Force Fields for Classical Molecular Dynamics simulations of Biomolecules

Emad Tajkhorshid

Theoretical and Computational Biophysics Group, Beckman Institute
Departments of Biochemistry and Pharmacology, College of Medicine
Center for Biophysics and Computational Biology
University of Illinois at Urbana-Champaign
Classical Force Field Parameters

- Topology and structure files
- Parameter files
- Where do all the numbers needed by an MD code come from?
- Where to find these numbers and how to change them if needed.
- How to make topology files for ligands, cofactors, special amino acids, …
- How to develop / put together missing parameters.
Classical Molecular Dynamics

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

Coulomb interaction

\[ U(r) = \epsilon_{ij} \left[ \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^{12} - \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^6 \right] \]
Classical Molecular Dynamics

Bond definitions, atom types, atom names, parameters, ....
Energy Terms Described in

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

\[ U(\vec{R}) = \sum_{\text{bonds}} k_i^{\text{bond}} (r_i - r_0)^2 + \sum_{\text{angles}} k_i^{\text{angle}} (\theta_i - \theta_0)^2 + \]

\[ \sum_{\text{dihedrals}} k_i^{\text{dihedral}} [1 + \cos (n_i \phi_i + \delta_i)] + \]

\[ \sum_{i} \sum_{j \neq i} 4 \epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i} \sum_{j \neq i} \frac{q_i q_j}{r_{ij}} \]

\[ U_{\text{bond}} = \text{oscillations about the equilibrium bond length} \]

\[ U_{\text{angle}} = \text{oscillations of 3 atoms about an equilibrium bond angle} \]

\[ U_{\text{dihedral}} = \text{torsional rotation of 4 atoms about a central bond} \]

\[ U_{\text{nonbond}} = \text{non-bonded energy terms (electrostatics and Lenard-Jones)} \]
Interactions between bonded atoms

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

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

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

<table>
<thead>
<tr>
<th>Chemical type</th>
<th>( K_{\text{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 Energy versus Bond length**

*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**

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

- \( K=10, n=1 \)
- \( K=5, n=2 \)
- \( K=2.5, N=3 \)

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

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

- \( q_i \): partial atomic charge
- \( D \): dielectric constant
- \( \epsilon \): Lennard-Jones (LJ, vdW) well-depth
- \( R_{\text{min}} \): LJ radius (\( R_{\text{min}} / 2 \) in CHARMM)

Combining rules (CHARMM, Amber)

\[ R_{\text{min},ij} = R_{\text{min},i} + R_{\text{min},j} \]

\[ \epsilon_{i,j} = \sqrt{\epsilon_i \cdot \epsilon_j} \]
Note that the effect is long range.

From MacKerell
Charge Fitting Strategy

CHARMM- Mulliken*          AMBER(ESP/RESP)

Partial atomic charges

\[
\begin{array}{c}
\text{0.5} \quad \text{C} = \text{O} \\
\text{-0.5} \\
\text{0.35} \\
\text{H} \quad \text{N} \\
\text{-0.45}
\end{array}
\]

*Modifications based on interactions with TIP3 water
CHARMM Potential Function

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

- **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"
Looking at 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 (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 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
Break Desired Compound into 3 Smaller Ones

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 MacKerell
From top_all22_model.inp

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

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.
Creation of topology for central model compound

<table>
<thead>
<tr>
<th>ATOM</th>
<th>Type</th>
<th>Charge</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>CT3</td>
<td>-0.27</td>
</tr>
<tr>
<td>H11</td>
<td>HA3</td>
<td>0.09</td>
</tr>
<tr>
<td>H12</td>
<td>HA3</td>
<td>0.09</td>
</tr>
<tr>
<td>H13</td>
<td>HA3</td>
<td>0.09</td>
</tr>
<tr>
<td>C2</td>
<td>C</td>
<td>0.51</td>
</tr>
<tr>
<td>O2</td>
<td>O</td>
<td>-0.51</td>
</tr>
<tr>
<td>N3</td>
<td>NH1</td>
<td>-0.47</td>
</tr>
<tr>
<td>H3</td>
<td>H</td>
<td>0.31</td>
</tr>
<tr>
<td>N4</td>
<td>NR1</td>
<td>0.16</td>
</tr>
<tr>
<td>C5</td>
<td>CEL1</td>
<td>-0.15</td>
</tr>
<tr>
<td>H51</td>
<td>HEL1</td>
<td>0.15</td>
</tr>
<tr>
<td>C6</td>
<td>CT3</td>
<td>-0.27</td>
</tr>
<tr>
<td>H61</td>
<td>HA</td>
<td>0.09</td>
</tr>
<tr>
<td>H62</td>
<td>HA</td>
<td>0.09</td>
</tr>
<tr>
<td>H63</td>
<td>HA</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Bonds:
- C1-H11-C1-H12-C1-H13-C1-C2-C2-O2-C2-N3-N3-H3
- N3-N4-C5-H51-C5-C6-C6-H61-C6-H62-C6-H63
- N4-C5 (DOUBL{E} only required for MMFF)

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 Charge Assignment

- Most important aspect for ligands

- Different force fields might take different philosophies
  - AMBER: RESP charges at the HF/6-31G level
    - Overestimation of dipole moments
    - Easier to set up
  - CHARMM: Interaction based optimization
    - TIP3P water representing the environment
    - Could be very difficult to set up

- Conformation dependence of partial charges
- Lack of polarization

- Try to be consistent within the force field

- pKa calculations for titratable residues
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>
Dihedral optimization based on QM potential energy surfaces (HF/6-31G* or MP2/6-31G*).

From MacKerell
Parameterization of unsaturated lipids

- All C=C bonds are cis, what does rotation about neighboring single bonds look like?

![Chemical structures and graphs illustrating the rotation of unsaturated lipids.](image-url)
DHA conformations from MD

- rotational barriers are extremely small
- many conformers are accessible w/ short lifetimes

Courtesy of Scott Feller, Wabash College
Dynamics of saturated vs. polyunsaturated lipid chains

- $sn_1$ stearic acid = blue
- $sn_2$ DHA = yellow
- 500 ps of dynamics

Movie courtesy of Mauricio Carrillo Tripp

Courtesy of Scott Feller, Wabash College
Lipid-protein interactions

- Radial distribution around protein shows distinct layering of acyl chains

Courtesy of Scott Feller, Wabash College
Lipid-protein interactions

- Decomposition of non-bonded interaction shows rhodopsin is strongly attracted to unsaturated chain
- All hydrophobic residues are stabilized by DHA

<table>
<thead>
<tr>
<th>resname</th>
<th>$U_{DHA}$</th>
<th>$U_{stearic}$</th>
<th>ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHE</td>
<td>-44.9</td>
<td>-22.6</td>
<td>2.0</td>
</tr>
<tr>
<td>ILE</td>
<td>-30.0</td>
<td>-10.1</td>
<td>3.0</td>
</tr>
<tr>
<td>VAL</td>
<td>-24.0</td>
<td>-9.6</td>
<td>2.5</td>
</tr>
<tr>
<td>LEU</td>
<td>-23.1</td>
<td>-13.0</td>
<td>1.8</td>
</tr>
<tr>
<td>MET</td>
<td>-22.8</td>
<td>-9.7</td>
<td>2.4</td>
</tr>
<tr>
<td>TYR</td>
<td>-18.6</td>
<td>-10.4</td>
<td>1.8</td>
</tr>
<tr>
<td>ALA</td>
<td>-11.4</td>
<td>-3.0</td>
<td>3.8</td>
</tr>
<tr>
<td>TRP</td>
<td>-10.3</td>
<td>-2.4</td>
<td>4.2</td>
</tr>
</tbody>
</table>

Courtesy of Scott Feller, Wabash College
Origin of protein:DHA attraction

• Flexibility of the DHA chain allows solvation of the rough protein surface to occur with little intra-molecular energy cost

Courtesy of Scott Feller, Wabash College
Major Recent Developments

• New set of lipid force field parameters for CHARMM (CHARMM32+)
  – Pastor, B. Brooks, MacKerell

• Polarizable force field
  – Roux, 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

![Chemical structure of retinal]

**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></th>
<th>$\phi_i$</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>
<td></td>
</tr>
<tr>
<td>C_6-C_7=C_8-C_9</td>
<td>39.98</td>
<td>2.0</td>
<td>180.00</td>
<td></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>
<td></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>
<td></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>
<td></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>
<td></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>
<td></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>
<td></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>
<td></td>
</tr>
<tr>
<td>C_14=C_15=N_16=C_16</td>
<td>28.76</td>
<td>2.0</td>
<td>180.00</td>
<td></td>
</tr>
</tbody>
</table>

Tajkhorshid et al., 1999.

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

500 nm
~50 kcal/mole
Classical Retinal Isomerization

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

\[ + \sum_i \sum_p q_p \frac{1}{r_{ip}} + \sum_A \sum_p Z_A q_p \frac{1}{r_{Ap}} \]

\[ + V_{QM-MM}^{MM} + V_{MM-MM}^{MM} \]
Ab Initio QM/MM Excited State MD Simulation

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!