

CHODERA LAB

HOW CAN BIOPHYSICS PLAY A ROLE IN THE ERA OF GENOMICS AND BIOMEDICAL BIG DATA?

MODELING



FOLDING
@HOME

amazon
webservices EC2

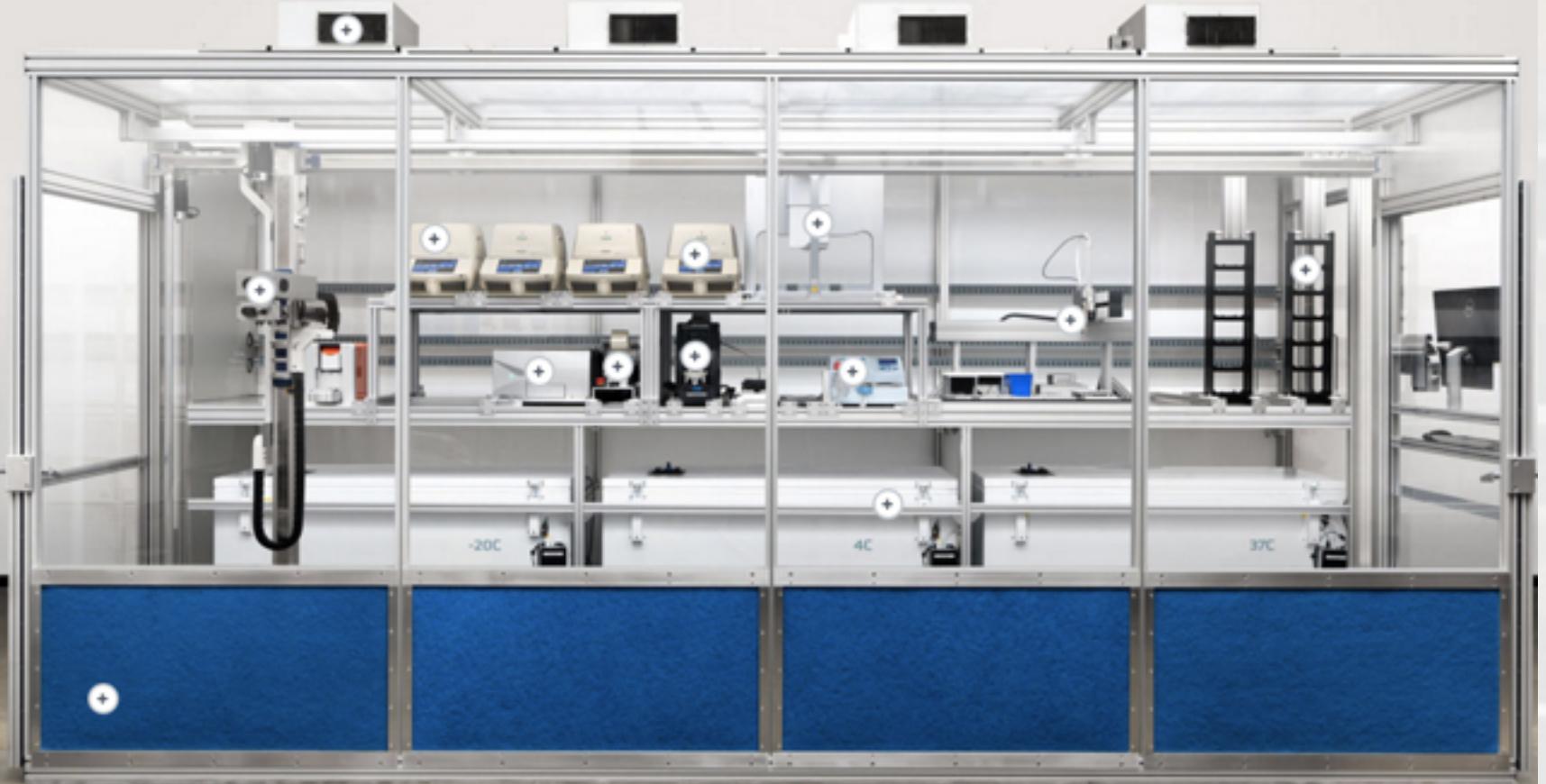
$$V(\mathbf{q}) = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2$$

$$+ \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

AUTOMATION



CHODERA LAB, Z17

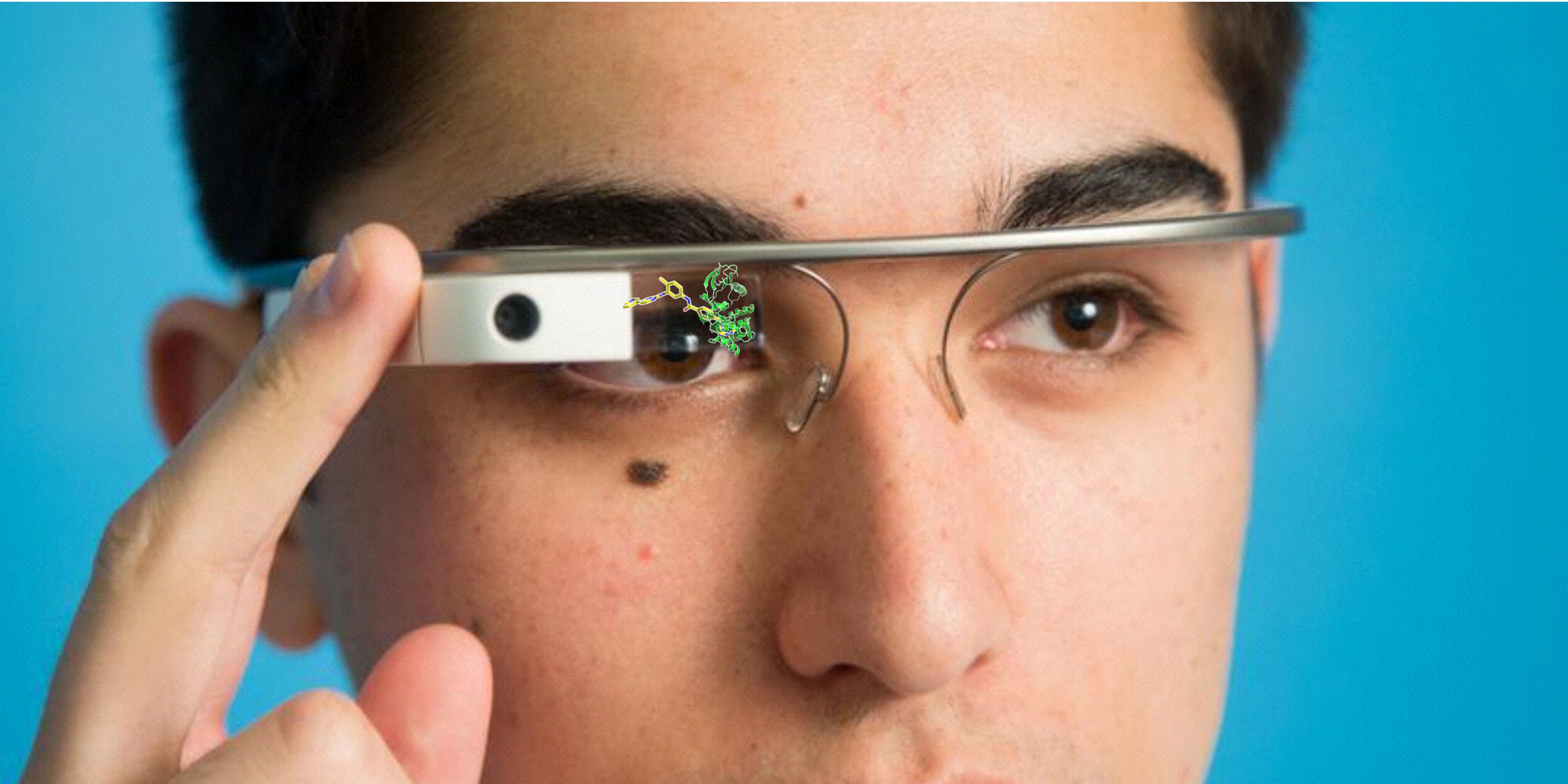


TRANSCRIPTIC CLOUD WETLAB

WHAT DOES THE FUTURE OF DRUG DESIGN LOOK LIKE?



WHAT DOES THE FUTURE OF DRUG DESIGN LOOK LIKE?





YANK: AN OPEN-SOURCE, COMMUNITY-ORIENTED PLATFORM FOR GPU-ACCELERATED FREE ENERGY CALCULATIONS



NVIDIA GTX-1080 (\$650) 9 TFLOP/S SINGLE PRECISION

OpenMM speedup (GTX-1080) over 12-core Xeon X5650 CPU for DHFR

method	natoms	gromacs CPU	OpenMM GPU	speedup
GB/SA	2,489	2.54 ns/day	789 ns/day	311 x
RF	23,558	18.8 ns/day	572 ns/day	30.4 x
PME	23,558	6.96 ns/day	337 ns/day	48.4 x

<http://openmm.org> OpenMM 7.1.0 development snapshot benchmark
gromacs benchmarks from <http://biowulf.nih.gov/apps/gromacs-gpu.html>

YANK

<http://www.getyank.org>

A GPU-accelerated Python framework for exploring algorithms for alchemical free energy calculations

Note

YANK is now in Early Access for its 1.0 release! YAML syntax should be fully operational while we prepare the underlying Python API. The program has not yet been extensively validated. Use at your own risk!

Features

- Modular Python framework to facilitate development and testing of new algorithms
- GPU-accelerated via the [OpenMM toolkit](#) (see [benchmarks](#))
- [Alchemical free energy calculations](#) in both [explicit](#) and [implicit](#) solvent
- Hamiltonian exchange among alchemical intermediates with [Gibbs sampling framework](#)
- General [Markov chain Monte Carlo](#) framework for exploring enhanced sampling methods
- Built-in equilibration detection and convergence diagnostics
- Support for AMBER [prmtop/inpcrd](#) files
- Support for absolute binding free energy calculations
- Support for transfer free energies (such as hydration or partition free energies)

A free, open-source, extensible platform
for free energy calculations and ligand design

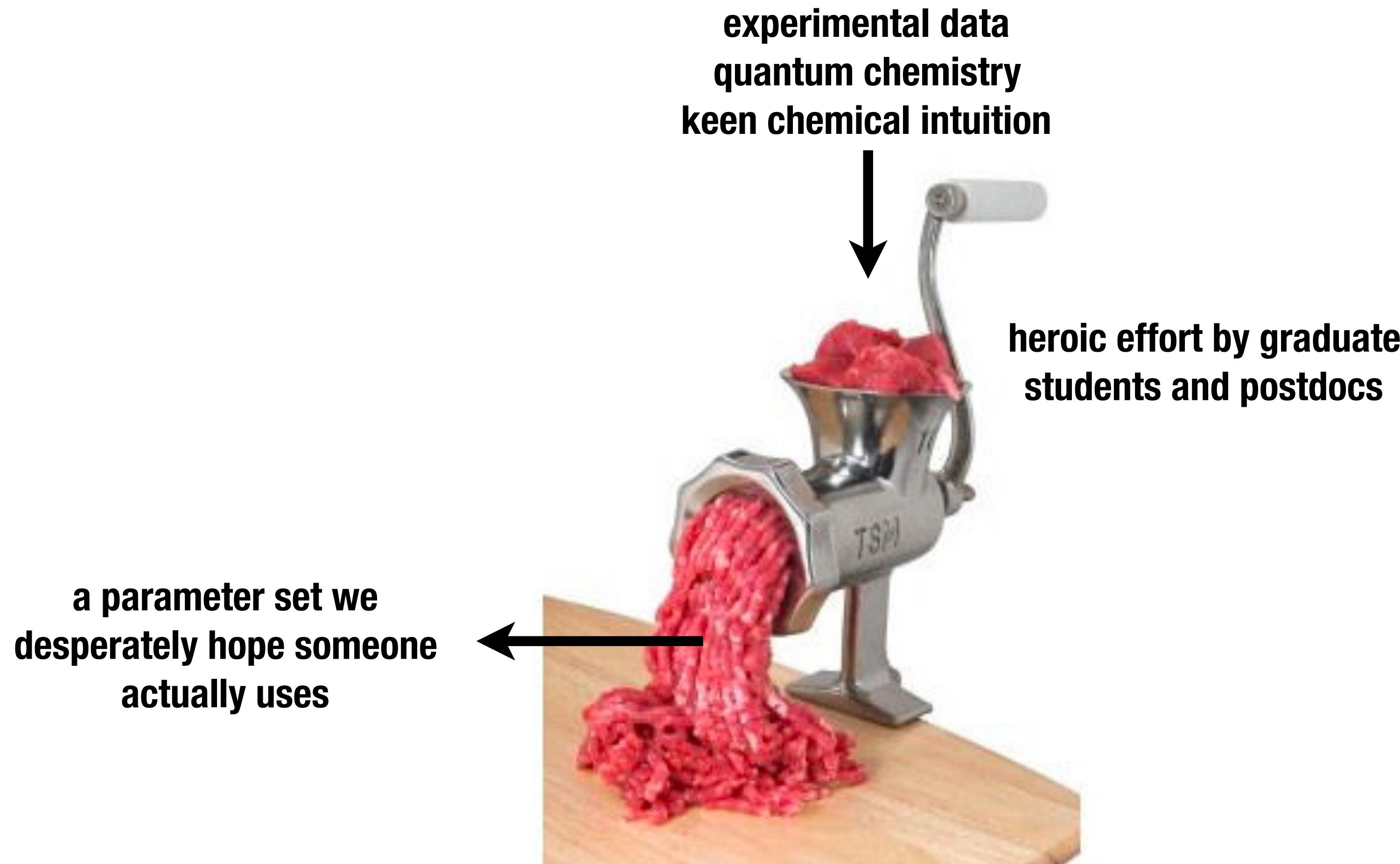


LEVI NADEN



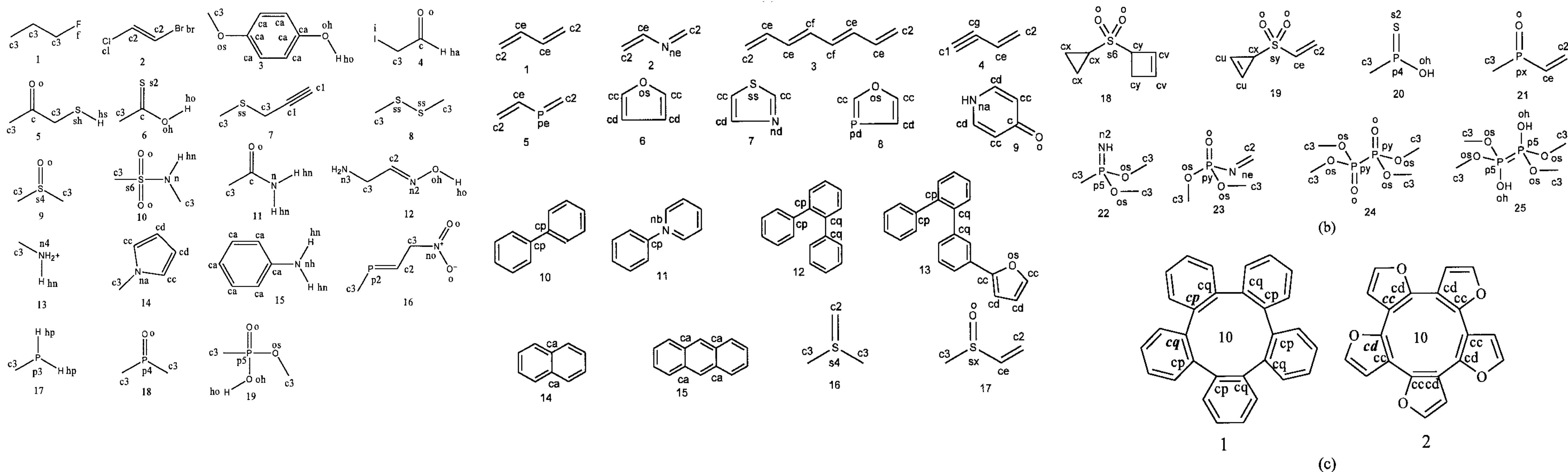
ANDREA RIZZI

HOW ARE FORCEFIELDS MADE?



AS DRUG DISCOVERY EXPLORES NEW PARTS OF CHEMICAL SPACE, HOW CAN FORCEFIELDS KEEP UP?

The Generalized Amber Forcefield (GAFF) was parameterized with this chemical universe:



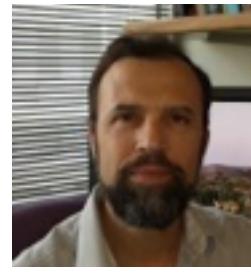
Extension of this universe is nontrivial because parameter fitting code never released!

THE OPEN FORCEFIELD GROUP

<https://github.com/open-forcefield-group>



DAVID MOBLEY
UCI



MICHAEL GILSON
UCSD



MICHAEL SHIRTS
UNIVERSITY OF COLORADO, BOULDER



CHRISTOPHER BAYLY
OPEN EYE SCIENTIFIC



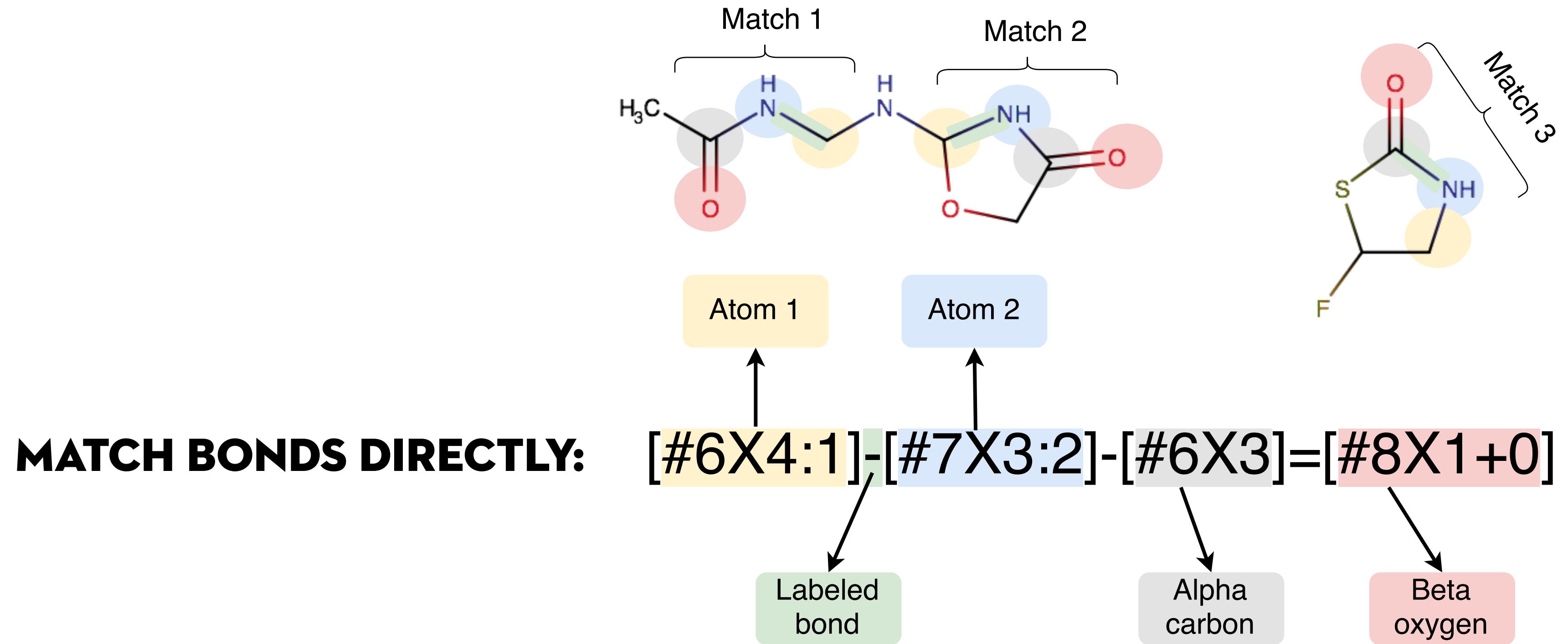
JOHN CHODERA
SKI/MSKCC



KENNETH KROENLEIN
NIST TRC

OPEN SOURCE, OPEN SCIENCE, OPEN DATA

THE SMIRKS NATIVE OPEN FORCE FIELD (SMIRNOFF) ESCAPES THE HEGEMONY OF ATOM TYPES



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

<SMIRNOFF>

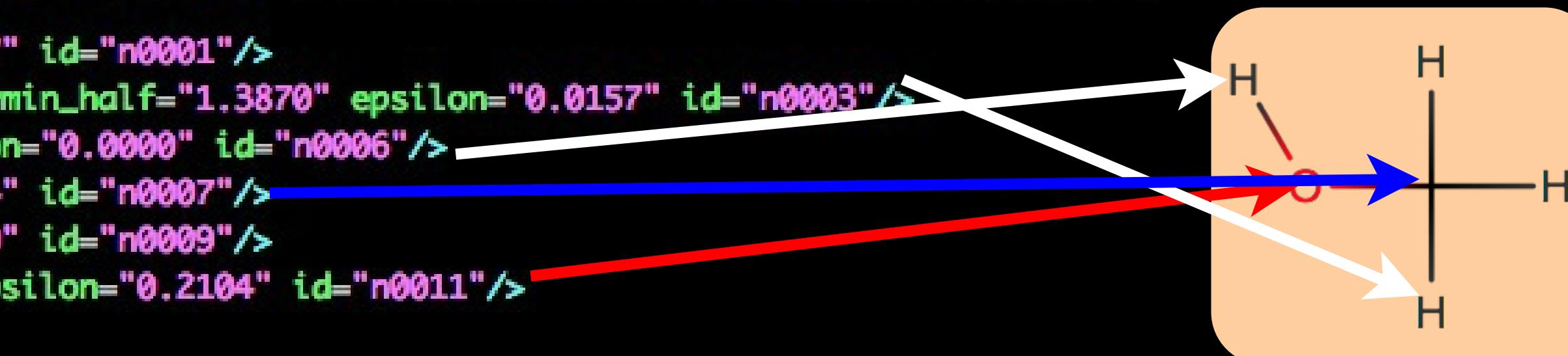
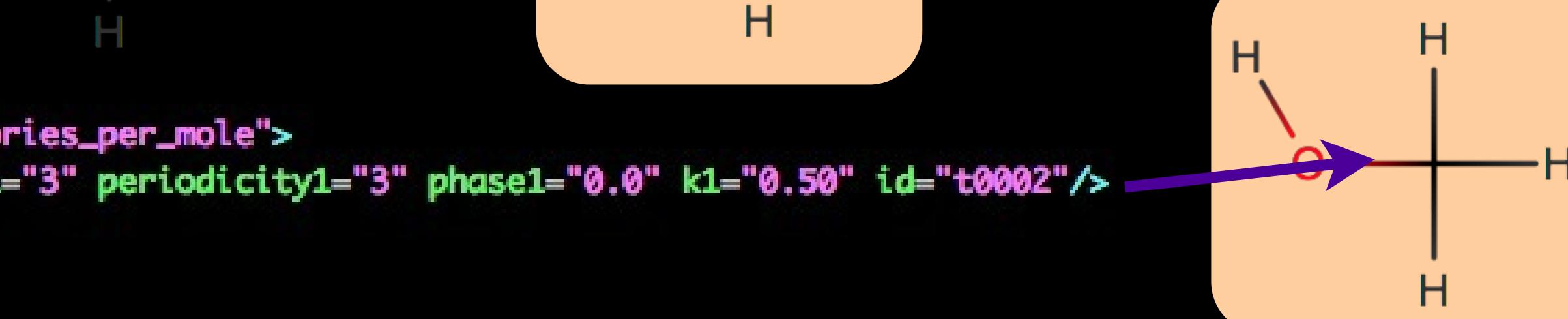
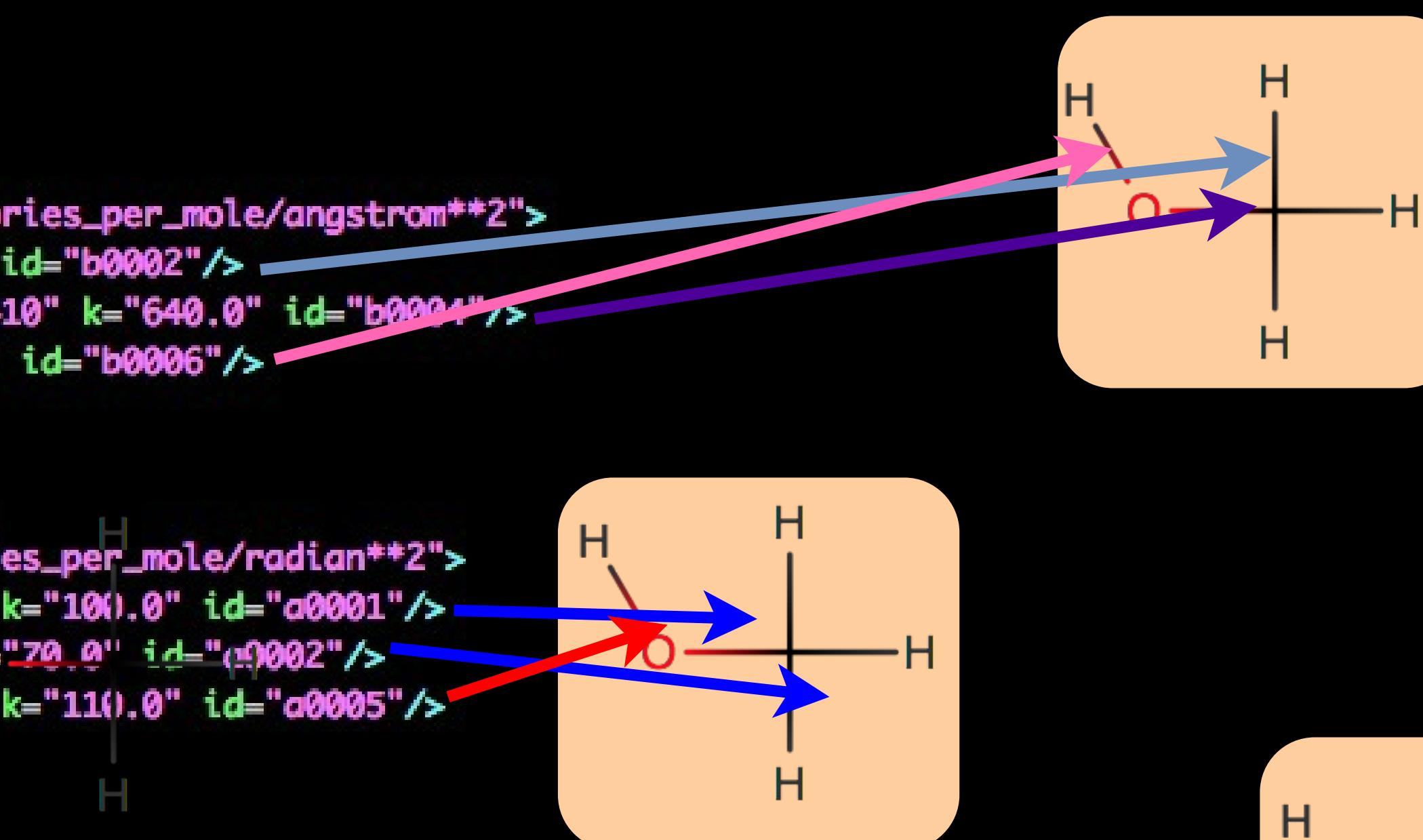
<HarmonicBondForce length_unit="angstroms" k_unit="kilocalories_per_mole/angstrom**2">
    <Bond smirks="[#6X4:1]-[#1:2]" length="1.090" k="680.0" id="b0002"/>
    <Bond smirks="[#6X4:1]-[#8&amp;X2&amp;H1:2]" length="1.410" k="640.0" id="b0004"/>
    <Bond smirks="[#8X2:1]-[#1:2]" length="0.960" k="1106.0" id="b0006"/>
</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" id="a0001"/>
    <Angle smirks="[#1:1]-[#6X4:2]-[#1:3]" angle="109.50" k="70.0" id="a0002"/>
    <Angle smirks="[#6X4:1]-[#8X2:2]-[#1:3]" angle="108.50" k="110.0" id="a0005"/>
</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" phasel="0.0" k1="0.50" id="t0002"/>
</PeriodicTorsionForce>

<NonbondedForce coulomb14scale="0.833333" lj14scale="0.5" sigma_unit="angstroms" epsilon_unit="kilocalories_per_mole">
    <!-- sigma is in angstroms, epsilon is in kcal/mol -->
    <Atom smirks="[#1:1]" rmin_half="1.4870" epsilon="0.0157" id="n0001"/>
    <Atom smirks="[$([#1]-[#6]-[#7,#8,#9,#16,#17,#35]):1]" rmin_half="1.3870" epsilon="0.0157" id="n0003"/>
    <Atom smirks="[#1$(*-[#8]):1]" rmin_half="0.0000" epsilon="0.0000" id="n0006"/>
    <Atom smirks="[#6:1]" rmin_half="1.9080" epsilon="0.1094" id="n0007"/>
    <Atom smirks="[#8:1]" rmin_half="1.6837" epsilon="0.1700" id="n0009"/>
    <Atom smirks="[#8X2+0$(*-[#1]):1]" rmin_half="1.7210" epsilon="0.2104" id="n0011"/>
</NonbondedForce>

</SMIRNOFF>
```



UP

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

README FILE[README.shtml](#)**File List**

[parm_Frosst_note.pdf](#),
[mmff94_am1bcc_atypes.txt](#),
[parm_Frosst.frcmod](#),
[parm_Frosst.pcp](#),
[zinc.sdf](#),
[zinc_p_f_types.txt](#),
[zinc_am1bcc_aty...](#),
[zinc_am1bcc_bty...](#),
[mmff94.sdf](#),
[mmff94_p_f_type...](#),

CCL Supporters

CCL Supporting
Members

**CCL Paid
Services**

Use CCL Paid
Services to support
its operation

The README File

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

Credit

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

Daniel McKay, contributed between (1997-2010)

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

This presents a molecular mechanics force field (FF) extending the AMBER FF to bioorganic small molecules of pharmaceutical interest. The presented **parm@Frosst** FF enables the simulation of biomolecules (enzymes, DNA, peptides, etc.) in the presence of complex organic molecules such as inhibitor and cofactors. As such it can be used as a small-molecule supplement to the AMBER parm9x or ff99 biomolecular force fields, as an alternative to e.g. gaff. The development took place at Merck Frosst Canada, a subsidiary of Merck & Co, between 1992 and 2010 in the context of numerous drug-discovery projects. As a result, **parm@Frosst**, when used to extend one of the "standard" AMBER force fields such as ff99sb, could successfully parameterize approximately 85% of the Merck corporate collection (of small molecules) in 2009 (personal communication to CIB from V. Hornak). Merck & Co generously cleared this material to be released to the scientific community. John Irwin and Brian Schochet generously permitted us to use a fraction of the ZINC dataset (zinc.docking.org). This data repository contains enough information to 1) implement the **parm@Frosst** force field and validate the implementation 2) validate the atom and bond typing of an implementation of the AM1BCC charge model as originally published.

File List for **parm_at_Frosst** Directory

- [parm_Frosst_note.pdf](#) [321kB] : Note about the history and the content of the **parm@Frosst** FF.
- [mmff94_am1bcc_atypes.txt](#) [101kB] : The MMFF94 set AM1-BCC atom types. One type per line. This exactly matches the atom ordering found in the mmff94.sdf file.
- [mmff94_am1bcc_btypes.txt](#) [35kB] : The MMFF94 set AM1-BCC bond types. One type per line. This exactly matches the bond list found in mmff94.sdf file.
- [parm_Frosst.frcmod](#) [151kB] : The AMBER FF parameters specific to small bioorganic.
- [parm_Frosst.pcp](#) [6kB] : The typing rules in PATTY format.
- [zinc.sdf](#) [20500kB] : ZINC set: three dimensional structures in the MDL SD format from a subset of the ZINC dataset.
- [zinc_p_f_types.txt](#) [622kB] : The ZINC set **parm@Frosst** atom types. One type per line. This exactly matches the atom ordering found in the zinc.sdf file.
- [zinc_am1bcc_atypes.txt](#) [1275kB] : The ZINC set AM1-BCC atom types. One type per line. This exactly matches the atom ordering found in the zinc.sdf file.
- [zinc_am1bcc_btypes.txt](#) [449kB] : The ZINC set AM1-BCC bond types. One type per line. This exactly matches the bond list found in the zinc.sdf file.
- [mmff94.sdf](#) [1628kB] : MMFF94 validation : three dimensional structures in the MDL SD format.
- [mmff94_p_f_types.txt](#) [48kB] : The **parm@Frosst** atom types of MMFF94 set. One type per line. This exactly match the atom ordering found in the mmff94.sdf file.

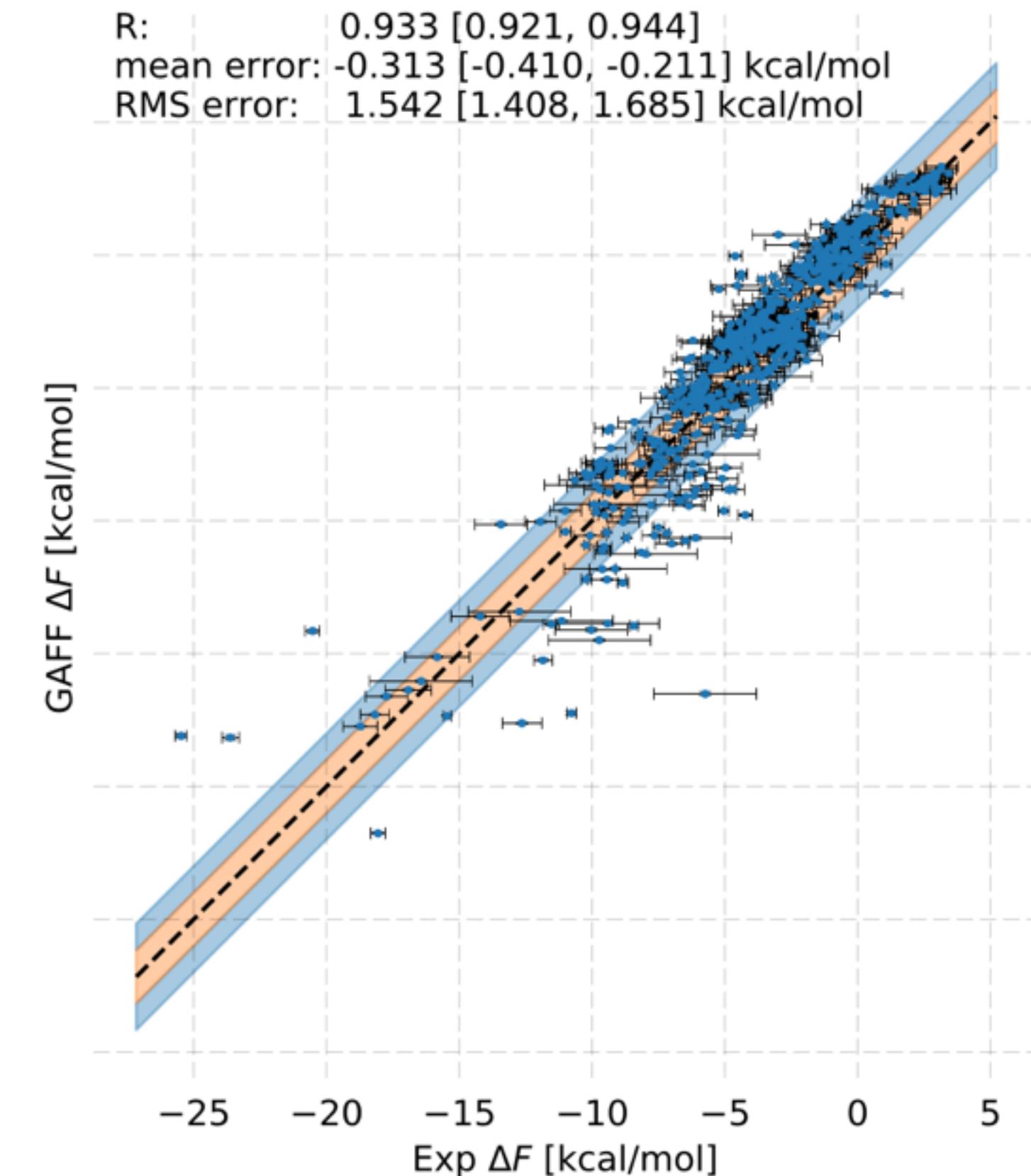
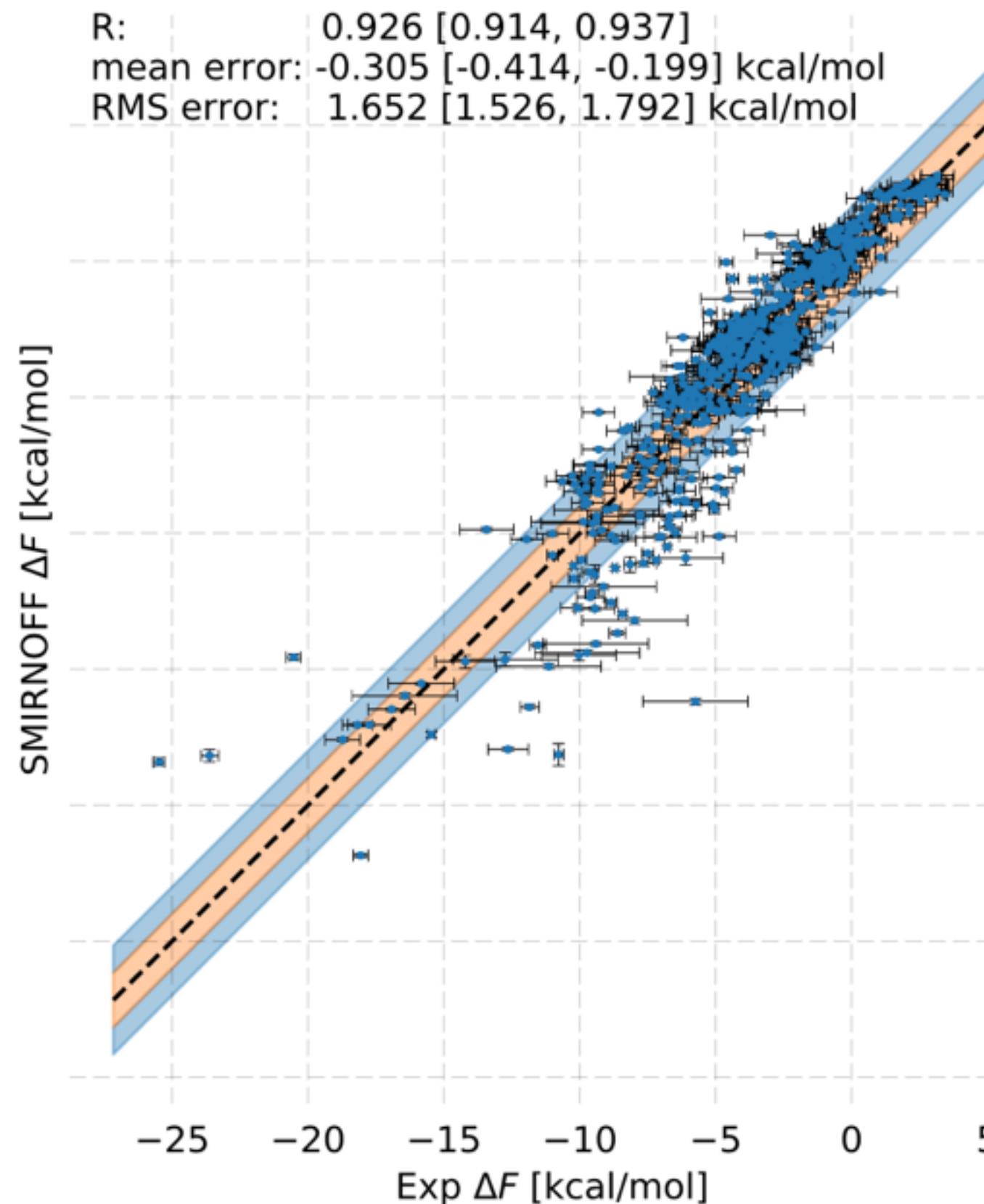
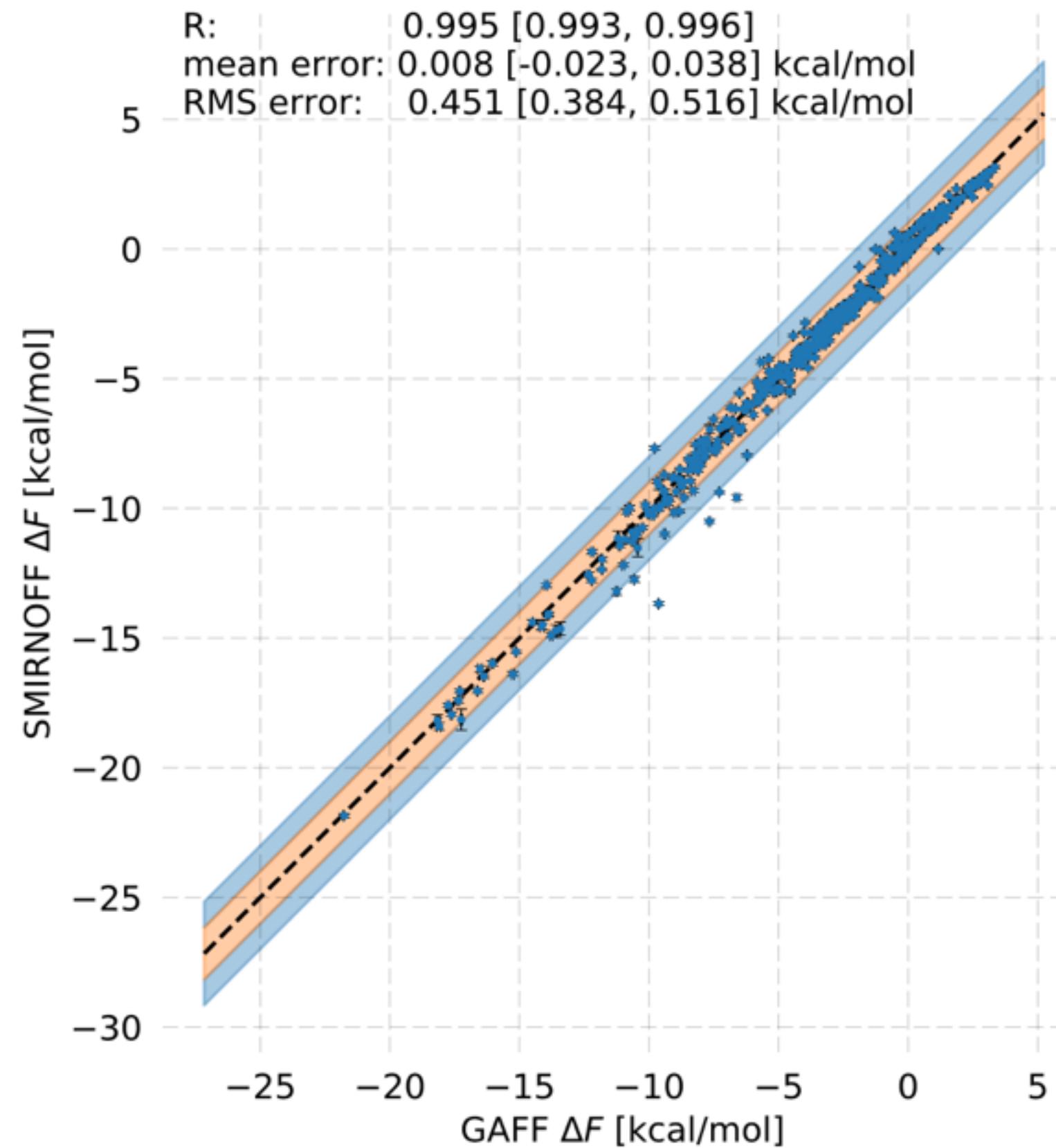
You can also retrieve all the files listed above as a compressed tar ([parm_at_Frosst.tgz](#)) or a zip ([parm_at_Frosst.zip](#)) archive.

NOTE THAT E-MAIL ADDRESSES HAVE BEEN MODIFIED!!!

All @ signs were changed to .. to fight spam. Before you send e-mail, you need to change .. to @

For example: change joe..big123comp.com to joe@big123comp.com

SMIRNOFF@FROSST DOES AS WELL AS GAFF BUT IS 300 LINES INSTEAD OF 6000+ LINES!



FREESOLV HYDRATION BENCHMARK SET

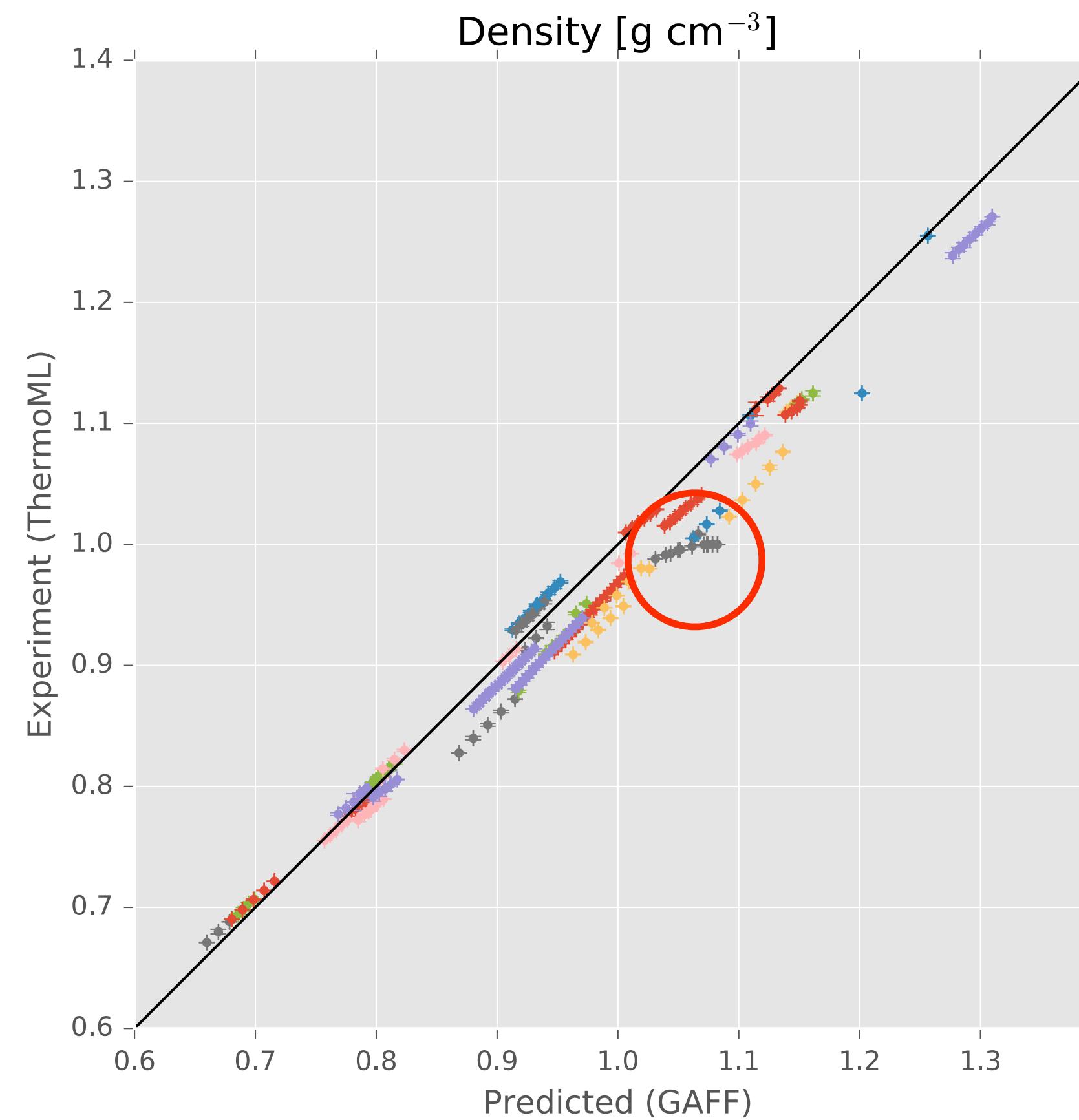
<http://github.com/open-forcefield-group/smirnoff99Frosst>

ANDREA RIZZI

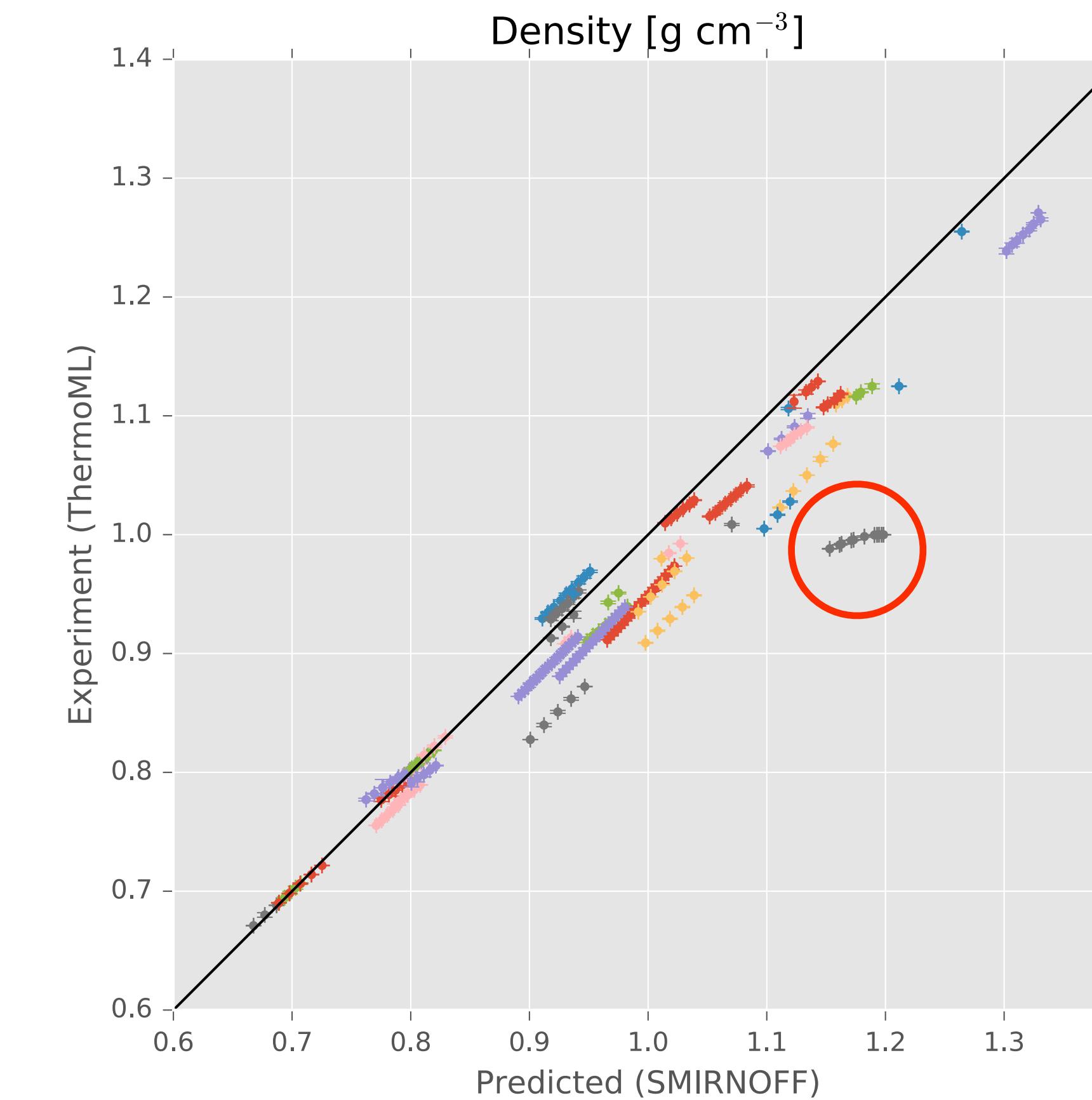


SMIRNOFF@FROSST DOES AS WELL AS GAFF BUT IS 300 LINES INSTEAD OF 6000+!

GAFF



SMIRNOFF



water

THERMOML BENCHMARK SET

SMIRNOFF99FROSST HAS GOOD COVERAGE

NO METALS, METALOIDS, OR BORON (YET)

TYPES MOST OF DRUGBANK THAT DOESN'T HAVE THESE ATOMS

TYPES MOST OF EMOLECULES (EXACT NUMBERS FORTHCOMING)

USING SMIRNOFF IS EASY

```
# Load the smirnoff99frosst forcefield
from openforcefield.typing.engines.smirnoff import ForceField
forcefield = ForceField('smirnoff99frosst.ffxml')
# Create an OpenMM System object to simulate
system = forcefield.createSystem(topology)
```

**CURRENTLY BASED ON OECHEM, BUT ADDING RDKIT SUPPORT
HEADED BY SHUZHE WANG (NOW IN RINIKER GROUP)**

**WOULD IT BE USEFUL TO MAKE SMIRNOFF99FROSST AVAILABLE IN
RDKIT AS AN MMFF ALTERNATIVE?
WOULD WE NEED IT AT THE C++ LAYER?**

**ALSO ADDING NATIVE SMALL MOLECULE SUPPORT (VIA AMBERTOOLS)
INTO OPENMM USING RDKit (PYTHON APP LAYER ONLY)**

BAYESIAN MACHINE LEARNING CAN AUTOMATICALLY GENERATE SMARTS/SMIRKS FOR CHEMICAL PERCEPTION

PARENT TYPES

```
% atom types
[#1] hydrogen
[#6] carbon
[#7] nitrogen
[#8] oxygen
[#9] fluorine
[#15] phosphorous
[#16] sulfur
[#17] chlorine
[#35] bromine
[#53] iodine
```

X

DECORATORS

```
% total connectivity
X1 connections-1
X2 connections-2
X3 connections-3
X4 connections-4
% total-h-count
H0 total-h-count-0
H1 total-h-count-1
H2 total-h-count-2
H3 total-h-count-3
% formal charge
+0 neutral
+1 cationic+1
-1 anionic-1
% aromatic/aliphatic
a aromatic
A aliphatic
```

=

PROPOSED CHILD TYPES

```
[#6X4:1] tetrahedral carbon
[#6:1]~[#7] carbon nitrogen-adjacent
```

INDEX	ATOMS	MOLECULES	
1 :	464	42	
2 :	0	0	
3 :	232	42	
4 :	107	42	
TOTAL :	803	42	

TYPE NAME
c_hydrogen
c_carbon
c_carbon neutral
c_oxygen

SMARTS	REF	TYPE	FRACTION OF REF TYPED MOLECULES	MATCHED
[#1]	HC		244 /	244 (100.000%)
[#6]				
[#6&+0]	CT		232 /	232 (100.000%)
[#8]	OH		68 /	68 (100.000%)

THE FUTURE OF FORCEFIELD PARAMETERIZATION?

