Question 5

We are trying to predict the structure of a protein, for which we cannot find a structure with a similar sequence

a) Mention an alternative method. Do you trust this method? Why / why not?
b) Why is it difficult to predict a protein structure from the sequence only?

After a protein model was generated, its quality needs to be assessed

c) Mention a molecular check that can be performed. What do we measure and what should the outcome be?
d) What question does a docking algorithm try to answer?
e) What is wrong with looking at a single protein structure?

Secondary structure prediction

structural (experimental) data are usually not available for:

- new sequences
- unknown folds
- artificial sequences

→ prediction of the secondary structure of a sequence without knowledge on the 3D-structure

---

The Protein Databank can be used to perform a general database search to find sequences that have a similar amino acid composition to the target sequence with the unknown structure.

When does the alignment indicate homology?

If at least 30% of the residues in the unknown sequence and the known sequence are identical, homology modeling can be done automatically.

If less than 30% aa are identical, additional methods and sources of information are necessary to obtain a valuable 3D-model.

→ f.e. by secondary structure prediction, threading ("inverse" protein folding)

① Template selection (database search and fold assignment)

- which experimentally known structure can be used as template?
- proteins have to be related to the polypeptide sequence of the target sequence (BLAST, FASTA search; 30% sequence identity required)

Alternative method:

Threading or structural fold recognition: predicts the structural fold of a protein by fitting the protein sequence to all folds and select the best-fitting fold. Can be used to identify structurally similar proteins even without detectable sequence similarity.

Structure prediction from sequence

Basic principles for modeling

- proteins consist of areas of defined secondary structure
- secondary structure elements are conserved

Statistical evaluation of many sequences shows a good correlation between amino acid sequence and 3D-structure

More than 90% of all amino acids in proteins can be assigned to secondary structures, less than 10% are "random coils".

Ab initio-based methods

make use of the sequence information only

predict secondary structures based on statistical calculations.

prediction of short-range interactions (α-helix) is easy, but prediction of long-range interactions (β-sheets) is extremely difficult.

Why is it difficult to predict a protein structure from the sequence only?

Number of entries in PDB is large, most of the protein structures are redundant. Among the unique protein structures, there are only a limited number of protein folds available (800) compared to ~1 Mio. unique protein sequences => the protein structures are much more conserved.

Wie kann man die Güte eines Proteinmodels überprüfen?

The final homology model has to be evaluated to make sure that the structural features of the model are consistent with the physicochemical rules. This involves checking anomalies in $\phi - \psi$ angles, bond lengths, close contacts, planarity, symmetry axes, ...

Evaluation programs: Procheck, WHATIF, ANOLEA, Verify 3D

Keep in mind: the evaluation tests performed by these programs only check the stereochemical correctness, regardless of the accuracy of the model.

The resulting structure must obey the obvious structural features of protein structures:

- Secondary structure elements (α-helices, β-sheets)
- hydrophilic outside, hydrophobic inside
- H-bonding; salt bridges
- ψ, ϕ distributions, side-chain rotamers

What question does a docking algorithm try to answer?

In the field of molecular modeling, docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex.

Docking is frequently used to predict the binding orientation of small molecule drug candidates to their protein targets in order to predict the affinity and activity of the small molecule.

Docking - computational simulation of a candidate ligand binding to a receptor

focus of molecular docking : computationally simulate the molecular recognition process. Achieve an optimized conformation for both the protein and ligand.