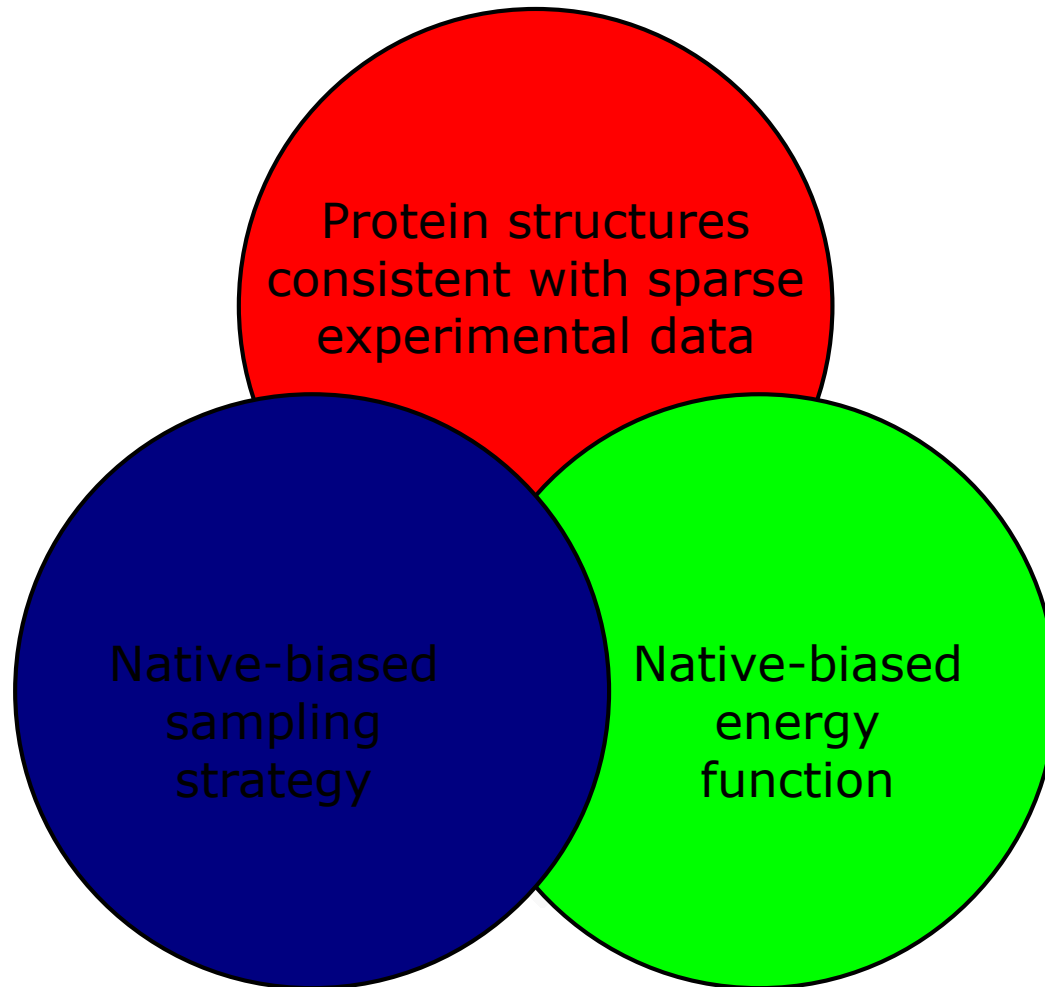


Combining Experimental Data with Rosetta Computational Models

Jessica A. Finn
jessica.a.finn@vanderbilt.edu

Combine strengths to build accurate models from limited data



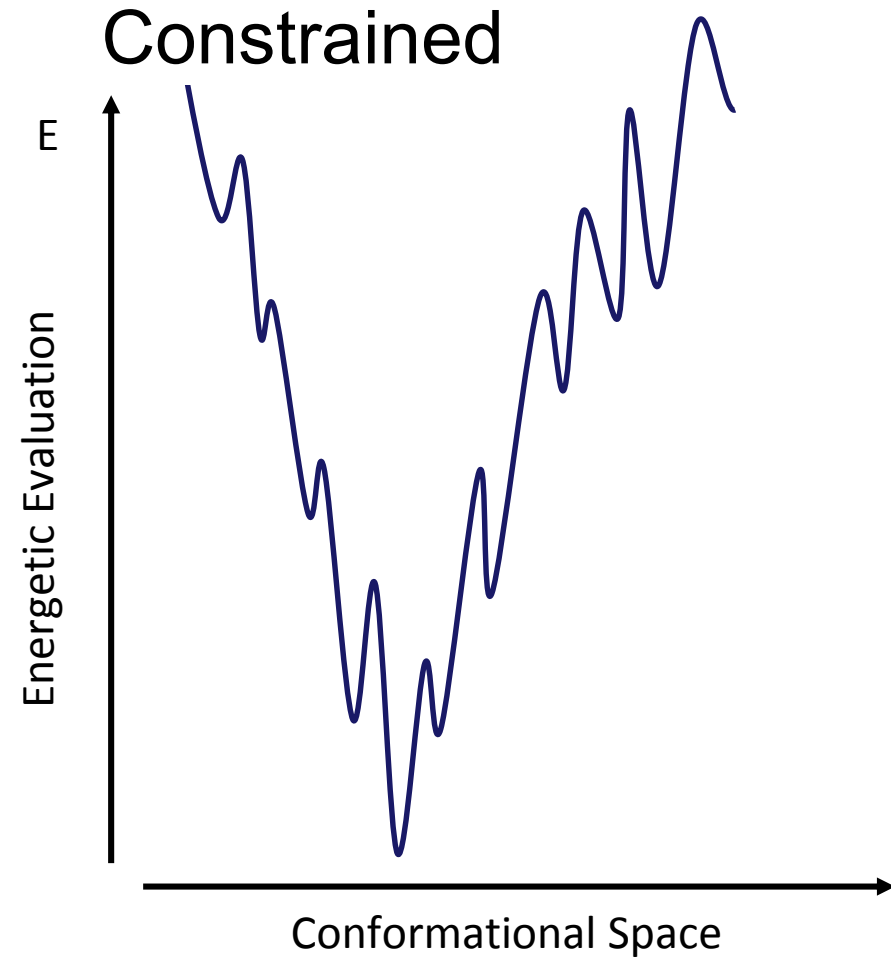
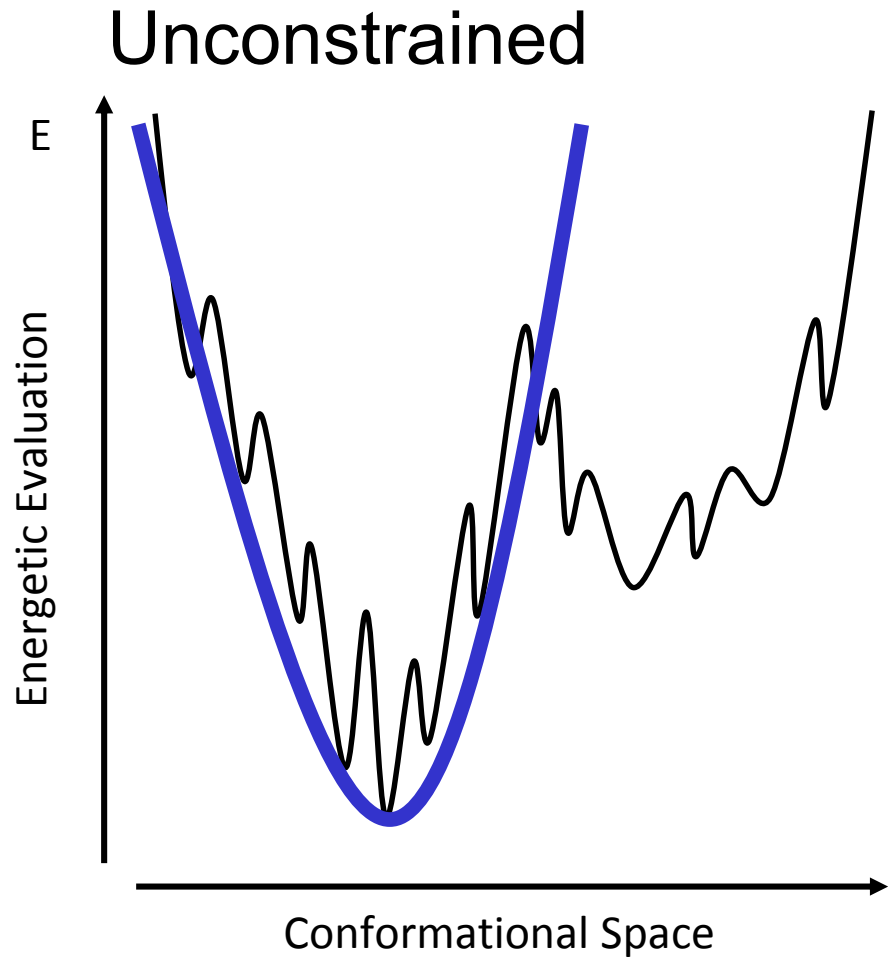
Complete
conformational
space



Experimental data takes many forms

- Distance measurements from chemical crosslinking
- Chemical shift or residual dipolar coupling (RDC) data from NMR
- Small angle X-ray scattering (SAXS) data
- Low-resolution electron density map from electron microscopy (EM) or X-ray crystallography

Rosetta Constraints alter the energy function



Rosetta Constraints are provided as a human-readable text file



What is Measured

How it is Scored



Many constraint types exist

```
AtomPair NE2 13 V3 32 HARMONIC 0.0 0.2
```

```
Angle CD2 13 NE2 13 ZN 32 HARMONIC 2.09 0.35
```

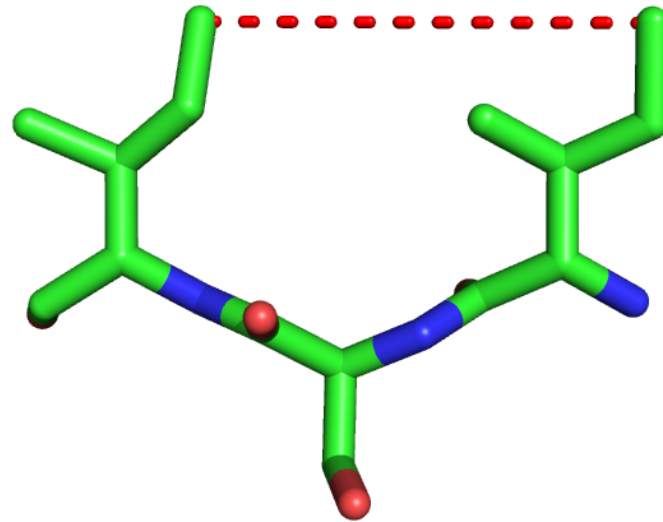
```
Dihedral CG 13 CD2 13 NE2 13 ZN 32 CIRCULARHARMONIC 3.14 0.35
```

- Distance (AtomPair)
- Angle
- Dihedral
- Coordinate
- MultiConstraint
- AmbiguousConstraint
- KofNConstraint
- Harmonic
- CircularHarmonic
- Flat-bottomed Harmonic
- Gaussian
- Sigmoid
- Bounded

Measurement types

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

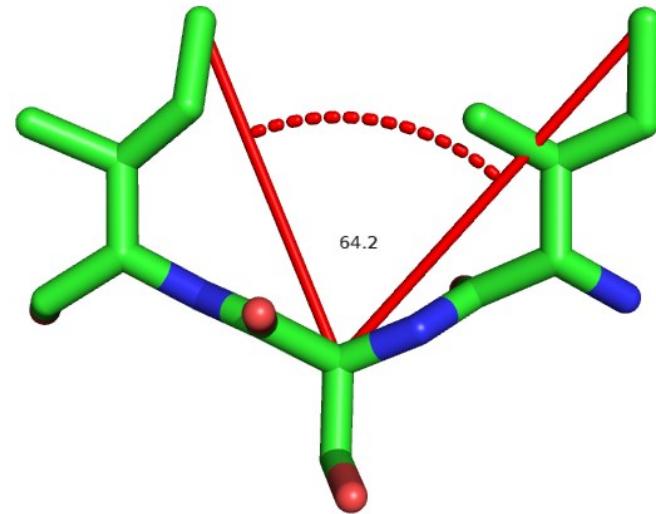
- Distance (AtomPair)
- Angle
- Dihedral
- Coordinate
- MultiConstraint
- AmbiguousConstraint
- KofNConstraint



Measurement types

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

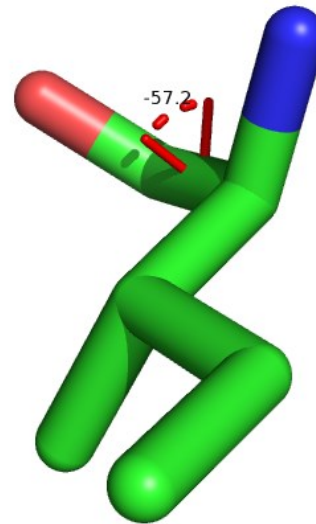
- Distance (AtomPair)
- Angle
- Dihedral
- Coordinate
- MultiConstraint
- AmbiguousConstraint
- KofNConstraint



Measurement types

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

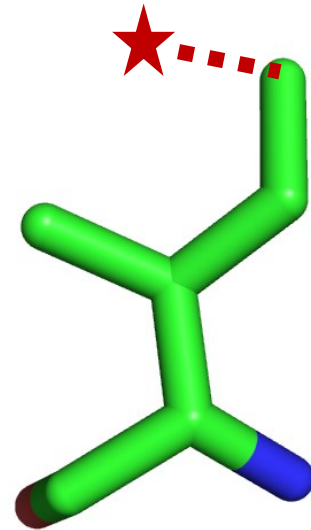
- Distance (AtomPair)
- Angle
- **Dihedral**
- Coordinate
- MultiConstraint
- AmbiguousConstraint
- KofNConstraint



Measurement types

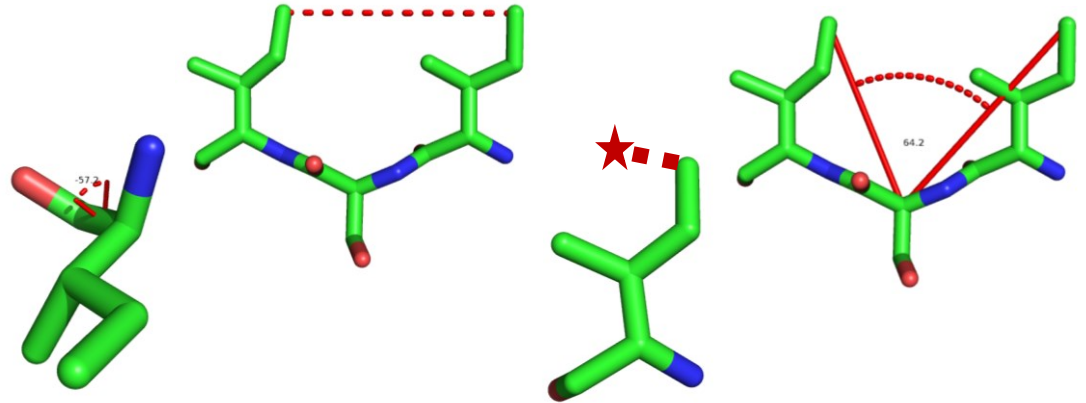
```
CoordinateConstraint CA 1 CD1 54 X Y Z HARMONIC 0.0 0.2
```

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



Measurement grouping

- MultiConstraint
- AmbiguousConstraint
- KofNConstraint



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

Scoring types

AtomPair CD1 52 CD1 54 HARMONIC 6.0 1.0

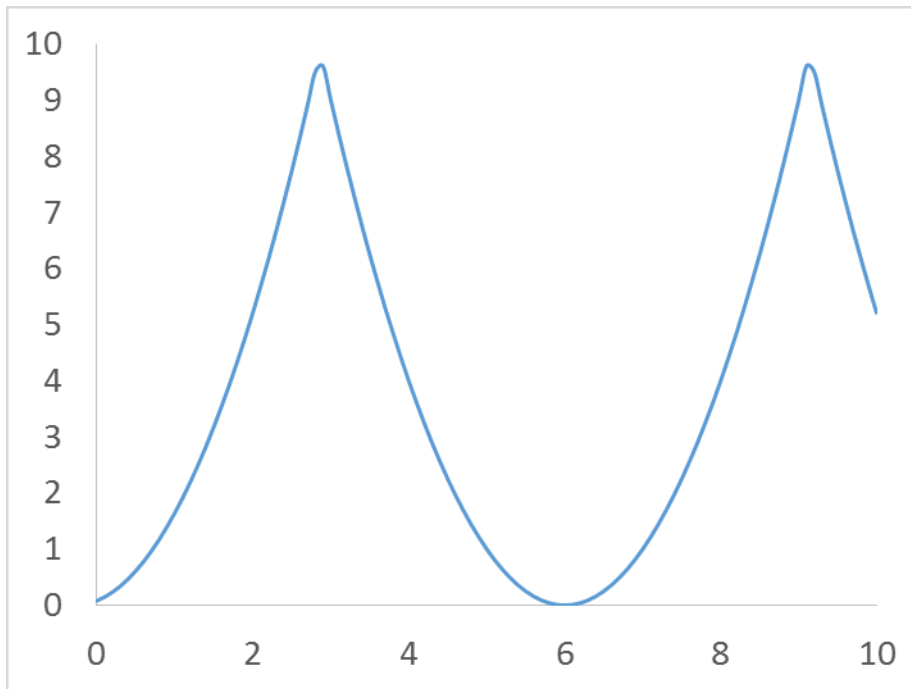
HARMONIC 6.0 3.0



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

Scoring types

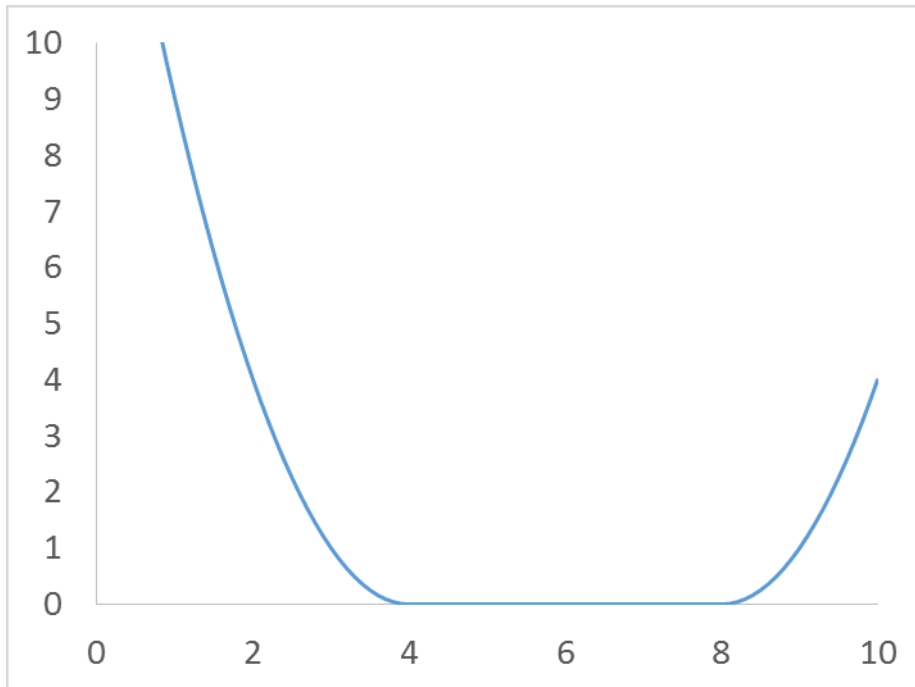
```
AtomPair CD1 52 CD1 54 CIRCULARHARMONIC 6.0 1.0
```



- Harmonic
- **CircularHarmonic**
- Flat-bottomed Harmonic
- Gaussian
- Sigmoid
- Bounded

Scoring types

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

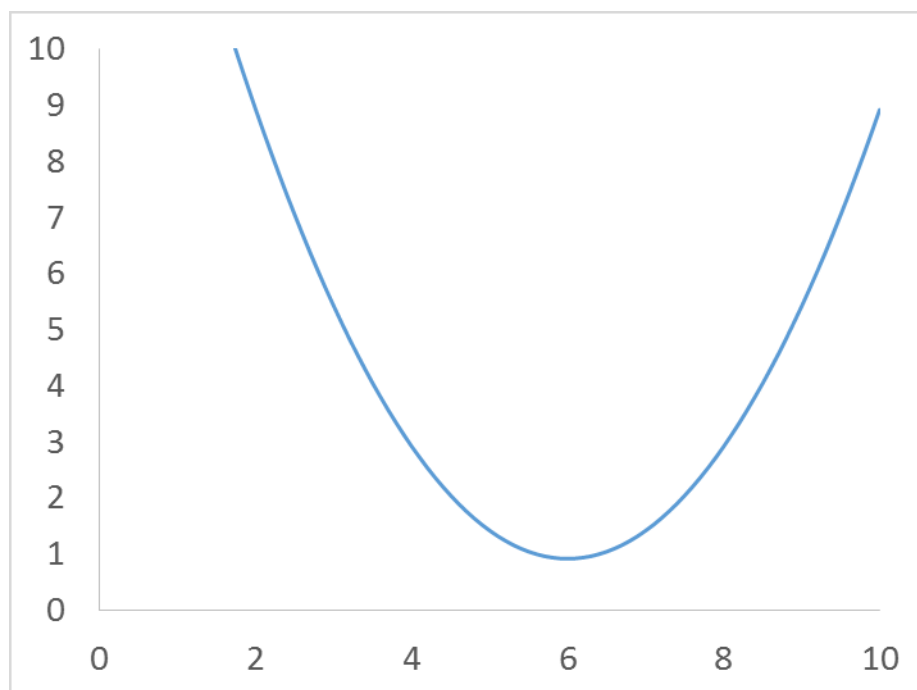


- Harmonic
- CircularHarmonic
- **Flat-bottomed Harmonic**
- Gaussian
- Sigmoid
- Bounded

Scoring types



```
AtomPair CD1 52 CD1 54 GAUSSIANFUNC 6.0 1.0 TAG
```

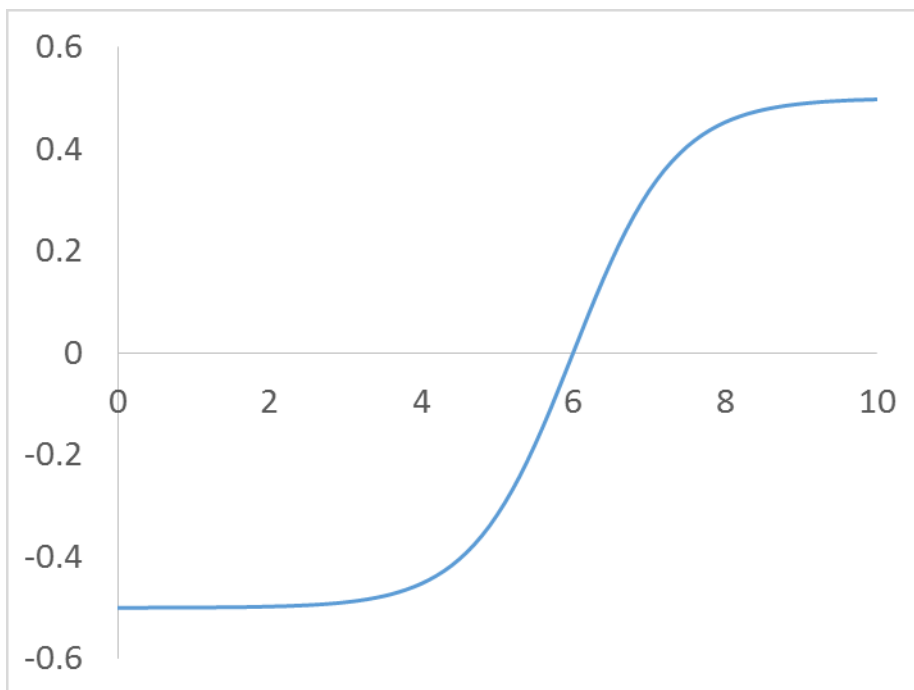


- Harmonic
- CircularHarmonic
- Flat-bottomed Harmonic
- Gaussian
- Sigmoid
- Bounded

Scoring types



AtomPair CD1 52 CD1 54 SIGMOID 6.0 5

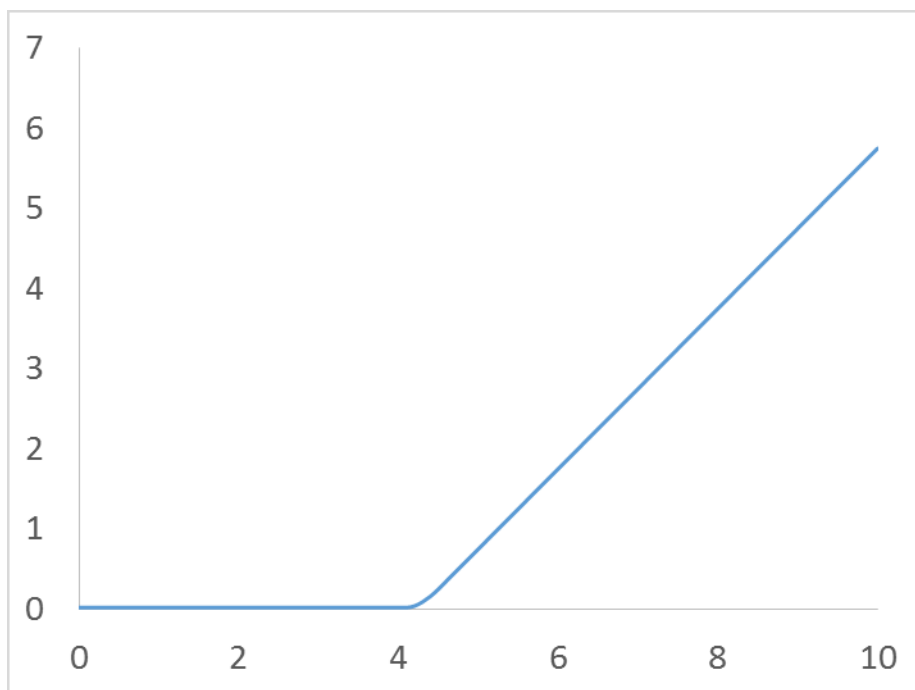


- Harmonic
- CircularHarmonic
- Flat-bottomed Harmonic
- Gaussian
- Sigmoid
- Bounded

Scoring types



AtomPair CD1 52 CD1 54 BOUNDED 0.0 4.0 1.0 0.5 TAG



- Harmonic
- CircularHarmonic
- Flat-bottomed Harmonic
- Gaussian
- Sigmoid
- **Bounded**



Using constraints

- Constraints in the Pose
 - Constraint file
- Constraints in the Scorefunction
 - Weights file
- Constraint flags
 - constraints:cst_file these_are_my.constraints
 - score:weights ref2015_cst.wts

Evaluating constraints

- Constraints are added to the score table as additional terms that influence the per-residue and total pose score

```
# All scores below are weighted scores, not raw scores.
#BEGIN_POSE_ENERGIES_TABLE ex2_wcons/ex2-11wt5_renum_0001.pdb
label fa_atr fa_rep fa_sol fa_intra_rep fa_elec pro_close hbond_sr_bb hbond_lr_bb hbond_bb_sc hbond_sc dslf_fa13 atom_pair_constraint coordin
ate_constraint angle_constraint dihedral_constraint rama omega fa_dun p_aa_pp ref res_type_constraint total
weights 0.8 0.44 0.75 0.004 0.7 1 1.17 1.17 1.17 1.1 1 1 1 1 1 0.2 0.5 0.56 0.32 1 1 NA
pose -829.409 100.163 448.378 2.06505 -91.2806 1.27809 -21.2574 -85.1707 -31.8975 -25.8795 -0.49984 0 0 0 2.64806 -10.6774 30.0743 194.658 -3
9.6382 5.37176 0 -351.073
GLN_p:NtermProteinFull_1 -0.85772 0.03751 0.83886 0.00531 -0.23938 0 0 0 0 -0.26803 0 0 0 0 0 1e-05 1.55705 0 -1.17797 0 -0.10437
VAL_2 -2.94747 0.56225 0.86563 0.01307 -0.08275 0 0 0 0 0 0 0 0 0 -0.19291 0.02784 0.09158 -0.3427 0.74484 0 -1.26061

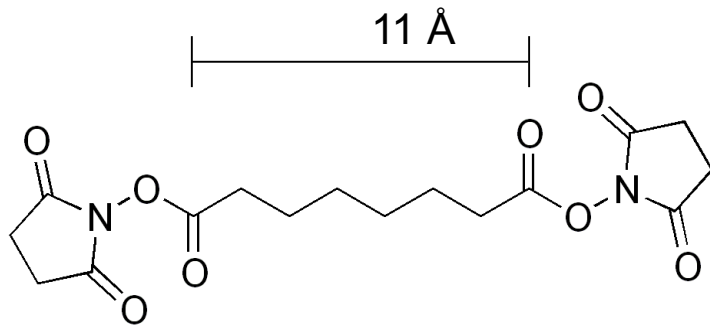
TYR_95 -7.42573 0.55873 3.34635 0.02421 -0.85737 0 0 0 0 -1.0491 0 0 0 0 0 -0.16257 0.2708 1.27367 -0.25728 0.1317 0 -4.14661
CYS_96 -3.46966 0.44618 1.94383 0.00189 -0.41055 0 0 0 -0.33043 0 0 0 0 0 0.10474 0.08028 0.02248 0.28275 -0.10582 0.35499 0 -1.07931
ALA_97 -3.18307 0.31768 1.39462 0.00066 -0.33975 0 0 0 0 0 0 0 0 0 0.31225 0.01153 0.05291 0 -0.18724 0.59294 0 -1.02746
ARG_98 -7.07128 0.40448 4.41041 0.01096 0.16306 0 0 0 0 0 0 0 0 0 0.51615 -0.14578 1e-05 1.4683 0.00759 -0.14916 0 -0.38525
SER_99 -2.44766 0.21928 2.21077 0.00066 -0.47777 0 0 0 0 -0.65771 0 0 0 0 0.32459 -0.01578 0.00721 0.51072 0.02569 0.17658 0 -0.12341
GLY_100 -1.69609 0.22908 1.29936 0.00011 -0.07953 0 0 0 -0.39086 0 0 0 0 0 0.01596 0.01856 0.05754 0 -0.7431 0.14053 0 -1.14847
GLY_101 -1.31452 0.19452 0.92981 0.00012 0.23747 0.00276 0 0 0 0 0 0 0 0 0 -0.06338 0.00377 0 -0.80654 0.14053 0 -0.67546
PRO_102 -1.6779 0.12523 0.55782 0.00178 0.03195 0.02426 0 0 0 0 0 0 0 0 0 -0.11974 0.19276 0.18944 -0.07634 -0.21929 0 -0.97003
TYR_103 -3.21624 0.36513 1.75051 0.02354 0.05388 0 0 0 0 0 0 0 0 0 0.0114 -0.13507 0.46186 1.3329 -0.14439 0.1317 0 0.63522
```

Examples of incorporating experimental data as Constraints

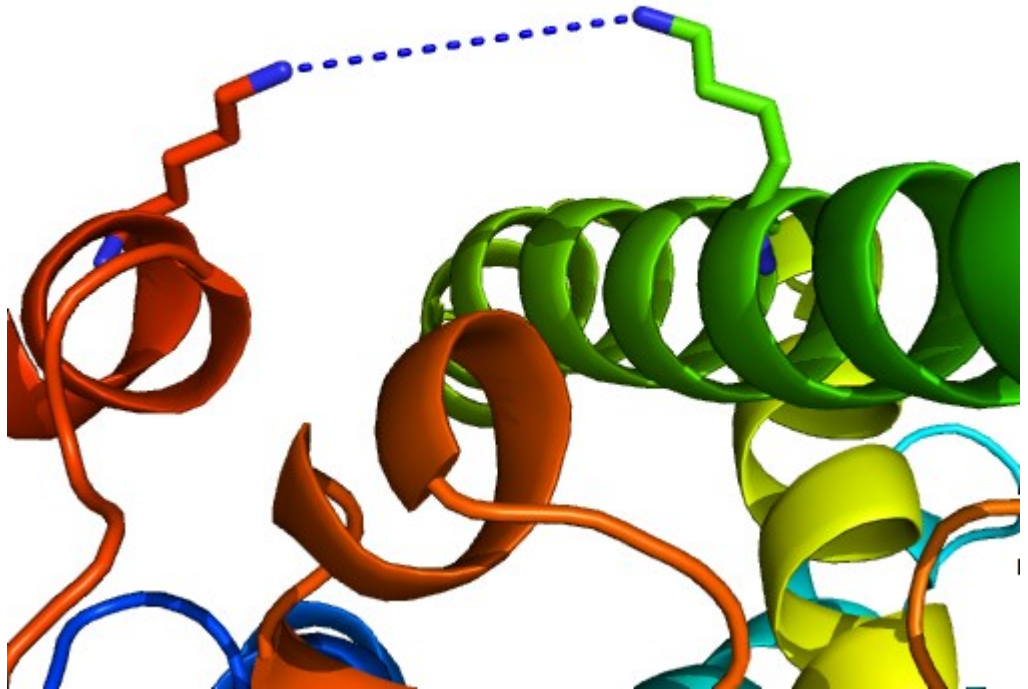


- Chemical crosslinking distances
- DEER EPR distances
- NMR NOE distances

Crosslinks – direct distance measurement



Disuccinimidyl suberate (DSS)

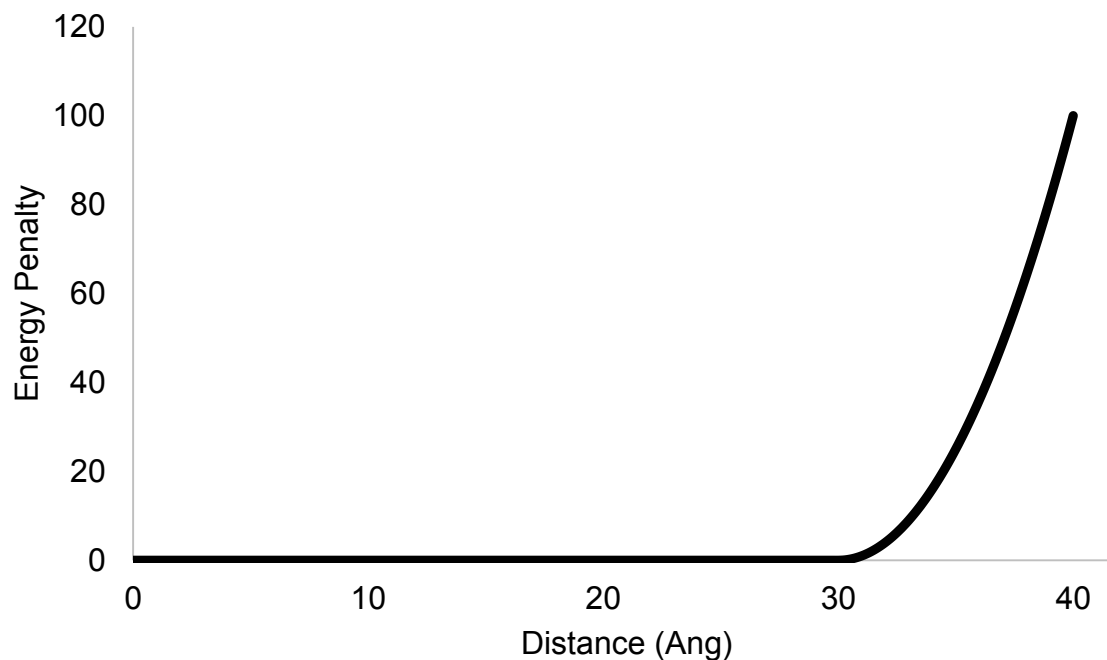


Kahraman et al. PLOS One, 2013, 8(9) e73411
rosetta.demos/protocol_capture/xl_driven_protein_docking/

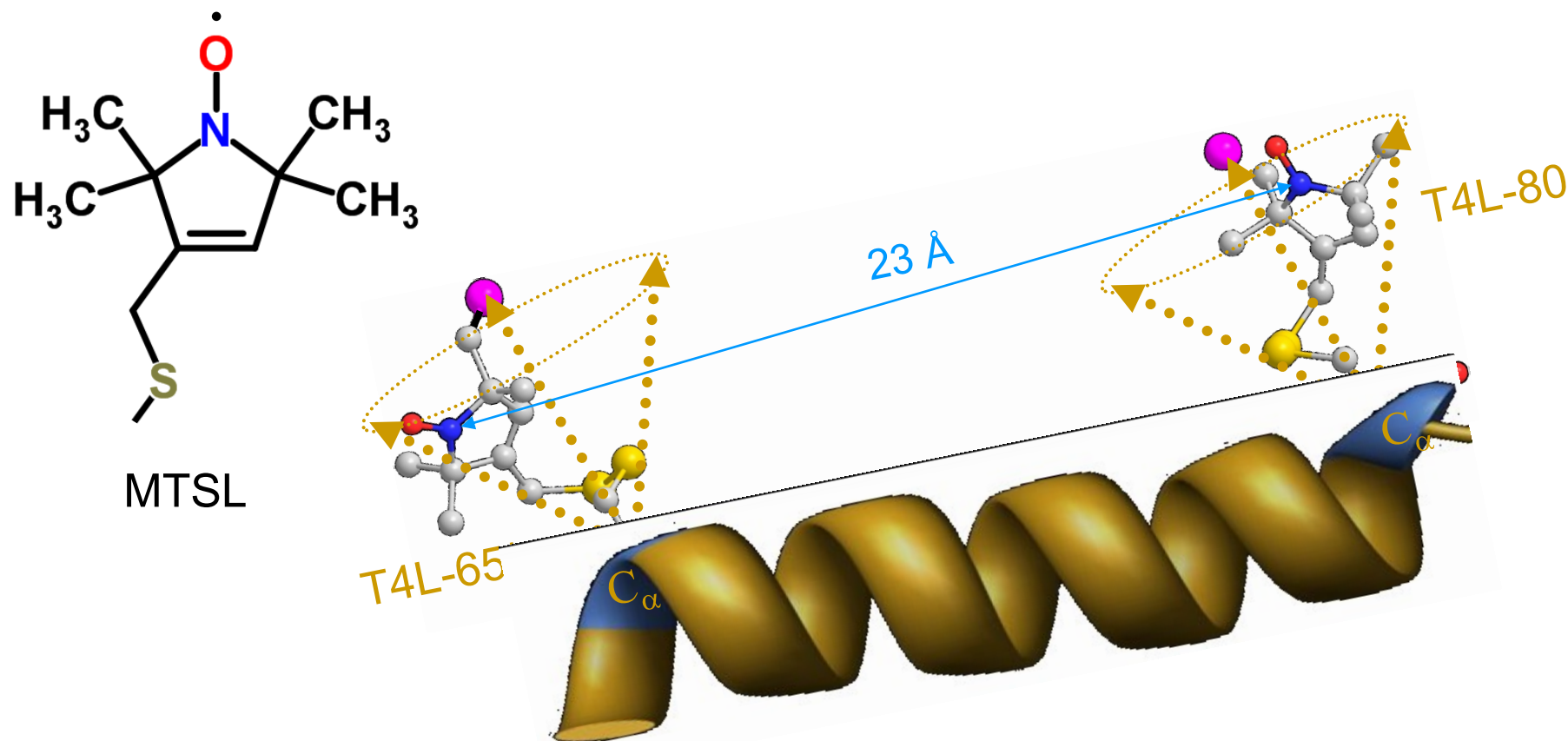
Chemical crosslinking gives distance restraint information



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

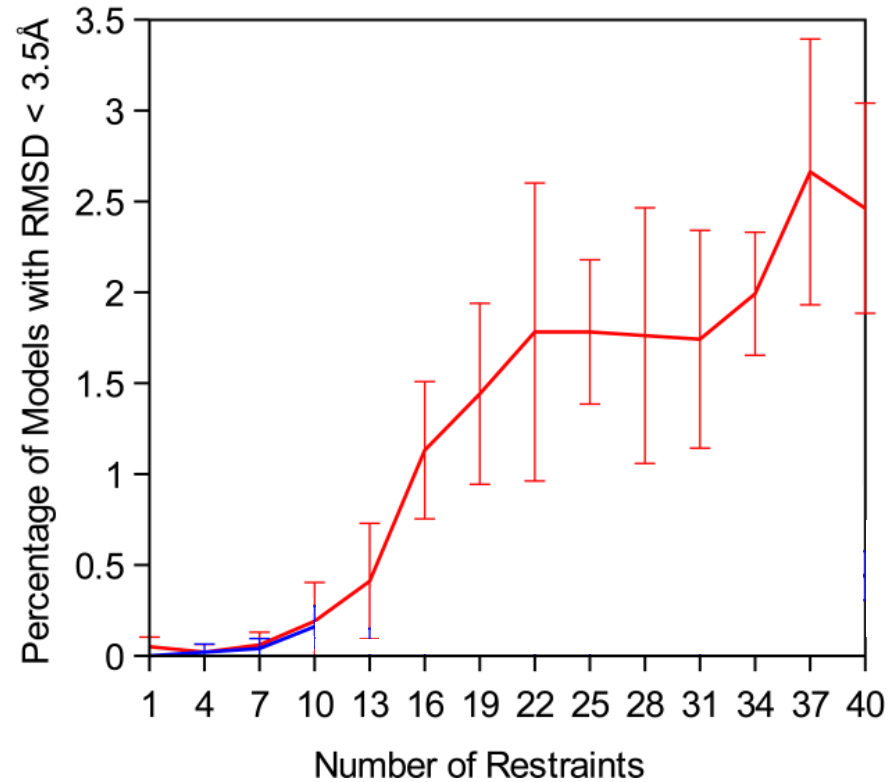
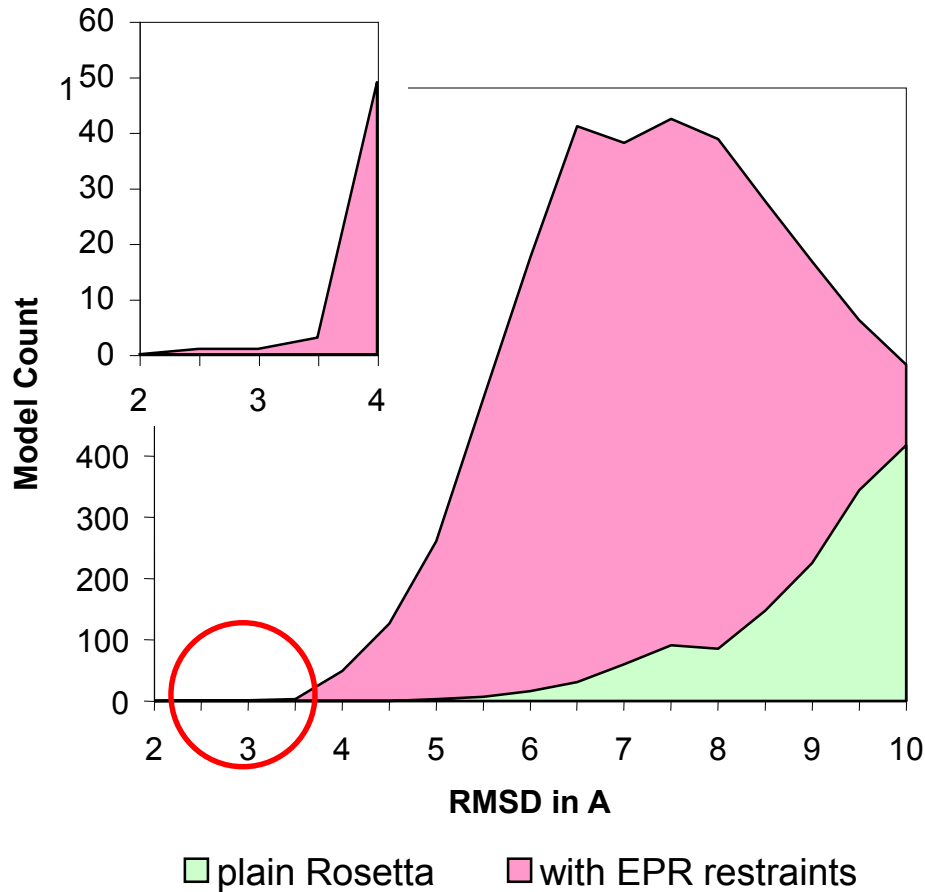


Double Electron-Electron Resonance (DEER) also gives distance information



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

Experimental data influences sampling and model quality





NMR - NOE distance constraints

Simple AtomPair constraints:

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]

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

Other examples of incorporating experimental data

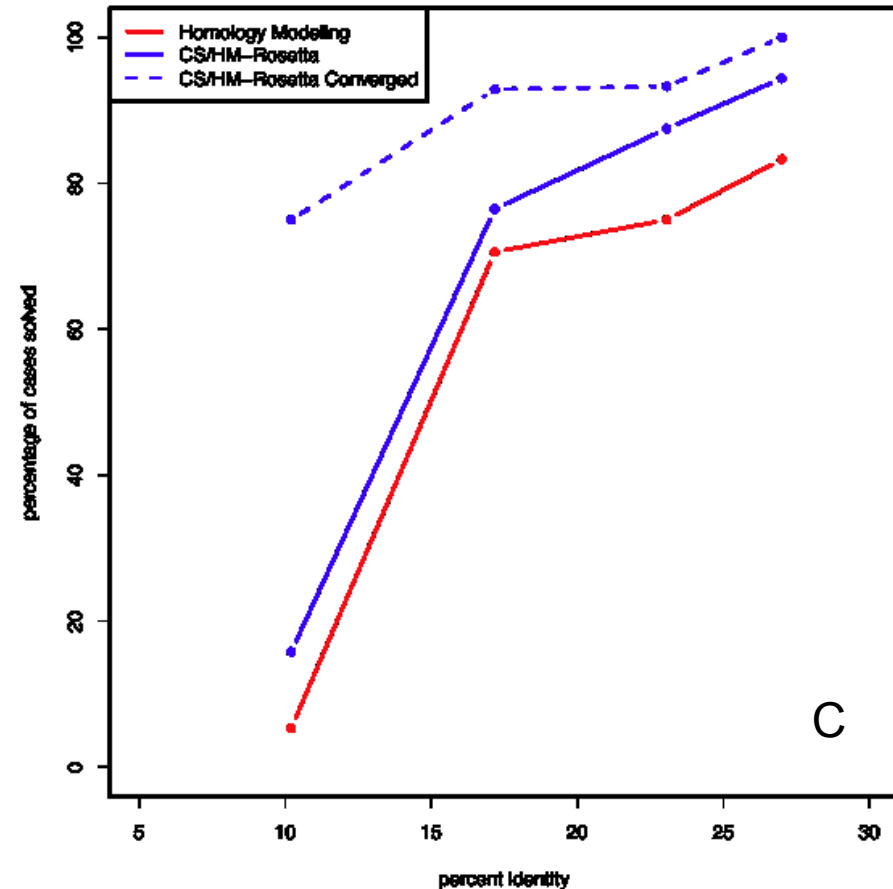
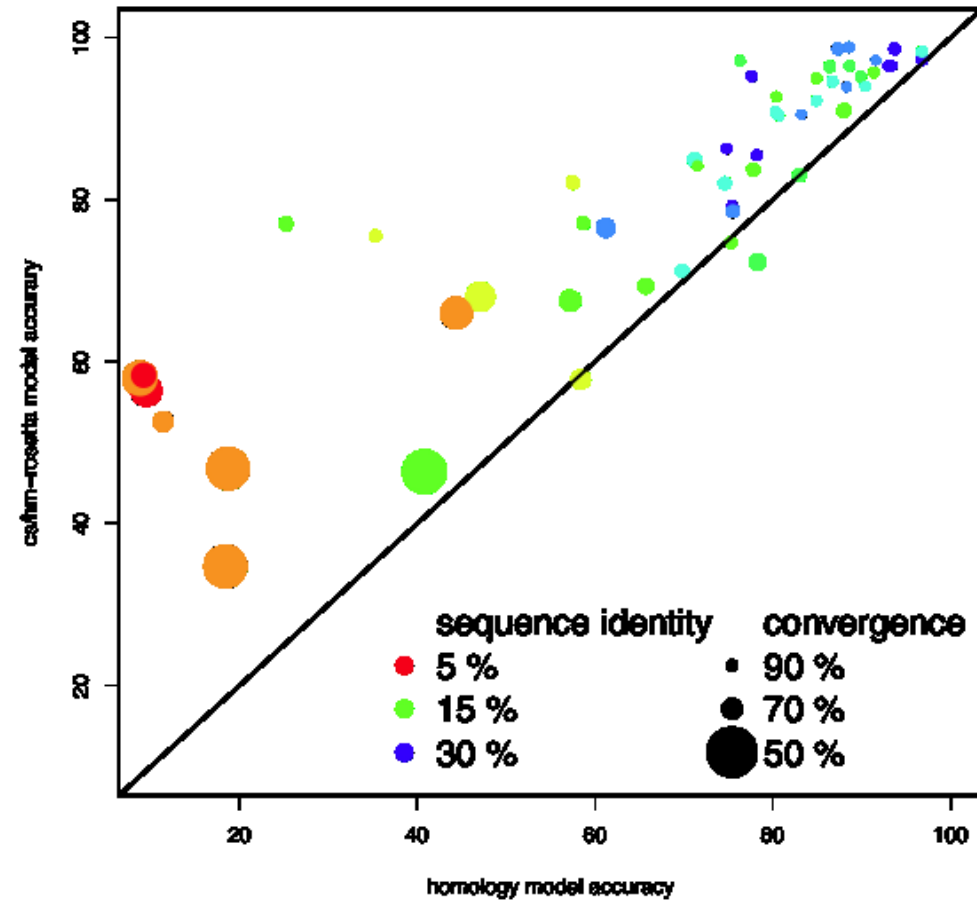


- Chemical Shift Homology Modeling (CS-HM Rosetta)
- Chemical shifts influencing fragment selection
- RDC restraints
- Electron density restraints

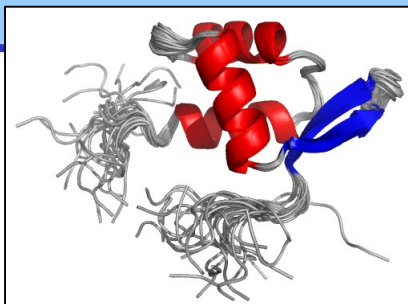
CS-HM Rosetta improves the performance of homology modeling



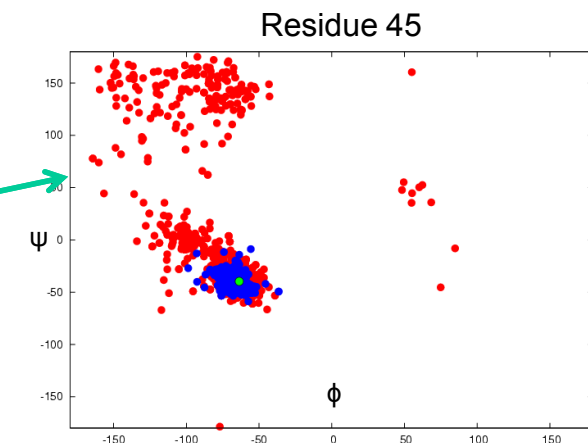
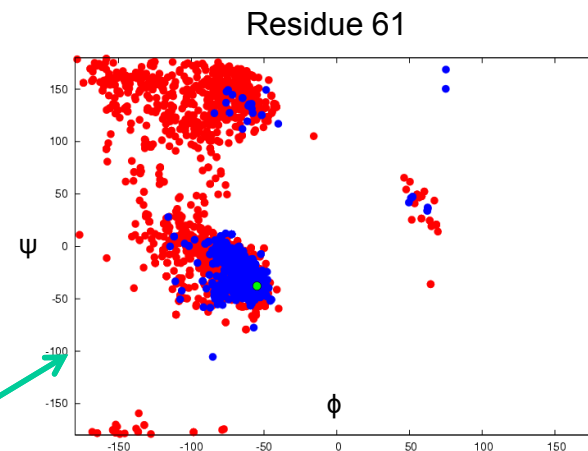
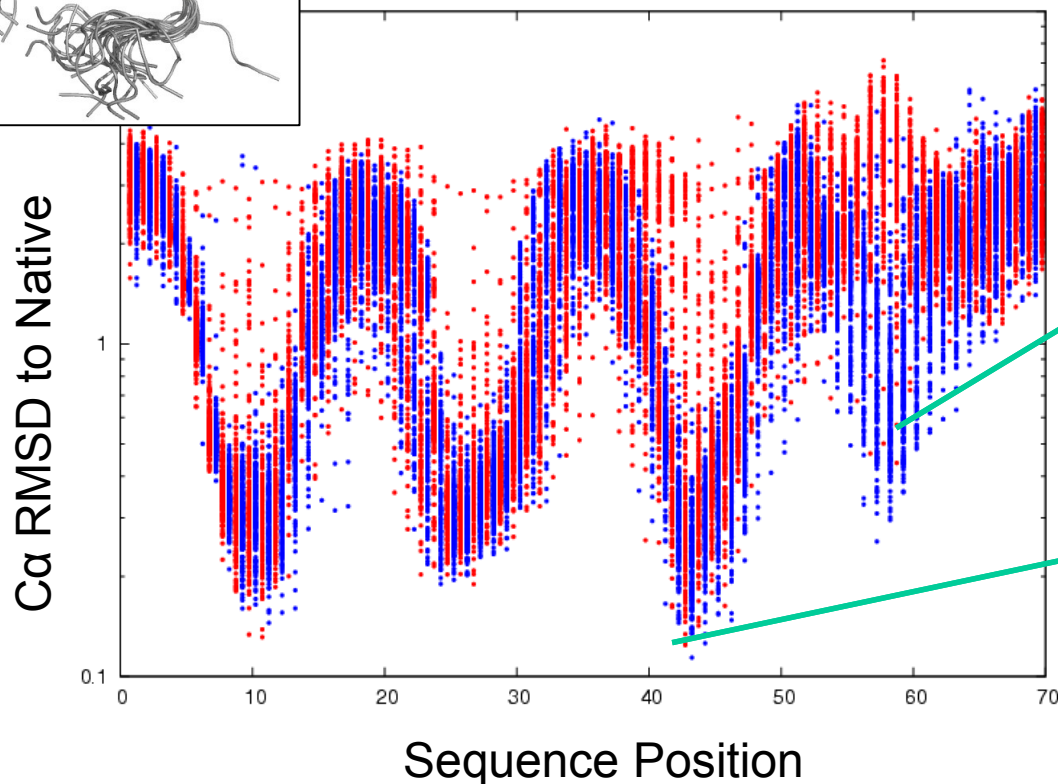
accuracy of cs/hm-rosetta and homology models



Chemical Shifts help fragment picking



Fragment Comparison



Red: Rosetta Fragments, picked by Sequence profile & Secondary Structure Prediction
Blue: CSRosetta Fragments, picked by CS Comparison & Sequence Matching

RDC restraints

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

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

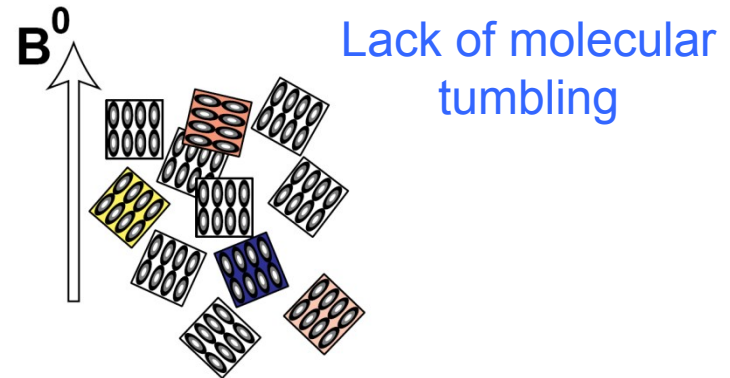
-rdc:fixDa [value]

-rdc:fixR [value]

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

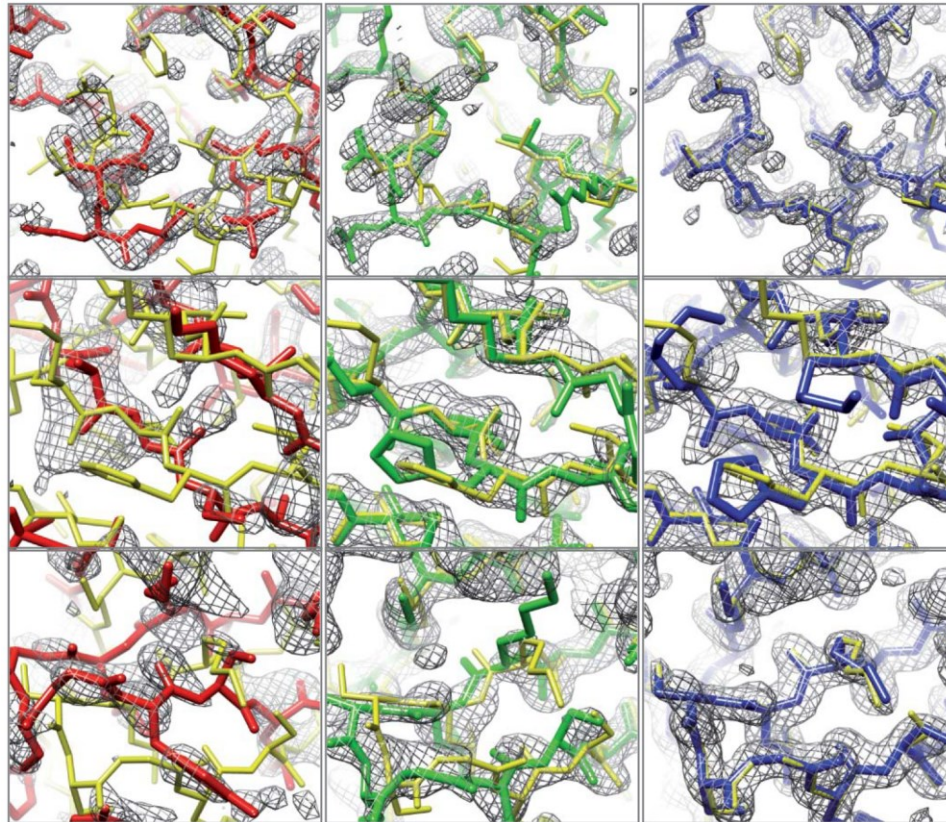
-out:level 999 |grep Qbax || Da || Rh ; exactly the same output as PALES

-rdc:print_rdc_values calc.rdc



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

Electron density restraints (EM, X-ray)



DiMaio F. (2011) Nature

Initial molecular
replacement



Improve Phases

Model rebuilding and
energy optimization

Questions?