

Background on computational chemistry

Tom Jackson

August 24, 2022

Abstract

This is a (disorganized) summary of entry-level material I've compiled for my own use.

1	First pass at an outline	2
1	Introduction	2
1.1	What to calculate?	2
2	Classical treatments	2
2.1	Molecular mechanics	2
2.2	Molecular dynamics	3
2.3	Proteins, specifically	3
2.4	Larger scales	3
3	Quantum treatments	3
2	Misc. topics	5
1	Renormalization group for polymers in solution	5

Chapter 1

First pass at an outline

1 Introduction

Approaches to fundamental complexity of solving the quantum many-body problem.

Nobels: 1998 to [Walter Kohn](#) for DFT and to [John Pople](#) for ab initio methods with Gaussian orbitals. 2013 to [Martin Karplus](#), [Michael Levitt](#) and [Arieh Warshel](#) for “multiscale models” — what are the relevant scales?

1.1 What to calculate?

Fundamental premise of chemistry is that structure is the key to everything. Is it too much of a stretch to lump conformational dependence of biomolecules in with that?

Charge density (multipole moments). Spectra. Reactivity — reaction pathways?

Quantitative structure-activity (or “property,” in general) relationship ([QSAR](#)): pure machine learning approach for empirical relationships, with all the usual caveats. “Matched molecular pair analysis” ([MMPA](#)) = restricting the problem to effects of single-atom substitution or other small structural changes.

2 Classical treatments

2.1 Molecular mechanics

Nuclear configuration as (heuristically) classical “balls on springs,” with configurational potentials/[force fields](#) incorporating quantum effects.

- *Covalent bonding terms*. Natural coordinates are bond length, angle (three-atom, relative to other bonds) and dihedral angle/torsion.
 - Morse potential (diatomic approx) for length; harmonic approximation is good at biological temperatures.
 - Torsional terms have multiple minima; also “improper” (=fudge?) terms to e.g. keep benzene rings planar.
- *Non-bonding terms*. Source of computational bottleneck; $O(N^2)$ scaling versus $O(N)$.
 - Electrostatic: Important in practice for macromolecular stability; r^{-1} falloff, at least before effective screening. [Particle-mesh Ewald](#) and [fast multipole](#) approaches similar in that they use a mean-field-ish effective approximation for long-range contributions, treating short-range interactions exactly.

- van der Waals: Lennard-Jones potential, frequently truncated at short distances. Falls off as r^{-6} at least.
- *Solvent*. **Implicit solvation** models solvent as continuum with effective properties. Can also add **explicit waters**.

To reiterate, specifying a “force field” involves mass, (partial) charge (and maybe polarizability) and LJ parameters for each atom, and equilibrium bond potential parameters for each pair and triplet of atoms. Atoms in different groups/environments treated differently. Of course many bells and whistles in **practical implementations**.

Inputs are largely empirical: geometry from x-ray structures; energetics from observed surface and hydration energies. Feedback loop in tweaking parameters to match observations. Can also make use of ab initio QM calculations.

Bigger question is generalizability of force field parameters, in that they can be set from “training data” dissimilar to what they’re applied to (e.g. gas- vs. solid-phase, small-molecule vs macromolecules). How desirable is parameter interpretability? **IFF** held as a relatively rigorous approach to force field design.

Immediate application for approx Bayes/experimental design.

2.2 Molecular dynamics

Dynamics needed to recover properties at $T > 0$. Again, purely classical equations of motion.

2.3 Proteins, specifically

2.4 Larger scales

Claim shift to larger scales accompanied by a shift from energy dominance (QM) to entropy dominance (nanostructures and biology).

Continuum methods like Navier-Stokes, finite-element methods for elasticity.

Dissipative particle dynamics (DPD) — distinct from Langevin in that momentum is still conserved.

Phase-field models — Goldenfeld RG ref.

3 Quantum treatments

- *Born-Oppenheimer approximation*: electronic and nuclear degrees of freedom are independent; nuclear positions may be regarded as static/quenched on electronic timescales; more specifically, treat nuclear kinetic energy as a perturbation governed by ratio of electron to nuclear masses.
- *Relativistic effects* can generally be neglected outside of heavy atoms (transition metals).
- *Hartree-Fock*, self-consistent field or molecular orbital approximation: includes the above, and adds the assumption that electron degrees of freedom are non-interacting (with each other); i.e. many-body wavefunction is a single Slater det of one-particle wavefunctions, the molecular orbitals.

Can also be viewed as a mean-field method: two-body Coulomb term is replaced by a single-particle approximation, the “Fock matrix.” Rest is variational perturbation theory: write molecular orbitals as linear combinations in some basis, and find coefficients minimizing approximate one-particle Hamiltonian.

Basis choices: “linear combination of atomic orbitals”

What sets effective charge?

- Hydrogenic/Schrodinger orbitals: from solving the single-particle Coulomb problem. Not used.
 - Slater orbitals: not to be confused with Slater determinants. Retain the $r^\ell e^{-\alpha r}$ radial dependence of hydrogenic orbitals, but drop the Laguerre polynomial (so no radial nodes). Also disused, in favor of
 - Gaussian orbitals (or “primitives,” since we’re not solving the one-particle Schrodinger equation anymore): $r^\ell e^{-\alpha r^2}$ dependence; several Gaussians can approximate more physical radial dependence, plus have the key property that overlap integrals are analytically tractable — specifically, product of two Gaussian orbitals at different loci can be written as a finite sum of Gaussians centered on the line connecting loci. Means all multi-center integrals can be decomposed to single-center ones; speedup far outweighs cost of larger basis. Note that neither Slater nor Gaussian primitives have radial nodes; idea is you reproduce that by taking a combination with opposing signs.
 - For same reason (ease of integration), most software uses a cartesian Gaussian basis (i.e. $x^i y^j z^k e^{-\alpha(x^2+y^2+z^2)}$), rather than the physical spherical harmonics.
 - “Contractions” refers to linear combinations of Gaussians that are treated as a unit (same coefficient) during variational optimization.
 - “Polarization” from including basis elements with higher ℓ than in occupied orbitals. “Diffuse” elements likewise refers to small- α Gaussians, meant to capture tail dependence.
- *Møller-Plesset perturbation theory* treats neglected two-body term in Rayleigh-Schrodinger perturbation theory (note still only using a single Slater det). Not guaranteed to be convergent...

What’s the breakdown between lookup tables, analytic fomulae and numerical integration?

Chapter 2

Misc. topics

1 Renormalization group for polymers in solution

“Protein folding and ligand binding are thermodynamically closer to crystallization, or liquid-solid transitions as these processes represent freezing of mobile molecules in condensed media.”

What’s new since De Gennes, Des Cloizeaux et al.?

References