

b) Give an example of a kind of protein for which it is very difficult to determine the structure experimentally. Explain why.

gegeben war ein Protein mit Anker in einer Membran und einer heme-group. Warum ist es gerade bei diesem Protein schwierig, die Struktur aufzuklären?

•) Big proteins / protein complexes

- don't crystallize well / resolution should be very high
- give an interpretable NMR spectrum

•) Membrane bound proteins

- Don't crystallize as without special tricks
- Are not in solution

•) Very flexible / partially unstructured proteins

- Don't crystallize (or in only one possible conformation)
- Mutually inconsistent nuclear distances

c) Mention a way to obtain a model of the 3D structure with a computer. Indicate why this should work.

(Mit welchen 2 Methoden kann man die Tertiärstruktur vorhersagen → in silico?) → Mit "in silico" bezeichnet man Vorgänge, die im Computer ablaufen.

There are 3 computational approaches to protein 3-dimensional structural modeling and prediction.

Homology modeling, threading, ab-initio prediction.

knowledge-based methods

simulation based

Homology modeling (comparative modeling): predicts protein structures based on sequence homology with known structures. Principle behind it: If two proteins share a high enough sequence similarity, they are likely to have very similar three-dimensional structures. If one of the protein sequences has a known structure, then the structure can be copied to the unknown protein with a high degree of confidence.

1. Search for homologous proteins in the Protein Data Bank (PDB)
2. Align sequences (automatically): at least 50%
3. Determine structurally conserved regions / variable regions
4. Determine coordinates: backbone, SCR, SVR, side chains (side chain refinement)

## Homology modeling: restrictions

- loops are not as structured as secondary structure elements
- identical sequences of up to 8 amino acids found in both alpha-helix and beta-strand

## Threading and fold recognition

There are only small number of protein folds available ( $< 1000$ ), compared to millions of protein sequences. (protein structures tend to be more conserved than protein sequences). Consequently, many proteins can share a similar fold even in the absence of sequence similarities.

Threading (structural fold recognition) predicts the structural fold of an unknown protein sequence by fitting the sequence into a structural database and selecting the best-fitting fold. The comparison emphasizes matching of secondary structures, which are most evolutionarily conserved.

→ This approach can identify structurally similar proteins even without detectable sequence similarity.

Protein threading: we don't find completely new structures so often. Usually a protein looks similar to something we know

Which structure fits to my sequence? Put your sequence over all known protein folds. Calculate their energies. Which structure gives the best energy?

The ab initio prediction method attempts to produce all-atom protein models based on sequence information alone without the aid of known protein structures. The ab initio prediction programs are designed using the energy minimization principle. These algorithms search for every possible conformation to find the one with the lowest global energy.

Ab initio prediction algorithms: Their prediction accuracies are too low to be considered practically useful.

d) In order to understand the workings of a protein, is it enough to know what it looks like? Why? No. There is no further information

about motifs, bonds,  
RasMol, Swiss-PDBViewer, CPK — visualization programs