MELD: an integrative structural biology tool

Laufer group

Alberto Perez

CASP XIII  perez@chem.ufl.edu  NMR-guided
Data rich: the data guide towards a narrow structural ensemble

Available conformational space

Restraints dominate energy surface

Very narrow ensemble

Agrees with data

simulation time

temperature
Different datasets might lead to different biologically relevant conformations
Complement sparse, ambiguous and/or noisy data with physics
Many possible restraints and only a few are correct
With sparsity and ambiguity in the data many distinct conformational envelopes are visited.
Each NMR peak has several interpretations — one of them explains the peak (ambiguity)

But, there could be false positive peaks (noise). And, some peaks are missing.
MELD uses a Bayesian inference approach to incorporate data into simulations

\[
p(x|D) = \frac{p(D|x)p(x)}{p(D)} \sim p(D|x)p(x)
\]

\[
p(x) \sim \exp[-\beta E_{\text{force}}(x)]
\]

\[
p(D_i|x) \sim \exp[-\beta E_{\text{restraint}}(x)]
\]

We use Hamiltonian Replica Exchange to enhance sampling

- High Temperature / Weak Restraints
- Low Temperature / Strong Restraints
Restraints sculpt the folding landscape creating faster folding funnels

MELD

Restraints sculpt the folding landscape creating faster folding funnels

\[ \frac{P_1}{P_2} = \frac{P_{1,\text{data}}}{P_{2,\text{data}}} \]

\[ Z \geq Z_{\text{data}} \]

\[ P_{1,\text{data}} \geq P_1 \quad P_{2,\text{data}} \geq P_2 \]

CASP-NMR protocol

- Talos data enforced at 60%. (Flat bottom harmonic restraints)
- Use psipred sse prediction and enforce at 60%.
- Use co-evolutionary data (gremlin) + metagenomic data. Enforce at 70%.
- NOESY peaks:
  - Removed any peak that could be explained by an $\text{abs}(\text{Res}_i - \text{Res}_j) < 4$
  - Traced ambiguous hydrogens to heavy atoms
  - Enforced at 90%.
- N1008 real data: enforced at 60%.
- Starting Structures (15): Templates from Baker-RosettaServer, Quark, Zhang-Server
We used three metrics to select structures from REMD ensembles:

- Population / free energy
- Restraints satisfied
- Lowest restraint energy
NMR data improves our predictions
NMR data could improve the best models even more.
• 246 residues. Two domains, not closely interacting. Our models are too compact.
The Laufer team
• Residues 39 to 364 (all) 7.2 Å

• 68 - 350 —> 5.2 Å

• Much larger than we could attempt without data