Open Force Field Roadmap

Open Force Consortium
Kickoff meeting, 7 Jan 2019

What are the ultimate goals of the Open Force Field Initiative?

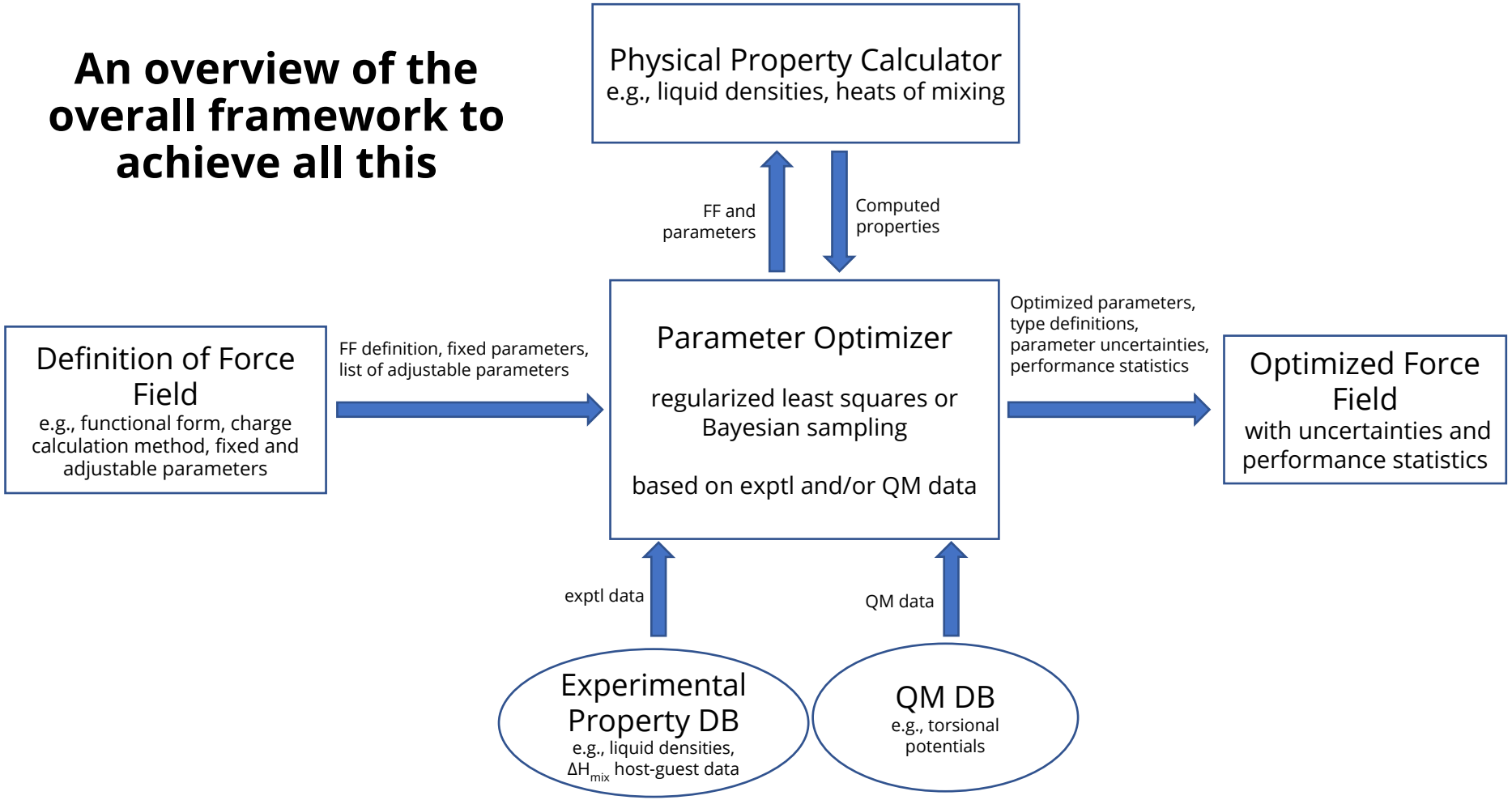
- Develop the highest possible accuracy classical atomistic force fields for biomolecular and biocompatible molecules
- Develop an open, scalable, extensible toolkit for automatically parameterizing classical atomistic force fields from experimental data
- Generate/curate open datasets necessary for producing high-accuracy biomolecular and biocompatible forcefields
- Answer important unresolved scientific questions about cost benefit of adding different types of detail to the molecular energy functions
- Provide consistency across different biomolecule and polymer classes

What are the goals of the Open Force Field Consortium?

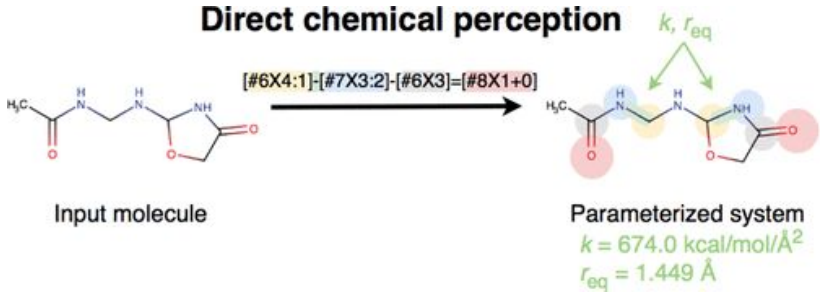
- The **Consortium** is this particular academic-industrial collaboration meeting here today
- Deliver small molecule force fields that:
  - Have extremely high coverage of chemical space
  - Are completely open
  - Are better than the existing art
  - Are consistent with current biomolecular force fields
- Deliver the tools, datasets, and benchmarking capability necessary to improve the above force fields

The OpenFF Consortium operates in synergy with the other efforts of the Open Force Field Initiative funded through different routes

An overview of the overall framework to achieve all this



What is ready now?



SMIRNOFF99frosst force field [Mobley et al. J. Chem. Theory Comput., 2018, 14 (11), pp 6076–6092](#)

Atom typing via direct chemical perception, **preliminary benchmarks of density, dielectric constants,** and **hydration free energies.** Comparable in accuracy to GAFF, ready to be optimized

Open Force Field Toolkit 0.1 https://github.com/openforcefield/openforcefield
Can parameterize a small molecule, combine it with protein and solvent, run in AMBER, CHARMM, NAMD, GROMACS, Desmond. Currently **requires OpenEye Toolkit** but **RDKit support nearly in (version 1.0)**

Planned deliverables and efforts for the first year

Deliverables:

- Fully open force field toolkit for parameter assignment with support for major small molecule input formats (mol2, SDF, etc)
- A small molecule force field with refit selected SMIRNOFF99Frosst torsions, Lennard-Jones, and bond/angle terms for **fixed SMIRNOFF types** using regularized least squares (ForceBalance) optimization, compatible with AMBER biomolecular force fields.
- An automatic benchmarking suite assessing:
  - Condensed phase properties
  - Compatibility with biomolecular force FF
- An open, validated database of experimental data and QM used for the refitting

Main Efforts:

- Tools for fragmentation, torsion drives, and torsion fitting
- Plug-in architecture for physical property calculation (for benchmarking and fitting)
- Hierarchical modeling for accelerating condensed phase parameter searches
- Force Field Strike Team to rapidly patch major flaws and develop long-term solutions
- Development of Bayesian inference infrastructure for second-generation force field

Planned deliverables and efforts for the second year

Deliverables:

- Regular versioned improvements of small molecule force fields:

  - Integrating additional experimental data each cycle
    - More complex mixture data: partition coefficients, chemical potentials
    - Host-guest binding thermodynamics
  - Starting to use the Bayesian inference framework
    - Sample parameter space broadly, escaping local minima
    - Refine/expand atom types
    - Provide next-generation BCC models

Main Efforts:

- Investigate effects of different functional forms using Bayesian inference: are they worth it?

Planned deliverables and efforts year 3 and beyond

Deliverables:

- Continued versioned improvements of small molecule force fields:

  - Improved functional forms where the evidence supports it
    - Off-site charges / polarizability / better LJ function
  - An easy-to-use optimization workflow to incorporate your own data internally and new potential energy models or components
  - Integrating additional experimental data each cycle
    - NMR couplings, protein-ligand binding free energies, protein crystals
  - Bayesian inference framework
    - Quantify parameter uncertainty
    - Provide statistical evidence for accuracy/complexity trade off
  - What else are we missing?

Planned Initiative efforts beyond the Consortium

- Completely self-consistent biomolecular, small molecule, and polymer force fields
  - Easy treatment of covalent ligands
  - Support for nonnatural amino acids
  - Biopolymer-similar support (peptoids, etc.)
  - Improved protein/nucleic acids/lipids, especially in heterogeneous systems

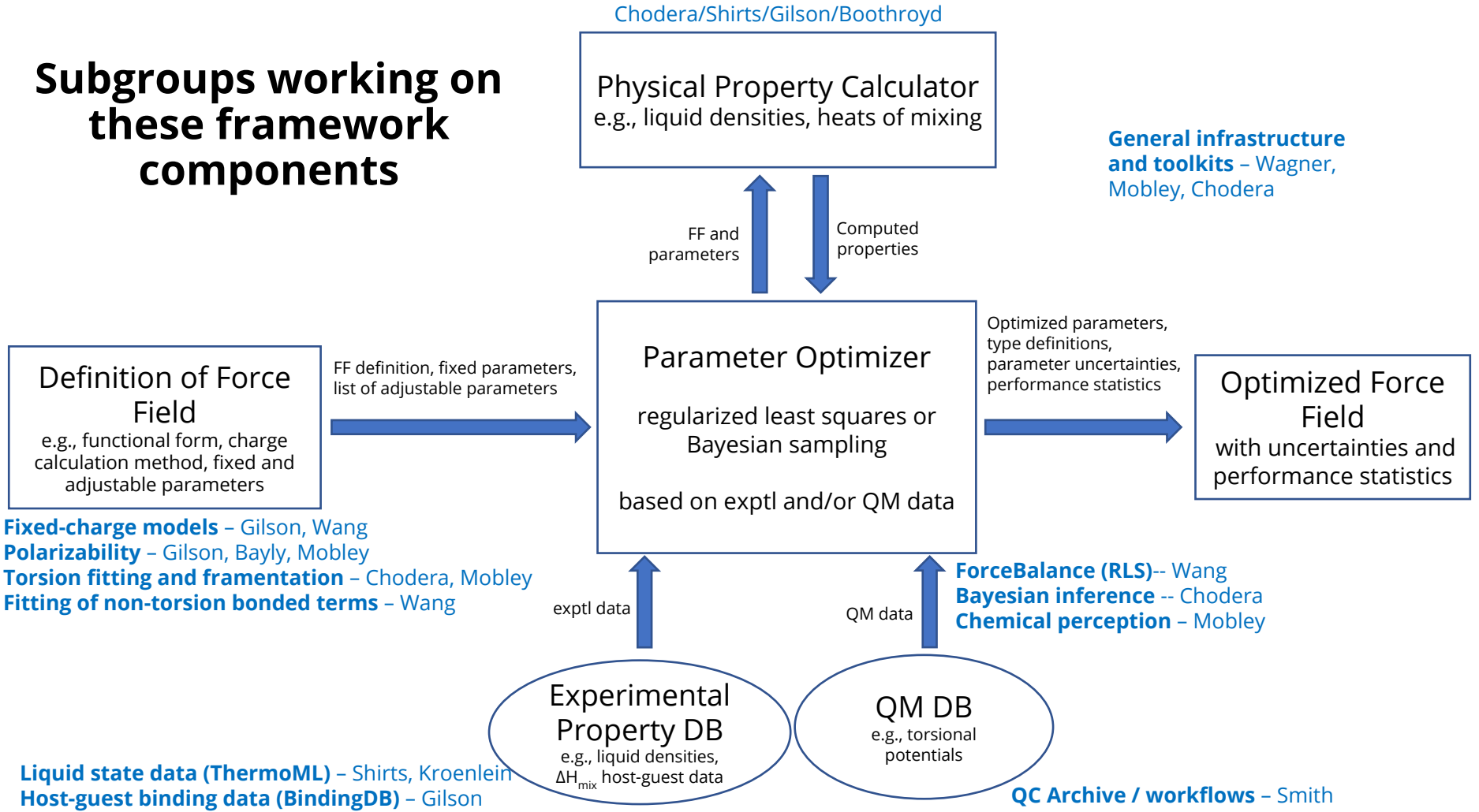
- Next-next generation force fields:
  - How do we combine these efforts with machine learning or neural network potentials? Is there a way to combine the physical picture with the informatics picture to better include multibody effects / reactive systems?

- NIH (in revision) and DARPA (to be submitted) proposals for these areas

- Collaborating with other force field developers on automated benchmarking, SMIRNOFF support, etc.

- Other things you are interested in that management won't fund for now?

Subgroups working on these framework components



General infrastructure and toolkits – Wagner, Mobley, Chodera

Fixed-charge models – Gilson, Wang
Polarizability – Gilson, Bayly, Mobley
Torsion fitting and framentation – Chodera, Mobley
Fitting of non-torsion bonded terms – Wang

ForceBalance (RLS)-- Wang
Bayesian inference -- Chodera
Chemical perception – Mobley

Liquid state data (ThermoML) – Shirts, Kroenlein
Host-guest binding data (BindingDB) – Gilson

QC Archive / workflows – Smith