Force Fields for MD simulations

• Topology/parameter files
• Where do the numbers an MD code uses come from?
• How to make topology files for ligands, cofactors, special amino acids, …
• How to obtain/develop missing parameters.
Classical Molecular Dynamics

\[ U(r) = \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}} \]

Coulomb interaction

\[ U(r) = \varepsilon_{ij} \left[ \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^{6} \right] \]

van der Waals interaction
Classical Molecular Dynamics

Bond definitions, atom types, atom names, parameters, ....
Energy Terms Described in the CHARMM Force Field

- Bond
- Angle
- Dihedral
- Improper
The Potential Energy Function

\[ U(\vec{R}) = \sum_{bonds} k_{i}^{bond} (r_{i} - r_{0})^2 + \sum_{angles} k_{i}^{angle} (\theta_{i} - \theta_{0})^2 + \]

\[ \sum_{dihedrals} k_{i}^{dihedral} [1 + \cos(n_{i}\phi_{i} + \delta_{i})] + \]

\[ \sum_{i \neq j} 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \sum_{i \neq j} \frac{q_{i}q_{j}}{\varepsilon r_{ij}} \]

\[ U_{bond} = \text{oscillations about the equilibrium bond length} \]
\[ U_{angle} = \text{oscillations of 3 atoms about an equilibrium bond angle} \]
\[ U_{dihedral} = \text{torsional rotation of 4 atoms about a central bond} \]
\[ U_{nonbond} = \text{non-bonded energy terms (electrostatics and Lenard-Jones)} \]
Interactions between bonded atoms

\[ V_{\text{angle}} = K_{\theta} (\theta - \theta_o)^2 \]

\[ V_{\text{bond}} = K_b (b - b_o)^2 \]

\[ V_{\text{dihedral}} = K_{\phi} (1 + \cos(n\phi - \delta)) \]
\[ V_{bond} = K_b (b - b_o)^2 \]

<table>
<thead>
<tr>
<th>Chemical type</th>
<th>( K_{bond} )</th>
<th>( b_o )</th>
</tr>
</thead>
<tbody>
<tr>
<td>C(-)C</td>
<td>100 kcal/mole/Å²</td>
<td>1.5 Å</td>
</tr>
<tr>
<td>C(=)C</td>
<td>200 kcal/mole/Å²</td>
<td>1.3 Å</td>
</tr>
<tr>
<td>C(=)C</td>
<td>400 kcal/mole/Å²</td>
<td>1.2 Å</td>
</tr>
</tbody>
</table>

**Bond angles** and **improper terms** have similar quadratic forms, but with softer spring constants. The force constants can be obtained from vibrational analysis of the molecule (experimentally or theoretically).
Dihedral Potential

\[ V_{dihedral} = K_\phi (1 + \cos(n\phi - \delta)) \]

\[ \delta = 0^\circ \]
Nonbonded Parameters

\[
\sum_{\text{nonbonded}} \frac{q_i q_j}{4\pi D r_{ij}} + \varepsilon_{ij} \left[ \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^{6} \right]
\]

q_i: partial atomic charge  
D: dielectric constant  
\(\varepsilon\): Lennard-Jones (LJ, vdW) well-depth  
R_{\text{min}}: LJ radius (R_{\text{min}}/2 in CHARM MM)  
Combining rules (CHARMM, Amber)  
\[R_{\text{min},i,j} = R_{\text{min},i} + R_{\text{min},j}\]  
\[\varepsilon_{i,j} = \text{SQRT}(\varepsilon_i \ast \varepsilon_j)\]
Electrostatic Energy versus Distance

Note that the effect is long range.

From MacKerell
Charge Fitting Strategy

CHARMM- Mulliken*  AMBER(ESP/RESP)

Partial atomic charges

\[
\begin{array}{cc}
0.5 \text{C} & -0.5 \\
0.35 \text{H} & -0.45 \text{N}
\end{array}
\]

*Modifications based on interactions with TIP3 water
CHARMM Potential Function

\[ U(\vec{R}) = \sum_{\text{bonds}} k_i^{\text{bond}} (r_i - r_0)^2 + \sum_{\text{angles}} k_i^{\text{angle}} (\theta_i - \theta_0)^2 + \sum_{\text{dihedrals}} k_i^{\text{dihedral}} \left[ 1 + \cos \left( n_i \phi_i + \delta_i \right) \right] + \sum_{i \neq j} 4\epsilon_{ij} \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 + \sum_{i \neq j} \frac{q_i q_j}{\epsilon r_{ij}} \]

- **PDB file**: geometry
- **PSF file**: Topology
- **Parameter file**: parameters
File Format/Structure

• The structure of a pdb file
• The structure of a psf file
• The topology file
• The parameter file
• Connection to potential energy terms
### Structure of a PDB file

<table>
<thead>
<tr>
<th>index</th>
<th>name</th>
<th>resname</th>
<th>chain</th>
<th>resid</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>segname</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATOM</td>
<td>22</td>
<td>N</td>
<td>ALA</td>
<td>B</td>
<td>-4.073</td>
<td>-7.587</td>
<td>-2.708</td>
<td>BH</td>
</tr>
<tr>
<td>ATOM</td>
<td>23</td>
<td>HN</td>
<td>ALA</td>
<td>B</td>
<td>-3.813</td>
<td>-6.675</td>
<td>-3.125</td>
<td>BH</td>
</tr>
<tr>
<td>ATOM</td>
<td>24</td>
<td>CA</td>
<td>ALA</td>
<td>B</td>
<td>-4.615</td>
<td>-7.557</td>
<td>-1.309</td>
<td>BH</td>
</tr>
<tr>
<td>ATOM</td>
<td>25</td>
<td>HA</td>
<td>ALA</td>
<td>B</td>
<td>-4.323</td>
<td>-8.453</td>
<td>-0.704</td>
<td>BH</td>
</tr>
<tr>
<td>ATOM</td>
<td>26</td>
<td>CB</td>
<td>ALA</td>
<td>B</td>
<td>-4.137</td>
<td>-6.277</td>
<td>-0.676</td>
<td>BH</td>
</tr>
<tr>
<td>ATOM</td>
<td>27</td>
<td>HB1</td>
<td>ALA</td>
<td>B</td>
<td>-3.128</td>
<td>-5.950</td>
<td>-0.907</td>
<td>BH</td>
</tr>
<tr>
<td>ATOM</td>
<td>28</td>
<td>HB2</td>
<td>ALA</td>
<td>B</td>
<td>-4.724</td>
<td>-5.439</td>
<td>-1.015</td>
<td>BH</td>
</tr>
<tr>
<td>ATOM</td>
<td>29</td>
<td>HB3</td>
<td>ALA</td>
<td>B</td>
<td>-4.360</td>
<td>-6.338</td>
<td>0.393</td>
<td>BH</td>
</tr>
<tr>
<td>ATOM</td>
<td>30</td>
<td>C</td>
<td>ALA</td>
<td>B</td>
<td>-6.187</td>
<td>-7.538</td>
<td>-1.357</td>
<td>BH</td>
</tr>
<tr>
<td>ATOM</td>
<td>31</td>
<td>O</td>
<td>ALA</td>
<td>B</td>
<td>-6.854</td>
<td>-6.553</td>
<td>-1.264</td>
<td>BH</td>
</tr>
<tr>
<td>ATOM</td>
<td>32</td>
<td>N</td>
<td>ALA</td>
<td>B</td>
<td>-6.697</td>
<td>-8.715</td>
<td>-1.643</td>
<td>BH</td>
</tr>
<tr>
<td>ATOM</td>
<td>33</td>
<td>HN</td>
<td>ALA</td>
<td>B</td>
<td>-6.023</td>
<td>-9.463</td>
<td>-1.751</td>
<td>BH</td>
</tr>
<tr>
<td>ATOM</td>
<td>34</td>
<td>CA</td>
<td>ALA</td>
<td>B</td>
<td>-8.105</td>
<td>-9.096</td>
<td>-1.934</td>
<td>BH</td>
</tr>
<tr>
<td>ATOM</td>
<td>35</td>
<td>HA</td>
<td>ALA</td>
<td>B</td>
<td>-8.287</td>
<td>-8.878</td>
<td>-3.003</td>
<td>BH</td>
</tr>
<tr>
<td>ATOM</td>
<td>36</td>
<td>CB</td>
<td>ALA</td>
<td>B</td>
<td>-8.214</td>
<td>-10.604</td>
<td>-1.704</td>
<td>BH</td>
</tr>
<tr>
<td>ATOM</td>
<td>37</td>
<td>HB1</td>
<td>ALA</td>
<td>B</td>
<td>-7.493</td>
<td>-11.205</td>
<td>-2.379</td>
<td>BH</td>
</tr>
<tr>
<td>ATOM</td>
<td>38</td>
<td>HB2</td>
<td>ALA</td>
<td>B</td>
<td>-8.016</td>
<td>-10.861</td>
<td>-0.665</td>
<td>BH</td>
</tr>
<tr>
<td>ATOM</td>
<td>39</td>
<td>HB3</td>
<td>ALA</td>
<td>B</td>
<td>-9.245</td>
<td>-10.914</td>
<td>-1.986</td>
<td>BH</td>
</tr>
<tr>
<td>ATOM</td>
<td>40</td>
<td>C</td>
<td>ALA</td>
<td>B</td>
<td>-9.226</td>
<td>-8.438</td>
<td>-1.091</td>
<td>BH</td>
</tr>
<tr>
<td>ATOM</td>
<td>41</td>
<td>O</td>
<td>ALA</td>
<td>B</td>
<td>-10.207</td>
<td>-7.958</td>
<td>-1.667</td>
<td>BH</td>
</tr>
</tbody>
</table>

>>> It is an ascii, fixed-format file <<<  

“No connectivity information”
Checking file structures

- PDB file
- Topology file
- PSF file
- Parameter file
Parameter Optimization Strategies

Check if it has been parameterized by somebody else

- Literature
- Google

**Minimal optimization**
- By analogy (i.e. direct transfer of known parameters)
- Quick, starting point

**Maximal optimization**
- Time-consuming
- Requires appropriate experimental and target data

**Choice based on goal of the calculations**
- Minimal
  - database screening
  - NMR/X-ray structure determination
- Maximal
  - free energy calculations, mechanistic studies, subtle environmental effects
Getting Started

- Identify previously parameterized compounds
- Access topology information – assign atom types, connectivity, and charges – annotate changes

CHARMM topology (parameter files)

top_all22_model.inp (par_all22_prot.inp)
top_all22_prot.inp (par_all22_prot.inp)
top_all22_sugar.inp (par_all22_sugar.inp)
top_all27_lipid.rtf (par_all27_lipid.prm)
top_all27_na.rtf (par_all27_na.prm)
top_all27_na_lipid.rtf (par_all27_na_lipid.prm)
top_all27_prot_lipid.rtf (par_all27_prot_lipid.prm)
top_all27_prot_na.rtf (par_all27_prot_na.prm)
toph19.inp (param19.inp)

NA and lipid force fields have new LJ parameters for the alkanes, representing increased optimization of the protein alkane parameters. Tests have shown that these are compatible (e.g. in protein-nucleic acid simulations). For new systems it is suggested that the new LJ parameters be used. Note that only the LJ parameters were changed; the internal parameters are identical.

www.pharmacy.umaryland.edu/faculty/amackere/force_fields.htm
When creating a covalent link between model compounds move the charge on the deleted H into the carbon to maintain integer charge (i.e. methyl ($q_C=-0.27$, $q_H=0.09$) to methylene ($q_C=-0.18$, $q_H=0.09$))
From top_all22_model.inp

RESI PHEN 0.00 ! phenol, adm jr.

GROUP
ATOM CG CA -0.115 !
ATOM HG HP 0.115 ! HD1 HE1

GROUP
ATOM CD1 CA -0.115 ! CD1--CE1
ATOM HD1 HP 0.115 ! // \ 

GROUP
ATOM CD2 CA -0.115 ! CD2==CE2 HH
ATOM HD2 HP 0.115 ! // \ 

GROUP
ATOM CE1 CA -0.115 ! HD2 HE2
ATOM HE1 HP 0.115

GROUP
ATOM CE2 CA -0.115
ATOM HE2 HP 0.115

GROUP
ATOM CZ CA 0.110
ATOM OH OH1 -0.540
ATOM HH H 0.430

BOND CD2 CG CE1 CD1 CZ CE2 CG HG CD1 HD1
BOND CD2 HD2 CE1 HE1 CE2 HE2 CZ OH OH HH
DOUBLE CD1 CG CE2 CD2 CZ CE1

HG will ultimately be deleted. Therefore, move HG (hydrogen) charge into CG, such that the CG charge becomes 0.00 in the final compound.

Use remaining charges/atom types without any changes.

Do the same with indole

Top_all22_model.inp contains all protein model compounds. Lipid, nucleic acid and carbohydrate model compounds are in the full topology files.

From MacKerell
Creation of topology for central model compound

Start with alanine dipeptide. Note use of new aliphatic LJ parameters and, importantly, atom types.

NR1 from histidine unprotonated ring nitrogen. Charge (very bad) initially set to yield unit charge for the group.

Note use of large group to allow flexibility in charge optimization.

From MacKerell
Partial Atomic Charge Determination
Method Dependent Choices

1. RESP: HF/6-31G overestimates dipole moments (AMBER)

2. Interaction based optimization (CHARMM)

For a particular force field do NOT change the QM level of theory. This is necessary to maintain consistency with the remainder of the force field.

From MacKerell
Starting charges??
Mulliken population analysis
Analogy comparison

Final charges (methyl, vary $q_C$ to maintain integer charge, $q_H = 0.09$)
interactions with water (HF/6-31G*, monohydrates!)

From MacKerell
Comparison of analogy and optimized charges

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Analogy</th>
<th>Optimized</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>CT3</td>
<td>-0.27</td>
<td>-0.27</td>
</tr>
<tr>
<td>H11</td>
<td>HA3</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>H12</td>
<td>HA3</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>H13</td>
<td>HA3</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>C2</td>
<td>C</td>
<td>0.51</td>
<td>0.58</td>
</tr>
<tr>
<td>O2</td>
<td>O</td>
<td>-0.51</td>
<td>-0.50</td>
</tr>
<tr>
<td>N3</td>
<td>NH1</td>
<td>-0.47</td>
<td>-0.32</td>
</tr>
<tr>
<td>H3</td>
<td>H</td>
<td>0.31</td>
<td>0.33</td>
</tr>
<tr>
<td>N4</td>
<td>NR1</td>
<td>0.16</td>
<td>-0.31</td>
</tr>
<tr>
<td>C5</td>
<td>CEL1</td>
<td>-0.15</td>
<td>-0.25</td>
</tr>
<tr>
<td>H51</td>
<td>HEL1</td>
<td>0.15</td>
<td>0.29</td>
</tr>
<tr>
<td>C6</td>
<td>CT3</td>
<td>-0.27</td>
<td>-0.09</td>
</tr>
<tr>
<td>H61</td>
<td>HA</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>H62</td>
<td>HA</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>H63</td>
<td>HA</td>
<td>0.09</td>
<td>0.09</td>
</tr>
</tbody>
</table>

![Chemical structure](image)
Dihedral optimization based on QM potential energy surfaces (HF/6-31G* or MP2/6-31G*).

From MacKerell
Retinal Proteins -- Rhodopsins

- Covalently linked to a lysine
- Usually protonated Schiff base
- all-trans and 11-cis isomers
Unconventional chemistry
Isomerization Barriers in retinal

\[
\text{DFT/6-31G}^{**}
\]

**TABLE 2** The parameter set B used for the torsional potentials of the main polyene chain of the retinal Schiff base

<table>
<thead>
<tr>
<th>Bond</th>
<th>( k_i ) (kcal/mol)**</th>
<th>( n_i )</th>
<th>( \delta_i ) (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C_5 = C_6 - C_7 = C_8 )</td>
<td>11.24</td>
<td>2.0</td>
<td>180.00</td>
</tr>
<tr>
<td>( C_5 = C_6 - C_7 = C_9 )</td>
<td>39.98</td>
<td>2.0</td>
<td>180.00</td>
</tr>
<tr>
<td>( C_7 = C_8 = C_9 = C_{10} )</td>
<td>17.03</td>
<td>2.0</td>
<td>180.00</td>
</tr>
<tr>
<td>( C_8 = C_9 = C_{10} = C_{11} )</td>
<td>37.28</td>
<td>2.0</td>
<td>180.00</td>
</tr>
<tr>
<td>( C_9 = C_{10} = C_{11} = C_{12} )</td>
<td>22.50</td>
<td>2.0</td>
<td>180.00</td>
</tr>
<tr>
<td>( C_{10} = C_{11} = C_{12} = C_{13} )</td>
<td>35.08</td>
<td>2.0</td>
<td>180.00</td>
</tr>
<tr>
<td>( C_{11} = C_{12} = C_{13} = C_{14} )</td>
<td>28.30</td>
<td>2.0</td>
<td>180.00</td>
</tr>
<tr>
<td>( C_{12} = C_{13} = C_{14} = C_{15} )</td>
<td>29.46</td>
<td>2.0</td>
<td>180.00</td>
</tr>
<tr>
<td>( C_{13} = C_{14} = C_{15} = N_{16} )</td>
<td>30.43</td>
<td>2.0</td>
<td>180.00</td>
</tr>
<tr>
<td>( C_{14} = C_{15} = N_{16} = C_\circ )</td>
<td>28.76</td>
<td>2.0</td>
<td>180.00</td>
</tr>
</tbody>
</table>

Tajkarshid et al., 1999.

\[ E_{\text{dihedral}}^{i} = (1/2)k_i[1 + \cos(n_i \varphi_i - \delta_i)]. \]
Coupling of electronic excitation and conformational change in bR
Inducing isomerization

500 nm

\sim 50 \text{ kcal/mole}
Classical Retinal Isomerization in Rhodopsin

Twist Propagation
\[ \hat{H} = \sum_i \frac{1}{2} p_i^2 + \sum_i \sum_A \frac{Z_A}{r_{iA}} + \sum_{i>j} \frac{1}{r_{ij}} + \sum_{A>B} \frac{Z_A Z_B}{r_{AB}} \]

\[ + \sum_i \sum_p \frac{q_p}{r_{ip}} + \sum_A \sum_p \frac{Z_A q_p}{r_{Ap}} \]

\[ + V_{QM-MM}^{MM} + V_{MM-MM}^{QM} \]
Quantum mechanical (QM) treatment of the chromophore, and force field (MM) treatment of the embedding protein.
QM/MM calculation of ATP hydrolysis
Coarse grain modeling of lipids

150 particles

9 particles!