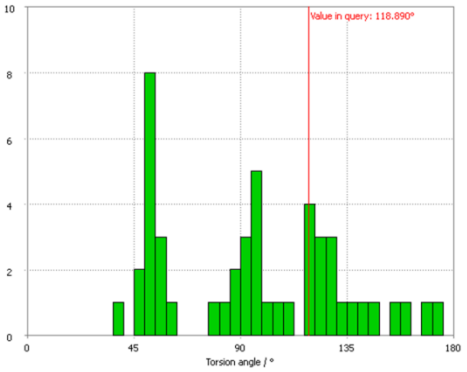
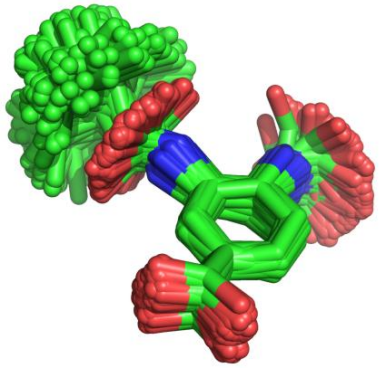




Generating Small Molecule Conformations from Structural Data

Patrick McCabe

mccabe@ccdc.cam.ac.uk

The Cambridge Crystallographic Data Centre,

12 Union Road,

Cambridge,

UK

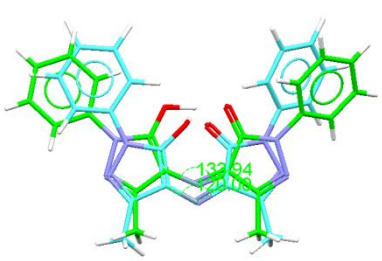
Outline

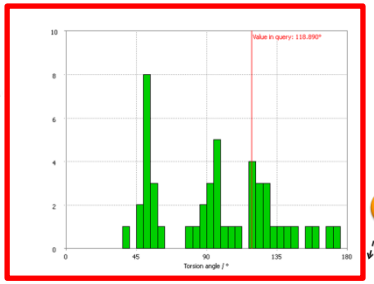
- The need for conformer generation
- Conformer generator description
- Test set
- Results
- Acknowledgements

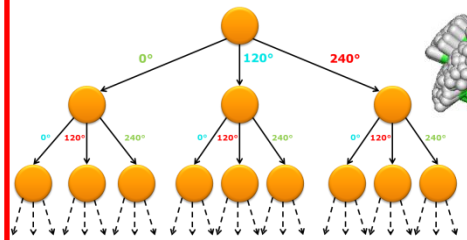
Why Conformer Generation is Important

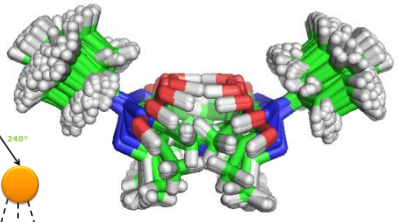
- Anywhere molecule flexibility is important
  - crystal structure prediction
  - drug design: most small molecule drugs work via physical interaction with specific biological target
  - function of drug intimately related to its 3D structure
- How do we model 3D flexible ligands?
  - solve dynamical equations of motion for ligand
  - use ensemble of pre-determined conformations

The CSD Conformer Generator









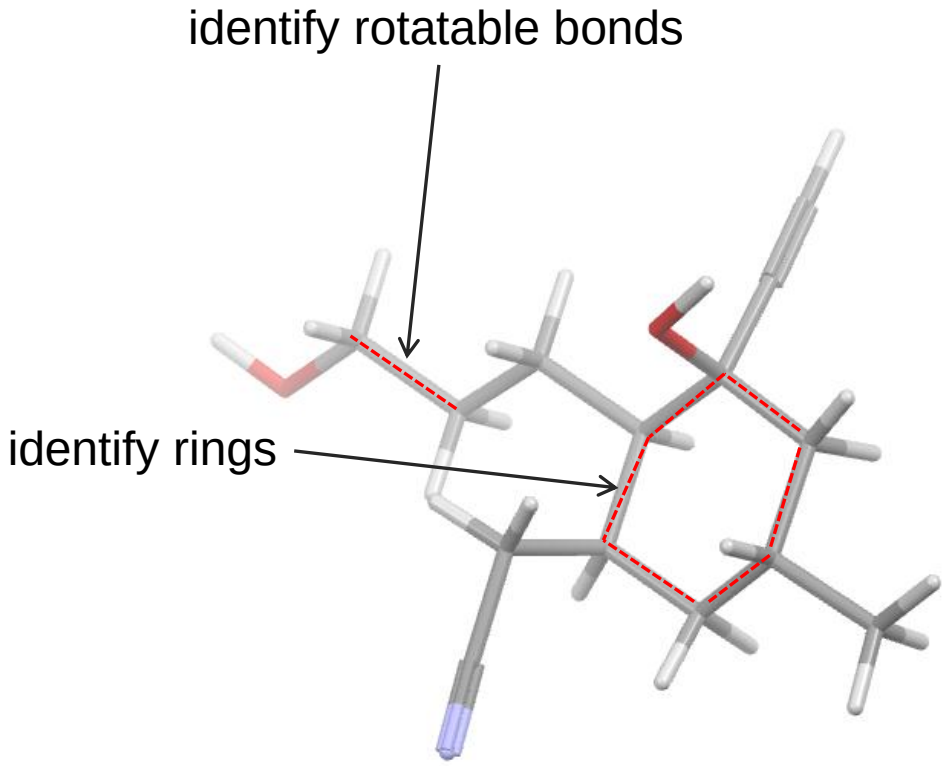
molecule → (minimise) → get rotamer and ring distributions → sample and score conformers → select diverse subset

- Knowledge based conformer generator

Knowledge Based Conformer Generator

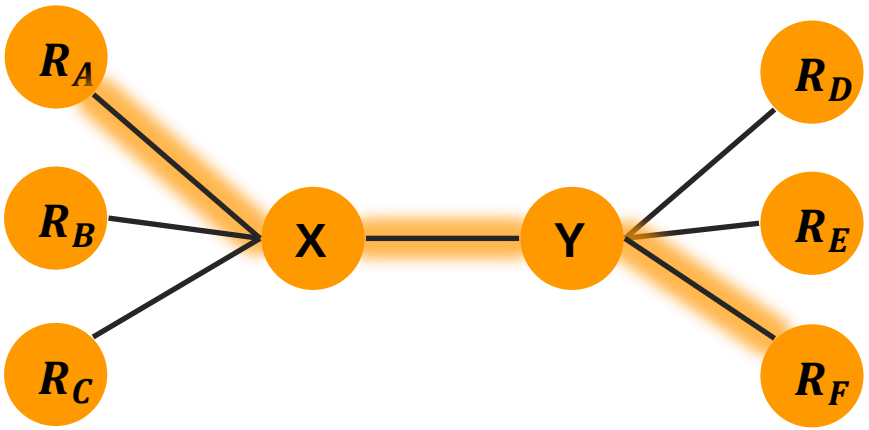
- Libraries built containing geometric preferences of molecules
- Bond length, bond valence angle, torsion angle distributions
- Unfused and fused ring system templates
- Validated against 3291 druglike molecules (not part of library)

Mapping Library Data onto Molecule



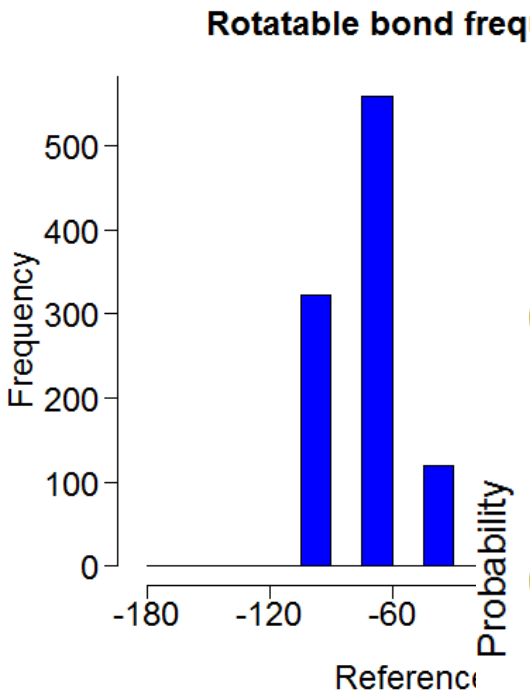
- Assign user defined fragment distributions to rotatable bonds
- Assign remaining rotatable bonds with CSD rotamer distributions
- Assign rings having alternative conformations with template distributions

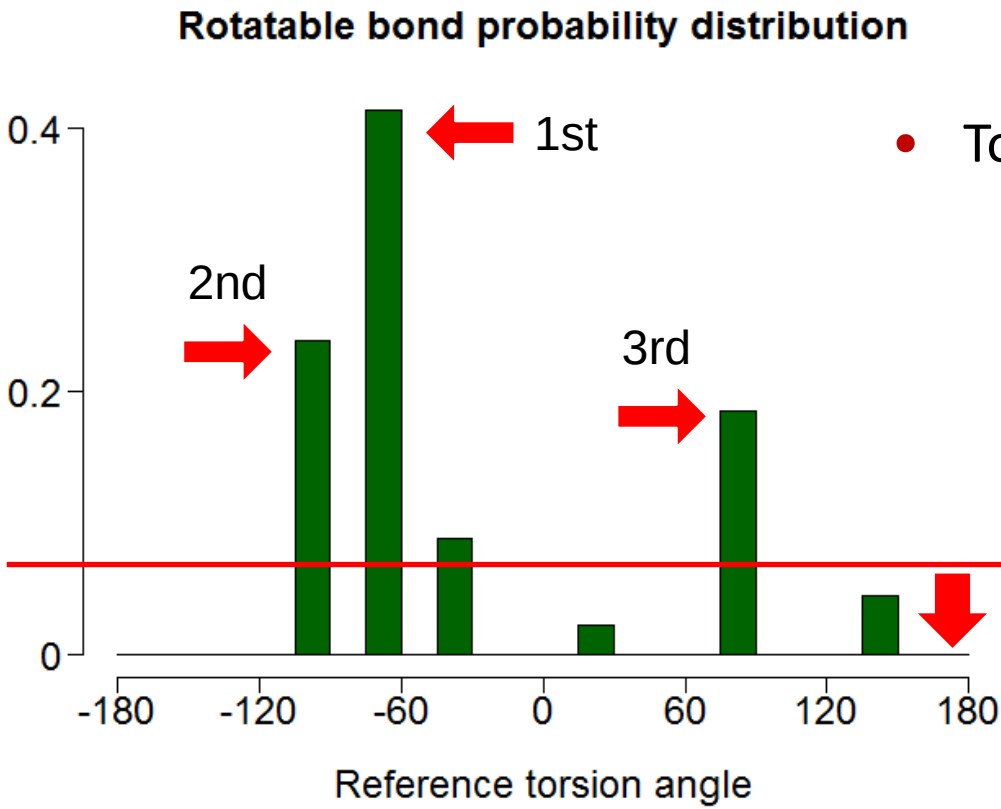
Rotamers



- Geometry around central bond X-Y could be defined by any of up to 9 torsions
- Simplify by defining 1 reference torsion
- Each central bond contributes to 1 rotamer distribution
- $-180 < \tau \leq 180$ capture asymmetric distributions

Structural Data Assigned

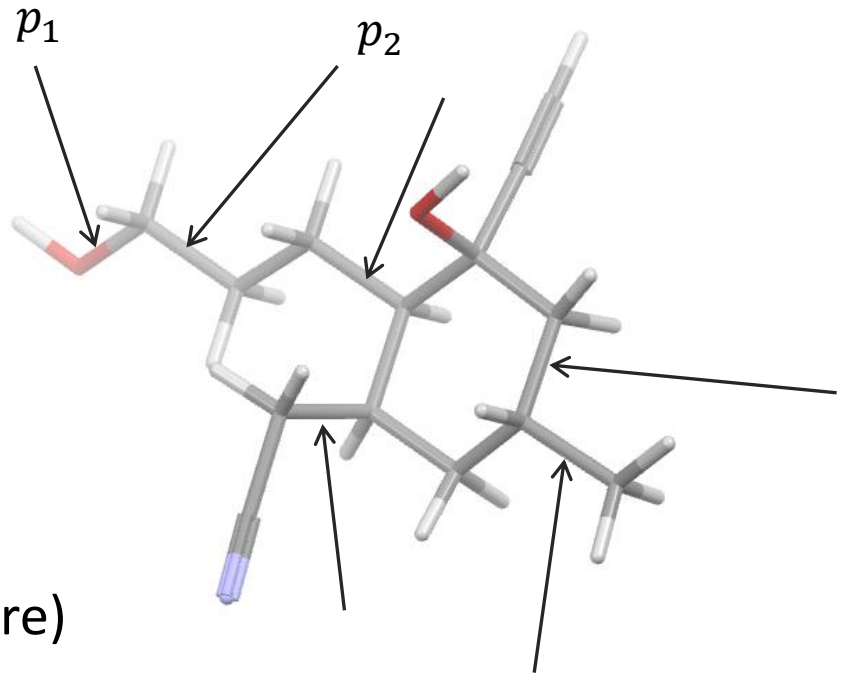




- Torsion angles selected in probability order

Unusual to reduce conformational space

Attribute Probabilities



- Torsion frequencies → probability
- Ring frequencies → probability
- Overall conformer probability (score)

$$p = p_1\, p_2 \cdots p_n$$

- In reality work with $\ln(p) = \ln(p_1)+\ldots+\ln(p_n)$

Torsion Distributions

- τ_1 and τ_2 independent $\Rightarrow$ knowing τ_1 tells us nothing about τ_2

		τ_2		
		120°	150°	180°
τ_1	90°	256	188	0
	120°	0	0	55

- τ_1 set to 90 $\Rightarrow$ τ_2 set to 120 or 150
- τ_1 set to 120 $\Rightarrow$ τ_2 set to 180
- Library of pre-determined dependent torsion angles

Attribute Probabilities

- τ_j and τ_i not independent

$$p = p_1\, p_2 \cdots p_{\tau_i} p(\tau_j|\tau_i) \cdots p_n$$

- No distribution for torsion i
  - 360° divided into 12 bins
  - $\mathrm{p}(\tau_i) = \frac{1}{12}$

- No ring template $\Rightarrow$ use input ring conformation

Conformer Probability Range

- $p_{min} \leq p_{conformer} \leq p_{max}$

- Maximum theoretical conformer probability

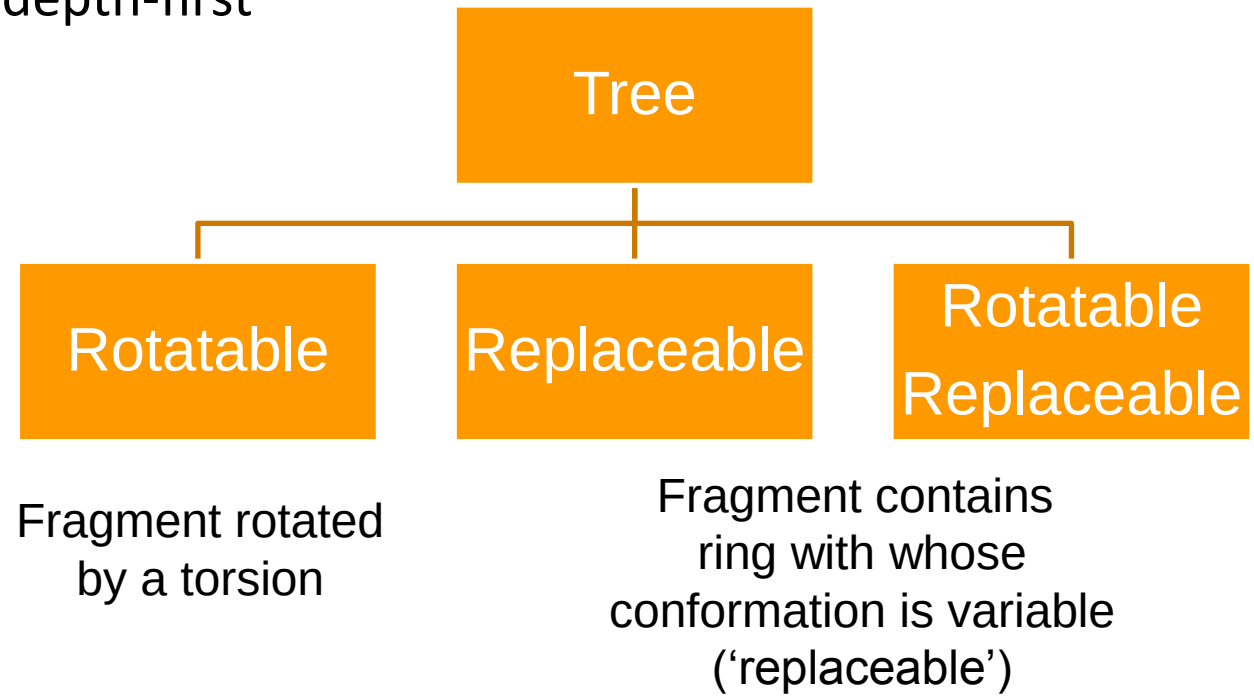
$$\ln p_{max} = \sum_{\substack{\text{t in \{ torsions,} \\ \text{ring templates \}}}} \ln \max(p_t)$$

- Similarly for minimum theoretical probability p_{min}

- To reduce conformational space

Molecule Representation

- Molecule is modelled as directed tree of connected fragments
- Root atom automatically chosen
- Tree is built depth-first



Generating Conformers

- An example molecule
  - 5 rotatable bonds; limit each to 12 torsion values
  - 2 flexible rings; each limited to e.g. 5 template conformations

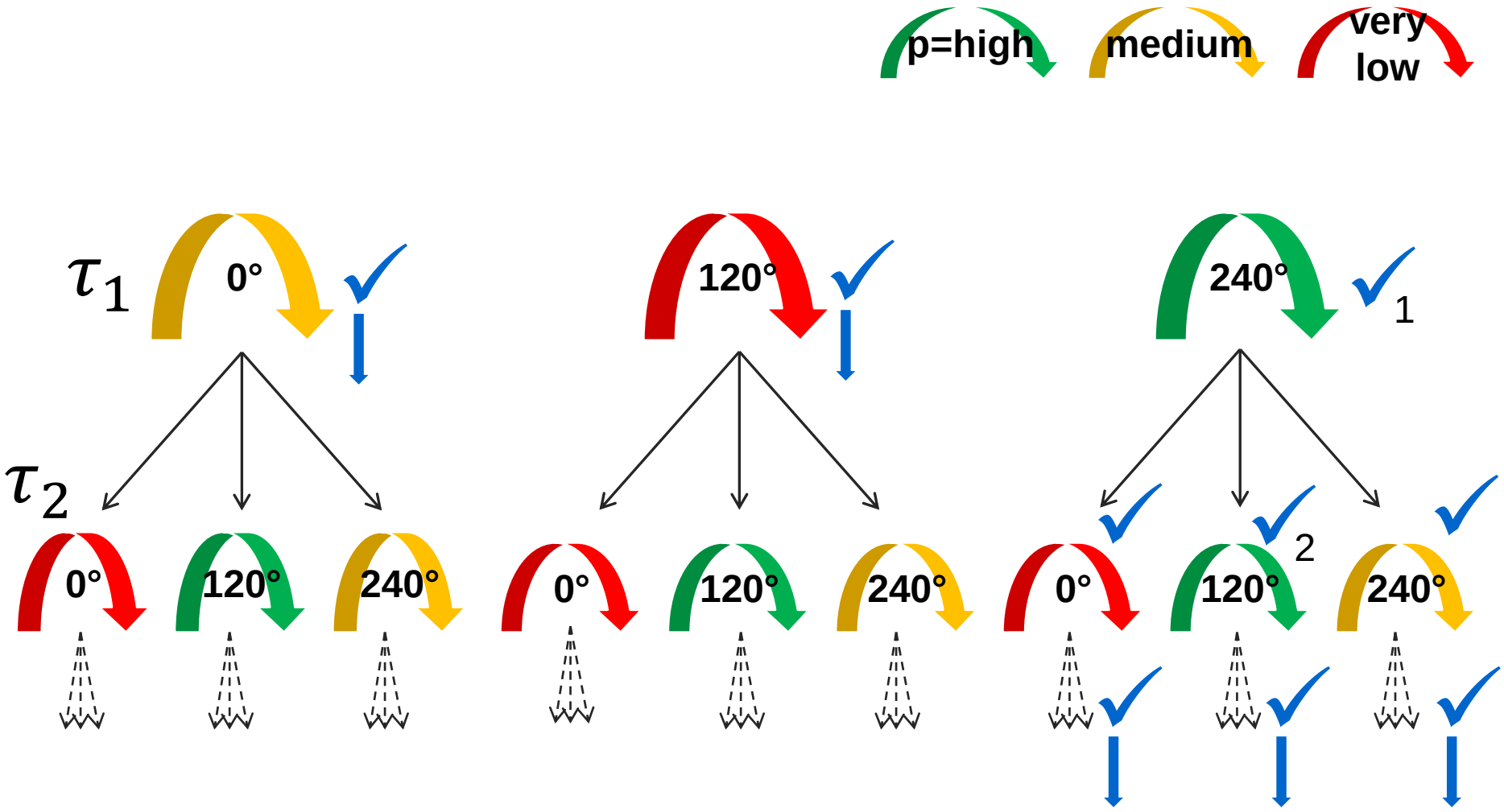
- 12 x 12 x 12 x 12 x 12 x 5 x 5 = 6,220,800 conformers

- Pruning techniques based on
  - number of unusual torsions
  - total number of conformations
  - probability range
  - clashing atoms

Traversing conformer tree

- No pruning
- 12 x 12 x 12 x 12 x 12 x 5 x 5 = 6,220,800 conformers

Traversing conformer tree - without pruning

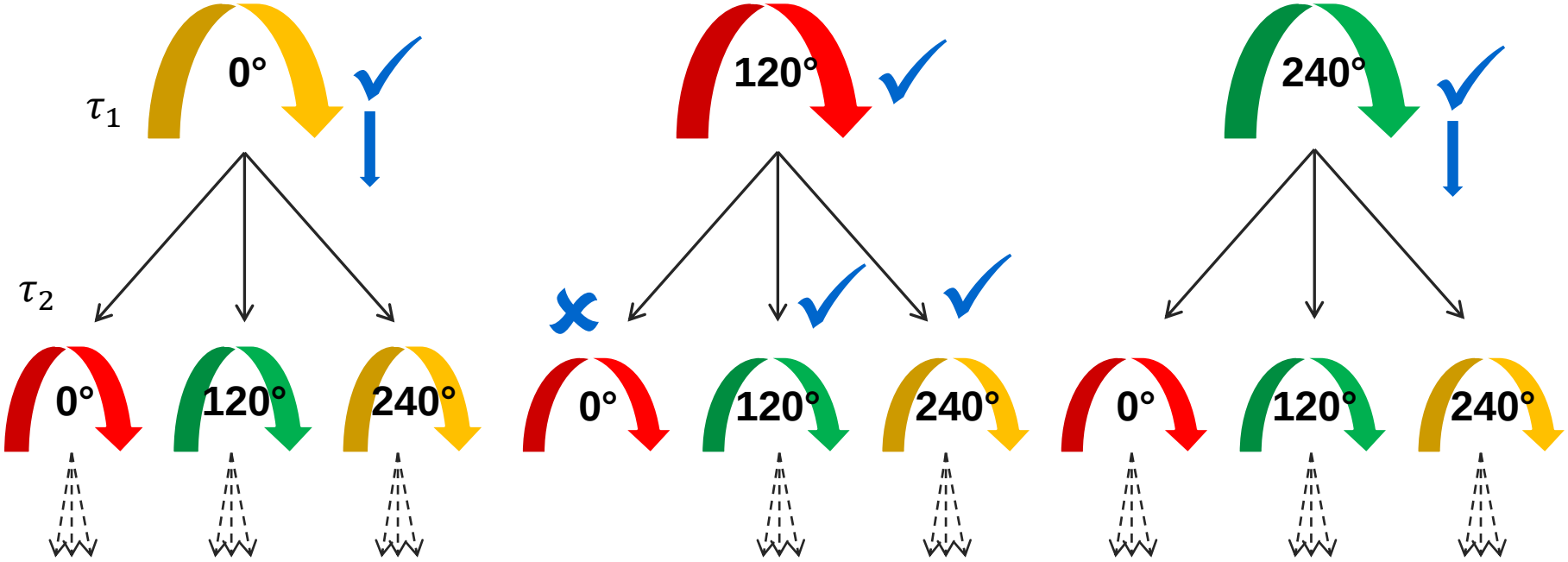


Pruning The Conformer Tree

- Number of unusual torsions
- Total number of conformations
- Probability range
- Clashing atoms

Pruning: Number of unusual torsions

- 1 unusual torsion allowed
  - unusual = very low probability

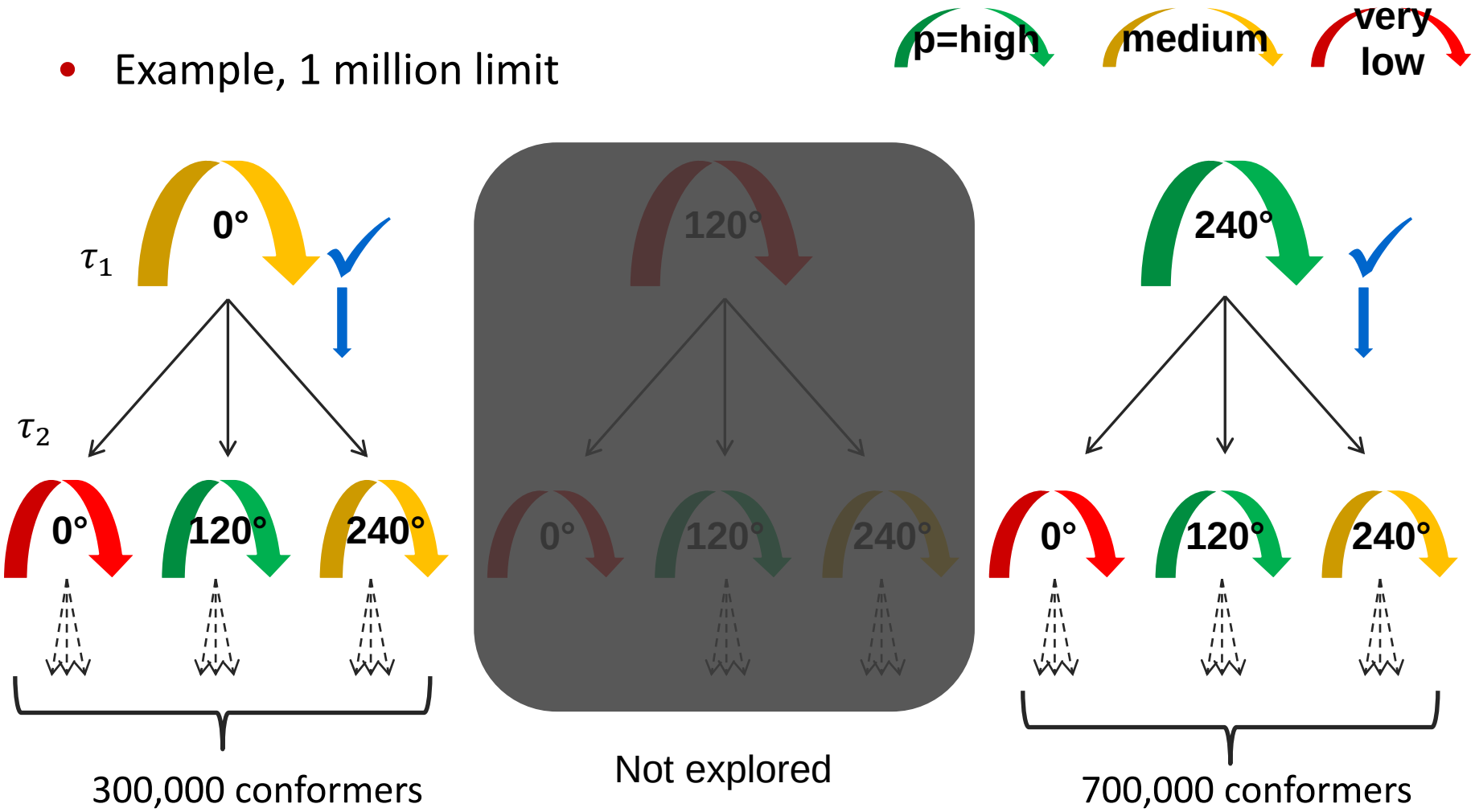


Pruning The Conformer Tree

- Number of unusual torsions
- Total number of conformations
- Probability range
- Clashing atoms

Pruning: Total number of conformations

- Example, 1 million limit



Pruning The Conformer Tree

- Number of unusual torsions
- Total number of conformations
- Probability range
- Clashing atoms

Pruning: Probability Range

- In principle

$$p_{min} \leq p \leq p_{max}$$

- In practice, introduce cut-off probability

$$p_{min} \leq \boxed{p_{cut} \leq p \leq p_{max}}$$

- If $p < p_{cut}$ then search branch eliminated

- If $\ln p < \ln p_{cut}$ then search branch eliminated

$$\ln p_{cut} = \lambda \ln p_{min} + (1 - \lambda) \ln p_{max}, 0 < \lambda \leq 1$$

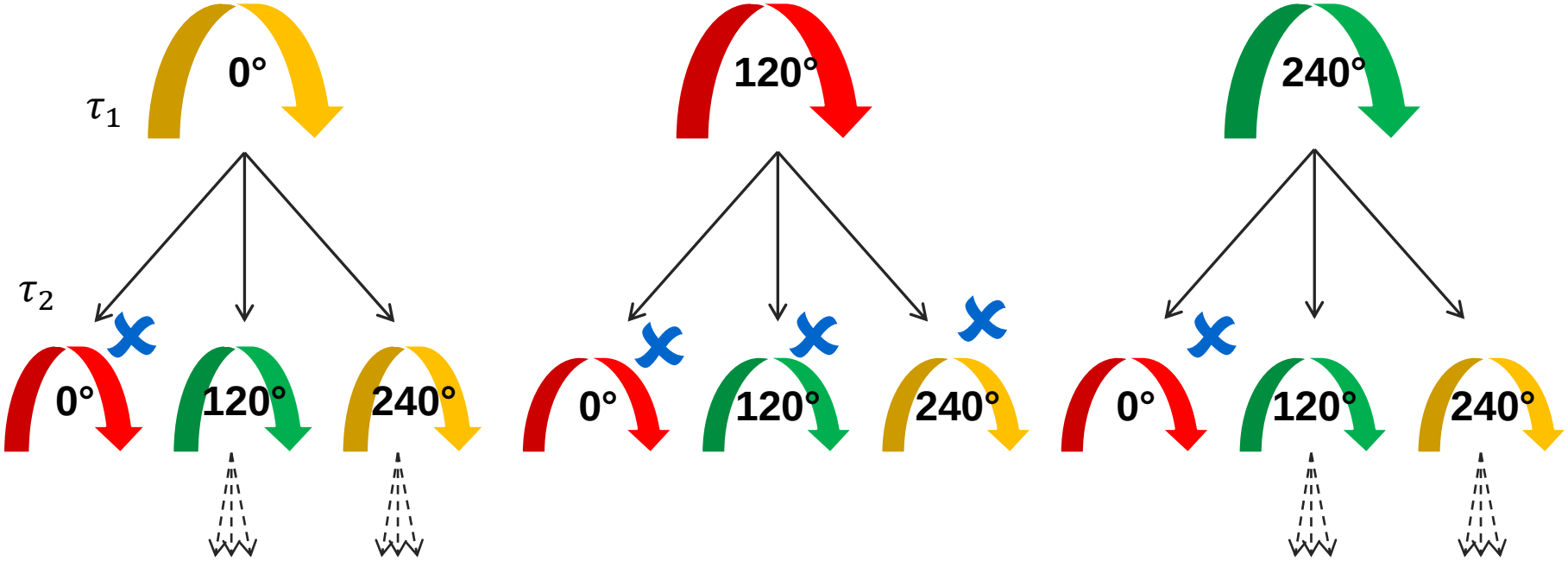
$\lambda = 0 \Rightarrow$ all branches eliminated

$\lambda = 1 \Rightarrow$ no branches eliminated

Pruning: Probability Range

Example $\lambda = 0.5$

$\ln p < \ln p_{cut}(\lambda = 0.5) \Rightarrow$ ✗



Pruning The Conformer Tree

- Number of unusual torsions
- Total number of conformations
- Probability range
- Clashing atoms

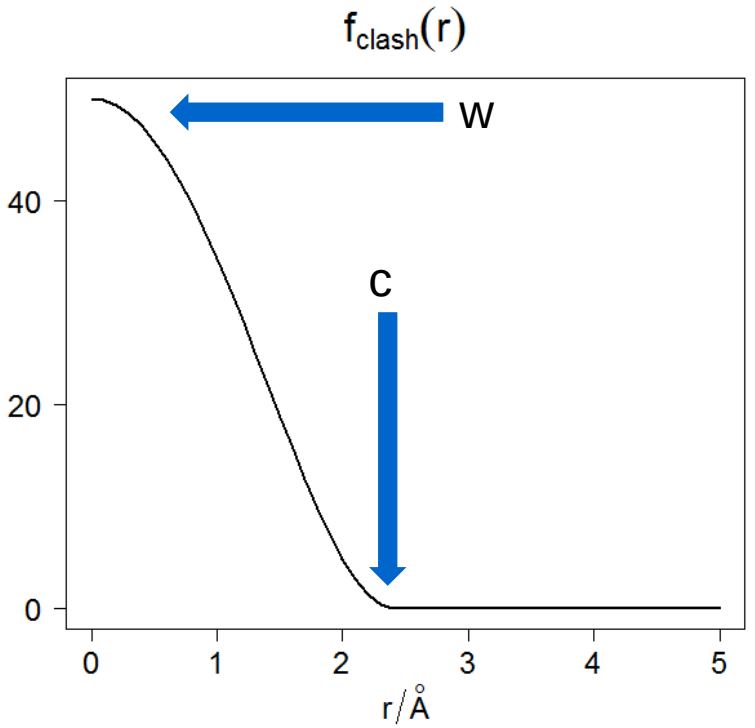
Pruning: Intramolecular clashes

- Calculated dynamically during incremental conformer construction ⇒ early rejection of clashing conformations

$$f_{clash}(r) = \begin{cases} w\left(1 - \frac{r^2}{c^2}\right)^2, & r \leq c \\ 0, & r > c \end{cases}$$

$$c = vdw_1 + vdw_2 - \alpha$$

- α allows fine tuning

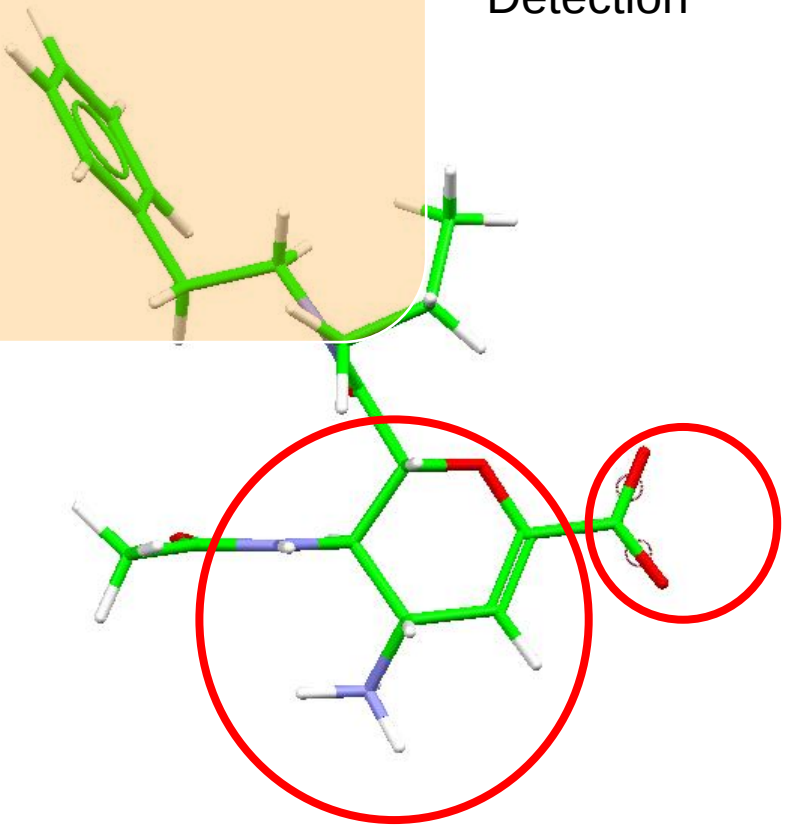


Intramolecular clashes

- bounding volumes do not clash ⇒ atoms can not clash

Detection

- bounding volumes do clash ⇒ check for atom-bounding volume clash

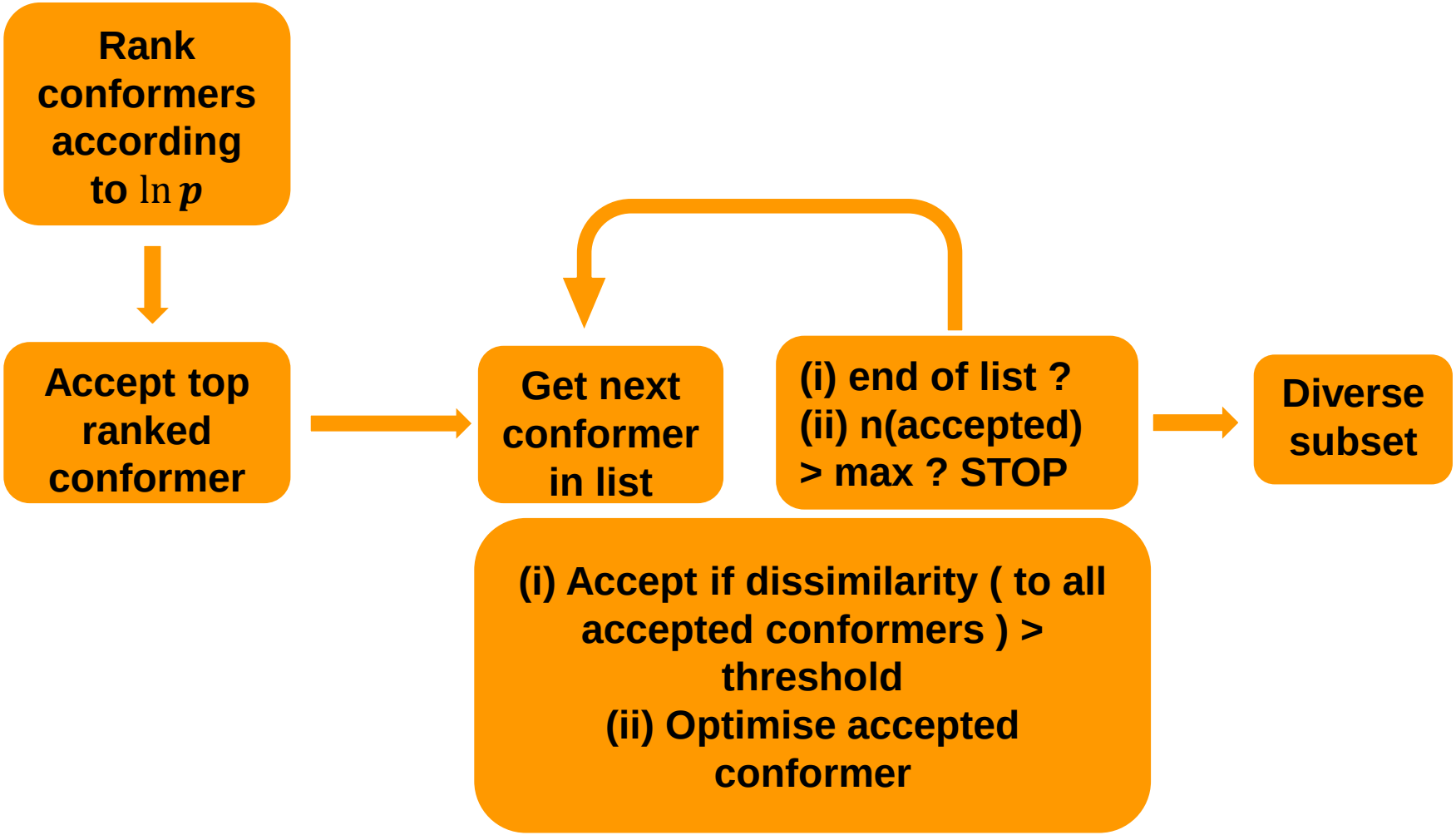


- If bounding volume-atom clash ⇒ evaluate all atom-atom clashes

Evaluation

Bounding volumes

Selecting Diverse Subset



Conformer Dissimilarity

- Heavy atom overlay RMSD – too slow to calculate
- Torsion Dissimilarity used as pre-screen
- Depends on difference between corresponding torsions
- Central torsions more significant than peripheral torsions

$$\tau_d = \sqrt{\sum_{\text{torsions i}} w_i(\tau_i(A) - \tau_i(B))^2}$$

conformers A, B

- w_i weighting accounting for importance of torsion

Conformer Dissimilarity

$$\tau_d = \sqrt{\sum_{\text{torsions i}} w_i(\tau_i(A) - \tau_i(B))^2}$$

conformers A, B

- $\tau_d \geq torsion\ threshold \Rightarrow$ dissimilar
- $\tau_d < torsion\ threshold \Rightarrow$ decision based on atom RMSD
  - $RMSD \geq RMSD\ threshold \Rightarrow$dissimilar
  - $RMSD < RMSD\ threshold \Rightarrow$similar
- Example $\tau_d = 100$, atom RMSD = 0.5
  - < 4% conformer pairs have τ_d > 100, atom RMSD < 0.5

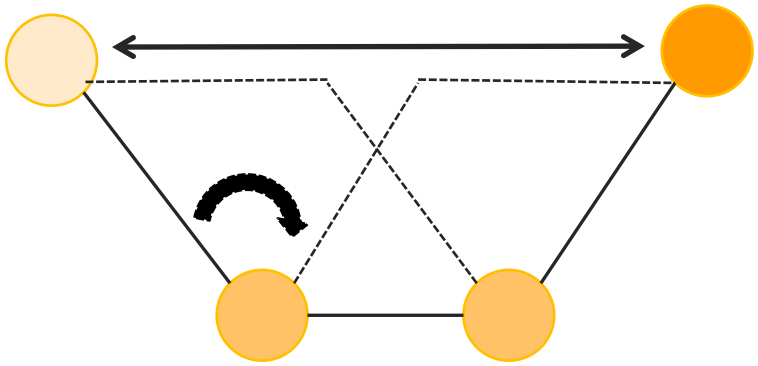
Post Optimisation

- Clash term may nevertheless still allow some clashes
- Allow further relaxation – torsion only

$$f_{relax} = f_{clash}(r) + f_{torsion\ restraint}(\tau)$$

$$f_{torsion\ restraint}(\tau) = c - k\ exp(v\cos(\tau - \tau_0))$$

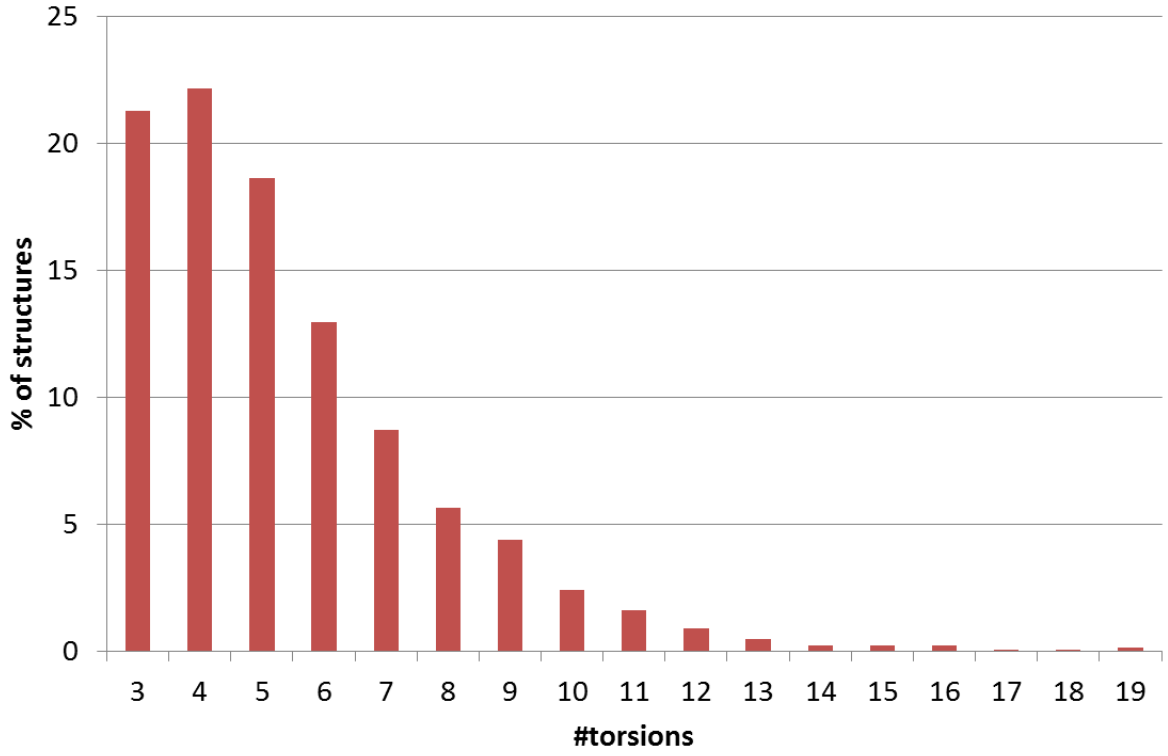
- τ_0 initial torsion



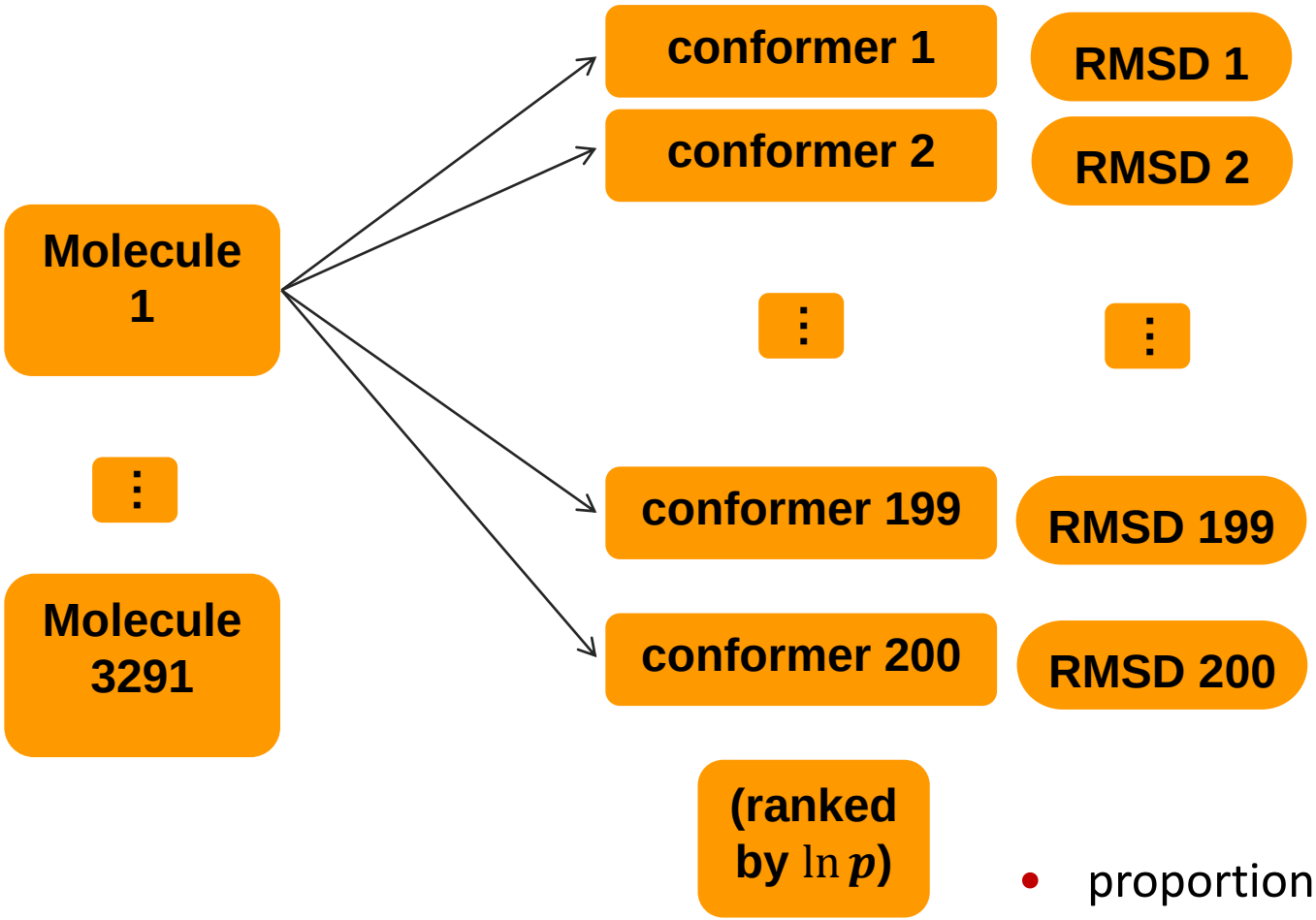
Test Set and Results

Test Set Description

- 3291 Corina structures based on CSD structures
- All excluded from geometry data libraries
- Average of 5-6 rotatable bonds per structure



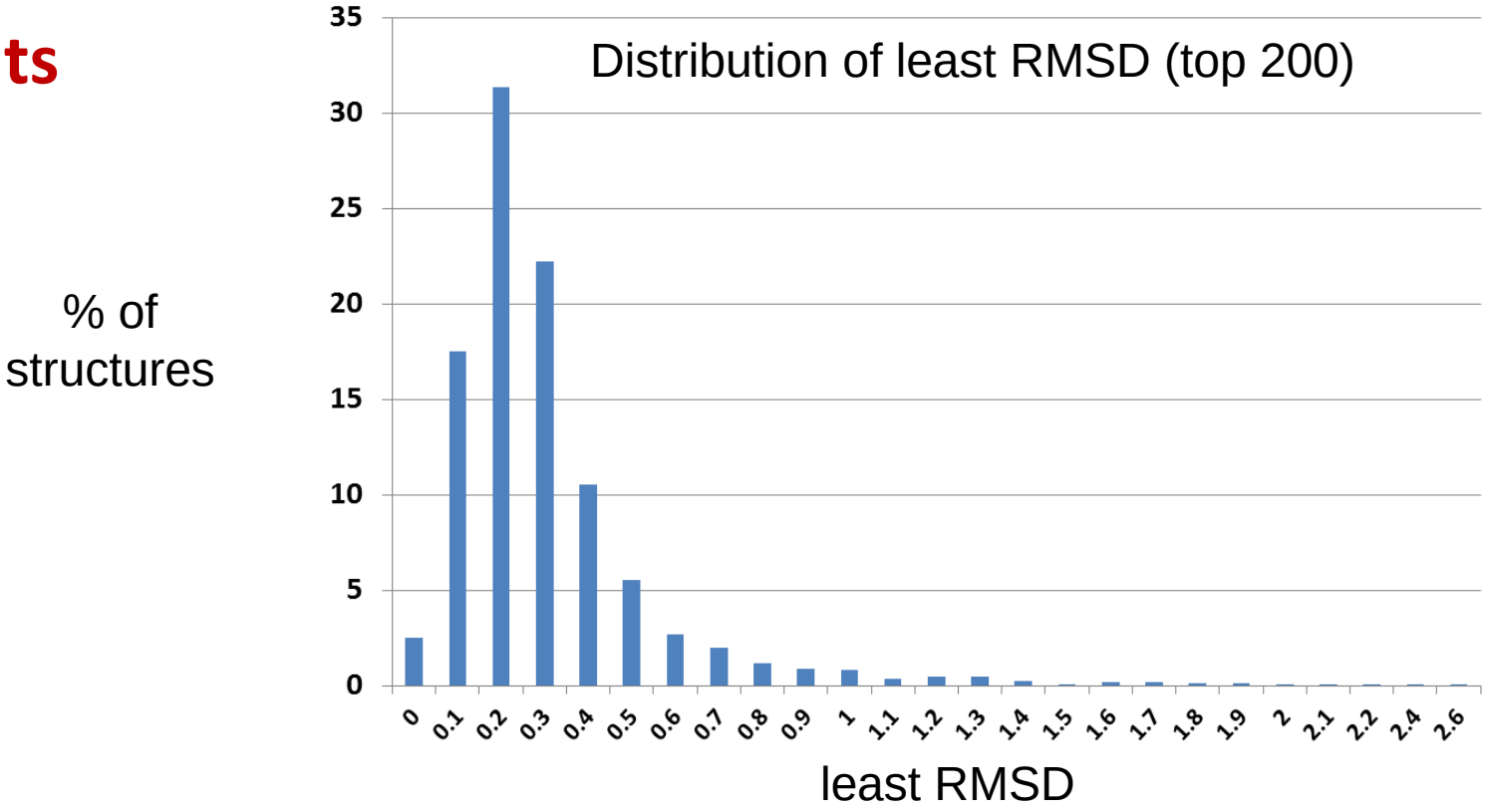
Results



- Average of least RMSD?
- proportion least RMSD< 0.5Å?

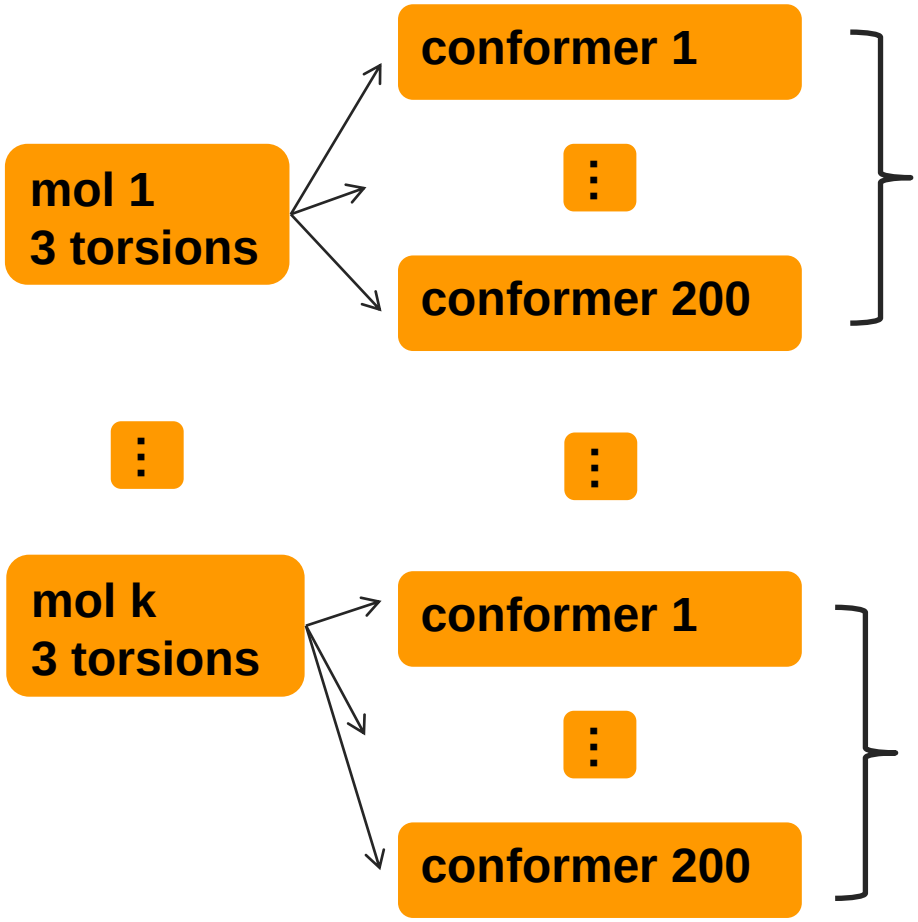
In top (1,50,100,200)

Results



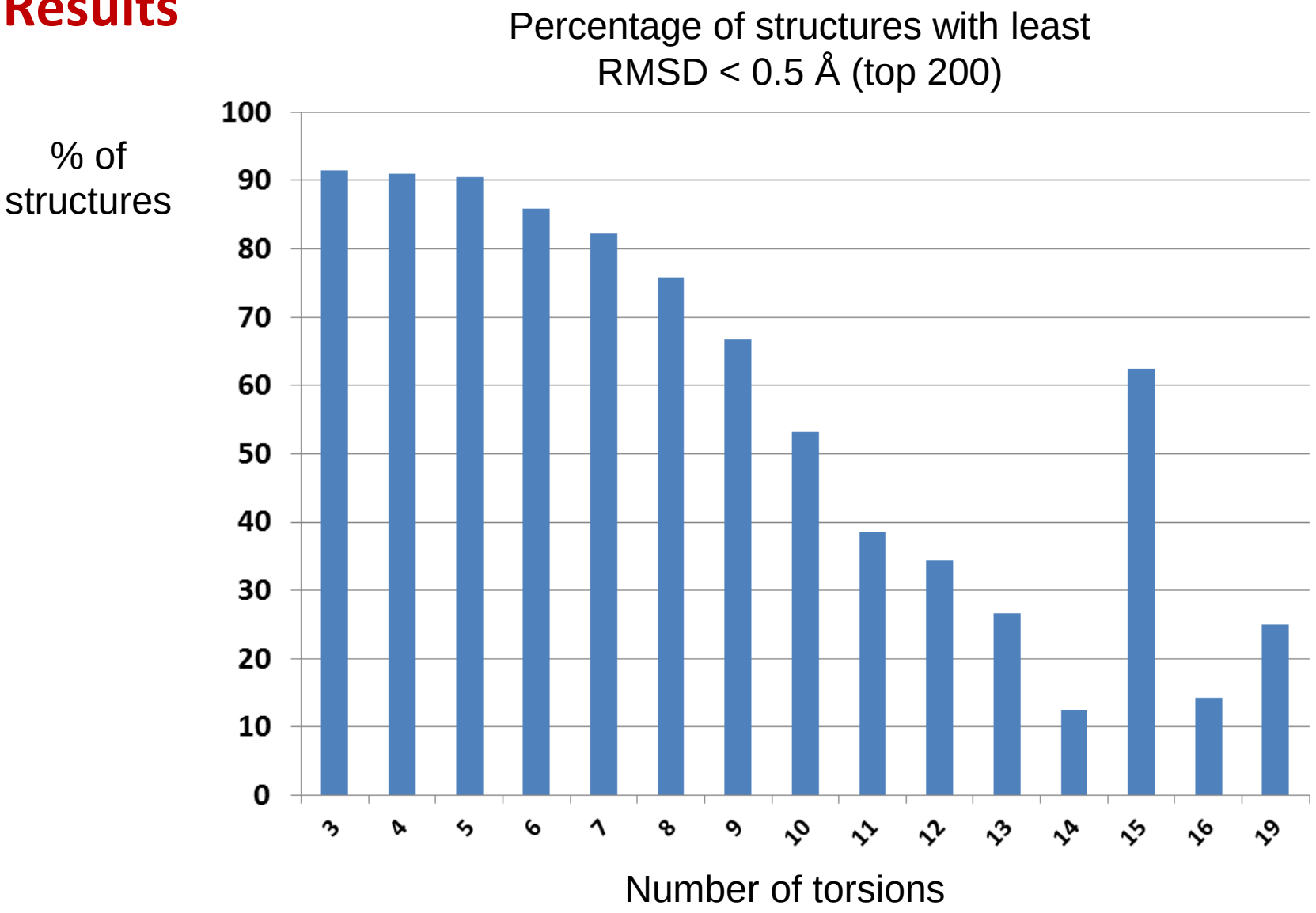
	Top n conformers				
n	1	25	50	100	200
least RMSD < 0.5 Å (%)	24	73	78	82	84
< least RMSD >	1.12	0.45	0.41	0.38	0.36

Results by Number of Torsions



- Structures with 3 torsions
- Repeat for structures with 4,5… torsions

Results


Percentage of structures with least RMSD < 0.5 Å (top 200)
% of structures
100
90
80
70
60
50
40
30
20
10
0
3
4
5
6
7
8
9
10
11
12
13
14
15
16
19
Number of torsions

Summary

- CSD knowledge based conformer generator
- Utilises probabilities of torsions and flexible rings
- Pruning techniques and efficient clash checking described
- Clustering
- RMSD results

Acknowledgements

- Oliver Korb
- Jason Cole
- Robin Taylor