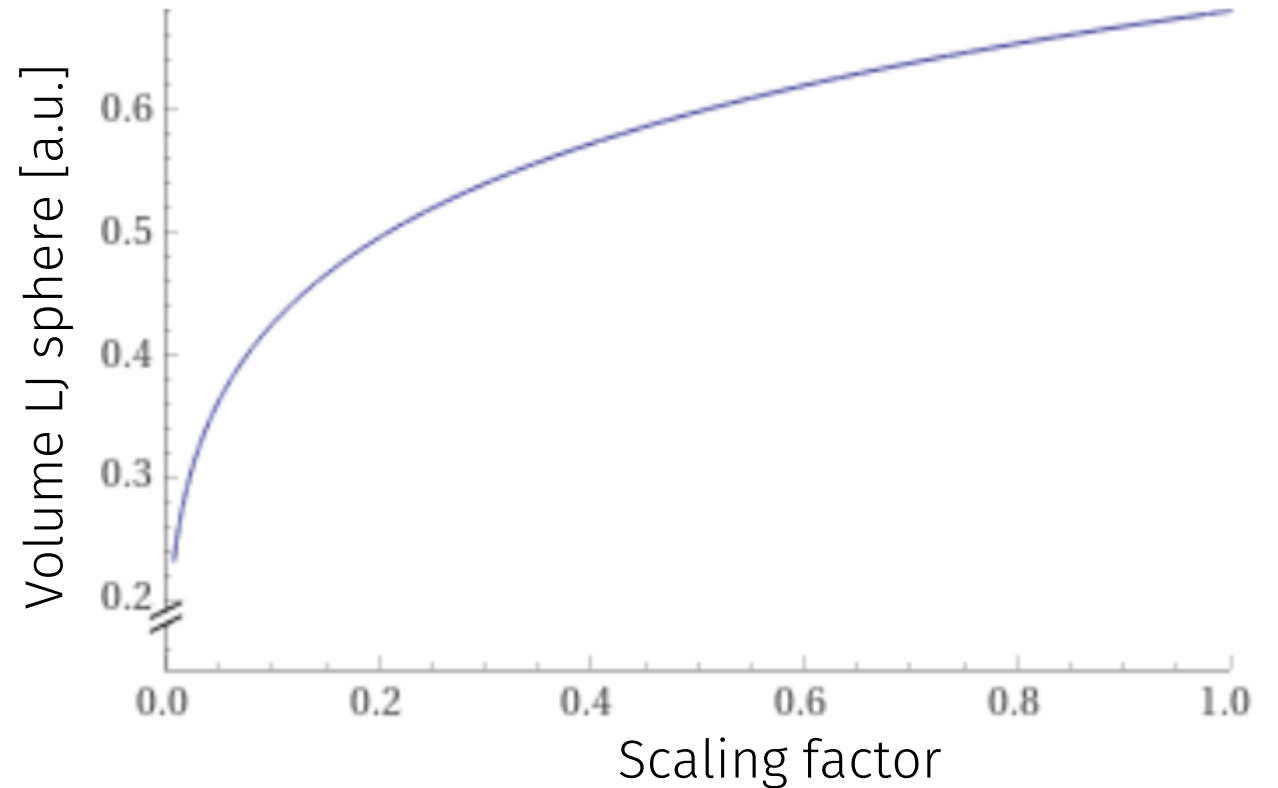Linear pathway not necessarily efficient / rarely "effectively linear"

$$V_{\mathrm{LJ}}(r) = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right]$$

- Avoid constrained/restrained configurations   Why?

- Choose low-change path: large changes mean large derivatives   Why is that bad?

- Change parameters to create effectively linear results

- Restrict number of intermediates (=mixed states)

- Beware electrostatics: keep net charge   Why?

Consider NVT:
Why is the free energy of solvation NOT simply the free energy differences with solute-solvent interactions turned off?
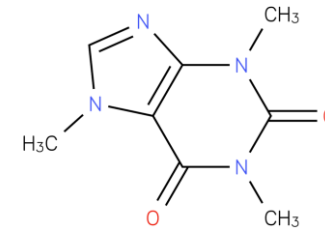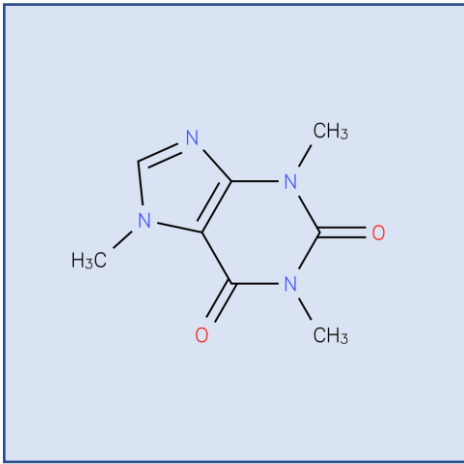
Empirical Force Fields for Biological Macromolecules: Overview and Issues

ALEXANDER D. MACKERELL, JR.

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland,
20 Penn Street, Baltimore, Maryland 21201

Abstract: Empirical force field-based studies of biological macromolecules are becoming a common tool for investigating their structure–activity relationships at an atomic level of detail. Such studies facilitate interpretation of experimental data and allow for information not readily accessible to experimental methods to be obtained. A large part of the success of empirical force field-based methods is the quality of the force fields combined with the algorithmic advances that allow for more accurate reproduction of experimental observables. Presented is an overview of the issues associated with the development and application of empirical force fields to biomolecular systems. This is followed by a summary of the force fields commonly applied to the different classes of biomolecules; proteins, nucleic acids, lipids, and carbohydrates. In addition, issues associated with computational studies on "heterogeneous" biomolecular systems and the transferability of force fields to a wide range of organic molecules of pharmacological interest are discussed.

Molecular Dynamics

Ergodicity: time average = ensemble average

$$\frac{\mathbf{F}}{m} = \mathbf{a} = \frac{d}{dt}\mathbf{v} = \frac{d^2}{dt^2}\mathbf{r}$$

Stepwise integration
- Long steps
- Reversible = conservation of energy
- No high derivatives
- Access to instantaneous velocities
- Metric: global error: long term error from integration

- Two Taylor expansions, one forward, one backward

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\Delta t^2 + \frac{1}{6}\mathbf{j}(t)\Delta t^3 + \mathcal{O}(\Delta t^4)$$

$$\mathbf{x}(t - \Delta t) = \mathbf{x}(t) - \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\Delta t^2 - \frac{1}{6}\mathbf{j}(t)\Delta t^3 + \mathcal{O}(\Delta t^4)$$

- Add them up

$$\mathbf{x}(t + \Delta t) + \mathbf{x}(t - \Delta t) = 2\mathbf{x}(t) + \mathbf{a}(t)\Delta t^2 + \mathcal{O}(\Delta t^4)$$

$$\mathbf{x}(t + \Delta t) + \mathbf{x}(t - \Delta t) = 2\mathbf{x}(t) + \mathbf{a}(t)\Delta t^2 + \mathcal{O}(\Delta t^4)$$

- No instantaneous velocities: finite differences required
- Needs two past positions and initial boundary conditions

$$\mathbf{x}(t_0 + \Delta t) = \mathbf{x}(t_0) + \mathbf{v}(t_0)\Delta t + \mathbf{a}(t_0)\Delta t^2$$

- Global error $\mathcal{O}(\Delta t^2)$
- Time reversible

- Like Verlet, explicitly calculate velocities via a half-step

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\Delta t^2$$

$$\mathbf{v}(t + \frac{1}{2}\Delta t) = \mathbf{v}(t) + \frac{1}{2}\mathbf{a}(t)\Delta t$$

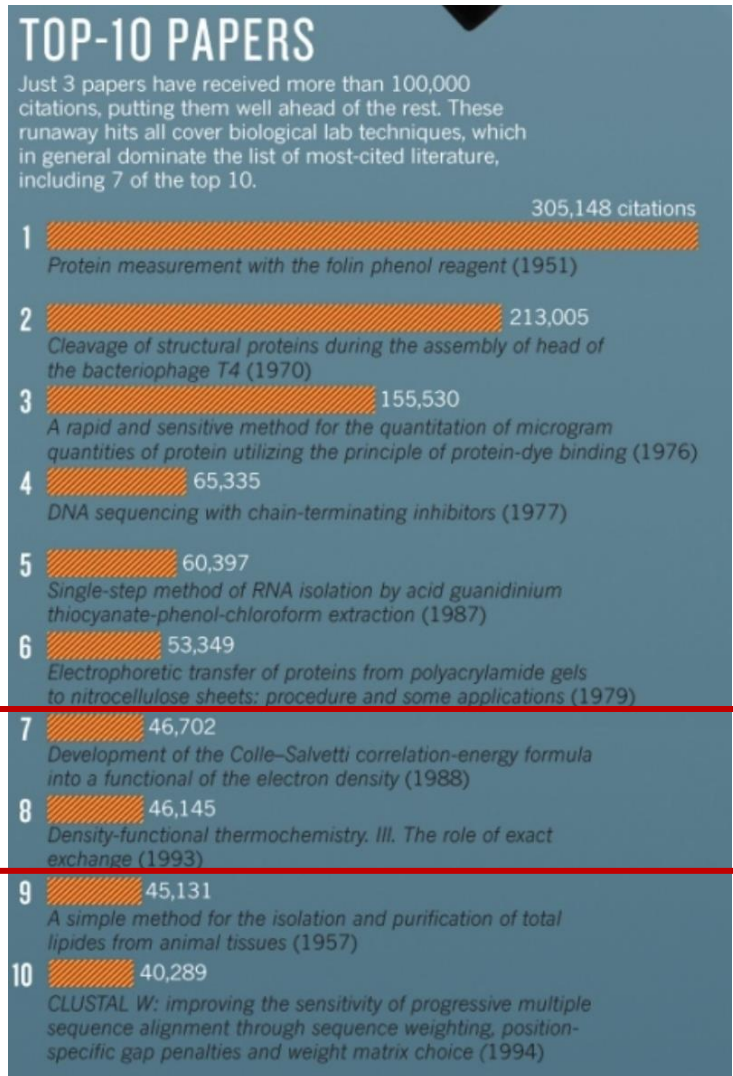$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t + \frac{1}{2}\Delta t) + \frac{1}{2}\mathbf{a}(t + \Delta t)\Delta t$$

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \frac{1}{2}\left[\mathbf{a}(t + \Delta t) + \mathbf{a}(t)\right]\Delta t$$

- Initial velocities: flying ice cube
- Interfaces: thermalisation
- Ensembles
    - NVT, NPT, NVE, …
- Periodic boundary conditions and minimum image convention

Density Functional Theory

TOP-10 PAPERS
Just 3 papers have received more than 100,000 citations, putting them well ahead of the rest. These runaway hits all cover biological lab techniques, which in general dominate the list of most-cited literature, including 7 of the top 10.

1 — 305,148 citations
Protein measurement with the folin phenol reagent (1951)

2 — 213,005
Cleavage of structural proteins during the assembly of head of the bacteriophage T4 (1970)

3 — 155,530
A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding (1976)

4 — 65,335
DNA sequencing with chain-terminating inhibitors (1977)

5 — 60,397
Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction (1987)

6 — 53,349
Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications (1979)

7 — 46,702
Development of the Colle–Salvetti correlation-energy formula into a functional of the electron density (1988)

8 — 46,145
Density-functional thermochemistry. III. The role of exact exchange (1993)

9 — 45,131
A simple method for the isolation and purification of total lipides from animal tissues (1957)

10 — 40,289
CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice (1994)

Density Functional Theory extremely influential

1926: Schrödinger's equation
1927: Thomas-Fermi
1965: Kohn-Sham

R. Van Noorden, B. Maher, R. Nuzzo, *Nature*, **2014**.

Known: Obtaining observables from quantum mechanics (Born-Oppenheimer)

1. System defines Hamiltonian $\hat{H} = \hat{H}(Z_i, \mathbf{R}_i, N_e, \sigma)$
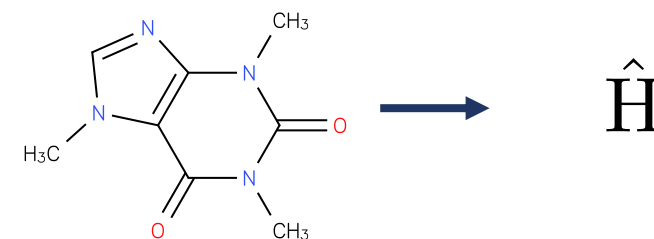
    $$\hat{H} = \hat{T} + \hat{V}$$

    $$\hat{H}$$

2. Solving Schrödinger's equation yields wave function Ψ

    $$\hat{H}\Psi = E\Psi$$

    $$\hat{H} \longrightarrow \Psi$$

3. Operator yields observable $\hat{A}$

    $$\langle\hat{A}\rangle = \langle\Psi|\hat{A}|\Psi\rangle$$
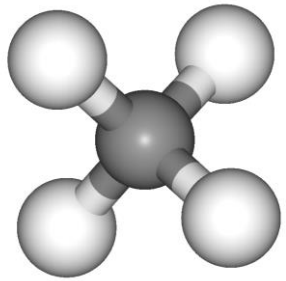
    $$\Psi \longrightarrow \text{property}$$

System defines Hamilton operator $\hat{H}$

$$\hat{H} = \hat{T} + \hat{V} = \textcolor{red}{\hat{T}_e} + \textcolor{blue}{\hat{V}_{en}} + \textcolor{orange}{\hat{V}_{ee}} + \hat{V}_{nn}$$

$$= \textcolor{red}{-\sum_i \frac{\hbar^2}{2m_e} \nabla^2_{\mathbf{r}_i}}$$ Kinetic energy of the electrons

$$\textcolor{blue}{-\sum_{i,j} \frac{Z_i e^2}{4\pi\varepsilon_0 |\mathbf{R}_i - \mathbf{r}_j|}}$$ Coulomb nuclei-electrons

$$\textcolor{orange}{+\sum_{i,j>i} \frac{e^2}{4\pi\varepsilon_0 |\mathbf{r}_i - \mathbf{r}_j|}}$$ Coulomb electrons-electrons

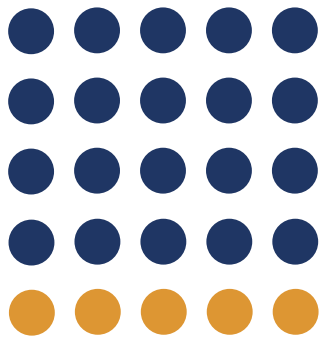$$+\sum_{i,j>i} \frac{Z_i Z_j e^2}{4\pi\varepsilon_0 |\mathbf{R}_i - \mathbf{R}_j|}$$ Coulomb nuclei-nuclei

Resulting wave function $\Psi = \Psi(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \ldots, \mathbf{r}_n)$ Impractical: 3N-dimensional!

Example 79

$$\Psi = \Psi(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_n)$$

Methane: 10 electrons
Grid with 5 points along each dimension

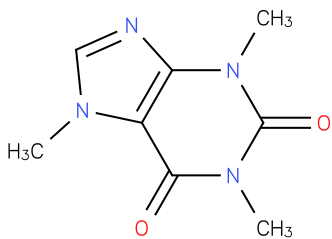Wave function: $5^{3 \times 10}$ grid points   6 ZB   7 times ALL data

Replace 3N-dimensional wave function Ψ by a 3-dimensional electron density ρ

$$\Psi = \Psi(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \ldots, \mathbf{r}_n) \quad \longrightarrow \quad \rho = \rho(\mathbf{r})$$

E.B. Wilson: Same information!

$\mathbf{R}_i$ Nuclear coordinates from local maxima

Z_i Nuclear charges: Kato's Theorem

$\longrightarrow \quad \hat{H} = \hat{H}(Z_i, \mathbf{R}_i, N_e, \sigma)$ N_e, σ Number electrons: $N_e = \displaystyle\int \mathrm{d}\mathbf{r}\, \rho(\mathbf{r})$

$\longrightarrow \quad \Psi$

$$\rho \longrightarrow Z_i, \mathbf{R}_i, N_e, \sigma \longrightarrow \hat{H} \longrightarrow \Psi$$

P.O. Löwdin, *Int. J. Quant. Chem.* **1986**. T. Kato, *Commun. Pure Appl. Math.*, **1957**.