Combining Experimental Data with Rosetta Computation Models

Rocco Moretti

(rocco.moretti@vanderbilt.edu)

Laboratory of Jens Meiler

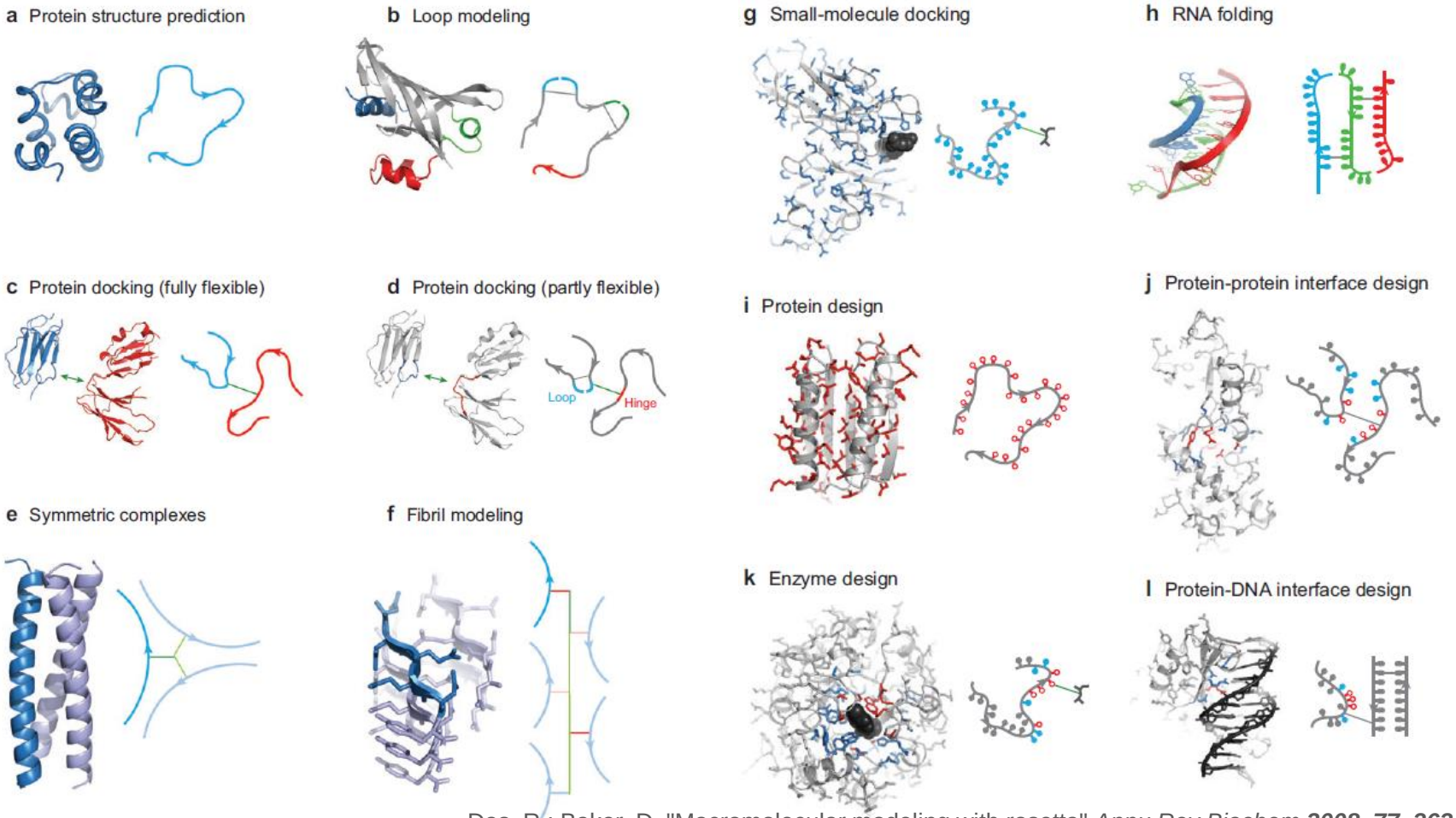
VANDERBILT UNIVERSITY

Center for Structural Biology
and the Department of Chemistry

Outline

- Part 1: Introduction, Crosslinking & EPR
  - Overview of Rosetta
  - Incorporating Experimental Data into Rosetta
  - Restraints (Rosetta “Constraints”)
  - Constraints and crosslinking
  - RosettaEPR
- Part 2: NMR (CSRosetta) – Nik Sgourakis
- Part 3: Electron Density – Frank Dimaio & Ray Wang

Rosetta: A Unified Framework for Tackling Molecular Modeling



Das, R.; Baker, D. "Macromolecular modeling with rosetta" *Annu Rev Biochem* ***2008, 77, 363-82.***

www.rosettacommons.org

- RosettaCommons
- Central organization for development of Rosetta
- Over 150 developers from over 23 groups in 6 countries

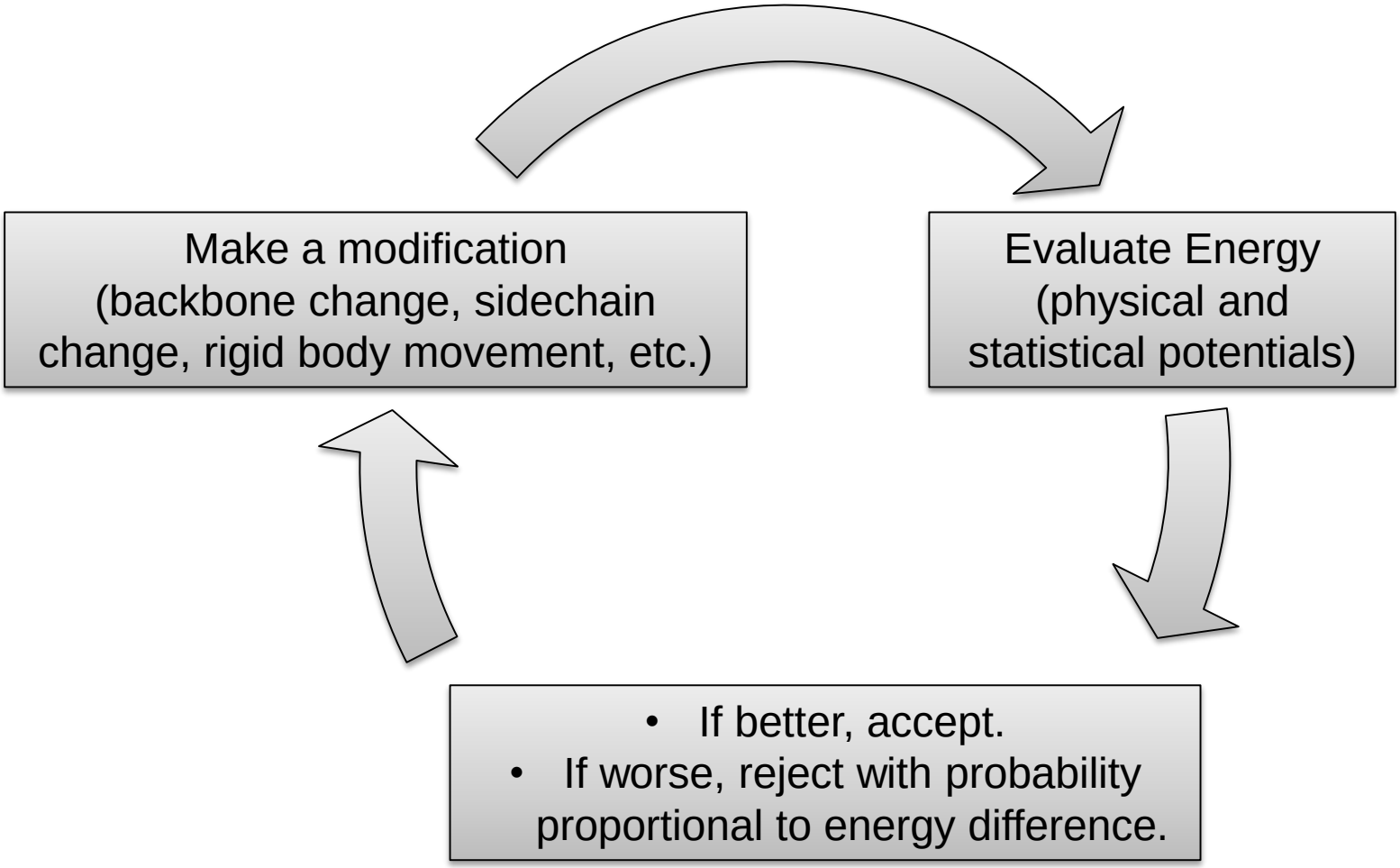
- Regular updates and bug fixes
  - New version (almost) every week.
- Freely available to academic users

- Multiple interfaces
  - Command line Rosetta
  - PyRosetta
  - FoldIt
  - Rosetta@Home

RosettaCon 2013, Leavenworth, WA, USA

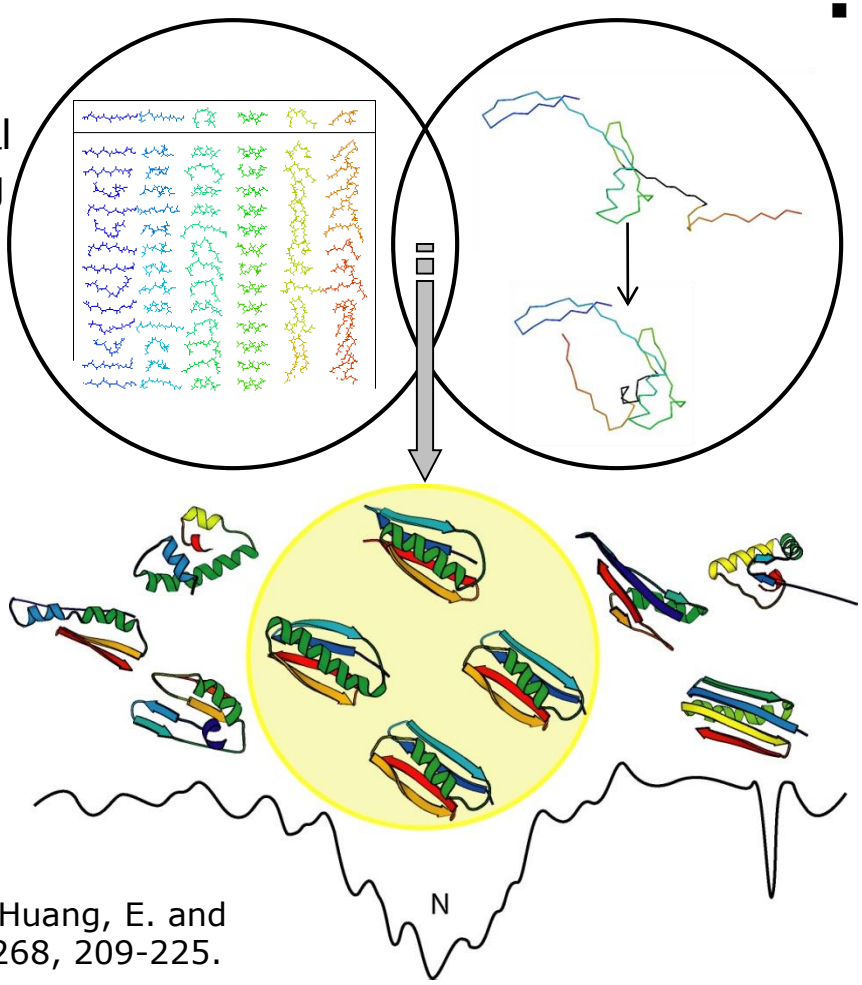
Kaufmann, K. W.; et al. "Practically Useful: What the Rosetta Protein Modeling Suite Can Do for You" *Biochemistry* ***2010.***

Rosetta uses Metropolis–Hastings Monte Carlo



Sampling and Scoring for Protein Folding Simulation

- Local Sequence Bias
  - Approximate local interactions using the distribution of conformations seen for similar sequences in known protein structures
- ***Monte Carlo energy minimization***
  Select broadest minima using cluster analysis



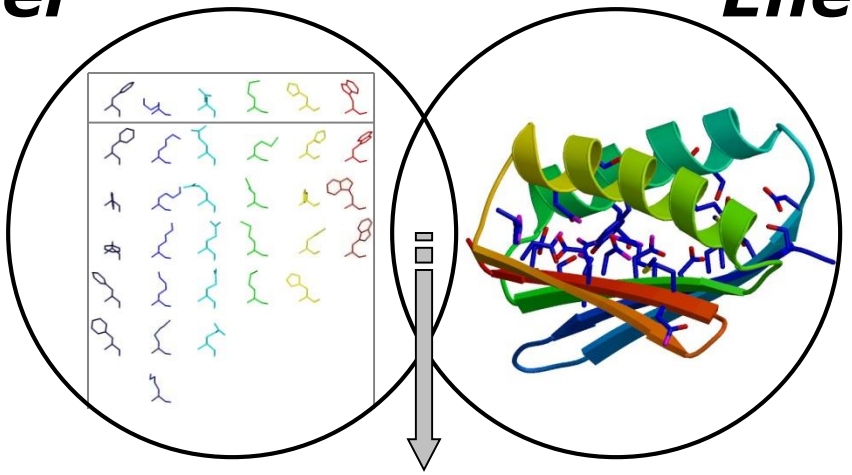
- Energy evaluation of non-local interactions using knowledge-based energy function
  - Steric overlap
  - Residue environment
  - Pair wise interactions
  - Strand pairing
  - Compactness
  - Secondary Structure Packing

Simons, K. T., Kooperberg, C., Huang, E. and Baker, D. (1997) *J. Mol. Biol.,* 268, 209-225.

Sampling and Scoring for Side Chain Repacking and Design

Local Rotamer Bias

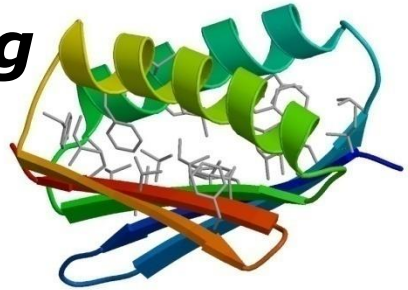
Approximate interactions within sidechain using the distribution of sidechain confor-mations (rotamers) seen in known protein structures



Energy function

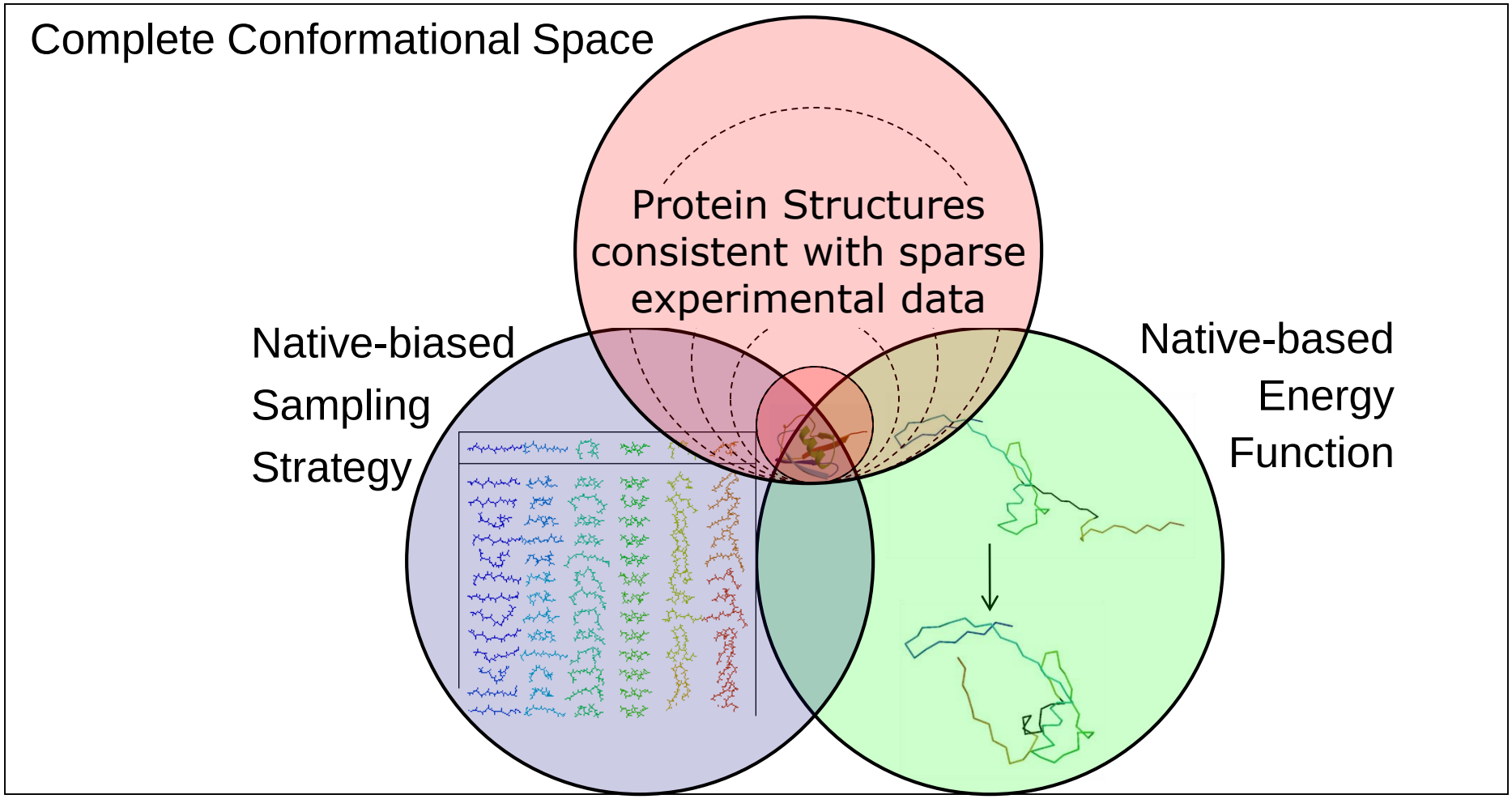
Statistically derived potential function
- ***VDW interaction***
- ***solvation***
- ***hydrogen bonding potential***
- ***pair wise interactions***
- ***rotamer probability***

Simulated Annealing Monte Carlo energy minimization



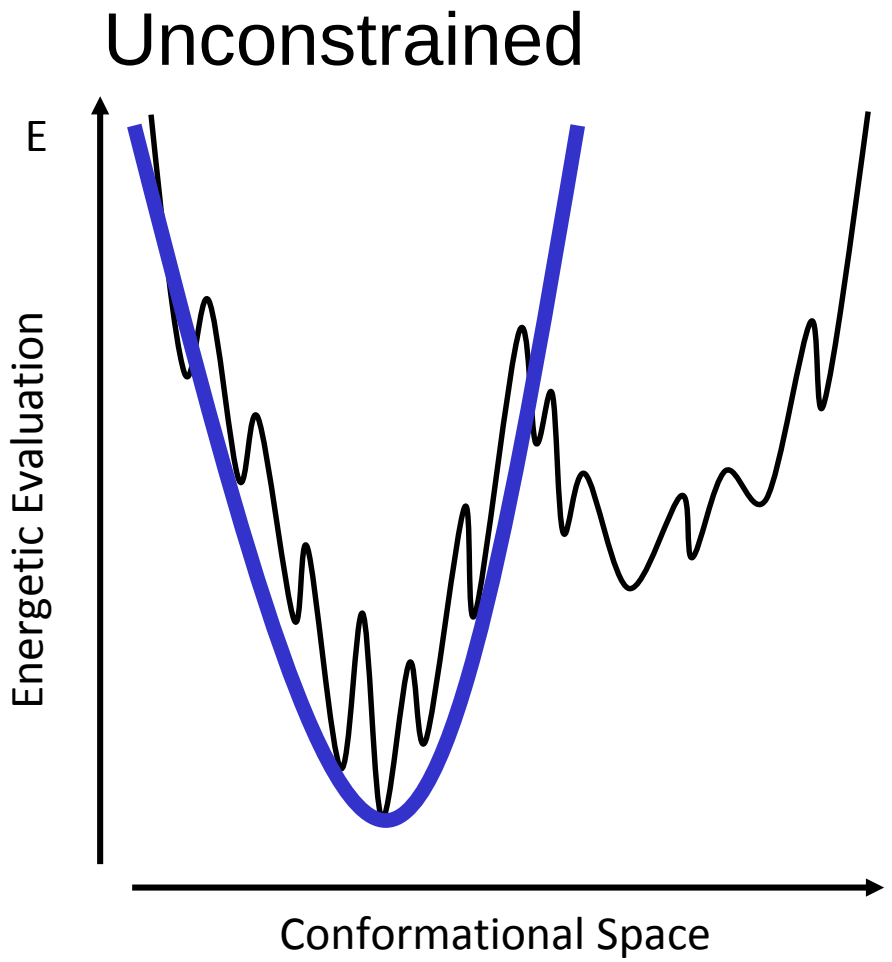
Dahiyat, B. I. and Mayo, S. L. (1997) *Science,* 278, 82-7
Dunbrack, R. L., Jr. and Karplus, M. (1993) *J Mol Biol,* 230, 543-74.
Kuhlman, B., et. al. (2003) *Science,* 302, 1364-1368.

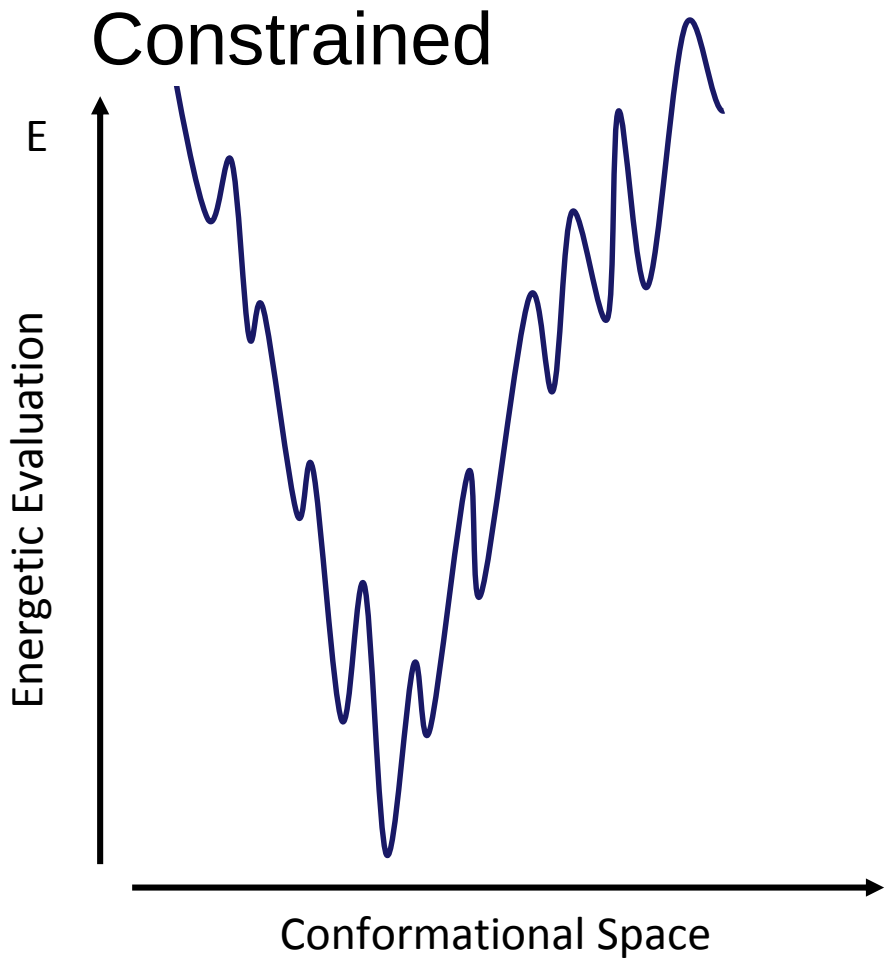
Combining Strengths: Building Accurate Models from Limited Data



“Constraints” (Restraints) in Rosetta

“Constraints” alter the energy function





Separation of Measure and Scoring

```
AtomPair NE2 13 V3 32 HARMONIC 0.0 0.2
Angle CD2 13 NE2 13 ZN 32 HARMONIC 2.09 0.35
Dihedral CG 13 CD2 13 NE2 13 ZN 32 CIRCULARHARMONIC 3.14 0.35
```

- Distance (AtomPair)
- Angle
- Dihedral
- Coordinate
- Site

- AmbiguousConstraint
- KofNConstraint

- Harmonic
- CircularHarmonic
- Flat-bottomed Harmonic
- Gaussian
- Sigmoid
- Spline

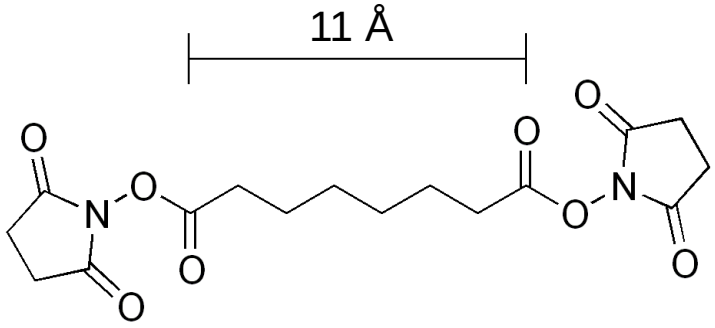
https://www.rosettacommons.org/docs/latest/constraint-file.html

Rosetta with crosslinking data

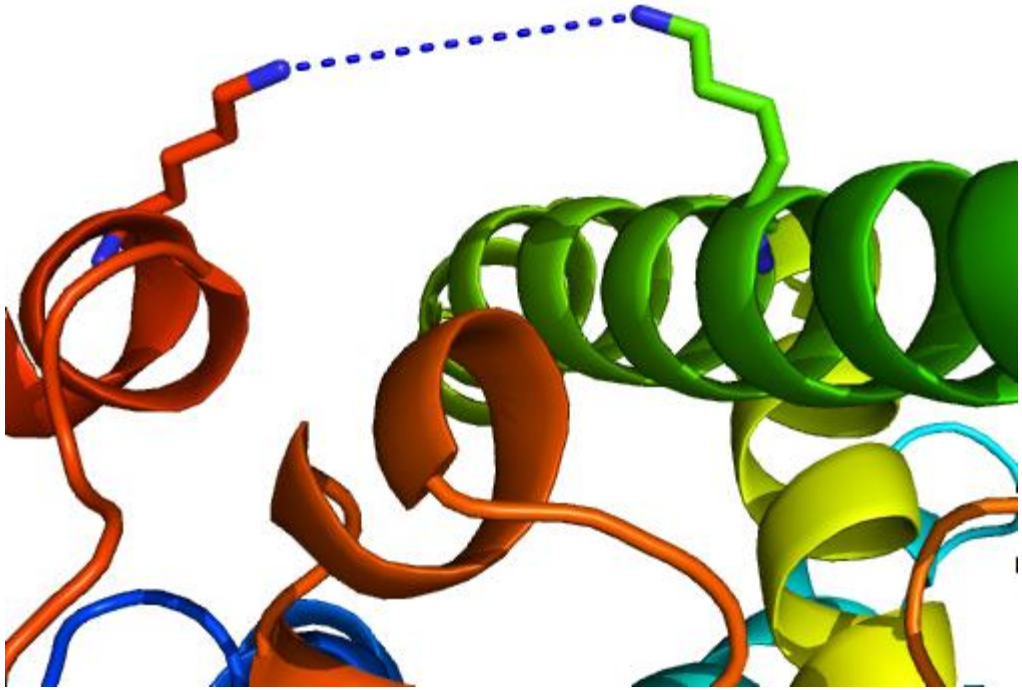
Based on “Cross-Link Guided Molecular Modeling with Rosetta”
Kahraman, Herzog, Leitner, Rosenberger, Aebersold, and Malmstöm
PLOS One, 2013, 8(9) e73411

Rosetta/demos/protocol_capture/2012/XL_driven_protein_docking/

Chemical crosslinking gives distance restraint information

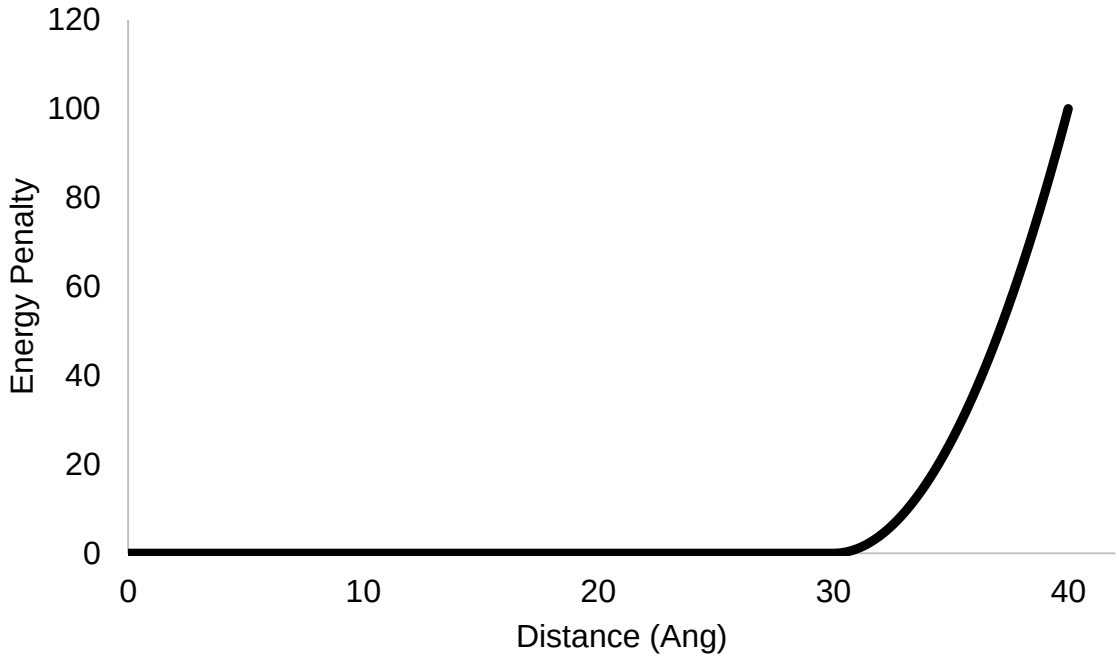


Disuccinimidyl suberate (DSS)



Chemical crosslinking gives distance restraint information

```
AtomPair CB 67 CB 124 FLATHARMONIC 15.0 1.0 15.0
```

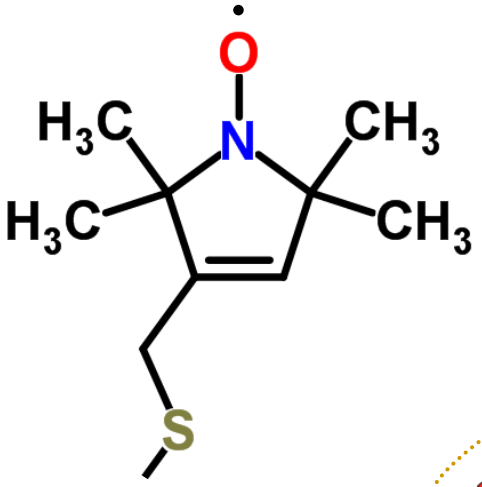


demo

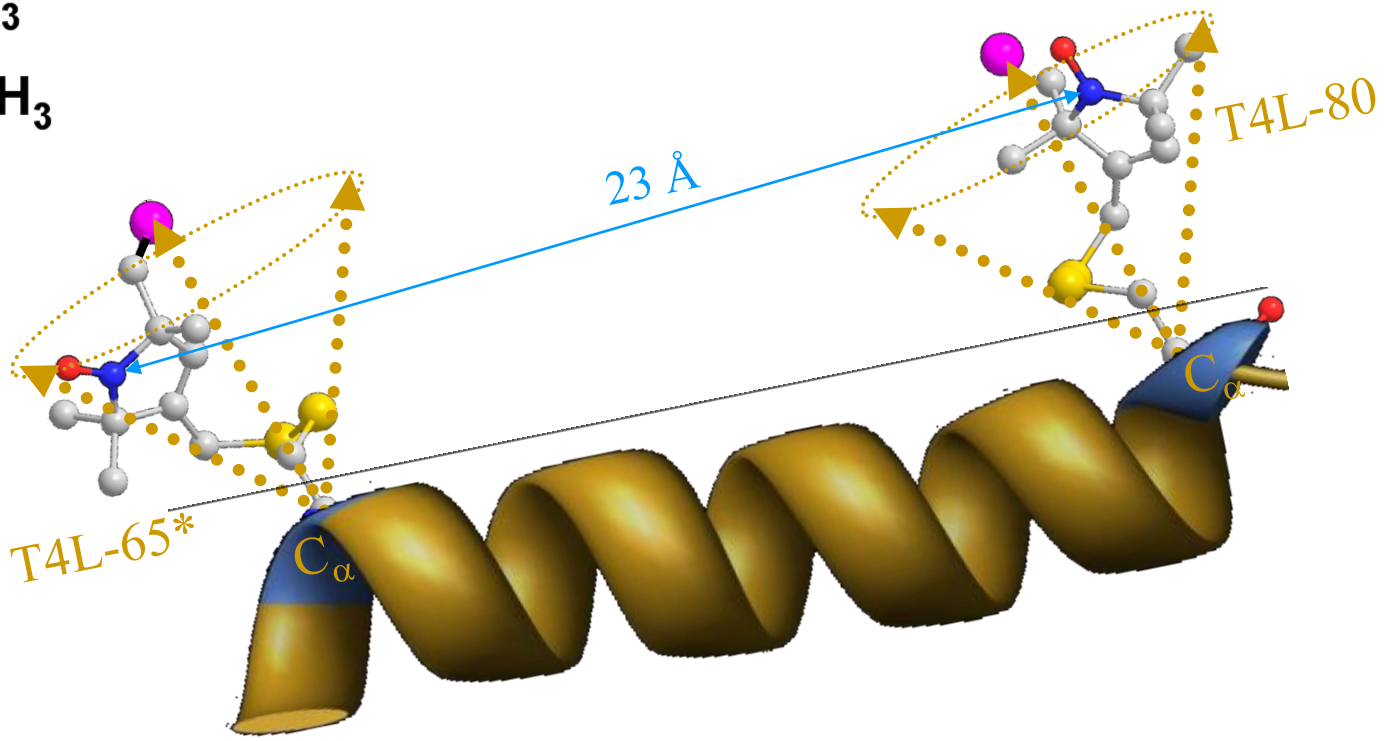
RosettaEPR

Based on “RosettaEPR: An integrated tool for protein structure determination from sparse EPR data.”
Hirst, Alexander, Mchaourab, and Meiler
J. Struct. Biol. 2011, 173; 506-514

Double Electron-Electron Resonance gives distance information

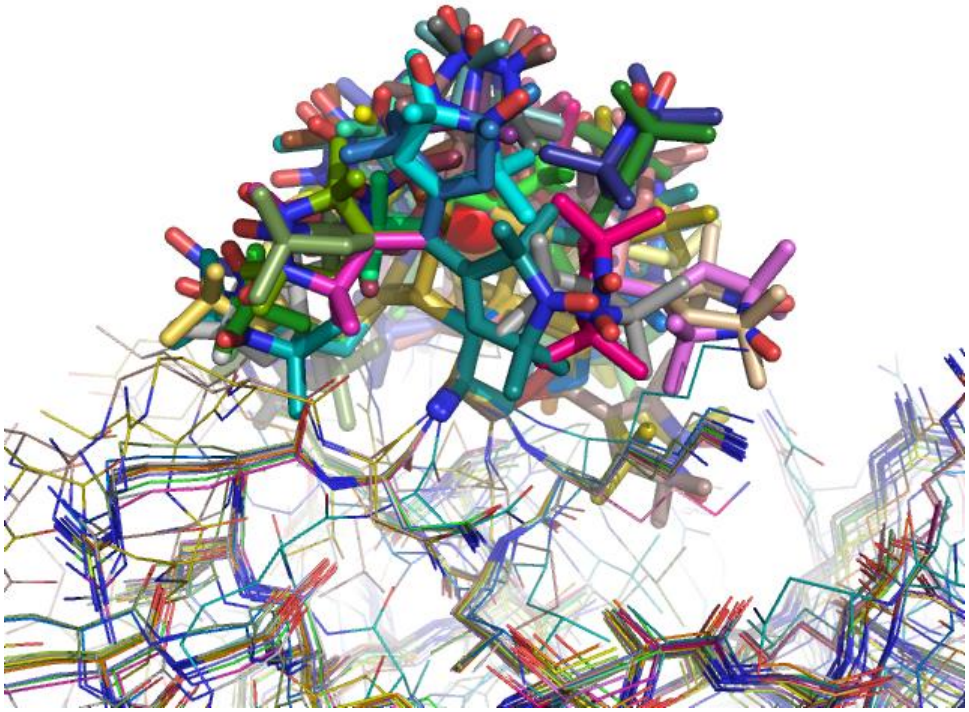


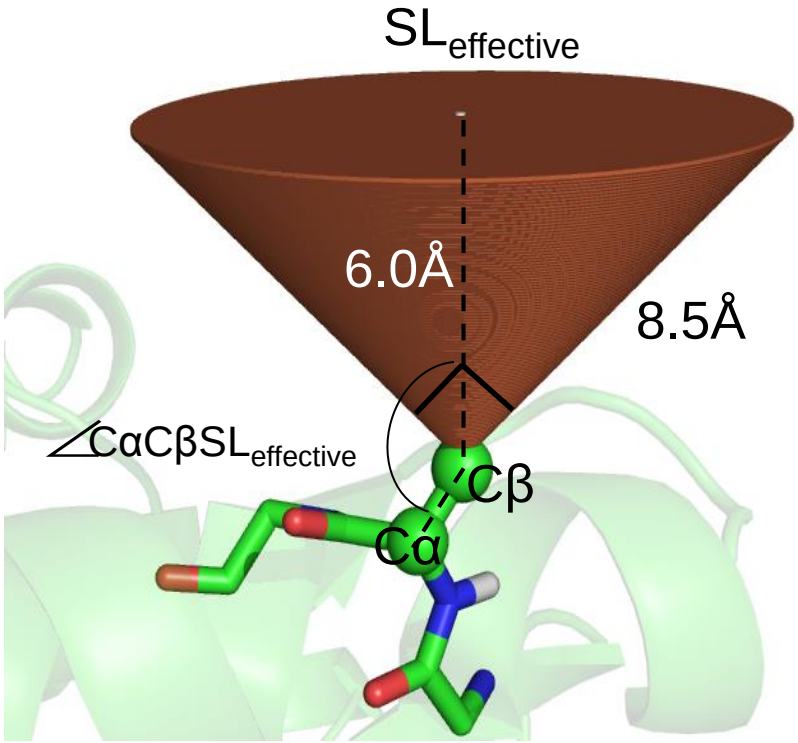
MTSL



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

Conformations of the spin label are modeled as a conical distribution

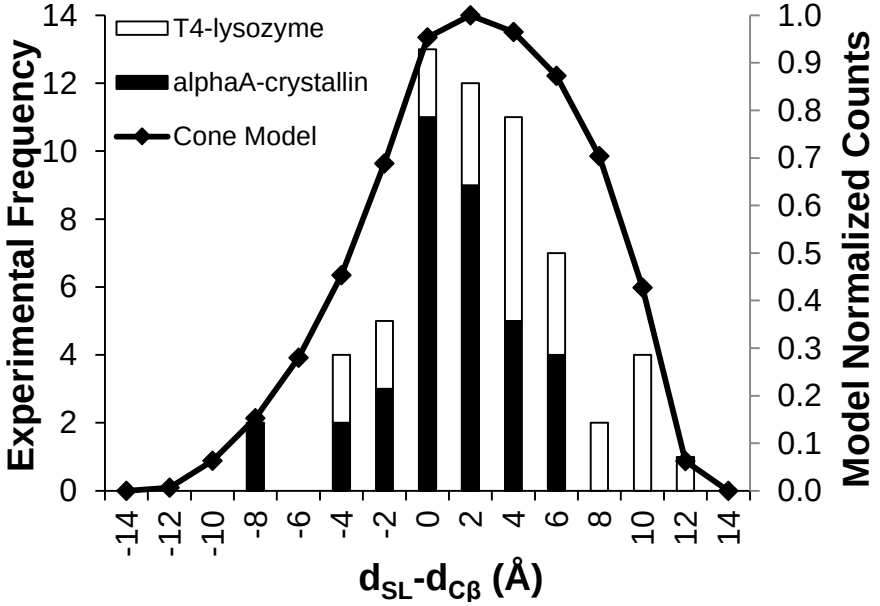


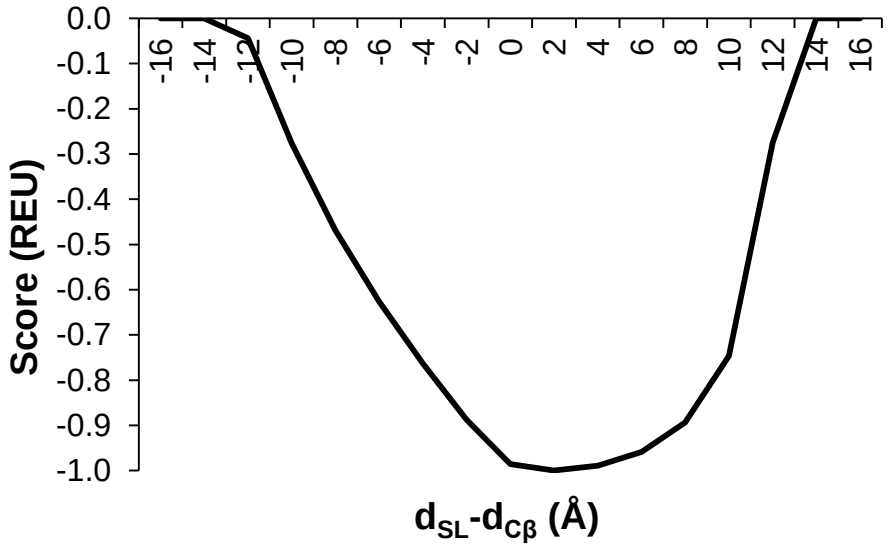


Alexander, N.; Al-Mestarihi, A.; Bortolus, M.; McHaourab, H.; Meiler, J. "De Novo High-Resolution Protein Structure Determination from Sparse Spin-Labeling EPR Data" *Structure* ***2008, 16, 181-95.***
S. J. Hirst, N. Alexander, H. S. McHaourab and J. Meiler; "RosettaEPR: an integrated tool for protein 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



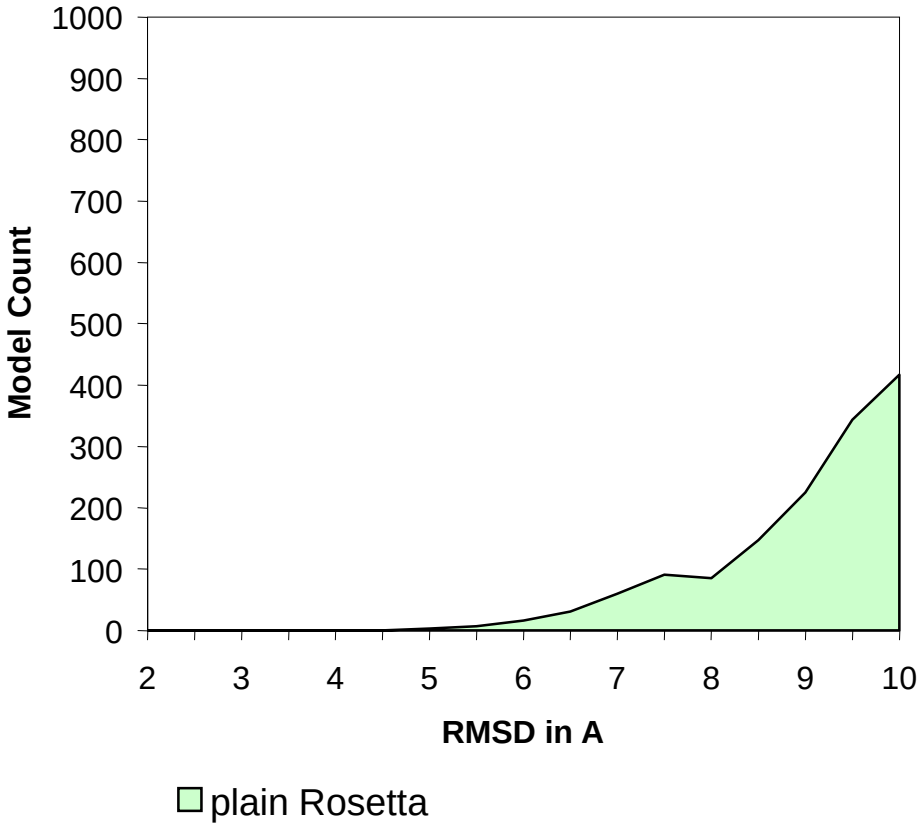


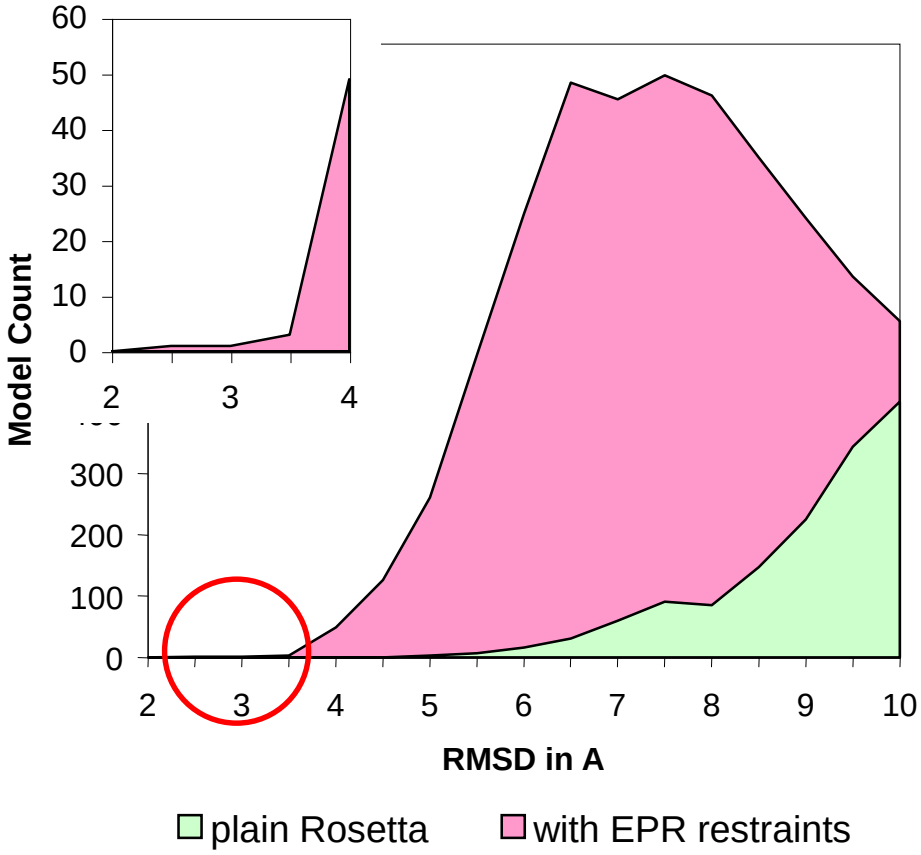
Alexander, N.; Al-Mestarihi, A.; Bortolus, M.; McHaourab, H.; Meiler, J. "De Novo High-Resolution Protein Structure Determination from Sparse Spin-Labeling EPR Data" *Structure* ***2008, 16, 181-95.***

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.

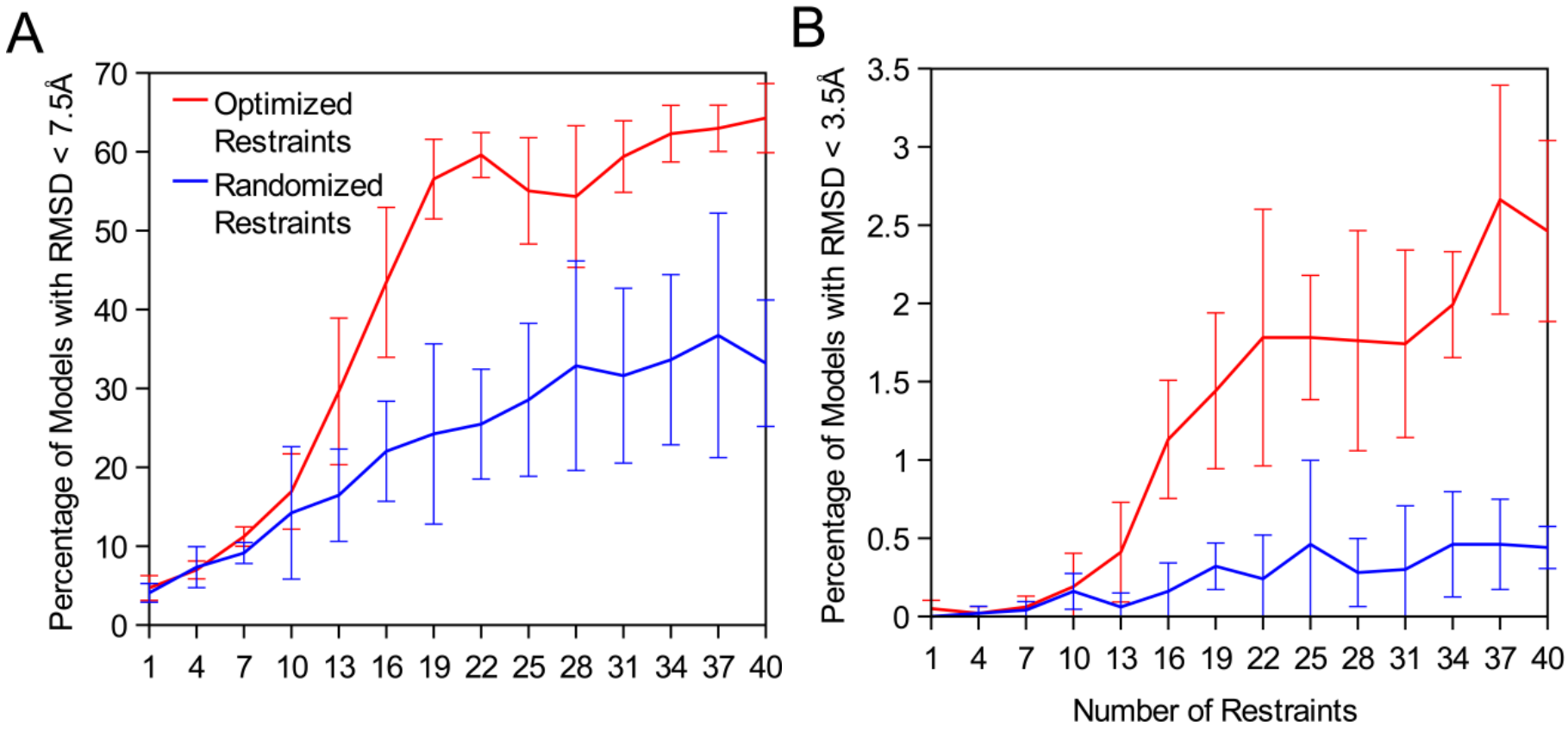
Influence of Experimental Data on Sampling and Model Quality

- RMSD histogram





Improvement in model quality requires a limited number of distance restraints



demo