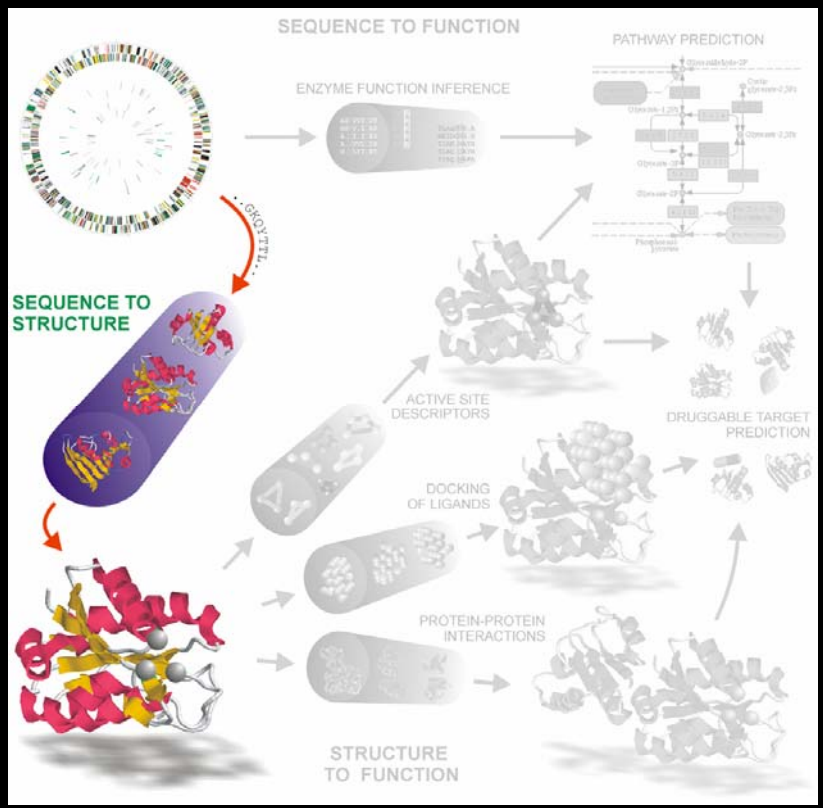
Prediction and refinement of NMR structures from sparse experimental data

Jeff Skolnick
Director
Center for the Study of Systems Biology
School of Biology
Georgia Institute of Technology

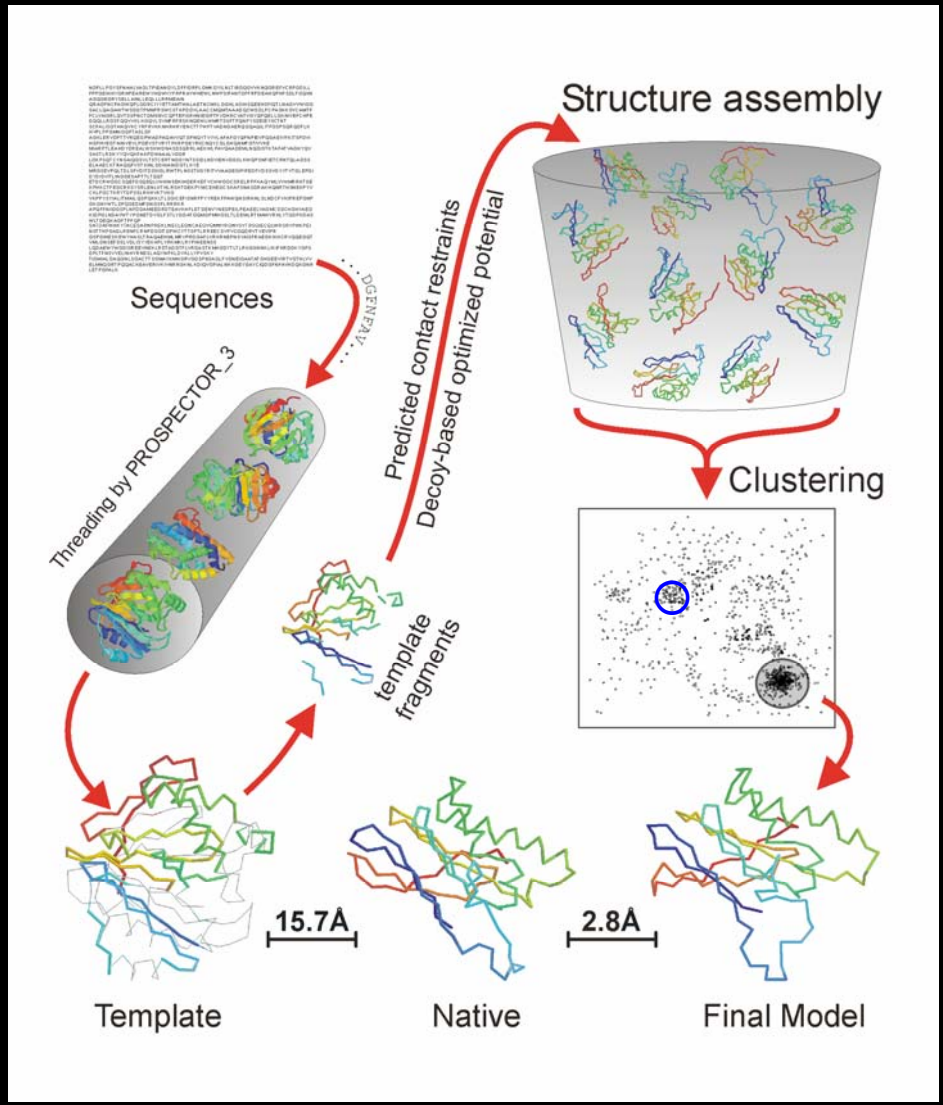
Overview of talk



- General methodology
- Structure prediction on weakly homologous proteins with/without experimental restraints
- Refinement of NMR structures and comparison with X-ray structures
- Conclusions

General Methodology

*TASSER:T*hreading/*ASSE*mbly/*R*efinement



Recipe for protein structure prediction

- Protein Model

  At what level of detail should the protein be represented?

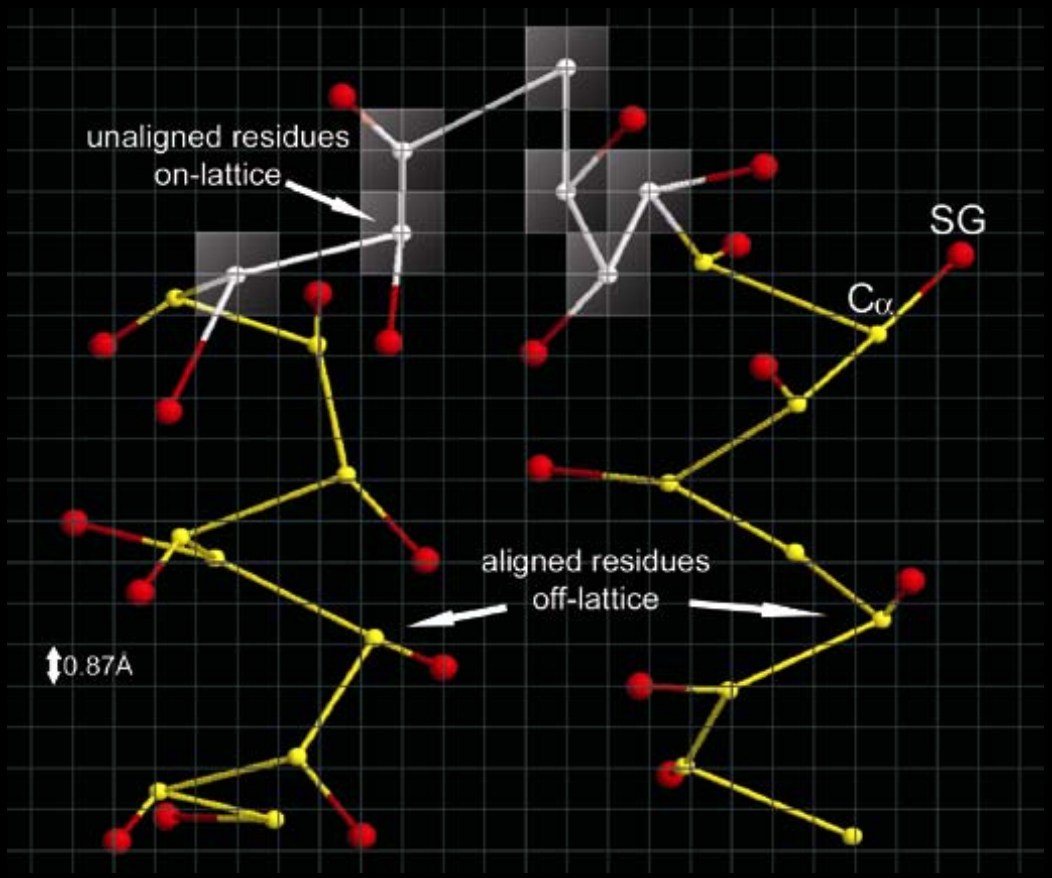
- Potential

  What types of “energy terms” should be used?

- Conformational search scheme

  How does one search for the native state?

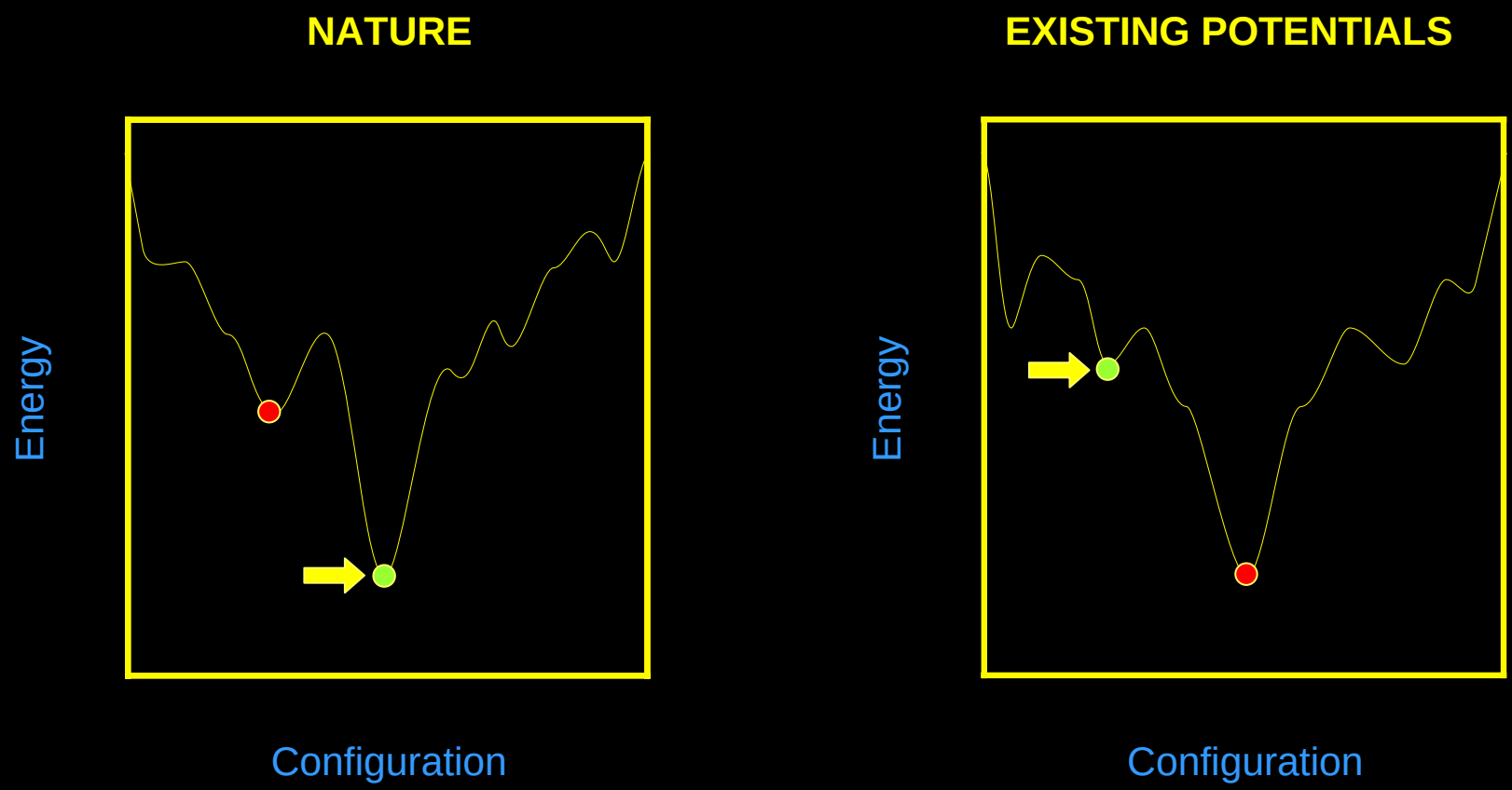
Protein Model



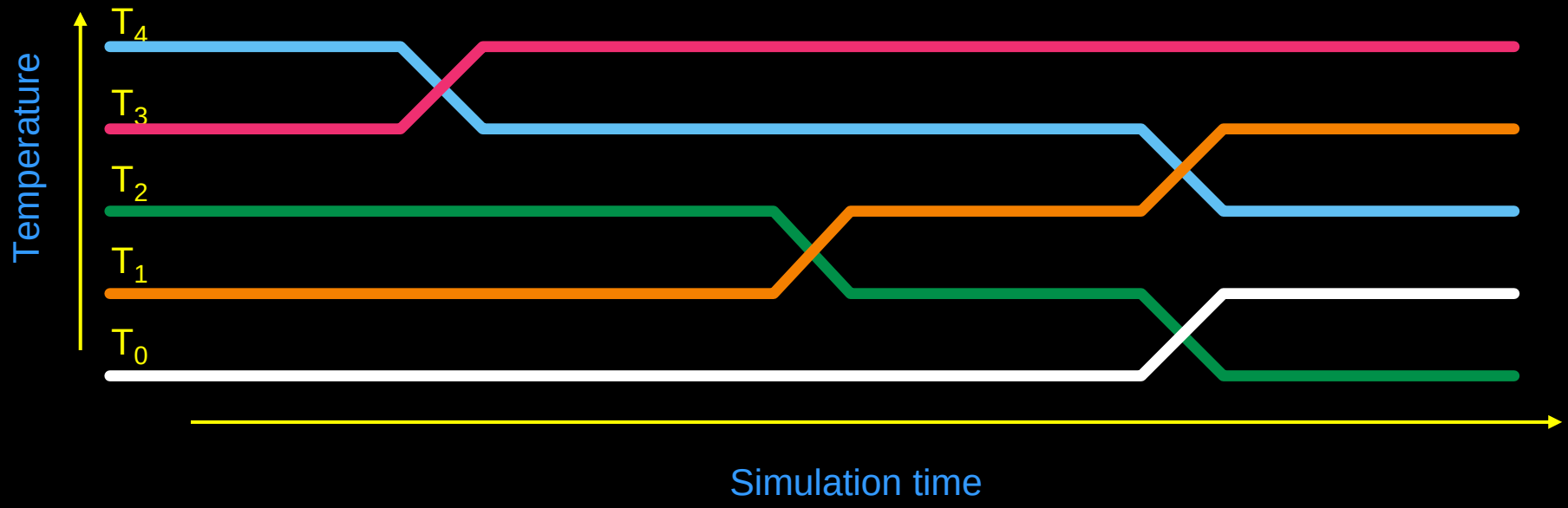
Potential:

- Generic terms that describe a protein such as *hydrogen bonding* and local conformational stiffness
- Predicted secondary structure
- Hydrophobic burial and pair interactions
- *Predicted consensus side chain contacts* from the template structures

Potential Energy Surface



CONFORMATIONAL SEARCH SCHEME:
Replica Exchange Monte Carlo



Use of sparse NMR restraints in TASSER

Large Scale Benchmark:

- 1375 non-homologous proteins between 40 and 200 residues that cover the PDB at 35 % sequence identity and which represent all secondary structure classes (α, β, α/β).
- Exclude all templates whose sequence identity to the target sequences >30%. *Weakly homologous limit*
- Compare TASSER runs with and without sparse NMR-derived constraints

Restraint Selection

- Randomly select N/8 restraints with N the number of residues in the protein such that their distribution is more or less uniform.
- Don’t want to select N/8 restraints such that many are clustered in on region of the structure.

How are the restraints implemented?

- The strength of a restraint depends on the sequence separation, with those having the largest sequence separation twice that of those with the shortest separation in sequence.
- The strength of the restraints at the end of the simulation is double that at the beginning (to avoid kinetic traps due to the distant-in-sequence restraints)

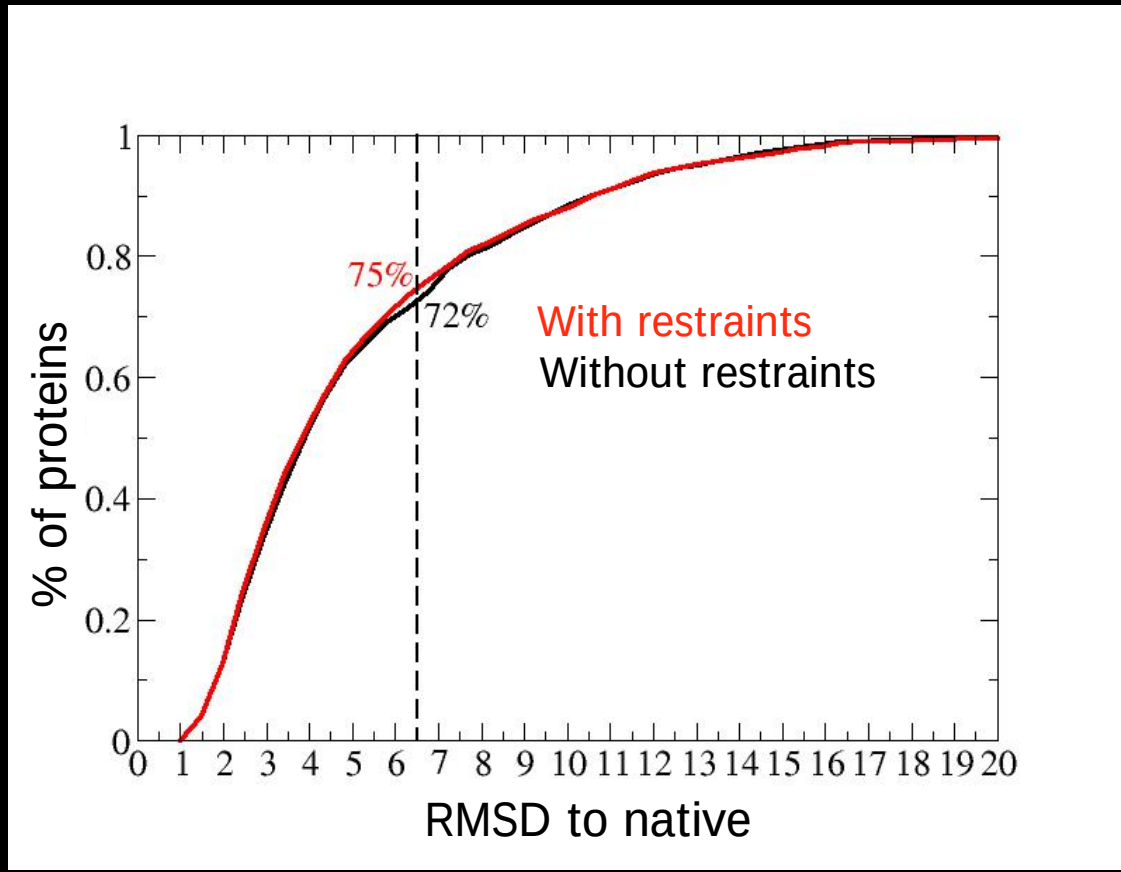
How are restraints implemented?

- The strength of a restraint depends on the sequence separation, with those having the largest sequence separation twice that of those with the shortest separation in sequence.

- The strength of the restraints at the end of the simulation is double that at the beginning (to avoid kinetic traps due to the distant-in-sequence restraints)

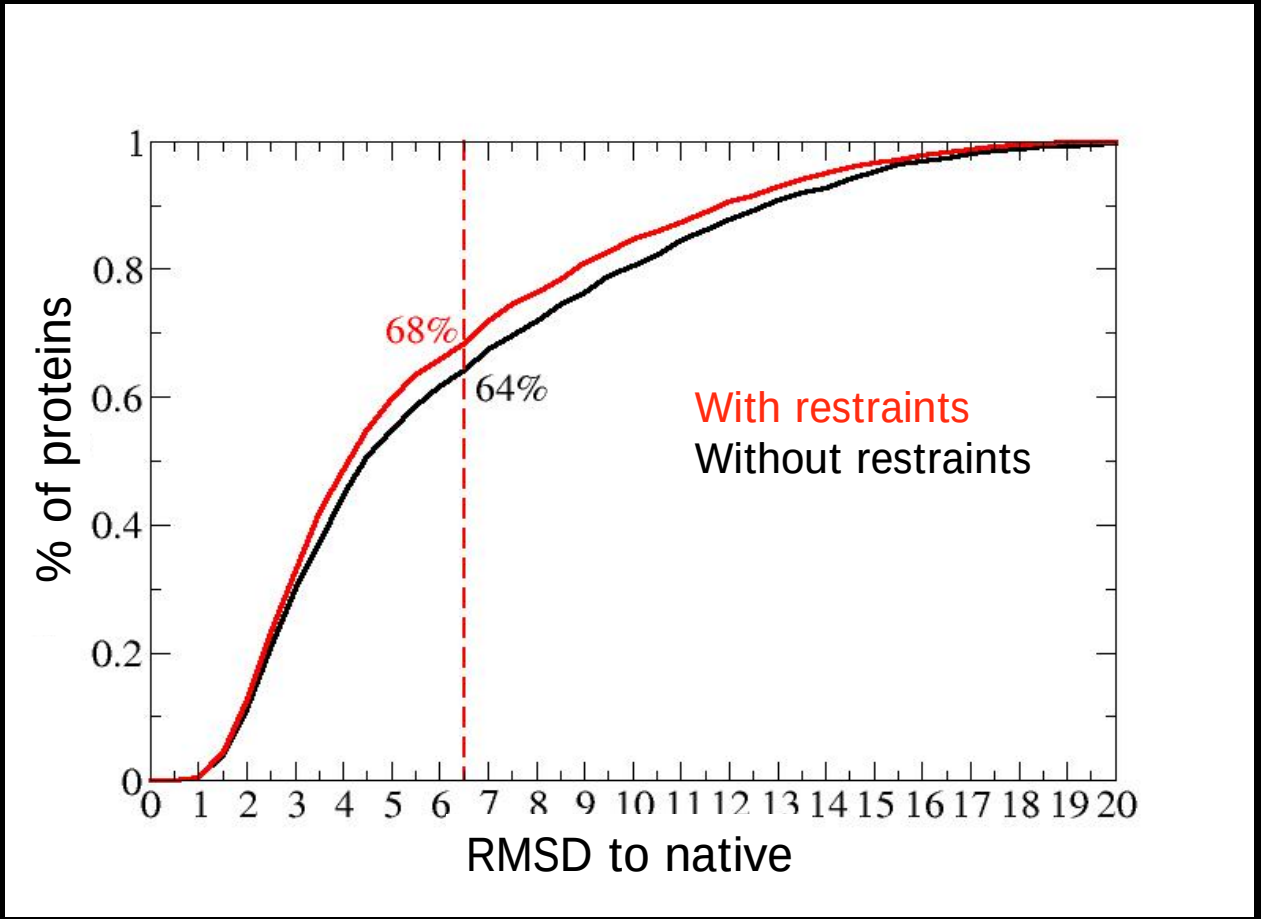
Results:

Comparison of the best of top 5 clusters

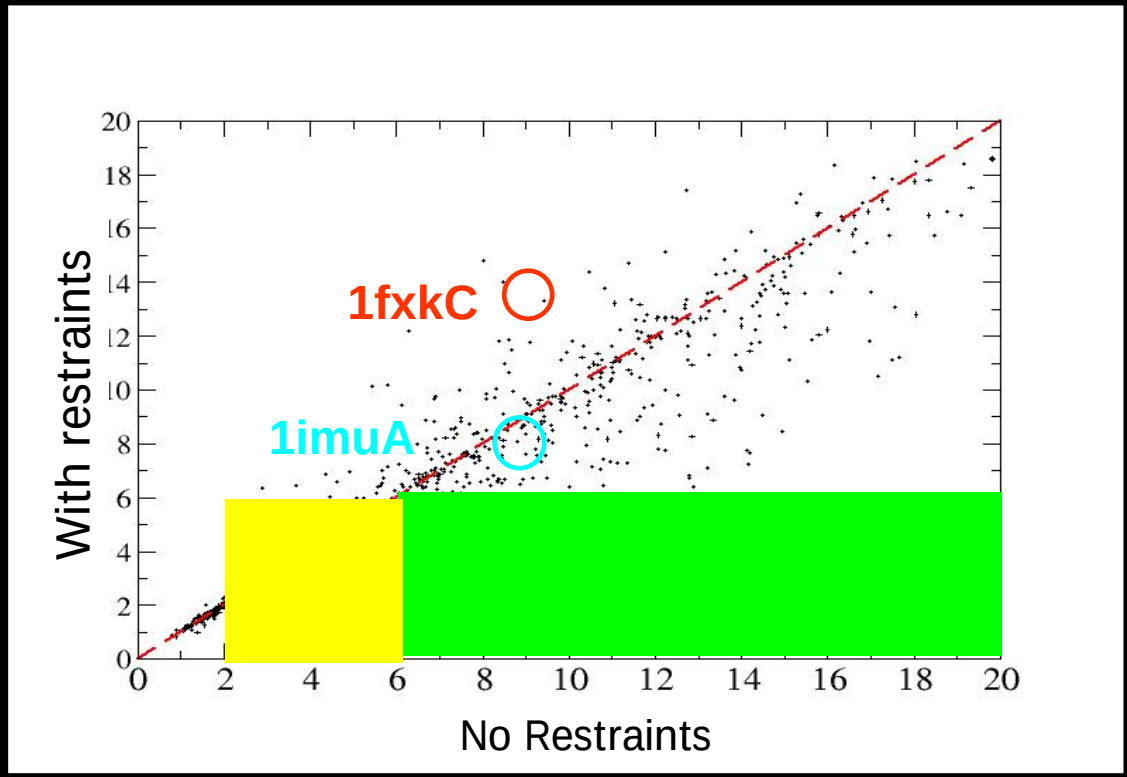


Addition of restraints slightly improves the quality of the results

Comparison of most populated cluster with and w/o restraints

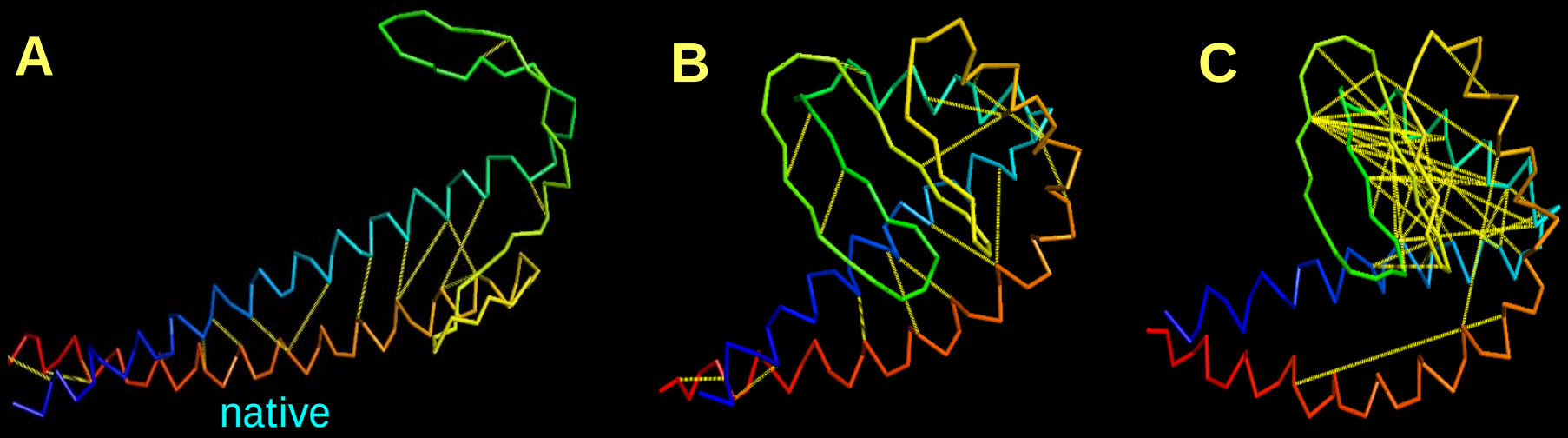


Comparison of the RMSD of the first cluster:



- Restraints decease the RMSD to native on average by 0.5 Å
- Biggest effect is for the poorly predicted proteins w/o exact restraints (>6 Å
  "Green Zone"
- See improvement in the 2-6 Å range of models
  "Yellow Zone"
- Little to no improvement for models <2 Å -resolution of TASSER models

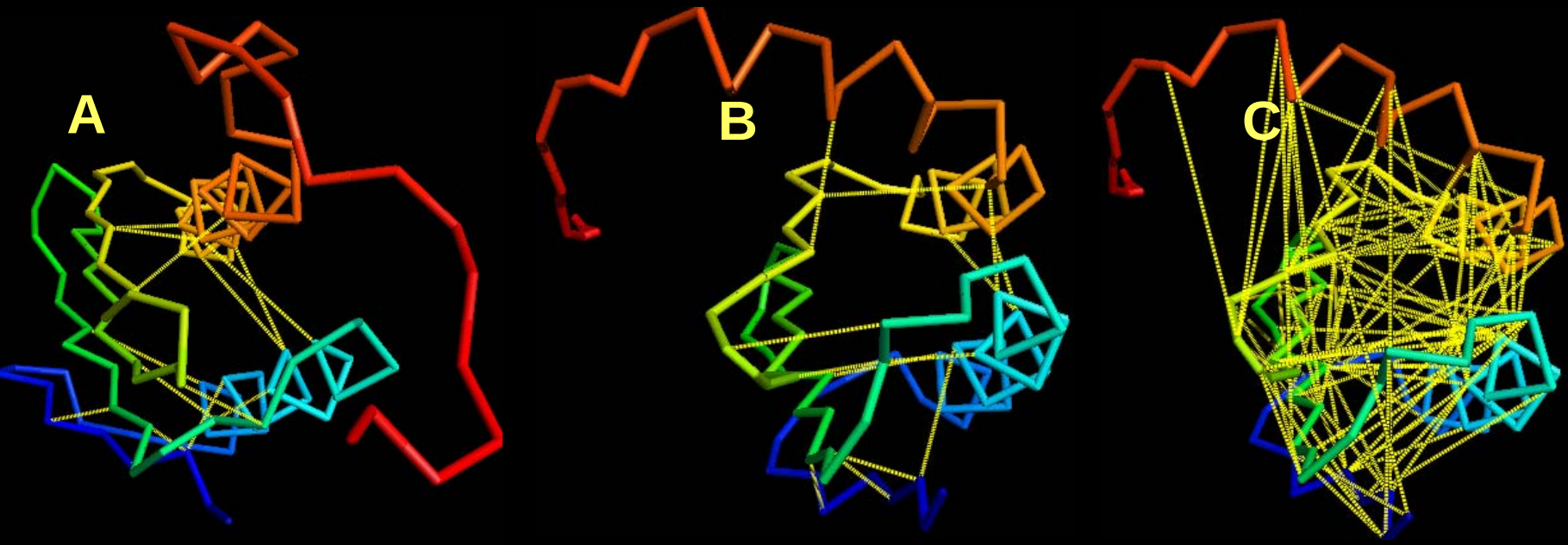
1fxkC: Model with restraints is worse than without restraints



A. Native State with the sparse restraints shown (dotted yellow lines). Green and yellow hairpins are far from each other.

B First cluster structure after adding restraints (RMSD from native=17 Å). The structure satisfies all the restraints. The yellow and green hairpins make a beta sheet, which is packed against the two long helixes. The population of this structure increases after inclusion of exact restraints.

1imuA : No improvement with restraints added



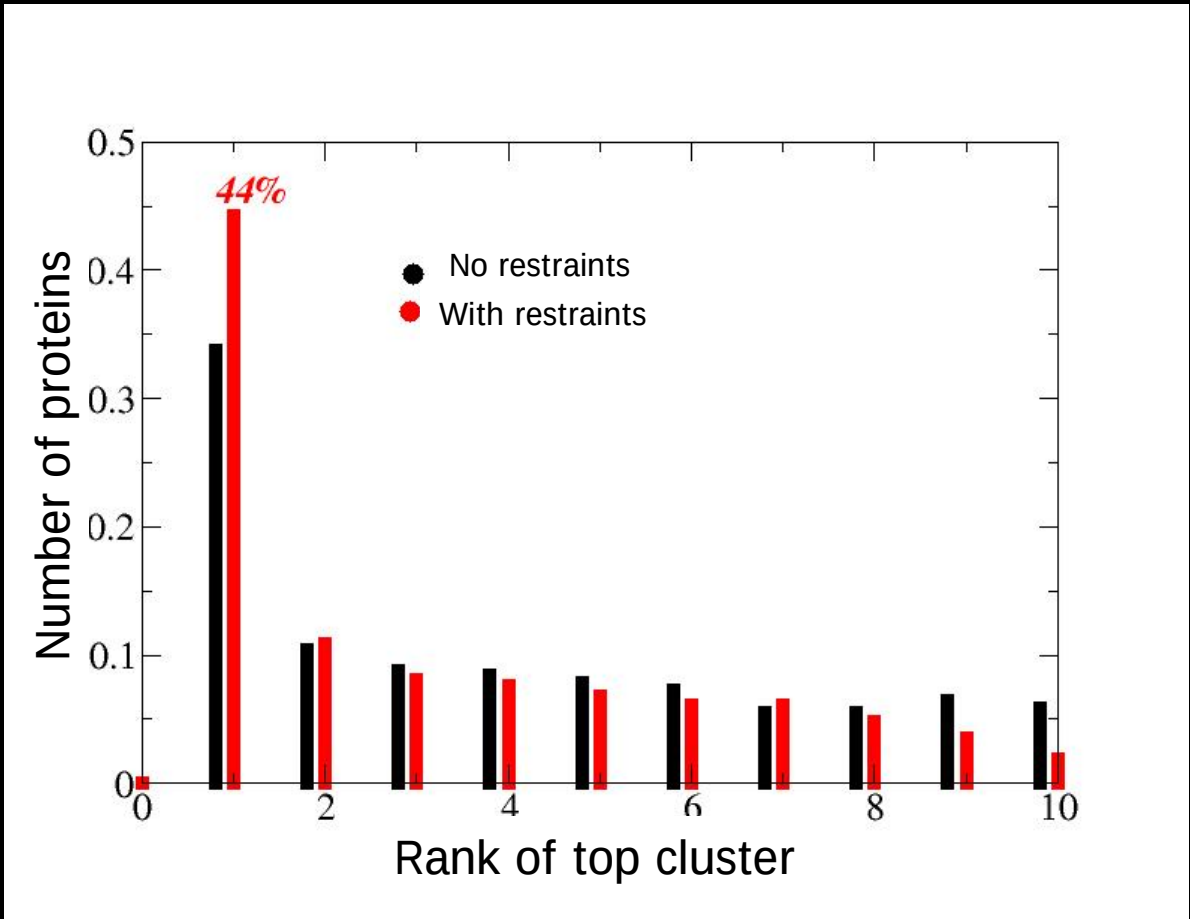
A. Native State showing the selected sparse restraints (dotted yellow lines). C-terminus (in red) is highly unstructured.

B. First cluster structure after adding restraints (RMSD=10 Å). The core is very similar to native, and most differences seem related to the C-terminal helix.

C. Predicted structure with all the predicted restraints

Thus, the predicted restraints incorrectly locate the C-terminal helix!

Cluster Rank Distribution



Addition of restraints improves identification of top cluster as the first cluster (for 44% of all proteins, the first and top cluster coincide). In addition, the number of proteins with top cluster rank decreases monotonically

TASSER refinement of NMR structures

S. Lee, Y. Zhang and J. Skolnick. TASSER-based refinement of NMR structures. Proteins 2006: **63**: 451-456.

Benchmark set

- 61 nonhomologous proteins with both X-ray and NMR structures.
- Require that there be multiple NMR structures in the PDB file
- Basically the same set as used by Garbuzhnskiy et al, Proteins, 60, 139 (2005).

Method

- Identify the consensus region of the NMR models by superimposing models 2 to N onto the first NMR model.
- Calculate the average distance to residue j in all the models.
- Chose the cutoff so that 90% of the NMR structure is defined as the “template”.

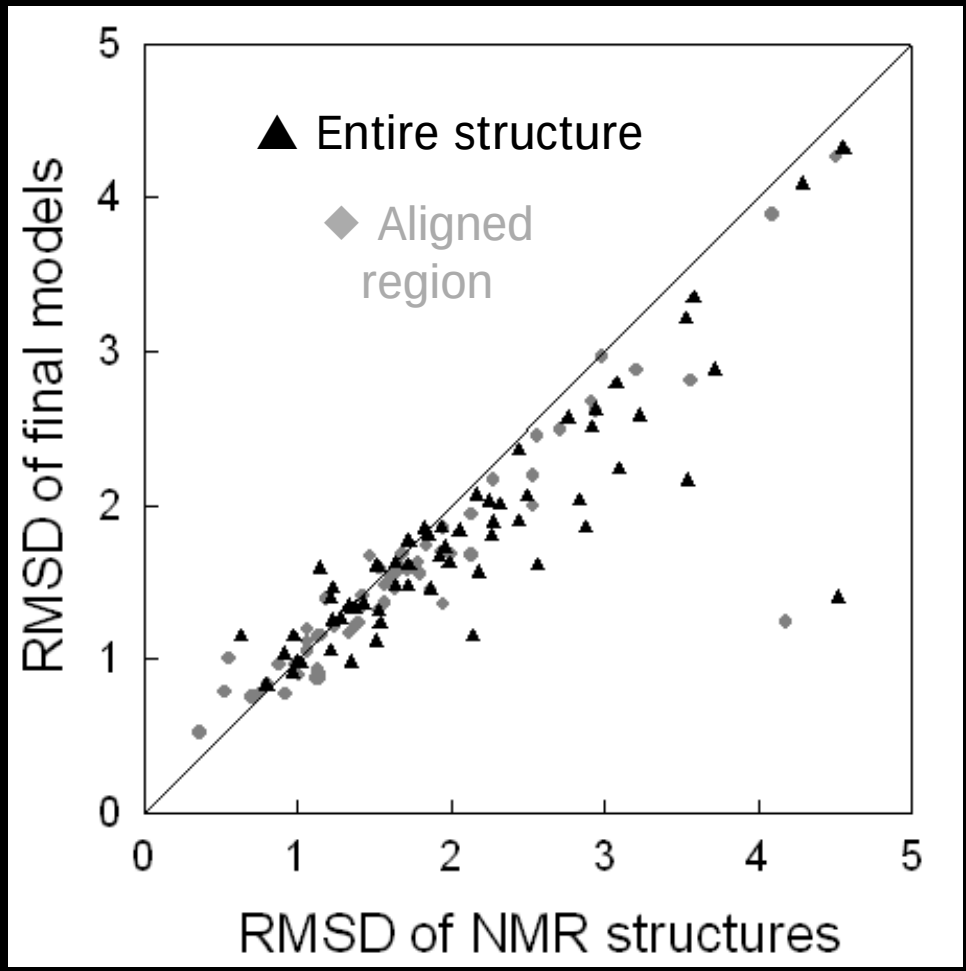
Results:

Table I. Comparison of final models from TASSER with X-ray structures

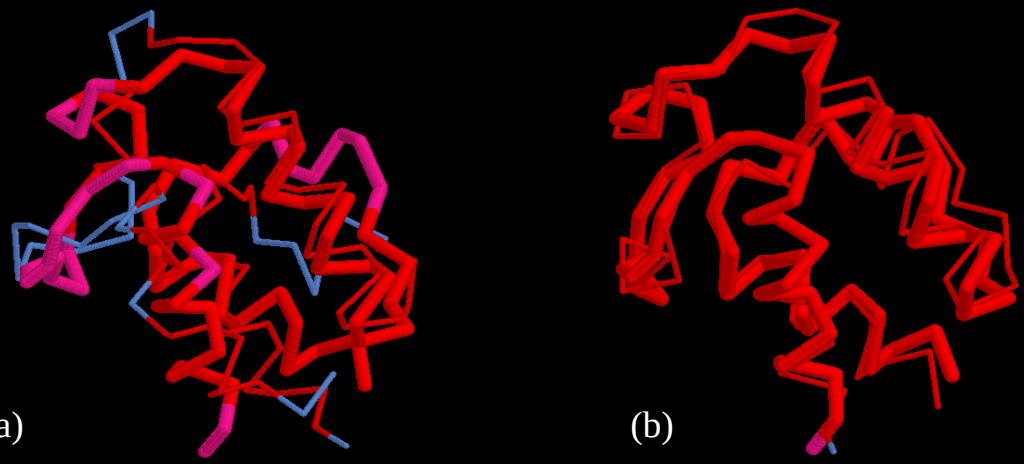
RMSD NMR/X-ray	RMSD to X-ray (ali)[a]		RMSD to X-ray (all)[b]		TM-score to X-ray (all)[c]	
	NMR	Model	NMR	Model	NMR	Model
<1	0.664	0.814	0.860	1.014	0.940	0.915
<2	1.222	1.179	1.428	1.377	0.885	0.887
<3	1.495	1.384	1.797	1.601	0.850	0.861
<4	1.600	1.474	1.959	1.709	0.841	0.853
<5	1.731	1.556	2.080	1.785	0.828	0.846

[a]NMR and final TASSER models over the same aligned regions;
[b] NMR and TASSER models over the entire chain.
TM-score to X-ray structures: [c]NMR and TASSER models over the entire chain.

RMSD to X-ray structure



Representative example of TASSER refinement

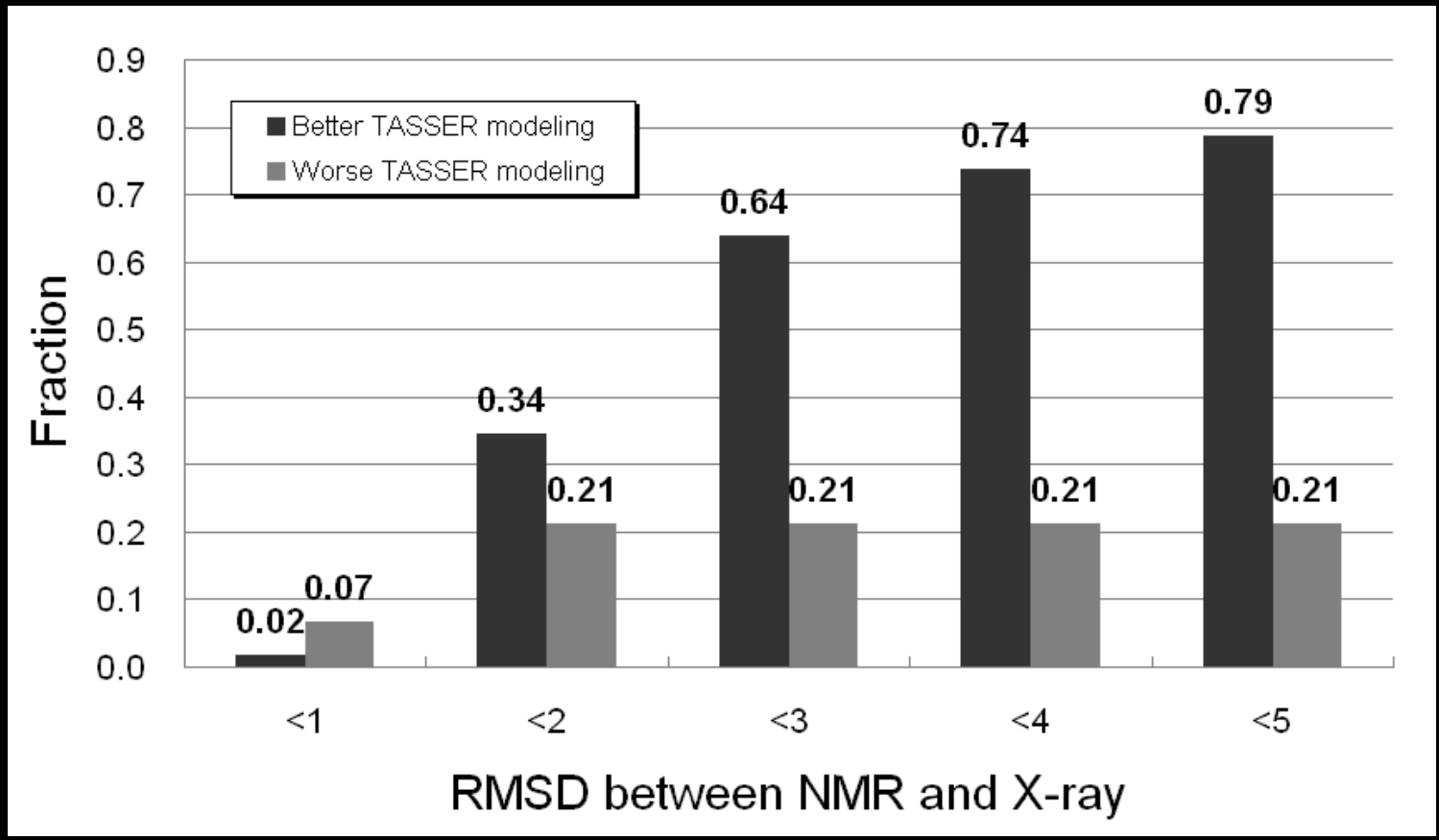


a. Superposition of NMR structure to NATIVE (RMSD = 4.3 Å)

b. Superposition of TASSER model to Native (RMSD=1.4 Å).

Red indicates residues <5 Å after superposition.

Histogram of better/worse RMSD relative to the initial X-ray/NMR RMSD



Average RMSD of TASSER refined models to X-ray s 1.79 Å
Average RMSD of NMR to X-ray structures = 2.1 Å

Conclusions

Sparse NMR Restraints

- Expect about 70% of nonhomologous proteins to have an acceptable model
- Effect of sparse restraints on best of top 5 predicted structures is very minor
- Sparse exact restraints DO improve the quality of the most populated structure, with the biggest effect for structures whose RMSD > 6 Å in the absence of exact restraints
- Also see improvement in the 2- 4 Å range
- On average little improvement in structures whose RMSD < 2 Å in the absence of exact restraints- reflects resolution of the model

NMR-X-ray Studies

- If the NMR-X-ray structure has a RMSD >2 Å from native, TASSER models systematically move closer to the X-ray structure
- If the NMR-X-ray structure RMSD <2 Å, then 21/34 models (60%) move closer. This reflects the inherent resolution of TASSER.

Software dissemination

- TASSER is a mature suite of programs that is ready for wide dissemination to assist in the determination of protein structures by NMR
- Have a TASSER-Lite is available as a web based service

  http://cssb.biology.gatech.edu/skolnick/webservice/tasserlite/index.html
- Will be releasing a version for downloading for academic users

Acknowledgements

Sparse NMR Restraints

Jose Borreguero GA Tech

Yang Zhang University of Kansas

NMR to X-ray Refinement

Seung Yup Lee GA Tech

Yang Zhang University of Kansas