Disclaimer:
The presented results are preliminary and subject to change.

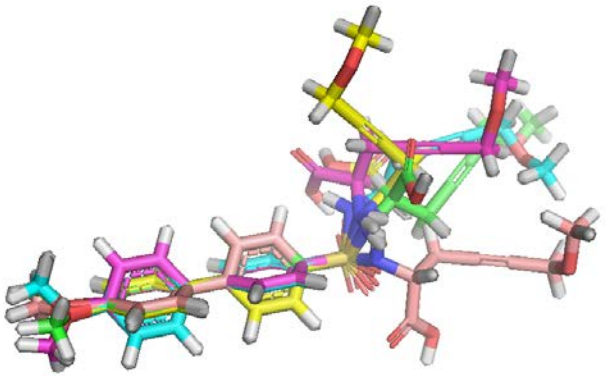
Teaching Distance Geometry About Experimental Torsion Preferences:

Using What We Know

Sereina Riniker, ETH Zürich, Switzerland

Gregory Landrum, Novartis Institutes for BioMedical Research, Basel, Switzerland

Conformer Generation



3D structure of a molecule
Ensemble/bundle of (low-energy) conformers

Applications
- Molecular docking
- 3D virtual screening
- 3D quantitative structure-activity relationships (QSAR)
- 3D pharmacophore modelling

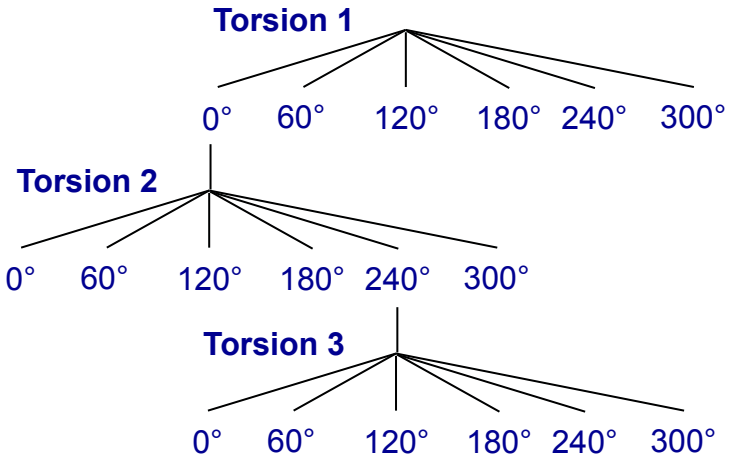
Requirements
- Fast, computationally efficient
- "Reasonable looking"
- Diverse → increased chance to find biologically relevant conformation

Validation
How well are crystal structures reproduced?
- Small-molecule crystal structures
- Protein-ligand complexes

Conformer Generators

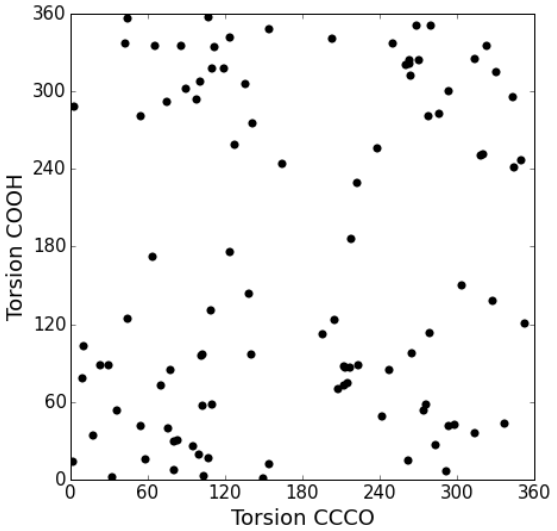
Methods to search conformational space:

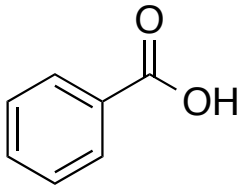
Systematic Search
Sample each rotatable bond in discrete intervals



+ all conformers are sampled
- limited to few rotatable bonds

Stochastic Search
Sample conformational space randomly





+ usable for highly flexible molecules
- not all conformers may be sampled

Conformer Generators

Knowledge-based methods:

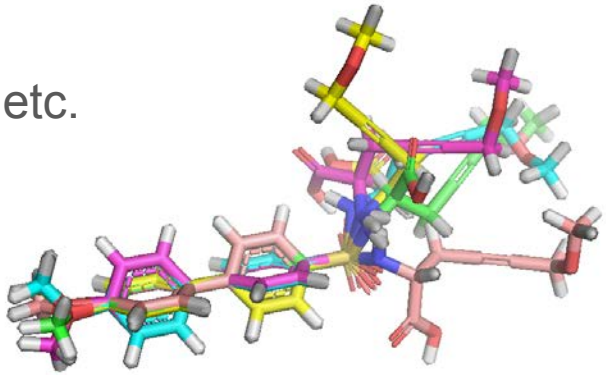
- Predefined libraries of torsional angles and ring conformations
- Cut molecules into fragments then reassemble
- Typically **systematic** search (stochastic also possible)

Corina, Confac, ConfGen, CAESAR, OMEGA, COSMOS, CONFECT, TrixX, FROG2, etc.

Distance-geometry (DG) method:

- **Stochastic** search
- Minimum degree of empirical information incorporated

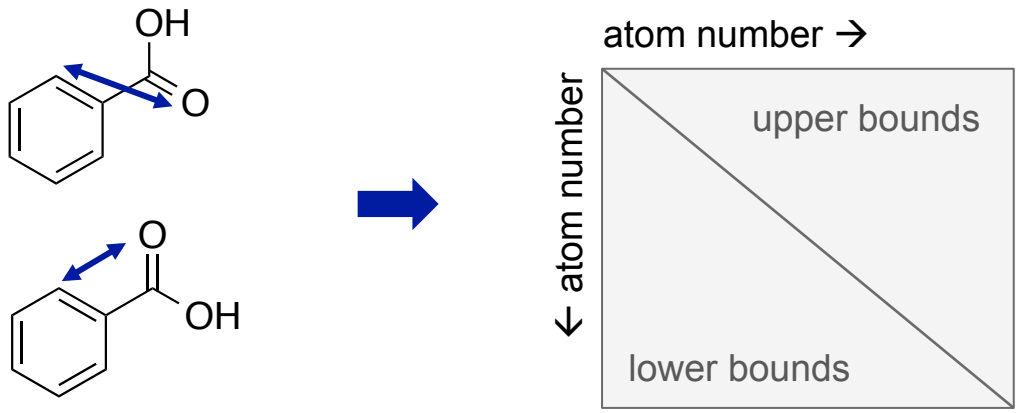
Implementation in the RDKit, BALLOON, DGEOM, etc.



Distance Geometry

Mathematical basis for geometric theory of molecular conformation
Special case of geometric algebra

Basic assumption:
It is possible to adequately define the set of all possible conformations of a molecule by means of purely geometric constraints.



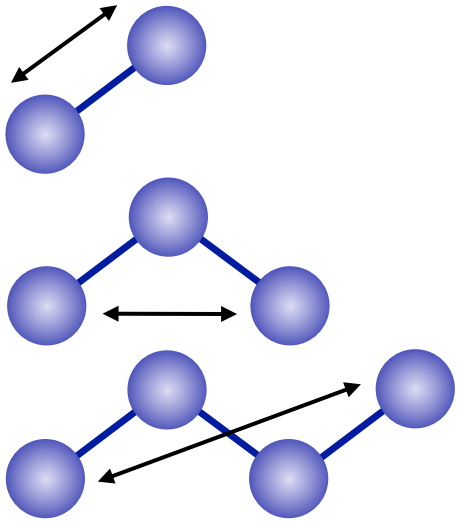
Matrix of distance constraints:
Lower and upper bounds on the distances between pairs of atoms

J. M. Blaney, J. S. Dixon, *Distance geometry in molecular modeling*, in "Reviews in Computational Chemistry", Vol. 5, VCH Publishers (1994), pp. 299

Distance Geometry

Matrix of distance constraints: Lower and upper bounds for atom pairs

Initial constraints:



1-2 distances
From ideal bond lengths

1-3 distances
From ideal bond lengths and bond angles

1-4 distances
From ideal bond lengths/angles and torsional angles
→ Empirical information to restrict torsional angles

Other distances
Default upper bound: arbitrary large value
Default lower bound: sum of the van der Waals radii

Triangle inequality bounds smoothing:

Iteratively apply the triangle inequality to make distance bounds more specific

J. M. Blaney, J. S. Dixon, *Distance geometry in molecular modeling*, in "Reviews in Computational Chemistry", Vol. 5, VCH Publishers (1994), pp. 299

Distance Geometry

Procedure of algorithm:
RDKit implementation

1) **Initializing of the distance bounds matrix**

2) **Triangle inequality bounds smoothing**

3) **Generating a random distance matrix**
   Chosen distances satisfy the distance bounds matrix

4) **Calculating the metric matrix**
   Describes the conformation in N-1 dimensions for the N atoms

5) **Embedding**
   Determining of the 3 largest eigenvalues + eigenvectors for the metric matrix
   Projecting the high dimensional system into 3 dimensions

6) **Refining the coordinates**
   Minimizing a distance-error function $F = \sum (\text{distance errors})^2 + \sum (\text{chirality errors})^2$

7) **Pruning of conformations within RMS threshold**

J. M. Blaney, J. S. Dixon, *Distance geometry in molecular modeling*, in “Reviews in Computational Chemistry”, Vol. 5, VCH Publishers (1994), pp. 299

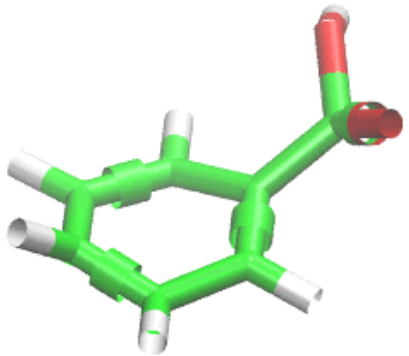
Distance Geometry

Advantages:

- Computationally efficient
- Based on 2D structure of molecule

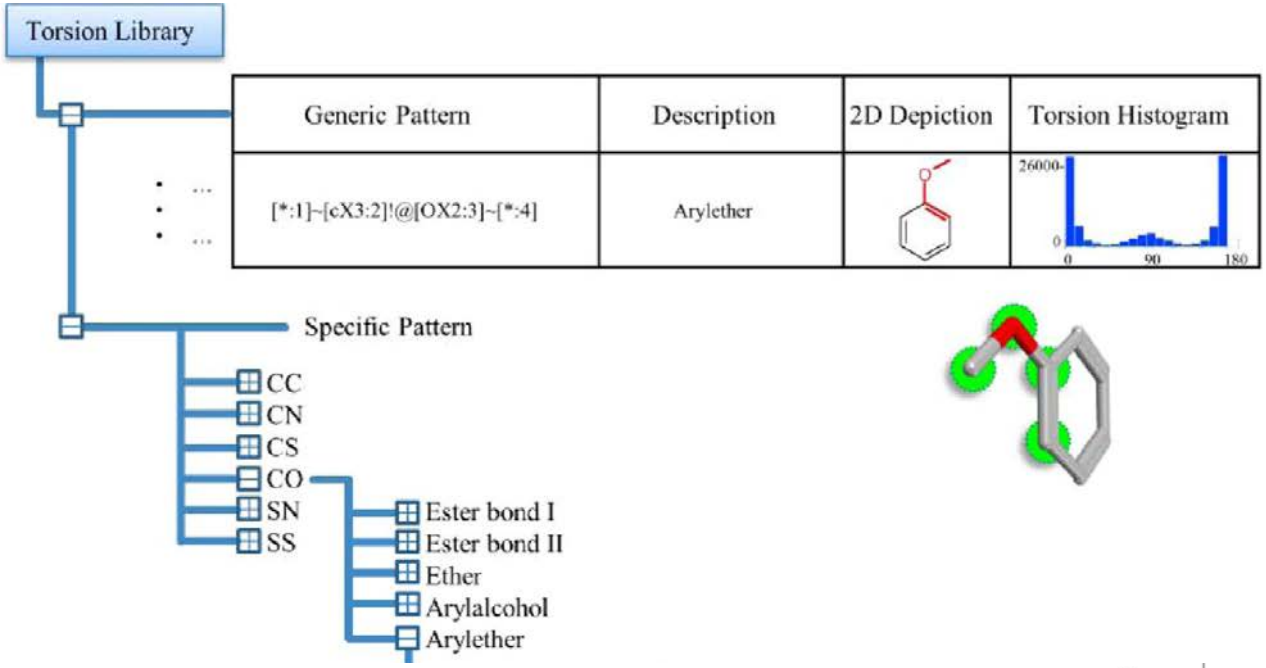
Disadvantage:

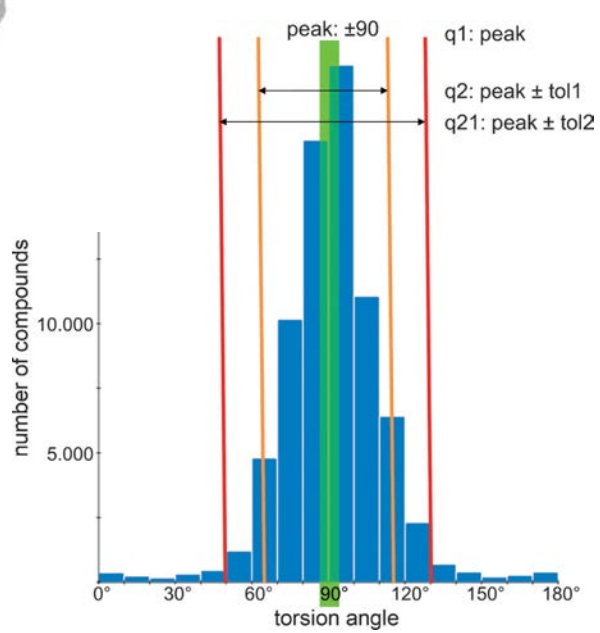
- Embedding can lead to “not-so-good-looking” conformations



Improve DG-conformers by enforcing experimental torsion-angle preferences

Experimental Torsion Preferences





392 specific patterns (SMARTS):

- Acyclic single bonds
- Hierarchically ordered
- Middle bonds: CC, CN, CS, CO, SN, SS
- For each pattern: Peaks + ranges

C. Schärfer, T. Schulz-Gasch, H.-C. Ehrlich, W. Guba, M. Rarey, M. Stahl, *J. Med. Chem.* (2013), 56, 2016.
C. Schärfer, T. Schulz-Gasch, J. Hert, L. Heinzlering, B. Schulz, M. Stahl, M. Rarey, *ChemMedChem* (2013), 8, 1690.

ETDG

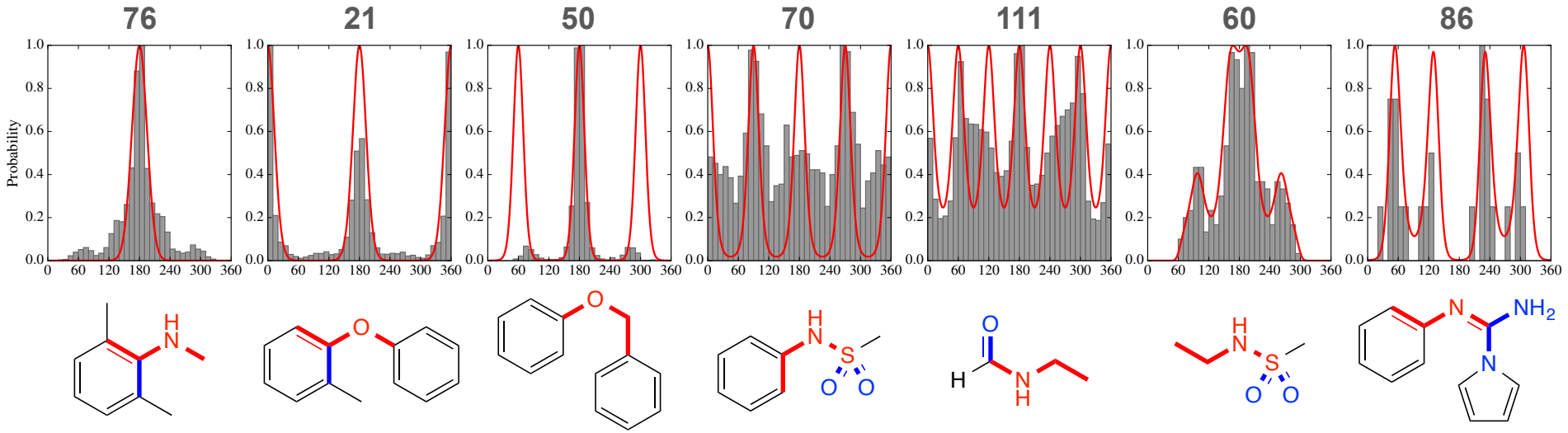
Incorporating experimental torsions (ET) into distance geometry (DG)

Three potential ways to do this:

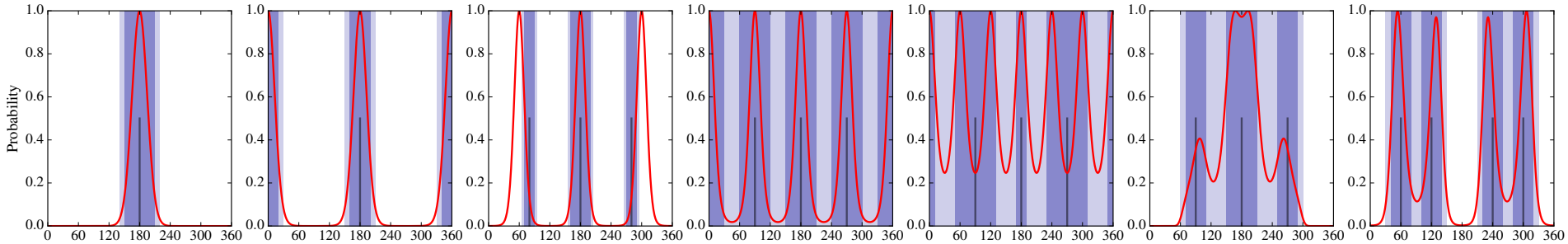
Adjust distance bounds matrix Upper/lower bounds based on experimental ranges	**Apply torsional angle constraints** Torsion angle constraints based on experimental ranges	**Apply torsional angle potentials** Torsion angle potentials based on fitted distributions
Prior to embedding	Additional minimization step after refinement	Additional minimization step after refinement
		
Bounds no longer correspond to the exact torsion ranges after embedding/refinement	**Series of constraints lead to problems during minimization**	

ETDG

Examples of fitted torsion potentials: $V_{tors}(x) = \sum_{i=1}^{6} K_i \left(1 + \cos(\delta_i)\cos(m_i x)\right)$



Comparison with torsion ranges from Schärfer *et al.*:



C. Schärfer, T. Schulz-Gasch, H.-C. Ehrlich, W. Guba, M. Rarey, M. Stahl, *J. Med. Chem.* (2013), 56, 2016.

ETDG

Data sets:

CSD data set: 1292 molecules
469 molecules from Hawkins *et al.* (*J. Chem. Inf. Model*. (2010), 50, 572)
823 additional molecules picked by same strategy

PDB data set: 239 molecules
79 molecules from Astex Diverse Set (*J. Med. Chem.* (2007), 50, 726) and
160 molecules from Hawkins *et al.* (*J. Chem. Inf. Model*. (2010), 50, 572)

Third data set only used for fine-tuning torsion-angle potentials: 431 molecules
Selected from CSD to contain torsion patterns not present in the other data sets

Parameters:

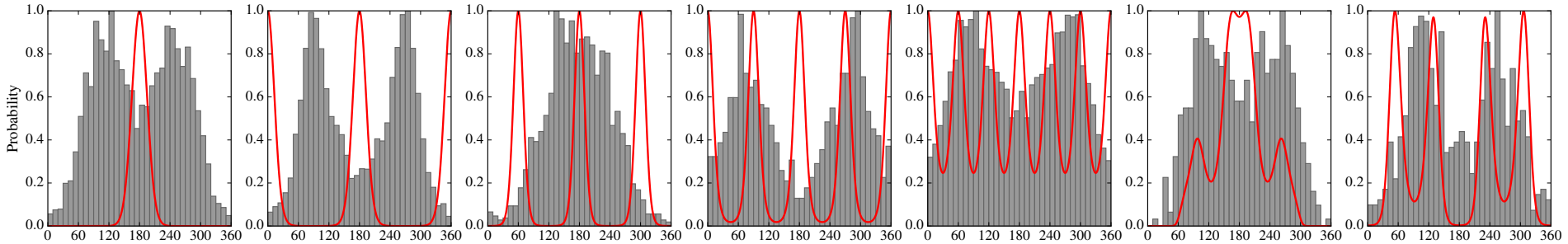
- Number of conformers (upper limit): n = 100
- RMS threshold = 1.0 Å

ETDG

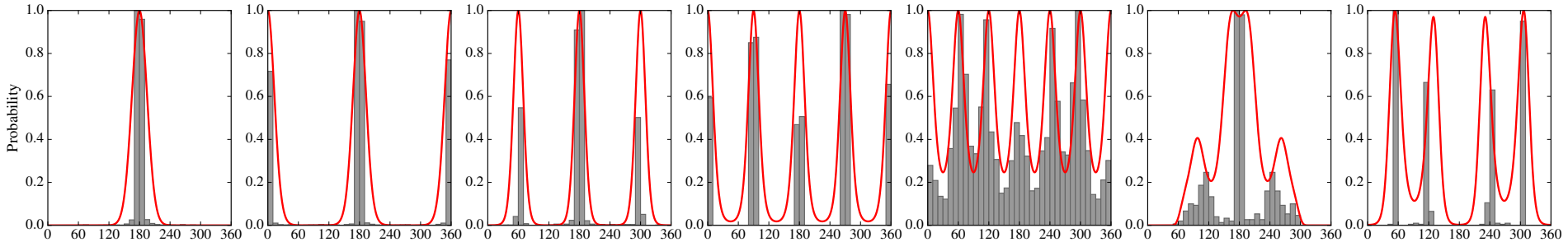
Distributions observed in a data set of 1962 molecules:

n = 100, RMS threshold = 1.0 Å

Conformations from DG



Conformations from ETDG

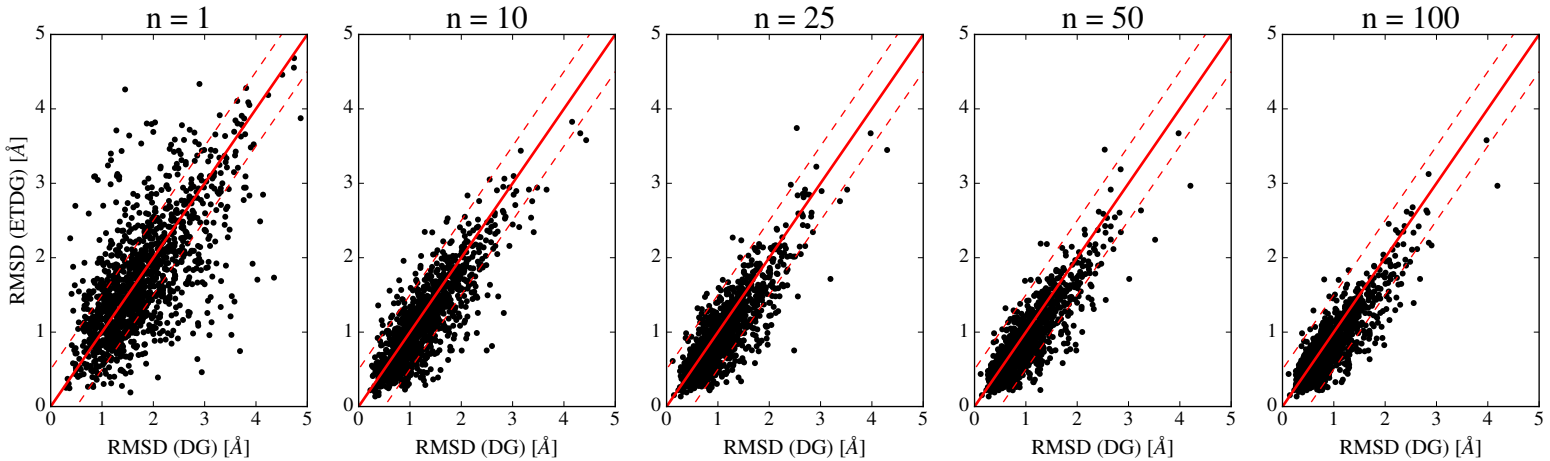


Torsion potentials restrict the torsional angles successfully.

Results

RMSD to crystal structure for best conformer:

Dataset: 1292 molecules from CSD

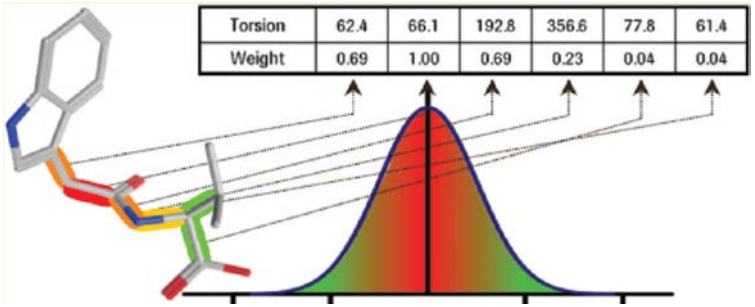


One-sided paired t-test:	p = 0.002476	p < 2.2e-16	p < 2.2e-16	p < 2.2e-16	p < 2.2e-16

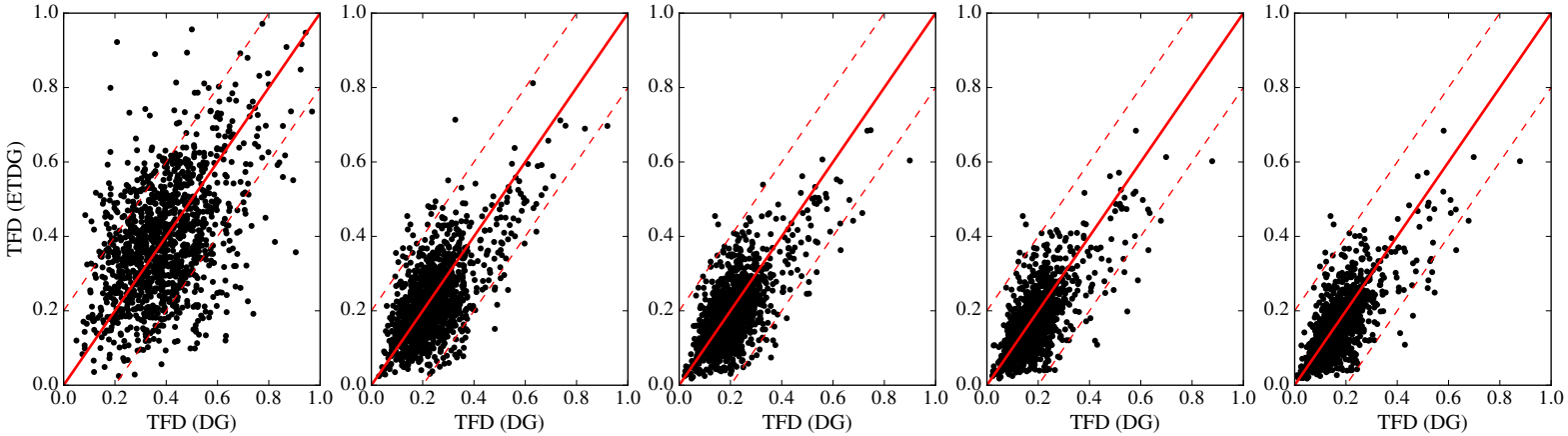
RMSD values improve – but not so much...

Results

TFD (torsion fingerprint deviation**) to crystal structure for best conformer:**



- Recording torsion angles around acyclic single bonds
- Conformations of ring systems represented by the sum of the absolute torsion values
- Topological weighting



TFD values improve – but not so much...

T. Schulz-Gasch, C. Schärfer, W. Guba, M. Rarey, *J. Chem. Inf. Model.* (2012), 52, 1499.

Results

ETDG: Aromatic rings and sp2-atoms in general “not good-looking”
Taken over from DG

Solution: Add “basic knowledge” terms to additional minimization step

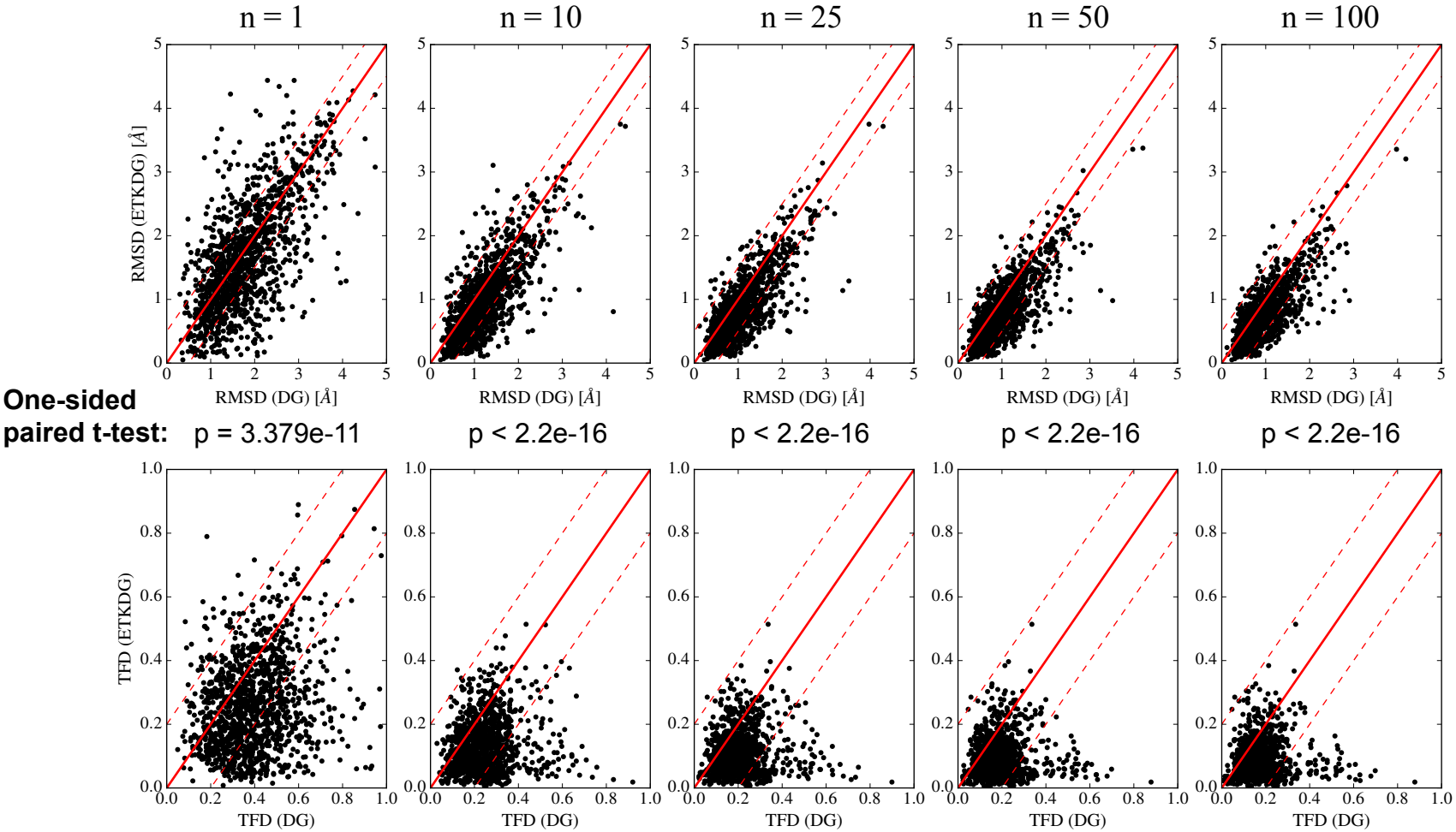
- Inversion (out-of-bend, improper dihedral) terms for sp2-hybridized atoms (N, O, C)
- Torsional potential with minima at 0°/180° for 4-, 5- or 6-membered rings where atoms are sp2-hybridized
- Bond angle potential with minima at 180° for triple bonds

Experimental torsions + knowledge distance geometry (ETKDG)

Results

RMSD / TFD to crystal structure for best conformer:

Dataset: 1292 molecules from CSD



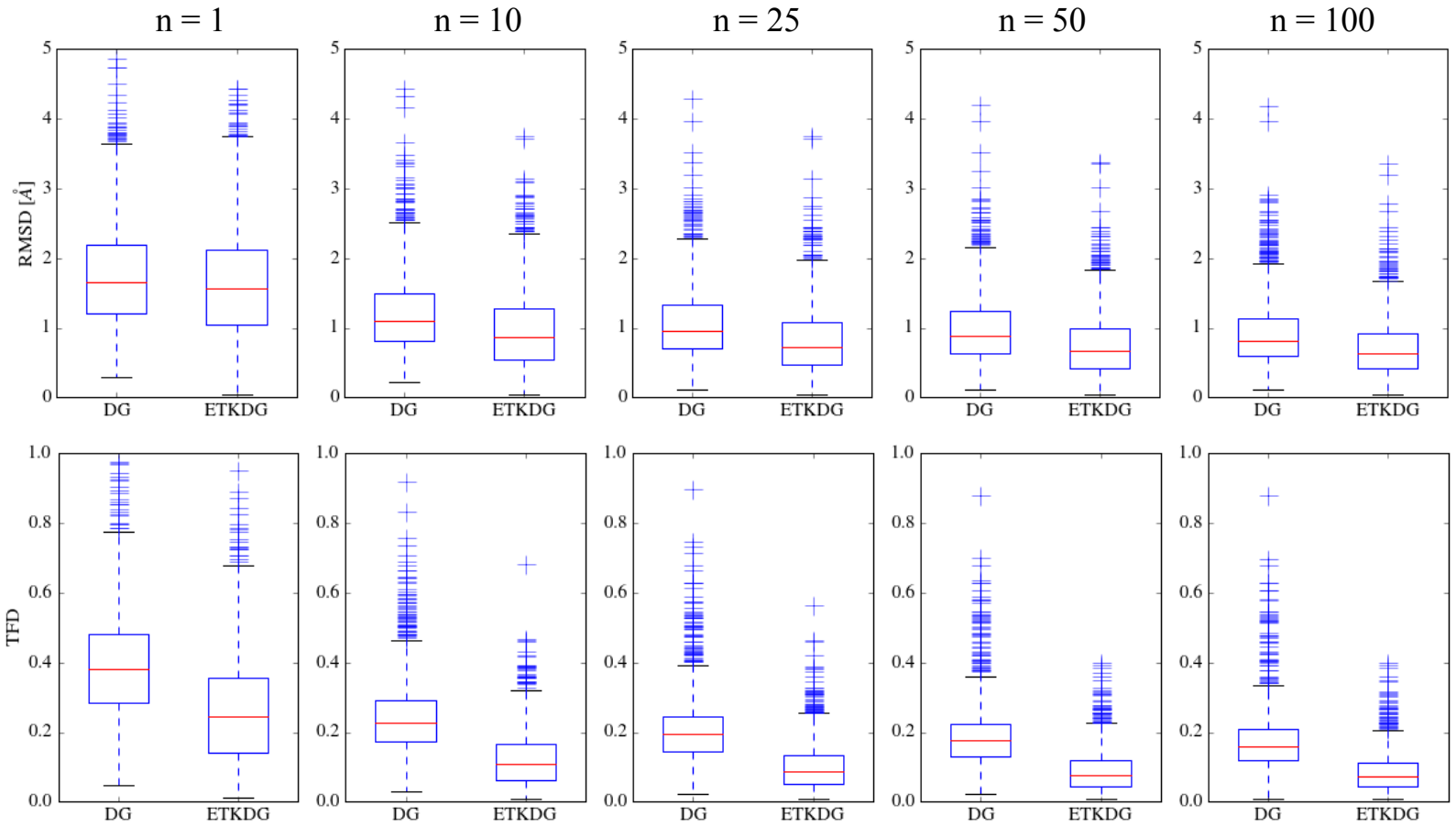
One-sided paired t-test: p = 3.379e-11 | p < 2.2e-16 | p < 2.2e-16 | p < 2.2e-16 | p < 2.2e-16

ETKDG outperforms DG both based on RMSD and TFD

Results

RMSD / TFD to crystal structure for best conformer:

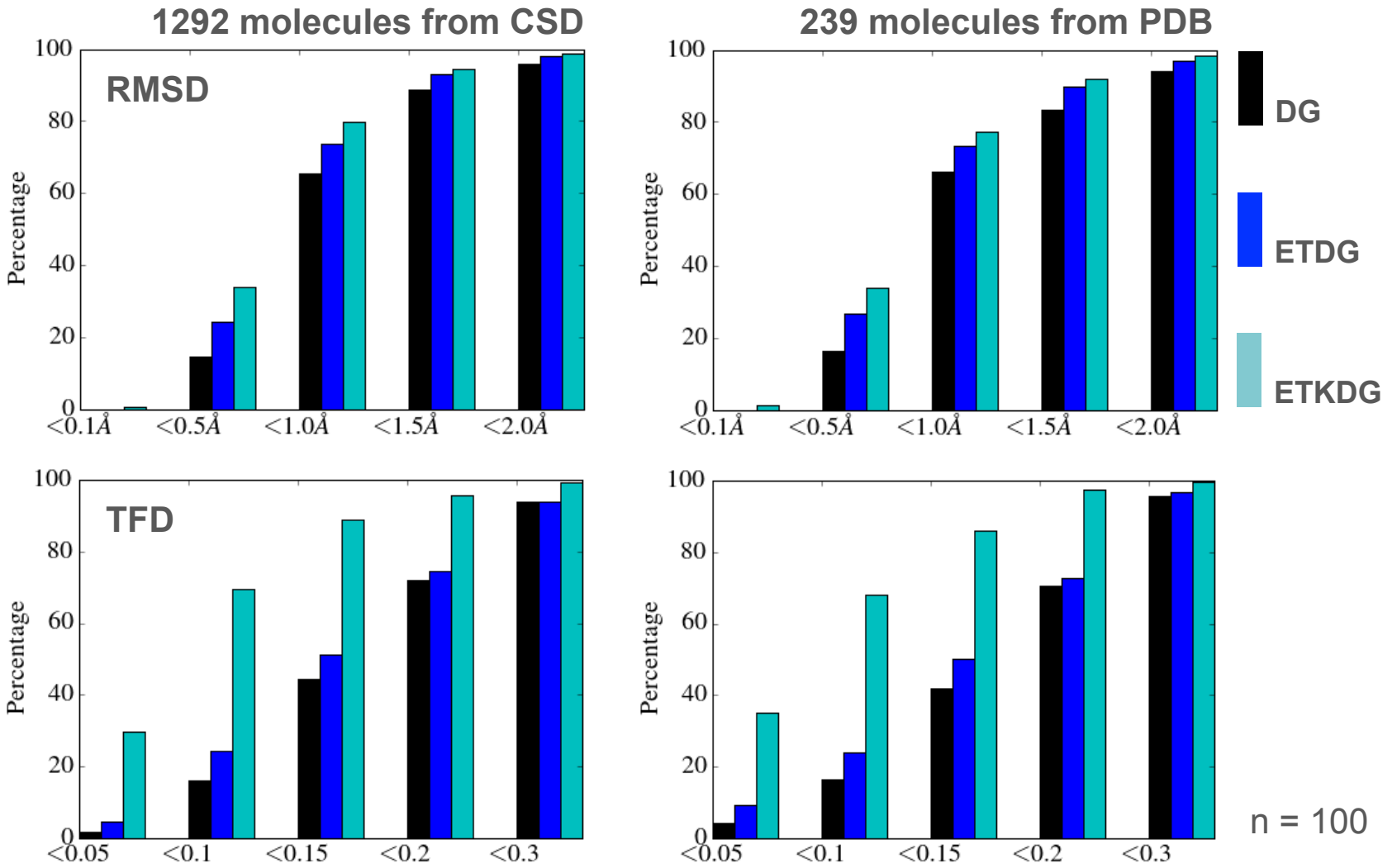
Dataset: 1292 molecules from CSD



ETKDG outperforms DG both based on RMSD and TFD

Results

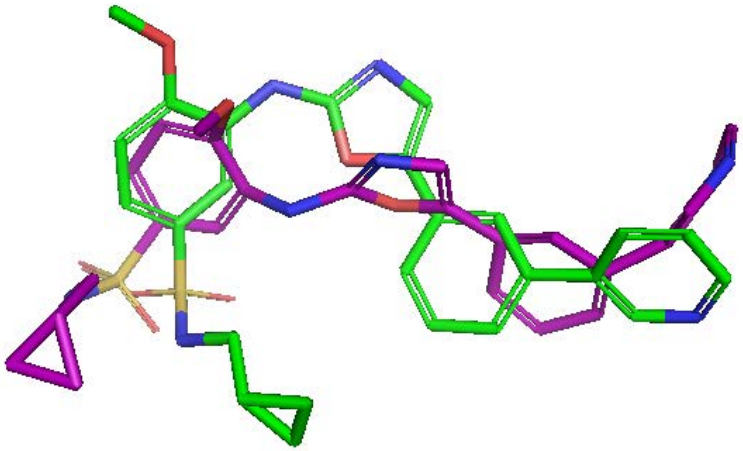
How well can the crystal structures be reproduced?

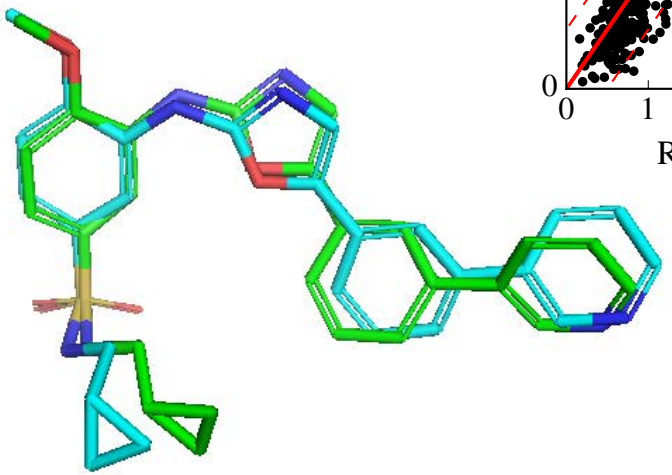


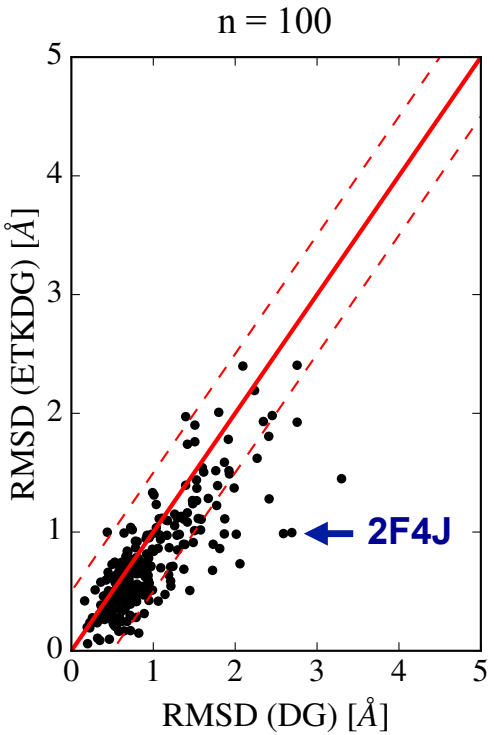
Results

Example where ETKDG gives big improvement

PDB 1Y6B







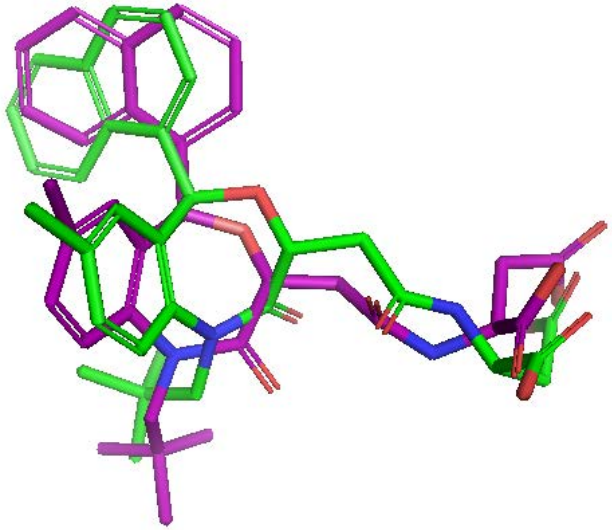
RMSD(DG) = 2.69 Å
RMSD(ETKDG) = 1.00 Å

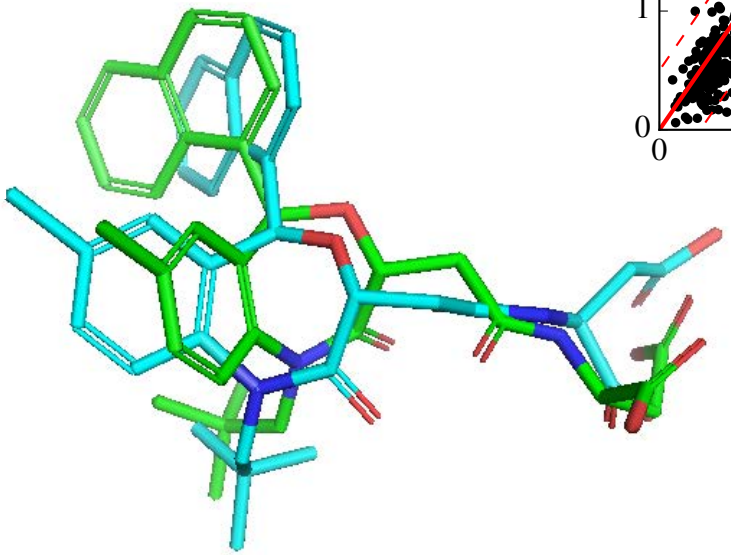
Best DG conformer **Crystal structure** **Best ETKDG conformer**

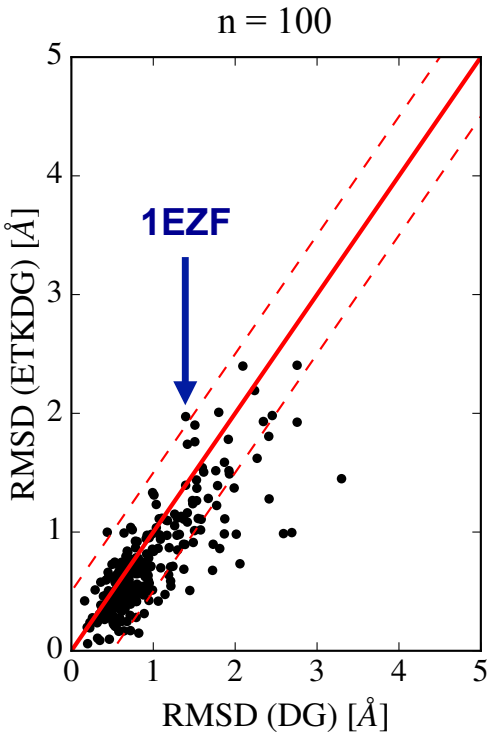
Results

Example where DG outperforms ETKDG

PDB 1EZF







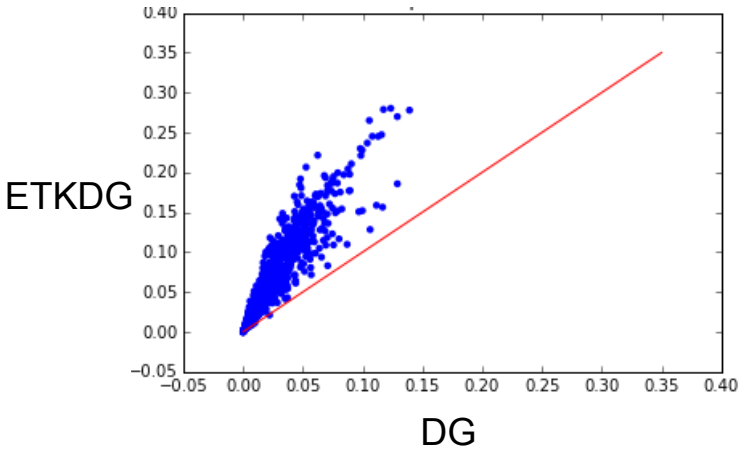
RMSD(DG) = 1.40 Å
RMSD(ETKDG) = 1.97 Å

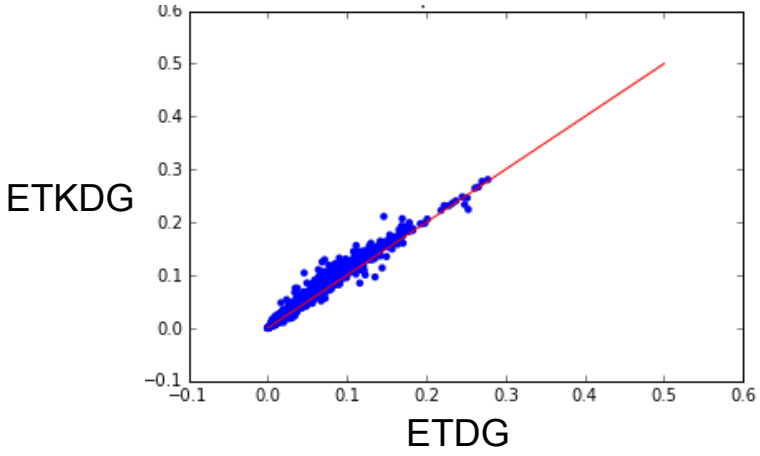
Best DG conformer **Crystal structure** **Best ETKDG conformer**

Timings

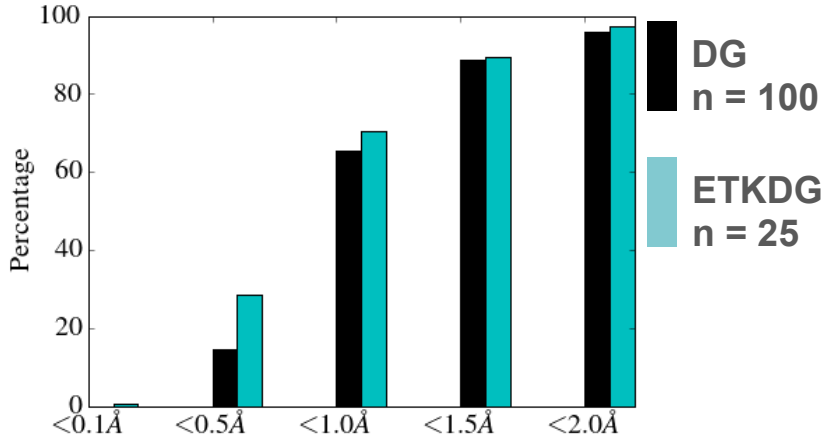
Time per conformer [ms]

n = 100, RMS threshold = 1.0 Å





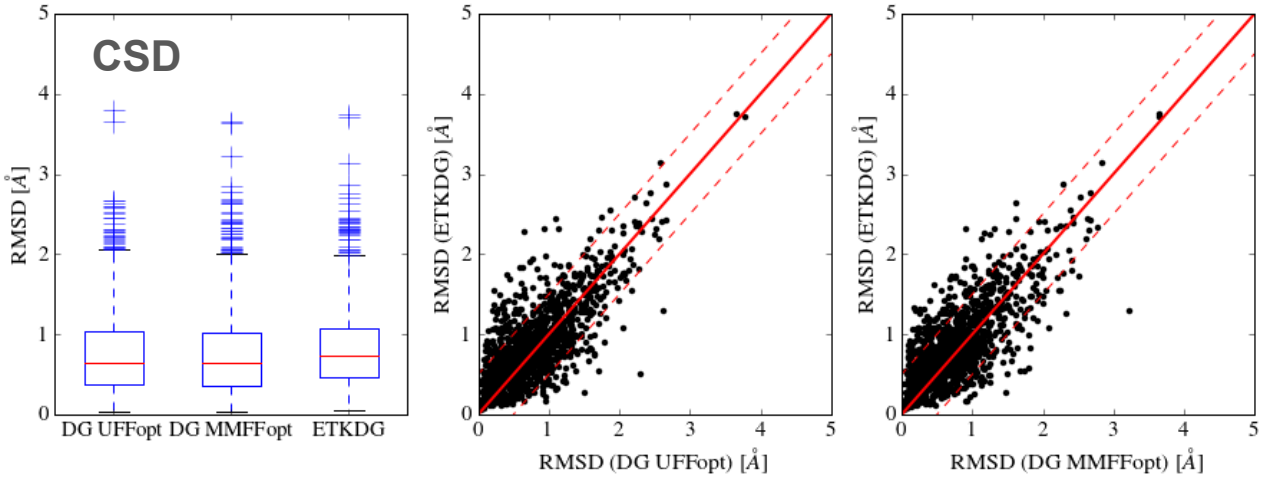
- **ET(K)DG require up to 2-3x more time per conformer**
- ETDG and ETKDG are comparable in speed
- **ETKDG requires 4x less conformers for better performance**

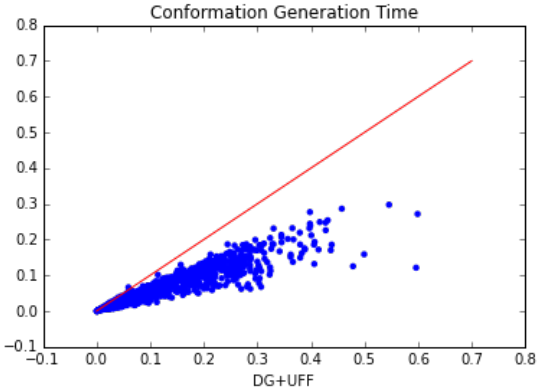


Results

Comparison of ETKDG versus FF-optimized DG:

RMSD analysis, n = 25

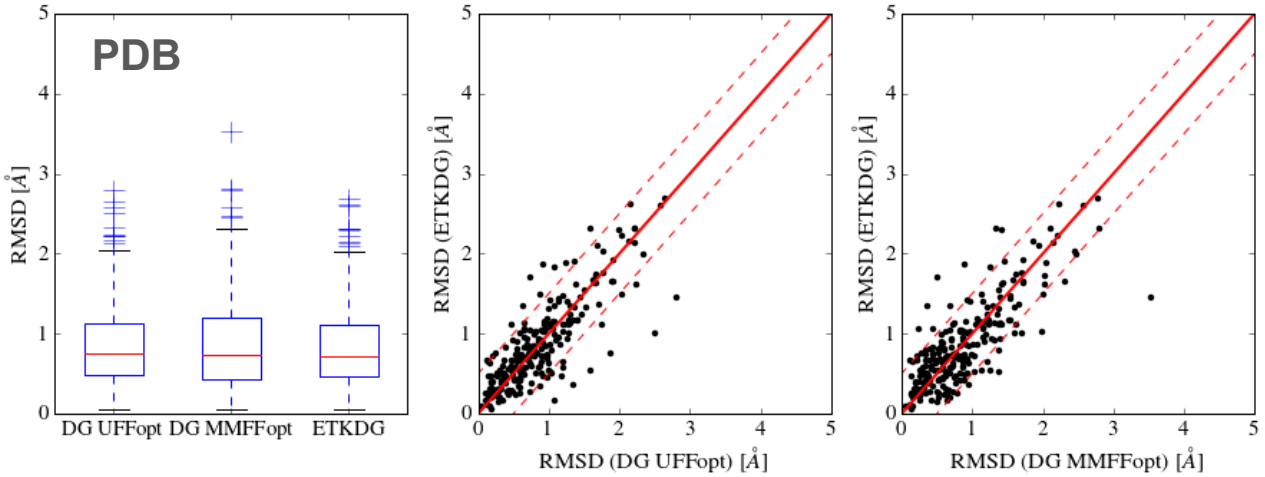




Two-sided paired t-test:

UFF: p = 2.795e-11

MMFF: p = 5.262e-12



Two-sided paired t-test:

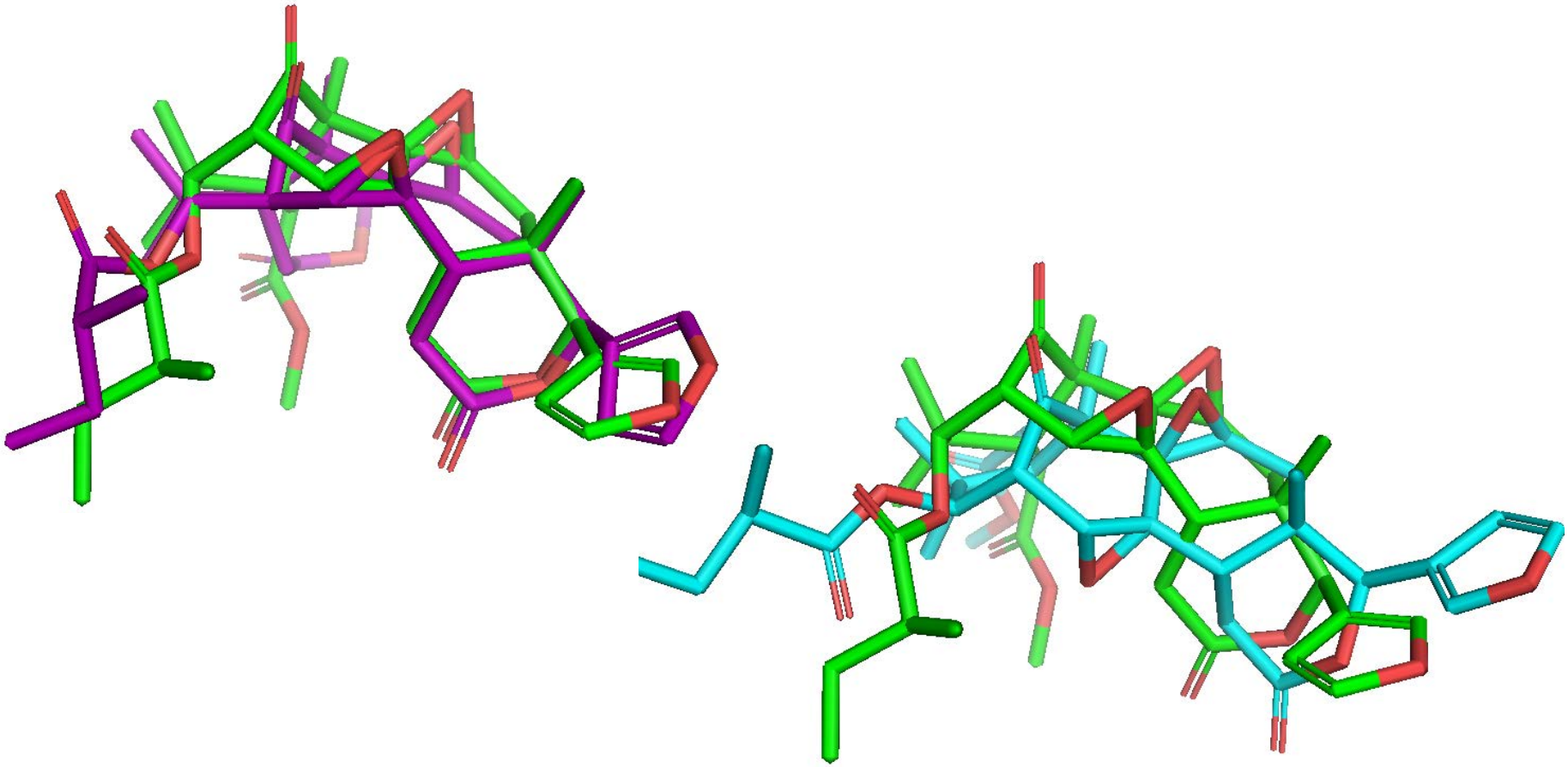
UFF: p = 0.4871

MMFF: p = 0.3694

Outlook

Fix chirality issues in DG

For highly constrained ring systems, chirality problems may occur



Summary

Distance Geometry for Conformer Generation

- Fast stochastic search of conformational space
- Minimal chemical information used
- Embedding step leads to distortions of aromatic rings and other sp2-centers

Experimental Torsion Angle Preferences

- 392 hierarchical SMARTS patterns for acyclic single bonds available (Schärfer *et al.* 2013)
- Fitting of torsion-angle potentials based on distributions for small molecule crystal data

ETKDG

- Additional minimization step with torsion-angle potentials
- ET terms alone are not sufficient, additional “basic knowledge” terms required for “good-looking” conformers
- ETKDG outperforms DG
- ETKDG is 2-3x slower per conformer but requires 4x less conformers
- ETKDG is an alternative to DG+FF-optimization