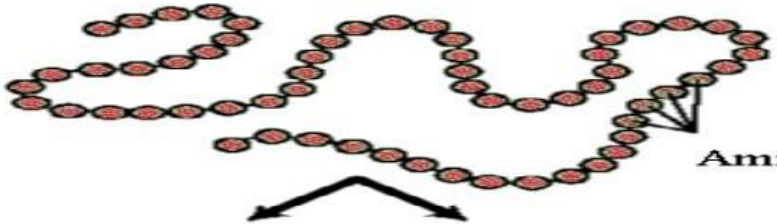
In silico protein structure prediction

Different Levels of Protein Structure



Primary protein structure
is sequence of amino acids

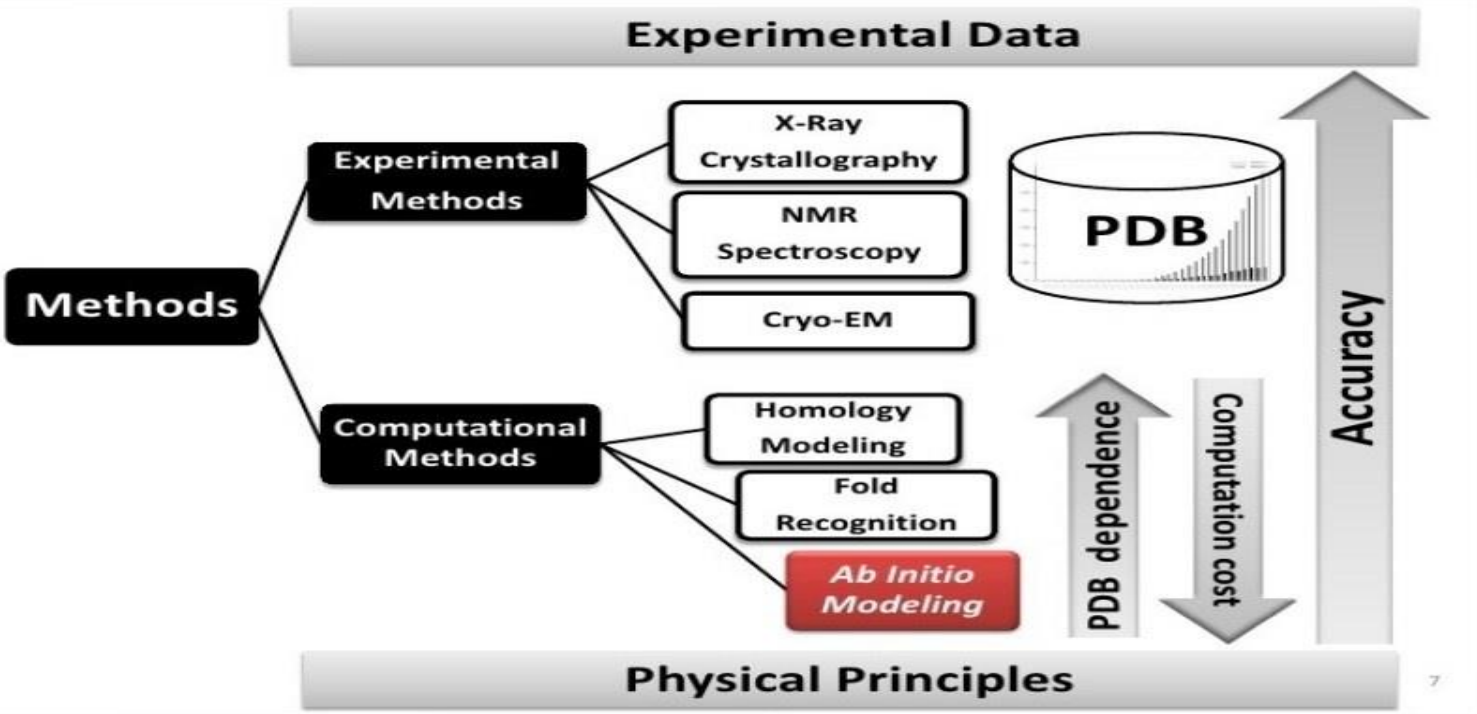
Secondary protein structure
local conformation
primarily stabilized by
hydrogen bonding

Tertiary protein structure
three dimensional conformation

Quaternary protein structure
Combination of multiple
polypeptide chains

METHODS FOR PROTEIN STRUCTURE PREDICTION

- 1. EXPERIMANTAL METHODS
- 2. COMPUTIONAL METHODS



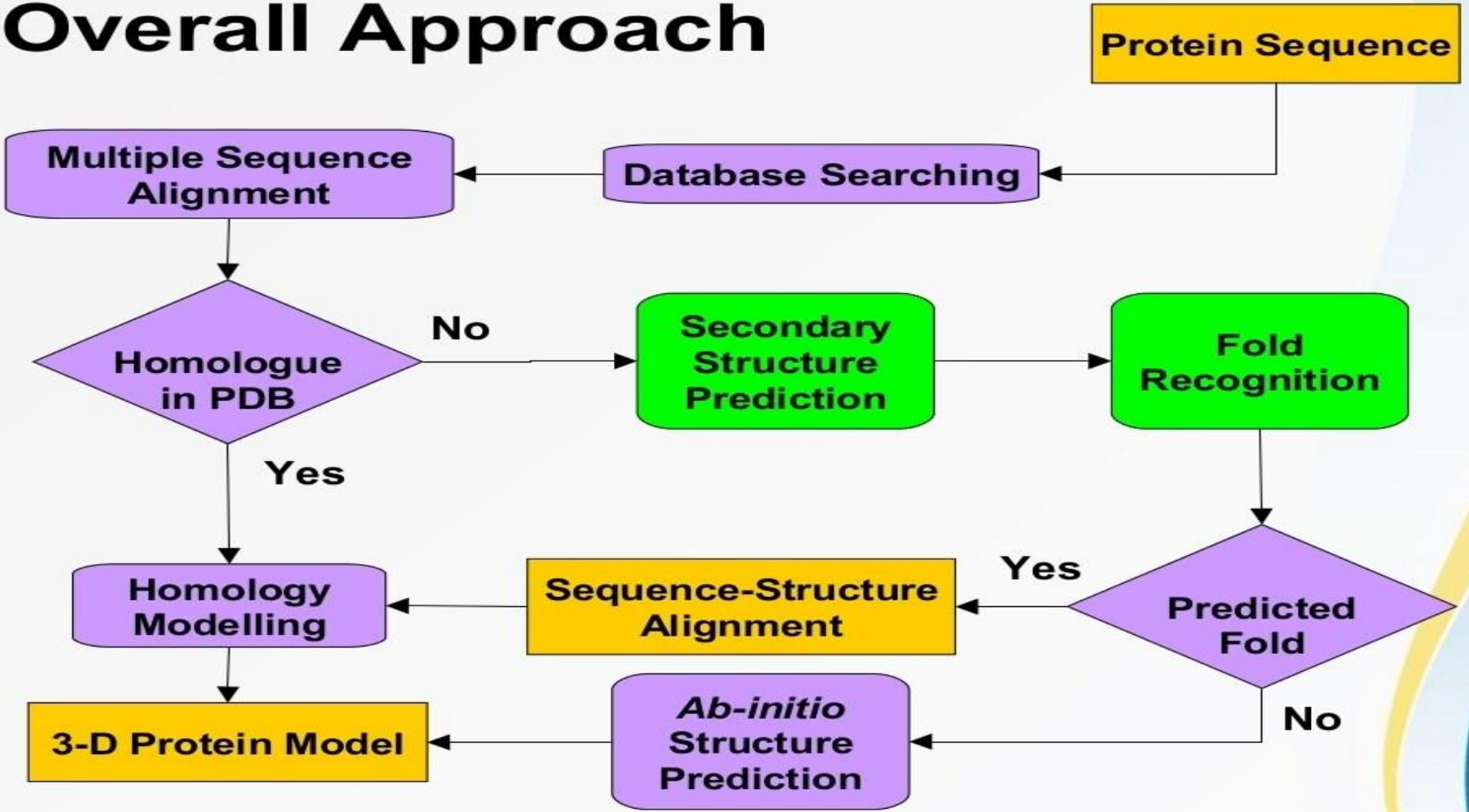
Experimental Protein Structure Determination

- X-ray crystallography
  - most accurate
  - in vitro
  - needs crystals
  - ~$100-200K per structure
  - time consuming and expansive.
- NMR
  - fairly accurate
  - *in vivo*
  - no need for crystals
  - limited to very small proteins
  - time consuming and hardly .
- Electron-microscopy
  - imaging technology
  - low resolution
  - not more observable.

Computational method

- Major Techniques
  - Template Modeling
    - Homology Modeling
    - Threading
    - Both are use known protein structure
  - Template-Free Modeling
    - *ab initio* Methods
      - Physics-Based
      - Knowledge-Based
      - without use of known protein structure

Overall Approach



Homology Modelling

- also called comparitive modeling.
- predict protein structures based on sequence homology with known structure.
- ***Principle:-***
- if two proteins share a high enough sequence similarity,they are likely to have very similar three dimensional structure.
- modeling server:-modbase,swiss-model etc.
- Fail in absence of homology

Homology Modelling

- six steps:-

1. template selection (BLAST and FASTA)
2. sequence alignment (T-coffee and PRALINE)
3. model building (CODA)

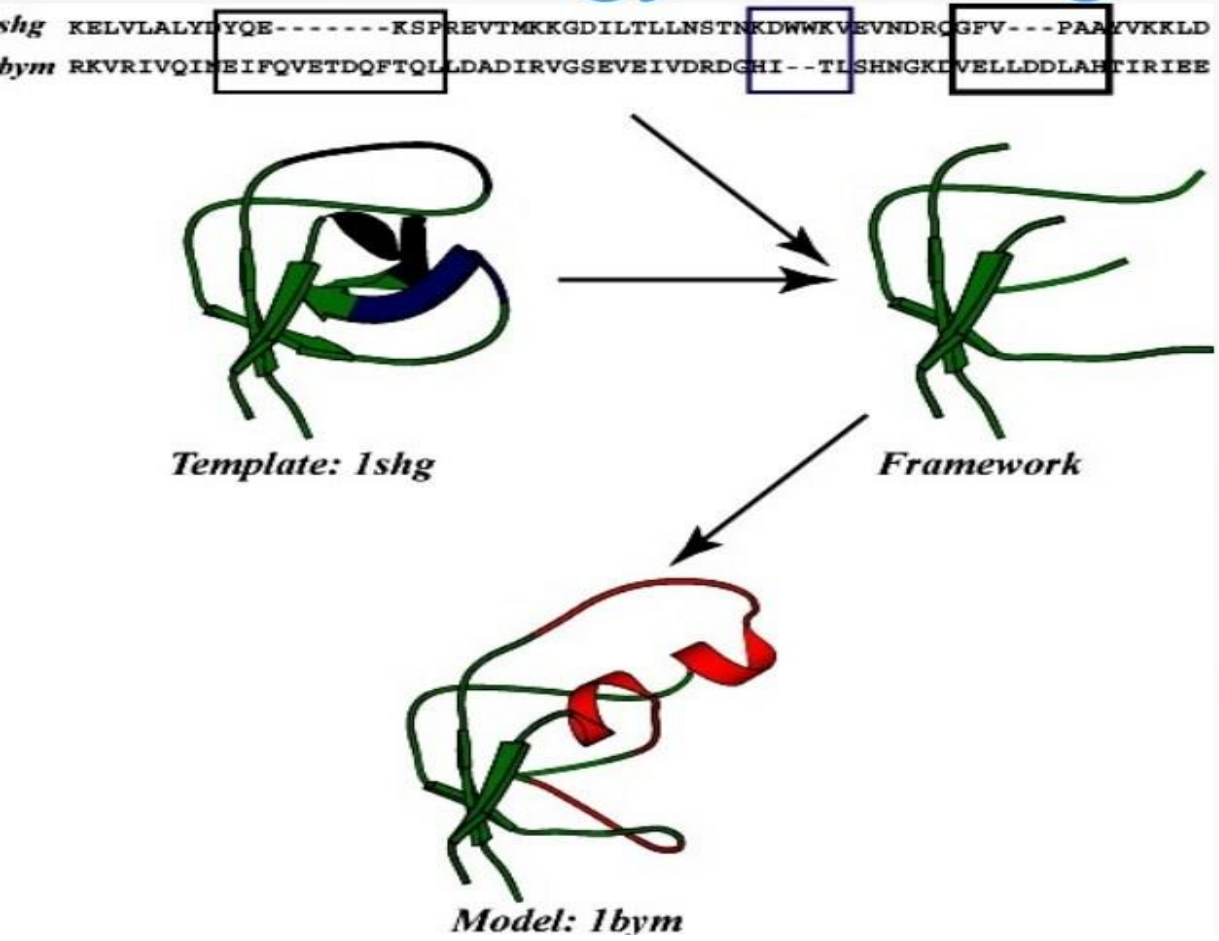
(a)backbone model building

(b)loop modeling

4. side chain refinement (SCWRL)
5. model refinement using energy function (GROMOS)
6. model evalution (PROCHECK & WHAT IF)

- Has been observed that even proteins with 30% sequence identity fold into similar structures
- Does not work for remote homologs (< 30% pairwise identity)

Homology Modeling: How it works



- *Find template*
- *Align target sequence with template*
- *Generate model:*
  - *add loops*
  - *add sidechains*
- *Refine model*

Threading

- Given:
  - sequence of protein 'P 'with unknown structure
  - Database of known folds
- Find:
  - Most plausible fold for 'P'
  - Evaluate quality of such arrangement
- Places the residues of unknown 'P' along the backbone of a known structure and determines stability of side chains in that arrangement

Threading and fold recognition

- predicts the structural fold of unknown protein sequences by fitting the sequence into a structural database and selecting the best fitting fold.
- we can identify structurally similar proteins even without detectablesequence similarity.
- two algorithms:-

1.pairwise energy based method

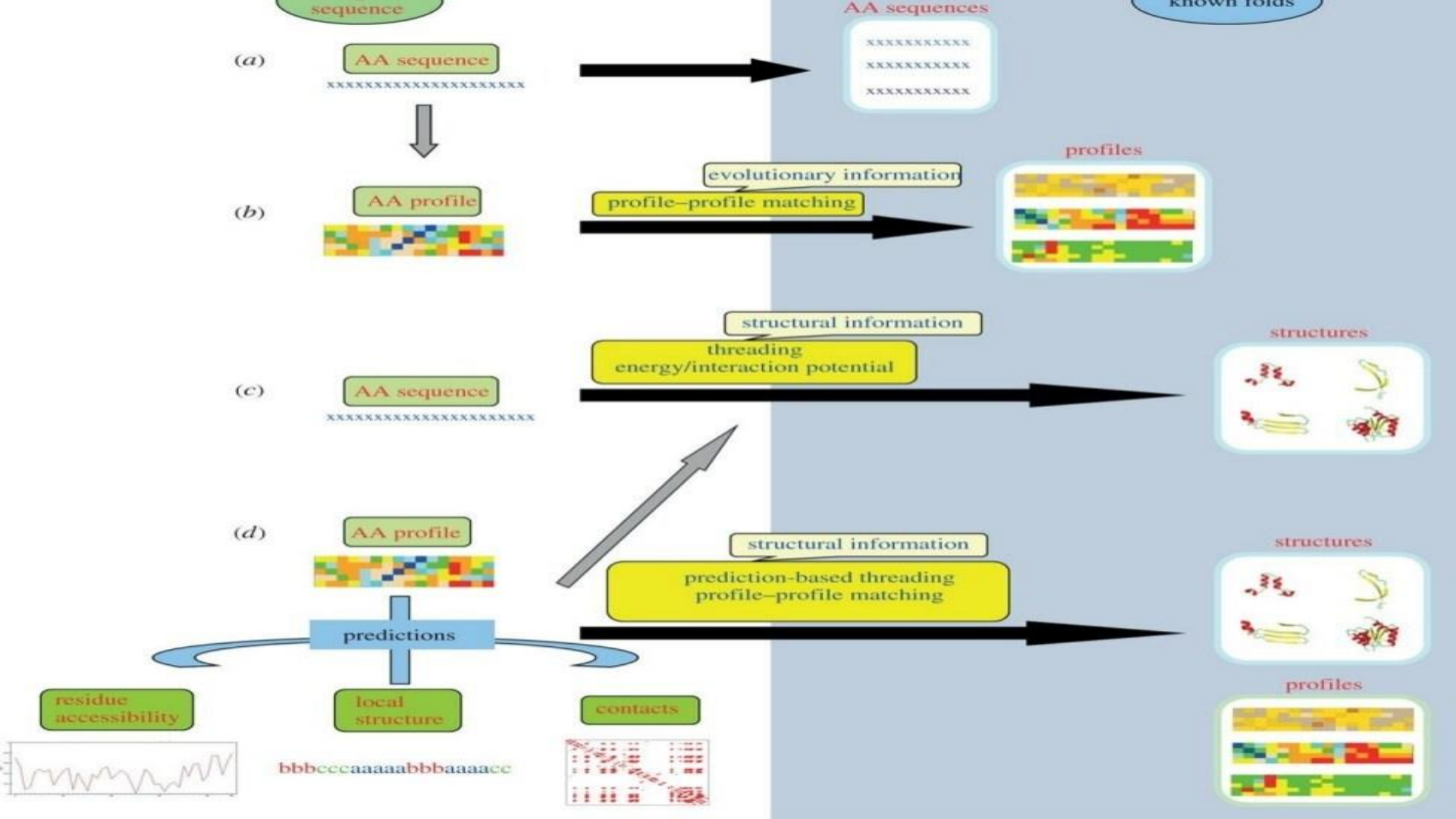
2.profile based method

1. ***Pairwise energy based method (threading)***

- Searched for a structural fold database to find the best matching structural fold using energy based criteria.
- Using dynamic programming and heuristic approaches.
- Calculate energy for raw model.
- Lowest energy fold that correspond to the structurally a group of most compatible fold.

2.profile based method (fold recognition)

- A profile is constructed for related protein structures.
- Generated by superimposition of the structures to expose corresponding residues.
- Secondary structure type,polarity,hydrophobicity.
- The protein fold to be predicted does not exist in the fold library,method will fail.


sequence
AA sequences
known folds
(a)
AA sequence
xxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxx
xxxxxxxxxxx
xxxxxxxxxxx
(b)
AA profile
evolutionary information
profile–profile matching
profiles
(c)
AA sequence
xxxxxxxxxxxxxxxxxxxx
structural information
threading
energy/interaction potential
structures
(d)
AA profile
predictions
structural information
prediction-based threading
profile–profile matching
structures
profiles
residue accessibility
local structure
contacts
bbbcccaaaaabbbaaaaacc

Ab Initio Protein Structure Prediction

- ***Ab initio* protein structure prediction** methods build protein 3D structures from sequence based on physical principles.
- **Importance**
  - The *ab initio* methods are important even though they are computationally demanding
  - *Ab initio* methods predict protein structure based on physical models, they are indispensable complementary methods to Knowledge-based approach.

    Knowledge-based approach would fail in following conditions:
    - Structure homologues are not available
    - Possible undiscovered new fold exists.

    ➢ **Anfinsen's theory**: Protein native structure corresponds to the state with the lowest free energy of the protein-solvent system.

Rosetta

- web server for protein 3d structure prediction.
- mini threading method.

➢ breaks down the quary sequence into many short segments (3 to 9).

➢ predicts the secondary structure of small segments using HMMSTR.

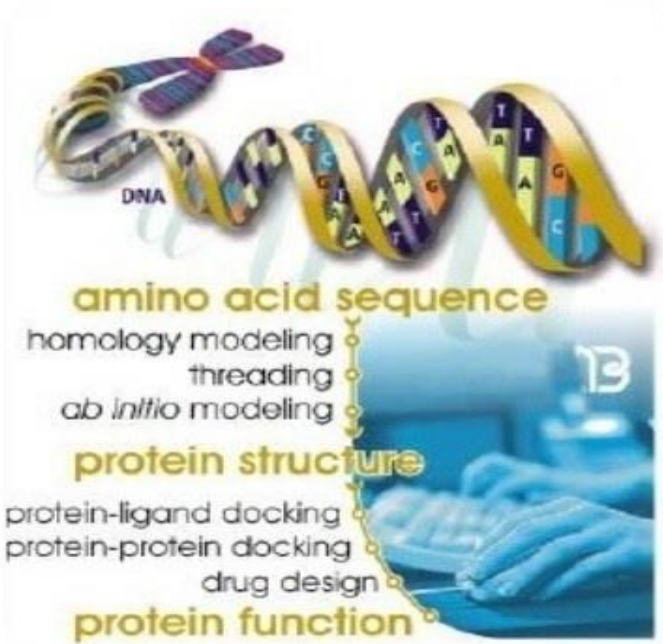
➢ Segments with assigned secondary structure are subsequently assembled into a 3D configuration.

➢ random combination of fragments, a large number of models are built and their over all energy potential calculated.

➢ conformation with lowest free energy is choosen as the best model.

Structure Prediction Approaches

- ***Ab-initio* fold prediction**
  - Not based on similarity to a sequence- structure
- **Threading (Fold Recognition)**
  - Requires a structure similar to a known structure
- **Homology modeling**
  - Based on sequence similarity with a protein for which a structure has been solved.
- **CASP**



Reference

- http://www.golujain/biology-protein-structure-in-cloud-computing
- http://infoman.teikav.edu.gr/~stpapad/EssentialBioinformatics.pdf
- https://en.wikipedia.org/wiki/Ab_Initio_Software
- Dictionary of bioinformatics/ by k. manikandakumar.