Why build model structures?

Knowledge of a protein's three-dimensional structure is vital to a full understanding of the molecular basis for its biological function.

We want to understand the function of all proteins encoded by a genome, therefore we would like to know all of their 3-D structures.

Experimental techniques for determining protein structure are relatively slow and expensive, so we look to modelling as a way of extending the set of 3-D structures.

Modelling can also be used in protein engineering when designing proteins for therapeutic applications.

Comparative modelling strategy

- identify a known structure that is predicted to be similar;

- align sequences;

- predict structurally conserved regions, and locations of insertions and deletions (sometimes called "indels");

- build model backbone structure
    — copy predicted conserved main chain regions from template structure,
    — remodel loops with insertions or deletions;

- add side chains to the modelled main chain;

- evaluate and refine model.

Comparative modelling and fold recognition

Comparative modelling (homology modelling):
Given:
- sequence of target protein with unknown structure
- known structure of a related protein
Predict:
- three-dimensional structure of target protein

Fold recognition:
Given:
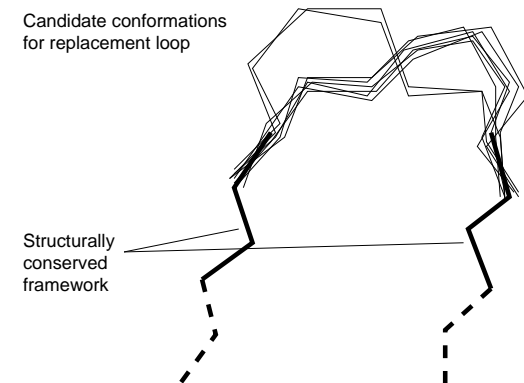- sequence of target protein with unknown structure
- library of known folds
Predict:
- known fold that is most compatible with the target protein's sequence

---

Fragment-fitting: an approach to remodelling loops

Candidate conformations for replacement loop

Structurally conserved framework

Fragment selection criteria

- steric overlap;

- packing
  — no protruding loops;
  — no internal cavities;

- disulphide bridges and salt bridges;

- solvent accessibility
  — avoid burying unpaired charges;

- sequence criteria
  — Gly and Pro residues
  — similarity between model's sequence and the sequences of the fragments in their native structures.

Energy calculations

Terms used in evaluating the energy of a conformation typically include:

- bond stretching
- bond angle bend
- terms penalising deviation from planarity, etc.
- torsion angles
- Van der Waals interactions
- hydrogen bonds
- electrostatics
- interactions with solvent, water and cosolutes

Side chain rotamers

There is an extremely large number of possible combinations of side chain conformations — infinite if we consider side-chain bonds to be continuously variable.

For practical purposes the search space can be discretised by considering a finite set of possible torsion angles for each side-chain.

The distribution of side chain conformations falls into statistically significant clusters. By using representative side chain conformations, or **rotamers**, the vast combinatorial search space can be greatly reduced.

Ponder, J.W. and Richards, F.M. (1987)
J. Mol. Biol., vol. 193, pp 775-791.

Fold recognition

The idea behind "threading":

Imagine a wire wound into the shape of a known protein's main chain "fold".

Imagine next that our new sequence is represented by beads that are "threaded", in order, onto the wire, and are pushed along the wire.

At each step, a score is calculated based on which residues are adjacent in space, which residues are buried, etc.

Repeat this process for each different known fold.

A high score indicates that the sequence is compatible with that fold.