Presenting: 'An evaluation of automated homology modelling methods at low target-template sequence similarity' by Dalton and Jackson

Vishvas Vasuki

October 5, 2007

Steps in homology modelling

- Template selection
- Alignment
- Model construction

Template selection

- Blast E-Values: .0001 E-Value means that the structures are pretty close.
- FASTA sequence identities
- For multiple templates: Ensure structural similarity among templates. Use 'Z-score'.

Alignment: Critical

- **Use sequence information only:** BLAST, FASTA and PSI-BLAST.
- **Use structure information too**: Use threading (3D-Coffee), build and align locale profiles (Staccato) or penalize indels in secondary structure locales (SAlign).

Model construction: Conserved region

- **Rigid-body assembly (Nest, Builder, Swiss-Model):** Copy structure of conserved core from the template. [2]
- **Segment-Matching (SegMod/ENCAD):** Assemble short (hexapeptide?) segments to fit atom positions inferred from the template.
- **Satisfaction of spatial restraints (Modeller):** Extract restraints on structure from template. Add additional restraints. (Possibly NMR data too) Minimize violations. Then, use energy functions. [2] This beats the rest by a small margin.

Model construction: Loop Modelling

- **Ab initio** (Modeller and Nest): Don't understand. But, they beat the rest by a small margin.
- **Database** (Builder and SegMod/ENCAD): Fragments from DB, energy minimization
- Database if Ab-initio fails (Swiss-Model)

Model construction: Sidechains

- Put it close to template sidechain. Use torsion angle libraries. Then use Energy minimization.

Tests and Results

- Test 3 new sequence-structure alignment programs: **3D-Coffee**, Staccato and SAlign.
- Find out if using multiple templates makes things better (and not worse).
- Test 5 homology modelling programs and their respective loop building methods: Builder, **Nest, Modeller**, SegMod/ENCAD and Swiss-Model.

Test data

- 'The SABmark database provided 123 targets with at least five templates from the same SCOP family and sequence identities 50%.' [1]
- Loops range in length from 5 to 16 residues. (These include 2 rei on either side of the unmatched region.)

Lessons

- 'There are two main areas of difficulty in homology modelling that are particularly important when sequence identity between target and template falls below 50%: sequence alignment and loop building.' [1]
- Use structure information in sequence alignment.
- Sequence alignment quality decreases with decreasing sequence identity.
- Overall, sequence alignment seems more critical than choice of modeller. (We can change this.)
- Ab initio loop modellers seem to work **slightly** better. Loop accuracy decreases with increasing length.
- Be wary about using multiple, rather than single, templates.


James A. Dalton and Richard M. Jackson.
An evaluation of automated homology modelling methods at low target template sequence similarity.
Bioinformatics, 23(15):1901–1908, August 2007.


Michael Y. Galperin and Eugene V. Koonin.
Frontiers in Computational Genomics.
Horizon Scientific Press, 2003.