Overview of Classical Force Fields and Parameter Optimization Strategy

Part I - Overview of CHARMM FF and Parameter Optimization

Part II - Introduction to Quantum Chemistry Calculations (SPARTAN)
Application to parameterization of thioester

Part III - Knowledge-based and Hybrid FF
Application to protein structure prediction and folding studies

Perth 2004
Classical Force Fields

✓ Physics-based, full atom FF like CHARMM, AMBER, OPLS - Mechanisms, function, protein/RNA/DNA interactions…

• Knowledge-based, coarse-grained model -
  Protein structure prediction, folding kinetics, docking,…. 

• Hybrid force fields like Go + full atom FF -
  Mutational effects on protein folding kinetics,…. 

• Hybrid QM/MM approaches 
  Enzyme reactions, bond breaking and making, excited states,…

Each problem has a different goal and time scale!
General Considerations

• Description of molecules?
• Optimization of force field parameters?
• Training set of compounds/data?
• Test set of compounds/data?
• Limitations – questions you should not ask of your force field
Common Full Atom Force Fields

Class I
CHARMM, CHARMM (Accelyrs), AMBER, OPLS
ECEPP, GROMOS, SYBYL (Tripos)

Class II
MMFF94, UFF, …

Class III
QM/MM (CHARMM, AMBER, …)
Polarizable FF - Freisner/Berne (Schroedinger), AMOEBA (Tinker)

*Websites contain roadmaps of force field parameterization strategy. And they are different!!! So parameters from one cannot usually be used in another force field.
Overview and parameter optimization of CHARMM Force Field

Based on protocol established by

Alexander D. MacKerell, Jr, U. Maryland

See references: www.pharmacy.umaryland.edu/faculty/amackere/force_fields.htm

Class I Potential Energy function

\[ E_{\text{Total}} = \sum_{\text{bonds}} k_b (b - b_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 \]

\[ + \sum_{\text{dihedrals}} \frac{V}{2} [1 + \cos(n\phi - \delta)] \]

\[ + \sum_{\text{impropers}} k_\omega (\omega - \omega_0)^2 + \sum_{\text{Urey–Bradley}} k_u (r_{1,3} - r_{1,3,0})^2 \]

Non-bonded Interaction Terms

\[ + \sum_{\text{electrostatics}} \left( \frac{q_i q_j}{\varepsilon r_{ij}} \right) + \sum_{\text{VDW}} \varepsilon_{ij} \left[ \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^{6} \right] \]

From MacKerell
Class II force fields (e.g. MMFF) – Transferability, organic comb.

\[
\sum_{\text{bonds}} \left[ K_{b,2}(b - b_o)^2 + K_{b,3}(b - b_o)^3 + K_{b,4}(b - b_o)^4 \right] + \\
\sum_{\text{angles}} \left[ K_{\theta,2}(\theta - \theta_o)^2 + K_{\theta,3}(\theta - \theta_o)^3 + K_{\theta,4}(\theta - \theta_o)^4 \right] + \\
\sum_{\text{dihedrals}} \left[ K_{\phi,1}(1 - \cos \phi) + K_{\phi,2}(1 - \cos 2\phi) + K_{\phi,3}(1 - \cos 3\phi) \right] + \\
\sum K_{\chi} \chi^2 + \\
\sum_{\text{bonds bonds'}} \left[ K_{bb}(b - b_o)(b' - b_o') \right] + \sum_{\text{angles angles'}} \left[ K_{\theta\theta}(\theta - \theta_o)(\theta' - \theta_o') \right] + \\
\sum_{\text{bonds angles}} \left[ K_{b\theta}(b - b_o)(\theta - \theta_o) \right] + \\
\sum_{\text{bonds dihedrals}} \left[ (b - b_o) \left[ K_{\phi,b1} \cos \phi + K_{\phi,b2} \cos 2\phi + K_{\phi,b3} \cos 3\phi \right] \right] + \\
\sum_{\text{bonds' dihedrals}} \left[ (b' - b_o') \left[ K_{\phi,b'1} \cos \phi + K_{\phi,b'2} \cos 2\phi + K_{\phi,b'3} \cos 3\phi \right] \right] + \\
\sum_{\text{angles dihedrals}} \left[ (\theta - \theta_o) \left[ K_{\phi,\theta1} \cos \phi + K_{\phi,\theta2} \cos 2\phi + K_{\phi,\theta3} \cos 3\phi \right] \right] + \\
\sum_{\text{angles angles' dihedrals}} \left[ (\theta - \theta_o)(\theta' - \theta_o') \cos \phi \right]
\]

From MacKerell
Interactions between bonded atoms

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

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

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

Intermolecular 1,4 interactions: 1 or scaled > 1,4 interactions: 1

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

<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/Å(^2)</td>
<td>1.5 Å</td>
</tr>
<tr>
<td>C=C</td>
<td>200 kcal/mole/Å(^2)</td>
<td>1.3 Å</td>
</tr>
<tr>
<td>C=-C</td>
<td>400 kcal/mole/Å(^2)</td>
<td>1.2 Å</td>
</tr>
</tbody>
</table>

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

\[ \delta = 0^\circ \]

From MacKerell
From MacKerell

\[ V_{improper} = K_\varphi \left( \varphi - \varphi_o \right)^2 \]

\[ V_{Urey-Bradley} = K_{UB} \left( r_{1,3} - r_{1,3o} \right)^2 \]
Intermolecular parameters

\[ \sum_{\text{nonbonded}} \frac{q_i q_j}{4 \pi \varepsilon_0 D r_{ij}} + \varepsilon_{ij} \left[ \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^6 \right] \]

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

Combining rules (CHARMM, Amber)

- \( R_{\text{min},i,j} = R_{\text{min},i} + R_{\text{min},j} \)
- \( \varepsilon_{i,j} = \sqrt[\varepsilon_i \ast \varepsilon_j} \)

From MacKerell
Electrostatic Energy versus Distance

Interaction energy, kcal/mol

Distance, Å

From MacKerell
Charge Fitting Strategy

CHARMM- Mulliken* AMBER(ESP/RESP)

Partial atomic charges

*Modifications based on interactions with TIP3 water
Lennard-Jones Energy versus Distance

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

From MacKerell
Summary of Potential Terms
What about a new ligand not present in the CHARMM force field??????
- **Imidazole Glycerol Phosphate Synthase**: 5th step in Histidine Biosynthesis.
- **Branch point between Purine (nucleotide) and Histidine synthesis.**
HisH Mechanism

- HisH glutamine amidotransferase
- Conserved catalytic triad: CYS84, HIS178, GLU180

Thioester? Not in FF
Parameter Optimization Strategies

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

**Maximal optimization**  
Time-consuming  
Requires appropriate 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

**Manual or Automated Fitting Procedures**
Roadmap Charmm27 Optimization*

Initial Geometries Model Compounds?

Partial Atomic Charges

VDW Parameters

Bonds, Angles, Torsions, Improper

Condensed Phase MD Simulations

Parameterization Complete

Exp. Data
IR, X-ray, ...
Stat. Var.

Heat Vap, Rmin, ...

QM/MP2/6-31G*
Barriers, bonds, ...

HF/6-31G* hydrated
groups, TIP3W

Self-consistent
iteration

*based on MacKerell, JCC v21, 86,105 (2000)
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
Comparison of atom names (upper) and atom types (lower)

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.

CEL1/HEL1 from propene (lipid model compound). See top_all27_prot_lipid.rtf

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

From MacKerell
HG1 deleted from CYS and the charge was moved to SG (-0.23 +0.16=0.07) so that the SG charge becomes 0.07 in final compound and the group remains neutral

Changes annotated!
Partial Atomic Charge Determination

Additive Models: account for lack of explicit inclusion of polarizability via “overcharging” of atoms.

- RESP: HF/6-31G overestimates dipole moments (AMBER)
- Interaction based optimization (CHARMM, OPLS)
  - local polarization included
  - scale target interaction energies (CHARMM)
  - 1.16 for polar neutral compounds
  - 1.0 for charged compounds

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??
peptide bond
methyl
imidazole (N-N=C)?
Mulliken population analysis

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

From MacKerell
Model compound 1-water interaction energies/geometries

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Interaction Energies (kcal/mole)</th>
<th>Interaction Distances (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ab initio</td>
<td>Analogy</td>
</tr>
<tr>
<td>1) O2...HOH</td>
<td>-6.12</td>
<td>-6.56</td>
</tr>
<tr>
<td>2) N3-H..OHH</td>
<td>-7.27</td>
<td>-7.19</td>
</tr>
<tr>
<td>3) N4...HOH</td>
<td>-5.22</td>
<td>-1.16</td>
</tr>
<tr>
<td>4) C5-H..OHH</td>
<td>-3.86</td>
<td>-3.04</td>
</tr>
</tbody>
</table>

Energetic statistical analysis

<table>
<thead>
<tr>
<th></th>
<th>Ave. Difference</th>
<th>RMS Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction Energies</td>
<td>1.13</td>
<td>1.75</td>
</tr>
<tr>
<td>Distances</td>
<td>0.06</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Dipole Moments (debye)

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction Energies</td>
<td>5.69</td>
<td>4.89</td>
</tr>
<tr>
<td>Distances</td>
<td>6.00</td>
<td></td>
</tr>
</tbody>
</table>

*Ab initio* interaction energies scaled by 1.16.

MacKerell: Sanibel Conference 2003
Summary of Parameterization

1. **LJ (VDW) parameters** – normally direct transfer from available parameters is adequate, but should be tested by comparison to heats of vaporization, density, partial molar volumes, crystal simulations,... (MacKerell JCC 2002). Other solvents?

2. **Bond, angle, dihedral, UB and improper force constants**

   Vibrational spectra- Frequencies
   Conformational Energetics -
   Relative energies
   Potential energy surfaces

   Vibrations are generally used to optimize the bond, angle, UB and improper FCs while conformational energies are used for the dihedral FCs. However, vibrations will also be used for a number of the dihedral FCs, especially those involving hydrogens and in rings. (MacKerell 2003)
Vibrational Spectra of Model Compound 1 from MP2/6-31G* QM calculations

<table>
<thead>
<tr>
<th>#</th>
<th>Freq</th>
<th>Assign</th>
<th>%</th>
<th>Assign</th>
<th>%</th>
<th>#</th>
<th>Freq</th>
<th>Assign</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>62</td>
<td>tC2N</td>
<td>64</td>
<td>tN3N</td>
<td>46</td>
<td>21</td>
<td>1446</td>
<td>rNH</td>
<td>35</td>
</tr>
<tr>
<td>2</td>
<td>133</td>
<td>tC1H3</td>
<td>50</td>
<td>tN3N</td>
<td>18</td>
<td>22</td>
<td>1447</td>
<td>rC5H</td>
<td>47</td>
</tr>
<tr>
<td>3</td>
<td>148</td>
<td>tC1H3</td>
<td>46</td>
<td>tC6H3</td>
<td>25</td>
<td>23</td>
<td>1527</td>
<td>dCH3</td>
<td>77</td>
</tr>
<tr>
<td>4</td>
<td>154</td>
<td>dC2NN</td>
<td>44</td>
<td>dN3NC</td>
<td>28</td>
<td>24</td>
<td>1532</td>
<td>dCH3</td>
<td>88</td>
</tr>
<tr>
<td>5</td>
<td>205</td>
<td>tC6H3</td>
<td>59</td>
<td>tN4C</td>
<td>22</td>
<td>25</td>
<td>1599</td>
<td>dCH3a'</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>333</td>
<td>tN4C</td>
<td>73</td>
<td>tC2N</td>
<td>22</td>
<td>26</td>
<td>1610</td>
<td>dCH3a</td>
<td>71</td>
</tr>
<tr>
<td>7</td>
<td>361</td>
<td>dC1CN</td>
<td>45</td>
<td>dN4CC</td>
<td>21</td>
<td>27</td>
<td>1612</td>
<td>dCH3a'</td>
<td>30</td>
</tr>
<tr>
<td>8</td>
<td>446</td>
<td>rC=O</td>
<td>32</td>
<td>dN4CC</td>
<td>20</td>
<td>28</td>
<td>1613</td>
<td>dCH3a</td>
<td>70</td>
</tr>
<tr>
<td>9</td>
<td>568</td>
<td>wNH</td>
<td>77</td>
<td></td>
<td></td>
<td>29</td>
<td>1622</td>
<td>dCH3a'</td>
<td>57</td>
</tr>
<tr>
<td>10</td>
<td>586</td>
<td>dC1CN</td>
<td>21</td>
<td>dC2NN</td>
<td>20</td>
<td>30</td>
<td>1782</td>
<td>sN=C</td>
<td>71</td>
</tr>
<tr>
<td>11</td>
<td>618</td>
<td>wC=O</td>
<td>83</td>
<td>wNH</td>
<td>28</td>
<td>31</td>
<td>1901</td>
<td>sC=O</td>
<td>78</td>
</tr>
<tr>
<td>12</td>
<td>649</td>
<td>rC=O</td>
<td>27</td>
<td>dN4CC</td>
<td>19</td>
<td>32</td>
<td>3250</td>
<td>sCH3</td>
<td>76</td>
</tr>
<tr>
<td>13</td>
<td>922</td>
<td>sC1-C</td>
<td>62</td>
<td></td>
<td></td>
<td>33</td>
<td>3258</td>
<td>sC5-H</td>
<td>78</td>
</tr>
<tr>
<td>14</td>
<td>940</td>
<td>wC5H</td>
<td>80</td>
<td></td>
<td></td>
<td>34</td>
<td>3280</td>
<td>sCH3</td>
<td>99</td>
</tr>
<tr>
<td>15</td>
<td>1031</td>
<td>rCH3'</td>
<td>33</td>
<td>sC5-C</td>
<td>31</td>
<td>35</td>
<td>3330</td>
<td>sCH3a</td>
<td>75</td>
</tr>
<tr>
<td>16</td>
<td>1114</td>
<td>rCH3</td>
<td>66</td>
<td></td>
<td></td>
<td>36</td>
<td>3372</td>
<td>sCH3a'</td>
<td>100</td>
</tr>
<tr>
<td>17</td>
<td>1139</td>
<td>rCH3'</td>
<td>76</td>
<td>wC=O</td>
<td>20</td>
<td>37</td>
<td>3377</td>
<td>sCH3a'</td>
<td>73</td>
</tr>
<tr>
<td>18</td>
<td>1157</td>
<td>rCH3</td>
<td>61</td>
<td>wC5H</td>
<td>21</td>
<td>38</td>
<td>3403</td>
<td>sCH3a</td>
<td>99</td>
</tr>
<tr>
<td>19</td>
<td>1234</td>
<td>sC5-C</td>
<td>33</td>
<td>sN-N</td>
<td>32</td>
<td>39</td>
<td>3688</td>
<td>sN-H</td>
<td>100</td>
</tr>
<tr>
<td>20</td>
<td>1269</td>
<td>sN-N</td>
<td>36</td>
<td>rCH3'</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Frequencies in cm⁻¹. Assignments and % are the modes and their respective percents contributing to each vibration.

From MacKerell
Comparison of the scaled ab initio, by analogy and optimized vibrations for selected modes

<table>
<thead>
<tr>
<th></th>
<th>g98</th>
<th>Analogy</th>
<th>Optimized</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>Freq</td>
<td>Assi</td>
</tr>
<tr>
<td>sN=N</td>
<td>30</td>
<td>1782</td>
<td>sN=C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>1646</td>
<td>sN=C</td>
</tr>
<tr>
<td>sN-N</td>
<td>19</td>
<td>1234</td>
<td>sC5-C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>1269</td>
<td>sN-N</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>1802</td>
<td>sN-N</td>
</tr>
<tr>
<td>dC2NN</td>
<td>4</td>
<td>154</td>
<td>dC2NN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dN3NC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dN4CC</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>586</td>
<td>dC1CN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dC2NN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rC=O</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NH1-NR1 from 400/1.38 to 550/1.36
NR1=CEL1 from 500/1.342 to 680/1.290:
C-NH1-NR1 from 50.0/120.0 to 50.0/115.0;

From MacKerell
Dihedral optimization based on QM potential energy surfaces (HF/6-31G* or MP2/6-31G*).

From MacKerell
Potential energy surfaces on compounds with multiple rotatable bonds

1) Full geometry optimization
2) Constrain $n-1$ dihedrals to minimum energy values or trans conformation
3) Sample selected dihedral surface
4) Repeat for all rotatable bonds dihedrals
5) Repeat 2-5 using alternate minima if deemed appropriate

From MacKerell
Note that the potential energy surface about a given torsion is the sum of the contributions from ALL terms in the potential energy function, not just the dihedral term.

From MacKerell
Model 1, Surface 1, Energy components

From MacKerell
Creation of full compound

1) Obtain indole and phenol from top_all22_model.inp
2) Rename phenol atom types to avoid conflicts with indole (add P)
3) Delete model 1 terminal methyls and perform charge adjustments
   i) Move HZ2 charge (0.115) into CZ2 (-0.115 -> 0.00) total charge on deleted C1 methyl (0.00) onto CZ2 (0.00 -> 0.00)
   ii) Move HPG charge (0.115) into CPG (-0.115 -> 0.00) and move total charge on the C6 methyl (0.18) onto CPG (0.00 -> 0.18)
4) Add parameters by analogy (use CHARMM error messages)
5) Generate IC table (IC GENErate)
6) Generate cartesian coordinates based on IC table (check carefully!)

From MacKerell
Chemistry of Thioesters

Most important example in biology of a thioester is acetyl coA, an intermediate used by nature in the biosynthesis of numerous organic compounds.

\[
\begin{align*}
\text{O} \\
\| \\
-C - S - C-
\end{align*}
\]

Experimental Data*

C-S (1.75 Å), O=C-S-C (~-4), C-S-C-H (low barrier)
R-C-S (~113), R-C-O (~123), S-C-O (~124)

*Arch.Bioch.Biophys. Zacharias et al. v222,22-34,1983
Thioester Linkage in Photoactive Yellow Protein - PDB

Carbon of resname HC4
thioester linkage = 1.77 angstroms
Sulfur of resid 69
HG1 deleted from CYS and the charge was moved to SG (-0.23 + 0.16 = 0.07) so that the SG charge becomes 0.07 in final compound and the group remains neutral.

This can be improved!!
Quantum Chemical Calculations for New CF Parameters

Classical Potentials:

\[ V(\vec{R}) = \sum_{i \in \text{bonds}} k_{i,\text{bond}} (r_{i,\text{bond}} - r_{i,\text{bond}}^{\text{eq}}) + \sum_{i \in \text{angles}} k_{i,\text{angle}} (\theta_{i,\text{angle}} - \theta_{i,\text{angle}}^{\text{eq}}) + \]

\[ + \frac{1}{2} \sum_{i \neq j \atop i, j \in \text{atoms}} \frac{q_i q_j}{|r_i - r_j|} + \frac{1}{2} \sum_{i \neq j \atop i, j \in \text{atoms}} V_{LJ}^i (|r_i - r_j|) \]

+ ....

QM Operators:

\[ \hat{H} \Psi_{e,n} = E \Psi_{e,n} \]
\[ \hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}_{e,n} \]

Born Oppenheimer Approximation:

\[ \Psi_{e,n} = \chi_n \psi_{e,n} \]

\[ \hat{H}_{\text{electronic}} (\vec{R}) \psi_{i,\text{electronic}} (\vec{r}_{\text{electronic}} ; \vec{R}) = E_{i,\text{electronic}} (\vec{R}) \psi_{i,\text{electronic}} (\vec{r}_{\text{electronic}} ; \vec{R}) \]
Electronic Hamiltonian

\[ \hat{H}_{\text{electronic}} = \sum_i \hat{T}(i) - \sum_{iA} \frac{Z_A}{r_{iA}} + \sum_{i,j} \frac{1}{r_{ij}} + \sum_{A,B} \frac{Z_A Z_B}{R_{AB}} \]

- Kinetic Energy of electrons
- Electron-nucleus attraction
- Electron-electron repulsion
- Nuclear-nuclear repulsion
The Never-Ending Contraction

\[ \chi_k^{AO,CBF} = \sum_{p=1}^{N_k} C_{pk}^{Prim->CBF} \chi_p^{AO,primitive} (r, R) \]

One-particle basis set

\[ \phi_i^{MO} = \sum_{k=1}^{N_{CBF}} c_{ik}^{MO} (R) \chi_k^{AO,CBF} (r, R) \]

Molecular orbitals are orthogonal contractions of AOs

\[ \psi_\mu (r, R) = A \left[ \prod_i \phi_i^{MO} (r, R) \right] \]

Antisymmetrized products of MOs

\[ \Psi_{elec} (r, R) = \sum_{\mu=1}^{N_{AP}} c_{\mu}^{CI} (R) \psi_\mu (r, R) \]

Total electronic wfn is contraction of APs

Basis Function Overview

\[ \phi_S(r) = \sum_{i=1}^{N} c_i, S e^{-\alpha_i \cdot S f^2 r^2} \]

S-type basis function is composed of sum of primitive gaussian functions. N is the number gfs and is called the degree-of-contraction. C, \( \alpha \), and f are the contraction coefficients, exponents, and scale factors.

It takes at least 3 Gaussians to approximate an S orbital

\[ e^{-r}, e^{-0.6r^2}, e^{-1.5r^2} \]
Basis Set Pople Classification

Minimal Basis Set - STO-3G
One BF per occupied orbital on an atom
e.g. H 1s, 1st row elements 1s,2s,2px,2py,2pz like orbitals

Split Valence Basis Set - Inner and Outer’ basis functions
3-21G: 1s - 3G; 2s, 2p - 2G; 2s’,2p’ - 1G
6-31G: 1s - 6G; 2s, 2p - 3G; 2s’, 2p’ - 1G

Polarization
Add higher angular momentum functions
e.g. 6-31G* = 6-31G+d for 1st row
6-31G** = 6-31G+(d,p) 1st row and +p for H
Semiempirical Methods and Number of Parameters in Method

MNDO Modified Neglect of Differential Overlap 10
AM1 Austin Model 1 13
PM3 Parametric Model number 3 13

Perturbation Method to Treat Electron Correlation: MP2
Improvement over HF, RHF and UHF

Molecular Properties from Wavefunctions
Dipole moment, partial charges, vibrations, ....
Choice of Basis Set in Spartan

<table>
<thead>
<tr>
<th>Level</th>
<th>Basis Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMFF94</td>
<td></td>
</tr>
<tr>
<td>PM3</td>
<td></td>
</tr>
<tr>
<td>6-31G*</td>
<td></td>
</tr>
</tbody>
</table>
PM3
Spartan

Bond angle C-CO-S bend

Bond stretch C=O
CH3 stretches
Dihedral Draw to Check Barriers
<table>
<thead>
<tr>
<th>Atom type</th>
<th>Mulliken 6-31G**</th>
<th>ESP 6-31G**</th>
<th>Charmm</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT2</td>
<td>-0.63</td>
<td>-0.08</td>
<td>-0.11</td>
</tr>
<tr>
<td>HA</td>
<td>0.22</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>HA</td>
<td>0.20</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>HA</td>
<td>0.22</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>0.11</td>
<td>-0.29</td>
<td>-0.07</td>
</tr>
<tr>
<td>CC</td>
<td>0.36</td>
<td>0.64</td>
<td>0.55</td>
</tr>
<tr>
<td>O</td>
<td>-0.50</td>
<td>-0.50</td>
<td>-0.55</td>
</tr>
<tr>
<td>CT2</td>
<td>-0.58</td>
<td>-0.46</td>
<td>-0.18</td>
</tr>
<tr>
<td>HA</td>
<td>0.20</td>
<td>0.14</td>
<td>0.09</td>
</tr>
<tr>
<td>HA</td>
<td>0.21</td>
<td>0.14</td>
<td>0.09</td>
</tr>
<tr>
<td>HA</td>
<td>0.22</td>
<td>0.14</td>
<td></td>
</tr>
</tbody>
</table>

No waters in QM calc.
## Results on Fragment

<table>
<thead>
<tr>
<th>Bonds</th>
<th>Ab initio 6-31G*</th>
<th>Minimization</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT2 – S</td>
<td>1.81 Å</td>
<td>1.85 Å</td>
</tr>
<tr>
<td>CC – S</td>
<td>1.78 Å</td>
<td>1.79 Å</td>
</tr>
<tr>
<td>CC – CT2</td>
<td>1.51 Å</td>
<td>1.54 Å</td>
</tr>
<tr>
<td>CC – O</td>
<td>1.19 Å</td>
<td>1.23 Å</td>
</tr>
<tr>
<td><strong>Angles</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S – CC – CT2</td>
<td>114.3 °</td>
<td>117.1 °</td>
</tr>
<tr>
<td>CC – S – CT2</td>
<td>100.2 °</td>
<td>107.7 °</td>
</tr>
<tr>
<td>S – CC – O</td>
<td>122.2 °</td>
<td>122.9 °</td>
</tr>
</tbody>
</table>
Classical Force Fields

- Physics-based, full atom FF like CHARMM, AMBER, OPLS - Mechanisms, function, protein/RNA/DNA interactions…

✔ Knowledge-based, coarse-grained model -
  - Protein structure prediction, folding kinetics, docking,…

✔ Hybrid force fields like Go + full atom FF -
  - Mutational effects on protein folding kinetics,…

- Hybrid QM/MM approaches
  - Enzyme reactions, bond breaking and making, excited states,…

Each problem has a different goal and time scale!
Protein Structure Prediction

1-D protein sequence
SISSIRVKSKRIQLG….

3-D protein structure

Ab Initio protein folding

Ab Initio Folding: the Energy Function

\[ E = E_{\text{backbone}} + E_{\text{residue/residue}} \]
Ab initio Structure Prediction – Prediction without Homology

• Reduced Representation: $C_\alpha, C_\beta, O$.

• Interaction Potentials: AMH and Contact averaged over MS

• Non-additive HB
**Ab initio Interaction Potentials**

\[ E_{AMC} = E_{AM} + E_C = (E_{short} + E_{medium}) + E_{long} \]

Associate similar sequence/structure fragments in protein database

\[
E_{AM} = -\frac{\varepsilon}{a} \sum_{\mu=1}^{N} \sum_{j-12 \leq i \leq j-3} \left\{ \gamma_{AM} [P_i, P_j, P_i^\mu, P_j^\mu, x(|i-j|)] \exp \left[ \frac{-(r_{ij} - r_{ij}^\mu)^2}{2\sigma_{ij}^2} \right] \right\}
\]

Long range interactions mimic pair distribution functions

\[
E_{long} = -\frac{\varepsilon}{a} \sum_{i<j-12} \sum_{k=1}^{3} \left\{ \gamma_{long} [P_i, P_j, k] c_k(N) U \left[ r_{min}(k), r_{max}(k), r_{ij} \right] \right\}
\]

Weights \( \gamma \) learned from training set
Optimization Strategy for Coarse Grained Energy Function

Energy Landscape Theory

When $<\frac{E_s}{E}>$ is maximum
the energy landscape is optimally funneled.

Vary the parameters to obtain discrimination

Optimization over an Ensemble of Folds

Energy Landscapes of Prediction Energy Function

Q fraction of native contacts
CASP5 Wolynes-Schulten Prediction

T0170

FF domain of HYPA/FBP11

Blue – NMR structure
Orange – Predicted Structure
Global RMSD 2.67 Å.

Hubbard Plot

T0170 contact map

1h40 NMR structure

T0170 prediction
CASP5 Result
Functional Annotation Requires Structures of $\sim Q=0.4$

\[ Q = \frac{1}{N_{\text{pairs}}} \sum_{i,j} \exp \left[ -\frac{(r_{ij} - r_{ij}^N)^2}{2\sigma_{ij}^2} \right] \]
Hybrid Force Field Calculations
Future Directions?

- Protein Folding Kinetics and Mutational Studies
  - Go Potentials
  - Charmm + Go Potentials

T. Pogorelov and Z. Luthey-Schulten, Biophysical J. 87 (2004)
LAMBDA REPRESSOR
MUTATIONS Q33Y and A37G
Folding of lambda-repressor
T. Pogorelov and Z. Luthey-Schulten, Biophysical J. 87 (2004)
Free-Energy Potentials for Fast and Slow Mutants
Acknowledgements

• Rommie Amaro – HisH/HisF
• Felix Autenrieth - Cyt c2 Parameterization
• Taras Pogorelo - Go + CHARMM