open
forcefield

@openforcefield

www.openforcefield.org

5th open force field workshop

June 28, 2022 | Virtual meeting

Speakers

David Mobley
Jeff Wagner
Lily Wang
Diego Nolasco

Feedback?

Feel free to use the Chat and Q&A features in Zoom to ask questions during the talks. Speakers will answer questions in the Q&A after they're done with their presentations.

We will host a brief open discussion after the conclusion of talks. Use the "raise hand" feature and a host will enable your mic when it's your turn.

We're here because of you – and because we share a common set of tractable problems

INDUSTRY

AbbVie

BASF

Bayer

Boehringer-Ingelheim

Bristol Myers Squibb

Cresset

Janssen

GlaxoSmithKline

Merck KGaA

OpenEye

Pfizer

Roche

Vertex

Eli Lilly

... and others

Open Molecular Software Foundation

PROJECT STAFF

Jeff Wagner
Technical Lead

Lily Wang
Science Lead

...

ACADEMIC

John Chodera (MSKCC)

Michael Gilson (UC San Diego)

David Mobley (UC Irvine)

Michael Shirts (CU Boulder)

Plus affiliates:

- **Danny Cole** (Newcastle)
- **Lee-Ping Wang** (UCD)
- **Dennis Della Corte** (BYU)
- **MolSSI** (Virginia Tech)

Openness is central, because our problems are too big to tackle alone or in small groups

Open source Python Toolkit: modern infrastructure for building/using force fields

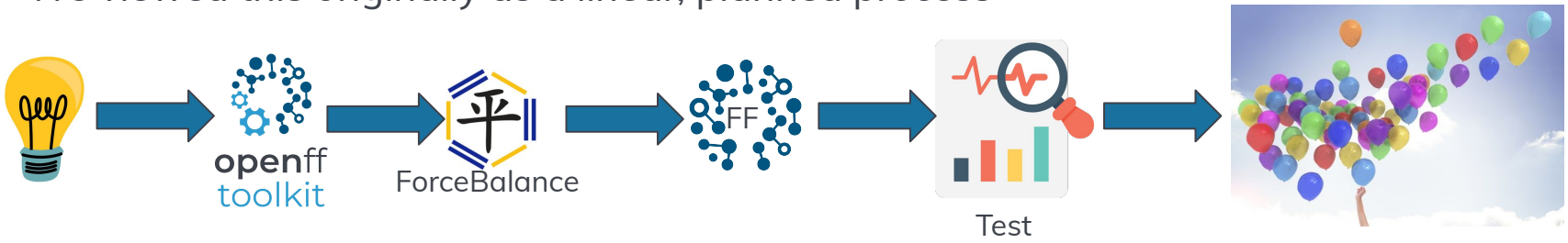
Open curated QM / physical property/binding datasets: data to build on

Open infrastructure: Extend our tools; run your own benchmarks

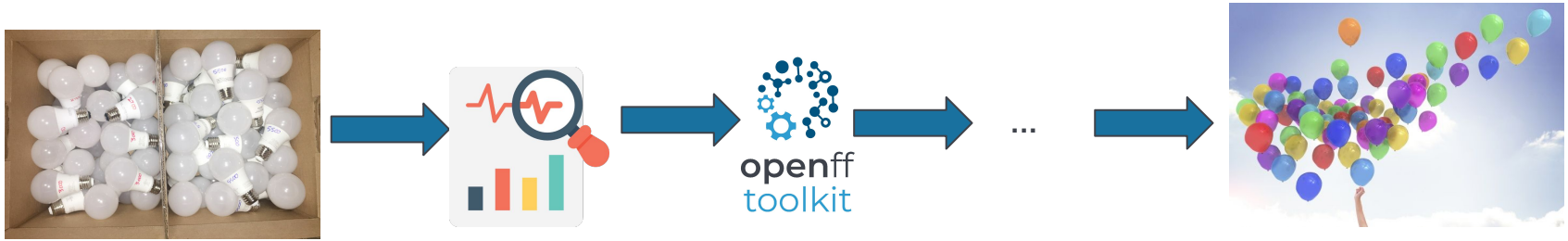
Open science: Everything done in the open; everyone can get involved

We've moved to trying a variety of ideas in parallel, then rolling the best into our force fields

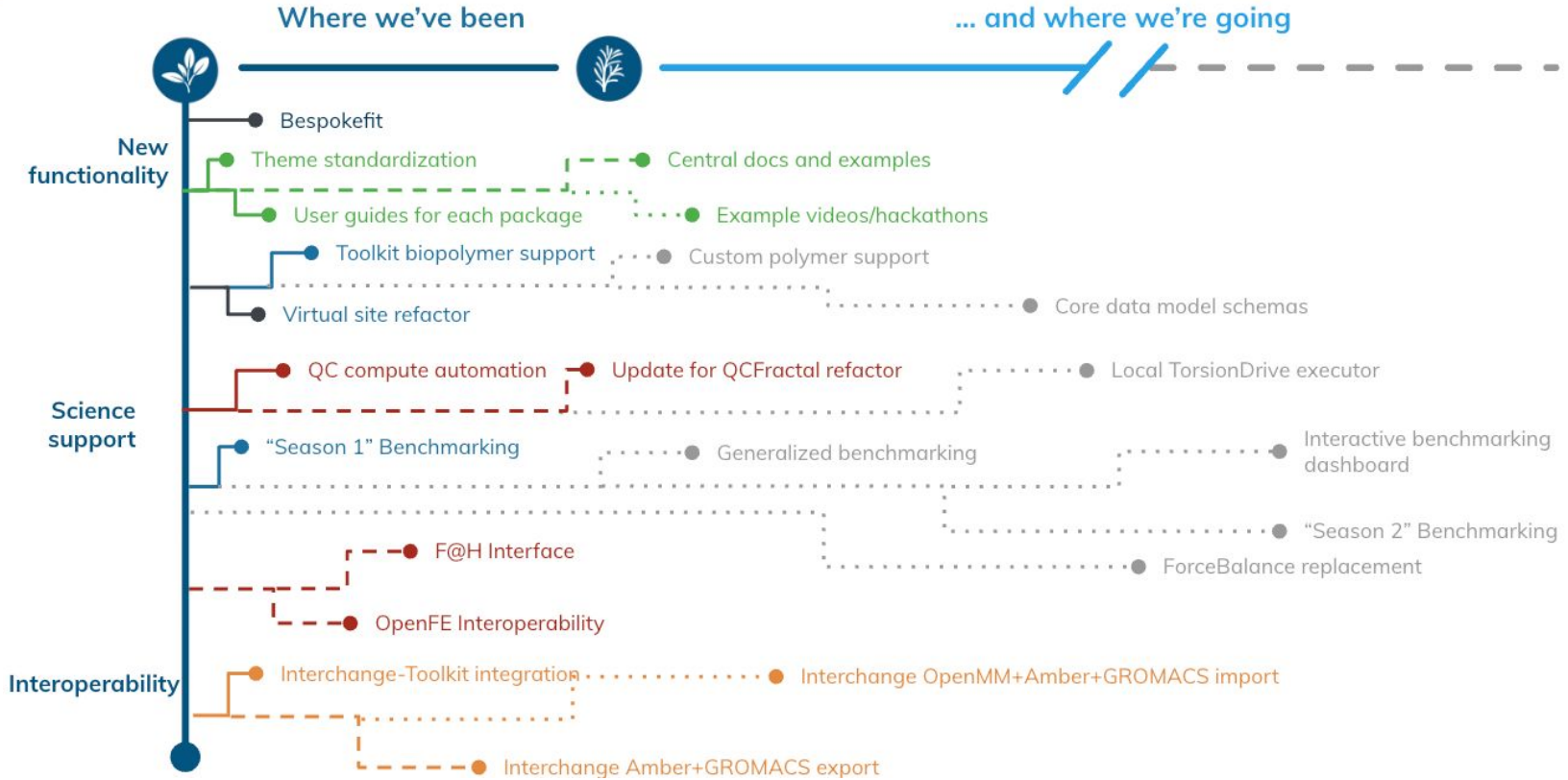
- We viewed this originally as a linear, planned process



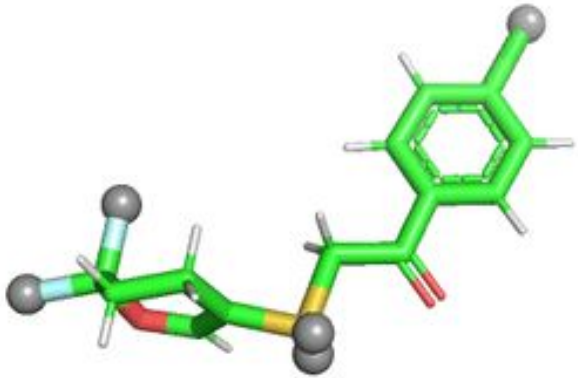
- We're learning it works better as a parallel process where the best ideas may be unexpected and apparent only later

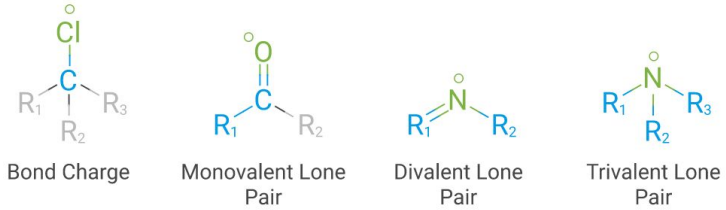


Before we can do the science, we need infrastructure, so you'll hear about great infrastructure...

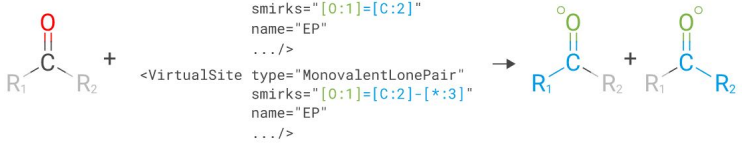


...including virtual site support for off-site charges





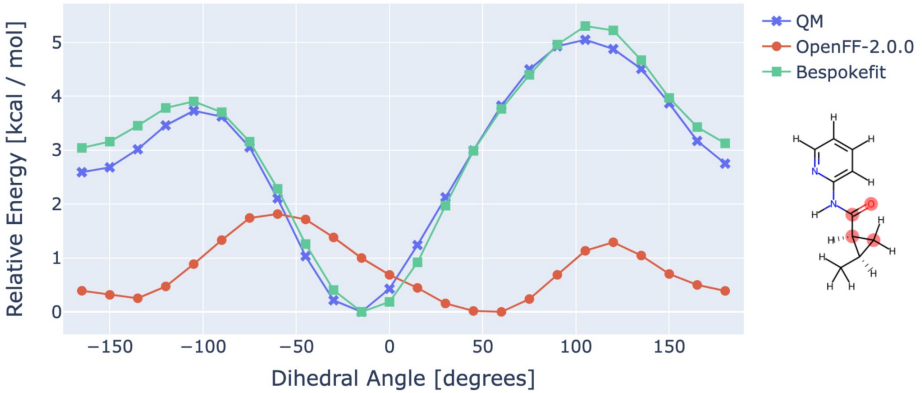
Examples of each type of virtual site with 'orientation' atoms colored blue and 'parent' atoms colored green.



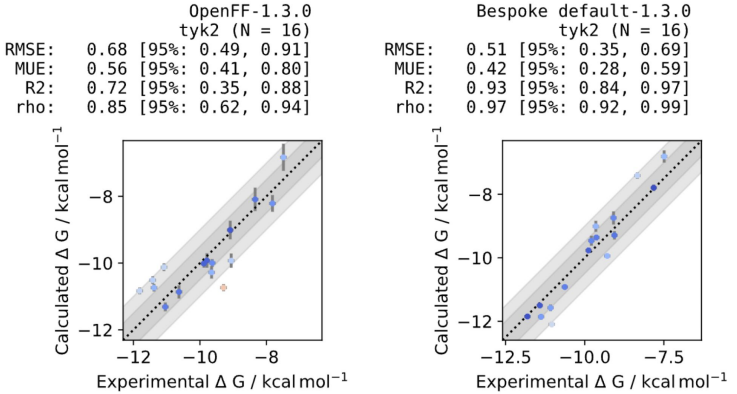
The last parameter to match a particular parent atom wins. Here the monovalent lone parameter would be assigned rather than the bond charge parameter as it appears later in the parameter list.

As well as our BespokeFit release for custom, more accurate torsional parameters

- Can retrain torsion parameters to bespoke torsion scans generated for ‘fragments’ of original molecule
- Very fast with GFN-XTB or similar; also works with QM method of choice

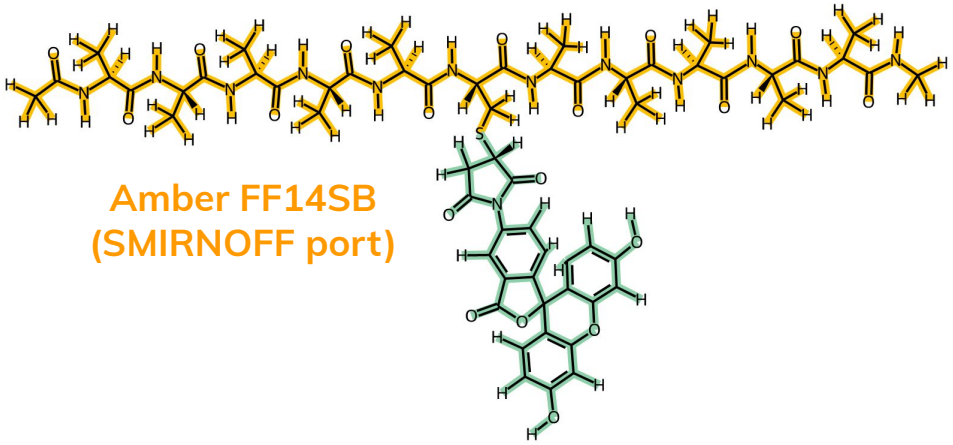


Force Field	Number of unique torsion parameters
OpenFF-1.0.0	157
OpenFF-1.2.0	163
OpenFF-1.3.0	167
OpenFF-2.0.0	167
OPLS3	48,142
OPLSe	146,669



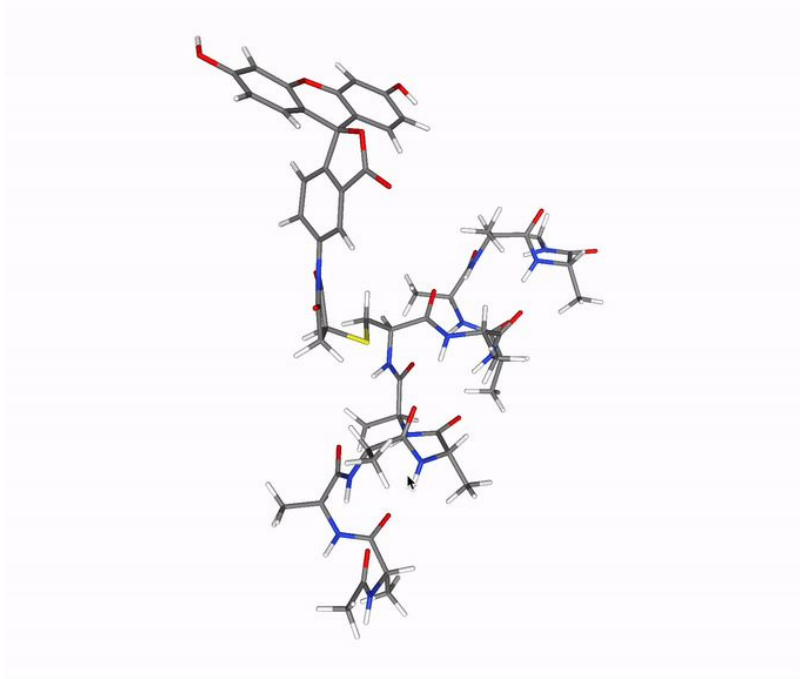
And of course you'll hear about biopolymer support for our Rosemary release

Handling post-translational modifications will no longer be terrible



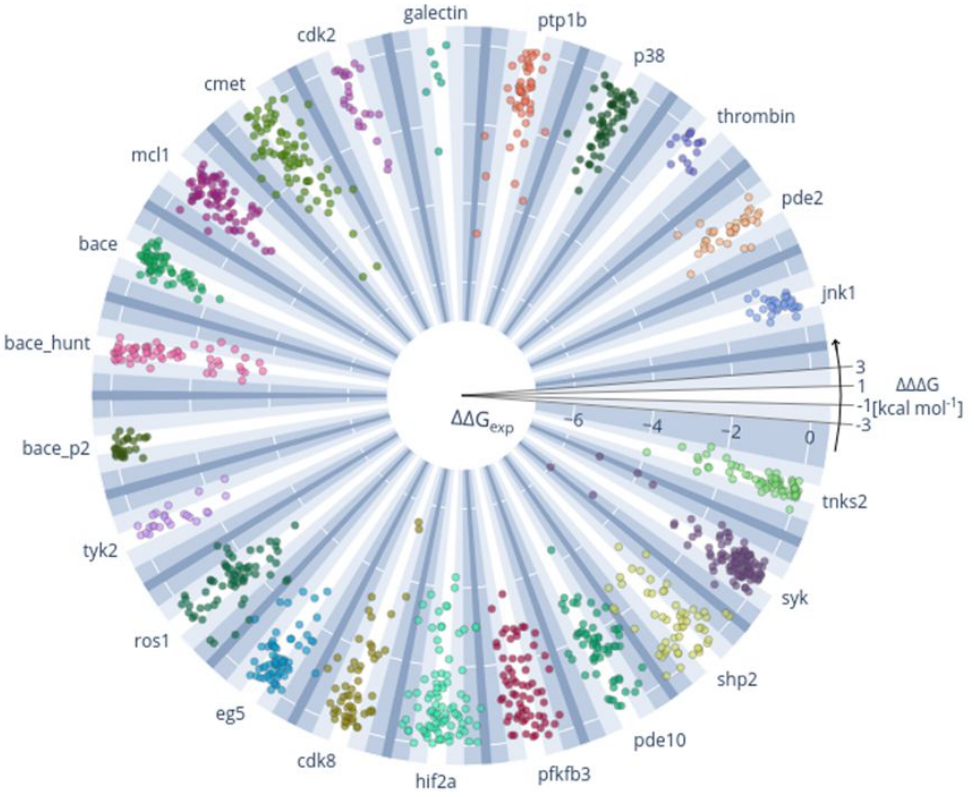
Amber FF14SB (SMIRNOFF port)

Sage



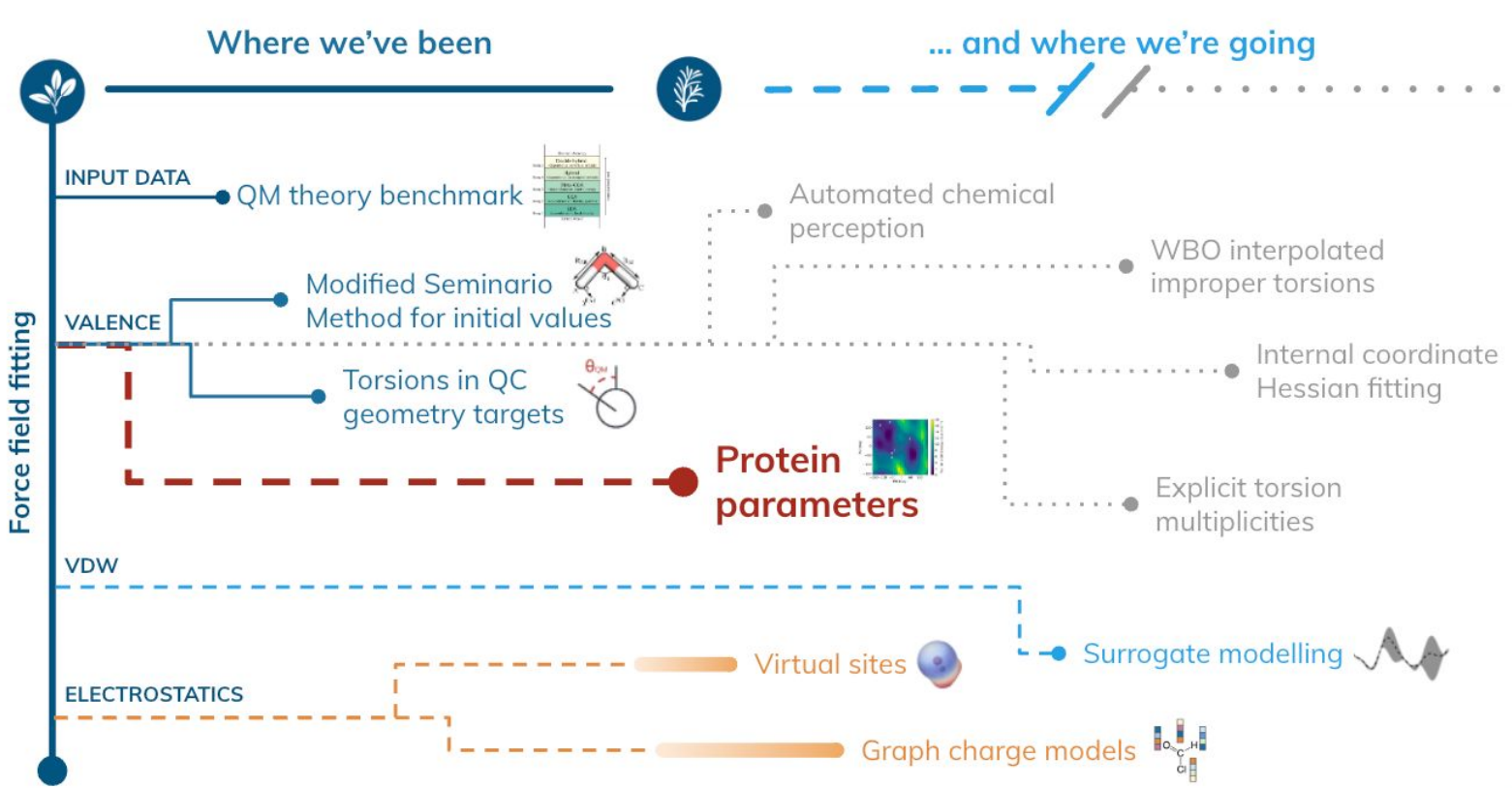
We have an ongoing need to benchmark on binding free energies, leading to a push to use Folding@Home

- Overview over all calculations performed
- Radial: exp. ΔΔG in kcal/mol
- Polar: difference between calc. and exp. ΔΔG, ΔΔΔG in kcal/mol



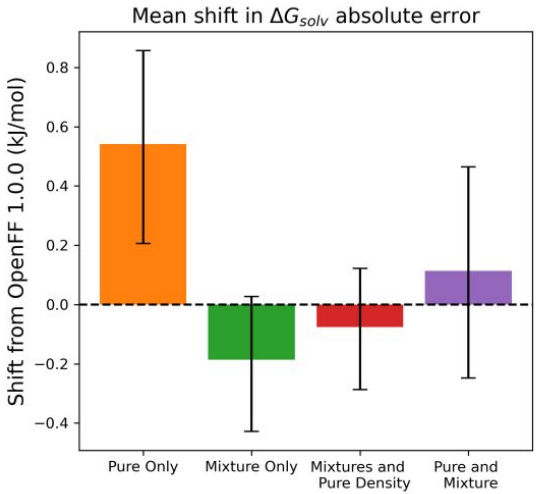
Protein ligand benchmark
https://github.com/openforcefield/protein-ligand-benchmark

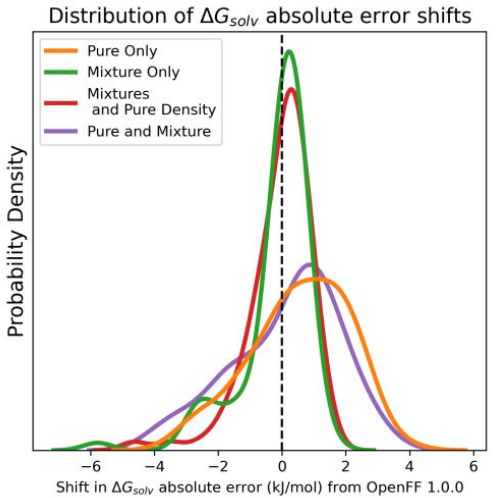
Meanwhile, there's a ton of great science going on



In Sage, fitting to mixture data substantially improved force field predictive power, a totally new result

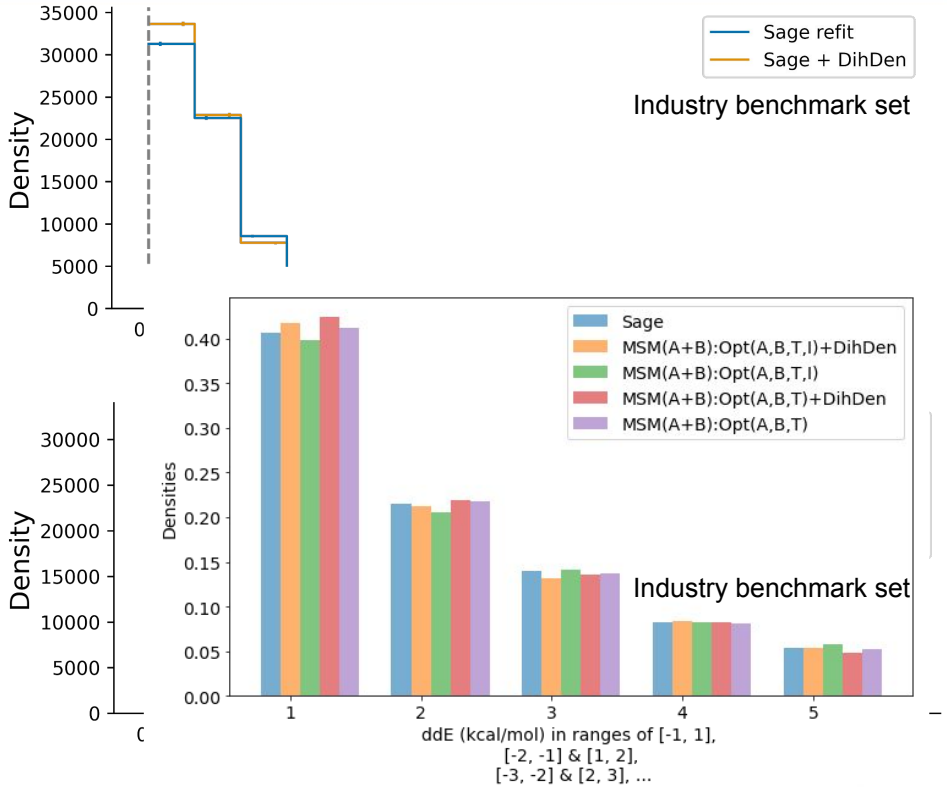
- Benchmarked refit vdW parameters against solvation free energies + transfer free energies
  - **Subset** of **FreeSolv** and **MNSol**
- Training to mixture data significantly improves performance relative to training to pure data only, or pure + mixture





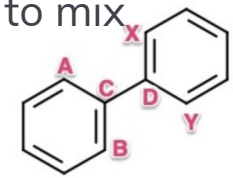
Simple improvements to the force field fitting process provide accuracy benefits - coming out soon

- Objective function ignored dihedral deviations in optimized geometries; updating provides better FFs
- New initial guesses from the modified Seminario method (MSM) improved accuracy, give more physical values
- Starting to fit impropers improves RMSD and TFD, maintains good ddE
- Cumulatively these improve accuracy substantially

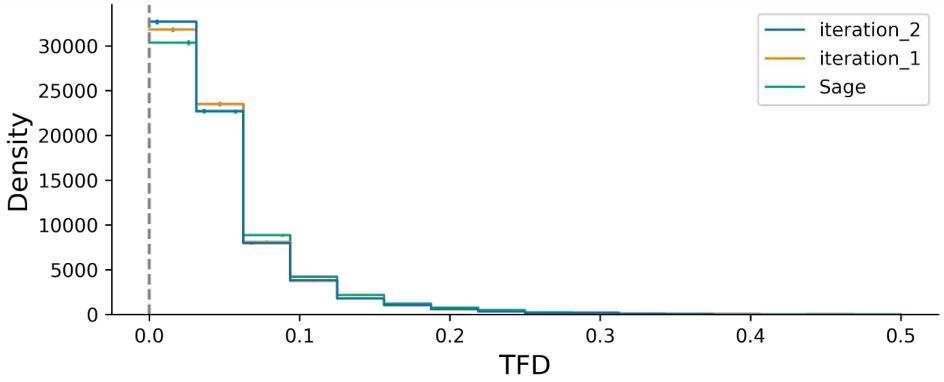


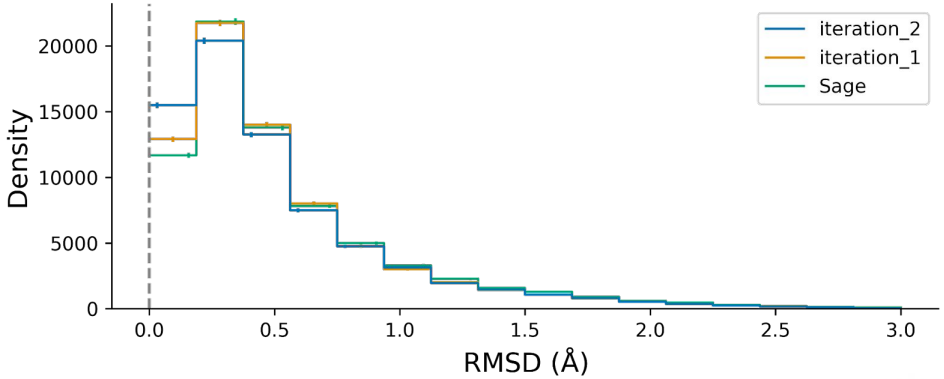
We’re finishing a fix to a problem with torsion multiplicities which will further improve accuracy

- Multiplicity: Number of torsion parameters applied to a central bond
- We realized it’s incorrect to mix multiplicity for torsions

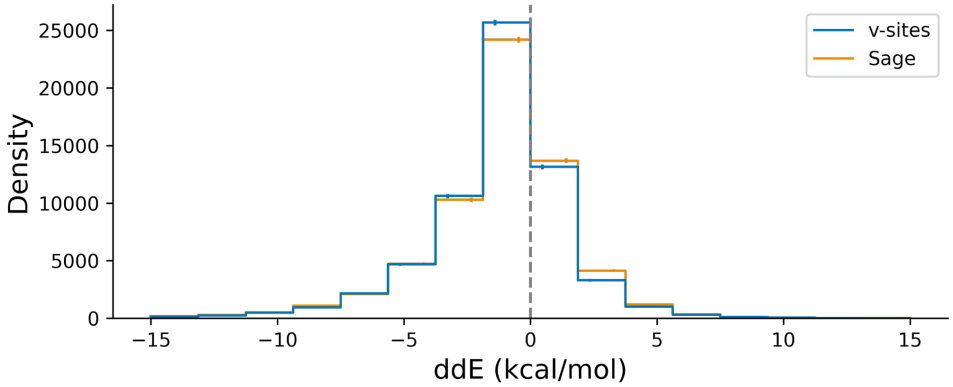


- e.g. t116: [*:1]-[#16X2,#16X3+1:2]-[#6:3]-[#1:4]
- Refits on optimization data are improved; TDs and full refit in progress

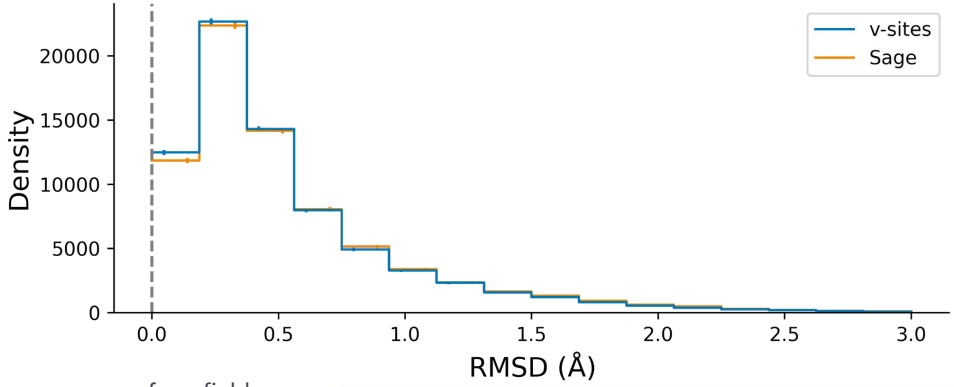


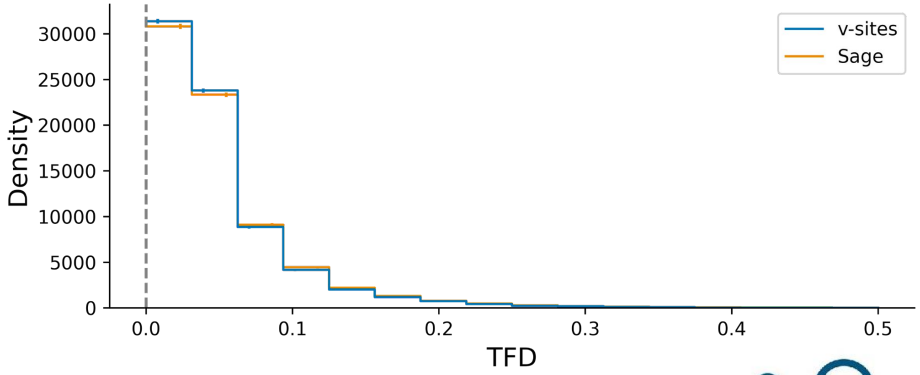


Preliminary tests of virtual sites improve accuracy broadly, even for gas phase energetics



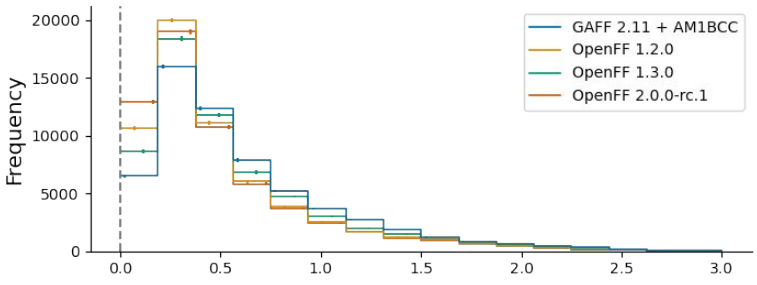
Benchmarks show modest improvements

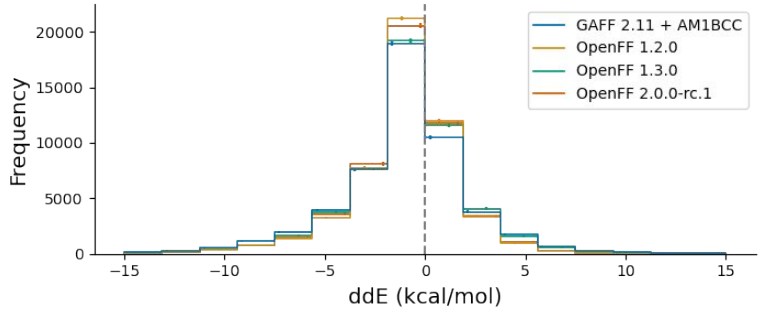


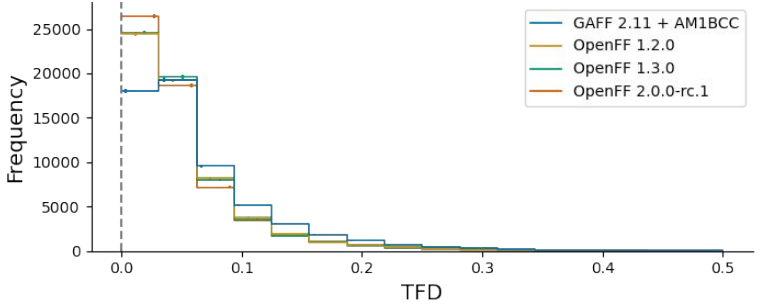


Automated benchmarking with industry has been great, indicates progress relative to QM

- **SAGE** showed excellent performance when benchmarked against the **Public** OpenFF Industry Benchmark Season 1 v1.0

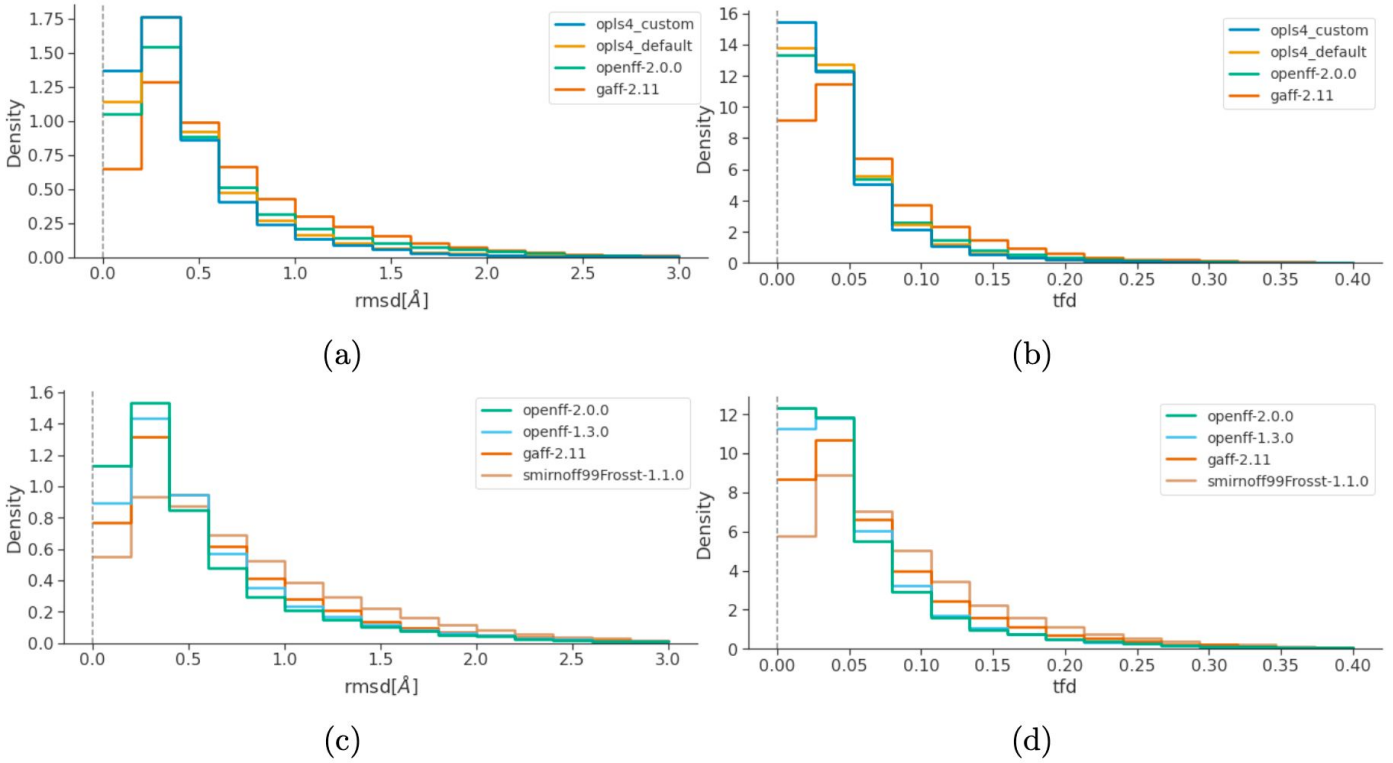






* 95% bootstrapped confidence intervals

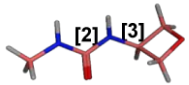
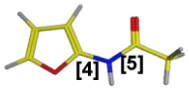
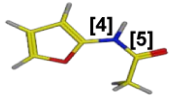
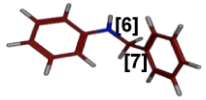
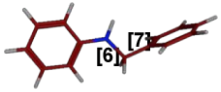
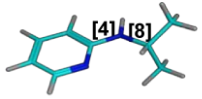
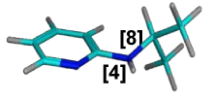
Benchmarking on proprietary datasets shows similar performance, significant improvement

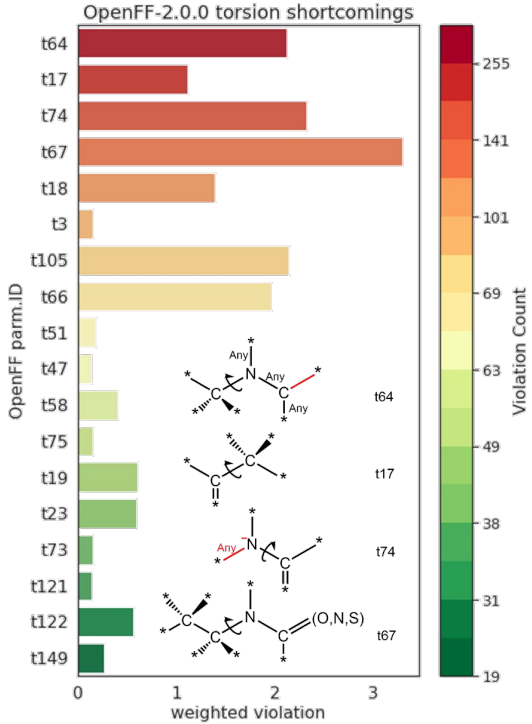


We've made marked improvement in force field accuracy, at least based on gas phase conformers

Percentages of conformations "correctly" predicted (RMSD and |ddE| within threshold)

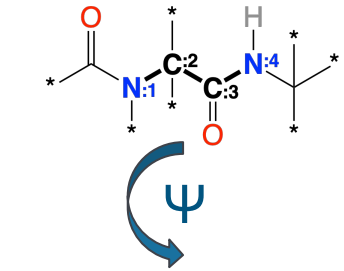
We can use benchmark data to look for trouble spots

Mol ID	QM min.	MM min.	$dE_{OpenFF-2.0.0}$	Torsion Parm.	Fig. Label
RCH-00632			+5.7 kcal/mol	[1] t47, t48	(a)
RCH-00689			+3.0 kcal/mol	[2] t75, t79 [3] t64, t67	(b)
RCH-00697			+2.7 kcal/mol	[4] t74 [5] t75, t77, t78	(c)
RCH-00333			+2.1 kcal/mol	[6] t64, t66 [7] t17	(d)
RCH-00658			+2.3 kcal/mol	[4] t74 [8] t64	(e)

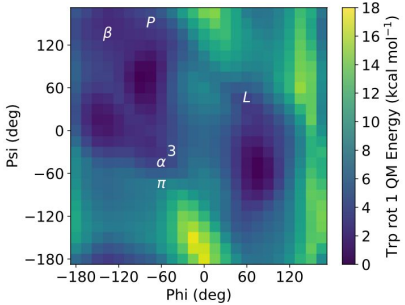


And we'll give a quick update on status of the protein force field and its parameterization

openff toolkit



[#6X3](=O)-[#7X3:1]-[#6X4:2]
-[#6X3:3](=O)-[#7X3H1:4]-[#6X4]




Iván Pulido


Jeff Wagner


Chapin Cavender

Software

Adding biopolymer support to OpenFF Toolkit

Selecting parameters

Encoding protein-specific parameters with SMARTS

QC Data

Systematic generation of torsion and optimization data for amino acids

Infrastructure in 2022

Our Organizational Values (in order of priority!)

1. Product leadership
   - Accurate force fields
   - Accessible infrastructure
   - Broad interoperability
2. Proximity with users
   - Custom solutions
   - Joint development
3. Operational excellence
   - Transparency
   - Predictability

Infrastructure in 2022

Our Organizational Values

1. Product leadership
    - Accurate force fields
    - Accessible infrastructure
    - Broad interoperability
2. Proximity with users
    - Custom solutions
    - Joint development
3. **Operational excellence**
    - **Transparency**
    - **Predictability**

Making major efforts predictable and transparent

Project Plan: Adoption of Interchange backend as export machinery for ForceField.create_openmm_system

JW MT Created by Jeffrey Wagner
Last updated: Jun 06, 2022 • 4 min read • 16 people viewed

Primary Driver	@Matt Thompson
Approver	@Jeffrey Wagner
Supporting Drivers	@Jeffrey Wagner
Stakeholders	@Jeffry Setiadi @Lily Wang @Michael Shirts Richard Gowers @Joshua Horton Irfan Alibay, Hyesu Jang,
Decision authority	Milestone completion: Unanimity of Primary Driver and Approver Significant spec changes: Unanimity of Primary Driver and Approver in upon consultation of stakeholders
Discussion/ Notification venue	The #developers slack channel, with a 3-day feedback period for any major decision, in which all stakeholders are tagged
Objective	Replace current Toolkit code that generates OpenMM systems with Interchange in a way that does not unduly affect users or FF developers

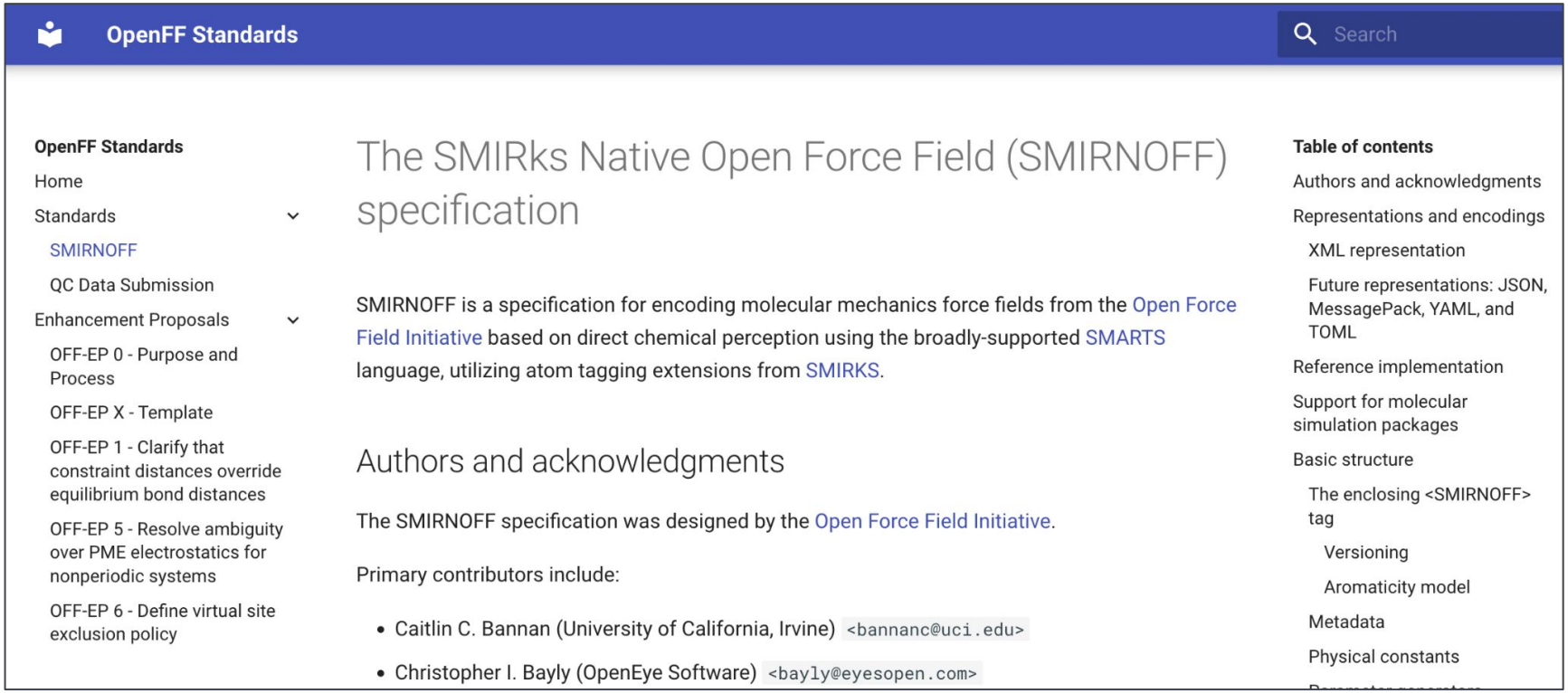
Protein-Ligand Benchmarks Automation via Folding@Home

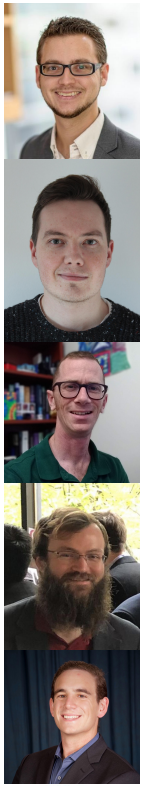
JW Created by David Dotson
Last updated: Mar 08, 2022 by Jeffrey Wagner • 2 min read • 18 people viewed

Primary Driver	@David Dotson
Approvers	@Jeffrey Wagner (OpenFF) @John Chodera (Chodera Lab, and F@H, and COVID Moonshot) @Richard Gowers (OpenFE)
Supporting Drivers	@Lorenzo D'Amore
Stakeholders	@Simon Boothroyd Irfan Alibay @Richard Gowers David Swenson @Jeffry Setiadi @Lorenzo D'Amore @Lily Wang Antonia Mey @David Mobley
Project Manager	@Diego Nolasco
Decision authority	Unanimity of Primary Driver and two Approvers (absences are vetos), only in synchronous meetings Veto authority: Primary Driver, any Approver

https://openforcefield.atlassian.net/wiki/spaces/IN/pages/2225111061/Project+Plan+Adoption+of+Interchange+backend+as+export+machinery+for+ForceField.create+openmm+system
https://openforcefield.atlassian.net/wiki/spaces/IN/pages/1447231531/Protein-Ligand+Benchmarks+Automation+via+Folding+Home

We have standards!





(source) https://github.com/openforcefield/standards
(published) https://openforcefield.github.io/standards/standards/smirnoff/

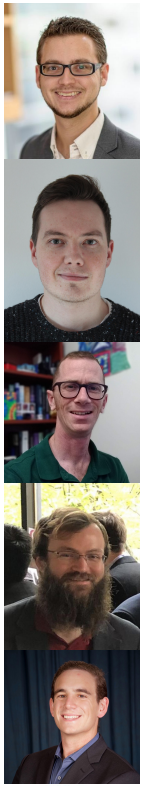
We have standards!

openforcefield / standards Public

- OFF-EP 0006 ✓ — 2
  #35 by SimonBoothroyd was merged on Apr 11 • Approved
- OFF-EP 0005b ✓ — 1 — 62
  #34 by jchodera was merged on Apr 20 • Approved
- OFF-EP 0005 ✓ — 1 — 13
  #30 by mattwthompson was closed on Mar 31 • Approved
- OFF-EP 0004 ✓ — 1 — 7
  #24 opened on Dec 1, 2021 by mattwthompson • Changes requested
- OFF-EP 0003 ✓ — 1 — 15
  #23 opened on Nov 23, 2021 by mattwthompson • Approved
- OFF-EP 0002 ✓ — 1 — 24
  #22 opened on Nov 23, 2021 by mattwthompson • Changes requested
- OFF-EP 0001 ✓ — 1 — 23
  #21 by mattwthompson was merged on Dec 3, 2021 • Changes requested
- OFF-EP 0000 — 16
  #1 by SimonBoothroyd was merged on May 26, 2021 • Changes requested — 1 task done

OpenFF Standards

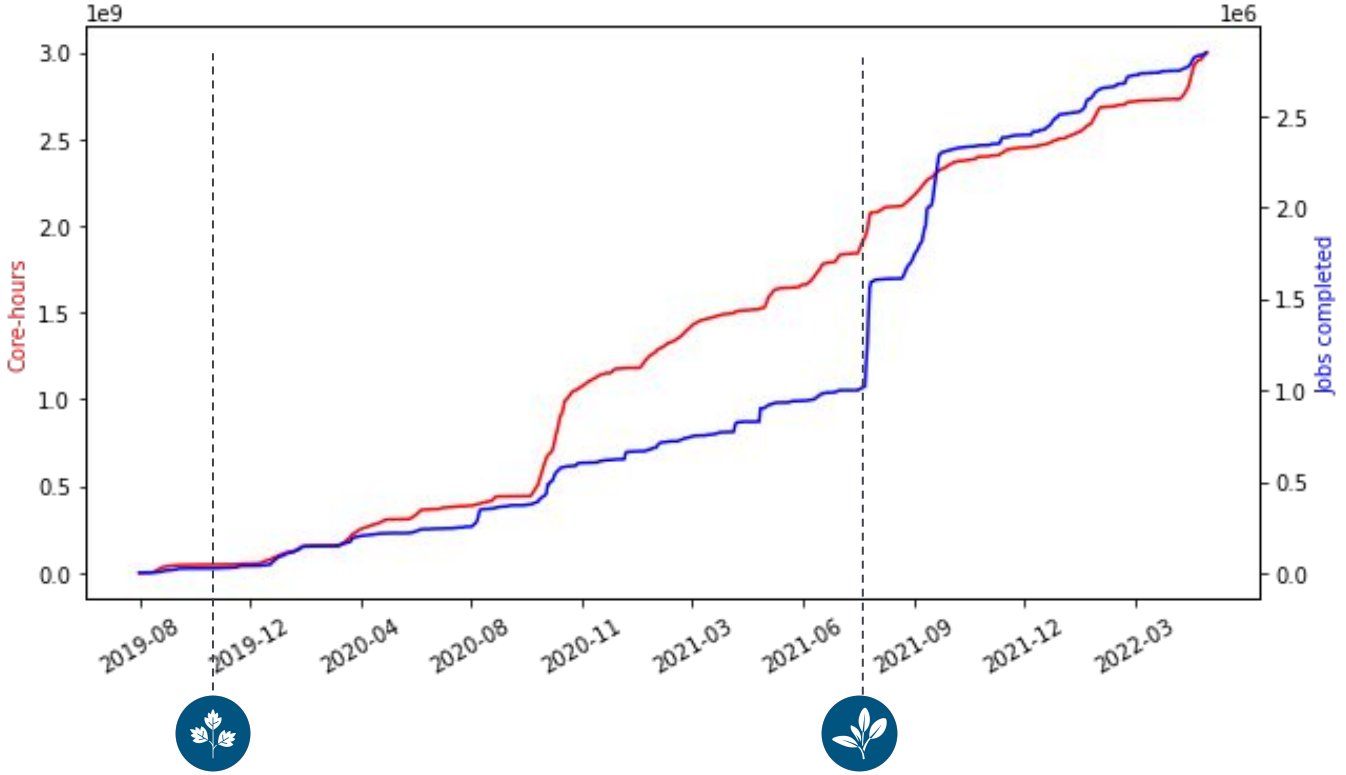
This repository aims to store and document the various standards employed across the Open Force Field Consortium. These standards encompass both the specification of core data models and formats, such as the SMIRNOFF specification, as well as standard operating procedures such as those detailing how OpenFF curated QC data sets are to be submitted to the QCArchive.

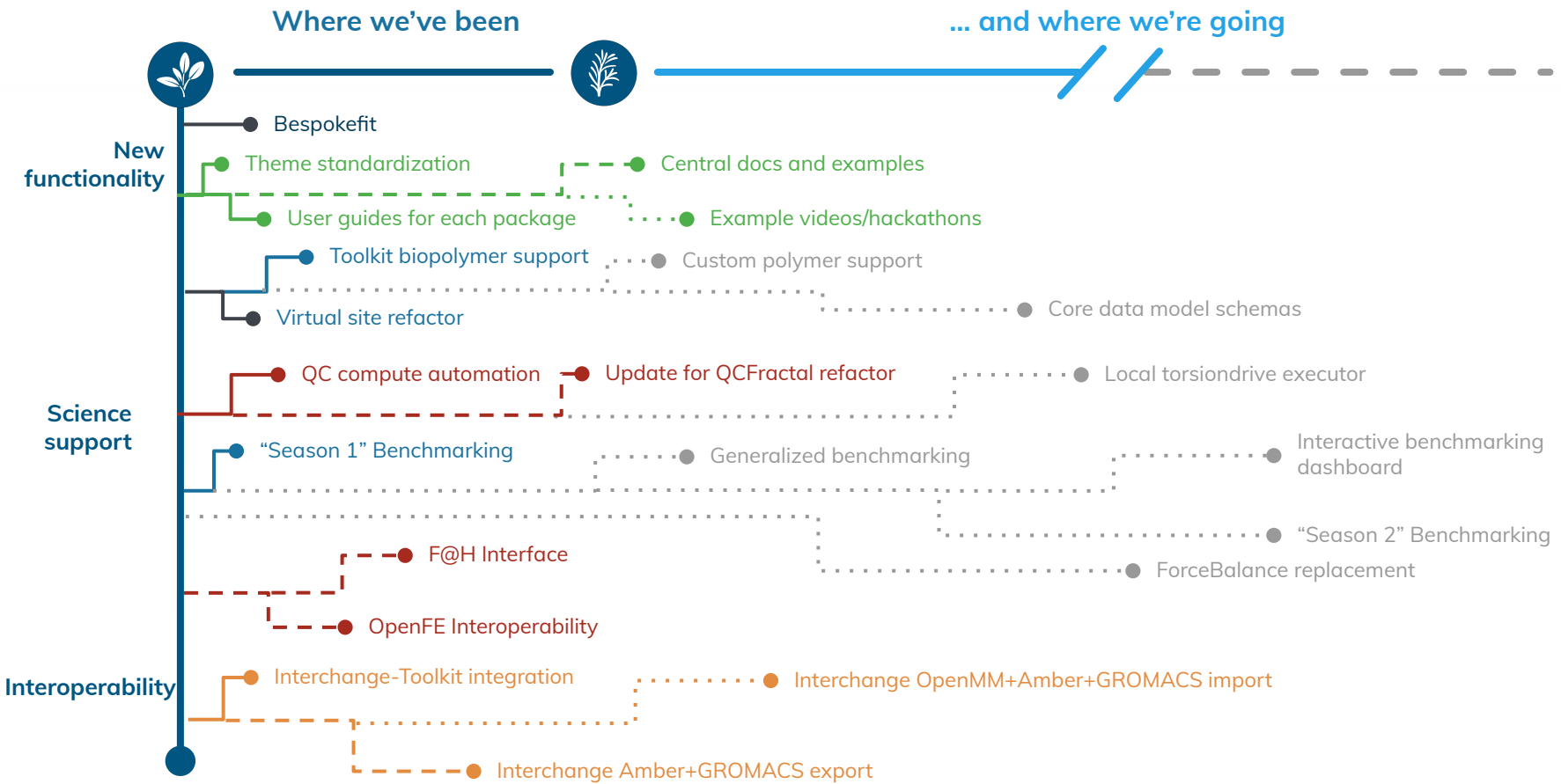


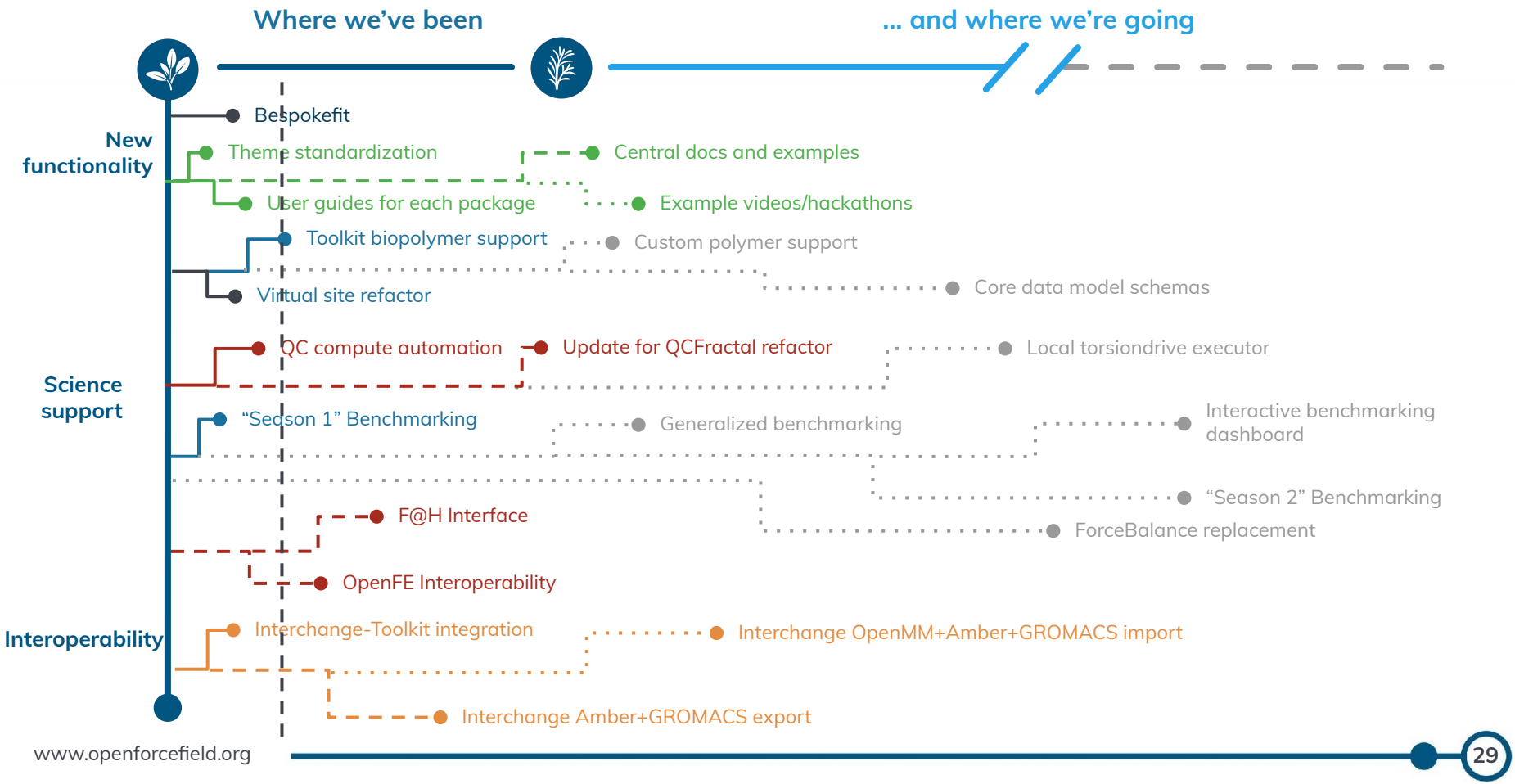
(source) https://github.com/openforcefield/standards

(published) https://openforcefield.github.io/standards/standards/smirnoff/

QC compute has continued scaling




Where we've been
... and where we're going
New functionality
Bespokefit
Theme standardization
Central docs and examples
User guides for each package
Example videos/hackathons
Toolkit biopolymer support
Custom polymer support
Virtual site refactor
Core data model schemas
Science support
QC compute automation
Update for QCFractal refactor
Local torsiondrive executor
"Season 1" Benchmarking
Generalized benchmarking
Interactive benchmarking dashboard
"Season 2" Benchmarking
F@H Interface
ForceBalance replacement
OpenFE Interoperability
Interoperability
Interchange-Toolkit integration
Interchange OpenMM+Amber+GROMACS import
Interchange Amber+GROMACS export


Where we've been
... and where we're going
New functionality
Science support
Interoperability
Bespokefit
Theme standardization
Central docs and examples
User guides for each package
Example videos/hackathons
Toolkit biopolymer support
Custom polymer support
Virtual site refactor
Core data model schemas
QC compute automation
Update for QCFractal refactor
Local torsiondrive executor
"Season 1" Benchmarking
Generalized benchmarking
Interactive benchmarking dashboard
"Season 2" Benchmarking
ForceBalance replacement
F@H Interface
OpenFE Interoperability
Interchange-Toolkit integration
Interchange OpenMM+Amber+GROMACS import
Interchange Amber+GROMACS export

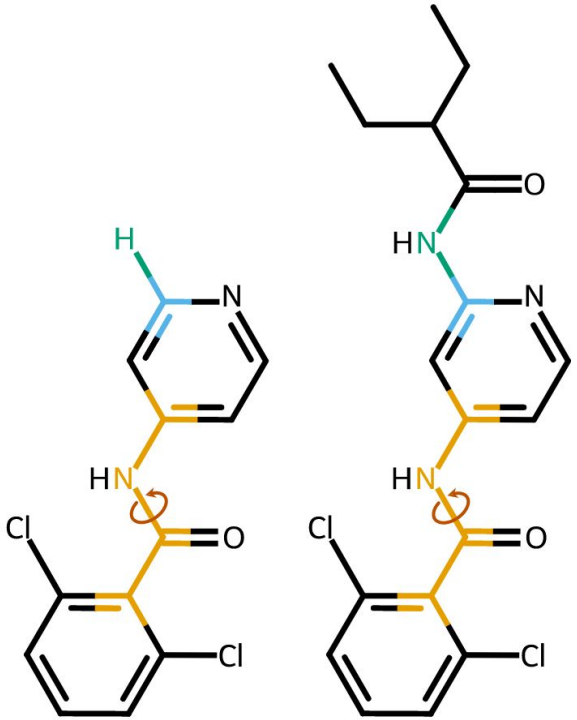
Infrastructure in 2022

Our Organizational Values

1. Product leadership
    - Accurate force fields
    - Accessible infrastructure
    - Broad interoperability
2. **Proximity with users**
    - **Custom solutions**
    - **Joint development**
3. Operational excellence
    - Transparency
    - Predictability

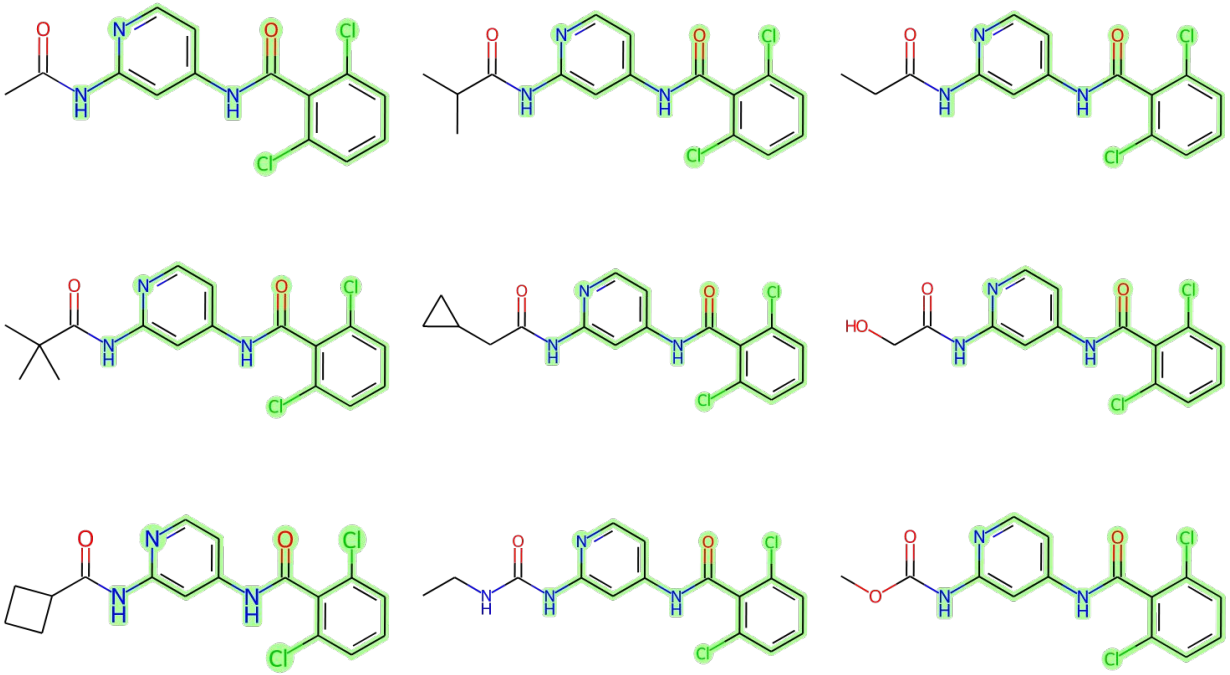
BespokeFit has been released

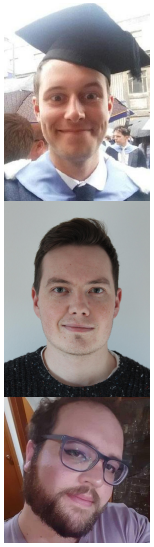
```
[#6H0X3x2r6+0a:1](
    :;@[#6H0X3x2r6+0a]
        (-;!@[#17H0X1x0!r+0A])
    :;@[#6H1X3x2r6+0a]
        (-;!@[#1H0X1x0!r+0A])
    :;@[#6H1X3x2r6+0a]
        (-;!@[#1H0X1x0!r+0A])
    :;@[#6H1X3x2r6+0a]
    -;!@[#1H0X1x0!r+0A]
)
(
    :;@[#6H0X3x2r6+0a]
    -;!@[#17H0X1x0!r+0A]
)
-;!@[#6H0X3x0!r+0A:2]
    (=;!@[#8H0X1x0!r+0A])
-;!@[#7H1X3x0!r+0A:3]
    (-;!@[#1H0X1x0!r+0A])
-;!@[#6H0X3x2r6+0a:4](
    :;@[#6H1X3x2r6+0a]
        (-;!@[#1H0X1x0!r+0A])
    :;@[#6H1X3x2r6+0a]
        (-;!@[#1H0X1x0!r+0A])
    :;@[#7H0X2x2r6+0a]
    :;@[#6H0X3x2r6+0a,#6H1X3x2r6+0a]
    -;!@[#1H0X1x0!r+0A,#7H1X3x0!r+0A]
)
:;@[#6H1X3x2r6+0a]
-;!@[#1H0X1x0!r+0A]
```



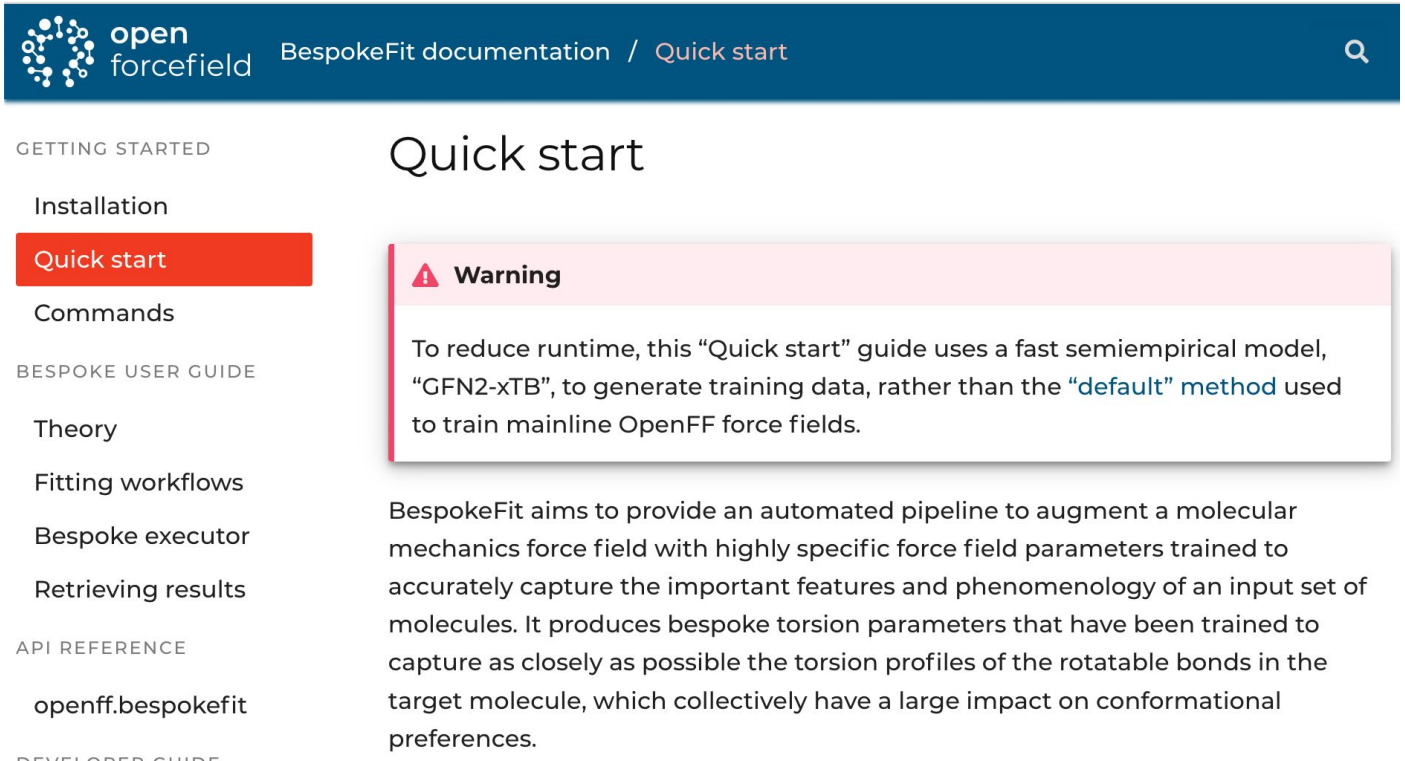
https://docs.openforcefield.org/projects/bespokefit/en/latest/getting-started/quick-start.html

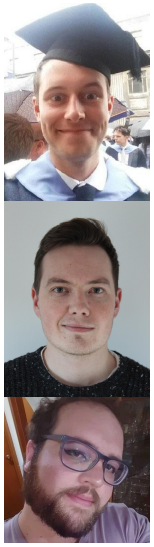
BespokeFit has been released





BespokeFit has been released



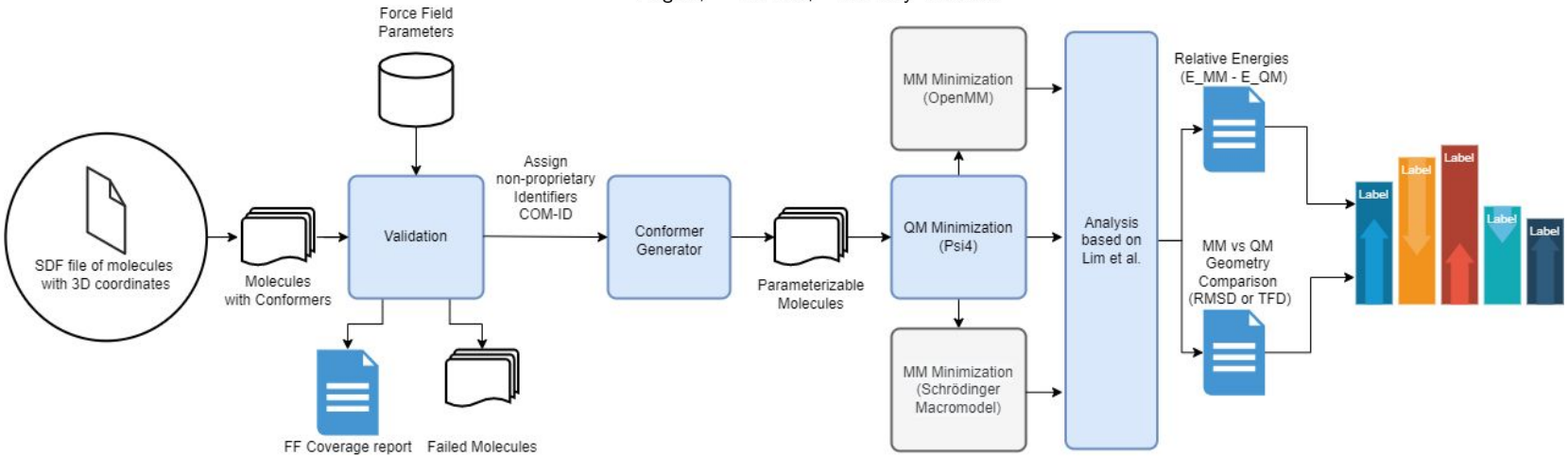


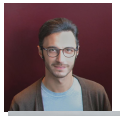
https://docs.openforcefield.org/projects/bespokefit/en/latest/getting-started/quick-start.html

Partner benchmarking manuscript headed for submission

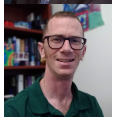
Collaborative assessment of molecular geometries and energies from the Open Force Field

Lorenzo D'Amore,[†] David F. Hahn,[‡] David L. Dotson,[¶] Joshua T. Horton,[§] Ian Craig,[∥] Pieter J. in 't Veld,[∥] Thomas Fox,[⊥] Alberto Gobbi,[#] Sirish Kaushik Lakkaraju,[@] Xavier Lucas,[△] Katharina Meier,[▽] David L. Mobley,[††] Arjun Narayanan,[‡‡] Christina E.M. Schindler,[¶¶] William Swope,[#] Jeffrey Wagner,[¶,§§] Bai Xue,[|||] and Gary Tresadern*,[‡]







Python toolkits help us find our weak spots

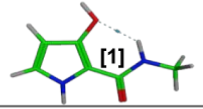
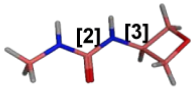
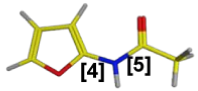
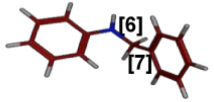
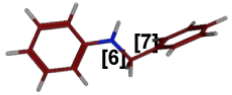
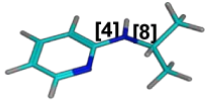
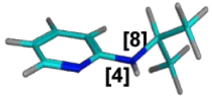
Mol ID	QM min.	MM min.	$dE_{OpenFF-2.0.0}$	Torsion Parm.	Fig. Label
RCH-00632			+5.7 kcal/mol	[1] t47, t48	(a)
RCH-00689			+3.0 kcal/mol	[2] t75, t79 [3] t64, t67	(b)
RCH-00697			+2.7 kcal/mol	[4] t74 [5] t75, t77, t78	(c)
RCH-00333			+2.1 kcal/mol	[6] t64, t66 [7] t17	(d)
RCH-00658			+2.3 kcal/mol	[4] t74 [8] t64	(e)

Figure 5: Molecular fragments of the Roche dataset containing concerning torsions. Global minima conformers optimized with QM and MM are shown with the concerning torsion(s) marked in brackets. Relative (dE) energies calculated according to equation 1 and torsion parameter(s) associated with corresponding concerning torsion(s) are reported.

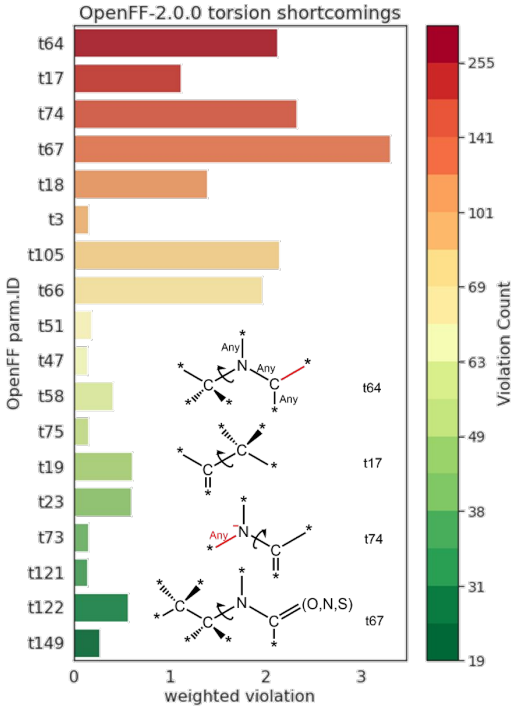


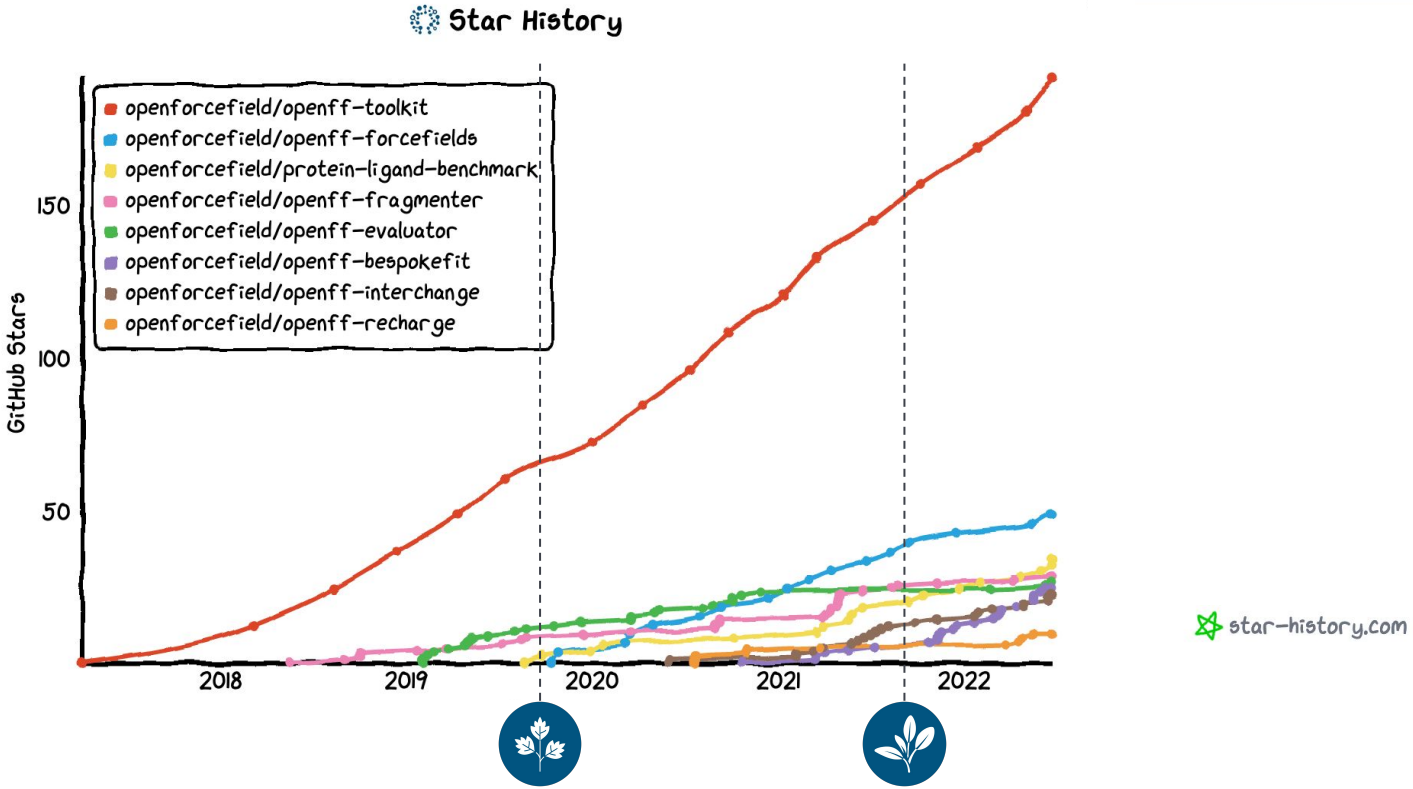
Figure 6: Analysis of torsion violations in the Public OpenFF Industry Dataset. Inset: 2D sketch chemistry match of selected torsion parameters. Elements in red color (bond, charge) may or may not exist, meaning that the corresponding atom can be either tri- or dicoordinated, respectively

Infrastructure in 2022

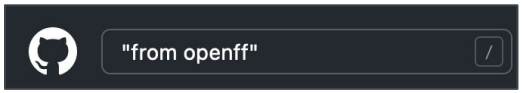
Our Organizational Values

1. **Product leadership**
   - **Accurate force fields**
   - **Accessible infrastructure**
   - **Broad interoperability**
2. Proximity with users
   - Custom solutions
   - Joint development
3. Operational excellence
   - Transparency
   - Predictability

Our projects are becoming widely-adopted



Our projects are becoming widely-adopted



1,259 code results

Sort: Best match

eduardocerqueira/seeker

seeker/snippet/build_waterbox.py

```
7  from openmm.app import Simulation
8  from openff.evaluator.protocols import coordinates
9  from openff.evaluator.utils import packmol
10 from openff.toolkit.typing.engines.smirnoff import ForceField
11 from openff.units import unit
12 from openff.toolkit.topology import Molecule, Topology
```

Python Showing the top nine matches Last indexed 4 days ago

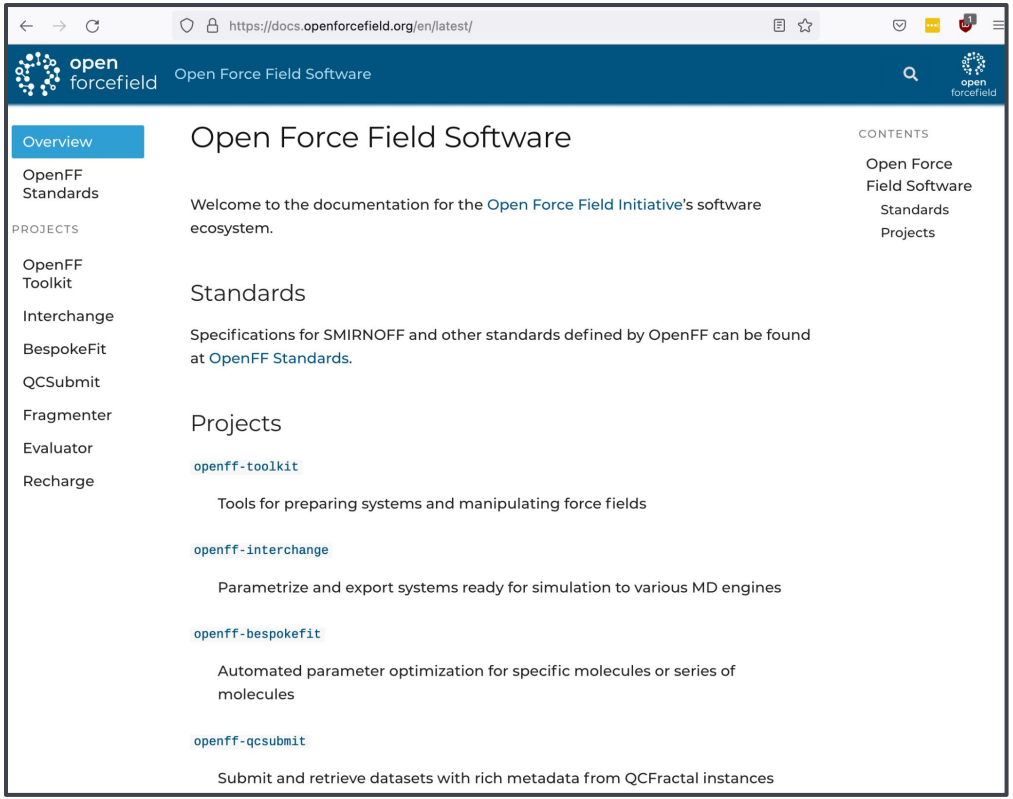
yoshiatsu163/myoff

/openff2lmp.py

```
1  # Use the OpenFF Toolkit to generate a minimal chemical topology
2  from openff.toolkit.topology import Molecule
3  from openff.toolkit.utils.toolkits import ToolkitRegistry, AmberToolsToolkitWrapper
```

Python Showing the top four matches Last indexed on Nov 3, 2021

Documentation - Central Docs



[docs.openforcefield.org](https://docs.openforcefield.org/en/latest/)

https://docs.openforcefield.org/en/latest/

Documentation - “Molecular Gastronomy”

Cookbook: Every way to make a `Molecule`

Every pathway through the OpenFF Toolkit boils down to four steps:

1. Using other tools, assemble a graph of a molecule, including all of its atoms, bonds, bond orders, formal charges, and stereochemistry[1]
2. Use that information to construct a `Molecule`
3. Combine a number of `Molecule` objects to construct a `Topology`
4. Call `ForceField.create_openmm_system(topology)` to create an OpenMM `System` (or, in the near future, an OpenFF `Interchange` for painless conversion to all sorts of MD formats)

So let’s take a look at every way there is to construct a molecule! We’ll use zwitterionic L-alanine as an example biomolecule with all the tricky bits - a stereocenter, non-zero formal charges, and bonds of different orders.

From SMILES

SMILES is the classic way to create a `Molecule`. SMILES is a widely-used compact textual representation of arbitrary molecules. This lets us specify an exact molecule, including stereochemistry and bond orders, very easily — though they may not be the most human-readable format.

The `Molecule.from_smiles()` method is used to create a `Molecule` from a SMILES code.

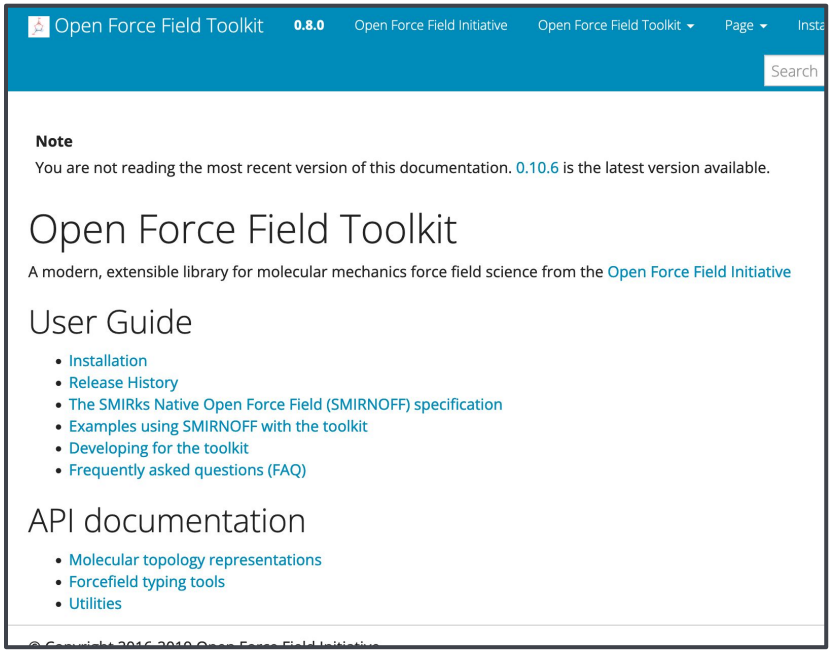
Implicit hydrogens SMILES

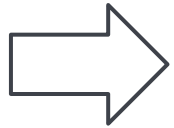
```
zw_l_alanine = Molecule.from_smiles("C[C@H]([NH3+])C(=O)[O-]")

zw_l_alanine.visualize()
```

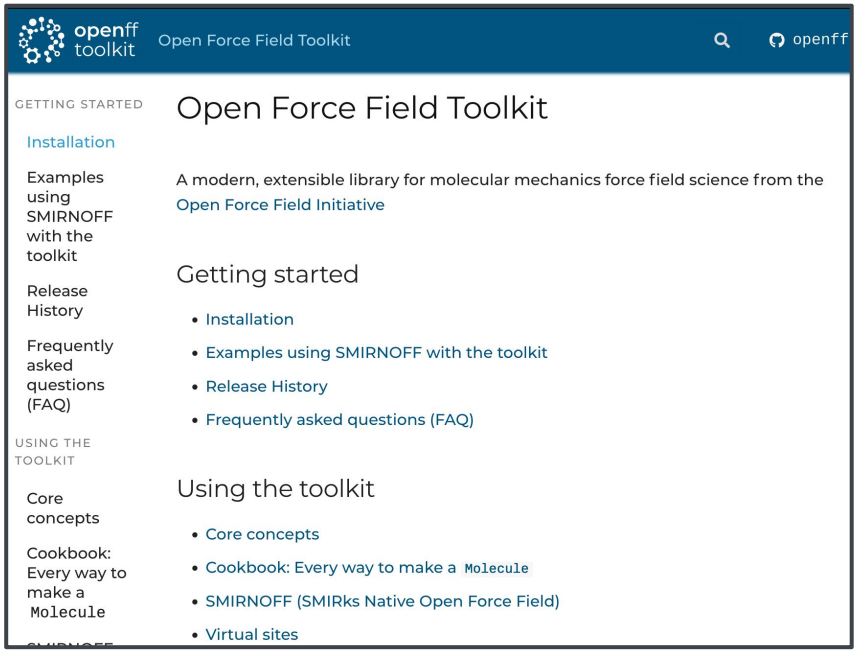
```
ModuleNotFoundError: No module named 'constraint'
```

CONTENTS

Documentation - Theme standardization







Documentation - Theme standardization

generate_conformers (*toolkit_registry=ToolkitRegistry containing The RDKit, AmberTools, Built-in Toolkit, n_conformers=10, rms_cutoff=None, clear_existing=True*)

Generate conformers for this molecule using an underlying toolkit. If n_conformers=0, no toolkit wrapper will be called. If n_conformers=0 and clear_existing=True, molecule.conformers will be set to None.

Parameters: **toolkit_registry** : *openforcefield.utils.toolkits.ToolkitRegistry or openforcefield.utils.toolkits.ToolkitWrapper, optional, default=None*
`ToolkitRegistry` or `ToolkitWrapper` to use for SMILES-to-molecule conversion

n_conformers : *int, default=1*
The maximum number of conformers to produce

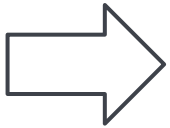
rms_cutoff : *simtk.Quantity-wrapped float, in units of distance, optional, default=None*
The minimum RMS value at which two conformers are considered redundant and one is deleted. Precise implementation of this cutoff may be toolkit-dependent. If None, the cutoff is set to be the default value for each ToolkitWrapper (generally 1 Angstrom).

clear_existing : *bool, default=True*
Whether to overwrite existing conformers for the molecule

Raises: **InvalidToolkitRegistryError**
If an invalid object is passed as the toolkit_registry parameter

Examples

```
>>> molecule = Molecule.from_smiles('CCCCCC')
>>> molecule.generate_conformers()
```



```
generate_conformers(
    toolkit_registry=GLOBAL_TOOLKIT_REGISTRY,
    n_conformers=10,
    rms_cutoff=None,
    clear_existing=True,
    make_carboxylic_acids_cis=True
)
```

Generate conformers for this molecule using an underlying toolkit.

If `n_conformers=0`, no toolkit wrapper will be called. If `n_conformers=0` and `clear_existing=True`, `molecule.conformers` will be set to `None`.

Parameters

`toolkit_registry` (`openff.toolkit.utils.toolkits.ToolkitRegistry` *or* `openff.toolkit.utils.toolkits.ToolkitWrapper`, *optional, default=None*) – `ToolkitRegistry` or `ToolkitWrapper` to use for SMILES-to-molecule conversion

`n_conformers` (`int`, *default=1*) – The maximum number of conformers to produce

`rms_cutoff` (*openmm.unit.Quantity-wrapped float, in units of distance, optional, default=None*) – The minimum RMS value at which two conformers are considered redundant and one is deleted. Precise implementation of this cutoff may be toolkit-dependent. If `None`, the cutoff is set to be the default value for each `ToolkitWrapper` (generally 1 Angstrom).

Transition to OpenFF-units

openforcefield / openff-units Public

This package provides a common unit registry for all OpenFF packages to use in order to ensure consistent unit definitions across the software ecosystem.

The unit definitions are currently sourced from the NIST 2018 CODATA, but may be updated in future versions as new CODATA updates are made.

While this repository is based on Pint, the main classes (`Unit`, `Quantity`, and `Measurement`) have been slightly modified in order to provide non-dynamic, more readily serialisable representations.

OpenMM Interoperability

For compatibility with OpenMM units, a submodule (`openff.units.openmm`) with conversion functions (`to_openmm`, `from_openmm`) is also provided.

```
>>> from openff.units import unit
>>> from openff.units.openmm import to_openmm, from_openmm
>>> distance = 24.0 * unit.meter
>>> converted = to_openmm(distance)
>>> converted
Quantity(value=24.0, unit=meter)
>>> type(converted)
<class 'openmm.unit.quantity.Quantity'>
>>> roundtripped = from_openmm(converted)
>>> roundtripped
<Quantity(24.0, 'meter')>
>>> type(roundtripped)
<class 'openff.units.units.Quantity'>
```

OpenFF Toolkit-Interchange native interoperability

```
from openff.toolkit import ForceField
ff = ForceField('ff14sb_off_impropers_0.0.3.offxml')
```

```
# Make the topology periodic
from openff.units import unit
top.box_vectors = [10., 10., 10.] * unit.nanometer
```

```
interchange = ff.create_interchange(top)
```

```
interchange
```

```
Interchange with 6 potential handlers, periodic topology with 2667 atoms.
```

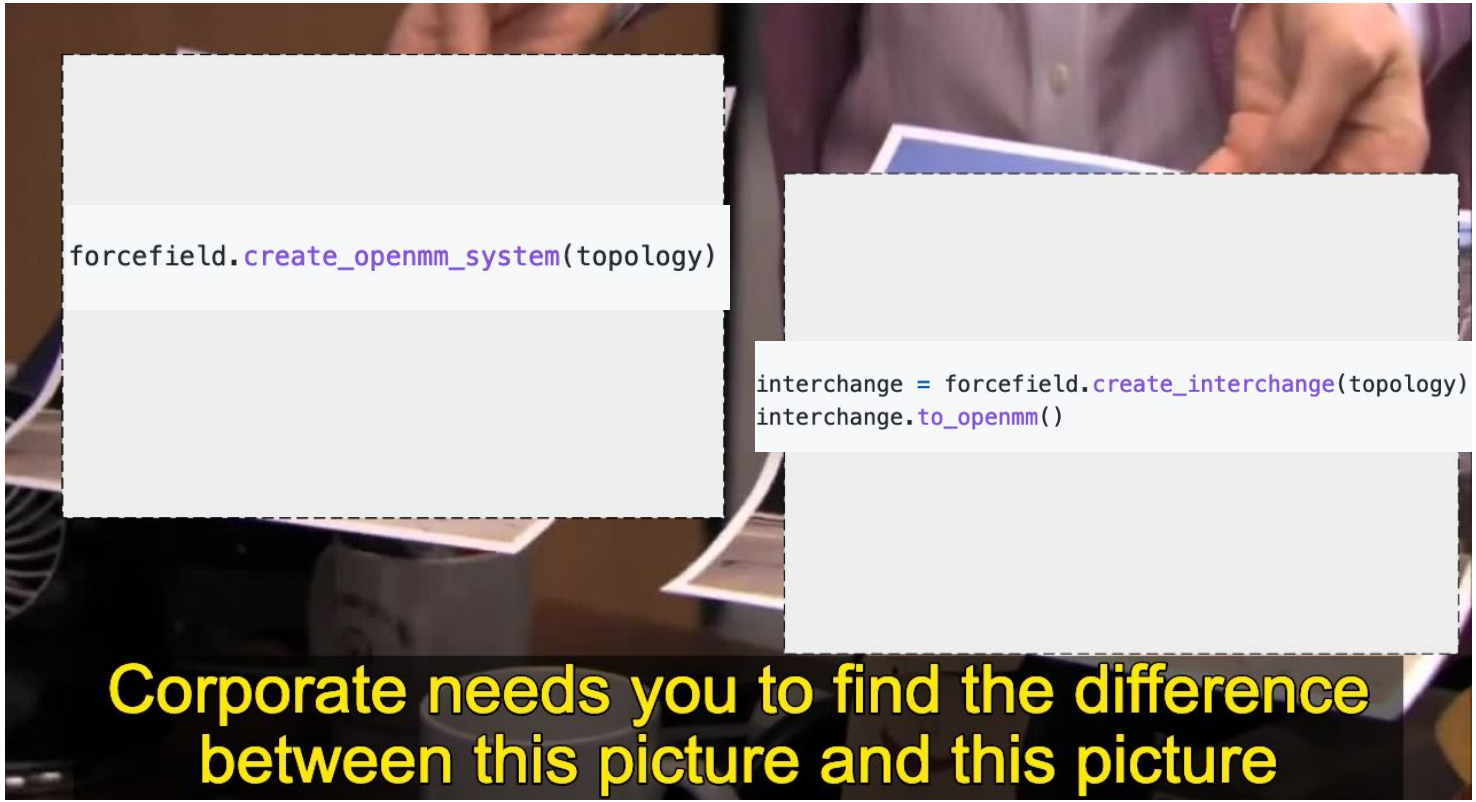
```
# Interchange contains the underlying parameters
[*interchange['Bonds'].potentials.values()][0]
```

```
Potential(parameters={'k': <Quantity(868.0, 'kilocalorie / angstrom ** 2 / mole')>, 'length': <Quantity(1.01, 'angs
trom')>}, map_key=None)
```

```
# The Interchange can export to an OpenMM system, which is pretty much the
# same as you'd get by using ForceField.create_openmm_system. It also contains
# an OpenFF Topology object (a copy of the one that made it) that can be turned
# into an OpenMM Topology
(interchange.to_openmm(), interchange.topology.to_openmm())
```

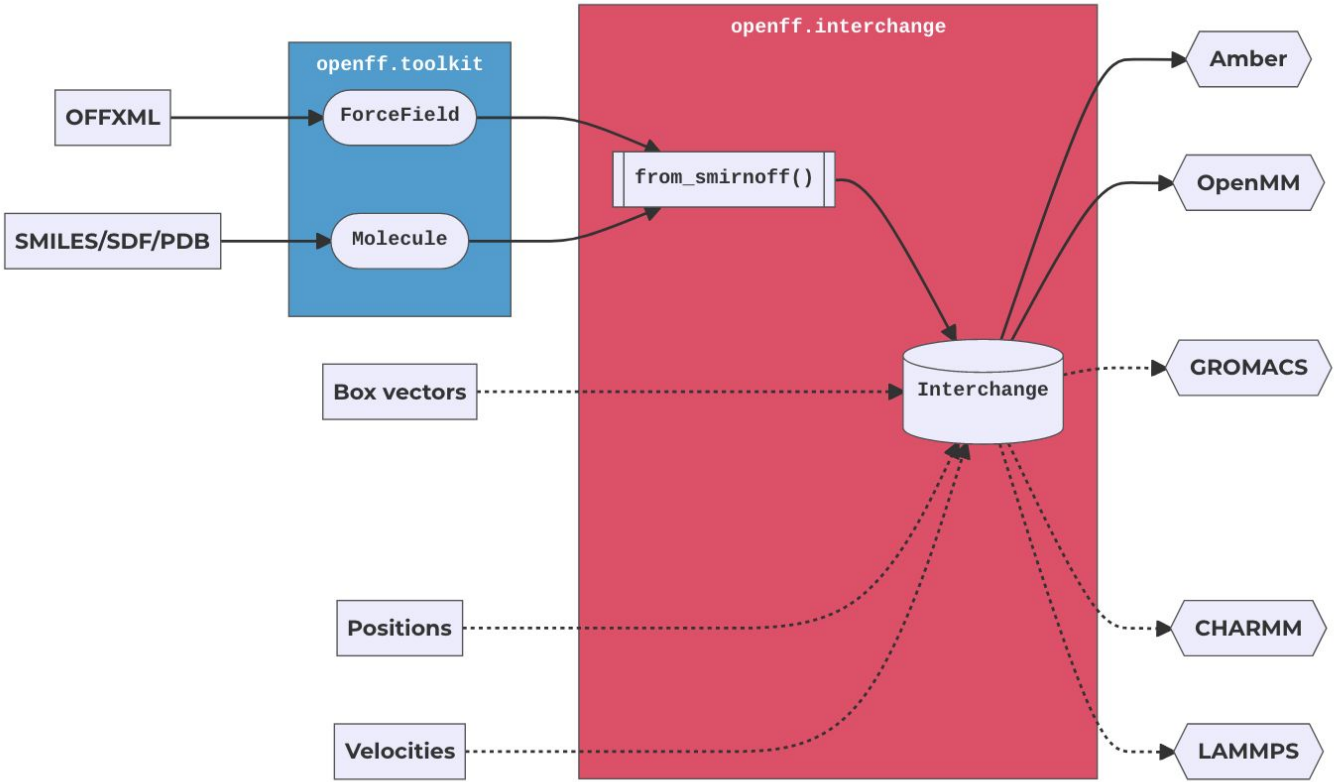
```
(<openmm.openmm.System; proxy of <Swig Object of type 'OpenMM::System *' at 0x180346960> >,
 <Topology; 2 chains, 172 residues, 2667 atoms, 2686 bonds>)
```

New code paths...

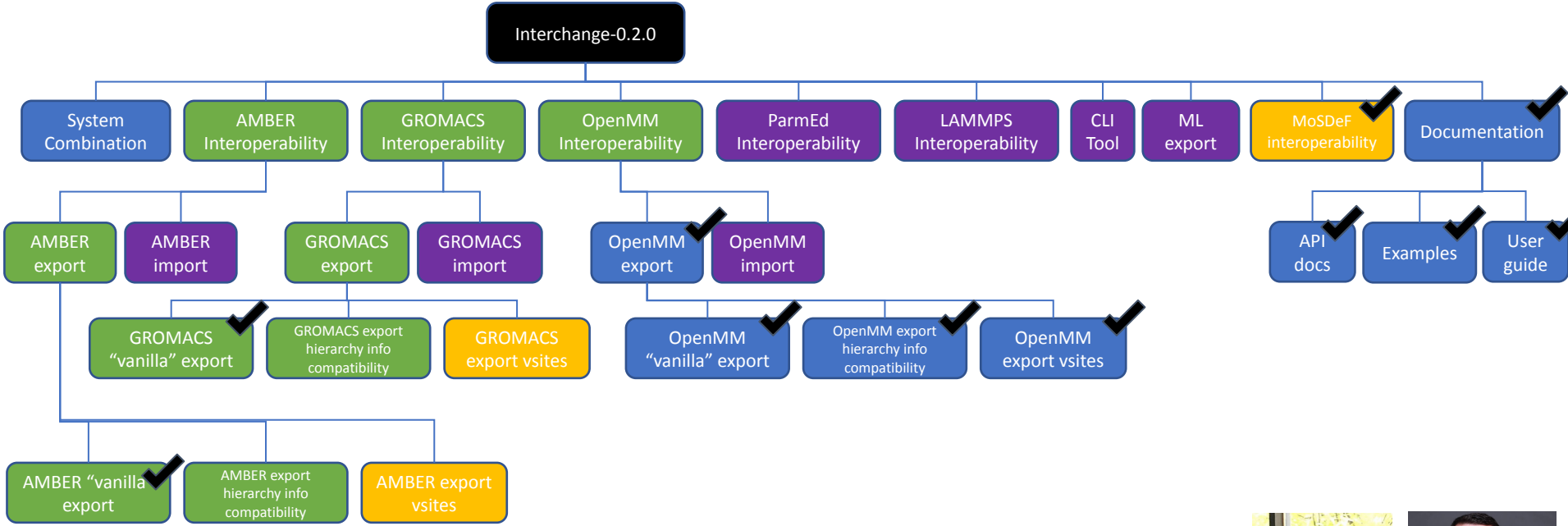


...same old behavior

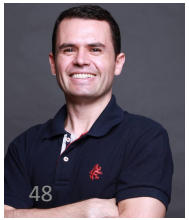
Integrating Interchange into the OpenFF Toolkit



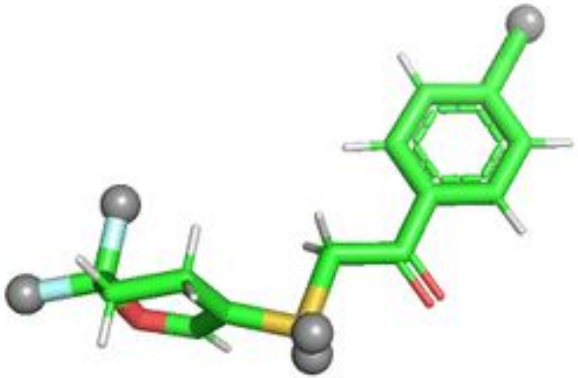
Must | Should | Could | Want

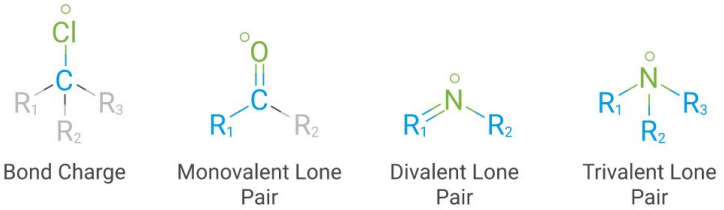


We will make <u>exporters</u> before <u>importers</u>

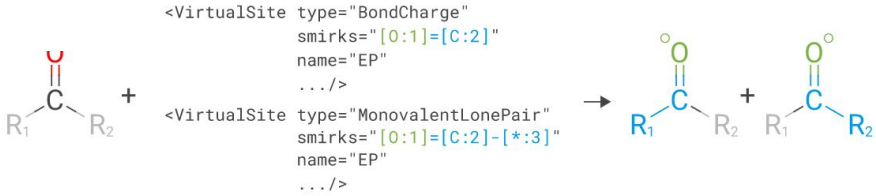


Extensive virtual site refactor



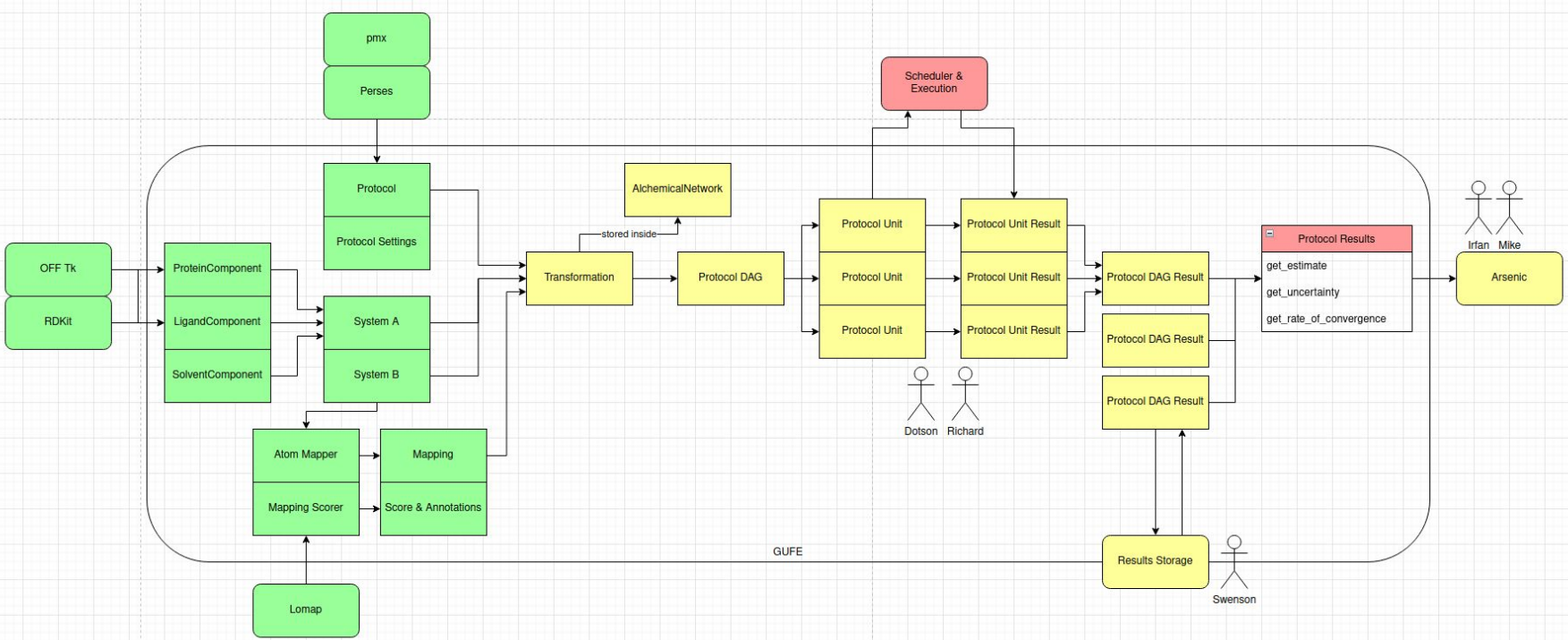


Examples of each type of virtual site with 'orientation' atoms colored blue and 'parent' atoms colored green.

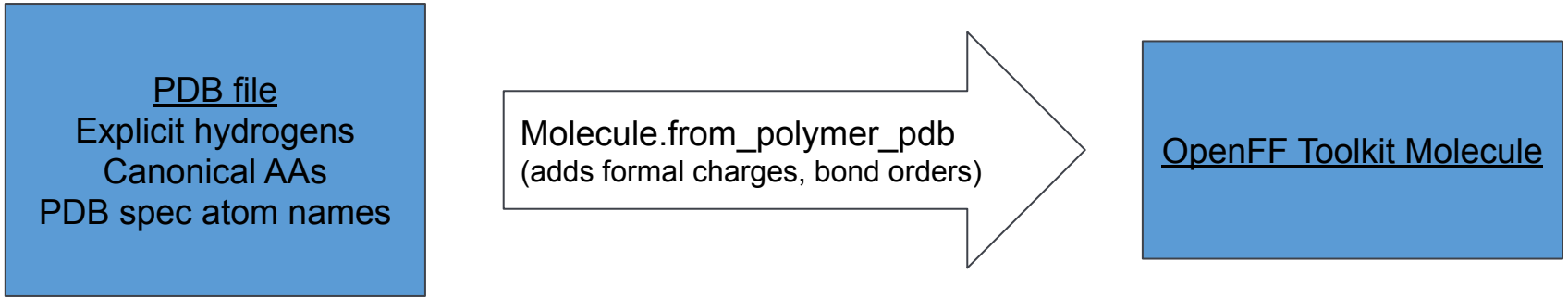


The last parameter to match a particular parent atom wins. Here the monovalent lone parameter would be assigned rather than the bond charge parameter as it appears later in the parameter list.

The F@H interface and OpenFE interoperability



OpenFF Toolkit Refactor - Supporting proteins

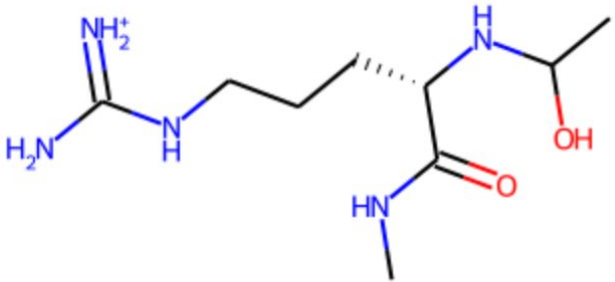



```
from openff.toolkit import Molecule, Topology
# Load structures and perceive info
# This PDB file must have explicit hydrogens
protein = Molecule.from_polymer_pdb('openff-toolkit/openff/toolkit/data/proteins/T4-protein.pdb')
```

Sometimes a toolkit will just... do the right thing

```
mol = Chem.MolFromPDBFile('openff-toolkit/openff/toolkit/data/proteins/MainChain_ARG.pdb')
mol.RemoveAllConformers()
mol
```

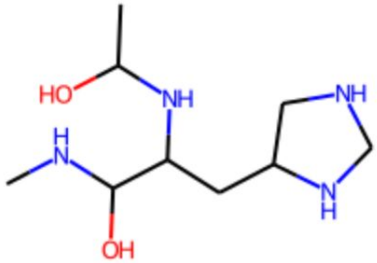
RDKit



But other times it won't

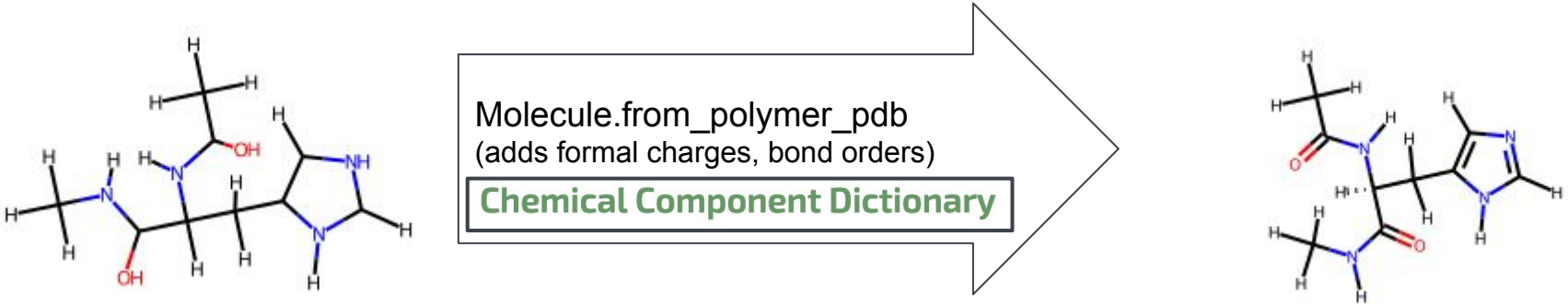
```
mol = Chem.MolFromPDBFile('openff-toolkit/openff/toolkit/data/proteins/MainChain_HID.pdb')
mol.RemoveAllConformers()
mol
```

RDKit



PDB files don’t contain complete chemical information!

CCD to the rescue!



The Chemical Component Dictionary is as an external reference file describing all residue and small molecule components found in PDB entries. This dictionary contains detailed chemical descriptions for standard and modified amino acids/nucleotides, small molecule ligands, and solvent molecules. Each chemical definition includes descriptions of chemical properties such as stereochemical assignments, chemical descriptors (SMILES & InChI), systematic chemical names, and idealized coordinates (generated using Molecular Networks' Corina, and if there are issues, OpenEye's OMEGA).

OFFMols are getting a PDB hierarchy API

```
# We offer lightweight residue and chain iteration functionality
protein.residues[:5]
```

```
[HierarchyElement ('A', '0', 'MET') of iterator 'residues' containing 19 atom(s),
 HierarchyElement ('A', '1', 'ASN') of iterator 'residues' containing 14 atom(s),
 HierarchyElement ('A', '2', 'ILE') of iterator 'residues' containing 19 atom(s),
 HierarchyElement ('A', '3', 'PHE') of iterator 'residues' containing 20 atom(s),
 HierarchyElement ('A', '4', 'GLU') of iterator 'residues' containing 15 atom(s)]
```

```
# The underlying atoms can be accessed using the "particles" iterator
[*protein.residues[4].atoms]
```

```
[Atom(name=N, atomic number=7),
 Atom(name=H, atomic number=1),
 Atom(name=CA, atomic number=6),
 Atom(name=HA, atomic number=1),
 Atom(name=CB, atomic number=6),
 Atom(name=HB2, atomic number=1),
 Atom(name=HB3, atomic number=1),
 Atom(name=CG, atomic number=6),
 Atom(name=HG2, atomic number=1),
 Atom(name=HG3, atomic number=1),
 Atom(name=CD, atomic number=6),
 Atom(name=OE1, atomic number=8),
 Atom(name=OE2, atomic number=8),
 Atom(name=C, atomic number=6),
 Atom(name=O, atomic number=8)]
```

PDB hierarchy info will be interoperable

```
# Here's an example of where hierarchy info goes when converting an OFFMol
# to OpenMM
omm_top = protein.to_topology().to_openmm()
omm_top.residues
```

```
<bound method Topology.residues of <Topology; 1 chains, 164 residues, 2634 atoms, 2654 bonds>>
```

```
[*omm_top.atoms()][100]
```

```
<Atom 100 (H51x) of chain 0 residue 5 (MET)>
```

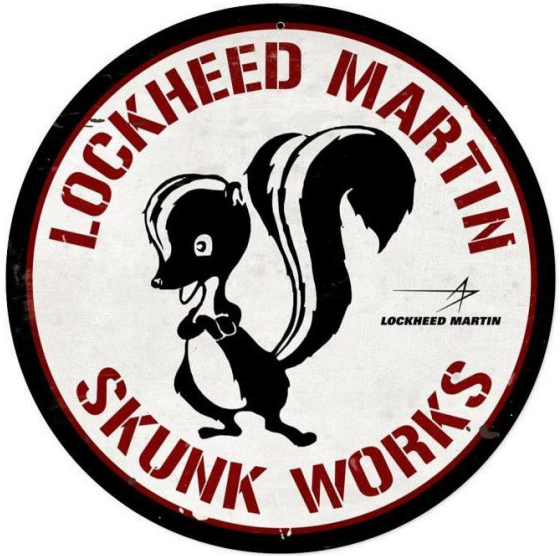
```
# Here's where hierarchy info goes in RDKit
rd_protein = protein.to_rdkit()
res_info = rd_protein.GetAtomWithIdx(100).GetPDBResidueInfo()
print(res_info.GetResidueName(), res_info.GetResidueNumber(), res_info.GetChainId())
```

```
MET 5 A
```

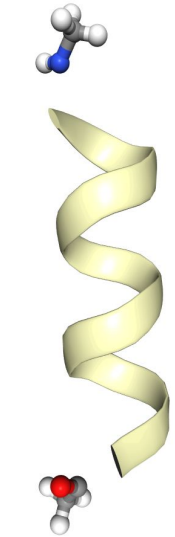
The power of SMIRNOFF

🚧 **This code is not production-ready**

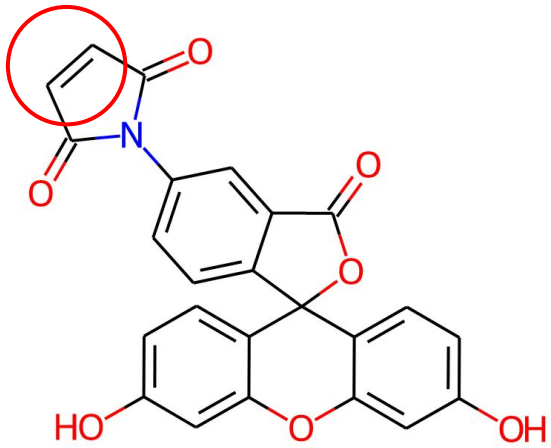
This code is a glimpse into the possibilities of the kinds of automated NCAA parametrization that future OpenFF force fields will provide. We present it as a preview, not as a recommended procedure. It uses unproven force fields and a lot of ad hoc code. We recommend waiting for OpenFF 3.0.0 Rosemary.



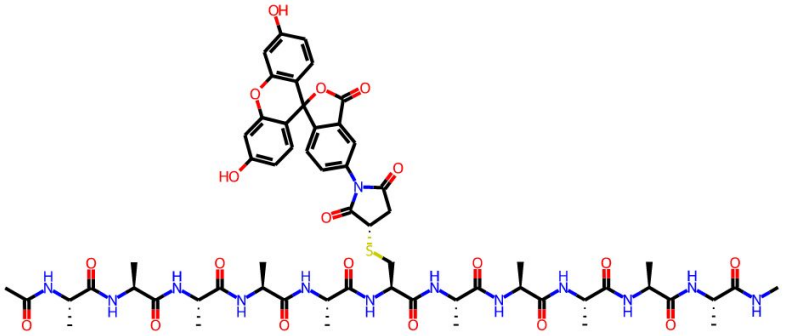
The power of SMIRNOFF



AAAAACAAAAA



Fluorescein-5-maleimide



The power of SMIRNOFF - The fine print

SMIRNOFF parameter specification is hierarchical

Parameters that appear later in a SMIRNOFF specification override those which come earlier if they match the same pattern... This hierarchical structure means that **a typical parameter file will tend to have generic parameters early in the section for each force type, with more specialized parameters assigned later.**

Multiple SMIRNOFF representations can be processed in sequence

Multiple SMIRNOFF data sources (e.g. multiple OFFXML files) can be loaded in sequence. If these files each contain unique top-level tags (such as <Bonds>, <Angles>, etc.), the resulting force field will be independent of the order in which the files are loaded. **If, however, the same tag occurs in multiple files, the contents of the tags are merged, with the tags read later taking precedence over the parameters read earlier, provided the top-level tags have compatible attributes.** The resulting force field will therefore depend on the order in which parameters are read.

This behavior is intended for limited use in appending very specific parameters, such as parameters specifying solvent models, to override standard parameters.

Franken-Force field

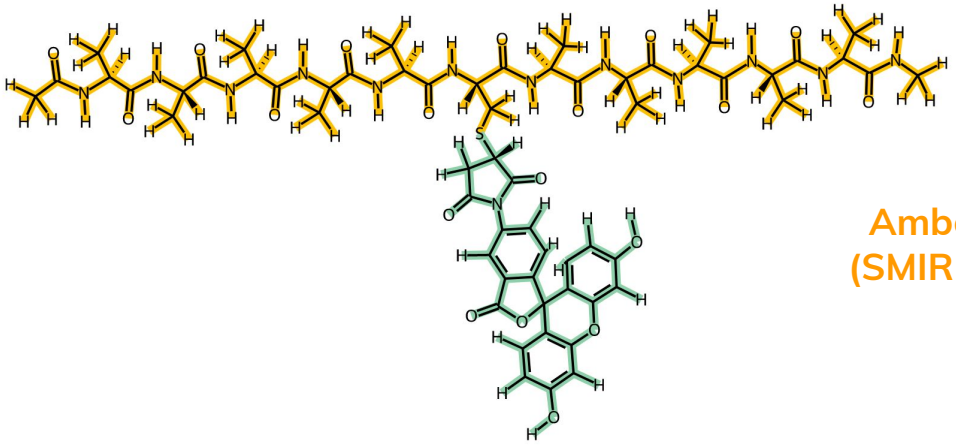
🚧 **This code is not production-ready**

In this notebook, we use the SMIRNOFF port of the Amber ff14SB force field. This is currently the only mainstream protein force field that works with the OpenFF Toolkit, but it's slow and hasn't been rigourously checked against the original force field. The Open Force Field Initiative recommends waiting for OpenFF 3.0.0 "Rosemary", which will include protein parameters, before using this in production work.

```
sage_ff14sb = ForceField('openff-2.0.0.offxml', 'ff14sb_off_impropers_0.0.3.offxml')
```

The power of SMIRNOFF

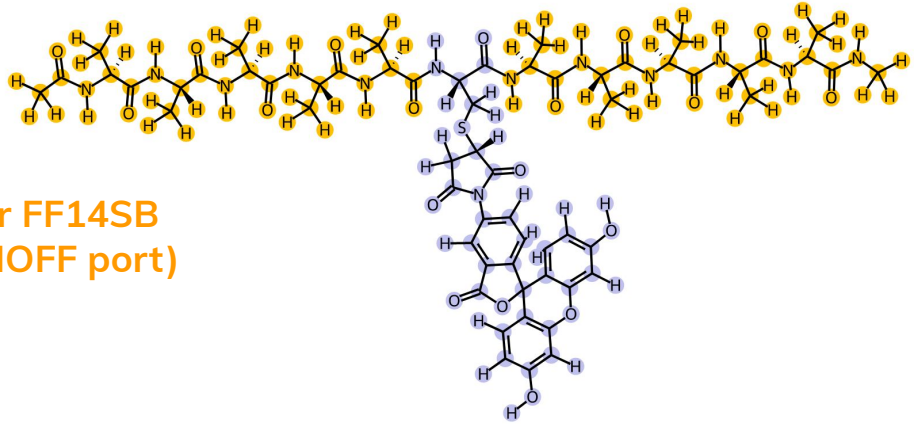
Bonds



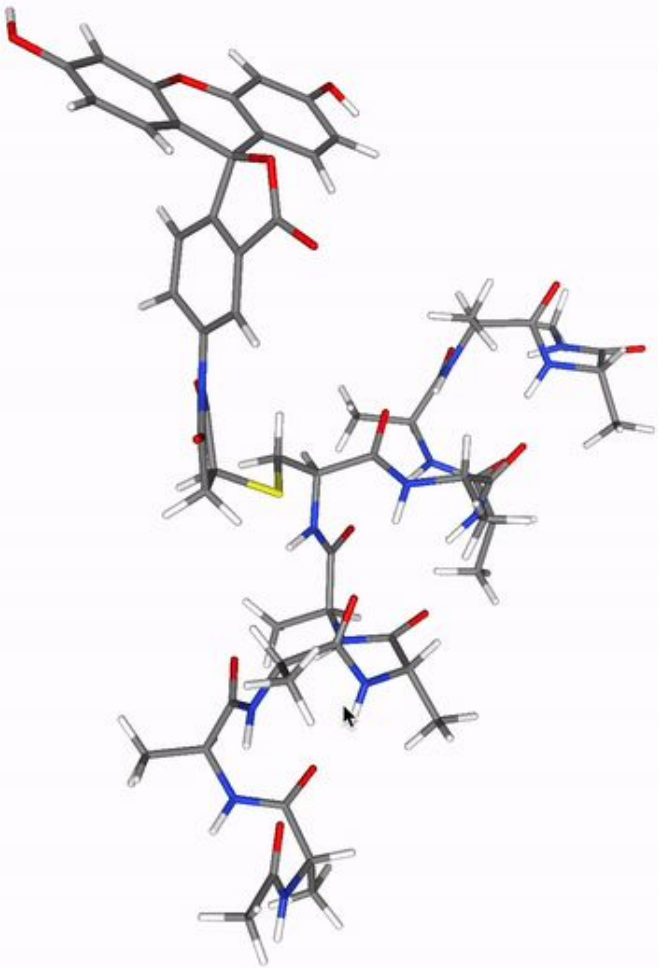
Sage

Amber FF14SB
(SMIRNOFF port)

Partial charges



AM1BCC
(some assembly required)



Where we've been

... and where we're going

2016

SMIRNOFF99Frosst

Initial SMIRNOFF port of the parm@Frosst force field

2019

Parsley

Retrained valence parameters against a redesigned QC data set

2021

Sage

Retrained vdW parameters against physical property data + retrained valence parameters

202X

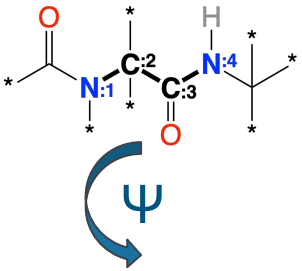
Rosemary

Self consistent **biopolymer** + small molecule force field

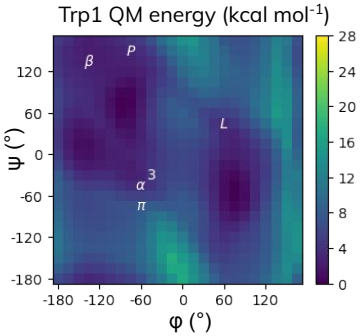
Rosemary

A protein force field

openff toolkit



[#6X3](=O)-[#7X3:1]-[#6X4:2]
-[#6X3:3](=O)-[#7X3H1:4]-[#6X4]



Iván Pulido

Jeff Wagner

Chapin Cavender

Software

Adding biopolymer support to OpenFF Toolkit

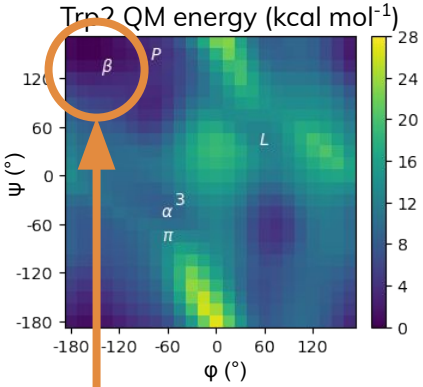
Selecting parameters

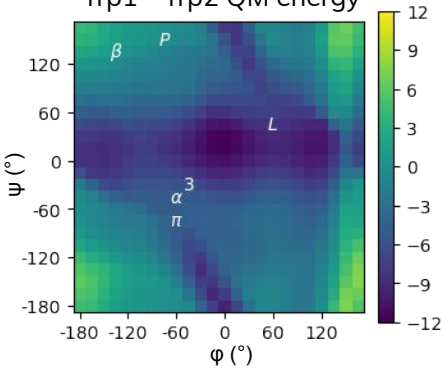
Encoding protein-specific parameters with SMARTS

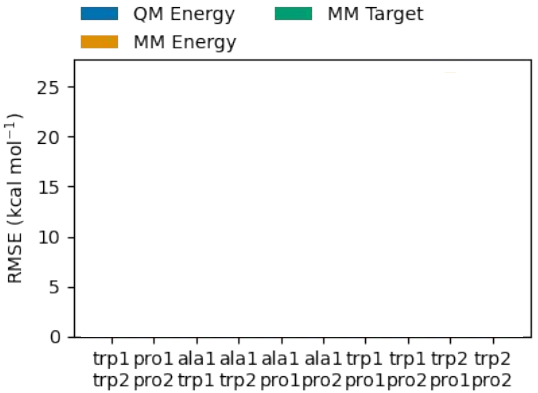
QC Data

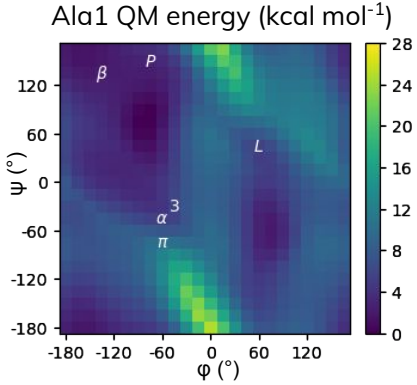
Systematic generation of torsion and optimization data for amino acids

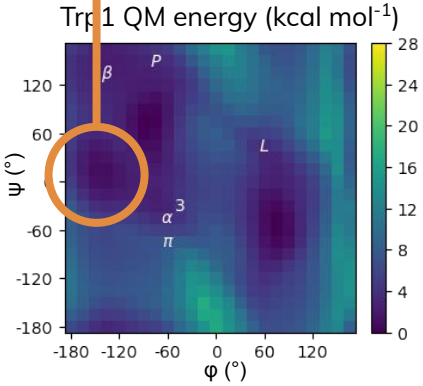
Do side-chains impact protein backbone torsions?

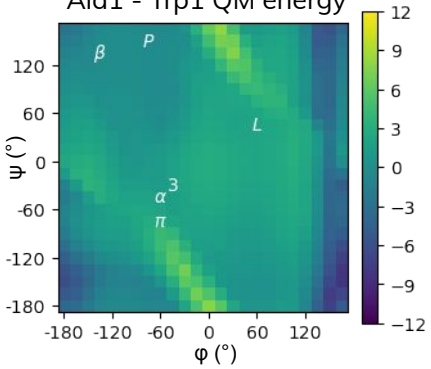












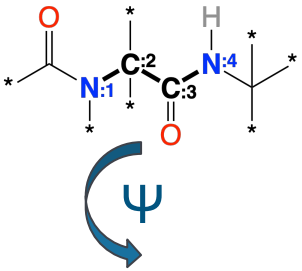
Yes → train individual torsions for each side-chain

Use different rotamers for validation

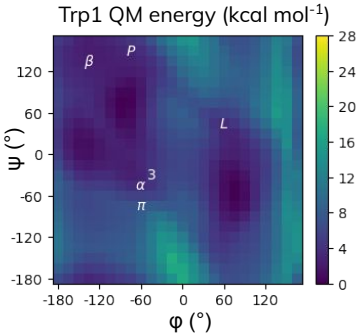
Rosemary

A protein force field

openff toolkit



[#6X3](=O)-[#7X3:1]-[#6X4:2]
-[#6X3:3](=O)-[#7X3H1:4]-[#6X4]




Iván Pulido


Jeff Wagner


Chapin Cavender

FOLDING @HOME

Software

Adding biopolymer support to OpenFF Toolkit

Selecting parameters

Encoding protein-specific parameters with SMARTS

QC Data

Systematic generation of torsion and optimization data for amino acids

Fitting

Fit protein-specific torsions to QC data, and re-fit valence and LJ parameters

Benchmarking

Curate and benchmark to NMR observables

Where we've been

... and where we're going

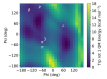
Force field fitting

VALENCE

VDW

ELECTROSTATICS

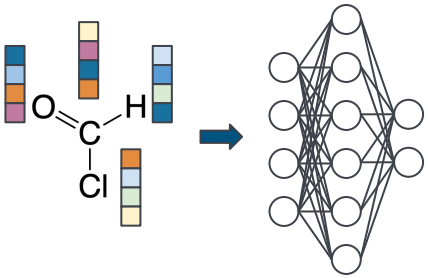
Protein parameters



Graph charge models

Graph charges

Neural networks for fast electrostatics

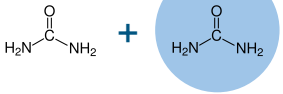


PyTorch

+

openff toolkit

ESP / EF



RESP2 δ=0.6

Simon Boothroyd

Hypothesis → **Software** ⫽→ **Training** → **Testing**

A graph convolutional model can produce fast, conformer-independent, high quality charges

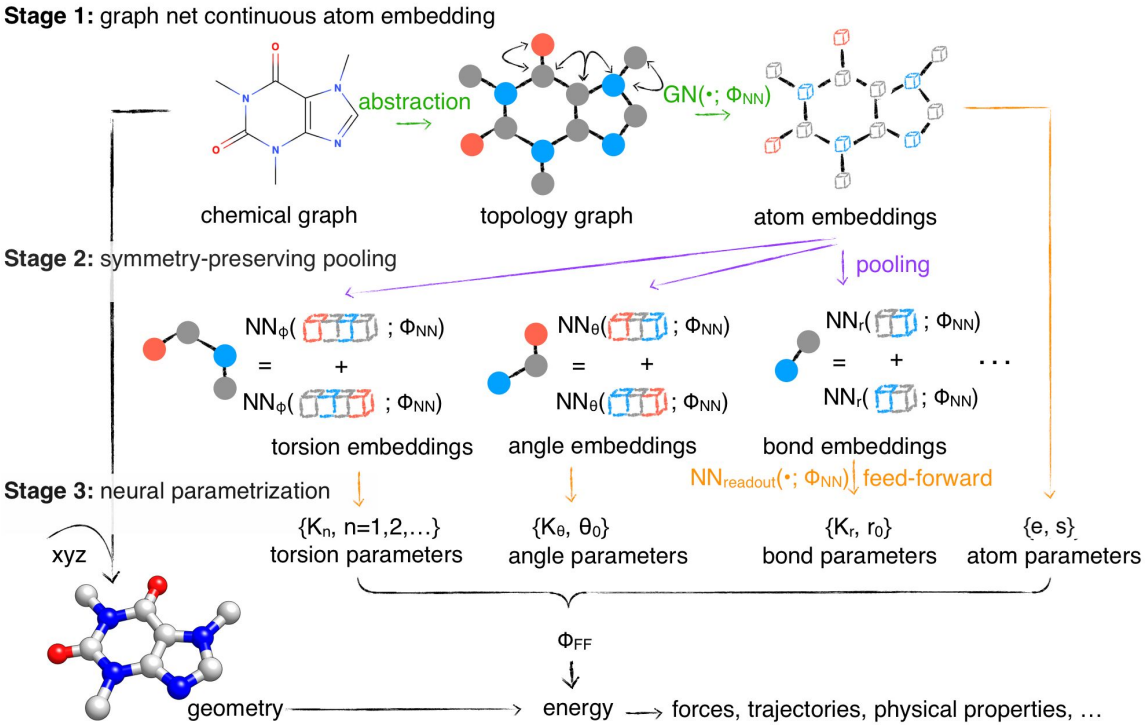
Support for graph models added to the OpenFF toolkit. Infrastructure for applying graph models to molecules created

A GNN will be trained to similar quality as AM1BCC (ELF10)

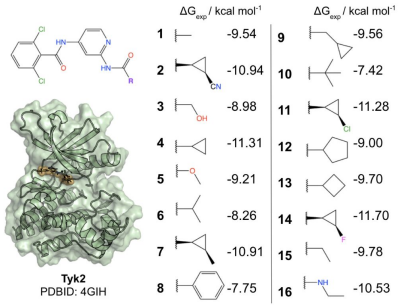
Trained parameters will be benchmarked against experimental and physical property data

Espaloma

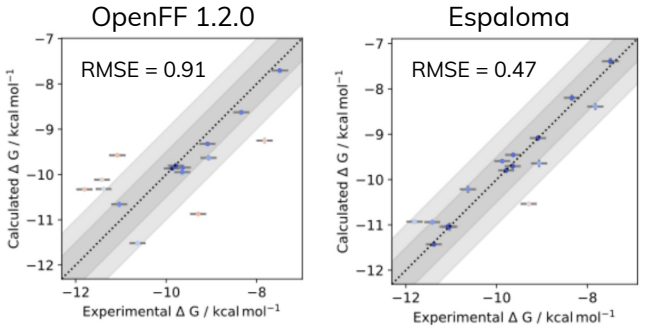
Neural networks for force field parameters



Absolute binding free energy performance



Yuanqing Wang



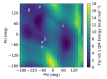
Where we've been

... and where we're going

Force field fitting

VALENCE

Protein parameters



VDW

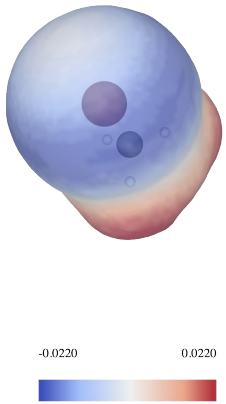
Virtual sites

ELECTROSTATICS

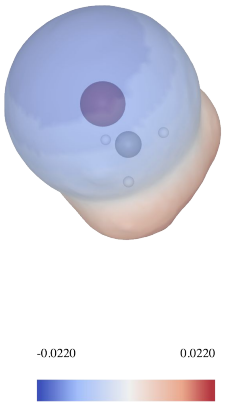
Graph charge models

ESP of C-Br has a sigma hole at HF/6-31G*

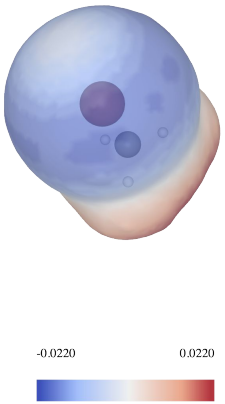
HF/6-31G*



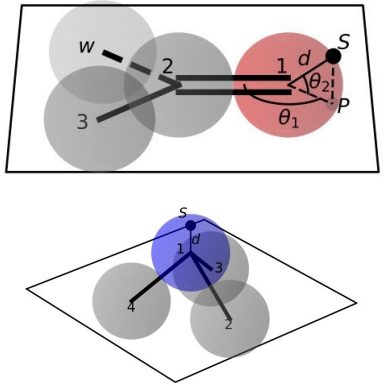
Sage



With virtual sites



Virtual sites



+

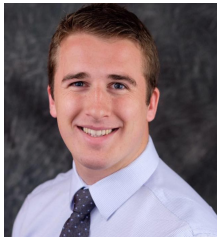
openff
toolkit

ESP / EF

+

RESP2 δ=0.6

Simon Boothroyd



Owen Madin

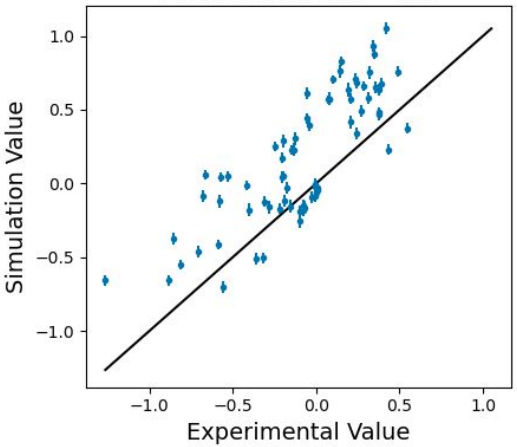
Trevor Gokey

Hypothesis → **Software** ⫽→ **Training** → **Testing** → **Interoperability**

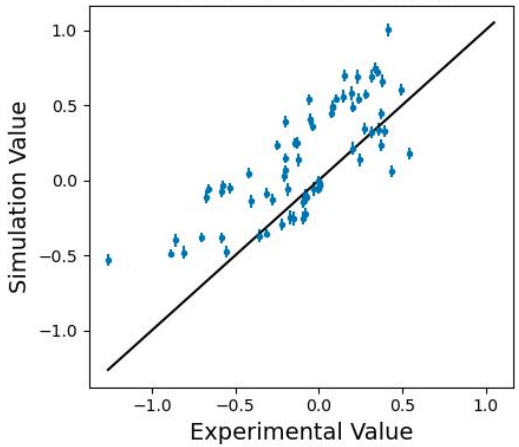
Hypothesis	Software	Training	Testing	Interoperability
The inclusion of off-site charges should improve the accuracy of a force fields electrostatic interactions	Virtual site support added to the OpenFF toolkit. Support for training to QC ESP + EF data	Virtual sites will be trained against ESP / EF QC data, based on input from the Cole group	Trained parameters will be benchmarked against experimental and physical property data	In order to include into mainline force field need major simulation packages to support proposed v-sites

$\Delta H_{mix}(x)$ of mixtures containing Chlorine

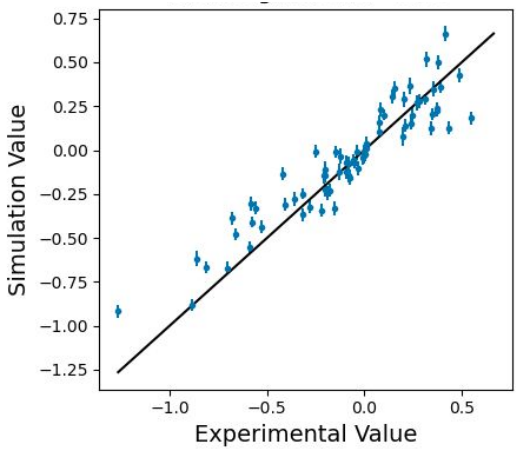
OpenFF 1.3.0



OpenFF 2.0.0



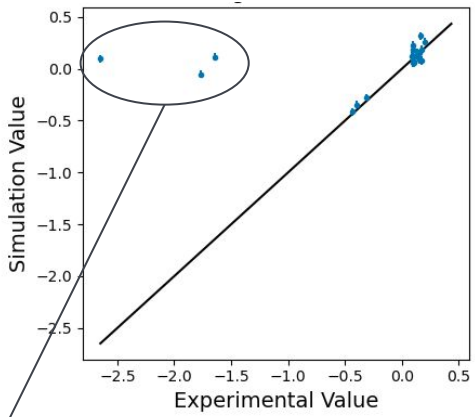
FF + new BCCs / v-sites



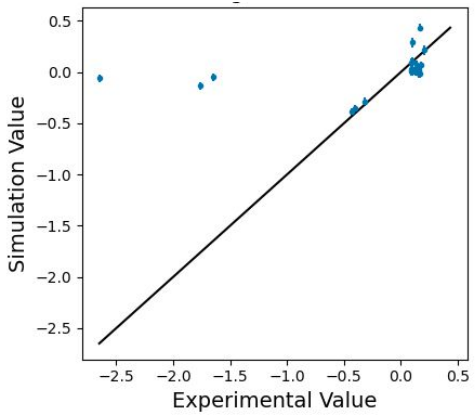
Proof of concept: halogens and pyridines

$\Delta H_{mix}(x)$ of Heteroaromatics

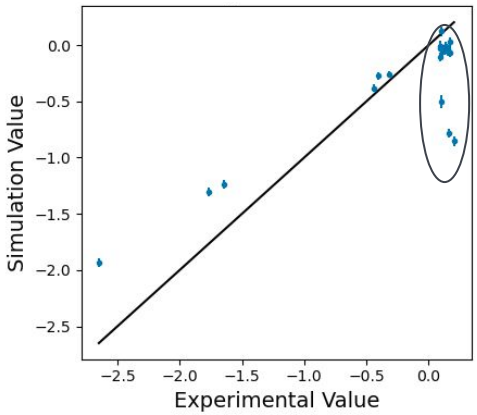
OpenFF 1.3.0

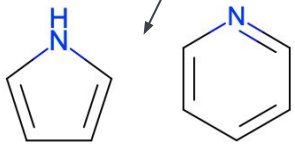


OpenFF 2.0.0

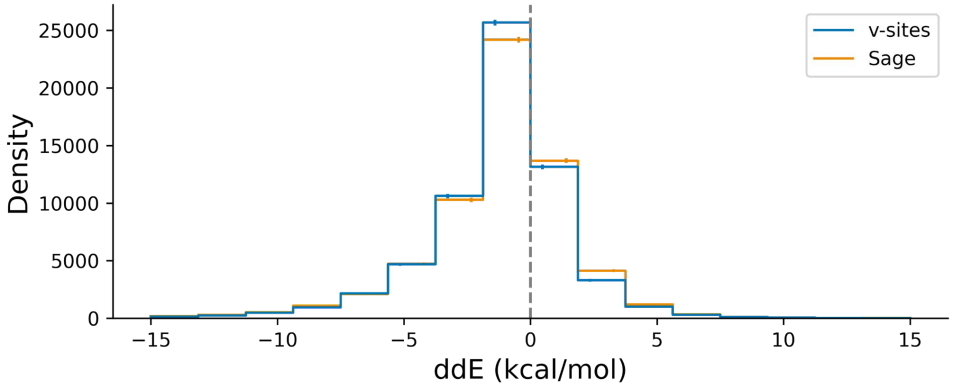


FF + new BCCs / v-sites

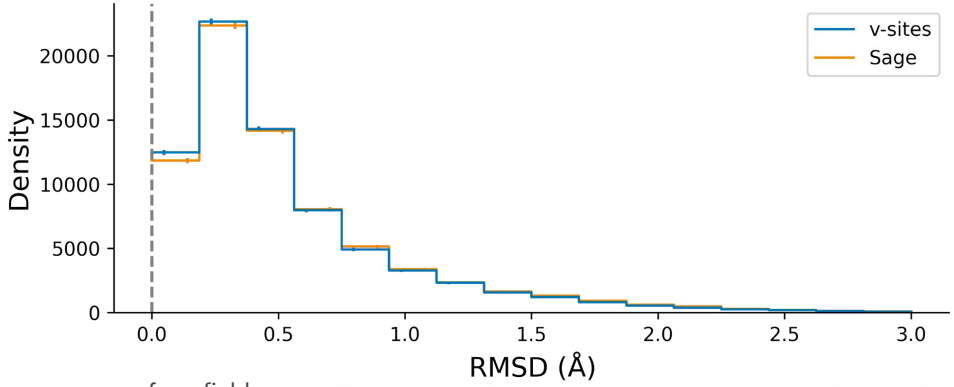


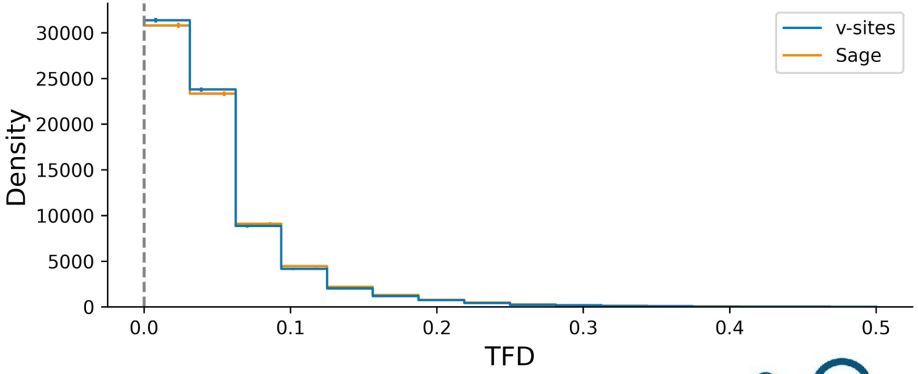


Proof of concept: halogens and pyridines



Benchmarks show modest improvements





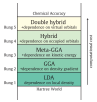
Where we've been

... and where we're going

Force field fitting

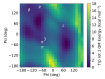
INPUT DATA

QM theory benchmark



VALENCE

Protein parameters



VDW

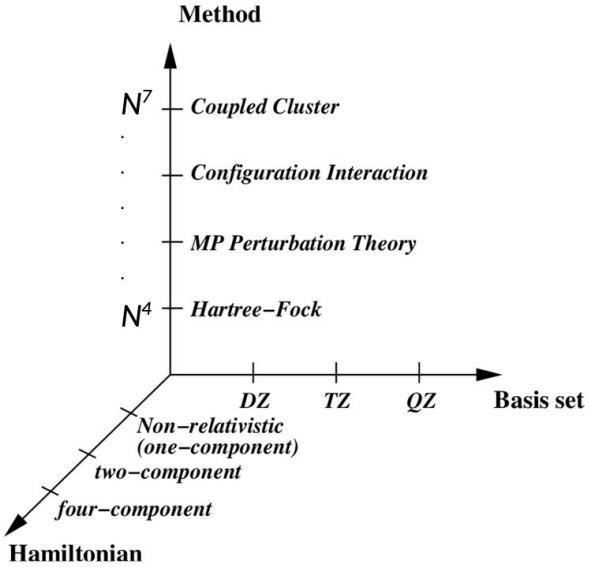
ELECTROSTATICS

Virtual sites

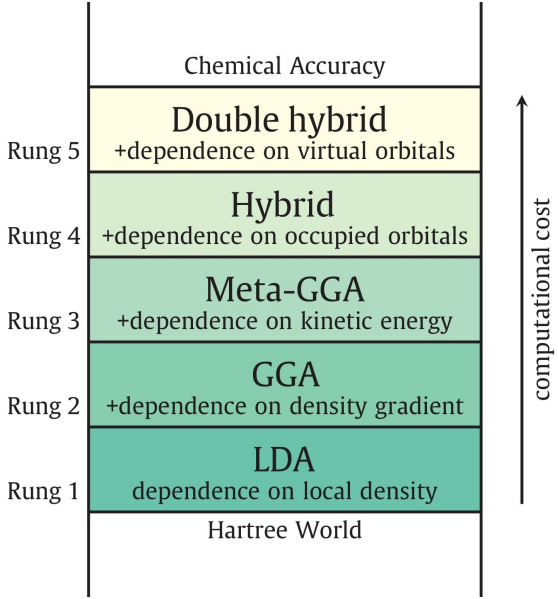
Graph charge models

QM data theory

Balancing accuracy and speed



The three axes along of WFT methods:
Basis set, **Hamiltonian**, and the treatment of **electron correlation**.
Courtesy: Timo Fleig, urn:nbn:de:hbz:061-20070312-091913-8

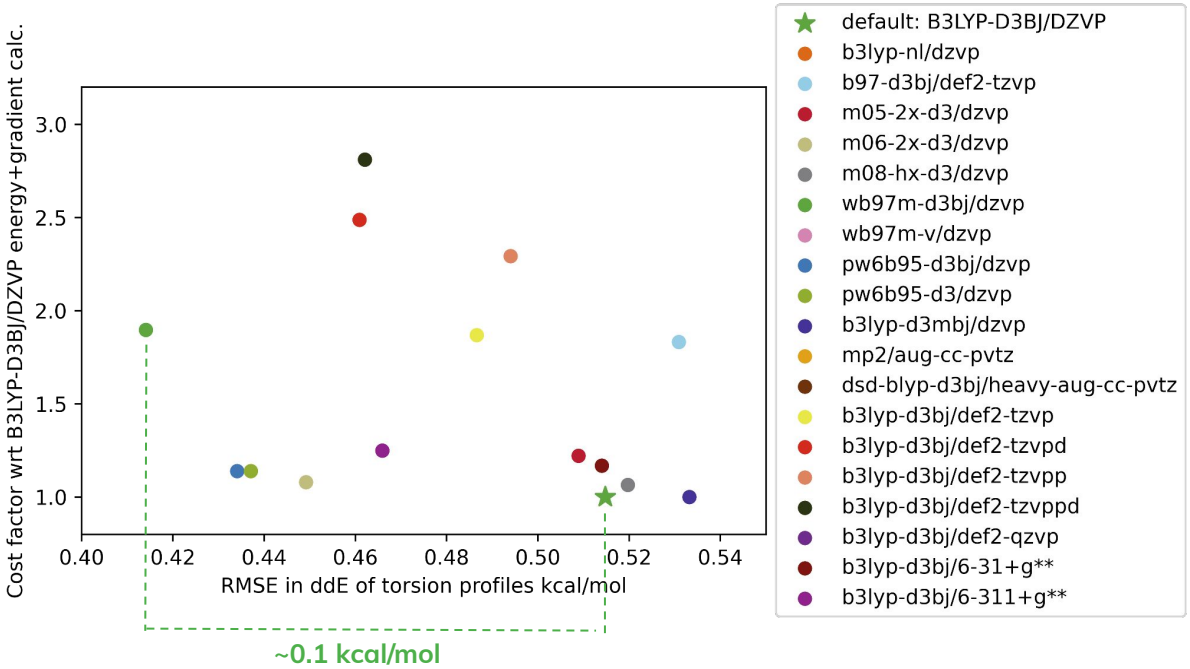


Jacob's ladder of density functional approximation for **exchange correlation** energy
Courtesy: 10.1016/j.ccr.2015.03.019

Pavan Behara

Hyesu Jang

B3LYP-D3BJ/DZVP best balance of quality and speed



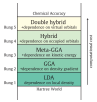
Where we've been

... and where we're going

Force field fitting

INPUT DATA

QM theory benchmark

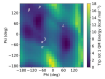


VALENCE

Modified Seminario Method for initial values

Torsions in QC geometry targets

Protein parameters



VDW

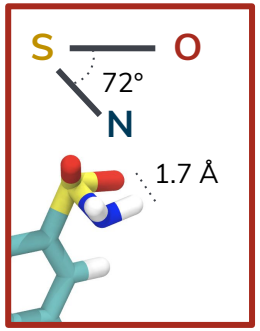
ELECTROSTATICS

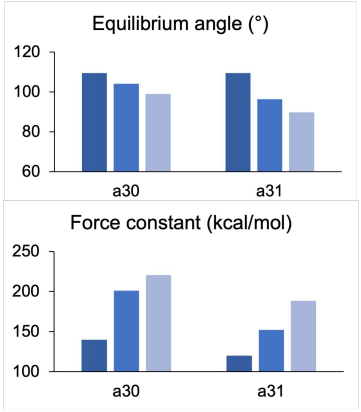
Virtual sites

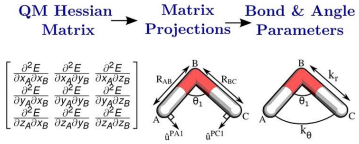
Graph charge models

Fitting protocol

Modified Seminario Method for initial values







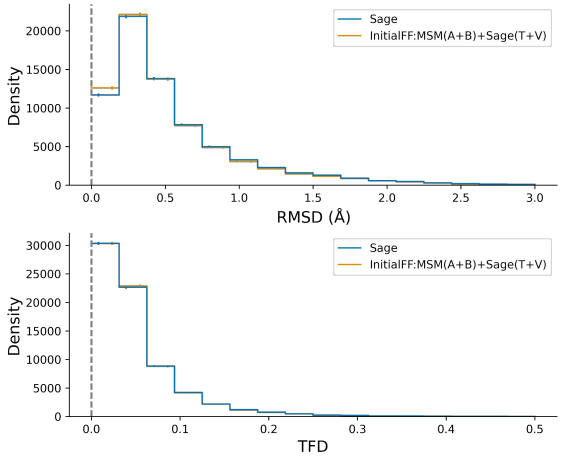
Josh Horton

Pavan Behara

Problem
incorrect sulfonamide valence angles in simulation

Diagnosis
equilibrium valence angles **decreased unphysically** between releases

Proposal
Use modified Seminario method to derive initial bond and angle values from QM Hessian matrix

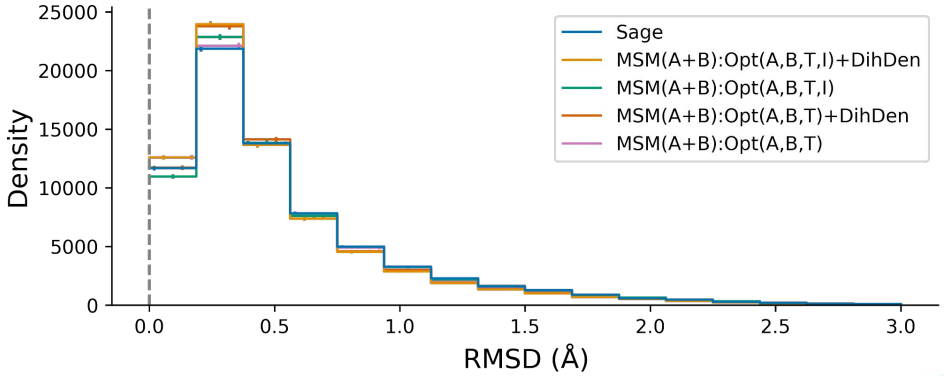


Fitting protocol

Updates for better fits

- ***Fitting metrics***: including dihedral deviations fitting to QM geometries [**DihDen**]
- ***Starting points***: Modified Seminario Method for angles and bonds [**MSM(A+B)**]
- Changing both and optimizing all valence parameters [**MSM(A+B):Opt(A, B, T, I)+DihDen**] shows improvements in **RMSD** and **TFD**

Pavan Behara



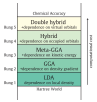
Where we've been

... and where we're going

Force field fitting

INPUT DATA

QM theory benchmark

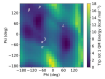


VALENCE

Modified Seminario Method for initial values

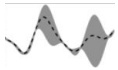
Torsions in QC geometry targets

Protein parameters



VDW

Surrogate modelling



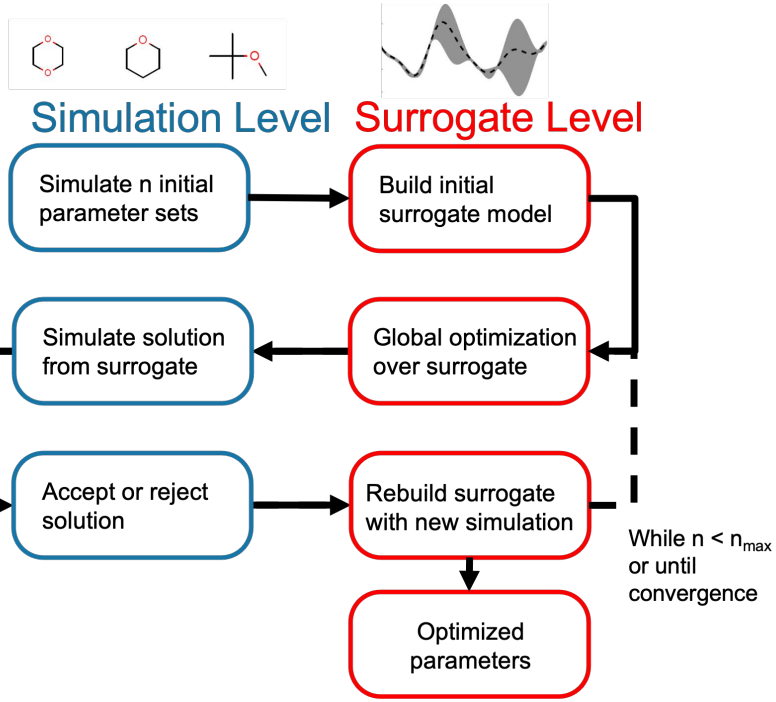
Virtual sites

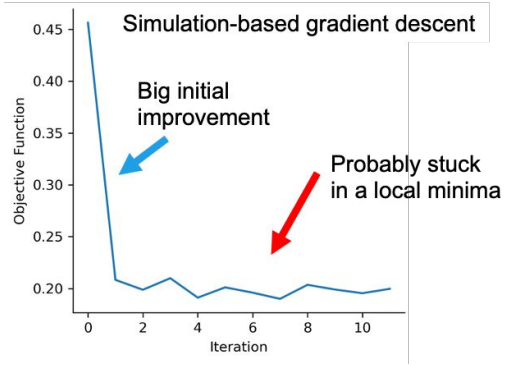
ELECTROSTATICS

Graph charge models

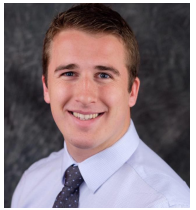
Surrogate modelling

Escaping local minima with global optimization

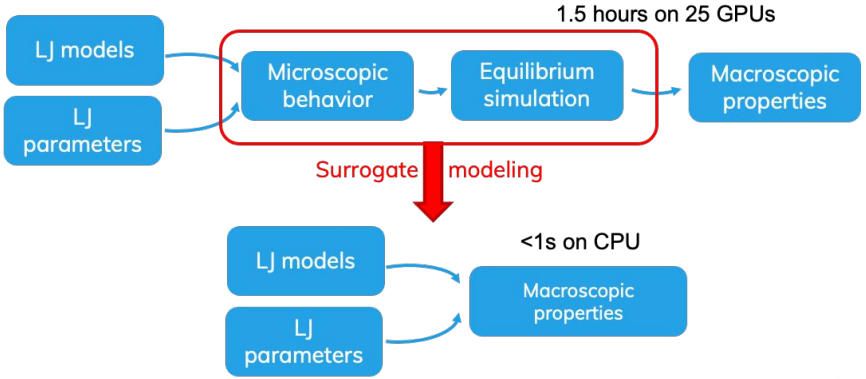




Simon Boothroyd

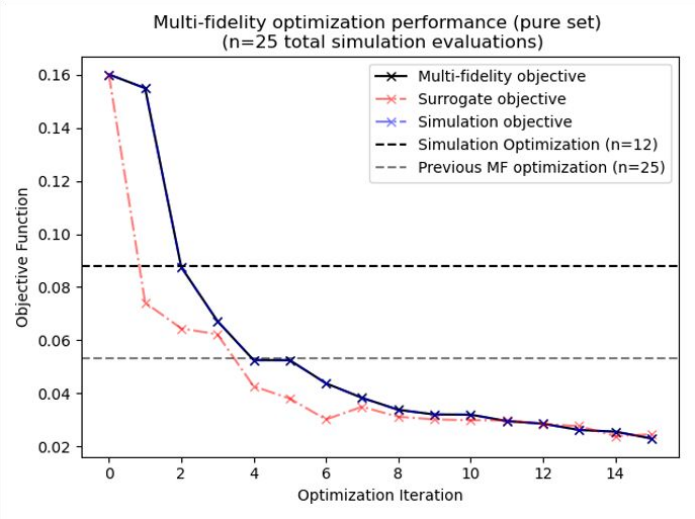


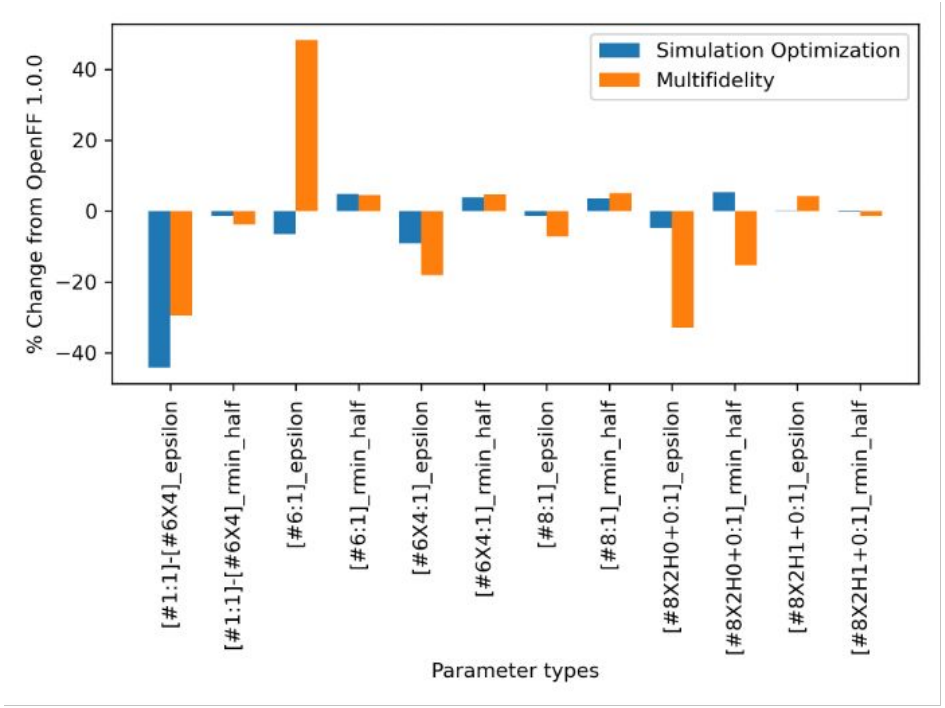
Owen Madin



Surrogate modelling

Escaping local minima with global optimization





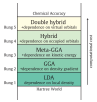
Where we've been

... and where we're going

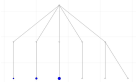
Force field fitting

INPUT DATA

QM theory benchmark



Automated chemical perception

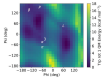


VALENCE

Modified Seminario Method for initial values

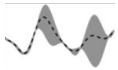
Torsions in QC geometry targets

Protein parameters



VDW

Surrogate modelling

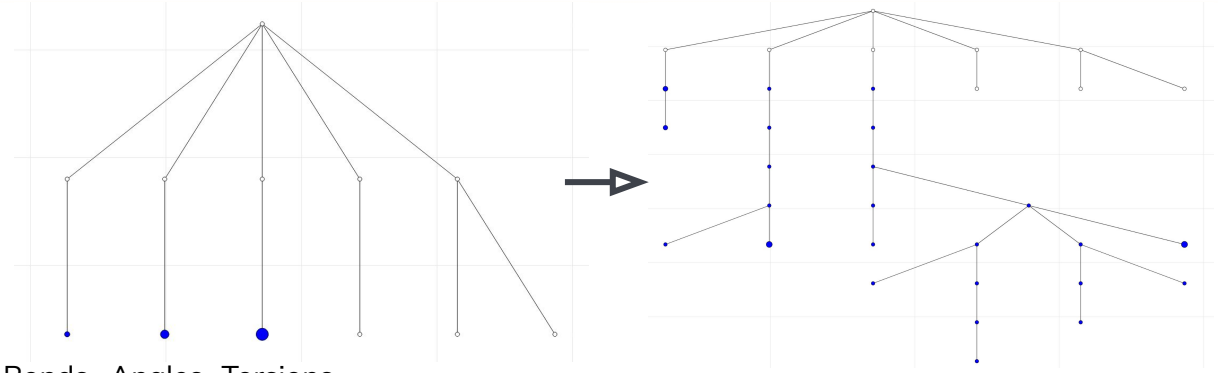


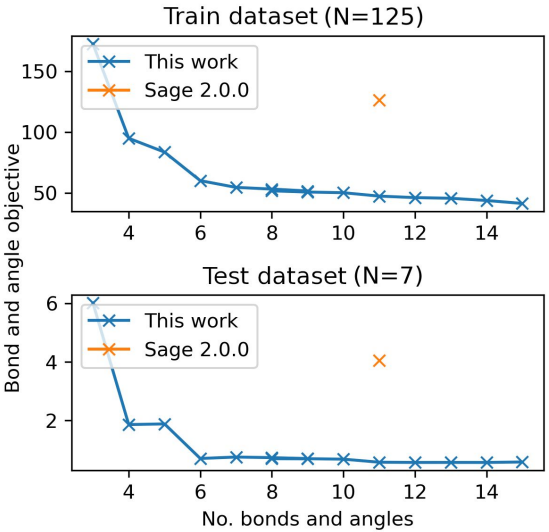
ELECTROSTATICS

Virtual sites

Graph charge models

Automated chemical perception





Trevor Gokey

Find new parameters based on QM geometry (bonds, angles, torsions)

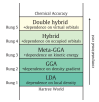
Where we've been

... and where we're going

Force field fitting

INPUT DATA

QM theory benchmark

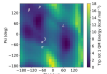


VALENCE

Modified Seminario Method for initial values

Torsions in QC geometry targets

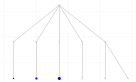
Protein parameters



VDW

ELECTROSTATICS

Automated chemical perception

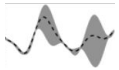


WBO interpolated improper torsions

Internal coordinate Hessian fitting

Explicit torsion multiplicities

Surrogate modelling



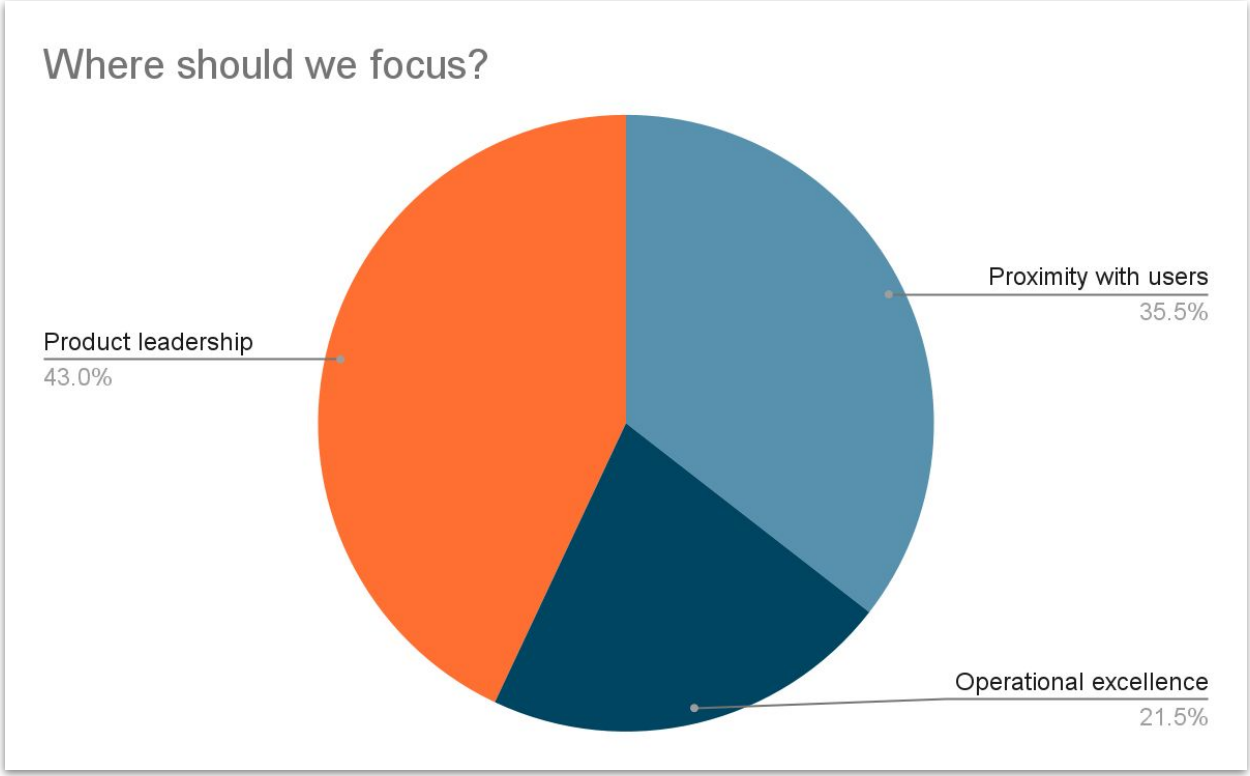
Virtual sites

Graph charge models

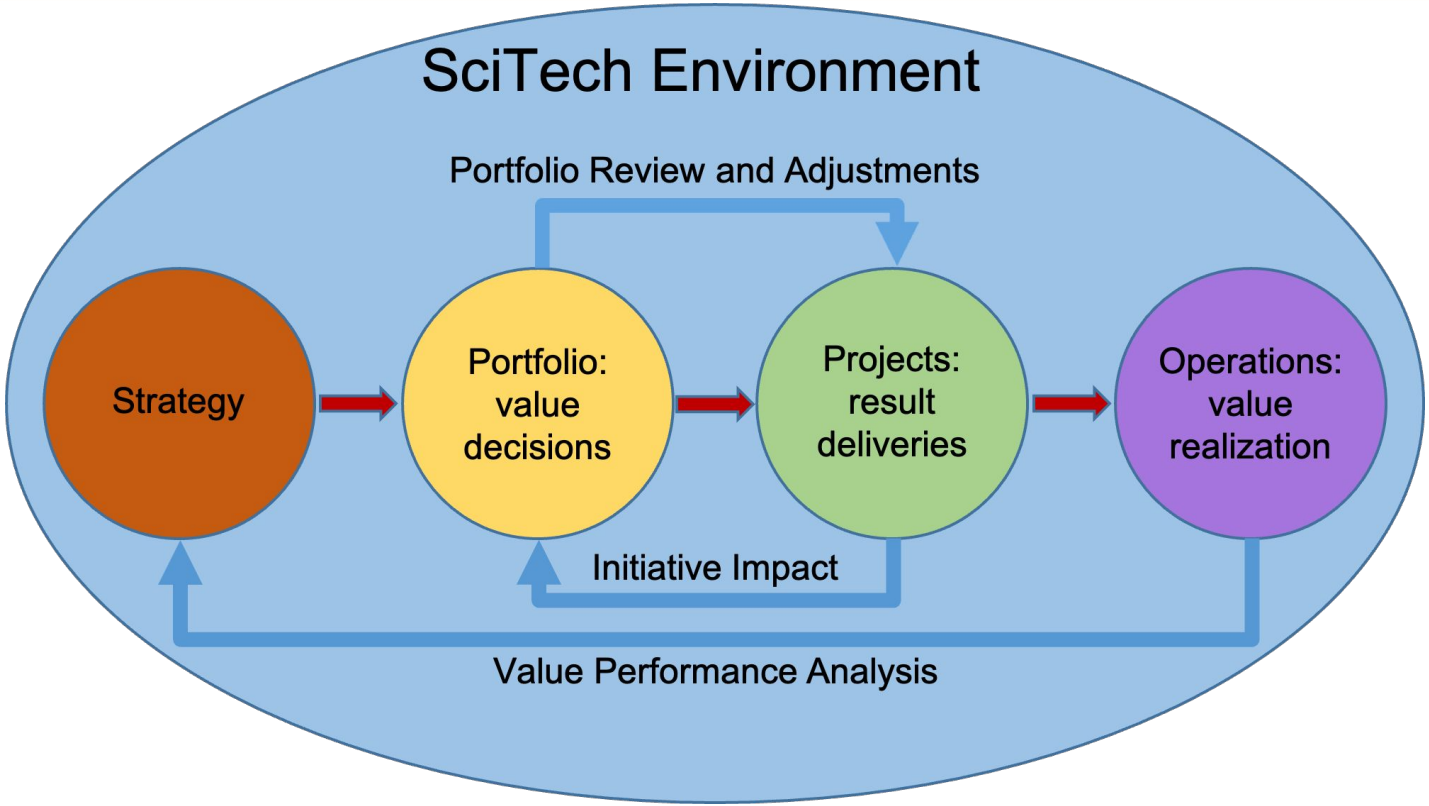
Where we've been ... where we're going ... where we can go

SMIRNOFF99Frosst Parsley **Sage** Rosemary OpenFF

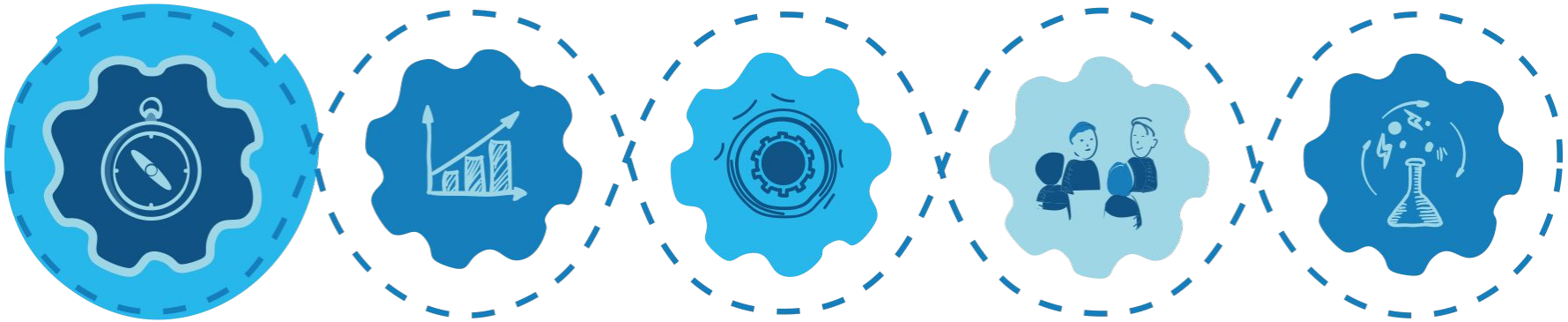
What are our organizational value disciplines?



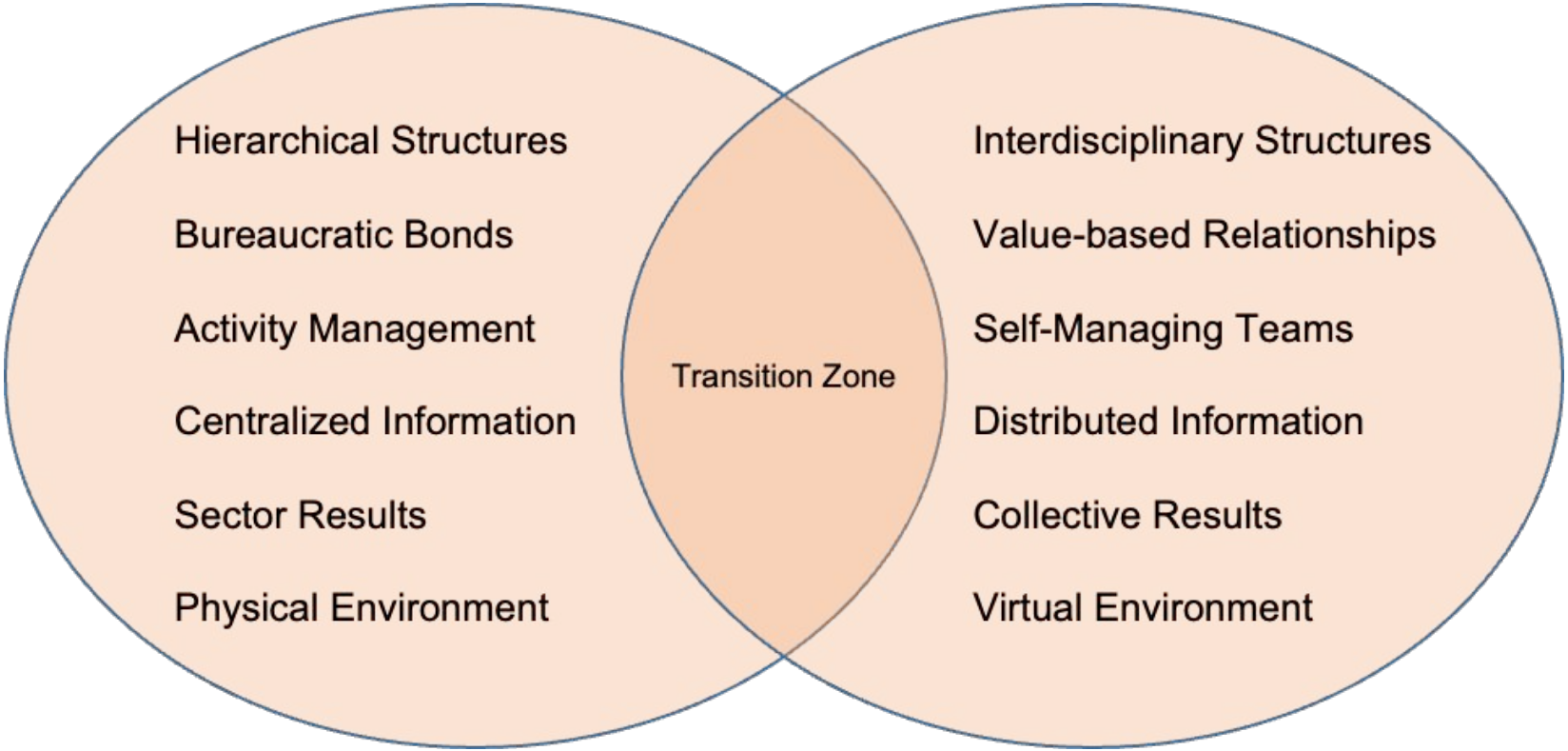
Why should we have a strategy?



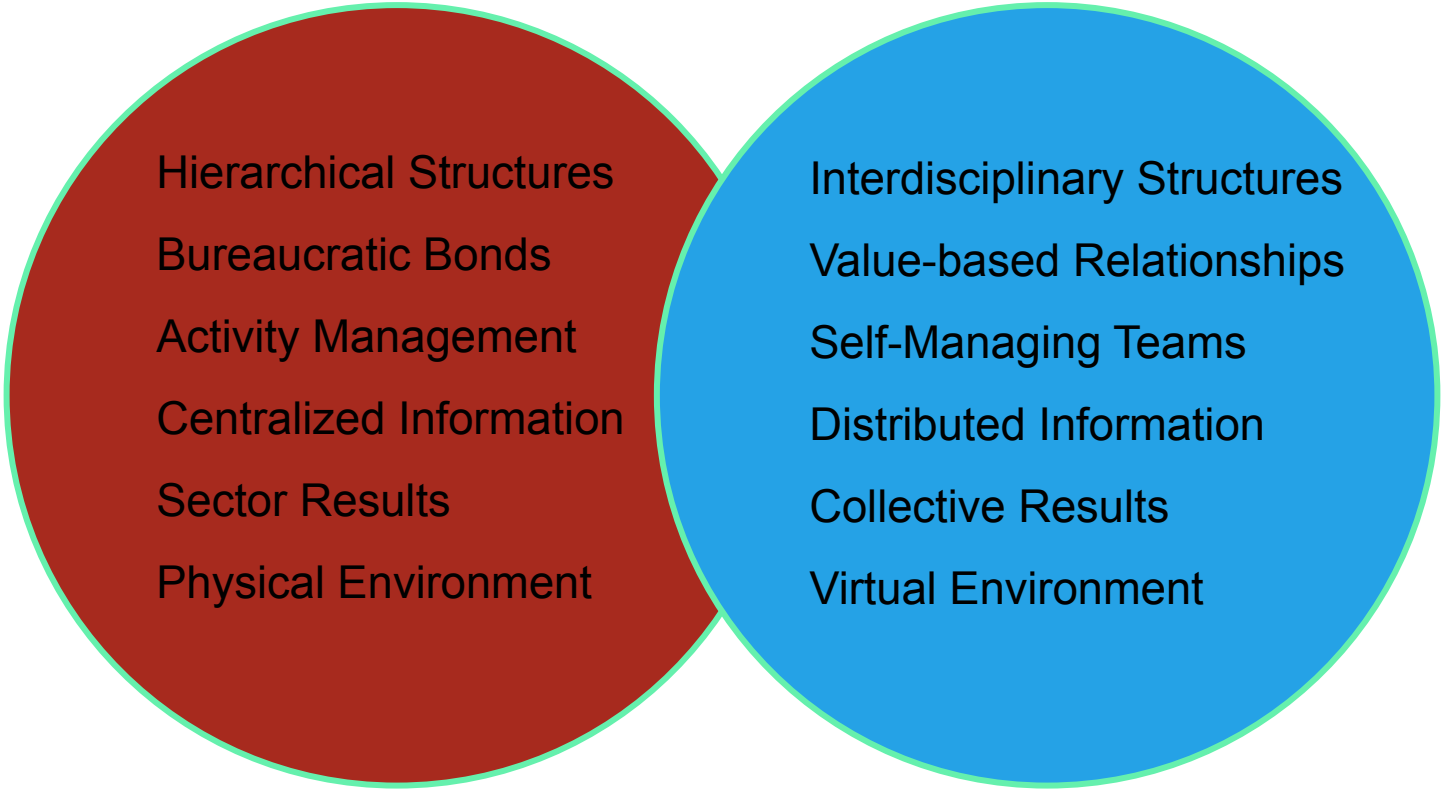
What is a project for?



How does strategy affect our daily lives?



How does strategy affect our daily lives?



Where we've been

... where we're going

... where we can go together

SMIRNOFF99Frosst

Parsley

Sage

Rosemary

OpenFF

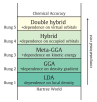
Where we've been

... and where we're going

Force field fitting

INPUT DATA

QM theory benchmark

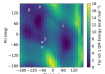


VALENCE

Modified Seminario Method for initial values

Torsions in QC geometry targets

Protein parameters



VDW

ELECTROSTATICS

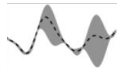
Automated chemical perception

WBO interpolated improper torsions

Internal coordinate Hessian fitting

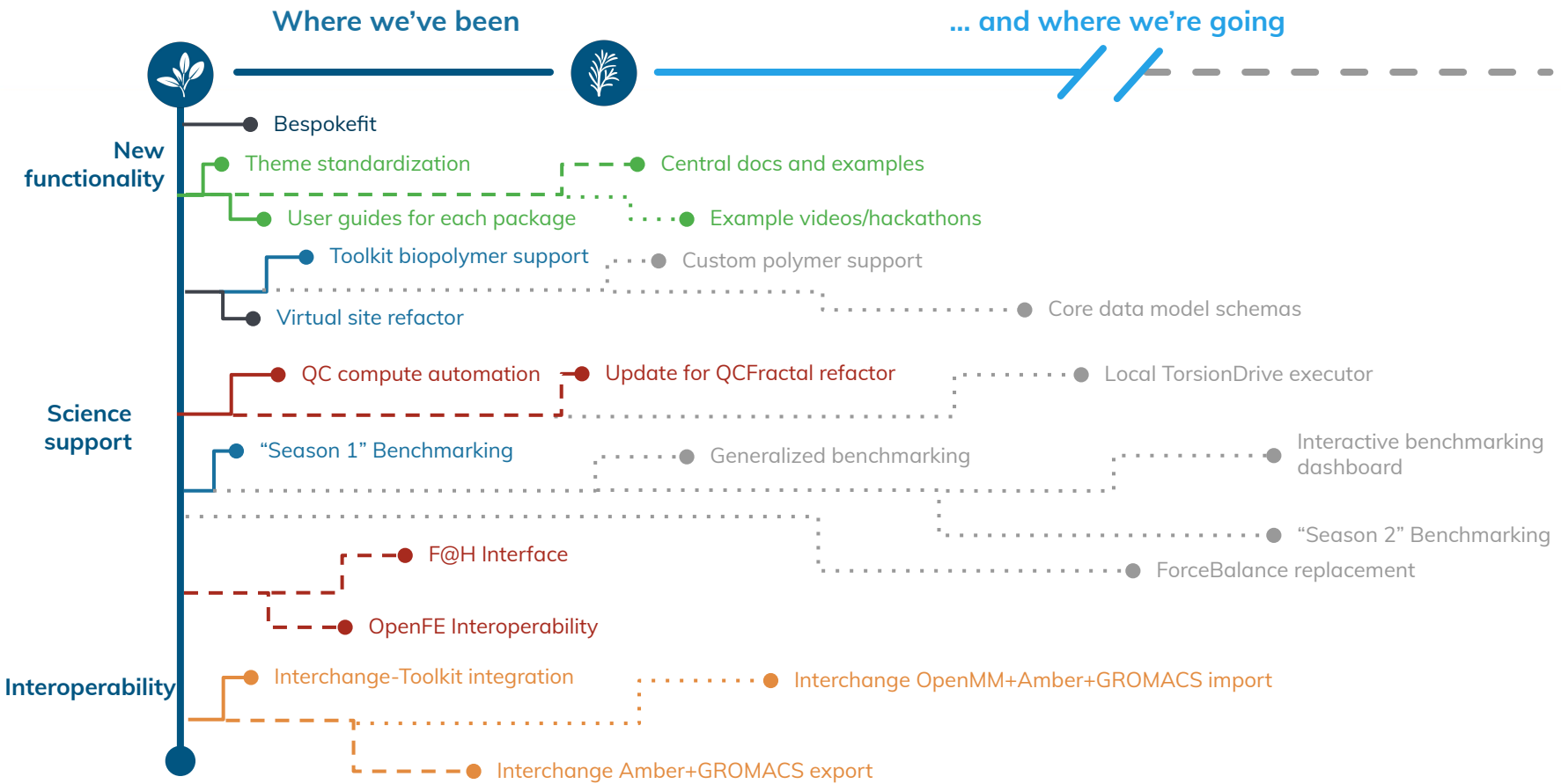
Explicit torsion multiplicities

Surrogate modelling



Virtual sites

Graph charge models


Where we've been
... and where we're going
New functionality
Bespokefit
Theme standardization
Central docs and examples
User guides for each package
Example videos/hackathons
Toolkit biopolymer support
Custom polymer support
Virtual site refactor
Core data model schemas
Science support
QC compute automation
Update for QCFractal refactor
Local TorsionDrive executor
"Season 1" Benchmarking
Generalized benchmarking
Interactive benchmarking dashboard
"Season 2" Benchmarking
F@H Interface
ForceBalance replacement
OpenFE Interoperability
Interoperability
Interchange-Toolkit integration
Interchange OpenMM+Amber+GROMACS import
Interchange Amber+GROMACS export

Discussion questions

- Who is interested in non-protein polymer support in force fields?
- What should be our best practice for combining parameterized components before Interchange can import from external formats? ParmEd or OpenMMForceFields?
- Where should we plan to go after graph charges and virtual sites?
- What should OpenFF look like in 3 years? How many people? Continued focus on core products or keep developing new products?