Contact-assisted prediction using Rosetta:

Ts/Tx category

Baker Group

Sergey Ovchinnikov, Hahnbeom Park, David Kim, Yuan Liu, and David Baker
Mainly done by a stateless one

Sergey Ovchinnikov
Rosetta & experimental data
Overview of the method (both Ts,Tx)

CASP10 pipeline

Clusters from alignments and server models

Ab initio models

Clusters from nonlocal fragments

Interpretation of given experimental data

External Model diversification

Iterative Rosetta Hybridize

Final model
Ts: simulated NMR contacts
Starting models: $\beta$-proteins

Rosetta FastRelax with non-ambiguous contacts as strong restraints

Non-ambiguous backbone contacts  All contacts
Starting models: $\alpha$-proteins

Rosetta $ab\ initio$ with non-ambiguous contacts as sigmoidal restraints

Non-ambiguous backbone contacts

All contacts
Iterative hybridization (also used in Tx)

- Structural hybridization of multiple structures
- Fragment insertion at random position

GDTTS=48

Restraint ambiguity adaptation

Non-ambiguous
Max ambiguity=5
All
3. Model diversification during iteration

**Aim:** To jump out from local minima (due to the ambiguous information)

**Methods:** components used in the Refinement category

“wiggling”

“loop modeling”
Ts results

- **all-βs** were more tractable
- Not very good at **multi-domain** targets

Comparison to the best among others

- **all-α**
- **α/β & α+β**
- **all-β**

Ts810: interaction w/un-modeled domain
Ts804: multi-dom, trimer
Ts794: multi-dom
Ts826: multi-dom, TM
What went right

- Rosetta Ab initio + Strategy on handling ambiguity helped

Ts777
GDT-TS: 67.0 (Z=3.5)
Best among others: 44.8

Ts827
GDT-TS: 57.2 (Z=2.8)
Best among others: 48.3
What went wrong

- Insufficient sampling on domain orientation
- Failure on deconvoluting intra- and inter-domain information

Ts794

GDT-TS: 48.6 (Z=0.9)
Best among others: 70.5
Tx: cross-linkage contacts
Tx results

- Too difficult
- Targets stood out had other reasons

- provided contacts
  (residue pairs that any atom pairs < 25 Å)

- residue pairs in the native structure
  that are apart < 25 Å
Impressive results?

Tx812: starting model was already good...

GDT-TS: 47.4 (Z=2.6)
T0808 BAKER: 44.2 – Co-evolution
Best among others: 40.2
Impressive results?

Tx808

contactScore (contS): 58.1
Best among others : 48.0

Native

D1 – from Homology model

Submission

Contact comparison
Summary

• **Ts category:**
  Combined with Rosetta, simulated NMR contact information was enough to bring the modeling procedure to the level of **homology modeling** (\(GDT\text{-}TS > 60\) for 79% of targets, similar to the results with Rosetta and sparse NMR data)

• **Tx category:**
  Crosslink contact (Tx) information is too sparse and too low resolution (25Å) to sufficiently guide Rosetta modeling

• **Future directions:**
  multi-domain & larger proteins
Acknowledgements

Institute of Protein Design, Univ. of Washington
Sergey Ovchinnikov
David Kim
Yuan Liu
Binchen Mao
David Baker

CASP participants, assessors, and organizers