Parameterization Perspective IV:
Data for Parameterization and Validation - Physical Properties

Michael R. Shirts, University of Colorado Boulder
(with many other people's work listed as we go!)

THE NEXT STEPS!

Benchmarking the force fields that are produced

Bringing in data for future parameterization

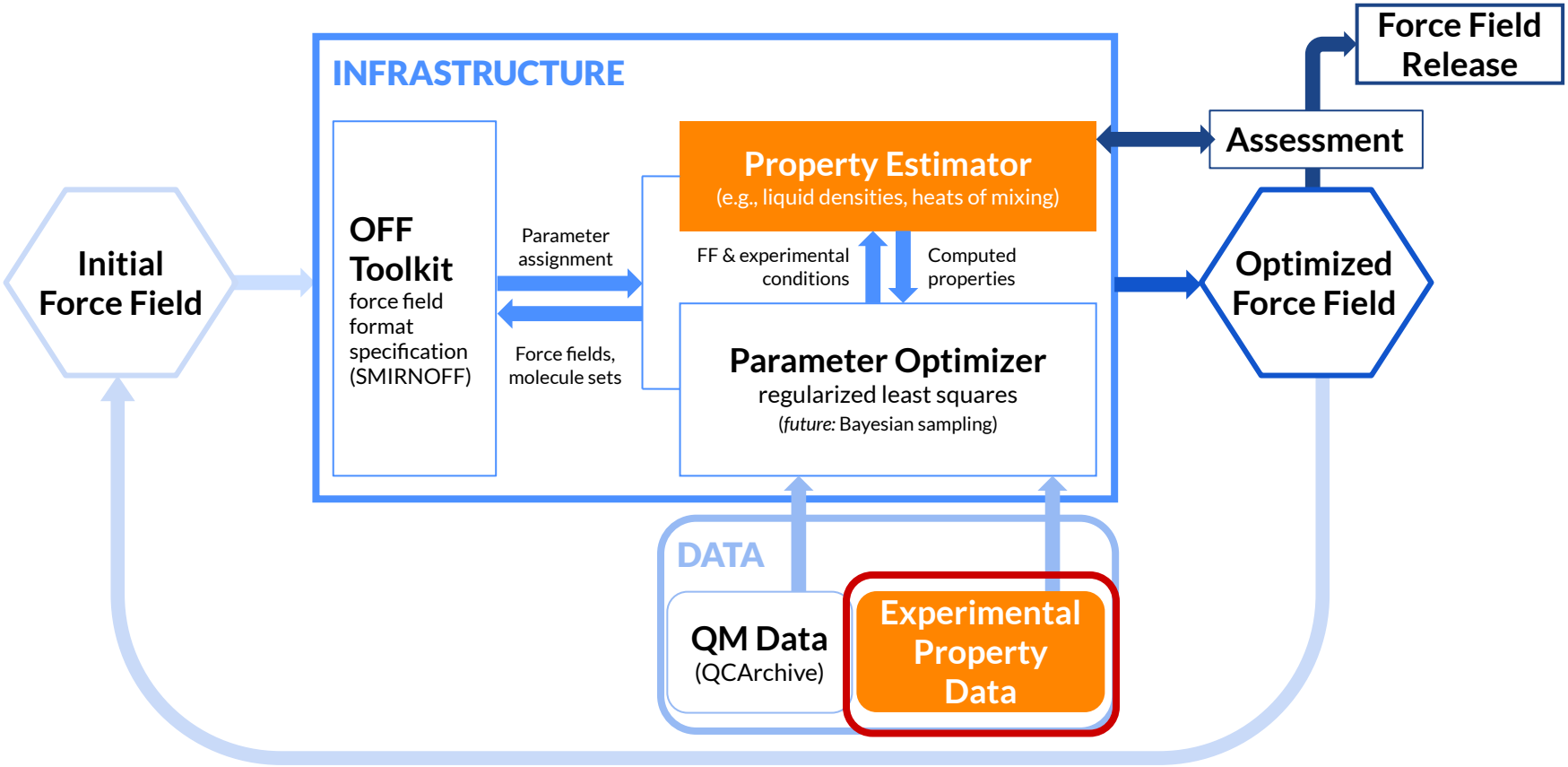
GENERAL STRATEGY

Parameterize on cheaper quantities to simulate

Benchmark on more expensive quantities

Over time, shift data from benchmarking to parameterization

CURRENT PLANS FOR PHYSICAL PROPERTIES



INITIAL PHYSICAL PROPERTIES USED FOR FIRST RELEASE

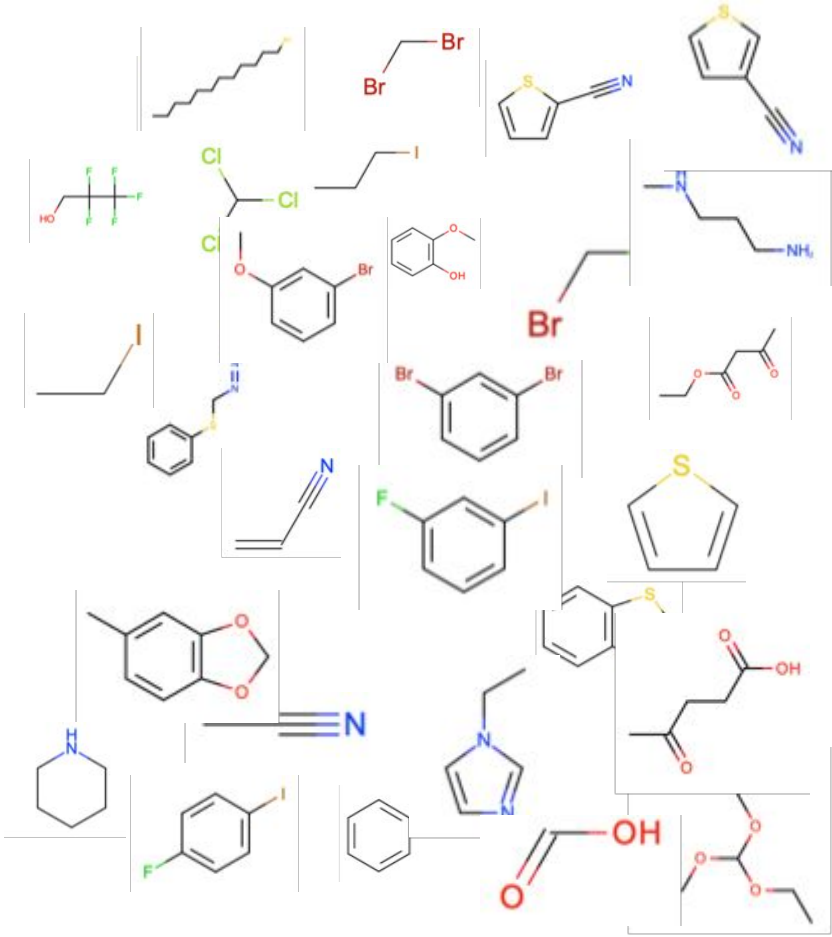
Used densities and ΔH_{vap} for pure fluids near room temperature

Just optimized **existing** LJ types

Only SMIRKS that are involved in at least two ΔH_{vap} and two densities, (about half of the total) are allowed to vary.

Just 30 compounds, 58 data points

Uncertainties are taken for average uncertainties for density and ΔH_{vap} over all of ThermoML



WE WANT COMPLEX FLUID MEASUREMENTS THAT ARE SUFFICIENTLY EASY TO COMPUTE

We **hypothesize** that getting the composition-dependent properties right in binary and ternary liquids will go quite a ways towards getting protein-ligand interactions right.

These calculations are significantly easier to converge,in simulation and significantly easier to make precise experimentally

WHAT PROPERTIES WILL WE USE AND WHY?

$G(T,P,N_i,N_j \ldots)$ is sufficient to characterize all thermodynamics in an NPT system

$dG/dP\ (T,P,N_i,N_j \ldots) = V(T,P,N_i,N_j \ldots)$

Obtained from densities of mixtures, excess volumes, excess density

$-d(\beta G)/d\beta\ (T,P,N_i,N_j \ldots) = H(T,P,N_i,N_j \ldots)$

Obtained from ΔH mixing, ΔH vaporization, ΔC_P mixing (2nd derivative)

$dG/dN_i = \mu_i(T,P,N_i,N_j \ldots)$

Obtained from chemical potentials, solvation ΔG, activity coefficients

dG/dE = related to dielectric constant when external electric field E is applied.

dG/dA = related to surface tension when interfacial surface A exists.

ThermoML HAS A WIDE RANGE OF FLUID MIXTURE DATA COLLECTED FROM THE LITERATURE

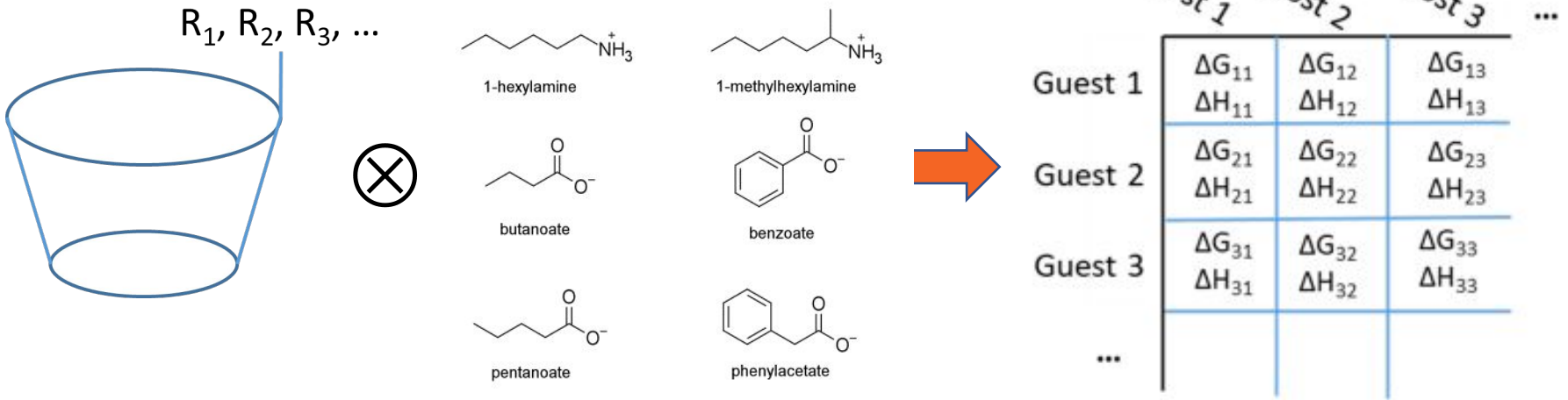
Property	Number of Data Points (in Thousands)		
	Pure	Binary	Ternary
Density	63.6	224.5	97.3
Vapor Pressure	26.3	57.3	7.2
Enthalpy of Vaporization	0.5	0	0
Enthalpy of Mixing	-	30.2	2.1
Dielectric Coefficient	1	3.2	0.2
Surface Tension	2.6	5.4	1
Activity Coefficient	30.3	3.1	4.3
Heat Capacity	14.1	16.7	2.9

Other physical property data we are planning on using

Host-guest binding affinities (μ_{ligand}(1 host, solvent))

The Gilson lab has extensive experience synthesizing, measuring (Katy Kellet), and simulating (David Slochower) these systems

NOT used in parameterization for now, will be used in benchmarking/validation



Other physical property data we are planning on using

Ligand binding affinities (μ_{ligand}(1 protein, solvent))

Semi-automated proten ligand binding energies with David Hahn (OpenFF@Janssen) with Vytas Gapsys (Bert de Groot, Max Plank) with pre-prepared systems from Gapsys and de Groot

Not ready for Oct 1st “Parsley” release

INVESTIGATING OTHER DATA SOURCES

ThermoML doesn't have older/"simpler" data

We are still trying to work out some of conditions for releasing and using uncertainties for the data

We may need help filling the gap for simper systems

- Talking with BYU/AIChE about Design Institute for Physical Properties (DIPPR) database of molecular properties
- Extensive predictive correlations for pure fluids
- Discussing getting access to the a subset of the validated data used as input (already nearly free data for ~60 fluids, more potentially available)

EXAMPLES OF DIPPR DATA

DIPPR® Database Interface Manage Account Log off

Constant Properties For: BENZENE

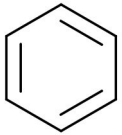
CAS Name	BENZENE	**Molecular Formula**	C6H6
IUPAC Name	BENZENE	**Structural Formula**	-CHCHCHCHCHCH-
CAS RN®*	71-43-2	**Release Date**	1/01/1983
DIPPR ID	501	**Family**	n-Alkylbenzenes
SMILES Formula	c1ccccc1	**Sub Family**	

* CAS Registry Number® is a Registered Trademark of the American Chemical Society

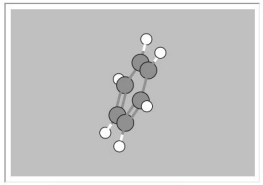
Synonyms:

BICARBURET OF HYDROGEN | CARBON OIL | COAL NAPHTHA | CYCLOHEXATRIENE | BENZOL | MINERAL NAPHTHA | MOTOR BENZOL | PHENYL HYDRIDE | BENZOLENE | PYROBENZOLE

2D Structure



3D Structure



Advanced 3D view

Property	Value	Units	Data Type	Uncertainty	Source Type	Note
Molecular Weight	78.11184	kg/kmol				
Critical Temperature	562.05	K	Smoothed	< 0.2%	Evaluated	
Critical Pressure	4895000	Pa	Smoothed	< 0.2%	Evaluated	
Critical Volume	0.256	m^3/kmol	Smoothed	< 3%	Evaluated	
Critical Compressibility Factor	0.268	unitless	Defined		Staff	
Acentric Factor	0.2103	unitless	Defined		Staff	
Normal Boiling Point	353.24	K	Experimental	< 1%	Evaluated	
Melting Point	278.68	K	Experimental	< 1%	Evaluated	
Triple Point Temperature	278.68	K	Predicted	< 1%	Staff	49
Triple Point Pressure	4764.22	Pa	Predicted	< 3%	Staff	
Liquid Molar Volume	0.0894764	m^3/kmol	Experimental	< 1%	Staff	
Ideal Gas Enthalpy of Formation	82880000	J/kmol	Experimental	< 3%	Evaluated	

DIPPR® Database Interface Manage Account Log off

Database References for: **Liquid Density of BENZENE**

(DIPPR ID: 501)

Primary Accepted references appear in **Blue Bold.**

Acceptable references appear in Green Plain.

Rejected references appear in *Red Italic*

Data Set	Reference	Note	Data Type	Uncertainty
27390	Campbell, A.N.; Chatterjee, R.M.; "Orthobaric Data of Certain Pure Liquids in the Neighborhood of the Critical Point"; *Can. J. Chem.* 1968 , 46 , 4		Experimental	< 1%

Temperature (K)	Liquid Density
378.8	10.062

Data Set	Reference	Note	Data Type	Uncertainty
27397	Smith, B.D.; Srivastava, R.; *Thermodynamic Data for Pure Compounds. Part A. Hydrocarbons and Ketones;* Elsevier, Amsterdam, 1986		Smoothed	< 1%

Temperature (K)	Liquid Density
284	11.373

Data Set	Reference	Note	Data Type	Uncertainty
27400	Steele, W.V.; Archer, D.G.; Chirico, R.D.; Collier, W.B.; Hossenlopp, I.A.; Nguyen, A.; Smith, N.K.; Gammon, B.E.; "The Thermodynamic Properties of Quinoline and Isoquinoline"; *J. Chem. Thermodyn.* 1988 , 20		Experimental	< 1%

Temperature (K)	Liquid Density
298.18	11.184

PHYSICAL PROPERTY BENCHMARKING FOR PARSLEY

Exact choices of molecules to be decided in the next week

- Pure fluids: (~40 diverse molecules), each at 2 temperatures.
  - Dielectric constants (if > 12)
  - Density
  - Heat of vaporization
- For binary mixtures (~15 molecules, where data exists, two compositions)
  - Excess volumes (two compositions, one T)
  - Heats of mixing (two compositions, one T)
- If time, run FreeSolv hydration energies
- Some host-guest calculations (2 hosts, 5 guests)
- NO protein-ligand binding calculations by Oct 1st (hopefully soon after)

OTHER FUTURE DATA FOR BENCHMARKING (OR EVEN PARAMETERIZATION?)

Partition coefficients between solvents

Relative solubilities

Speed of sound (function of both temperature and pressure dependence)

Strain energies from CSD structures (or even matching X-ray data?)

NMR data

Collecting NEW simple liquid data (density, heat of mixing)

WHAT ELSE? COME TO THE BREAKOUT SESSION AFTER LUNCH!