Computational Molecular Biophysics

Computational Biophysics, GRS Jülich

SS 2013
### Computational Considerations

#### Number of terms for different energy contributions

<table>
<thead>
<tr>
<th>(n)</th>
<th>(N_{\text{atoms}})</th>
<th>(E_{\text{str}}) ((%)</th>
<th>(E_{\text{bend}}) ((%)</th>
<th>(E_{\text{tors}}) ((%)</th>
<th>(E_{\text{vdw}} / E_{\text{el}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>32</td>
<td>31 (5%)</td>
<td>30 (10%)</td>
<td>81 (14%)</td>
<td>405 (70%)</td>
</tr>
<tr>
<td>20</td>
<td>62</td>
<td>61 (3%)</td>
<td>60 (6%)</td>
<td>171 (8%)</td>
<td>1710 (83%)</td>
</tr>
<tr>
<td>50</td>
<td>152</td>
<td>151 (1%)</td>
<td>300 (3%)</td>
<td>441 (4%)</td>
<td>11025 (93%)</td>
</tr>
<tr>
<td>100</td>
<td>302</td>
<td>301 (1%)</td>
<td>600 (1%)</td>
<td>891 (2%)</td>
<td>44550 (96%)</td>
</tr>
<tr>
<td>(N)</td>
<td>((N-1))</td>
<td>2((N-2))</td>
<td>3((N-5))</td>
<td>(\frac{1}{2}N(N-1)-3N+5)</td>
<td></td>
</tr>
</tbody>
</table>

\(\sim N\) \(\sim N\) \(\sim N\) \(\sim N^2\)

#### Computational cost for non-bonded interactions

- Number of atoms: 302 \(\sim 96\%\)
- 1000 \(\sim 98.8\%\)
- 10000 \(\sim 99.88\%\)
Non-Bonded Interactions

Approaches for reducing computational costs

• Spherical cutoffs
  • Suitable for van der Waals interactions (~1/r^6)

• Particle Mesh Ewald
  • Preferred method in biological simulations

• Fast multipole schemes

• Continuum solvation methods
Non-Bonded Interactions

Computational Bottleneck

- Pairwise calculation $\sim N^2$ ($N =$ number of atoms)

Cutoff-Schemes

- Neglect long range interactions beyond some distance
- Poor approximation for highly charged systems

Figure 10.1. CPU time per step (energy plus force evaluation) for water clusters of various sizes modeled in CHARMM when cutoffs are used (at 10, 12, and 15 Å) versus no cutoffs (see left vertical scale) by direct calculation, and corresponding times required for 1 ns trajectories assuming $10^6$ steps of 1 fs (see right vertical scale). The number of variables is nine times the number of water molecules. Timings were obtained on a single Xeon/3GHz processor of a Dell Linux machine.
Non-Bonded Interactions

Cutoff Schemes

Applied to either energy or force function of the non-bonded potential.

- **Truncation**
  - Abruptly defines values to be zero at the cutoff $b$.
  - Do no alter values for distances $r < b$.

- **Switching**
  - Begin to change values at a nonzero value $a < b$.
  - Leave values unchanged for $r < a$.

- **Shift functions**
  - Alter functions more gradually for all $r < b$.

- **Atom-based cutoffs**
- **Group-based cutoffs:**
  - Distances between group centers.
Non-Bonded Interactions

Cutoff Schemes

Guidelines

• Short-range energies and forces should be altered as minimally as possible (while satisfying other criteria).
• Energies should be altered gradually rather than abruptly to avoid the introduction of artificial minima (where the potential energy and gradient values are suddenly zero).
• Avoid introducing large forces around the cutoff region.
• Should alter the function in a way to approximately conserve energy.
Cutoff Schemes

Formula appear on the following 2 slides

Non-Bonded Interactions

Altered potential

Errors in the forces
Non-Bonded Interactions

Cutoff Schemes

• Distance-dependent shift or switch function:

\[ E_{NB}(X) = \sum_{i,j} \omega_{ij} S(r_{ij}) \left[-\frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} + \frac{q_i q_j}{\epsilon r_{ij}}\right] \quad 0 \leq \omega_{ij} \leq 1 \]

\[ \hat{F}_{ij}(r_{ij}) = \omega_{ij} S(r_{ij}) F_{ij}(r_{ij}) \]

• Truncation:

\[ S(r) = \begin{cases} 
1 & r < b \\
0 & r \geq b 
\end{cases} \]

Discontinuity
Fails to conserve energy

• Potential switch

\[ S(r) = \begin{cases} 
1 & r < a \\
1 + y(r)^2 [2y(r) - 3] & a \leq r \leq b \\
0 & r > b 
\end{cases} \]

\[ y(r) = \frac{(r^2 - a^2)}{(b^2 - a^2)} \]

• Potential shift

\[ S_1(r) = [1 - (r/b)^2]^2, \quad \text{for } r \leq b, \]

\[ \text{CHARMM: electrostatics} \]

\[ S_2(r) = [1 - r/b]^2, \quad \text{for } r \leq b, \]

\[ S_2: \text{atom-based pot. shift} \]
Cutoff Schemes

- Potential shift: CHARMM van der Waals shift

\[ E_{LJ}(X) = \begin{cases} 
\sum_{i,j} \omega_{ij} \left[ \left( \frac{-A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} + C_{ij} r_{ij}^6 + D_{ij} \right) \right] & r \leq b \\
0 & r > b 
\end{cases} \]
Periodic Boundary Conditions

PBC

- Unit cell
- Infinitely many images in space

Figure 10.4. Periodic domain in 2D, showing the unit cell (center) and its images (surrounding replicas), and in 3D (rectangular), as used for a solvated protein (BPTI) system [1089].
Minimum Image Convention

- Only coordinates of the unit cell need to be recorded and propagated.
- As an atom leaves the unit cell, an image enters to replace it conserving the total number of particles.
- Cell size: solvated biomolecule with at least 10 Å layer of water.

- Each atom acts only with the closest periodic image of the other $N-1$ particles.
- Spherical cutoff: distance at most $L/2$, where $L$ is the dimension of the side of the box.
Ewald sum and crystallography

- Reciprocal term
  - The reciprocal lattice is an orthogonal system related to the orthogonal system associated with the atoms in the unit cell (real-space lattice).
  - This lattice is used to express scattering vector in crystals.
  - Ewald summation: two terms arise for the computation of the electrostatic energy:
    - Atom pairs in the direct lattice: Direct calculation
    - Atom pairs corresponding to interactions with images of the unit cell atoms: Fourier transformation
  - Fast Fourier transforms (FFT)
    - Ewald sum: evaluate the electrostatic energy corresponding to the reciprocal lattice.
  - Electron density distribution
    - Electrostatic energy in terms of charge density rather than point charges to make the infinite sum convergent.
Box 10.1: Some Tricks from X-Ray Crystallography

A crystal is a 3D periodic arrangement of atoms (see [1047], for example, for an introduction to crystallography, and Chapter 1 for an introduction to the phase problem). The structural subunit of crystals is the **unit cell**. The X-ray diffraction pattern by the electrons of a molecular crystalline system is interpreted in terms of both the **direct lattice**, the orthogonal system associated with the unit-cell atoms, and the **reciprocal lattice**, an auxiliary orthogonal system used to express the scattering vector in crystals. Specifically, the scattering of each atom \( j \) with respect to an electron at the origin of the cell it occupies is:

\[
F_{s_j} = f_j \exp(2\pi i x_j \cdot s),
\] 

(10.17)

where \( i = \sqrt{-1} \), \( f_j \) is an experimentally-measured amplitude, \( x_j = \{x_{j1}, x_{j2}, x_{j3}\} \) denotes the Cartesian position vector of atom \( j \) in the direct lattice, and \( s \) is the scattering vector corresponding to a triplet of indices (**indices of reflection**), such as \( s = \{h, k, l\} \), in the reciprocal lattice. Thus for \( N \) atoms in the unit cell, we write the complex-valued **structure factor** \( F_s \) as:
\[ F_s = \sum_{j=1}^{N} f_j \exp(2\pi i x_j \cdot s) = A_s + iB_s, \]  
\[ A_s = \sum_{j=1}^{N} f_j \cos(2\pi i x_j \cdot s), \quad B_s = \sum_{j=1}^{N} f_j \sin(2\pi i x_j \cdot s). \]  

From these quantities, the phase angles associated with \( F_s \) are computed as \( \phi_s = \tan^{-1}(B_s/A_s) \). The intensity of the diffracted X-ray beam is proportional to the square of the structure factor’s amplitude and depends only on the interatomic vectors, not on the actual, origin-dependent atomic coordinates:

\[ |F_s|^2 = A_s^2 + B_s^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} f_i f_j \cos[2\pi (x_i - x_j) \cdot s]. \]  

The relationship between crystallography and the electrostatic potential arises by relating the diffracted amplitude to a continuous distribution of electron density, \( \rho(x) \), expressed as electrons per unit volume. The scattered amplitude for the whole crystal, \( F'(s) \), from a small volume element \( dv \), results from an effective point charge of \( \rho(x) \, dv \) electrons:

\[ F(s) = \int_V \rho(x) \exp(2\pi is \cdot x) \, dv, \]
where the integration is taken over the volume $V$ of the unit cell. This is the Fourier transform of the electron density in one unit cell of the crystal, sampled at points $s$ on the reciprocal lattice. The inverse transformation yields the electron density via integration over the entire volume of the reciprocal space in which $s$ is defined:

$$\rho(x) = \int_{V^*} F(s) \exp\left(-2\pi is \cdot x\right) dv^*. \quad (10.22)$$

Crystallographers compute a discrete analog of this electron density to obtain the atomic arrangement in the crystal in which each point of the reciprocal lattice (where $F(s)$ is defined) has an associated weight of $F_s/V$ where $F_s$ is the structure factor defined in eq. (10.18):

$$\rho(x) = \frac{1}{V} \sum_{h,k,l=-\infty}^{\infty} F_s \exp\left(-2\pi is \cdot x\right). \quad (10.23)$$

This computation is performed using Fast Fourier transforms. The amplitudes of these structure factors are known from the observed intensities of the X-ray reflections, but the phase angles cannot be measured. This is the heart of the problem of deducing structure from X-ray diffraction patterns.
Ewald Summation

Coulomb energy in periodic domains

\[
E_{\text{coul}} = \frac{1}{2} \sum_{i,j=1}^{N} \sum_{\text{images } |n|} \frac{q_i q_j}{|r_{ij} + n|} = \frac{1}{2} \sum_{j=1}^{N} q_j \Phi(x_j),
\]

(Without constants)

where

\[
\Phi(x_j) = \sum_{i=1}^{N} \sum_{|n|} \frac{q_i}{|r_{ij} + n|}.
\]

\((\sum_{n}^{'})\) indicates that for \(n = 0\) we omit the \(i = j\) interaction.

- The sum is conditionally convergent since the terms decay of \(1/|n|\) like a harmonic series.

---

2A series \(S = \sum_{n=1}^{\infty} a_n\) is conditionally convergent if the infinite sum \(\sum_{n} a_n\) converges but \(\sum_{n} |a_n|\) diverges. It is also known that the sum for a conditionally convergent series depends on the order of summation. For \(a_n = (-1)^{n+1}/n\), the ‘alternating harmonic series’ \(1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots\) converges, but the harmonic series \(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots\) diverges, though \(a_n \to 0\) as \(n \to \infty\).
**Idea:** Splitting into short-range and long-range contributions, which converts a conditionally convergent sum into two terms that are rapidly and absolutely converging.

**Short range**
- Embedding each point charge in a screening potential:
- Gaussian with opposite sign: outside this range (width of the Gaussian) the net charge is zero => interaction between screened point charges is short-ranged and can be evaluated directly in real space.

**Long range**
- To recover the original point charge interaction, the effect of the screening potentials must be subtracted again.
- This compensating term is an interaction between Gaussian charge distributions evaluated efficiently in reciprocal space.
Idea: Splitting into short-range and long-range contributions

\[ \phi(r) \approx \frac{1}{r} = \frac{f(r)}{r} + \frac{1 - f(r)}{r} \]

Calculated in reciprocal (Fourier) space. Calculated in real space.
Ewald summation

- Convert the sum into an expression involving a sum of two absolutely and rapidly converging series in direct and reciprocal space.
- Conversion is accomplished by representing each point charge as a Gaussian charge density, producing an exponentially decaying function.
- This Gaussian transformation must be counteracted by an analogous subtraction to leave the net result of an effective point charge (delta function charge).
  - This cancelling distribution is summed in the reciprocal space and transformed back to real space because it is a smoothly-varying periodic function which can be represented by a rapidly convergent Fourier series.

\[
\Phi(r) \equiv \frac{1}{r} = \Phi_{\text{real}}(r) + \Phi_{\text{recip}}(r),
\]

where

\[
\Phi_{\text{real}}(r) = \frac{1}{r} - \frac{\text{erf}(\beta r)}{r} = \frac{\text{erfc}(\beta r)}{r},
\quad \Phi_{\text{recip}}(r) = \frac{\text{erf}(\beta r)}{r}.
\]

Short-ranged with singularity at origin                   long-ranged smooth
• **Error function**

\[
erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) \, dt,
\]

\[
erf(x) = \frac{2}{\sqrt{\pi}} \left( x - \frac{x^3}{3} + \frac{1}{2!} \frac{x^5}{5} - \frac{1}{3!} \frac{x^7}{7} + \cdots \right)
\]

\[
erfc(x) = 1 - erf(x)
\]

Complementary error function

• **Screening Gaussians** centered at each point charge (parameterized by \( \beta \))

\[
\rho_{G_j}(x) = -q_j \left( \frac{\beta}{\sqrt{\pi}} \right)^3 \exp[-\beta^2 |x|^2]
\]

• Instead of the cumulative point charge density described by a sum of delta functions

\[
\rho(x) = \sum_{j=1}^{N} q_j \delta(x - x_j)
\]

• a sum of localized densities is used to define the total Gaussian screening charge density

\[
\rho_G(x) \equiv \sum_{j=1}^{N} \rho_{G_j}(x) = -\sum_{j=1}^{N} q_j \left( \frac{\beta}{\sqrt{\pi}} \right)^3 \exp[-\beta^2 |x - x_j|^2]
\]
Ewald Summation

• Real space sum

• Given in terms of the complementary error function using Poisson’s equation for the electrostatic potential, due to the Gaussian charge cloud followed by integration.

• Solution of the Poisson equation

\[ \nabla^2 \Phi_{G_j}(r) = -4\pi \rho_{G_j}(r) \]

leads to

\[ \Phi_{G_j}(r) = \frac{q_j}{r} \text{erf}(\beta r) \]

• Thus, the sum for the Coulomb energy is converted into the pair of terms

\[ E_{\text{coul}} = \frac{1}{2} \sum_{j=1}^{N} \sum_{|n|} q_j \frac{\text{erfc}(\beta|x_{ij} + n|)}{|r_{ij} + n|} + \frac{1}{2} \sum_{j=1}^{N} \sum_{|n|} q_j \frac{\text{erf}(\beta|x_{ij} + n|)}{|r_{ij} + n|} \]

Though reliable, Ewald’s algorithm as corrected in [293] was still \( \mathcal{O}(N^2) \) in computational complexity. This is because the long-range reciprocal-space Fourier sum requires \( \mathcal{O}(N^2) \) to be a sufficiently accurate approximation for large \( \beta \) (see below).
Ewald Summation

• Ewald sum complexity

• Optimization of $\beta$ (Gaussian width)
  • As $\beta$ increases, the real space sum converges more rapidly, but the reciprocal sum more slowly.
  • Both sums are truncated in practice
  • A sufficiently large $\beta$ yields an accurate direct space sum with an appropriate truncation resulting in $\sim N$ rather than $\sim N^2$.
  • The reciprocal space sum still requires $\sim N^2$ work.
  • In total $\sim N^{3/2}$.

• Mesh interpolation
  • Trigonometric function values in the Fourier series used to represent the reciprocal space term can be evaluated through a smooth interpolation of the potential over a regular grid.
  • Particle-Mesh Ewald (PME): $\sim N(\log N)$.

• Variations
  • Particle-Particle Particle Mesh Ewald (P$^3$M):
    • Long range forces in large systems are split into a short-range, rapidly varying part (direct particle-particle sum) and a smooth, slowly varying remainder (particle-mesh).
Ewald Summation

• Resulting Ewald Summation

\[ E_{\text{coul}} = E_{\text{real}} + E_{\text{recip}} + E_{\text{cor, self}} + E_{\text{cor, ex}} + E_{\text{cor, \epsilon}} . \]

• \( E_{\text{real}} \): real-space sum of electrostatic energy due to the point charges screened by oppositely charged Gaussians.

• \( E_{\text{recip}} \): associated cancelling term summed in reciprocal space using smooth interpolation of Fourier-series values.
Ewald Summation

• Resulting Ewald Summation

\[ E_{\text{coul}} = E_{\text{real}} + E_{\text{recip}} + E_{\text{cor,self}} + E_{\text{cor,ex}} + E_{\text{cor,\varepsilon}}. \]

• \( E_{\text{cor, self}} \): Subtracts the self-interaction term (each point charge and its corresponding charge cloud).

• \( E_{\text{cor, ex}} \): Subtracts the Coulomb contribution from the nonbonded energies included in the Coulomb energy since they are separately accounted for in bonded, bond-angle, and possibly dihedral-angle terms.

• \( E_{\text{cor, \varepsilon}} \): Accounts for non-infinite dielectric medium
Ewald Summation

- **Practical implementation**

- Gaussian width
  - Real space term should acquire the desired accuracy tolerance
  - $\approx 10 / L$ ($L$ ranges roughly from 60 to 100 Å).

- Grid size and accuracy
  - Coarse interpolation grid: 1-2 Å (moderate accuracy, $10^{-4}$ relative force error)
  - Finer grid: -0.5 Å (very high accuracy, $10^{-10}$ relative force error)
Fast Multipole Moment (FMM)

- **Splitting into near- and far field**
  - Near field: exact calculation
  - Far field: dividing space into boxes
    - Interaction between all molecules in one box with all molecules in another is approximated as interactions between multipoles located at the box centers.
    - The further away from each other two boxes are, the larger the boxes can be for a given accuracy, thereby reducing the formal $N^2$ scaling to linear scaling.
  - Cross-over point when FMM method become faster than PME: $\sim 10^5$ atoms.
  - Works best when particles are relatively uniformly distributed.
  - Maximum error is significantly larger than for Ewald methods.
  - No requirement of periodicity, thus applicable to large non-periodic systems.
Fast Multipole Moment (FMM)

„Naive“ Multipole Method

\[ U_{AC1} \]
\[ U_{AC2} \]
\[ U_{AC3} \]
\[ U_{AC4} \]

\[ 1/4 U_{AB} \]

\[ O(M^2) \]

FMM Method

\[ U_{AC} \]

\[ O(M^{1}) \]
Reducing computational costs

Long range interaction

- Ewald summation method  Scaling: $O(N^{3/2})$
- Particle mesh Ewald techniques  $O(N \log N)$
- Fast multipole methods  $O(N)$

$N = \text{number of atoms}$
Parameterization

- Non-bonded and torsion parameters affect performance most, stretching and bending are less important.
- Aim to reproduce
  - exp. structures
  - exp. thermodynamics
  - QM derived properties

---

Roadmap Charmm27 Optimization*

- Initial Geometries Model Compounds?
- Partial Atomic Charges
- VDW Parameters
- Bonds, Angles, Torsions, Impropers
- Condensed Phase MD Simulations
- Parameterization Complete

*based on MacKerell, JCC v21, 86,105 (2000)
Parameterization

- Parameter set and functional forms are self-consistent, thus parameters are not transferable from one force field to another.

<table>
<thead>
<tr>
<th>Force field</th>
<th>$R_0$ (Å)</th>
<th>$k$ (mdyn/Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM2</td>
<td>1.523</td>
<td>1.402</td>
</tr>
<tr>
<td>MM3</td>
<td>1.525</td>
<td>1.413</td>
</tr>
<tr>
<td>MMFF</td>
<td>1.508</td>
<td>1.418</td>
</tr>
<tr>
<td>AMBER</td>
<td>1.526</td>
<td>1.410</td>
</tr>
<tr>
<td>OPLS</td>
<td>1.529</td>
<td>1.410</td>
</tr>
</tbody>
</table>

- Force constants may differ by almost a factor of two, but this is of course related to the stiffness of the bond.
- The associated energy contributions are small regardless of the force constant, i.e. the minimum energy geometry is insensitive to the exact value of the force constant.
• **Class I / Harmonic / Diagonal / Classical FF**
  • Designed to treat large systems such as proteins or DNA
  • Simple functional form (only harmonic functions for bonded interactions, no cross terms, Lennard-Jones potential for van der Waals interactions).
  • CHARMM, AMBER, OPLS, GROMOS, …

• **Class II / Second generation FF**
  • Reproducing small to medium sized molecules with high accuracy.
  • Cross terms, cubic or quartic expansions for bond and angle potentials.
  • Reproduce also vibrational frequencies
  • MMFF94 (Merck), MM2/MM3/MM4 (Allinger, organic molecules), …

• **Class III**
  • Refinements allowing parameters to depend on neighboring atom types
  • e.g. for modeling hyperconjugation
  • including electronic polarization effects.
• Class I potential energy function

\[
E_{\text{Total}} = \sum_{\text{bonds}} k_b (b - b_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \frac{V}{2} [1 + \cos(n\phi - \delta)] + \sum_{\text{impropers}} k_\omega (\omega - \omega_0)^2 + \sum_{\text{Urey-Bradley}} k_u (r_{1,3} - r_{1,3,0})^2
\]

Non-bonded Interaction Terms

\[
+ \sum_{\text{electrostatics}} \left( \frac{q_i q_j}{\varepsilon r_{ij}} \right) + \sum_{\text{VDW}} \varepsilon_{ij} \left[ \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^{6} \right]
\]
• Class II potential energy function

\[
\sum_{\text{bonds}} \left[ K_{b,2} (b - b_o)^2 + K_{b,3} (b - b_o)^3 + K_{b,4} (b - b_o)^4 \right] \\
+ \sum_{\text{angles}} \left[ K_{\theta,2} (\theta - \theta_o)^2 + K_{\theta,3} (\theta - \theta_o)^3 + K_{\theta,4} (\theta - \theta_o)^4 \right] \\
+ \sum_{\text{dihedrals}} \left[ K_{\phi,1} (1 - \cos \phi) + K_{\phi,2} (1 - \cos 2\phi) + K_{\phi,3} (1 - \cos 3\phi) \right] \\
+ \sum_{\text{improper}} K_x \chi^2 \\
+ \sum_{\text{bonds bonds'}} K_{bb'} (b - b_o) (b' - b_o') + \sum_{\text{angles angles'}} K_{\theta\theta'} (\theta - \theta_o) (\theta' - \theta_o') \\
+ \sum_{\text{bonds angles}} K_{b\theta} (b - b_o) (\theta - \theta_o) \\
+ \sum_{\text{bonds dihedrals}} (b - b_o) \left[ K_{\phi,b1} \cos \phi + K_{\phi,b2} \cos 2\phi + K_{\phi,b3} \cos 3\phi \right] \\
+ \sum_{\text{bonds' dihedrals}} (b' - b_o') \left[ K_{\phi,b1} \cos \phi + K_{\phi,b2} \cos 2\phi + K_{\phi,b3} \cos 3\phi \right] \\
+ \sum_{\text{angles dihedrals}} (\theta - \theta_o) \left[ K_{\phi,\theta1} \cos \phi + K_{\phi,\theta2} \cos 2\phi + K_{\phi,\theta3} \cos 3\phi \right] \\
+ \sum_{\text{angles angles' dihedrals}} (\theta - \theta_o) (\theta' - \theta_o') \cos \phi
\]
Polarizable Force Fields (Class III)

- **Based on electronic structure theory**
  - X-Pol: Explicit polarization theory, fragment based electronic structure method (J. Gao)

- **Based on induced dipoles**
  - CFF/ind, ENZYMIX
  - DFF90 (P. Th. van Duijnen)
  - PIPF (J. Gao)

- **Based on point charges**
  - PFF: Polarizable force field (R. A. Friesner)
  - CHARMM PFF (S. Patel, C. L. Brooks III)
  - CHARMM PFF: classical Drude oscillators (A. MacKerell, B. Roux)
  - AMBER PFF (J. Caldwell)

- **Based on distributed multipoles**
  - AMOEBA (P. Ren, J. W. Ponder)

- **Based on density**
  - Gaussian electrostatic model (GEM) (T. Darden)
  - Kim-Gordon approach (J. Hutter)
Simplified Force Fields

• **United atom force fields**
  - Do not consider hydrogen atoms explicitly
  - Option for AMBER, CHARMM, GROMOS, DREIDING
  - Reduce the effort by a factor ~2-3.

  ![All atom vs. United atom](image)

• **Coarse-grained FF**
  - Atom groups as single particles (pseudo atoms or beads)
  - E.g. amino acid: 1-2 beads on the backbone, 0-3 beads for side chain.
  - MARTINI

“CH$_2$ atom”
Coarse-grained Force Fields

- MARTINI force field: Amino acid representation

Coarse-grained representation of all amino acids. Different colors represent different particle types.

Luca Monticelli; Senthil K. Kandasamy; Xavier Periole; Ronald G. Larson; D. Peter Tieleman; Siewert-Jan Marrink; *J. Chem. Theory Comput.* 2008, 4, 819-834.
Coarse-grained Force Fields

• MARTINI force field

• **Empirical Valence Bond (EVB)**
  - A. Warshel
  - Chemical reactions, e.g. enzymatic catalysis, proton transfer, …
  - Multi-state EVB: G. A. Voth

---

**Figure 2:** Panel (a) Grotthuss shuttling through a small chain of four water molecules; Panel (b) The MS-EVB states required to represent the process in panel (a); Panel (c) the Eigen cation $\text{H}_9\text{O}_4^+$ (left) and Zundel cation $\text{H}_5\text{O}_2^+$
Knowledge-Based Force Fields

- Knowledge-based FF as alternative to physics-based FF
  - Energy functions derived from databases of known protein structures
  - Probabilities are calculated that residues appear in specific configurations and that residues appear together in a particular relative geometry.
  - Probabilities are converted into an effective potential energy using the Boltzmann equation.
    \[
    \Delta G = -k_B T \ln \frac{p(c_1, c_2)}{p(c_1)p(c_2)},
    \]
  - \(c_1\): amino acid type, \(c_2\): structure descriptor of the same or neighboring residue, \(p\): relative occurrence in the data base.
  - Pro: Any behavior seen in protein structures can be modeled.
  - Con: Phenomenological functions cannot predict behavior not present in the training data set.
  - Examples: Go models (P. Wolynes), Rosetta (D. Baker)
<table>
<thead>
<tr>
<th>Force field</th>
<th>Types</th>
<th>$E_{str}$</th>
<th>$E_{bend}$</th>
<th>$E_{oop}$</th>
<th>$E_{vdw}$</th>
<th>$E_{el}$</th>
<th>$E_{cross}$</th>
<th>Molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMBER</td>
<td>41</td>
<td>P2</td>
<td>P2</td>
<td>imp.</td>
<td>12–6</td>
<td>charge</td>
<td>none</td>
<td>proteins, nucleic acids, carbohydrates</td>
</tr>
<tr>
<td>CFF91/93/95</td>
<td>48</td>
<td>P4</td>
<td>P4</td>
<td>P2</td>
<td>9–6</td>
<td>charge</td>
<td>ss,bb,st, sb, bt, btb</td>
<td>general</td>
</tr>
<tr>
<td>CHARMM</td>
<td>29</td>
<td>P2</td>
<td>P2</td>
<td>imp.</td>
<td>12–6</td>
<td>charge</td>
<td>none</td>
<td>proteins</td>
</tr>
<tr>
<td>COSMIC</td>
<td>25</td>
<td>P2</td>
<td>P2</td>
<td>Morse</td>
<td>12–6</td>
<td>charge</td>
<td>none</td>
<td>general</td>
</tr>
<tr>
<td>CVFF</td>
<td>53</td>
<td>P2 or Morse</td>
<td>P2</td>
<td>P2</td>
<td>12–6</td>
<td>charge</td>
<td>ss,bb,sb, btb</td>
<td>general</td>
</tr>
<tr>
<td>DREIDING</td>
<td>37</td>
<td>P2 or Morse</td>
<td>P2(cos)</td>
<td>P2(cos)</td>
<td>12–6 or Exp–6</td>
<td>charge</td>
<td>none</td>
<td>general</td>
</tr>
<tr>
<td>EAS</td>
<td>2</td>
<td>P2</td>
<td>P3</td>
<td>none</td>
<td>Exp–6</td>
<td>none</td>
<td>none</td>
<td>alkanes</td>
</tr>
<tr>
<td>ECEPP</td>
<td>fixed</td>
<td>fixed</td>
<td>fixed</td>
<td>fixed</td>
<td>12–6 and 12–10</td>
<td>charge</td>
<td>none</td>
<td>proteins</td>
</tr>
<tr>
<td>EFF</td>
<td>2</td>
<td>P4</td>
<td>P3</td>
<td>none</td>
<td>Exp–6</td>
<td>none</td>
<td>ss,bb,sb, st, btb</td>
<td>alkanes</td>
</tr>
<tr>
<td>ENCAD</td>
<td>35</td>
<td>P2</td>
<td>P2</td>
<td>imp.</td>
<td>12–6</td>
<td>charge</td>
<td>none</td>
<td>proteins, nucleic acids</td>
</tr>
<tr>
<td>ESFF</td>
<td>97</td>
<td>Morse</td>
<td>P2(cos)</td>
<td>P2</td>
<td>9–6</td>
<td>charge</td>
<td>none</td>
<td>all elements, proteins, nucleic acids, carbohydrates</td>
</tr>
<tr>
<td>GROMOS</td>
<td>P2</td>
<td>P2</td>
<td>P2(imp.)</td>
<td>12–6</td>
<td>charge</td>
<td>none</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>Force Fields</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MM2</strong></td>
<td>71</td>
<td>P3</td>
<td>P2+P6</td>
<td>P2</td>
<td>Exp–6</td>
<td>dipole</td>
<td>dipole or charge</td>
<td>sb</td>
</tr>
<tr>
<td><strong>MM3</strong></td>
<td>153</td>
<td>P4</td>
<td>P6</td>
<td>P2</td>
<td>Exp–6</td>
<td>dipole</td>
<td>charge</td>
<td>sb,bb,bb</td>
</tr>
<tr>
<td><strong>MM4</strong></td>
<td>56</td>
<td>P6</td>
<td>P6</td>
<td>imp.</td>
<td>Exp–6</td>
<td>none</td>
<td>quad, polar</td>
<td>ss,bb,bb,tt,tt</td>
</tr>
<tr>
<td><strong>MMFF</strong></td>
<td>99</td>
<td>P4</td>
<td>P3</td>
<td>P2</td>
<td>14–7</td>
<td>charge</td>
<td>none</td>
<td>bb</td>
</tr>
<tr>
<td><strong>MOMEC</strong></td>
<td>P2</td>
<td>P2</td>
<td>P2</td>
<td>Exp–6</td>
<td>none</td>
<td>quad, polar</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td><strong>NEMO</strong></td>
<td>fixed</td>
<td>fixed</td>
<td>none</td>
<td>Exp–6</td>
<td>quad, polar</td>
<td>none</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td><strong>OPLS</strong></td>
<td>41</td>
<td>P2</td>
<td>P2</td>
<td>imp.</td>
<td>12–6</td>
<td>charge</td>
<td>ss,bb,bb,bb,bb,bb</td>
<td>Hydrocarbons</td>
</tr>
<tr>
<td><strong>PFF</strong></td>
<td>P2</td>
<td>P2</td>
<td>P2</td>
<td>imp.</td>
<td>12–6</td>
<td>polar</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td><strong>PROSA</strong></td>
<td>41</td>
<td>P2</td>
<td>P2</td>
<td>imp.</td>
<td>12–6</td>
<td>polar</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td><strong>QMFF</strong></td>
<td>32</td>
<td>P4</td>
<td>P4</td>
<td>P2</td>
<td>9–6</td>
<td>charge</td>
<td>ss,bb,bb,bb,bb,bb,bb</td>
<td>general</td>
</tr>
<tr>
<td><strong>SDFF</strong></td>
<td>P4</td>
<td>P4</td>
<td>P4</td>
<td>9–6</td>
<td>charge</td>
<td>ss,tt</td>
<td>none</td>
<td>C, N, O compounds</td>
</tr>
<tr>
<td><strong>TraPPE</strong></td>
<td>fixed</td>
<td>P2</td>
<td>fixed</td>
<td>12–6</td>
<td>charge</td>
<td>none</td>
<td>none</td>
<td>C, N, O compounds</td>
</tr>
<tr>
<td><strong>TRIPOS</strong></td>
<td>31</td>
<td>P2</td>
<td>P2</td>
<td>cos($n\theta$)</td>
<td>12–6</td>
<td>charge</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td><strong>UFF</strong></td>
<td>126</td>
<td>P2 or Morse</td>
<td>cos($n\theta$)</td>
<td>imp.</td>
<td>12–6 and 12–10</td>
<td>charge</td>
<td>none</td>
<td>proteins</td>
</tr>
</tbody>
</table>

Notation: Pn: Polynomial of order n; Pn(cos): polynomial of order n in cosine to the angle; cos($n\theta$): Fourier term(s) in cosine to the angle; Exp–6: exponential + $R^{-6}$; $n – m$: $R^{-n} + R^{-m}$ Lennard-Jones type potential; quad: electric moments up to quadrupoles; polar: polarizable; fixed: not a variable; imp.: improper torsional angle; ss: stretch–stretch; bb: bend–bend; sb: stretch–bend; st: stretch–torsional; bt: bend–torsional; tt: torsional–torsional; btb: bend–torsional–bend.
### Force Fields

**Class I:**  
*AMBER (Assisted Model Building and Energy Refinement)*

<table>
<thead>
<tr>
<th>Force Field</th>
<th>Component Types</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ff94, ff96, ff98</td>
<td>proteins</td>
<td>ff94 biased towards helices, ff96 favoring extended conformations too much</td>
</tr>
<tr>
<td>f99</td>
<td>proteins, RNA, DNA</td>
<td>good nonbonded interactions, overstabilization of α helices</td>
</tr>
<tr>
<td>ff99SB</td>
<td>proteins, RNA, DNA</td>
<td>fixed overstabilization of helices</td>
</tr>
<tr>
<td>ff03</td>
<td>proteins, RNA, DNA</td>
<td>partial charges of f99 reparameterized, better balance between α helix and β sheet</td>
</tr>
<tr>
<td>ff99SB-ILDN</td>
<td></td>
<td>Improved side-chain torsional potentials for Ile, Leu, Asp, and Asn.</td>
</tr>
<tr>
<td>ff99SB-ILDN-NMR</td>
<td></td>
<td>NMR-based torsional modifications</td>
</tr>
<tr>
<td>ff03*, f99SB*</td>
<td>proteins, RNA, DNA</td>
<td>modified backbone potential for a better balance between helix and coil</td>
</tr>
<tr>
<td>GAFF</td>
<td>any molecule</td>
<td>general Amber force field, used for ligands</td>
</tr>
<tr>
<td>GLYCAM</td>
<td>sugars</td>
<td>carb</td>
</tr>
</tbody>
</table>
### Force Fields

- **Class I: ** **CHARMM** *(Chemistry at HARvard Molecular Mechanics)*

<table>
<thead>
<tr>
<th>Force Field</th>
<th>Functions</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHARMM19</td>
<td>proteins, lipids</td>
<td>united atom</td>
</tr>
<tr>
<td>CHARMM22</td>
<td>proteins</td>
<td>all-atom, parameterized for TIP3P (adapted for CHARMM), though often used with implicit solvent, tends to favor π helices.</td>
</tr>
<tr>
<td>CHARMM22-CMAP</td>
<td>proteins</td>
<td>CMAP backbone dihedral angle corrections</td>
</tr>
<tr>
<td>CHARMM27</td>
<td>RNA, DNA, lipids, sugars</td>
<td>based on CHARMM22/CMAP</td>
</tr>
</tbody>
</table>
**Force Fields**

- **Class I:**

  **GROMOS** (GROningen MOlecular Simulation package)
  **OPLS** (Optimized Potentials for Liquid Simulations)

<table>
<thead>
<tr>
<th>Force Field</th>
<th>Applications</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROMOS87, GROMOS96, GROMOS05, GROMOS11</td>
<td>alkanes, proteins, sugars</td>
<td>united atom Parameter sets: 54, 53, 45, 43, traditionally used with SPC water</td>
</tr>
<tr>
<td>OPLS-UA</td>
<td>proteins</td>
<td>united atom, TIP3P/TIP4P water models, optimized to fit exp. properties of liquids.</td>
</tr>
<tr>
<td>OPLS-AA</td>
<td>proteins</td>
<td>all atom, TIP3P/TIP4P water models, optimized to fit exp. properties of liquids.</td>
</tr>
<tr>
<td>OPLS-2001, OPLS-2005</td>
<td>proteins, sugars</td>
<td>based on OPLS-AA</td>
</tr>
</tbody>
</table>
### Class III: Polarizable force fields

<table>
<thead>
<tr>
<th>Force Field</th>
<th>Application</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHARMM-FQ</td>
<td>proteins</td>
<td>fluctuating charge model or charge equilibration model</td>
</tr>
<tr>
<td>CHARMM-CHEQ</td>
<td>proteins</td>
<td>Drude model</td>
</tr>
<tr>
<td>CHARMM-Drude</td>
<td>proteins</td>
<td>Drude model</td>
</tr>
</tbody>
</table>
ff99SB*-ILDN and CHARMM22* perform consistently well in reproducing the experimental data in the test set.

• **Deficiencies / Need for improvement**

  • Determination of partial charges
  • Improvement of electrostatic potentials (e.g. use of distributed multipole analysis)
  • Methods for solvent representation.
  • Resolve conflicting results by different models and potentials.
  • Van der Waals parameter are strongly environment dependent

  • Inclusion of polarization.
App1: Derivatives of MM Energy Functions

- Derivatives for forces

\[ r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \]

\[ \frac{\partial \psi}{\partial x_i} = \frac{\partial \psi}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial x_i} \]

\[ \frac{\partial r_{ij}}{\partial x_i} = \frac{(x_i - x_j)}{r_{ij}} \]

Thus, for the Lennard-Jones potential:

\[ \frac{\partial \psi}{\partial r_{ij}} = \frac{24 \varepsilon}{r_{ij}^2} \left[ -2 \left( \frac{\sigma}{r_{ij}} \right)^{12} + \left( \frac{\sigma}{r_{ij}} \right)^6 \right] \]

The force in the x direction acting on atom i due to its interaction with atom j is given by:

\[ f_{x_i} = (x_i - x_j) \frac{24 \varepsilon}{r_{ij}^2} \left[ 2 \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^6 \right] \]