendants represents the inferred most recent common ancestor of the descendants.
- The edge lengths in some trees may be interpreted as time estimates.
- Rooted trees can show temporal evolutionary direction.
- Expensive.

Unrooted:

- Only the relatedness of the leaf nodes.
- Do not require the ancestral root to be known or inferred.
- Less expensive.

Uses:

In bioinformatics, such as

- Rooted and unrooted trees can be used to show phylogenetic relationships between sequences.

For further help: Bc150403244@vu.edu.pk