CHARMM force fields, parameterization strategies and future/ongoing force field developments

Alexander D. MacKerell, Jr.

University of Maryland, Baltimore

School of Pharmacy

MMTSB/CTBP

2006 Summer Workshop

Potential energy function
(mathematical equations)

Empirical force field

equations and parameters
relate chemical structure and conformation to energy

Common “additive” empirical force fields

Class I

CHARMM
CHARMm (Accelrys)
AMBER
OPLS/AMBER/Schrödinger
ECEPP (free energy force field)
GROMOS

Class II

CFF95 (Accelrys)
MM3
MMFF94 (CHARMM, Macromodel, MOE, elsewhere)
UFF, DREIDING

State of the art additive force fields are typically all-atom models

All atoms, including all hydrogens, explicitly represented in the model.

Lone pairs included on hydrogen bond acceptors in some force fields.

e.g., CHARMM22 and 27, AMBER94….03, OPLS/AA

Extended or united atom models

(omit non-polar hydrogens)

CHARMM PARAM19 (proteins)

- often used with implicit solvent models
  - ACE, EEF, GB variants
- improper term to maintain chirality
- loss of cation - pi interactions

OPLS

AMBER

GROMOS

Transition State Force Field Parameters

Same approach as standard force field parameterization

Require target data for transition state of interest: *ab initio*

Metal Force Field Parameterization

Only interaction parameters or

include intramolecular terms

Parameterization of QM atoms for QM/MM calculations

Polarizable “non-additive” force fields

Include explicit term(s) in the potential energy function to treat induction/polarization of the charge distribution by the environment. Still under development.

CHARMM
- Drude (MacKerell, Roux and coworkers)
- PIPF (Gao and coworkers)
- Cheq (Brooks and coworkers)

AMBER

Friesner/Berne et al. (Schrödinger Inc.)

TINKER

Class I Additive Potential Energy Function

Intramolecular (internal, bonded terms)

$$\sum_{bonds} K_b\left(b-b_o\right)^2 + \sum_{angles} K_\theta\left(\theta-\theta_o\right)^2 + \sum_{torsions} K_\phi\left(1+\cos(n\phi-\delta)\right)$$

$$+ \sum_{impropers} K_\varphi\left(\varphi-\varphi_o\right)^2 + \sum_{Urey-Bradley} K_{UB}\left(r_{1,3}-r_{1,3,o}\right)^2$$

Intermolecular (external, nonbonded terms)

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

Class II force fields (e.g. MM3, MMFF, UFF, CFF)

$$\sum_{bonds}\left[K_{b,2}\left(b-b_o\right)^2+K_{b,3}\left(b-b_o\right)^3+K_{b,4}\left(b-b_o\right)^4\right]$$

$$+\sum_{angles}\left[K_{\theta,2}\left(\theta-\theta_o\right)^2+K_{\theta,3}\left(\theta-\theta_o\right)^3+K_{\theta,4}\left(\theta-\theta_o\right)^4\right]$$

$$+\sum_{dihedrals}\left[K_{\phi,1}\left(1-\cos\phi\right)+K_{\phi,2}\left(1-\cos 2\phi\right)+K_{\phi,3}\left(1-\cos 3\phi\right)\right]$$

$$+\sum_{impropers}K_{\chi}\chi^2$$

$$+\sum_{bonds}\sum_{bonds'}K_{bb'}\left(b-b_o\right)\left(b'-b_o'\right)+\sum_{angles}\sum_{angles'}K_{\theta\theta'}\left(\theta-\theta_o\right)\left(\theta'-\theta_o'\right)$$

$$+\sum_{bonds}\sum_{angles}K_{b\theta}\left(b-b_o\right)\left(\theta-\theta_o\right)$$

$$+\sum_{bonds}\sum_{dihedrals}\left(b-b_o\right)\left[K_{\phi,b1}\cos\phi+K_{\phi,b2}\cos 2\phi+K_{\phi,b3}\cos 3\phi\right]$$

$$+\sum_{bonds'}\sum_{dihedrals}\left(b'-b_o'\right)\left[K_{\phi,b'1}\cos\phi+K_{\phi,b'2}\cos 2\phi+K_{\phi,b'3}\cos 3\phi\right]$$

$$+\sum_{angles}\sum_{dihedrals}\left(\theta-\theta_o\right)\left[K_{\phi,\theta 1}\cos\phi+K_{\phi,\theta 2}\cos 2\phi+K_{\phi,\theta 3}\cos 3\phi\right]$$

$$+\sum_{angles}\sum_{angles'}\sum_{dihedrals}\left(\theta-\theta_o\right)\left(\theta'-\theta_o'\right)\cos\phi$$

Merck Molecular FF:
Force field for drug-like molecules

MMFF is a force field designed for pharmaceutical compounds as well as biological molecules. It may be considered one of the better general FFs, although its quality in treating proteins etc. is worse than CHARMM and other biological FFs. Therefore, MMFF is good for computing drug-receptor interactions but not for extensive minimizations etc. of proteins. The tutorial MMFF_Interaction gives an example of reading a drug molecule in Mol2 format, reading a protein structure and calculating the interaction energy. See mmff_inter_energy.inp

Intramolecular energy function and corresponding force field parameters

$$\sum_{bonds} K_b\left(b - b_o\right)^2 + \sum_{angles} K_\theta\left(\theta - \theta_o\right)^2 + \sum_{torsions} K_\phi\left(1 + \cos(n\phi - \delta)\right)$$

$$+ \sum_{impropers} K_\varphi\left(\varphi - \varphi_o\right)^2 + \sum_{Urey-Bradley} K_{UB}\left(r_{1,3} - r_{1,3,o}\right)^2 + \sum_{\phi,\psi} V_{CMAP}$$

Equilibrium terms

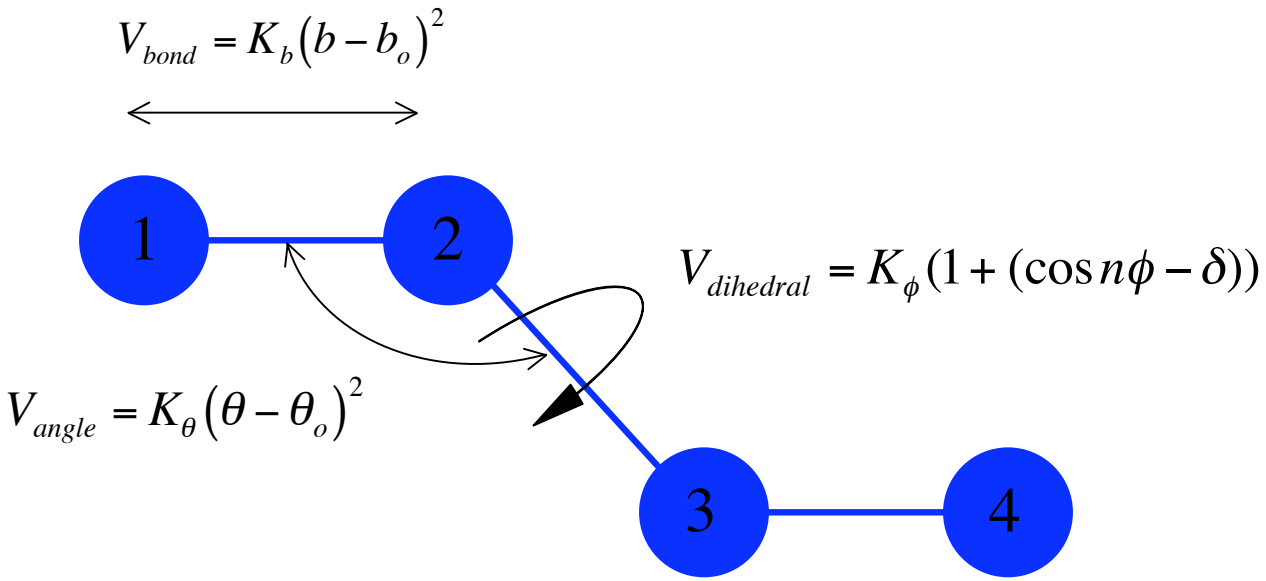
- b_o: bonds
- θ_o: angles
- n: dihedral multiplicity
- δ_o: dihedral phase
- ω_o: impropers
- $r_{1,3o}$: Urey-Bradley

Force constants

- K_b: bonds
- K_θ: angles
- K_ϕ: dihedral
- K_ω: impropers
- K_{UB}: Urey-Bradley

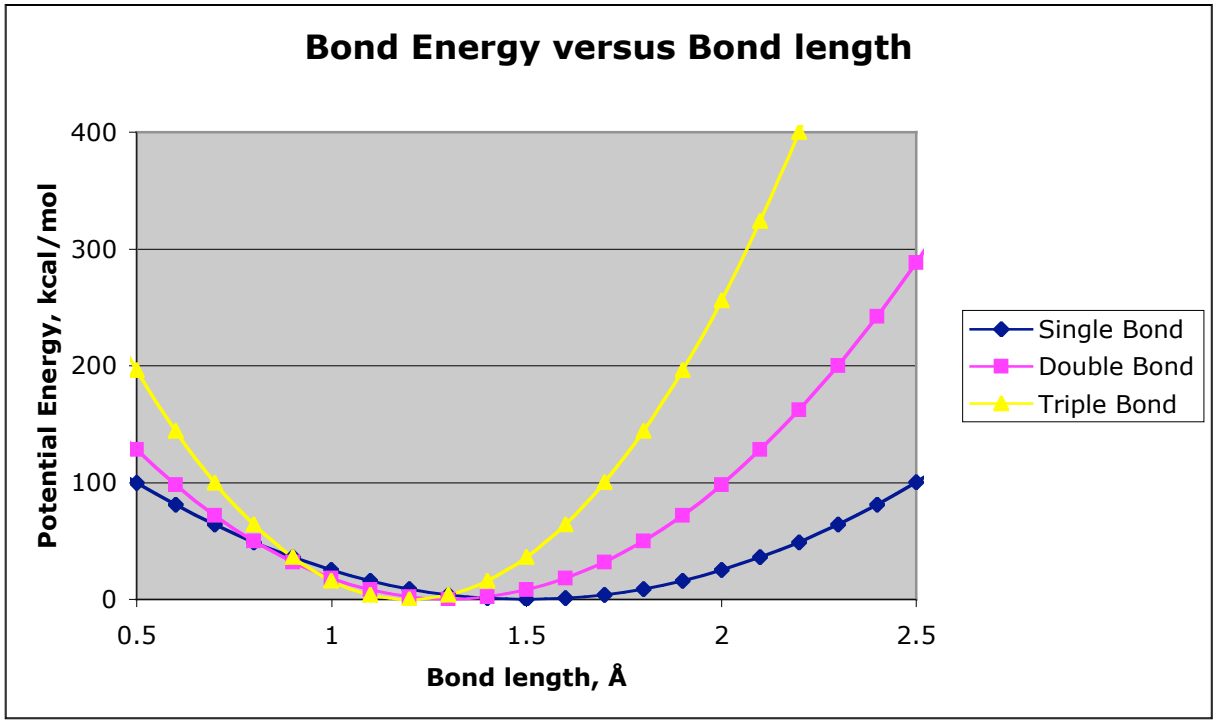
Aka. Internal or bonded terms

Diagram of intramolecular energy terms

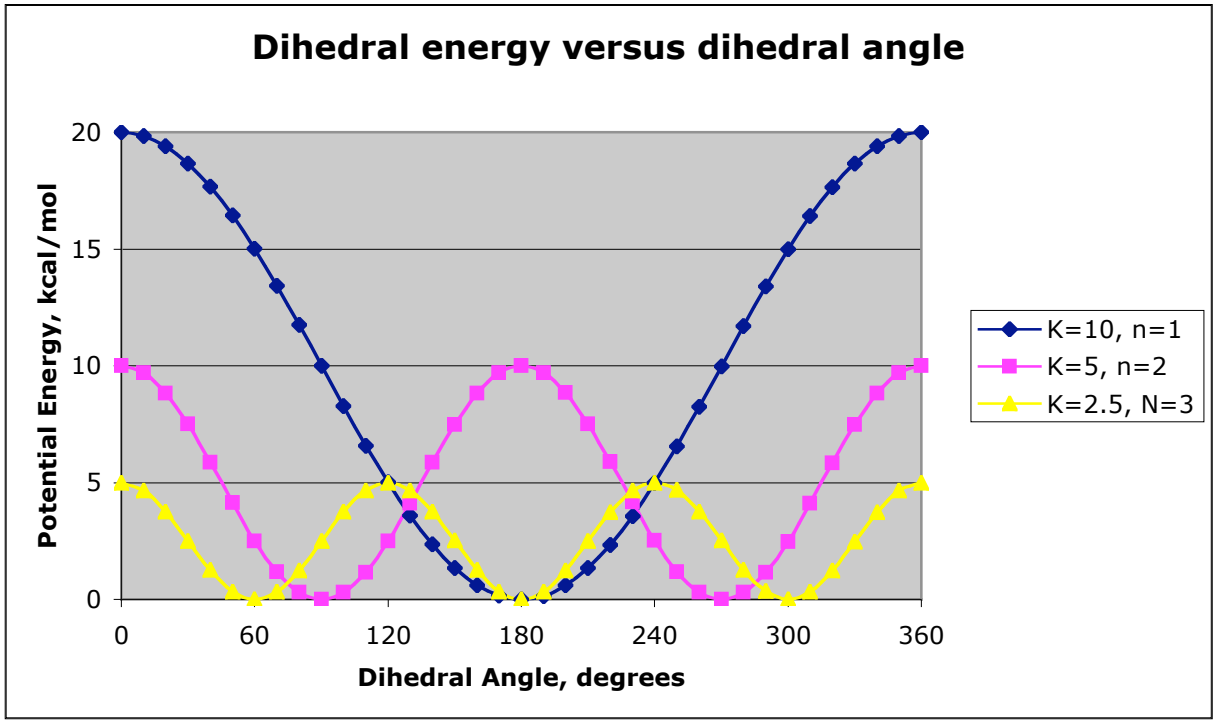


$$V_{bond} = K_b\left(b - b_o\right)^2$$

Chemical type	K_{bond}	b_o
C-C	100 kcal/mole/Å^2	1.5 Å
C=C	200 kcal/mole/Å^2	1.3 Å
C=-C	400 kcal/mole/Å^2	1.2 Å

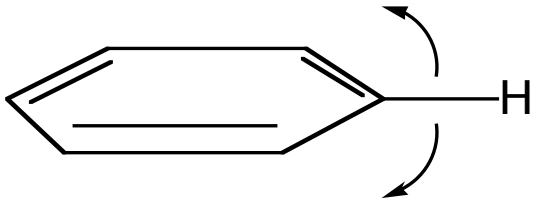


$$V_{dihedral} = K_{\phi}(1 + (\cos n\phi - \delta))$$

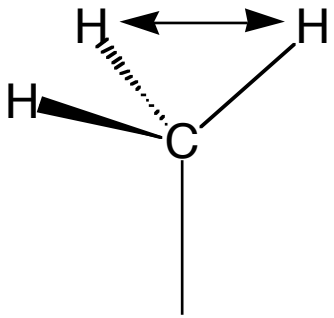


$\delta = 0^{\circ}$

Note use of a Fourier series for a dihedral



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



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

2D dihedral energy correction map to the CHARMM 22 ϕ,ψ backbone (CMAP)

ϕ,ψ grid-based energy correction via bicubic interpolation

$$V_{CMAP} = f(\phi,\psi) = \sum_{i=1}^{4}\sum_{j=1}^{4} c_{ij}\left(\frac{\phi-\phi_L}{\Delta_\phi}\right)^{i-1}\left(\frac{\psi-\psi_L}{\Delta_\psi}\right)^{j-1}$$

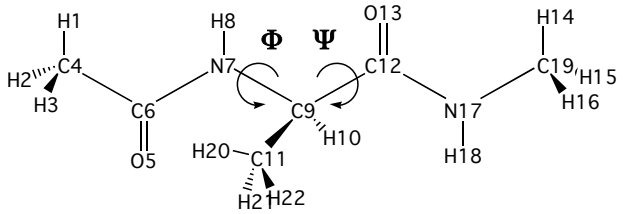
Smooth first derivatives, continuous second derivatives

Grid rectangle coefficients, c_{ij}

1) Corner grid points

2) First derivatives: $\frac{\partial f}{\partial \phi}$, $\frac{\partial f}{\partial \psi}$

3) Cross derivatives: $\frac{\partial f^2}{\partial \phi \partial \psi}$



Use bicubic spline interpolation to determine derivatives

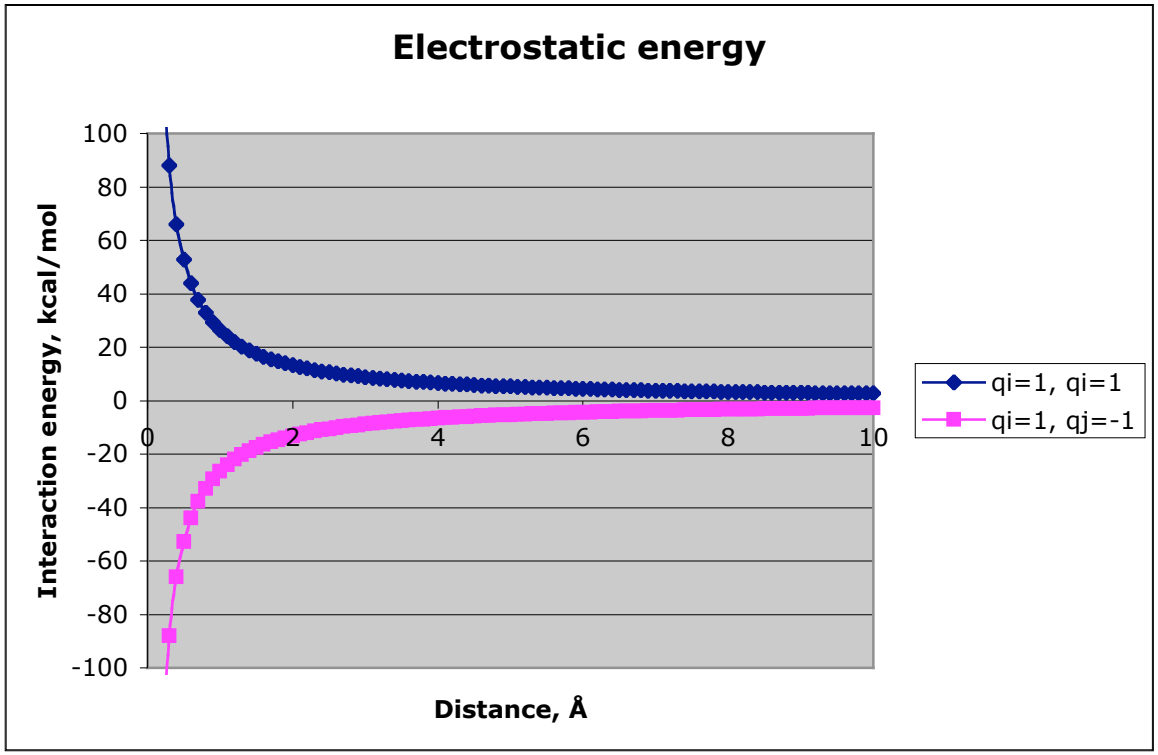
Additive intermolecular energy function and corresponding parameters

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

q_i: partial atomic charge
D: dielectric constant
ε: Lennard-Jones (LJ, vdW) well-depth
R_{min}: LJ radius ($R_{min}/2$ in CHARMM)
Combining rules (CHARMM, Amber)

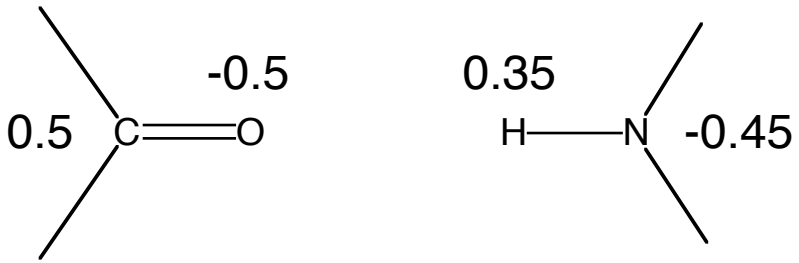
$R_{min\ i,j} = R_{min\ i} + R_{min\ j}$
$\varepsilon_{i,j} = SQRT(\varepsilon_i * \varepsilon_j)$

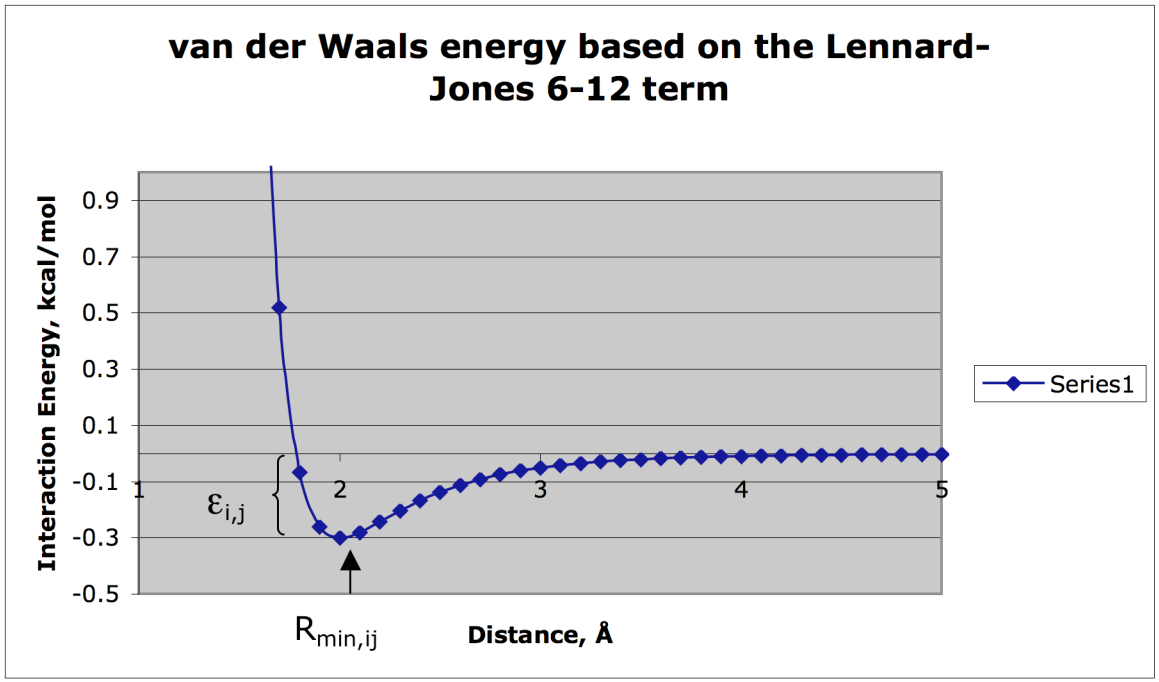
Aka. Nonbonded or external terms



Treatment of hydrogen bonds???

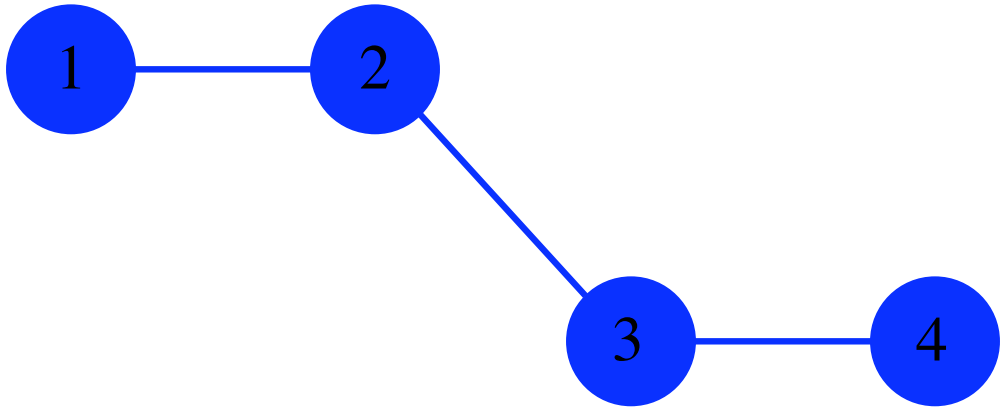
Partial atomic charges





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

Example of nonbond exclusions



nonbond (intermolecular) interactions between bonded atoms are treated with special rules

1,2 interactions: 0

1,3 interactions: 0

1,4 interactions: 1 or scaled

> 1,4 interactions: 1

Alternate intermolecular terms for the electrostatic (additive) or vdW interactions

$$V_{Hbond} = \sum_{Hbonds} \varepsilon_{HB} \left[\left(\frac{R_{HB,A-H}}{r_{A-H}} \right)^{12} - \left(\frac{R_{HB,A-H}}{r_{A-H}} \right)^{10} \right] * \cos\left(\theta_{A-H-D}\right)$$

$$V_{vdw} = \sum_{vdw} \varepsilon_{ij} \left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{9} - \left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right]$$

$$V_{vdw} = \sum_{vdw} \varepsilon_{ij} \left(e^{\frac{-aR_{\min,ij}}{r_{ij}}} - \left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right)$$

Limitation of additive force fields

The use of Coulomb's law with fixed atomic charges to treat the electrostatic interactions is a major simplification in current force fields. It is well known that the electron distribution of a molecule (and, thus, the atomic charges) changes as a function of the electrostatic field around the molecule. This is ignored in additive force fields. To compensate for this omission, the atomic charges are "enhanced" to mimic the polarization of molecules that occurs in a polar, condensed phase environment (e.g. aqueous solution, TIP3P water model dipole moment = 2.35 versus gas phase value of 1.85). This approximation has worked well in the current additive force fields; however, in many cases these models fail. To overcome this, next generation force fields are being developed that explicitly treat electronic polarization.

Methods to include electronic polarization in force fields

Fluctuating charge (CHEQ)

Induced dipoles (PIPF, Berne/Friesner, AMBER)

Classical Drude Oscillator

All methods require that the perturbation of the electronic distribution due to the surrounding electrostatic field be optimized in an iterative fashion. This is due to the change in the "charge distribution" of a system leading to a new electrostatic field which then requires additional re-adjustment of the charge distribution (SCF: self-consistent field calculation). Matrix diagonalization may also be used, but is frequently inaccessible due to the large number of atoms in biological systems. In the end the need to perform an SCF calculation leads to a large increase in computational demands. Special methods to minimize this limitation in MD simulations have been developed (see below).

Fluctuating Charge Model (CHEQ)

Polarization is based on the movement of charge, q, between bonded atoms i and j in response to the surrounding electrostatic field. The extent of charge movement is based on the relative electronegativity, χ, and hardness, J, of the bonded atoms. The electrostatic energy is then obtained from the Coulombic interactions between the relaxed charges.

$$V(q_{ij}) = \chi_{ij} q_{ij} + \frac{1}{2} J_{ij} q_{ij}^2$$

$$\chi_{ij} = \chi_i^{'} + \chi_j^{'} \qquad\qquad J_{ij} = J_i^{'} + J_j^{'} + 2J_{ij}^{'}$$

Electronegativity: attraction of an atom for electrons
Hardness: work needed to transfer charge (resistance to charge movement)

Induced Dipole Model

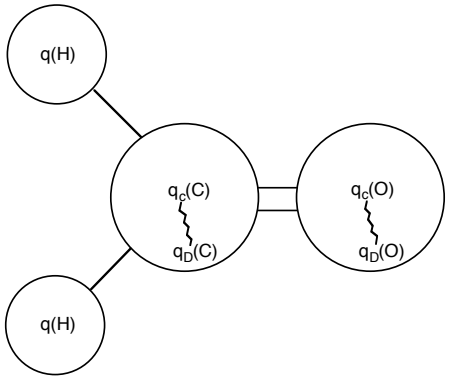
Each atom, i, carries a charge, q_i, and a dipole moment, μ_i, such that electrostatic interactions between atoms i and j include:

charge-charge interactions: $1/r_{ij}$
charge-dipole interactions: $1/r_{ij}^2$
dipole-dipole interactions: $1/r_{ij}^3$

Polarization included via relaxation of dipole moments in the electrostatic field, E_i, where α_i is the polarizability of atom i

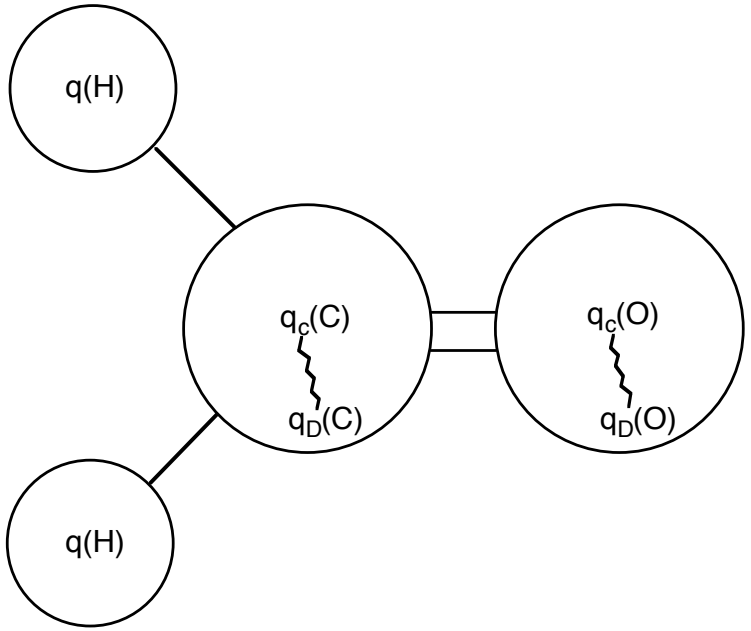
$$\mu_i = \alpha_i\left(E_i^0 + E_i^{induced}\right) = \alpha_i\left(E_i^0 + \sum_{i \neq j} T_{ij}\mu_j\right)$$

Classical Drude Oscillator



To each atom, i, add a virtual particle (Drude) attached to the atomic core via a harmonic spring and place a charge, q_D, on the Drude. The Drudes then relax their positions with respect the surrounding electrostatic field with the relative positions of the Drudes with respect to their parent atom along with the respective charges of each yielding an induced dipole moment on each atom. The electrostatic energy is then obtained from the Coulombic interactions between the atomic and Drude charges.

Classical Drude oscillator



$$\alpha(A) = q_{\mathrm{D}}^2(A) \Big/ k_{\mathrm{D}}$$

$$q_D(A) = \sqrt{\alpha(A) \Big/ k_D}$$

$$q_c(A) = q(A) - q_D(A)$$

$$U_{Drude} = \sum_{A<B}^{N,N_{\mathrm{D}}} \frac{q_{\mathrm{D}}(A) \cdot q_c(B)}{\left|\mathbf{r}_{\mathrm{D}}(A) - \mathbf{r}(B)\right|} + \sum_{A<B}^{N_{\mathrm{D}}} \frac{q_{\mathrm{D}}(A) \cdot q_{\mathrm{D}}(B)}{\left|\mathbf{r}_{\mathrm{D}}(A) - \mathbf{r}_{\mathrm{D}}(B)\right|} + \frac{1}{2} \sum_{A}^{N_{\mathrm{D}}} k_{\mathrm{D}} \left|\mathbf{r}_{\mathrm{D}}(A) - \mathbf{r}(A)\right|^2$$

MD Simulations with polarizable force fields: Extended-Langrangian

SCF calculation of induced dipole moments are computationally to demanding for MD simulations. As an alternative the polarization is treated as a dynamic variable that is propagated during the MD trajectory. This is done such that the electronic degrees of freedom being propagated in the MD simulation stay close to the Born-Oppenheimer approximation (e.g. equivalent to the SCF result). For example, in the Drude model, the Drude particle is assigned part of the mass of the parent atom (e.g. 0.5 amu) and then the Drude is propagated as an atom at each step of the MD simulation with the relative momentum of the Drude with respect to the parent atom "cooled" to 0 K, thereby approaching the Born-Oppenheimer approximation.

Potential energy function versus a force field

A potential energy function is merely an equation that relates structure to energy (and forces etc.). However, the equation alone is useless until the parameters that have to be input into the equations (see above) have been optimized to represent real chemical systems. Once this has been attained one has a force field that may be used for energy minimization, MD simulations and so on. In the remainder of this lecture the methods used to optimize parameters for new molecules will be presented. This will be done primarily in the context of additive force fields currently in use in CHARMM. However, the majority of the concepts may be transferred directly to next generation polarizable force field. The major difference will be in the optimization of the electrostatic parameters.

Extension of the additive CHARMM force fields for drug like molecules

1) Decompose molecule into molecular fragments

2) Identify molecular fragments already in the CHARMM force fields

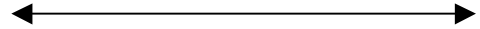
3) Create RTF information for full molecule and molecular fragments (ie. Model compounds) not available (toppar stream file).

4) Identify missing parameters, obtain initial guesses for the new parameters based on analogy to available parameters and place in the toppar stream file.

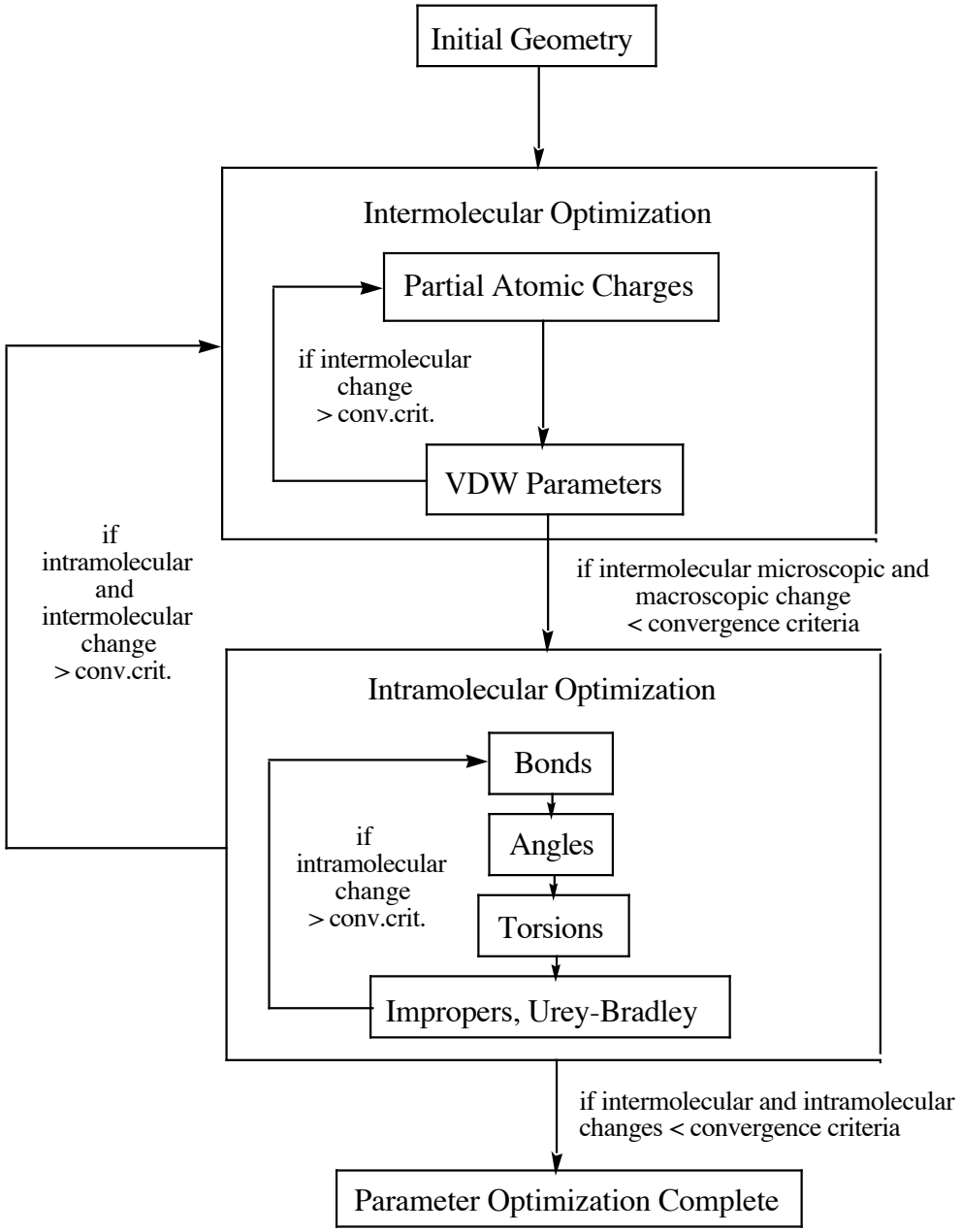
5) Optimize new parameters based on QM data
   i) Geometries and vibrational spectra at MP2/6-31G* (MP2/6-31+G* for anions)
   ii) Conformational energies for rotation of selected dihedrals at MP2/6-31G* (MP2/6-31+G* for anions)
   iii) Partial atomic charges based on reproduction of HF/6-31G* water-model compound interaction energies

6) Perform tests to reproduce experimental data on new molecule if available (structures of many small molecules are available in the Cambridge Structural Database).

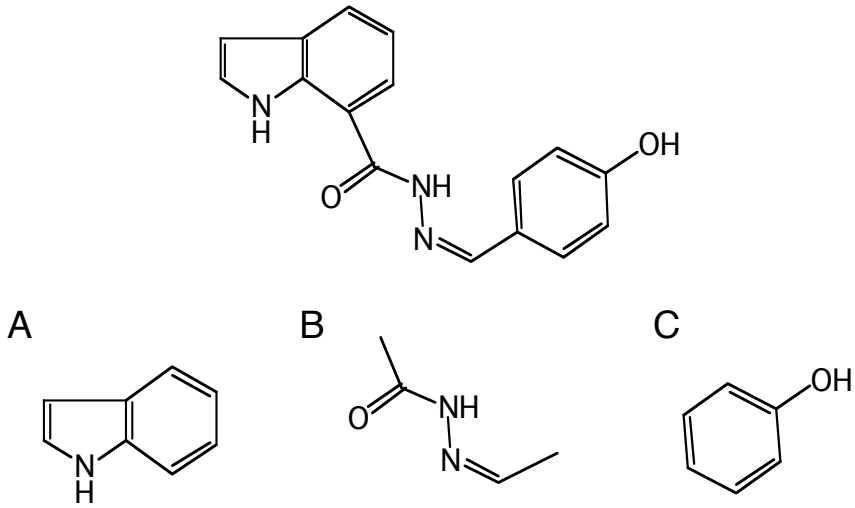
An iterative approach is required to obtain self-consistent parameters

Intramolecular  Intermolecular

The nonbond/intermolecular parameters will impact the resulting geometries, vibrations and conformational energies. Thus, it is necessary to apply an iterative approach where once intramolecular parameters are optimized, the intermolecular parameters are optimized following which the intramolecular parameters must be rechecked and so on in an iterative fashion to all values of the parameters converge. Typically, this only requires one or two iterations, but it may be more with highly flexible molecules.



Deconstruct target molecule into molecular fragments for parameter assignment and optimization



A) Indole
B) Hydrazine (model compound 1)
C) Phenol

Linking model compounds: When creating a covalent link between model compounds move the charge on 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)

1) Identify previously parameterized model compounds in the CHARMM FF
2) Access topology information
   i) Assign atom types
   ii) Connectivity (bonds)
   iii) Charges

In CHARMM toppar and stream subdirectory search for compounds representative of the molecular fragments

Phenol: stream/toppar_all22_prot_model.str (RESI PHEN)
Indole: stream/toppar_all22_prot_model.str (RESI INDO)
Model B not available: create RTF
Identify appropriate parent toppar files that contain the necessary residues and parameters (protein and lipid, as the lipid includes C=C moieties).

top_all27_prot_lipid.rtf
par_all27_prot_lipid.prm

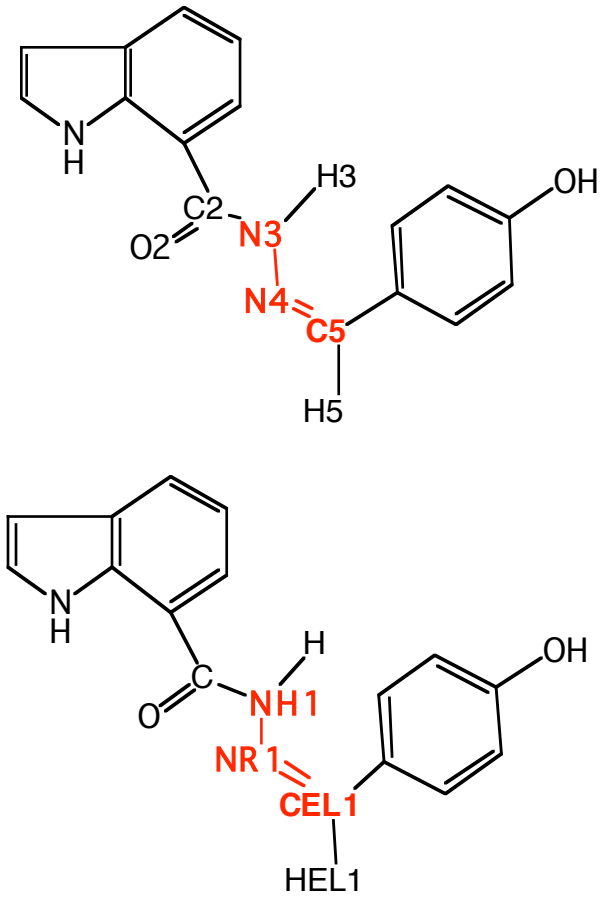
Toppar stream file

see top_mmtsb_example.str

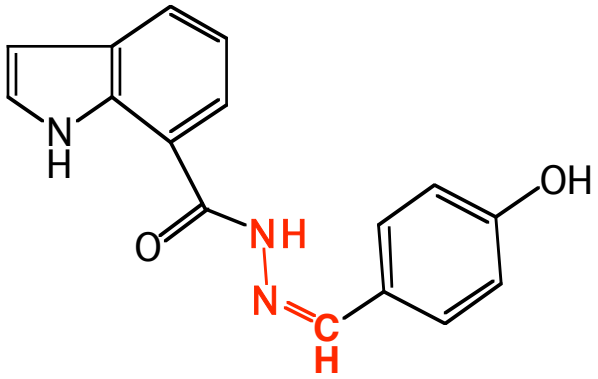
Instead of appending new topology and parameter information to the original rtf and parameter file, create a toppar stream file that contains only the information required for the new molecules. This preserves the integrity of the original files and makes dealing with logistic issues much easier (ie. parameters to be optimized).

Limitation (if not using new flexible parameter reader): need to include MASS specifications for new atom types in the original topology file. Remaining information can be in the toppar stream file. Note that that nonbond parameters for new atom types can be in the toppar stream file, although this will lead to warnings when the parent parameter file is read.

Comparison of atom names (upper) and atom types (lower)



Identify internal parameters to be optimized.
Only optimize new parameters!



Bonds (list doesn't include lipid-protein alkane nomenclature differences)
NH1-NR1, NR1-CEL1
Angles
NR1-NH1-H, NR1-NH1-C, NH1-NR1-CEL1
NR1-CEL1-CTL3, NR1-CEL1-HEL1
Dihedrals
CTL3-C-NH1-NR1, C-NH1-NR1-CEL1, O-C-NH1-NR1,
NH1-NR1-CEL1-HEL1, NH1-NR1-CEL1-CTL3
H-NH1-NR1-CEL1, NR1-CEL1-CTL3-HAL3

Let CHARMM identify missing parameters during IC and energy calls. Add explicit terms if wildcards are used for dihedrals to increase quality of agreement. ONLY include new parameters; do NOT optimize available parameters as this will negatively impact other aspects of the force field. If necessary, create a new atom type for a selected atom to allow for new parameters to be required and optimized.

read rtf card append
(see top_mmtsb_example.str)

```
Resi Mod1        0.00   ! Model compound B
! based on a combination of peptide and lipid alkane/alkene parameters.
Group
ATOM C1   CTL3  -0.27
ATOM H11  HAL3   0.09
ATOM H12  HAL3   0.09
ATOM H13  HAL3   0.09
GROUP
ATOM C2   C      0.58
ATOM O2   O     -0.50
GROUP
ATOM N3   NH1   -0.32
ATOM H3   H      0.33
ATOM N4   NR1   -0.31   !new atom
ATOM C5   CEL1  -0.25
ATOM H5   HEL1   0.29
ATOM C6   CTL3  -0.09
ATOM H61  HAL3   0.09
ATOM H62  HAL3   0.09
ATOM H63  HAL3   0.09

BOND C1 H11 C1 H12 C1 H13 C1 C2 C2 O2 C2 N3 N3 H3
BOND N3 N4 C5 H5  C5 C6 C6 H61 C6 H62 C6 H63
DOUBLE N4 C5 ! (DOUBLE only required for MMFF)
! IC table omitted
Patch first none last none
```

read param card append
(see top_mmtsb_example.str)

```
read param card append
* append parameters for MMTSB/CTBP workshop
*

!most parameters omitted due to space limitations
BONDS
!Model compound 1
NH1  NR1   550.000     1.3600  !from NR1  CPH1
CTL3 C     250.000     1.4900  !from CT3  C
NR1  CEL1  680.000     1.290   !from CEL1  CEL2

ANGLES
!Model compound 1
HAL3 CTL3 C      33.00    109.50   30.00   2.16300 !from HA CT3 C
NH1  C    CTL3   80.00    116.50 ! from NH1 C CT3, mp2 angle: 112.9

DIHEDRALS
!Model Compound 1
CTL3 C    NH1  H      2.5000  2   180.00 ! from H    NH1  C    CT3
CTL3 C    NH1  NR1    1.6000  1     0.00 ! from CT3  C    NH1  CT1
CTL3 C    NH1  NR1    4.0000  2   180.00

!No IMPRoper or NONBond parameters are needed.

end !end append parameters
Return
```

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                   ! HG--CG       CZ--OH
ATOM CD2  CA    -0.115 !        \       /       \
ATOM HD2  HP     0.115 !       CD2==CE2      HH
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      0.11
ATOM OH   OH1   -0.54
ATOM HH   H       0.43
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
```

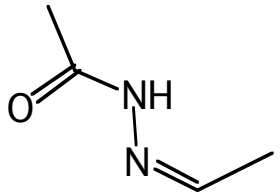
Top_all22_model.inp contains all protein model compounds. Lipid, nucleic acid and carbohydate model compounds are in the full topology files (to toppar stream files/2004).

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 and include these in the toppar stream file.

Creation of topology for central model compound




```
Resi  Mod1 ! Model compound 1
Group          !specifies integer charge group of atoms (not
essential)
ATOM C1    CT3   -0.27
ATOM H11   HA3    0.09
ATOM H12   HA3    0.09
ATOM H13   HA3    0.09
GROUP
ATOM C2    C      0.51
ATOM O2    O     -0.51
GROUP
ATOM N3    NH1   -0.47
ATOM H3    H      0.31
ATOM N4    NR1    0.16   !new atom
ATOM C5    CEL1  -0.15
ATOM H51 HEL1    0.15
ATOM C6    CT3   -0.27
ATOM H61   HA     0.09
ATOM H62   HA     0.09
ATOM H63   HA     0.09
BOND C1 H11 C1 H12 C1 H13 C1 C2 C2 O2 C2 N3 N3 H3
BOND N3 N4 C5 H51 C5 C6 C6 H61 C6 H62 C6 H63
DOUBLE N4 C5 (DOUBLE only required for MMFF)
```

Start with alanine dipeptide.
Note use of new 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.

Parameters by analogy versus optimized parameters

In the following slides various aspects of the parameter optimization process will be given. In slides with results, data labeled “Analogy” represent the results for parameters obtained by analogy to other parameters while the optimized results are those following optimization of the parameters.

Charmm scripts to generate model compounds

Create charmm inputs to generate and minimize models compounds

gen_model_b.inp
gen_full_drug.inp

The scripts involve the compound being generated (ie. created) in Charmm and the structure energy minimized. During this step the program will identify missing parameters which is useful for creation of the list of required parameters in the toppar stream files. Note the creation of multiple conformations to allow for comparison of their energies and geometries and the creation of input files for the Gaussian QM program (gauss subdirectory).

Intermolecular Optimization Target Data

A number are methods are available to obtain the charges and LJ parameters as shown below. For the charges, CHARMM is based on the reproduction of QM minimum interaction energies and geometries along with dipole moments. Final tests are performed to reproduce condensed phase properties, although such data is typically not available for drug-like molecules.

- Local/Small Molecule
  - Experimental
    - Interaction enthalpies (MassSpec)
    - Interaction geometries (microwave, crystal)
    - Dipole moments
  - Quantum mechanical
    - Mulliken Population Analysis
    - Electrostatic potential (ESP) based
      - CHELPG (g03: POP=(CHELPG,DIPOLE))
      - Restricted ESP (AMBER)
    - Dimer Interaction Energies and Geometries (OPLS, CHARMM)
    - Dipole moments

- Global/condensed phase (all experimental)
  - Pure solvents (heats of vaporization, density, heat capacity, isocompressibility)
  - Aqueous solution (heats/free energies of solution, partial molar volumes)
  - Crystals (heats of sublimation, lattice parameters, interaction geometries)

CHARMM Partial Atomic Charge Determination

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

Adjust charges to reproduce HF/6-31G* minimum interaction energies and distances between the model compound and water

scale target HF/6-31G* interaction energies

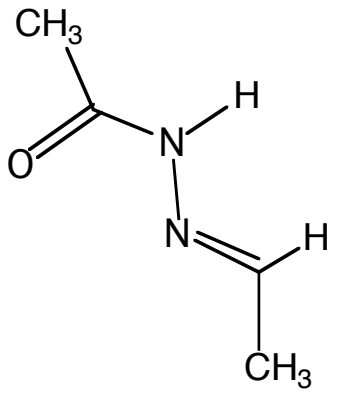
1.16 for polar neutral compounds

1.0 for charged compounds

Empirical distances should be ~0.2 Å shorter the HF/6-31G*

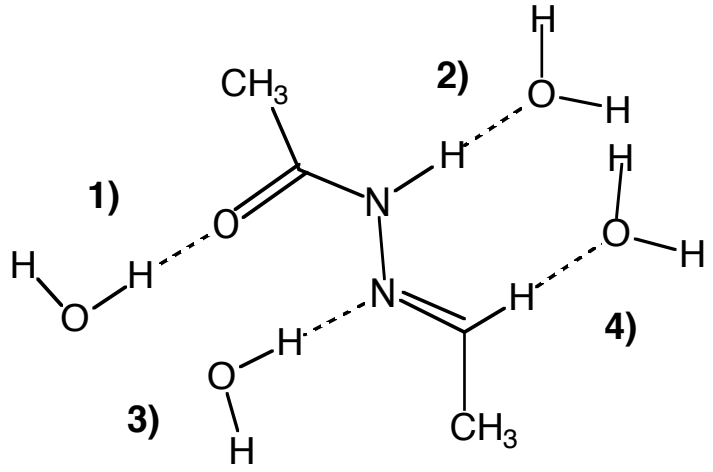
Empirical Dipole moments should be ~10 to 20% large than HF/6-31G* values

For a particular force field do NOT change the QM level of theory for determination of electrostatic parameters. This is necessary to maintain consistency with the remainder of the force field. Thus, use HF/6-31G* for CHARMM additive force fields



Starting charges??
- peptide bond
- methyl
- imidazole (N-N=C)?
- Mulliken population analysis
- Merz-Kollman ESP charges

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

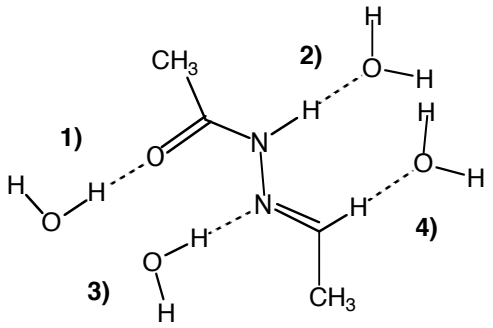


Model compound B-water interaction energies/geometries

see water_model_b.inp

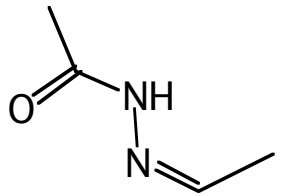
	Interaction Energies (kcal/mole)			Interaction Distances (Å)		
	QM	Analogy	Optimized	QM	Analogy	Optimized
1) O2...HOH	-6.12	-6.56	-6.04	2.06	1.76	1.78
2) N3-H..OHH	-7.27	-7.19	-7.19	2.12	1.91	1.89
3) N4...HOH	-5.22	-1.16	-5.30	2.33	2.30	2.06
4) C5-H..OHH	-3.86	-3.04	-3.69	2.46	2.51	2.44
Energetic statistical analysis						
Ave. Difference		1.13	0.06			
RMS Difference		1.75	0.09			
Dipole Moments (debye)		HF/6-31G*	CHARMM			
		5.69	6.47			

Ab initio interaction energies scaled by 1.16.



Comparison of analogy and the final optimized charges

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



Note charge on C6 methyl carbon. Non-integer charge is typically placed on the adjacent aliphatic carbon.

LJ (vdw) parameters

Direct transfer from available parameters is generally adequate

Test via

- Heat of vaporization
- Density (Molecular Volume)
- Partial molar volume
- Crystal simulations

For details of LJ parameter optimization see Chen, Yin and MacKerell, JCC, 23:199-213 (2002)

Intramolecular optimization target data

Listed below are the types of target data for the internal parameters. For most drug molecules the amount of experimental data is minimal, requiring the use of QM data. (MP2/6-31G* or MP2/6-31+G* for anions). However, for geometries it is often possible to do surveys of the Cambridge Structural Database for a type of linkage to obtain target geomtries.

Geometries (equilibrium bond, angle, dihedral, UB and improper terms)
- microwave, electron diffraction, *ab initio*
- small molecule x-ray crystallography (CSD)
- crystal surveys of geometries

Vibrational spectra (force constants)
- infrared, raman, *ab initio*

Conformational energies (force constants)
- microwave, *ab initio*

Bonds and angles for model compound B

In gen_model_b.inp, look at geometries after minimization using the IC FILL, IC PRINT commands and compare data with target data. Alternatively, the QUICK commands may be used to obtain the CHARMM geometries for comparison.

	MP2/6-31G*		CSD		Analogy	Optimized
Bond lengths	1	2	1	2		
C-N[a]	1.385	1.382	1.37±0.03	1.35±0.01	1.342	1.344
N-N	1.370	1.366	1.38±0.02	1.37±0.01	1.386	1.365
N=C	1.289	1.290	1.29±0.02	1.28±0.01	1.339	1.289
Angles						
C-N-N	120.8	122.4	120.7±5.8	119.7±2.9	124.5	121.4
N-N=C	116.0	116.6	114.5±5.3	115.8±1.6	119.6	115.6
N=C-C	119.9	120.0	120.7±4.7	121.2±2.2	122.4	121.0

The MP2/6-31G* results are for the 1) all-trans and 2) 0˚, 180˚, 180˚ global minimum energy structures. The Cambridge structural database results represent mean±standard deviation for all structures with R-factor < 0.1 and 1) the N7 and C10 sites undefined and 2) the N7 and C10 sites explicitly protonated. A) Not optimized as part of the present study.

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, NH1- NR1-CEL1 from 50.0/120.0 to 50.0/115.0, NR1-CEL1-CT3 from 48.0/123.5 to 48.0/122.5. For planar systems keep the sum of the equilibrium angle parameters equal to 360.0

Bond, angle, dihedral, UB and improper force constants

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

Vibrations are generally used to optimize the bond, angle, UB and improper FCs and, initially, all the dihedrals. Conformational energies associated with rotations about flexible bonds are then used for optimization of the dihedral parameters (K, n and δ) for only dihedrals containing all non-hydrogen atoms.

See model_b_molvib.inp and model_b_molvib_g03: CHARMM scripts to obtain vibrational spectra including assignment of normal modes to frequencies for the empirical and QM levels of theory, respectively.

Vibrational Spectra of Model Compound B from MP2/6-31G* QM calculations (see model_b_molvib_g03.inp)

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

Frequencies in cm^{-1}. Assignments and % are the modes and there respective percents contributing to each vibration.

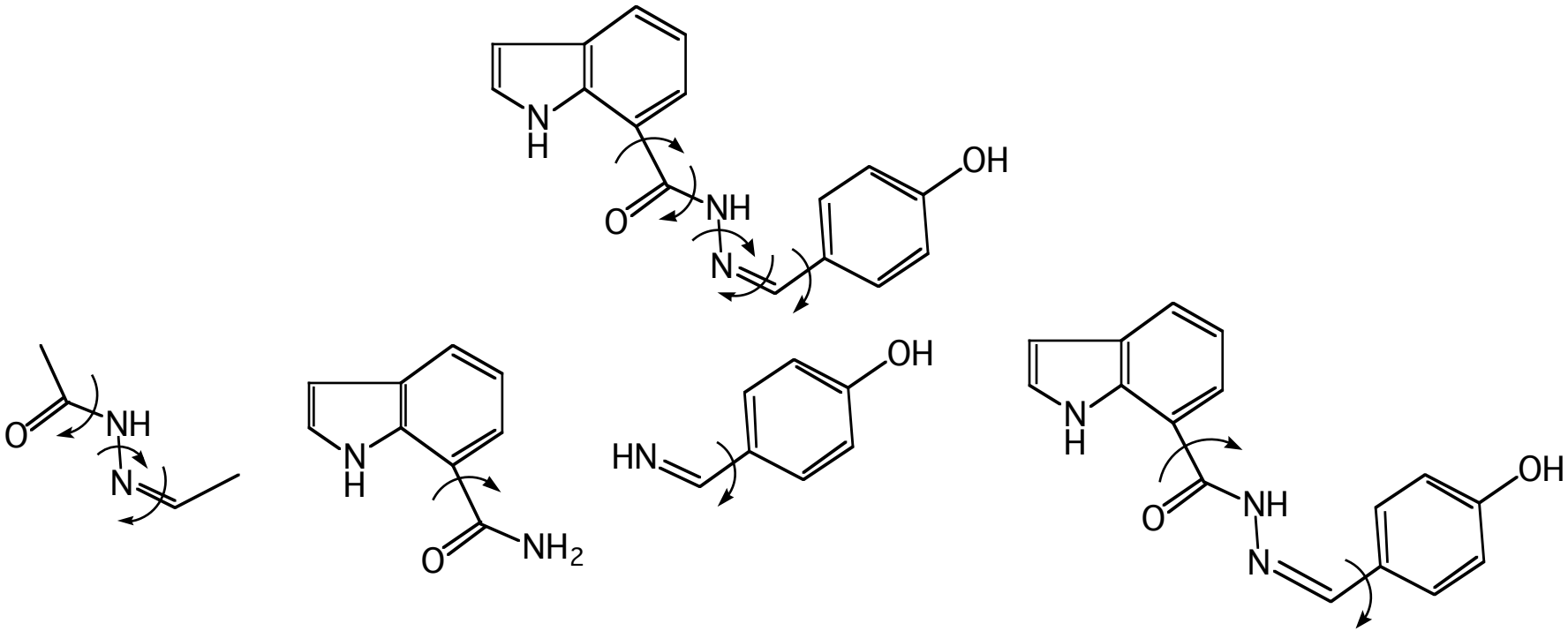
Comparison of the scaled ab initio, by analogy and optimized vibrations for selected modes

g98				Analogy				Optimized		
#	Freq	Assi	%	#	Freq	Assi	%	#	Freq	Assi
sN=C										
30	1782	sN=C	71	21	1228	sN=C	37	31	1802	sN=C
						rC5H	36			sN-N
				30	1646	sN=C	28			
						sC5-C	24			
						rC5H	18			
sN-N										
19	1234	sC5-C	33	20	1113	sN-N	53	20	1200	rNH
		sN-N	32			rNH	26			sN-N
20	1269	sN-N	36							rC5H
		rCH3'	18					23	1395	dCH3
										sN-N
								31	1802	sN=C
										sN-N
dC2NN										
4	154	dC2NN	44	5	207	dC2NN	36	4	158	dC2NN
		dN3NC	28			tN4C	31			dN3NC
		dN4CC	16							dN4CC
10	586	dC1CN	21	12	607	dC1CN	26	11	574	dC1CN
		dC2NN	20			dC2NN	25			dC2NN
		rC=O	18							dN4CC

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,

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

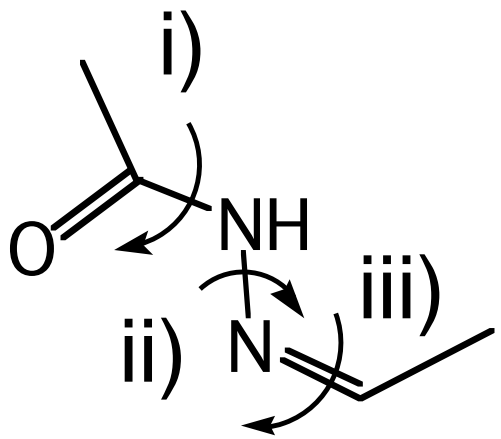
Final optimization of selected dihedrals (typically those containing only non-hydrogen atoms along a rotatable bond) are based on the reproduction of QM potential energy surfaces. This assures that both the relative energy and location of minima are correctly treated as are the barriers to rotation.



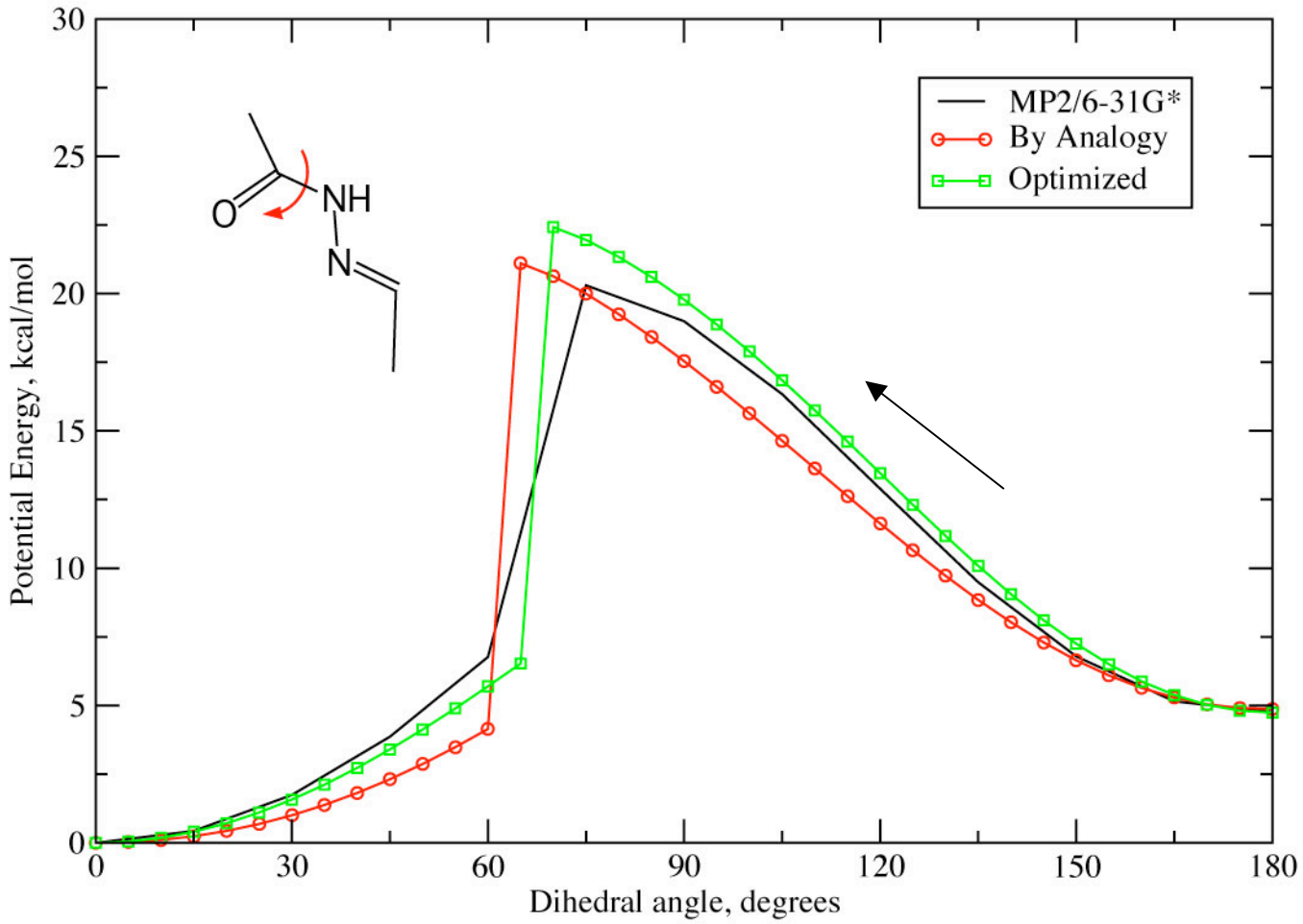
Note that additional model compounds may be required.

Potential energy surfaces on compounds with multiple rotatable bonds.

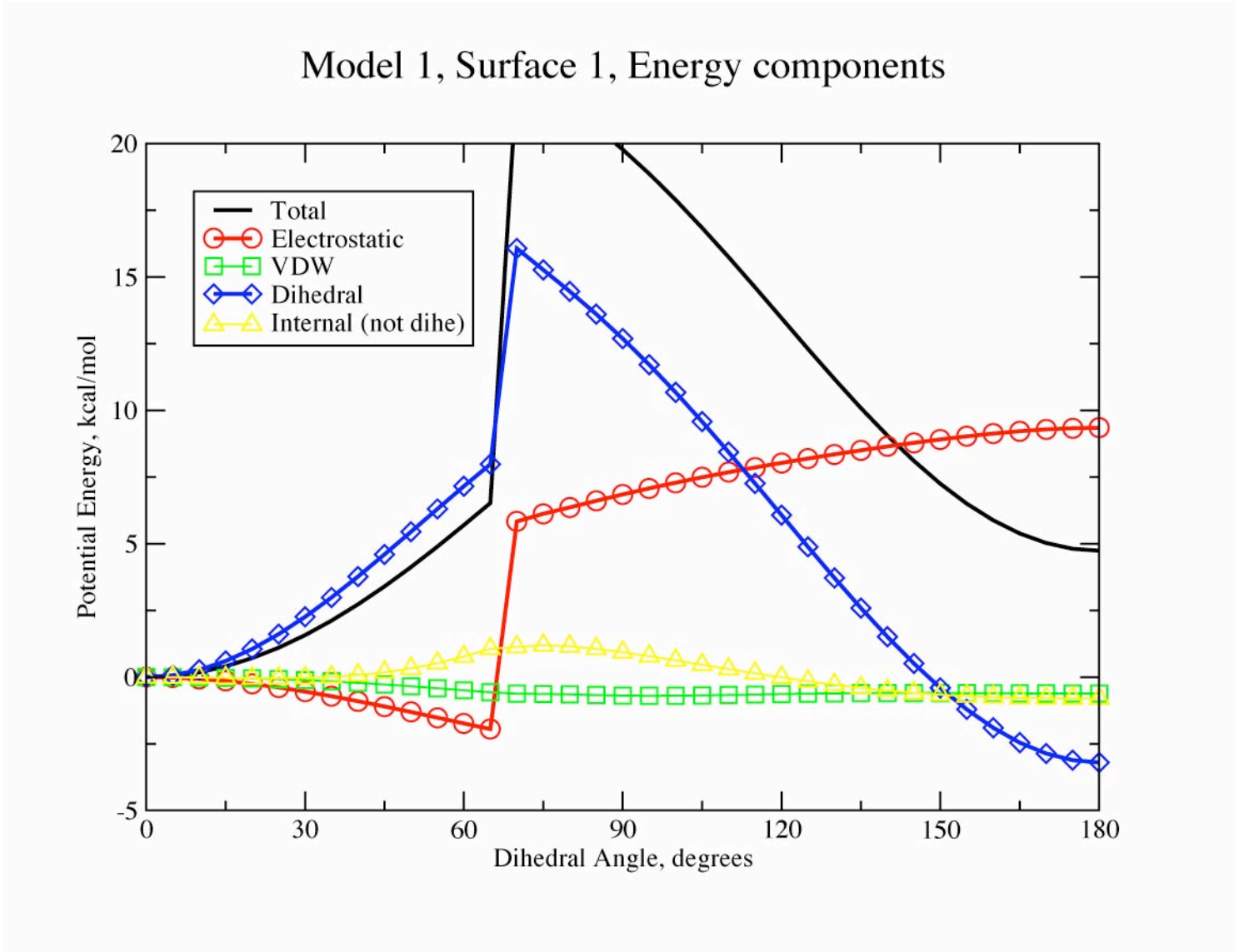
Run model_b_surf_all_one.inp followed by model_b_surf_all_two.inp to obtain energy surfaces



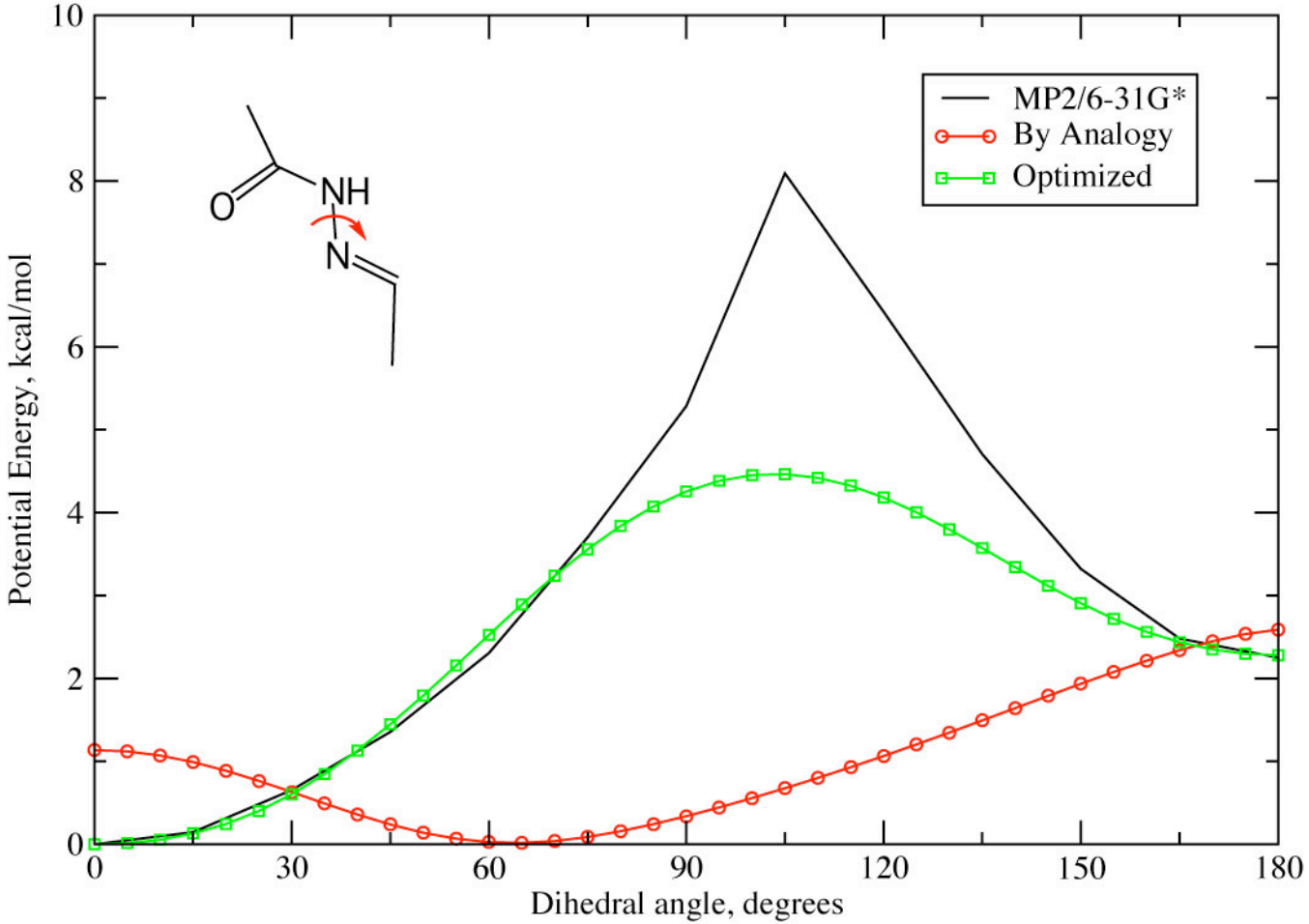
1) Full geometry optimization
2) Constrain n-1 dihedrals to minimum energy values or trans conformation
3) Sample selected dihedral surface
4) Repeat for all rotatable bonds
5) Repeat 2-4 using alternate minima if deemed necessary



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. This is the reason why parameter optimization is an iterative process as described above.

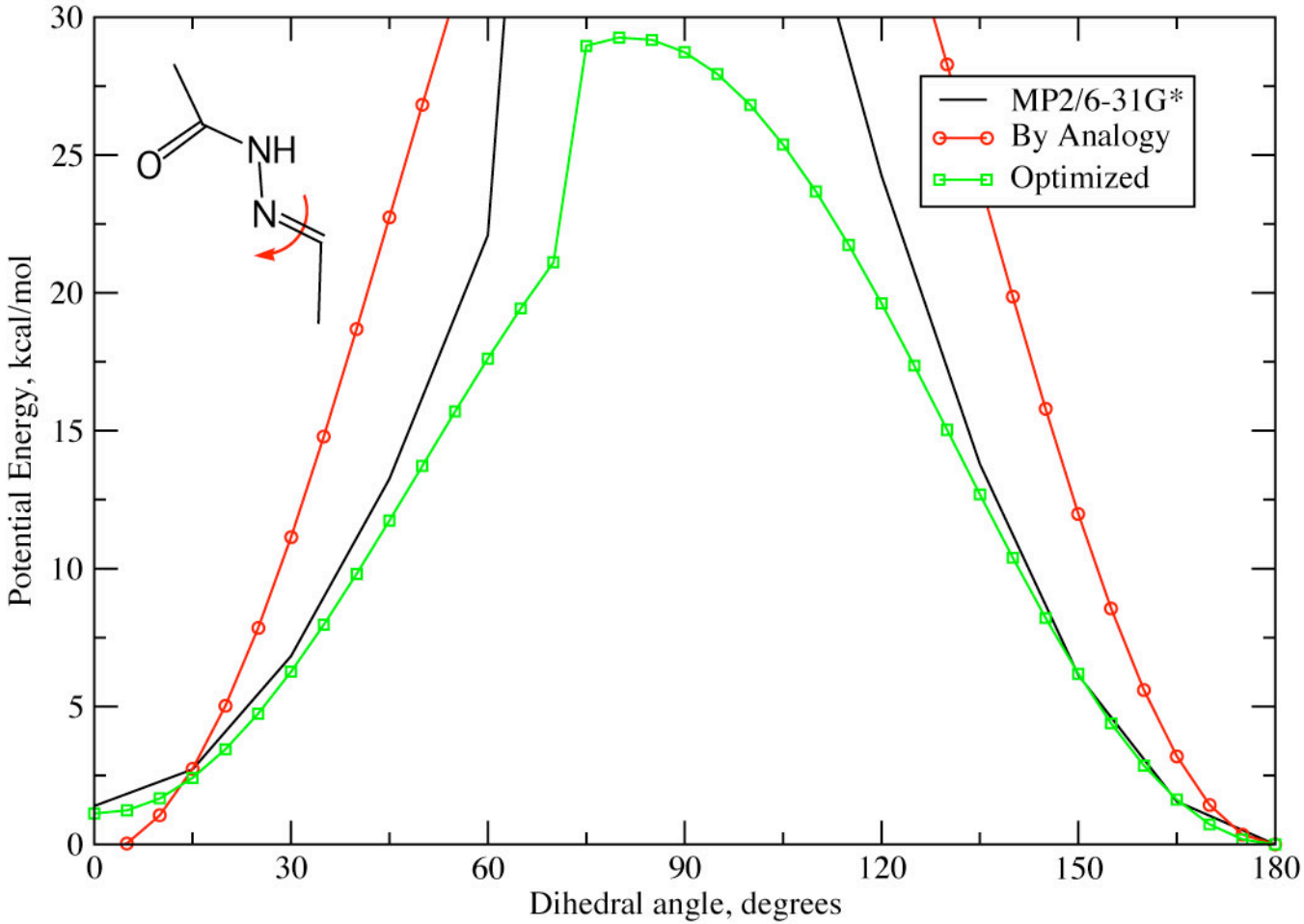


Model Compound 1, Surface 2



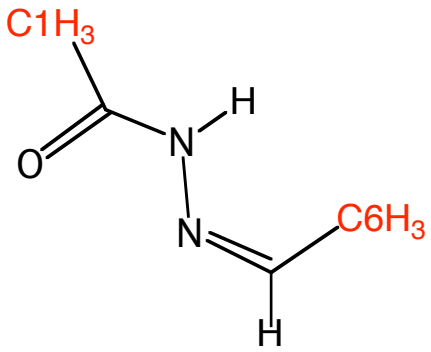
Note the emphasis on fitting the low energy region of the surface as this region is sampled in MD simulations. However, if studies are targeting rotation about that bonds this emphasis must be taken into account when interpreting results.

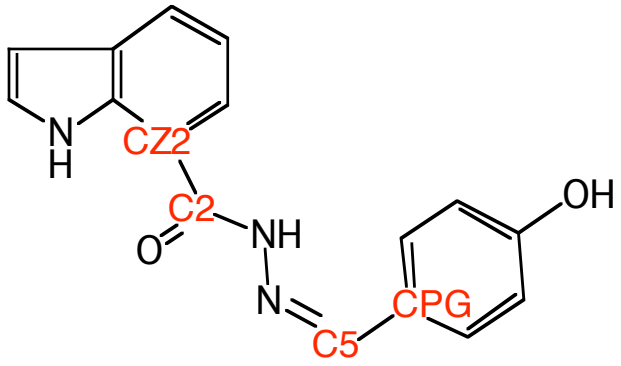
Model Compound 1, Surface 3



Creation of full drug compound

1) Rename phenol atom types to avoid conflicts with indole (add P to atom type)
2) Delete model 1 terminal methyls, indole and phenol HZ2 and HPG hydrogens, respectively, and perform charge adjustments
   i) Move HZ2 charge (0.115) into CZ2 (-0.115 -> 0.000) total charge on deleted C1 methyl (0.00) onto CZ2 (0.00 -> 0.00)
   ii) Move HPG charge (0.115) into CPG (-0.115 -> 0.000) 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!)





Check novel connectivities

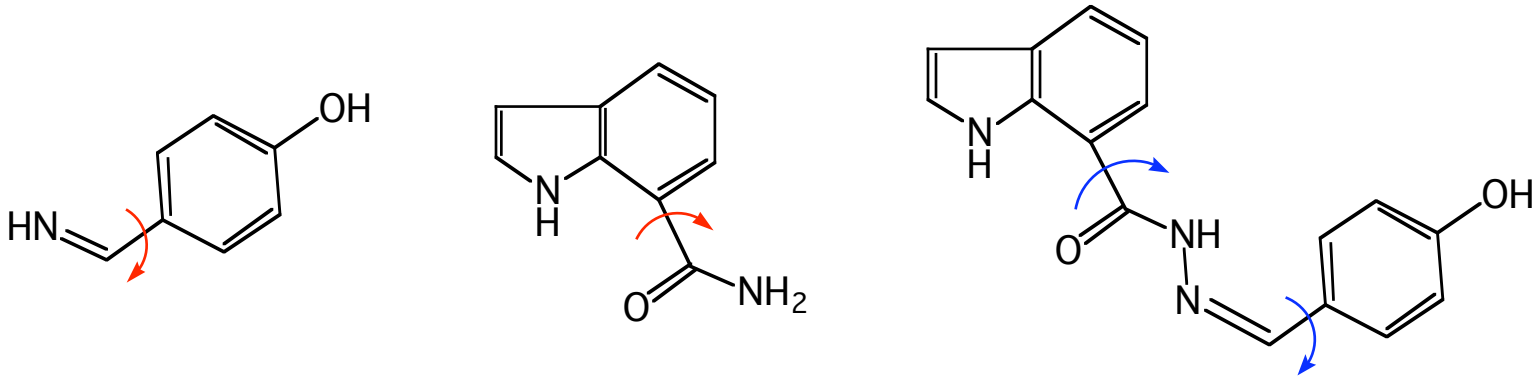
1) dihedrals !
2) bonds
3) angles

MP2/6-31G*

versus

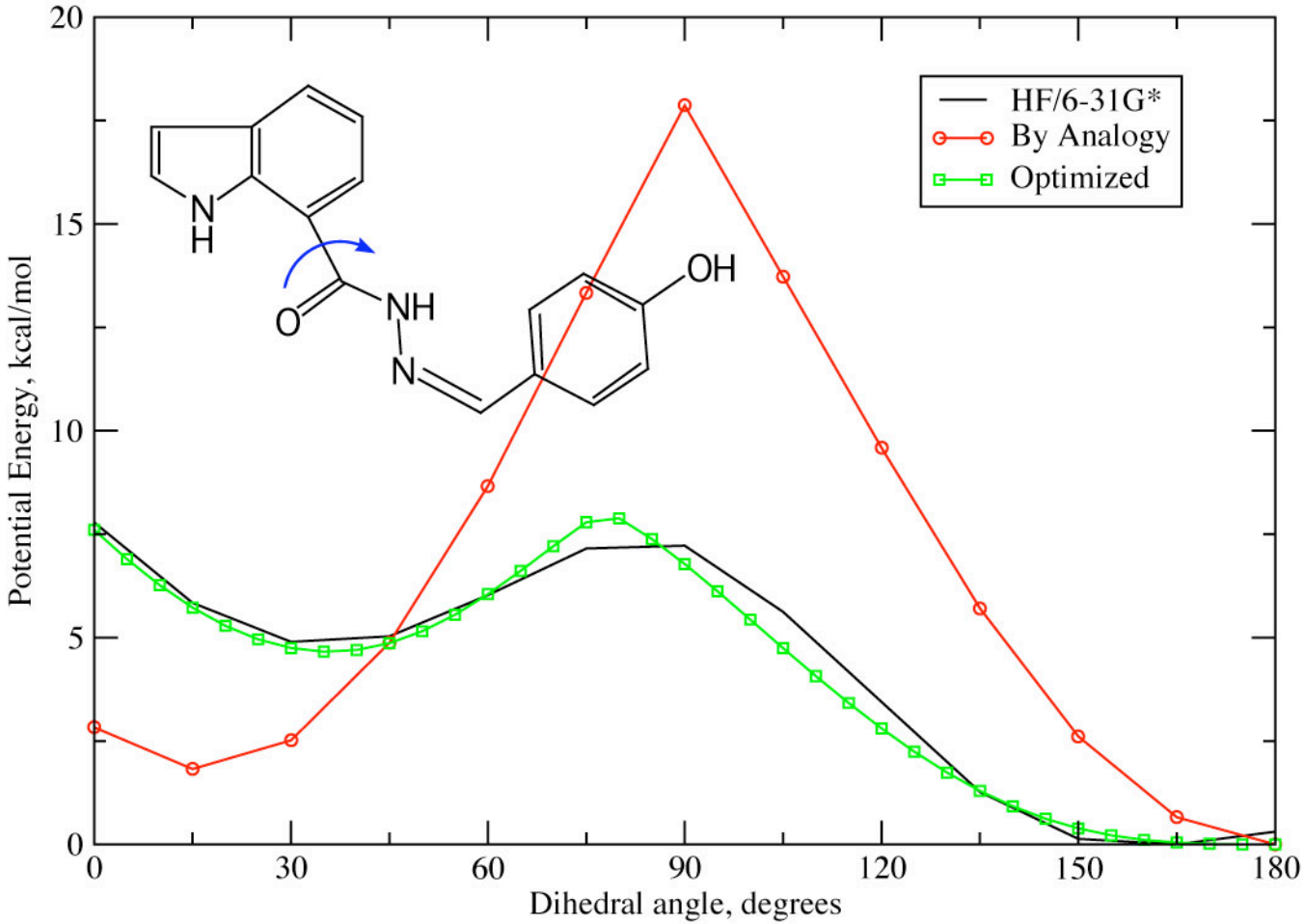
HF/6-31G*

MP2 data is preferable to HF for conformational energies; however, for a large compound doing MP2 calculations may not be feasible. Therefore, perform HF calculations and use the results as the target data; it will typically yield accurate location of minima while the barrier heights will be less reliable as will the relative energies of local minima. But, hey, its better than nothing!

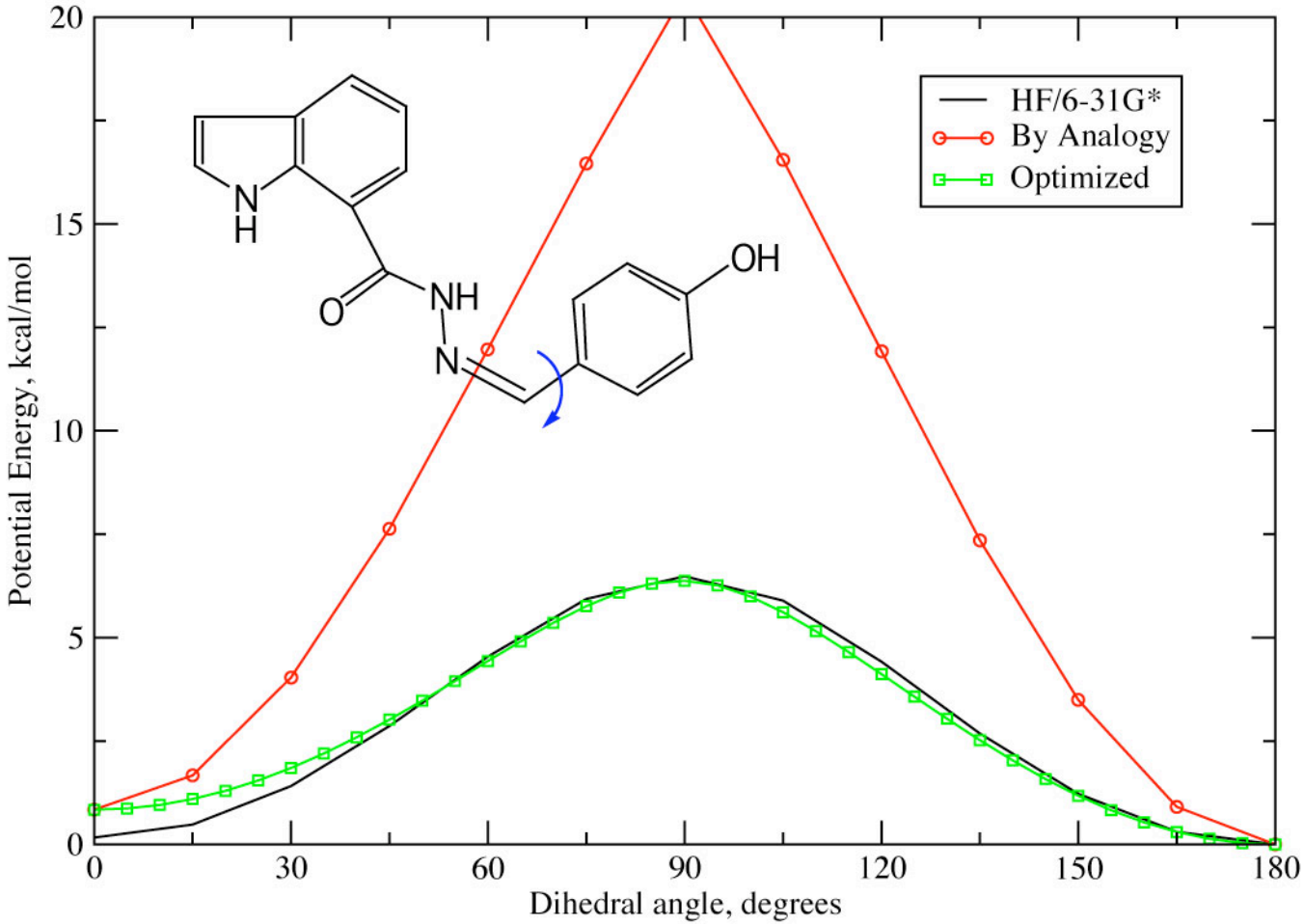


See model_2_surf_all.inp

Model Compound 2, Surface 1

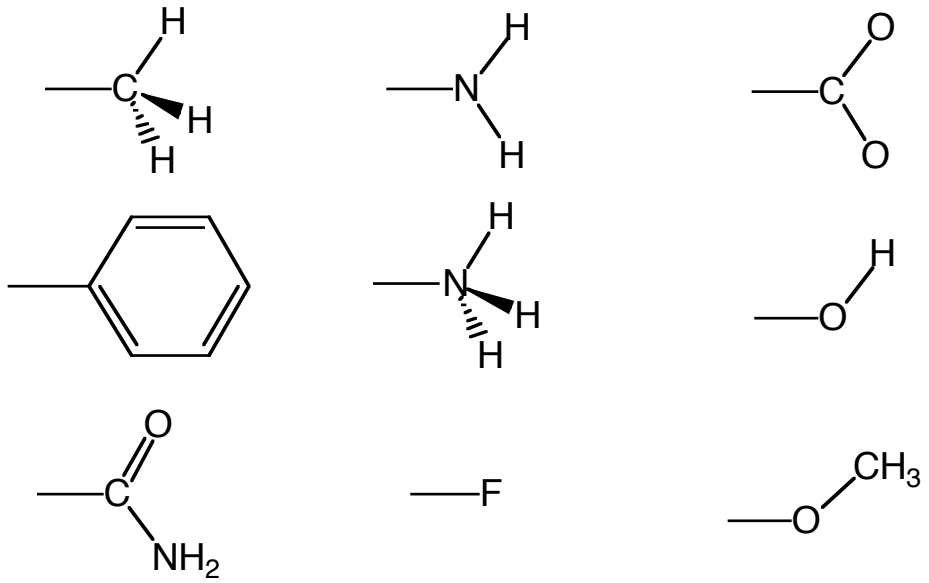


Model Compound 2, Surface 2



Lead Optimization

Addition of simple functional groups is generally straightforward once the full compound parameters have been optimized.



1) Delete appropriate hydrogens (i.e. at site of covalent bond)
2) Shift charge of deleted hydrogen into carbon being functionalized.
3) Add functional group
4) Offset charge on functionalized carbon to account for functional group charge requirements
    1) Aliphatics: just neutralize added functional group, q_H=0.09
    2) Phenol OH: q_C=0.11, q_O=-0.54, q_H=0.43
    3) Aliphatic OH: q_C=-0.04, q_O=-0.66, q_H=0.43
    4) Amino: q_C=0.16, q_{CH}=0.05, q_N=-0.30, q_H=0.33
    5) Carboxylate: q_C=-0.37, q_{CO}=-0.62, q_O=-0.76
5) Internal parameters should be present. Add by analogy if needed.
6) Optimize necessary parameters.

Perform above via the CHARMM PATCH (PRES) command

Summary

1) Junk in, junk out: Parameter optimization effort based on application requirements.
2) Follow standard protocol for the force field of interest (higher level QM is not necessarily better).
3) Careful parameter optimization of lead molecules
4) Simple substitutions often require minimal or no optimization.

Acknowledgements

MacKerell lab members
Igor Vorobyov
Deva Priyakumar
Shijun Zhong

NIH
DoD HPC
NPACI
PSC Terascale Computing

Class I potential energy function

$$V_T = \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} K_\phi \left[1 + \cos(n\phi - \delta)\right]$$
$$+ \sum_{1,3 pairs} K_{ub} (S - S_0)^2 + \sum_{improper} K_w (w - w_0)^2$$
$$+ \sum_{nonbonded} \varepsilon_{ij} \left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi D r_{ij}}$$

Amber
CHARMM
GROMOS
OPLS