

C7790 Introduction to Molecular Modelling

TSM Modelling Molecular Structures

Lesson 23 Molecular Mechanics I

PS/2020 Distant Form of Teaching: Rev1

Petr Kulhánek

kulhanek@chemi.muni.cz

National Centre for Biomolecular Research, Faculty of Science
Masaryk University, Kamenice 5, CZ-62500 Brno

Context

macroworld

states

(thermodynamic properties, G, T,...)

phenomenological thermodynamics

equilibrium (equilibrium constant)

kinetics (rate constant)

free energy
(Gibbs/Helmholtz)



partition function

statistical thermodynamics

microstates

(mechanical properties, E)

microstate \neq microworld

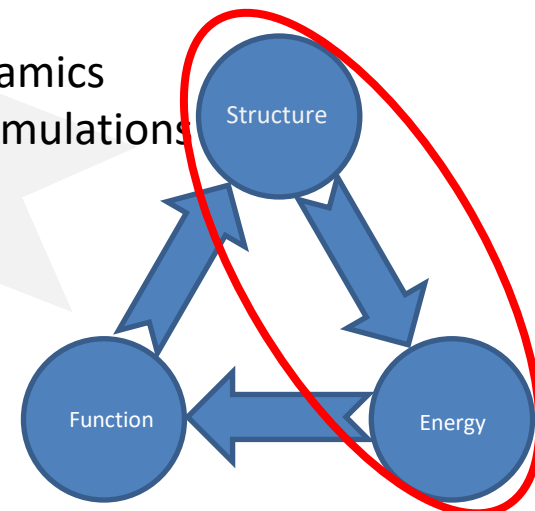
microworld

Description levels (model chemistry):

- quantum mechanics
 - semiempirical methods
 - ab initio methods
 - post-HF methods
 - DFT methods
- **molecular mechanics**
- coarse-grained mechanics

Simulations:

- molecular dynamics
- Monte Carlo simulations
- docking
- ...



Molecular Mechanics

Schrodinger equation - quantum mechanical description

$$\hat{H}\psi_k^{\mathbf{R}}(\mathbf{r}_e) = E_k(\mathbf{R})\psi_k^{\mathbf{R}}(\mathbf{r}_e)$$

approximation

electron motions is omitted

(electron motions is implicitly included in empirical parameters)

$$E_k(\mathbf{R}) = \underbrace{E_{bonds} + E_{angles} + E_{torsions}}_{\text{bonded contributions}} + \underbrace{E_{ele} + E_{vdw} + \dots}_{\text{non-bonded contributions}}$$

bonded contributions

non-bonded contributions

Classical physics - mechanical description

Nomenclature

Molecular mechanics uses classical mechanics to model molecular systems. The **Born–Oppenheimer approximation is assumed valid**, and the potential energy of all systems is calculated as a function of the atomic coordinates using force fields.

Coarse-grained modeling, coarse-grained models, aim at simulating the behavior of complex systems using their coarse-grained (simplified) representation. Coarse-grained models are widely used for molecular modeling of biomolecules at **various granularity levels**.

In the context of chemistry and molecular modelling, a **force field** is a **computational method** that is used to estimate the forces between atoms within molecules and between molecules.

More precisely, the force field refers to the **functional form** and **parameter sets** used to calculate the **potential energy of a system** of atoms or coarse-grained particles.

$$E_k(\mathbf{R}) = E_{bonds} + E_{angles} + E_{torsions} + E_{ele} + E_{vdw} + \dots$$

+ parameters

[https://en.wikipedia.org/wiki/Force_field_\(chemistry\)](https://en.wikipedia.org/wiki/Force_field_(chemistry))

https://en.wikipedia.org/wiki/Coarse-grained_modeling

https://en.wikipedia.org/wiki/Molecular_mechanics

Bonded Contributions

Bonded Contributions

Main contributions

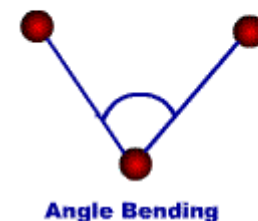
Bond stretching

$$E_{bonds} = \sum_{b=1}^{bonds} \frac{1}{2} K_b (d_b - d_{b0})^2$$



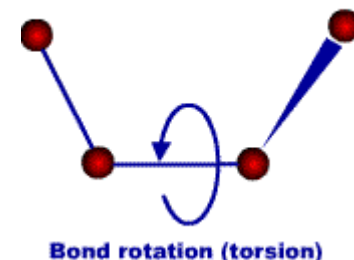
Angle bending

$$E_{angles} = \sum_{a=1}^{angles} \frac{1}{2} K_a (\theta_a - \theta_{a0})^2$$



Bond rotation

$$E_{torsions} = \sum_{t=1}^{torsions} \sum_n \frac{V_{t,n}}{2} (1 + \cos[n\varphi_t - \delta_{t,n}])$$



 empirical parameters

Bond stretching

- The simplest description of a valence bond deformation is provided by **the harmonic approximation**.
- Valence bond deformation is then effectively described in the limit of **Hooke's law**.

Harmonic potential

$$E_b(r) = \frac{1}{2} K (r - r_0)^2$$

stiffness (force constant) \swarrow \nwarrow equilibrium bond distance

$$\frac{dE_b(r)}{dr} = -F = F_S = K(r - r_0)$$

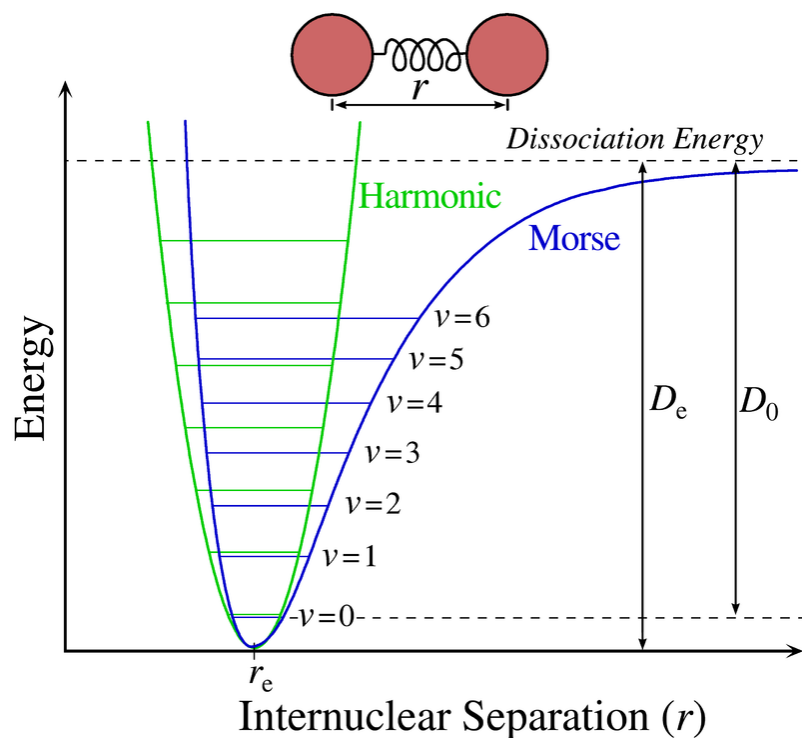
restoring force \swarrow \nwarrow deformation force

$F_S = K\Delta r$

- **Hooke's law** is a law of physics that states that the force (F_S) needed to extend or compress a spring by some distance (Δr) scales linearly with respect to that distance. The scaling factor (K) is a stiffness constant.
- Higher order potentials such as a cubic potential can be used, but they exhibit catastrophic behavior at longer stretching distances. Why?

Bond stretching, cont.

- The harmonic potential does not have dissociation limit. Thus, **force fields employing the harmonic approximation cannot describe reactivity.**
- The reactivity is difficult to describe even with Morse potential properly, a noticeable exceptions are **ReaxFF (reactive force field)** and **EVB (empirical valence bond).**



Harmonic potential

$$V(r) = \frac{1}{2} K (r - r_0)^2$$

Morse potential

$$V(r) = D_e \left(1 - e^{-a(r-r_0)} \right)^2$$

Disadvantage of Morse potential

- more parameters are needed
- more computationally demanding

Angle bending

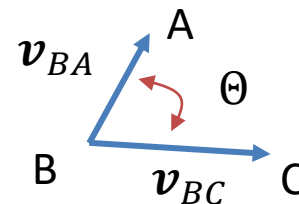
- The simplest description of a valence angle deformation is provided by **the harmonic approximation** similarly to a valence bond.

$$E_a = \frac{1}{2} K (\Theta - \Theta_0)^2$$

↑
stiffness (force constant)

↙
equilibrium valence angle

$$\cos(\Theta) = \frac{\mathbf{v}_{BA} \cdot \mathbf{v}_{BC}}{|\mathbf{v}_{BA}| |\mathbf{v}_{BC}|}$$



- Alternative forms

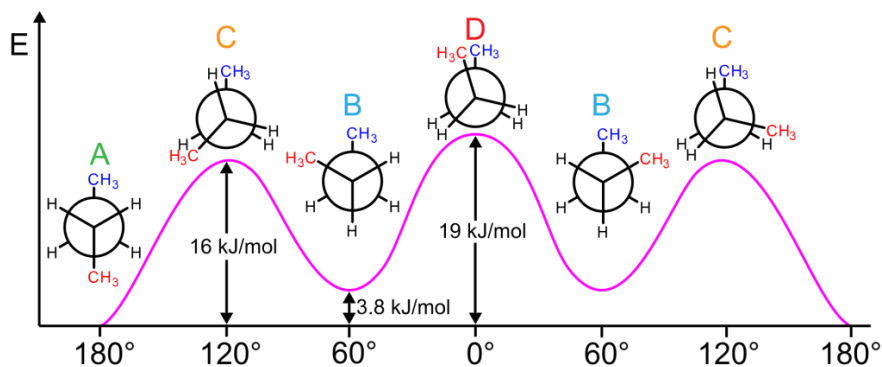
- GROMOS Forcefield

$$E_a = \frac{1}{2} K [\cos(\Theta) - \cos(\Theta_0)]^2$$

this form avoids singularities in gradient calculations for straight angles (180°)

- Higher order potentials
- **Cross-terms**, which describes coupling between bond stretching and angle bending.

Bond rotation



Bond rotations can be described by torsion (dihedral) angles.

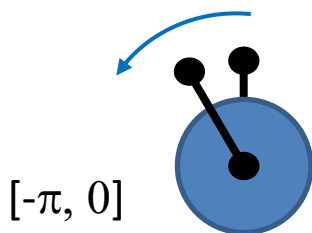
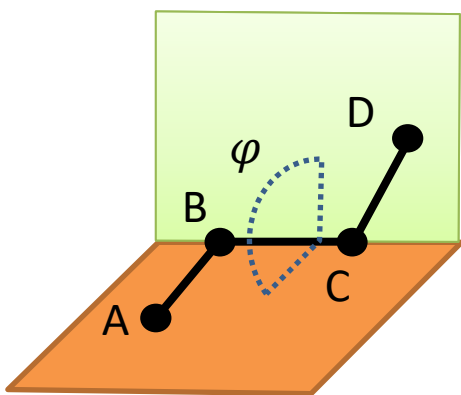
Torsion angle (dihedral angle) is signed angle between two planes A-B-C and B-C-D.

only $[0, \pi]$

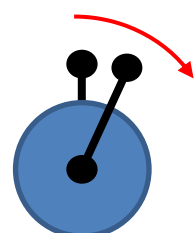
$$\cos(\Theta) = \frac{\mathbf{n}_{ABC} \cdot \mathbf{n}_{BCD}}{|\mathbf{n}_{ABC}| |\mathbf{n}_{BCD}|}$$

normal vectors (perpendicular to the plane)

$$\mathbf{n}_{ABC} = \mathbf{v}_{BA} \times \mathbf{v}_{BC}$$



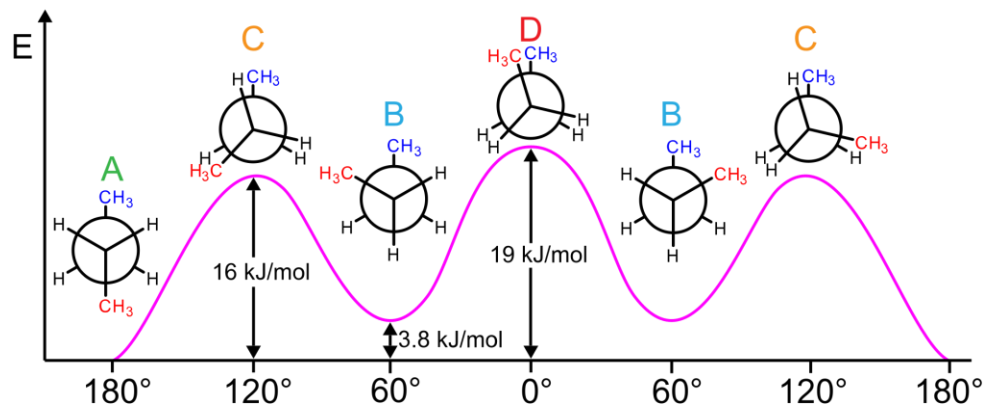
(-) anti clock-wise



(+) clock-wise

$[0, \pi]$

Bond rotation, cont.



torsional energy is a **periodic function**

- Torsional energy can be approximated by a **Fourier series**.

$$E_t = \sum_{n=1}^K \frac{V_n}{2} [1 - \cos(n\varphi - \gamma_n)]$$

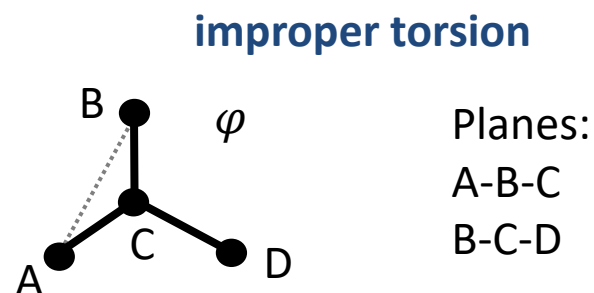
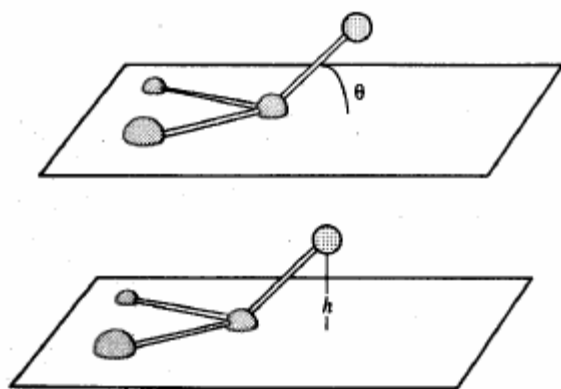
length of the Fourier series
amplitude
phase

This equation is for a single torsion!!!!

- Due to computational complexity, the length of the Fourier series is truncated. Typical lengths are up to $K = 4$.

Out-of-plane deformations

- Angle bending and bond rotations are not able to properly describe out-of-plane deformations.
- **Out-of-plane deformations** are observed in planar parts of molecules such as **aromatic rings or peptide bonds** (conjugated system).



- There are several geometrical descriptions of out-of-plane deformations (deviation from the plane or improper torsion).
- The deformation energy is expressed as either harmonic potential or a Fourier series.

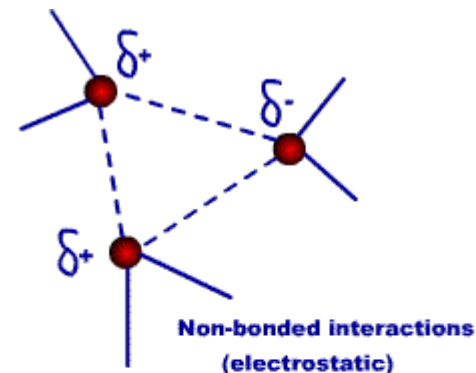
Non-Bonded (Non-covalent) Contributions

Non-bonded Contributions

Main contributions

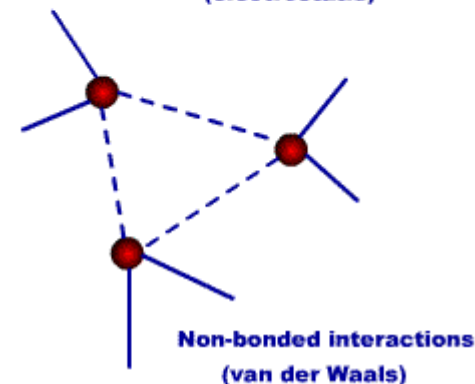
Electrostatic interactions

$$E_{ele} = \sum_{i=1}^N \sum_{j=i+1}^N \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}$$



van der Waals interactions

$$E_{vdw} = \sum_{i=1}^N \sum_{j=i+1}^N 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$



N – number of atoms

 empirical parameters

Excluded Interactions

- Non-bonded interactions are computed between ALL atoms except of those pairs, which are explicitly described by bonded terms.
- Non-bonded interactions are thus both of **intra and intermolecular origin**.
- Excluded interactions are:
 - 1-2 (bond stretching)
 - 1-2-3 (angle bending)
 - 1-2-3-4 (bond rotations) are not fully excluded but instead they are scaled

Electrostatic interactions

- Electron density and positively charged nuclei of atoms are **approximated by point charges**.
- Interaction between point charges is provided by Coulomb potential.

$$E_{ele} = \sum_{i=1}^N \sum_{j=i+1}^N \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}$$

- This simple model does not describe **induction** and **penetration effect**.
- Penetration electrostatic energy is always attractive and is consequence of overlap of smeared electron density and point nuclear charges at short distances. However, the effect of penetration energy is absorbed in LJ potential in the simple FF.
- Polarizable force fields
 - they try to include effect of induction employing polarizable electrostatic models (induced dipoles, Drude oscillator, fluctuating charges, etc...)
 - more computationally demanding, improvement is questionable

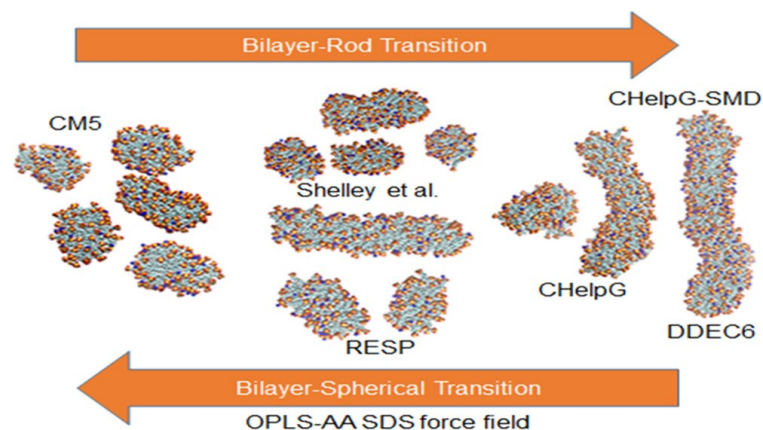
Point charges

Atomic partial charges are not observable. Rather, they are (crude) approximation of electron density and point nuclear charges.

Point charges must describe well electrostatic potential of molecules and their interactions.

Sources of charges are QM calculations:

- **Restrained ElectroStatic Potential Charges (RESP)** derived from ESP charges.
 - RESP corrects chemically non-intuitive charges on carbon atoms.
 - RESP/ESP procedure fails on large structures with buried atoms.
- CM5 charges
- CHelpG charges
- DDEC6 charges
- others ...

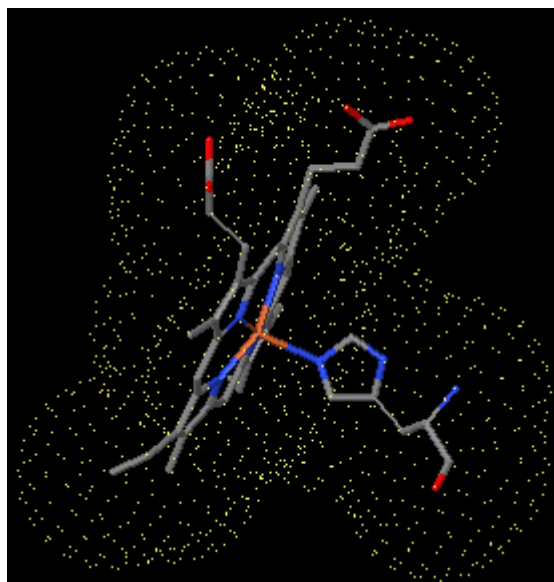


Abdel-Azeim, S. Revisiting OPLS-AA Force Field for the Simulation of Anionic Surfactants in Concentrated Electrolyte Solutions. *J. Chem. Theory Comput.* **2020**, *16* (2), 1136–1145.
<https://doi.org/10.1021/acs.jctc.9b00947>.

ESP charges

ESP charges (ElectroStatic Potential) are charges derived from electrostatic potential. The principle of charge calculation consists of two steps:

1. calculation of electrostatic potential V_{QM} from wave function on discretized molecular envelope (set of points)
2. finding point atomic charges that create electrostatic potential V_{PC} which is in the best agreement with the quantum mechanical potential (least squares method)



$$\Psi \rightarrow V_{QM,p} \quad V_{PC,p} \leftarrow q_k$$

$$\sum_p (V_{QM,p} - V_{PC,p})^2 = \min!$$

is searched by least squares method

ESP charges and their derivatives are used in **molecular mechanics** because, by their nature, they describe well **electrostatic properties** of molecules/system.

<http://biomodel.uah.es/Jmol/surfaces/inicio.htm>

van der Waals interactions

van der Waals interactions are very often described by **Lenard-Jones potential**.

Alternative forms:

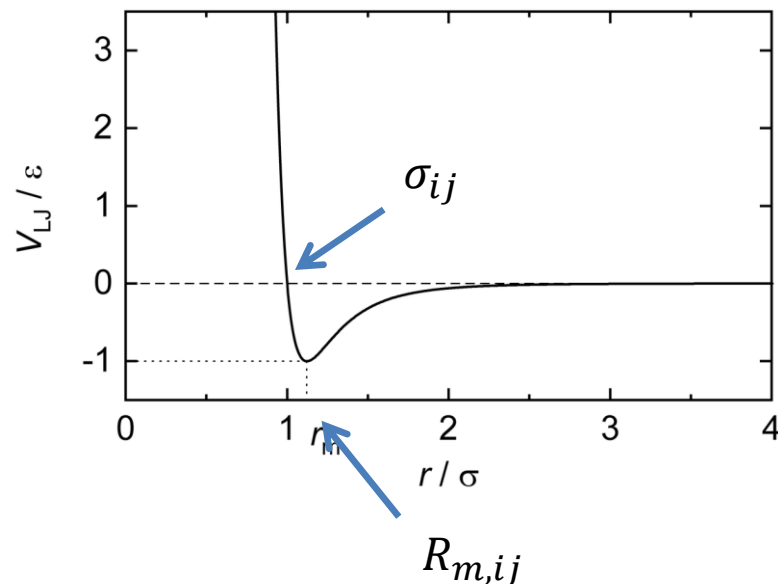
$$E_{vdw} = \sum_{i=1}^N \sum_{j=i+1}^N 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

$$E_{vdw} = \sum_{i=1}^N \sum_{j=i+1}^N \varepsilon_{ij} \left[\left(\frac{R_{m,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{m,ij}}{r_{ij}} \right)^6 \right]$$

$$E_{vdw} = \sum_{i=1}^N \sum_{j=i+1}^N \left[\left(\frac{A_{ij}}{r_{ij}} \right)^{12} - \left(\frac{B_{ij}}{r_{ij}} \right)^6 \right]$$

↑
repulsive part
(Pauli repulsion)

↑
attractive part
(dispersion)



"Better" models:

- exp-6 potential (however, it is unphysical at short distances)

Combining rules

In computational chemistry and molecular dynamics, the **combination rules** or **combining rules** are equations that provide the interaction energy between **two dissimilar non-bonded atoms**, usually for the part of the potential representing the van der Waals interaction.

Combining rules considerably reduce the number of needed parameters.

Lorentz-Berthelot rules

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \quad \text{geometric mean}$$

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \quad \text{arithmetic mean}$$

LB rules are rather inaccurate. However, their success relies on the fact that the number of unlike interactions is considerably higher than between like atoms.

Then, like parameters are optimized in such a way that resulting unlike parameters performs well. Side outcome is that like parameters alone do not perform well.

https://en.wikipedia.org/wiki/Combining_rules

FF Parameters

Atom Types

➤ Force Filed parameters are required for:

- each valence bond (K_b, r_0)
- each valence angle (K_a, Θ_0)
- each torsion angle (series of V_n, γ)
- each atom ($q_i, \epsilon_{ii}, \sigma_{ii}$)

This is impractical due large number of required parameters!!!

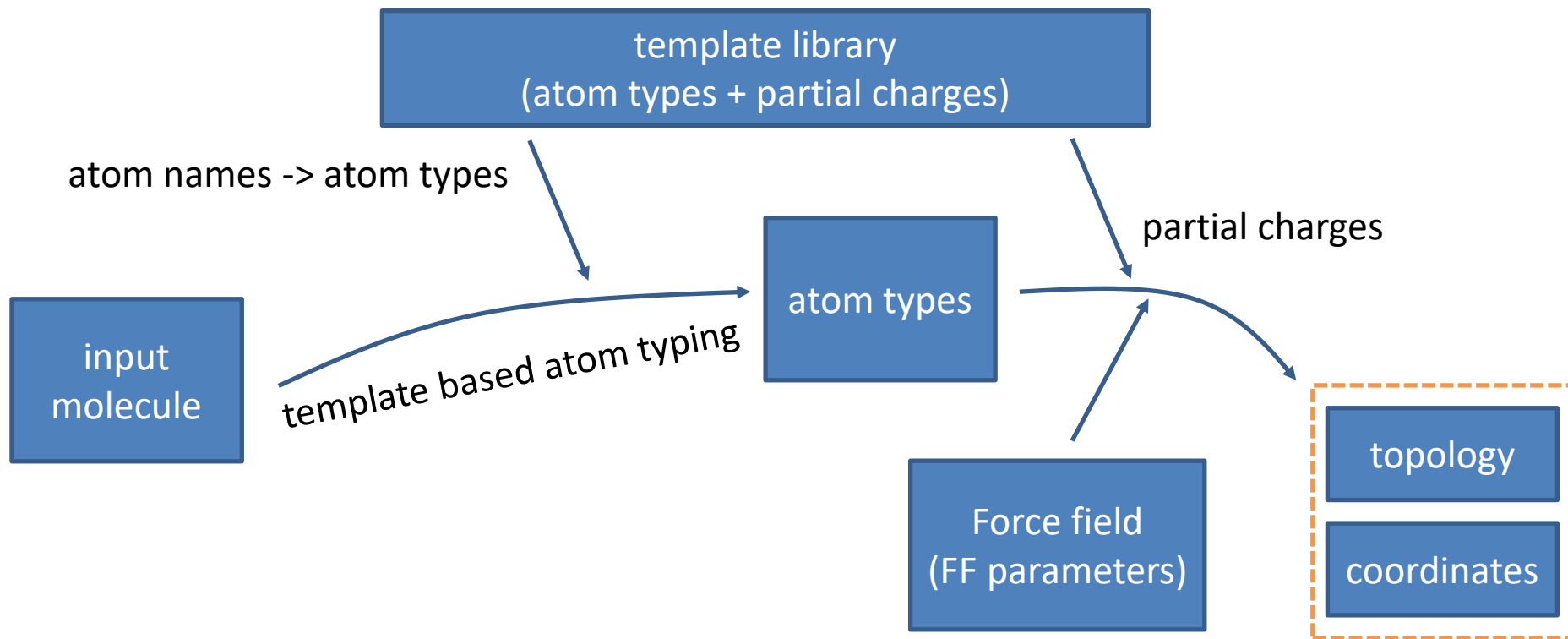
solution

Atom Types:

- Atoms types group atoms within the similar chemical environment

PARM99 for DNA,RNA,AA, organic molecules, TIP3P wat. Polariz.& LP incl.02/04/99			
C	12.01	0.616	! sp2 C carbonyl group
CA	12.01	0.360	sp2 C pure aromatic (benzene)
CT	12.01	0.878	sp3 aliphatic C
CV	12.01	0.360	sp2 arom. 5 memb.ring w/1 N and 1 H (HIS)
CW	12.01	0.360	sp2 arom. 5 memb.ring w/1 N-H and 1 H (HIS)
C*	12.01	0.360	sp2 arom. 5 memb.ring w/1 subst. (TRP)
CY	12.01	0.360	nitrile C (Howard et al.JCC,16,243,1995)
CZ	12.01	0.360	sp C (Howard et al.JCC,16,243,1995)
H	1.008	0.161	H bonded to nitrogen atoms
HC	1.008	0.135	H aliph. bond. to C without electrwd.group
H1	1.008	0.135	H aliph. bond. to C with 1 electrwd. group
HA	1.008	0.167	H arom. bond. to C without elctrwd. groups

Assigning Parameters - I

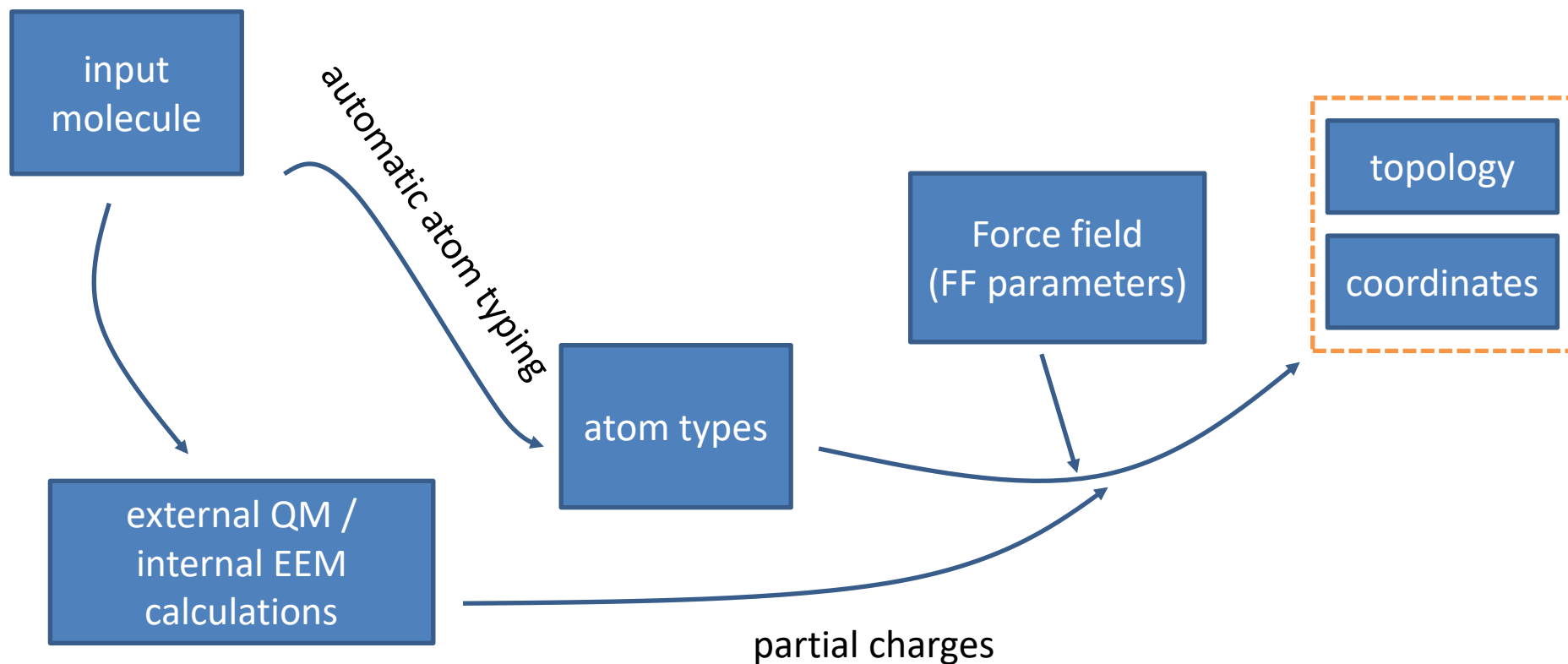


predefined template libraries for components of biomolecular structures

- aminoacid residues
- nucleotides

Assigning Parameters - II

- typically, small organic molecules
- partial atomic charges are calculated independently



Parameter optimization

Parameters can be derived from:

- experimental data
 - bulk properties (density, etc.)
 - thermodynamics (binding affinities)
 - experimental geometries (equilibrium bond lengths and valence angles)
 - vibrational spectra (force constants)
- computational data (QM calculations)
 - geometry (equilibrium values)
 - vibrational analysis (force constants)
 - potential energy scans for bond rotations

Transferability of parameters is key for general use in modelling.

Common Force Fields

- AMBER** (Assisted Model Building and Energy Refinement) – widely used for proteins and DNA.
- CHARMM** (Chemistry at HARvard Molecular Mechanics) – originally developed at Harvard, widely used for both small molecules and macromolecules
- GROMOS** (GRoningen MOlecular Simulation) – a force field that comes as part of the GROMOS software, a general-purpose molecular dynamics computer simulation package for the study of biomolecular systems. GROMOS force field A-version has been developed for application to aqueous or apolar solutions of proteins, nucleotides, and sugars. A B-version to simulate gas phase isolated molecules is also available.
- MMFF** (Merck Molecular Force Field) – developed at Merck for a broad range of molecules.
- OPLS** (Optimized Potential for Liquid Simulations) (variants include OPLS-AA, OPLS-UA, OPLS-2001, OPLS-2005, OPLS3e, OPLS4) – developed by William L. Jorgensen at the Yale University Department of Chemistry.
- GAFF** – a general AMBER force field was developed for rational drug design. GAFF is compatible with the AMBER force field, and it has parameters for almost all the organic molecules made of C, N, O, H, S, P, F, Cl, Br and I. As a complete force field, GAFF is suitable for study of a great number of molecules (such as database searching) in an automatic fashion.

[https://en.wikipedia.org/wiki/Force_field_\(chemistry\)](https://en.wikipedia.org/wiki/Force_field_(chemistry))