Protein Structure Analysis

Iosif Vaisman

2012

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 recognition

Ab initio Methods

Simplified models
- simplified alphabet (HP)
- simplified representation (lattice)

Build-up techniques
- quantum mechanics
- diffusion equations
- DFT

Stochastic searches
- Monte Carlo
- genetic algorithms

Genetic Algorithms Applications

HP Lattice Models

<table>
<thead>
<tr>
<th>Parents</th>
<th>Children</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 00 01 00 10</td>
<td>10 00 00 01 11</td>
</tr>
<tr>
<td>10 00 01 11</td>
<td>10 00 00 10</td>
</tr>
</tbody>
</table>

HP model
- HH $\varepsilon < 0$
- HP $\varepsilon = 0$
- PP $\varepsilon = 0$

HP+ model
- HH $\varepsilon < 0$
- HP $\varepsilon = 0$
- PP $\varepsilon = 0$
- NNC $\varepsilon > 0$
HP Lattice Models

Folding pathways

Hierarchical *ab initio* prediction

Lattice models
Knowledge-based scoring functions

*Ab initio* prediction using Rosetta

Target 77

Target 74

Target 56

Target 79

Ab initio prediction using Robetta

Quantum Chemistry Refinement of Protein Structures
**Density Functional Theory**

HK theorem: Each local one-particle potential corresponds to exactly one ground state density.

**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 recognition

**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**

- Wu and Zhang, 2010

**Threading**

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

**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 alignments.

- **Procedures:**
  Superposition of related 3D-structures.
  Generation of a multiple alignment.
  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.