Molecular Modeling

Prediction of Protein 3D Structure from Sequence

Vimalkumar Velayudhan

Jain Institute of Vocational and Advanced Studies

May 21, 2007

Outline

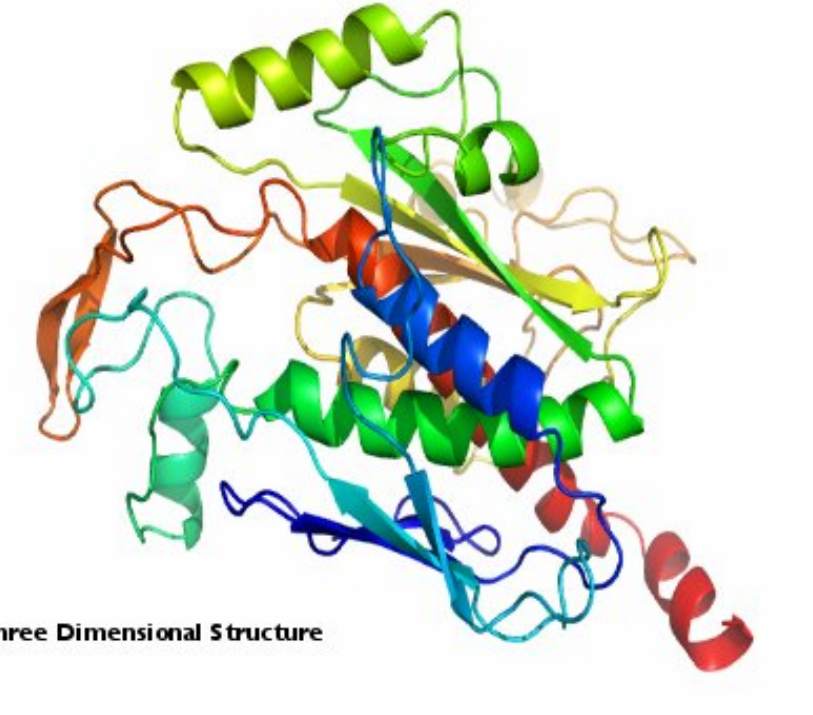
What is Molecular Modeling?

- The Prediction of a protein's three dimensional structure from its sequence
- A computational method based on our understanding of protein structures

Principle

The Sequence of a protein ie., the order of amino acids determine the 3D structure of the protein and hence, its function

Amino acid sequence	Gly	Phe	Leu	Gly	Ala	Ala	Gly	Ser	Thr	Met	Gly	Ala



Three Dimensional Structure

Methods in Molecular Modeling

Methods in Molecular Modeling

1. Homology Modeling[1]
2. Threading of Fold Recognition
3. *ab initio* Prediction

Accuracy

The accuracy of the methods are in the following order
Homology Modeling $>$ Threading $>$ *ab initio* Prediction

[1] Also known as Comparative Modeling

Homology Modeling

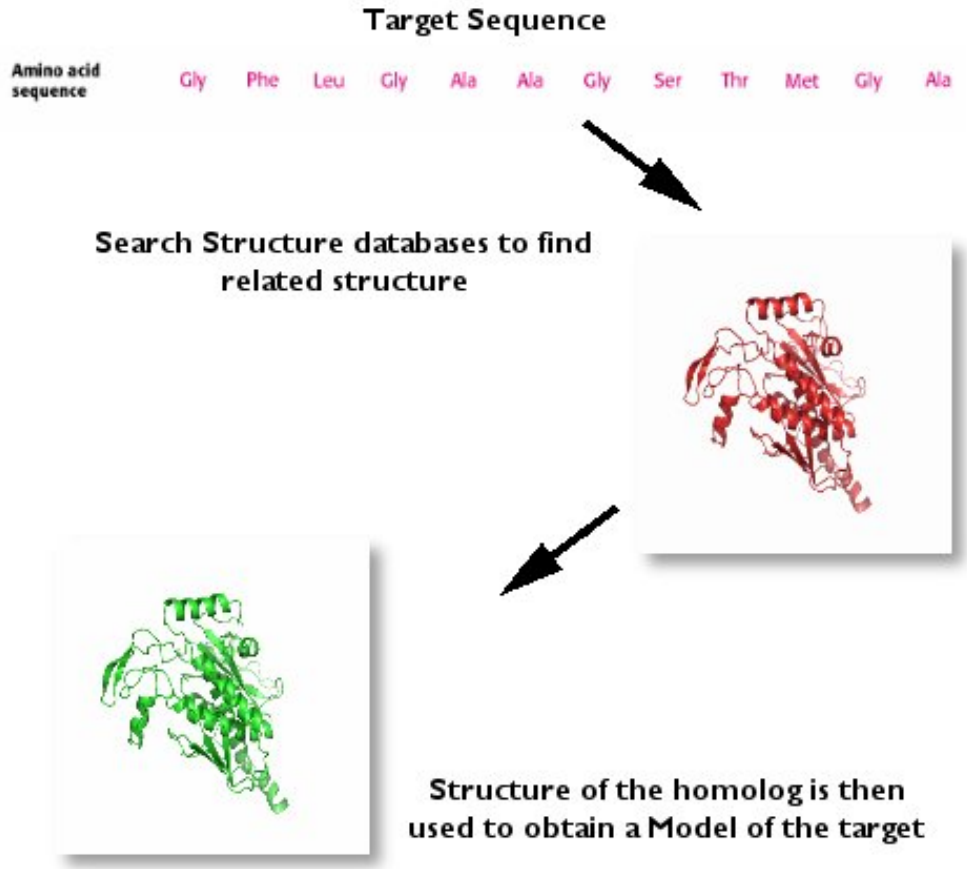
- Predicting the structure of a protein using the structure of its homolog[2]
- The protein whose structure is not known is referred to as the target and the structure of the homolog is referred to as the template

Basic Principle

Similar Sequences tend to adopt Similar Structures

[2] Proteins which share a common ancestor in evolution

Principle


Target Sequence
Amino acid sequence
Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly Ala
Search Structure databases to find related structure
Structure of the homolog is then used to obtain a Model of the target

Steps

Given the sequence of a protein, its structure can be predicted in the following steps

1. Template detection
2. Target –Template Alignment
3. Backbone generation
4. Modeling of Side-chains and Loops
5. Model Validation and Optimization

Template detection

- A template is a protein whose structure is already known[3]
- Target refers to the protein we would like to model - whose structure is unknown

Template Detection refers to the identification of a suitable template corresponding to the target by database similarity searches

- Can be performed by doing a BLAST[4] search[5] against the PDB[6] database

[3] by X-ray Crystallography, NMR or other techniques

[4] A program to search a database with a sequence to identify related sequences

[5] blastp - Protein-Protein BLAST

[6] Protein Data Bank - a worldwide repository for protein structures

Template detection

Target Protein Sequence
(Obtained from database like Swiss-Prot)

```
>CYB_HUMAN P00156 Cytochrome b.
MTPMRKINPLMKLINHSFIDLPTPSNISAWWNFGSLLGACLILQITTGLFLAMHYSPDAS
TAFSSIAHITRDVNYGWIIRYLHANGASMFFICLFLHIGRGLYYGSFLYSETWNIGIILL
LATMATAFMGYVLPWGQMSFWGATVITNLLSAIPYIGTDLVQWIWGGYSVDSPTLTRFFT
FHFILPFIIAALATLHLLFLHETGSNNPLGITSHSDKITFHPYYTIKDALGLLLFLLSLM
TLTLFSPDLLGDPDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILI
LAMIPILHMSKQQSMMFRPLSQSLYWLLAADLLILTWIGGQPVSYPFTIIGQVASVLYFT
TILILMPTISLIENKMLKWA
```

Template detection

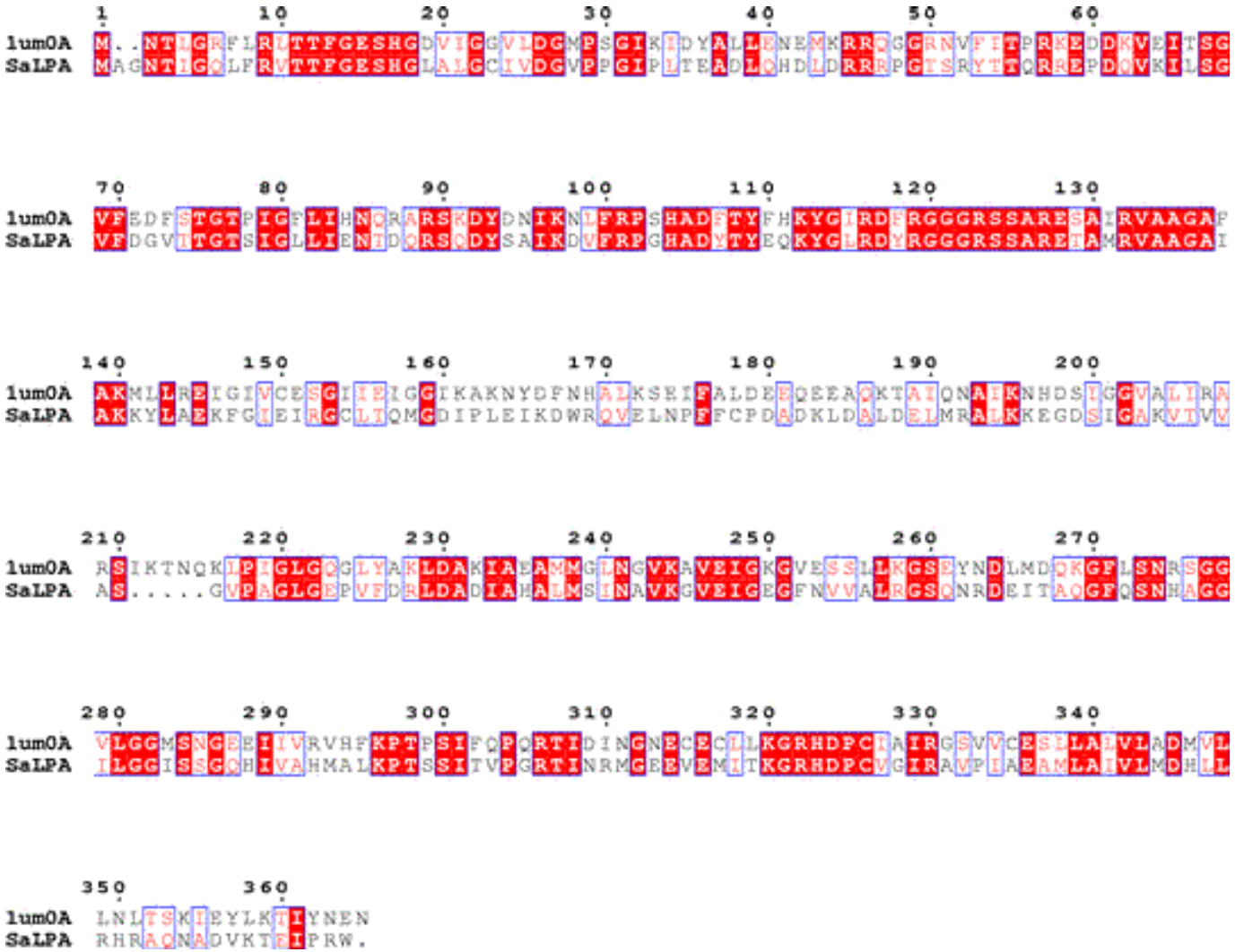
- Atleast 25% sequence identity is required between the query and the subjects from PDB
- A number of other parameters are also involved in the selecting the right template
  - Resolution - the higher[7] the better
  - The template should cover the entire length of the target
  - Gaps should be minimal

[7] Higher resolution will correspond to a low numeric value - Ex., 1.0 is better than 2.0

Target-Template Alignment

- An alignment is necessary to state which residues in the target correspond to which residue in the template
- The model-building program uses this information to build the backbone of the target
- Can be performed using programs like ClustalX, T-Coffee or the Modeling program itself

Target-Template Alignment

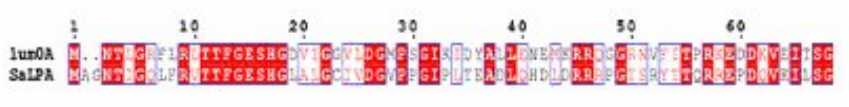


Backbone Generation

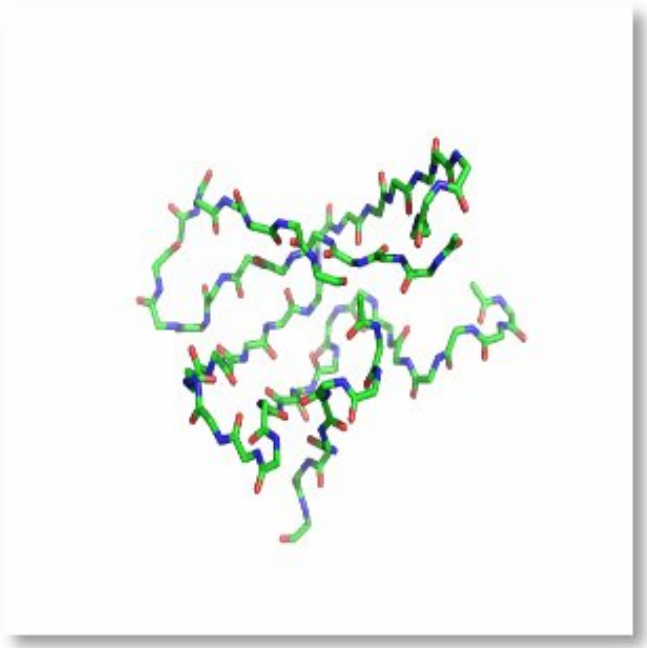
- Based on the information in the alignment, the model-building program generates the backbone of the target
- Backbone refers to the repeating N - C_α- C atoms of the polypeptide chain

Backbone Generation

Target - Template Alignment



Using the Alignment,
A backbone of the target
is generated

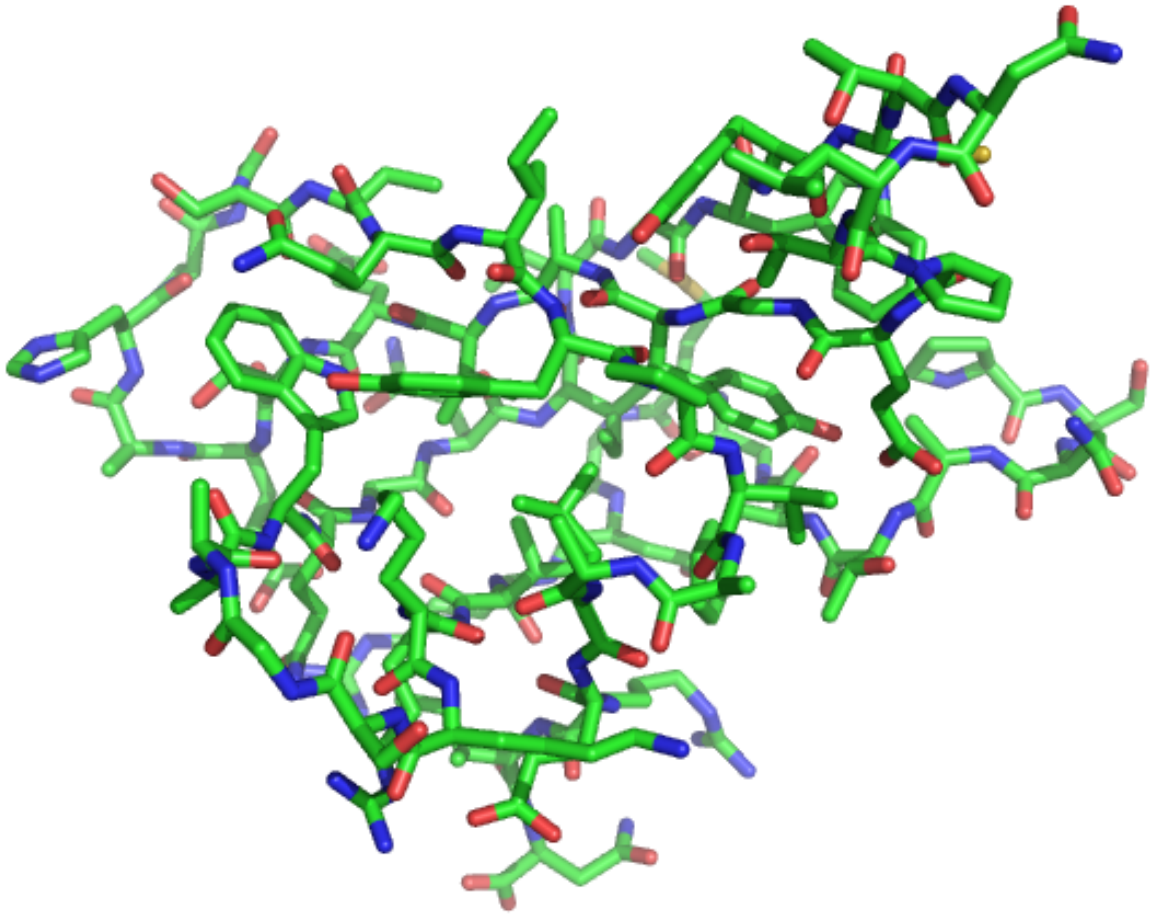


Modeling of Side-Chains and Loops

- Once the backbone is built, the side-chains of the residues are then modeled by
  - Using the information from the alignment if the alignment is conserved
  - If residues are different, an inbuilt library of side-chain conformations[8] is consulted
- The same principle is applied to model loop regions too

[8] Referred to as Side-chain rotamer library

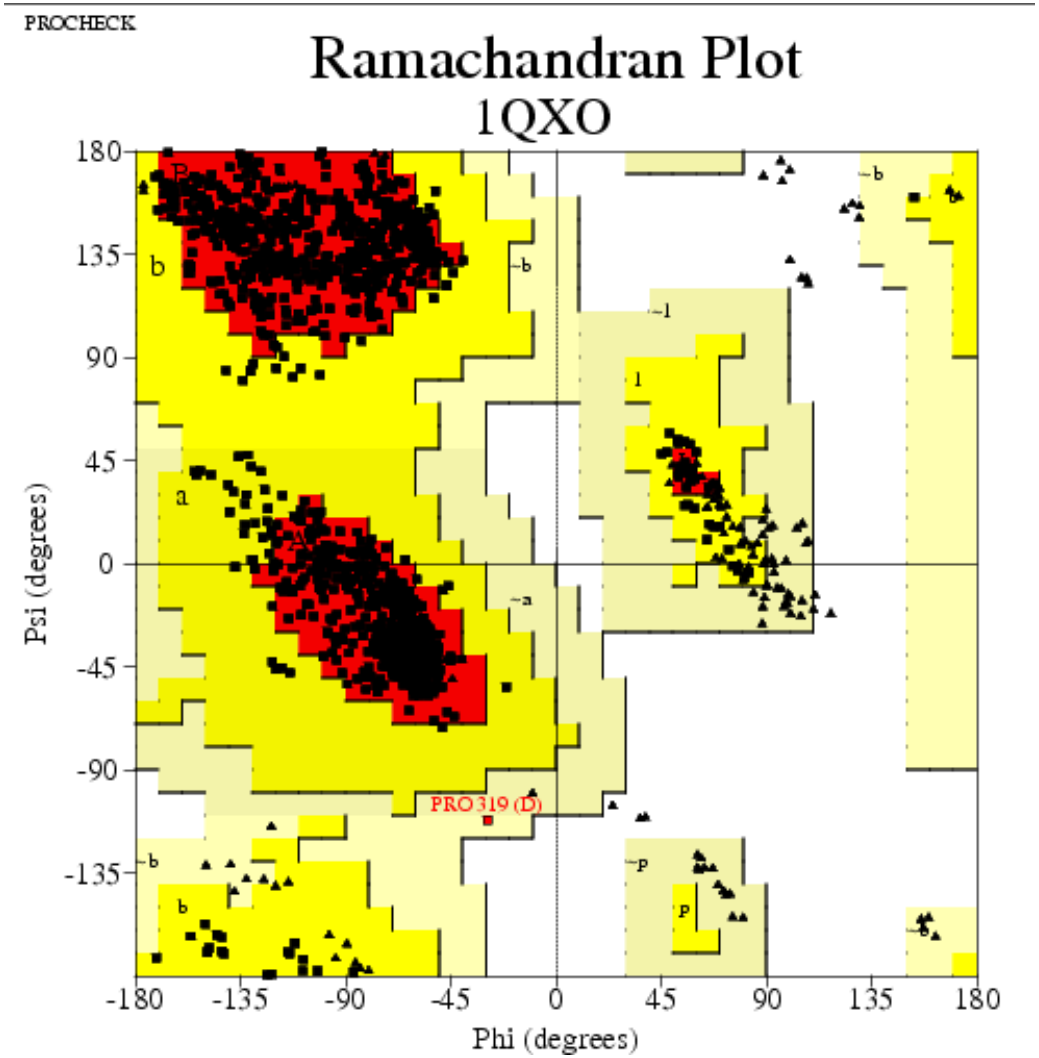
Modeling of Side-Chains and Loops



Model Validation and Optimization

- After the backbone, sidechains and loop regions have been modelled
- The model is checked for its quality
  - Done by assessing parameters like bond-lengths, bond-angles etc.,
- If there are problems at particular regions, the alignment can be verified and adjusted if required
- The steps are performed once again until a satisfactory model is obtained

Model Validation and Optimization



Guidelines

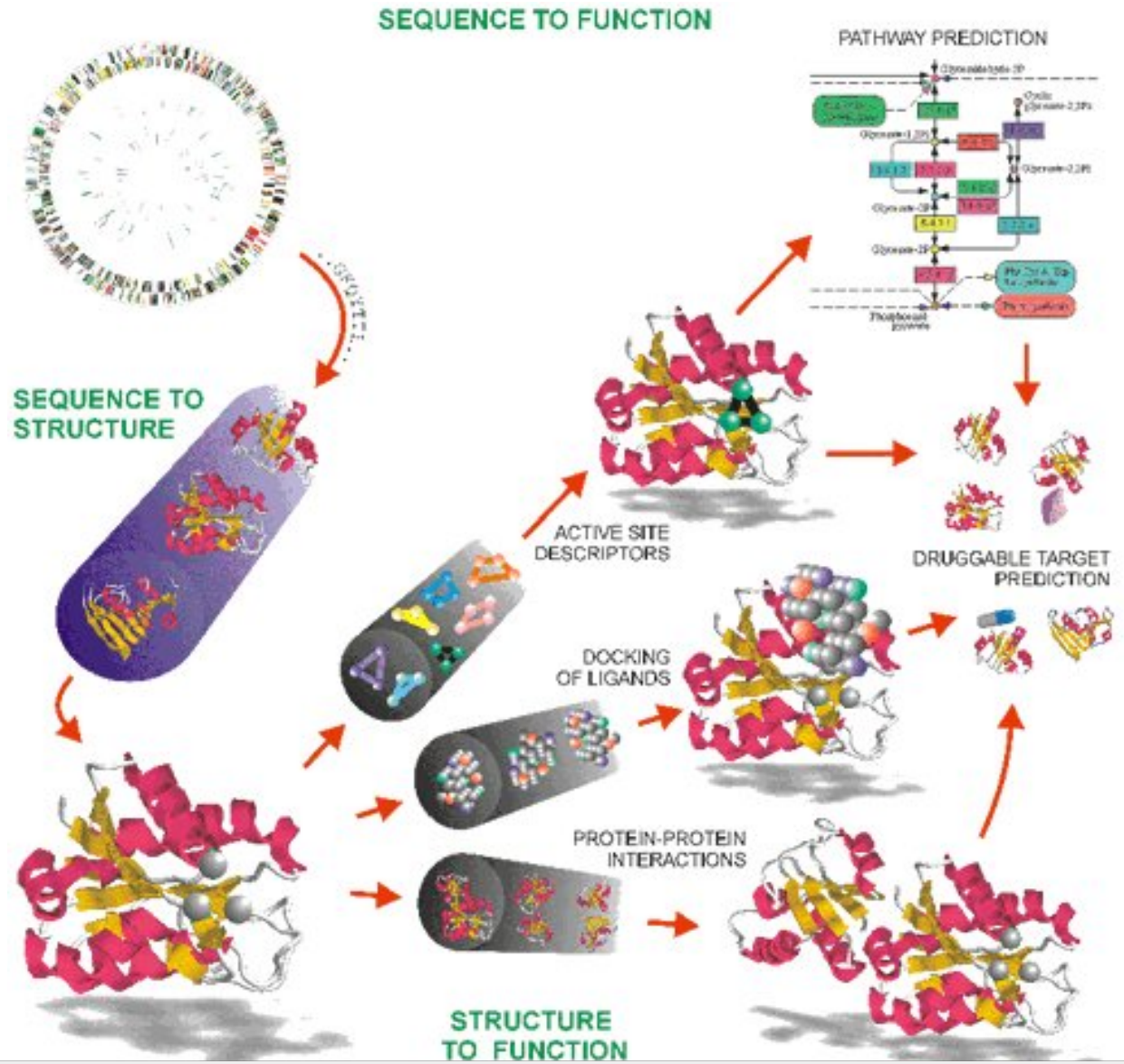
Alignment The most important step! - incorrect alignment will result in a model with errors

Quality of the template - good resolution and stereochemical quality

Percentage Identity between the target and template - the higher the better [9]

[9] 25% is the minimum

Applications



Modeling Programs

- Academic
  - MODELLER - http://www.salilab.org
  - Deepview and SWISS-MODEL - http://www.expasy.org/spdbv
  - CPH-Models - http://www.cbs.dtu.dk/services
- Commercial
  - Accelrys Insight II - http://www.accelrys.com

Further Reading

1. Structural Bioinformatics - Philip E Bourne, John Wiley & Sons Publications
2. Bioinformatics - From Genomes to Drugs, Thomas Lengauer, Wiley publications