A Brief History of Molecular Dynamics

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

David S. Cerutti
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.

\[
U = \sum_{\text{bond}} k \uparrow p (L \downarrow - L \downarrow 0) r^2 + \sum_{\text{ang}} k \downarrow a (\theta \downarrow - \theta \downarrow 0) r^2 + \sum_{\text{dihe}} k \downarrow v \cos (n \theta \downarrow - \phi \downarrow v) r^2 \\
+ \sum_{i,j} [k \downarrow c q \downarrow i q \downarrow j / r \downarrow ij - 4 \epsilon \downarrow ij (\sigma \downarrow ij / r \downarrow ij ) r^6 + 4 \epsilon \downarrow ij (\sigma \downarrow ij / r \downarrow ij ) r^12]
\]
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.
“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.

(I) Bonded Interactions  (II) Bonded Interactions  (III) Bonded Interactions
    vdw Interactions  vdw Interactions  vdw Interactions
    Electostatic Interactions  Electostatic Interactions  Electostatic Interactions
    + Solvent contributions  +  +
    Energy of the polarized wave function  Energy of the unpolarized wave function  Energy of the unpolarized wave function

• 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.

(I) Bonded Interactions + vdw Interactions + Electrostatic Interactions + Solvent contributions = Energy of the polarized wave function

(II) Bonded Interactions + vdw Interactions + Electrostatic Interactions = Energy of the unpolarized wave function

(III) Bonded Interactions + vdw Interactions + Electrostatic Interactions = Energy of the unpolarized wave function
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.

(IV) + (V)
What viable alternative is left?

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

And we can get close to Scheme (V), if our MM electrostatics can capture the polarization energy penalty.
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
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

A REsP-like system of equations

Extended REsP

Fitted charges

Restraints to hold charges to particular values

Charge perturbations

Restraints holding each perturbation charge to zero

Target potentials in vacuum

Target potentials in a new environment
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:

- Ramachandran

- ff14ipq

• 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 IPoIQ 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.
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.

- 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?

<table>
<thead>
<tr>
<th></th>
<th>Ala(5) Scores</th>
<th>K19 (α-helix)</th>
<th>GB1 (β-hairpin)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Orig</td>
<td>DFT1</td>
<td>DFT2</td>
</tr>
<tr>
<td>ff14ipq</td>
<td>1.3</td>
<td>2.6</td>
<td>1.5</td>
</tr>
<tr>
<td>Change polar H Lennard-Jones radius to 1.5Å, refit torsions</td>
<td>ff15ipq-05</td>
<td>1.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Add N-CA-C, CA-C-N, and C-N-CA for neutral, Glycine, (+) and (-) residues</td>
<td>ff15ipq-06</td>
<td>0.7</td>
<td>2.0</td>
</tr>
<tr>
<td>Decrease polar H Lennard-Jones radius to 1.3Å</td>
<td>ff15ipq-08</td>
<td>0.7</td>
<td>2.3</td>
</tr>
<tr>
<td>Add CA-N-H, CA-C=O for neutral, Glycine, (+) and (-) residues</td>
<td>ff15ipq-09</td>
<td>0.6</td>
<td>2.7</td>
</tr>
</tbody>
</table>
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.

\[ \text{Difference in QM Methods, } \Delta U(\text{DFT}) - U(\text{MP2}) \]
\[ \text{Poly-Proline II} \]

\[ \text{β-Sheet} \]
\[ \text{α-Helix} \]

\[ +1: \text{ Destabilized relative to QM} \]

\[ -1: \text{ Favored relative to QM} \]

\[ \phi \]
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 helicities in proteins. The peptide crystallizes with all solvent (water) accounted for in the unit cell.

- With the GPU-based pmemod 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

Simulation

X-ray structure
MD as a Consumer of HPC

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

Computing Process

<table>
<thead>
<tr>
<th>Create Training Data</th>
<th>Fit Model</th>
<th>Run Simulations</th>
<th>Analyze Simulations</th>
</tr>
</thead>
</table>

Human Process

<table>
<thead>
<tr>
<th>Create Training Data</th>
<th>Fit Model</th>
<th>Run Simulations</th>
<th>Analyze Simulations</th>
</tr>
</thead>
</table>
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