Homology Modeling

Benedikt Zacher, Robert Greil

Homology Modeling – Part 1: Template Recognition

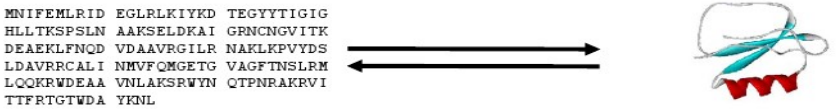
1. Overview over basic steps
2. Sequence similarity and Homology: The HSSP-curve
3. Overview over template recognition in HHpred, SWISS-MODEL and iTASSER
4. Backbone generation

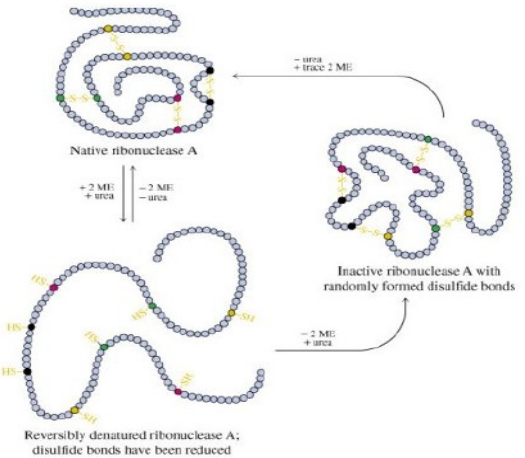
Introduction

- use structure of homolgous sequences for preditction
- Homology Modeling is based on two assumptions/observations:
  - The structure of a protein is uniquely determined by its amino acid sequence (Anfinson's paradigm)
  - Structure is more

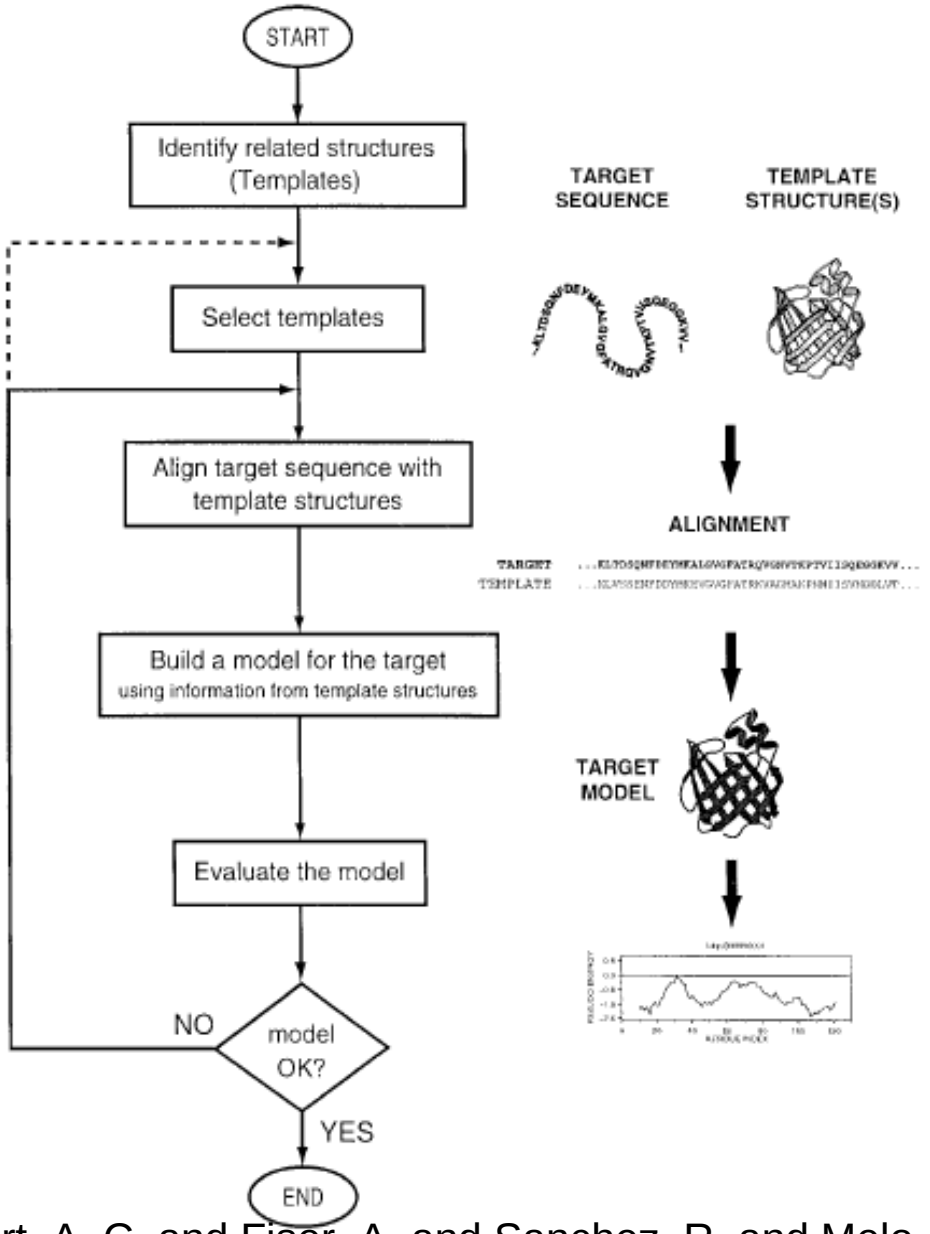
    evolutionary conserved

    than sequence





Workflow in homology modeling



Marti-Renom, M. A. and Stuart, A. C. and Fiser, A. and Sanchez, R. and Melo, F. and Sali, A. (2000) Comparative protein structure modeling of genes and genomes., Annu Rev Biophys Biomol Struct, 29, 291-325

Template recognition

- This initial step is the most crucial step in Homology modeling:
  - a wrong structure as model, will result in a very unlikely or even completely useless prediction
- can be „trivial“ for well-characterized protein families
- can be „impossible“ for completely new/unknown folds

Twilight zone of protein sequence alignments

HSSP curve:

- Originally defined by:

  Sander, C. & Schneider, R. (1991).

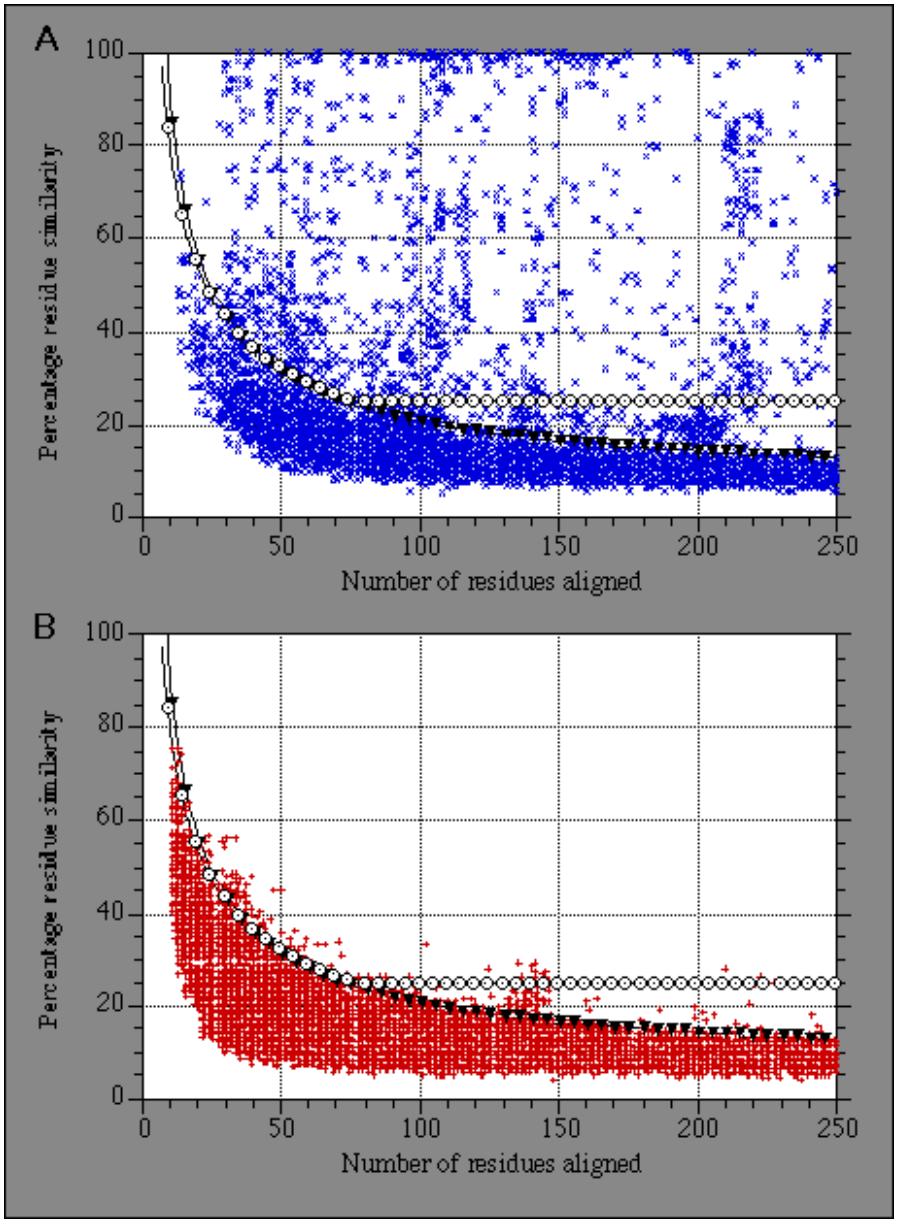
  Database of homology-derived structures and

  the structural meaning of sequence alignment.

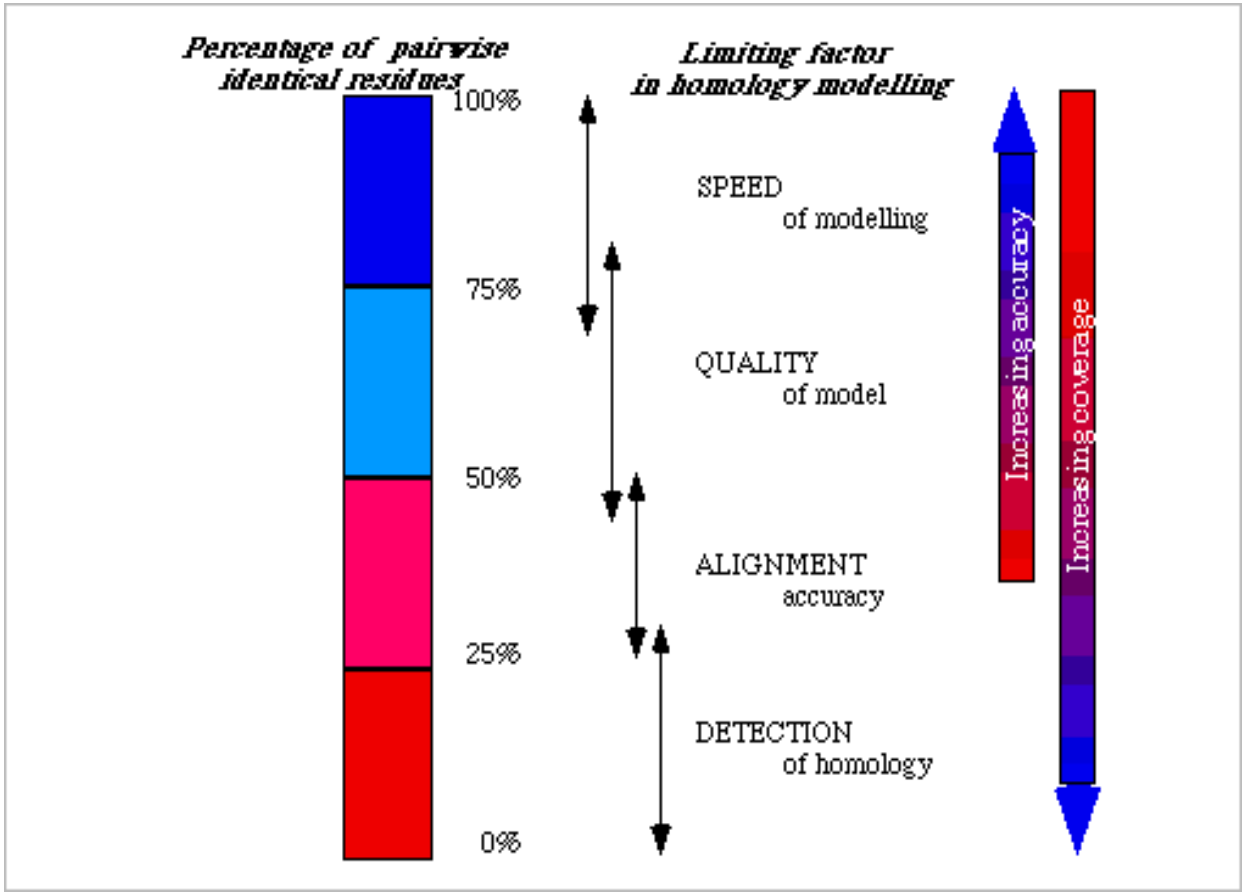
  Proteins, 9, 56-68.

source:

http://www.embl.de/~rost/Papers/1999_twilight/paper.html



Sequence identity and limitations



source: http://www.embl.de/~rost/Papers/Dfig/Cabios97/fig2.gif

Template recognition - HHpred

- 7 steps:

1. MSA is built with sequences found by PSI-BLAST
2. Profile-HMM is calculated from MSA and searched against a profile-HMM DB derived from PDB (Proteins structures are splitted in domains)
3. Results are re-ranked, using a neural network (NN)

   The NN is used to predict the TM score (measure to judge structural model), given the alignments.
4. Stets of MSA with successively lower sequence diversities are constructed.

   This is used to determine the colsest relative to the target sequence.
5. Generate target-template alignments and again rank them with a NN (TM score prediction).
6. Choose template
7. Run Modeller

Template Recognition – SWISS-MODEL

- Uses SWISS-MODEL template library, derived from PDB
  - Proteins are split in chains, which are then grouped
- 3 ways for template recognition:
  1. BLAST search
     - → If close homologs are available
  2. PSI-BLAST search
     - → for more remote sequences
  3. HMM-search
     - → for more remote sequences
     - → Non-redundant sets in the library are used to calculate HMMs

SWISS-MODEL

- Different extent of user interaction:

1. Automated mode

   if target-template similarity is high (only require AA sequence)

2. Alignment mode

   Submission of MSA including target protein and at least one protein with known PDB structure

3. Project mode

   User has full control over modeling parameters.

   - Manipulation of alignments
   - Selection and superposition template structure, etc.

Template recognition - iTASSER

- Uses MUSTER (= THREADING method) to find template
- Additional information from structure used:
  1. sequence profiles
  2. secondary structures
  3. structure fragment profiles
  4. solvent accessibility
  5. dihedral torsion angles
  6. hydrophobic scoring matrix

➔ Structural profile of template is generated, which the target is aligned to

- For „hard“ target, additional programs are used

  (LOMETS = FUGUE+HHSEARCH+PROSPECT+PPA+SP3)

Summary – Template recognition

- Most criticial step
- Templates are not necessarily complete Proteins, but domains and chains
- Either fully automated or user aided
  - ➔ User can manipulate alignments and select final target
- HHpred, SWISS-MODEL use BLAST and HMM profile searches
- iTASSER uses THREADING method (3D-1D alignment)

Backbone generation

- Follows template selection
- The residues of the target sequence are assigned to coordinates of aligned residues from the template structure
- This structural model is then optimized
- 2 possible gap situations in alignment:
  - gap in target sequence results in unassigned 3D coordinates in template structure
  - gap in template sequence results in uassigned target residues (i.e. They have no 3D coordinates assigned)
  - Both situations need to be accounted during model generation

sources

- Rost, B. and O'Donoghue, S. (1997) Sisyphus and prediction of protein structure. Comput Appl Biosci, 13(4), 345-356
- Rost, B. (1999) Twilight zone of protein sequence alignments. Protein Eng, 12(2), 85-94
- Hildebrand, A. and Remmert, M. and Biegert, A. and Soding, J. (2009) Fast and accurate automatic structure prediction with Hhpred. Proteins, 77, 128-132
- Arnold, K. and Bordoli, L. and Kopp, J. and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics, 22(2), 195-201
- Zhang, Y. (2009) I-TASSER: fully automated protein structure prediction in CASP8. Proteins, 77, 100-113
- Marti-Renom, M. A. and Stuart, A. C. and Fiser, A. and Sanchez, R. and Melo, F. and Sali, A. (2000) Comparative protein structure modeling of genes and genomes., Annu Rev Biophys Biomol Struct, 29, 291-325