

Protein Structure Determination from Limited Data

Jens Meiler



Center for Structural Biology
Departments of Chemistry, Pharmacology, and
Biomedical Informatics

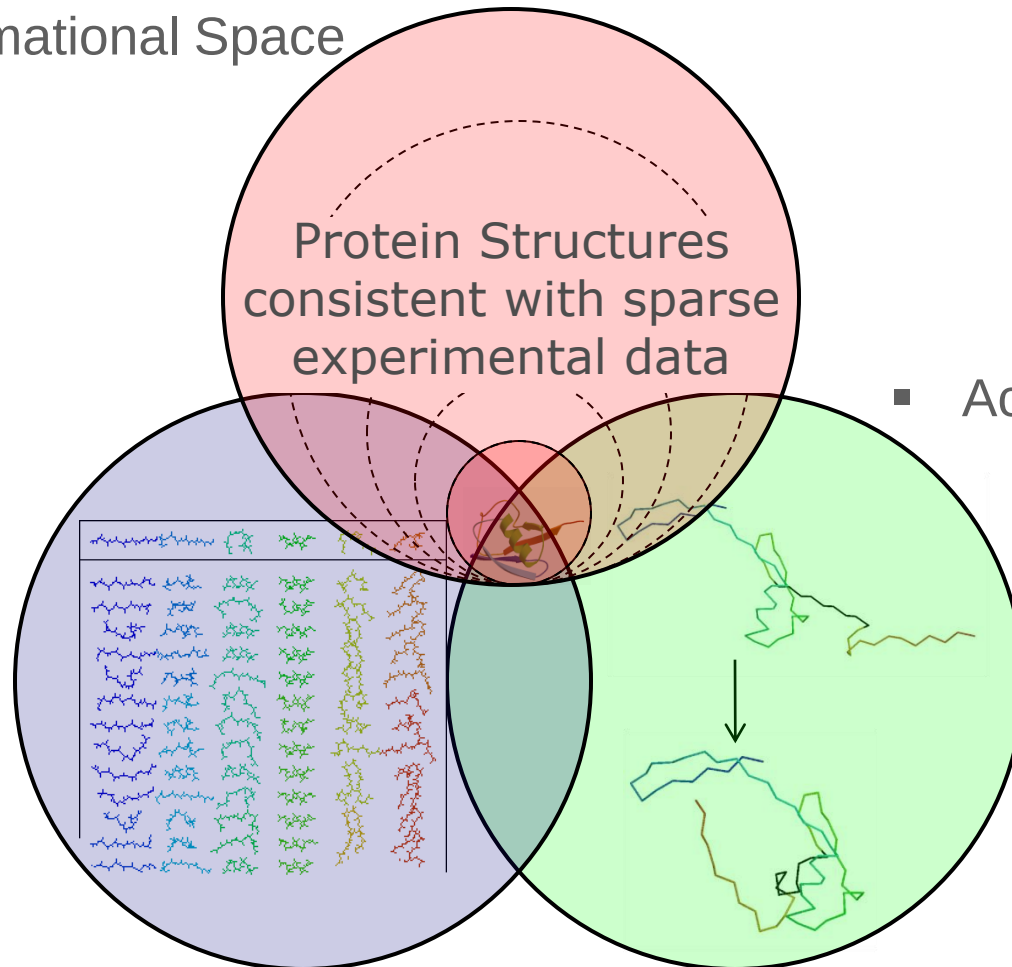
Combining Strengths: Building Accurate Models from Limited Data



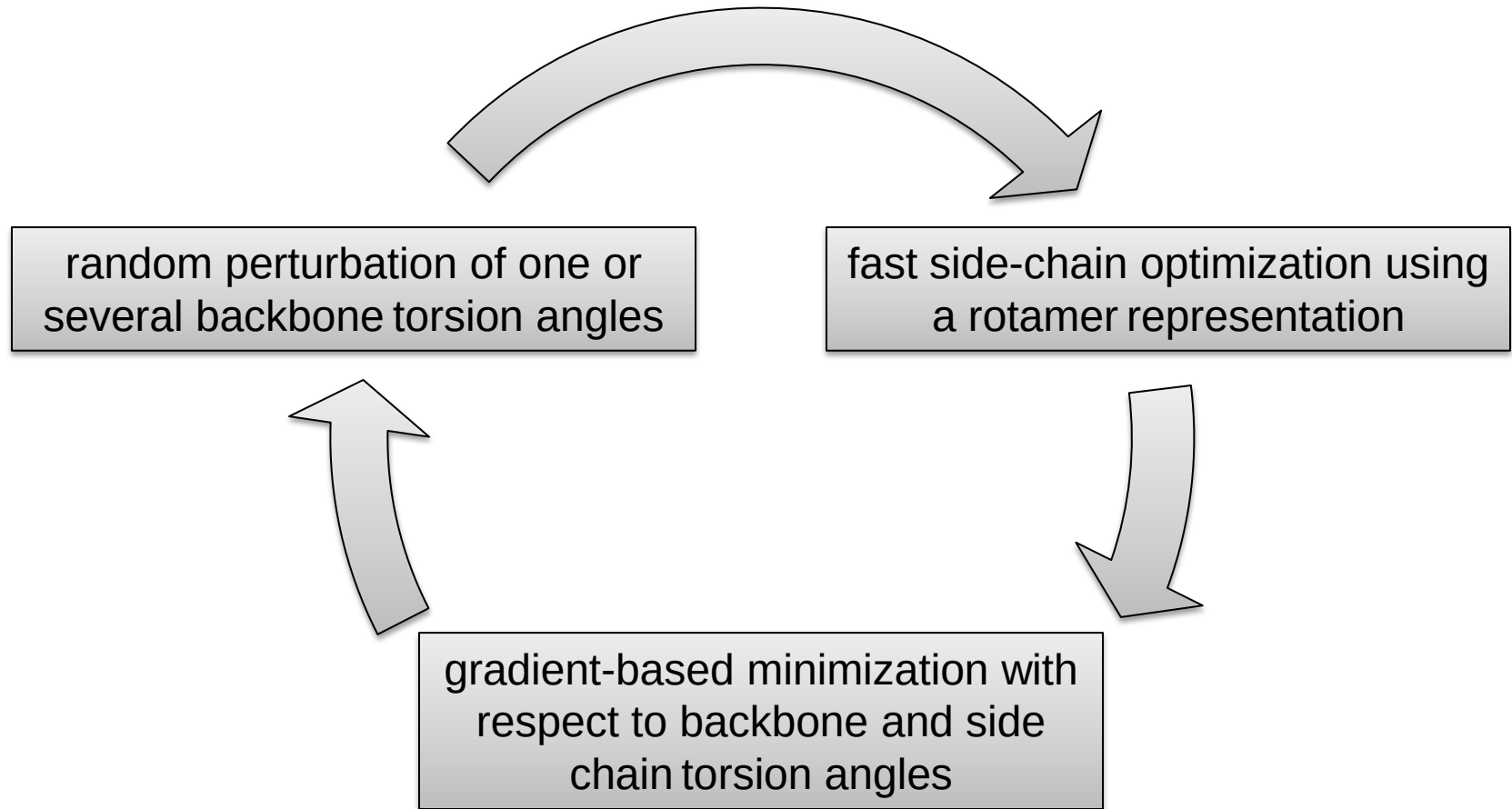
- Complete Conformational Space

- Efficient Sampling Strategy

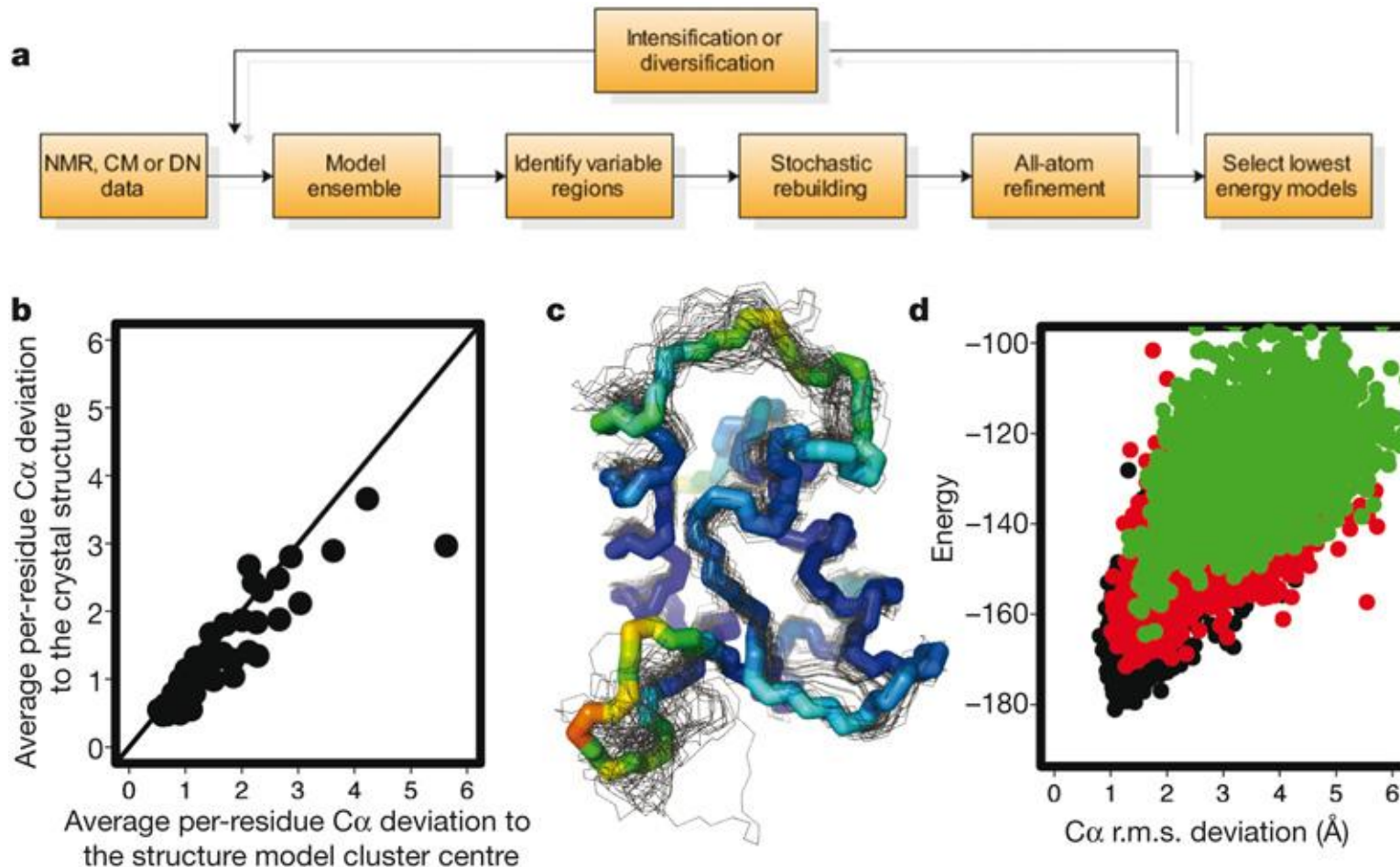
- Accurate Energy Function



Refinement Cycle with Side Chain Repacking and All Atom Minimization



Overview of the Rebuilding-and-Refinement Method.



Qian, B.;
Raman, S.;
Das, R.;
Bradley, P.;
McCoy, A. J.;
Read, R. J.;
Baker, D.
"High-resolution
structure
prediction and
the
crystallographic
phase problem"
Nature **2007**.

Improvement of model accuracy and molecular replacement

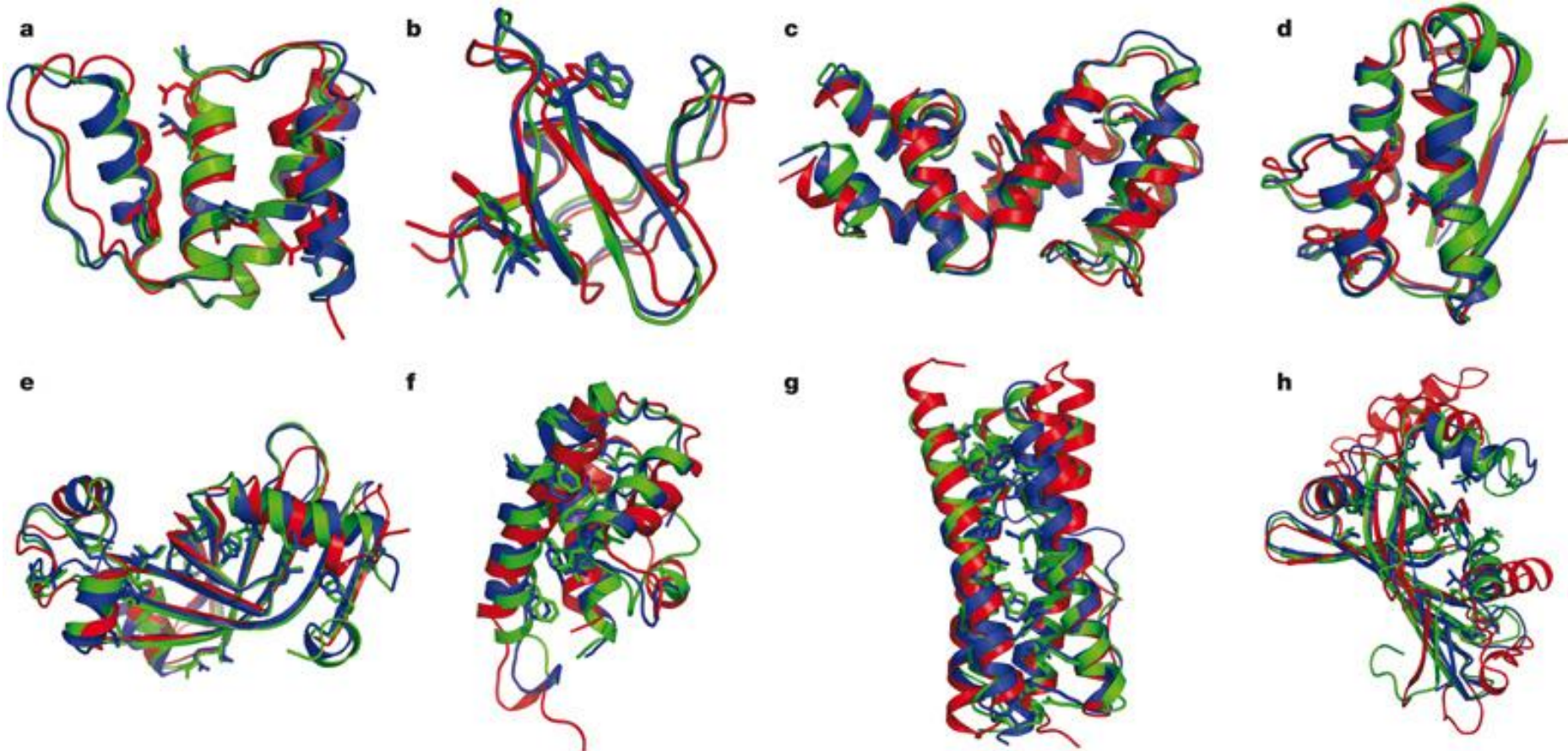


Table 1 | Improvement of model accuracy and molecular replacement by a rebuilding and refinement protocol

	X-ray structure	Starting model*	Length (n)†	Sequence identity to best template (%)‡	GDT-HA§		TFZ in molecular replacement		Auto-traced residues (backbone, side chain)*¶	
					Best template	Refined model	Best template	Refined model	Best template	Refined model
NMR	1hb6	2abd	86	N/A	0.58	0.79	4.1	11.3	12, 0	80, 80
	1who	1bmw	94	N/A	0.59	0.68	5.7	8.3	25, 12	47, 44
	1gnu	1kot	119	N/A	0.64	0.73	6.6	10.6	62, 53	82, 78
	1a19	1ab7	89 (2)	N/A	0.63	0.78	3.7	8.8	31, 20	48, 37
							4.5	12.5	14, 0	44, 35
	1fvk	1a24	189 (2)	N/A	0.49	0.69	3.4	6.9	66, 50	97, 91
							4.3	12.4	55, 43	85, 68
	1mzl	1afh	93	N/A	0.60	0.66	4.6	5.1	36, 29	58, 44
	1tvq	1xpw	143	N/A	0.63	0.74	4.3	6.7	15, 6	103, 86
	2snm	2sob	97	N/A	0.45	0.48	3.8	4.8	17, 16	43, 37
	1agr	1ezy	129	N/A	0.49	0.76	N/A#		N/A#	
	1abq	1awo	56	N/A	0.58	0.83	N/A☆		N/A☆	
CM	2hhz (T0331)	1ty9A	149	14.5	0.49	0.58	5.4	8.8	28, 24	68, 63
	2hr2 (T0368)	2c21C	158 (6)	14.8	0.57	0.67	6.0	5.4	37, 37	20, 14
	2hq7 (T0380)	2fhqA	145 (2)	25.4	0.58	0.69	4.4	6.6	47, 23	92, 83
							4.6	14.2	30, 17	60, 59
	2ib0 (T0385)	1jgcB	170 (2)	7.8	0.62	0.69	5.1	7.9	63, 37	56, 56
							5.8	15.5	50, 2	52, 52
	2hi0 (T0329_D2)	1rqlA	92 (2)	8.8	0.52	0.67	N/A#		N/A#	
	2hcf (T0330_D2)	1lvhB	75	14.1	0.51	0.65	N/A#		N/A#	
	2hi6 (T0357)**	1aco	132	8.4	0.45	0.52	N/A**		N/A**	
DN	2hh6 (T0283)	2b2j	112	3.6	0.22	0.64	5.4	9.0	26, 12	112, 112

Qian, B.; Raman, S.; Das, R.; Bradley, P.; McCoy, A. J.; Read, R. J.; Baker, D. "High-resolution structure prediction and the crystallographic phase problem" *Nature* **2007**.

Refinement of NMR Structures (a-d) and Comparative Models (e-h)



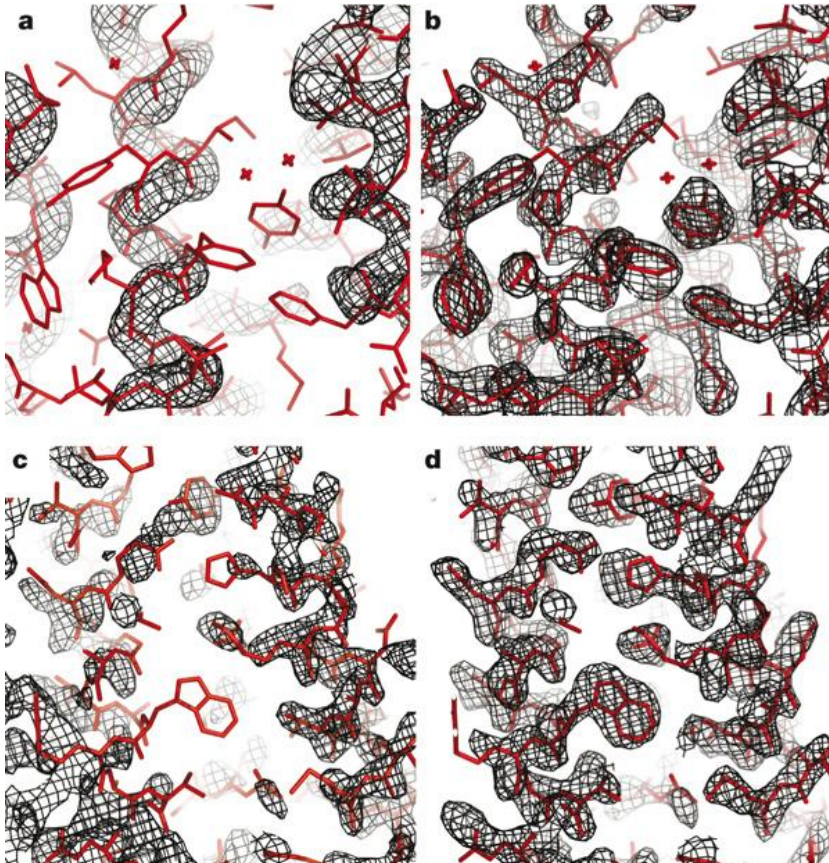
Qian, B.; Raman, S.; Das, R.; Bradley, P.; McCoy, A. J.; Read, R. J.; Baker, D. "High-resolution structure prediction and the crystallographic phase problem" *Nature* **2007**.

native crystal structure (blue), template/NMR structure (red), and the refined model (green)

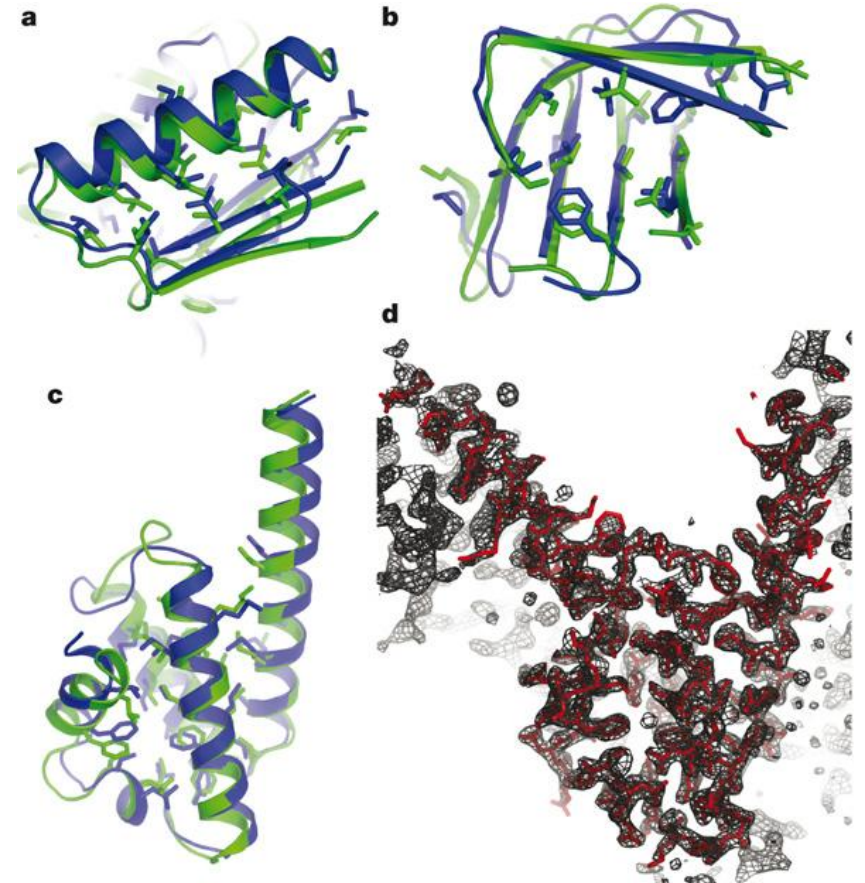
Phasing by refined NMR structures, comparative and *de novo* models



■ NMR (a $\Rightarrow$ b) / CM (c $\Rightarrow$ d)



■ *de novo* model (c $\Rightarrow$ d)

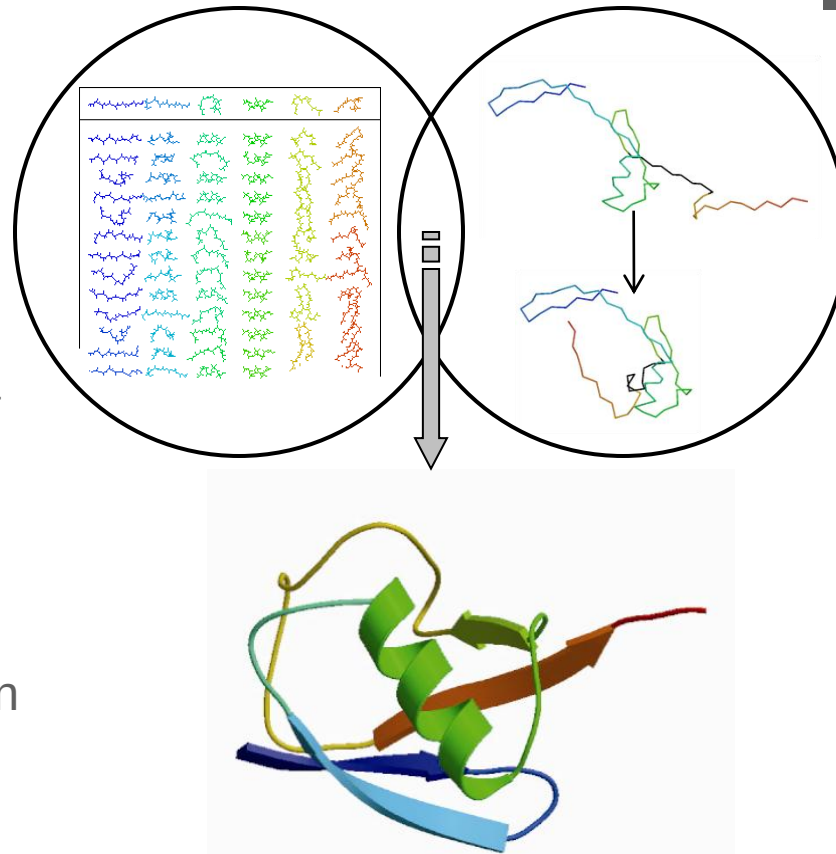


RosettaNMR: Usage of CSs, NOEs, and RDCs



- NMR data are used *in addition* to the Local Sequence Bias

- CS derived dihedral angle restrictions (via TALOS)
- Local NOE distance restraints
- RDC orientation restraints



- NMR data are used *in addition* to energy evaluation of non-local Interactions
- Long-range NOE distance restraints
- RDC orientation restraints

Bowers, P. M.; Strauss, C. E. M.; Baker, D., *J. Biomol. NMR* **2000**, 18, 311-318.
Rohl, C.; Baker, D., *J. Am. Chem. Soc.* **2002**, 124, (11), 2723-2729.

RosettaNMR: High-Resolution from “one” Restraint per Amino Acid



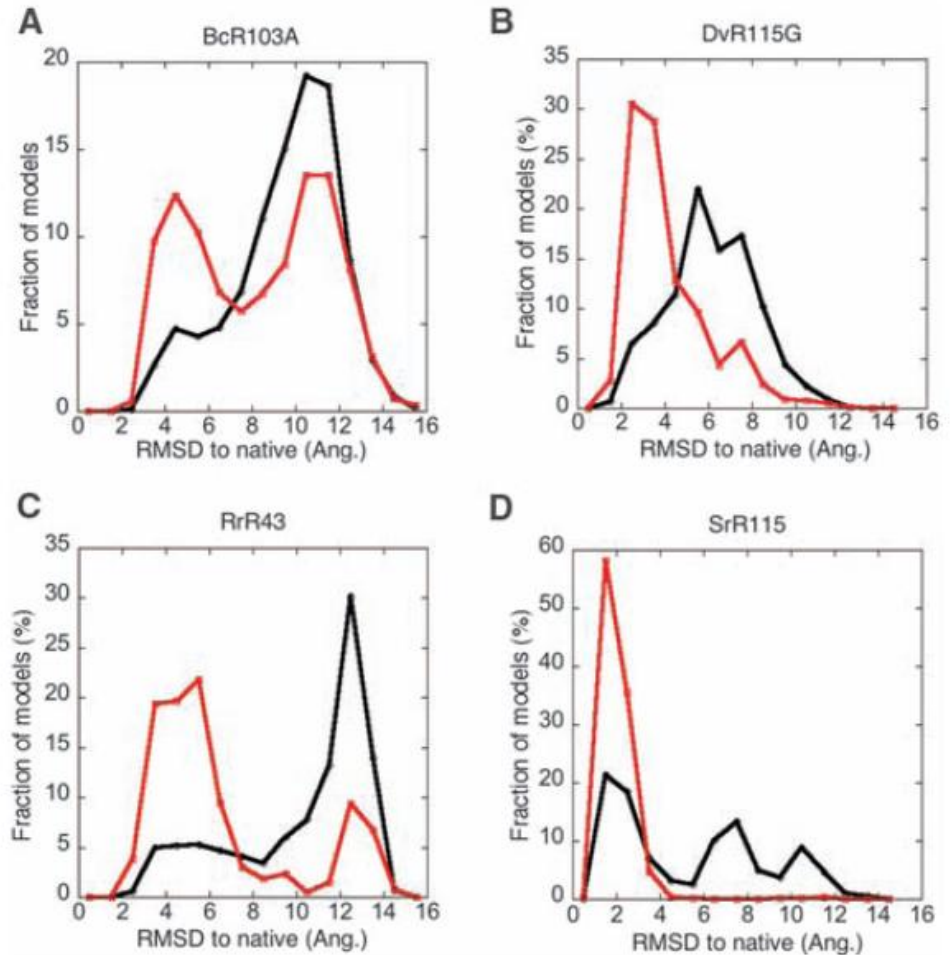
	Xray/NMR	RosettaNMR
■ Profilin I with CSs and 1.01 NOEs/AA		
■ Ubiquitin with 0.89 RDCs/AA		
■ Profilin I with CSs and 1.25 RDCs/AA		

Bowers, P. M.; Strauss, C. E. M.; Baker, D., *J. Biomol. NMR* **2000**, 18, 311-318.
Rohl, C.; Baker, D., *J. Am. Chem. Soc.* **2002**, 124, (11), 2723-2729.

Restraints Enable Dense Sampling of Structures Consistent with NMR

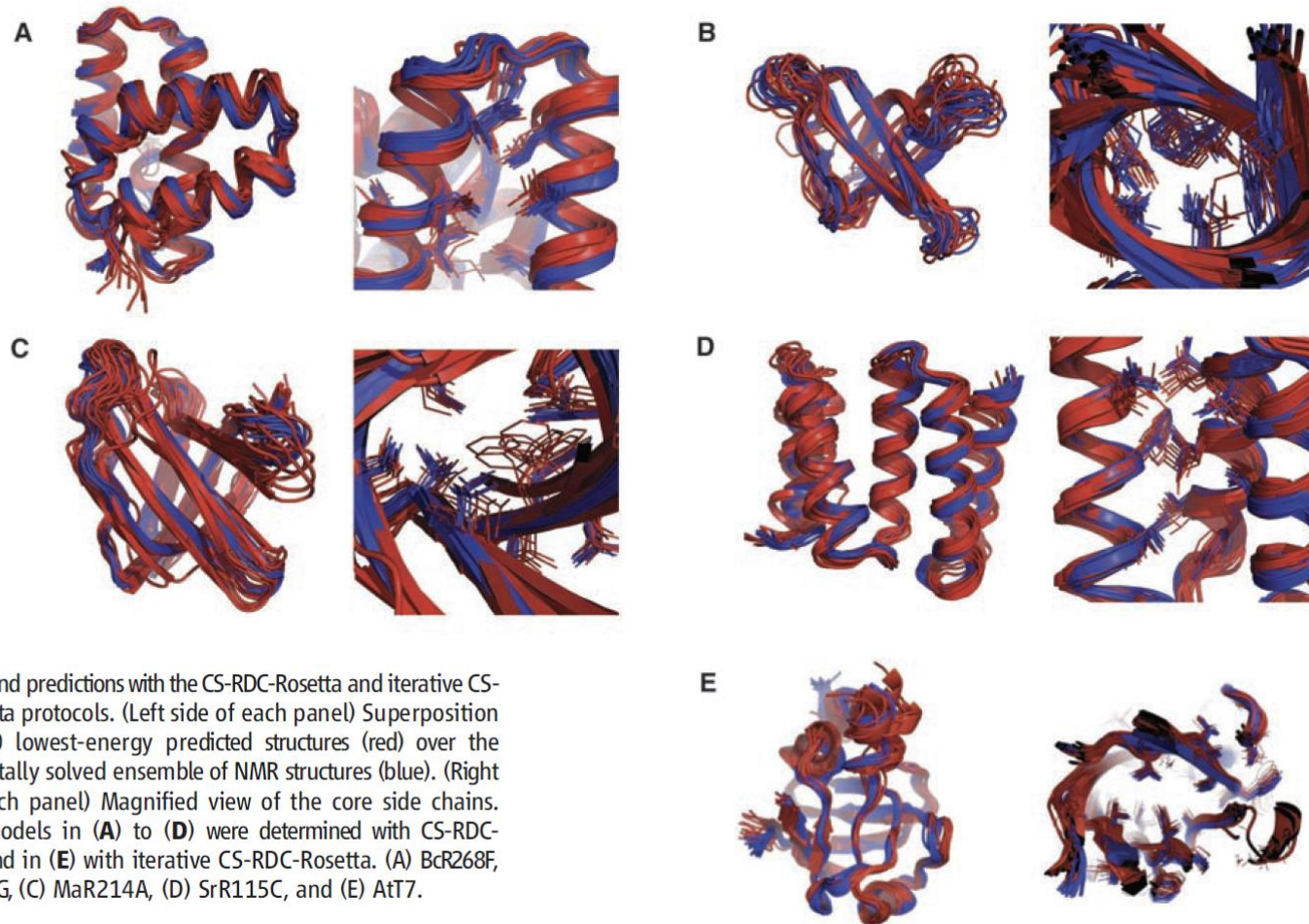


Fig. 1. Impact of RDC data on conformational search. Lines depict RMSD histograms of structures selected in the lowest 10th percentile of coarse-grained energy for ensembles generated with the use of CS-Rosetta (black) or CS-RDC-Rosetta (red). (A) BcR103A, (B) DvR115G, (C) RrR43, and (D) SrR115C. Ang., angstrom.

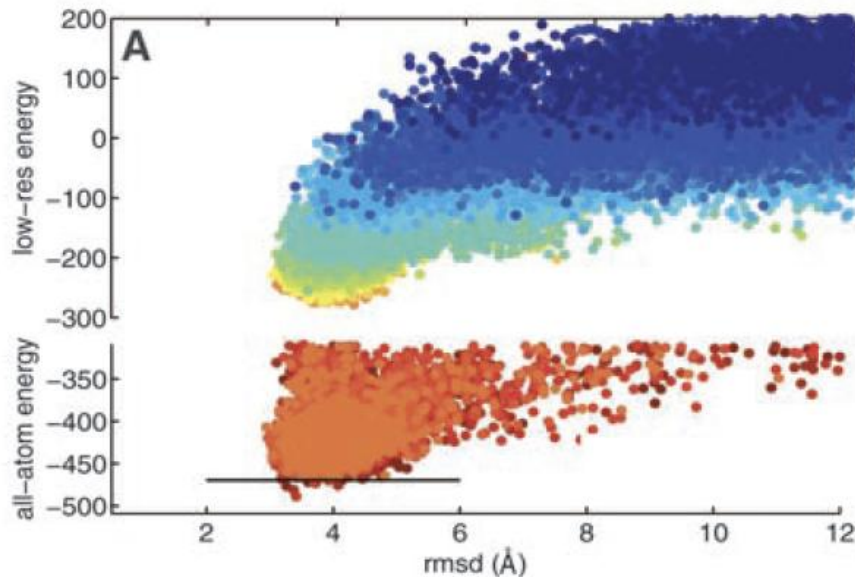


Raman, S.; Lange, O. F.; Rossi, P.; Tyka, M.; Wang, X.; Aramini, J.; Liu, G.; Ramelot, T. A.; Eletsky, A.; Szyperski, T.; Kennedy, M. A.; Prestegard, J.; Montelione, G. T.; Baker, D., NMR structure determination for larger proteins using backbone-only data. *Science* **2010**, *327* (5968), 1014-8.

Blind Structure Predictions with CS-RDC-Rosetta.



NMR Structure Determination for Proteins Using Backbone-Only Data



S. Raman, O. F. Lange, P. Rossi, M. Tyka, X. Wang, J. Aramini, G. Liu, T. A. Ramelot, A. Eletsky, T. Szyperski, M. A. Kennedy, J. Prestegard, G. T. Montelione and D. Baker; "NMR structure determination for larger proteins using backbone-only data"; *Science*; **2010**; Vol. 327 (5968): p. 1014-8.

Size Limit for *de novo* Protein Structure Prediction



■ Topology correct

■ Atomic detail



Rosetta *de novo*
(Bradley, 2005. Science)



Rosetta NMR with unassigned data
(Meiler, 2003, PNAS)



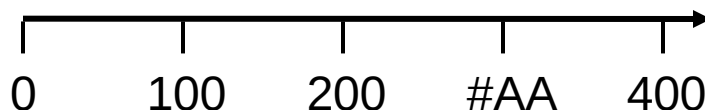
Rosetta NMR with sparse NOE/CS data
(Bowers, 2000, J. Bio. NMR)



Rosetta NMR with sparse RDC/CS data
(Rohl, 2002, JACS)



Rosetta NMR with CS data
(Shen, 2008, PNAS)



Size Limit for *de novo* Protein Structure Prediction



■ Topology correct

■ Atomic detail



Rosetta *de novo*
(Bradley, 2005, Science)



Rosetta NMR with unassigned data
(Meiler, 2003, PNAS)



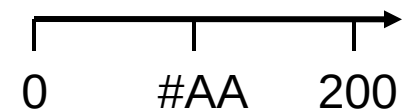
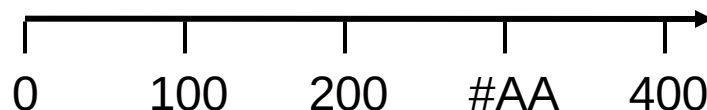
Rosetta NMR with sparse NOE/CS data
(Bowers, 2000, J. Bio. NMR)



Rosetta NMR with sparse RDC/CS data
(Rohl, 2002, JACS)



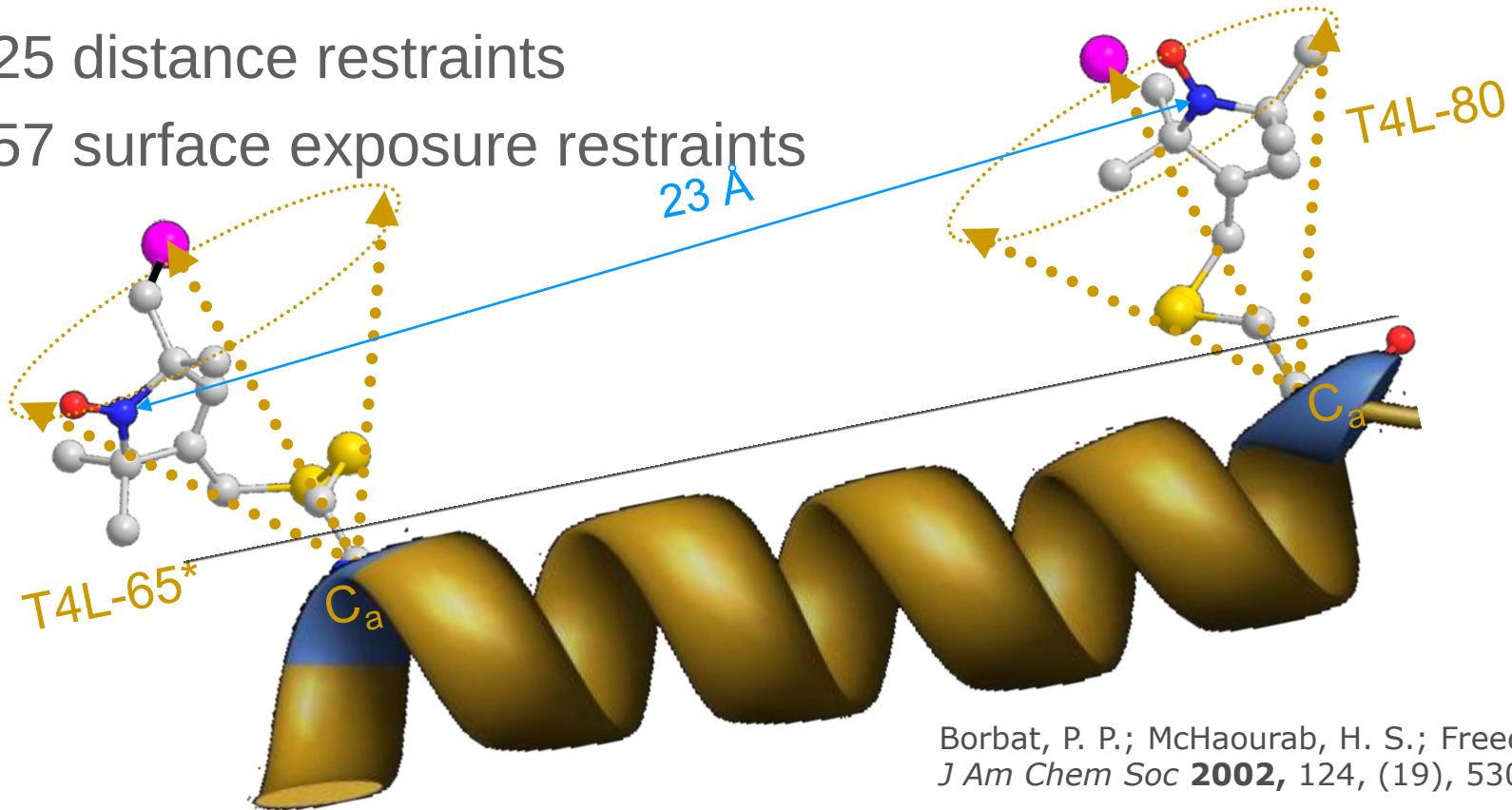
Rosetta NMR with backbone data
(Shen, 2008, PNAS; Raman, 2010, Science)



Structure Elucidation of Lysozyme from 25 Experimental EPR Distances



- 25 distance restraints
- 57 surface exposure restraints



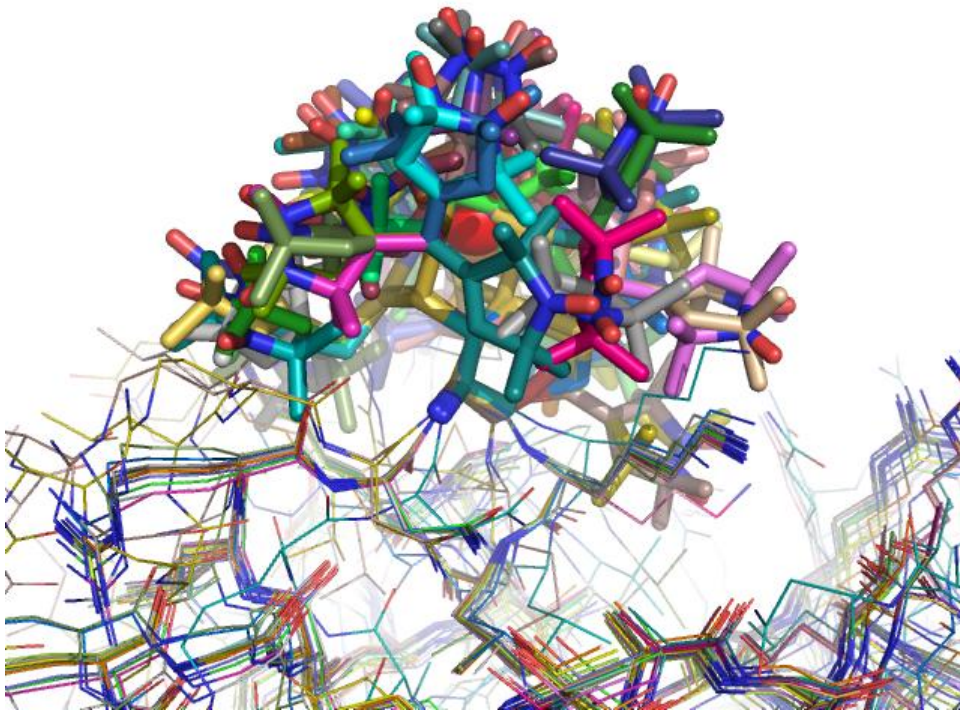
Borbat, P. P.; McHaourab, H. S.; Freed, J. H.,
J Am Chem Soc **2002**, 124, (19), 5304-14.

- Thanks to Hassane Mchaourab and his lab for experimental data

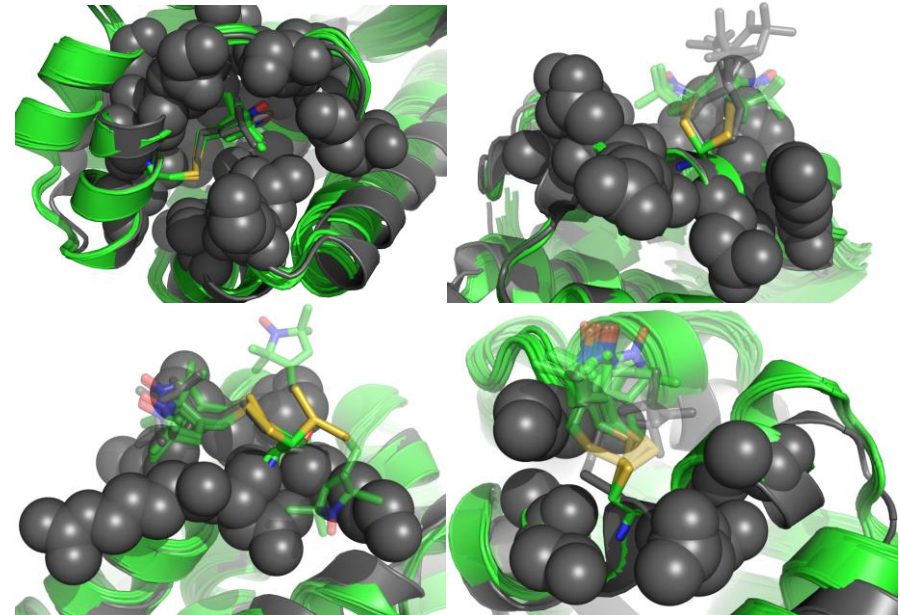
Rotamer Representation of MTS Spin Label in Rosetta



■ Rotamer Sampling



■ Rotamer Recovery

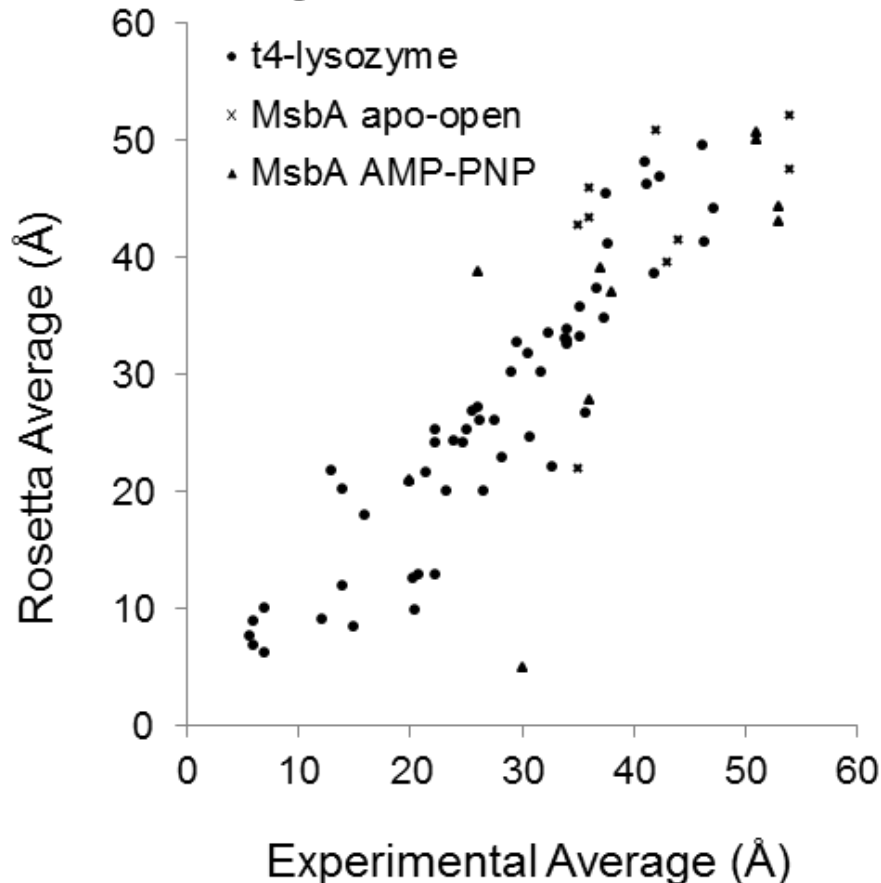


Nathan S. Alexander, Richard A. Stein, Hanane A. Koteiche, Kristian Kaufmann, Hassane Mchaourab, Jens Meiler
RosettaEPR: Rotamer library for spin label structure and dynamics, submitted to PLOS ONE

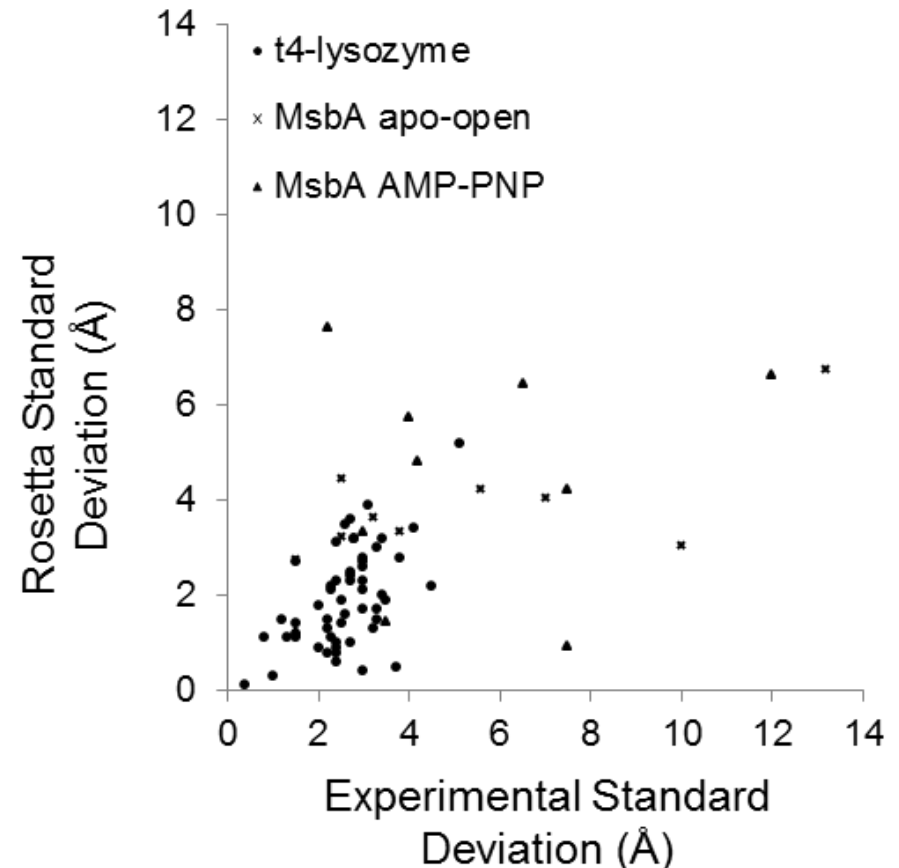
Rotamer Representation of MTS Spin Label in Rosetta



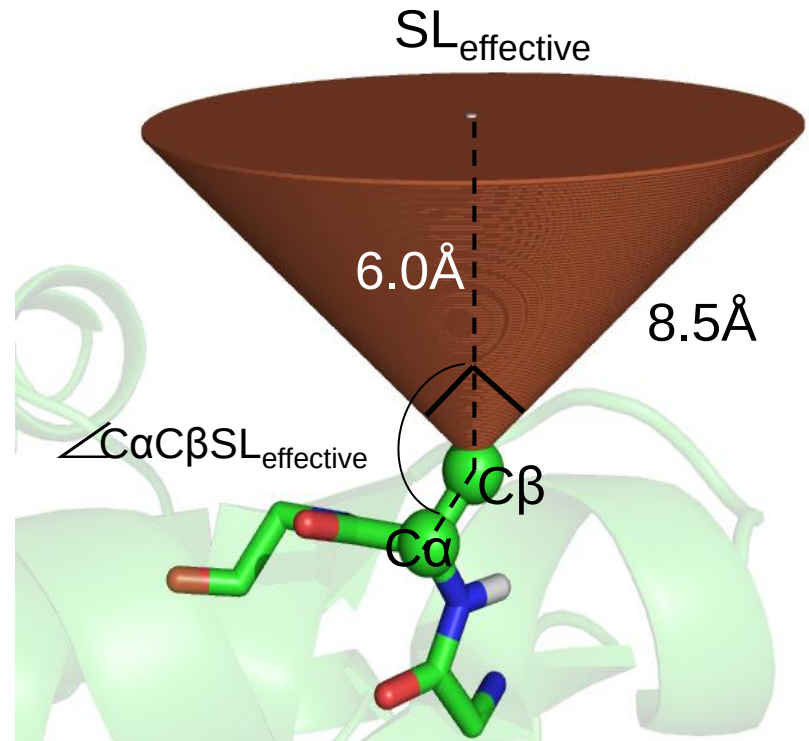
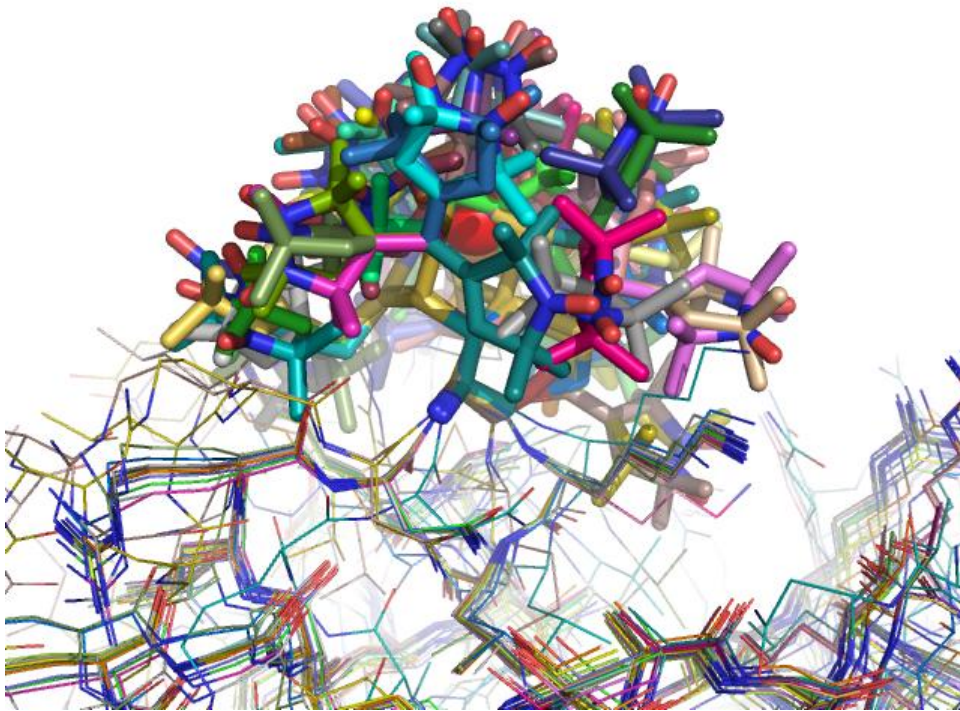
■ Average Distance



■ Width of Distribution



Conformations of the spin label are modeled as a conical distribution



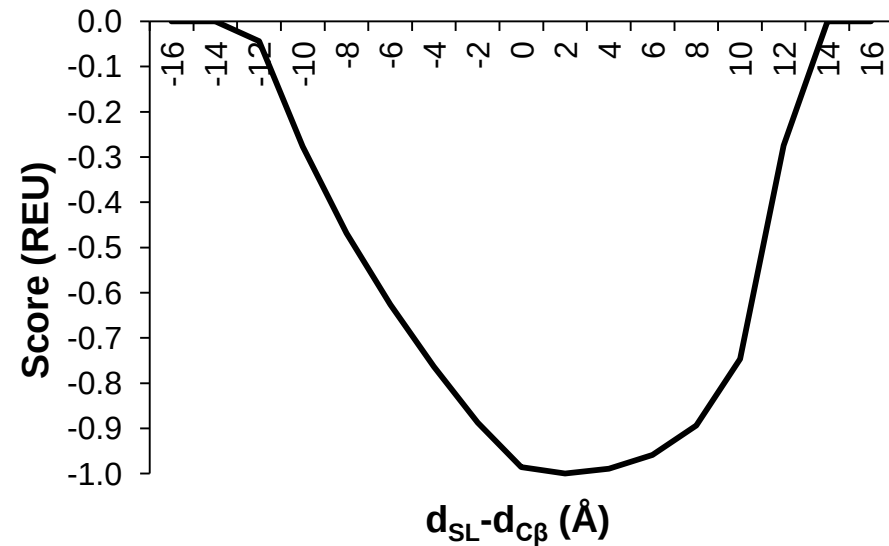
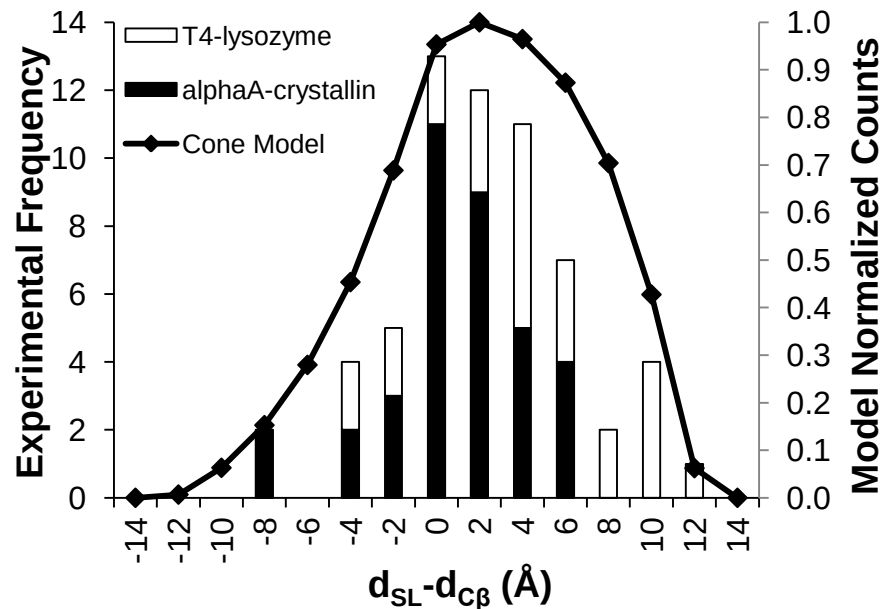
Alexander, N.; Al-Mestarihi, A.; Bortolus, M.; McHaourab, H.; Meiler, J. "De Novo High-Resolution Protein Structure Determination from Sparse Spin-Labeling EPR Data" *Structure* **2008**, *16*, **181-95**.

S. J. Hirst, N. Alexander, H. S. McHaourab and J. Meiler; "RosettaEPR: an integrated tool for protein structure determination from sparse EPR data"; *J Struct Biol*; **2011**; Vol. 173 (3): p. 506-14.

Statistics of $D_{SL} - D_{C\beta}$ calculated from cone model match experiment



Allows the creation of a scoring function indicating how well a protein model agrees with EPR distance data

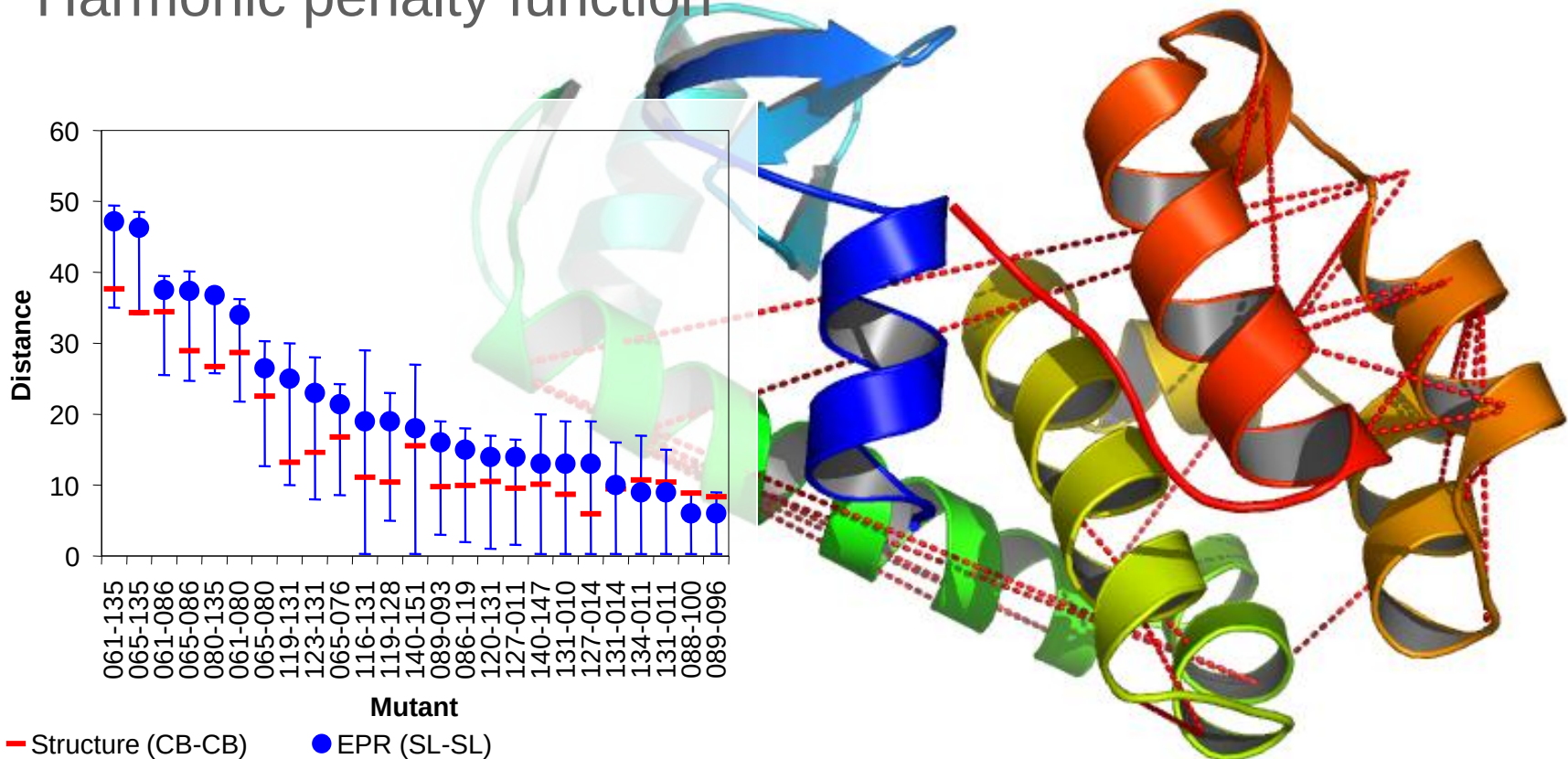


Alexander, N.; Al-Mestarihi, A.; Bortolus, M.; McHaourab, H.; Meiler, J. "De Novo High-Resolution Protein Structure Determination from Sparse Spin-Labeling EPR Data" *Structure* **2008**, *16*, 181-95.

S. J. Hirst, N. Alexander, H. S. McHaourab and J. Meiler; "RosettaEPR: an integrated tool for protein structure determination from sparse EPR data"; *J Struct Biol*; **2011**; Vol. 173 (3): p. 506-14.

25 Experimental Distance Restraints

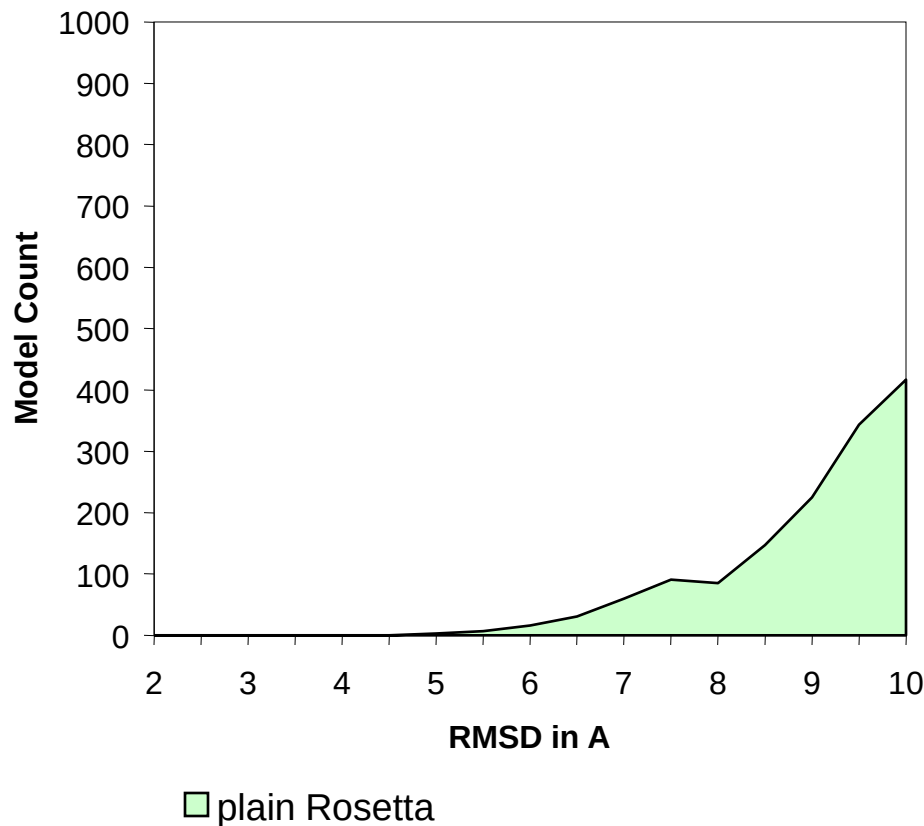
- Restraint: $(d_{\text{SL-SL}} - \sigma_{\text{SL-SL}} - 10\text{\AA}) \leq d_{\text{CB-CB}} \leq (d_{\text{SL-SL}} + \sigma_{\text{SL-SL}})$
- Harmonic penalty function



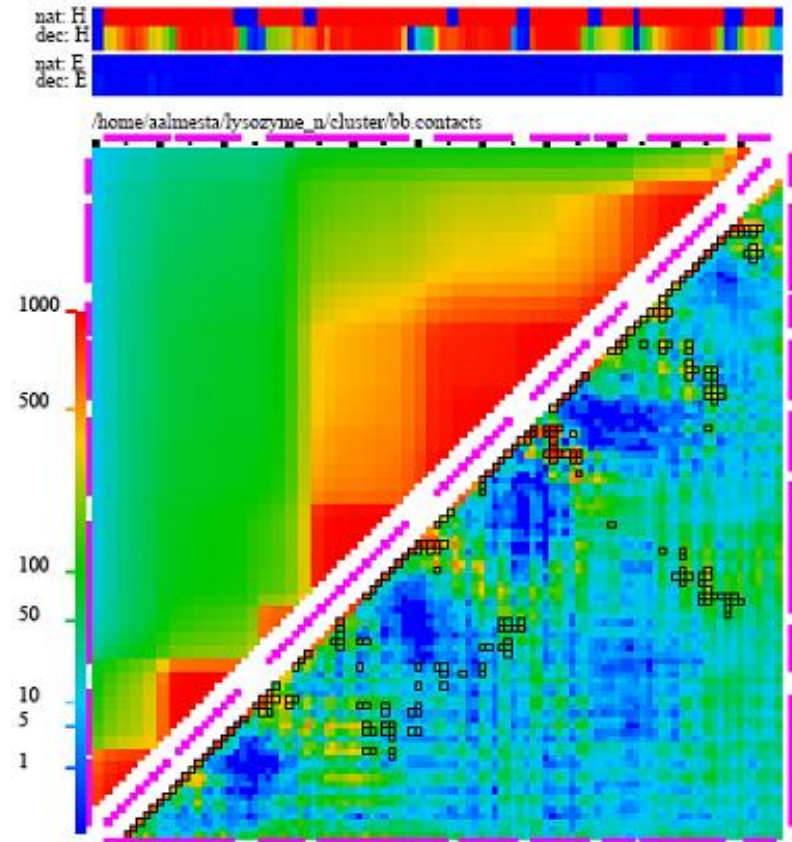
Influence of Experimental Data on Sampling and Model Quality



- RMSD histogram



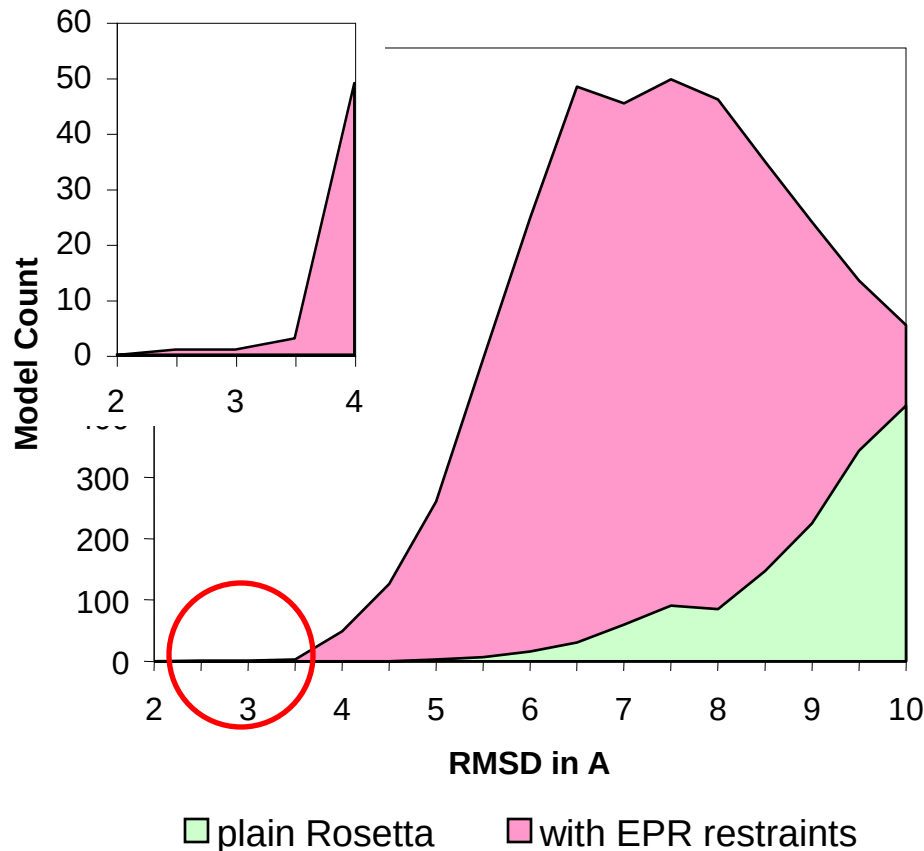
- C



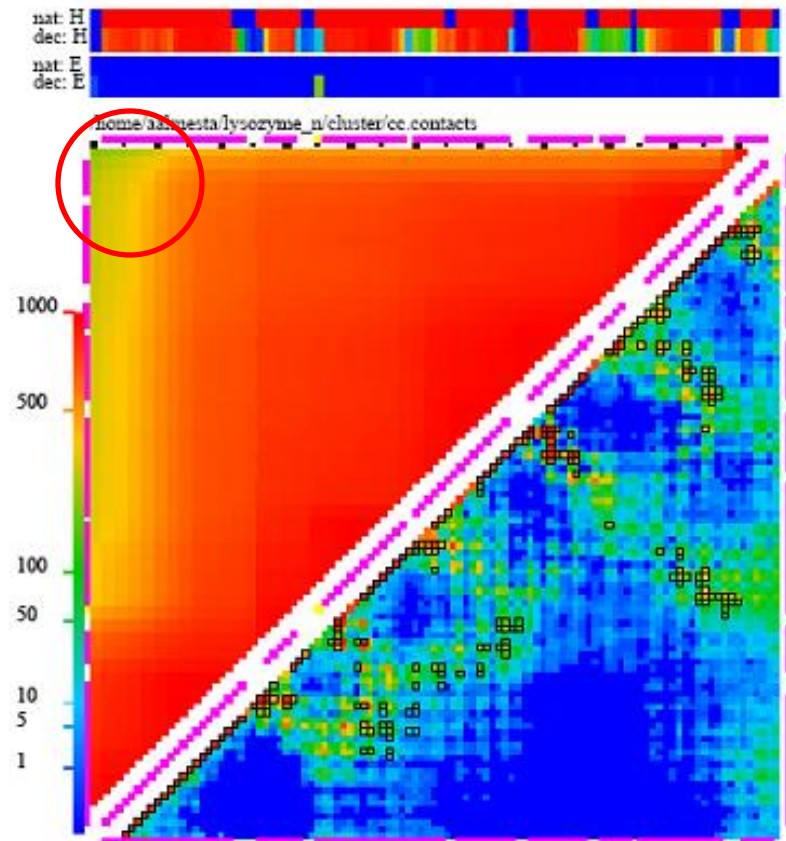
Influence of Experimental Data on Sampling and Model Quality



■ RMSD histogram



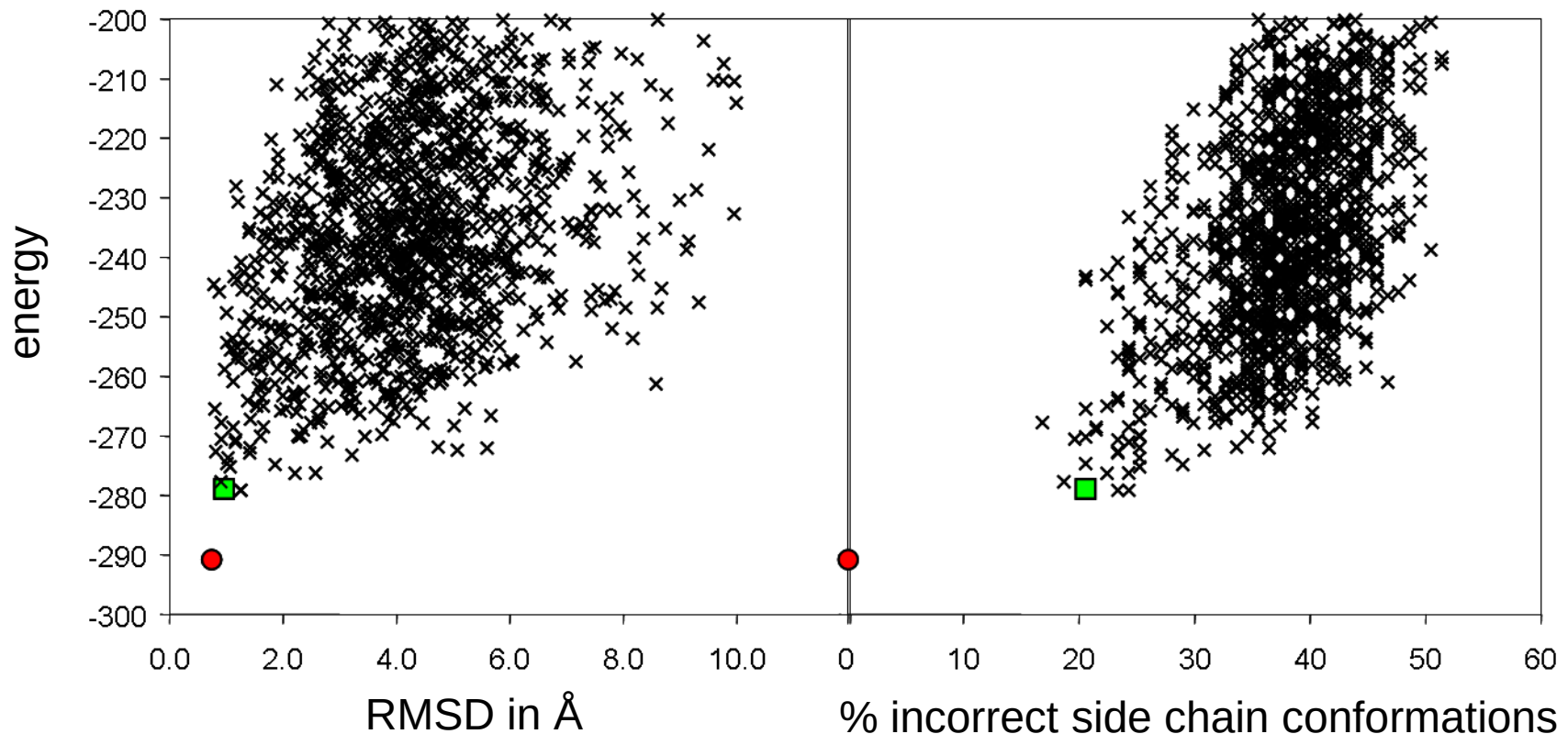
■ C



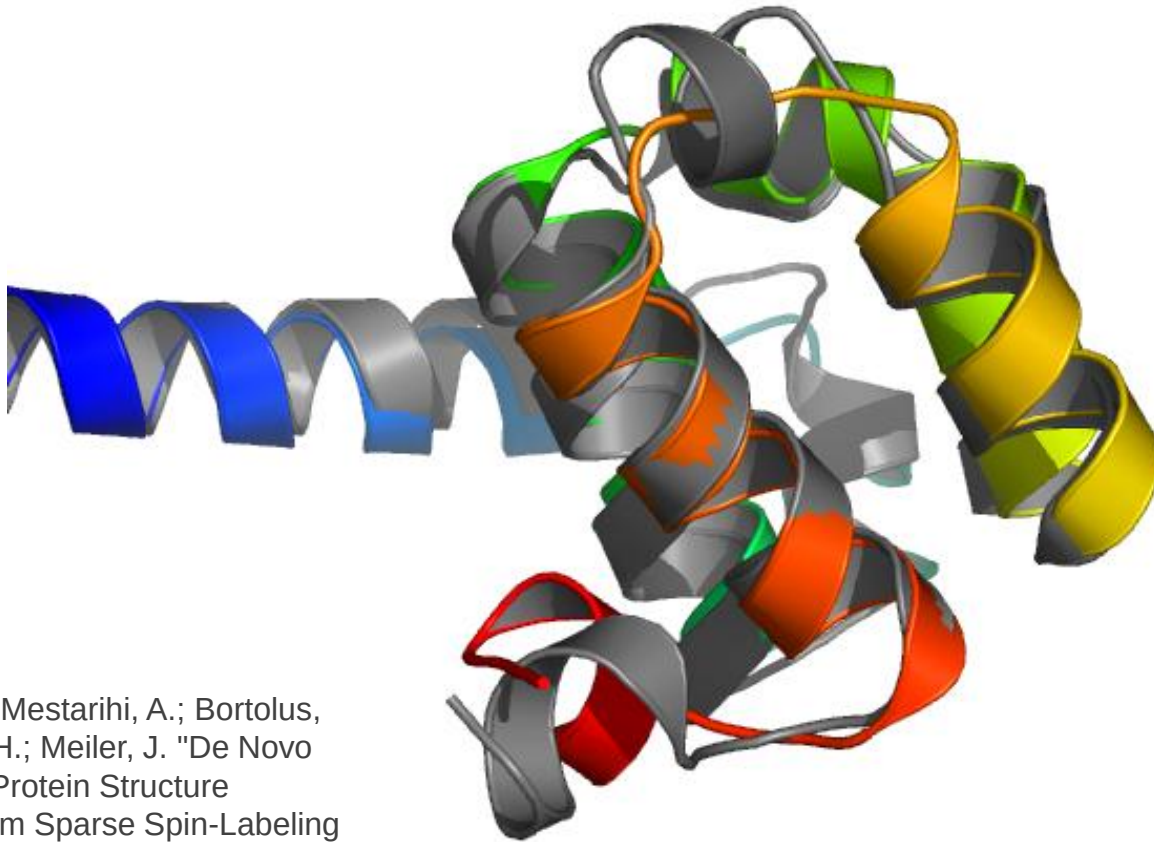
High Resolution Energy Refinement



- Lowest scoring ~11,000 models out of 500,000 were refined

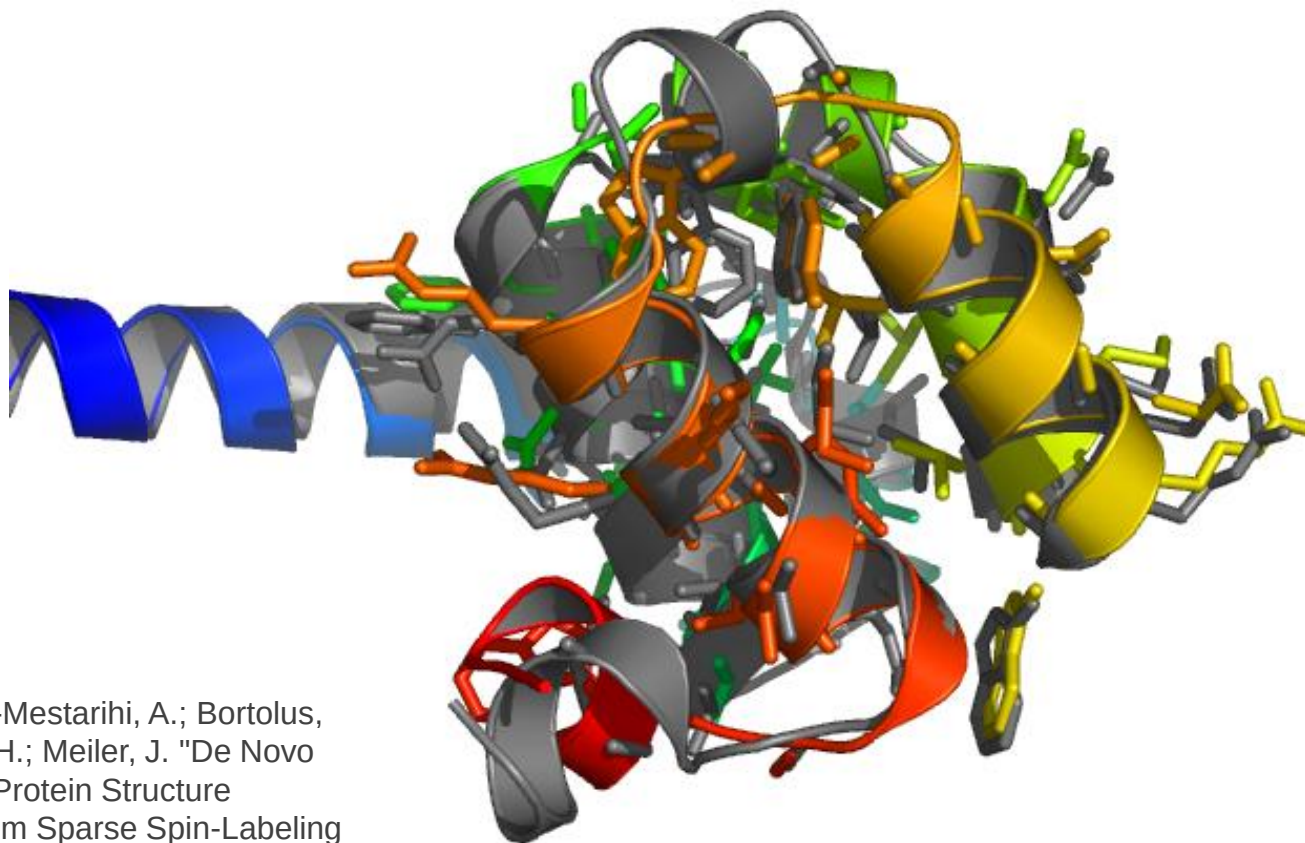


Backbone RMSD is 0.96Å



Alexander, N.; Al-Mestarihi, A.; Bortolus, M.; McHaourab, H.; Meiler, J. "De Novo High-Resolution Protein Structure Determination from Sparse Spin-Labeling EPR Data" *Structure* **2008**, *16*, 181-95.

All but Two Core Amino Acids have Correct Side Chain Conformation



Alexander, N.; Al-Mestarihi, A.; Bortolus, M.; McHaourab, H.; Meiler, J. "De Novo High-Resolution Protein Structure Determination from Sparse Spin-Labeling EPR Data" *Structure* **2008**, *16*, 181-95.