

The Open Force Field Initiative: A brief update on our progress and plans

The Open Force Field Group

openforcefield.org



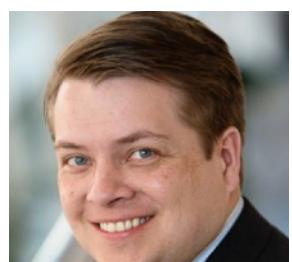
JOHN CHODERA
SLOAN KETTERING INSTITUTE



MICHAEL GILSON
UNIVERSITY OF CALIFORNIA, SAN DIEGO



DAVID MOBLEY
UNIVERSITY OF CALIFORNIA, IRVINE



MICHAEL SHIRTS
UNIVERSITY OF COLORADO, BOULDER



LEE-PING WANG
UNIVERSITY OF CALIFORNIA, DAVIS

CONSULTANTS/ADVISORS



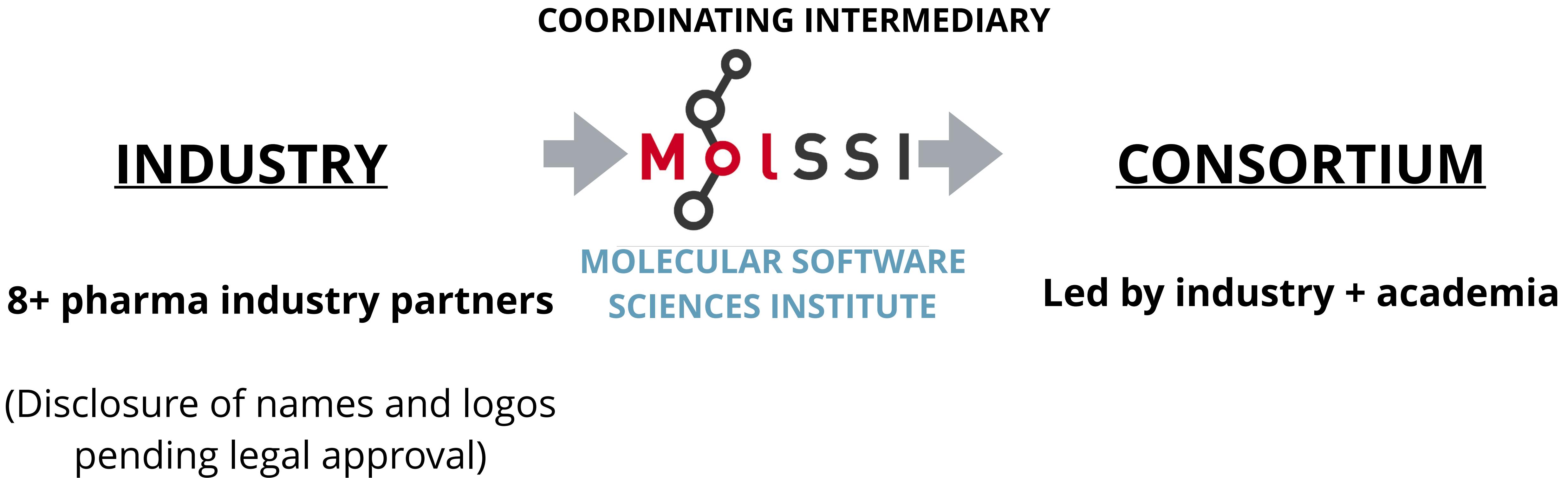
CHRISTOPHER BAYLY
OPENEYE SCIENTIFIC



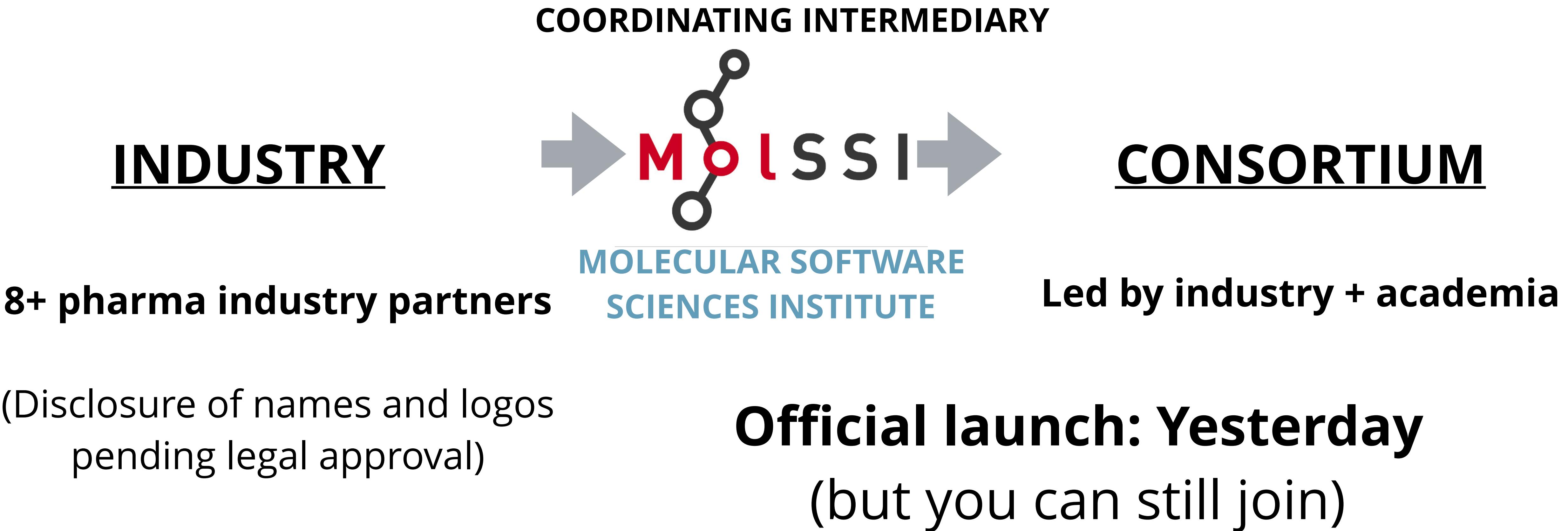
KENNETH KROENLEIN
**NIST THERMODYNAMICS RESEARCH
CENTER**

(NIST is a US federal agency)

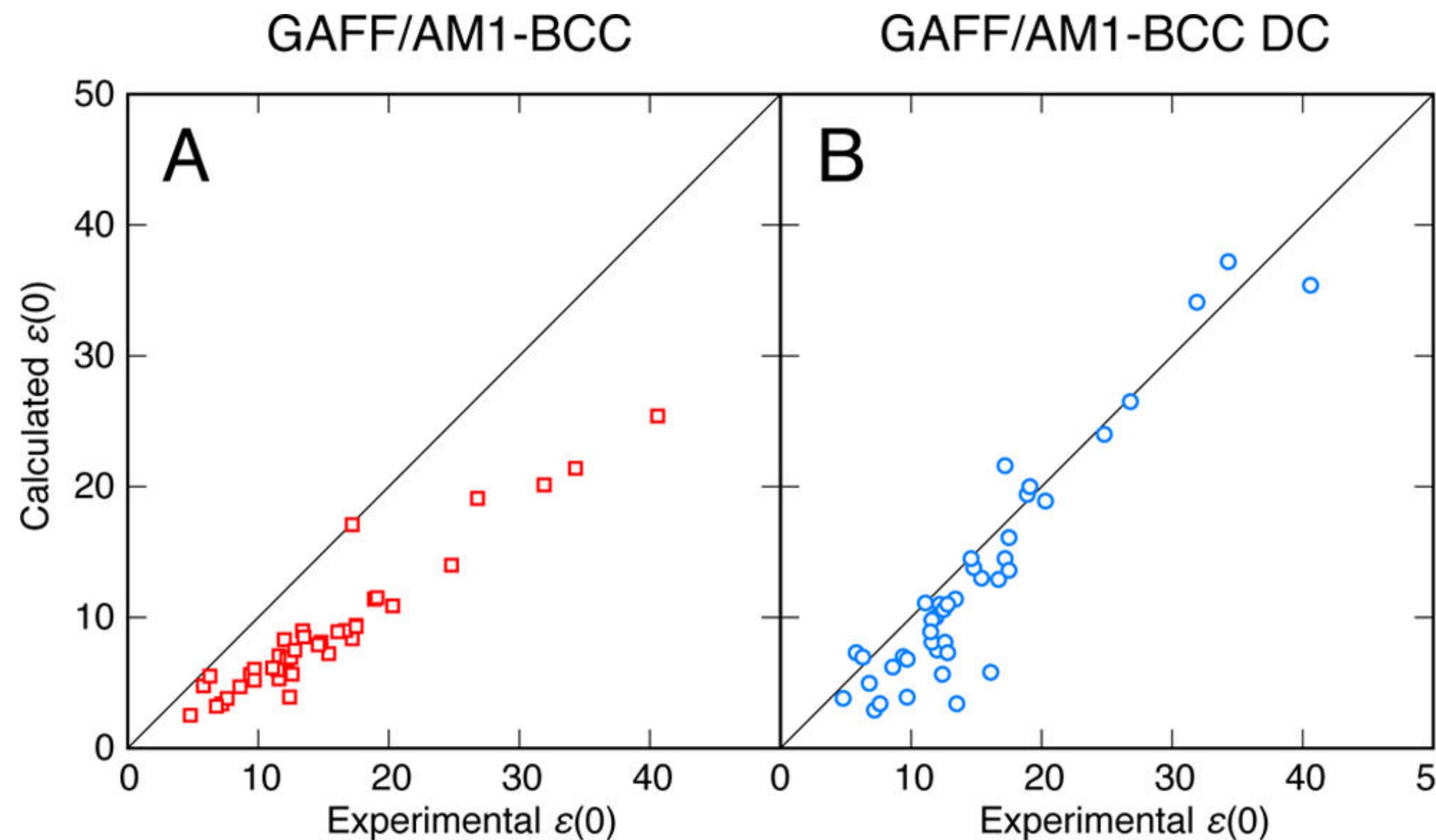
We're delighted to announce the Open Force Field Consortium, building new open force fields for the community



We're delighted to announce the Open Force Field Consortium, building new open force fields for the community



Why new force fields? We already know we can do better than today's force fields without new physics

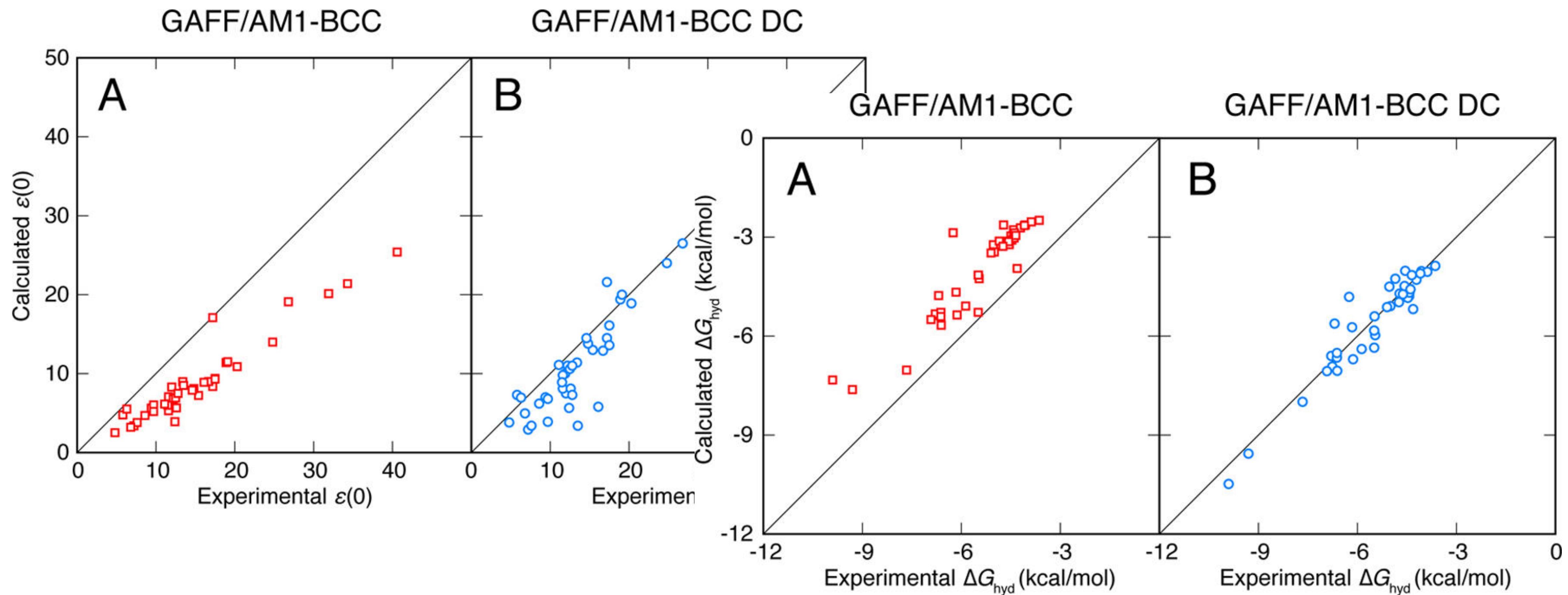


e.g. hydroxyl dielectrics
and hydration free energies

but many other functional groups
too; see e.g. 10.1021/ct800409d

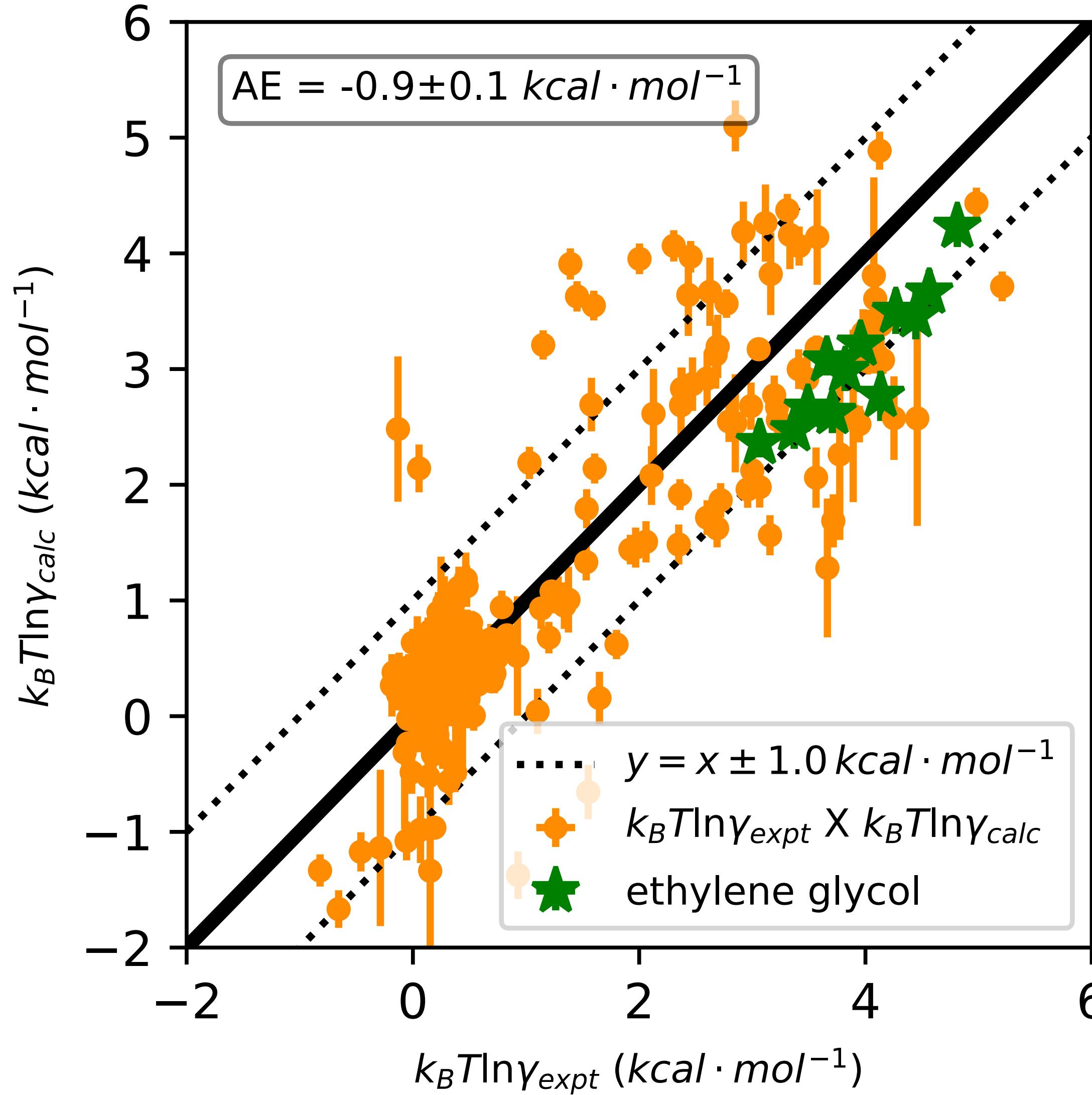
Many functional groups in today's force fields have serious systematic errors that await a general fix, because refitting is too hard

Why new force fields? We already know we can do better than today's force fields without new physics



Many functional groups in today's force fields have serious systematic errors that await a general fix, because refitting is too hard

Similar issues stare at us everywhere we look



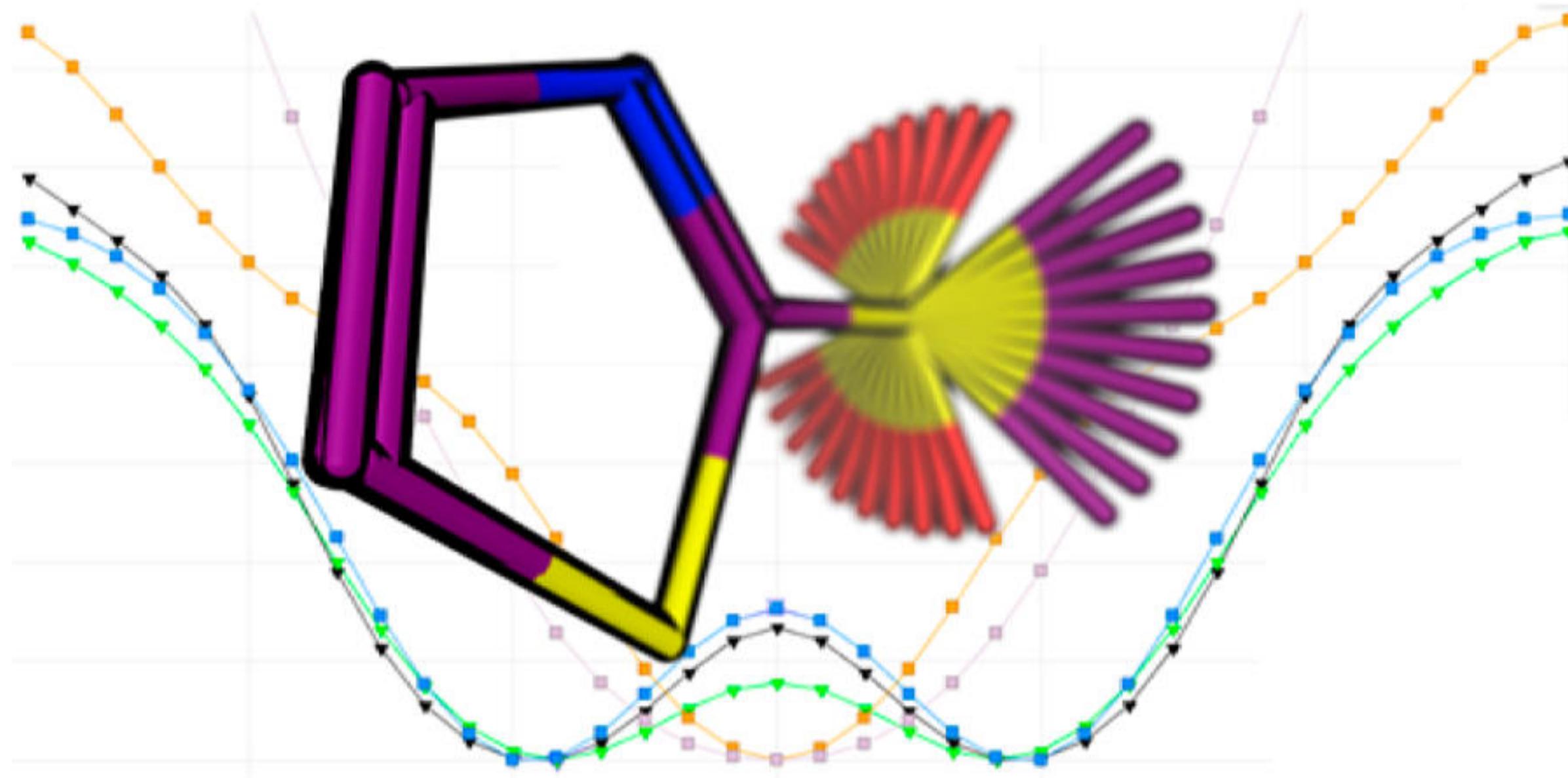
Infinite dilution activity coefficients are untapped for force field development and show clear systematic errors

They inform on relative solvation in different solvents

Here, ethylene glycol is poorly represented as a solvent

(Calculations of all the suitable activity coefficients from NIST's ThermoML done overnight on Orion)

We all have our own war stories, too



JOURNAL OF
**CHEMICAL INFORMATION
AND MODELING**

Article

pubs.acs.org/jcim

A Comparison of Quantum and Molecular Mechanical Methods to Estimate Strain Energy in Druglike Fragments

Benjamin D. Sellers,*[●] Natalie C. James, and Alberto Gobbi

Department of Discovery Chemistry, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, United States

Different force fields might not even agree on the location of a minimum

The consortium seeks to improve force field science, accelerate progress

Make force field fitting/extension straightforward by automating derivation of:

- Numerical parameters
- Classification of atom types, bond types, etc.
- Confidence metrics

Generate/curate open datasets necessary for producing high-accuracy biomolecular forcefields: We need data we can use to automatically fit FFs

Build new FFs

OpenFF is a truly open effort

OPEN SOURCE

Everyone will have access to the (free) toolkit for building/improving forcefields

OPEN DATA

Collected and curated datasets made available

OPEN SCIENCE

Everything done in the open on GitHub; all forcefields open

All code, datasets, and forcefields: <https://github.com/openforcefield>

parm@frosst is the starting point for a GAFF-like small molecule forcefield using SMARTS/SMIRKS

A Second Generation Force Field for the Simulation of Organic Molecules

http://www.ccl.net/cca/data/parm_at_Frosst/index.shtml

CCL An Informal AMBER Small Molecule Force Field: parm@Frosst

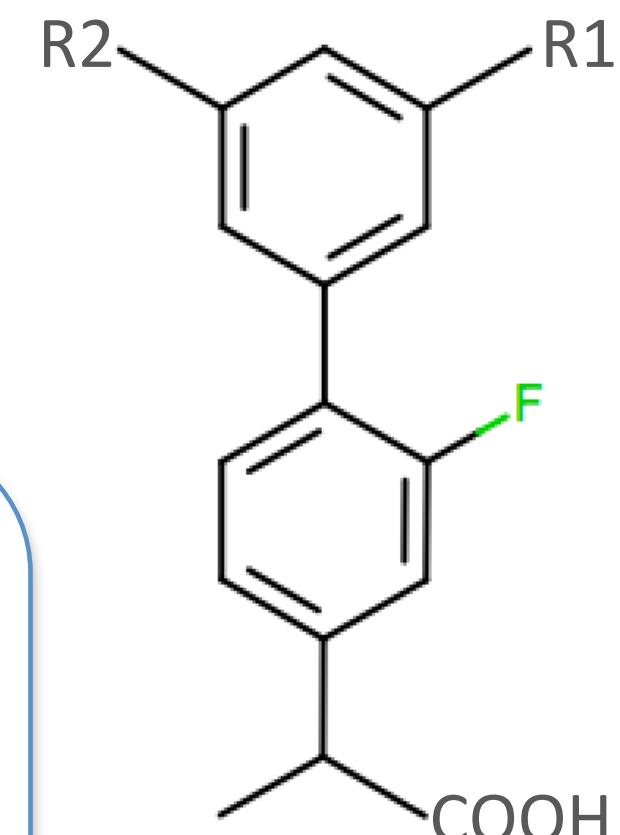
An Informal AMBER Small Molecule Force Field: **parm@Frosst**

Christopher Bayly, lead the effort between (1992-2010)

Daniel McKay, contributed between (1997-2010)

Jean-François Truchon, contributed between (2002-2010)

Volume 9, Issue 3, 8 February 1999, Pages 307-312



Bayly et al.'s parm@Frosst is an AMBER-family small molecule force field and sibling of GAFF

SMIRNOFF:

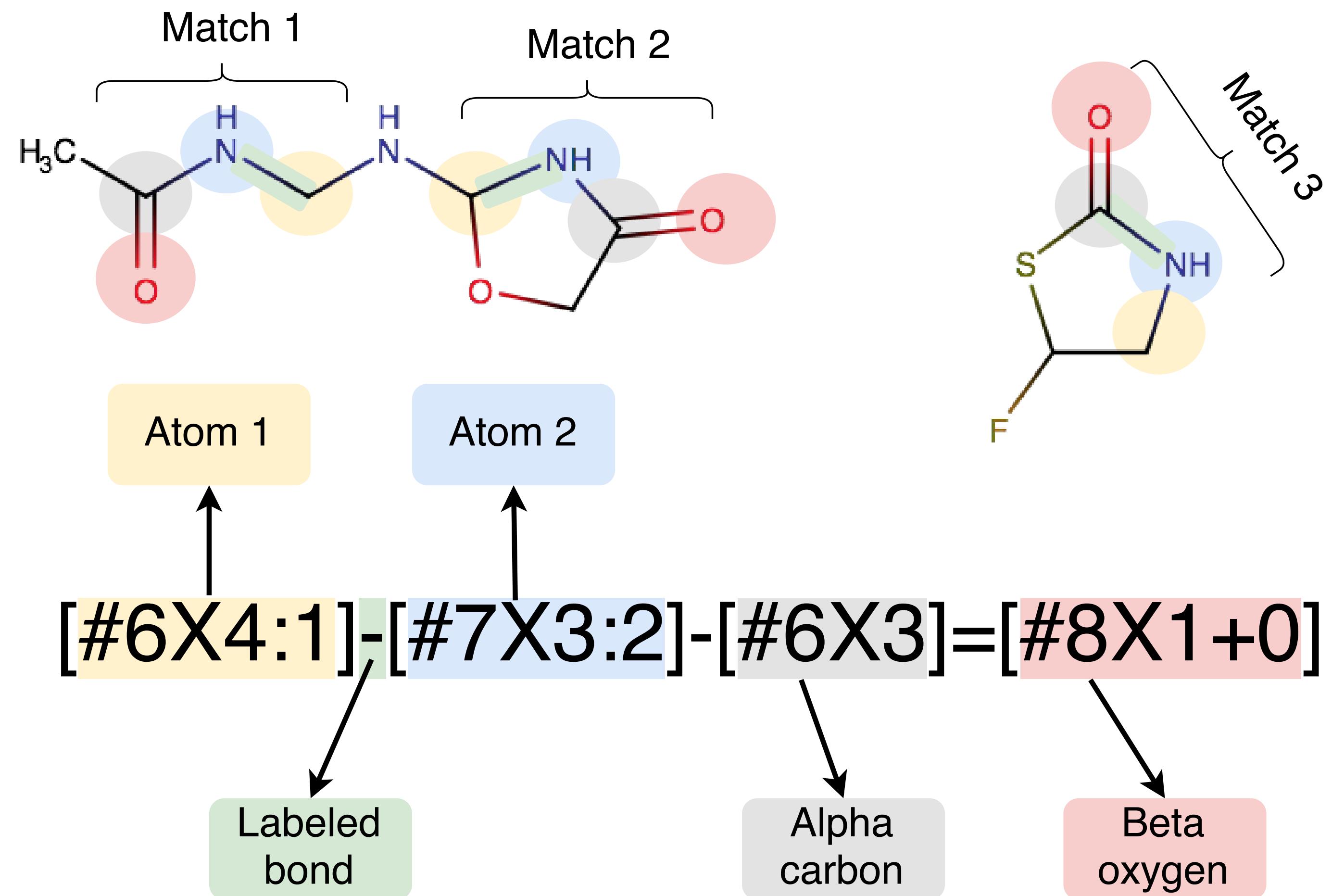
A modern starting point for forcefield parameterization

Preprint: <https://doi.org/10.1101/286542>

Software: <https://github.com/openforcefield/openforcefield>

The SMIRKS Native Open Force Field (SMIRNOFF) avoids the complexities of atom typing

Use of industry-standard SMARTS/SMIRKS chemical perception greatly simplifies tooling for parameter assignment while solving issues with extensibility and flexibility



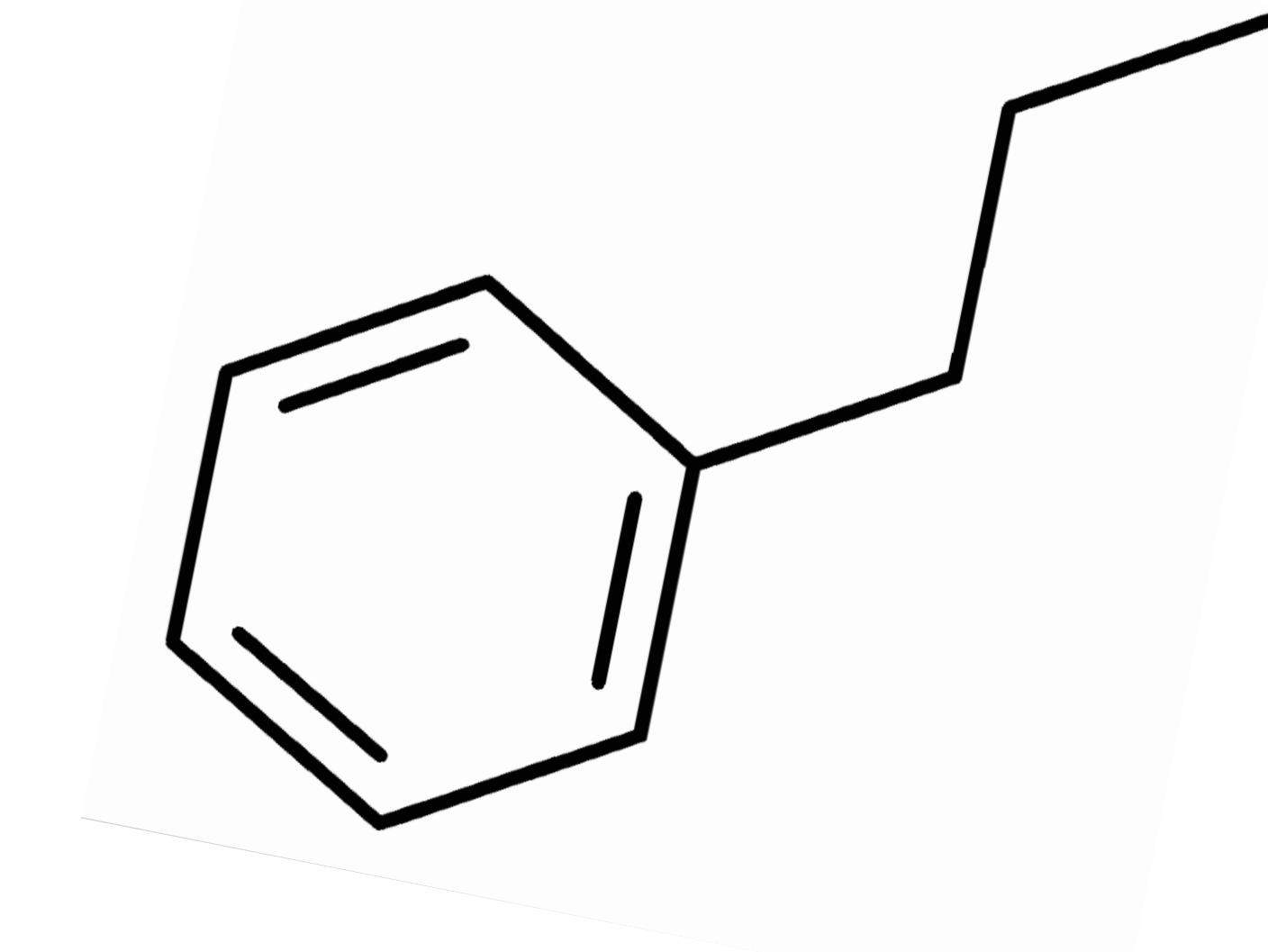
So, how would we use SMIRKS for a force field? Let's think of a carbon-carbon single bond

[#6:1]-[#6:2], length=1.526 angstroms, force
constant=620.0 kcal/(mol angstrom²)

Or, maybe we'd want a generic carbon-carbon bond:
[#6:1]~[#6:2] with its own parameters

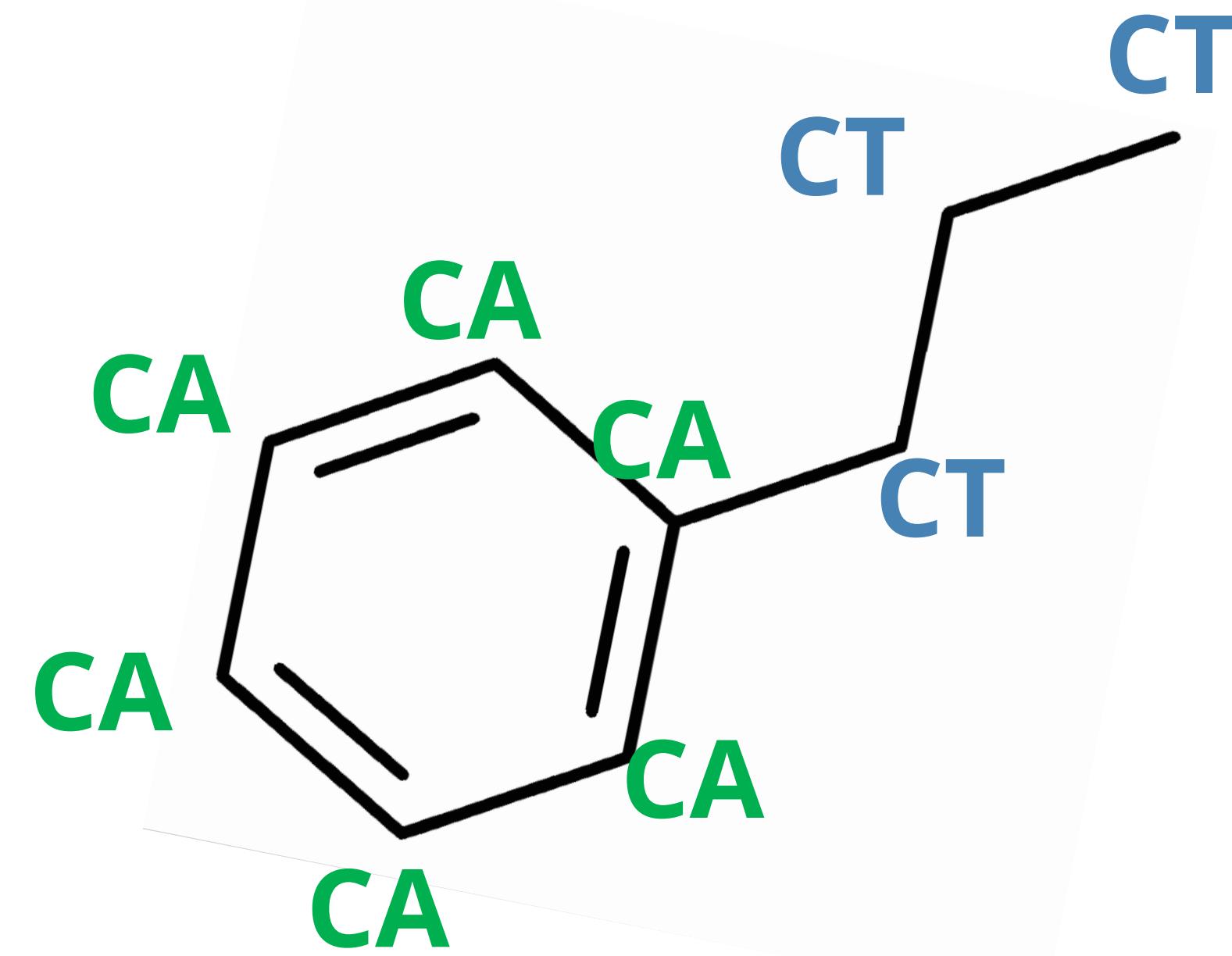
Perhaps a more specialized bond?
[#6X3:1]=[#6X3:2] with different parameters

Why is this a good thing? Let's think of atom typing or "chemical perception" which defines which parameters are used where



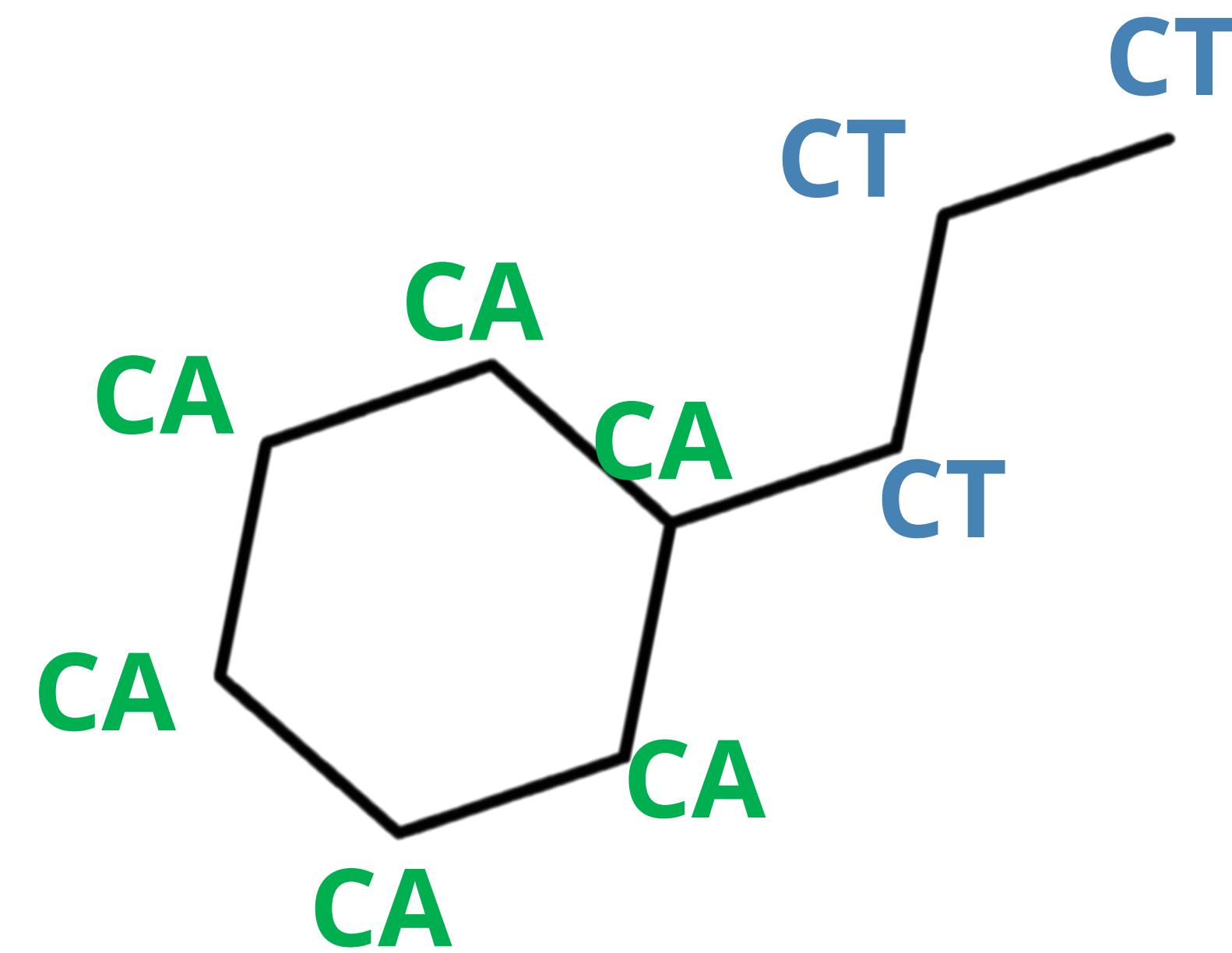
Aliphatic sp³ carbon (**CT**)
Aromatic sp² carbon (**CA**)

Why is this a good thing? Let's think of atom typing or "chemical perception" which defines which parameters are used where



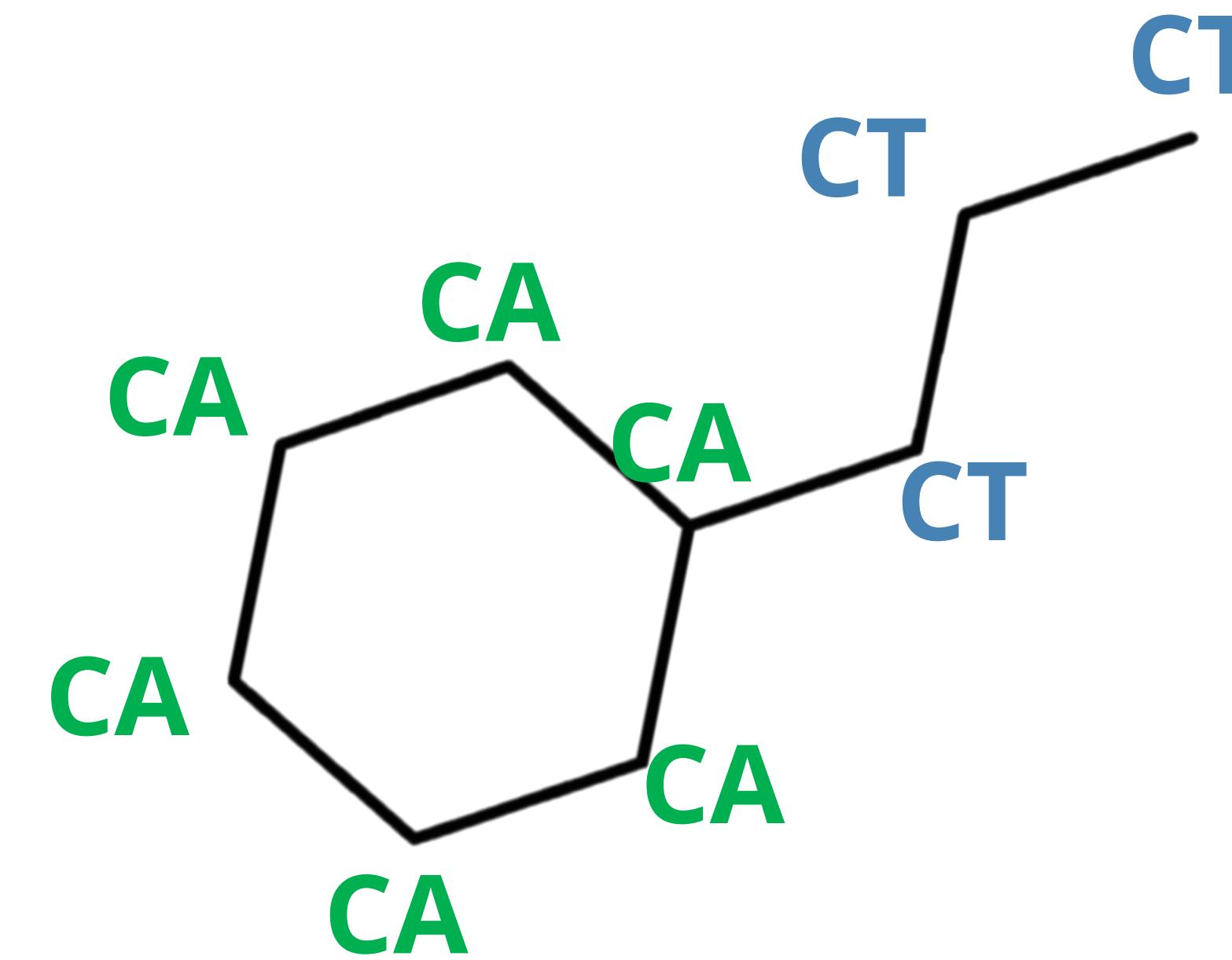
Aliphatic sp₃ carbon (**CT**)
Aromatic sp₂ carbon (**CA**)

Why is this a good thing? Let's think of atom typing or "chemical perception" which defines which parameters are used where



Aliphatic sp³ carbon (**CT**)
Aromatic sp² carbon (**CA**)

Why is this a good thing? Let's think of atom typing or "chemical perception" which defines which parameters are used where



Aliphatic sp³ carbon (CT)
Aromatic sp² carbon (CA)

X - CT - CT - X

Low Barrier Torsion

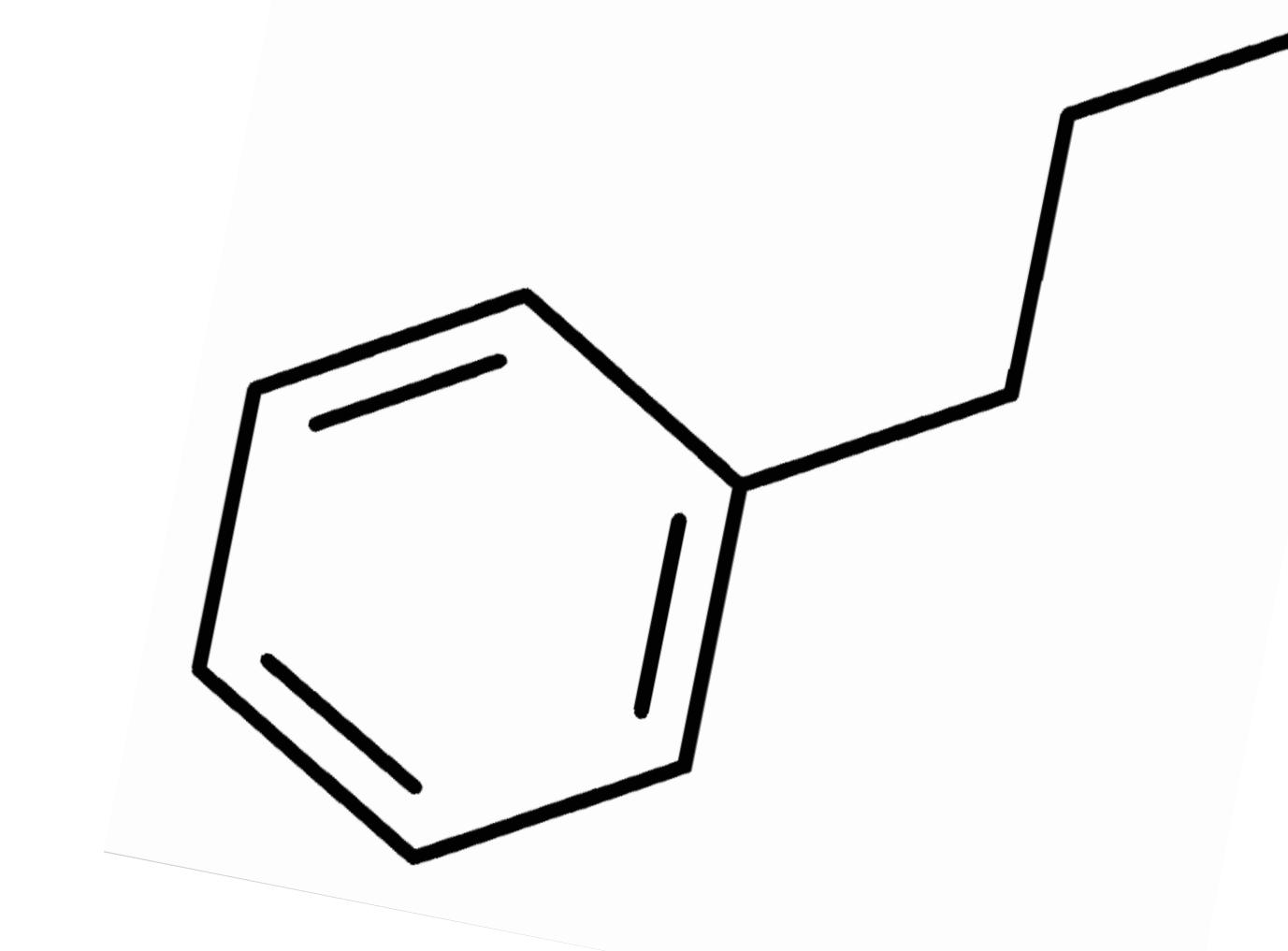
X - CT - CA - X

Low Barrier Torsion

X - CA - CA - X

High Barrier Torsion

Today's force fields mostly use indirect chemical perception



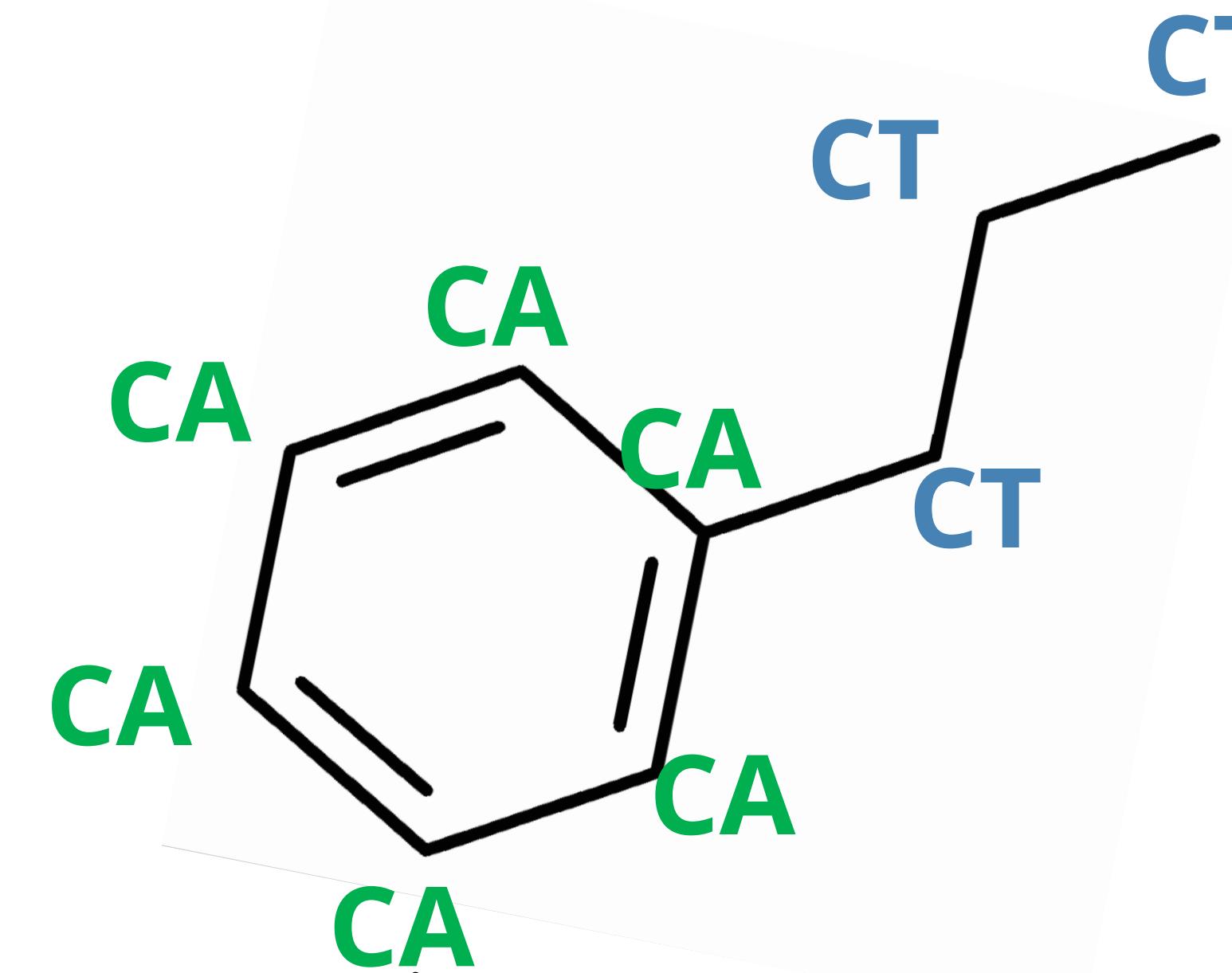
Aliphatic sp³ carbon (**CT**)
Aromatic sp² carbon (**CA**)

Some tool (or human) assigns atom types

From the atom types, parameters are assigned

Thus, atom types must encode all requisite chemistry and can't be fitted as part of the process

Today's force fields mostly use indirect chemical perception



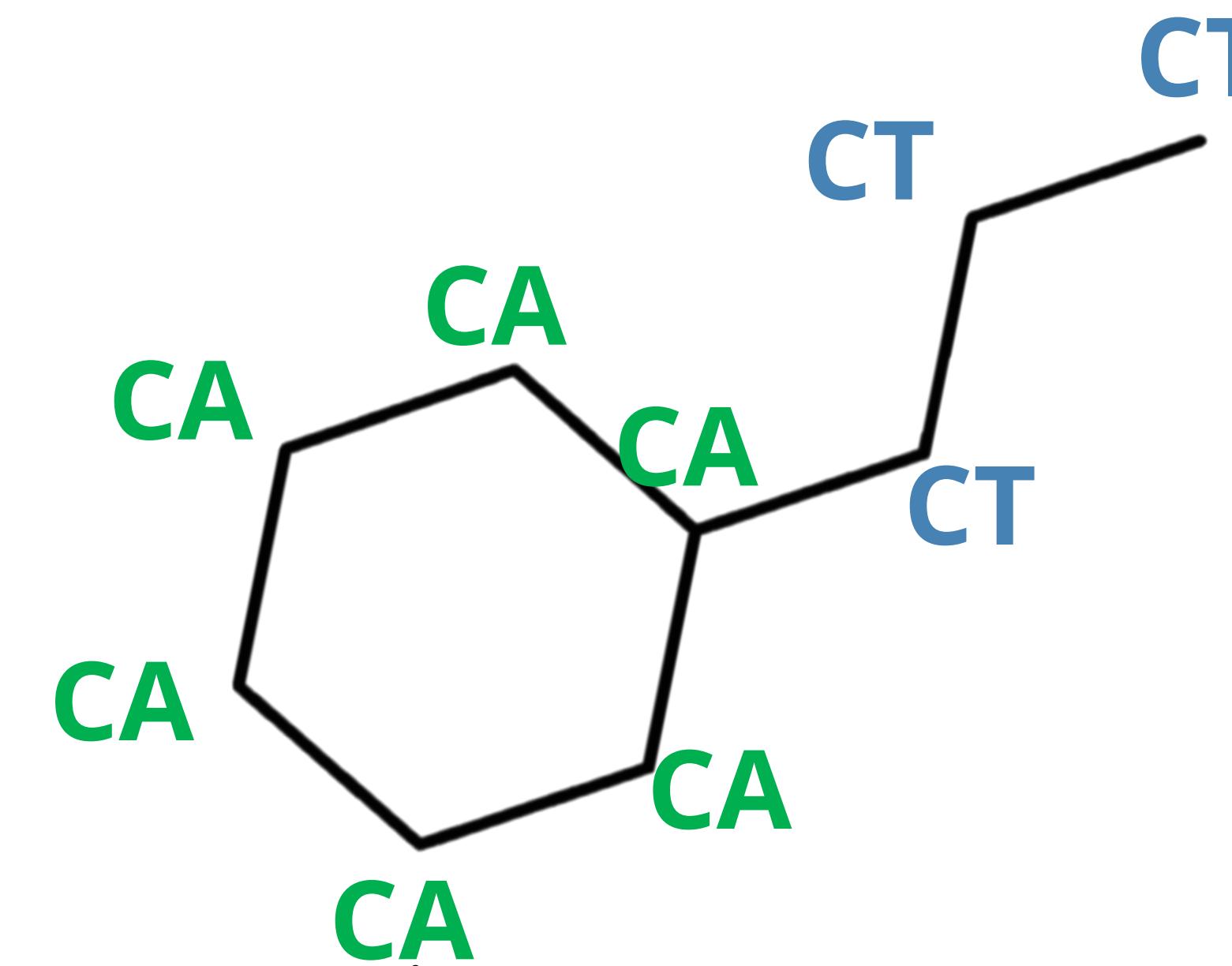
Aliphatic sp₃ carbon (**CT**)
Aromatic sp₂ carbon (**CA**)

Some tool (or human) assigns atom types

From the atom types, parameters are assigned

Thus, atom types must encode all requisite chemistry and can't be fitted as part of the process

Today's force fields mostly use indirect chemical perception



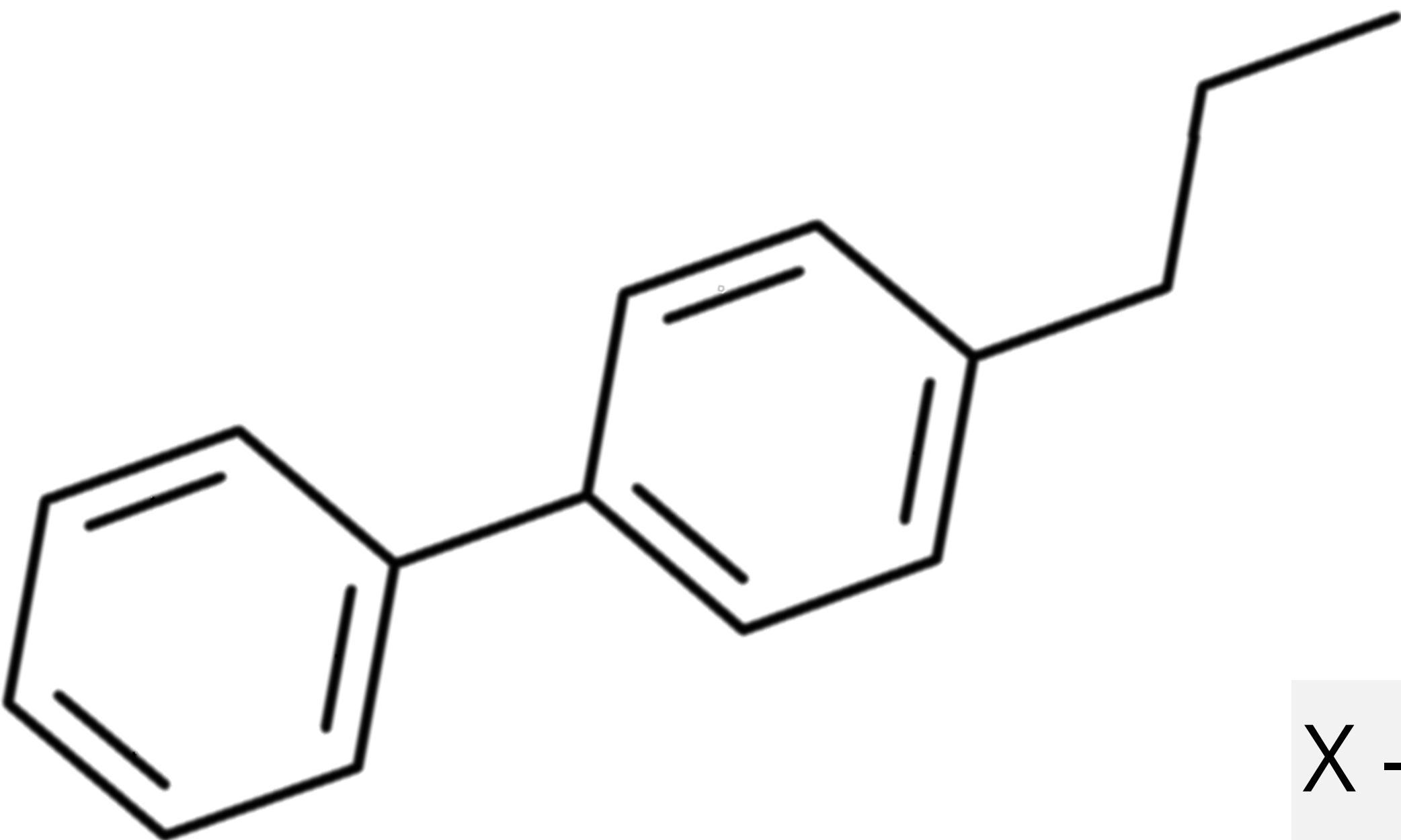
Aliphatic sp₃ carbon (**CT**)
Aromatic sp₂ carbon (**CA**)

Some tool (or human) assigns atom types

From the atom types, parameters are assigned

Thus, atom types must encode all requisite chemistry and can't be fitted as part of the process

This is a vital issue: Failing to capture the requisite chemistry leads to disaster



Aliphatic sp³ carbon (**CT**)
Aromatic sp² carbon (**CA**)

X -CT-CT-X

Low Barrier Torsion

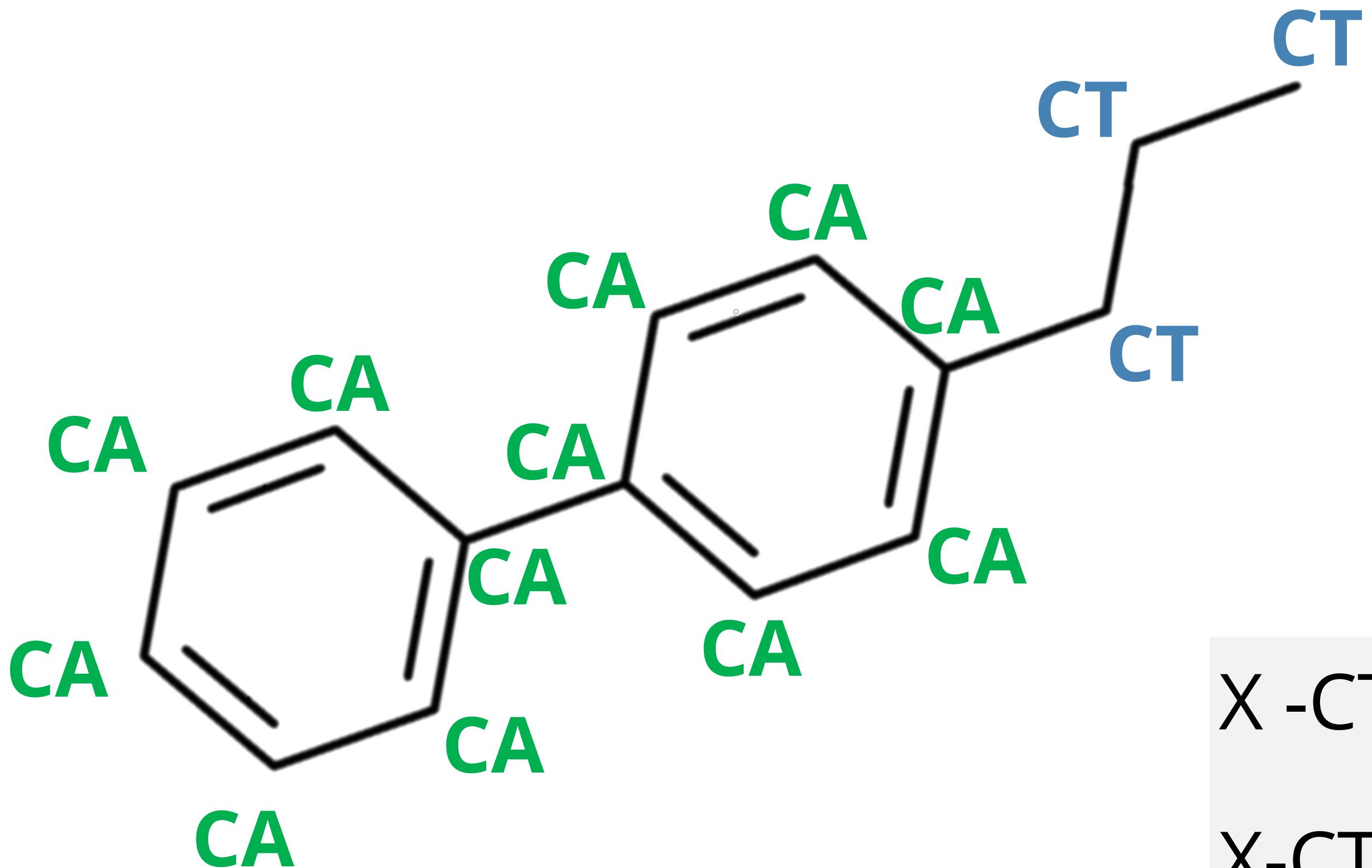
X-CT-CA-X

Low Barrier Torsion

X -CA-CA-X

High Barrier Torsion

This is a vital issue: Failing to capture the requisite chemistry leads to disaster



Aliphatic sp³ carbon (**CT**)
Aromatic sp² carbon (**CA**)

X -CT-CT-X

Low Barrier Torsion

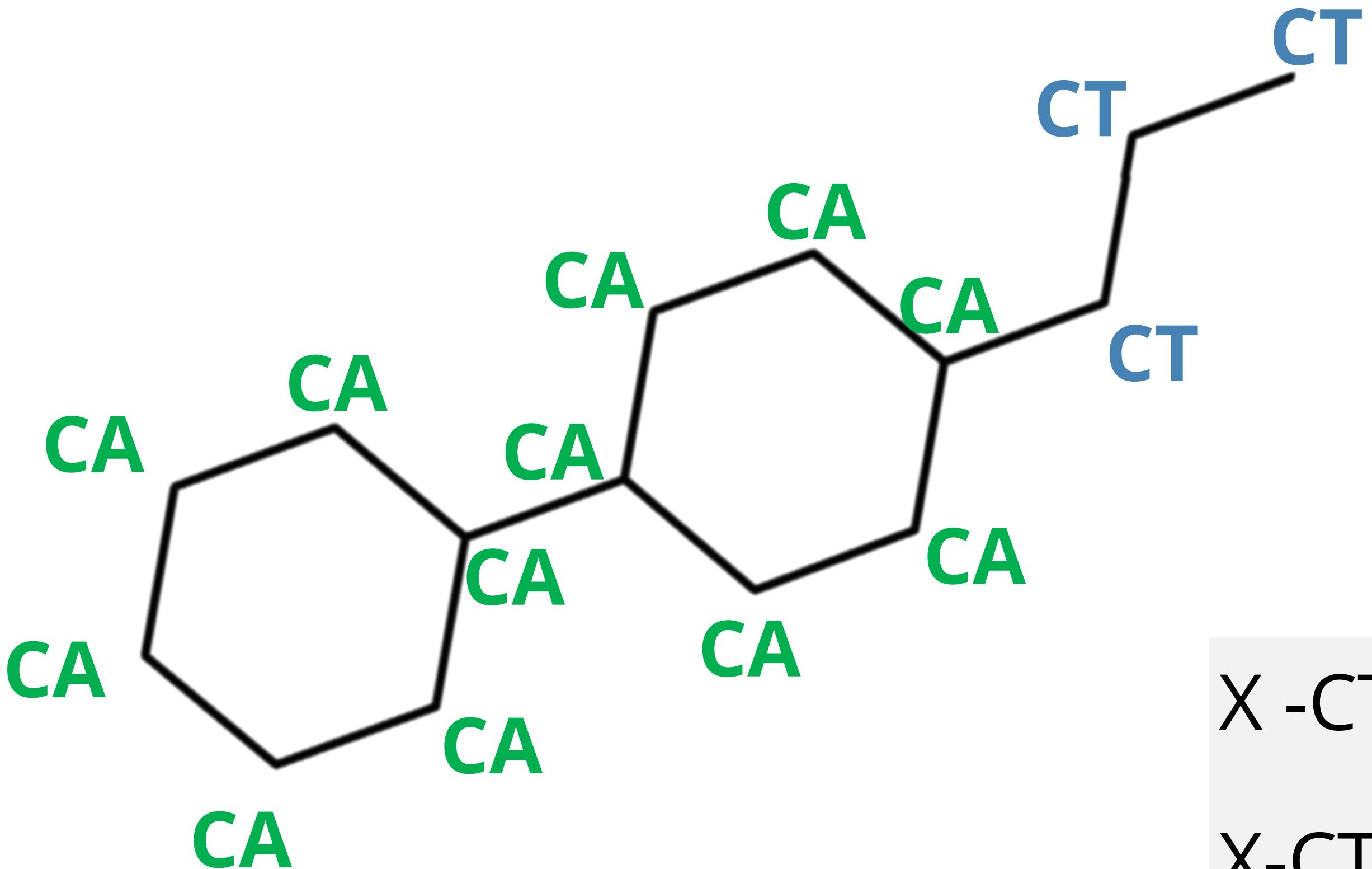
X-CT-CA-X

Low Barrier Torsion

X -CA-CA-X

High Barrier Torsion

This is a vital issue: Failing to capture the requisite chemistry leads to disaster



Aliphatic sp³ carbon (CT)
Aromatic sp² carbon (CA)

X -CT-CT-X

Low Barrier Torsion

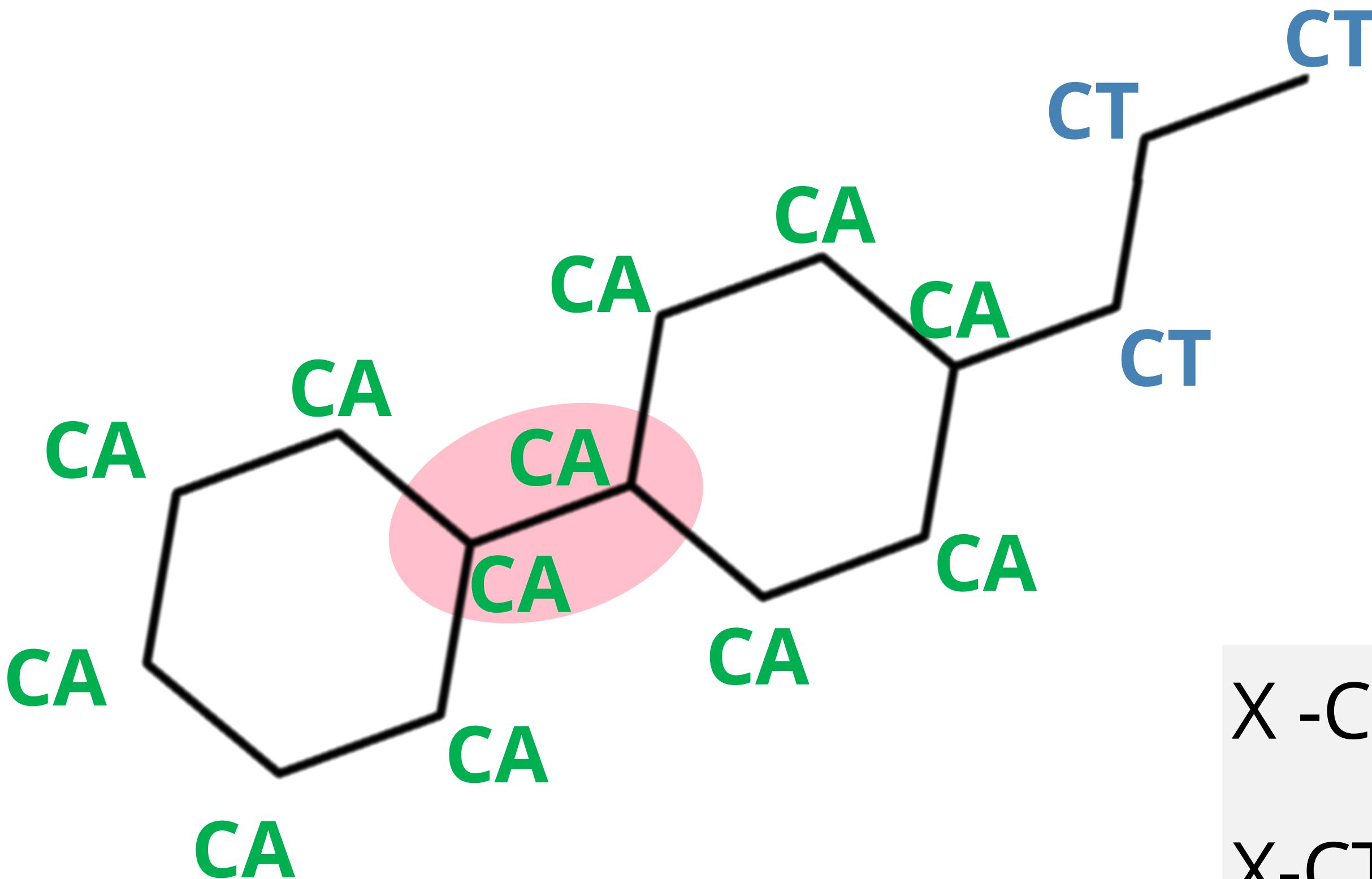
X-CT-CA-X

Low Barrier Torsion

X -CA-CA-X

High Barrier Torsion

This is a vital issue: Failing to capture the requisite chemistry leads to disaster



Aliphatic sp³ carbon (**CT**)
Aromatic sp² carbon (**CA**)

X -CT-CT-X

Low Barrier Torsion

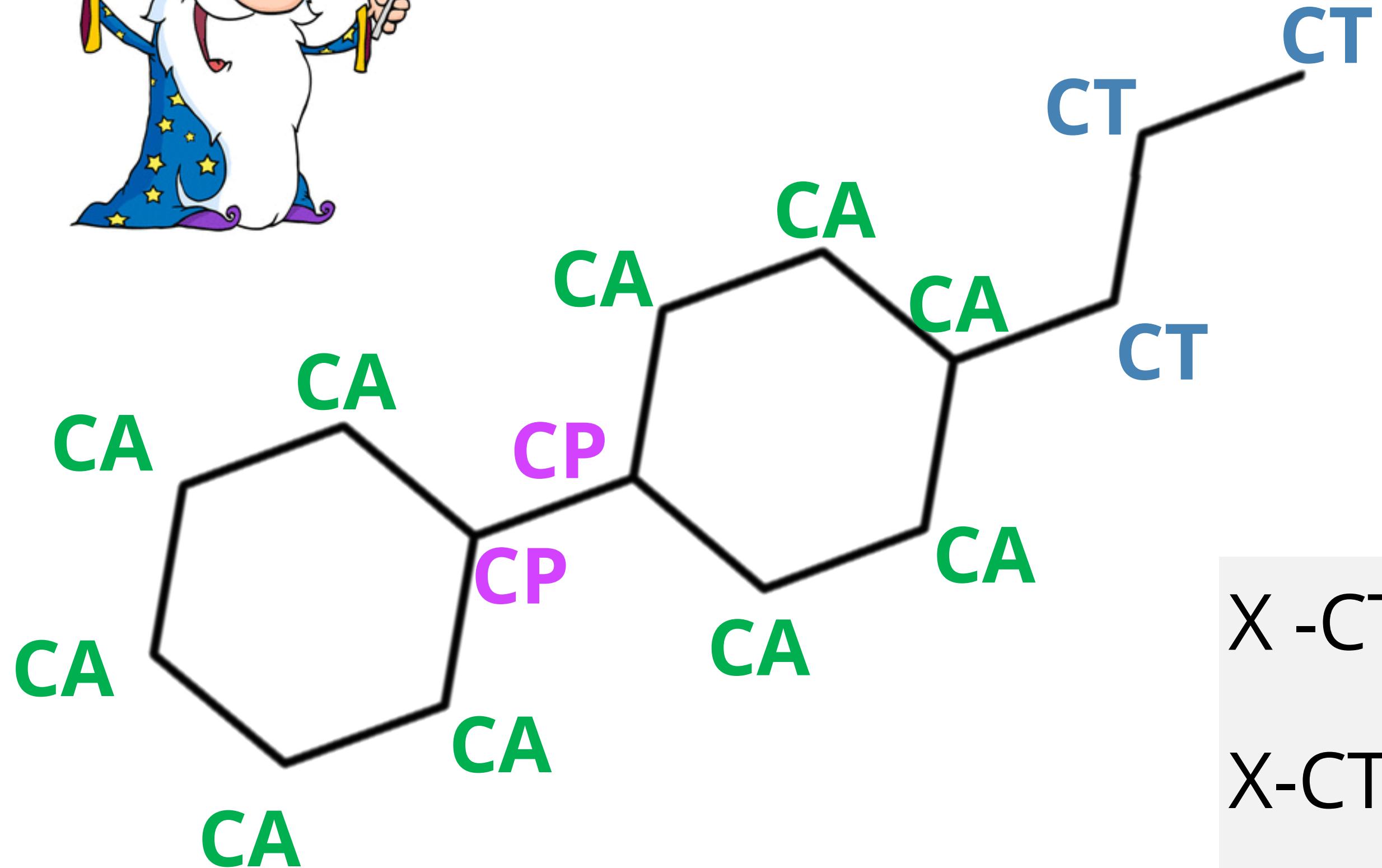
X-CT-CA-X

Low Barrier Torsion

X -CA-CA-X

High Barrier Torsion

One can fix this with more complex atom typing



Aliphatic sp₃ carbon (**CT**)
Aromatic sp₂ carbon (**CA**)
Aromatic carbon bridging
phenyl rings (**CP**)

X -CT-CT-X

Low Barrier Torsion

X-CT-CA-X

Low Barrier Torsion

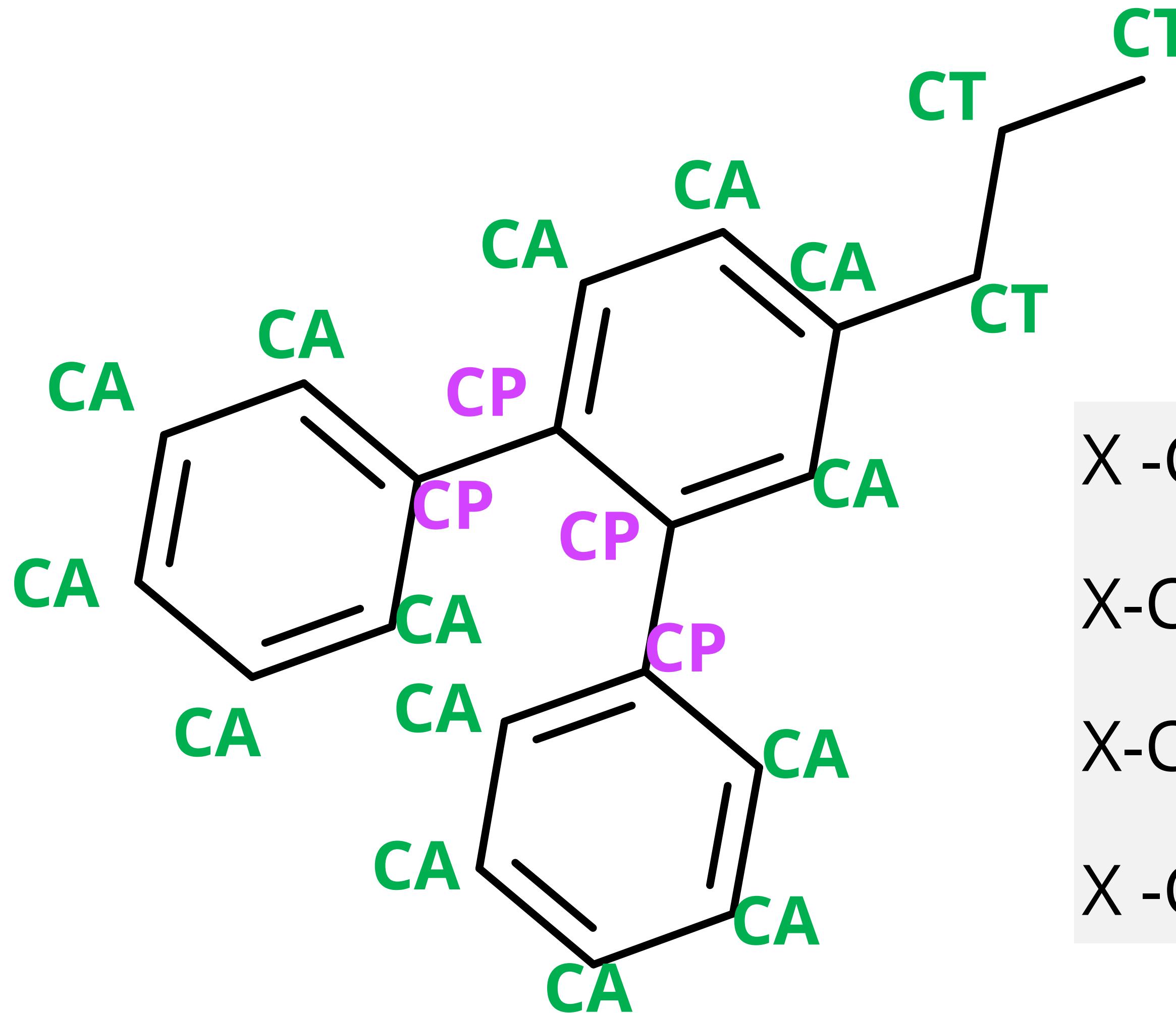
X-CP-CP-X

Low Barrier Torsion

X -CA-CA-X

High Barrier Torsion

Though, new chemistry can still pose challenges so you need an atom type for every special case



Aliphatic sp₃ carbon (**CT**)

Aromatic sp₂ carbon (**CA**)

Aromatic carbon bridging phenyl rings (**CP**)

X -CT-CT-X

Low Barrier Torsion

X-CT-CA-X

Low Barrier Torsion

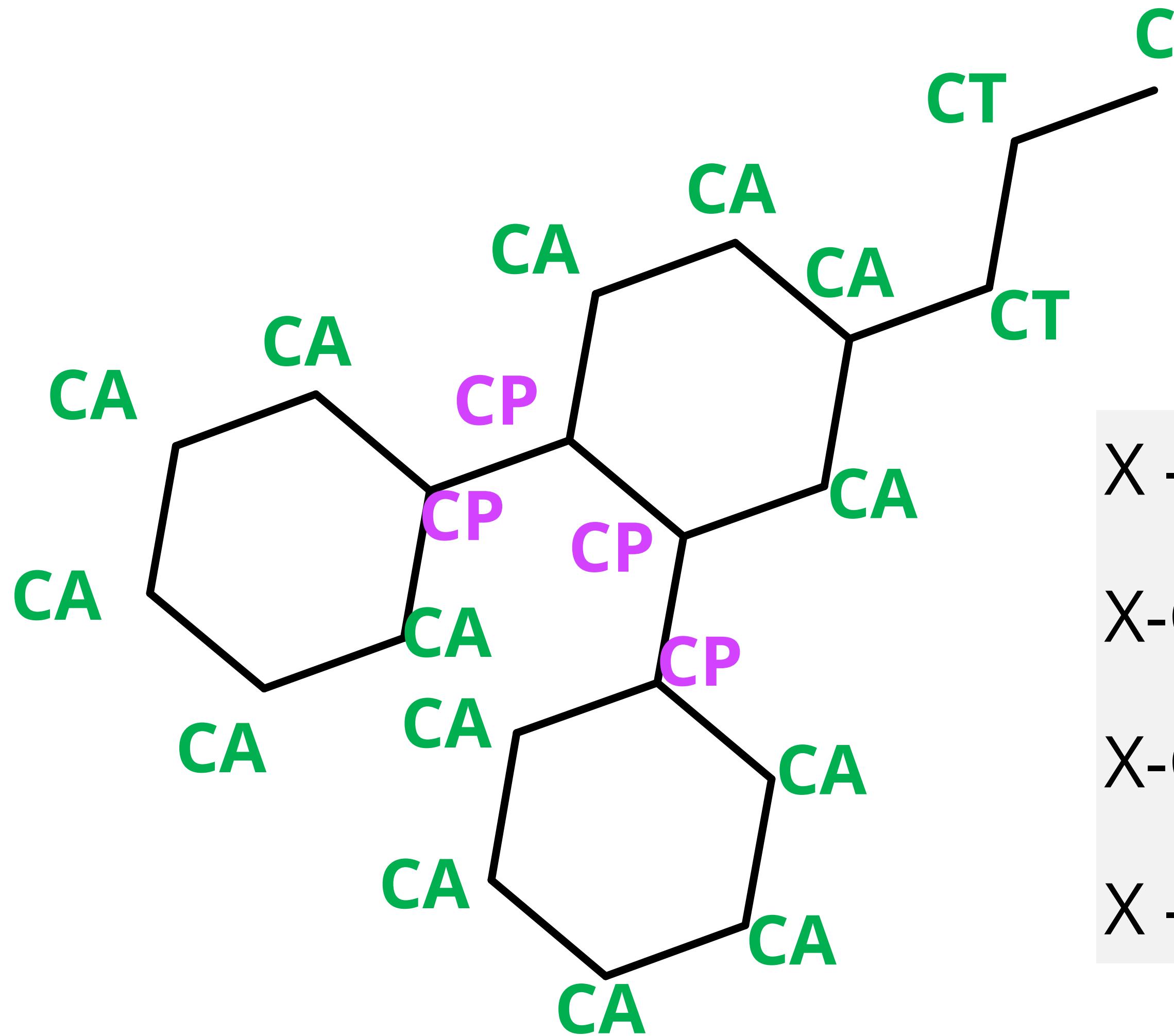
X-CP-CP-X

Low Barrier Torsion

X -CA-CA-X

High Barrier Torsion

Though, new chemistry can still pose challenges so you need an atom type for every special case



Aliphatic sp₃ carbon (**CT**)
Aromatic sp₂ carbon (**CA**)
Aromatic carbon bridging phenyl rings (**CP**)

X -CT-CT-X

Low Barrier Torsion

X-CT-CA-X

Low Barrier Torsion

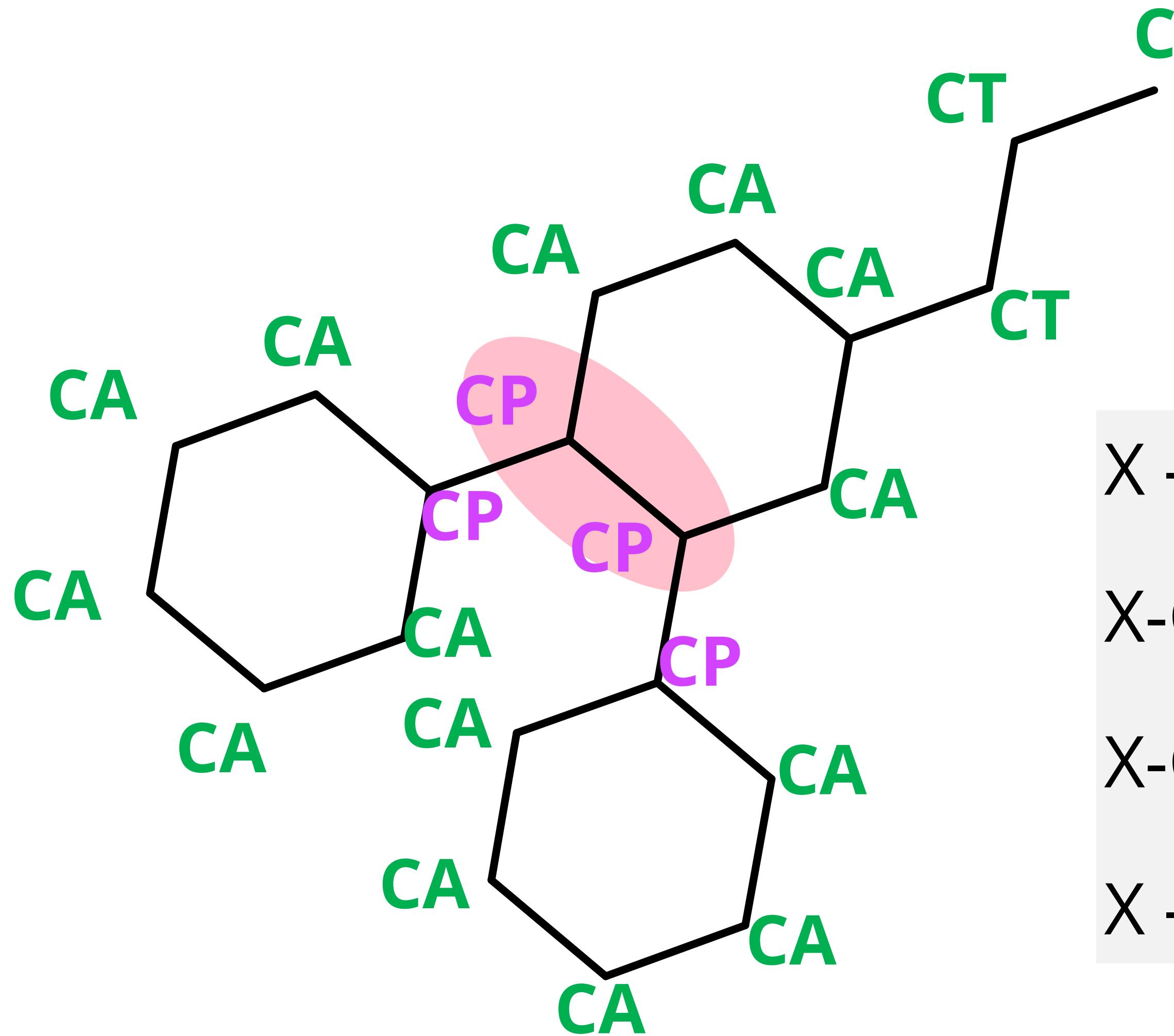
X-CP-CP-X

Low Barrier Torsion

X -CA-CA-X

High Barrier Torsion

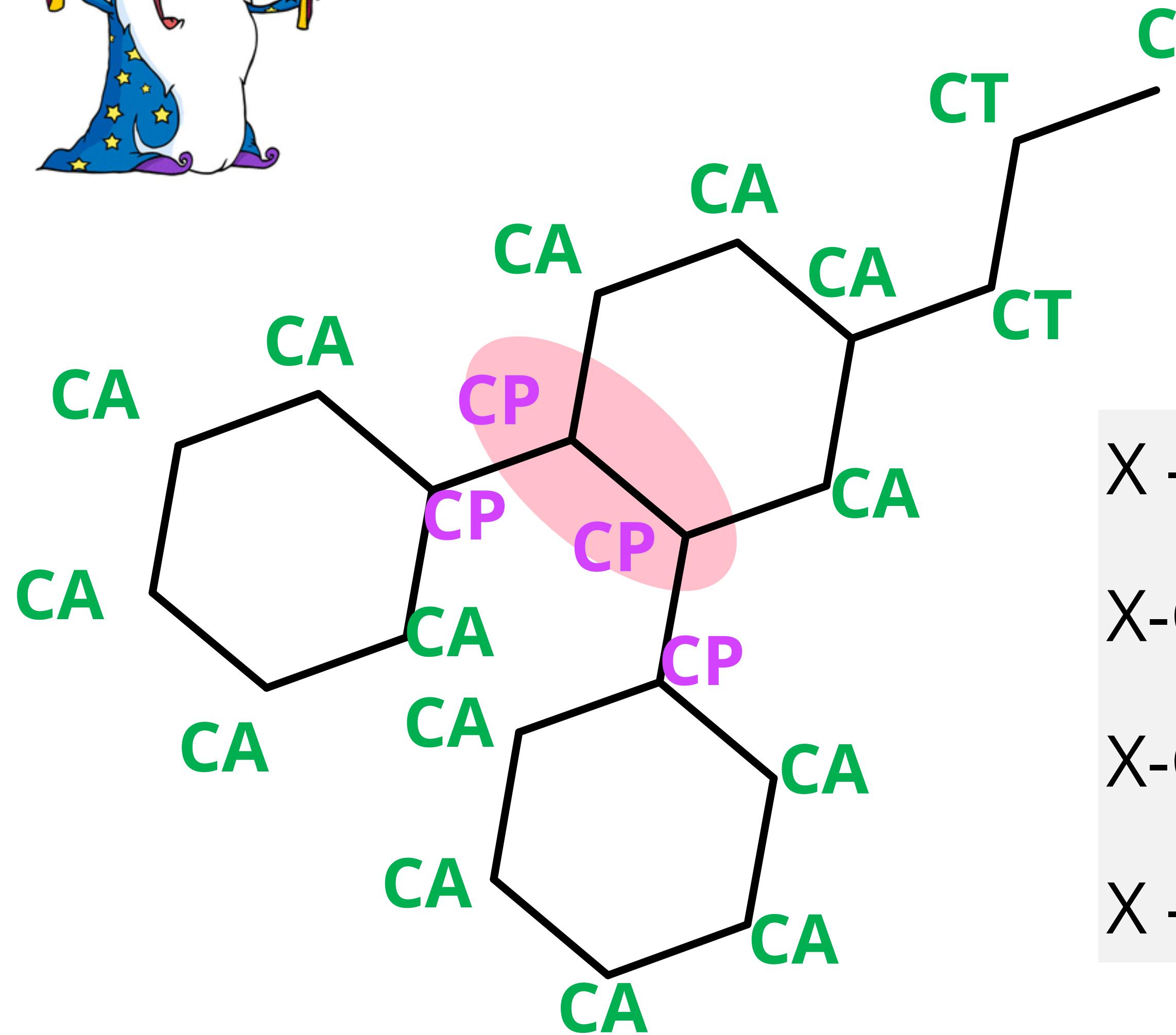
Though, new chemistry can still pose challenges so you need an atom type for every special case



Aliphatic sp³ carbon (**CT**)
Aromatic sp² carbon (**CA**)
Aromatic carbon bridging phenyl rings (**CP**)

X -CT-CT-X	Low Barrier Torsion
X-CT-CA-X	Low Barrier Torsion
X-CP-CP-X	Low Barrier Torsion
X -CA-CA-X	High Barrier Torsion

Though, new chemistry can still pose challenges so you need an atom type for every special case



Aliphatic sp³ carbon (**CT**)
Aromatic sp² carbon (**CA**)
Aromatic carbon bridging phenyl rings (**CP**)
Same as CP for multiple bridges (**CQ**)

X -CT-CT-X

Low Barrier Torsion

X-CT-CA-X

Low Barrier Torsion

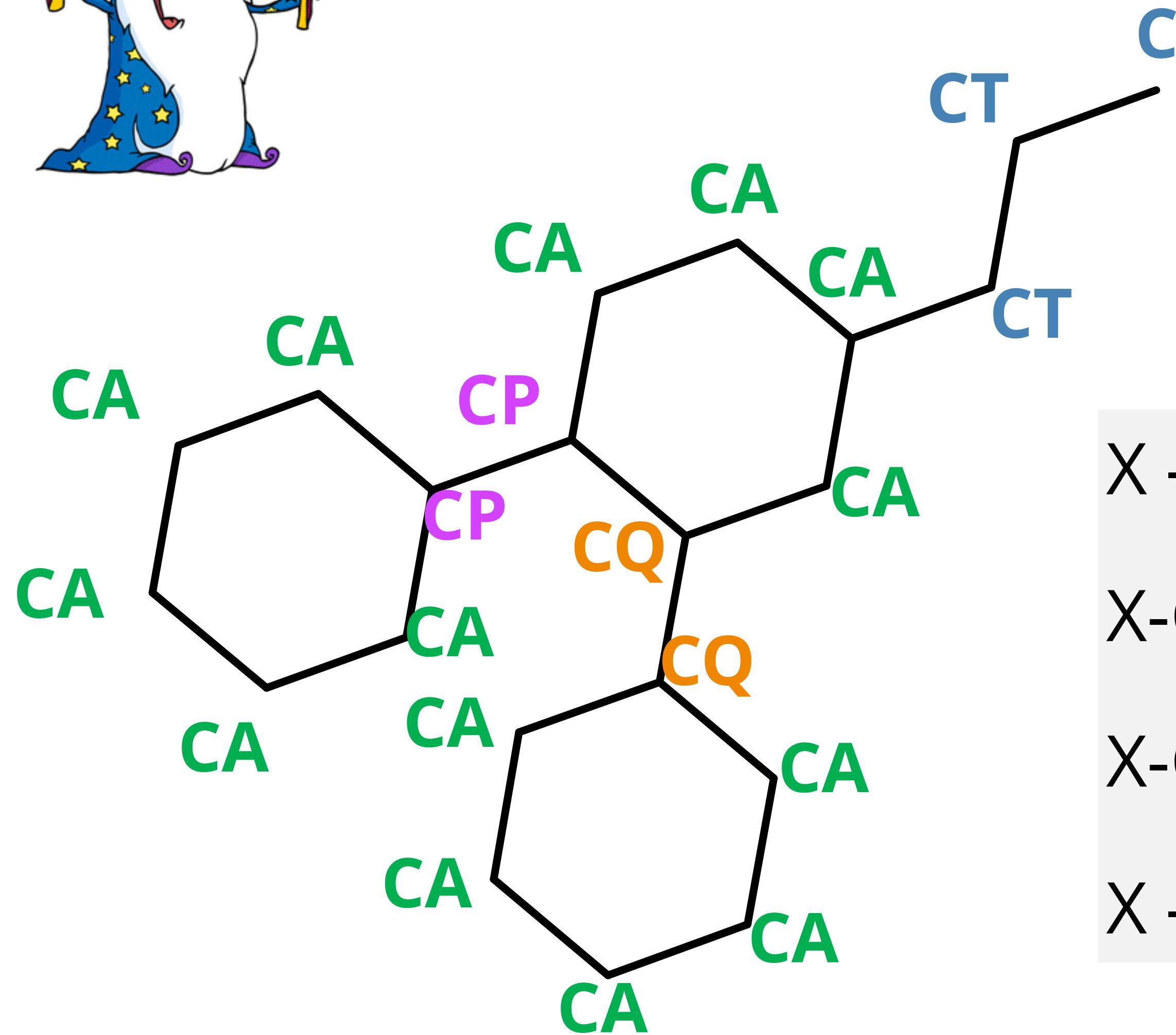
X-CP-CP-X

Low Barrier Torsion

X -CA-CA-X

High Barrier Torsion

Though, new chemistry can still pose challenges so you need an atom type for every special case



Aliphatic sp³ carbon (**CT**)

Aromatic sp² carbon (**CA**)

Aromatic carbon bridging phenyl rings (**CP**)

Same as CP for multiple bridges (**CQ**)

X -CT-CT-X

Low Barrier Torsion

X-CT-CA-X

Low Barrier Torsion

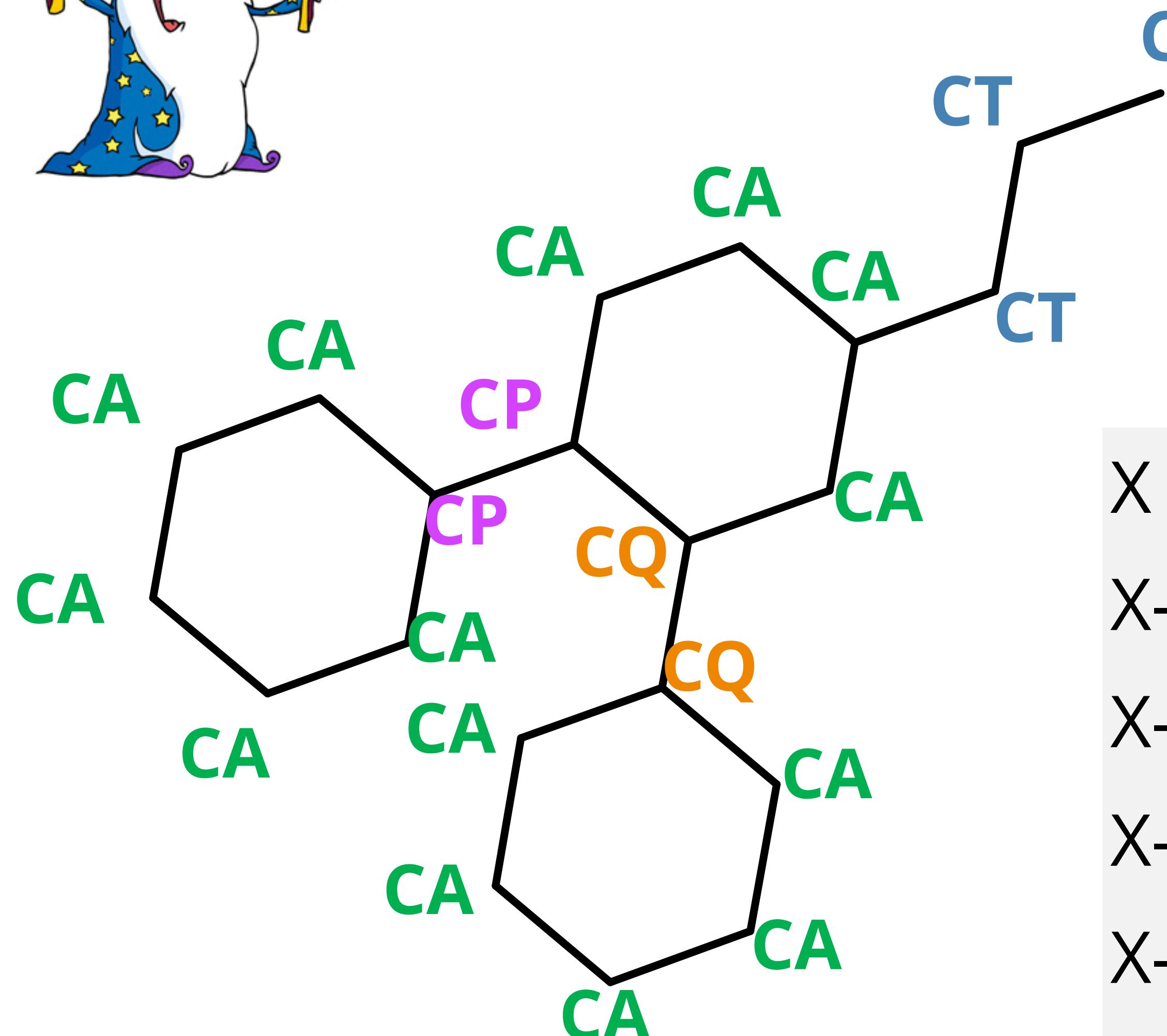
X-CP-CP-X

Low Barrier Torsion

X -CA-CA-X

High Barrier Torsion

Though, new chemistry can still pose challenges so you need an atom type for every special case



X -CT-CT-X	Low Barrier Torsion
X-CT-CA-X	Low Barrier Torsion
X-CP-CP-X	Low Barrier Torsion
X-CQ-CQ-X	Low Barrier Torsion
X-CP-CQ-X	Low Barrier Torsion
X -CA-CA-X	High Barrier Torsion

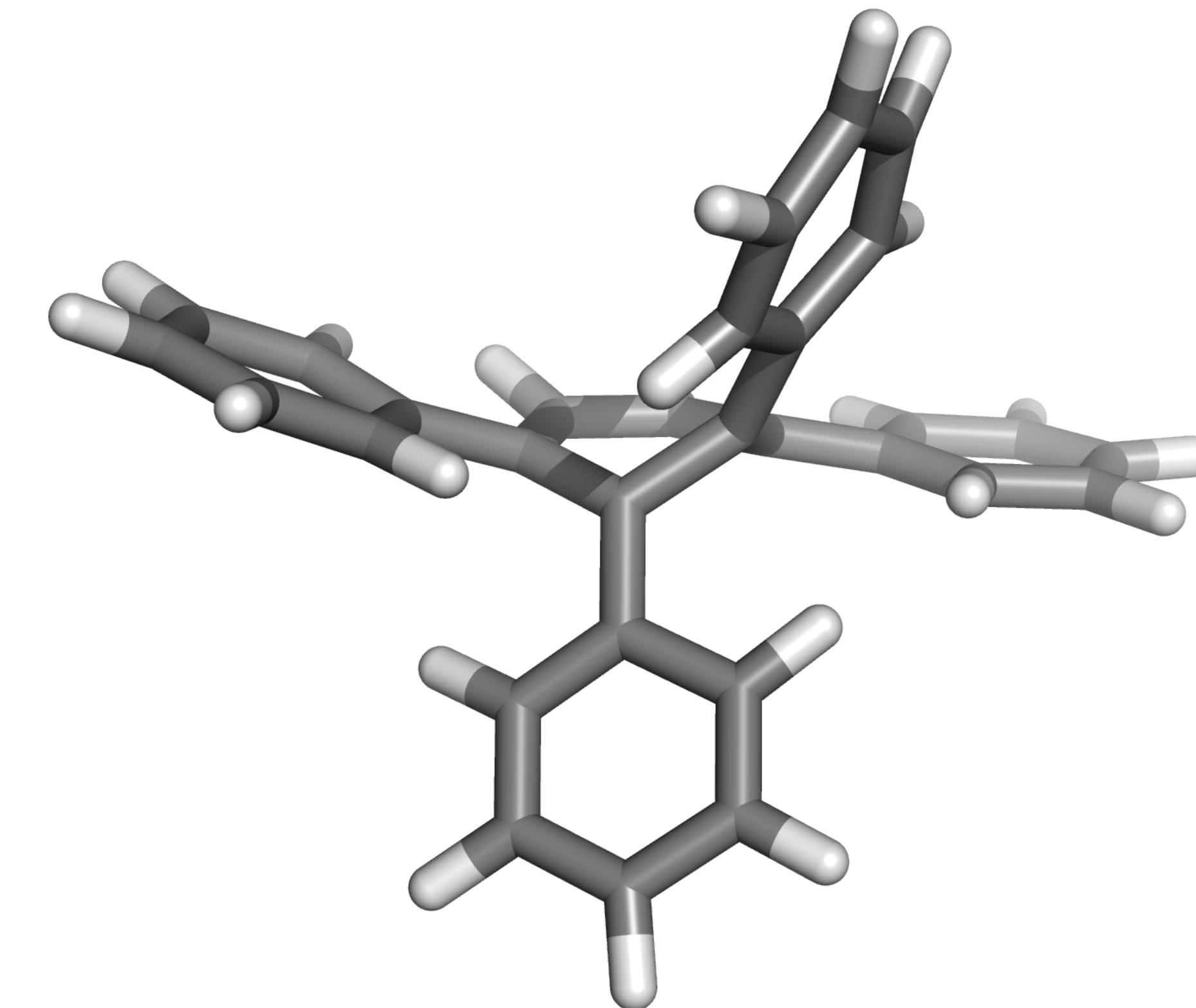
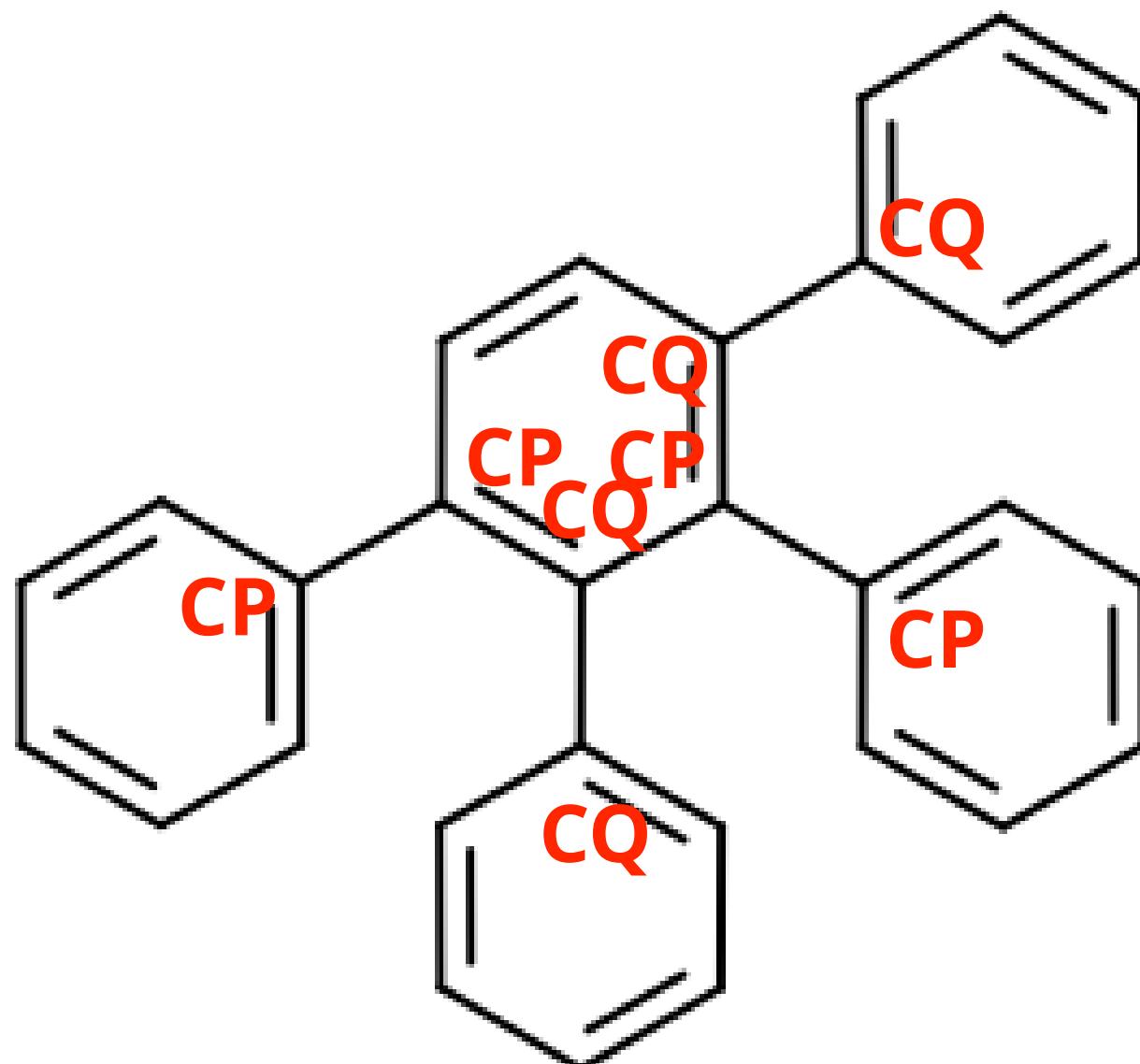
This produces many redundant parameters

CA-CA-SO	70.000 120.000 force ff94 CA-CA-CT
CA-CA-SH	70.000 120.000 std aromatic
CA-CA-SD	70.000 120.000 std aromatic
CA-CA-S	70.000 120.000 std aromatic
CA-CA-P	70.000 120.000 std aromatic
CA-CA-OS	70.000 120.000 ** Gro, JACS,V111,2152('89)
CA-CA-OH	70.000 120.000 ff94 CA-C-OH
CA-CA-O2	70.000 120.000 guess March 5 2009 anionic O
CA-CA-NL	70.000 120.000 guess
CA-CA-ND	70.000 120.000 calc B3PW91/6-31+G** Jan 30 2002
CA-CA-NC	70.000 120.000 quinoline, ff94 CA-CA-CT
CA-CA-NB	70.000 120.000 guess april 11 2000
CA-CA-NA	70.000 120.000 ff94 CA-CA-CT
CA-CA-N3	70.000 120.000 guess april 11 2000
CA-CA-N2	70.000 120.000 ff94 CA-CA-CT
CA-CA-N*	70.000 120.000 ff94 std aromatic
CA-CA-N	70.000 120.000 ff94 CA-CA-CT cb 6jan97
CA-CA-I	70.000 120.000 std sp2 carbon aug 15 2001
CA-CA-F	70.000 120.000 ff94 CA-C-OH
CA-CA-Cl	70.000 120.000 ff94 CA-C-OH
CA-CA-CW	70.000 120.000 amidopyridine, ff94 CA-CA-CT

parm@frosst has a few hundred lines of this type of redundancy

More than 60 identical parameters for CP alone

Downstream problems persist to this day, even in GAFF and GAFF2



- Torsions within the ring end up getting X-CP-CP-X values (rotatable single bond) rather than X-CA-CA-X

Ditching “atom types” for SMIRKS (“parameter types”) allows considerable simplification

For example, GAFF2 has
16 vdW types for carbon

c	1.8606	0.0988
cs	1.8606	0.0988
ca	1.8606	0.0988
cc	1.8606	0.0988
cd	1.8606	0.0988
ce	1.8606	0.0988
cf	1.8606	0.0988
cp	1.8606	0.0988
cq	1.8606	0.0988
cz	1.8606	0.0988
cu	1.8606	0.0988
cv	1.8606	0.0988
cg	1.9525	0.1596
ch	1.9525	0.1596
cx	1.9069	0.1078
cy	1.9069	0.1078

But this should be
three SMIRKS strings

[#6:1]	1.8606	0.0988
[#6X1:1]	1.9525	0.1596
[#6X3r3,#6X3r4:1]	1.9069	0.1078

Very relevant when attempting to automatically fit parameters — are there 32 parameters here, or 6?
(We would argue 6 — the atom types were introduced because of the need for angle or torsional complexity, usually)

SMIRNOFF parameters for methanol are simple

```
<?xml version="1.0"?>
<SMIRNOFF>

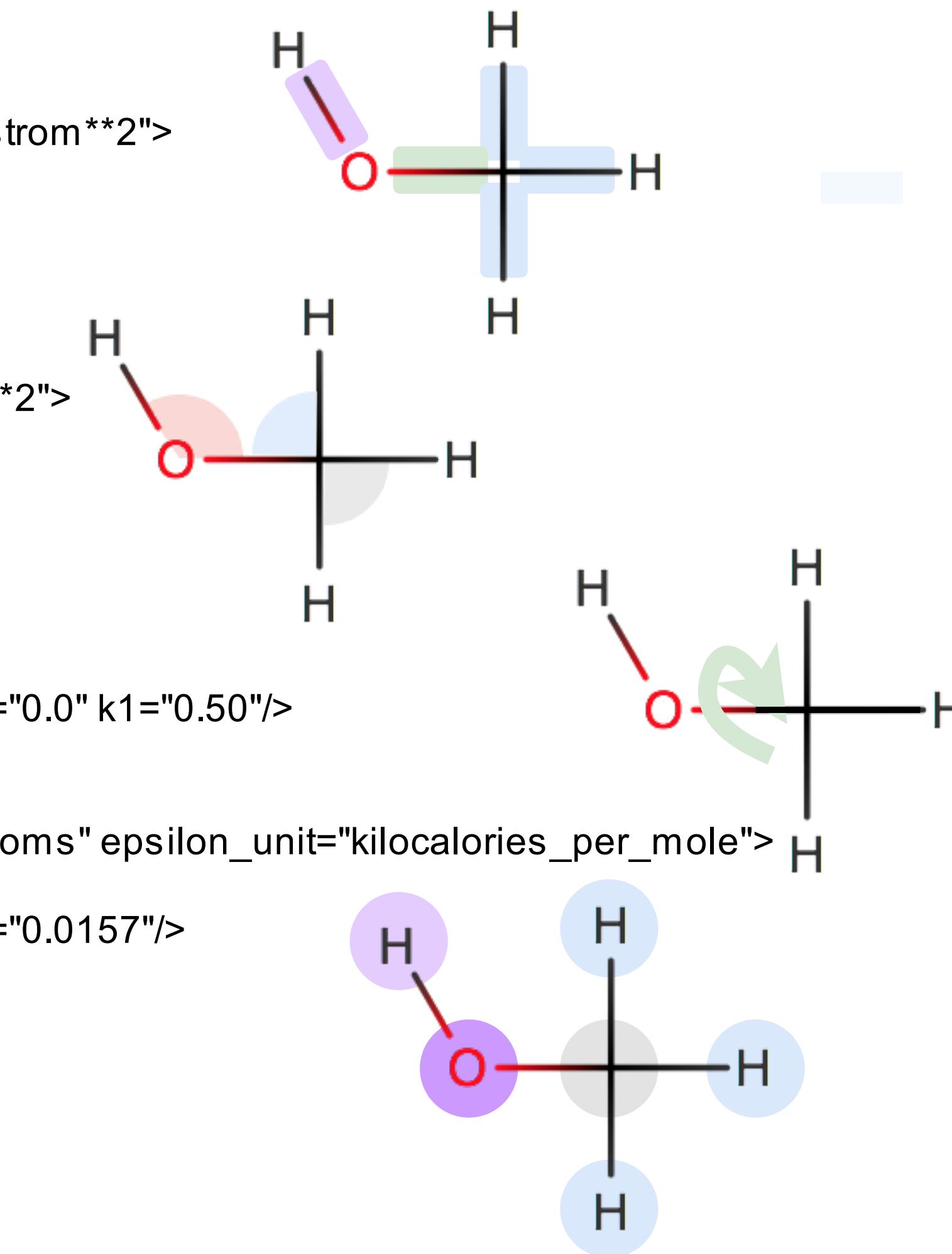
<HarmonicBondForce length_unit="angstroms" k_unit="kilocalories_per_mole/angstrom**2">
    <Bond smirks="[#6X4:1]-[#1:2]" length="1.090" k="680.0"/>
    <Bond smirks="[#6X4:1]-[#8&amp;X2&amp;H1:2]" length="1.410" k="640.0"/>
    <Bond smirks="[#8X2:1]-[#1:2]" length="0.960" k="1106.0"/>
</HarmonicBondForce>

<HarmonicAngleForce angle_unit="degrees" k_unit="kilocalories_per_mole/radian**2">
    <Angle smirks="[a,A:1]-[#6X4:2]-[a,A:3]" angle="109.50" k="100.0"/>
    <Angle smirks="[#1:1]-[#6X4:2]-[#1:3]" angle="109.50" k="70.0"/>
    <Angle smirks="[#6X4:1]-[#8X2:2]-[#1:3]" angle="108.50" k="110.0"/>
</HarmonicAngleForce>

<PeriodicTorsionForce phase_unit="degrees" k_unit="kilocalories_per_mole">
    <Proper smirks="*[a,A:1]-[#6X4:2]-[#8X2:3]-[#1:4]" idivf1="3" periodicity1="3" phase1="0.0" k1="0.50"/>
</PeriodicTorsionForce>

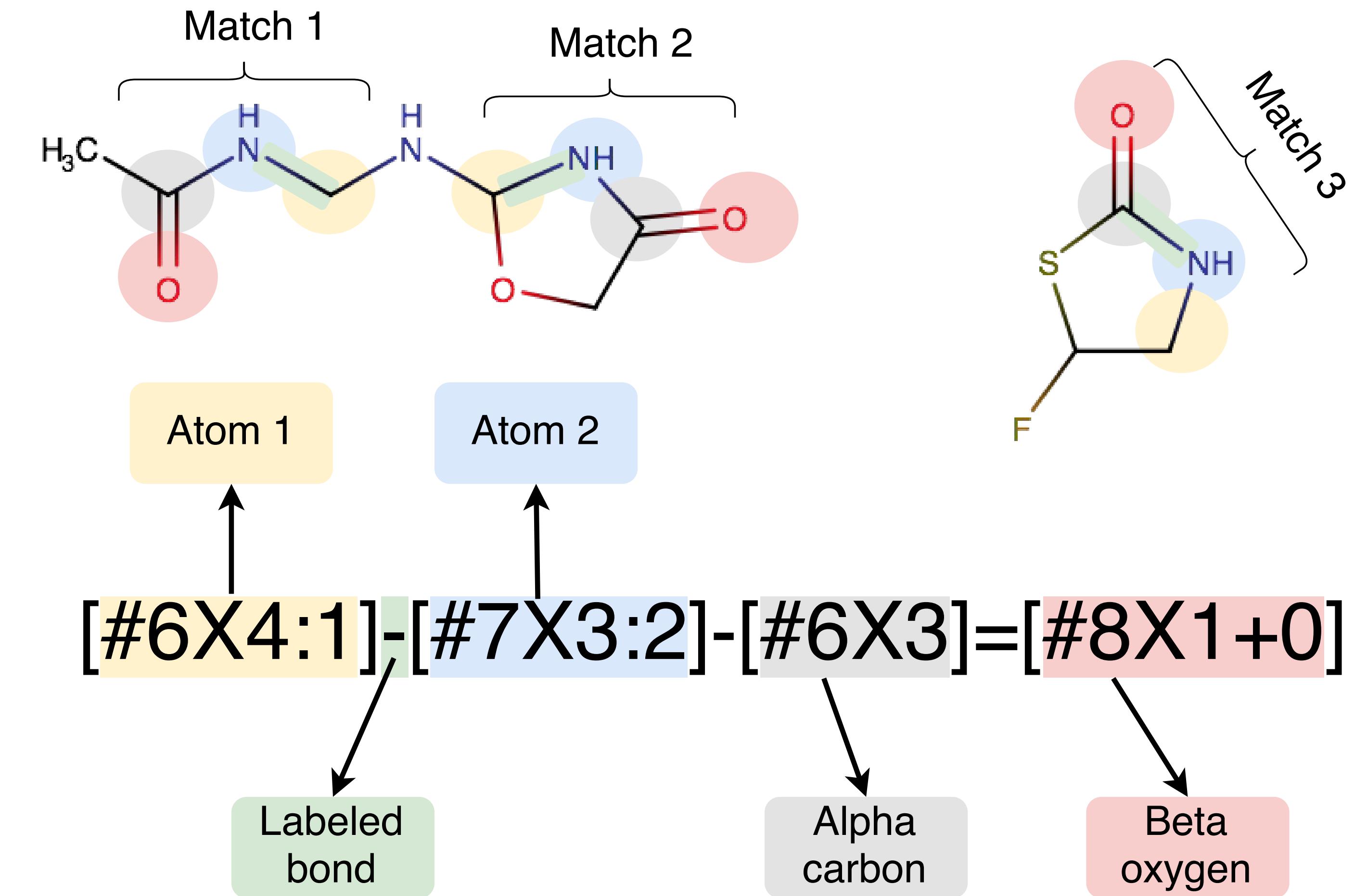
<NonbondedForce coulomb14scale="0.833333" lj14scale="0.5" sigma_unit="angstroms" epsilon_unit="kilocalories_per_mole">
    <Atom smirks="[#1:1]" rmin_half="1.4870" epsilon="0.0157"/>
    <Atom smirks="[$([#1]-[#6]-[#7,#8,#9,#16,#17,#35]):1]" rmin_half="1.3870" epsilon="0.0157"/>
    <Atom smirks="[#1$(*-[#8]):1]" rmin_half="0.0000" epsilon="0.0000"/>
    <Atom smirks="[#6:1]" rmin_half="1.9080" epsilon="0.1094"/>
    <Atom smirks="[#8:1]" rmin_half="1.6837" epsilon="0.1700"/>
    <Atom smirks="[#8X2+0$(*-[#1]):1]" rmin_half="1.7210" epsilon="0.2104"/>
</NonbondedForce>

</SMIRNOFF>
```

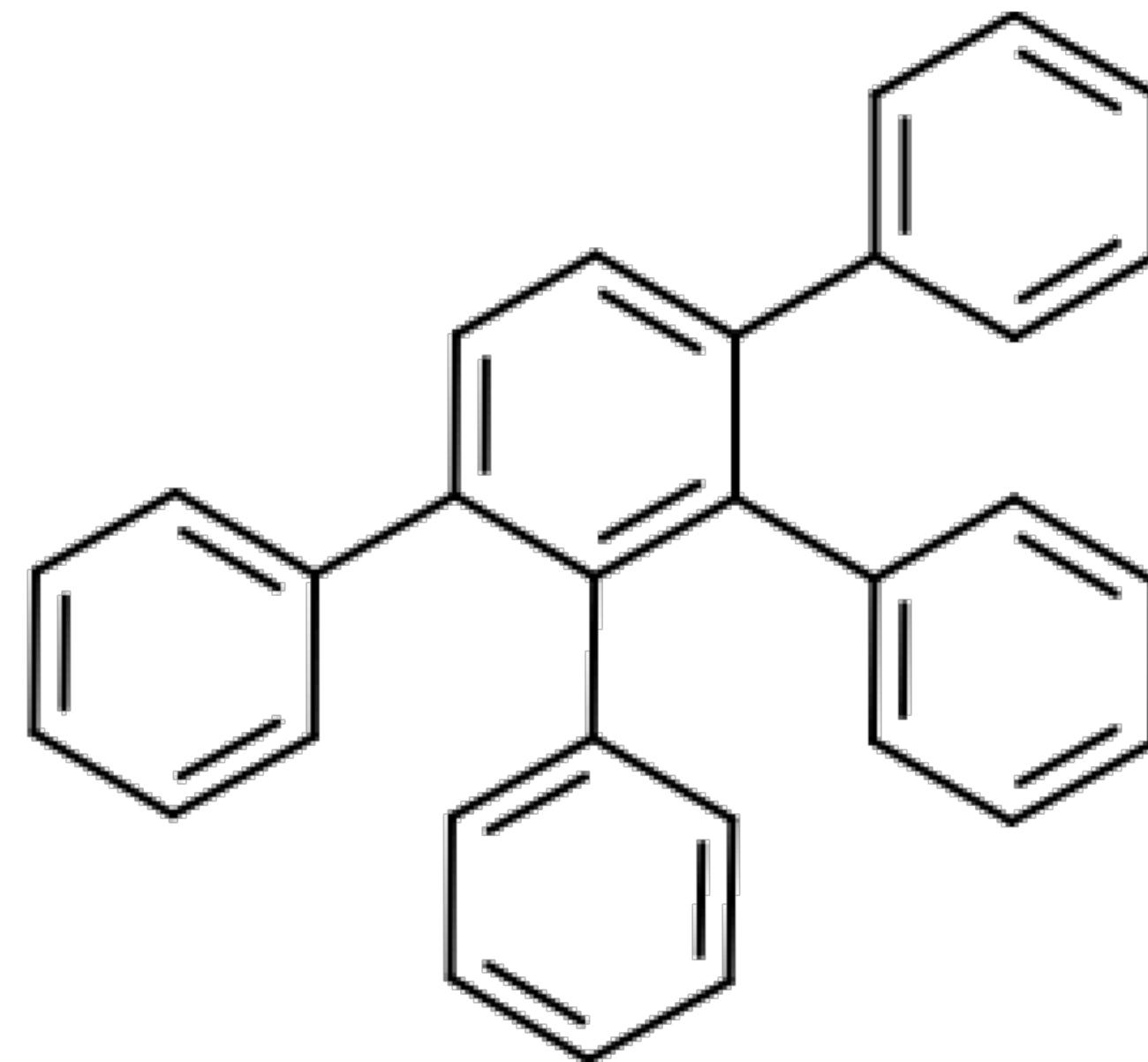


Using SMARTS/SMIRKS allows us to escape atom typing

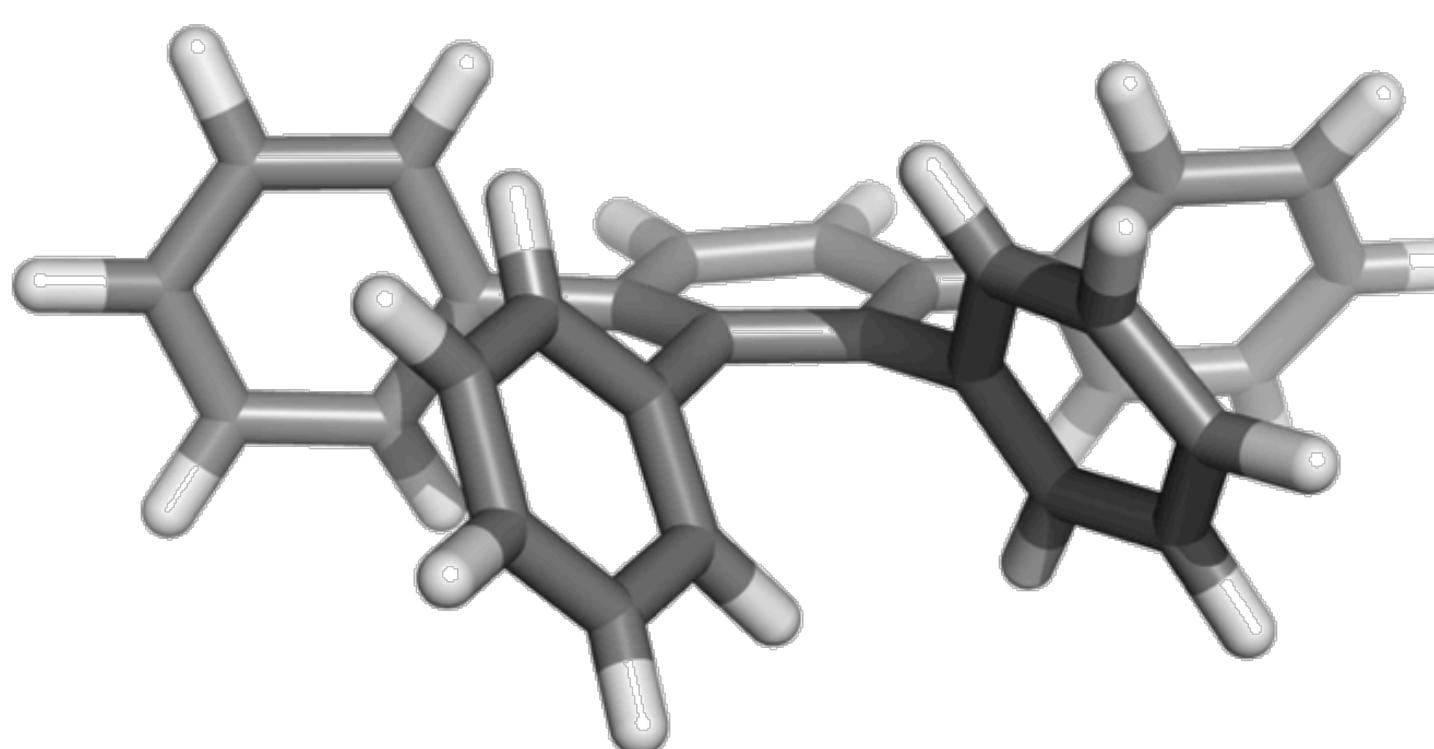
We use substructure searches on the molecule to assign parameters, rather than atom typing



Direct chemical perception utilizes bond order



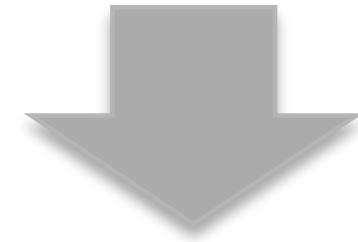
Torsion Chemistry	Barrier Height	Torsion Minimum
$[*:1]~[#6x3:2]:[#6x3:3]~[*:4]$	4	14.50
$[*:1]~[#6x3:2]-[#6x3:3]~[*:4]$	4	2.50
...		



We get the geometry right with no special treatment and far fewer parameters

smirnoff99Frosst is our adaptation of parm99+parm@Frosst into this format

Description	Force Field	Lines of parameters
Basic Amber FF:	parm99	720
Merck Frosst small mol:	parm@Frosst	2893
Total:		3613

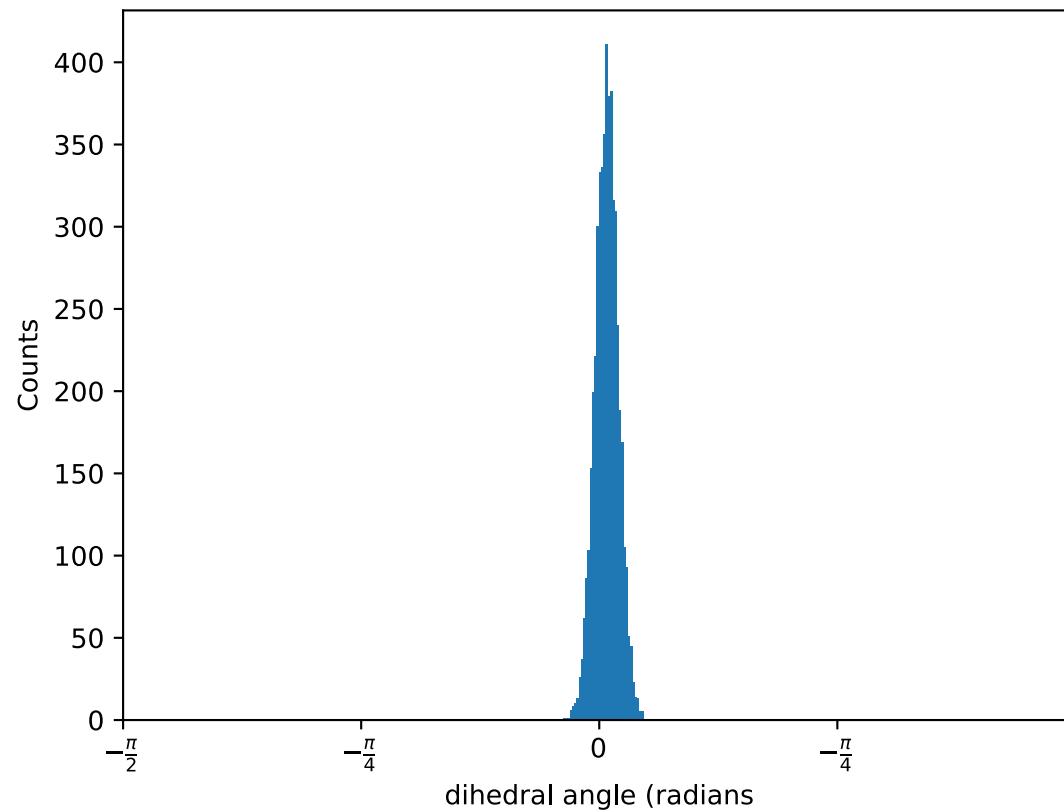
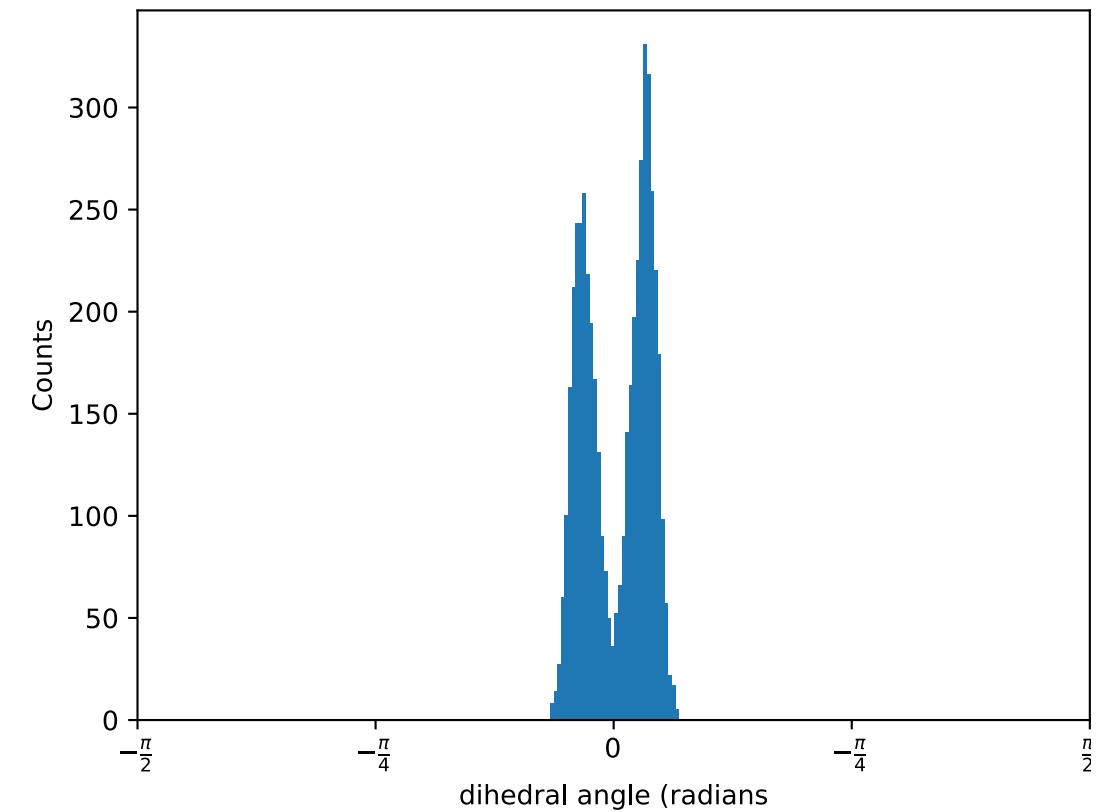
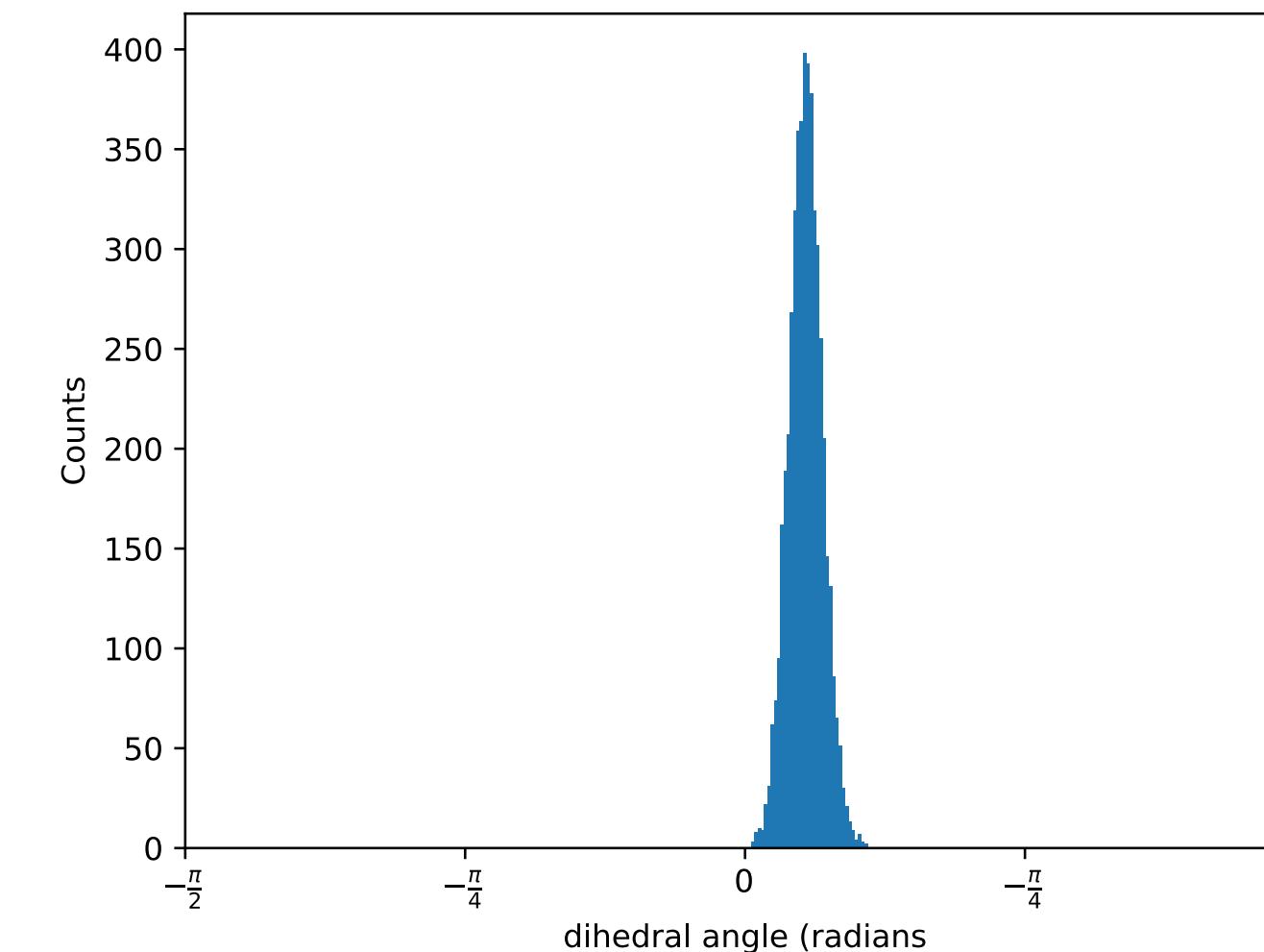
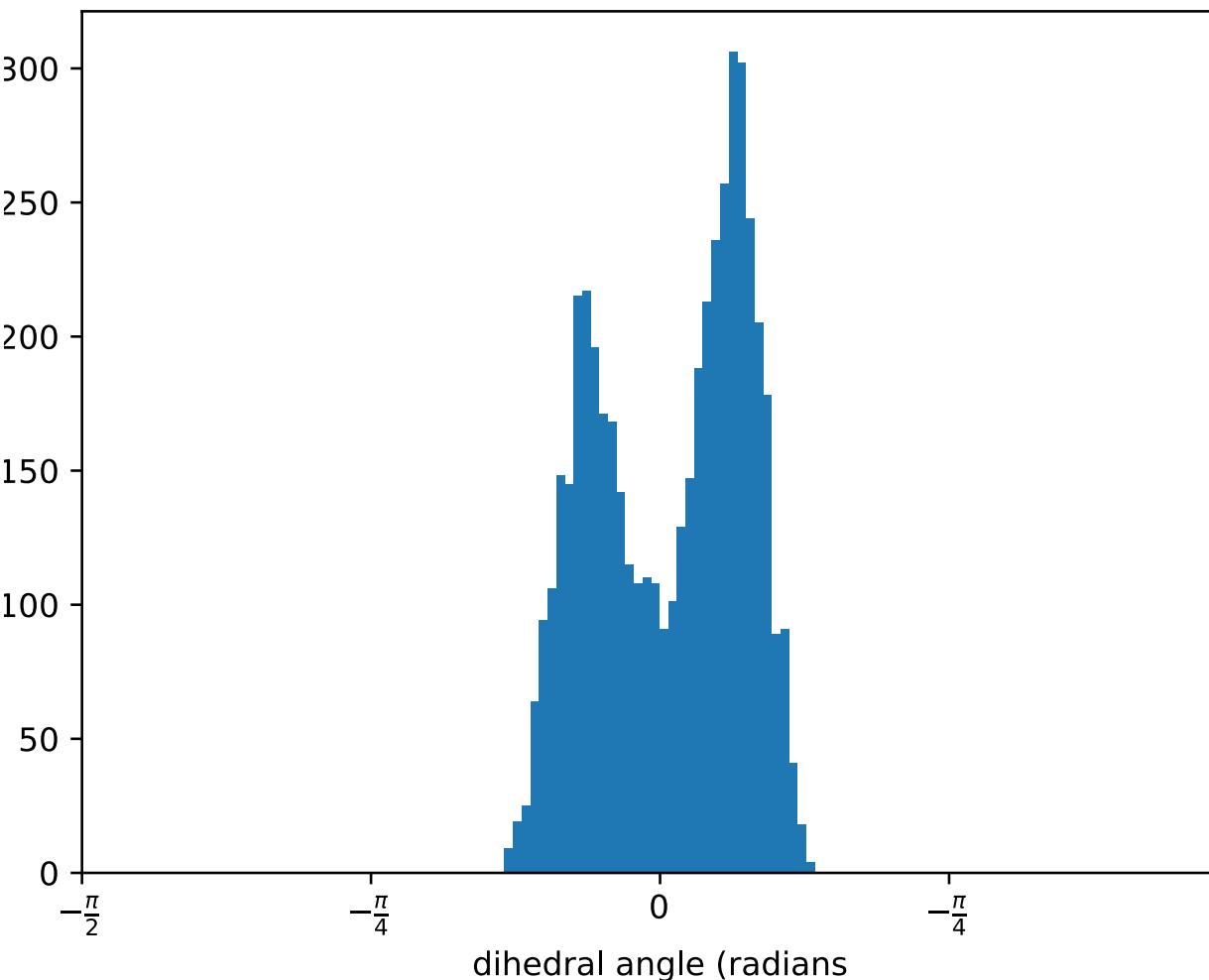
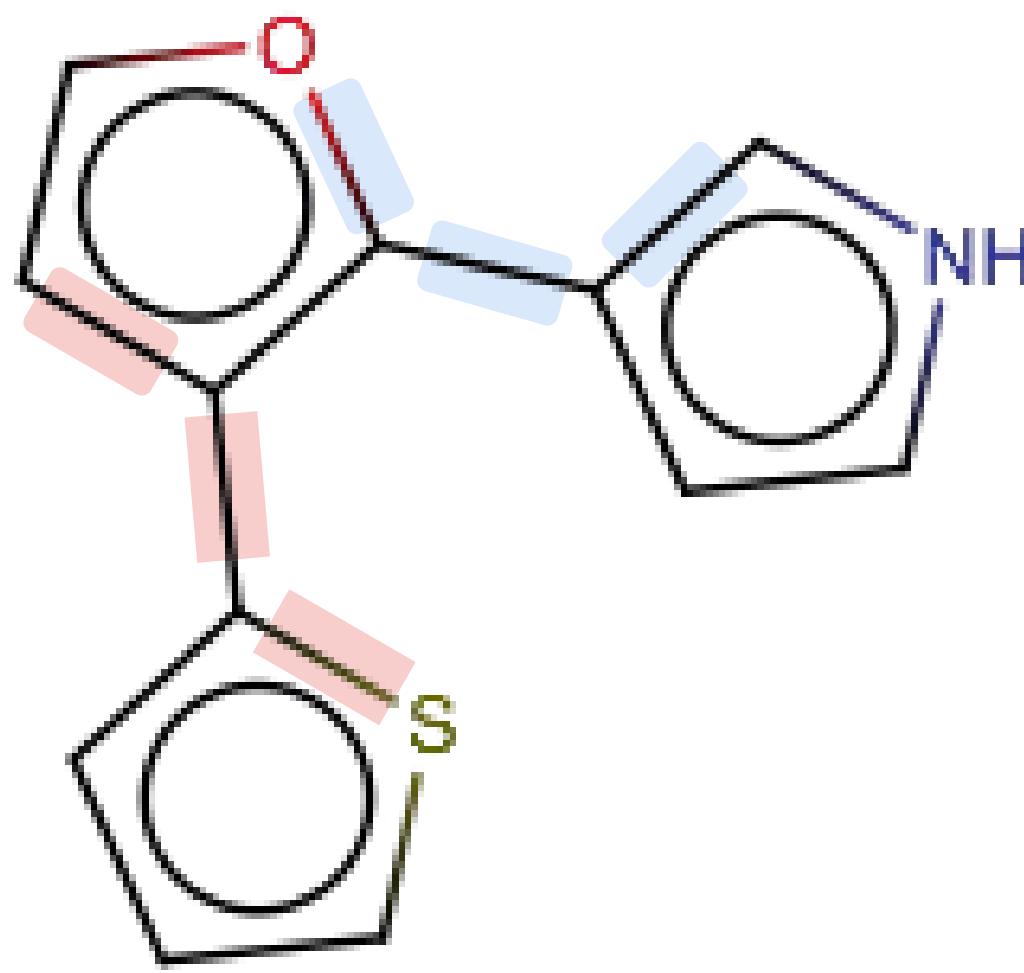


smirnoff99Frosst 332

	Chemical Space Coverage	
	smirnoff 99Frosst	parm @Frosst
Database		
DrugBank	99.7%	60%
ZINC	99.8%	52%
eMolecules	99.5%	--

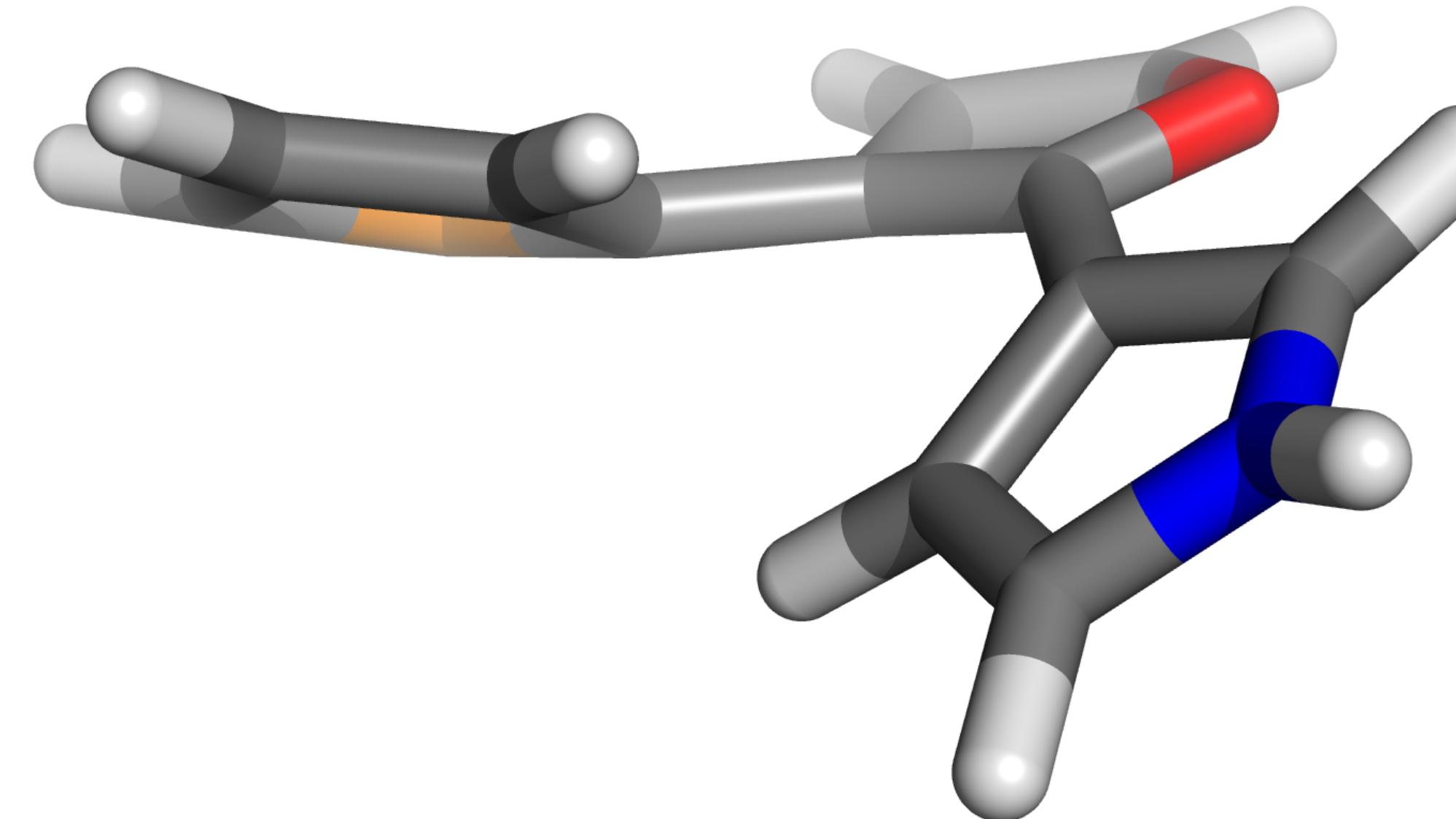
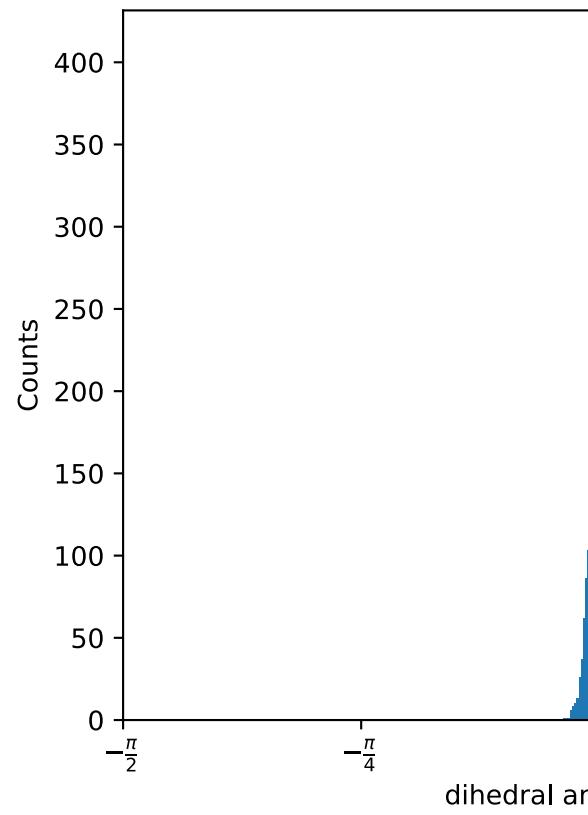
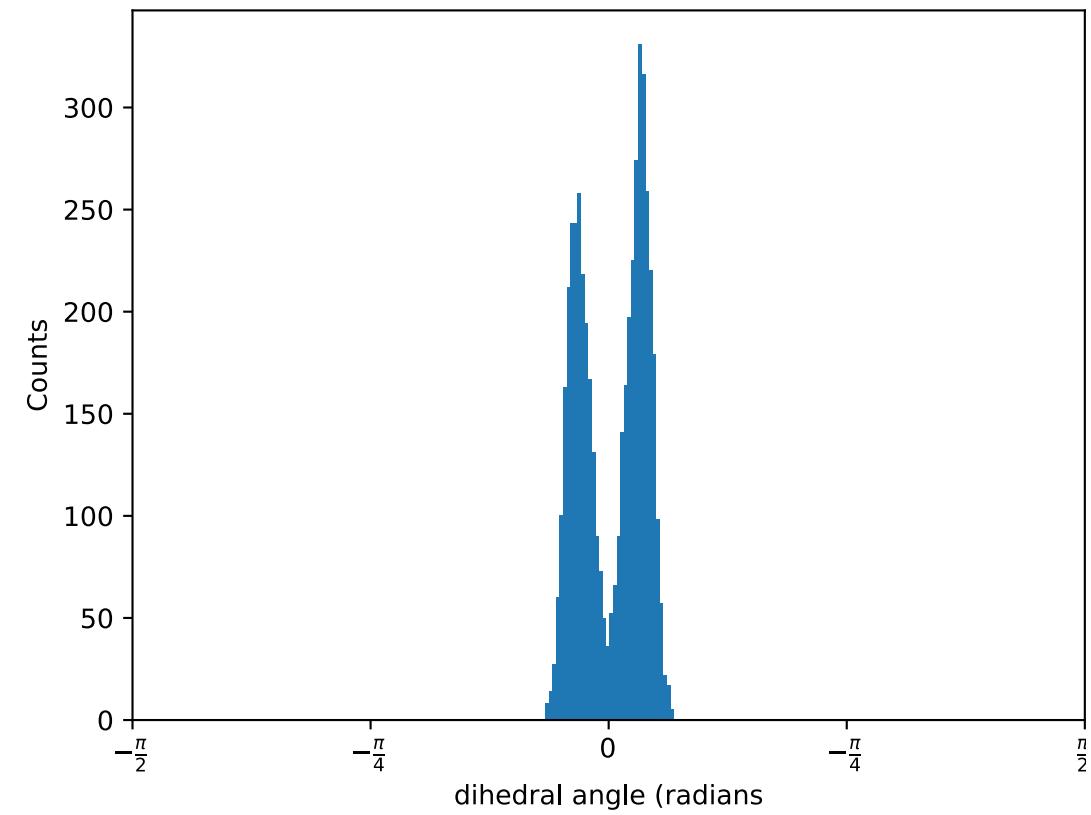
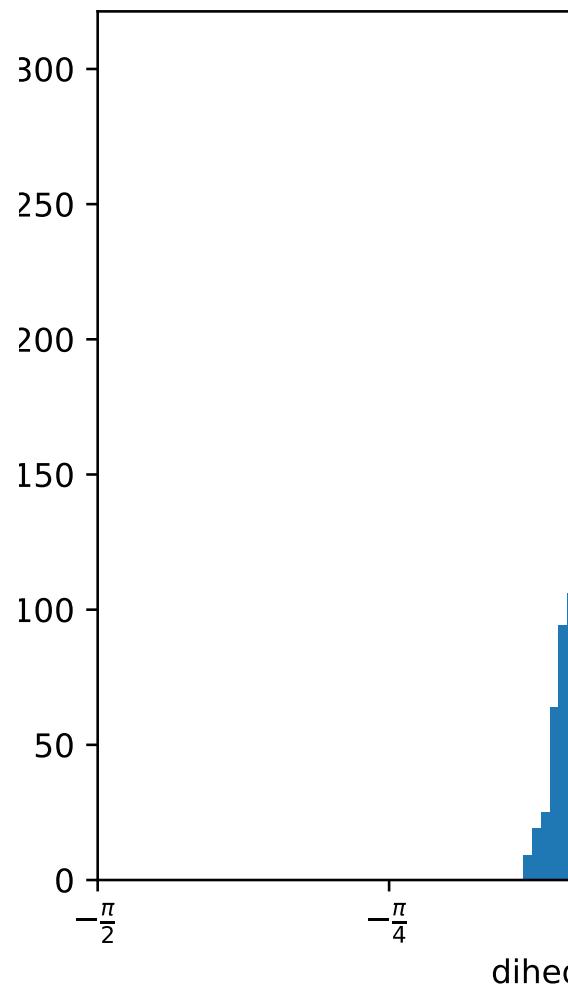
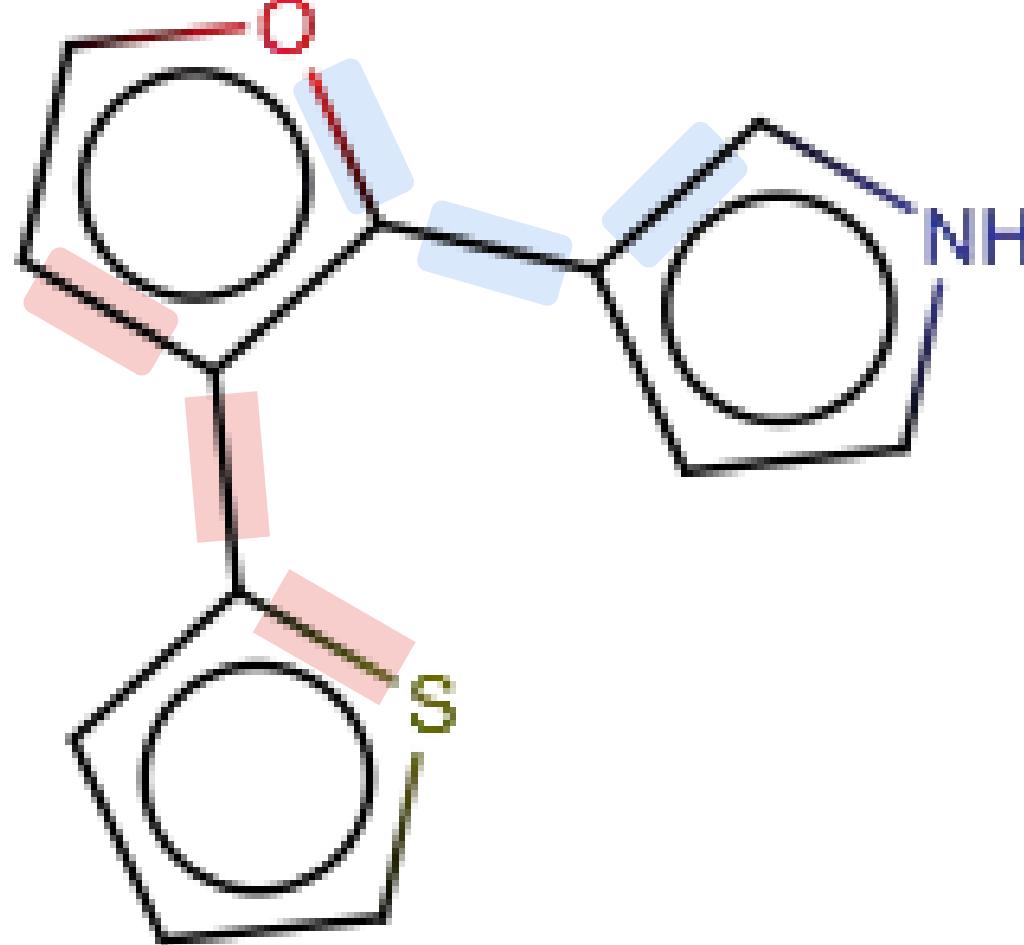
- Less than 1/10 the size of the original force field
- Removes redundancy
- Almost completely covers pharmaceutical chemical space

Out of the box it fixes a variety of problems, including siblings of the biphenyl problem

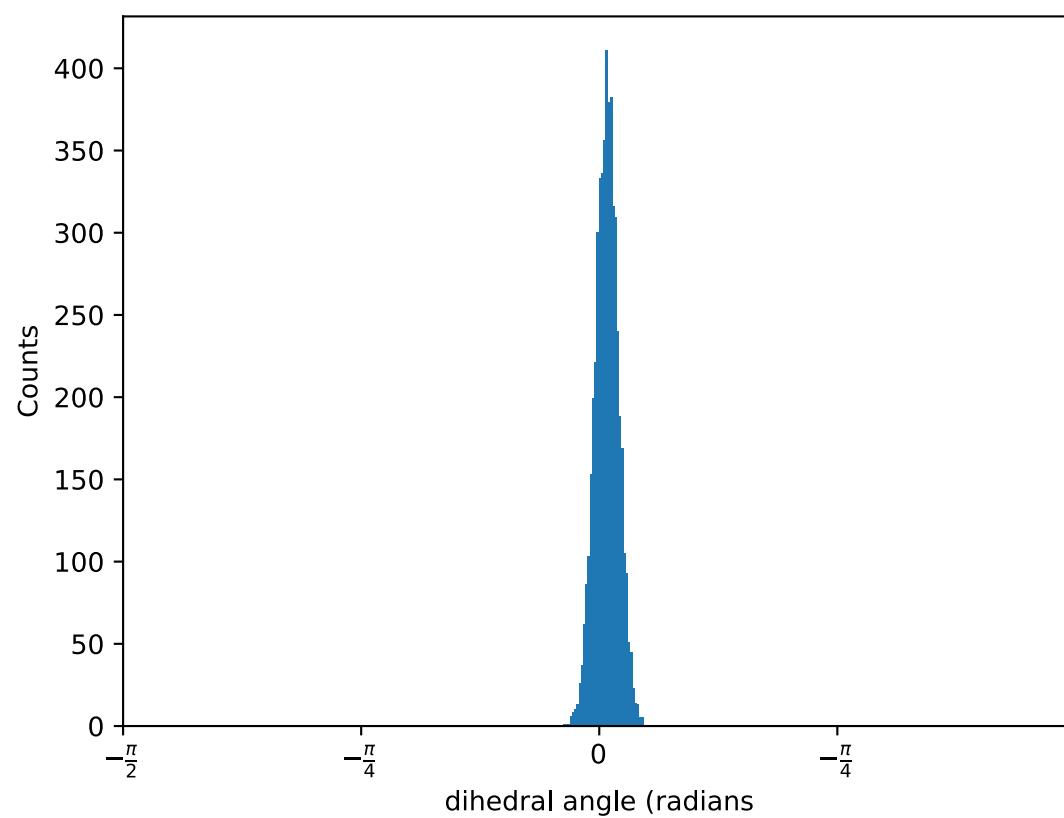
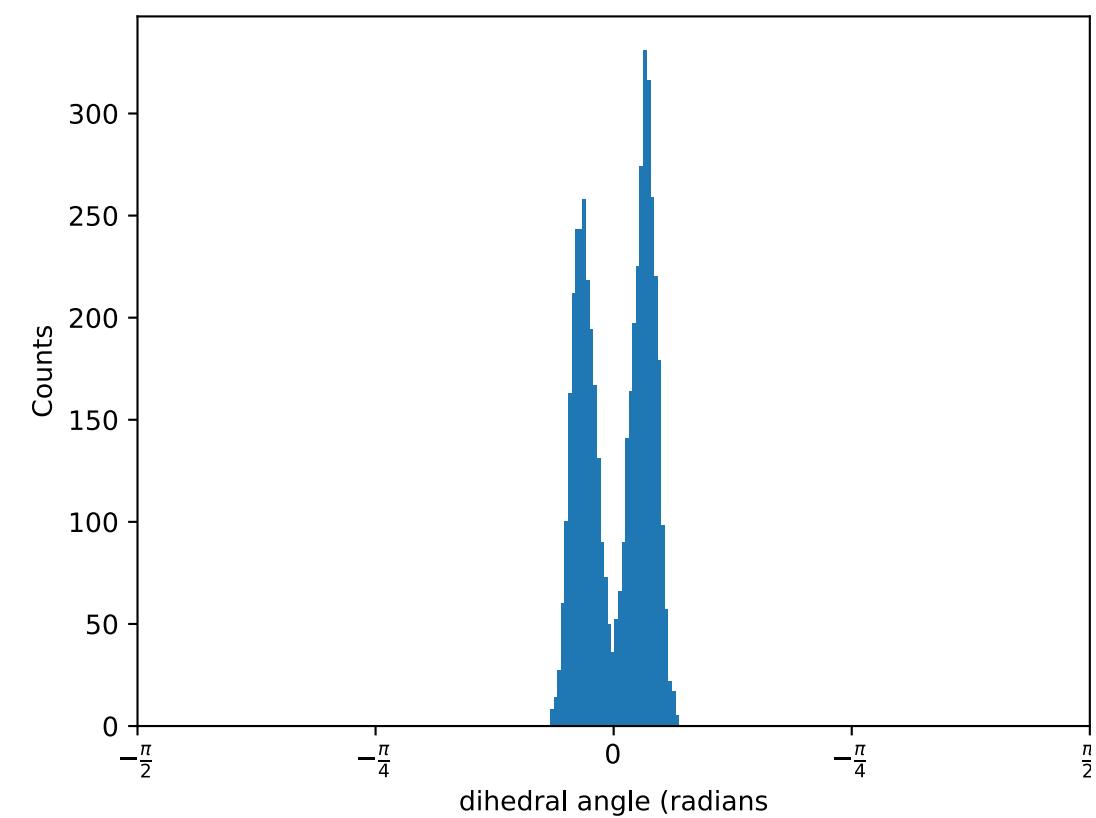
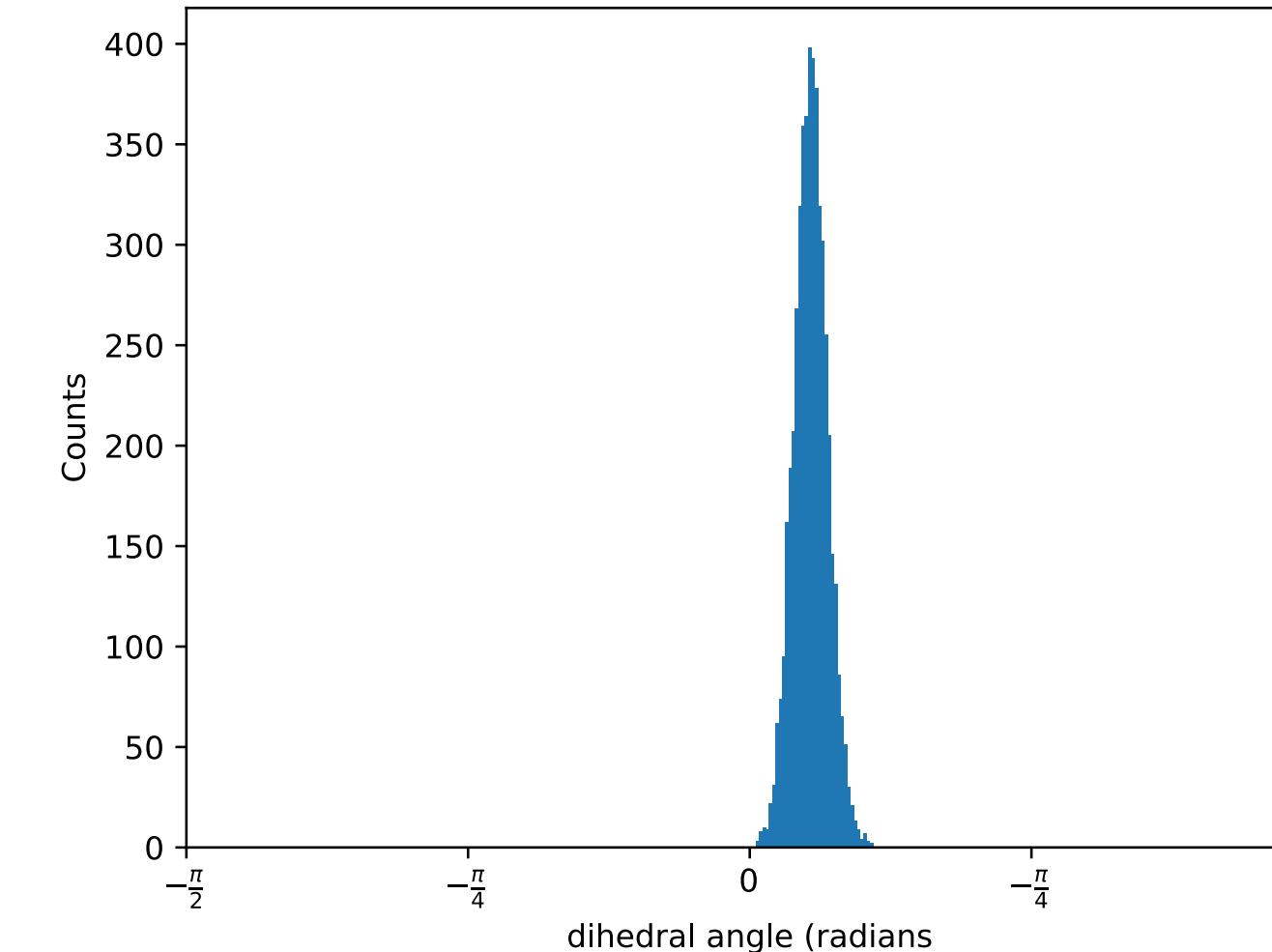
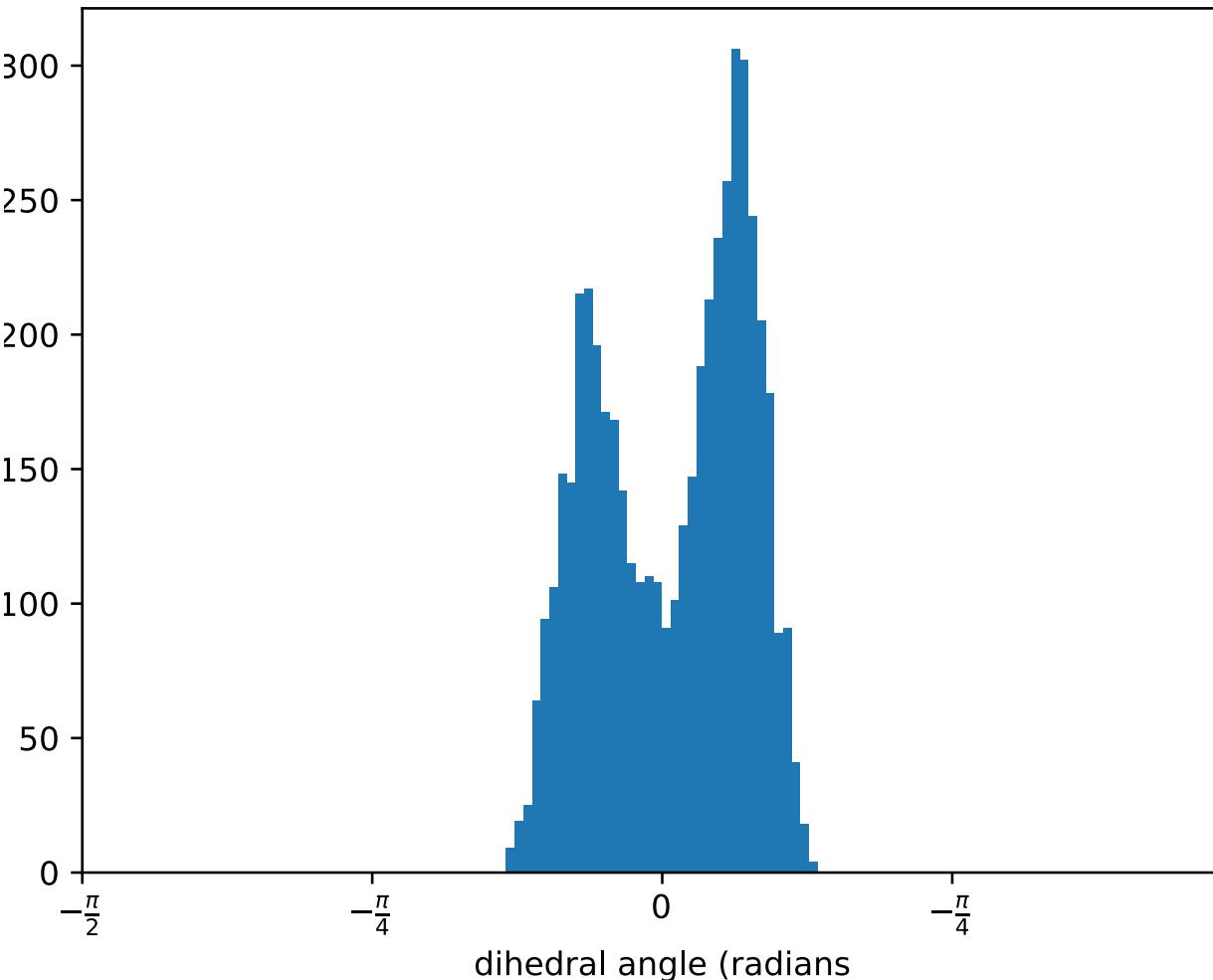
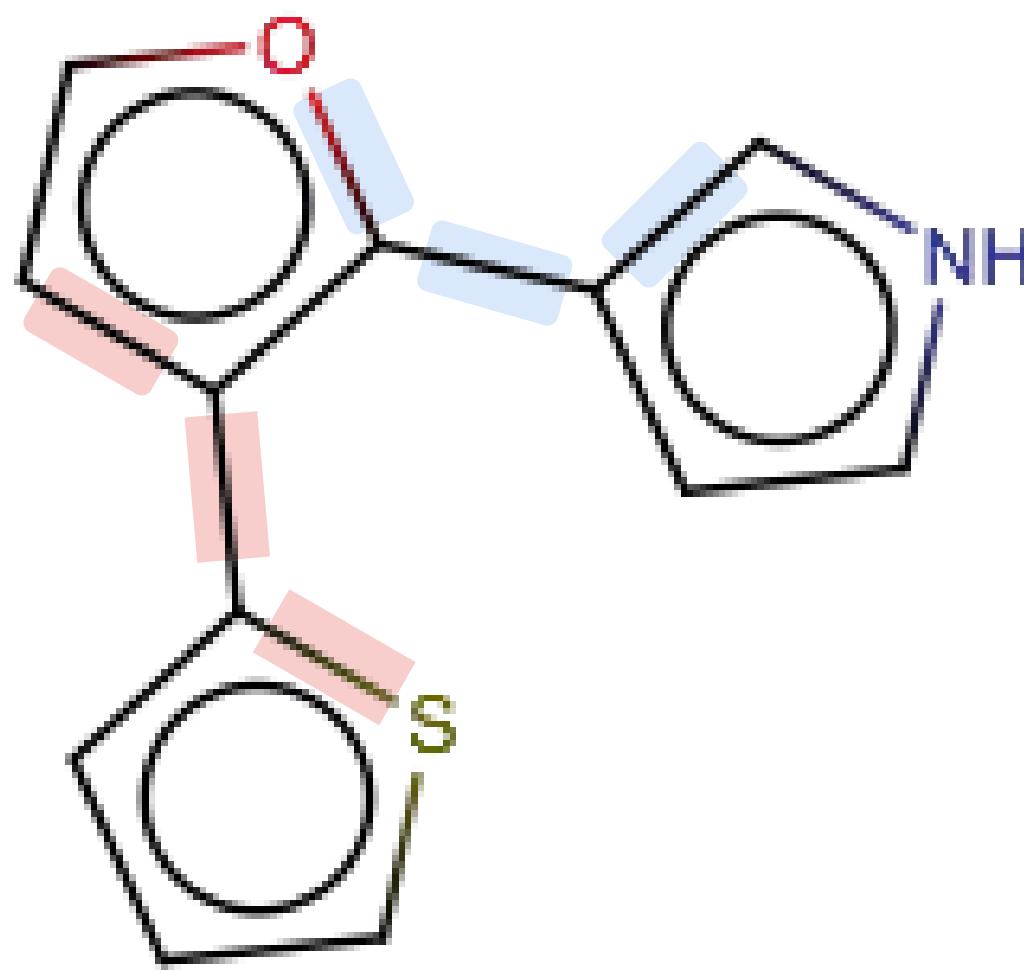


- SMIRNOFF makes these rotatable single bonds
- GAFF/GAFF2 make them essentially aromatic
- Central ring buckles due to steric strain

Out of the box it fixes a variety of problems, including siblings of the biphenyl problem



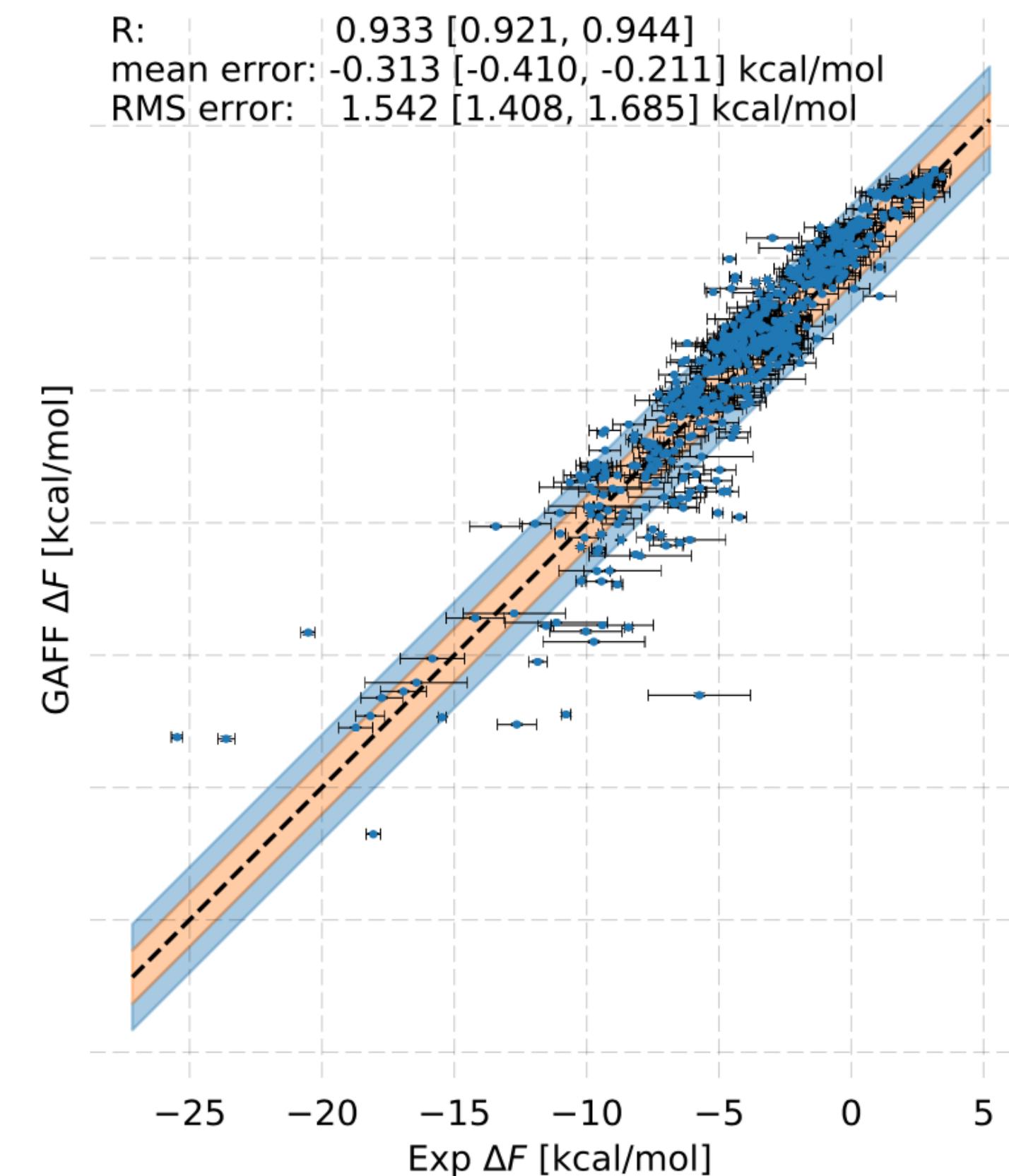
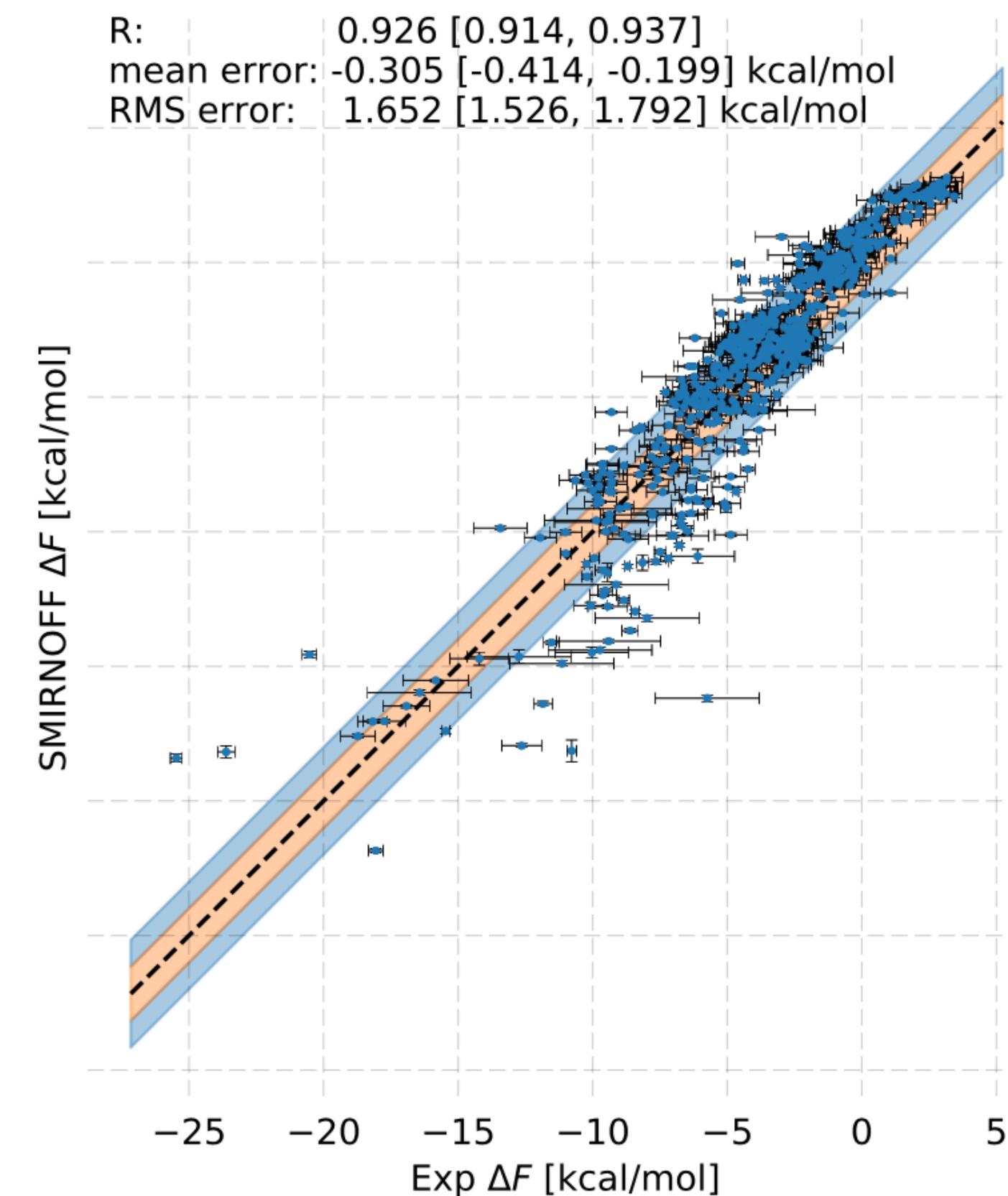
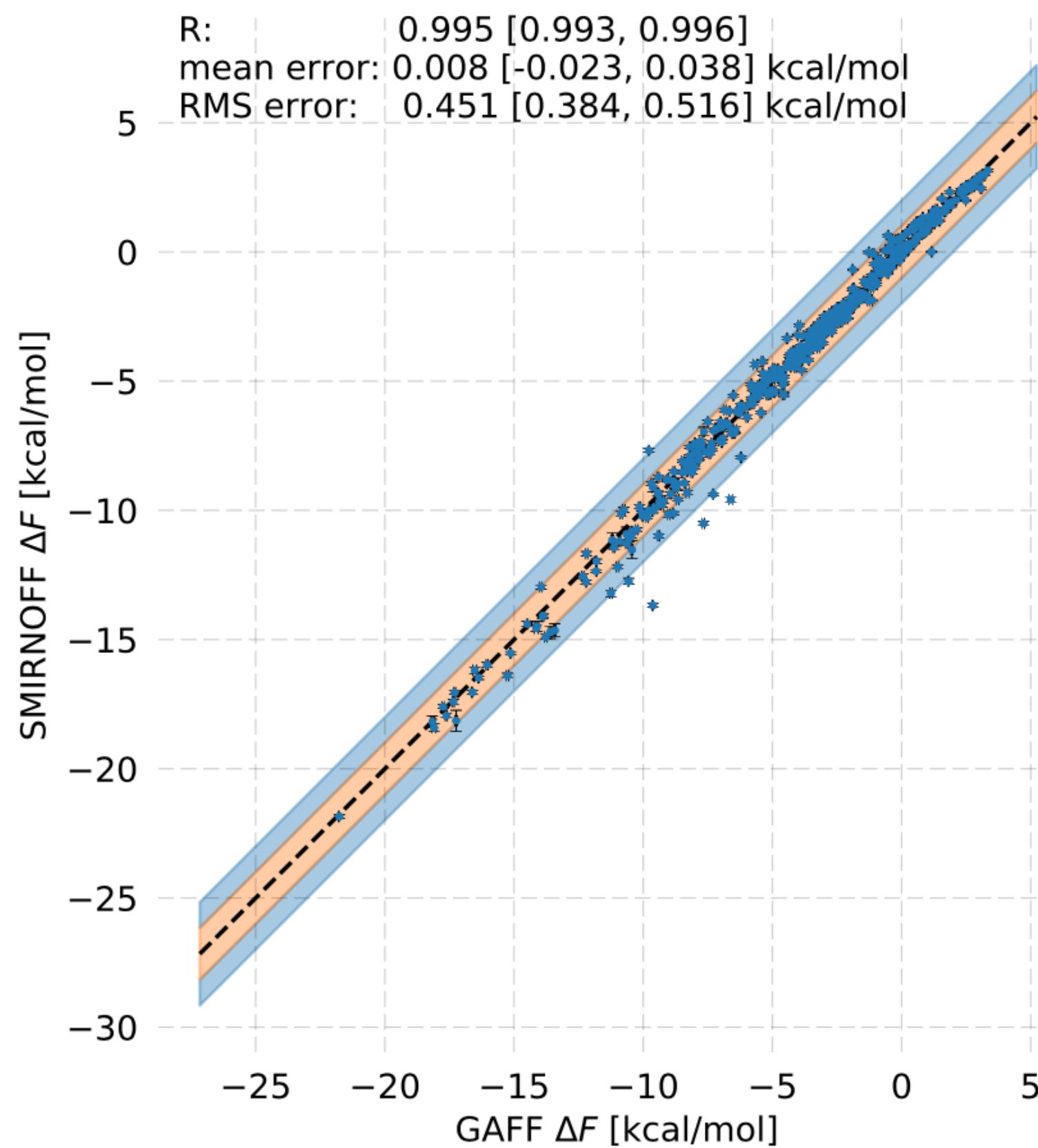
Out of the box it fixes a variety of problems, including siblings of the biphenyl problem



- SMIRNOFF makes these rotatable single bonds
- GAFF/GAFF2 make them essentially aromatic
- Central ring buckles due to steric strain

We think smirnoff99Frosst is a great starting point for parameterization work

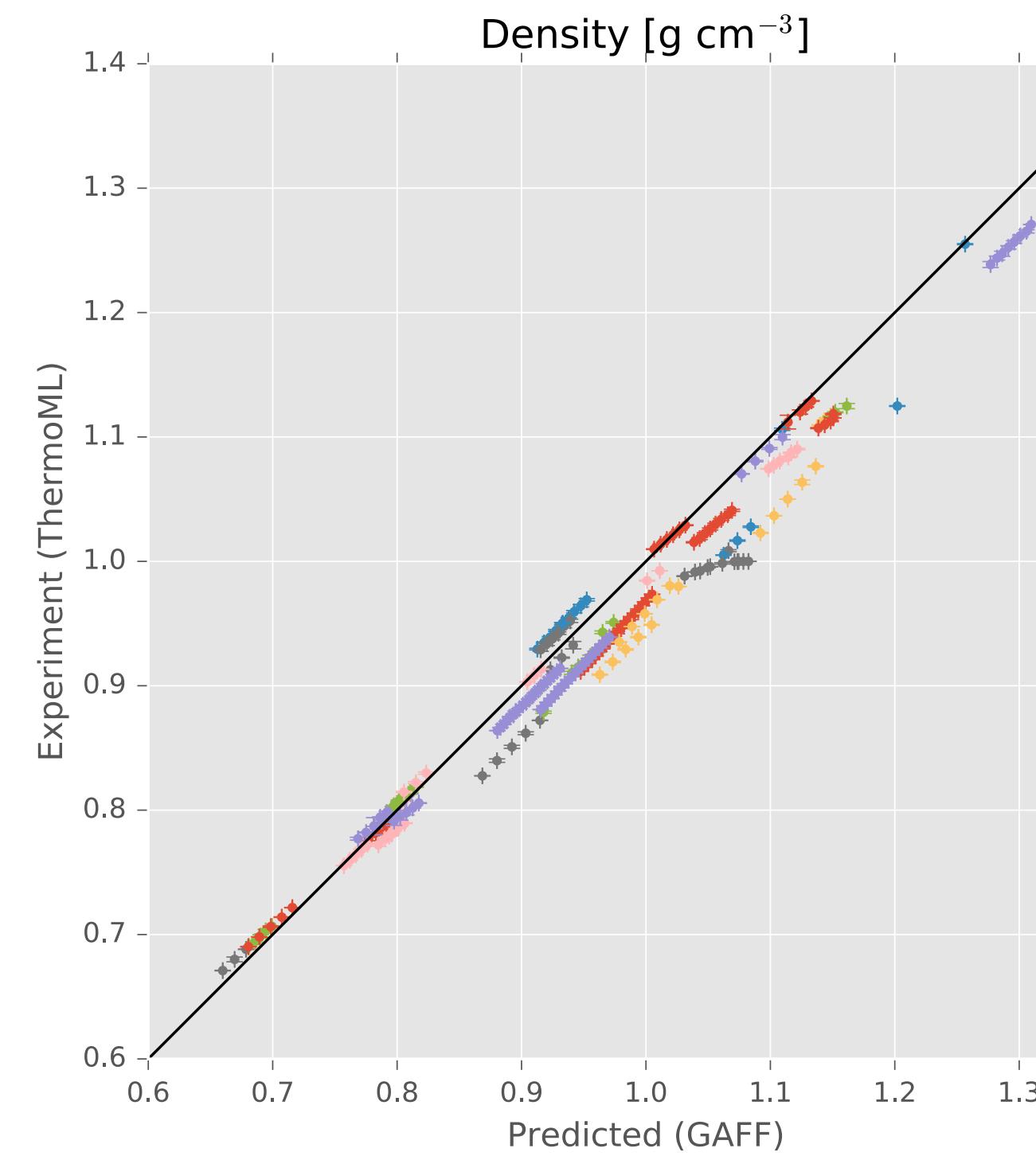
FreeSolv hydration free energy benchmark



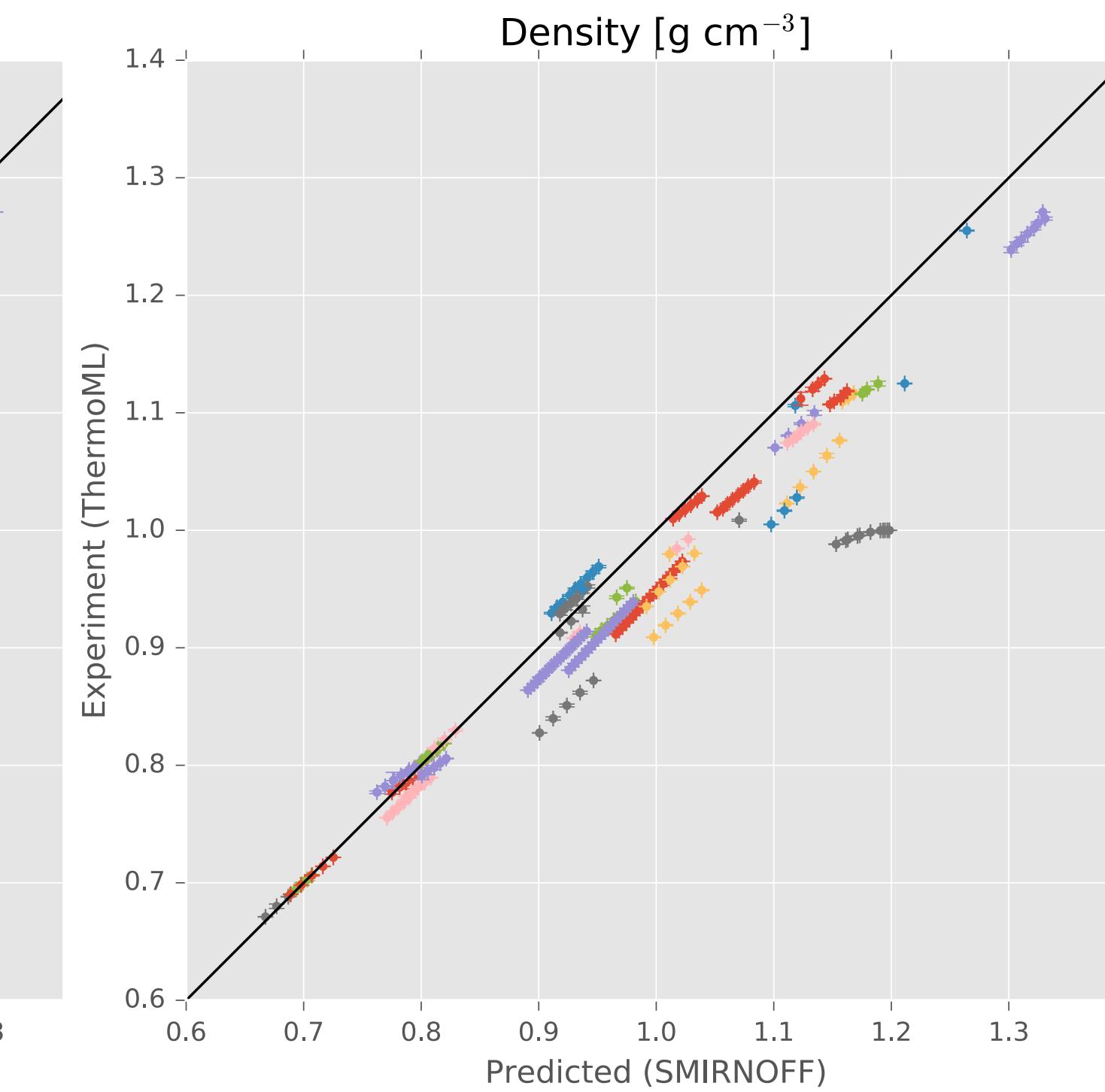
It is competitive with GAFF but with far fewer parameters

ThermoML Archive density/dielectric benchmark set

GAFF



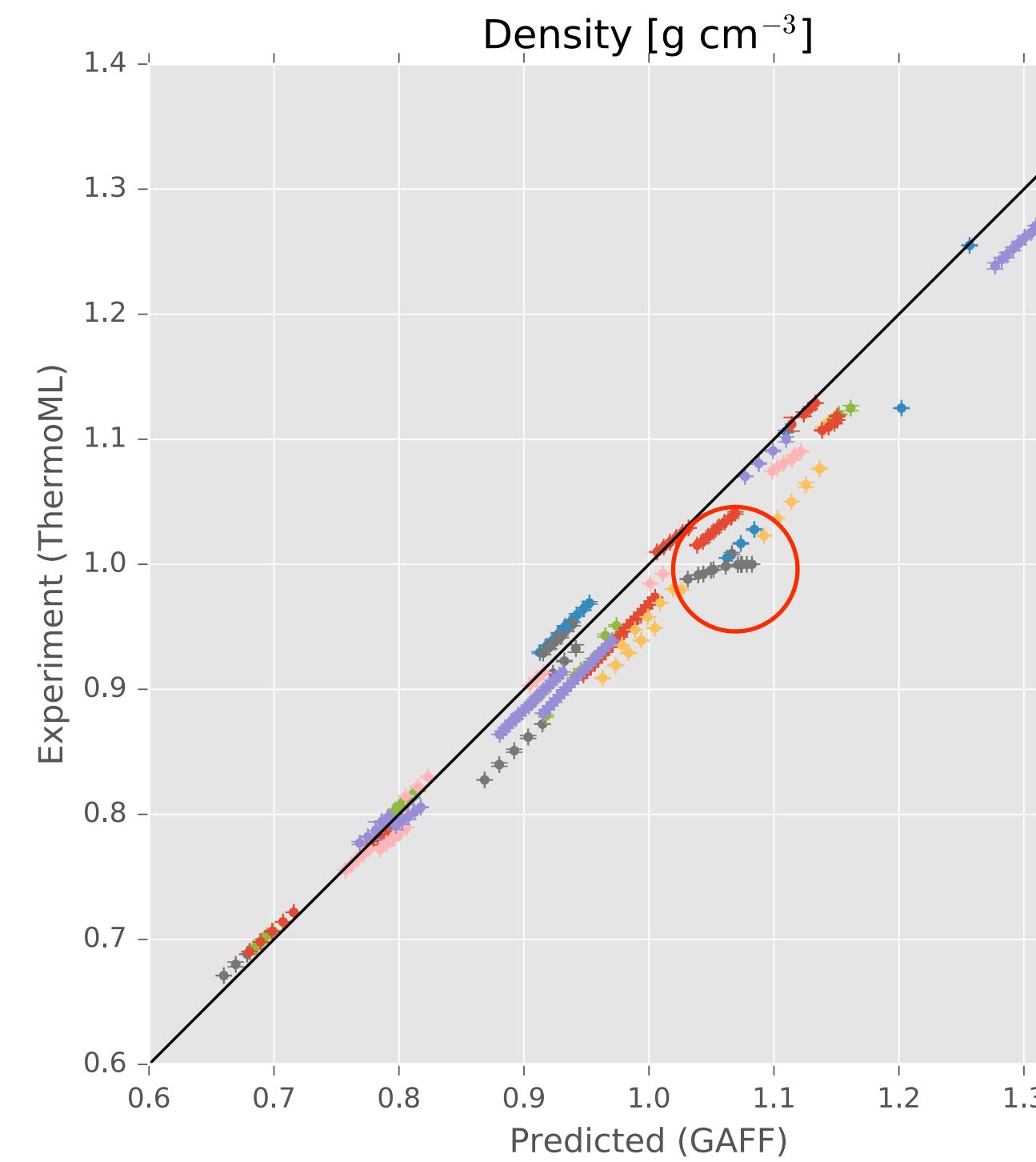
smirnoff99frosst



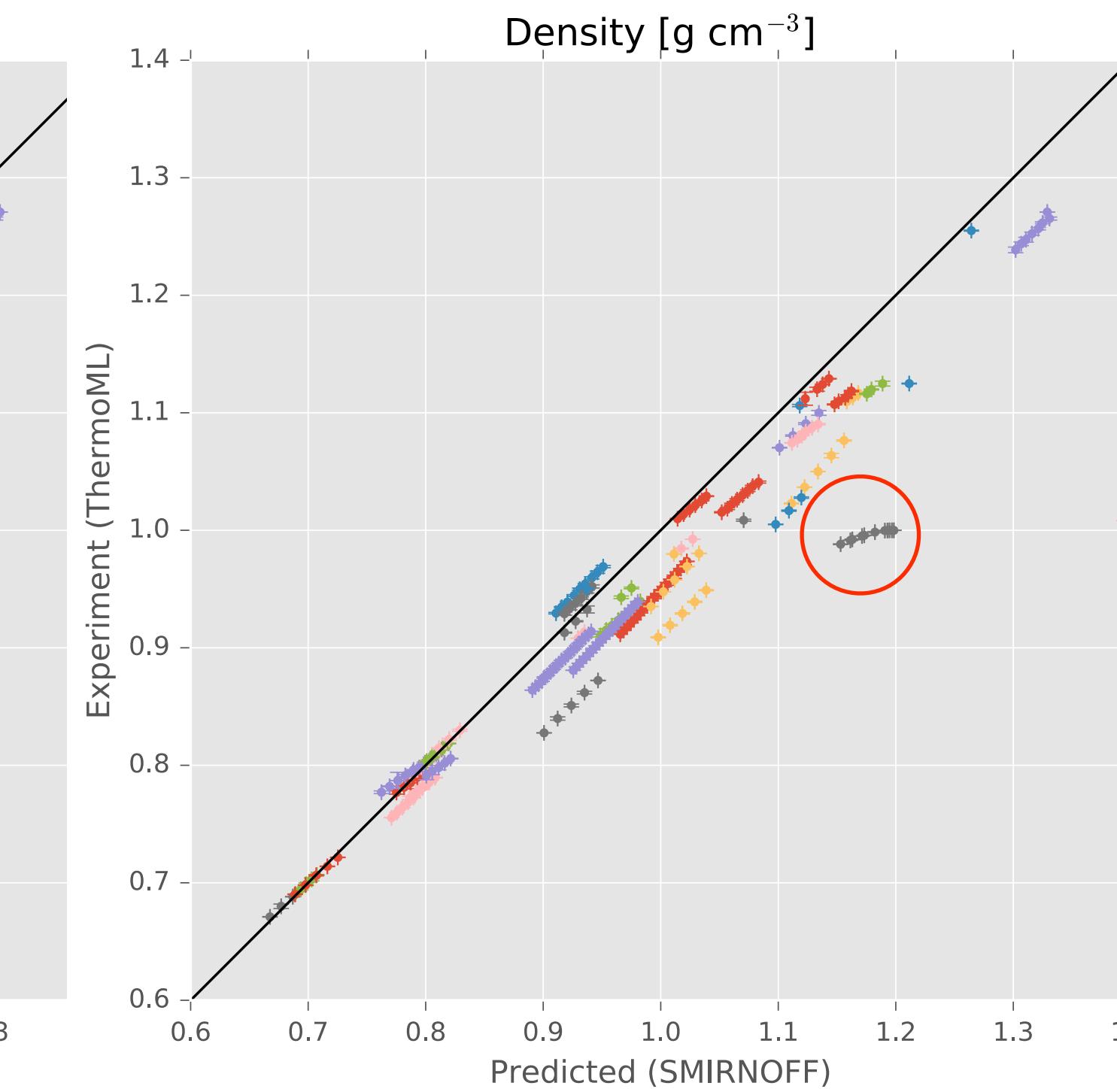
It is competitive with GAFF but with far fewer parameters

ThermoML Archive density/dielectric benchmark set

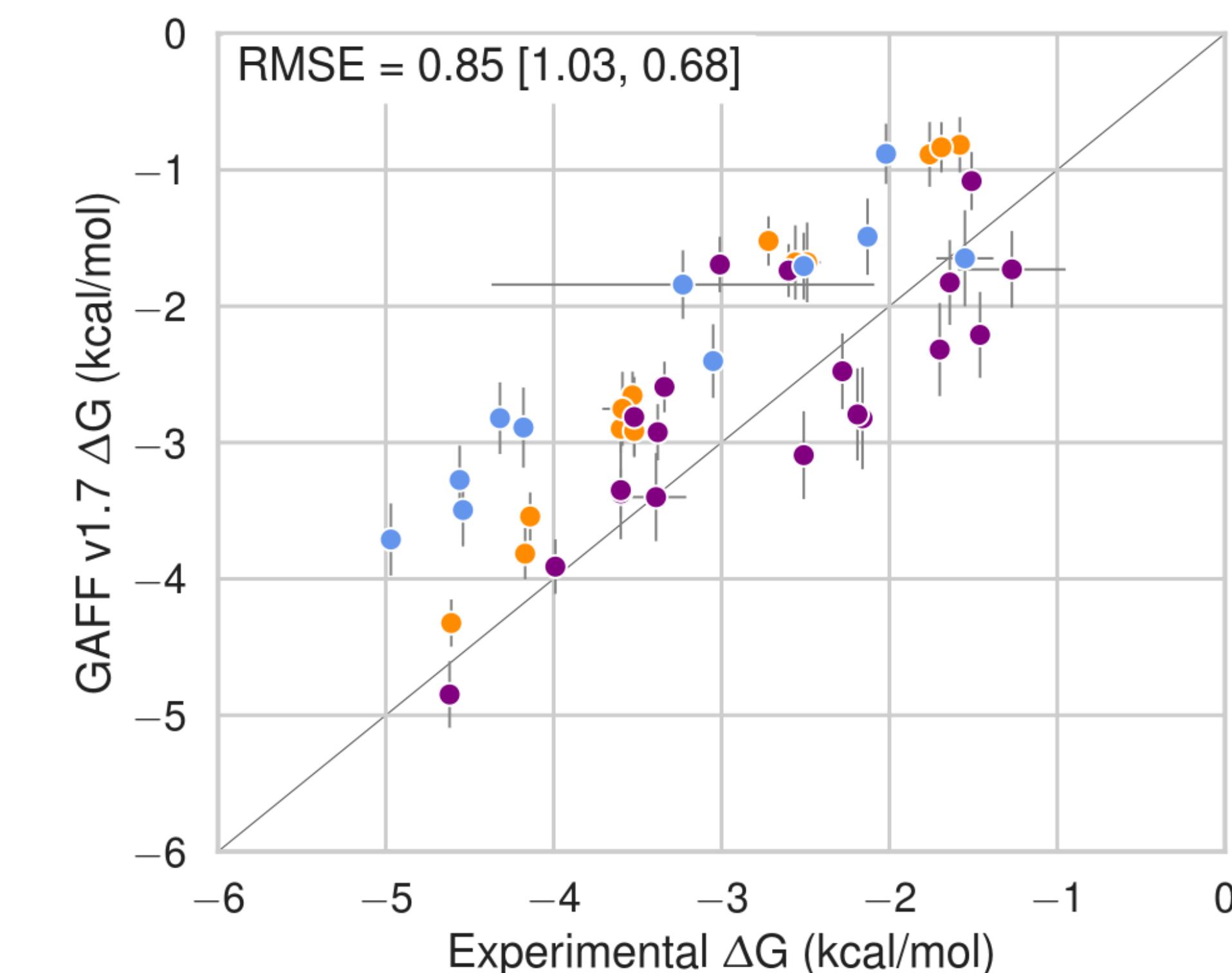
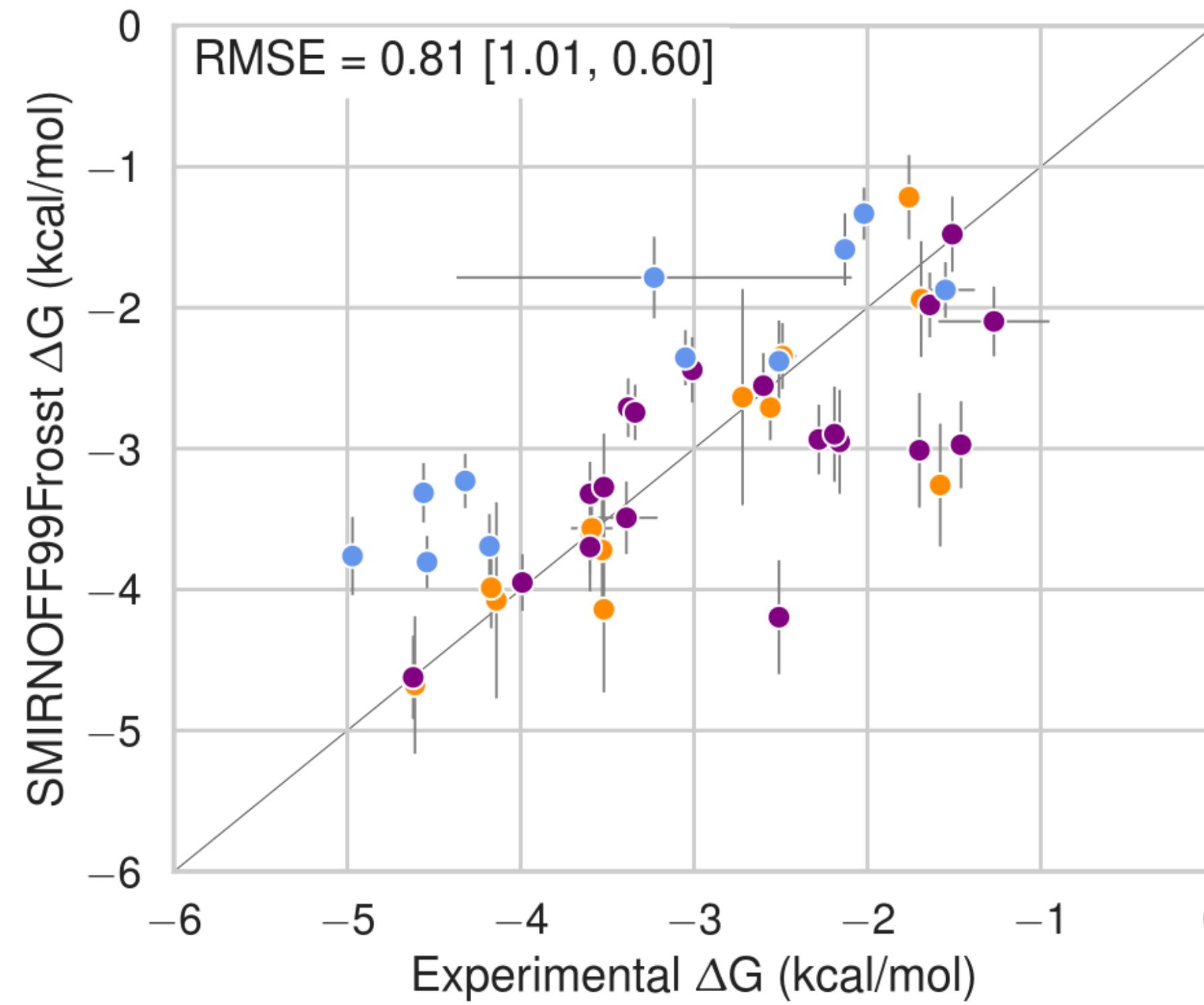
GAFF



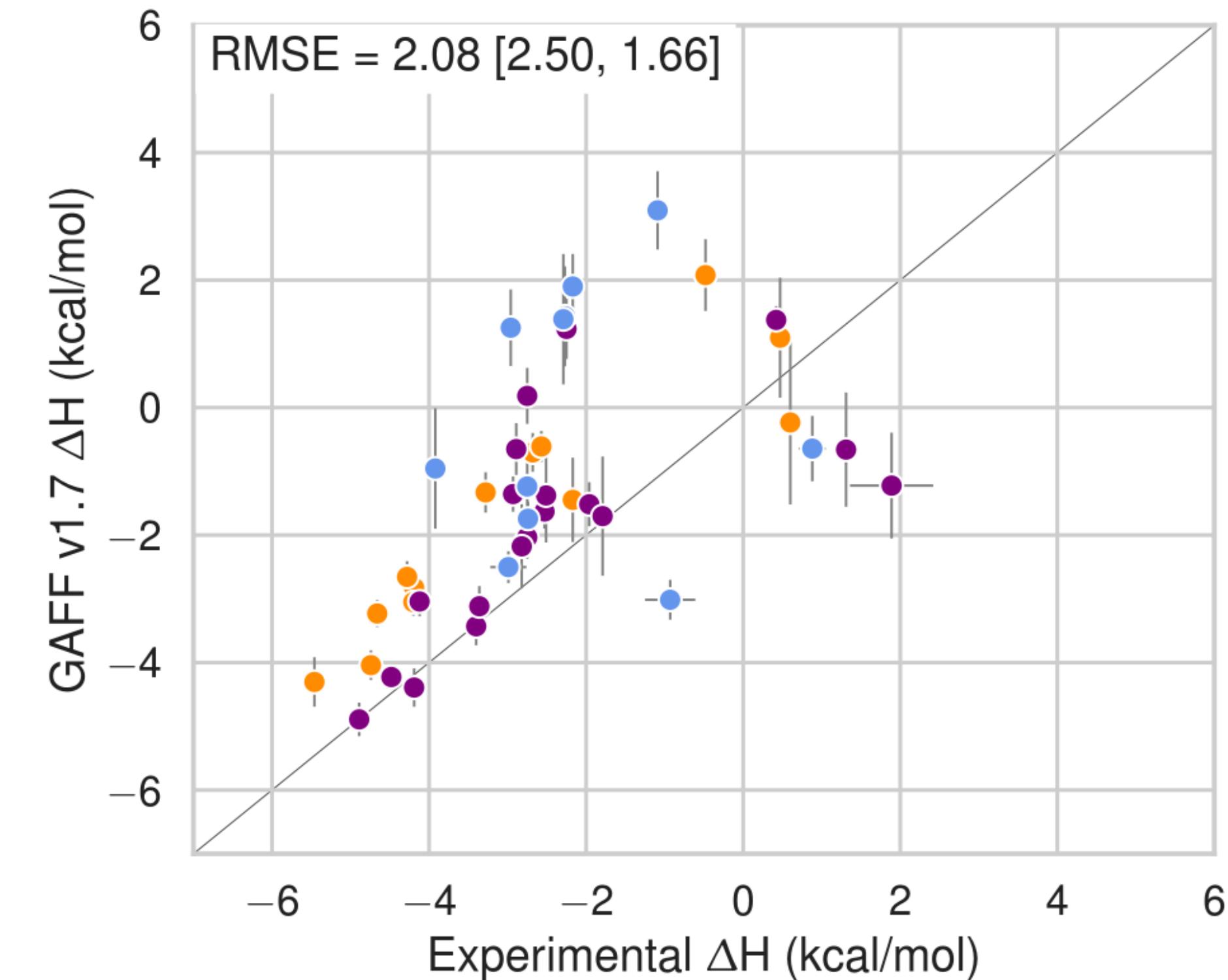
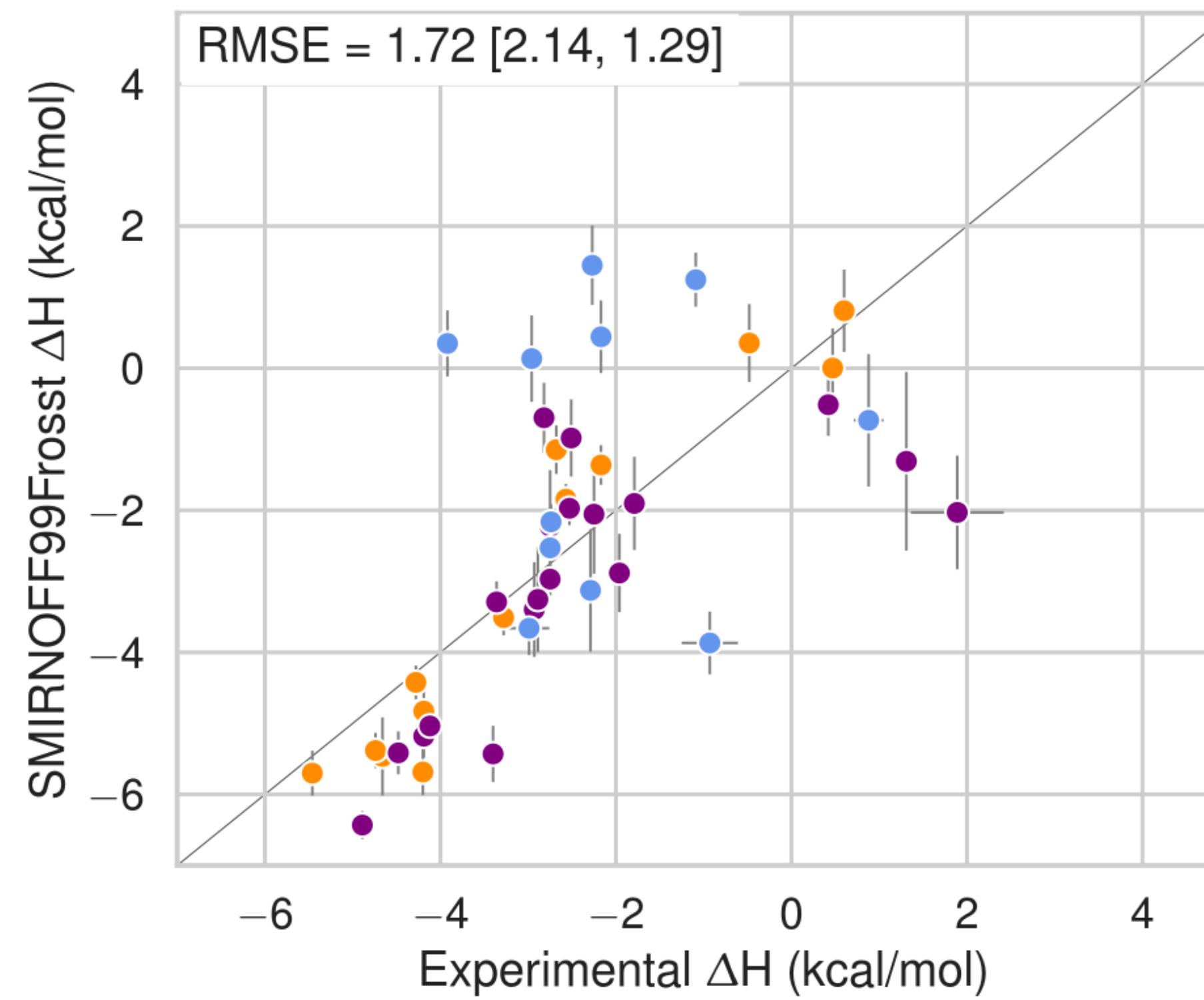
smirnoff99frosst



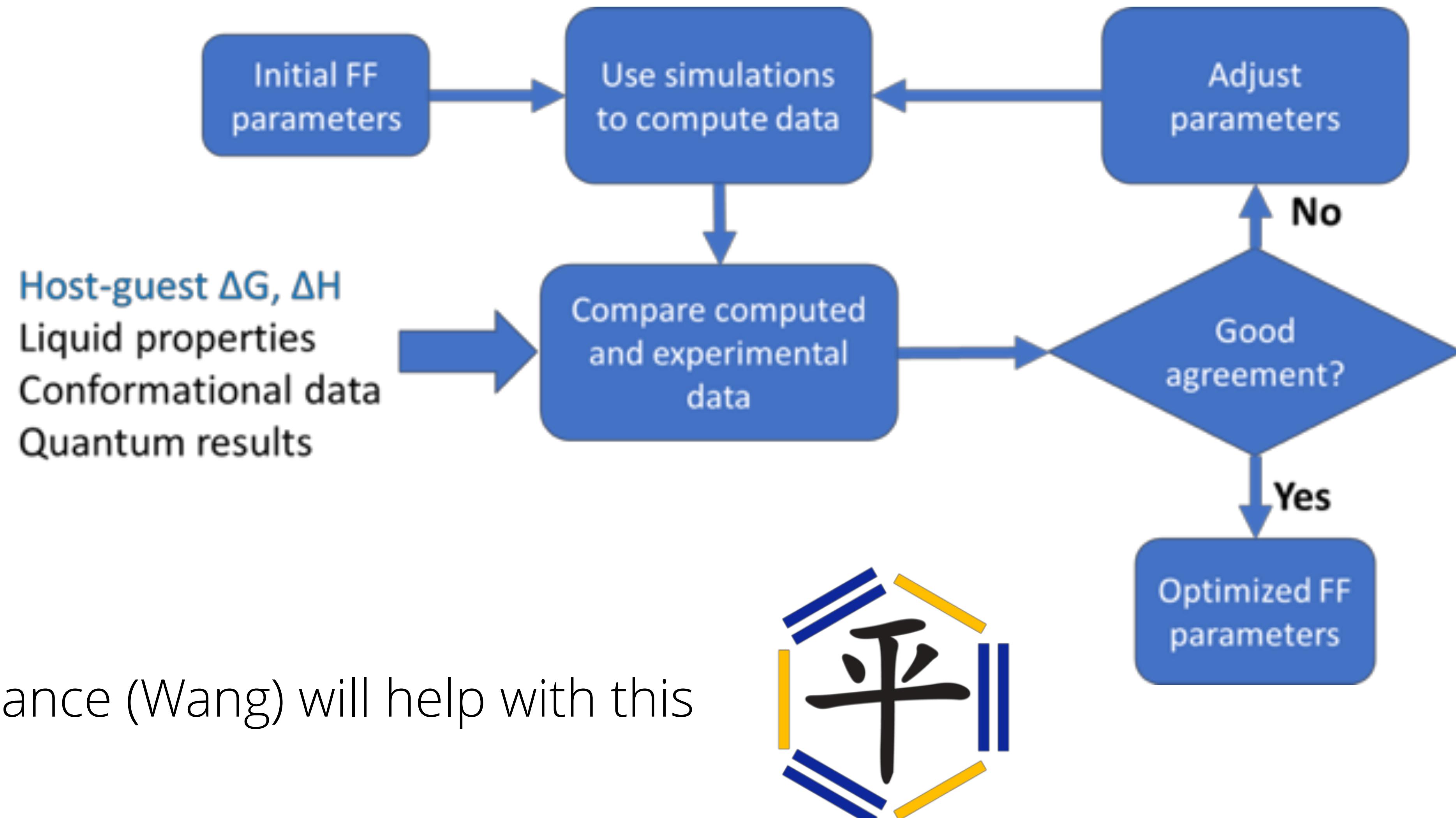
We're also interested in using binding data for fitting,
and it turns out to do fairly well on host-guest
binding also



It even works fairly well on enthalpies of binding



Automated refitting plays a key role in our plans



We plan stages of refitting and improvement

EASY

Bonds/angle refitting to high-level QM

Generation 1

Refit torsions to high-level QM for drug-like molecules

Valence type expansion

Small molecule Lennard-Jones improvements based on liquid property data

Lennard-Jones type expansion

Inclusion of host-guest thermodynamics in fitting

Refit BCCs to high-quality QM and liquid-phase data

Use partial bond orders in fitting process to simplify valence type complexity

Introduce off-site charges and BCCs to support them

Generation 2

Complete Lennard-Jones refit (requires breaking AMBER compatibility)

Bayesian parameter uncertainty propagation to quantify systematic error

Surrogate thermodynamic models to accelerate forcefield parameterization

Automated type refinement to penalize complexity

Selective polarizability

HARD

FORCEFIELD GENERATIONS

Generation 1: Improvement of an AMBER-compatible small molecule forcefield

Full compatibility with major simulation packages (Amber, CHARMM, gromacs, OpenMM, NAMD, ...)

Energetically compatible with AMBER biopolymer forcefields

Targeted improvements to smirnoff99Frosst to remedy known deficiencies and refit selected parameters (especially torsions) on generated QM data and curated datasets

Generation 2: A full refit for increased accuracy

Full compatibility with major simulation packages (Amber, CHARMM, gromacs, OpenMM, NAMD, ...)

Small molecule and biopolymer parameters via full consistent refit of Lennard-Jones parameters

Releases will improve accuracy by introducing off-site charges, making use of partial bond orders during parameterization, and including high-quality liquid mixture and host-guest data

There are lots of types of data we can use

- densities of neat liquids and miscible liquid mixtures
- enthalpies of mixing of miscible molecular liquids
- transfer free energies (partition and distribution coefficients, hydration free energies)
- host-guest binding thermodynamics (free energies and enthalpies)
- small molecule 1D/2D NMR data (chemical shifts, J-coupling constants, NOE/ROEs)
- dielectric constants of neat liquids (and possibly mixtures)
- speed of sound data
- small molecule crystal structures and primary reflection data (CCSD)
- protein-ligand binding free energies

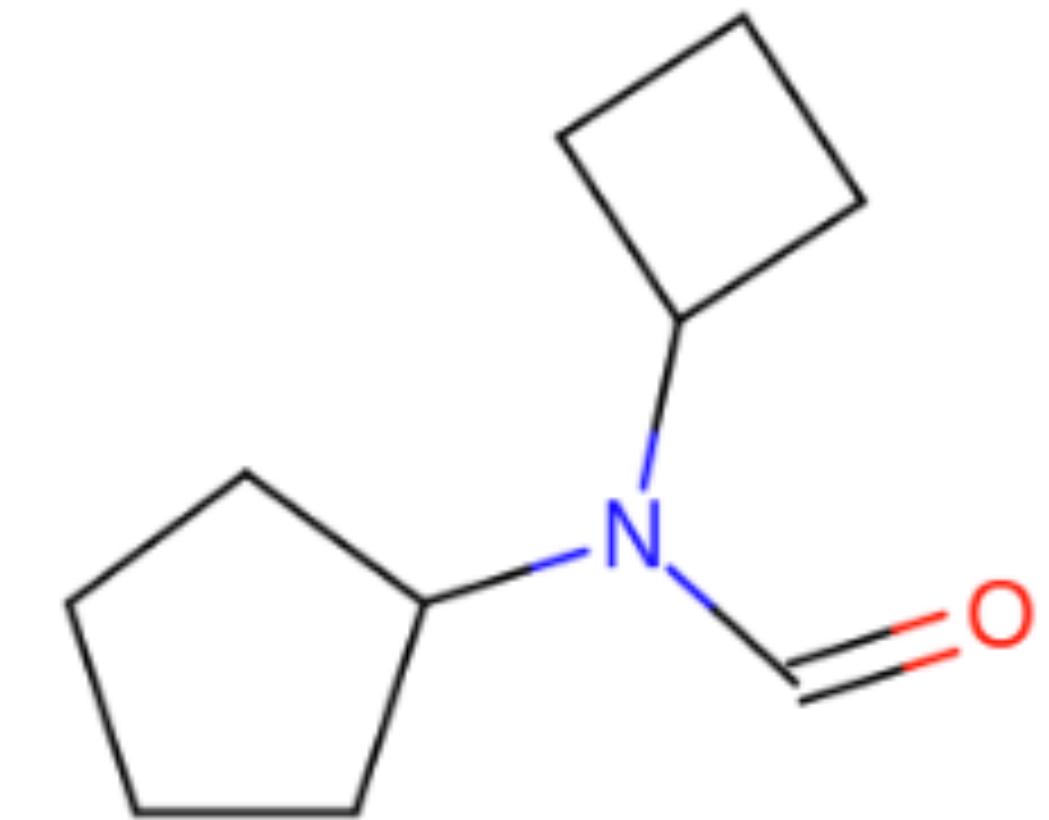
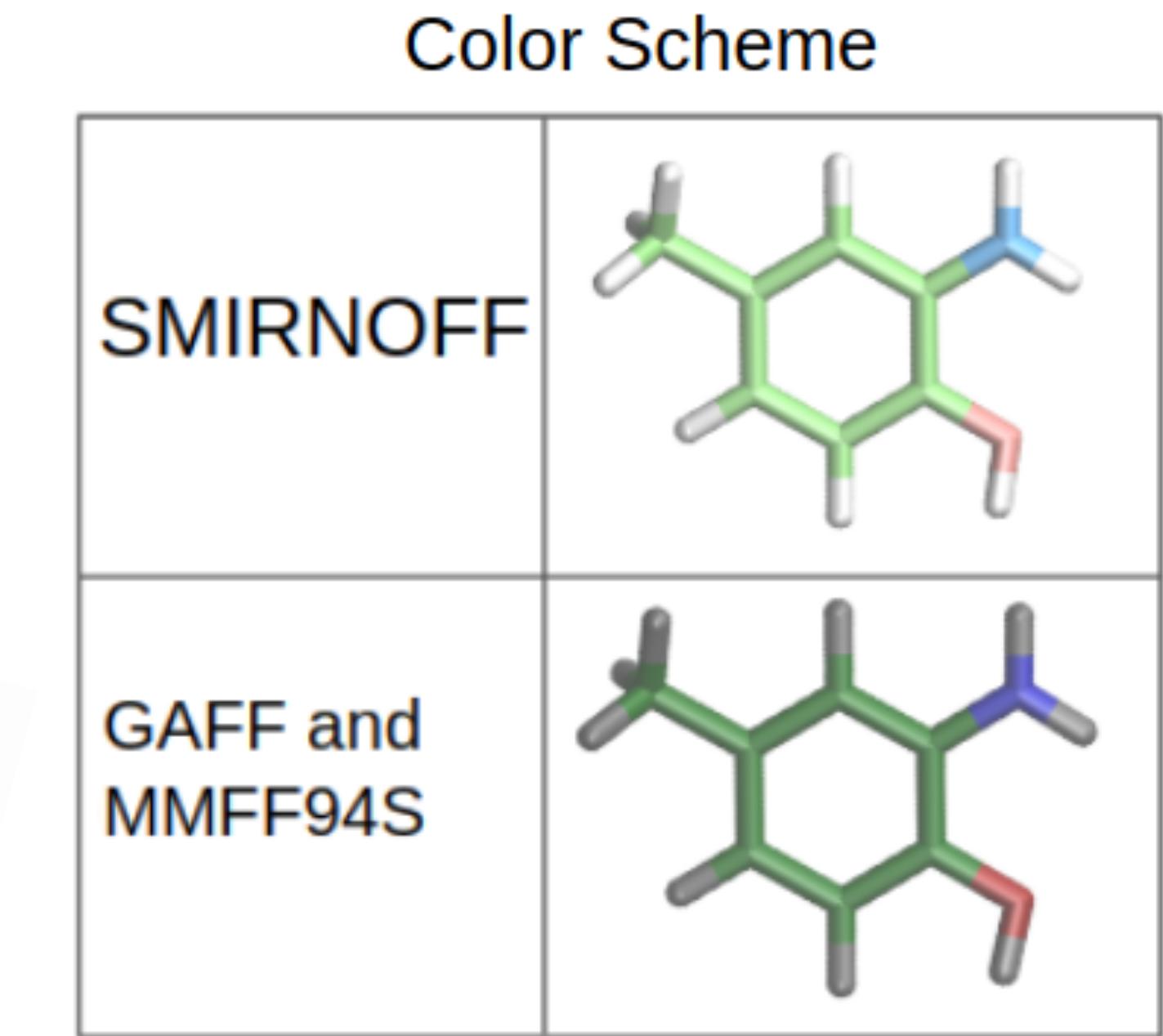
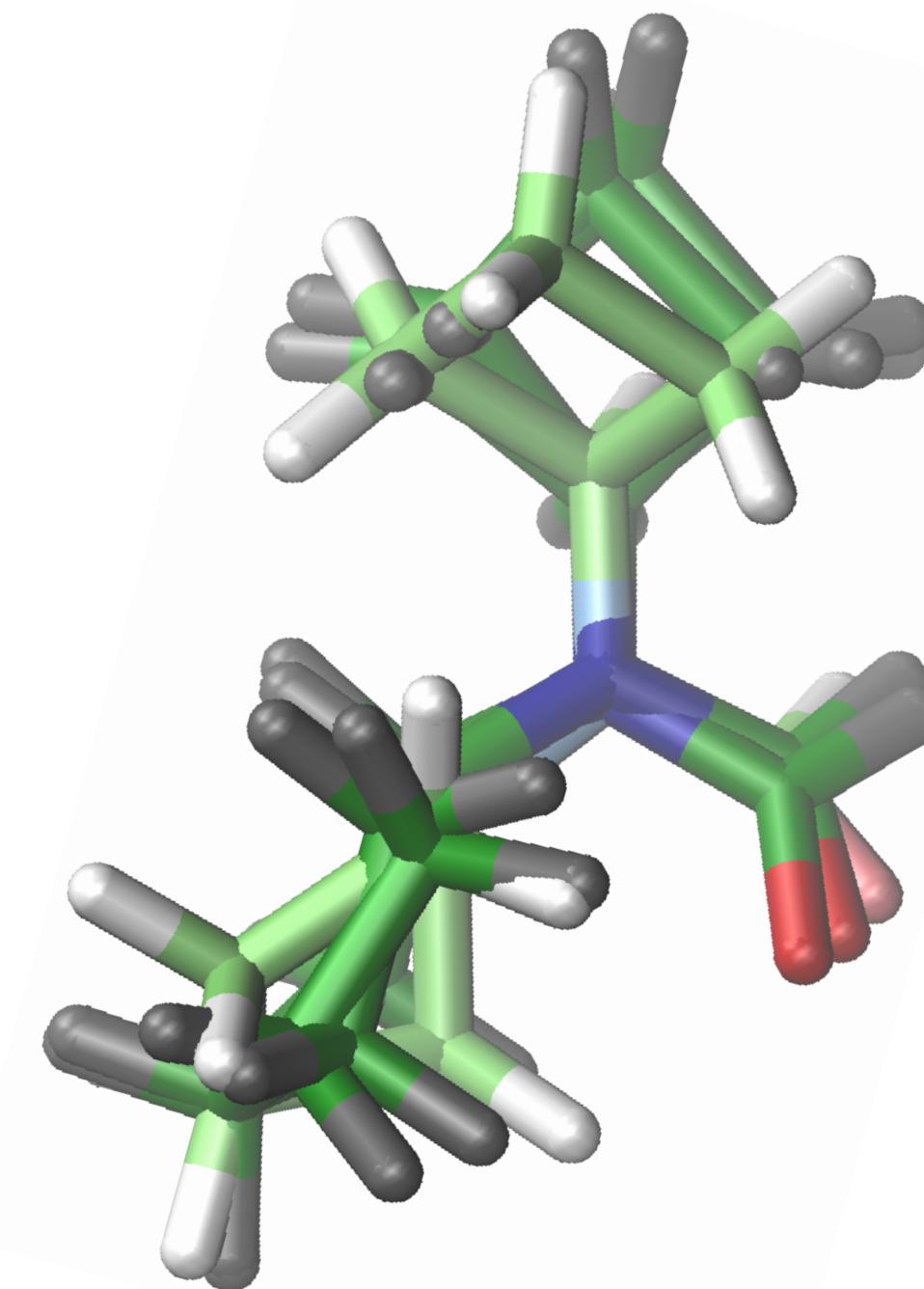
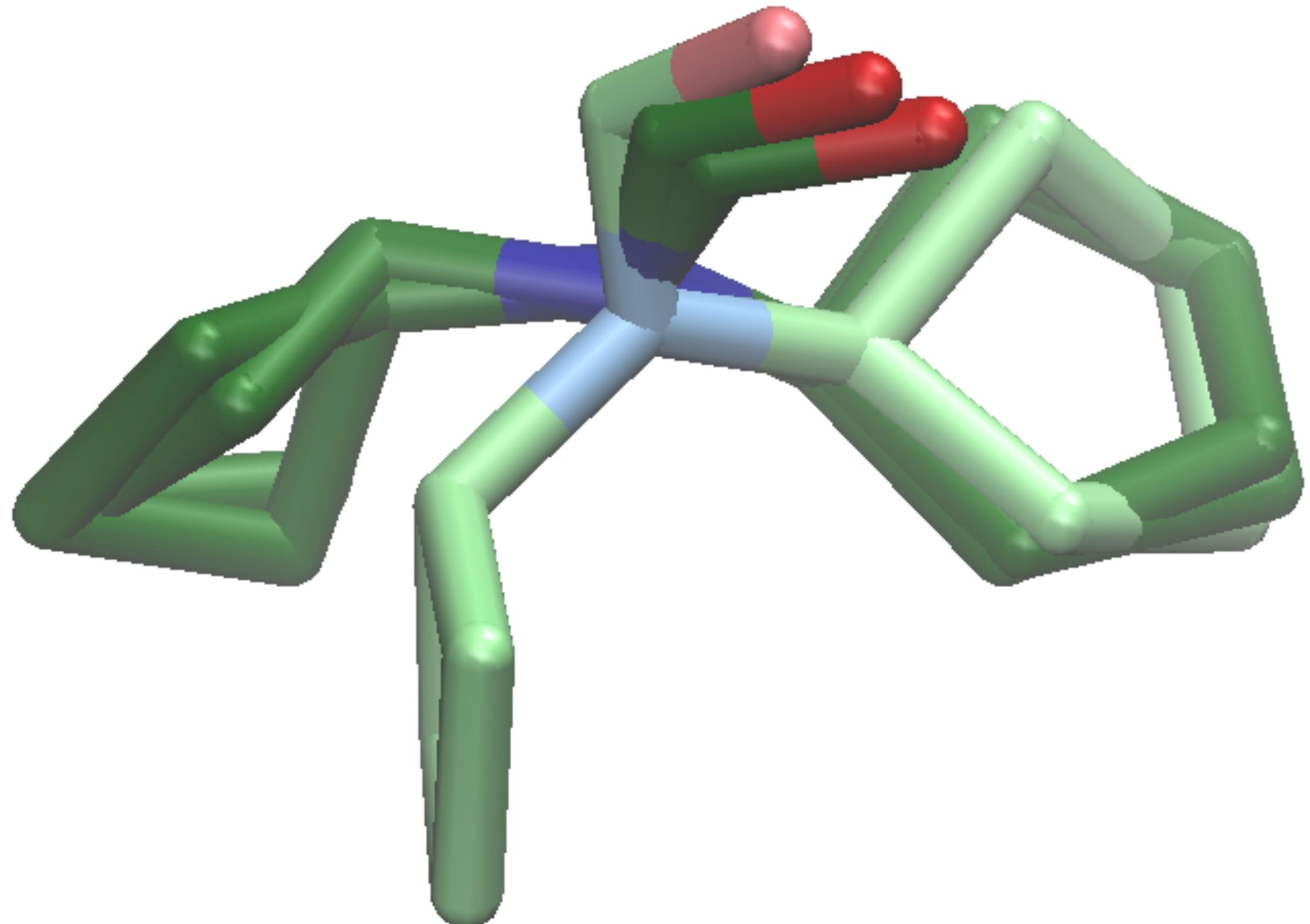
EXPERIMENTAL DATA

- QM electrostatic potentials near molecular surface
- QM equilibrium geometries and force constant matrices (Hessians)
- QM single-point energies for 1- and 2-torsion drives
- C6 dispersion coefficients
- statistic atomic and molecular polarizabilities

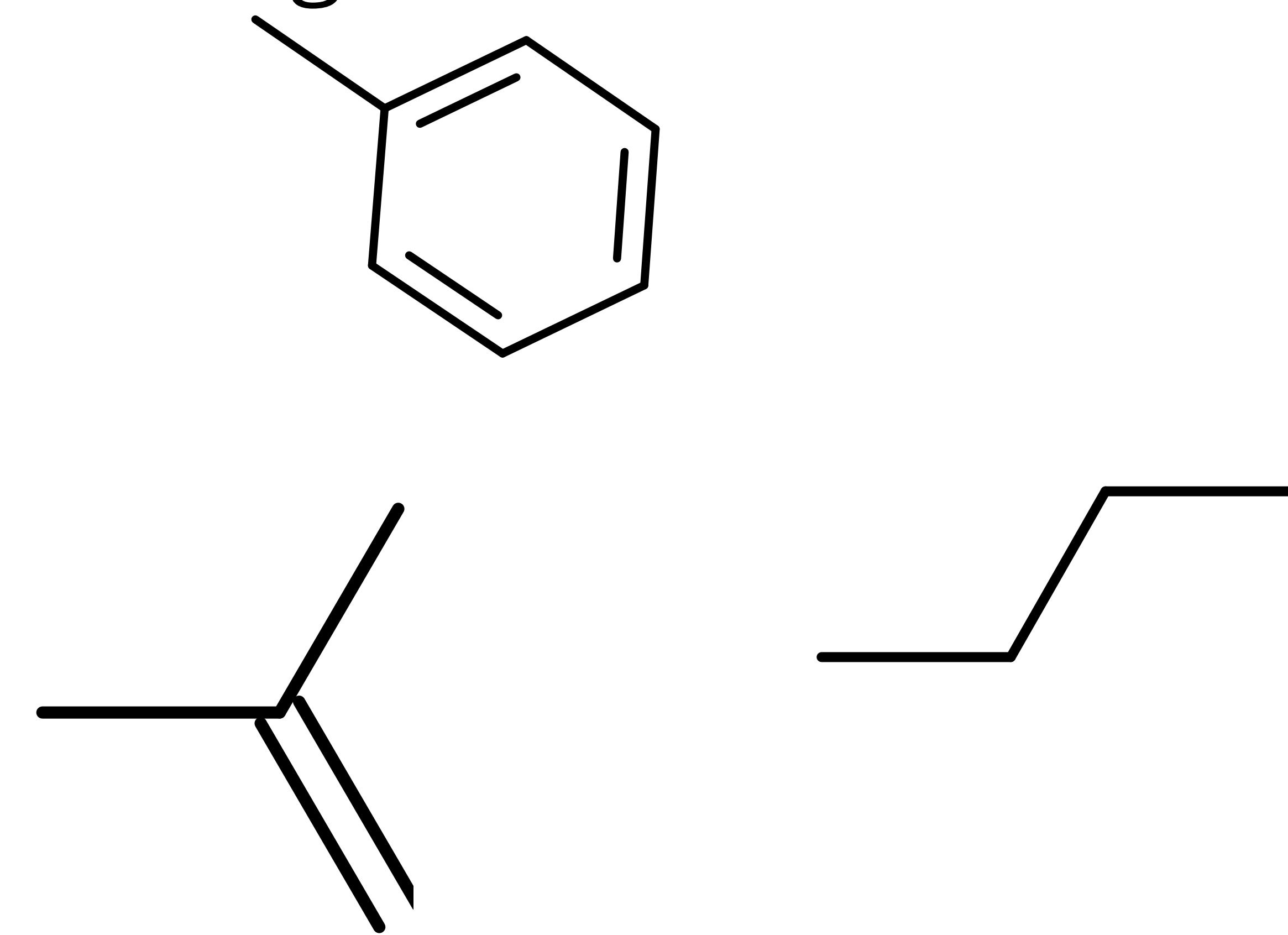
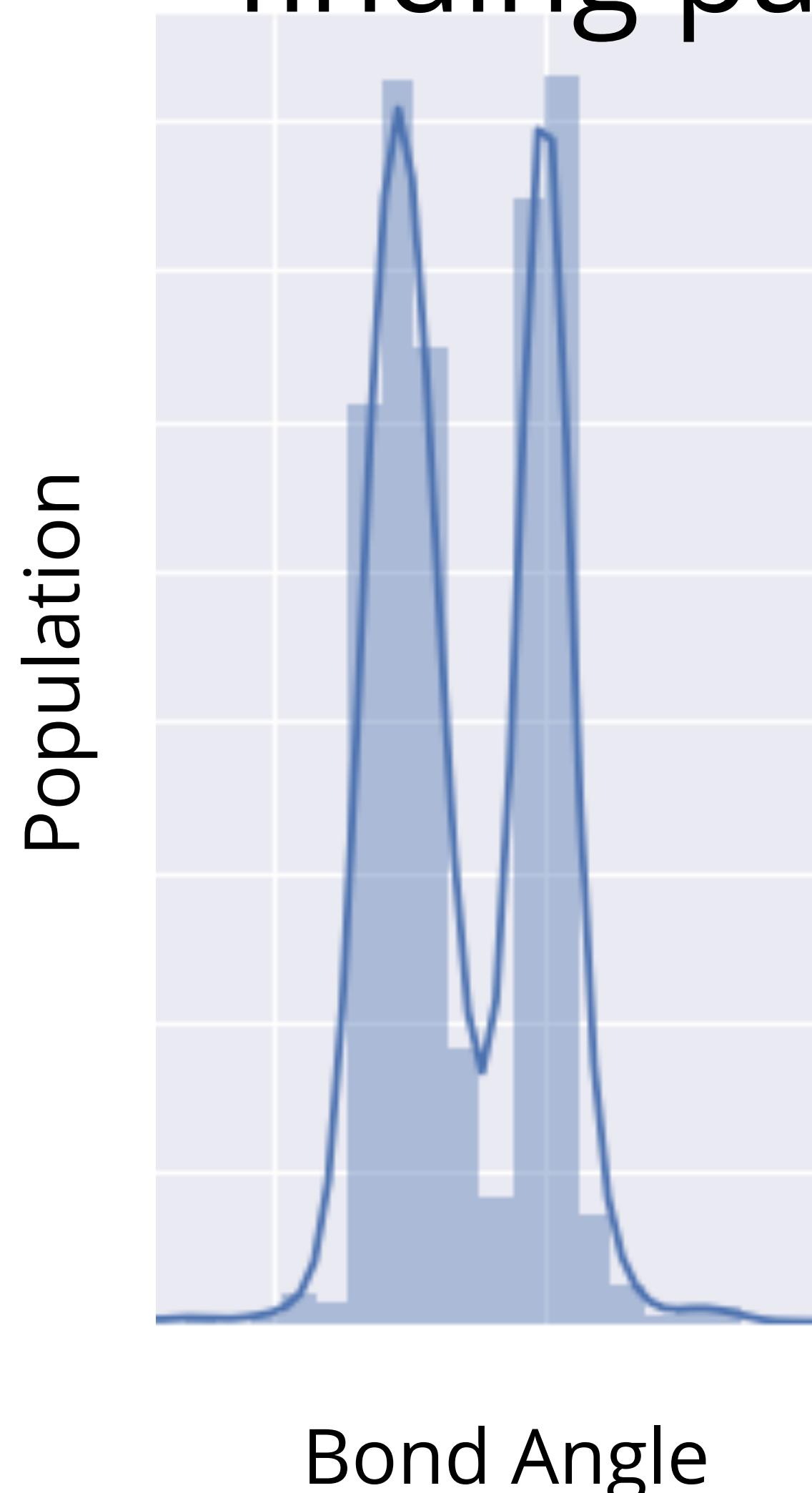
QM DATA

- primarily valence terms
- primarily Lennard-Jones
- primarily electrostatics

We're doing a vast comparison of current FFs to help prioritize chemistry for refinement



We want to automatically learn which chemistry is important in a data-driven manner. Example: Consider finding parameters for angles around carbon

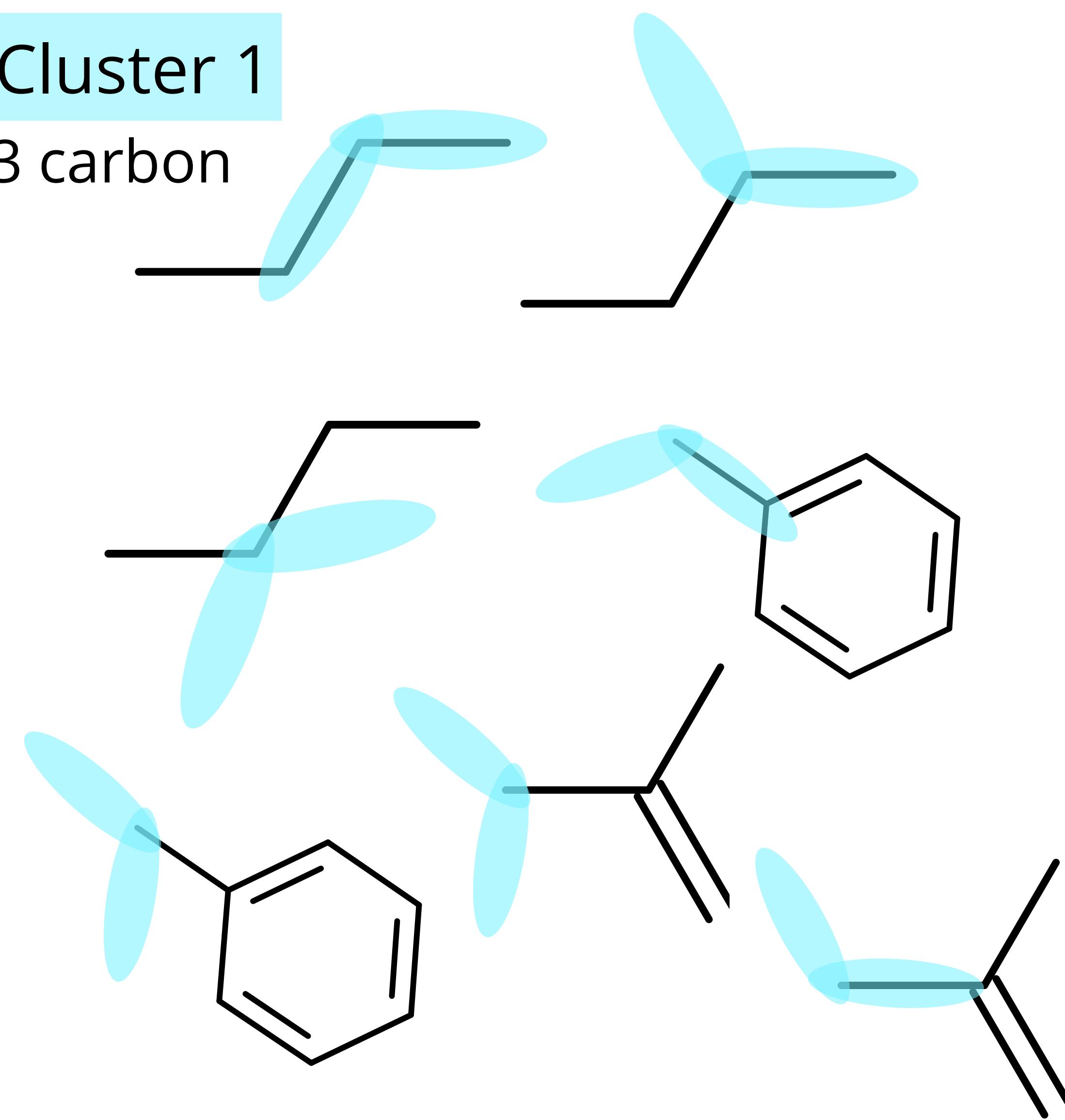


For the purpose of this exercise,
lets assume there are only 2 angle types

Step 1: cluster chemistry by which parameter it should have

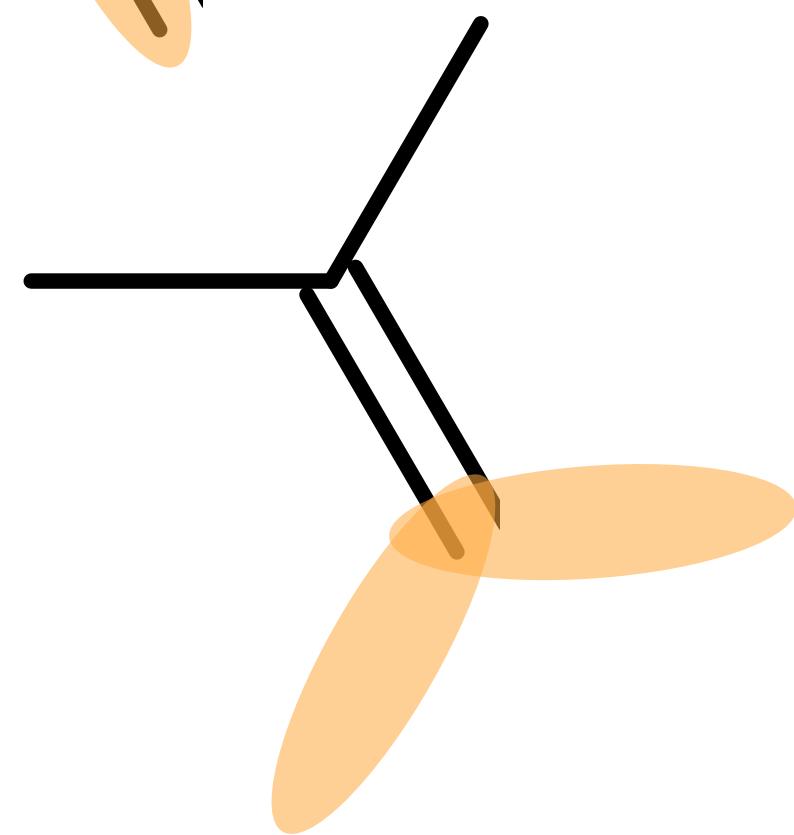
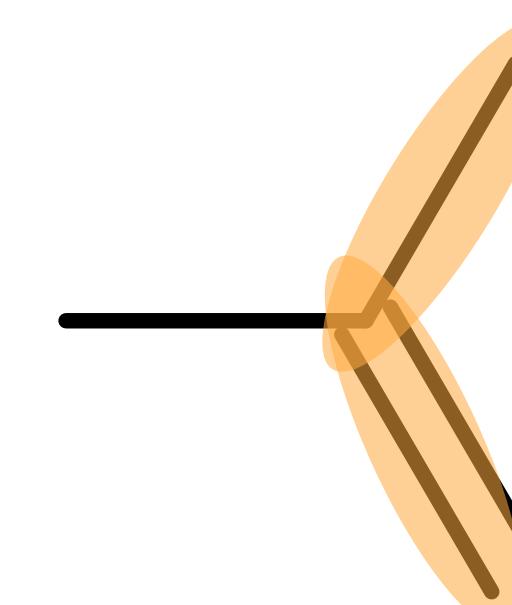
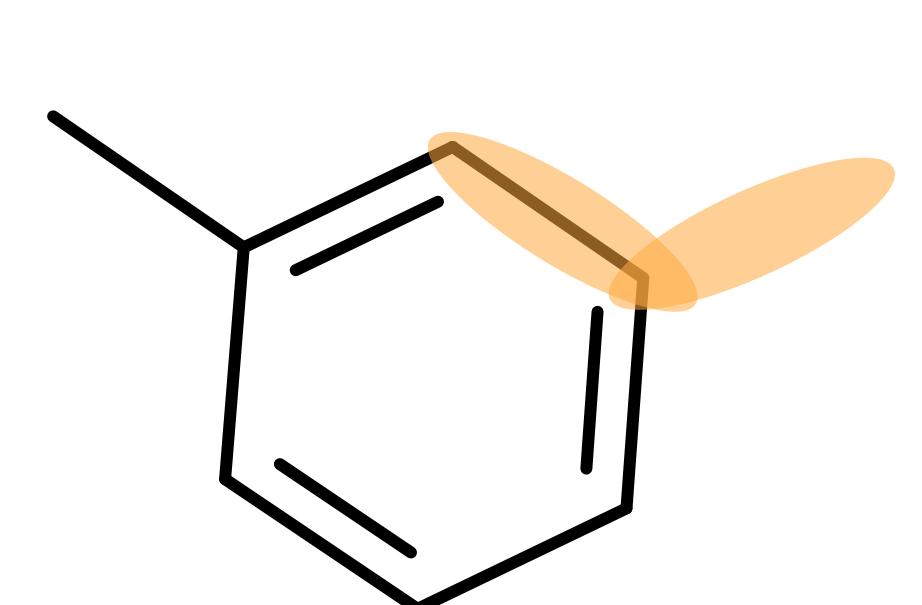
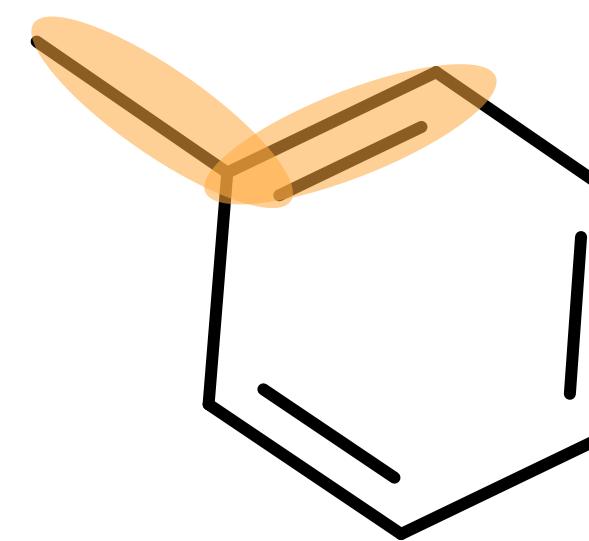
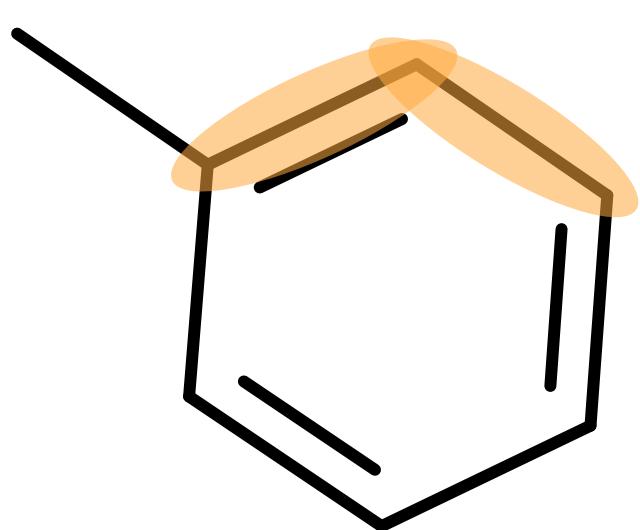
Cluster 1

Sp₃ carbon



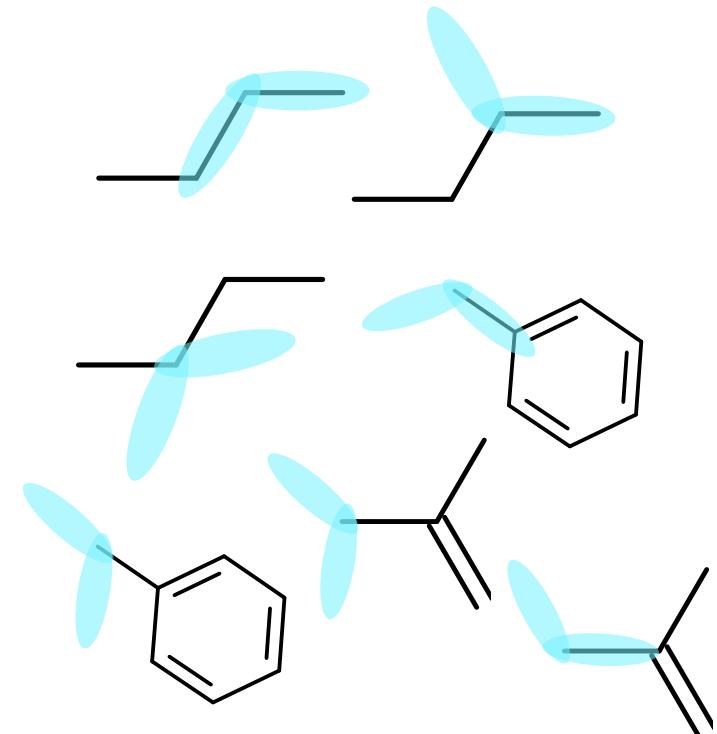
Cluster 2

Sp₂ carbon



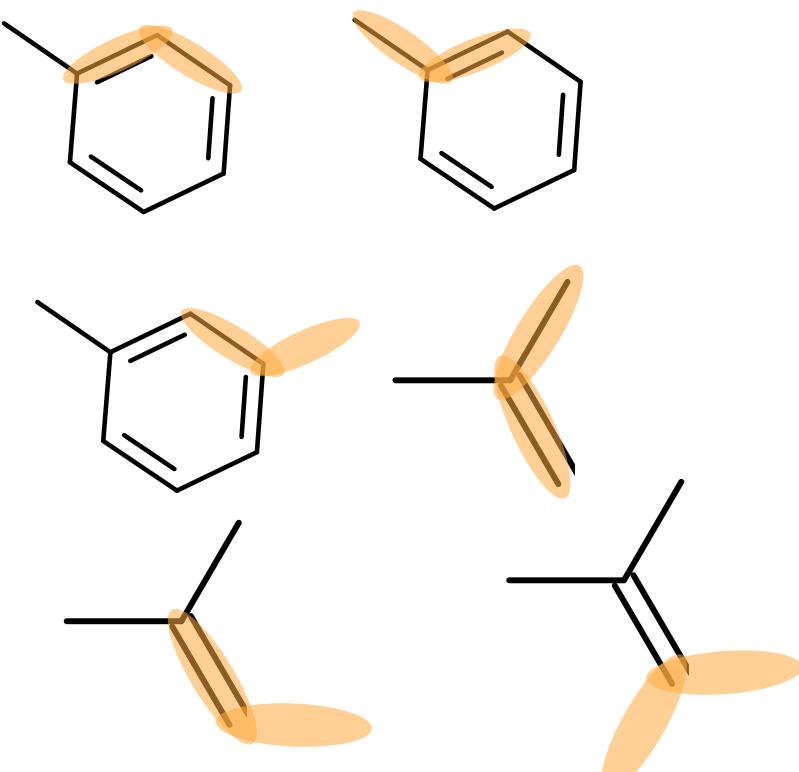
Step 2: Extract all possible SMIRKS patterns for each cluster

Cluster 1



[#1AH0X1x0!r+0,#6AH3X4x0!
r+0,#6aH0X3x2r6+0:1]-;![#6AH2X4x0!
r+0,#6AH3X4x0!r+0:2]-;![#1AH0X1x0!
r+0,#6AH0X3x0!r+0,#6AH2X4x0!r+0:3]

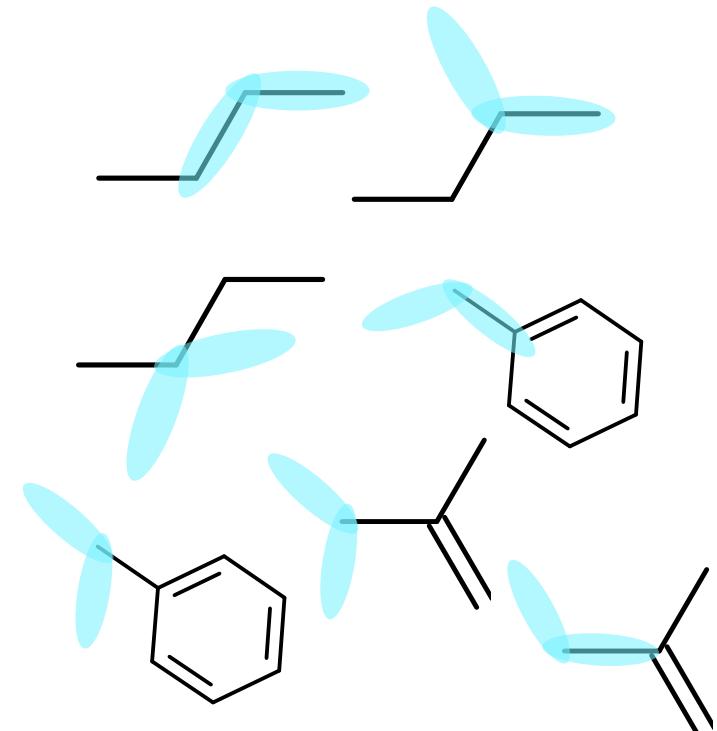
Cluster 2



[#1AH0X1x0!r+0,#6AH3X4x0!
r+0,#6aH1X3x2r6+0:1]-,:[#6AH0X3x0!
r+0,#6AH2X3x0!
r+0,#6aH0X3x2r6+0,#6aH1X3x2r6+0:2]-,:=[#1AH0
X1x0!r+0,#6AH0X3x0!r+0,#6AH2X3x0!
r+0,#6AH3X4x0!r+0,#6aH1X3x2r6+0:3]

Step 2: Extract all possible SMIRKS patterns for each cluster

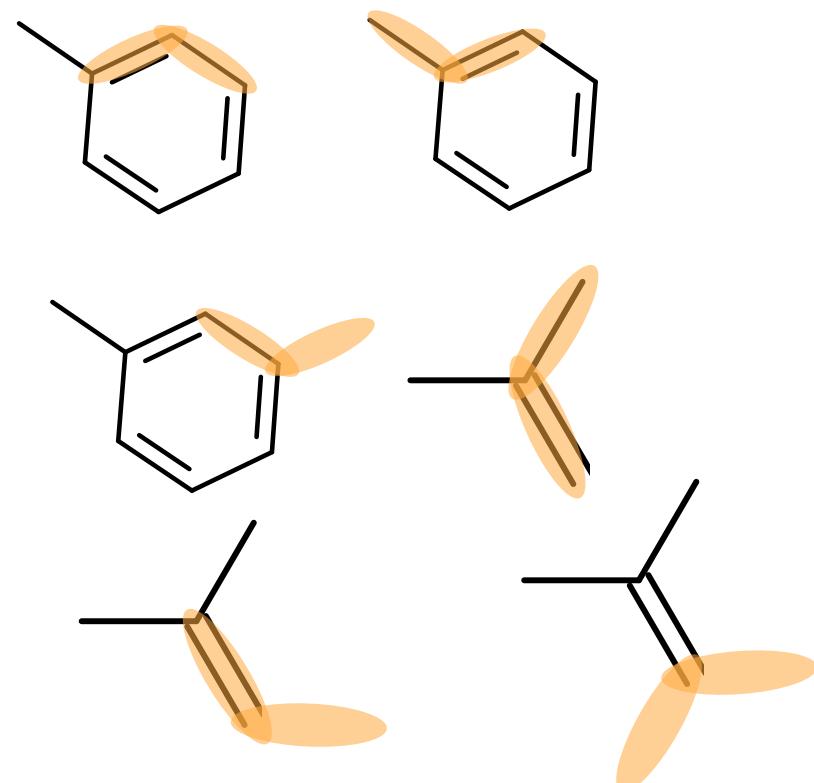
Cluster 1



[#1AH0X1x0!r+0,#6AH3X4x0!
r+0,#6aH0X3x2r6+0:1]-;![#6AH2X4x0!
r+0,#6AH3X4x0!r+0:2]-;![#1AH0X1x0!
r+0,#6AH0X3x0!r+0,#6AH2X4x0!r+0:3]

Warning:
do not try to read SMIRKS!

Cluster 2



[#1AH0X1x0!r+0,#6AH3X4x0!
r+0,#6aH1X3x2r6+0:1]-,:[#6AH0X3x0!
r+0,#6AH2X3x0!
r+0,#6aH0X3x2r6+0,#6aH1X3x2r6+0:2]-,:=[#1AH0
X1x0!r+0,#6AH0X3x0!r+0,#6AH2X3x0!
r+0,#6AH3X4x0!r+0,#6aH1X3x2r6+0:3]

Step 3: Systematically reduce SMIRKS to only the essentials

Cluster 1 $[:1]-[\#6\times 4]-[:3]$

Cluster 2 $[:1]\sim[\#6\times 3]\sim[:3]$

We would then fit parameters for each of these patterns

Step 3: Systematically reduce SMIRKS to only the essentials

Cluster 1 $[:1]-[\#6X4]-[:3]$

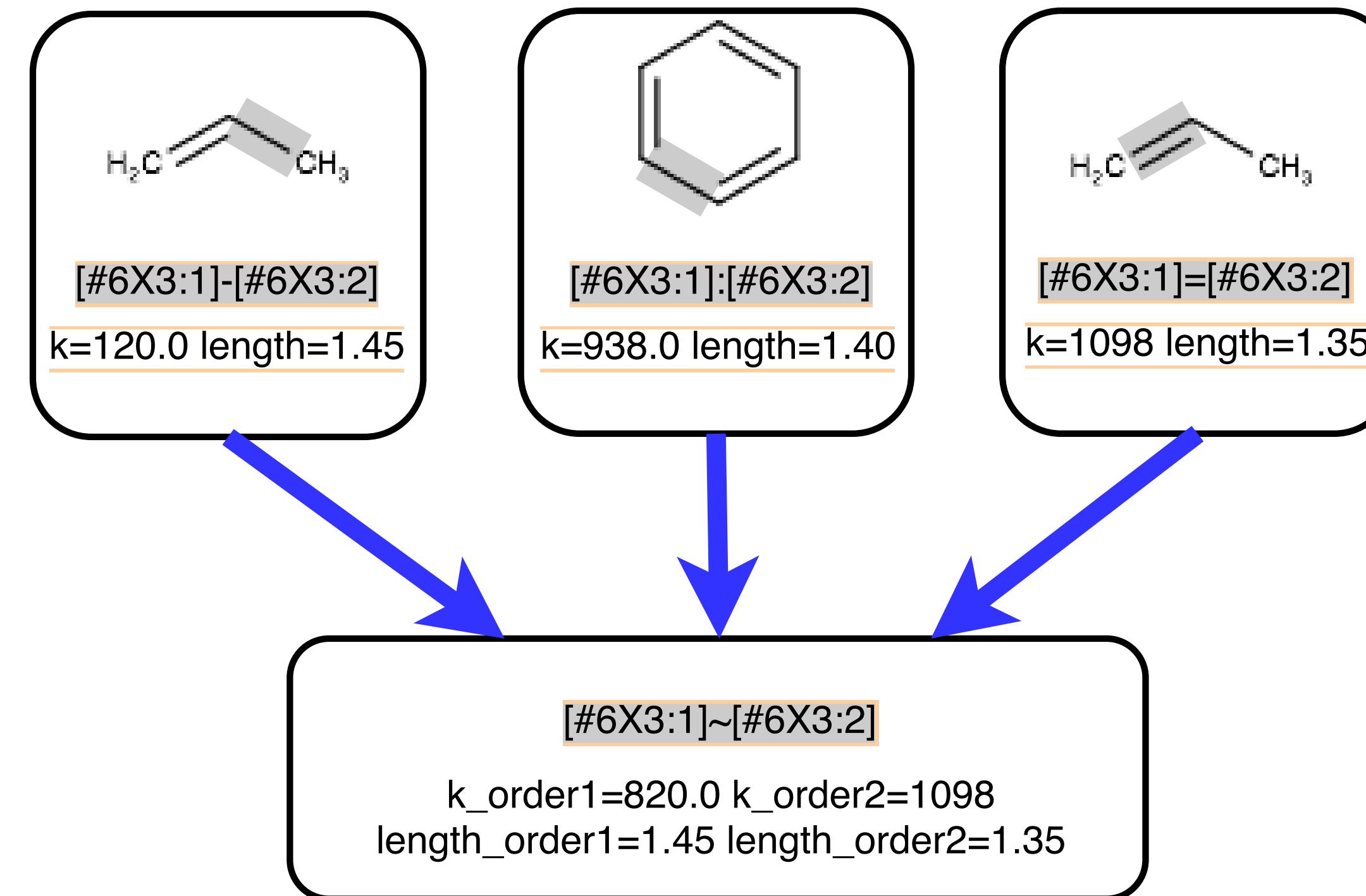
Readable, but more importantly,
they are as generic as possible

Cluster 2 $[:1]\sim[\#6X3]\sim[:3]$

We would then fit parameters for each of these patterns

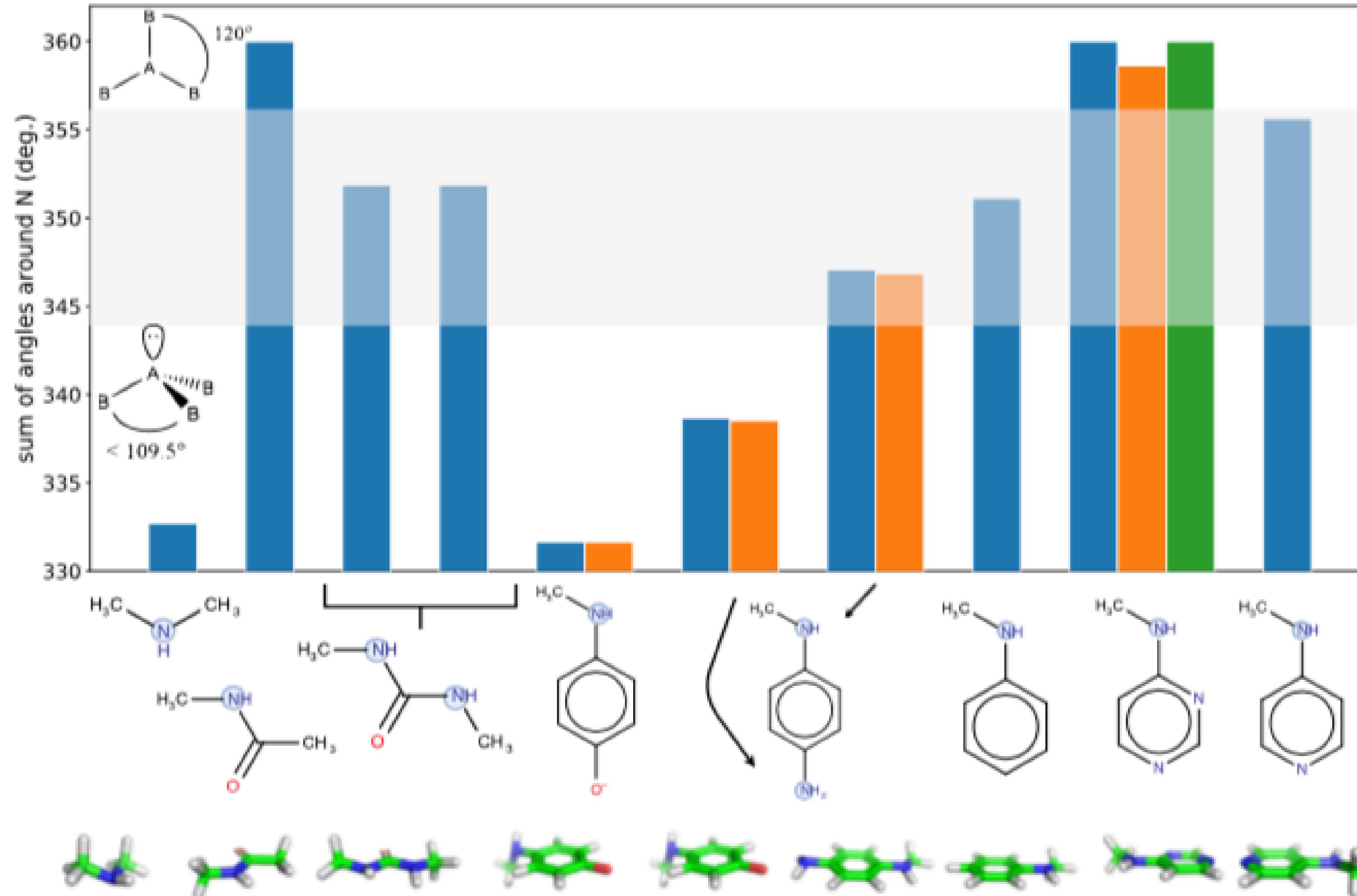
Some aspects are longer-term science
outside the scope of the
industry-funded consortium

We simplify further by building in more chemistry, like partial bond orders



We can use partial bond orders (Wiberg bond orders) to get bond orders specific to each molecule for interpolation

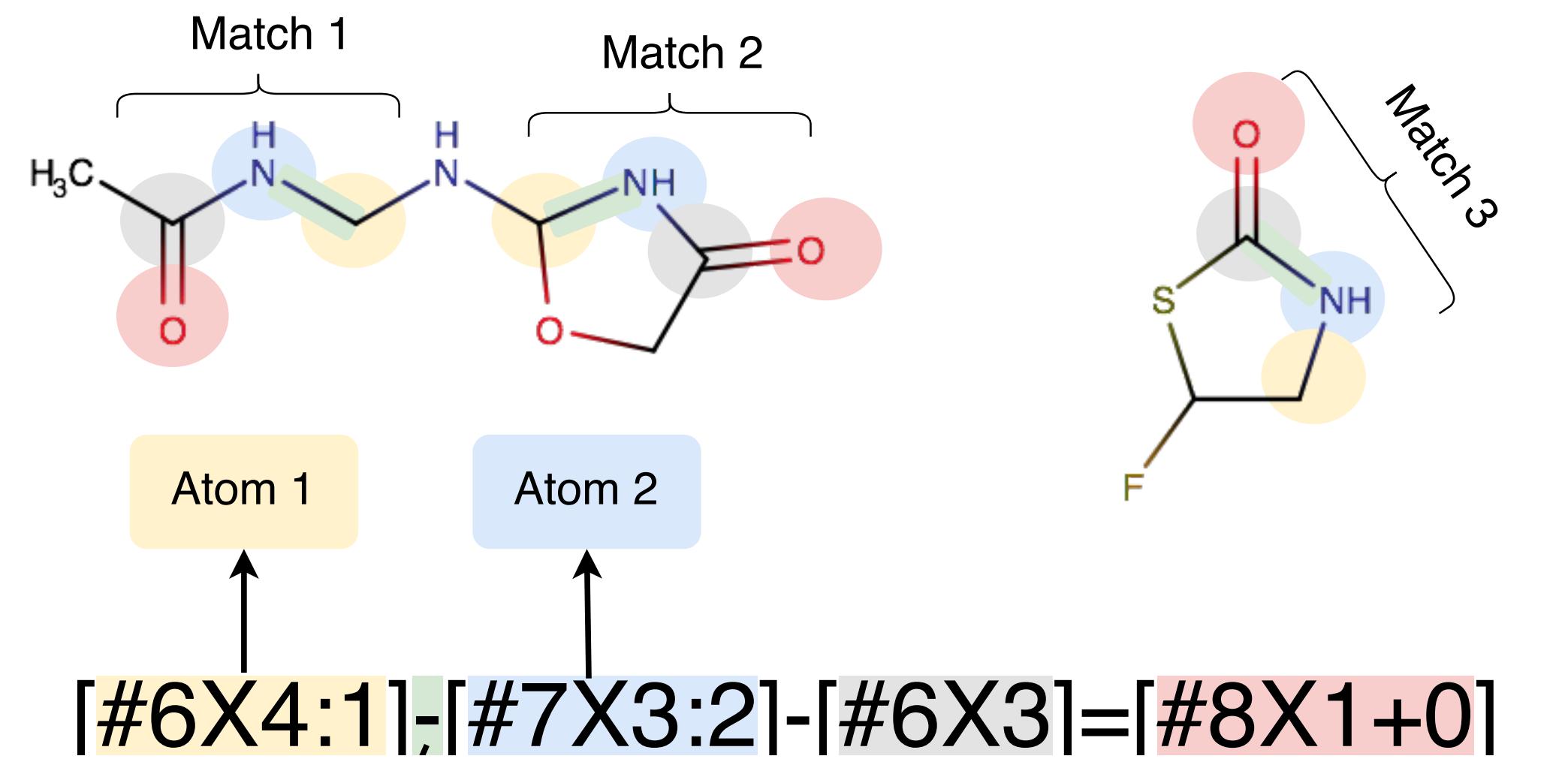
This opens up interesting avenues for new force field science, such as partial planarity



- We can interpolate based on partial bond order around the nitrogen center
- Connecting this to impropers should allow partially planar nitrogens

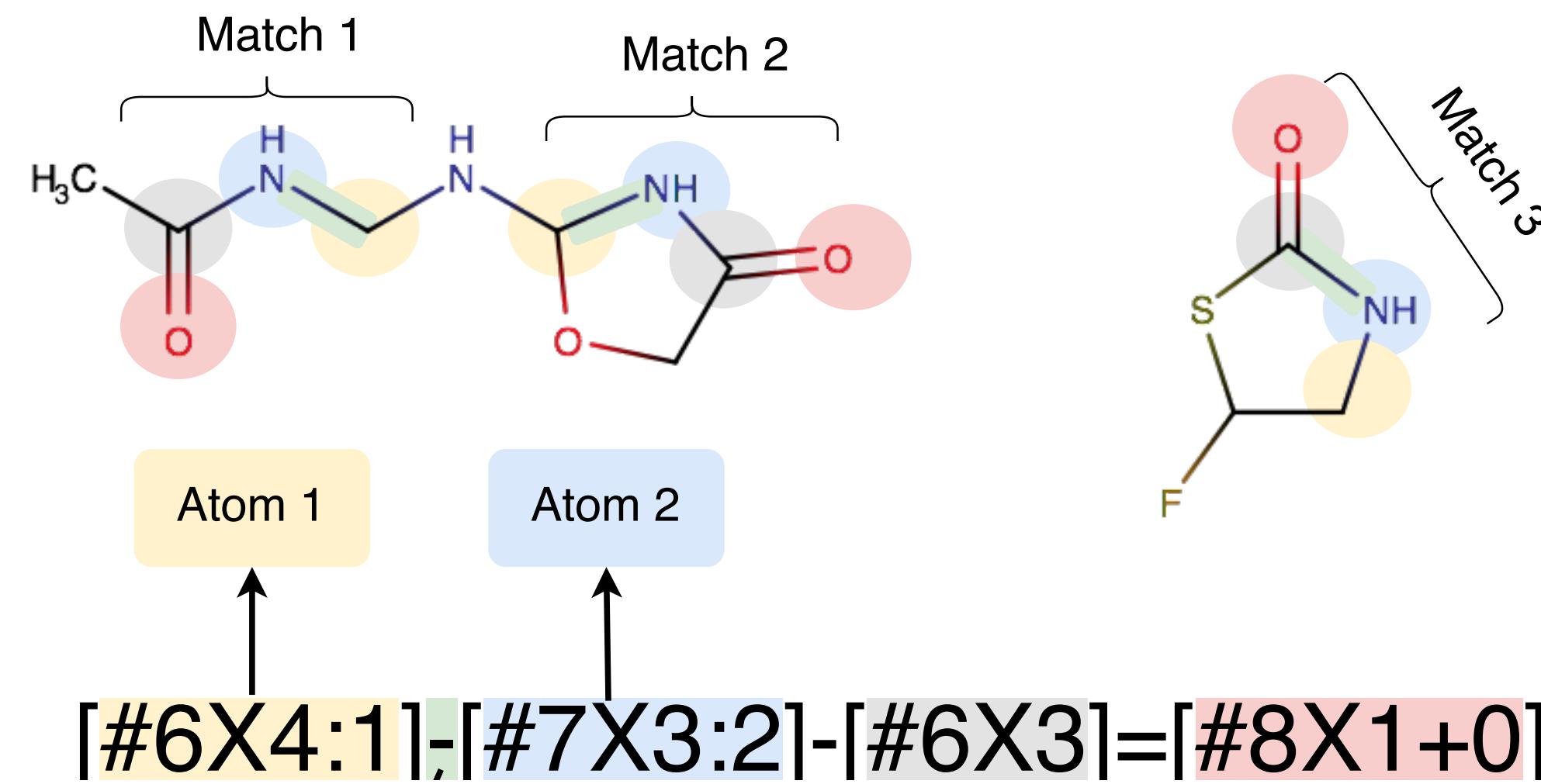
**You can use SMIRNOFF now, but it's an exciting new
avenue for force fields**

You can use SMIRNOFF now, but it's an exciting new avenue for force fields



SMIRKS provide a language for
direct chemical perception

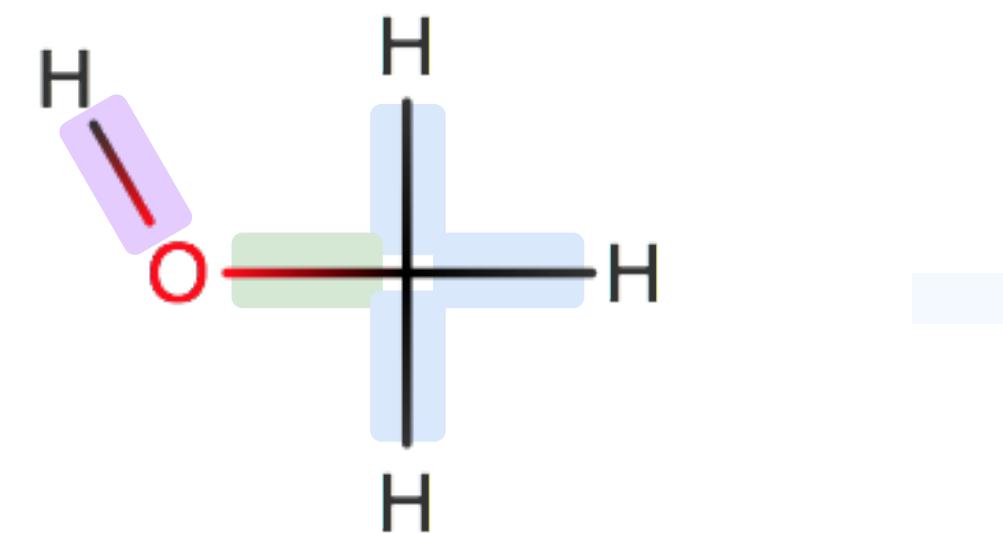
You can use SMIRNOFF now, but it's an exciting new avenue for force fields



SMIRKS provide a language for direct chemical perception

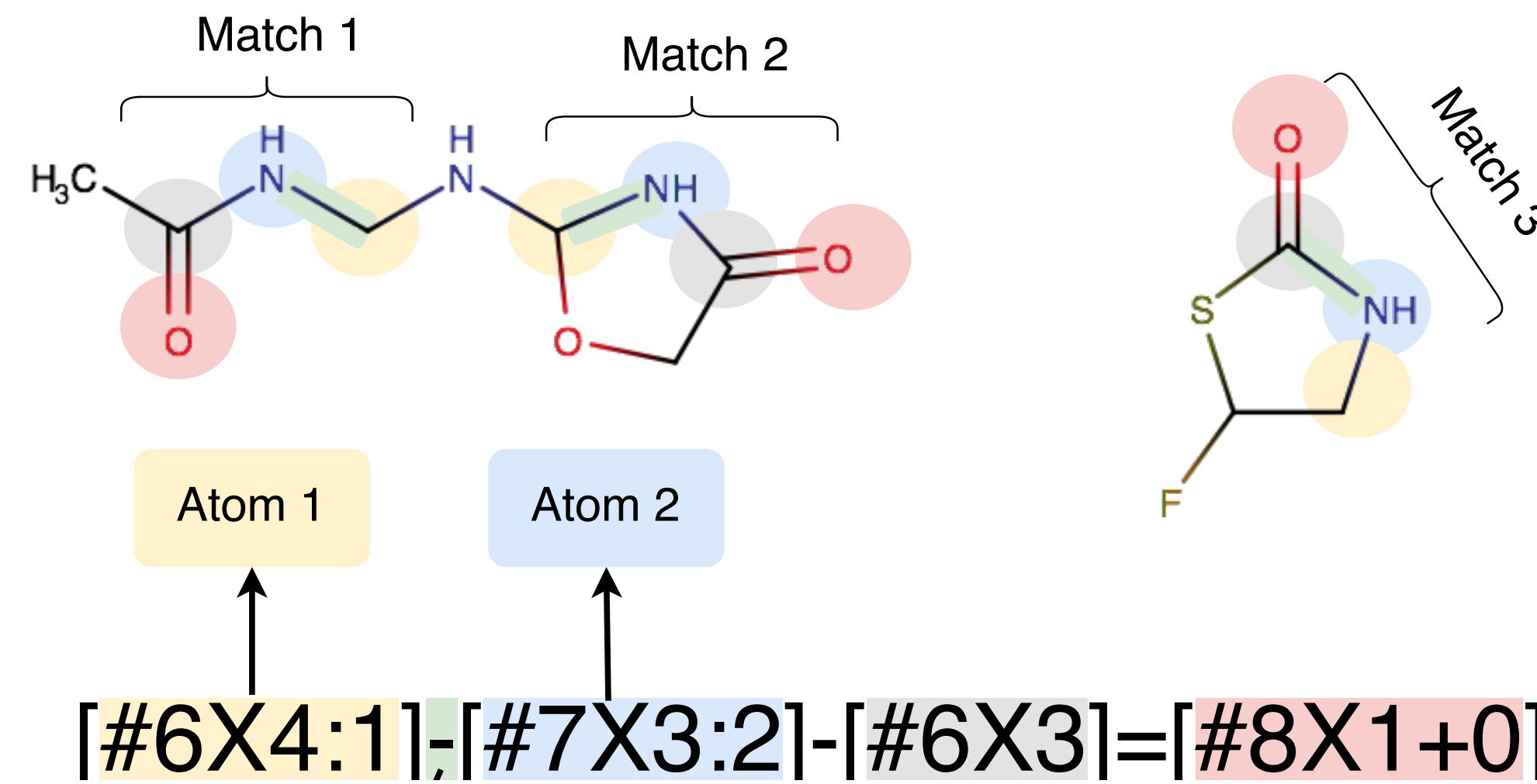
```
<?xml version="1.0"?>
<SMIRNOFF>

<HarmonicBondForce length_unit="angstroms" k_unit="kilocalories_per_mole/angstrom**2">
    <Bond smirks="[#6X4:1]-[#1:2]" length="1.090" k="680.0"/>
    <Bond smirks="[#6X4:1]-[#8&amp;X2&amp;H1:2]" length="1.410" k="640.0"/>
    <Bond smirks="[#8X2:1]-[#1:2]" length="0.960" k="1106.0"/>
</HarmonicBondForce>
```

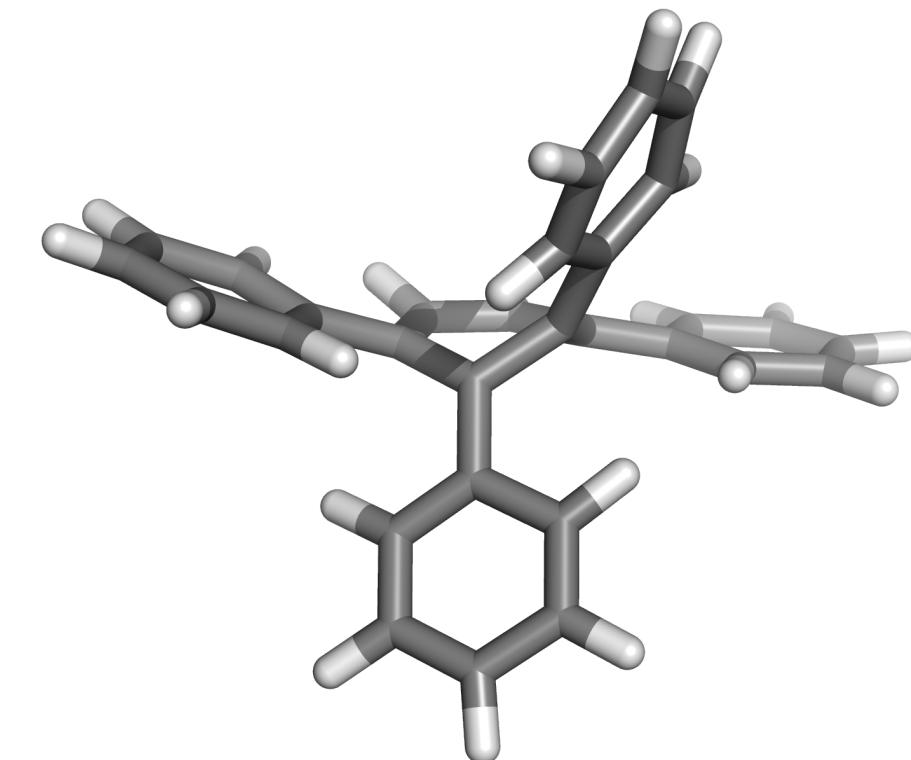


Thus we have a new force field format

You can use SMIRNOFF now, but it's an exciting new avenue for force fields



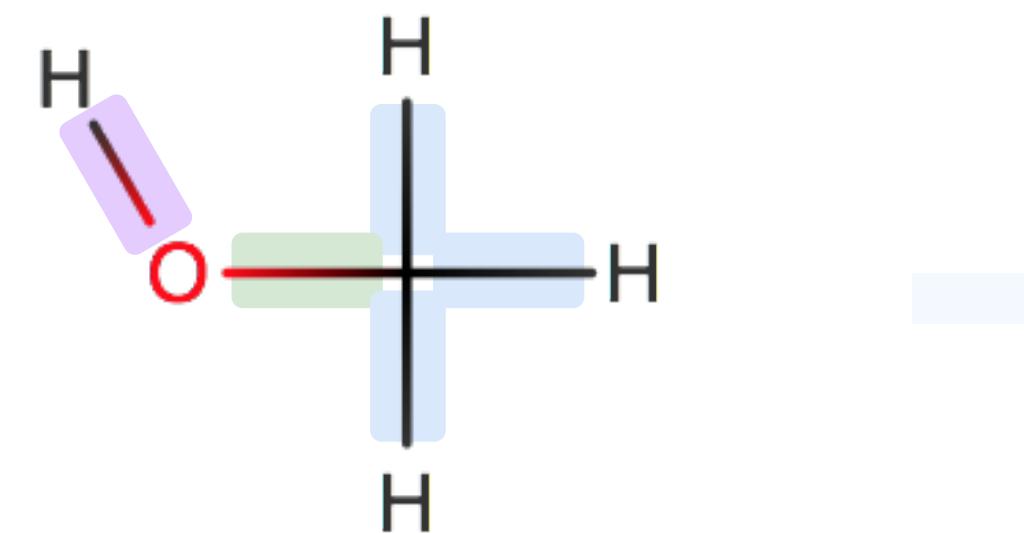
SMIRKS provide a language for direct chemical perception



And a new FF which fixes some problems with GAFF

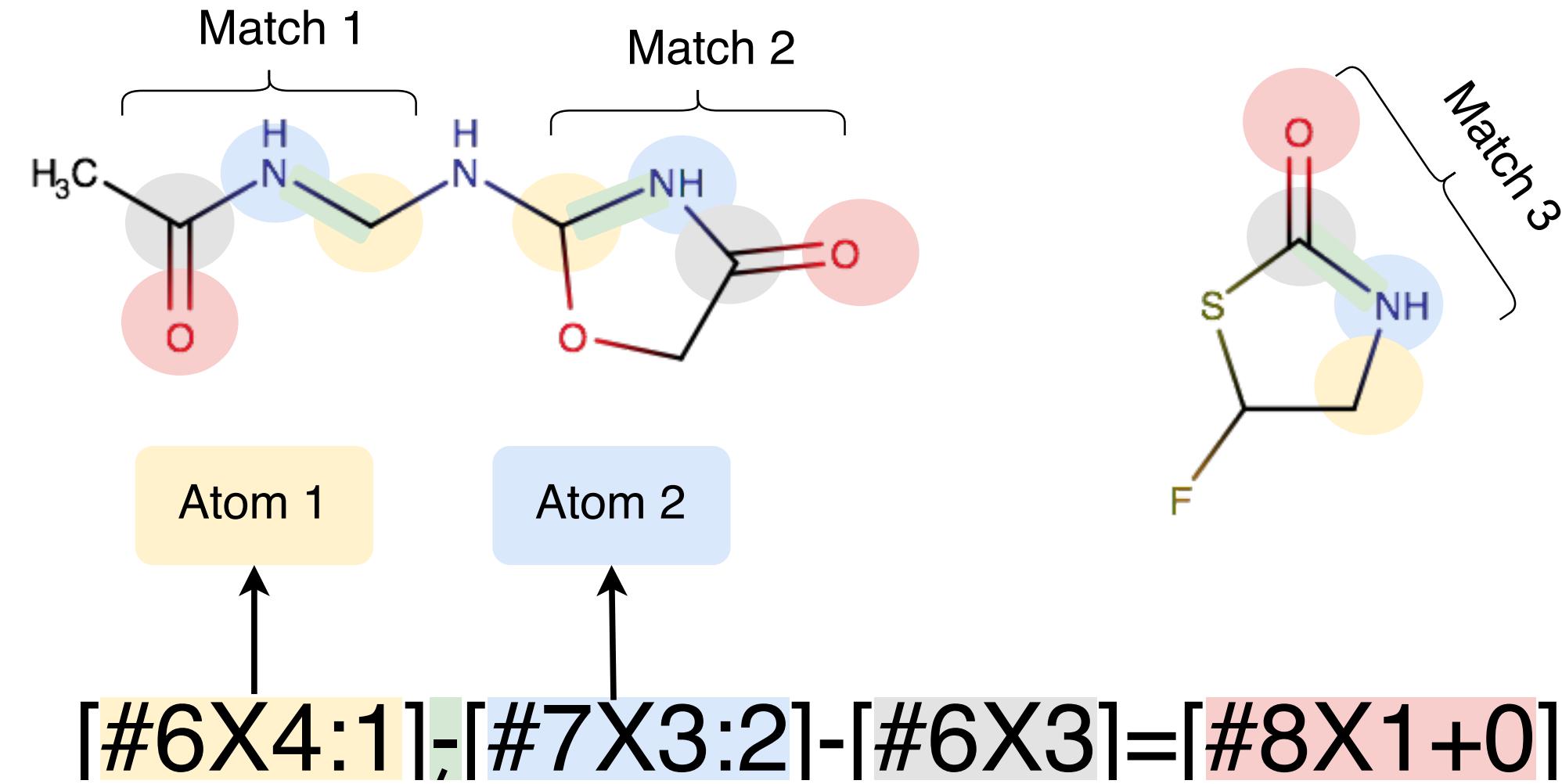
```
<?xml version="1.0"?>
<SMIRNOFF>

<HarmonicBondForce length_unit="angstroms" k_unit="kilocalories_per_mole/angstrom**2">
    <Bond smirks="#6X4:1-[#1:2]" length="1.090" k="680.0"/>
    <Bond smirks="#6X4:1-[#8&amp;X2&amp;H1:2]" length="1.410" k="640.0"/>
    <Bond smirks="#8X2:1-[#1:2]" length="0.960" k="1106.0"/>
</HarmonicBondForce>
```



Thus we have a new force field format

You can use SMIRNOFF now, but it's an exciting new avenue for force fields

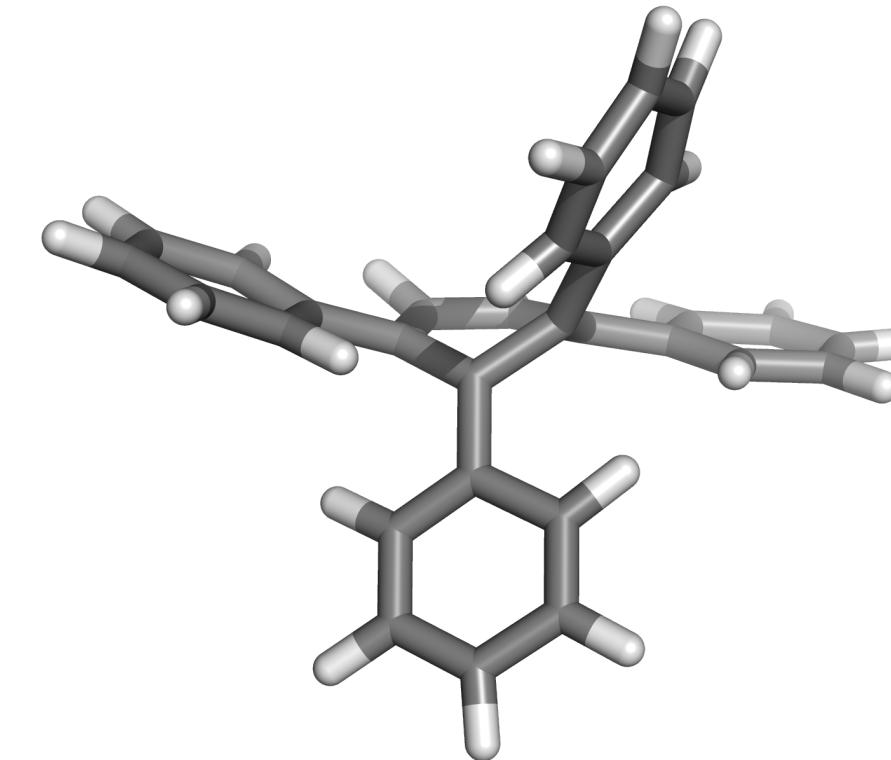


SMIRKS provide a language for direct chemical perception

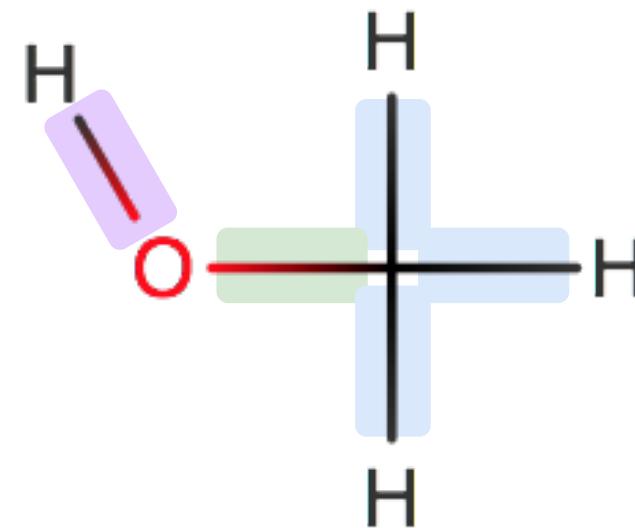
```
<?xml version="1.0"?>
<SMIRNOFF>

<HarmonicBondForce length_unit="angstroms" k_unit="kilocalories_per_mole/angstrom**2">
  <Bond smirks="#6X4:1-[#1:2]" length="1.090" k="680.0"/>
  <Bond smirks="#6X4:1-[#8&amp;X2&amp;H1:2]" length="1.410" k="640.0"/>
  <Bond smirks="#8X2:1-[#1:2]" length="0.960" k="1106.0"/>
</HarmonicBondForce>
```

Thus we have a new force field format



And a new FF which fixes some problems with GAFF



But much bigger plans to improve open FFs and FF infrastructure with help from the community

Acknowledgments

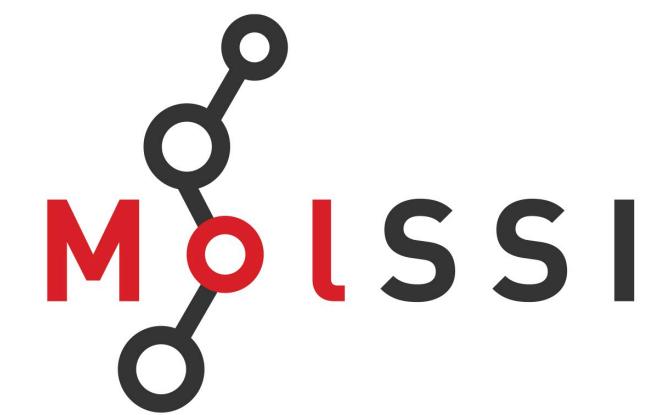
The Open Force Field Initiative and everyone involved

Data work especially from: Caitlin C. Bannan, Camila Zanette, Christopher I. Bayly, Victoria Lim, Jessica Maat, Jordan Ehrman, Josh Fass, Andrea Rizzi, Kyle A. Beauchamp, David R. Slochower, Lee-Ping Wang, Michael K. Gilson, Michael R. Shirts, John D. Chodera, Bryce Manubay, David L. Mobley, Owen Madin

OpenEye for supporting this by sending Christopher Bayly to UCI on sabbatical!

openforcefield.org

github.com/openforcefield/openforcefield



Acknowledgments

The Open Force Field Initiative and everyone involved

