

Molecular Mechanics and Force Fields

What Kind of Modeling?

- “Molecular Mechanics” refers to the **description of a molecular system as a mechanical body**
- Molecule = collection of particles (masses) centered at the nuclei, connected by “springs” (bonds)
- MM relies on building up a potential energy U , function of atomic positions and tuned by empirical data (aka “force field”, somehow improperly)
- U can be used in optimizations/simulations
- Such a description is usually appropriate to model molecular structures and processes, **but not bond-breaking events.**

Recall: Potentials

- Potential energy U is a **scalar field**:
it's described by a single value at each point in space
- Force F instead is a vector quantity, related to U :

$$F = -\nabla U$$

As in (non-linear) molecules with n atoms there are $p=3n-6$ degrees of freedom, we expect that U would depend on $3n-6$ independent variables

Often, especially in quantum mechanics modeling, U is also referred to as "PES" (Potential Energy Surface)

MM: Underlying Principles

- Molecular Mechanics relies on the validity of three principles:
 1. **Thermodynamic hypothesis**
 - Many macromolecules reach their native, folded state due to thermodynamic factors (energy minimization)
 2. **Additivity**
 - Molecular energy can be expressed as sum of terms related to simple physical forces
 3. **Transferability of potentials**
 - Developing potentials according to experimental behavior of specific chemical groups, makes such potentials applicable to macromolecules composed of such groups

Force Fields: Components

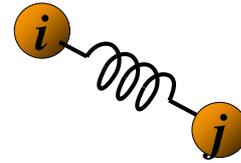
- U can be expressed as a sum of potentials derived from simple physical forces.
- Components of U can be divided into:
 $U = U_{local} + U_{nonlocal}$
 - local (bonded, i.e. related to bonds) terms
 - Mechanical-like strains related to displacements from ideal values of bond lengths/angles
 - Internal torsion flexibility
 - non-local (i.e. not related to bonds) terms
 - Van der Waals
 - Electrostatic
- Terms already present in local interactions are not counted in non-local ones.

FFs: Accuracy vs Speed

- Force fields needs not to be perfect to be useful! (T. Schlick)

Local Terms: Bond Length

$$U_{local} = U_{blen} + U_{bang} + U_{tors}$$



- The bond stretching component U_{blen} is usually taken as harmonic, so that bonds are treated as ideal springs:

$$U_{blen} = \sum_{i,j \in S_B} S_{ij} (r_{ij} - \underline{r}_{ij})^2$$

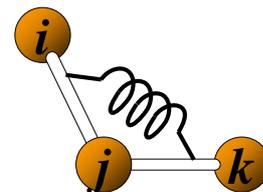
- Note that, for each term,

$$F = -dU/dr = -1/2 S (r_{ij} - \underline{r}_{ij}) \quad (\text{Hook's law})$$

- Harmonic potential can be used only for *small* deviations from reference values \underline{r}_{ij} in the set of bonds S_B
- More accurate formulations for larger deviations have been proposed (e.g. Morse Term, computationally heavier)
- Some force fields introduce explicit terms also for H-bonds

Local Terms: Bond Angle

$$U_{local} = U_{blen} + U_{bang} + U_{tors}$$



- The angle bending component U_{bang} is usually taken as harmonic in deviations from $\underline{\theta}_i$ or their corresponding cosines:

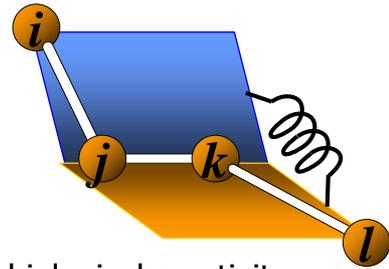
$$U_{bang-harm} = \sum_{i,j,k \in S_{BA}} K_{(H)ijk} (\theta_{ijk} - \underline{\theta}_{ijk})^2$$

$$U_{bang-trig} = \sum_{i,j,k \in S_{BA}} K_{(T)ijk} (\cos \theta_{ijk} - \cos \underline{\theta}_{ijk})^2$$

- The trigonometric form (bounded) is deemed more convenient in numeric treatment; By expanding $U_{bang-trig}$ by a Taylor series, we get $K_{(T)} \approx K_{(H)} \sin^2(\underline{\theta})$
- Determination of reference angles is often a challenging task
- For more accurate formulations, as in U_{blen} functions with odd powers should be avoided, as U can assume negative values during computations (nonsensical!)

Local Terms: Torsion

$$U_{local} = U_{blen} + U_{bang} + U_{tors}$$



- Taking into account rotational flexibility of atomic sequences is particularly important for biological reactivity
- Each component in U_{tors} is necessarily periodic, repeating for sure every 2π radians. So it can be expressed as a n-truncated series of Fourier terms like

$$U_{tors} = \sum_{ijkl \in S_{DA}} \sum_n (\frac{1}{2} V_{nijkl} [1 + \cos(n \tau_{ijkl} - \tau_{ijkl})])$$

- Typically, n is taken in [1 ... 6]. The most common values are 1, 2, 3.
- Sometimes, other harmonic terms are added, named "improper torsion terms", to account for enforcing planarity/chirality in particular groups. Improper torsion terms are adopted also in protein force fields.

Local "Cross Terms"

- Cross terms (i.e. involving simultaneously different molecular variables) are of primary importance only in force fields developed for small molecular systems (not proteins!).
- Cross terms are aimed at modeling correlations between bond-length/bond-angle potentials, as well as between dihedral/bond-angle and improper/improper-dihedral terms
- Cross terms typically show the following aspect:

- $U_{rr'} = S (r - \underline{r}) (r' - \underline{r}')$
- $U_{\theta\theta'} = K (\theta - \underline{\theta}) (\theta' - \underline{\theta}')$
- $U_{r\theta} = SK (r - \underline{r}) (\theta - \underline{\theta})$
- $U_{\tau\theta} = KV_{\tau\theta} (\theta - \underline{\theta}) \cos(\tau - \underline{\tau})$

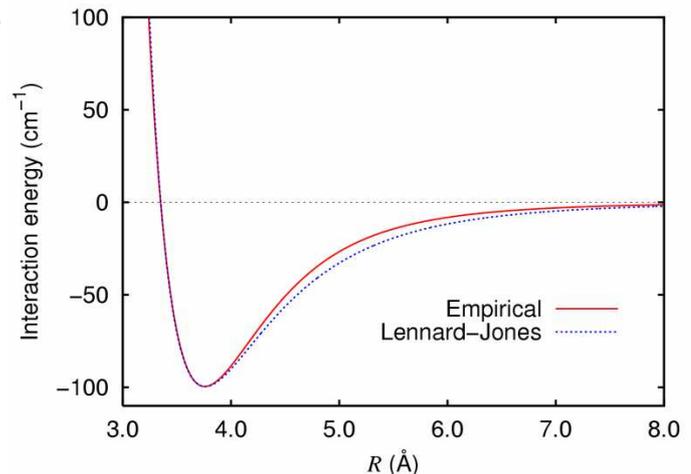
Non-local: Van der Waals

- For macromolecules, the Van der Waals potential is taken in the common 6/12 Lennard-Jones form, with attractive (6) and a repulsive (12) portions:

- $U_{LJ} = -A_{ij}/r_{ij}^6 + B_{ij}/r_{ij}^{12}$ (between atoms i and j)

- A and B depend on the types of interacting atoms

- As $r_{ij} \rightarrow \infty$, $U_{LJ} \rightarrow 0$ very rapidly, it is convenient to introduce a cutoff radius r_{cutoff} : only interactions within r_{cutoff} are calculated.



Non-local: Coulombic

- Coulomb's law: $F(r_{ij}) \propto -q_i q_j / r_{ij}^2$
- So, $U_{coul} = K_{coul} q_i q_j / \epsilon r_{ij}$ with $K_{coul} = 1/(4\pi\epsilon_0)$
- As $r_{ij} \rightarrow \infty$, $U_{coul} \rightarrow 0$ very slowly (unlike U_{LJ}), and this fact:
 - becomes essential in stabilizing macromolecules in solvent
 - creates computational problems, because evaluation of all pairwise terms takes $O(N^2)$ (anyway, $O(N)$ methods have been found)
- Further issues come from the expression of the (distance-dependent) dielectric function $\epsilon = \epsilon(r)$, as in a medium like water charges becomes somehow screened.

Parameterization

- One defined, a force field contains a lot of parameters, whose values must be properly chosen.
- Theoretically, each energy term could be calibrated according to specific experimental data from small molecules; in practice, approximation is unavoidable because of:
 - Scaling the system from small to large chemical groups
 - Interactions with solvent and counterions
- In summary, many different parameterizations are possible: anyway, only correct ones lead to reliable structural predictions

Oobatake-Crippen FF (1981)

- Example of 1st gen. FFs; operates on a “virtual bonds model”
- It considers the positions of C^αs: x_i is the position of the i-th C^α and the “virtual” bond length r_0 is 3.8 Å

- $U = U_b + U_{nb}$ $U_b = \frac{k_b}{2} \sum_{i=1}^{N-1} (r_{i,i+1} - r_0)^2$

- $U_{nb} = U_1 + U_2 + U_3$ one 4/6 LJ pot. + two **gaussian** pot.

$$U_1 = \frac{e_1}{n-m} \left[-m \left(\frac{r_1}{r} \right)^n + n \left(\frac{r_1}{r} \right)^m \right] \quad m=6, n=4$$

$$U_2 = e_2 e^{-\frac{1}{2} \left(\frac{r-r_2}{d_2} \right)^2}$$

$$U_3 = e_3 e^{-\frac{1}{2} \left(\frac{r-r_3}{d_3} \right)^2}$$

Constants $r_1, r_2, r_3, e_1, e_2, e_3, d_2$ and d_3 depends on the residue types.

Classical Force Fields

- CHARMM – www.charmm.org
 - a set of FFs for many-particle systems (mainly biomolecules) from Harvard
- AMBER – www.ambermd.org
 - a set of FFs for the simulation of biomolecules, developed at UCSF
- GROMACS – www.gromacs.org
 - FF used in the efficient molecular dynamics tool developed at University of Groningen, addressing biochemical molecules (proteins and also polymers)

U to Explore Vibrations (I)

- U depends on $p = 3n - 6$ independent variables; so far, for them we have used internal coordinates: let's call them collectively as vector q ; so we have $U = U(q)$
- Internal coordinates q can be related to the $3n$ Cartesian coordinates by the Wilson matrix B :
$$q = BX$$
- Given an equilibrium point q_e for the molecule, there $U(q)$ will present a minimum, so $\nabla U(q)|_{q=q_e} = 0$. Usually, $\nabla U(q)$ is called g

U to Explore Vibrations (II)

- Calling $\xi \equiv (q_1 - q_{e1} \dots q_{(3p-6)} - q_{e(3p-6)})^T$, the Taylor expansion of $U(q)$ about the point q_e is

$$U(q) - U(q_e) = \xi^T \nabla U(q)|_{q=q_e} + \frac{1}{2} \xi^T H(q)|_{q=q_e} \xi + \dots$$

H is the Hessian matrix of U: $H_{ij} = \partial^2 U / \partial q_i \partial q_j$

- Usually, matrix $H(q)|_{q=q_e}$ is called F
- At minima, eigenvalues of H are all positive.