FACULTY OF ENGINEERING AND TECHNOLOGY

Department of Biotechnology

Tertiary structure predictions: Homology modeling,

abc

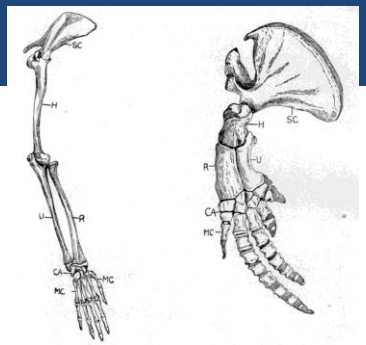
Homology Modeling

Based on two major observations:

1) **The structure of a protein is determined by its amino acid sequence**

2) **Structure is much more conserved than sequence during evolution**

What's homology modeling?

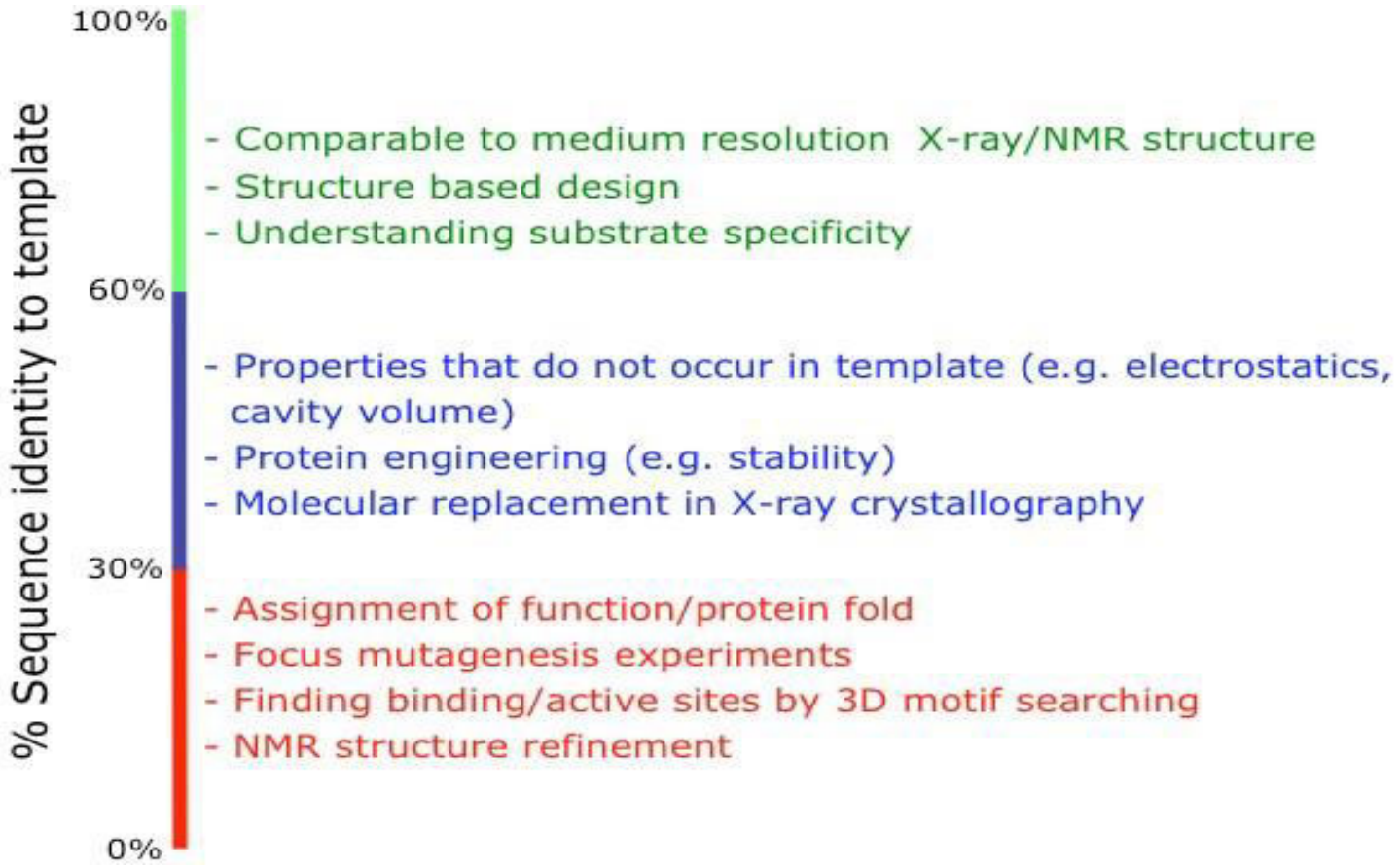


Predicts the three-dimensional structure of a given protein sequence (target) based on an alignment to one or more known protein structures (templates).

If similarity between the target sequence and the template sequence is detected, structural similarity can be assumed.

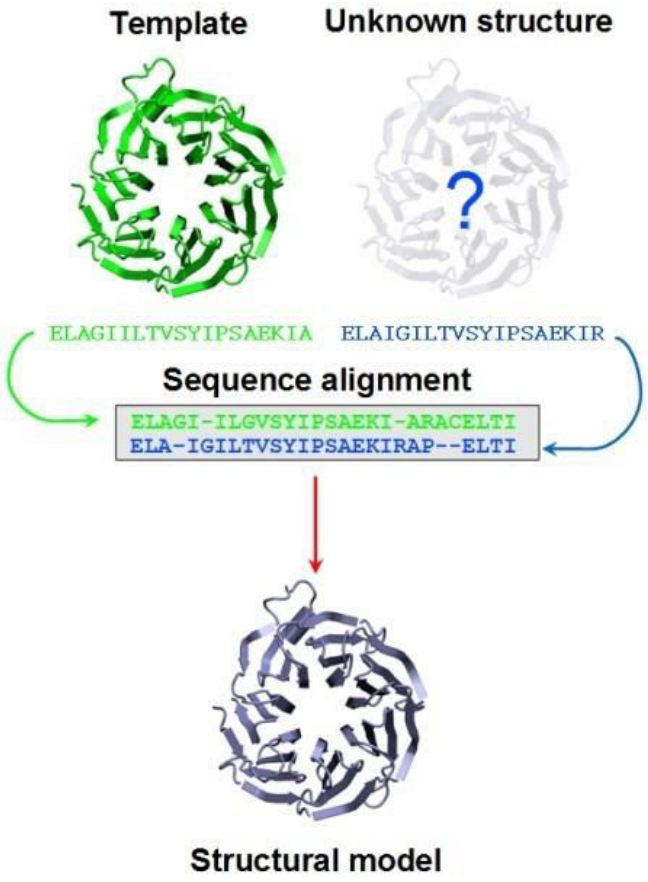
In general, 30% sequence identity is required to generate an useful model.

Accuracy and application of protein **structure**



In summary: homology modeling steps

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



Steps in Homology Modeling

Step 1: Template Recognition and Initial Alignment:-

In practice, one just feeds the query sequence to one of the countless BLAST servers on the web, selects a search of the PDB, and obtains a list of hits—the modeling templates and corresponding alignments

abc

2. An alignment matrix

By comparing thousands of sequences and sequence families, it became clear that the opening of gaps is about as unlikely as at least a couple of nonidentical residues in a row.

Opening penalty: for every new gap

Gap extension penalty: for every residue that is skipped in the alignment

abc

Step 2: Alignment Correction

Sometimes it may be difficult to align two sequences in a region where the percentage sequence identity is very low.

One can then use other sequences from homologous proteins to find a solution.

LTLTLTLT

TYTYTYTY

T YAYAYAYAVY

–LTLTLTLT–

TYTYTYTYT

–YAYAYAYAY

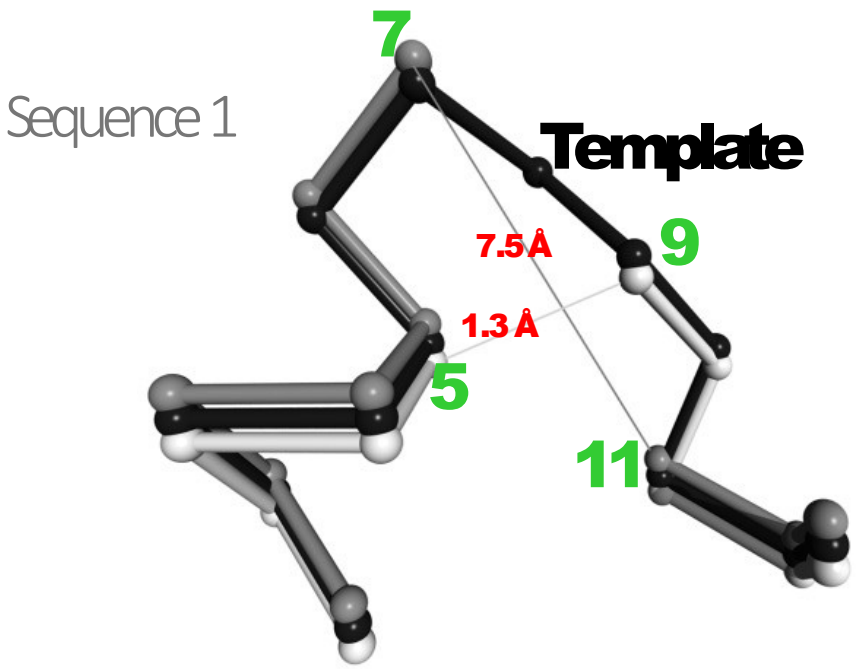
abc

		1	2	3	4	5	6	7	8	9	10	11	12	13
Template		PHE	ASP	ILE	CYS	ARG	LEU	PRO	GLY	SER	ALA	GLU	ALA	VAL
Model (bad)	1	PHE	ASN	VAL	CYS	ARG	ALA	PRO	---	---	---	GLU	ALA	ILE
Model (good)	2	PHE	ASN	VAL	CYS	ARG	---	---	---	ALA	PRO	GLU	ALA	ILE

2 is correct, because it leads to a small gap, compared to a huge hole associated with alignment 1.

XYZ

abc



abc

Step 3: Backbone Generation:-

One simply copies the coordinates of those template residues that show up in the alignment with the model sequence.

If two aligned residues differ, only the backbone coordinates (N, Cα, C and O) can be copied.

If they are the same, one can also include the side chain.

abc

Step 4: Loop Modeling:-

There are two main approaches to loop modeling:-

1). <u>Knowledge based</u>: one searches the PDB for known loops with endpoints that match the residues between which the loop has to be inserted and simply copies the loop conformation.

2). <u>Energy based</u>: as in true *ab initio* fold prediction, an energy function is used to judge the quality of a loop

abc

Step 5: Side-Chain Modeling:-

Comparing the side-chain conformations (rotamers) of residues that are conserved in structurally similar proteins

Similar torsion angle about the Cα –Cβ bond -

It is therefore possible to simply copy conserved residues entirely from the template to the model

-Rotamer libraries

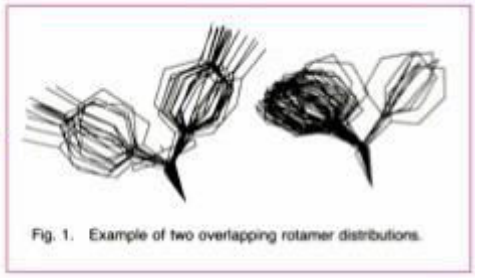
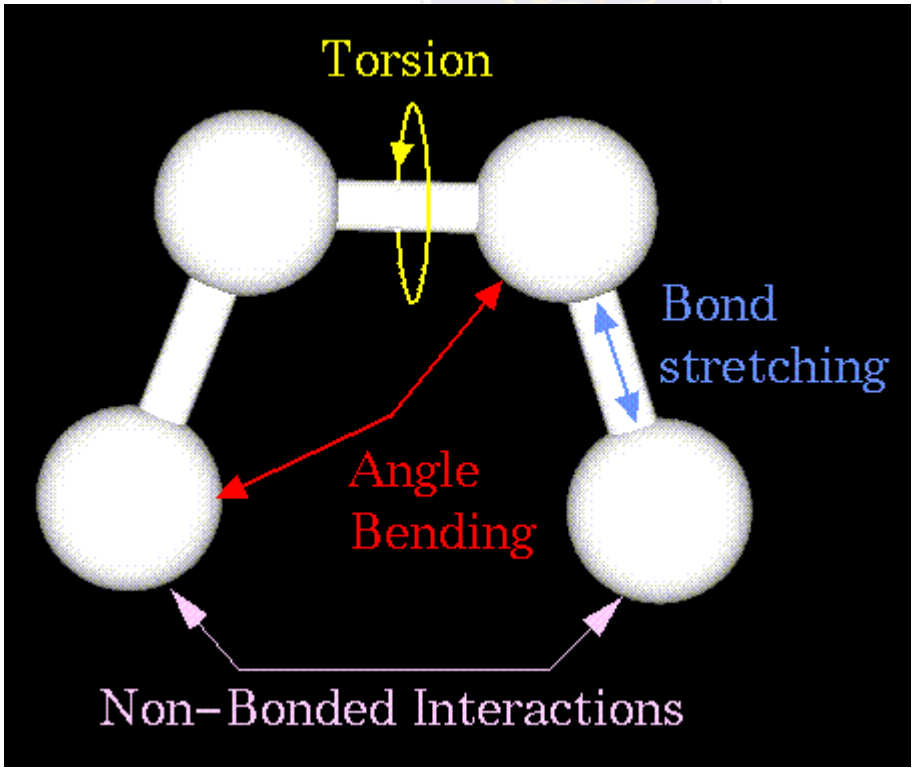


Fig. 1. Example of two overlapping rotamer distributions.

Step 6: Model Optimization:-

Energy = Stretching Energy +Bending Energy +Torsion Energy +Non-Bonded Interaction Energy

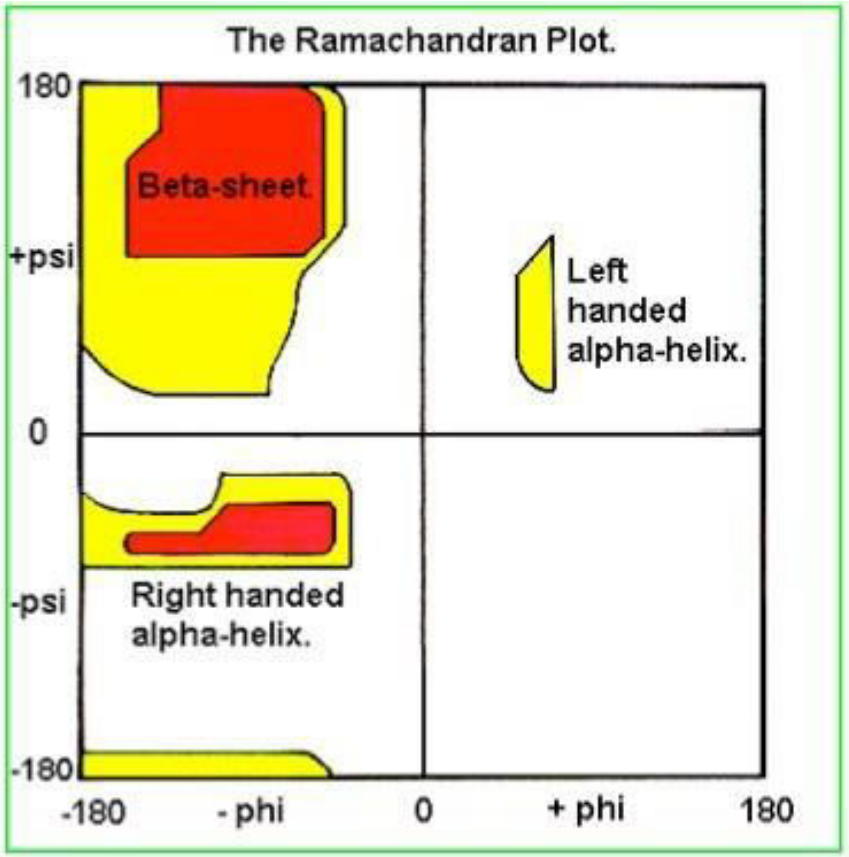


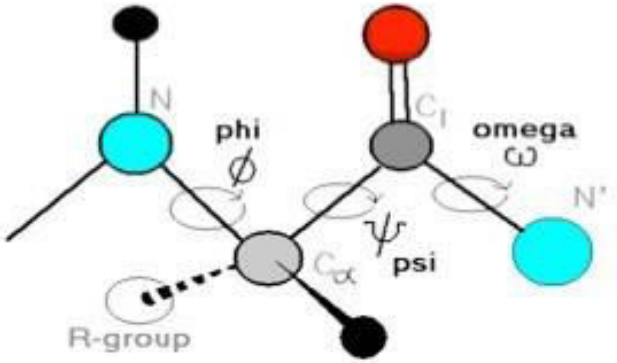
Step 7: Model Validation

Model should be evaluated for:

- **correctness of the overall fold/structure**
- **errors over localized regions**
- **stereochemical parameters: bond lengths, angles, etc**

Step 7: Model Validation





RAMA
UNIVERSITY
UTTAR PRADESH
KANPUR

MCQs

1. A
2. A
3. A
4. A
5. A
6. A
7. A
8. A
9. A
10. A