Conformational sampling of macrocycles in solution and in the solid state

Paul Hawkins, Ph.D.
Head of Scientific Solutions
Stanislaw Wlodek, Ph.D.
Senior Scientific Developer

The Iron Triangle

Fast, cheap, good: choose two

https://berkonomics.com/?p=2437

The Iron Triangle in conformation generation

Accurate

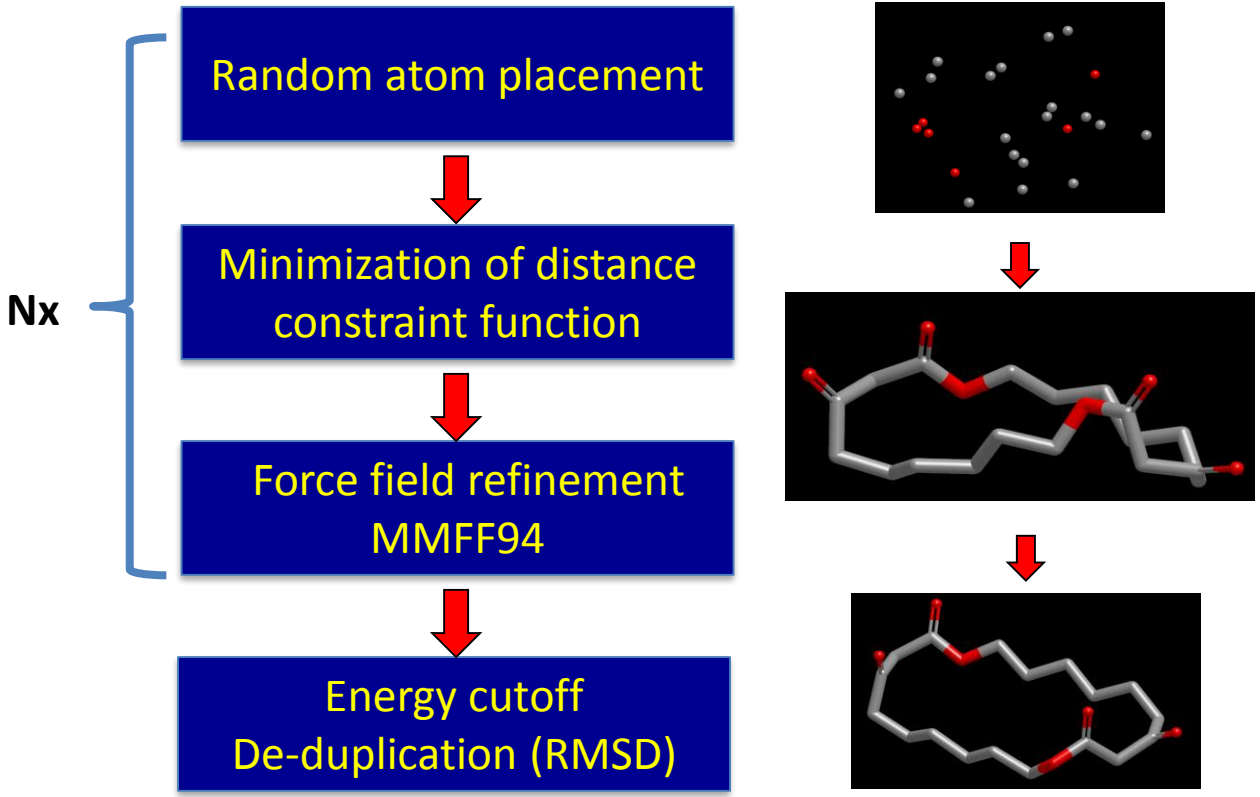
Fast

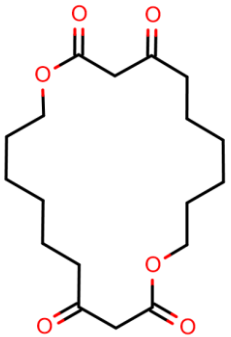
Small ensemble size

Sampling macrocycle conformations

- Molecular Dynamics (MD)
  - Pro: Physics-based model, explicit solvent possible
  - Con: Slow, 3D input required, stochastic
- Torsion sampling
  - Pro: Fast (?), could be deterministic
  - Con: Implicit solvent only, 3D input required
- Distance geometry (DG)
  - Pro: Fast, no 3D input
  - Con: Stochastic, implicit solvent only

OMEGA: Macrocycle sampling by DG





NO 3D structure required.

Spellmeyer *et al., J. Mol. Graph. Model.* **15**, 18 (1997).

Outline

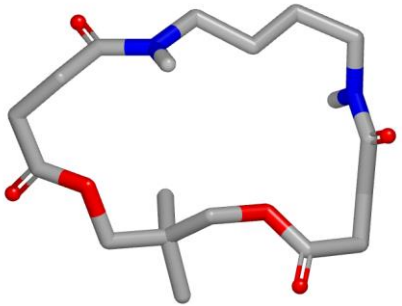
- Validation against the solid-state
- Breaking the Iron Triangle
- Modelling the solution state

Outline

- Validation against the solid-state
  - Reproducing precise, reliable experimental data
- Breaking the Iron Triangle
- Modelling the solution state

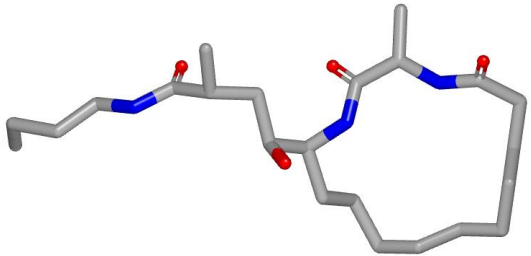
Validating against the solid-state

CSD



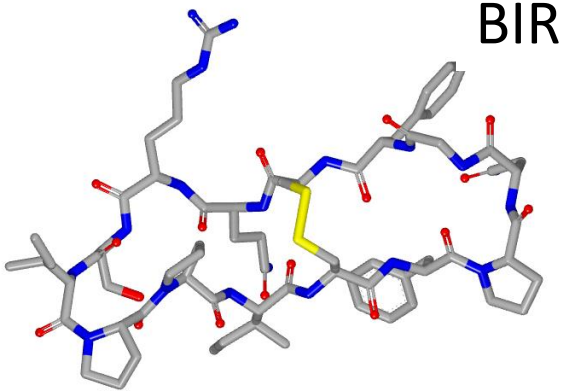
TRAIN against the CSD.
Very reliable conformations.

PDB



TEST against the PDB.
Biologically relevant structures.

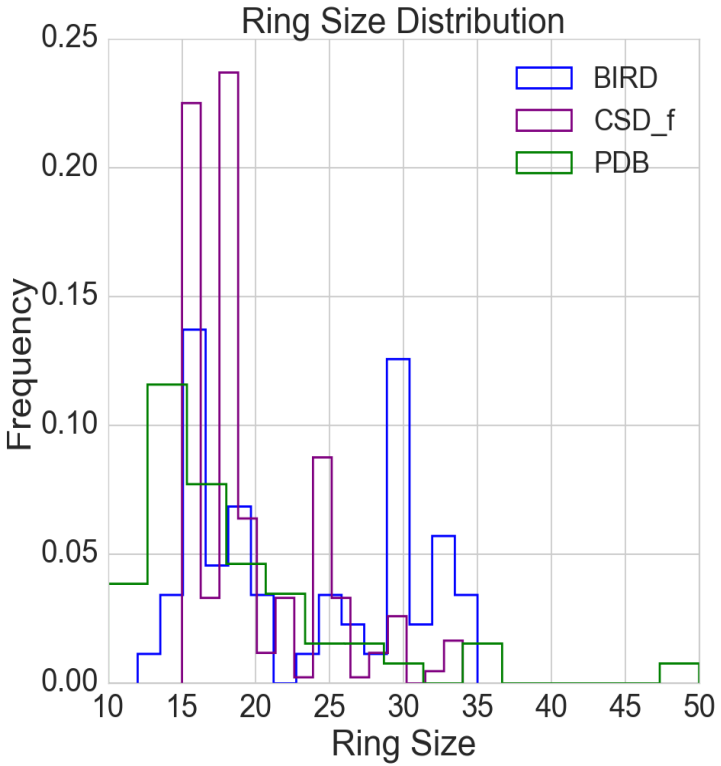
BIRD

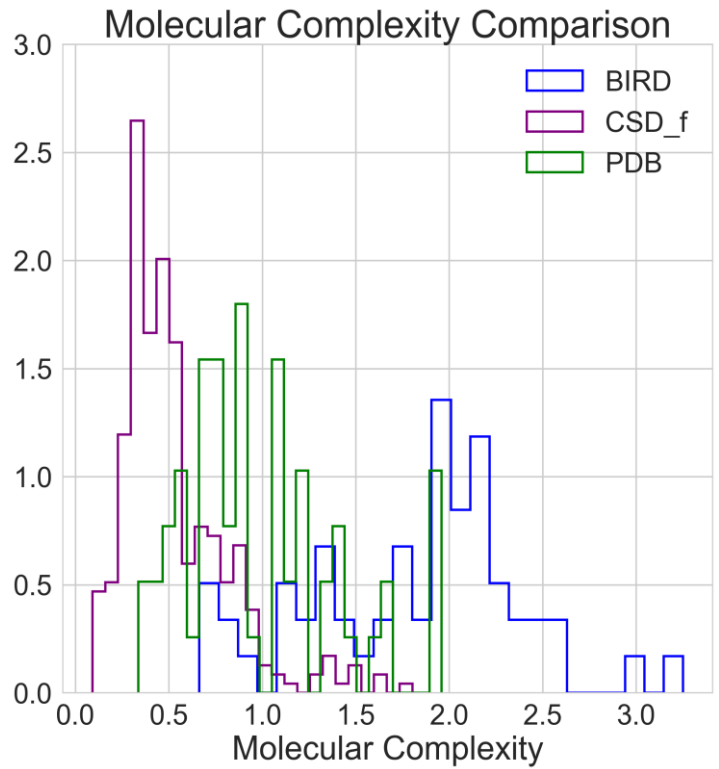


VALIDATE against BIRD.
Very challenging.

BIRD: http://www.wwpdb.org/data/bird

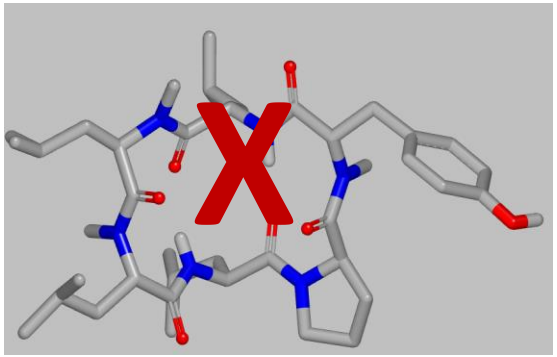
Basic chemical properties



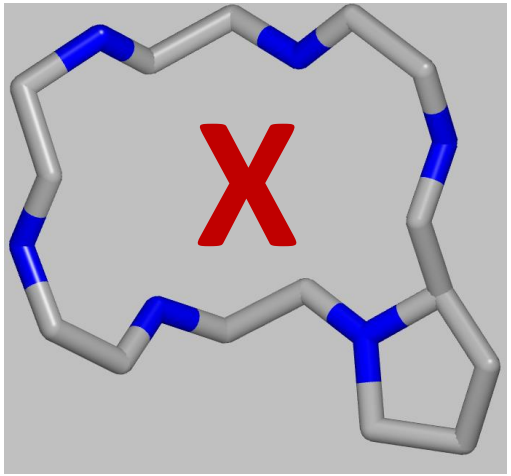


Measuring reproduction performance

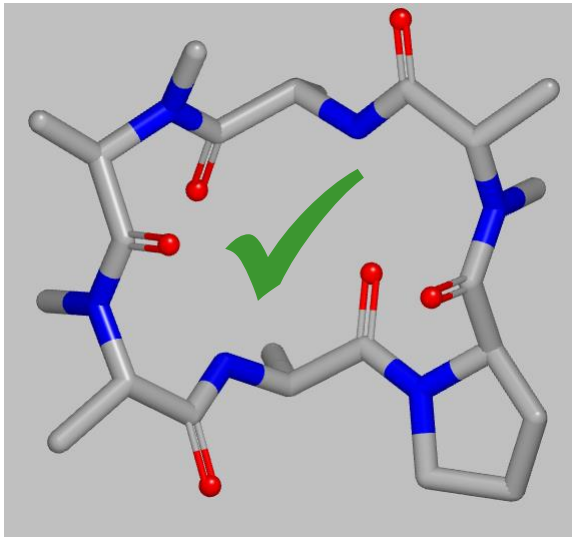
Whole molecule RMSD



“Ring only” RMSD



“Ring + beta atom” RMSD



Effect of:

DG attempts

Solvent model

RMSD

Ewindow

Max confs kept

Solvent modelling

- Poisson-Boltzmann

$$\nabla[\varepsilon(r)\nabla\phi(r)] - \varepsilon(r)\kappa(r)^2 = q\,(r)/k\,T$$

Numerical optimisation

Null model: Coulomb, ε = 1 (vacuum)

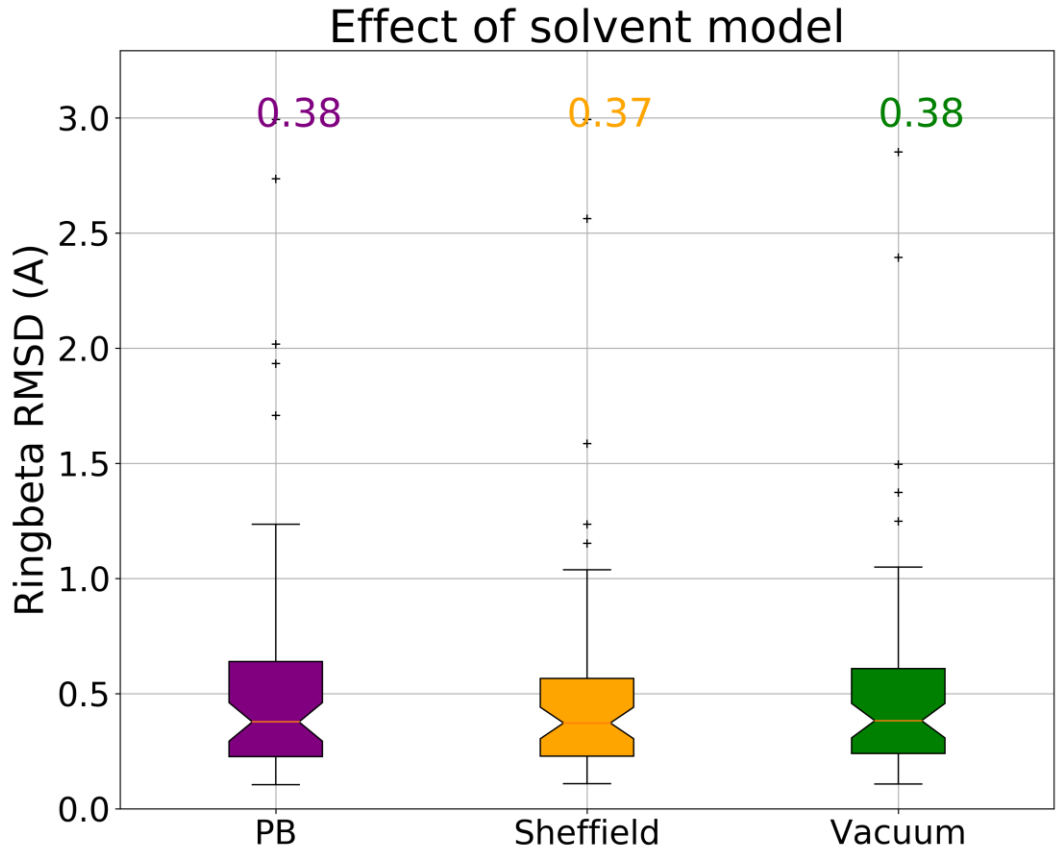
- Sheffield

$$E_{IJ}^{\text{pair}} = -\frac{f_\epsilon}{4\pi\epsilon_0}\frac{Q_I Q_J}{\sqrt{\sigma_I^{\text{B}}\sigma_J^{\text{B}}\text{e}^{-cR_{IJ}^2/\sigma_I^{\text{B}}\sigma_J^{\text{B}}} + R_{IJ}^2}}$$

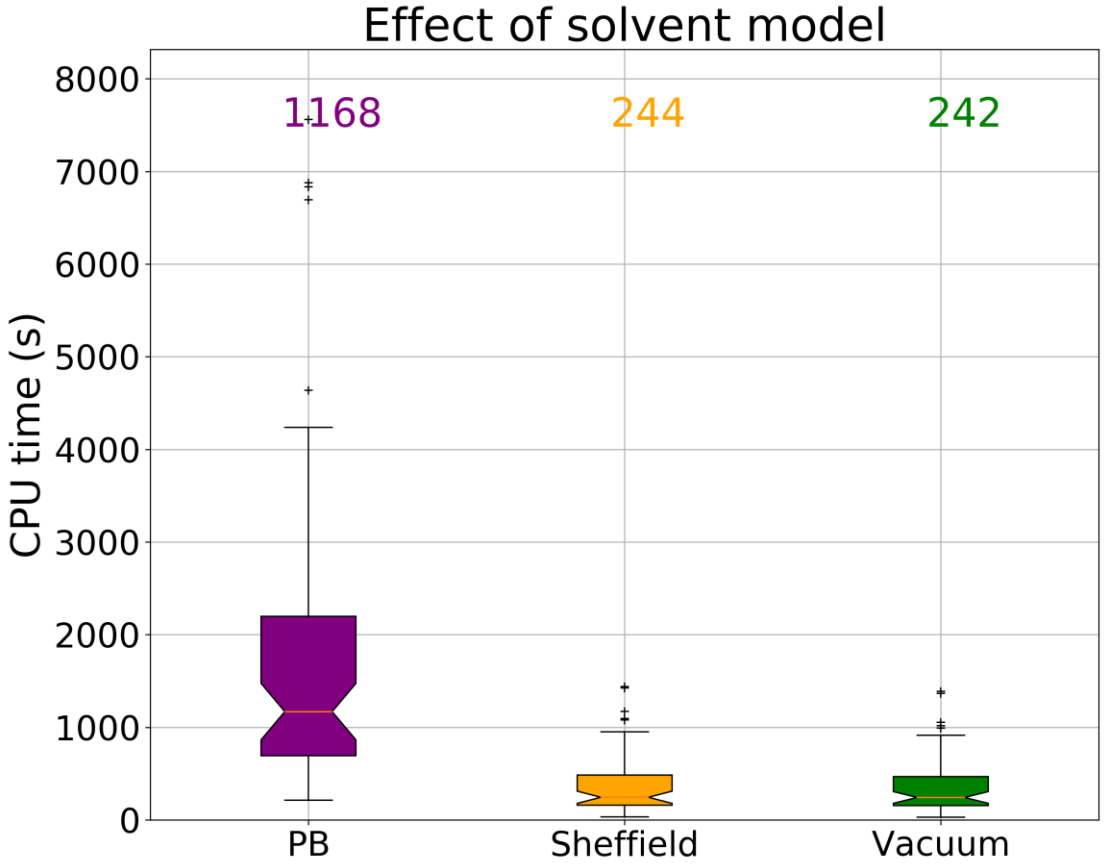
Analytical optimisation

Grant *et al.*, *Chem. Phys. Lett.*, **443**, 163 (2007).

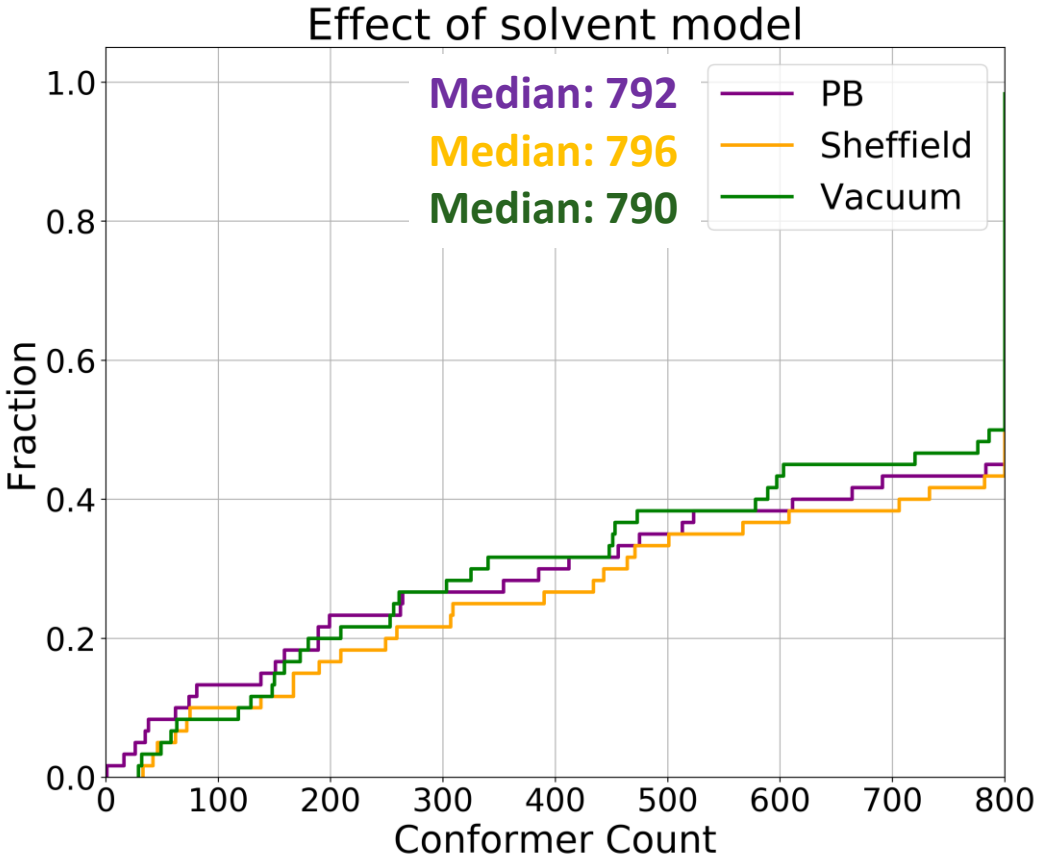
Vacuum v. Sheffield v. PB: Good?



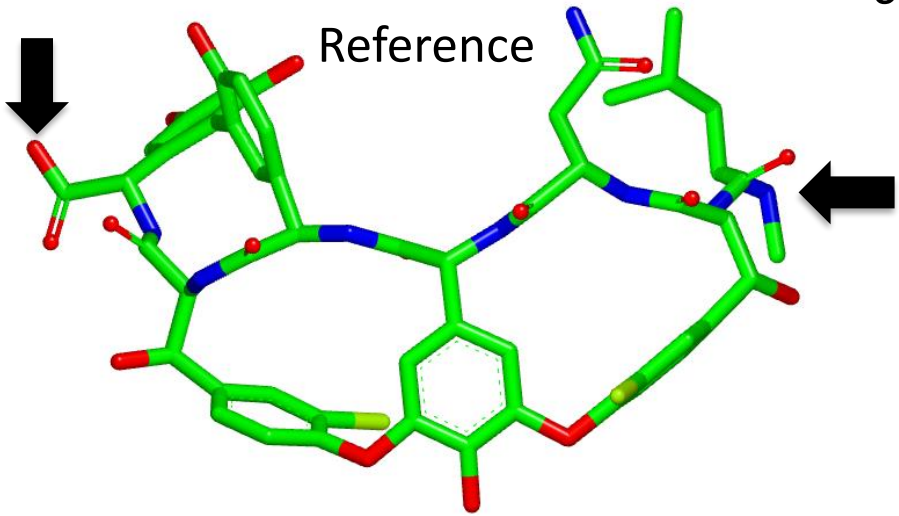
Vacuum v. Sheffield v. PB: Fast?



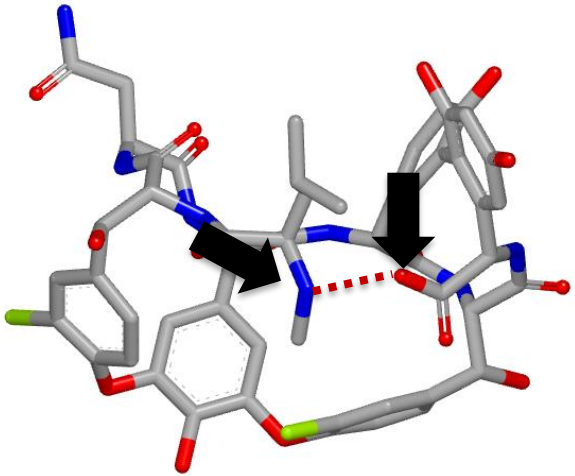
Vacuum v. Sheffield v. PB: Cheap?

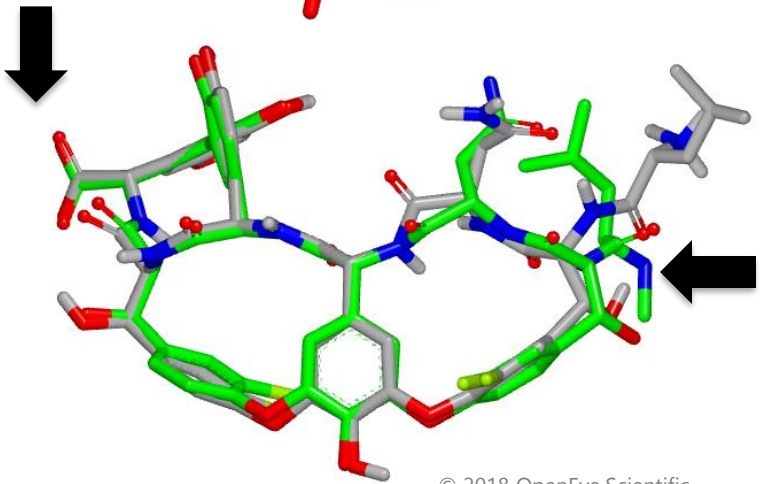


Does the null model win? The case of 1HHY



Vacuum
RMSD 2.8Å
63 conformers





Sheffield
RMSD 1.7Å
280 confs

Sheffield solvation: Fast & cheap & good

$$E_{IJ}^{\text{pair}} = -\frac{f_\epsilon}{4\pi\epsilon_0}\frac{Q_I Q_J}{\sqrt{\sigma_I^{\text{B}}\sigma_J^{\text{B}}\text{e}^{-cR_{IJ}^2/\sigma_I^{\text{B}}\sigma_J^{\text{B}}} + R_{IJ}^2}}$$

Parameter selection: A balancing act

Parameters after training:

DG attempts = 2000
Solvent model = Sheffield
RMSD = 0.5Å
Ewindow = 20 kcal/mol
# confs kept = 400

Outline

- Validation against the solid-state
- Breaking the Iron Triangle
- Modelling the solution state

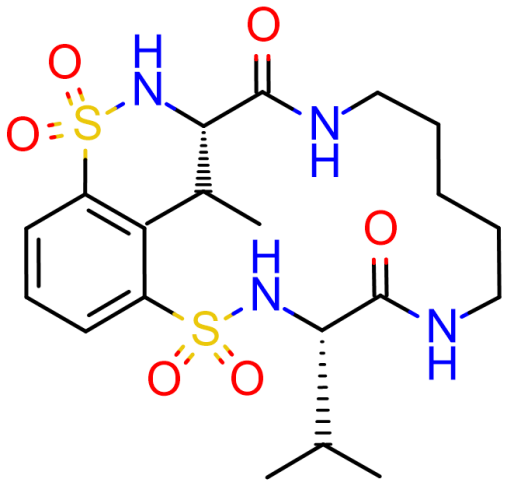
Multiple method comparison

Improving Accuracy, Diversity, and Speed with Prime Macrocycle Conformational Sampling

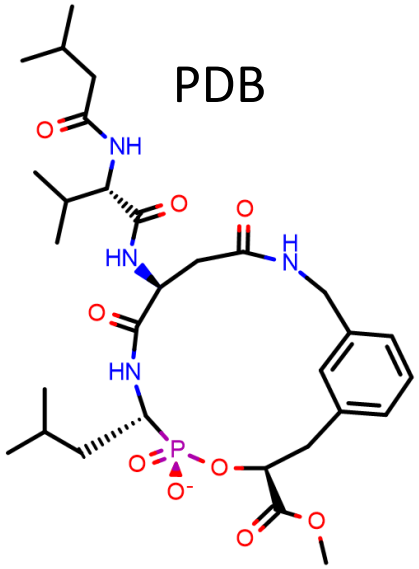
Dan Sindhikara,*[,‡] Steven A. Spronk,[§] Tyler Day,[‡] Ken Borrelli,[‡] Daniel L. Cheney,[§] and Shana L. Posy[§]

208 macrocycles
130 CSD, 60 PDB, 18 BIRD

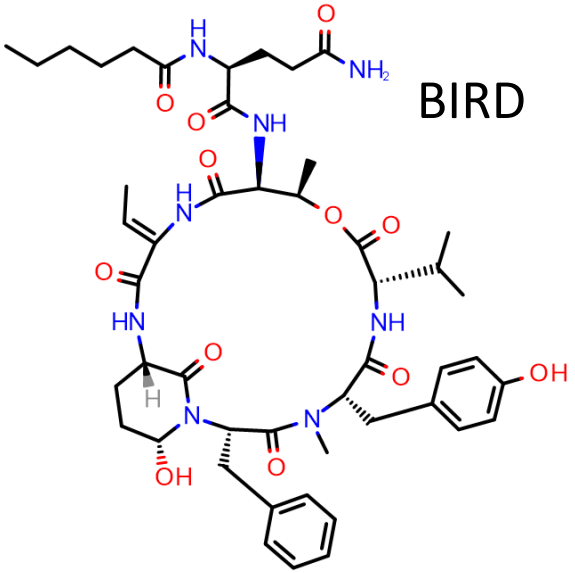
CSD



PDB



BIRD

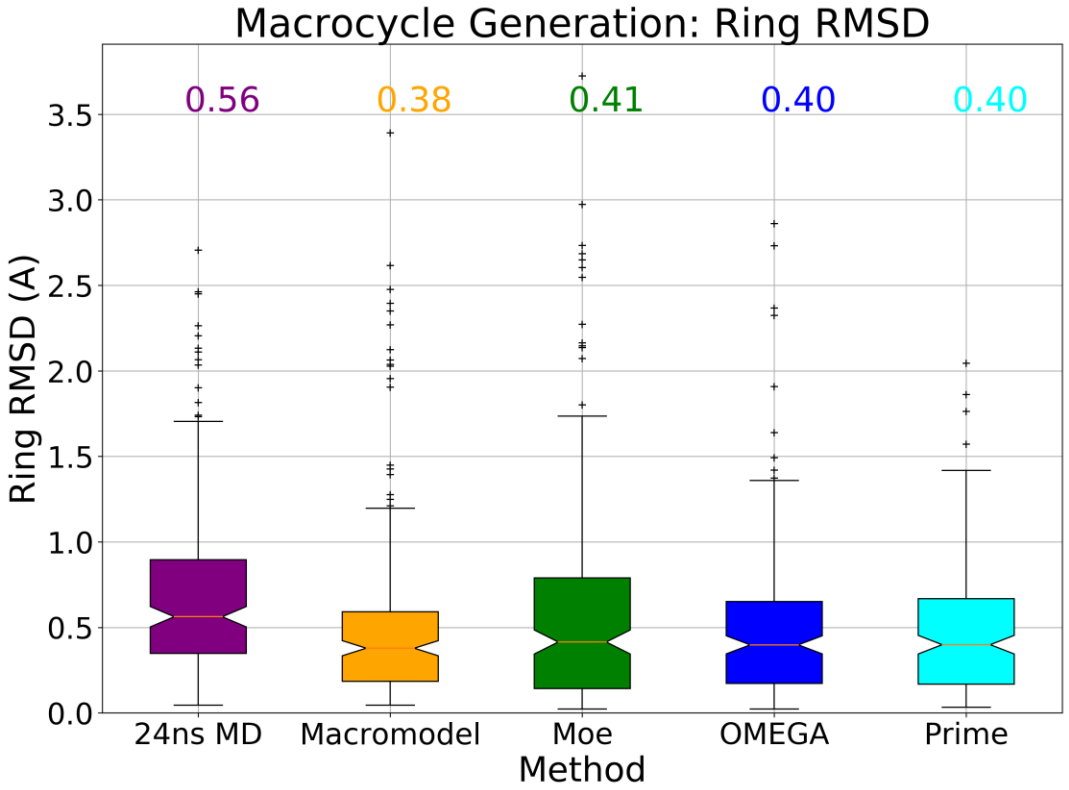


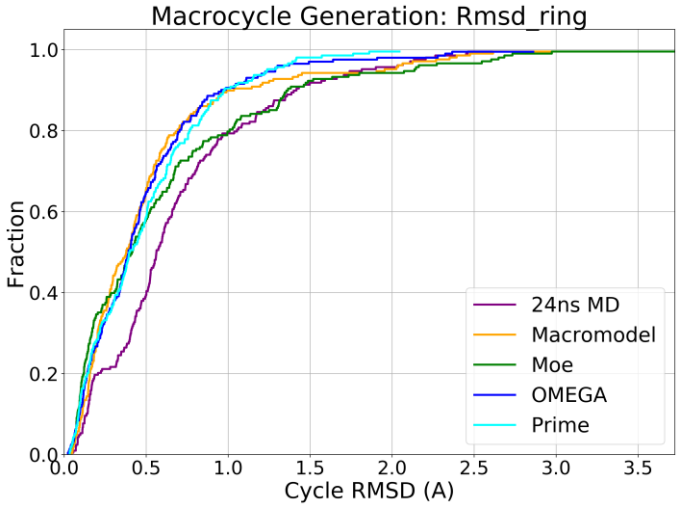
Sindhikara *et al.*, *J. Chem. Inf. Model.*, **57**, 1881 (2017).

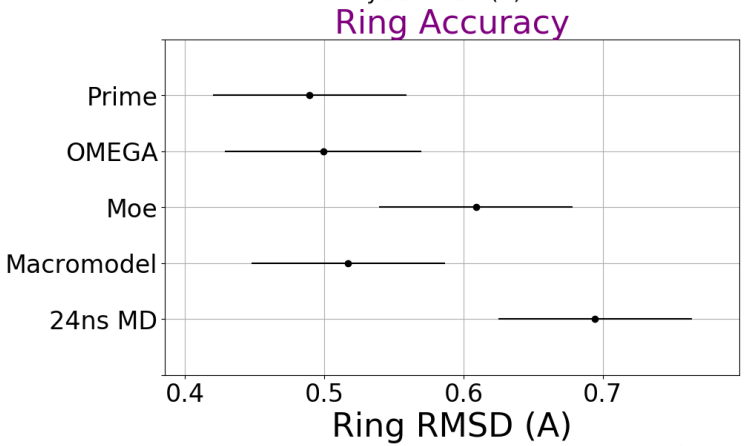
Methods compared

Method	Algorithm	Forcefield	Solvent	Requires 3D?
Macromodel	LowMode MD	OPLS05	GB/SA	YES
MD	MD	OPLS 2.1	Explicit	YES
Moe	LowMode MD	AMBER10	SRF	YES
Prime	Torsion sampling	OPLS05	Vacuum	YES
OMEGA	Distance geometry	MMFF94	Sheffield	NO

Accuracy of reproduction







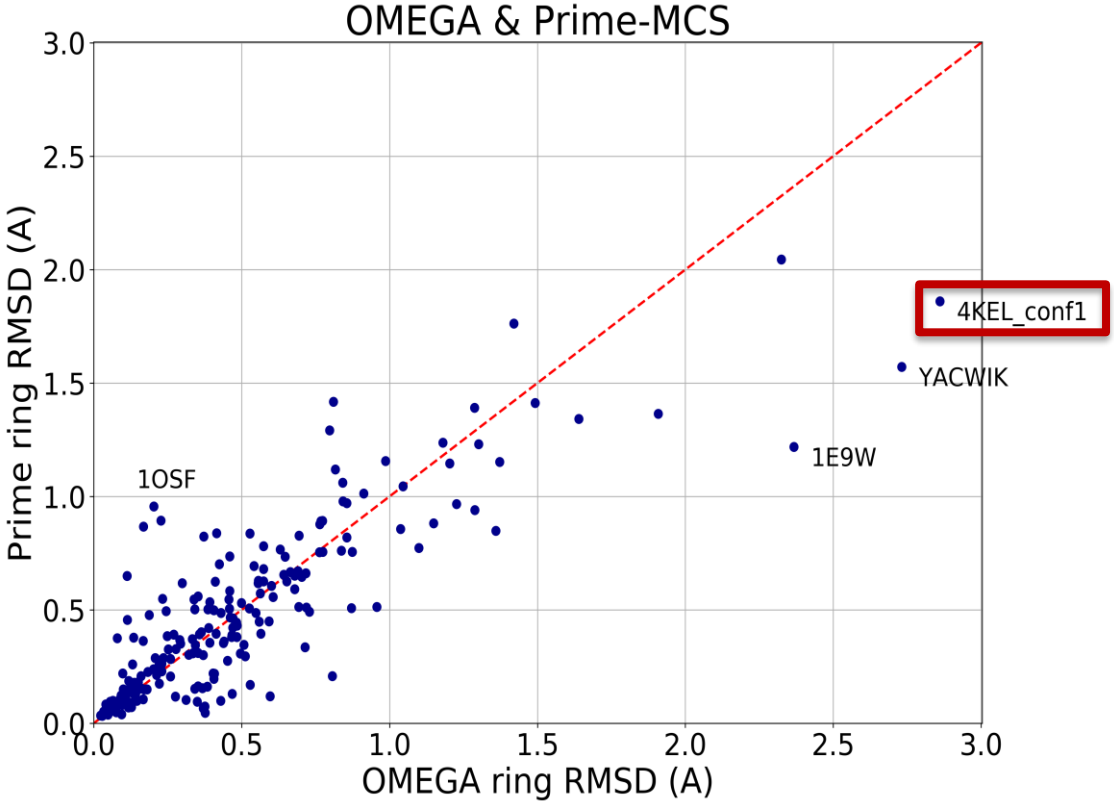
MD does not sample near the solid state well

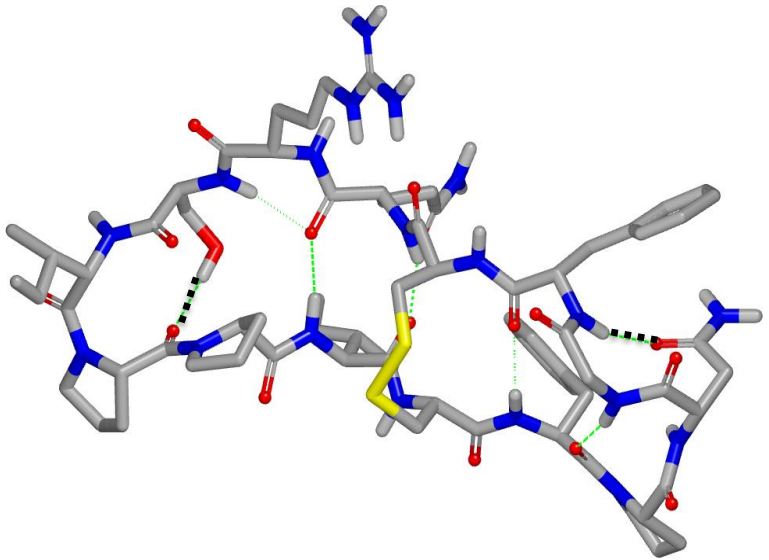
Method 1	Method 2	P < 0.05	Effect size
24ns MD	Macromodel	TRUE	0.33
24ns MD	Moe	FALSE	0.15
24ns MD	OMEGA	TRUE	0.42
24ns MD	Prime	TRUE	0.44
Macromodel	Moe	FALSE	0.16
Macromodel	OMEGA	FALSE	0.06
Macromodel	Prime	FALSE	0.06
Moe	OMEGA	FALSE	0.22
Moe	Prime	FALSE	0.23
OMEGA	Prime	FALSE	0.0

P < 0.05:
Is the difference consistent?

Effect size:
Does the difference make a difference?

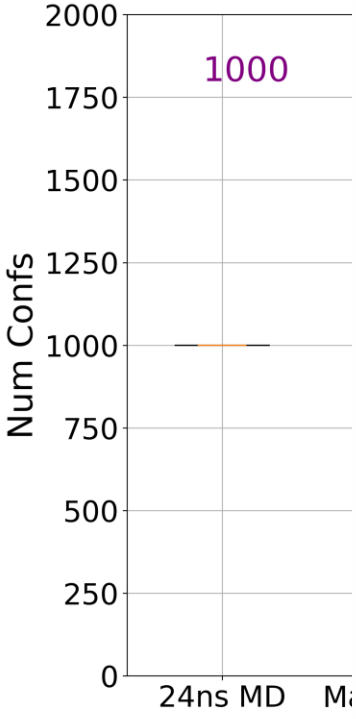
Intra-molecular H-bonds are difficult

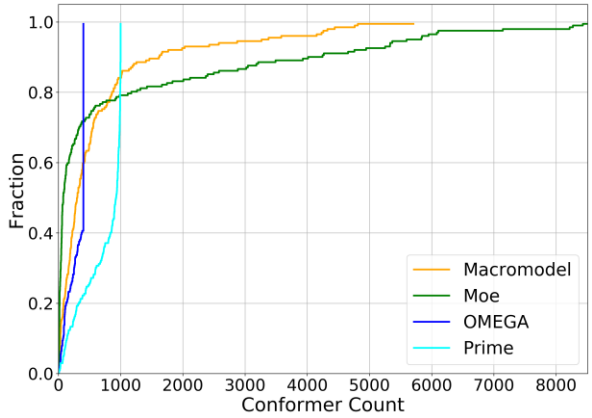


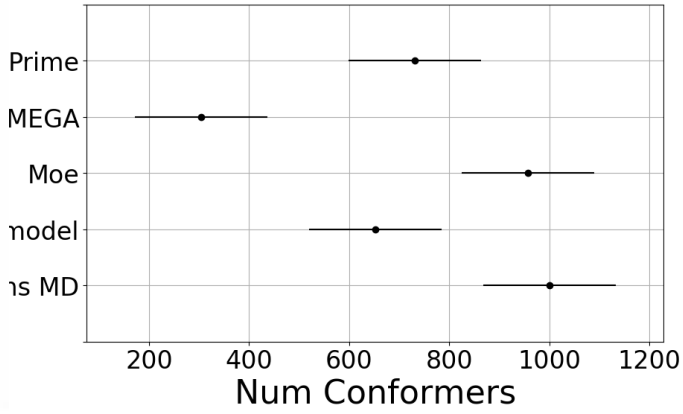


OMEGA is accurate

OMEGA is cheap







OMEGA is fast

Breaking the Iron Triangle: Summary

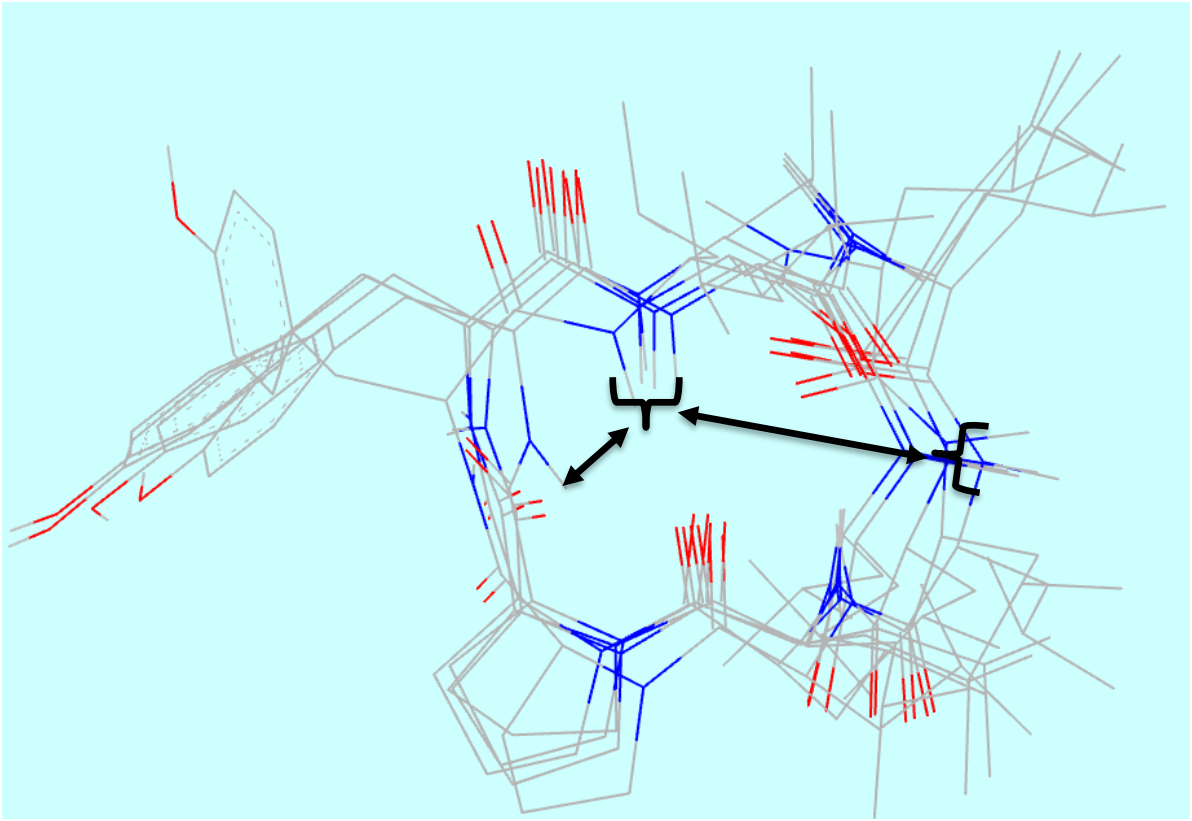
- Training and testing on carefully chosen datasets finds broadly transferable parameters
  - CSD **<->** PDB **->** BIRD (> 450 molecules)

- Comparison to other methods is important
  - OMEGA performs well
  - Informative failure cases found

Outline

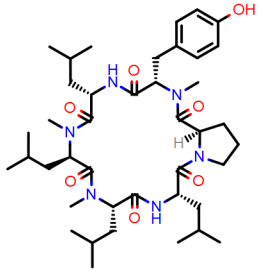
- Validation against the solid-state
- Breaking the Iron Triangle
- Modelling the solution state

Structures in solution: NMR

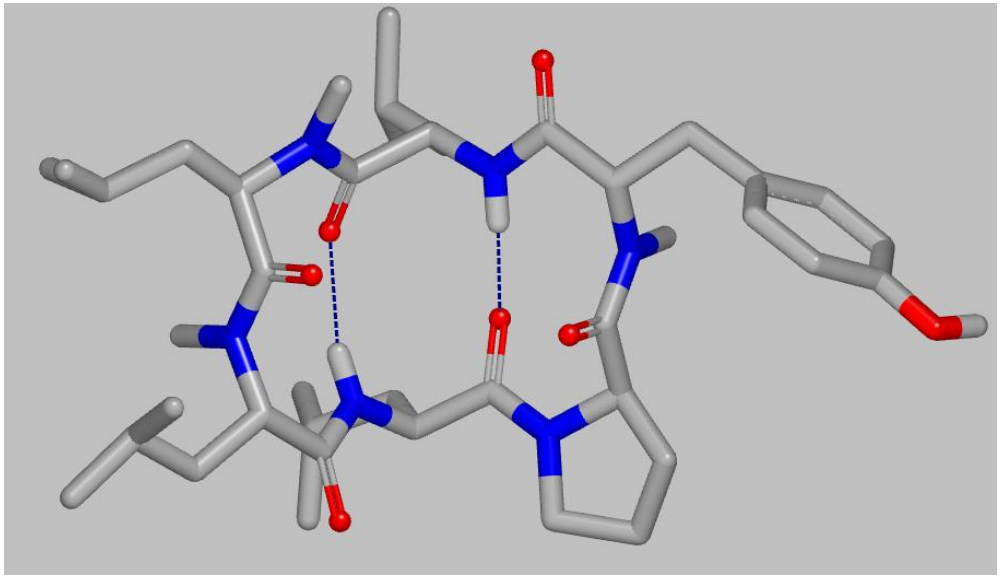


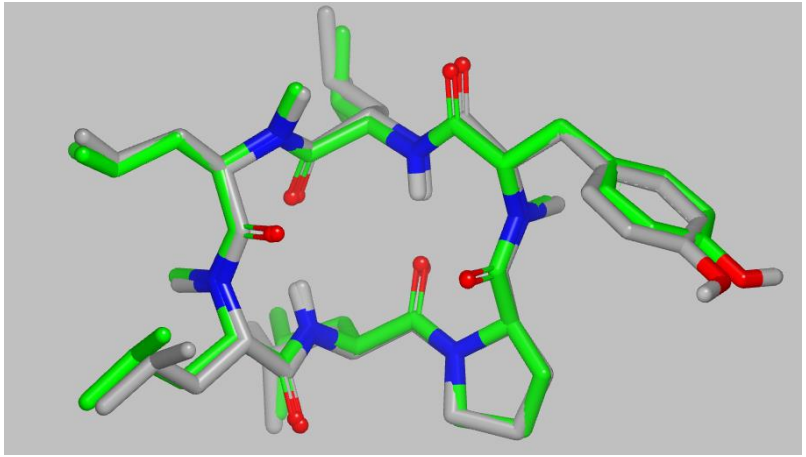
Inter-atomic (proton) distances & *J*-coupling. HIGHLY under-determined.

An easy case: the ‘Lokey peptide’



- H-bonds strongly affect conformation
  - Solid-state conformation easy to reproduce

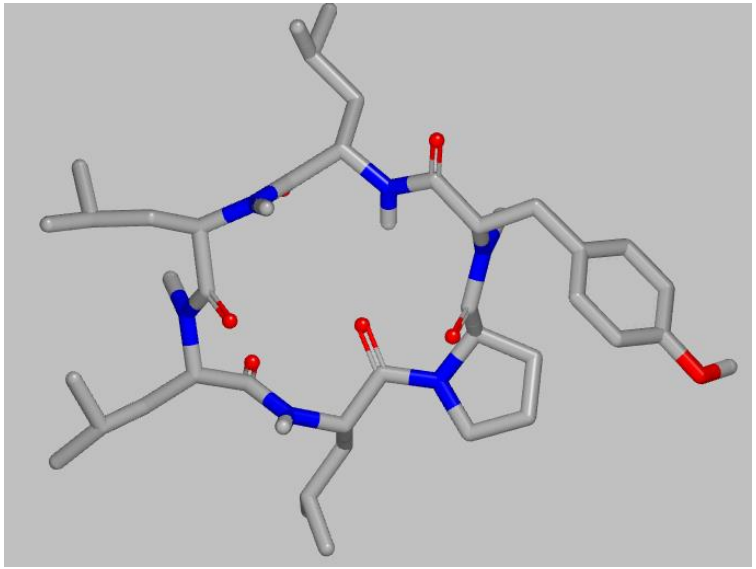




Ring_beta RMSD: 0.29Å

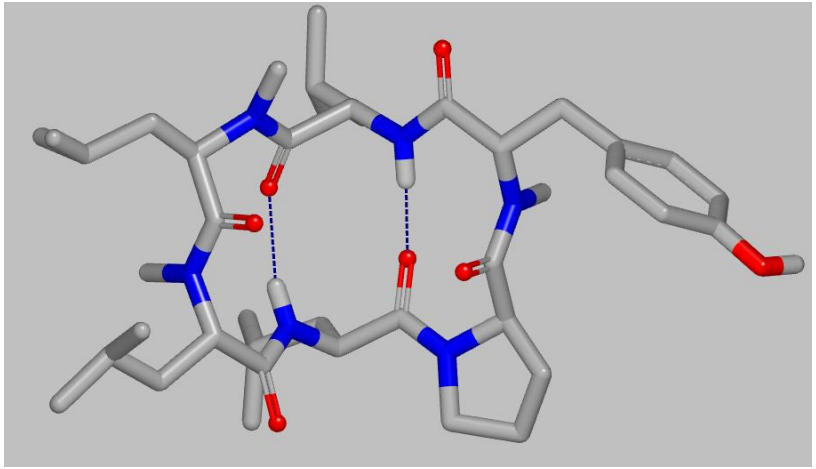
White *et al.*, *Nature Chem. Biol.*, **7**, 801 (2011).

Intra-molecular H-bonds driven by polarity



$CHCl_3$

H_20



Stabilised in HIGH polarity solvents

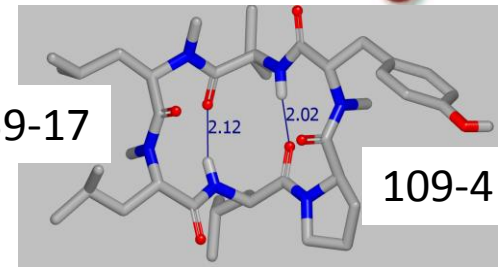
Stabilised in LOW polarity solvents

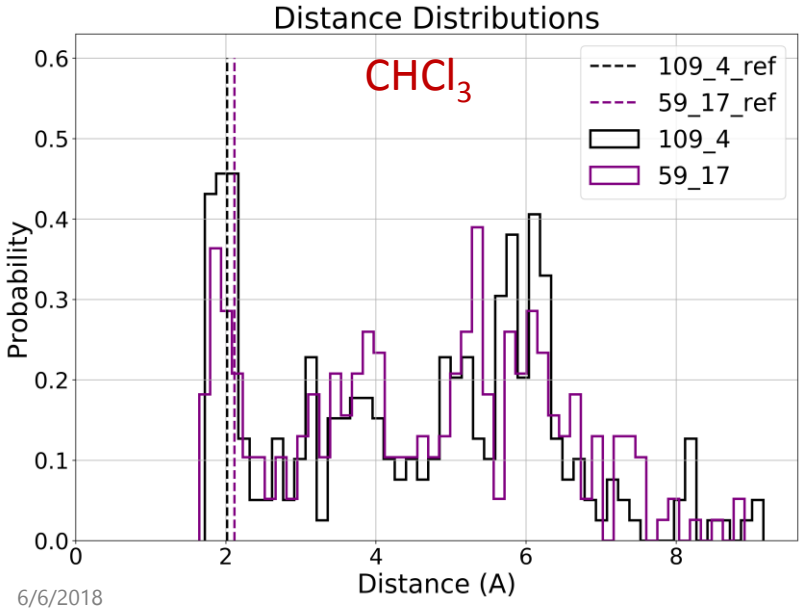
IMHB propensity increases as solvent polarity decreases.

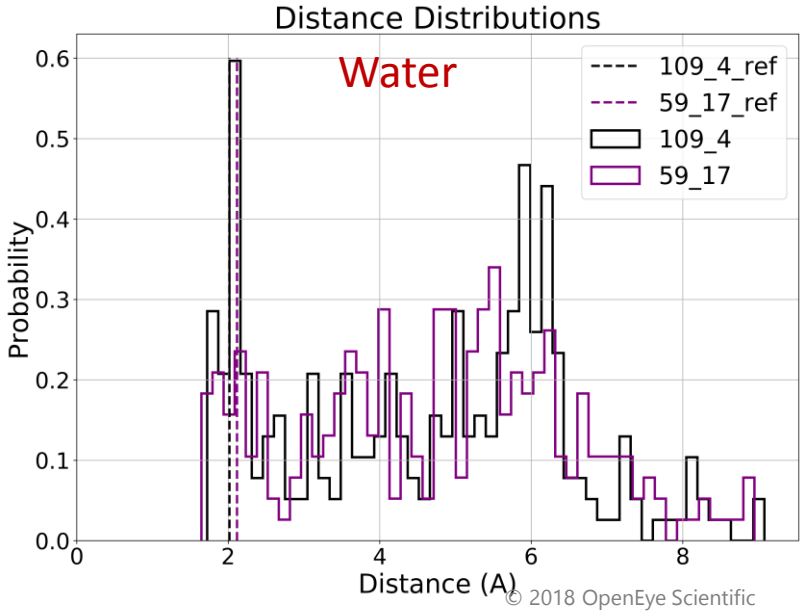
Modelling the solvent

	Water	$CHCl_3$
109-4 < 2.5Å	11%	19%
59-17 < 2.5Å	8%	12%

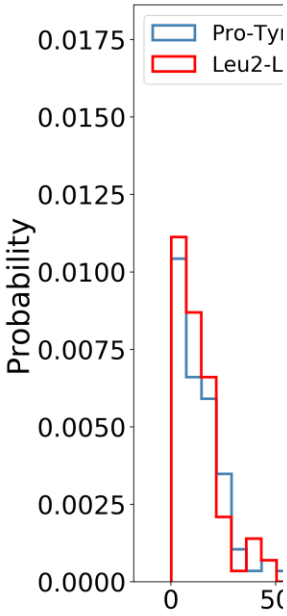
Simulation responds QUALITATIVELY correctly to change in solvent dielectric

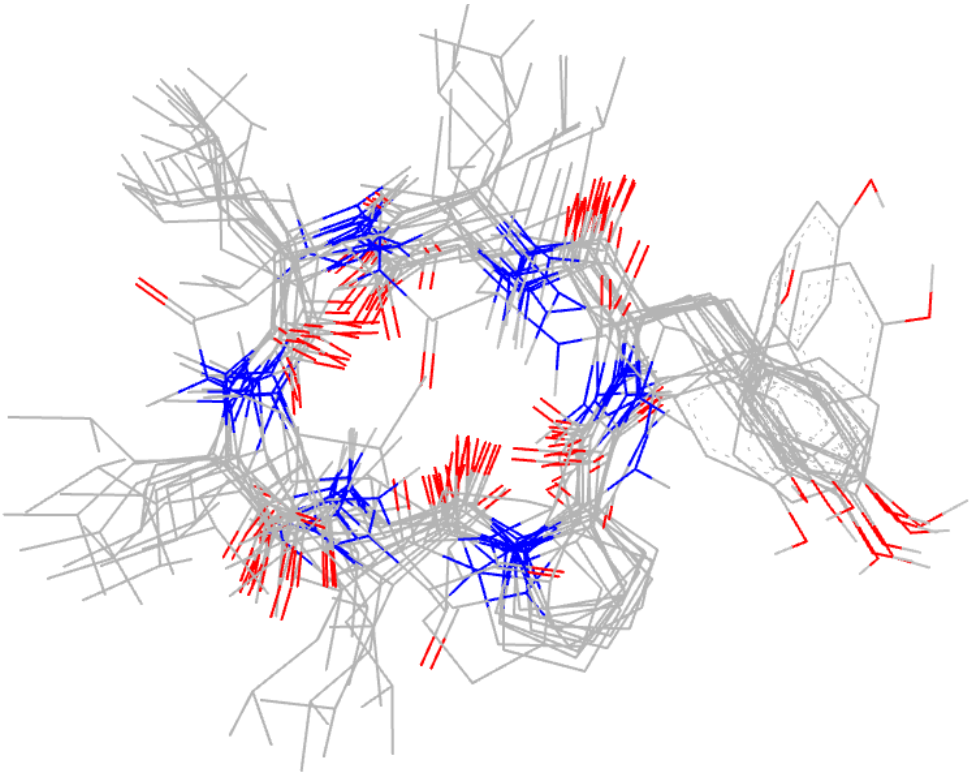


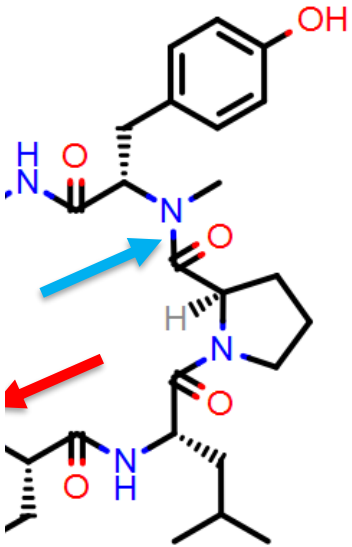




Torsion analysis: Now

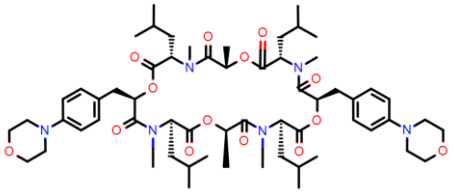




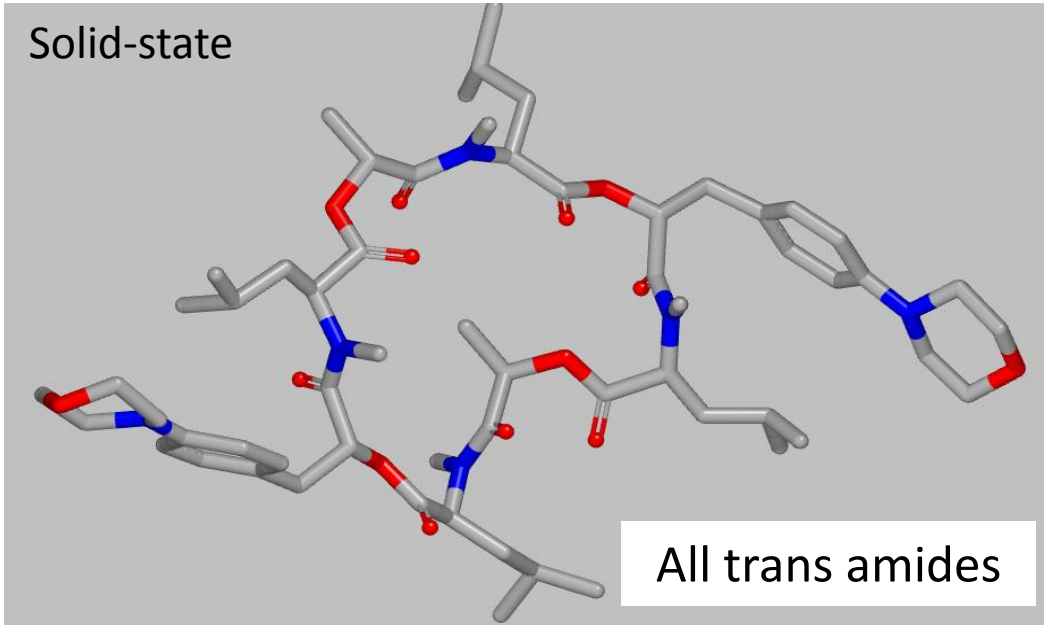


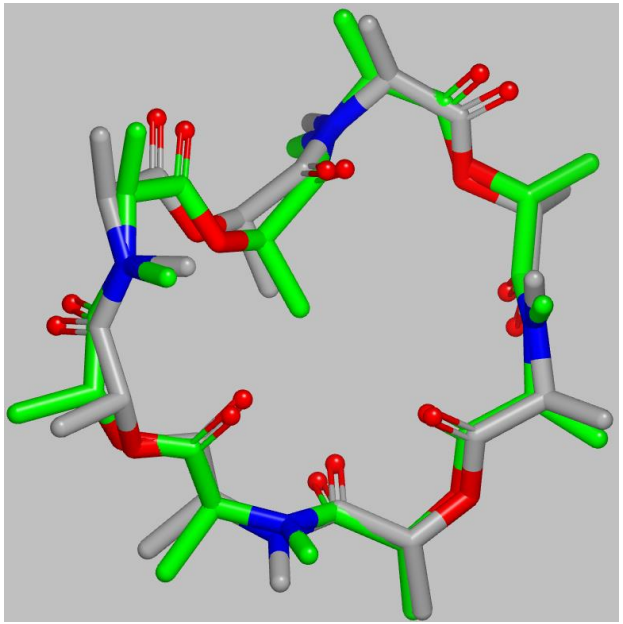
Torsion Analysis Reimagined

A harder case: Emodepside



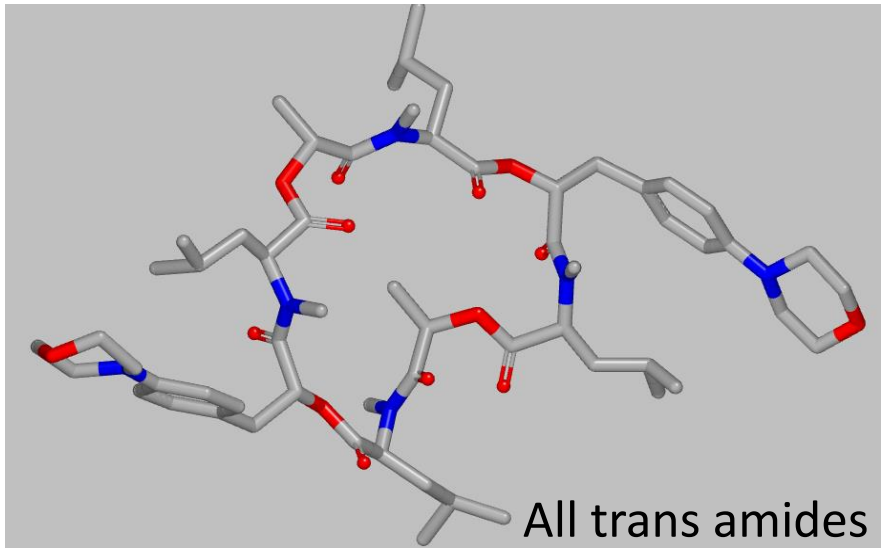
- No IMHBs, all amide N's are capped



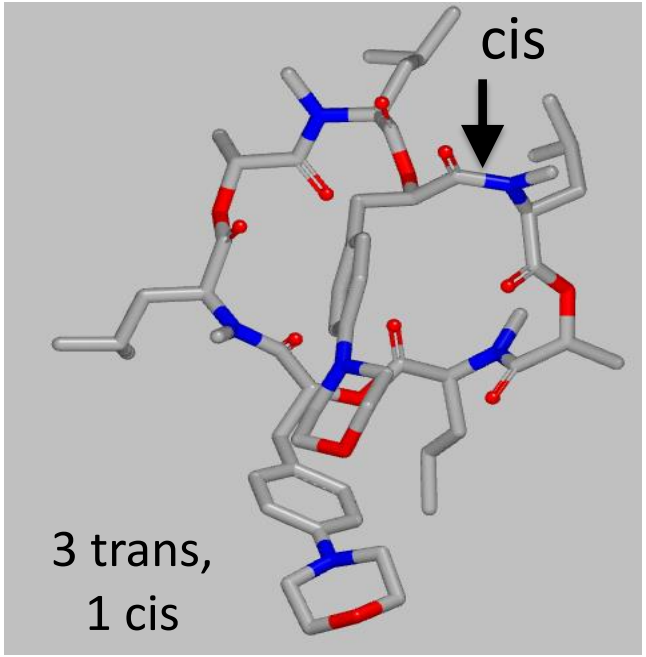


Ring_beta RMSD: 0.43Å

Conformational hetereogeneity in solution

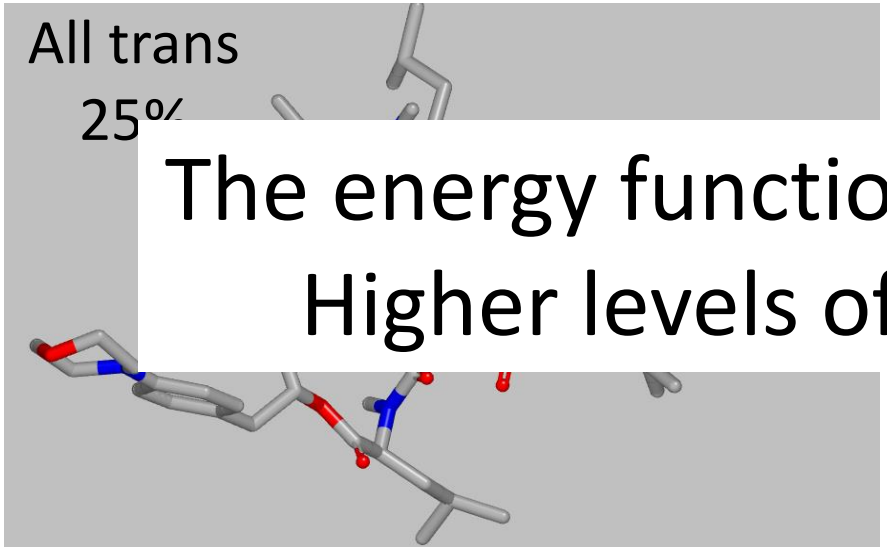


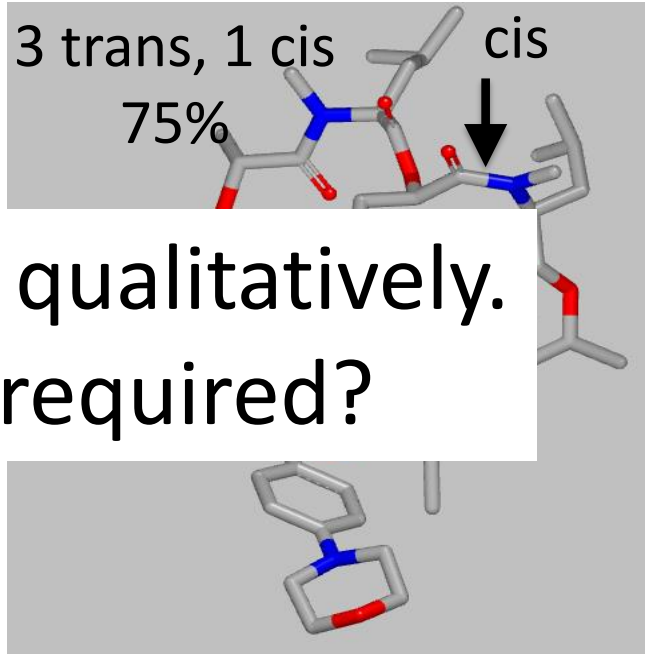
$CHCl_3$



Scherkenbeck *et al.*, *Curr. Topics Med. Chem.*, **2**, 759 (2002).

Testing the energy function in solution





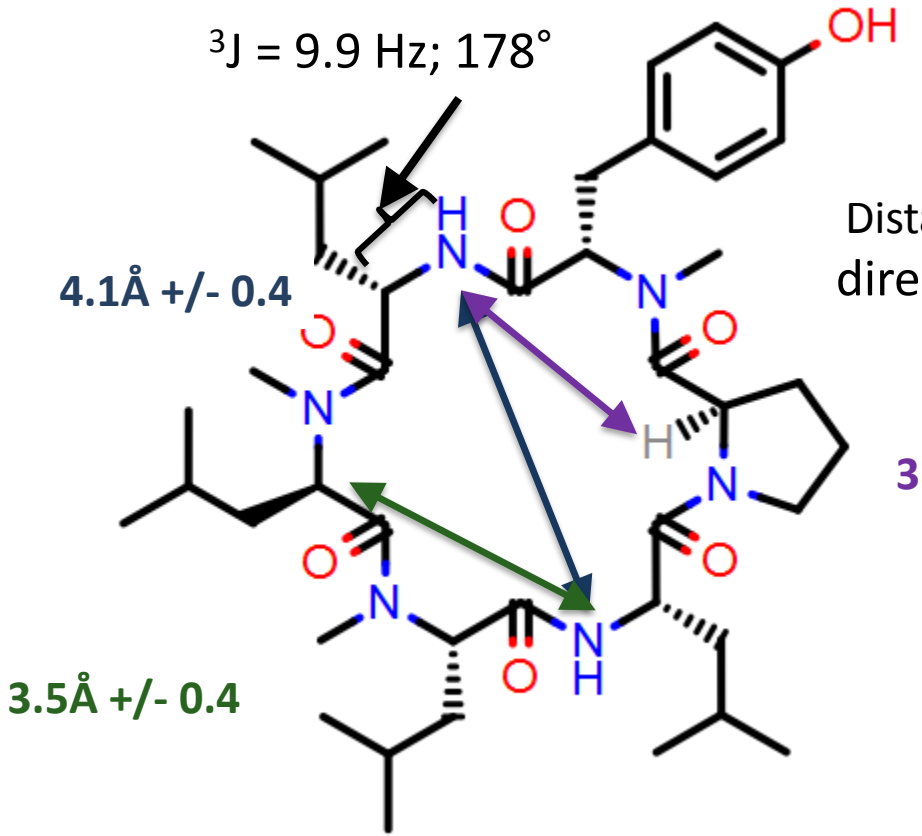
The energy function works qualitatively.
Higher levels of theory required?

3/10 most stable
BUT 60% of Boltzmann ensemble

7/10 most stable
BUT 40% of Boltzmann ensemble

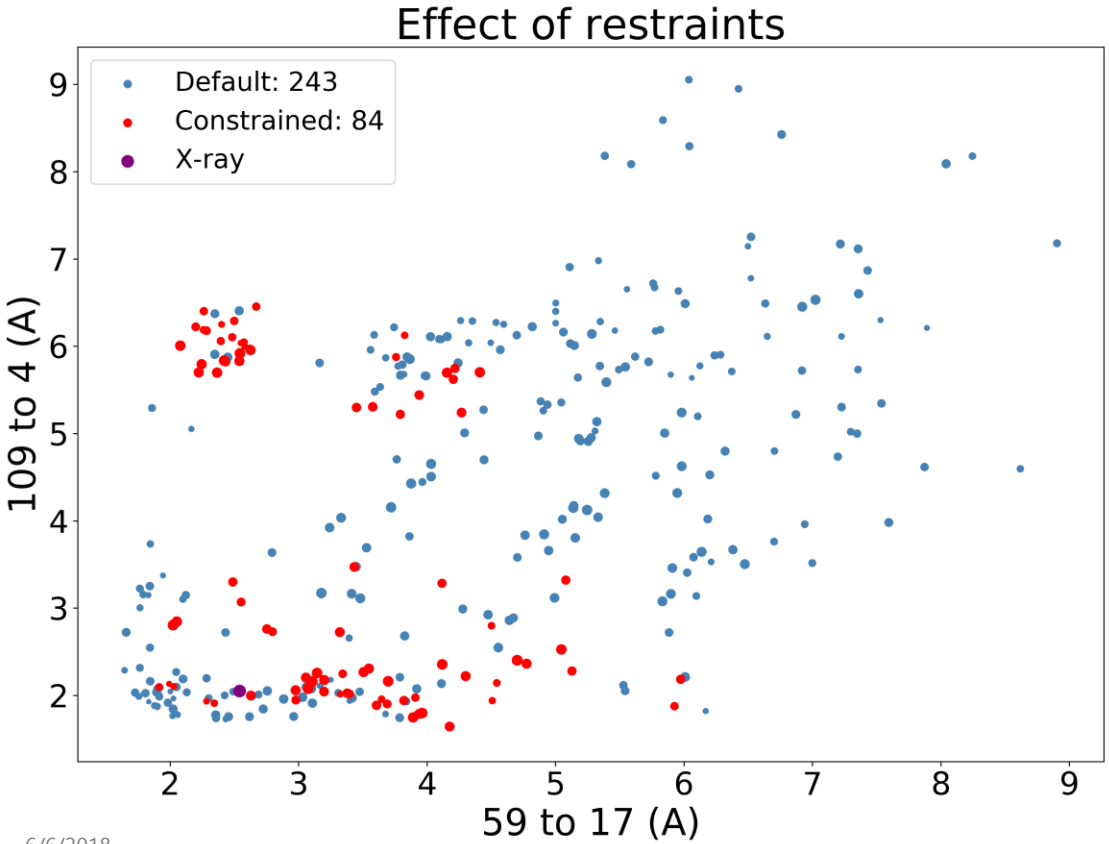
Emodepside in low dielectric ($CHCl_3$)

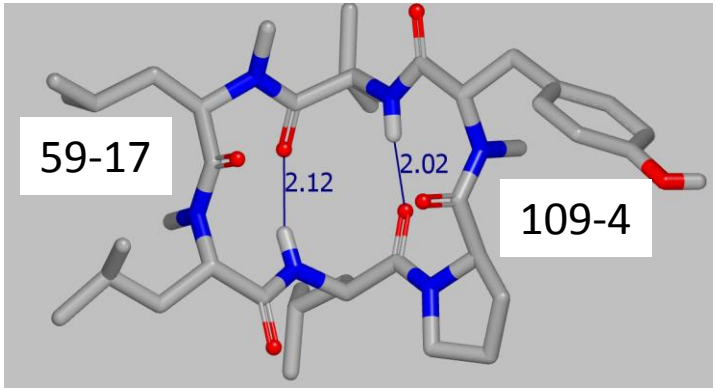
NMR data: distance & angle restraints



Distance & angle restraints work directly in OMEGA conformation generation

Incorporating NMR restraints: Lokey peptide





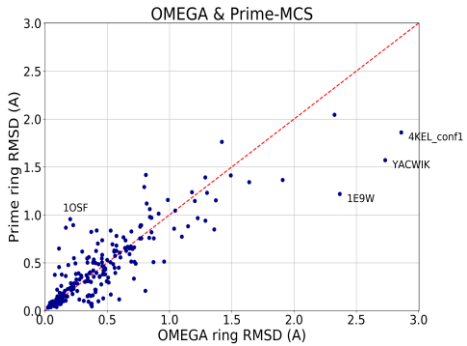
Experimental restraints focus sampling.

Lokey peptide in $CHCl_3$: unrestrained

Lokey peptide in $CHCl_3$: NMR restraints

Summary

- OMEGA works well for reproduction of the solid state



- Side-by-side comparisons drive future development

- Modelling the solution state is possible
  - More difficult than the solid-state

Acknowledgements

- OpenEye
  - Stan Wlodek, Krisztina Boda, Burt Leland
- Unnatural Products
  - Cameron Pye, Josh Schwochert
- BMS
  - Shana Posy, Steve Spronk