Combining Experimental Data with Rosetta Computation Models

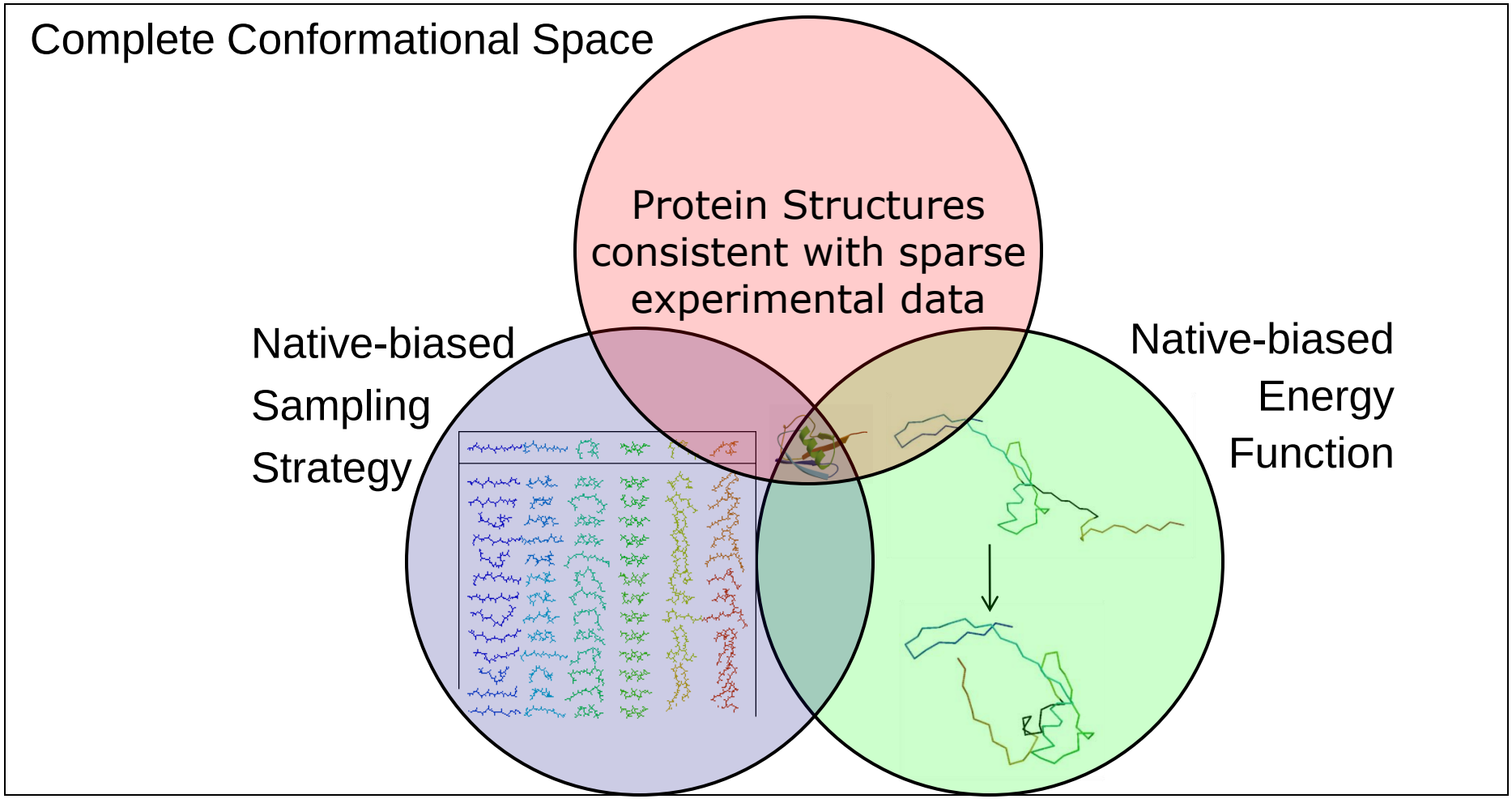
Georg Kuenze
(georg.kuenze@vanderbilt.edu)

Rocco Moretti
(rocco.moretti@vanderbilt.edu)

VANDERBILT UNIVERSITY

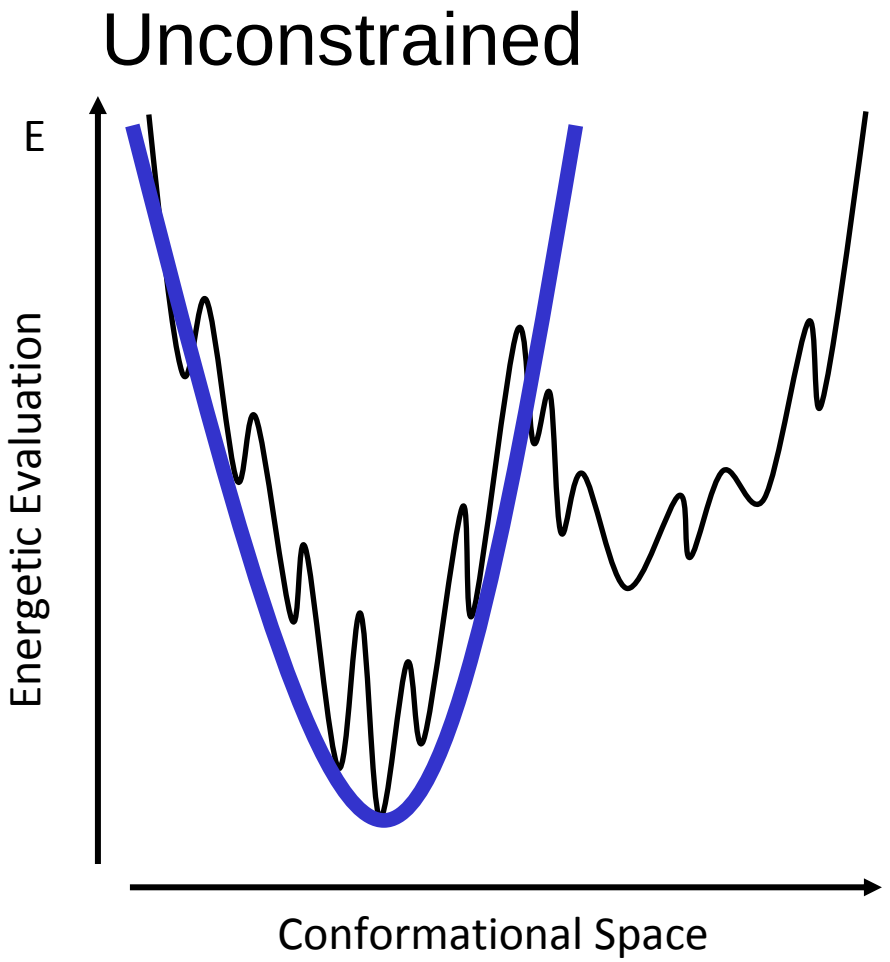
Center for Structural Biology
and the Department of Chemistry

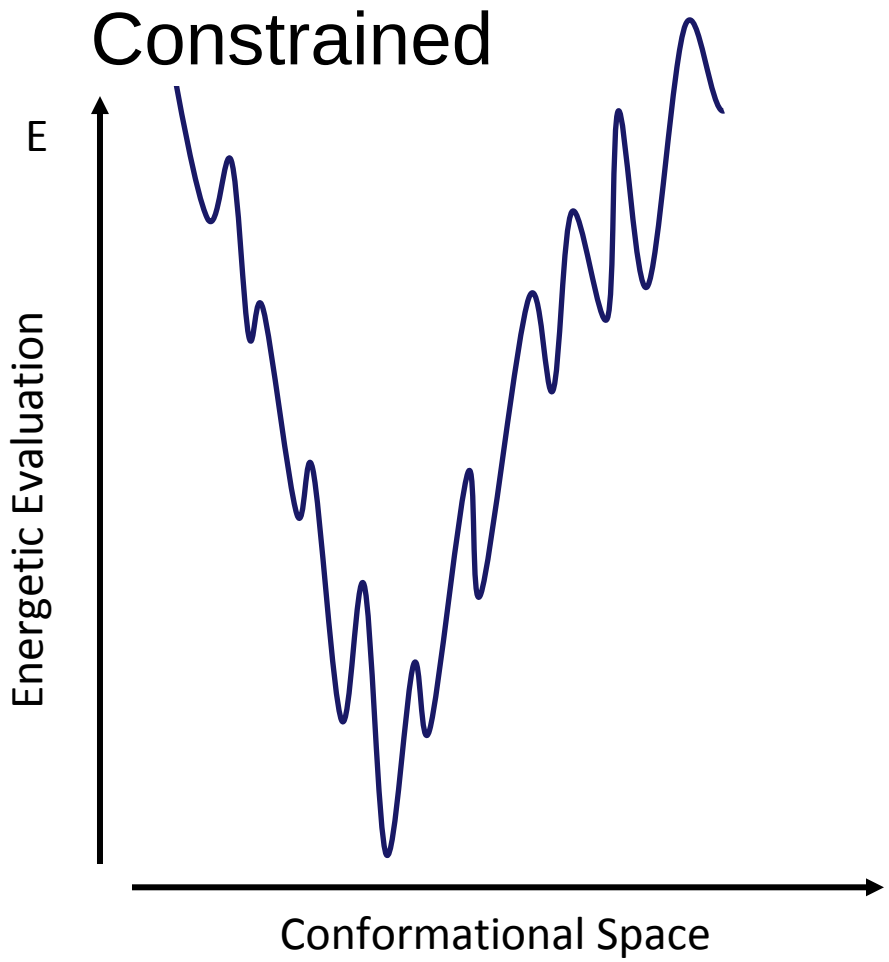
Combining Strengths: Building Accurate Models from Limited Data



“Constraints” (Restraints) in Rosetta

“Constraints” alter the energy function





Separation of Measurement 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
```

Separation of Measurement 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

- AmbiguousConstraint
- KofNConstraint

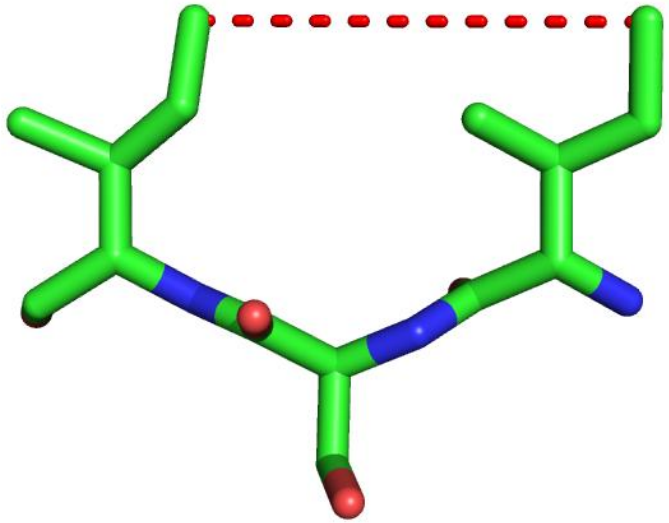
- Harmonic
- CircularHarmonic
- Flat-bottomed Harmonic
- Sigmoid
- Bounded

Measurement Types

```
AtomPair CD1 52 CD1 54 HARMONIC 6.0 0.2
```

- Distance (AtomPair)
- Angle
- Dihedral
- Coordinate

- AmbiguousConstraint
- KofNConstraint

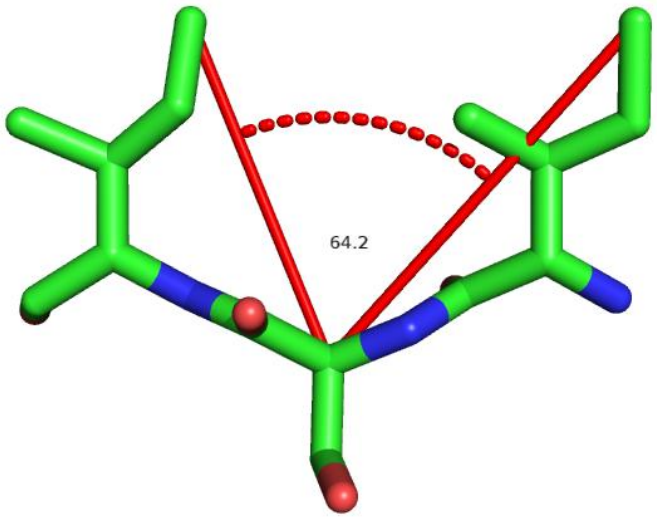


Measurement Types

```
Angle CD1 52 CA 53 CD1 54 HARMONIC 60.0 0.2
```

- Distance (AtomPair)
- Angle
- Dihedral
- Coordinate

- AmbiguousConstraint
- KofNConstraint

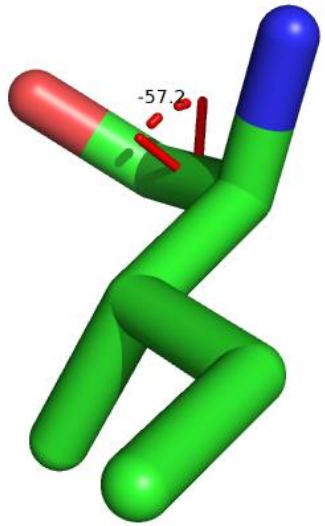


Measurement Types

```
Dihedral N 52 CA 52 C 52 O 52 HARMONIC 0.0 0.2
```

- Distance (AtomPair)
- Angle
- Dihedral
- Coordinate

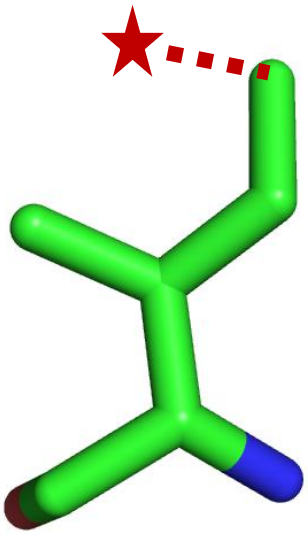
- AmbiguousConstraint
- KofNConstraint



Measurement Types

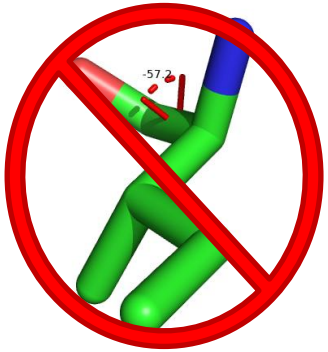
```
CoordinateConstraint CD1 54 CA 54 X Y Z HARMONIC 0.0 0.2
```

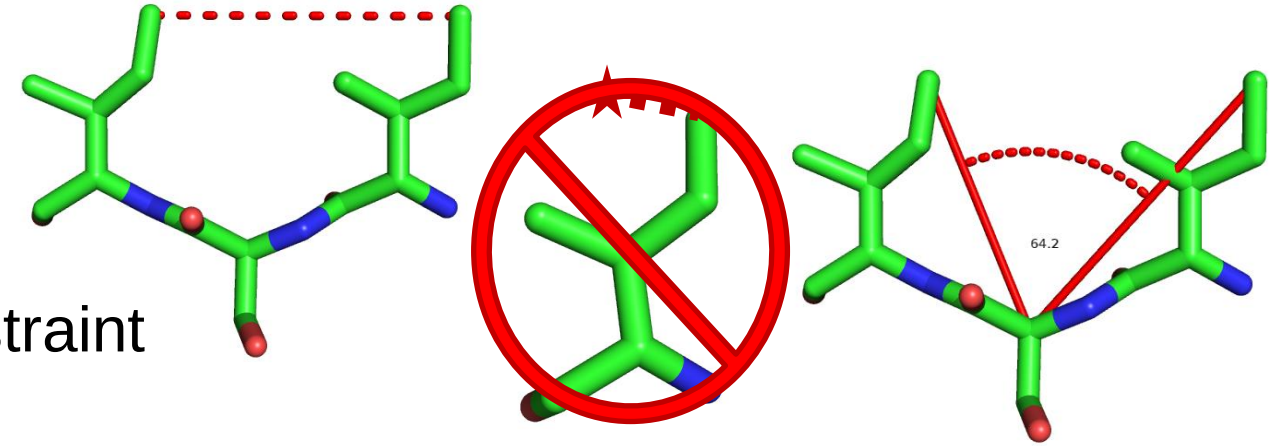
- Distance (AtomPair)
- Angle
- Dihedral
- Coordinate
- AmbiguousConstraint
- KofNConstraint



Measurement Types

```
KofNConstraint 2
AtomPair CD1 52 CD1 54 HARMONIC 6.0 0.2
Angle CD1 52 CA 53 CD1 54 HARMONIC 60.0 0.2
Dihedral N 52 CA 52 C 52 O 52 HARMONIC 0.0 0.2
CoordinateConstraint CA 1 CD1 54 HARMONIC 0.0 0.2
END
```





- AmbiguousConstraint
- KofNConstraint

Scoring Types

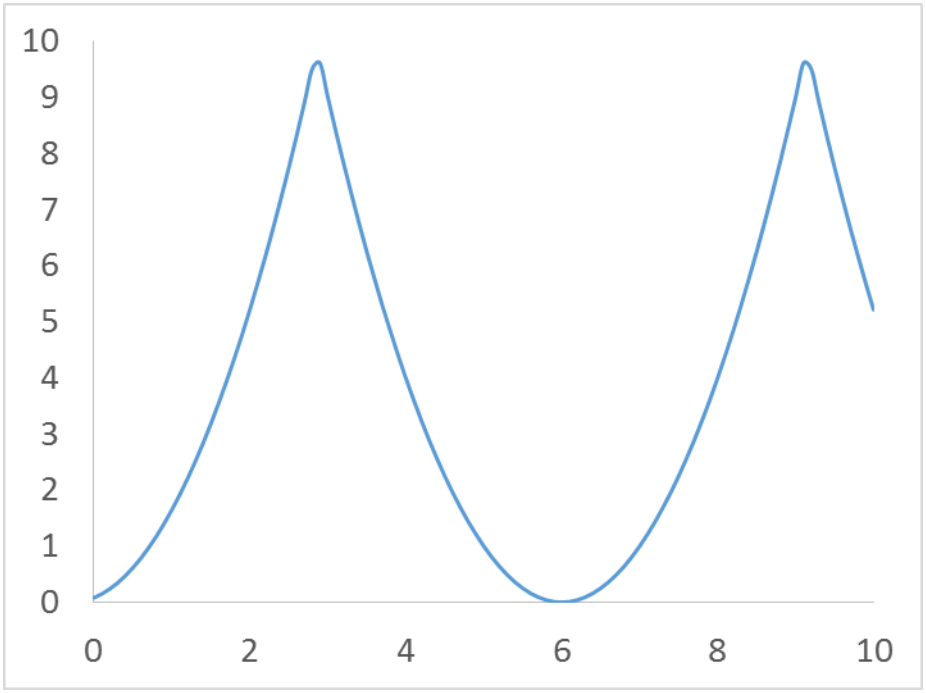
```
AtomPair CD1 52 CD1 54 HARMONIC 6.0 1.0
```

```
HARMONIC 6.0 3.0
```

- Harmonic
- CircularHarmonic
- Flat-bottomed Harmonic
- Sigmoid
- Bounded

Scoring Types

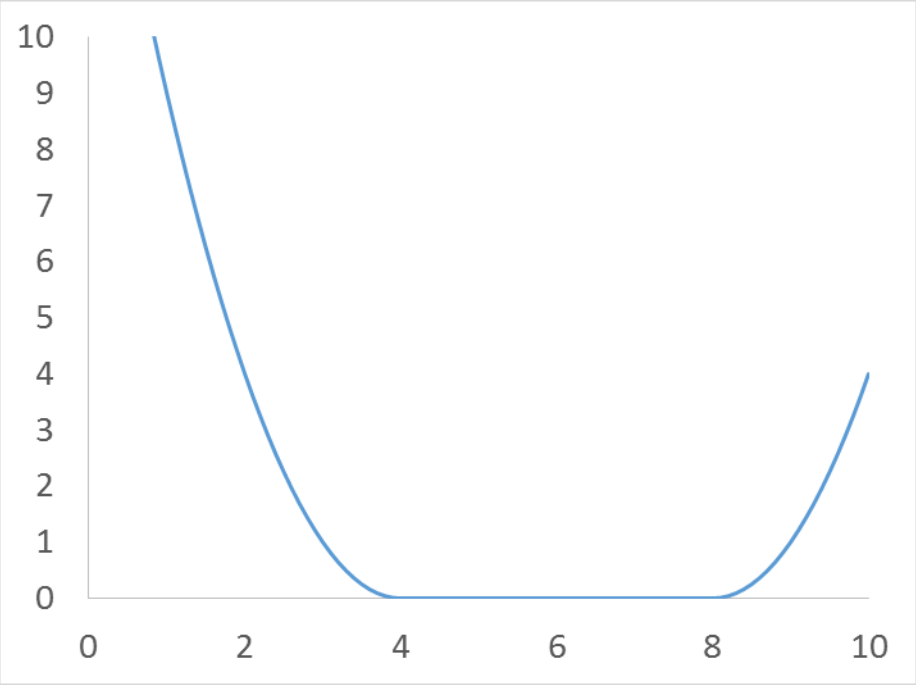
```
Dihedral N 52 CA 52 C 52 O 52 CIRCULARHARMONIC 6.0 1.0
```



- Harmonic
- CircularHarmonic
- Flat-bottomed Harmonic
- Sigmoid
- Bounded

Scoring Types

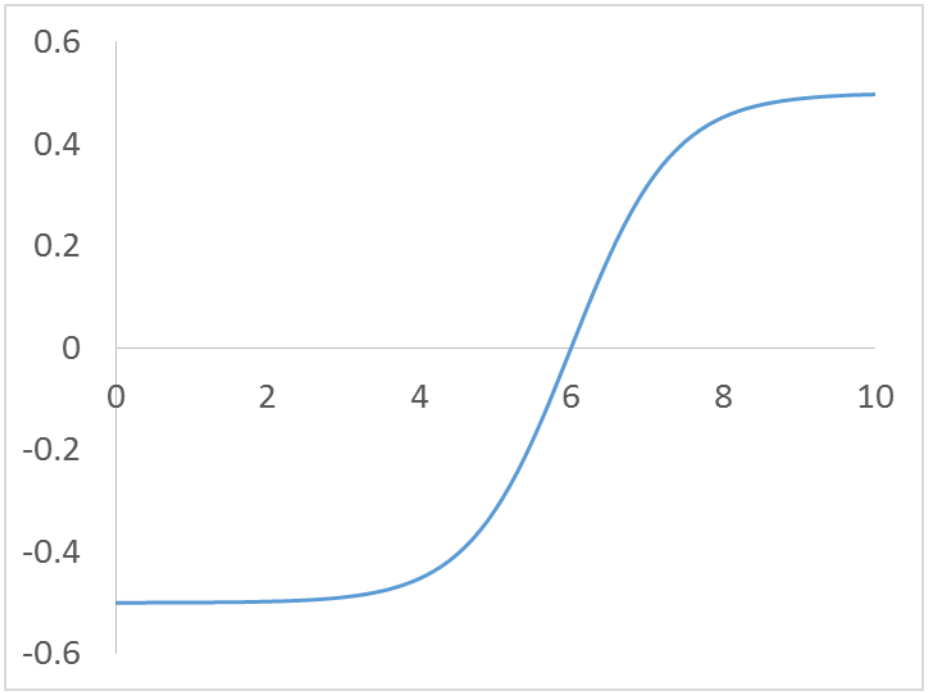
```
AtomPair CD1 52 CD1 54 FLAT_HARMONIC 6.0 1.0 2.0
```



- Harmonic
- CircularHarmonic
- Flat-bottomed Harmonic
- Sigmoid
- Bounded

Scoring Types

```
AtomPair CD1 52 CD1 54 SIGMOID 6.0 5
```

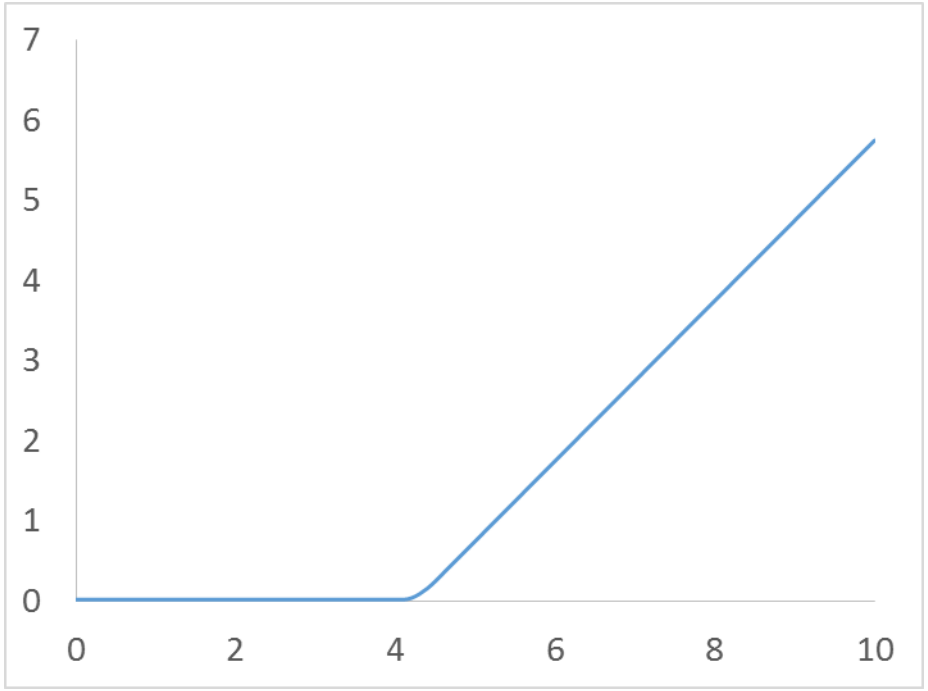


- Harmonic
- CircularHarmonic
- Flat-bottomed Harmonic
- Sigmoid
- Bounded

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

Scoring Types

```
AtomPair CD1 52 CD1 54 BOUNDED 0.0 4.0 1.0 0.5 TAG
```



- Harmonic
- CircularHarmonic
- Flat-bottomed Harmonic
- Sigmoid
- Bounded

Using Constraints: Two Parts

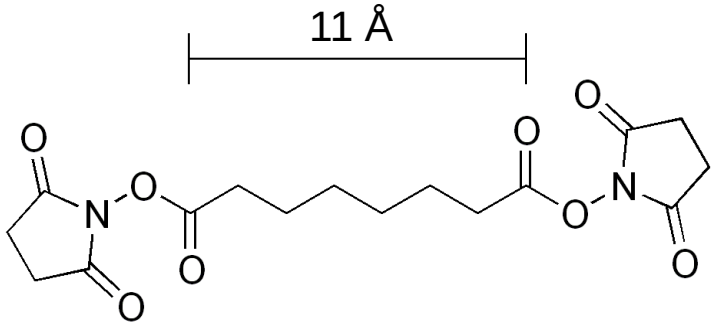
Constraints in the Pose

- Constraint file

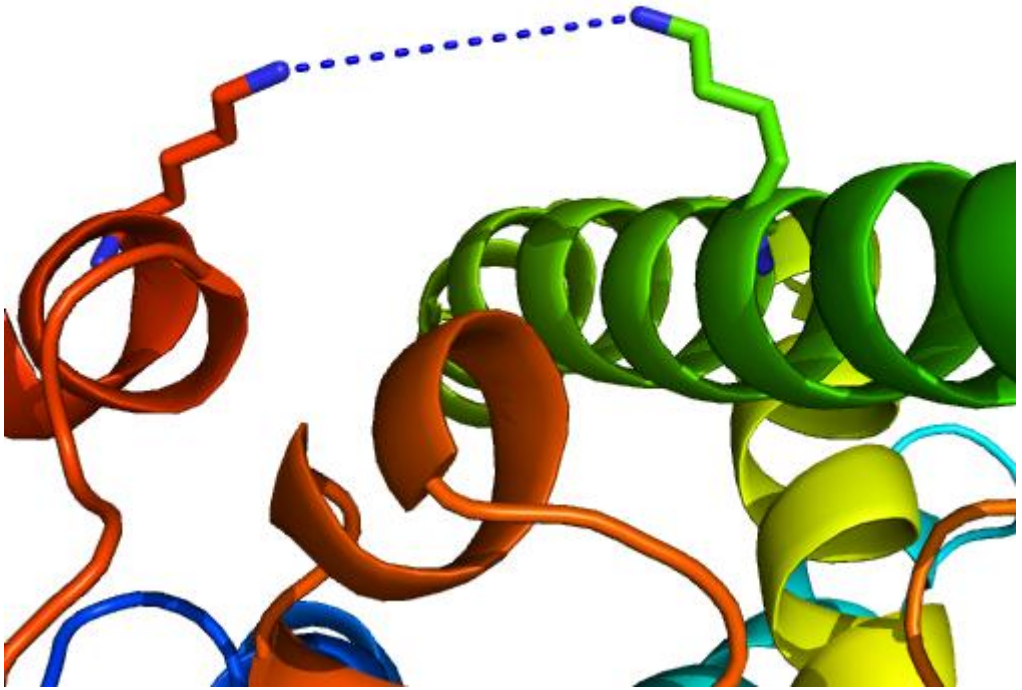
Constraints in the Scorefunction

- Weights File

Crosslinks – direct distance measurement



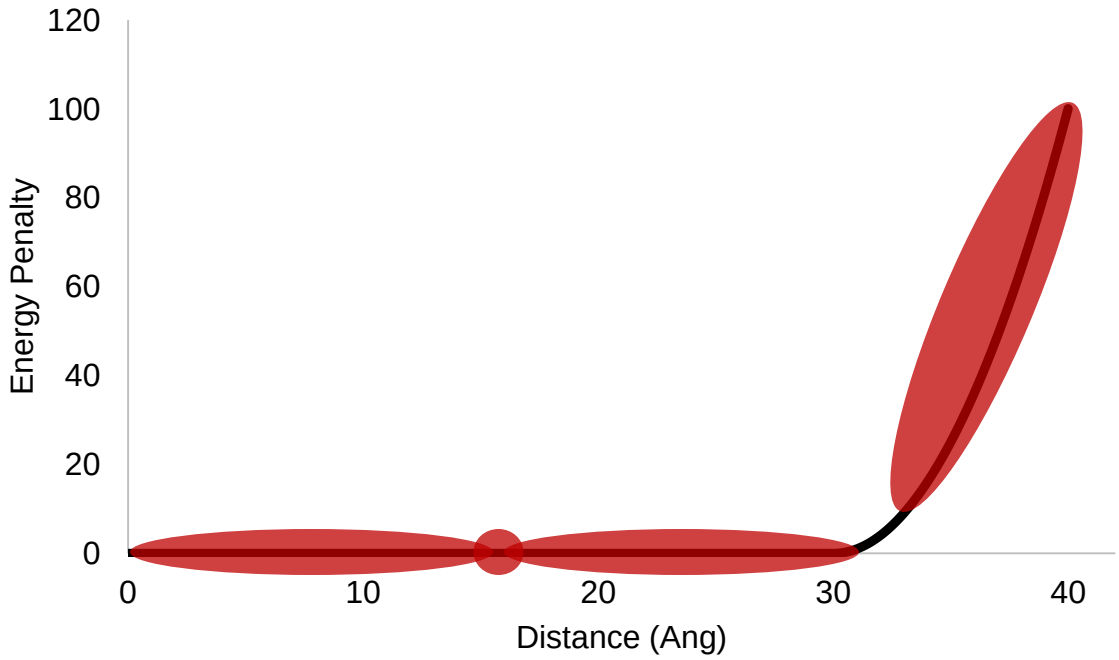
Disuccinimidyl suberate (DSS)



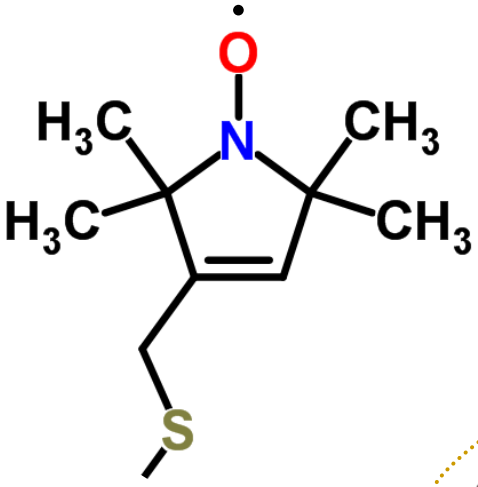
Kahraman et al. PLOS One, 2013, 8(9) e73411
rosetta/demos/protocol_capture/xl_driven_protein_docking/

Chemical crosslinking gives distance restraint information

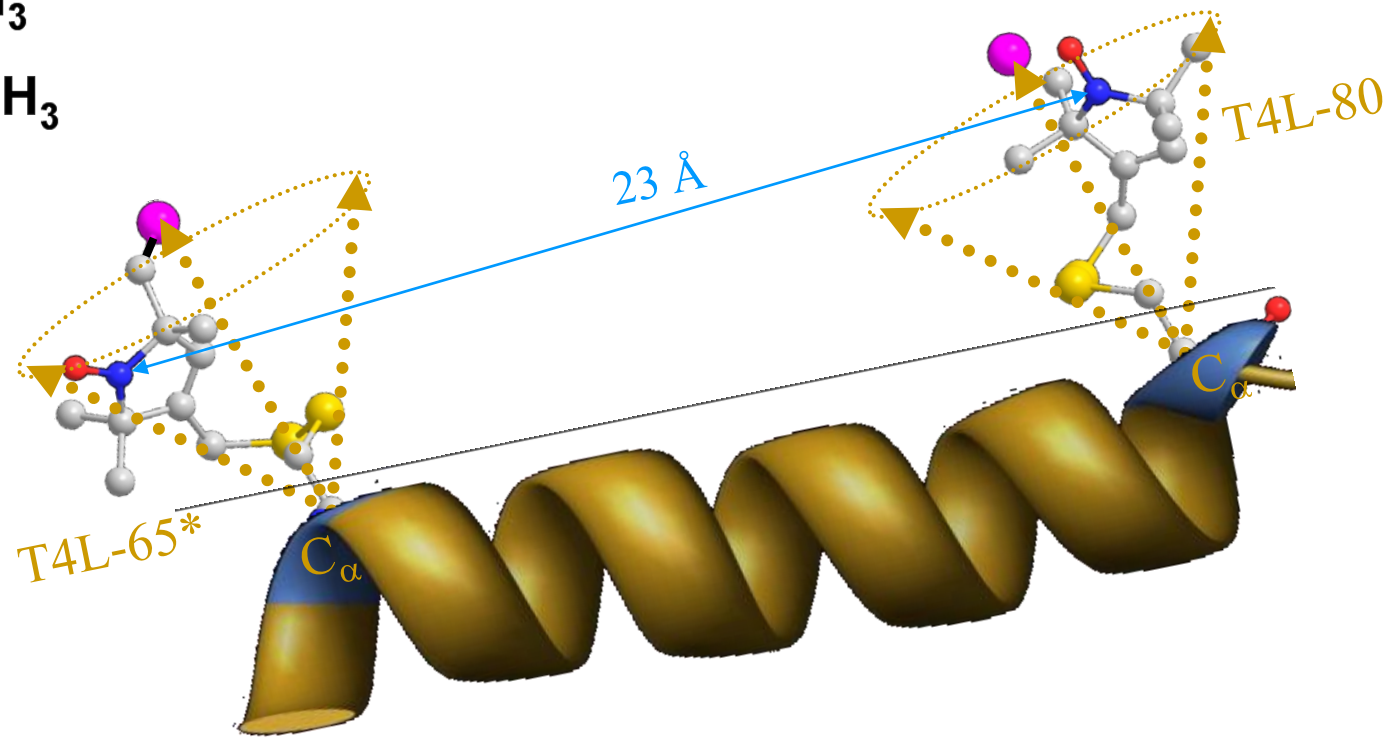
```
AtomPair CB 67 CB 124 FLATHARMONIC 15.0 1.0 15.0
```



Double Electron-Electron Resonance gives distance information

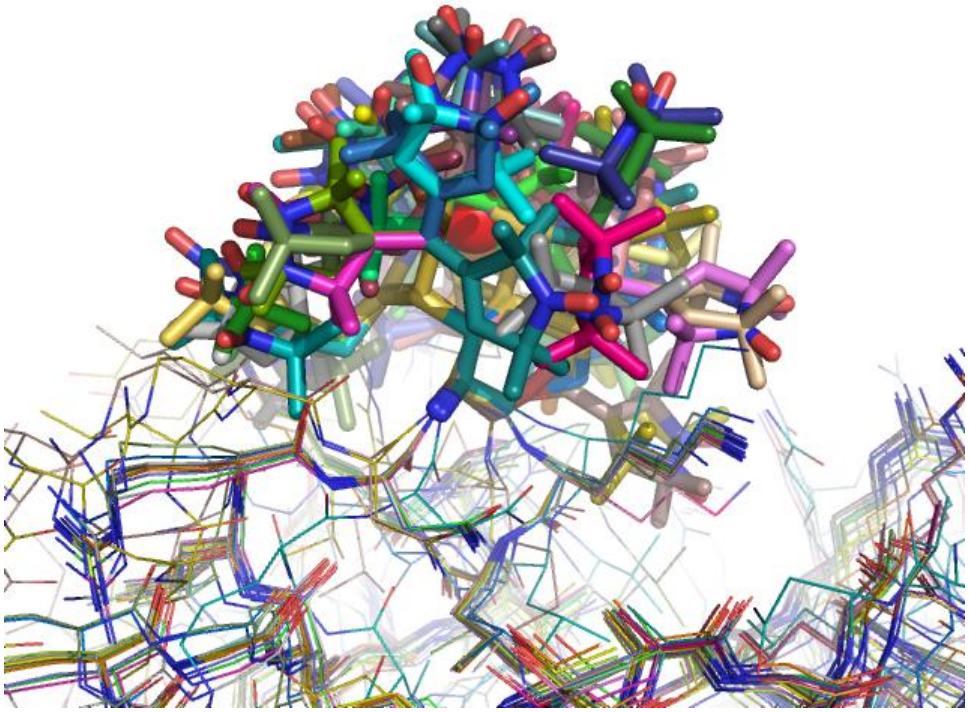


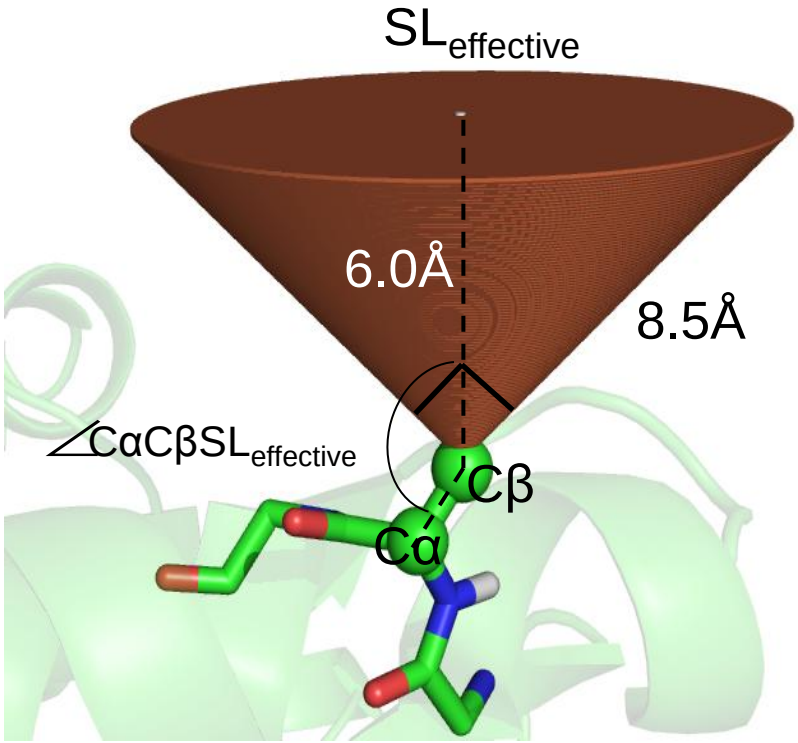
MTSL



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



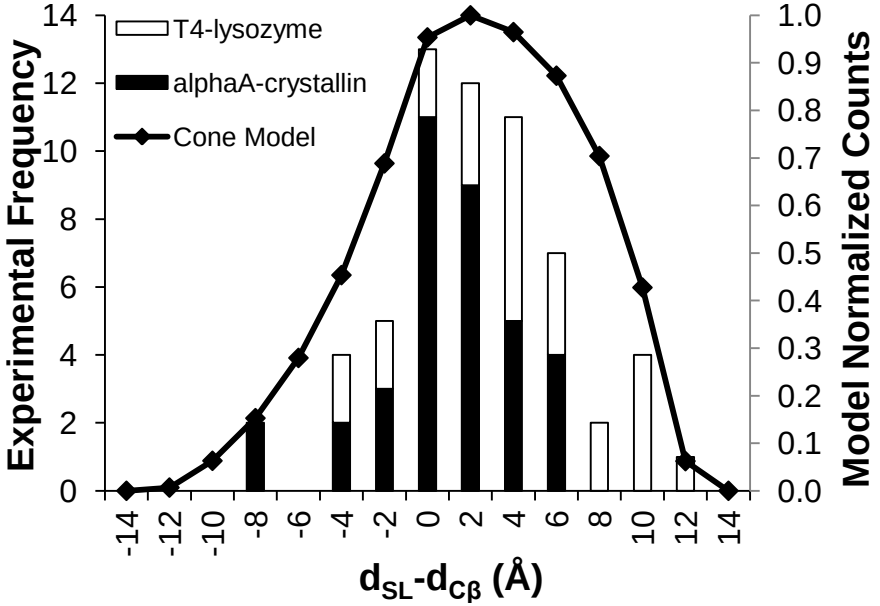


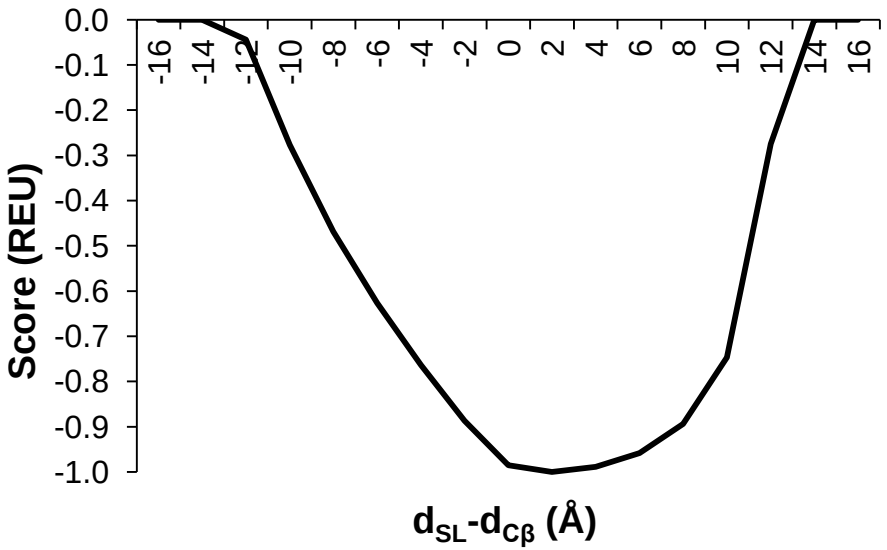
Alexander, N.; Al-Mestarihi, A.; Bortolus, M.; McHaourab, H.; Meiler, J. "De Novo High-Resolution Protein Structure Determination from Sparse Spin-Labeling EPR Data" *Structure* ***2008, 16, 181-95.***

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

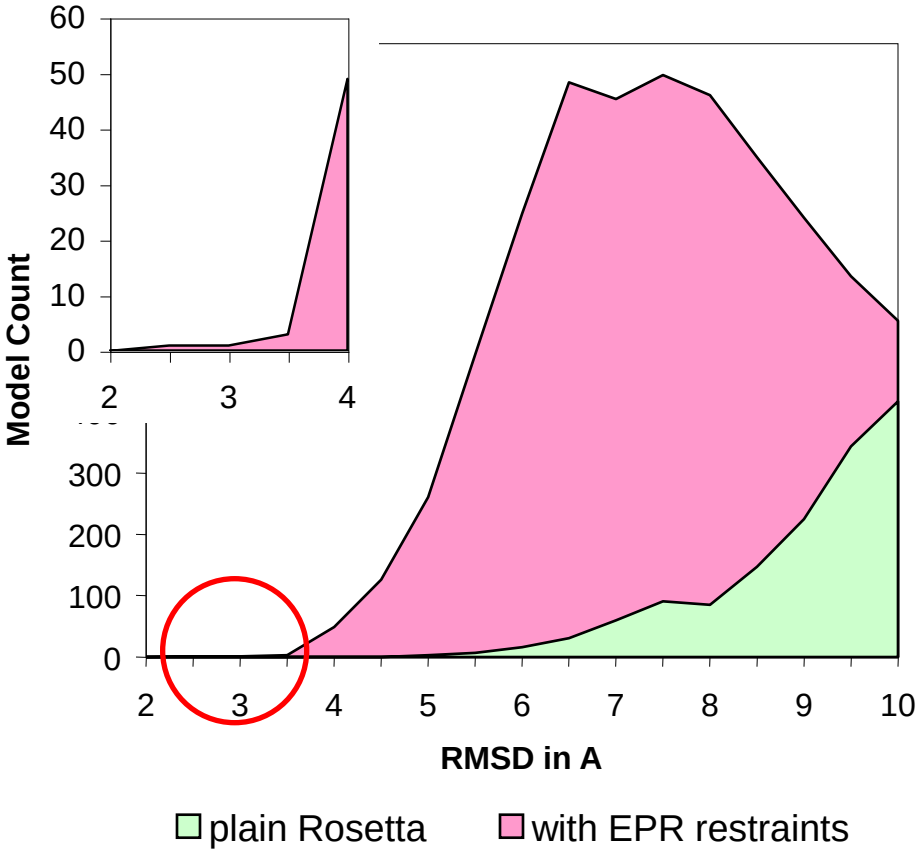


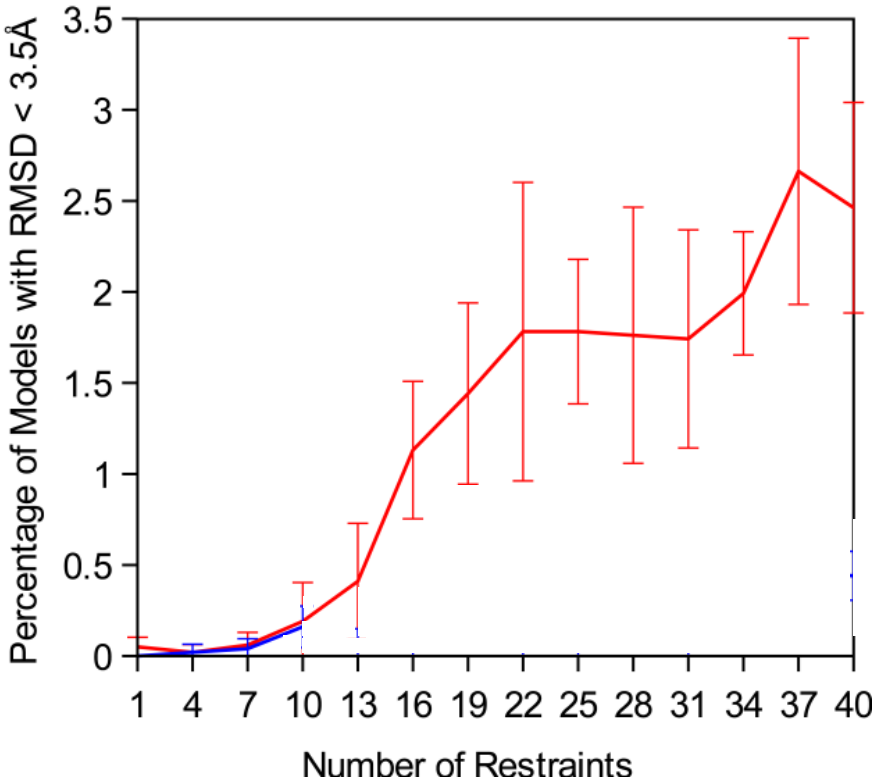



```
AtomPair CB 65 CB 80 SPLINE EPR_DISTANCE 16.0 4.0 0.5
```


Alexander *et al. Structure* ***2008, 16, 181-95.***
Hirst *et al. J Struct Biol*; **2011**; Vol. 173 (3): p. 506-14.

Influence of Experimental Data on Sampling and Model Quality





NMR - NOE distance constraints

Simple AtomPair contraints:

AtomPair H 95 H 105 BOUNDED 1.5 3.650 0.3 NOE; amide-amide

Proton Ambiguities:

AmbiguousNMRDistance H 56 QD1 71 BOUNDED 1.5 4.000 0.5 NOE; amide-methyl

Proton/Carbon Ambiguities:

AmbiguousNMRDistance H 55 QQG 94 BOUNDED 1.5 4.000 0.5 NOE; amide-methyl

AmbiguousNMRDistance QQD 25 QQG 108 BOUNDED 1.5 4.000 0.5 NOE; methyl-methyl

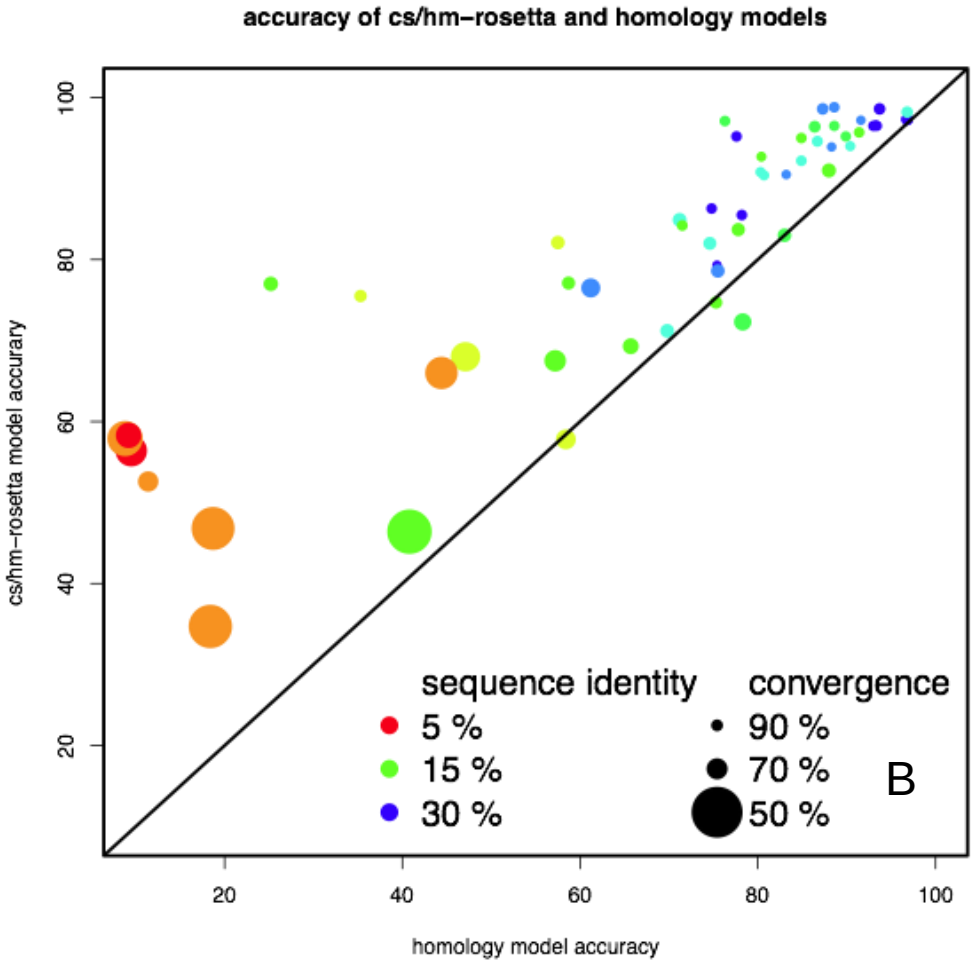
All restraints must be mapped into the centroid representation:

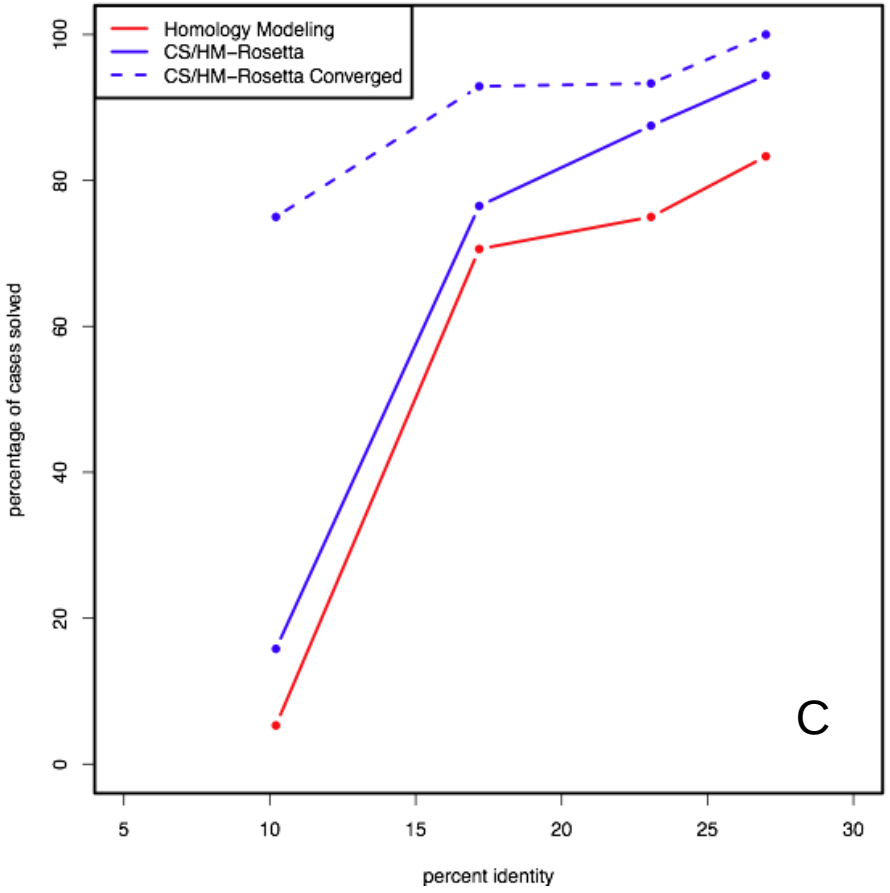
cat [fullatom Rosetta constraint file] | perl map_csts_to_centroid_simple.pl > [centroid Rosetta constraint file]

$$E(x) = \begin{cases} 0, & x \geq ll \cap x \leq ul \\ 2^{(x-ul)/0.3}, & x > ul \\ 2^{(ll-x)/0.3}, & x < ll \end{cases}$$

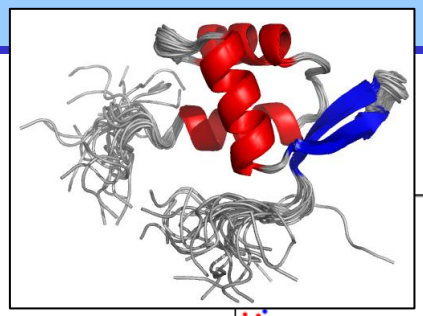
Non-Constraint-Based Experimental Data Incorporation

Chemical shifts improves the performance of sequence-alone homology modeling methods

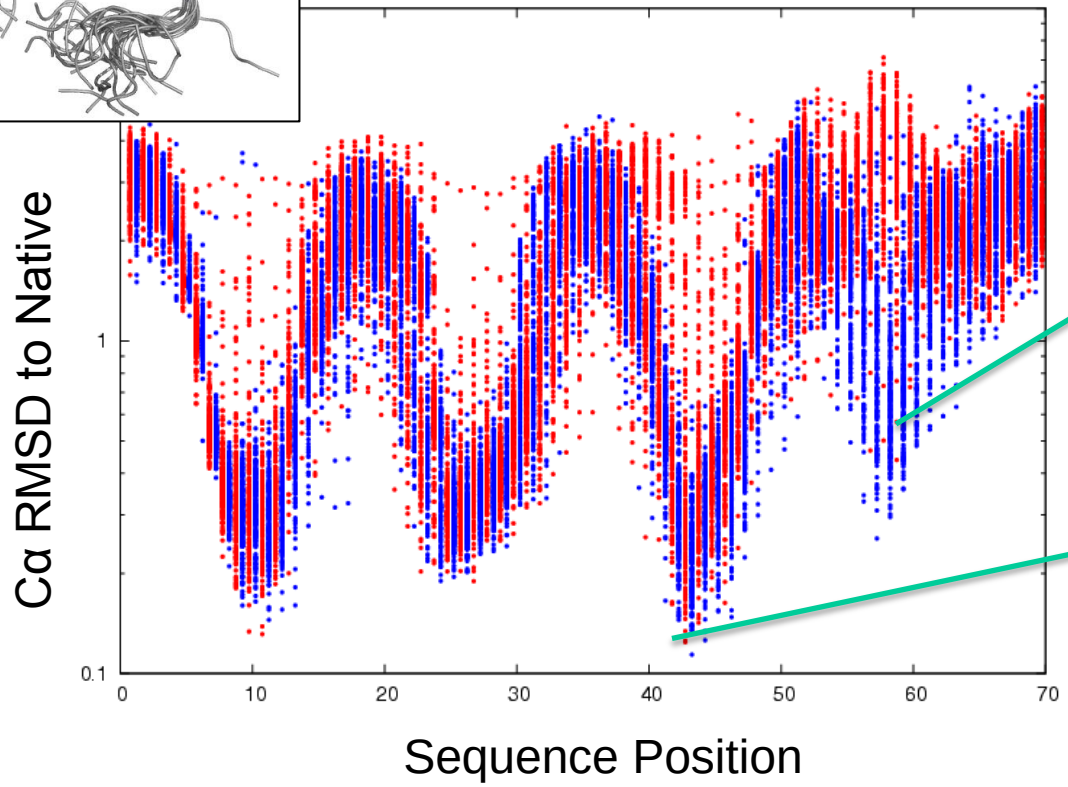




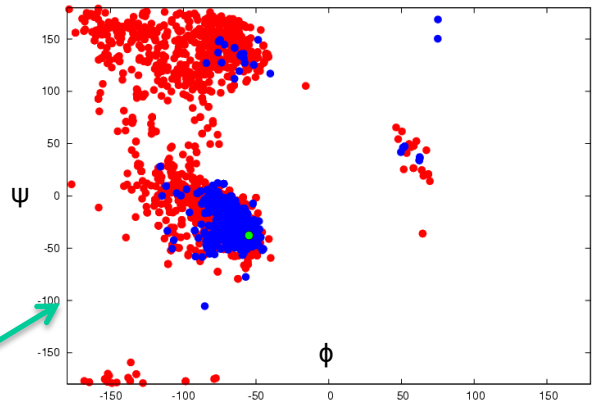
Chemical Shifts help fragment picking



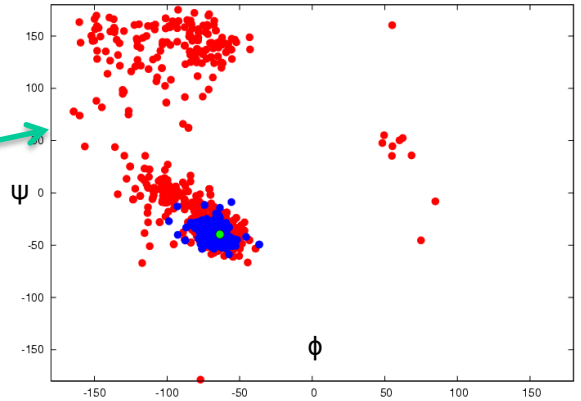
Fragment Comparison



Residue 61



Residue 45



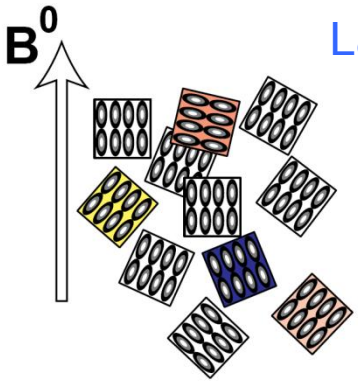
Red: Rosetta Fragments, picked by Sequence profile & Secondary Structure Prediction

Blue: CSRosetta Fragments, picked by CS Comparison & Sequence Matching

RDC restraints in Rosetta

Multiple bond vector support (data are automatically scaled relative to N-H)

```
3 N  3  H  6.64; amide
3 C  4  N  3.34; C'-N
3 C  3 CA  3.4; C'-Ca
```



Lack of molecular tumbling

Support of multiple alignment datasets

```
-in:file:rdc gel.rdc phage.rdc
```

2 types of scoring:

1) Singular value decomposition

```
-rdc:fit_method svd
```

2) Non-linear fitting of the 5 alignment tensor parameters

```
-rdc:fit_method nls
-rdc:fixDa [value]
-rdc:fixR [value]
```

Evaluation of Q-factors, al. tensor parameters calculated RDCs:

```
-out:level 999 |grep Qbax || Da  || Rh ; exactly the same output as PALES
-rdc:print_rdc_values calc.rdc
```

$$Q = \frac{RMS(D_{calc} - D_{obs})}{\sqrt{D_a^2(4+3R^2)/5}}$$

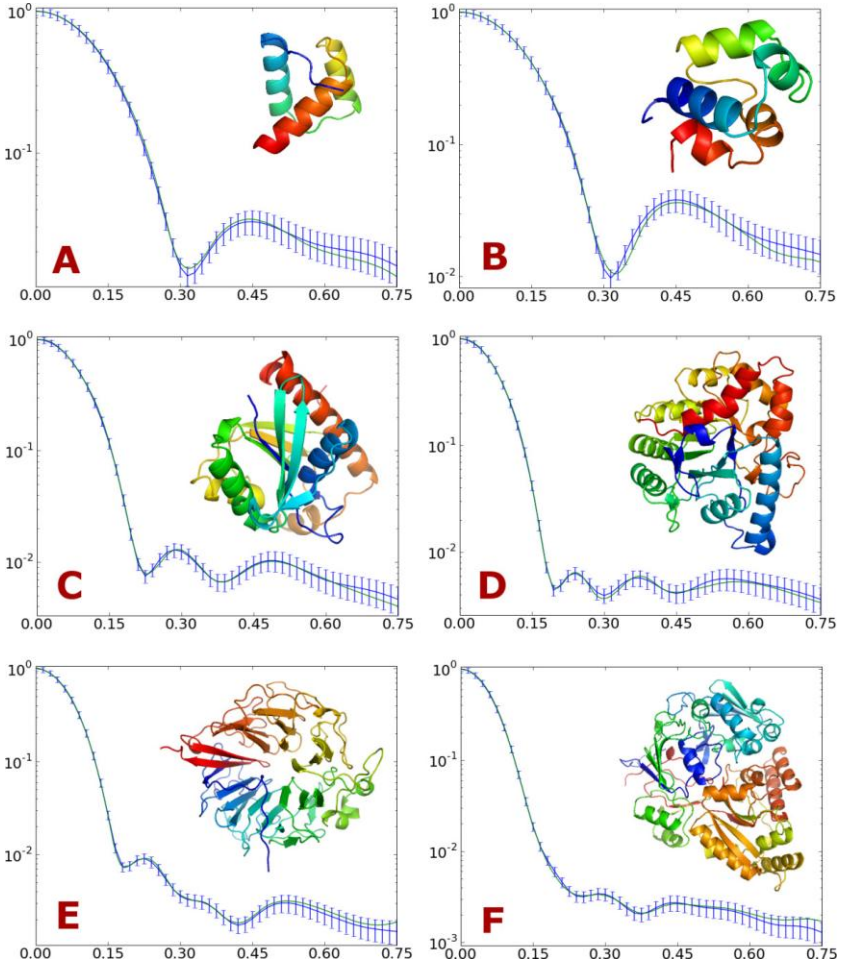
SAXS restrains in Rosetta

Uses a coarse-grain representation with residue-specific “form factors” (Stovgaard et al., BMC Bioinformatics, 2010):

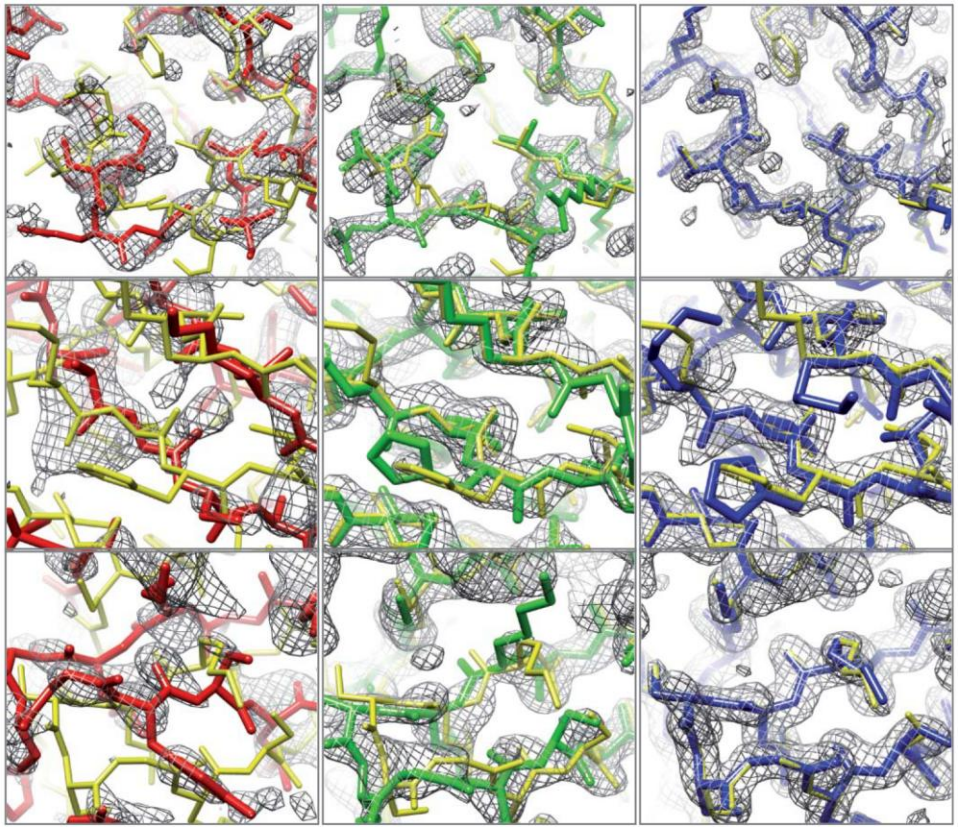
-residues:patch_selectors CENTROID_HA
-score:saxs:ref_spectrum saxs_sparse.dat
-score::patch patch_saxs

Patch file contains:
fastsaxs = 0.05

Data file:
q I(q) Delta
0.00771096 7554.24 70.6635
0.017006 7253.15 9.33698
0.0263011 6830.58 7.28595
0.0355961 6285.76 6.17379
0.0448912 5670.72 5.27
0.0541862 4985.43 4.56433
0.0634813 4285.38 3.83601
0.0727763 3587.02 3.27892
0.0820714 2926.51 2.97787



Electron Density (EM, X-ray)



DiMaio F. (2011) Nature

Initial molecular replacement → Improve Phases → Model rebuilding and energy optimization