BINF 731

Protein Structure Analysis

Iosif Vaisman

2020

Protein Modeling Methods

- ***Ab initio* methods:**
  solution of a protein folding problem
  search in conformational space
- **Energy-based methods:**
  energy minimization
  molecular simulation
- **Knowledge-based methods:**
  homology modeling
  fold recogniion

Knowledge

Knowledge is a.pattern that exceeds certain threshold of interestingness.

Factors that contribute to interestingness:
  coverage
  confidence
  statistical significance
  simplicity
  unexpectedness
  actionability

Knowledge-based methods

Finding patterns in known structures

Deriving rules (usually in the form of PMF)

Applying the rules

Fold Recognition

Pattern searching
  sequence patterns
  structure patterns
  residue composition patterns

Threading
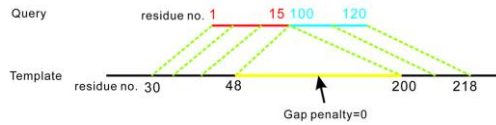 sequence-structure compatibility
  structure-sequence compatibility
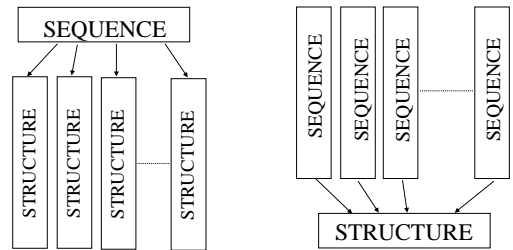
Threading

Sequence-structure
compatibility
(fold recognition)

Structure-sequence
compatibility
(inverse folding)

Segmental Threading

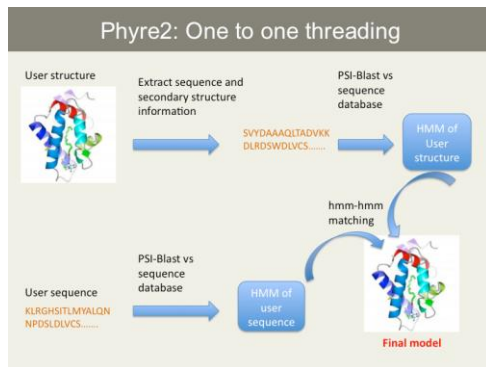Query   residue no. 1 15 100 120

Template   residue no. 30   48   200 218

Gap penalty=0

Segmental Threading

Continuous segments   **Discontinuous segments**

RMSD=5.5Å   RMSD=0.9Å   RMSD=8.5Å   RMSD=2.2Å

RMSD=2.4Å   RMSD=0.3Å   RMSD=3.0Å   RMSD=0.7Å

RMSD=12.2Å   RMSD=2.6Å   RMSD=3.5Å   RMSD=1.2Å

One to one threading

Phyre2: One to one threading

User structure → Extract sequence and secondary structure information → PSI-Blast vs sequence database

SVYDAAAQLTADVKK DLRDSWDLVCS......

HMM of User structure

hmm-hmm matching

User sequence → PSI-Blast vs sequence database → HMM of user sequence

KLRGHSITLMYALQN NPDSLDLVCS......
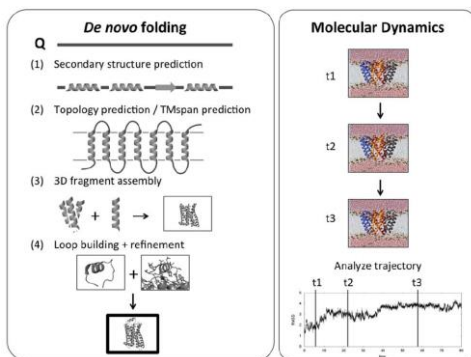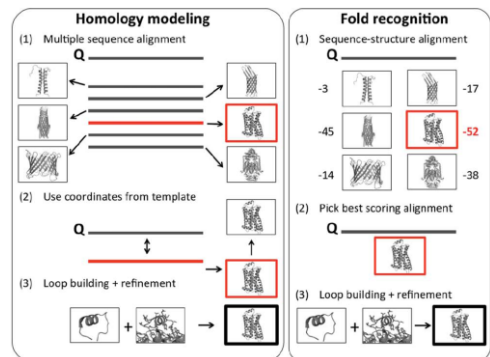
Final model

Threading

- Only the local environment is taken into account
- Non-local contacts are assumed with generic peptide
- No gaps are allowed in the alignment

Protein modeling techniques

De novo folding

Q
(1) Secondary structure prediction
(2) Topology prediction / TMspan prediction
(3) 3D fragment assembly
(4) Loop building + refinement

Molecular Dynamics

t1

t2

t3

Analyze trajectory

t1 t2 t3

Protein modeling techniques

Homology modeling

(1) Multiple sequence alignment
Q

(2) Use coordinates from template
Q

(3) Loop building + refinement

Fold recognition

(1) Sequence-structure alignment
Q
-3   -17
-45   -52
-14   -38

(2) Pick best scoring alignment
Q

(3) Loop building + refinement

Homology Modeling

- Identification of structurally conserved regions (using multiple alignment)
- Backbone construction (based on SCR)
- Loop construction (KB or conformational search)
- Side-chain restoration (KB, rotamer, or MM)
- Structure verification and evaluation
- Structure refinement (energy minimization)

Swiss-Model

- Method:
  Knowledge-based approach.

- Requirements:
  At least one known 3D-structure of a related protein.
  Good quality sequence alignements.

- Procedures:
  Superposition of related 3D-structures.
  Generation of a multiple a alignement.
  Generation of a framework for the new sequence.
  Rebuild lacking loops.
  Complete and correct backbone.
  Correct and rebuild side chains.
  Verify model structure quality and check packing.
  Refine structure by energy minimisation and molecular dynamics.