RaptorX: A Multiple-Template Approach for Protein Threading

Jian Peng and Jinbo Xu
Toyota Technological Institute at Chicago

Template-based Modeling for Protein Structure Prediction

Observation
- ~70,000 proteins in the Protein Data Bank (PDB)
- ~1000-1500 unique structural folds

Methodology
- Use known structures to predict the structure of a new protein

Query Sequence: DRVYIHPFADRVYIHPFA

The Best Match (alignment)
The Multiple-Template Approach

- Single template cannot always cover the whole sequence
- Template selection is difficult
- Alignments are usually noisy
- Hard to build a 3D model from inconsistent alignments

Our Approach

- Probabilistic-consistency refinement with template similarity
- Correct errors in alignments
- Generate a consistent multiple template/sequence alignment

Overall Algorithm

1. Input target sequence
2. Run BoostThreader to select top templates
3. Generate a probabilistic alignment matrix between any two templates using TMalign/MAtr
4. Generate a probabilistic alignment matrix for the target and each template using BoostThreader
5. Run probabilistic-consistency transformation to update probabilistic alignment matrices
6. Run progressive alignment and refinement to generate a multiple alignment of target and templates
Blind Evaluation in CASP9

50 highest TBM targets
direct head-to-head comparison only on GDT

- RaptorX-MSA
- RaptorX
- RaptorX-Boost

30min for a protein with 200 amino acids
3D & 2D structure prediction
Protein Alignment
Functional prediction
Updated Database
Email Notification

RaptorX Webserver

- raptorx.uchicago.edu
- 30min for a protein with 200 amino acids
- 3D & 2D structure prediction
- Protein Alignment
- Functional prediction
- Updated Database
- Email Notification