Molecular Mechanics

Empirical Potential
Force Field

Focused on biomolecules

Michal Otyepka
Department of Physical Chemistry
Faculty of Science, Palacky University Olomouc

4. 12. 2011

Why? There is QM.

Small molecule – tractable system at high-level QM

Medium size molecule – limited quality QM

(HF/6-31G(d); GGA/LDA - DFT)

Ordinary protein – intractable system by QM, but … SCC-DFT-B, PM6-DH …

Compromise

<table>
<thead>
<tr>
<th>MM</th>
<th>LDA</th>
<th>GGA</th>
<th>hybrid DFT</th>
<th>MP2</th>
<th>CCSD(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model size</td>
<td>100.000</td>
<td>10.000</td>
<td>1.000</td>
<td>100</td>
<td>10</td>
</tr>
</tbody>
</table>

Quality (reliability)

Computer cost

representative model
approximative method

incomplete model
reliable method

CCSD(T)/CBS the gold standard (thermochemical or higher accuracy, err. < 1 kcal/mol)
**Born-Oppenheimer approximation**

Separation of electronic and nuclear motions

Quantum electrons vs. Classical nuclei

\[ E = f(R) \]

Tractable by classical physics

molecular mechanics

---

**Bond Distance Deformation**

Behavior close minimum

Taylor expansion

\[ f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n. \]

\[ E(r) = E(r_0) + \frac{1}{2} \frac{\partial^2 E}{\partial r^2} (r-r_0) + \ldots \]

\[ E(r) = \frac{1}{2} k (r-r_0)^2 \]

---

**Bond as Spring**

\[ F = -\frac{\partial E}{\partial r} = -k(r-r_0) \]

\[ r - r_0 = x \]

deviation from equil. position

\[ F = -kx \]

Hooke’s law – strain is directly proportional to stress

\[ \text{Force (spring constant)} \]

\[ \frac{\partial^2 E}{\partial r^2} \]

Morse pot.

---

*Ut tensio, sic vis, meaning, “As the extension, so the force”.*
Different Bonds
Different Springs

various covalent bonds have various
\( r_0 \), bond distance
\( k \), force constant

<table>
<thead>
<tr>
<th>molecule</th>
<th>( k / \text{N} \cdot \text{m}^{-1} )</th>
<th>( r_0 / \text{pm} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{H}_2 )</td>
<td>510</td>
<td>74.1</td>
</tr>
<tr>
<td>( \text{H}^3\text{Cl} )</td>
<td>478</td>
<td>127.5</td>
</tr>
<tr>
<td>( \text{H}^7\text{Br} )</td>
<td>408</td>
<td>141.4</td>
</tr>
<tr>
<td>( \text{H}^{127}\text{I} )</td>
<td>291</td>
<td>160.9</td>
</tr>
</tbody>
</table>

Bond Types

similar bonds \( \text{X} - \text{Y} \) share similar behavior in all
molecules, are insensitive to environment
(context)
- Parameters are transferable
- Transferability: Application of empirical force field
parameters to molecules not explicitly included during the
parameter optimization.

To define similar bonds, to assign \( k \) a \( r_0 \)
Atom types are defined

Carbon Atom Types
(in biomolecules – parm99/AMBER)

PARM99 for 200,000+ AL organic molecules, TIP3P wat.,
C sp3 C aliphatic
CA sp2 C carbonyl group
CB sp2 C pure aromatic (benzene)
CC sp2 C 5- and 6-membered ring junction
CD sp2 C atoms in the middle of \( \text{C} = \text{CD} - \text{CD}= \text{C} \)
CK sp2 C 5-membered ring in purines
CM sp2 C pyrimidines in pos. 5 & 6
CN sp2 C aromatic 5-membered ring
CQ sp2 C in 5 mem. ring of purines between \( 2 \text{~N} \)
CR sp2 C arom as \( \text{CQ} \) but in HIS
CT sp3 aliphatic C
CV sp2 arom. 5 mem. ring w/1 N and 1 H
C* sp2 arom. 5 mem. ring w/1 subst.
CY nitrile C (Howard et al., JCC, 16, 243, 1995)
CZ sp C (Howard et al., JCC, 16, 243, 1995)
Bond Types

<table>
<thead>
<tr>
<th>Bond Type</th>
<th>k</th>
<th>r₀</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT-CT</td>
<td>310.0</td>
<td>1.526</td>
</tr>
<tr>
<td>CT-EC</td>
<td>340.0</td>
<td>1.000</td>
</tr>
<tr>
<td>CT-EQ</td>
<td>340.0</td>
<td>1.000</td>
</tr>
<tr>
<td>CT-EF</td>
<td>340.0</td>
<td>1.000</td>
</tr>
<tr>
<td>CT-EP</td>
<td>340.0</td>
<td>1.000</td>
</tr>
<tr>
<td>CT-EH</td>
<td>337.0</td>
<td>1.475</td>
</tr>
<tr>
<td>CT-EI</td>
<td>337.0</td>
<td>1.463</td>
</tr>
<tr>
<td>CT-CF</td>
<td>320.0</td>
<td>1.410</td>
</tr>
<tr>
<td>CT-CG</td>
<td>320.0</td>
<td>1.410</td>
</tr>
<tr>
<td>C+C</td>
<td>307.0</td>
<td>1.080</td>
</tr>
<tr>
<td>C+D</td>
<td>306.0</td>
<td>1.469</td>
</tr>
<tr>
<td>C+E</td>
<td>317.0</td>
<td>1.469</td>
</tr>
<tr>
<td>C+F</td>
<td>319.0</td>
<td>1.392</td>
</tr>
<tr>
<td>C+G</td>
<td>407.0</td>
<td>1.410</td>
</tr>
</tbody>
</table>

Database of parameters – AMBER force fields

How to Derive Parameters?

- From experiments
  - bond geometries
  - X-ray and neutron scattering, NMR, rotational spectroscopy
  - force constants
  - vibrational spectroscopies
- From calculations
  - fit of energy curves

Molecular Mechanics

potential energy as a function of nuclei positions

\[
E = f(R) = E_{\text{covalent}} + E_{\text{noncovalent}}
\]

\[
E_{\text{covalent}} = E_b + E_a + E_l
\]

\[
E_{\text{noncovalent}} = E_c + E_{vdw}
\]

Additive model
Simple form of the potential energy function
Internal Terms

\[ E_{\text{coulomb}} = E_p + E_a + E_t \]

Bond Stretching
Angle Bending
Torsion/Dihedral Rotation

Angle Bending

\[ E = \frac{k_t}{2} (\theta - \theta_0)^2 \]

- \( k_t = 80 \text{ kcal/mol deg}^2 \)
- \( \theta_0 = 122.9^\circ \)

Torsions/Dihedrals

\[ E = \frac{k_t}{2} \left( 1 + \cos(n\phi - \phi_0) \right) \]

- \( n = 3 \)
- \( \phi_0 = 180^\circ \)
- \( k_t = 2.9/2^{9} \text{ (IDIVF=9)} \)
- \( k_t = 2.9/2 \text{ (IDIVF=9)} \)
Dihedrals

Improper Torsion

Conformational Behavior

- Conformational behavior of biomacromolecules is given mostly by dihedral parameters (and non-bonded terms)
- Parameterization
  - Deduced/derived from experimental data (NMR, X-ray - distributions) – „knowledge based“ – CHARMM
  - QM profiles – QM in gas phase // QM in solvent ($\delta_i = 80$)
- CMAP – 2D dihedral energy correction map, CHARMM (some dihedrals are correlated)
The popular CHARMM 22 FF (C22) is shown on the left, giving a poor Ramachandran plot. In 2004, Alexander MacKerell added a grid-based φ/ψ term (CMAP) to the CHARMM 22 (C22) force field (Mackerell 2004 J. Comput. Chem. 25:1584). Ramachandran plot looks acceptable to a crystallographer.

Nonbonded Terms

$$E_{\text{noncovalent}} = E_c + E_{\text{vdW}}$$

Coulomb (electrostatic) interaction
Van der Waals forces

$$u(R_1,......R_n) = \sum_{j<k} u(R_j, R_k) + \sum_{i<k} u(R_i, R_j) + ...$$

$$u(R_1,......R_n) = \sum_{j<i} u(R_j, R_i)$$  model of pair potential

$$u(R_1,......R_n) - \sum_{i<j} u(R_j, R_i) = 0$$  effective pair potential
Non-Covalent Interactions

- H-bonds (bridges); blues-shifting H-bond
- Salt bridges
- π-π interaction, ion-π interaction
- Van der Waals forces
- Halogen bond
- Dihydrogen bonds
- *Hydrophobic effect*

The Four Basic NC Interactions

Coulombic (Electrostatic)
  - p. multipole – p. multipole (+, -)
  - Polarization
  - p. multipole - ind. multipole (-)
  - London (dispersion) forces
  - inst. multipole – inst. multipole (-)
  - Repulsion
    - (Pauli) electron exchange (+)

  (+) repulsive
  (-) attractive

Coulombic Interaction

- multipole-multipole interaction in monopole expansion
- atomic centred parcial charges (how to derive them?)
- Coulomb law

\[ E = \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r} \]
Electrostatics

Partial Charges

- Nonobservable!
- Mulliken – many drawbacks
- RESP – “Restrained ElectroStatic Potential fit” – based on fit of electrostatic potential of molecule
  - HF/6-31G* - overestimated dip. moments, compensation of missing induction (parm99)
  - B3LYP/cc-pVTZ/PCM(ε_r=4) … (ff03)
  - Polarization can be treated explicitly $E_{pol} = \frac{1}{2} \sum_i \mu_i \cdot \epsilon_i$

RESP
\[
\chi_j = \chi_{j}^{\text{resp}} + \chi_{j}^{\text{esp}} + \lambda g(q_j)
\]
\[
\chi_{j}^{\text{resp}} = \sum_i (v_i - \bar{v}_j)^2, \bar{v}_j = \frac{\sum_i q_i}{r_0}
\]
\[
\chi_{j}^{\text{esp}} = a \sum_i \left(q_i^2 + b^2 \right)^{1/2} - b
\]

**Induction/Polarization**

- In many classical FF neglected
- FF with polarisation
  - fluctuating charges, induced dipole, Drude oscillators
  - slow, small improvement

London Dispersion Forces

\[ E_{\text{disp}} \approx -\frac{3\alpha^4 a^3 I_A I_B}{4(I_A + I_B)} R^{-6} \]

This equation represents the London dispersion interaction between two atoms, where \( a \) and \( I \) are the static polarizabilities and ionization potentials of the respective atoms. The parameters \( a \) and \( I \) are the first ionization potentials of the atoms and \( R \) is the intermolecular distance.

Van der Waals Term

covers dispersion and repulsion int.

Lennard-Jones Potential

\[ u(r) = 4\varepsilon \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^{6} \]

\[ u(\sigma) = 0, \quad u(\infty) = \varepsilon, \quad \sigma_{\text{vdw}} / 2 \]

van der Waals radius

Why is LJ potential 12-6 computationally effective?

Enumeration of square is quick

\[ r^{-12} \approx (r^{-6})^2 \]

Repulsion increases exponentially!
Mixing Rules

- How to derive $\sigma$ and $\epsilon$ for atoms A and B?
- Lorentz/Berthelot mixing rule

$$\sigma_{ab} = \frac{\sigma_a + \sigma_b}{2}$$

$$\epsilon_{ab} = \sqrt[3]{\epsilon_a \epsilon_b}$$

$$u(\epsilon) = \left( \frac{\sigma}{\epsilon} \right)^2 - 2 \left( \frac{\sigma}{\epsilon} \right)^4 = \sqrt[3]{\epsilon} \epsilon \left( \frac{\sigma_a + \sigma_b}{2} \right)^2 - 2 \left( \frac{\sigma_a + \sigma_b}{2} \right)^4$$

Large Compensation of Errors – parm99

![Graph showing compensation of errors in pairwise-additive empirical force fields.](image)

Minimum – vdw complex

- Eikot
- Eiding
- Enrich
- Eriot

References:

Molecular Mechanics

**Empirical Potential Energy Function**

- Bonds
  $$E_b = \frac{k}{2} (r - r_0)^2$$
- Angles
  $$E_a = \frac{k}{2} (\theta - \theta_0)^2$$
- Torsions
  $$E_t = \frac{k}{2} (1 + \cos(n\phi - \phi_0))$$
- Non-bonded
  $$E_{\text{vdW}} = \frac{1}{4\pi\epsilon_0} \frac{9\epsilon q_i q_j}{r_{ij}}$$
  $$E_{\alpha} = -2\epsilon \left( \frac{\sigma_a}{r_{ij}} + \epsilon \right) \left( \frac{\sigma_b}{r_{ij}} + \epsilon \right)$$
Molecular Topology

defines, which bonds, angle, dihedrals etc. are applicable

### What is enumerated?

- **binding** – covalent bonds, angles, dihedrals only
- **coulomb** – 1-2, 1-3 neglected; 1-4 scaled (2.0), all (but PME)
- **vdW** – 1-2 and 1-3 neglected; 1-4 scaled (1.2), all (but ...)
  - to reduce number of non-covalent interactions – cutoff

### Cutoff

- Number of nonbonding terms \( \sim 2 \times N(N-1)/2 \)
  - for 10.000 atoms (smaller system) \( \sim 10^5 \) pairs
- vDW term quickly vanishes with distance \( (r^{-6}) \)
  - vDW inter. at 2.4x minimum distance is \( \sim 100 \times \) smaller than in minimum
  - cutoff for vdw interaction; pairs over \( r_{\text{max}} \) neglected
  - problem: a discontinuity is introduced (infinite derivations at \( r_{\text{max}} \))
    - **switching function**
Switching function

- Graph of van der Waals potential with and without the application of the switching function. With the switching function active, the potential is smoothly reduced to 0 at the cutoff distance. Without the switching function, there is a discontinuity where the potential is truncated.

Cutoff

- Cutoff for Electrostatics?
  - Electrostatic interaction of two monopoles vanishes slowly ($r^{-3}$)
  - At 10r is equal to 10% $E_{\text{el}}$ at r
  - When periodic boundaries (PBC) are used - Ewald summation can be used
  - PME - particle mesh Ewald
  - Interaction of two particles is summed at „short“ distances directly and at „long“ distances in Fourier space. Fast Fourier Transform is used

Cutoff for Electrostatics?

- Electrostatic interaction of two monopoles vanishes slowly ($r^{-3}$)
  - At 10r is equal to 10% $E_{\text{el}}$ at r
  - When periodic boundaries (PBC) are used - Ewald summation can be used
  - PME - particle mesh Ewald
  - Interaction of two particles is summed at „short“ distances directly and at „long“ distances in Fourier space. Fast Fourier Transform is used
More Approximative FF

- United Atoms
  - Non-polar hydrogens are not treated explicitly
  - E.g. –CH₃ is treated as one „pseudoatom“
  - Membrane simulations

- Coarse grained models (CG)
  - Beads
  - For very large systems
  - Rough approximation
  - Applications: membranes, protein complexes
  - CGFF – MARTINI
  - http://troll.med.unc.edu/ifoldrna/ (RNA CG)

Coarse Grained

- MARTINI
  - The forcefield has been parametrized in a systematic way, based on the reproduction of partitioning free energies between polar and apolar phases of a large number of chemical compounds.

  - The model is based on a four-to-one mapping, i.e. on average four heavy atoms are represented by a single interaction center. In order to keep the model simple, only four main types of interaction sites are defined: polar (P), non-polar (N), apolar (C), and charged (Q). Each particle type has a number of subtypes, which allow for an accurate representation of the chemical nature of the underlying atomistic structure.

  - Currently topologies are available for many lipids and surfactant molecules, including cholesterol, and for all amino acids. Scripts are furthermore available to build topologies for arbitrary peptides and proteins.
Common Force Fields

- organic molecules
  - (Allinger) MM2, MM3,
  - MMFF (Merck Molecular FF), CVFF ...
- biomacromolecules
  - AMBER
    - parm99, ff03, f10 ...
  - CHARMM
  - OPLS
2.2 The ff10 force field

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>amber.cif</td>
<td>This will load the files listed below</td>
</tr>
<tr>
<td>amber.in</td>
<td>basic force field parameters</td>
</tr>
<tr>
<td>amino50.in</td>
<td>topologies and charges for amino acids</td>
</tr>
<tr>
<td>amino50r.in</td>
<td>same, for N-terminal amino acids</td>
</tr>
<tr>
<td>amino50t.in</td>
<td>same, for C-terminal amino acids</td>
</tr>
<tr>
<td>mcs450.in</td>
<td>topologies and charges for nucleic acids</td>
</tr>
</tbody>
</table>

The FF10 force field collects a variety of updates and modifications to the general Amber force field, which is described below in Section 2.7. It incorporates the following additions and changes:

1. FF99SB for proteins. Several groups have noticed that ff99 and ff94 as well do not provide a good energy balance between helical and extended regions of peptide and protein backbones. Another problem is that many of the ff94 variants had incorrect treatment of glycine backbone parameters. FF99SB is the recent attempt to improve this behavior, and was developed in the Simmerling group[6]. It presents a careful reparameterization of the backbone torsion terms in ff99 and achieves much better balance of four basic secondary structure elements (PP, β, α, and γ). A detailed explanation of the parameterization, as well as an extensive comparison with many other variants of fixed-charge Amber force fields is given in the reference above. Briefly, dihedral terms parameters were obtained through fitting the energies of multiple conformations of glycine and alanine tetrapeptides to high-level ab initio QM calculations. We have shown that this force field provides much improved proportions of helical versus extended structures. In addition, it corrected the glycine sampling and should also perform well for β-turn structures, two things which were especially problematic with most previous Amber force field variants.

2. FF99SB and cboLJ3 for nucleic acids. The nucleic acid force fields have recently been updated from those in ff90 in order to address a tendency of DNA double helices to convert to fairly long simulations to extended forms in the rr and backbone torsion angles gauche+, trans and twisting of the sugar into the plane of the base[7]. Also, in order to improve the γ glycosidic torsional potential for RNA, recent work has proposed additional force field modifications[8] (see also an alternative set of γ modifications below). These force field modifications lead to better representations of both DNA and RNA, note that the γ modifications only apply to RNA.

3. Updated ion parameters. Recently, Iommy and Cheatham have created a more consistent set of parameters, ionic solvation free energies, radial distribution functions, ion-water interaction energies and crystal lattice energies and lattice constants for non-polarizable spherical ions[8, 10]. These have been separately parameterized for each of three popular water models, as indicated above. Please note: you need to load an additional ff99md file specific to the water model you are using; see Section 2.10.
these regions are $\Psi$ for $\alpha$-helices (right-handed), $\phi$ for $\beta$-sheets, $\gamma$ and $\gamma'$ for tight turns.

AMBER FF - proteins

- parm99 – error for Gly, sampling of disallowed region in RP
  - corrected parm99SB
- parm03 (ff03) – seems to overstabilize $\alpha$-helices

FF99, 99SB a 03

- chignolin
D(R)NA backbone

<table>
<thead>
<tr>
<th>Structure Type</th>
<th>Alpha</th>
<th>Beta</th>
<th>Gamma</th>
<th>Delta</th>
<th>Epsilon</th>
<th>Zeta</th>
<th>Chi</th>
</tr>
</thead>
<tbody>
<tr>
<td>A- DNA (fibres)</td>
<td>152</td>
<td>106</td>
<td>44</td>
<td>276</td>
<td>-78</td>
<td>157</td>
<td>-146</td>
</tr>
<tr>
<td>GGCCGGCC</td>
<td>185</td>
<td>56</td>
<td>91</td>
<td>-166</td>
<td>-75</td>
<td>-149</td>
<td></td>
</tr>
<tr>
<td>B- DNA (fibres)</td>
<td>136</td>
<td>38</td>
<td>139</td>
<td>-133</td>
<td>-157</td>
<td>-102</td>
<td></td>
</tr>
<tr>
<td>CGCGAATTCGCG</td>
<td>171</td>
<td>54</td>
<td>91</td>
<td>-169</td>
<td>-108</td>
<td>-117</td>
<td></td>
</tr>
<tr>
<td>Z- DNA (C residues)</td>
<td>137</td>
<td>56</td>
<td>82</td>
<td>-95</td>
<td>80</td>
<td>159</td>
<td></td>
</tr>
<tr>
<td>Z- DNA (G residues)</td>
<td>137</td>
<td>56</td>
<td>123</td>
<td>-104</td>
<td>-69</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>DNA - RNA decamer</td>
<td>175</td>
<td>55</td>
<td>82</td>
<td>-151</td>
<td>-75</td>
<td>-162</td>
<td></td>
</tr>
<tr>
<td>A- RNA</td>
<td>178</td>
<td>54</td>
<td>82</td>
<td>-153</td>
<td>-158</td>
<td>-158</td>
<td></td>
</tr>
</tbody>
</table>


AMBER FF - NA

- parm99 – α/γ flips (pathological irreversible γ-trans), slow degradation of B-DNA
- parmbcs0 – repar. α/γ
- parm99+parmbcs0 – resonably good description of RNA, but „ladder-like“ structures in A-RNA duplexes

AMBER FF - RNA

- parm99/parmbcs0 – „ladder-like“ structures
Parameterization

- Bonds, Angles – $r_0$, $k$ – from QM calculations / by analogy
- Dihedrals – significantly affects conformational behavior of molecules; before any productive application parameters should be thoroughly tested

Parameterization of Dihedrals

- Calculate QM profile
  - Enough reliable QM method (MP2/cc-pVTZ)
  - Which environment? Gas phase or solvent?
    - 1,2-dichloroethane (in gas phase – trans; in water – gauche)
- Calculate MM profile
  - without parameters of the parameterized torsion
  - Developed parameters will inherently include used partial charges and LJ parameters – straightforward transferability between FF is impossible
Parameterization of Dihedrals

- Traditional way
  - Derive "pure dihedral profile"
  - $E_{dih} = QM - MM_{QMgeom}$
  - Fit the profile by Fourier series

- Alternative way 
  - QM in water (CPCM), MM geometries are relaxed (PB), full FF
  - SP without parameterized torsion (PB)

MM Applications

- Minimization of molecules
- Conformational search
- Classical MD
- NMR, X-ray refinement
- Scoring functions (drug design)
- Normal mode analysis (entropy estimation)

Exploring PES

PES of larger molecules are complex

\[ E = f(R) \]

\[ \dim E \leq 3N - 6 + 1 \]
PES in 3D

Important Points on PES

\[ \frac{\partial E}{\partial z_i} = 0, \quad \text{stationary points} \]

\[ H = \begin{pmatrix}
\frac{\partial^2 E}{\partial z_1^2} & \frac{\partial^2 E}{\partial z_1 \partial z_2} & \cdots & \frac{\partial^2 E}{\partial z_1 \partial z_n} \\
\frac{\partial^2 E}{\partial z_2 \partial z_1} & \frac{\partial^2 E}{\partial z_2^2} & \cdots & \frac{\partial^2 E}{\partial z_2 \partial z_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 E}{\partial z_n \partial z_1} & \frac{\partial^2 E}{\partial z_n \partial z_2} & \cdots & \frac{\partial^2 E}{\partial z_n^2}
\end{pmatrix} \quad \text{Hess matrix} \]

\[ \begin{pmatrix}
\lambda_1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \lambda_n
\end{pmatrix} \quad \text{Eigenvalues of Hess matrix} \]
Reaction coordinate – PES dissection

PES vs thermodynamics and kinetics

Chemistry is a hiking on PES
Catalysis

Kinetically driven reaction

without catalyst

with catalyst

intermediate

reaction coordinate

reaction coordinate

Photochemistry

conical intersection

transition state

reactant

product

reaction coordinate

Applicability of MM

• "common" FF does not allow bond dissociation and formation
• Electronic states cannot be considered explicitly
• Exploring of conformational space
• But – combination with QM – QM/MM methods
Exploring PES

- Straightforward minimization
  - Simplex, Newton method ... - closest local minimum is often localized
- Grid based methods (e.g. Torsional driving)
  - not effective time consuming
- Monte Carlo
- Simulated annealing
- Genetic algorithms

Failure of MM

- Not able to describe halogen bond

\[ D \cdot X \ldots Y \]
\[ X - \text{halogen} \]
\[ \text{Cl} < \text{Br} < \text{I} \]

\[ \begin{align*}
  & \text{A} \\
  & 8.6 \text{ kJ/mol} \\
  & \text{B} \\
  & 10.1 \text{ kJ/mol}
\end{align*} \]

MM Failure

- The PM6-D2X optimized structure (red lines) of the inhibitor bound to the CK2a active site agrees well with the 2OXY X-ray structure (blue lines), B. whereas the AMBER optimized structure (red lines) shows substantial differences from the X-ray geometry (blue lines). The failure of the AMBER empirical potential is caused by its inability to describe halogen bonds (shown by black dotted lines).

• machinery of statistical thermodynamics
  
  \[ A = \frac{\sum A_i e^{-\beta \mu_i}}{Q} \]
  
  \[ S = \frac{U}{T} + k \ln Q \]
  
  \[ Q = Q_{\text{Hess}} Q_{\text{rot}} Q_{\text{vib}} Q_{\text{rel}} \]  
  
  QM or MM calculation from Hess matrix, moment of inertia from mass

\[ Q = nRT \ln \frac{q_m}{N} \]
\[ q_m = q_t, q_r, q_v, q_e \]
\[ q_t = \frac{V}{A^3}, \quad A = \frac{h}{\sqrt{2\pi mk_B T}} \]
\[ q_r = \sum (2J + 1) e^{-\beta \epsilon_{J(J+1)}} \]
\[ \theta_i = \frac{hcB}{k_B} \]
\[ q_v = \frac{e^{-\beta \epsilon_{J(J+1)}}}{1 - e^{-2\beta \epsilon_{J(J+1)}}} \]