MELD: an integrative structural biology tool

Laufer group

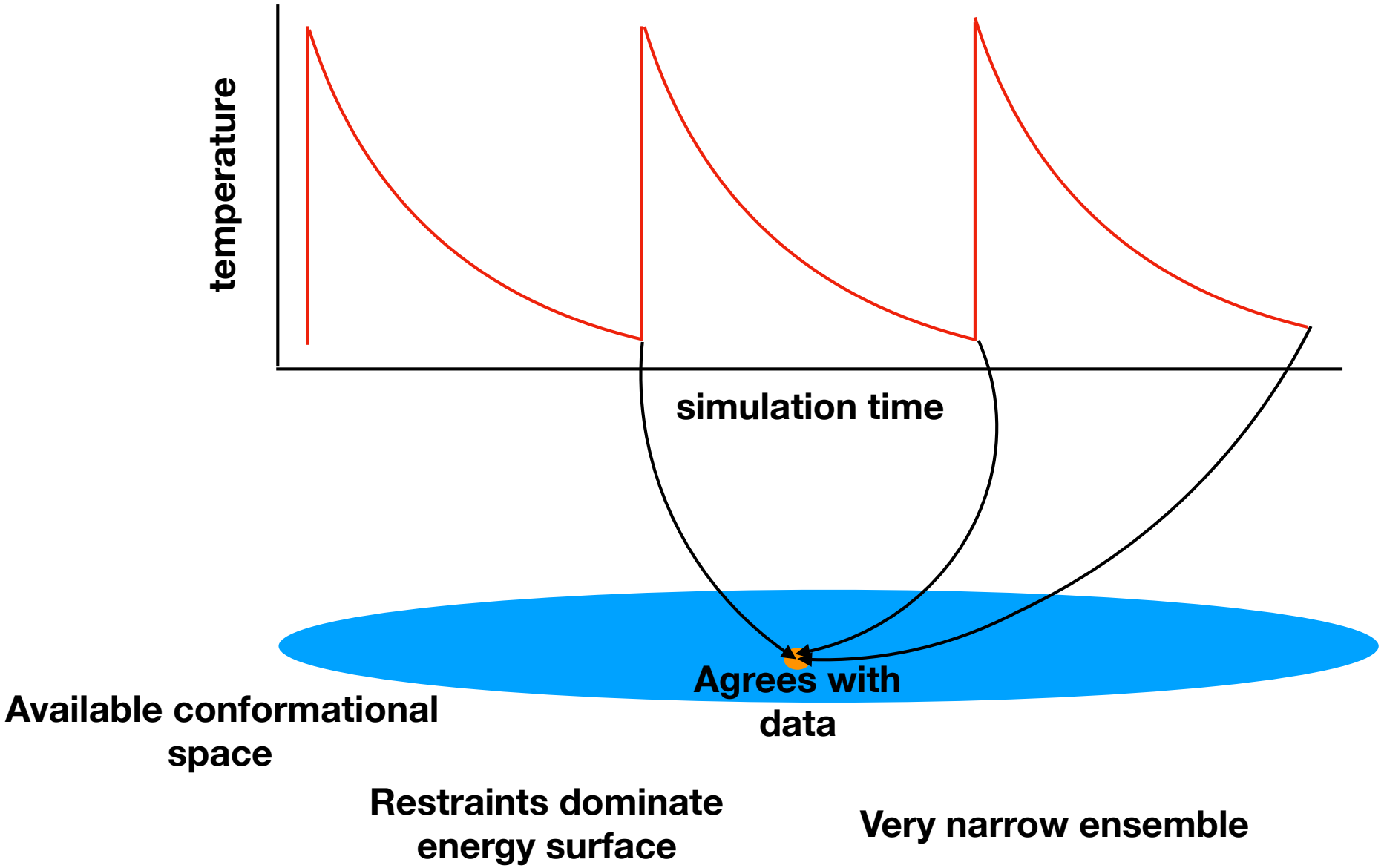
Alberto Perez

CASP XIII

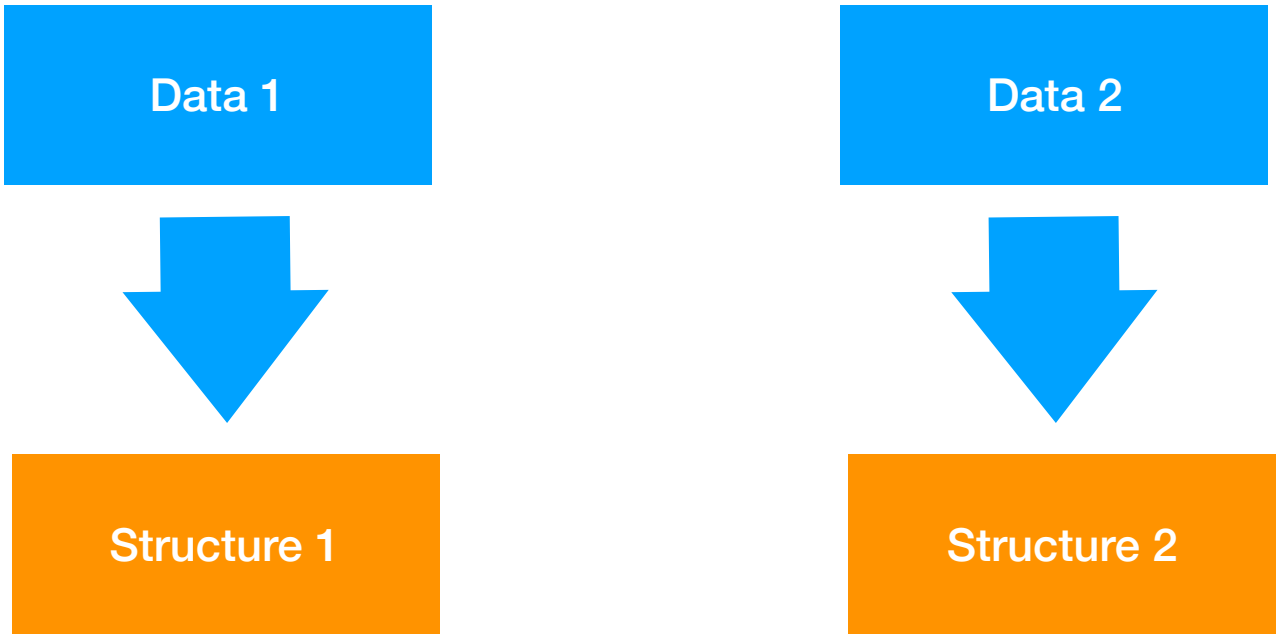
perez@chem.ufl.edu

NMR-guided

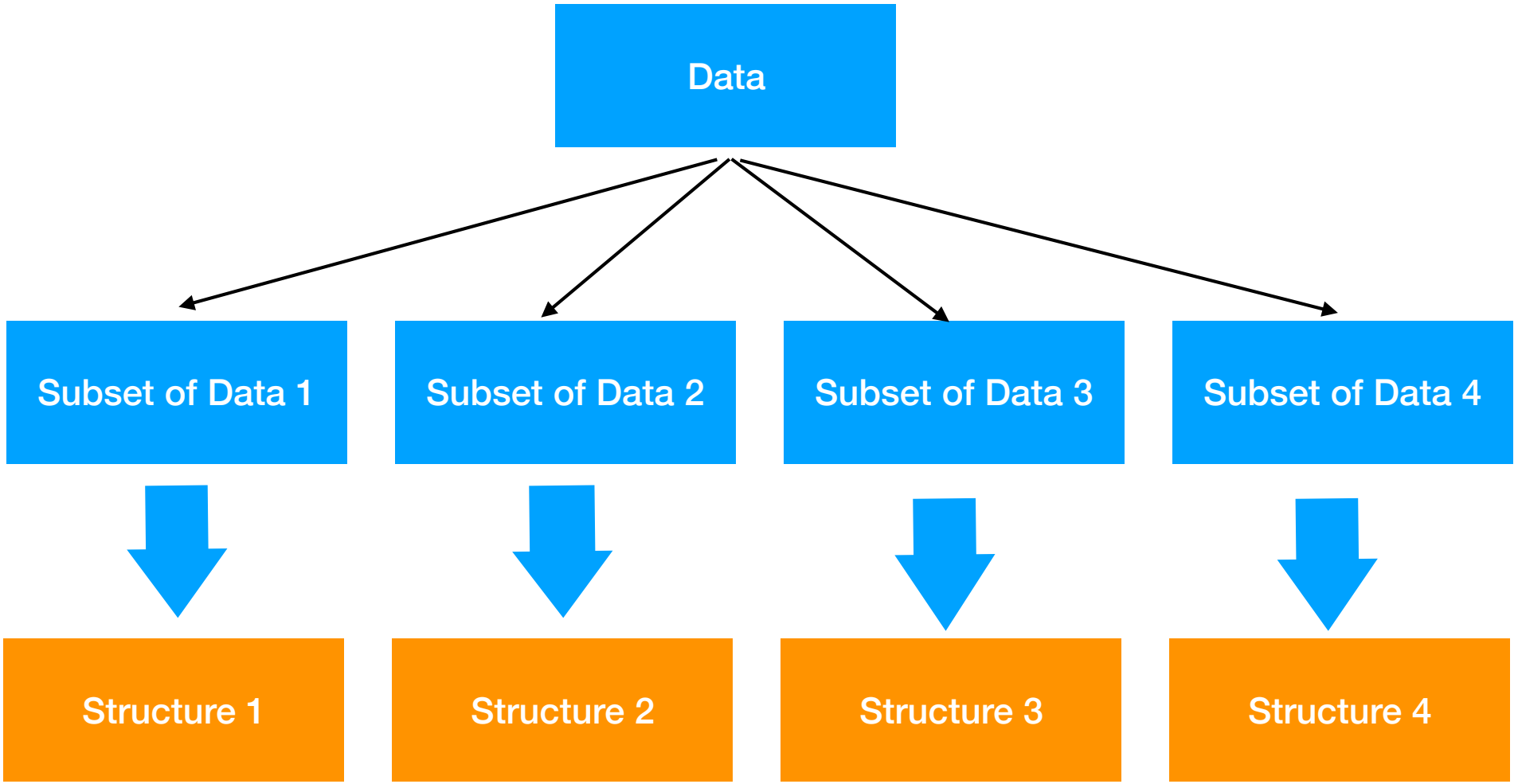
Data rich: the data guide towards a narrow structural ensemble



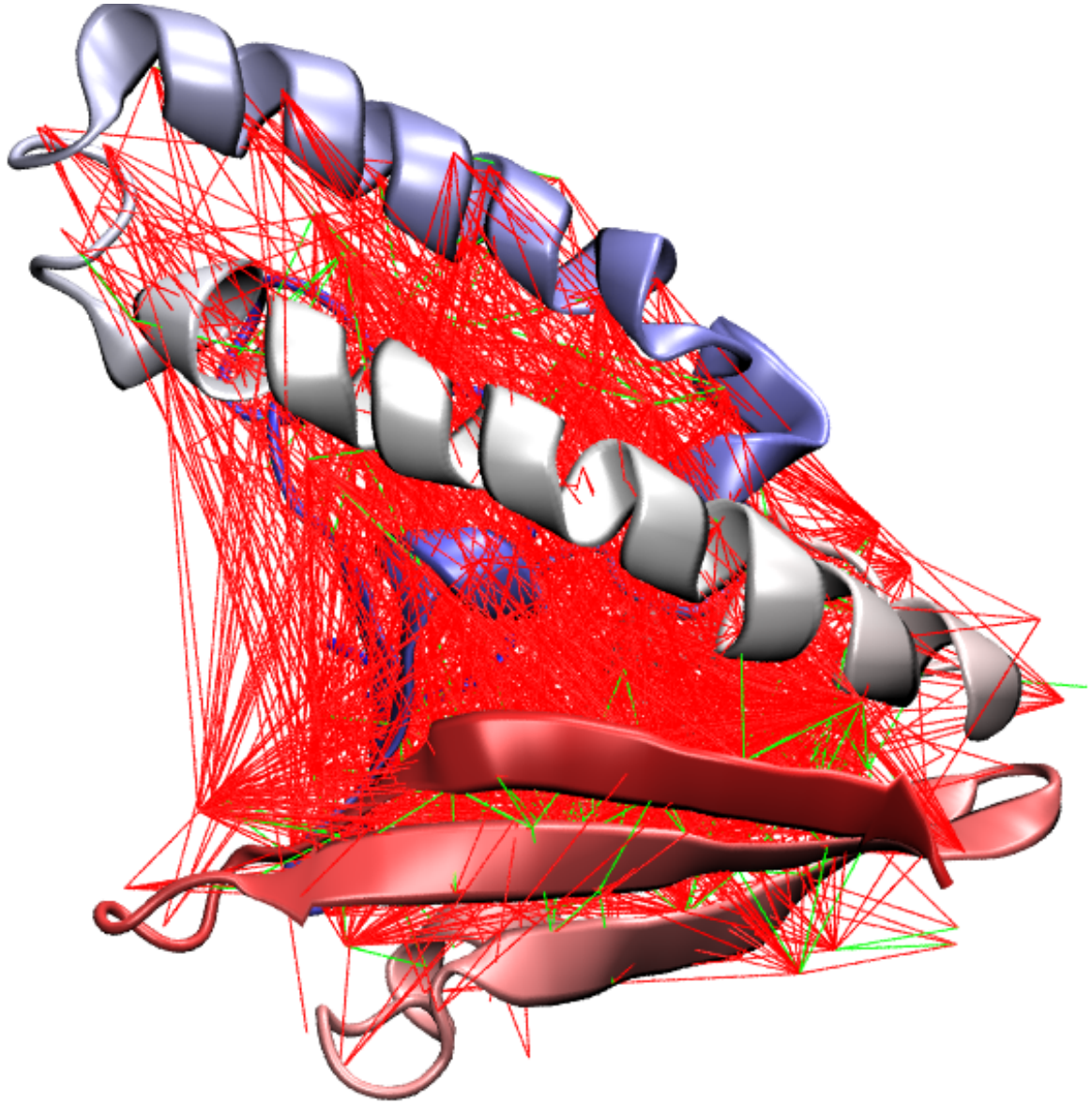
Different datasets might lead to different biologically relevant conformations



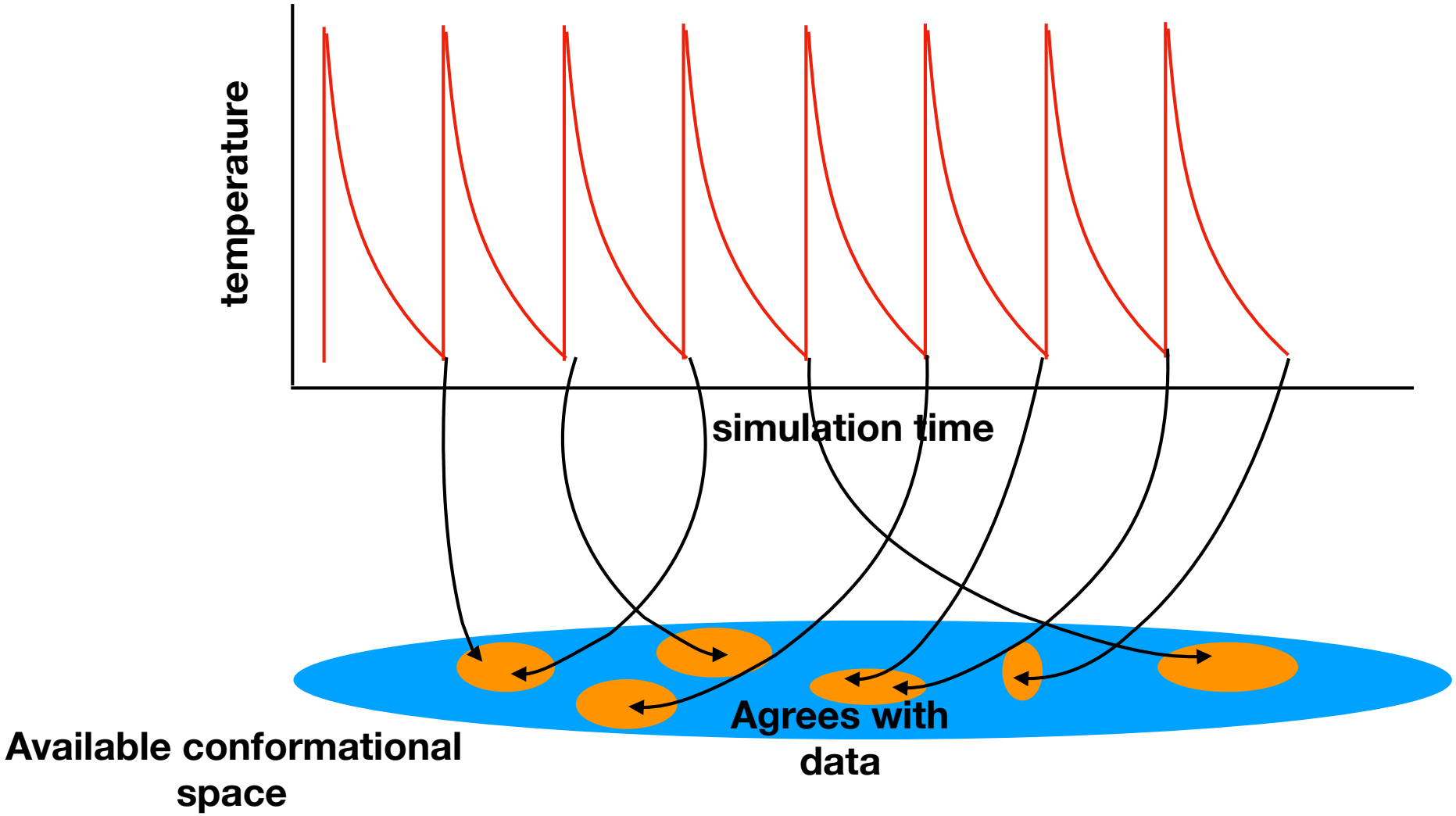
Complement sparse, ambiguous and/or noisy data with physics



Many possible restraints and only a few are correct

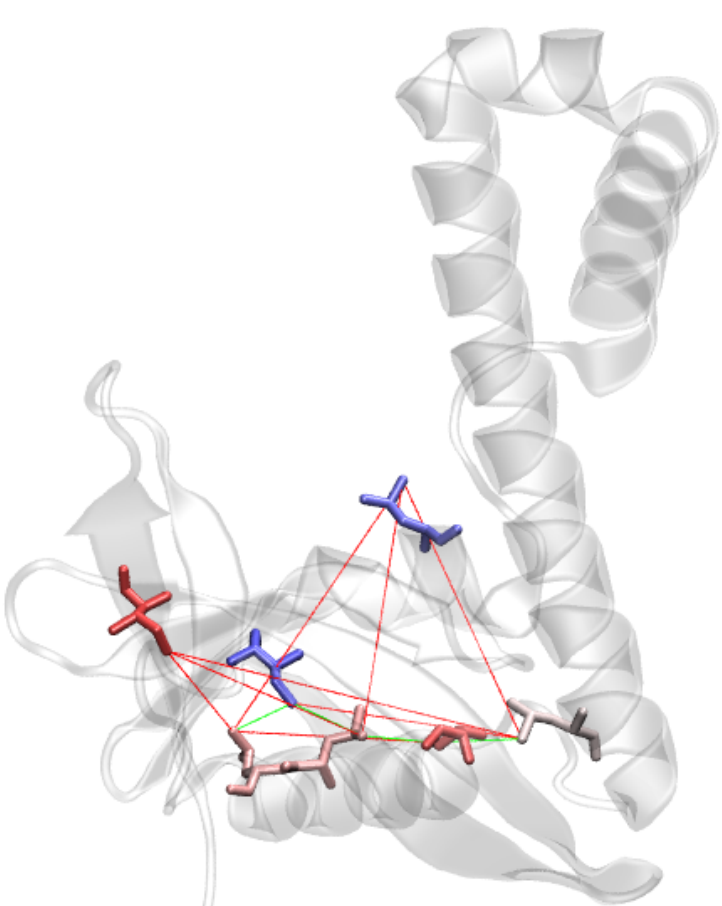


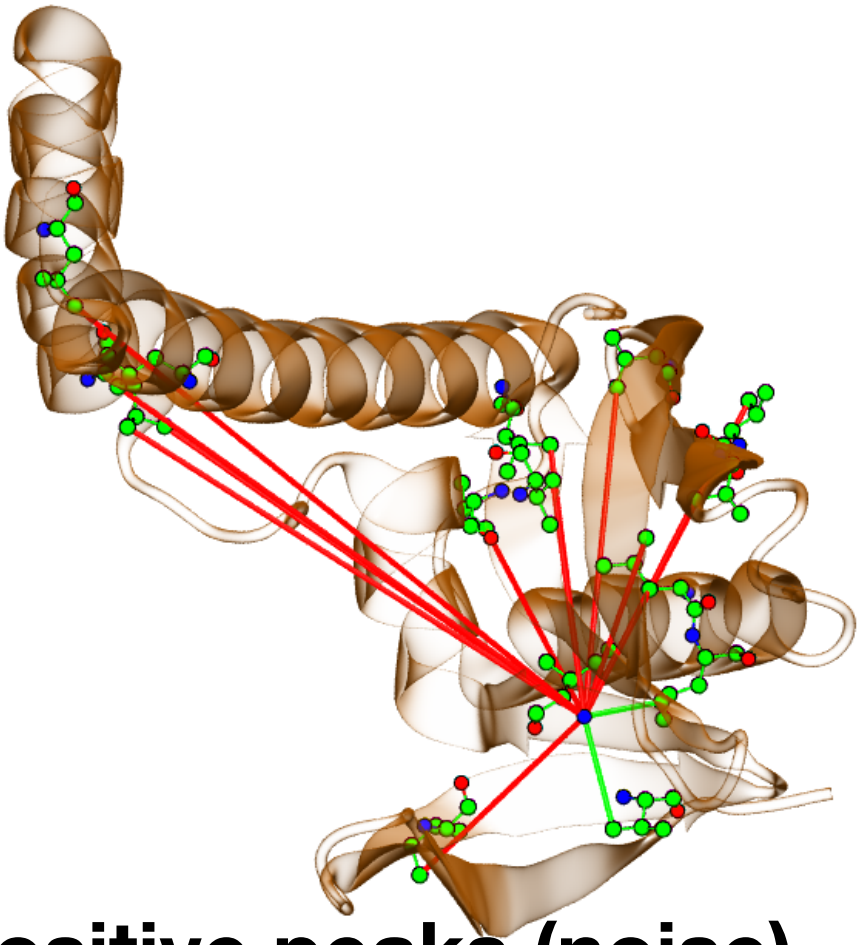
With sparsity and ambiguity in the data many distinct conformational envelopes are visited



Restraints guide to many different basins.

Each NMR peak has several interpretations — one of them explains the peak (ambiguity)



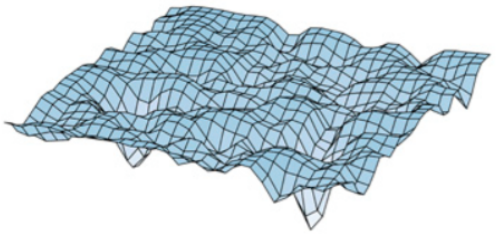


But, there could be false positive peaks (noise).
And, some peaks are missing.

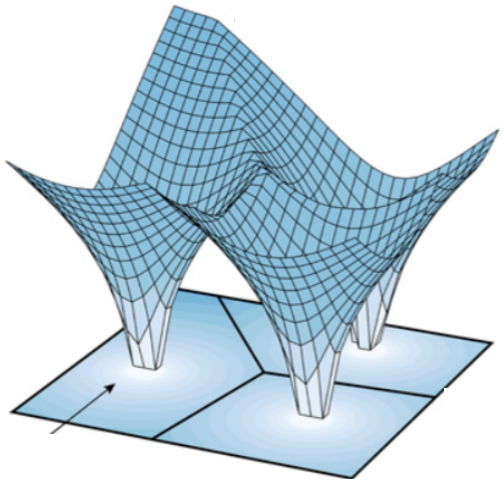
MELD uses a Bayesian inference approach to incorporate data into simulations

$$\overbrace{p(\mathbf{x}|\mathbf{D})}^{\text{posterior}} = \frac{p(\mathbf{D}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{D})} \sim \overbrace{p(\mathbf{D}|\mathbf{x})}^{\text{likelihood}}\overbrace{p(\mathbf{x})}^{\text{prior}}$$

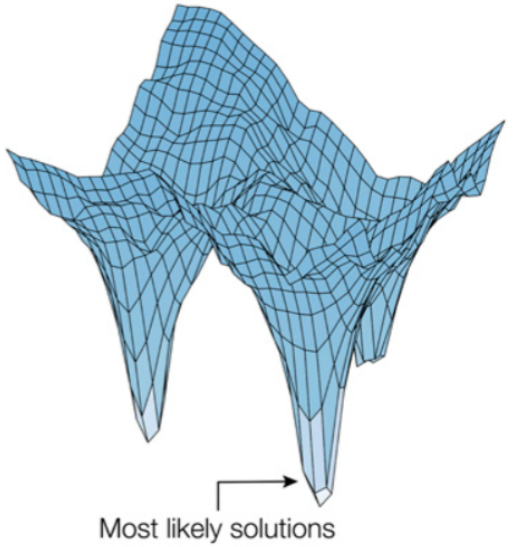
Force Field
(prior)



MELD Energy
(likelihood)



MELD Energy + Force Field
(posterior)

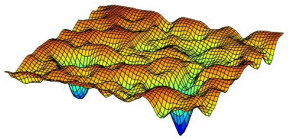


$$p(\mathbf{x}) \sim \exp[-\beta E_{\text{force field}}(\mathbf{x})]$$

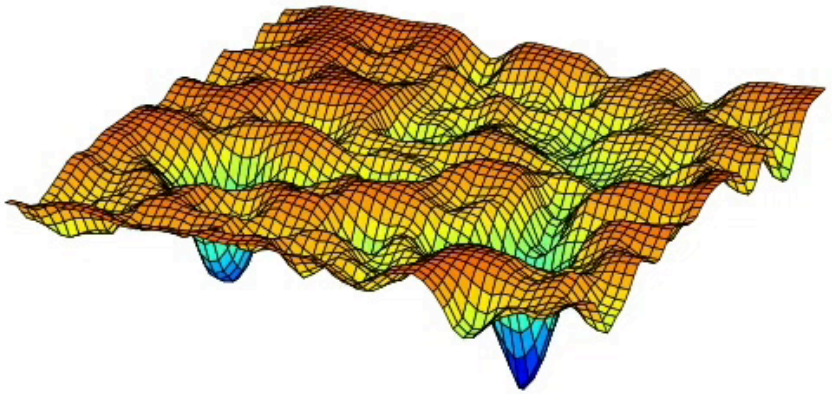
$$p(D_i|\mathbf{x}) \sim \exp\left[-\beta E_i^{\text{restraint}}(\mathbf{x})\right]$$

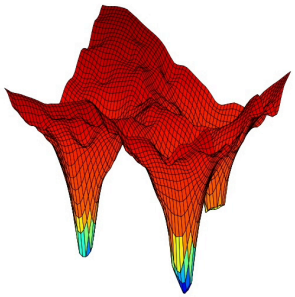
MacCallum*, Perez*, & Dill, Proc. Natl. Acad. Sci. U.S.A. 112, 6985–6990 (2015).

We use Hamiltonian Replica Exchange to enhance sampling



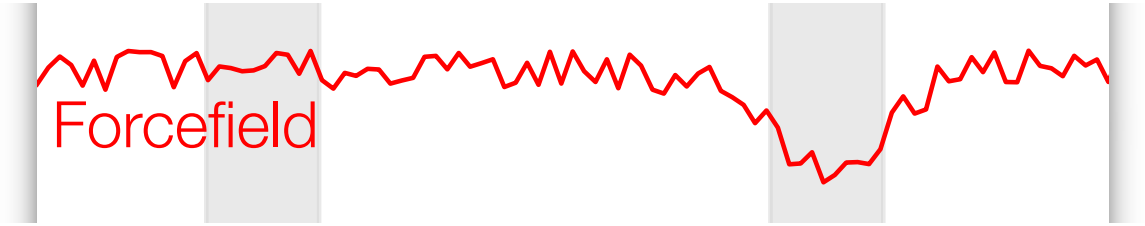
- High Temperature / Weak Restraints





- Low Temperature / **Strong Restraints**

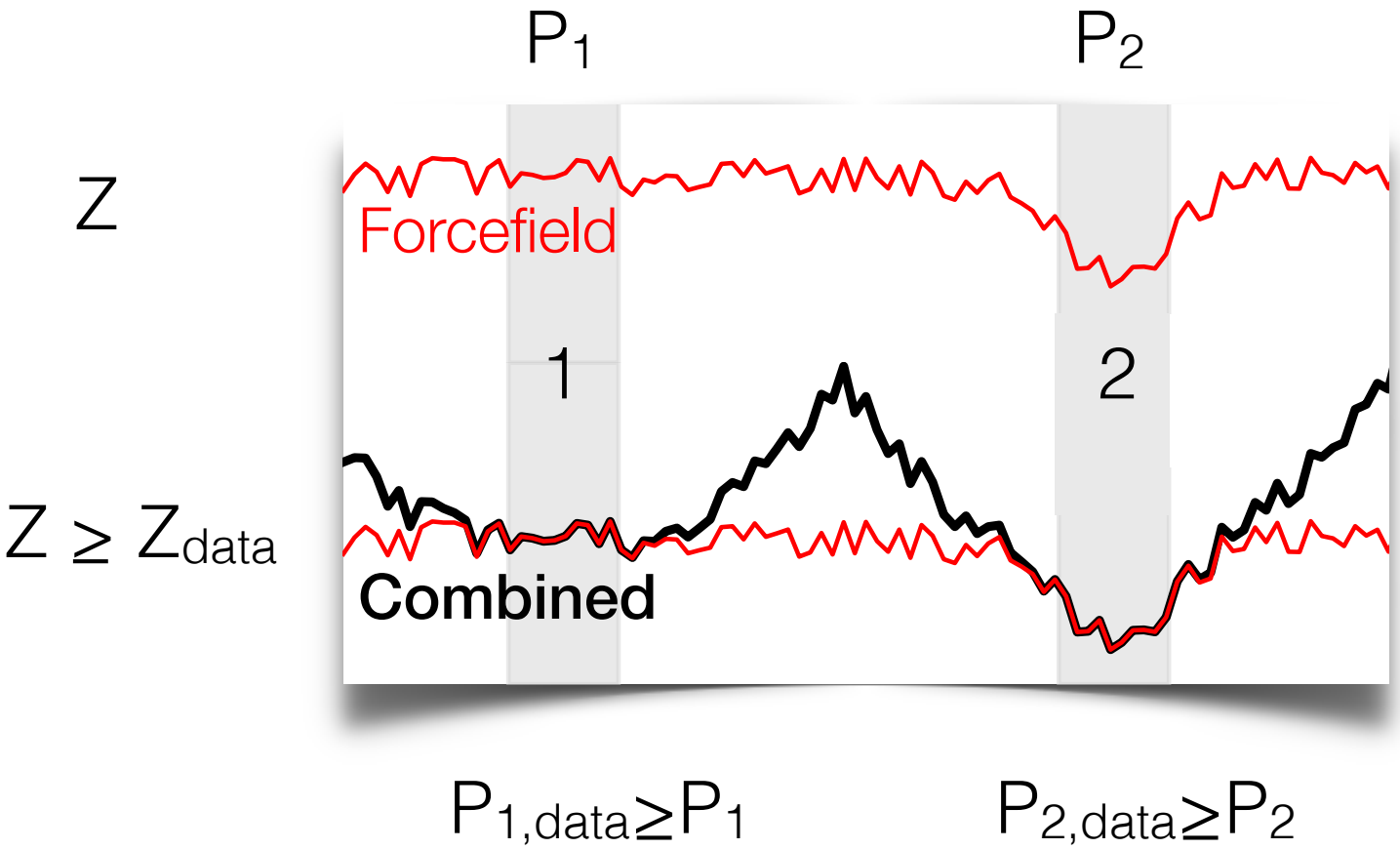
Restraints sculpt the folding landscape creating faster folding funnels



MELD

MacCallum*, Perez*, Dill. PNAS(2015)

Restraints sculpt the folding landscape creating faster folding funnels



$$\frac{P_1}{P_2} = \frac{P_{1,data}}{P_{2,data}}$$

MacCallum*, Perez*, & Dill, Proc. Natl. Acad. Sci. U.S.A. 112, 6985–6990 (2015).

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 abs(Res_i - 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

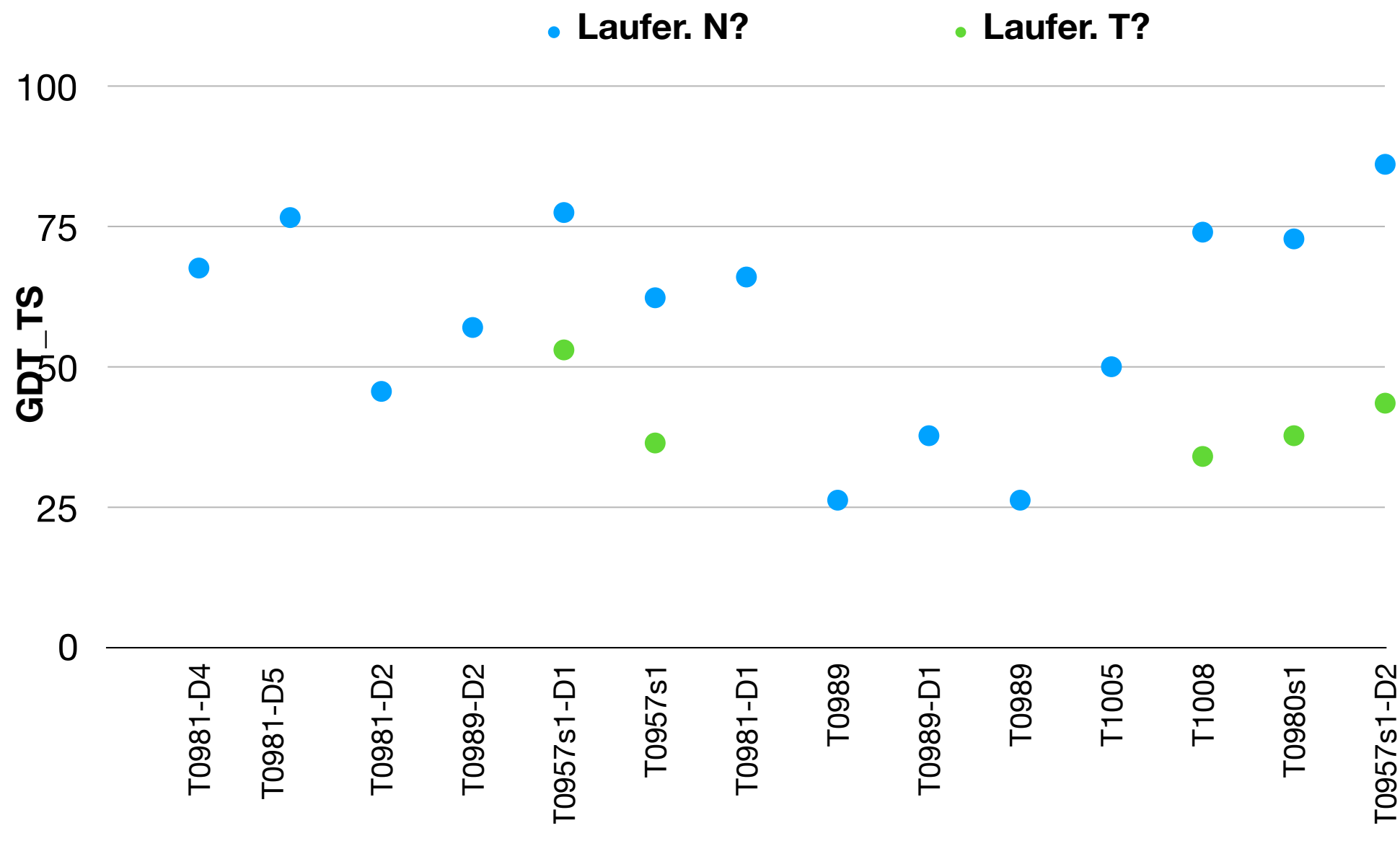
We used three metrics to select structures from REMD ensembles

Population / free energy

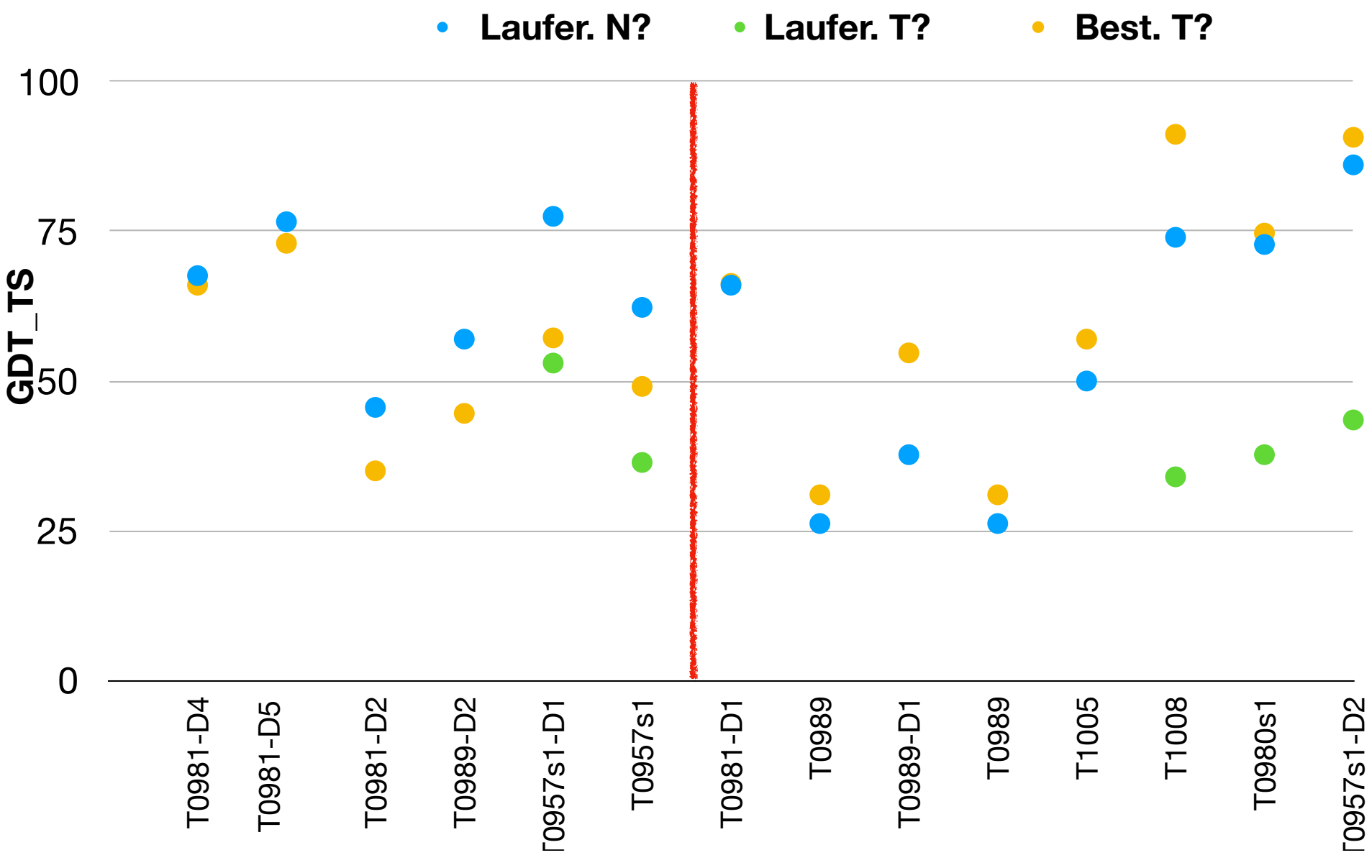
Restraints satisfied

Lowest restraint Energy

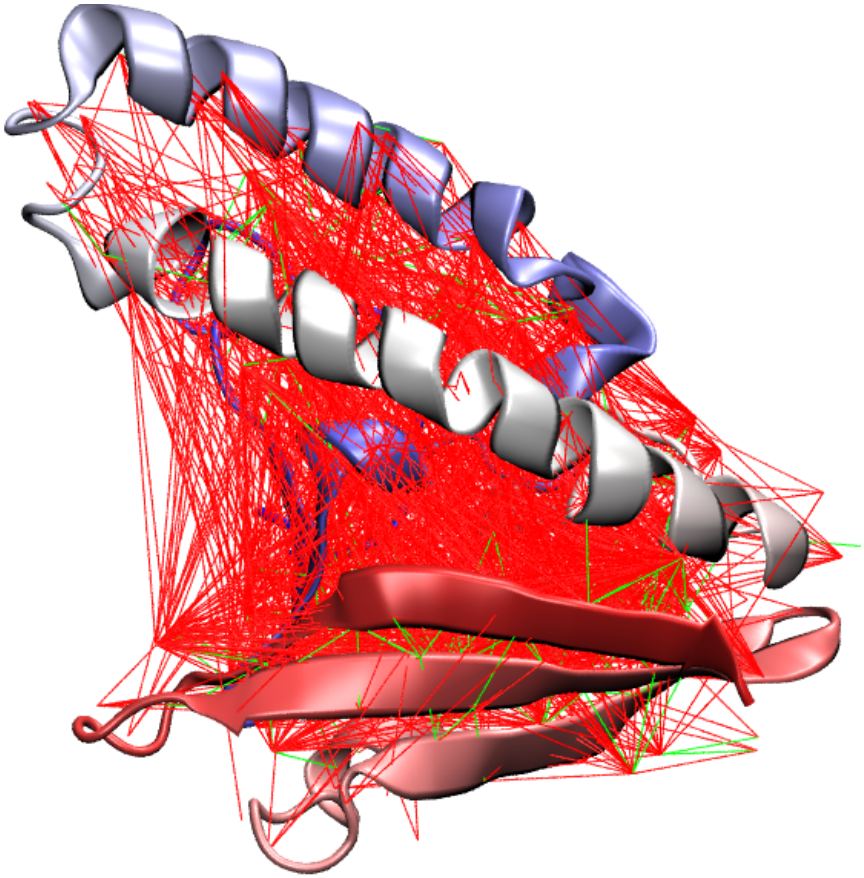
NMR data improves our predictions



NMR data could improve the best models even more



N0957s1



N0989

CASP:
Image redacted

- 246 residues. Two domains, not closely interacting. Our models are too compact

The Laufer team

EMBO

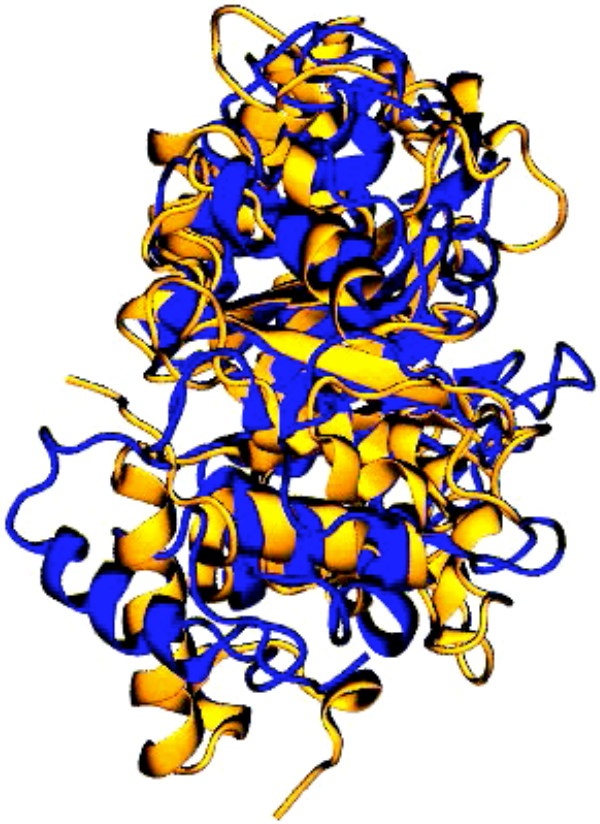
NATIONAL INSTITUTES OF HEALTH

NSF

BLUE WATERS
SUSTAINED PETASCALE COMPUTING

UF
UNIVERSITY of FLORIDA

N1005



- Residues 39 to 364 (all) 7.2 Å
- 68 - 350 —> 5.2Å
- Much larger than we could attempt without data