A Brief History of Molecular Dynamics

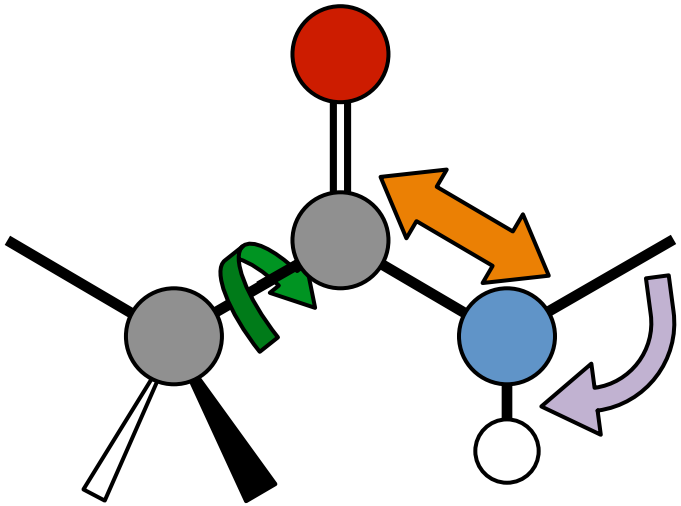
Or, an analysis of its place in high-performance computing

David S. Cerutti

CI Forum,
Michigan State University

The Universe According to MD

- The Newtonian approximation of chemistry is a canonization of the way freshmen and sophomore students are taught to think of molecules.



$$U=\sum_{p\in \text{bond}} k_p\,(L_i - L_0)^2 + \sum_{a\in \text{ang}} k_a\,(\theta_i - \theta_0)^2 + \sum_{v\in \text{dihe}} k_v \cos(n\theta_v - \varphi_v)$$

$$+\sum_{i,j} \left[\frac{k_c\, q_i\, q_j}{r_{ij}} - 4\epsilon_{ij}\left(\frac{\sigma_{ij}}{r_{ij}}\right)^6 + 4\epsilon_{ij}\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}\right]$$

Building on a remarkable success

- The Cornell charge set (parm94, Amber ff95) has been passed down through a long lineage: ff99, ff99SB, ff99SB-ILDN, ff14SB.
- Much faster computers enable QM calculations vastly more sophisticated than HF/6-31G*.
- Hydrogen steric models have changed since 1994. A great deal more effort, in aggregate, has gone toward modifying torsions than any other aspect of the Amber force field.
- Electrostatics are perhaps the most physically meaningful parameters in the force field, and are easy to derive in an automated fashion.

Building on a remarkable success

"Hartree-Fock calculations in vacuum yield more polarized charges than many post Hartree-Fock methods, making these charges suitable for condensed-phase calculations."

– Everyone who develops such charge sets

Building on a remarkable success

"Hartree-Fock calculations in vacuum yield more polarized charges than many post Hartree-Fock methods, making these charges suitable for condensed-phase calculations."

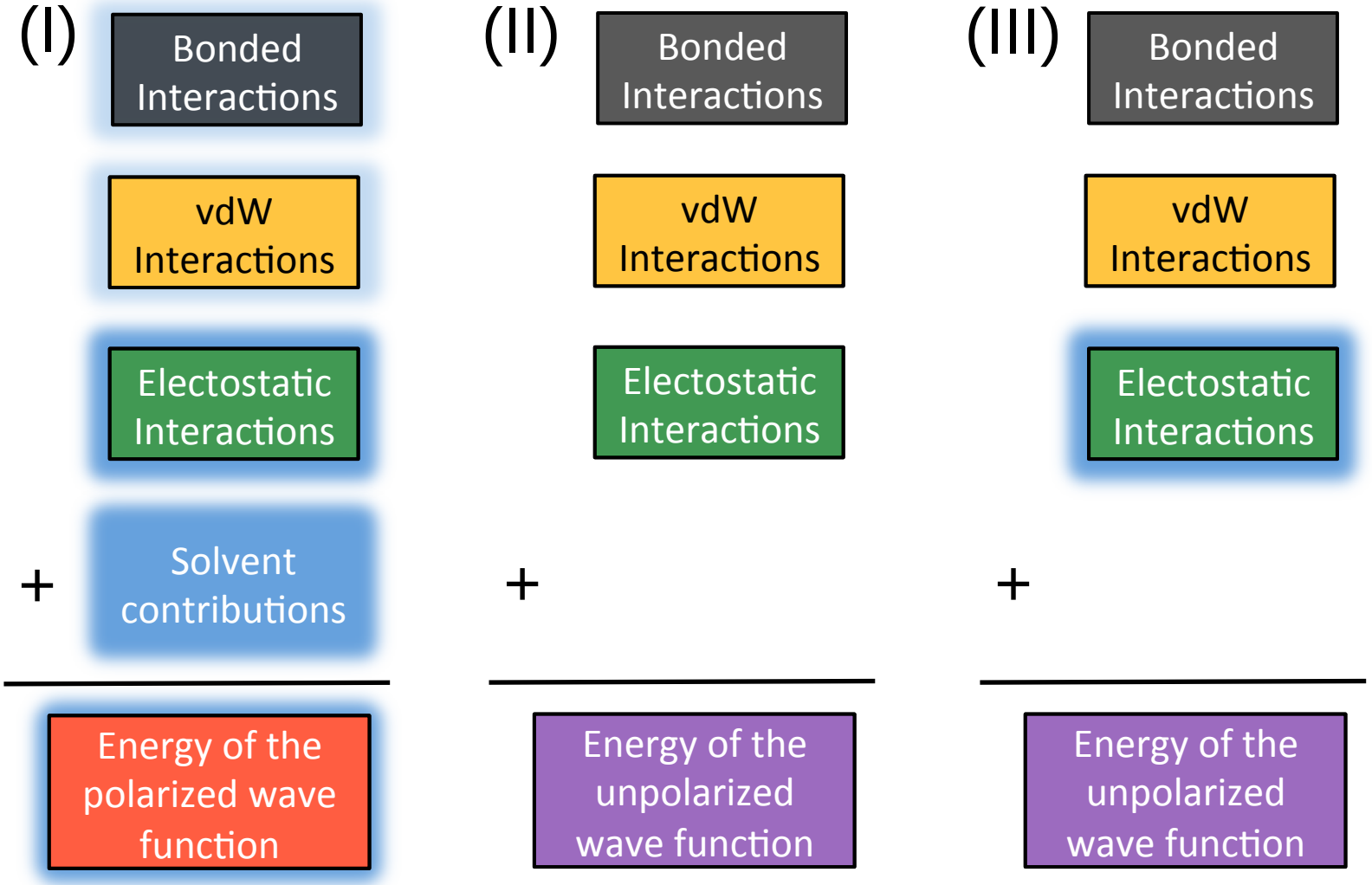
– Everyone who develops such charge sets

"If once you start down the dark path, forever will it dominate your destiny."

– Yoda

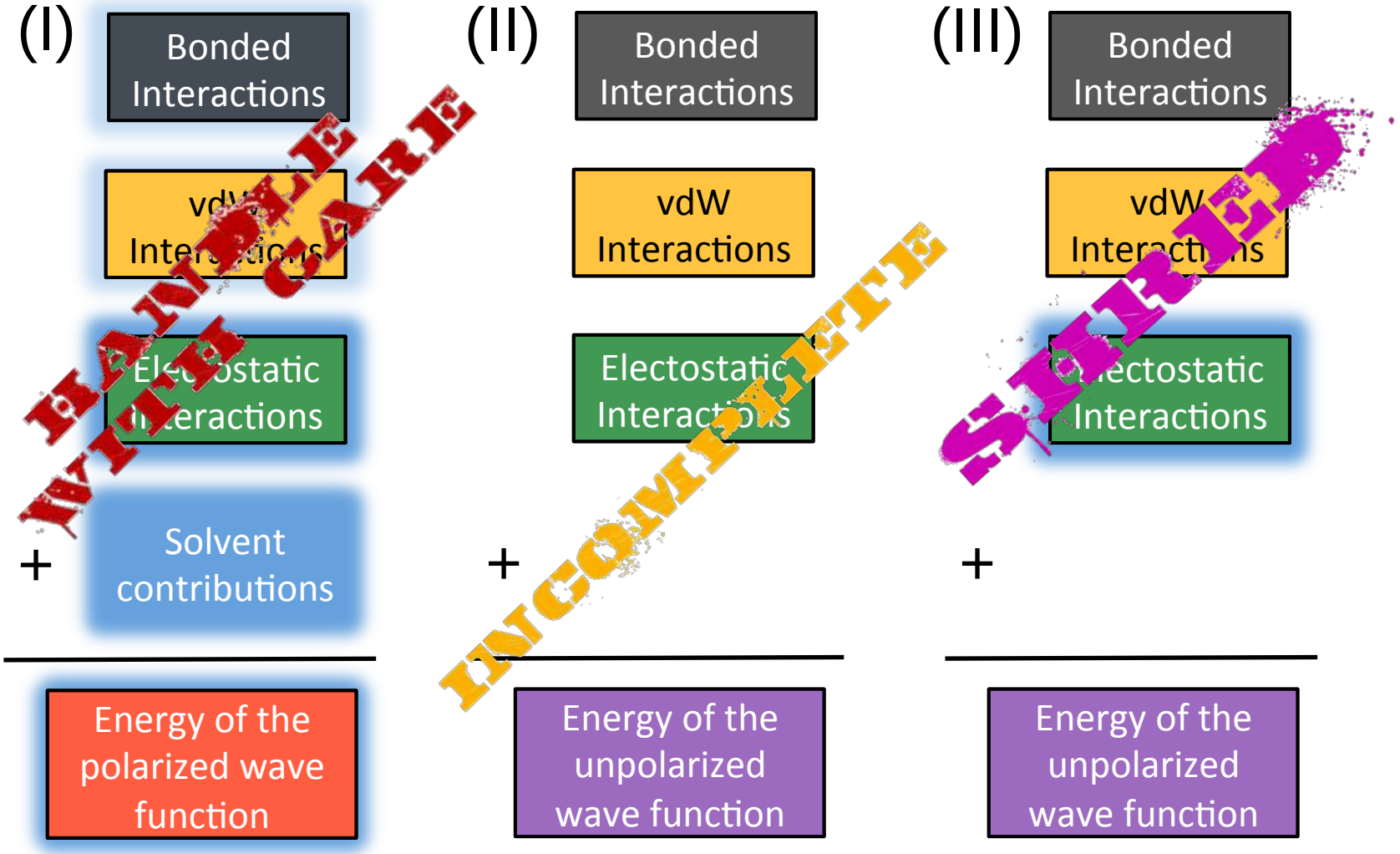
How did this go subtly wrong?

- Scheme (I) is pretty hard. Scheme (II) is only relevant in a vacuum. Scheme (III) is wrong.



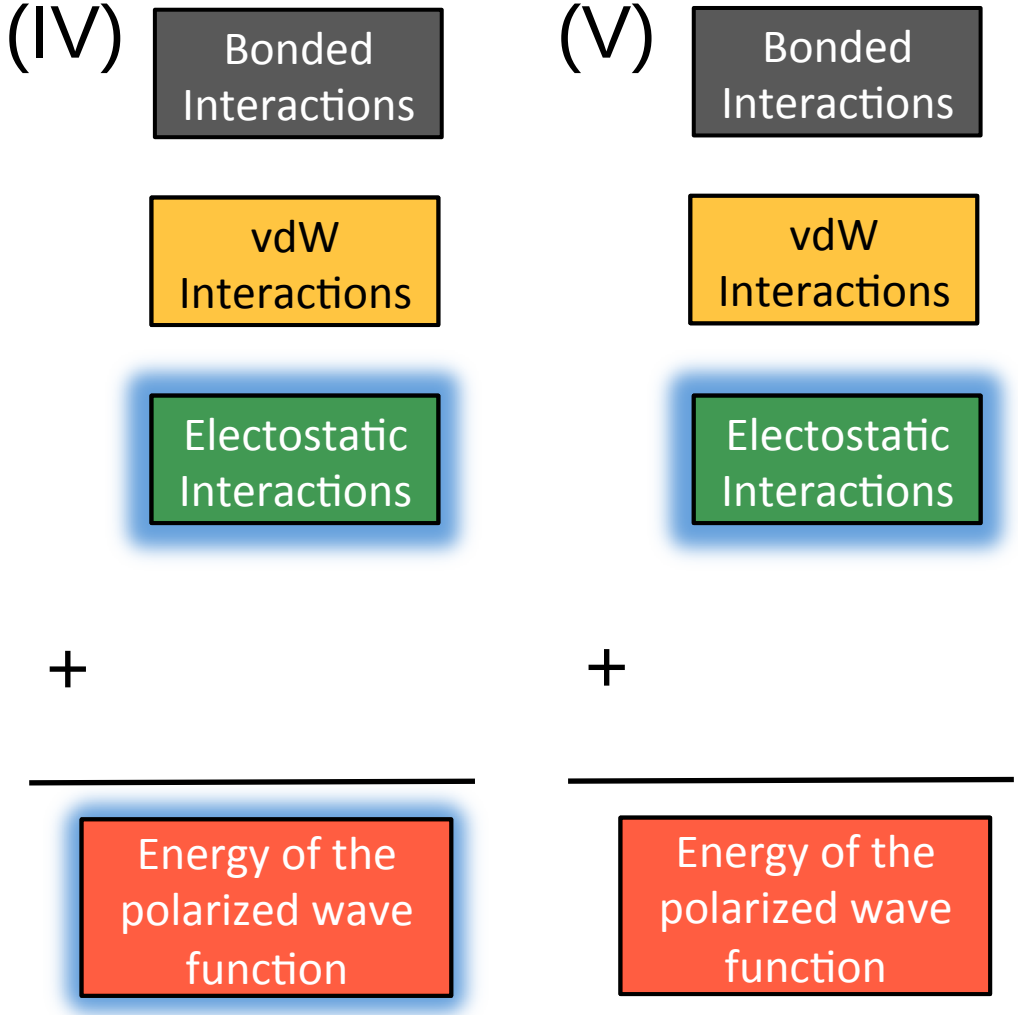
How did this go subtly wrong?

- Scheme (I) is pretty hard. Scheme (II) is only relevant in a vacuum. Scheme (III) is wrong.



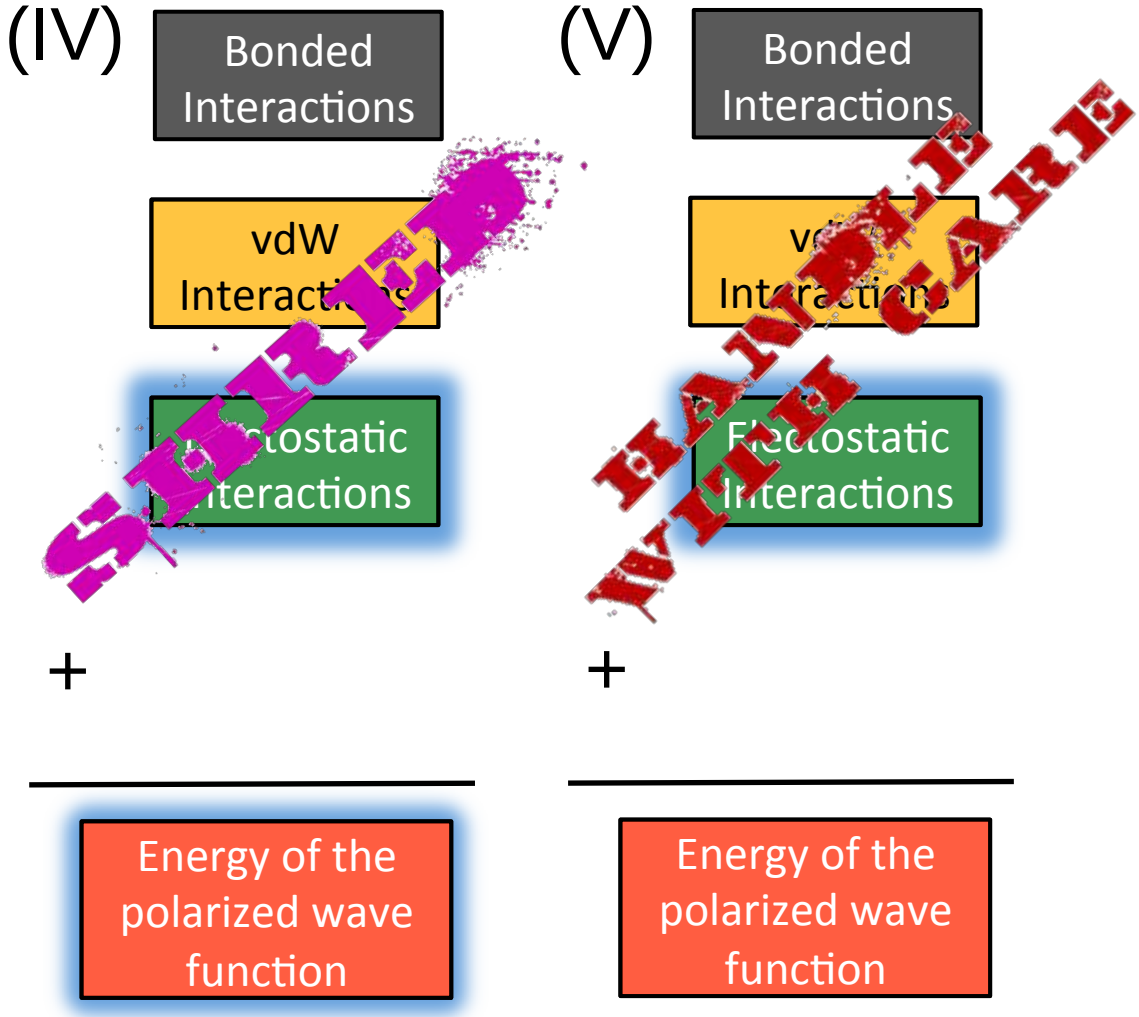
How did this go subtly wrong?

- Scheme (IV) is, again, wrong. (V) is, again, hard if performed with post Hartree-Fock methods.



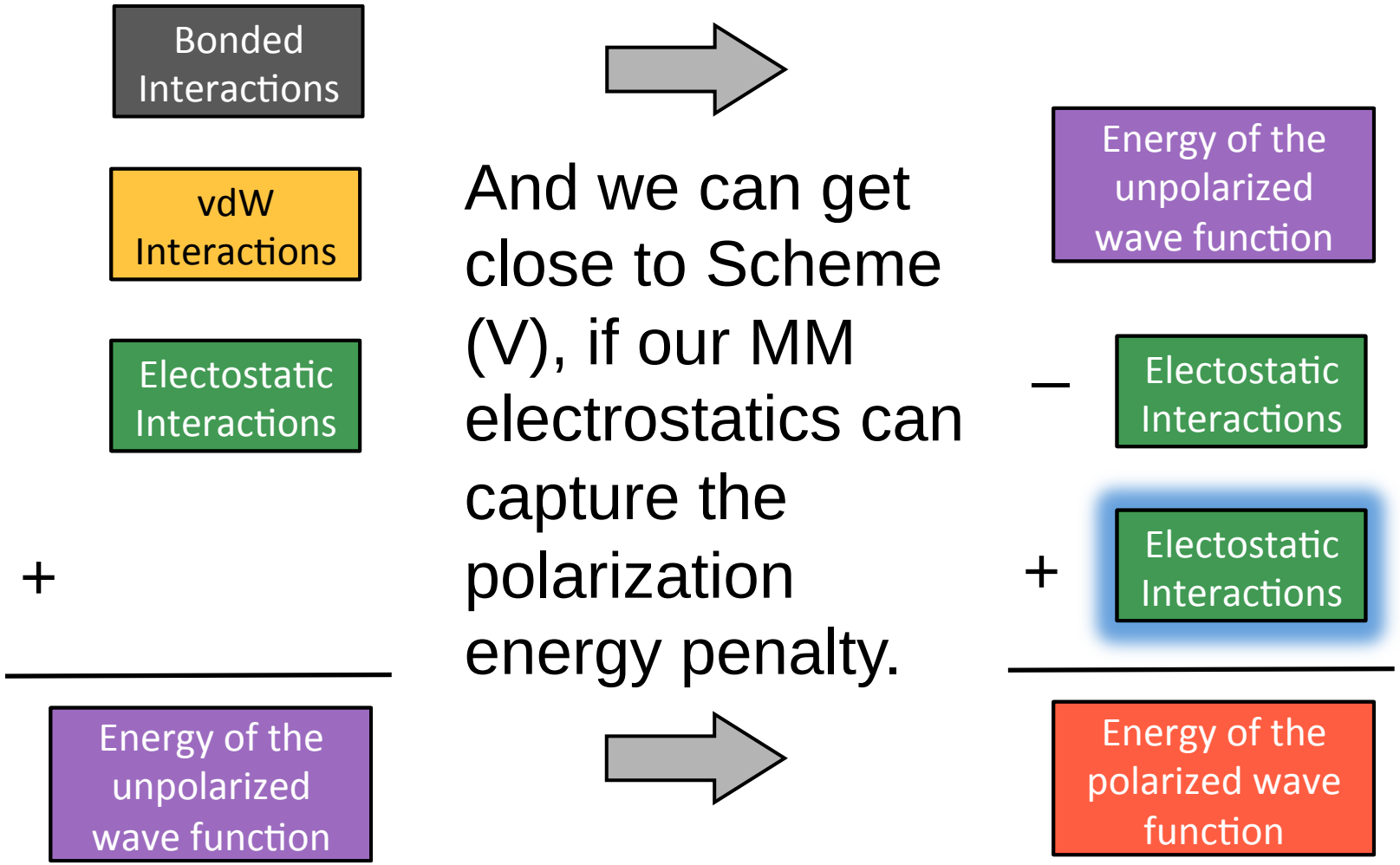
How did this go subtly wrong?

- Scheme (IV) is, again, wrong. (V) is, again, hard if performed with post Hartree-Fock methods.



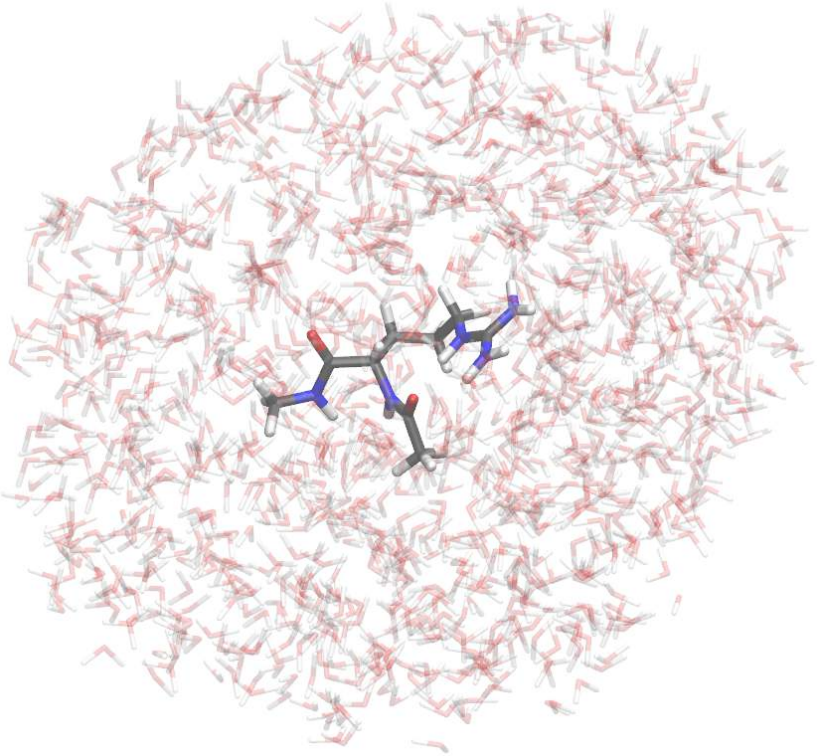
What viable alternative is left?

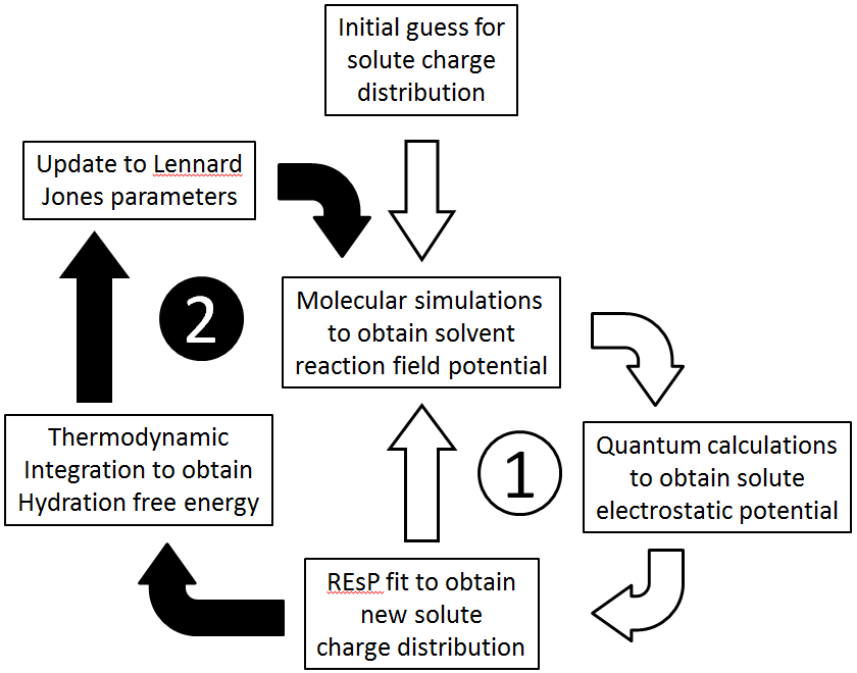
- Back to scheme (II). It's incomplete, but it's straightforward and not yet wrong.



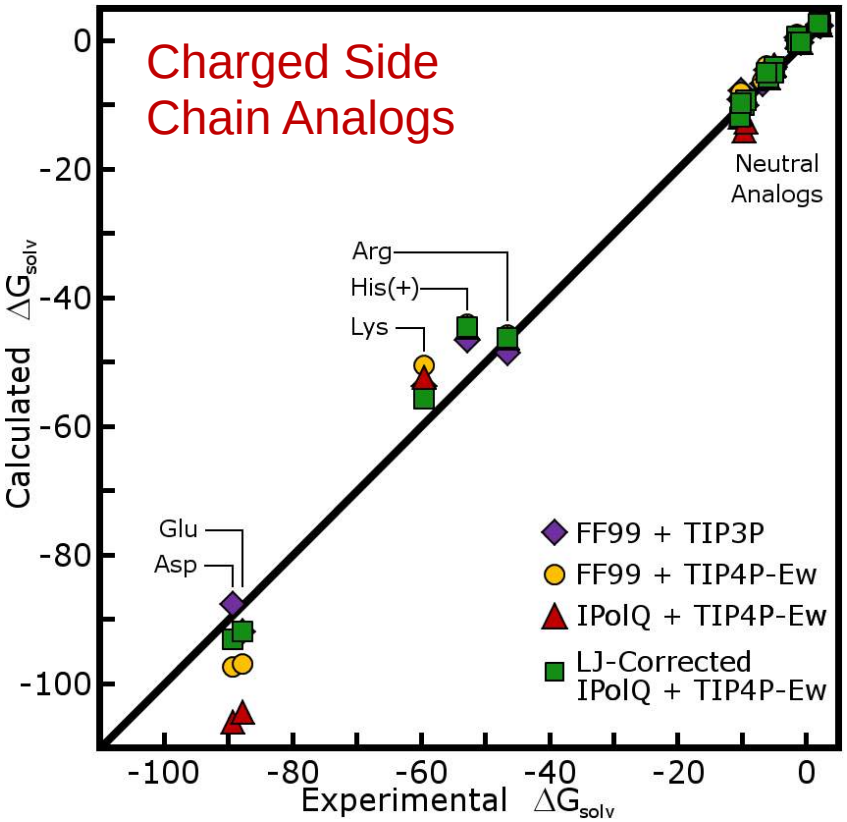
A new charge model: IPolQ

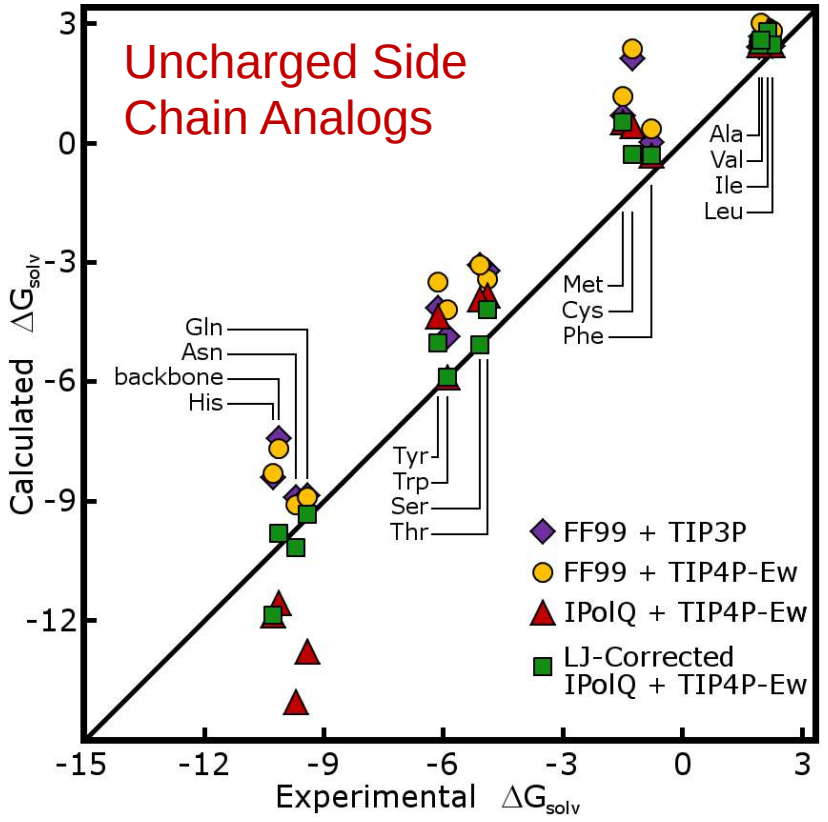
- The target potential is an average of two MP2 / cc-pvTZ calculations:
  - The molecular conformation in vacuum, and…
  - In a reaction field due to a bath of TIP4P-Ew water
- The fitting cycle is iterative





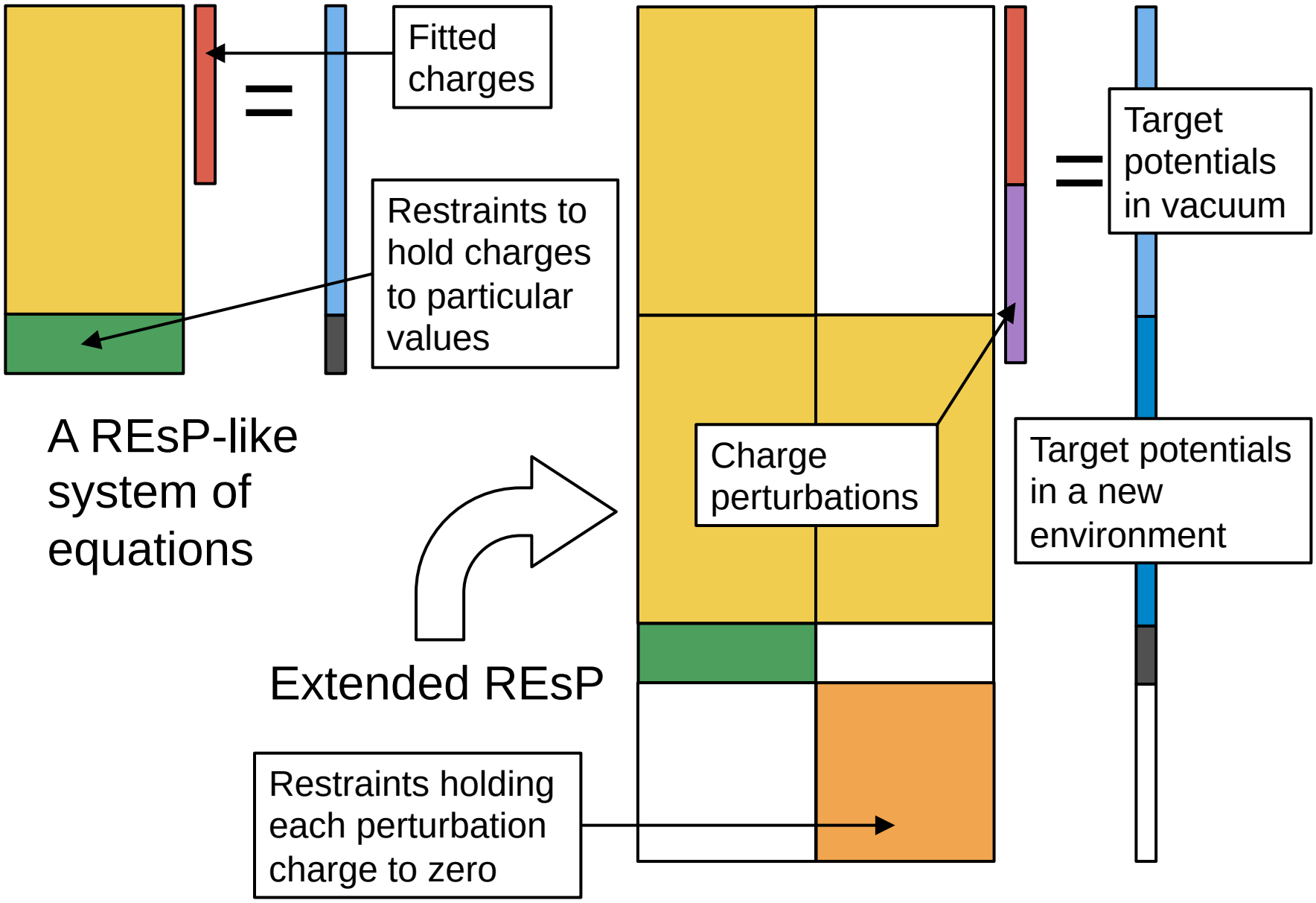
A New Steric Repulsion set





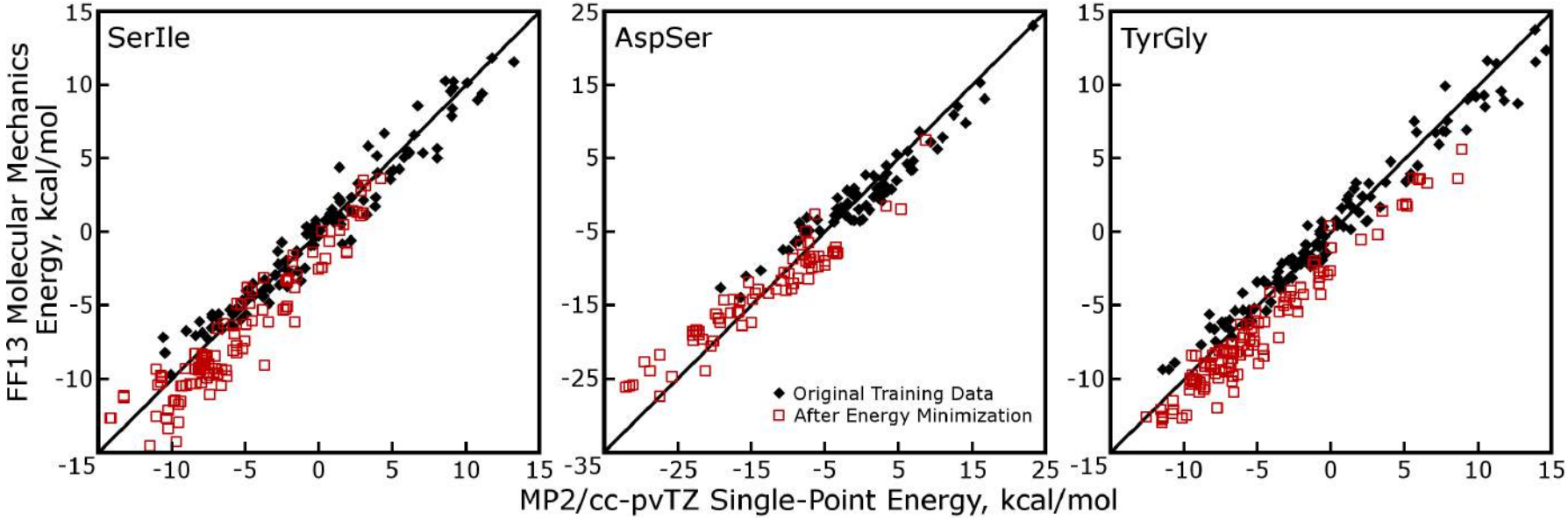
- In order to bring hydration free energies of side chain analogs into agreement with experiment, Lennard-Jones parameters of five polar heavy atom types were adjusted.

Two charge sets in one: Extended REsP



Let the Model Drive

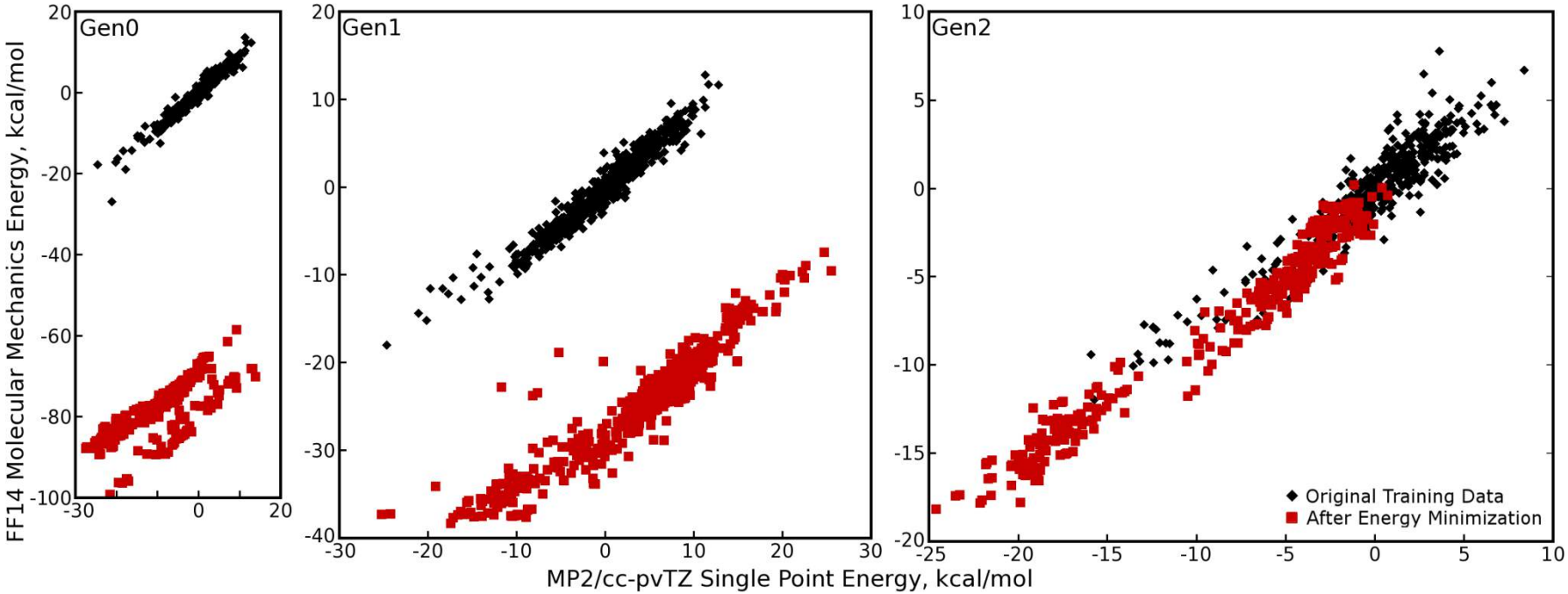
- A model may fit its training data and an independently generated test set. But what will it produce if allowed to drive energy minimizations?



- Any molecular model will make compromises against its benchmark, over- and under-estimating the energy in different regions. In actual simulations, the trend is always toward structures that are scored too favorably.

Show the Model its Mistakes

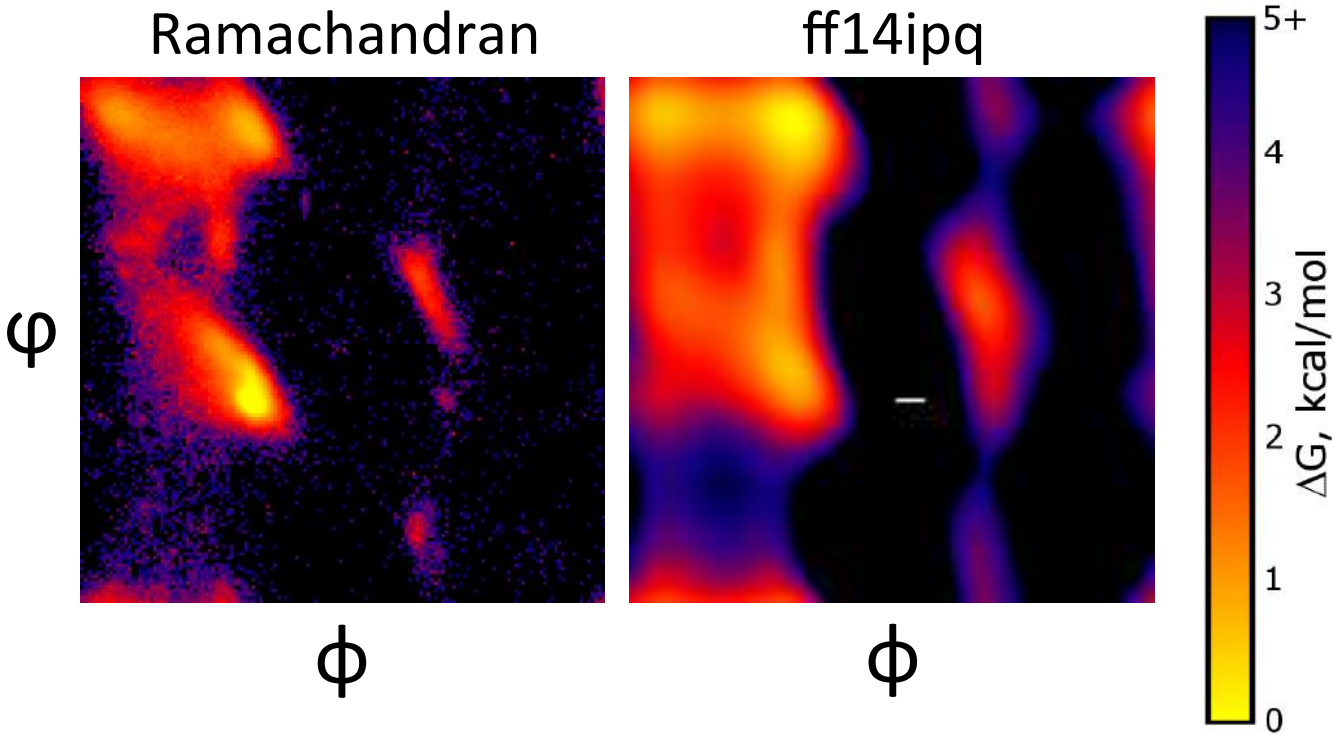
- Sometimes the errors are very pronounced. In those cases, re-introducing the results of model-guided structure optimizations back into the training set produces a much more reliable model.



- In the case of Lysine dipeptide (above), new mistakes are evidently found in the second generation, but the model appears to be fixed by the third.

ff14ipq: The first IPolQ protein model

- First, let's look at the backbone behavior:

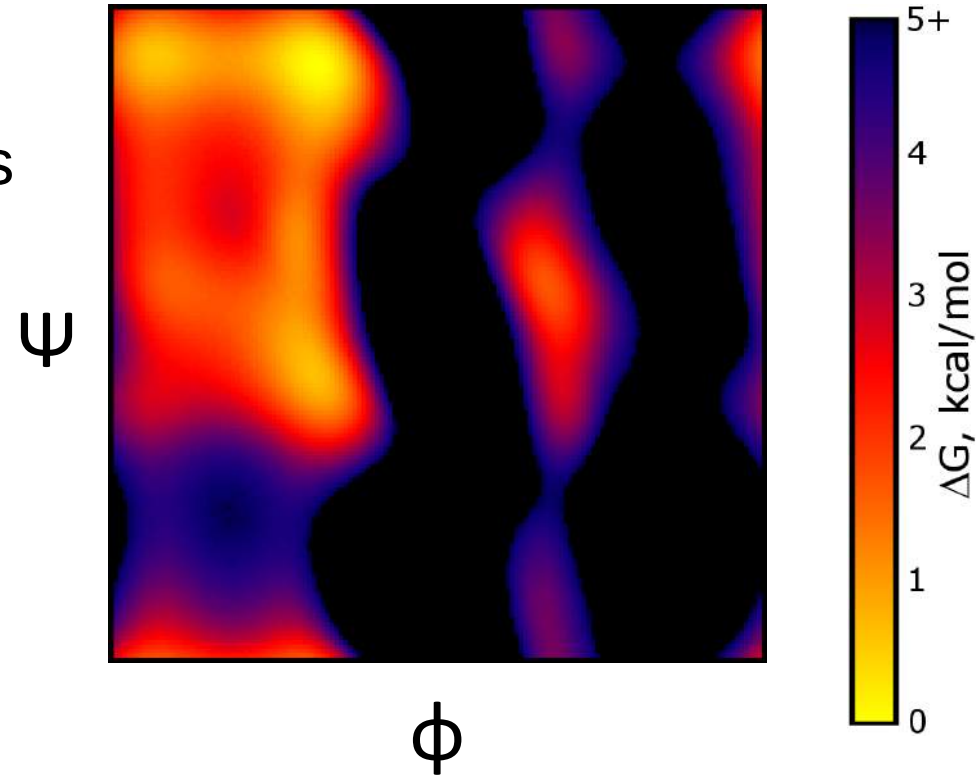


- The first generation of ff14ipq (solution phase charges paired with vacuum phase torsions) puts all of the minima in the right places.

ff14ipq: The first IPolQ protein model

- Subsequent generations refine the Alanine dipeptide PMF:

Generation 1:
- 28,000 MP2 energies
- Amino acid dipeptides, tripeptides, and tetrapeptides
- Among the largest data sets as of early 2013



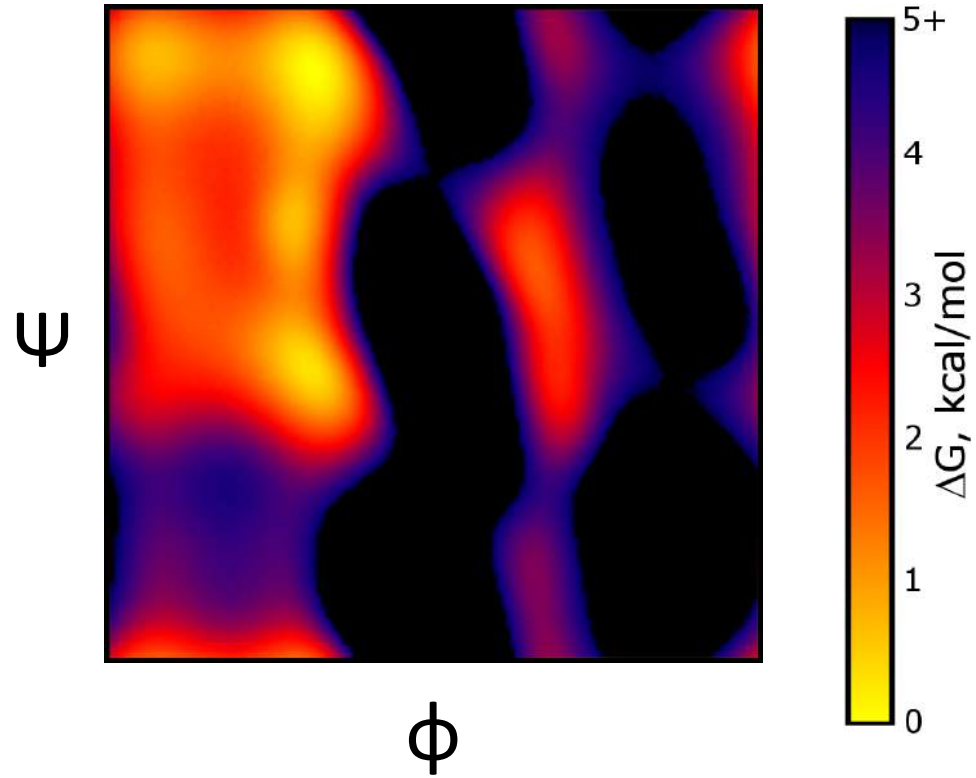
Amber ff14ipq, Generation 1

ff14ipq: The first IPolQ protein model

- Subsequent generations refine the Alanine dipeptide PMF:

Generation 3:

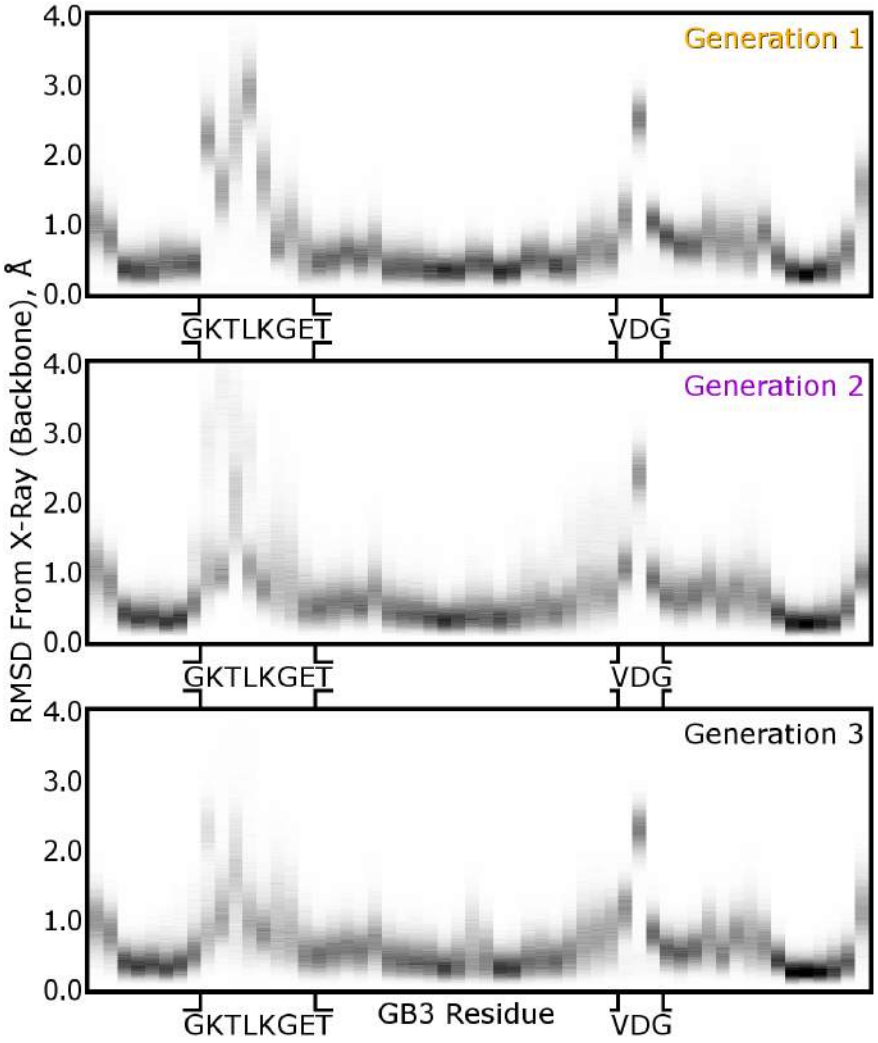
- 65,000 MP2 energies
- Force-field optimized results of generations 1 and 2
- Artificially low minima were also eliminated in many amino acids

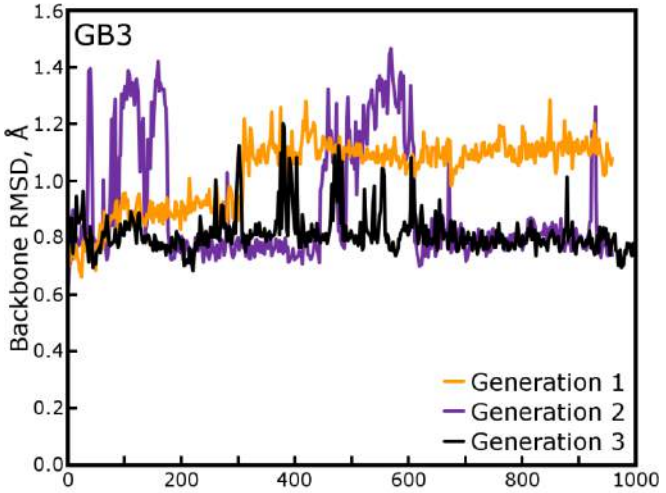


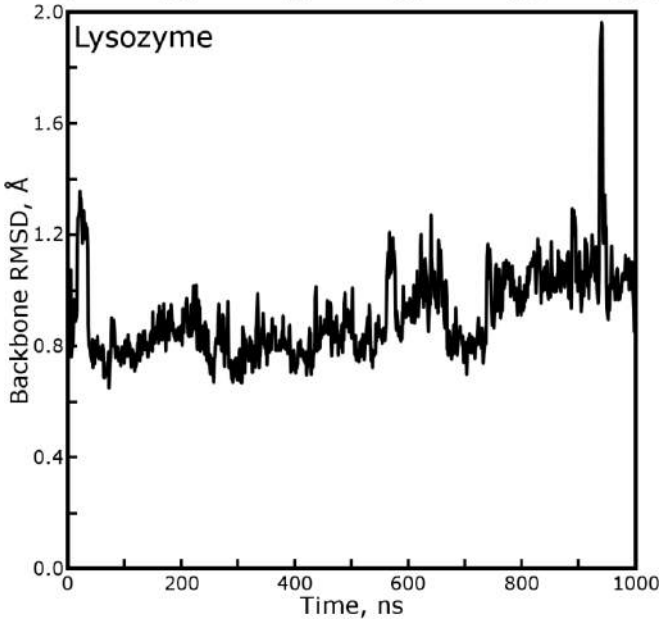
Amber ff14ipq, Generation 3

Protein simulations and stability mature

- What appear to be interest excursions may just be incomplete models.

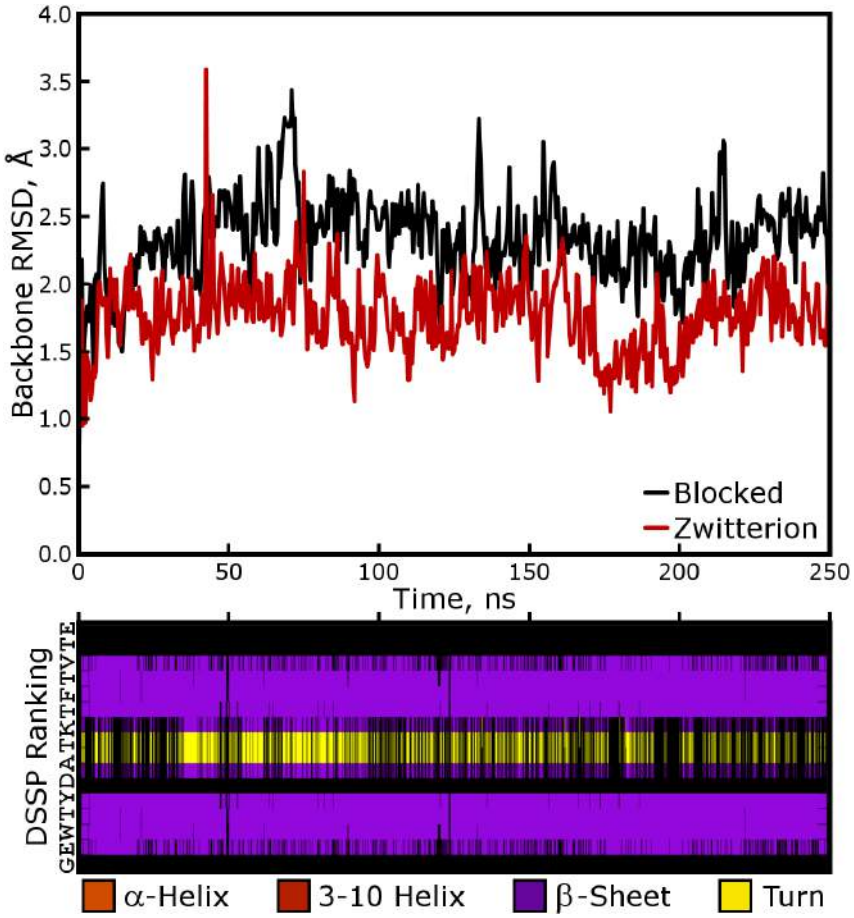




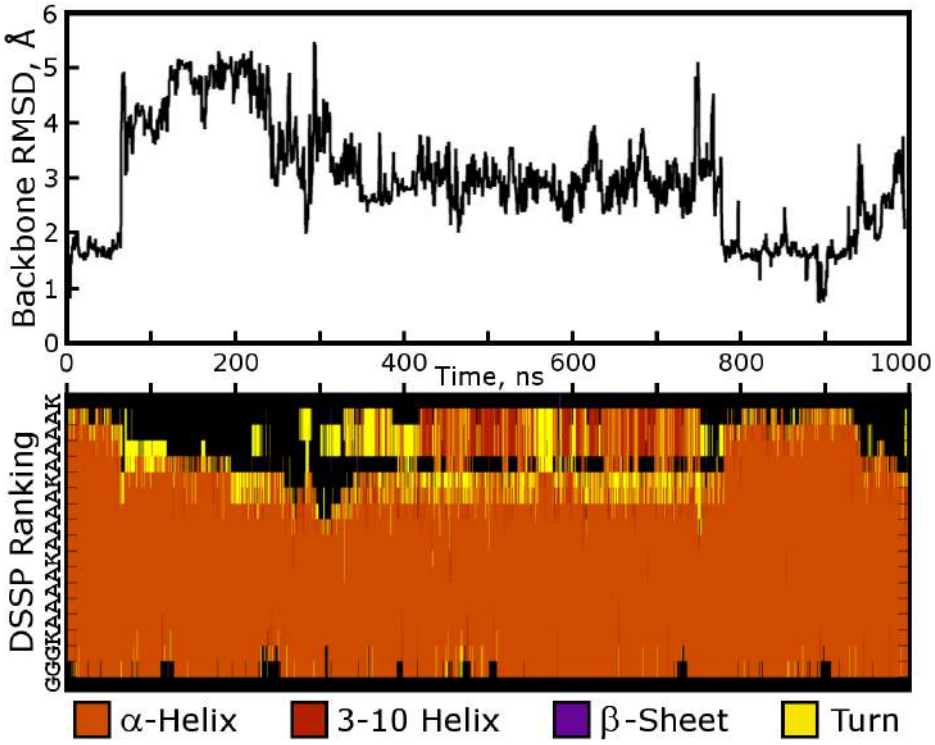


Stability, but too much?

- β-sheets are too stable; in K19, lysine head groups contact the backbone too much.



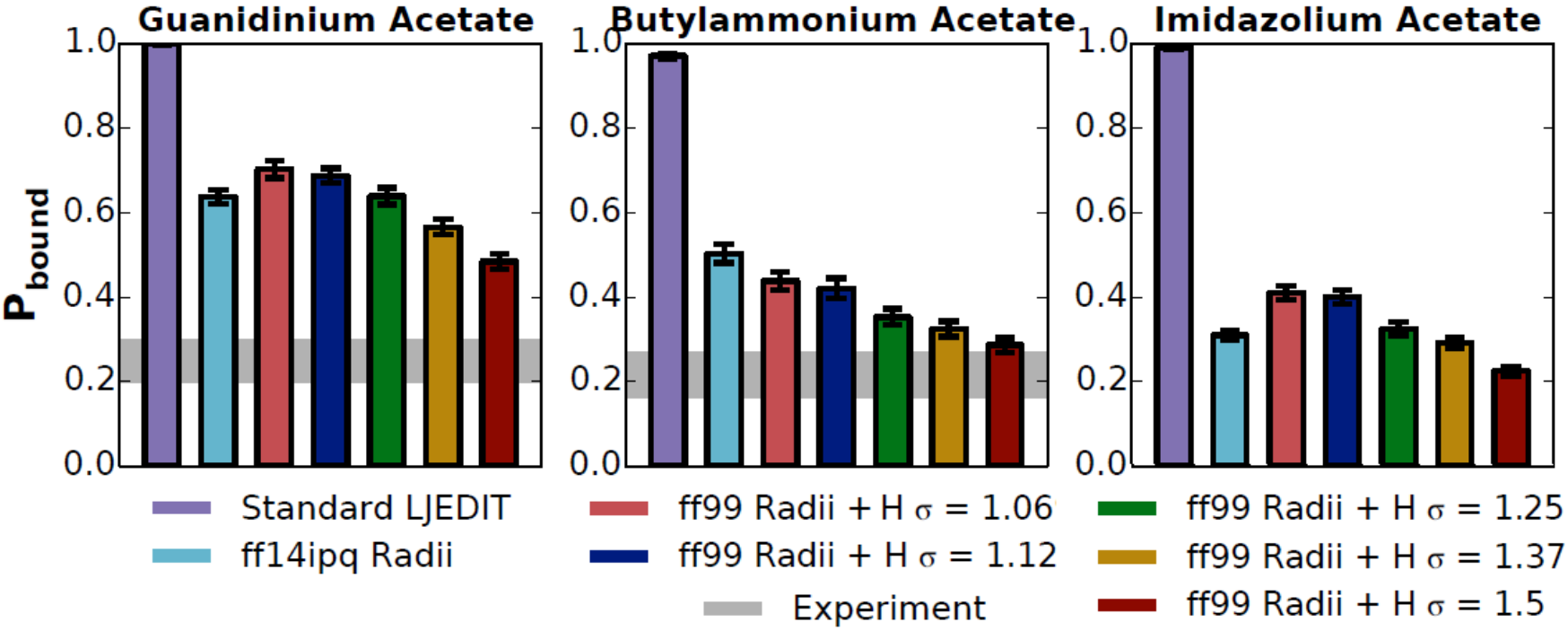
GB1 Hairpin, 300K



K19 Helix, 277K

ff14ipq: What was the problem?

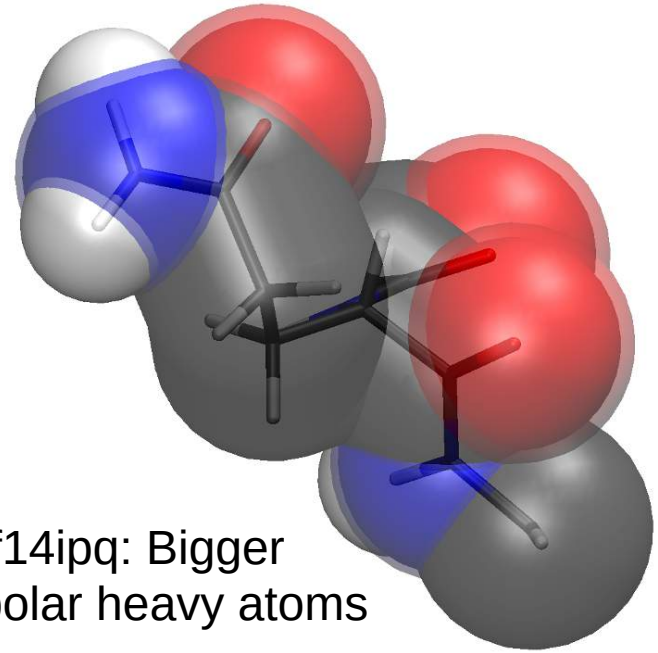
- The salt bridges are too stable: it seems that our polar atom Lennard Jones meddling did this.



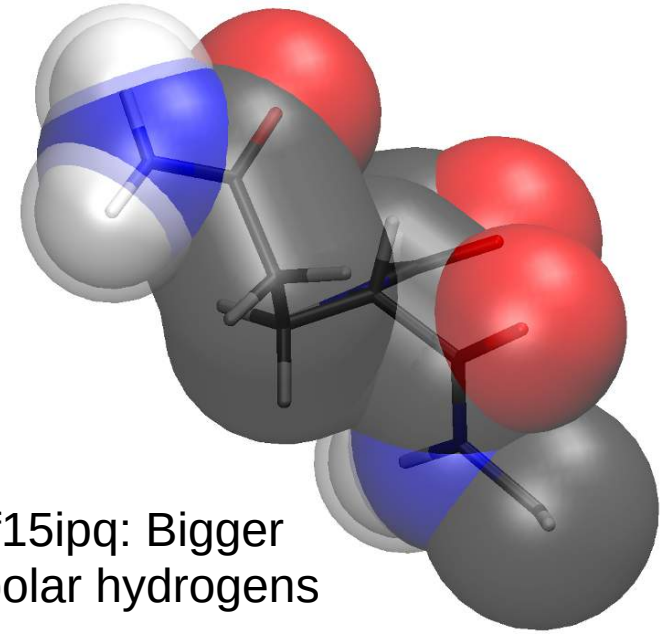
- The over-stabilizations tend to be on the order of 1 kcal/mol, similar to an estimate in the ff14ipq paper.

ff15ipq: The next IPolQ protein model

- Rather than make polar heavy atoms bigger, focus on the hydrogens. Nitrogens are central to the problems in hydration free energies.



ff14ipq: Bigger polar heavy atoms



ff15ipq: Bigger polar hydrogens

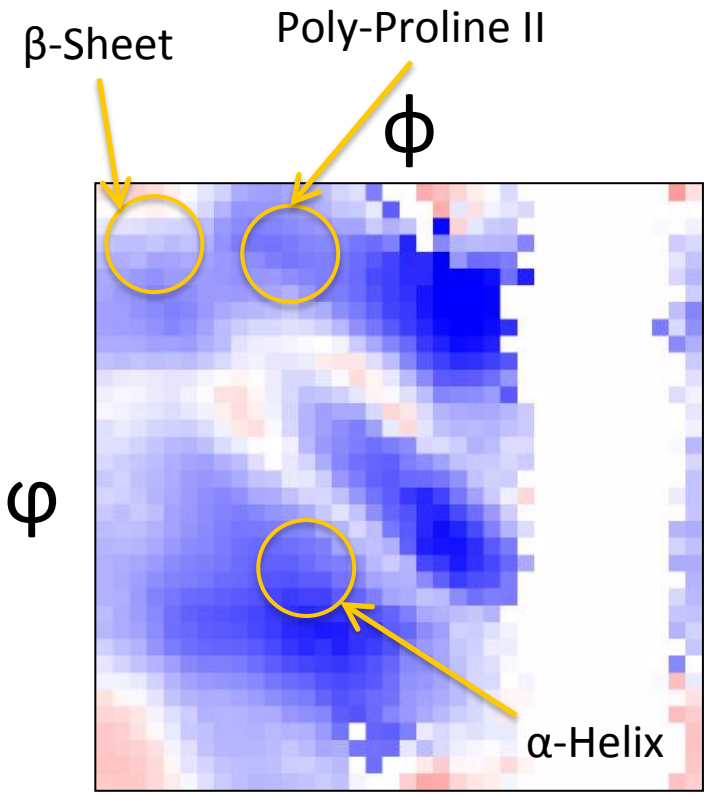
- Introduce angle fitting alongside dihedrals. This helps relieve a lot of strain that would otherwise spill over into fitted torsion parameters.

How effective is angle fitting?

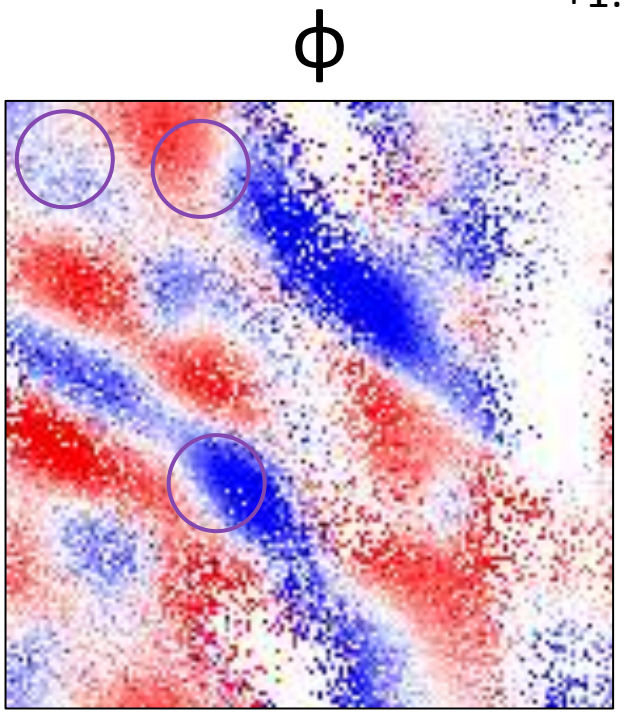
	Ala(5) Scores				K19 (α-helix)	GB1 (β-hairpin)
	Orig	DFT1	DFT2	KLL		
ff14ipq	1.3	2.6	1.5	1.4	Metastable at 277K	Completely stable at 300K
Change polar H Lennard-Jones radius to 1.5Å, refit torsions						
ff15ipq-05	1.5	2.5	1.5	1.5	Unstable	Completely stable
Add N-CA-C, CA-C-N, and C-N-CA for neutral, Glycine, (+) and (-) residues						
ff15ipq-06	0.7	2.0	0.8	0.7	Unstable	Unstable
Decrease polar H Lennard-Jones radius to 1.3Å						
ff15ipq-08	0.7	2.3	1.0	0.7	Too stable at 275K, begins to melt at 315K	Unstable
Add CA-N-H, CA-C=O for neutral, Glycine, (+) and (-) residues						
ff15ipq-09	0.6	2.7	1.0	0.7	Metastable at 275-285K, 300-325K t.b.d	Melts at 300K

How far does automation go?

- Automated parameter creation is bound to hit a wall: even if QM were perfect our ability to mimic it is not.



Difference in QM Methods,
U(DFT) – U(MP2)



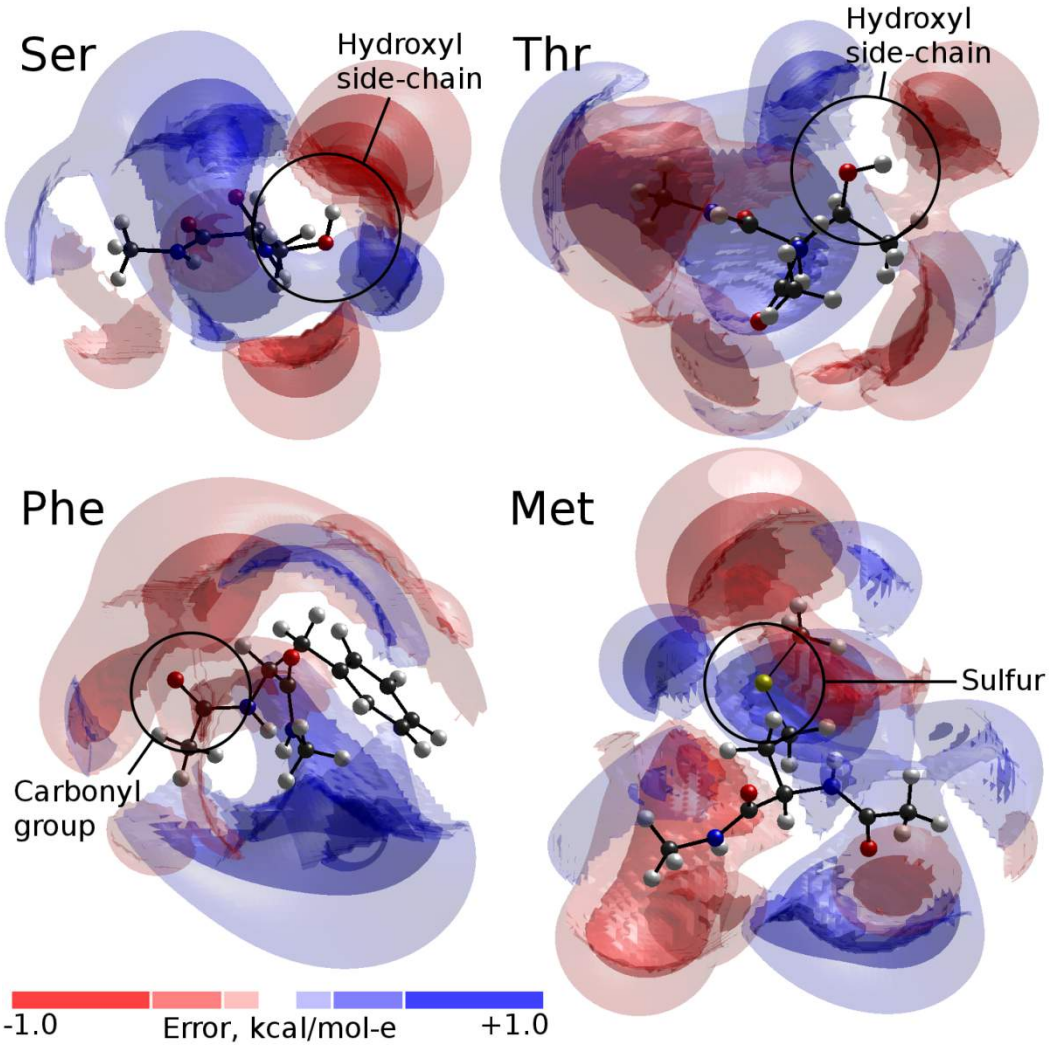
Error,
U(Tinker Toy) – U(MP2)

+1: Destabilized relative to QM

-1: Favored relative to QM

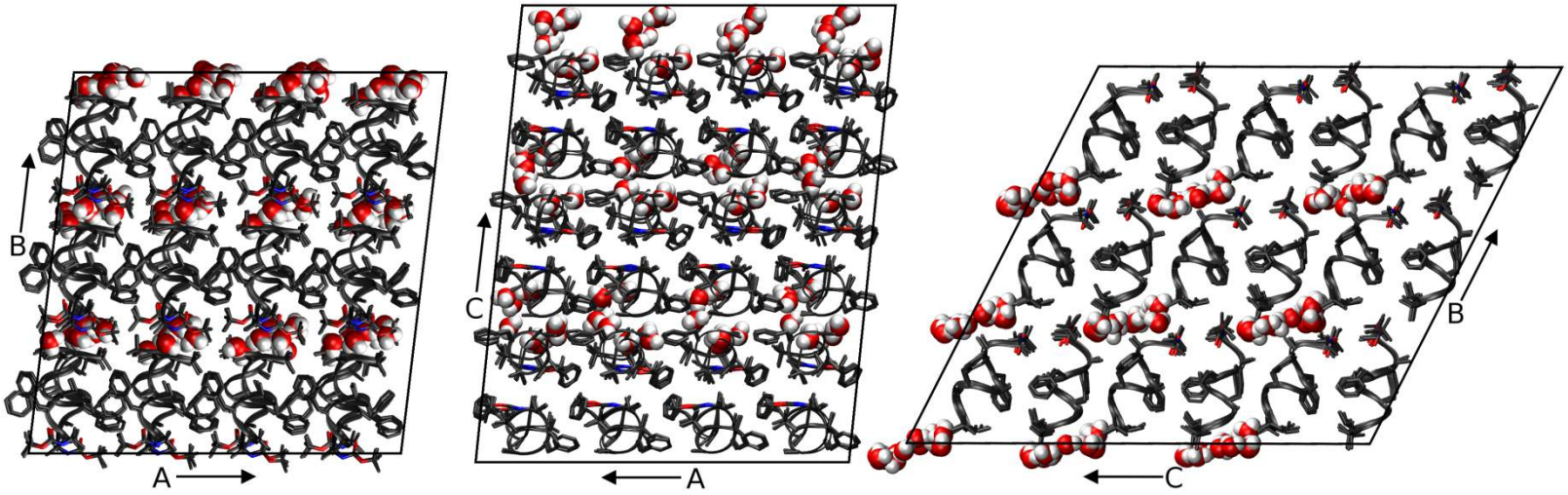
Electrostatic potential fitting

- The nuclear charges fit the quantum target with many compromises



A force field is but a means to an end

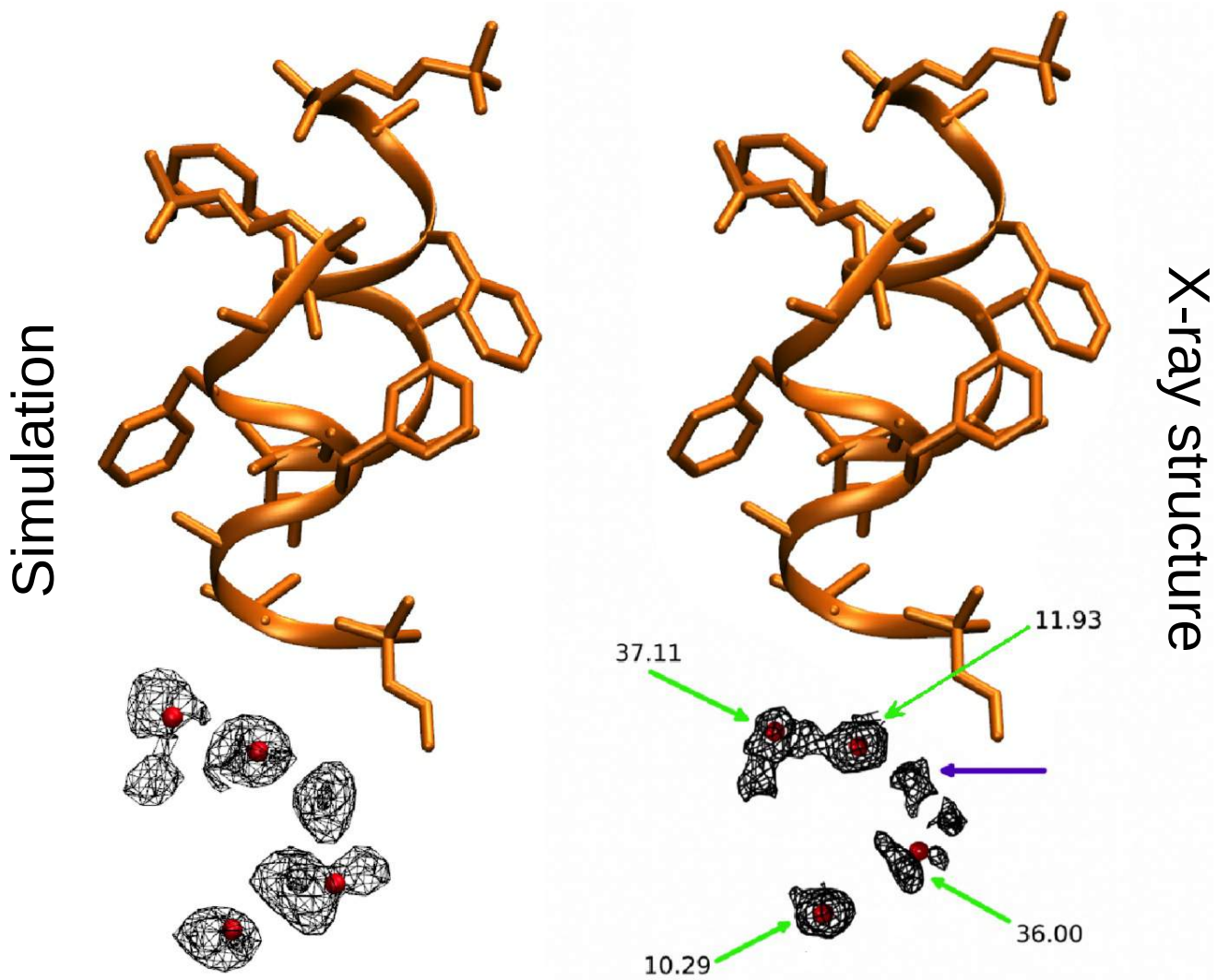
- The fav8 peptide was engineered to study aromatic stacking between helicies in proteins. The peptide crystallizes with all solvent (water) accounted for in the unit cell.



- With the GPU-based pmemd code, this small system was simulated for nearly 10 μs (10 billion time steps).

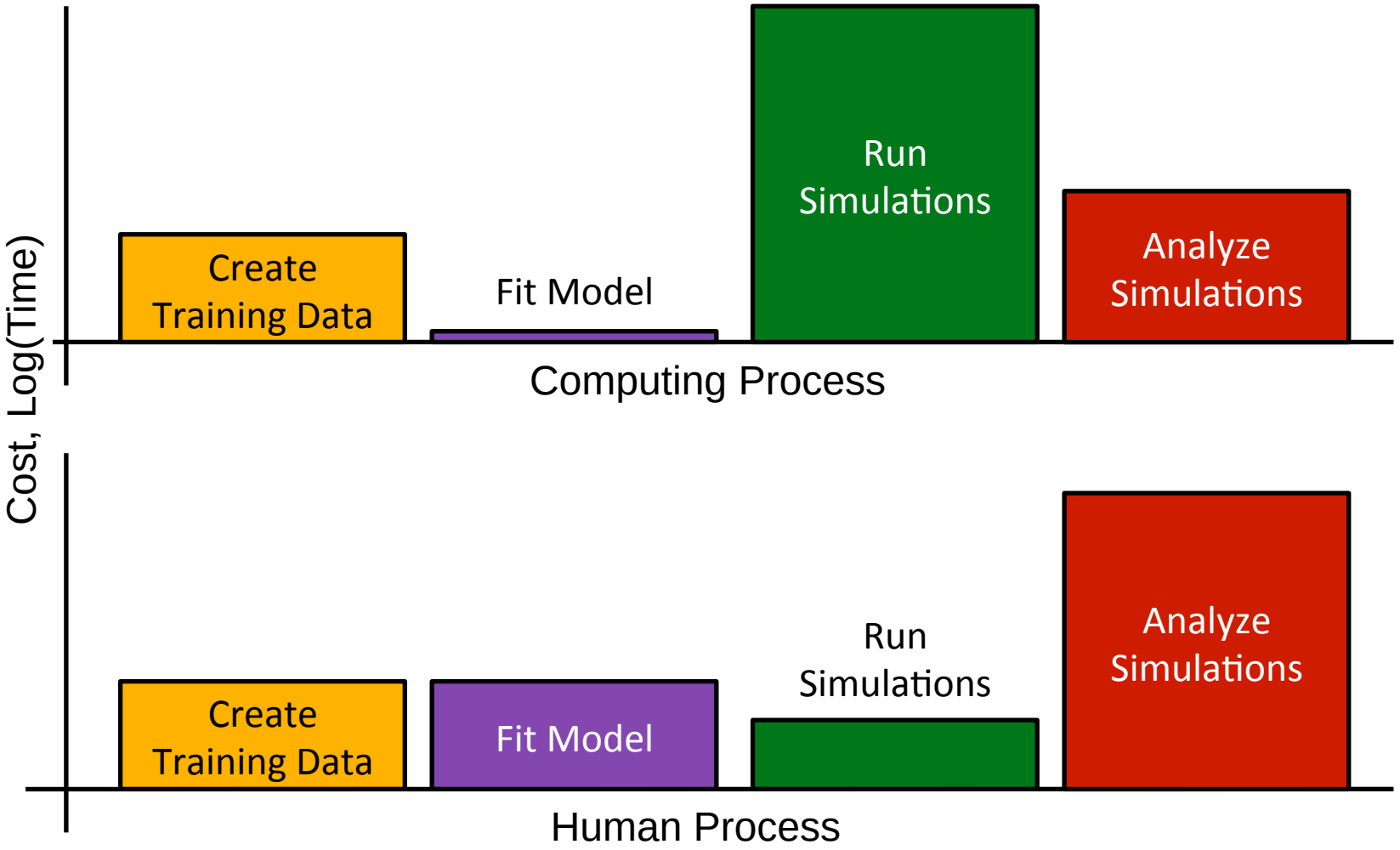
A force field is but a means to an end

- The simulated water density reproduces the natural electron density, not just the refined water positions



MD as a Consumer of HPC

- Unmistakable choke points on the path between parameter development and biochemical simulation:



Acknowledgements

- University of Pittsburgh
  - Karl Debiec
  - Professor Lillian Chong
- Rutgers, the State University of New Jersey
  - Professor David A. Case
  - Pawel Janowski
  - Jason Swails, Dan Roe
- International Business Machines
  - William C. Swope
  - Julia E. Rice