open
forcefield

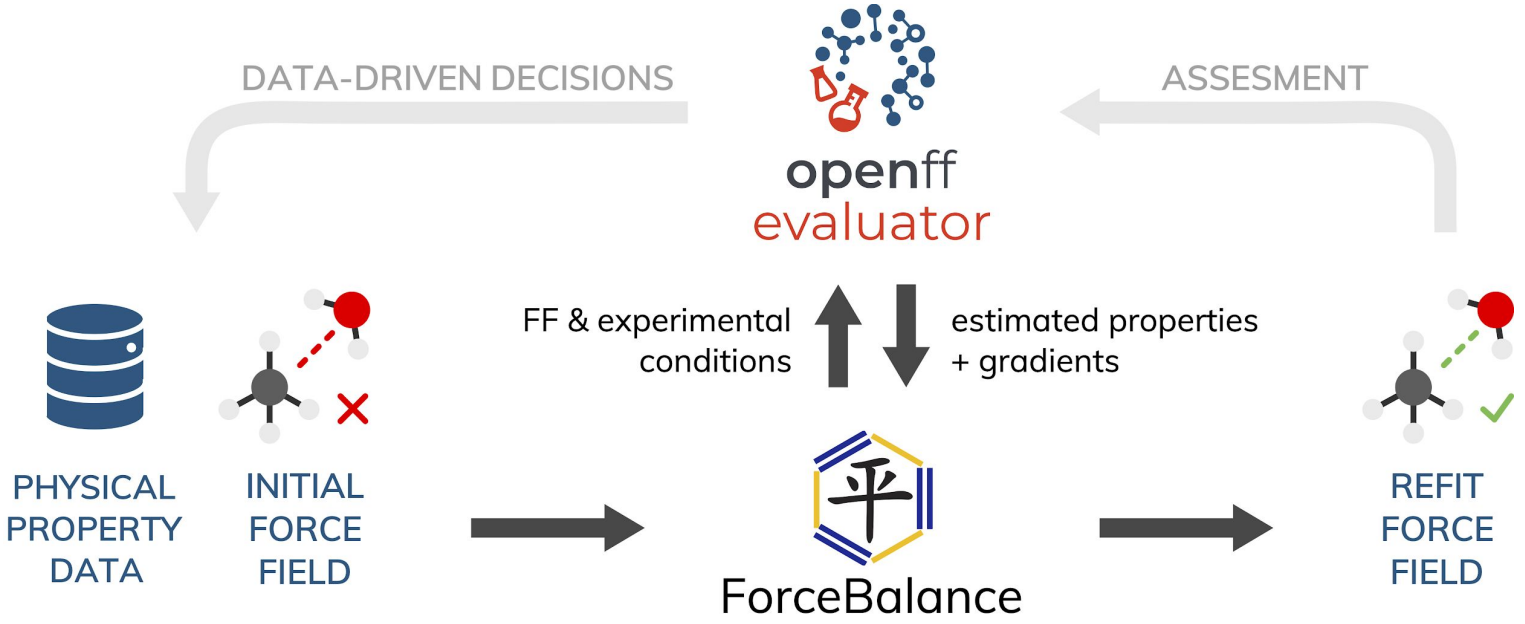
@openforcefield

www.openforcefield.org

Which physical properties should we be optimizing non-bonded interactions against?

Simon Boothroyd, XtalPi / Open Force Field Consortium Distinguished Postdoctoral Fellow

We can now routinely retrain non-bonded (VdW) interaction parameters against physical property data

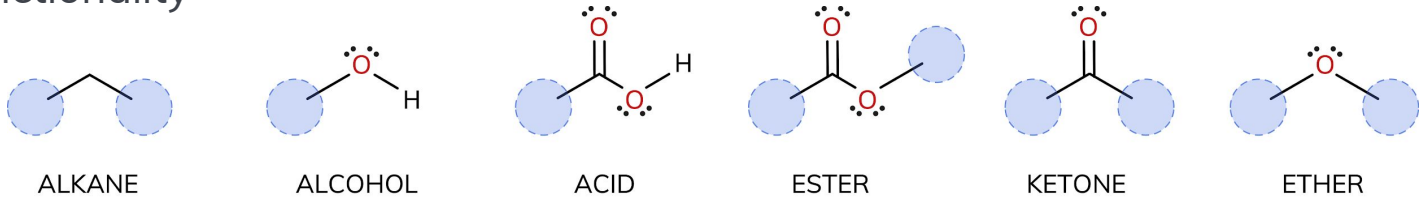


Which physical properties are most informative to train VdW parameters against?

- Historically training against pure properties only [ϱ_{pure}, H_{vap}]
  - Difficult to find **accurate** data sets of H_{vap} measurements
  - Change in polarisation not captured by **fixed charged** models
- Mixture properties [$H_{mix}(x)$, $\varrho(x)$] possibly offer significant benefits
  - **Significantly** more data to train / test against
  - Easier to build training sets containing **specific** interactions
  - Potential **correlation** between $H_{mix}(x)$ *and* ***binding free energies, solvation free energies, partition coefficients etc.***

Aim: refit and benchmark the Parsley VdW parameters against different combinations of properties

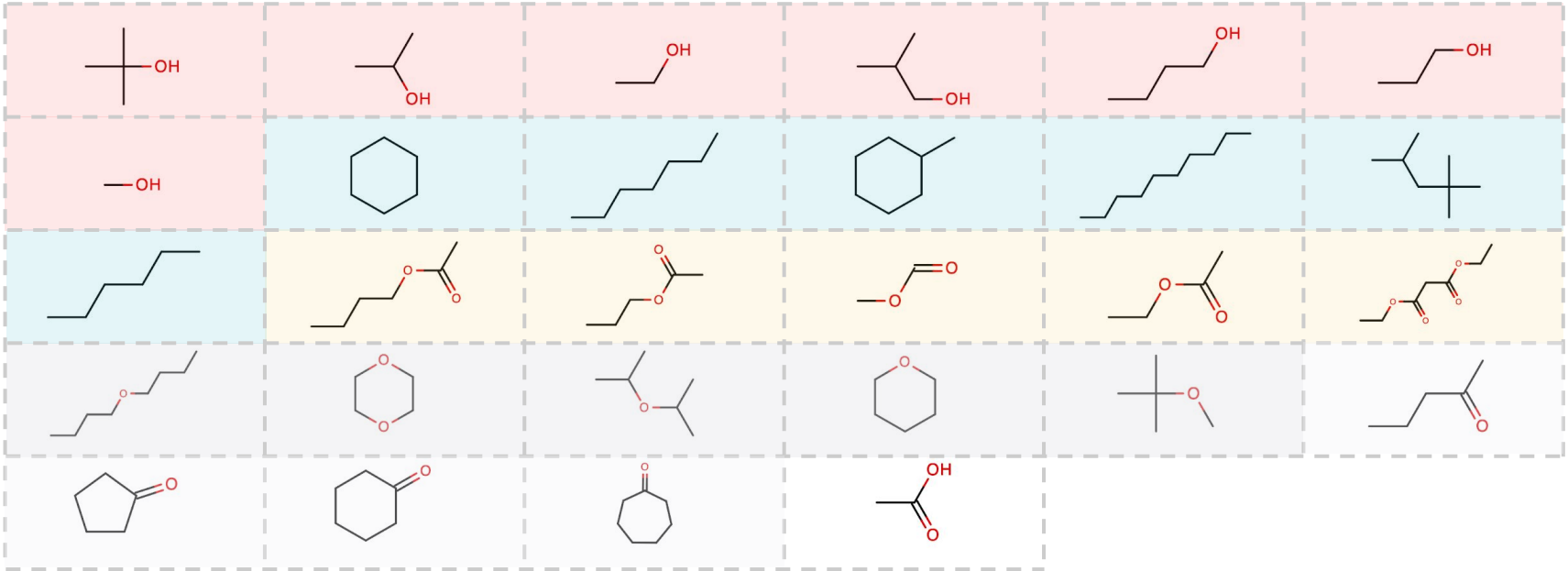
- Used the **`openff-1.0.0`** force field as the initial parameters
- Performed optimizations against different combinations of properties
  - $H_{mix}(x) + \varrho(x)$
  - $H_{mix}(x) + \varrho(x) + \varrho_{pure}$
  - $\varrho_{pure} + H_{vap}$
  - $H_{mix}(x) + \varrho(x) + \varrho_{pure} + H_{vap}$
- Test / train on a combination of alcohol, acid, ester, ether, ketone and alkane functionality



Optimisation

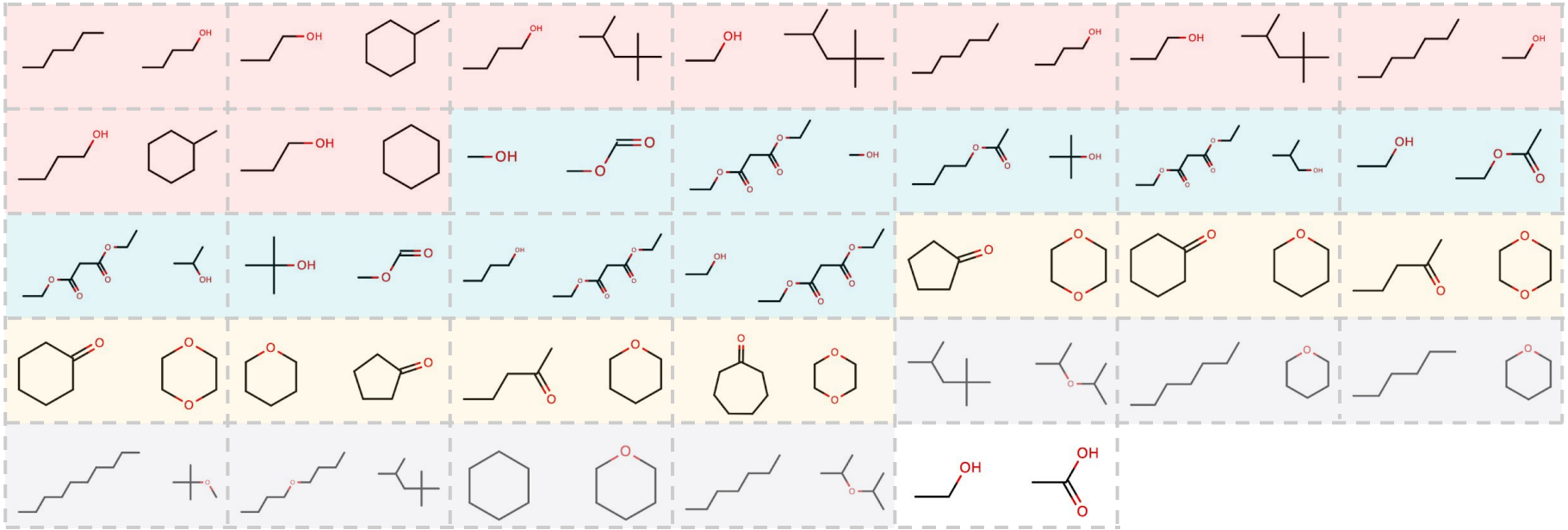
The pure training set contained a total of 56 data points

- ϱ_{pure} and H_{vap} measurement at T~298.15 K and p=1.0 atm

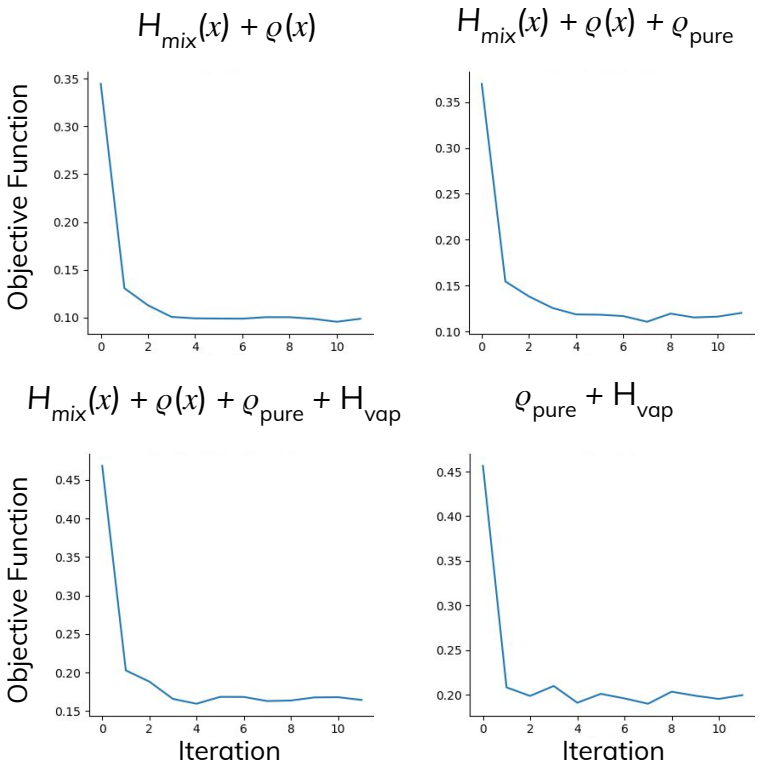


The mixture training set contained a total of 198 data points

- $H_{mix}(x)$ and $\varrho(x)$ measurements at $x \in [0.25, 0.50, 0.75]$, T~298.15 K p=1.0 atm



The objective function decreased by a factor of ~2-3 for each combination of properties

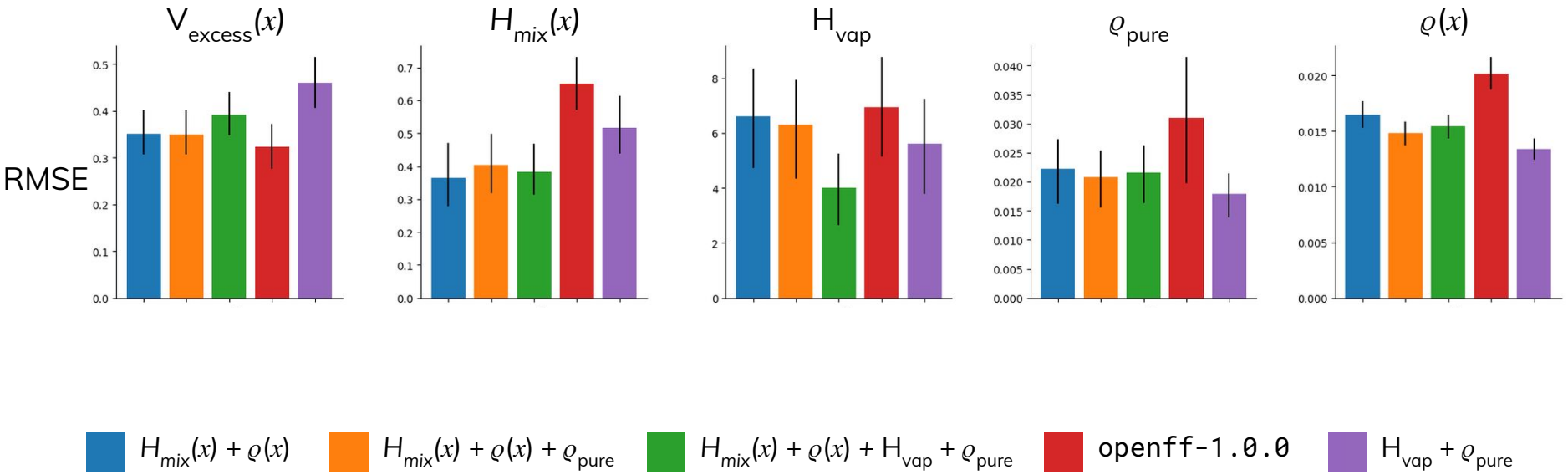


Benchmarking

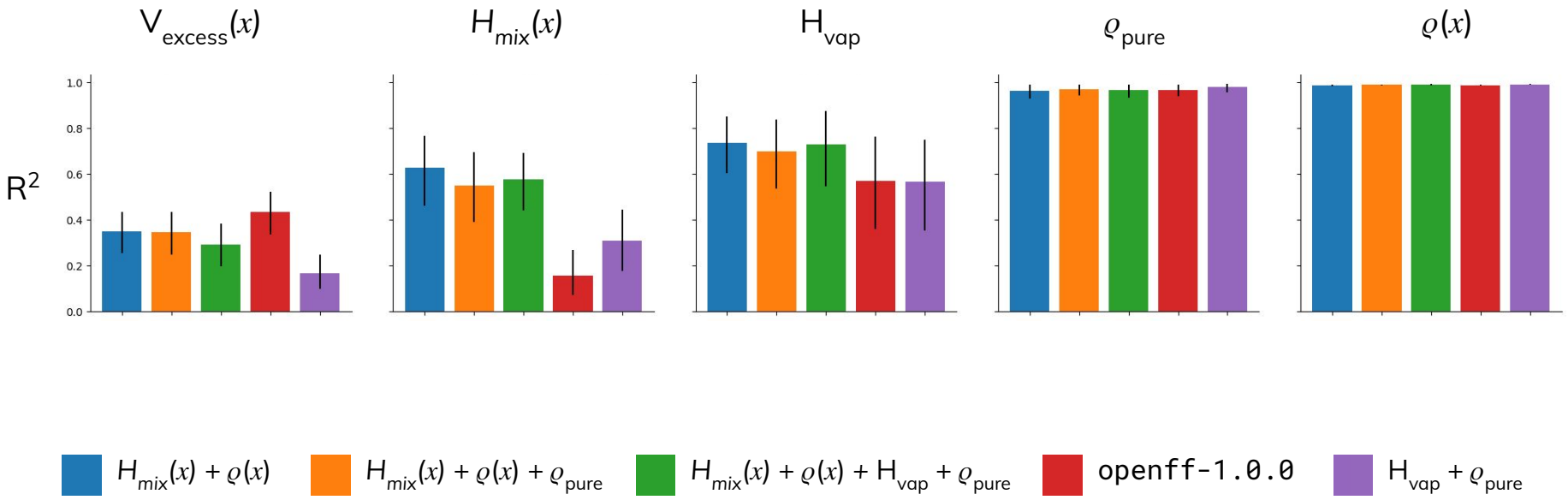
Benchmarking set ~5x larger than the training set

- Contained a total of **24 ρ_{pure} and 24 H_{vap}** pure data points
  - T~298.15 K, p=1.0 atm
  - ~ even split of alkane, alcohol, ester, ether and ketone functionality
- Contained ~**240 $H_{mix}(x)$**, **370 $\rho(x)$**, and **370 $V_{excess}(x)$** binary data points
  - $x \in$ [0.25, 0.50, 0.75], T~298.15 K, p=1.0 atm
  - contained almost all **permutation pairs** of pure functionalities

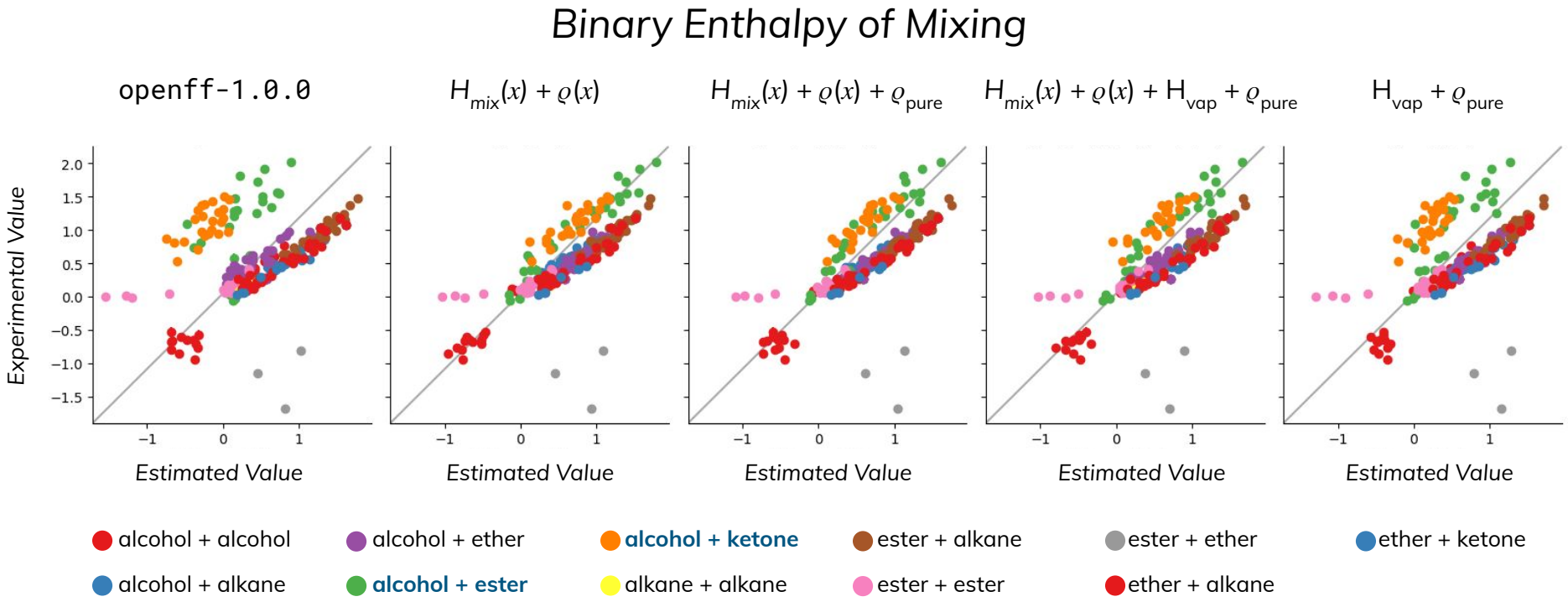
Properties improved almost universally after refitting



Properties improved almost universally after refitting



Most significant improvements observed in mixtures with large deviations from ideality



Conclusion and Next Steps

- The automated fitting infrastructure was used to **effortlessly** refit select VdW parameters of **four new** force fields
- Results of the study are suggestive that in future we should be training against **mixture physical property data**
- In the process of benchmarking against host-guest binding affinity data

All input scripts / tools can be found on GitHub

- Optimization + Benchmarking Input and Analysis Scripts
  - https://github.com/openforcefield/nistdataselection/
- OpenFF Evaluator
  - https://github.com/openforcefield/openff-evaluator
- ForceBalance
  - https://github.com/leeping/forcebalance

Questions?