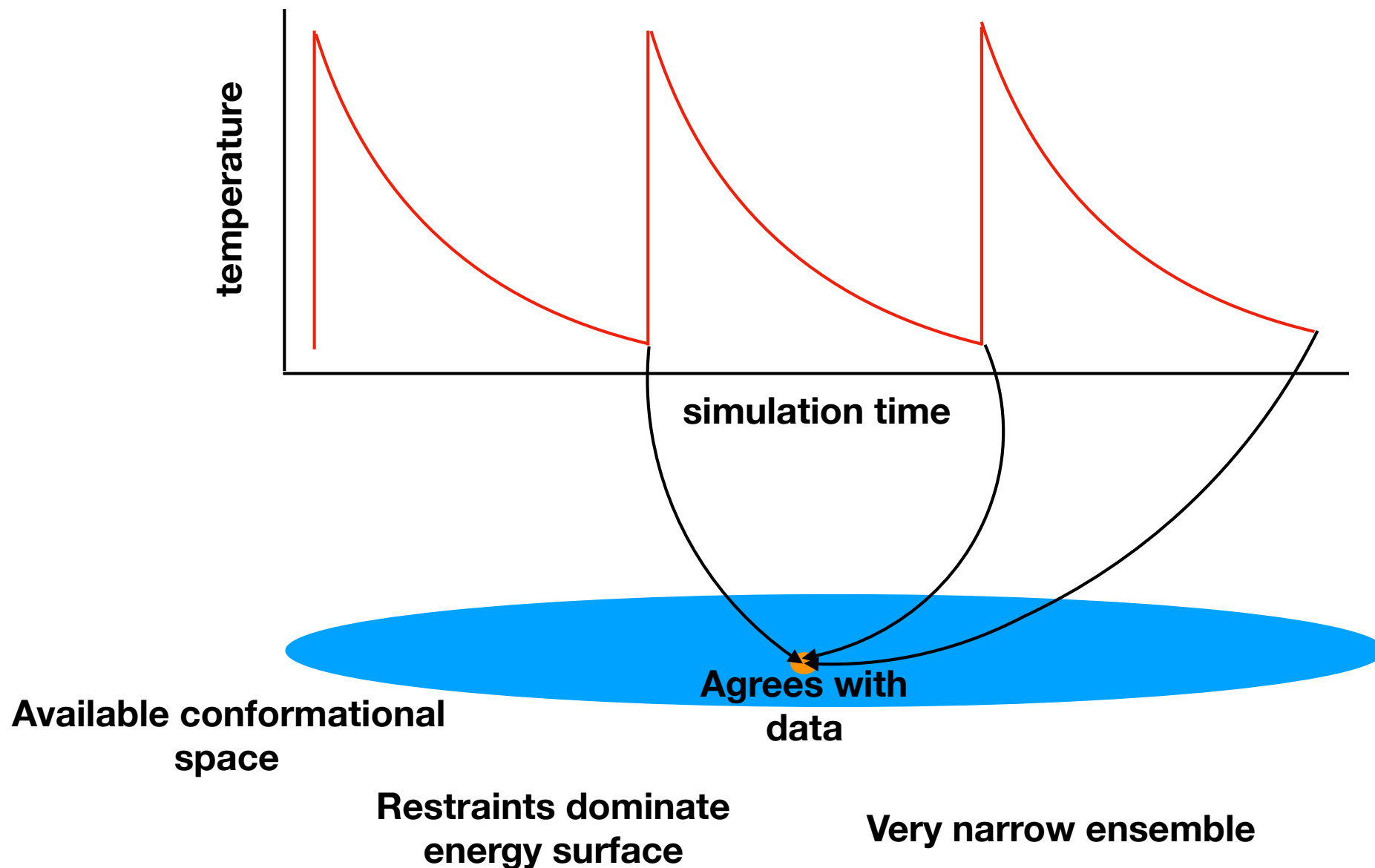


MELD: an integrative structural biology tool

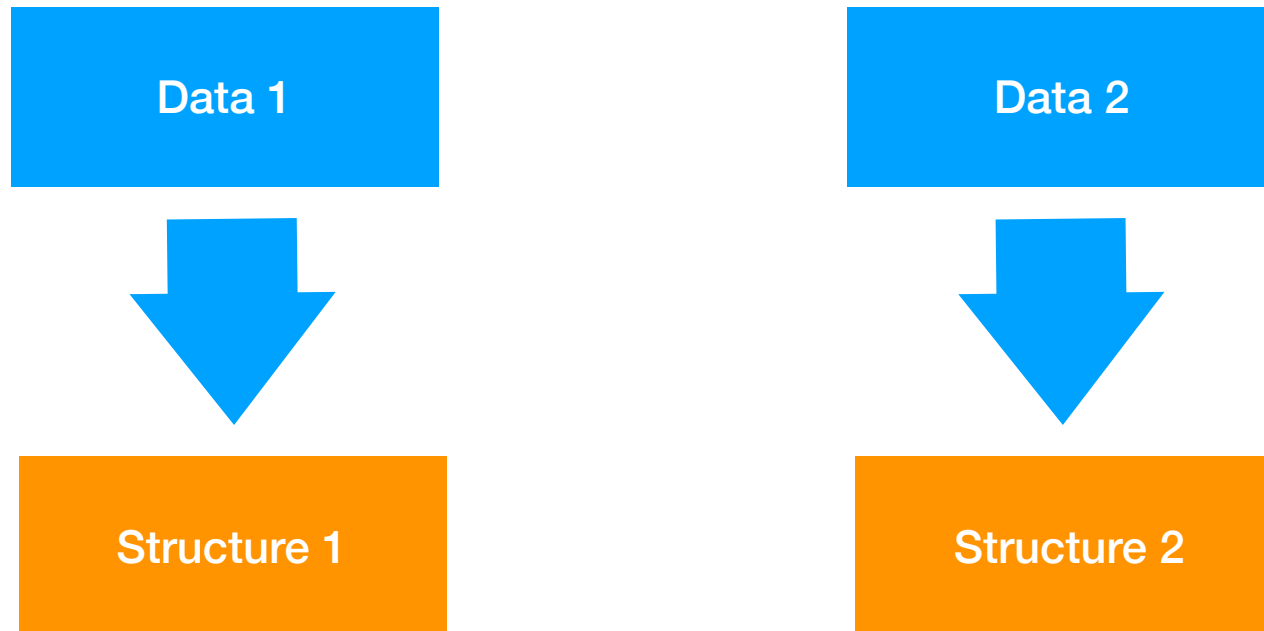
Laufer group

Alberto Perez

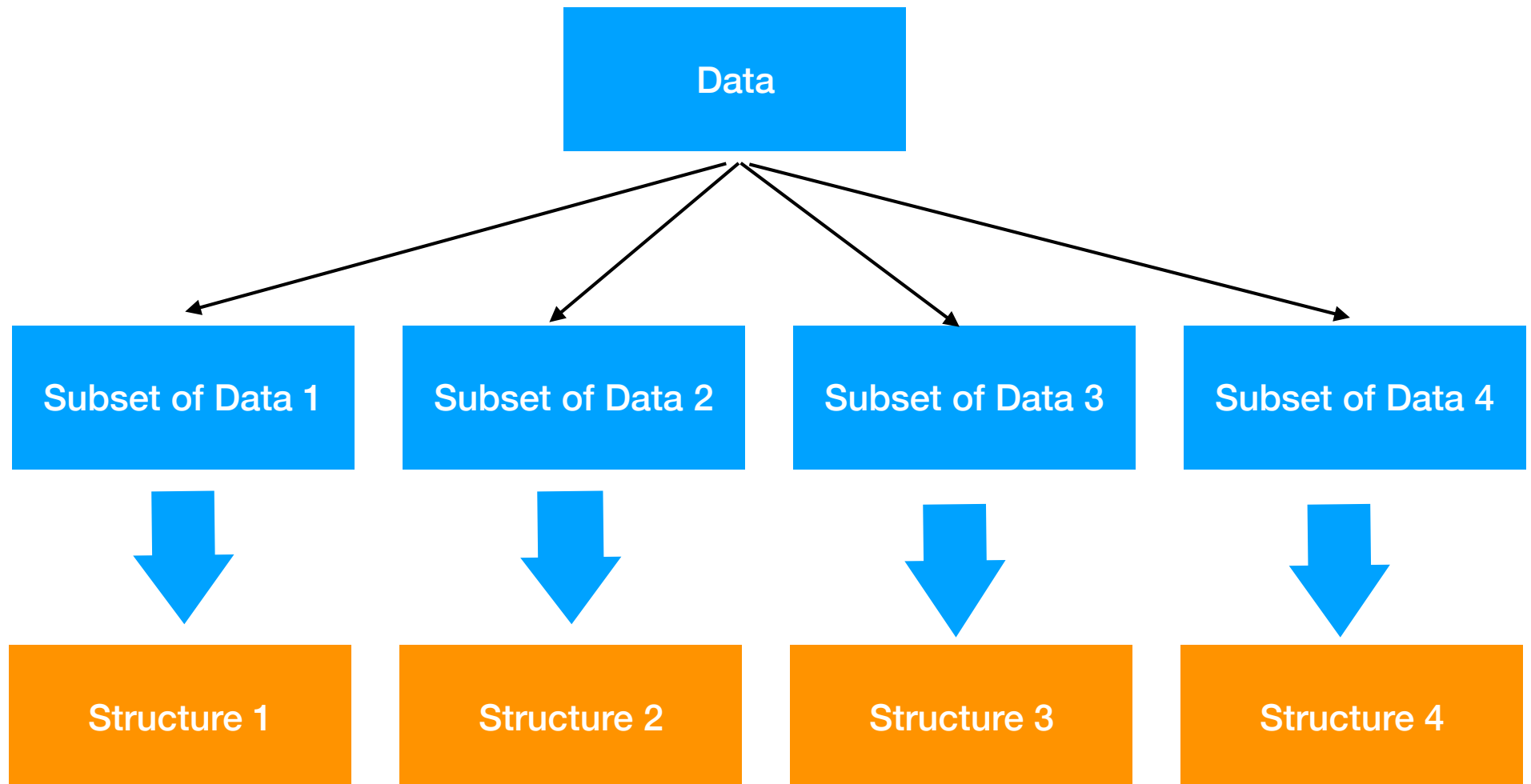
Data rich: the data guide towards a narrow structural ensemble



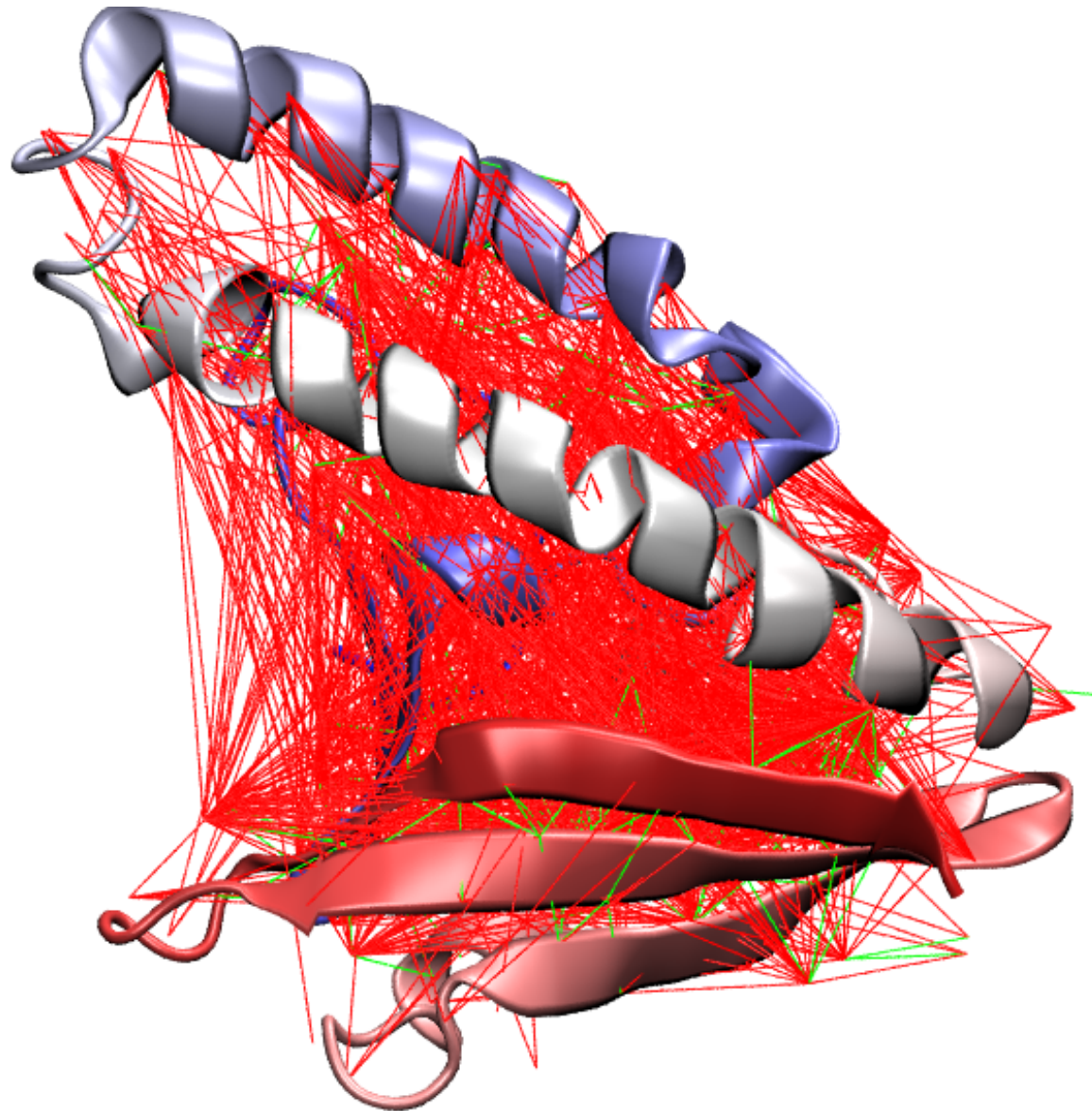
Different datasets might lead to different biologically relevant conformations



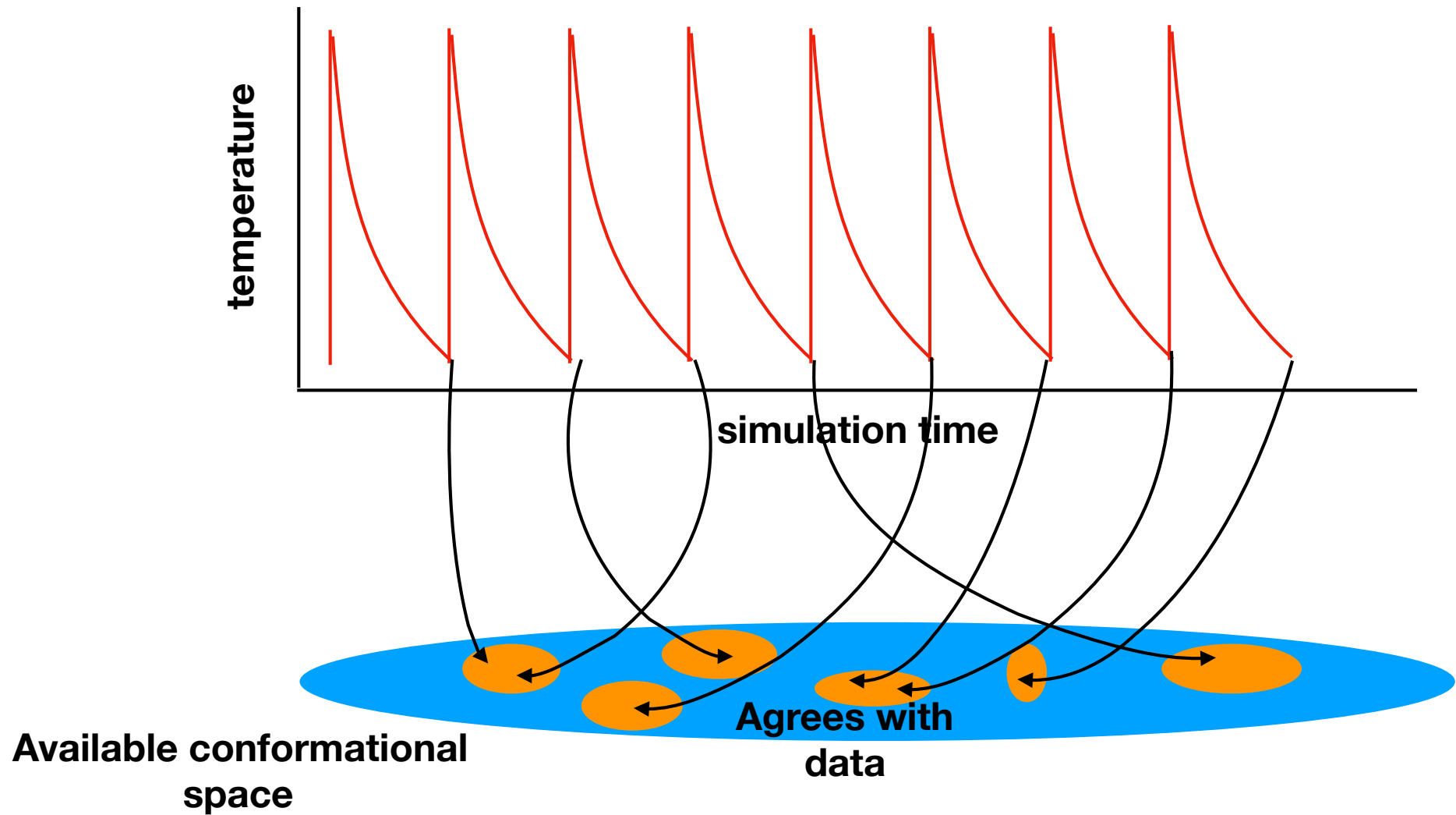
Complement sparse, ambiguous and/or noisy data with physics



**Many possible restraints
and only a few are correct**

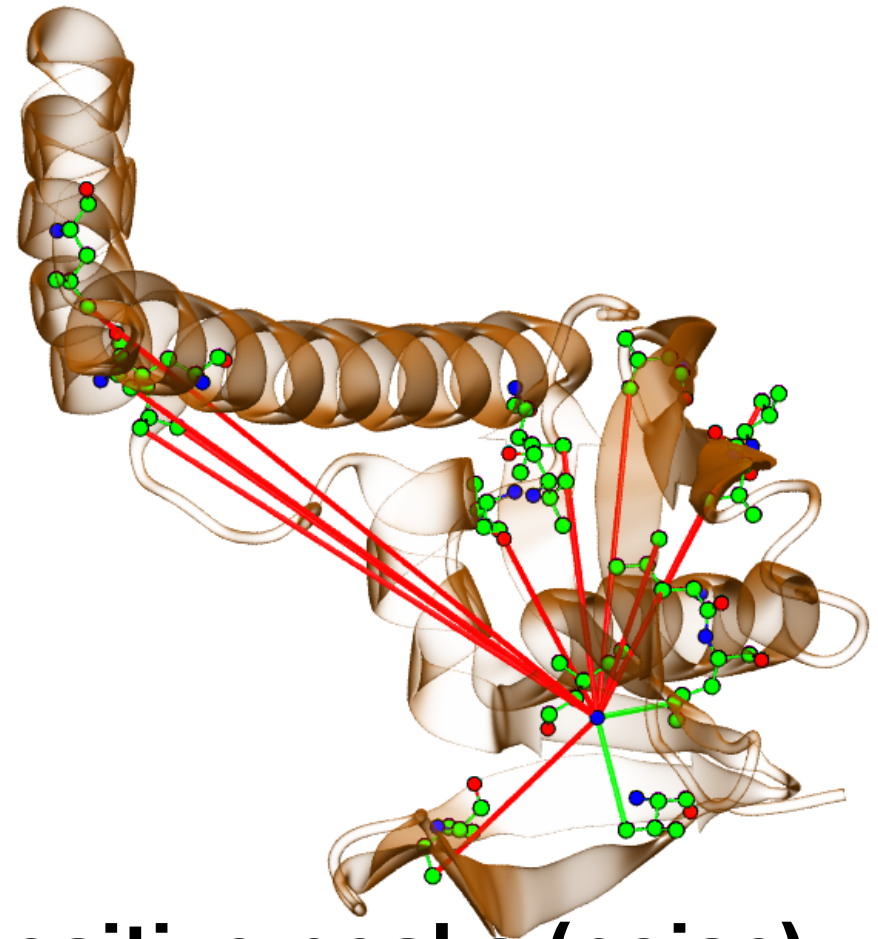
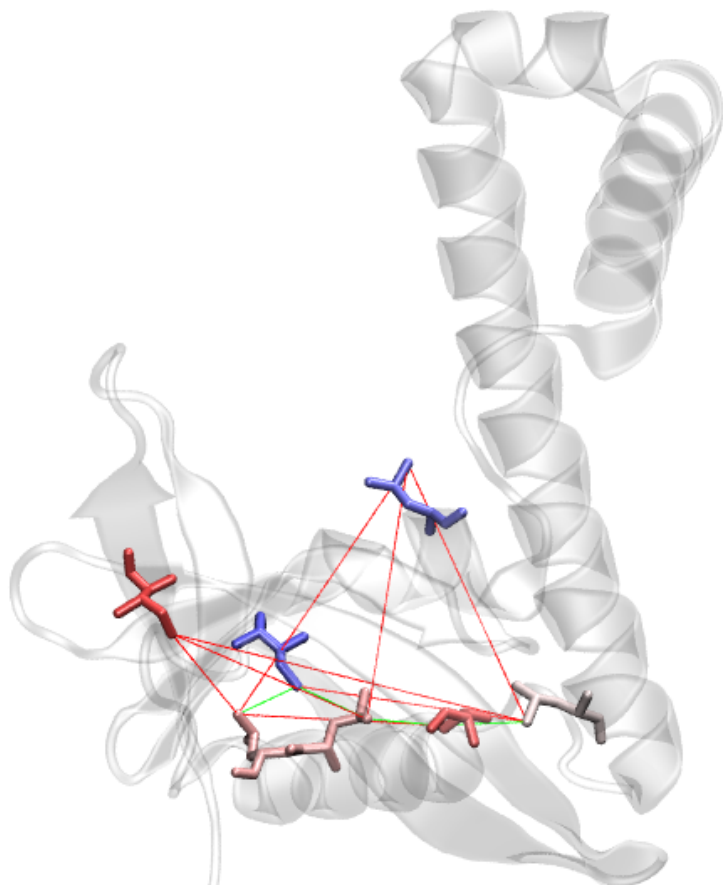


**With sparsity and ambiguity in the data
many distinct conformational envelopes
are visited**



Restrains guide to many different basins.

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

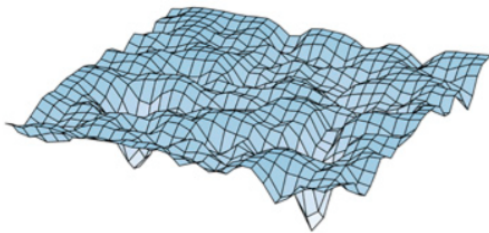


**But, there could be false positive peaks (noise).
And, some peaks are missing.**

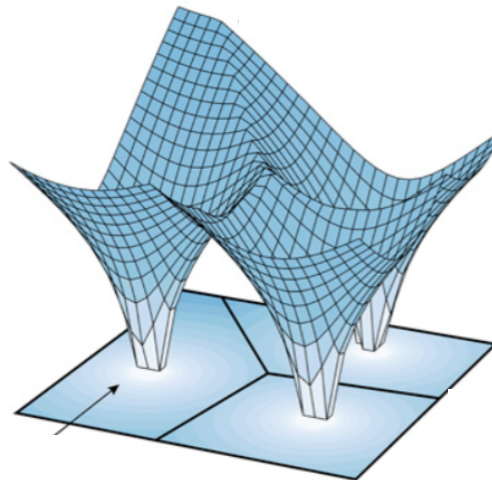
MELD uses a Bayesian inference approach to incorporate data into simulations

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

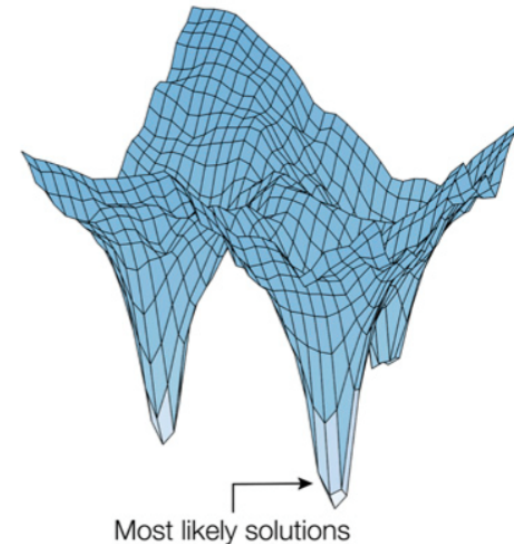
Force Field
(prior)



MELD Energy
(likelihood)



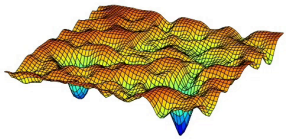
MELD Energy + Force Field
(posterior)



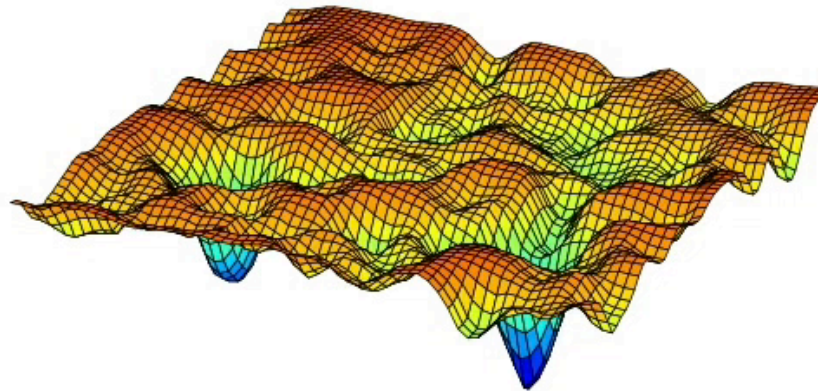
$$p(\mathbf{x}) \sim \exp[-\beta E_{\text{force field}}(\mathbf{x})]$$

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

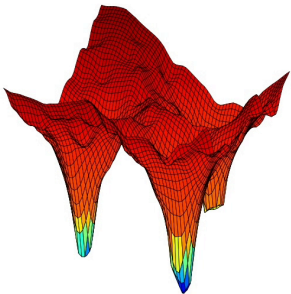
We use Hamiltonian Replica Exchange to enhance sampling



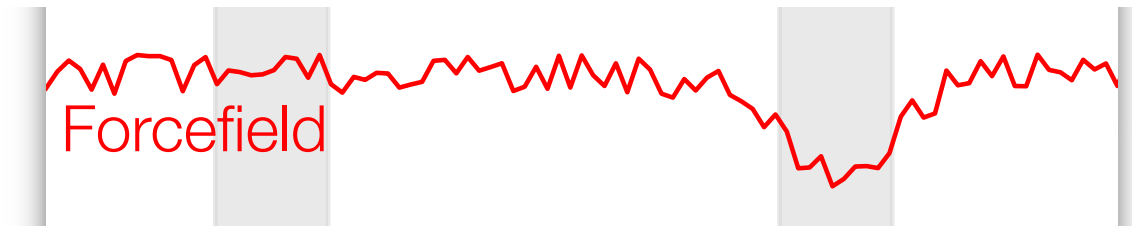
● High Temperature / Weak Restraints



● Low Temperature / **Strong Restraints**

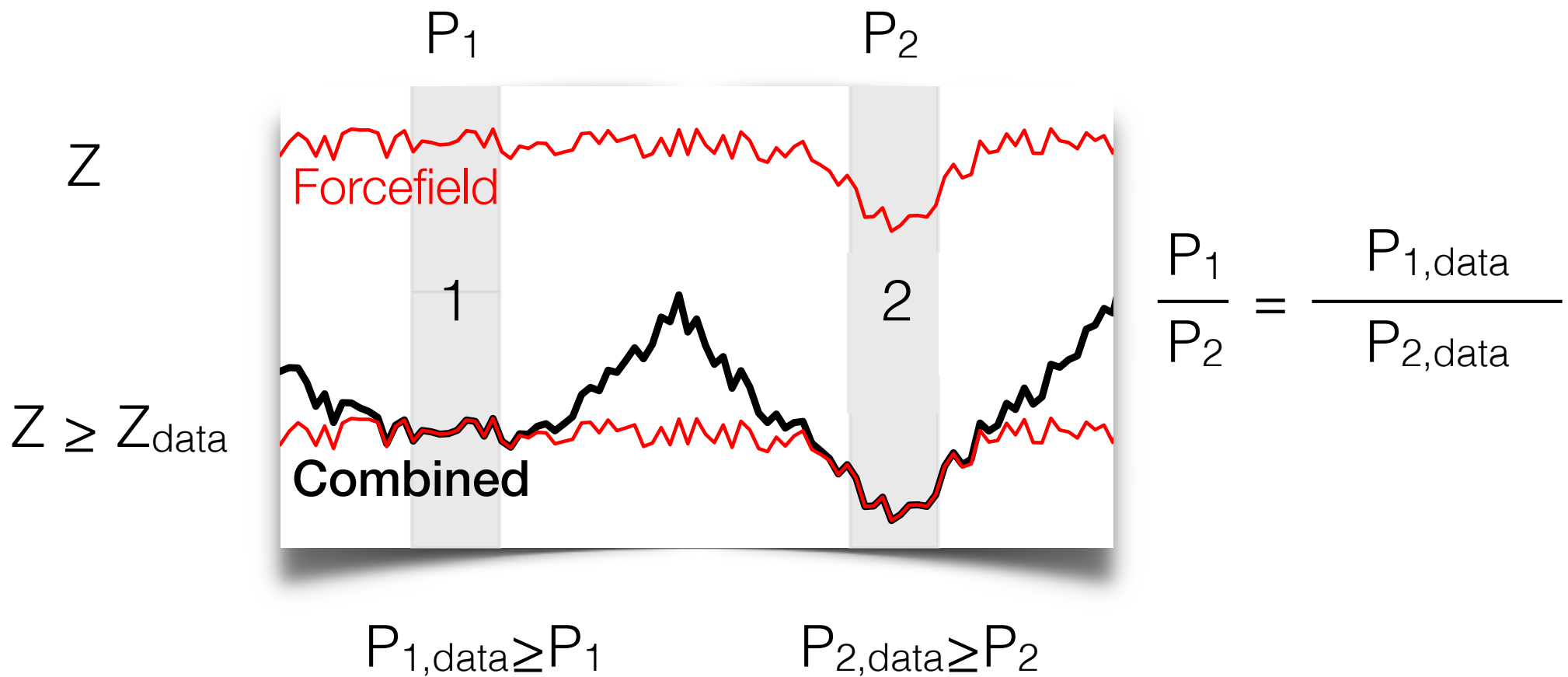


Restraints sculpt the folding landscape creating faster folding funnels



MELD

Restrains sculpt the folding landscape creating faster folding funnels



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

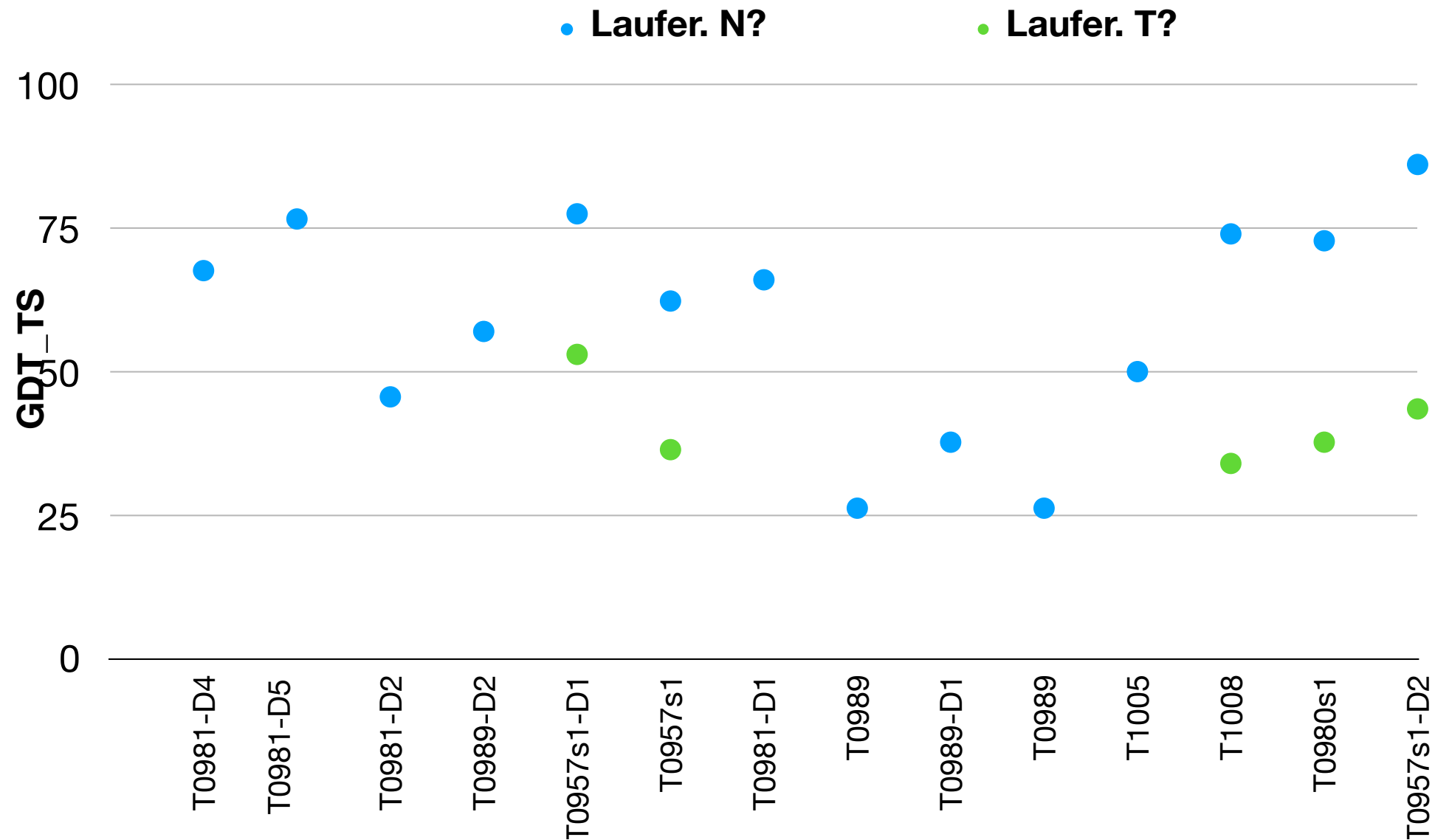
We used three metrics to select structures from REMD ensembles

Population / free energy

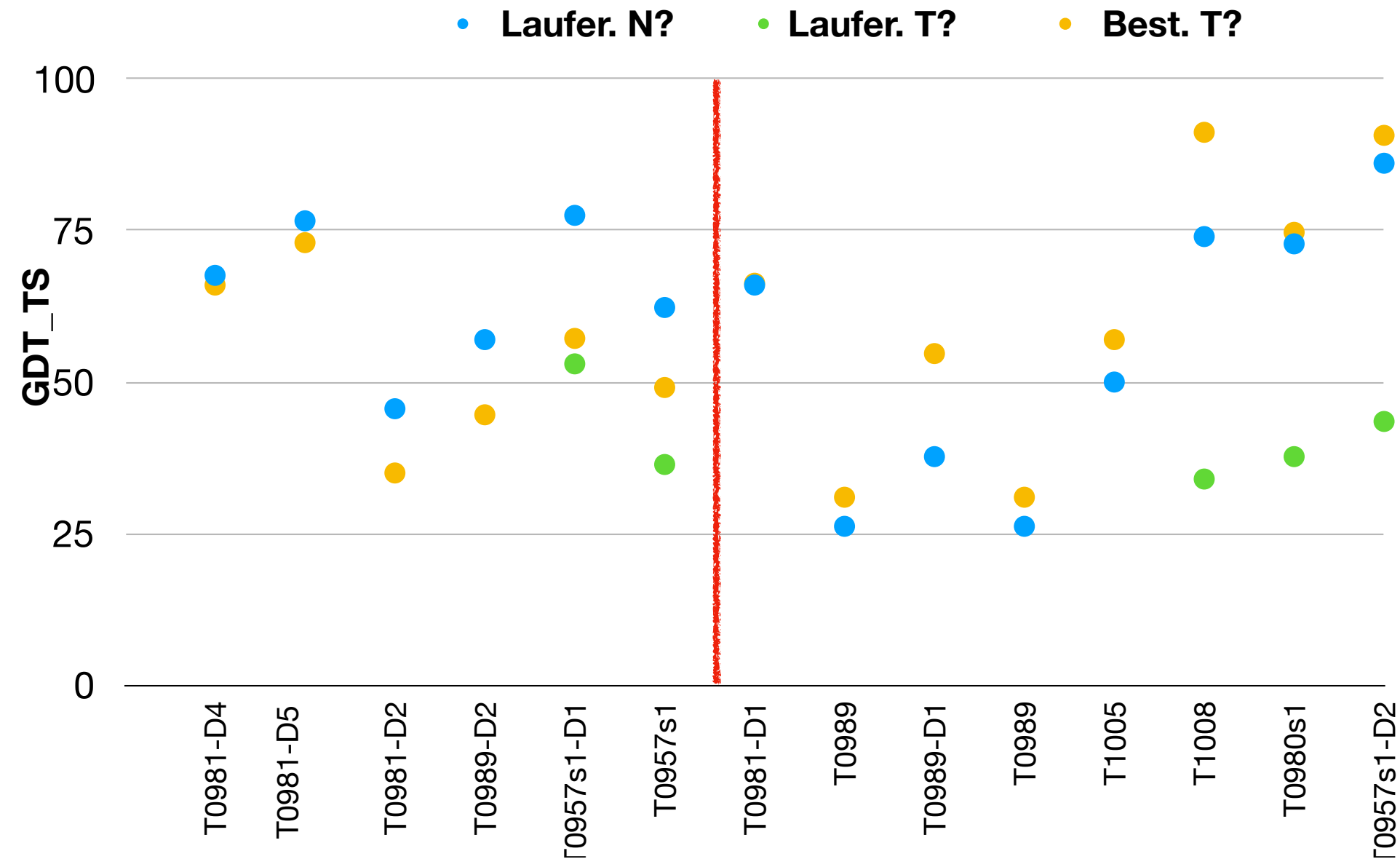
Restrains satisfied

Lowest restraint Energy

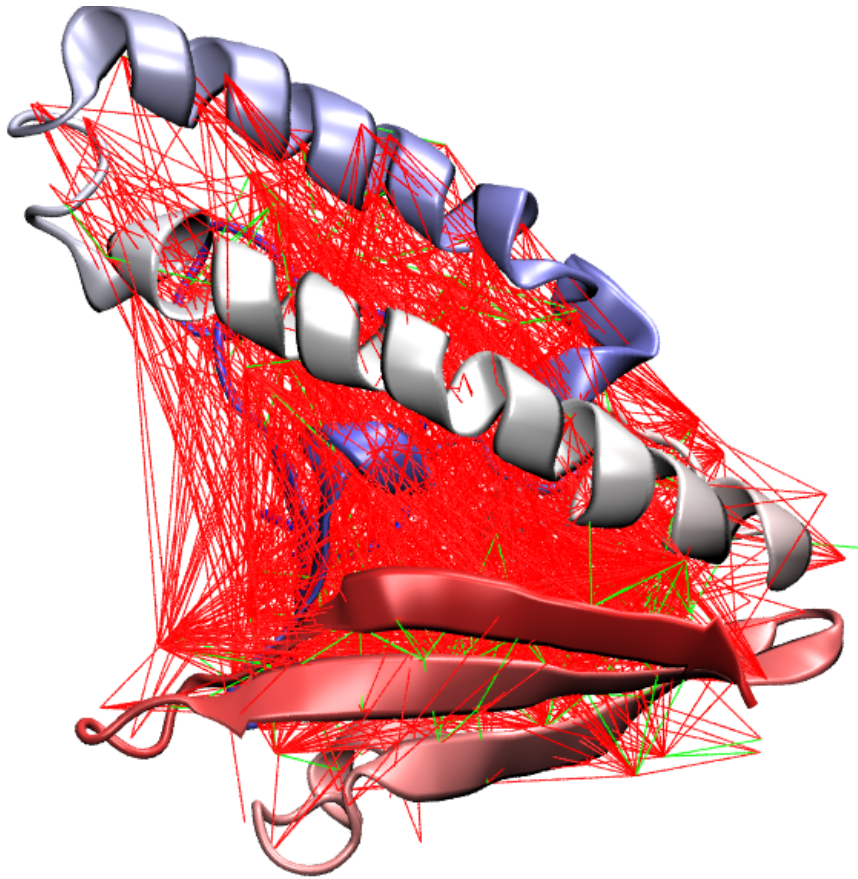
NMR data improves our predictions



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



Protein Structures and Populations with MELD x MD
 Alberto Perez¹, Cong Liu², Roy Nassar², James C. Robertson², Emiliano Brini², Ken A. Dill²

MELD x MD is a Bayesian inference method...
 $p(x|D) = \frac{p(D|x)p(x)}{p(D)} \sim p(D|x)p(x)$
 posterior: MELD energy + MD force field
 likelihood: MELD energy
 prior: MD force field

...that uses Hamiltonian, Temperature Replica Exchange Molecular Dynamics...

...to give native structures and populations.

Can MELD x MD fold small proteins with heuristics?
 • Targets: 125 amino acids were selected
 • Each replica starts from extended conformation
 • Heuristics inform restraints
 • Hydrophobic cores
 • PSIPRED secondary structure
 • Beta strand pairing

Can MELD x MD improve refinement?
 • Server predictions were selected to each replica
 • Targets: 250 amino acids were selected
 • Heuristic-informed restraints
 • Metagenome-derived co-evolutionary restraints

NMR Data
 • Ambiguous NOE contacts
 • Refinement target selected to all replicas
 • 4 MELD x MD predictions and 1 MD prediction

Laufer_abinito
 • Targets: 250 amino acids were selected
 • Similar to Laufer_100 with co-evolutionary data added
 • Co-evolutionary restraints
 • Metastable restraints
 • GREMLIN predicted contacts based on MSA sequences per length, score, and probability of contact being correct

Do speeded simulations improve convergence?
 • Can MELD x MD aid NMR data assigned predictions?
 • Can MELD x MD improve refinement?

Laufer_100 folded small proteins
 • CASP13
 • Prior CASPs
 • Fast folding proteins
 • Nonthreadables

Laufer folded larger proteins and predicted structures using NMR data

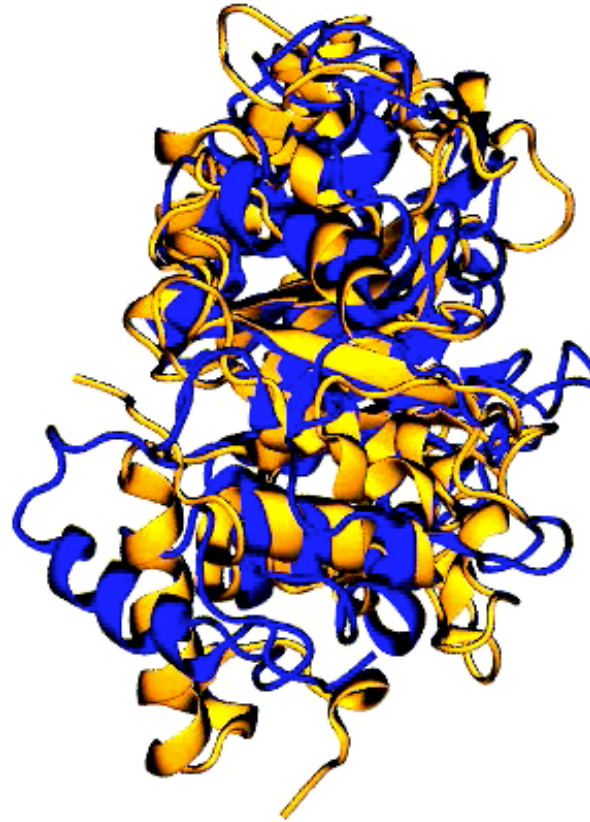
Laufer_abinito converges faster

Primary results show faster populations with Laufer_abinito vs. Laufer_100

Cluster Representations in RMSD (Å)
 Higher populations are associated with folding success.



N1005



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