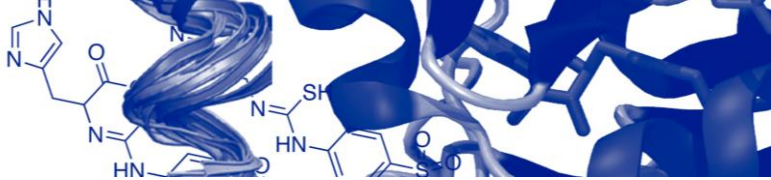
Computational Chemistry Group

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich



Unraveling Torsional Preferences:

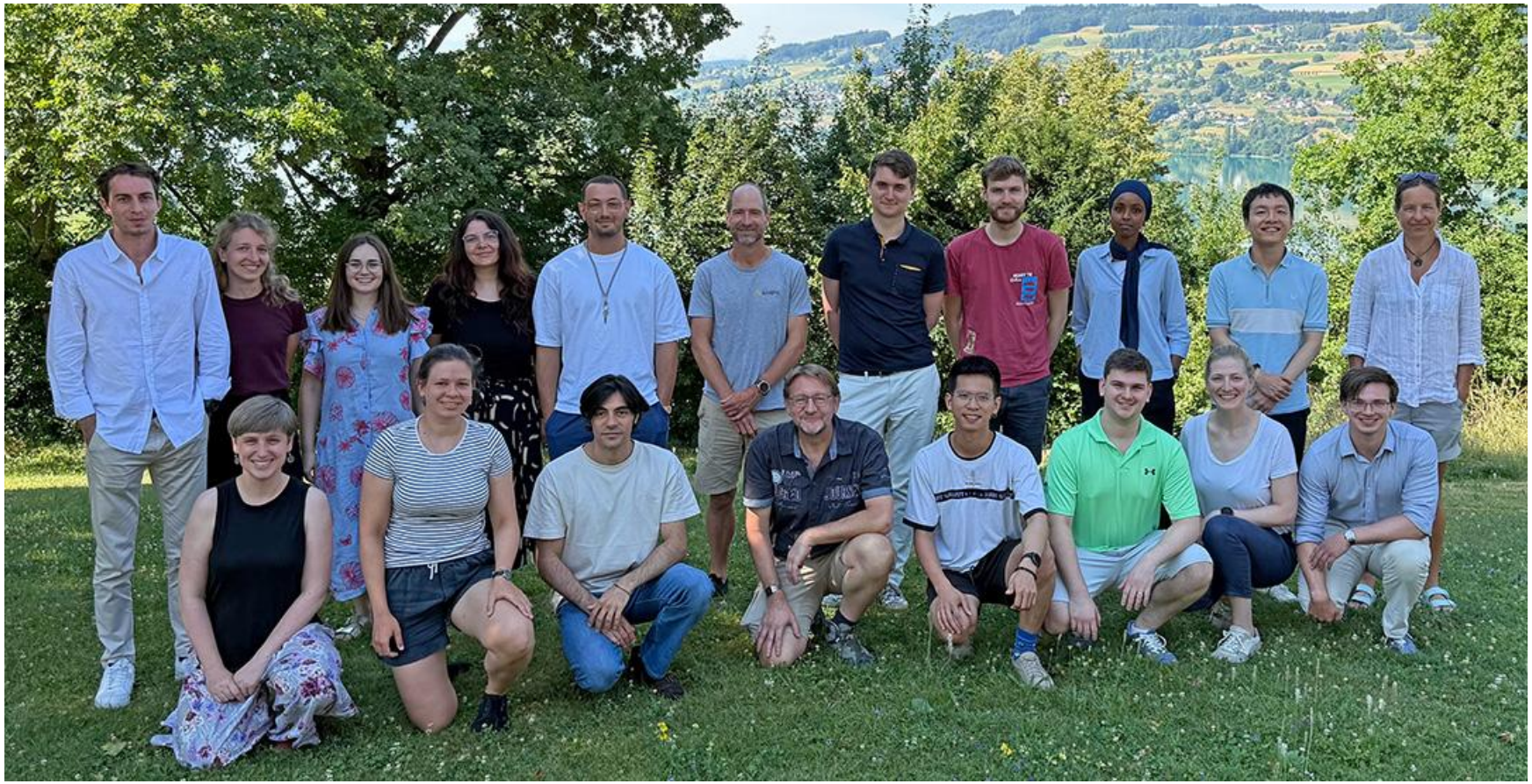
Torsion Motif Angle Distributions in Different Environments

RDKit UGM 2025

10.09.2025

jessica.braun@ethz.ch

Acknowledgements

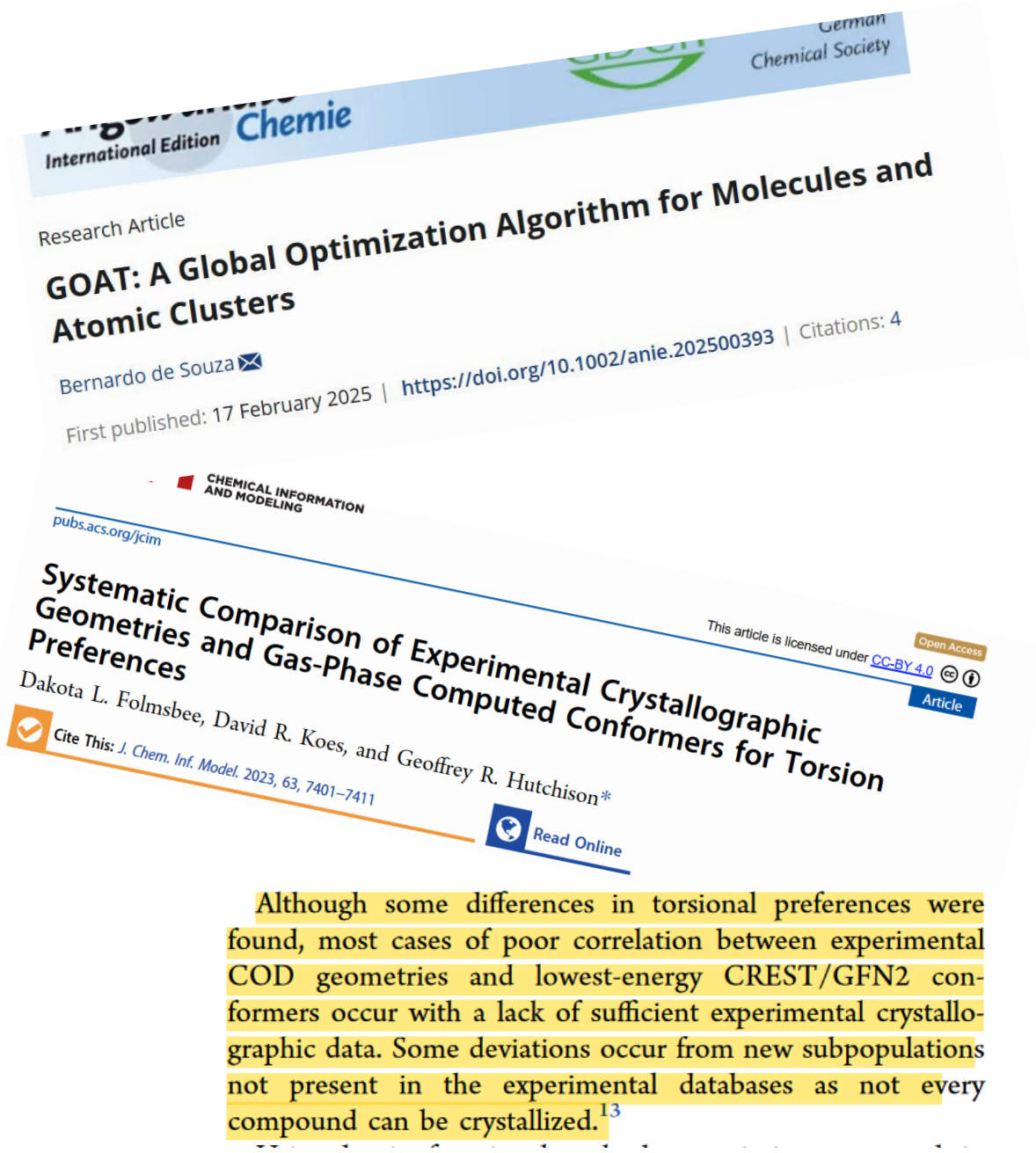


Sereina,
Greg,
Paul,
Djahan

Niels, Riccardo, Shu-Yu, Enrico, Carl

… the whole group!

“We shouldn't be surprised that a method which is designed to bias its output to reproduce crystal structure data doesn't perform well when asked to reproduce gas phase data.”



International Edition Chemie

German Chemical Society

Research Article

GOAT: A Global Optimization Algorithm for Molecules and Atomic Clusters

Bernardo de Souza

First published: 17 February 2025 | https://doi.org/10.1002/anie.202500393 | Citations: 4

CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim

Systematic Comparison of Experimental Crystallographic Geometries and Gas-Phase Computed Conformers for Torsion Preferences

Dakota L. Folmsbee, David R. Koes, and Geoffrey R. Hutchison*

This article is licensed under CC-BY 4.0

Open Access

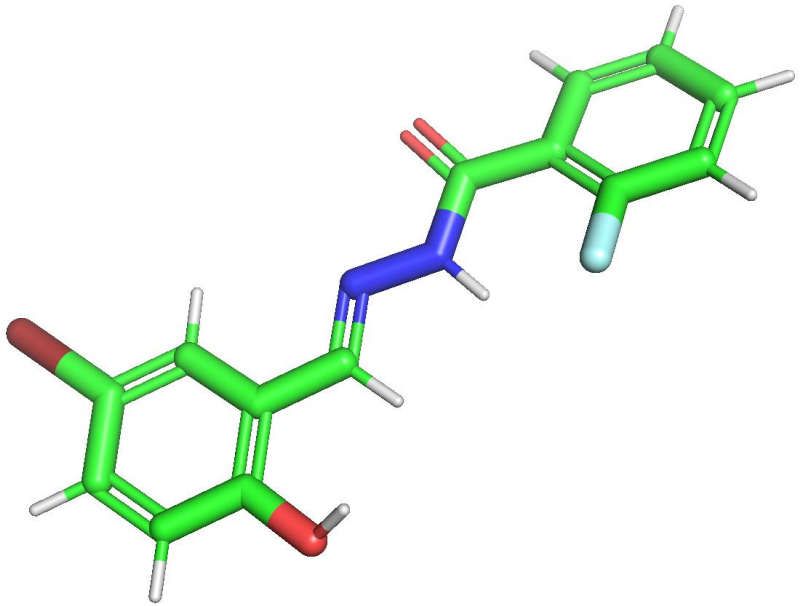
Article

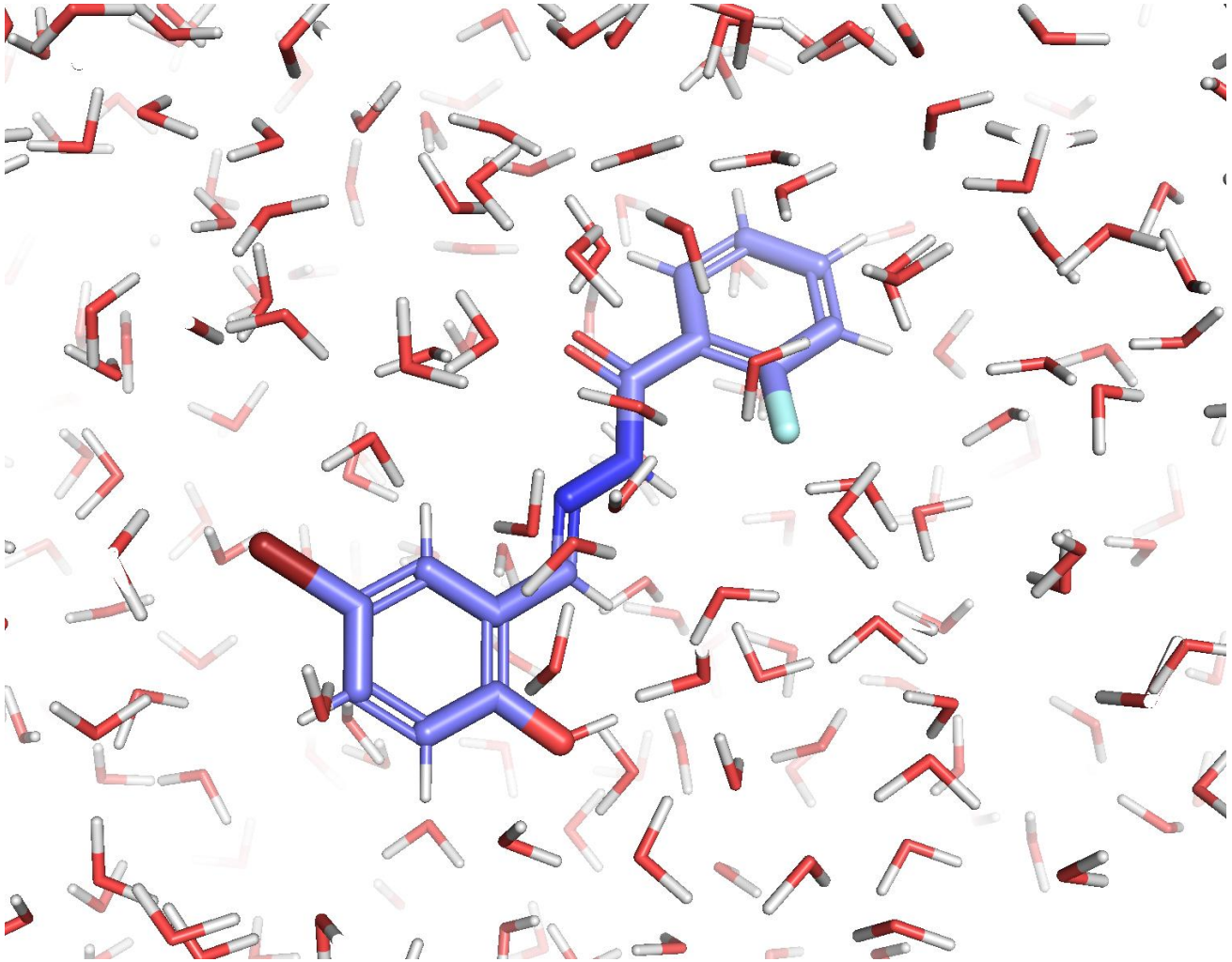
Cite This: J. Chem. Inf. Model. 2023, 63, 7401−7411

Read Online

Although some differences in torsional preferences were found, most cases of poor correlation between experimental COD geometries and lowest-energy CREST/GFN2 conformers occur with a lack of sufficient experimental crystallographic data. Some deviations occur from new subpopulations not present in the experimental databases as not every compound can be crystallized.[13]

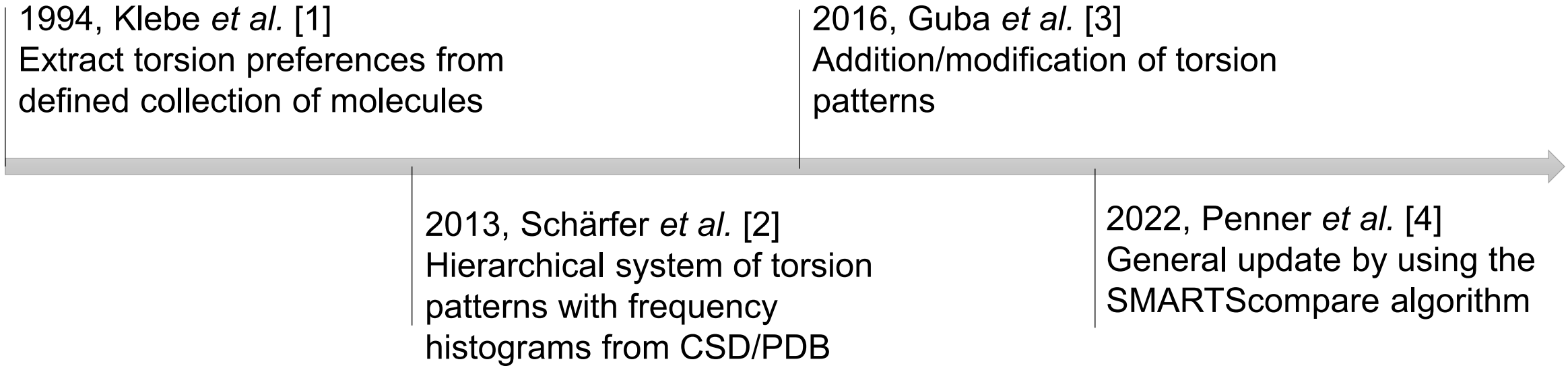
What Are We Comparing?





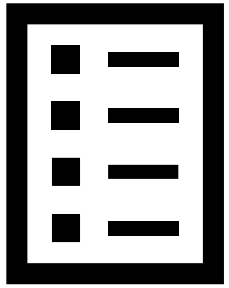
Torsion Motifs of the Torsion Library

- Hierarchical system of torsion patterns (SMARTS) covering druglike chemical space

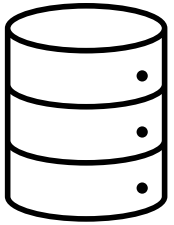


[1] *J. Comput. Aided Mol. Des.* 1994, 8, 583-606.
[2] *J. Med. Chem.* 2013, 56, 2016-28.
[3] *J. Chem. Inf. Model*. 2016, 56, 1, 1–5.
[4] *J. Chem. Inf. Model. 2022*, 62, 7, 1644–1653.

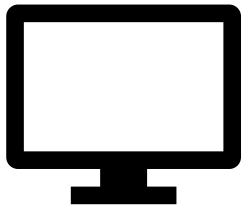
Experimental Setup



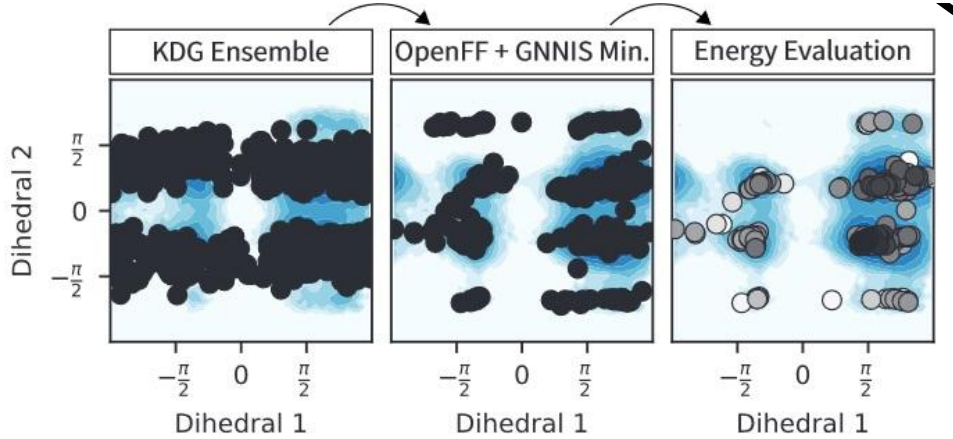
Hierarchical list of SMARTS patterns derived from the Torsion Library [3]

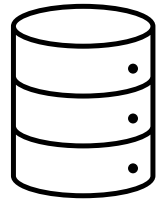


CSD 2024.03 [1]

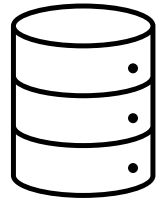


Rapid access to small molecule conformational ensembles with GNNIS [2]

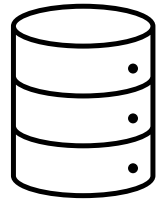




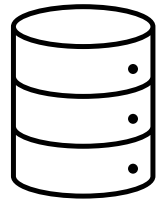
crystal



vacuum



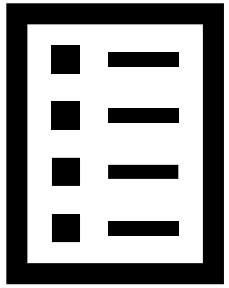
water



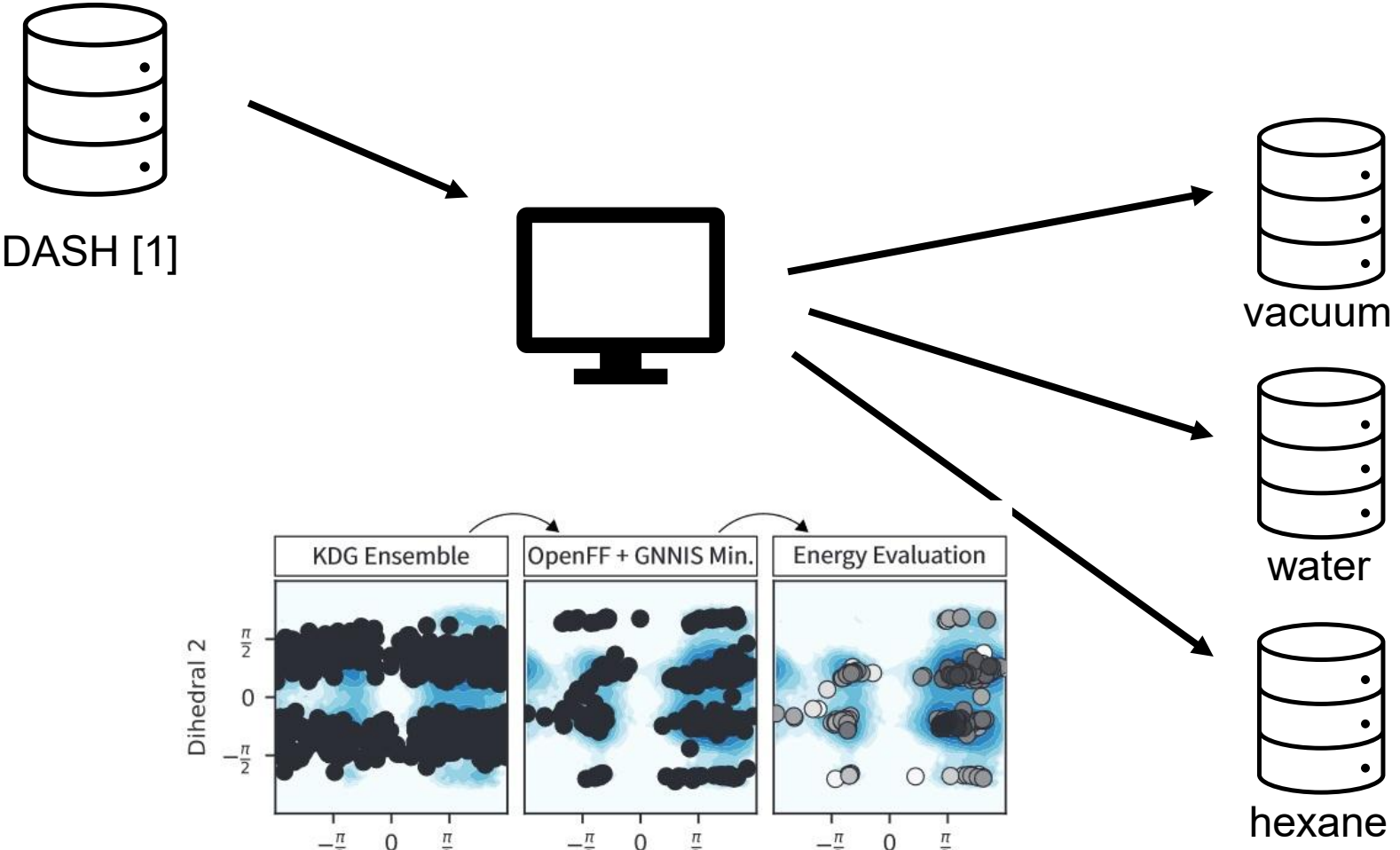
hexane

[1] *Acta Crystallogr B*. 2002, 58, 380-8.
[2] *J. Am. Chem. Soc*. 2025, 147, 16, 13264–13275

Experimental Setup



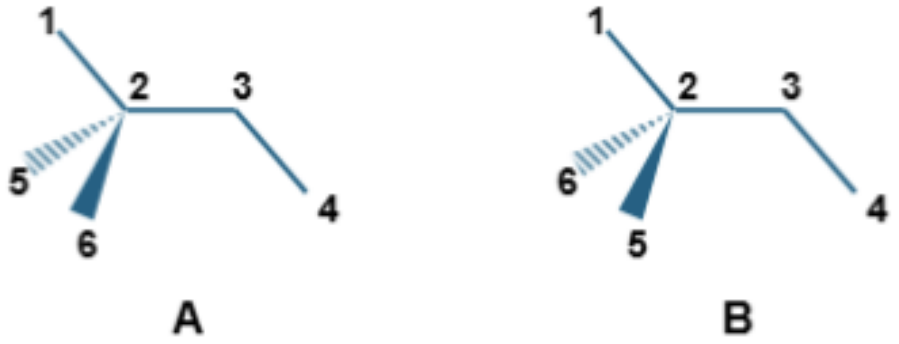
Hierarchical list of SMARTS patterns derived from the Torsion Library



[1] *Chem. Sci.*, 2024, 15, 10794-10802

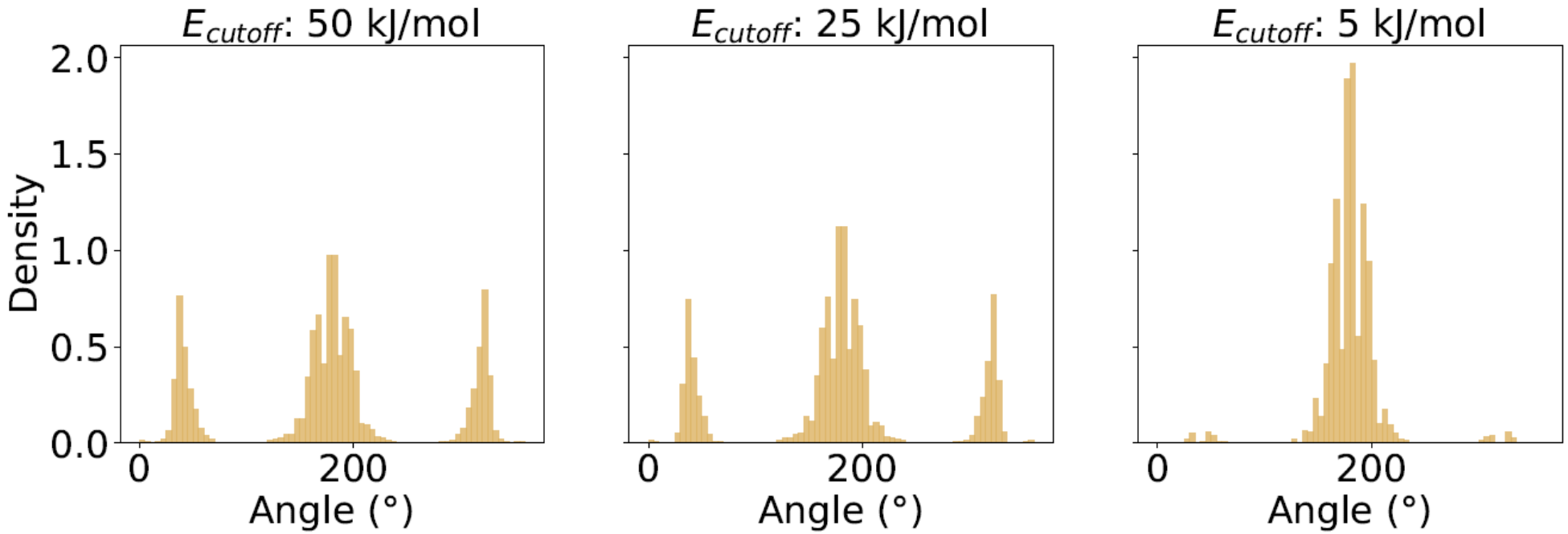
Notes on the Profile Extraction

- Accumulated torsion profiles (torsion motif angle distributions) are considered symmetric and modeled with symmetric functions.
- Torsion definitions in SMARTS patterns do not take symmetry into account, hence the mirror image would also match it. Although not observed, the mirror structure could always be constructed.



Note on the Energy Cutoff

SMARTS: [CX3H0:1]=[CX3H0:2]!@;-[CX3:3]=[CX3:4]



Overview: Comparisons

Environment 1	Environment 2		
	Vacuum	Hexane	Water
Crystal	X	X	X
Vacuum		X	X
Hexane			X

There are two comparison groups:

- Orange: comparisons between computational workflow and crystal.
- Blue: only comparisons between computational results.

Comparison of Torsion Profiles

- In literature, comparisons of torsion profiles are often case-based and expert-driven.
- Focus on characteristics of peaks:
  - Shift in peak positions
  - Changes in the relative peak heights
  - Peak appearances/disappearances
- The metric of change needs to be sensitive to both local and global changes in mass and shape.

Comparison Metrics

(2nd order) Wasserstein distance

Wasserstein distances are describing the minimal transport cost from one distribution to another:

$$W_p(\mu,\nu) = inf_{X\sim\mu,Y\sim\nu}\left(\boldsymbol{E}\left||X-Y|\right|^p\right)^{\frac{1}{p}}$$

p is the order of the Wasserstein distance.

Cramér-von Mises

$$\Delta(F,F_0) = \int_{-\infty}^{\infty}\left(F(x)-F_0(x)\right)^2 dF_0(x)$$

$$N\omega^2 = \frac{NM}{N+M}\int_{-\infty}^{\infty}\left(F_N(x)-G_M(x)\right)^2 dH_{N+M}(x)$$

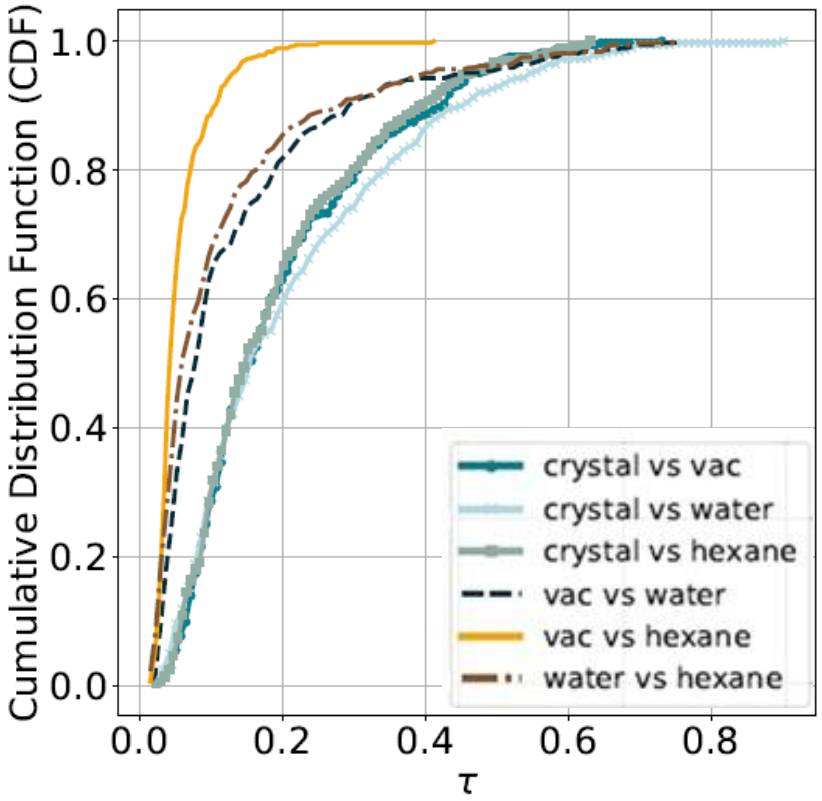
Comparison Metrics: Combined Score τ

- Take both metrics into account, by defining a combined score τ.

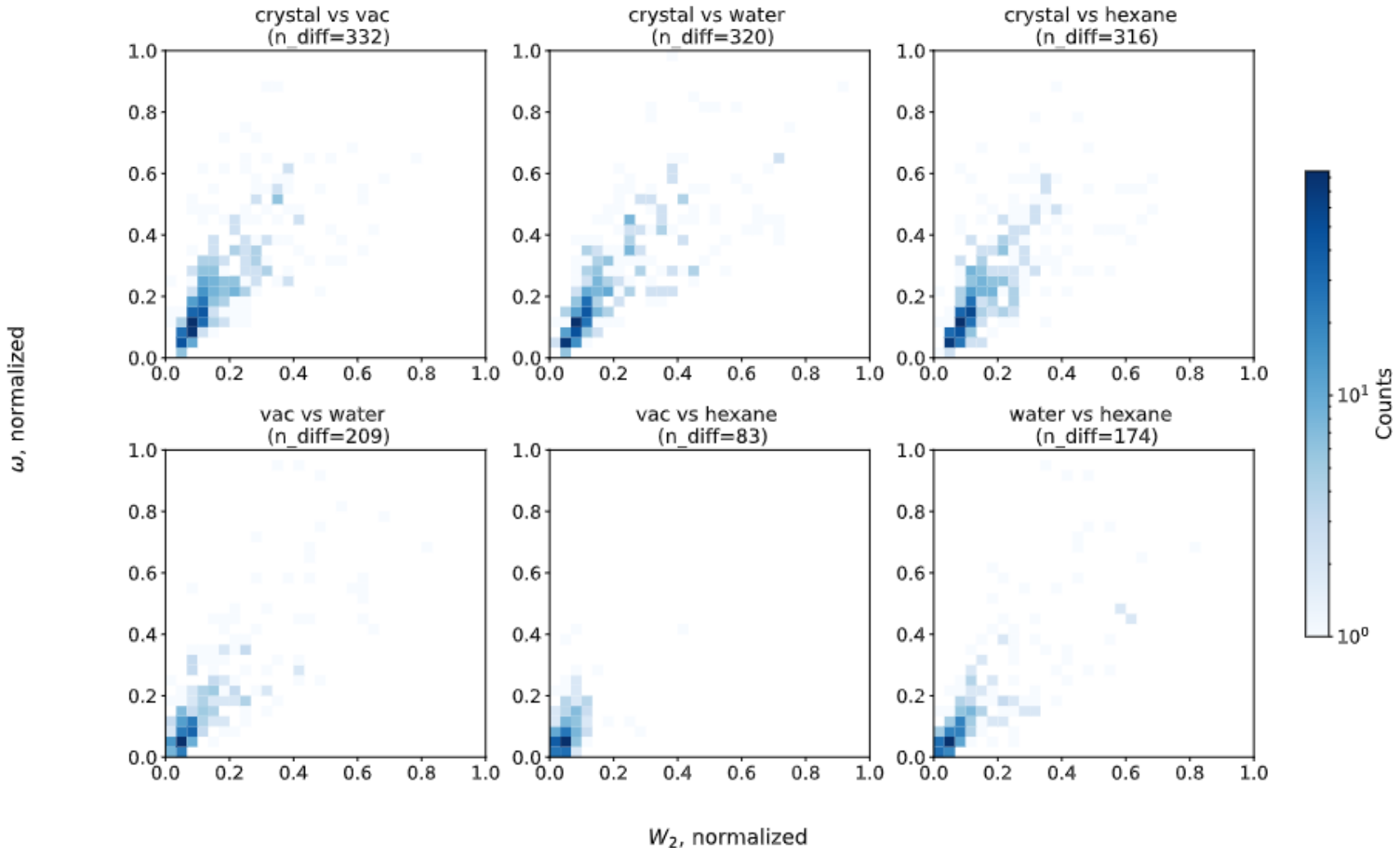
$$\tau = \frac{W_{2,norm} + \omega_{2,norm}}{2}$$

- Normalization performed by constructing the theoretical maximum value both metrics can take.

Results: TMADs Across Environments

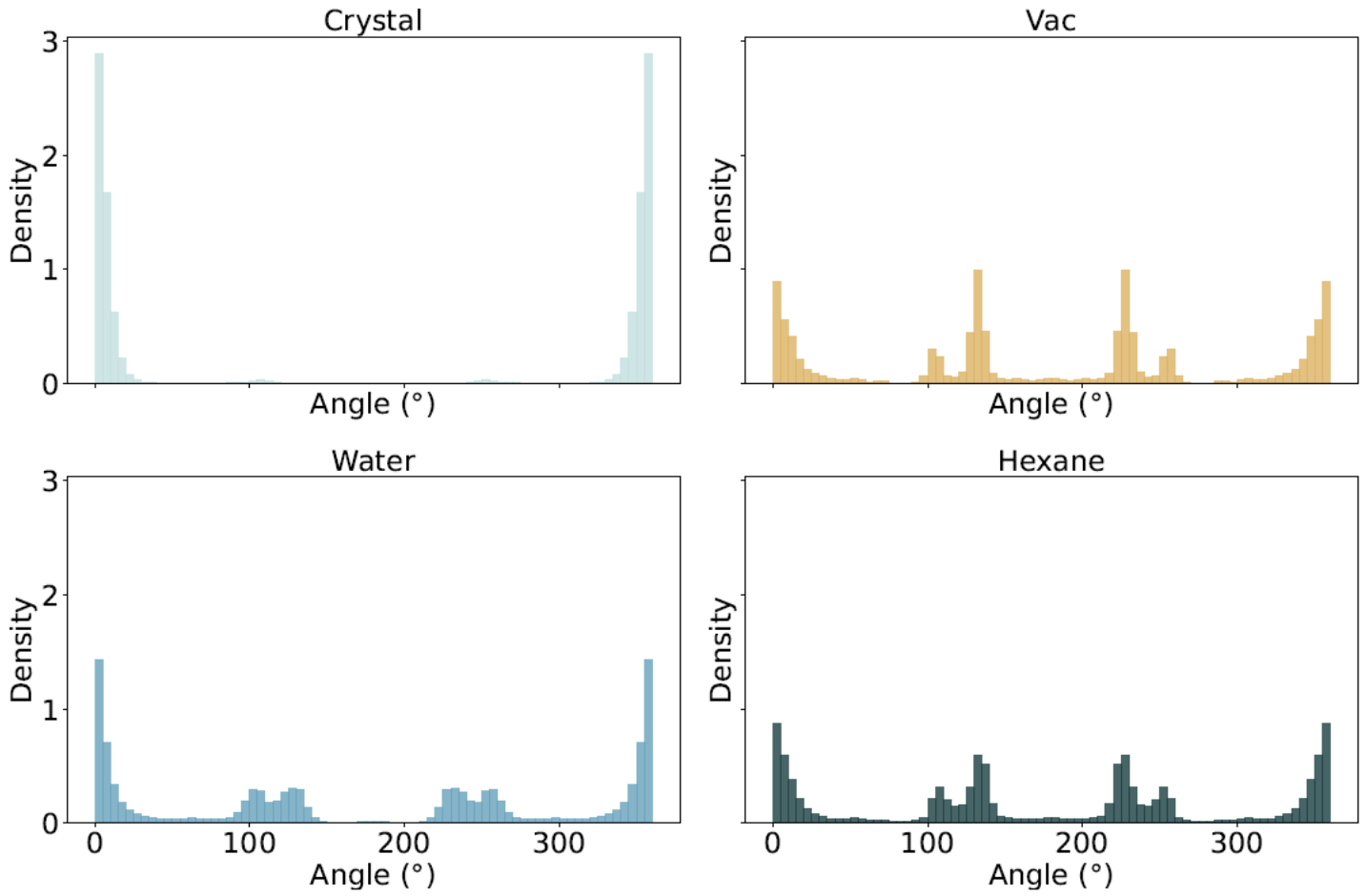


Results: TMADs Across Environments

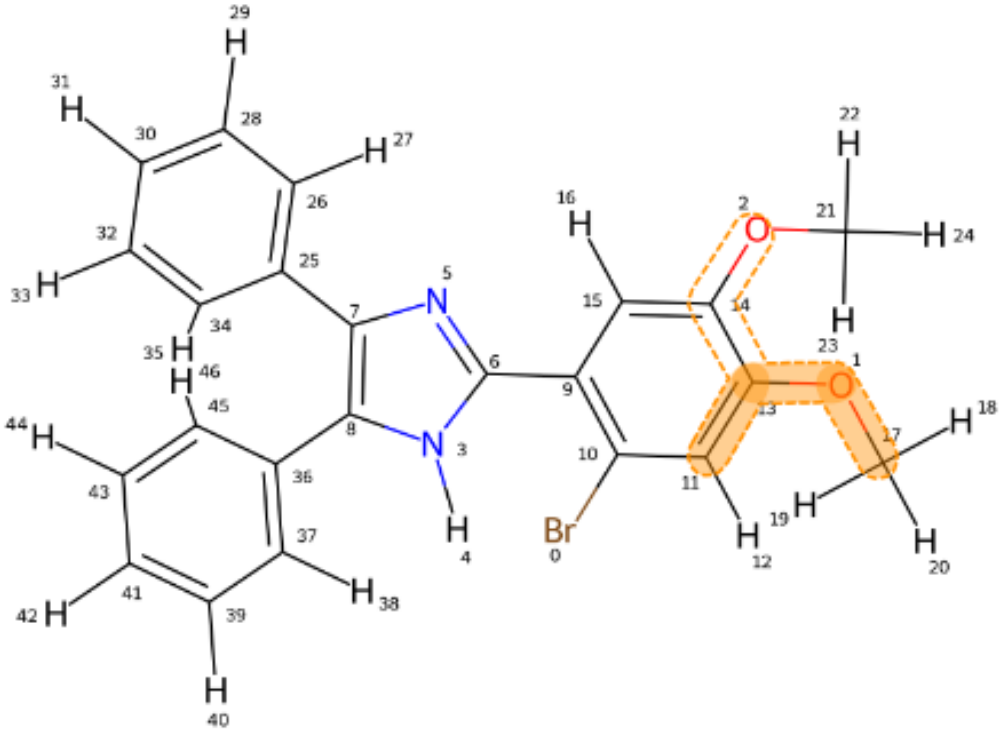


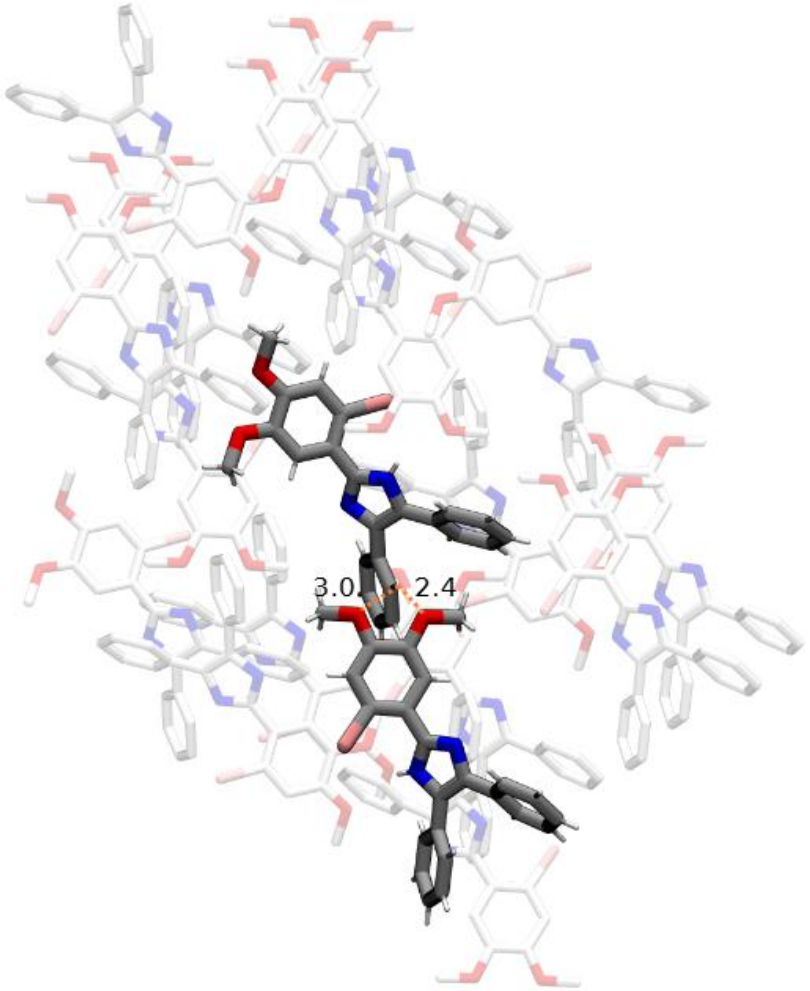
Results: Crystal Packing Effects

SMARTS: [cH1:1][c:2](cO)!@;-[O:3][C:4]



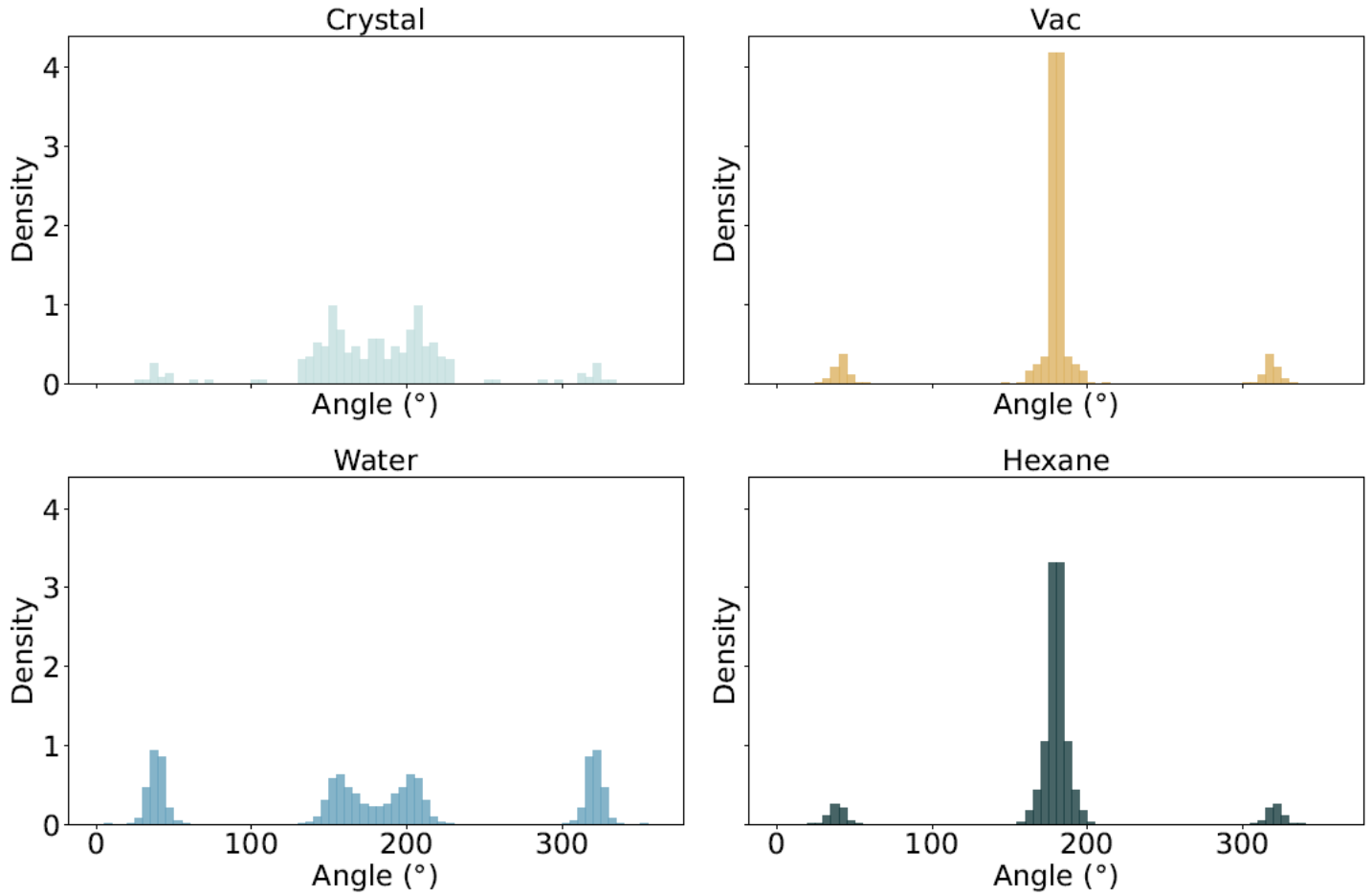
Results: Crystal Packing Effects



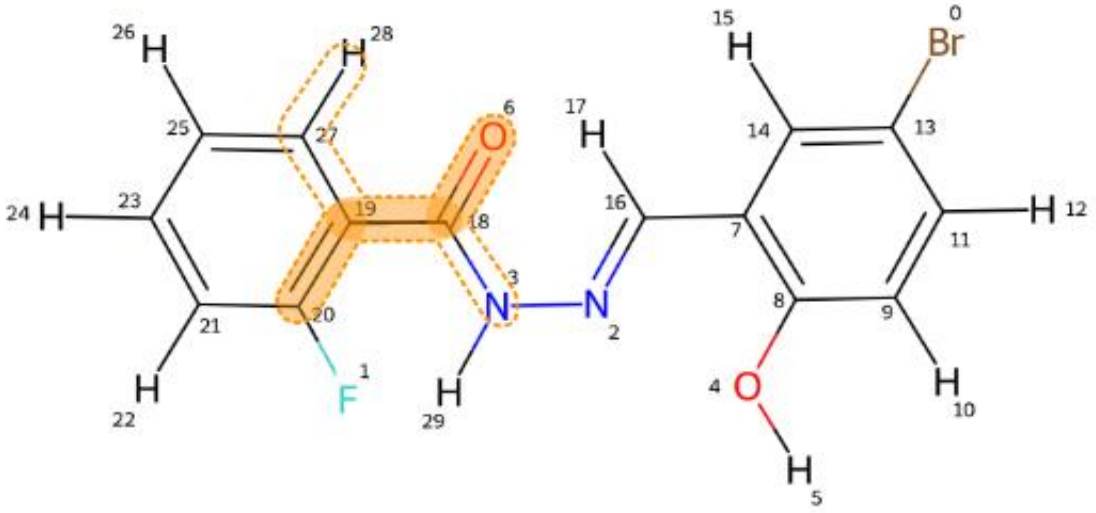


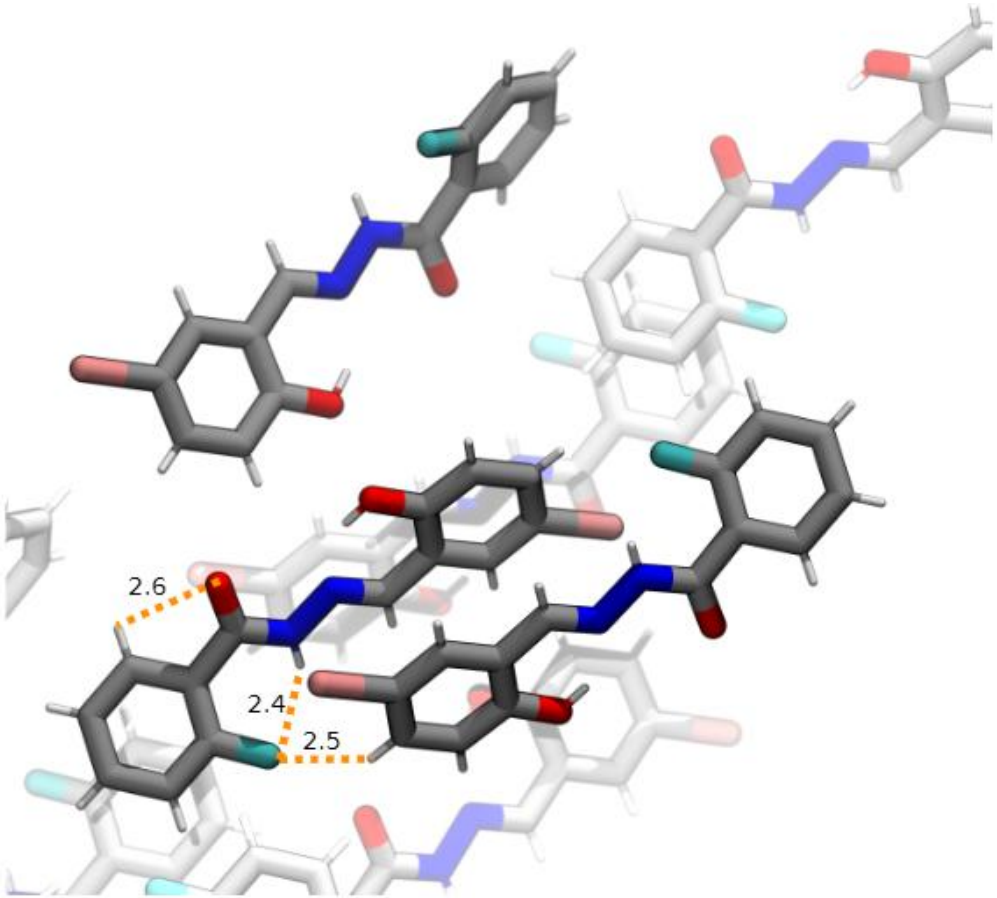
Results: Solute-Solvent Interactions In Water

SMARTS: [$([cH0]F):1][c:2]([cH1])!@;-[CX3:3]([NX3H1])=[O:4]

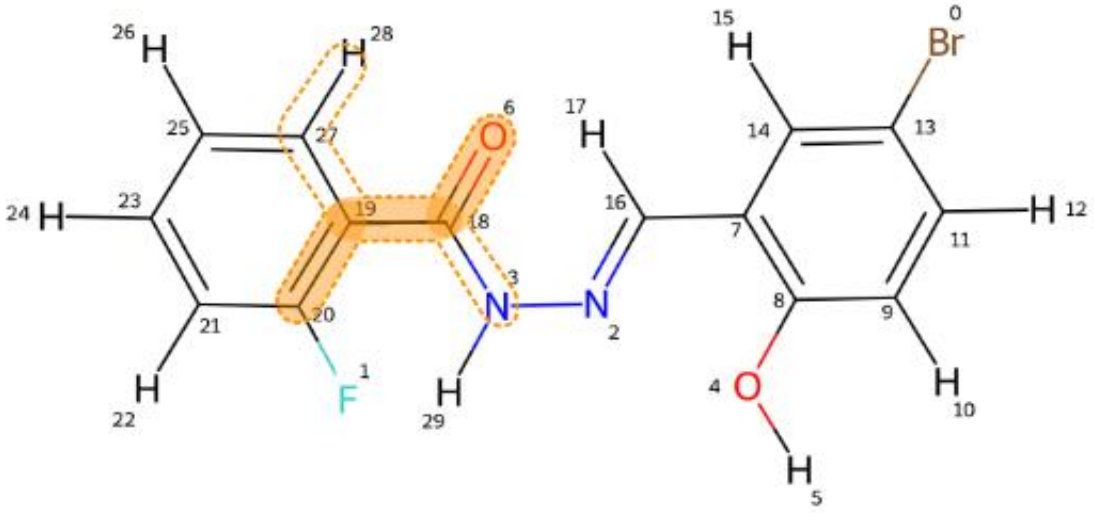


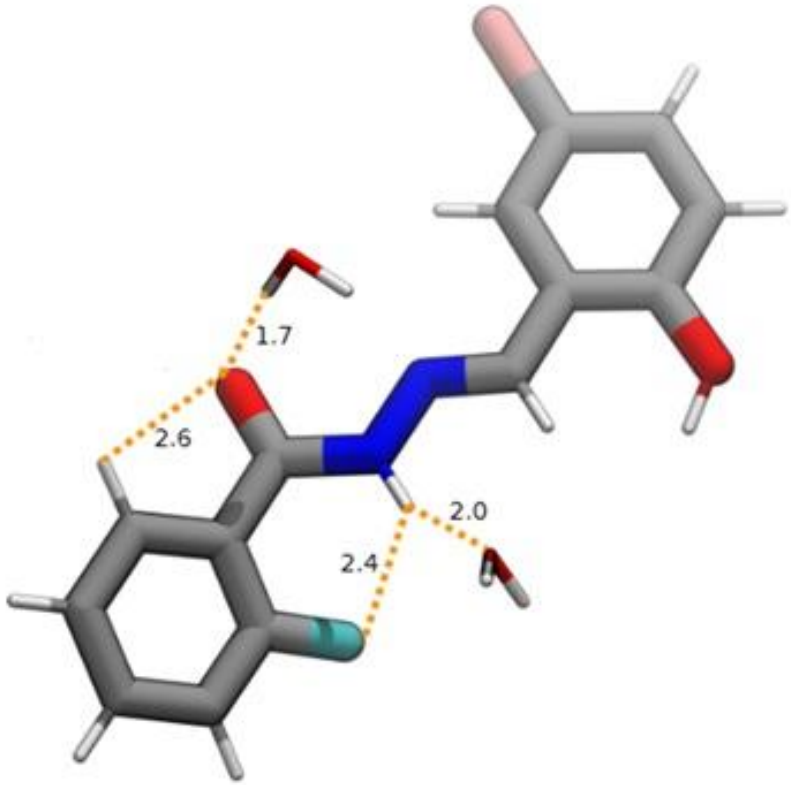
Results: Solute-Solvent Interactions In Water





Results: Solute-Solvent Interactions In Water

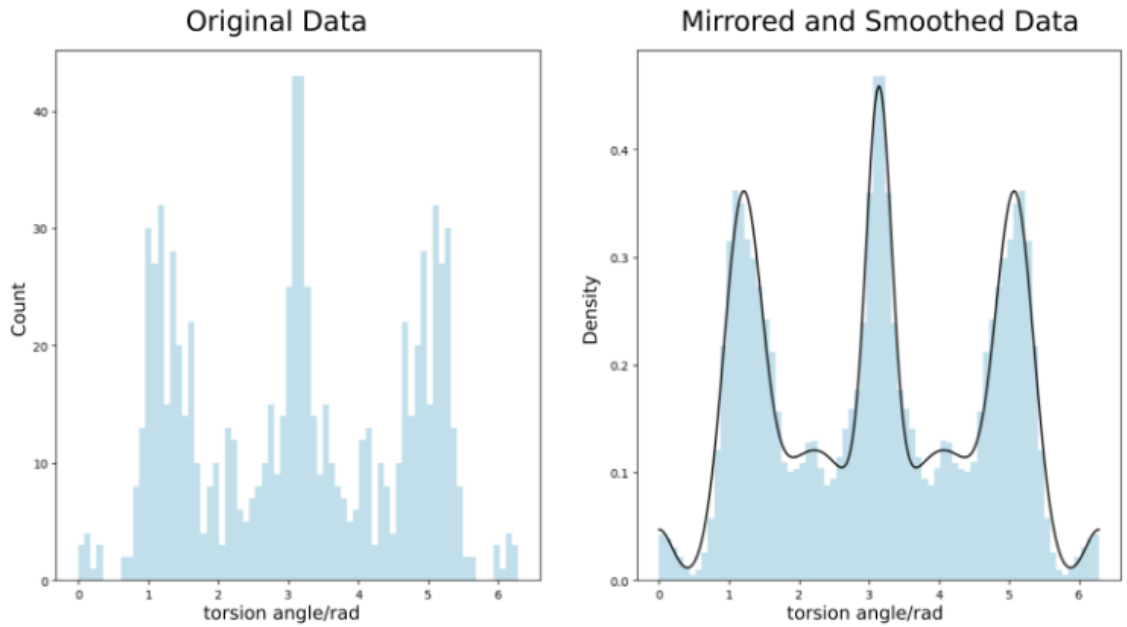




Automated Fitting of Torsion Distributions

All obtained torsion distributions were fitted with Wrapped Mirrored Gaussians:

$$P(x) = \sum_{i=1}^{3} a_i \left(\exp\left(-\left(\frac{x - b_i}{c_i} \right)^2 \right) + \exp\left(-\left(\frac{2\pi - x - b_i}{c_i} \right)^2 \right) + \exp\left(-\left(\frac{x - b_i + 2\pi}{c_i} \right)^2 \right) + \exp\left(-\left(\frac{4\pi - x - b_i}{c_i} \right)^2 \right) \right)$$



Conclusions

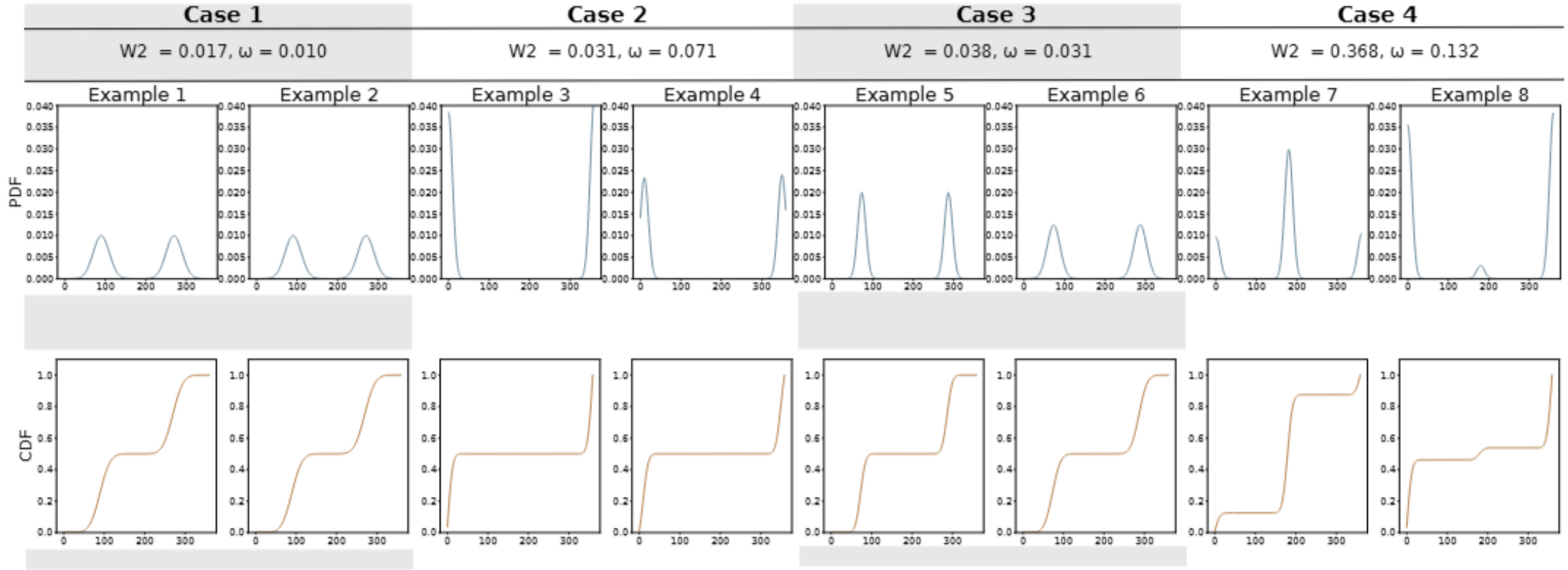
There are significant differences in the Torsion Motif Angle Distributions (TMADs) for the different environments.

Through case studies, we were able to rationalize observed trends including the higher similarity of torsion profiles in hexane and vacuum than hexane and water.

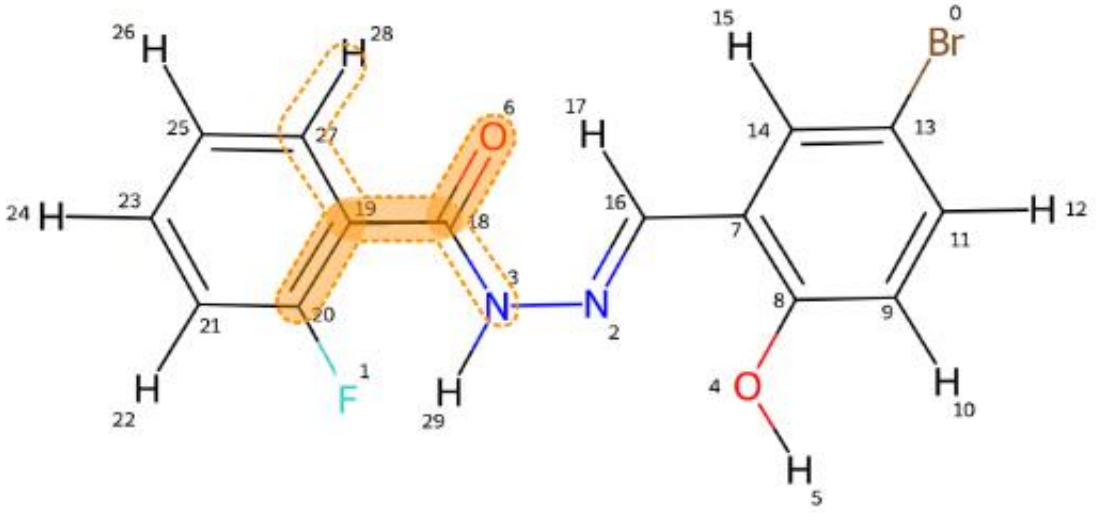
The experimental crystal environment showed the largest difference from the other computed environments.

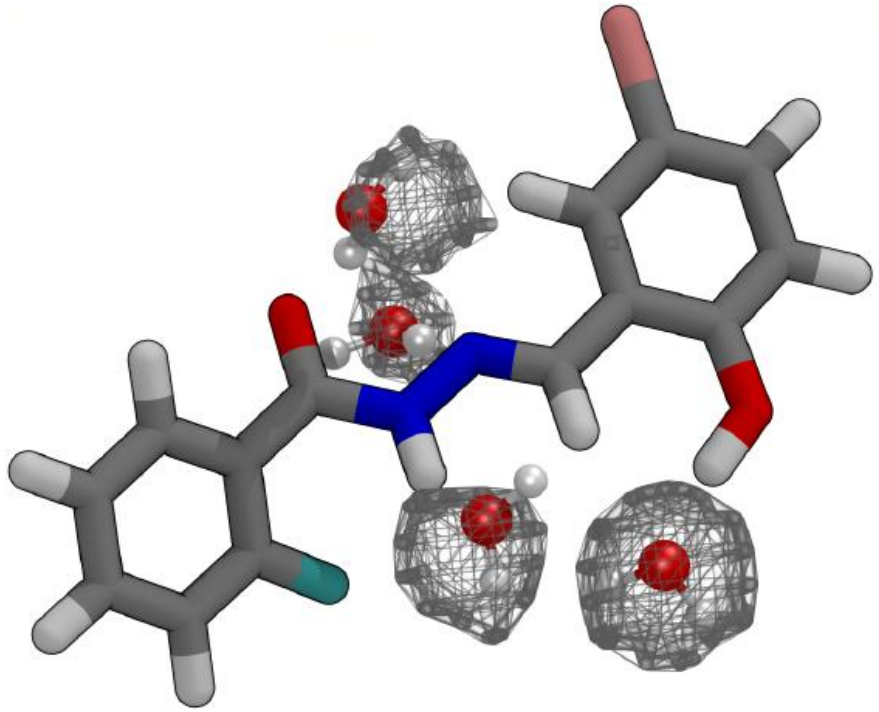
Thanks for listening! Questions?

Differences of W2 and ω



Results: Solute-Solvent Interactions In Water





Results: Data Set Differences