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.

Using known substructures in protein crystallography

Jones, T.A. and Thirup, S. (1986)
The EMBO Journal, vol. 5, pp 819-822.

Electron density map interpretation is made easier by fitting regular α-helices and strands into the map.

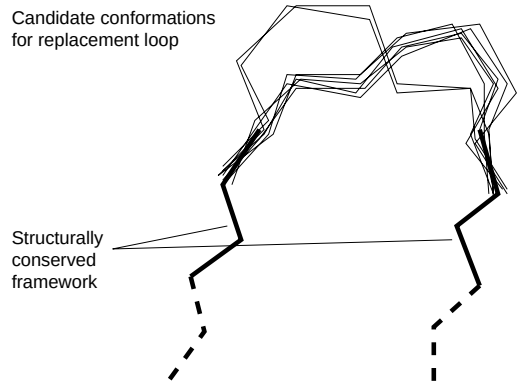
This building-block approach to protein modelling can be extended to include **all** main chain fragments.

For example, a model of retinol binding protein was built using fragments from only three other proteins. A model with Cα atoms matching within an R.M.S. error of 1Å was built using only 15 fragments.

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.

Fragment-fitting: an approach to remodelling loops



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.

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.

Cluster fragments

Kelley, L.A., Gardner, S.P. and Sutcliffe, M.J. (1996)
An Automated Approach For Clustering An Ensemble Of NMR-Derived Protein Structures Into Conformationally-Related Subfamilies.
Protein Engineering, vol. 9, pp 1063-1065.

RMS distances between pairs of structures or fragments.

Clusters built using average linkage.

Penalty function seeks to simultaneously minimise:

a) the number of clusters;
b) the spread across each cluster.

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