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.

- QM and QM/MM force fields/potential energy descriptions used for molecular simulations.

The Potential Energy Function

$$U(\vec{R}) = \underbrace{\sum_{bonds} k_i^{bond}(r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2}_{U_{angle}} +$$

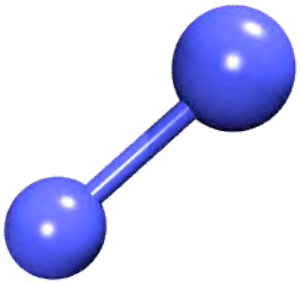
$$\underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i\phi_i + \delta_i)]}_{U_{dihedral}} +$$

$$\underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$

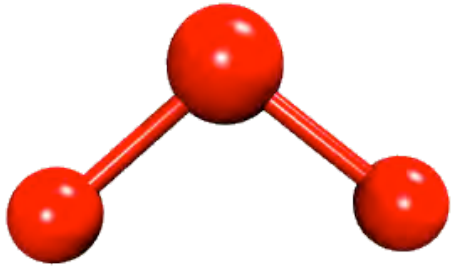
U_{bond} = oscillations about the equilibrium bond length
U_{angle} = oscillations of 3 atoms about an equilibrium bond angle
$U_{dihedral}$ = torsional rotation of 4 atoms about a central bond
$U_{nonbond}$ = non-bonded energy terms (electrostatics and Lenard-Jones)

Energy Terms Described in the CHARMm Force Field

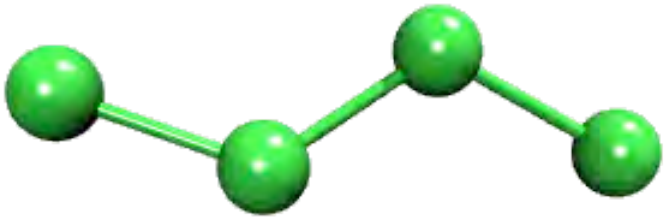
Bond



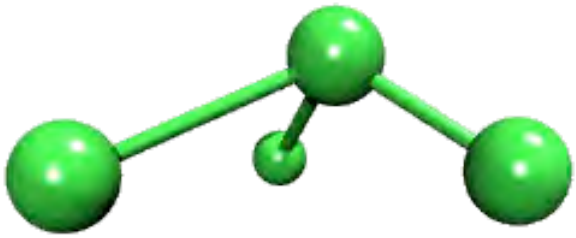
Angle



Dihedral



Improper



Classical Molecular Dynamics

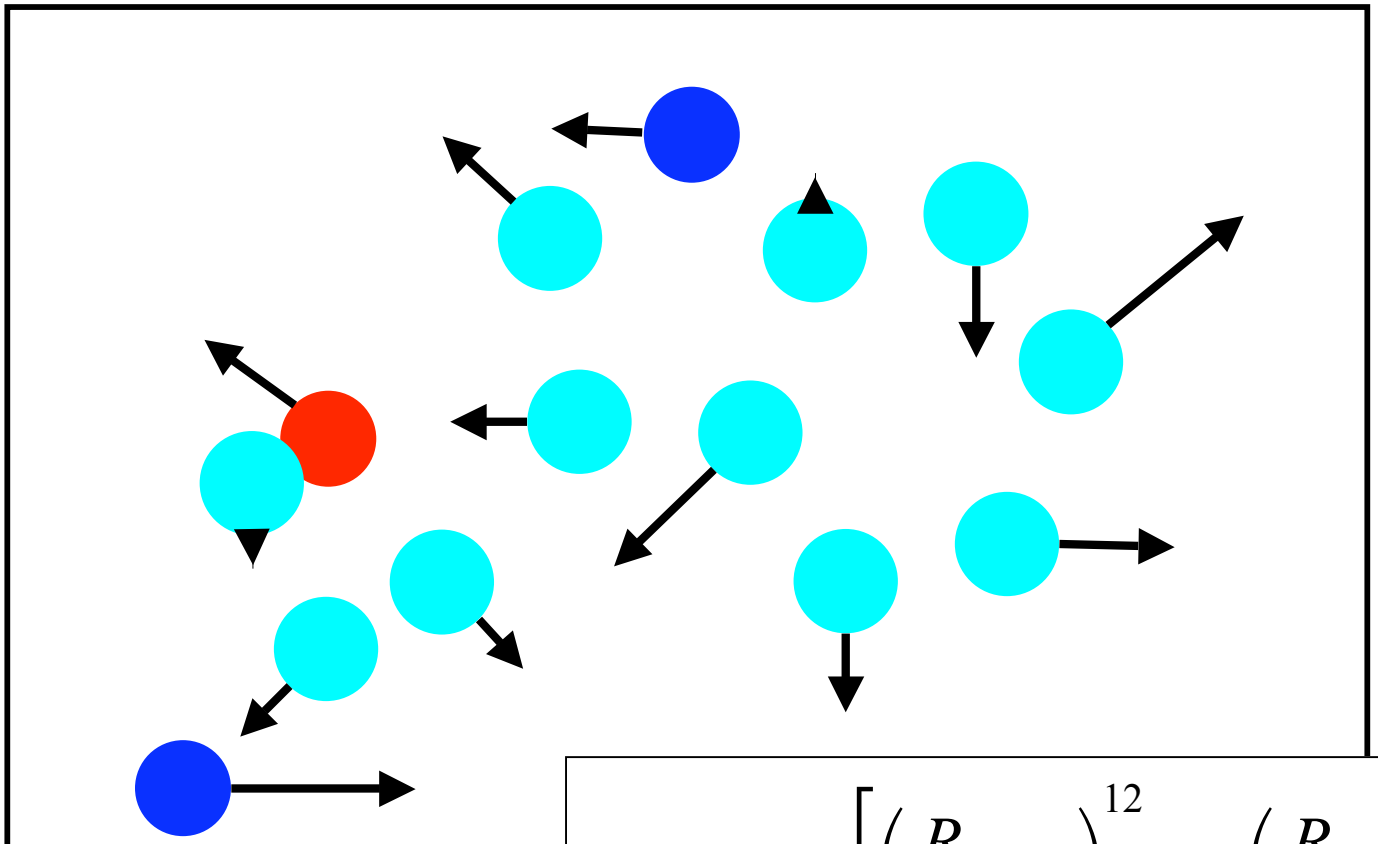
$$\boldsymbol{r}(t+\delta t) = \boldsymbol{r}(t) + \boldsymbol{v}(t)\delta t$$

$$\boldsymbol{v}(t+\delta t) = \boldsymbol{v}(t) + \boldsymbol{a}(t)\delta t$$

$$\boldsymbol{a}(t) = \boldsymbol{F}(t)/m$$

$$\boldsymbol{F} = -\frac{d}{d\boldsymbol{r}}U(\boldsymbol{r})$$

Classical Molecular Dynamics



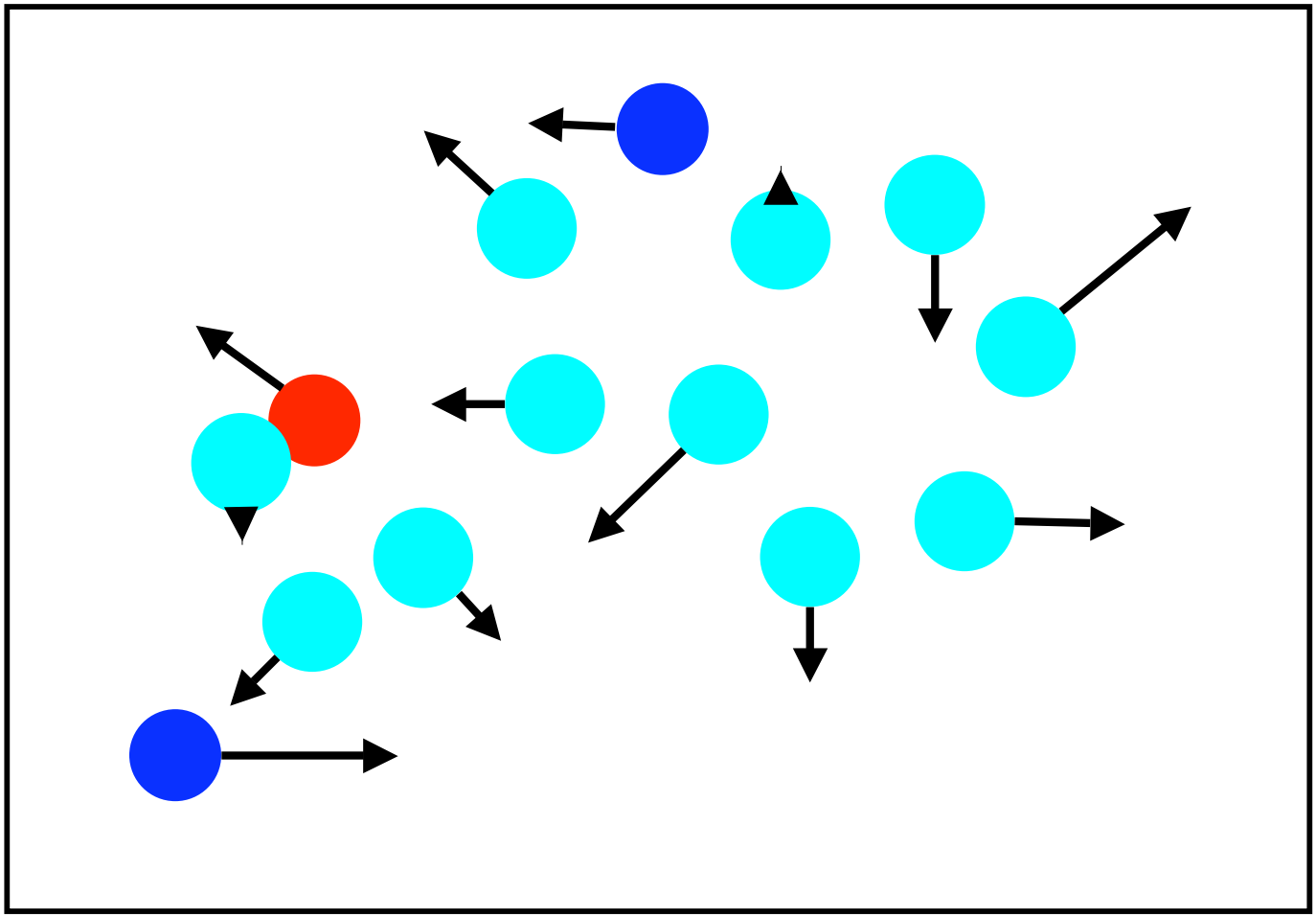
$$U(r) = \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}}$$

Coulomb interaction

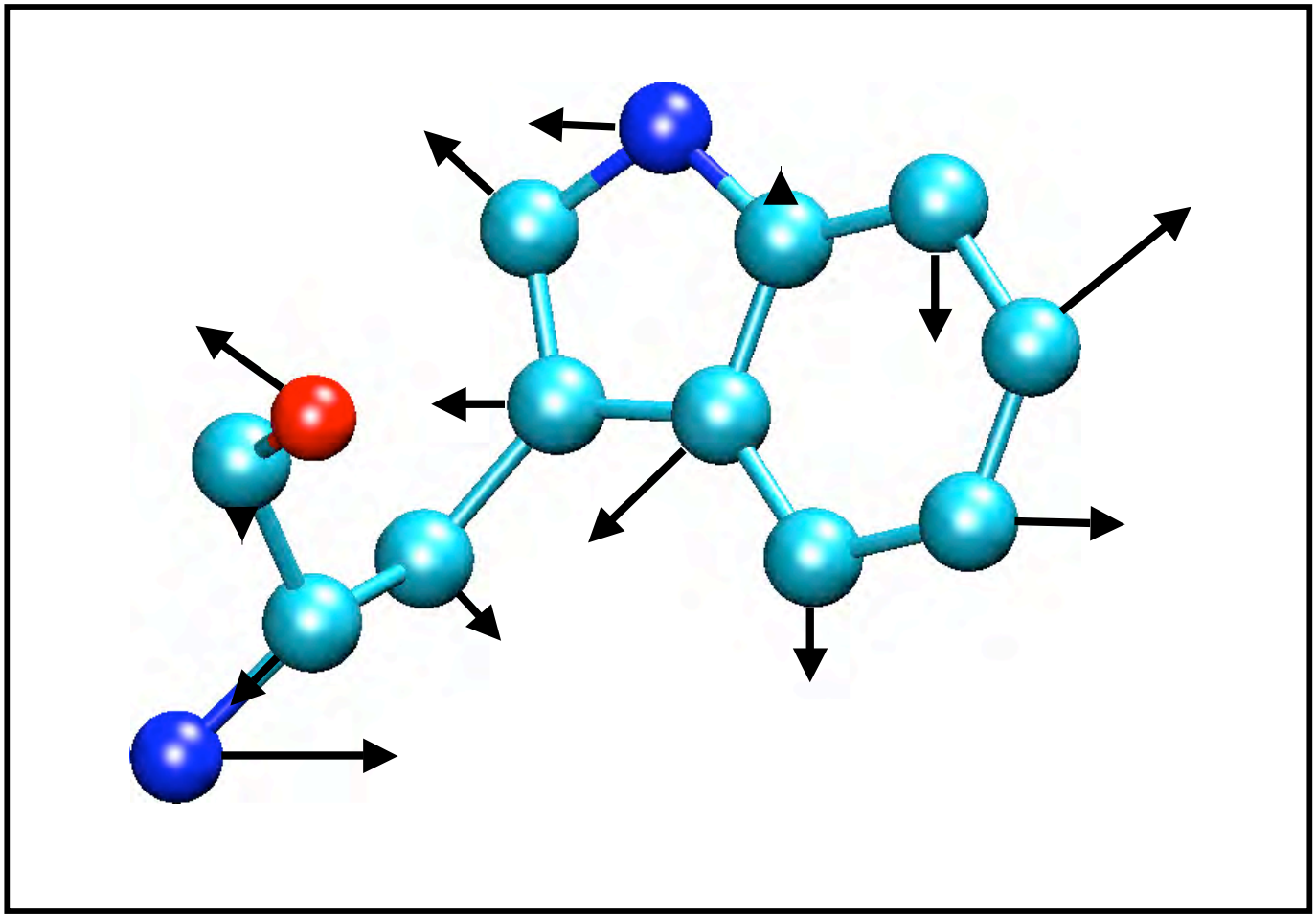
$$U(r) = \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]$$

van der Waals interaction

Classical Molecular Dynamics



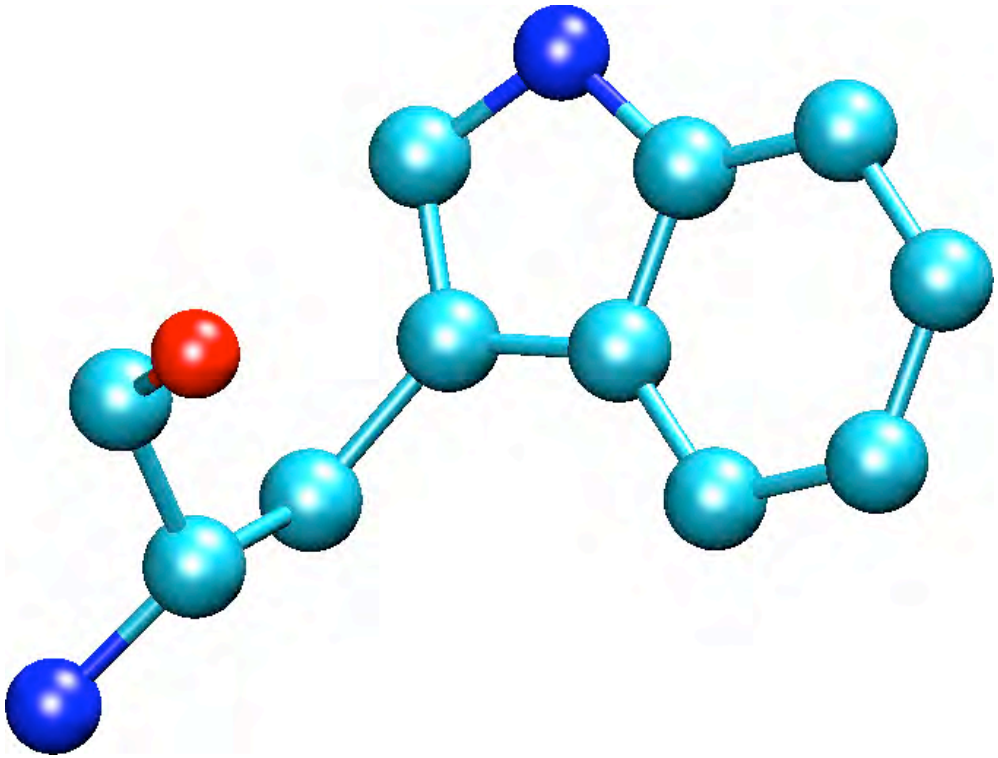
Classical Molecular Dynamics



Bond definitions, atom types, atom names, parameters, ….

What is a Force Field?

In molecular dynamics a molecule is described as a series of charged points (atoms) linked by springs (bonds).



To describe the time evolution of bond lengths, bond angles and torsions, also the non-bonding van der Waals and elecrostatic interactions between atoms, one uses a force field.
The force field is a collection of equations and associated constants designed to reproduce molecular geometry and selected properties of tested structures.

Energy Functions

$$U(\vec{R}) = \underbrace{\sum_{bonds} k_i^{bond}(r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2}_{U_{angle}} +$$

$$\underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i\phi_i + \delta_i)]}_{U_{dihedral}} +$$

$$\underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$

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

Parameter optimization of the CHARMM Force Field

Based on the protocol established by

Alexander D. MacKerell, Jr , U. Maryland

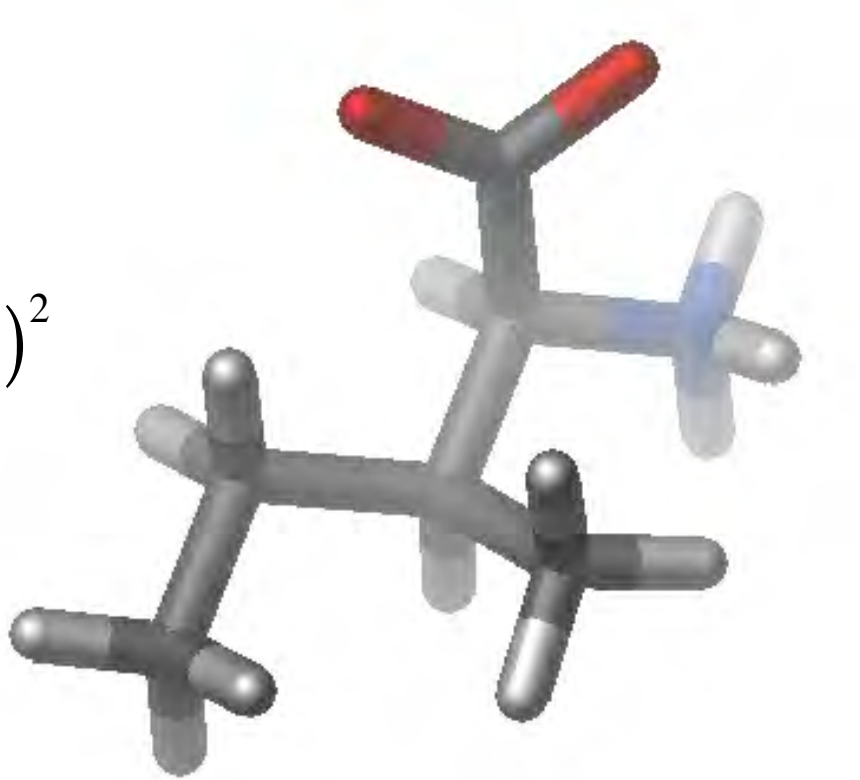
See references: www.pharmacy.umaryland.edu/faculty/amackere/force_fields.htm

Especially Sanibel Conference 2003, JCC v21, 86,105 (2000)

Interactions between bonded atoms

$$V_{angle} = K_\theta \left(\theta - \theta_o\right)^2$$

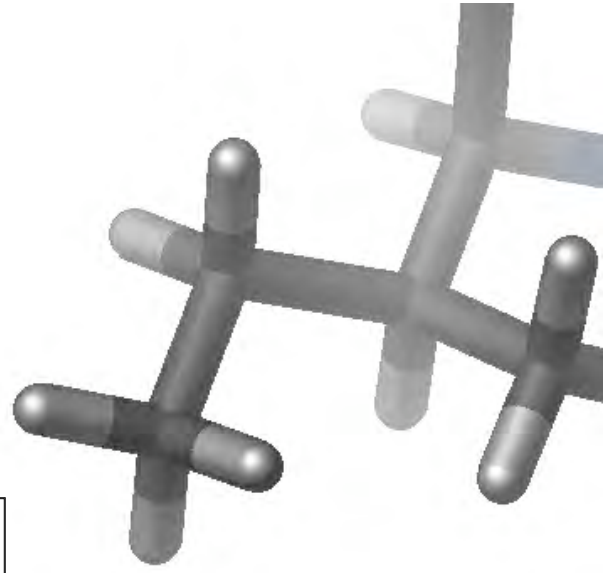
$$V_{bond} = K_b \left(b - b_o\right)^2$$

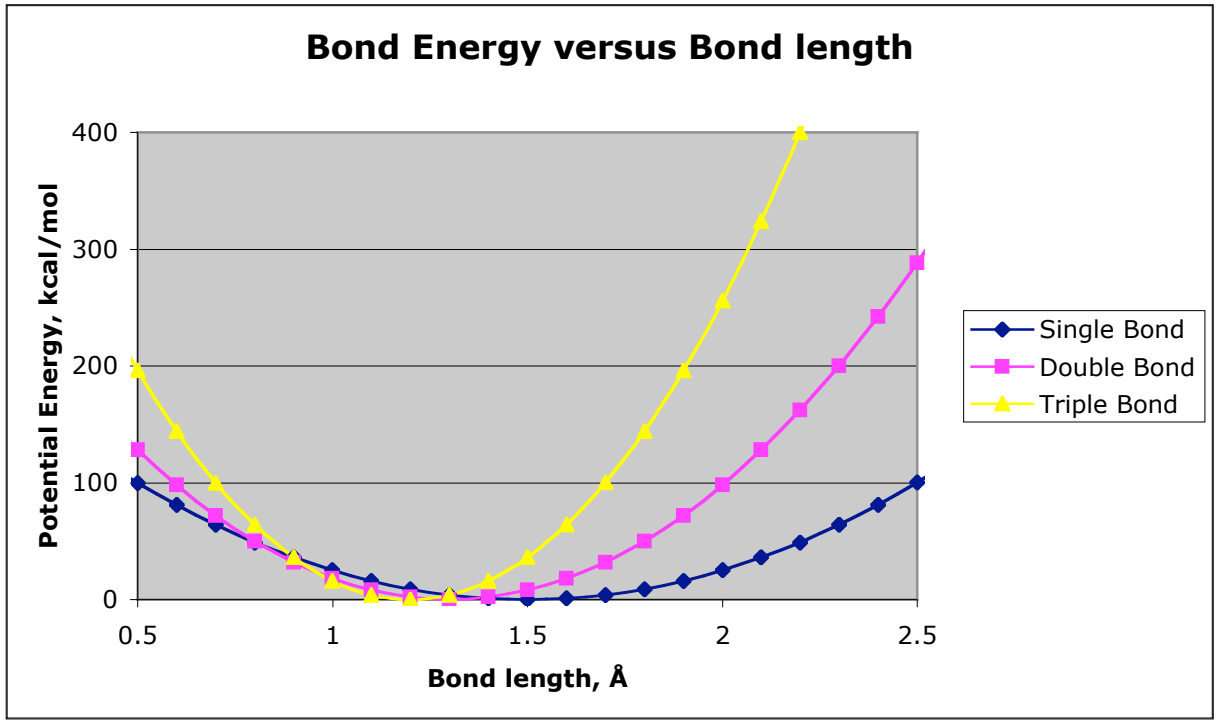


$$V_{dihedral} = K_\phi (1 + \cos(n\phi - \delta))$$

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

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

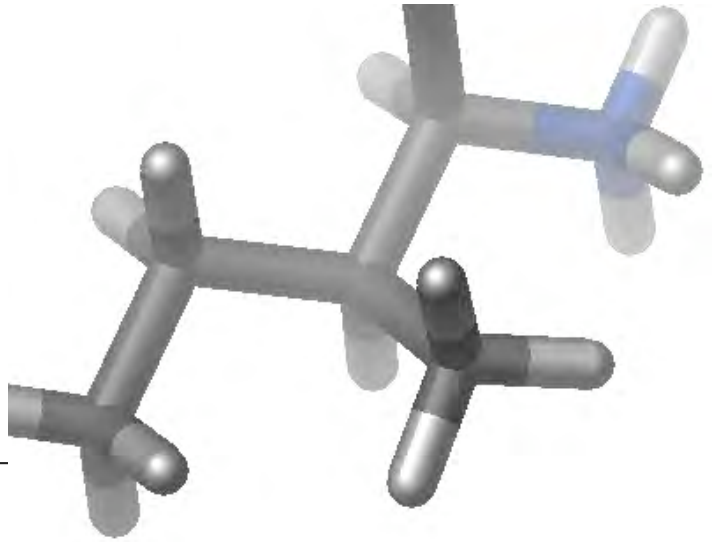


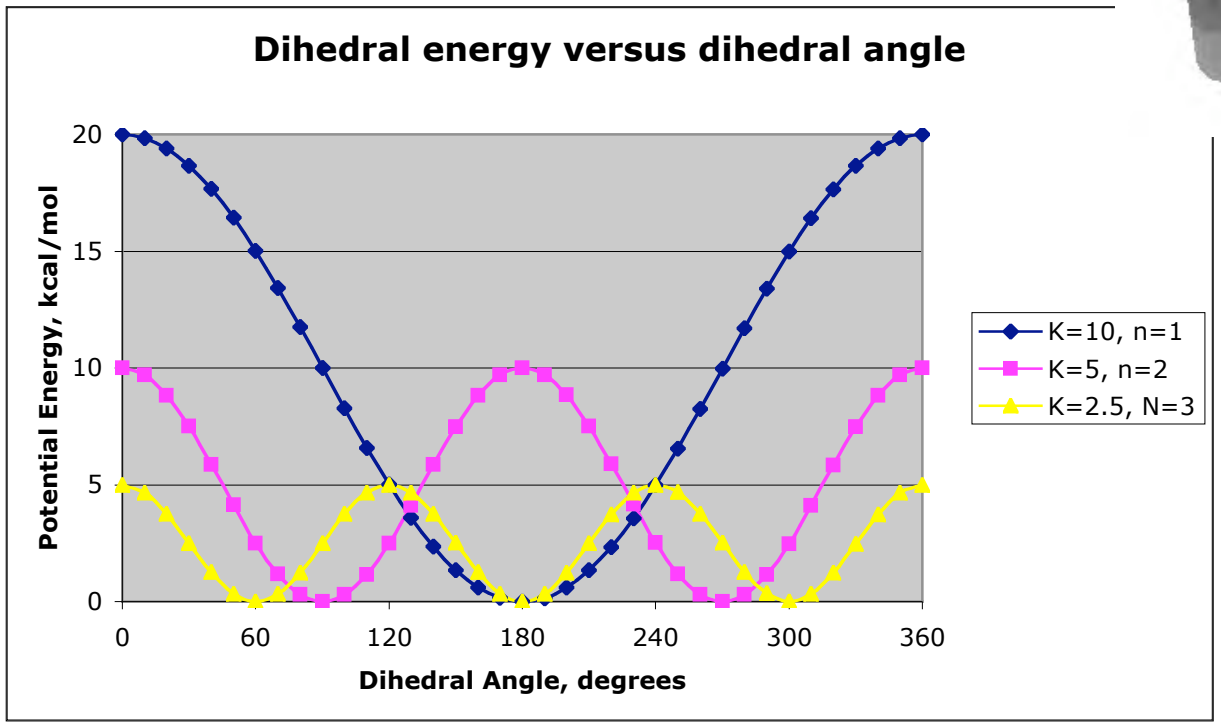


Bond angles and *improper* terms have similar quadratic forms, but with softer spring constants. The force constants can be obtained from vibrational analysis of the molecule (experimentally or theoretically).

Dihedral Potential

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





$\delta = 0^{\circ}$

Nonbonded Parameters

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

q_i: partial atomic charge

D: dielectric constant

ε: Lennard-Jones (LJ, vdW) well-depth

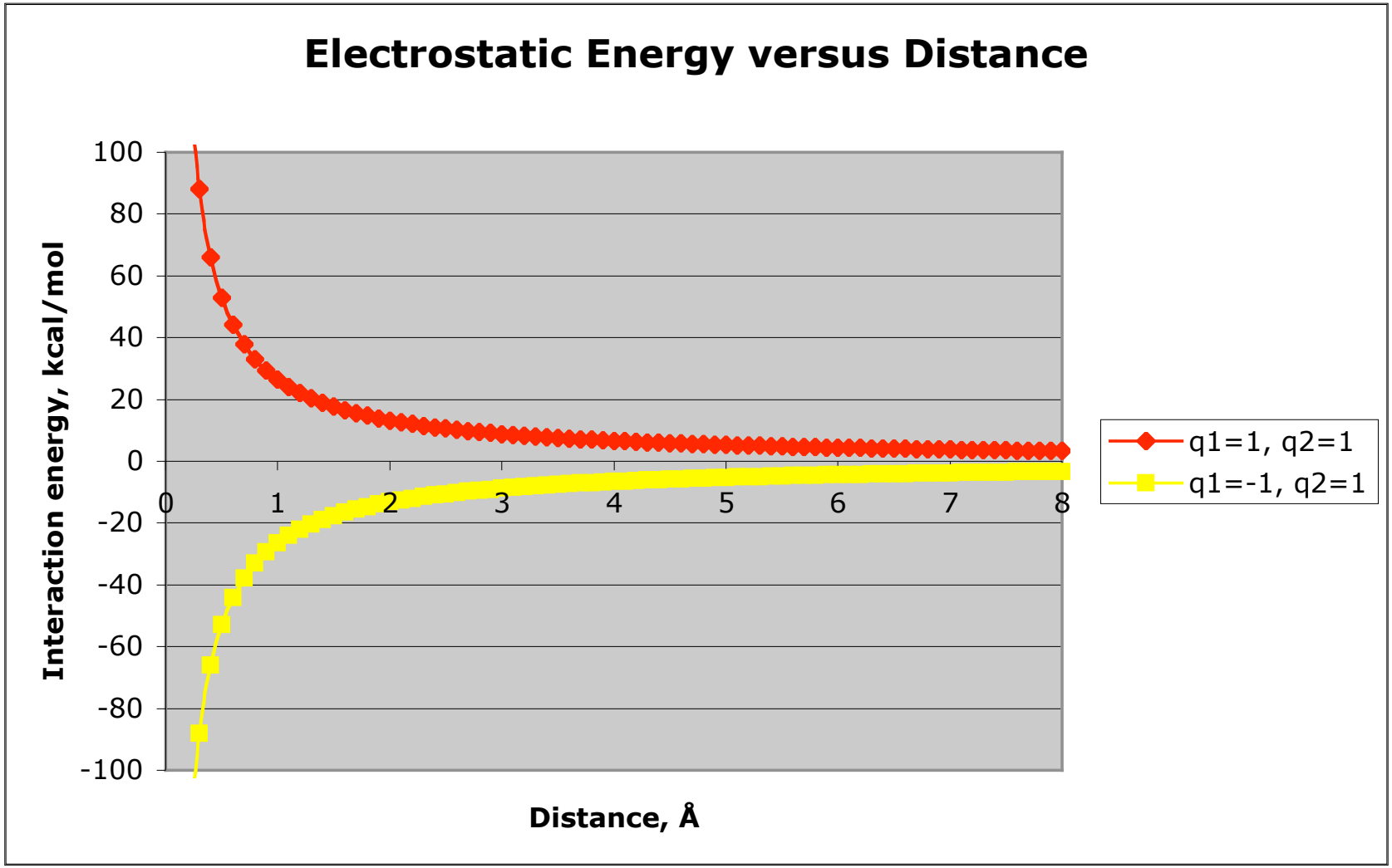
R_{min}: LJ radius ($R_{min}/2$ in CHARMM)

Combining rules (CHARMM, Amber)

$R_{min\ i,j} = R_{min\ i} + R_{min\ j}$

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

From MacKerell



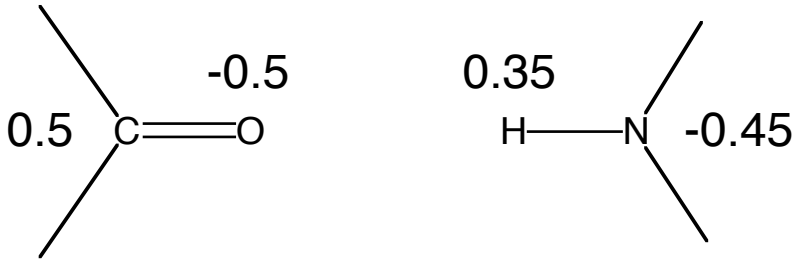
Note that the effect is long range.

From MacKerell

Charge Fitting Strategy

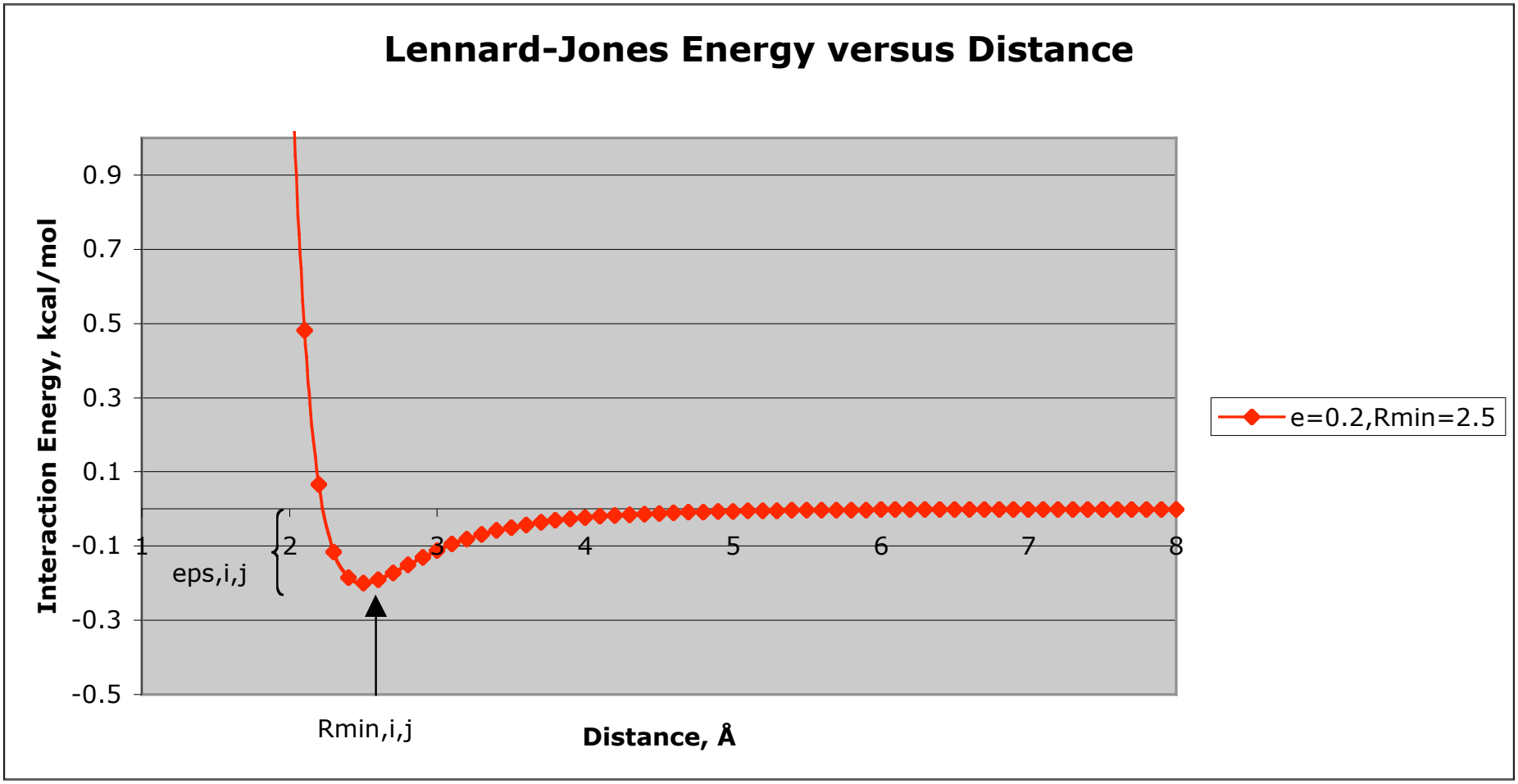
CHARMM- Mulliken* AMBER(ESP/RESP)

Partial atomic charges



*Modifications based on interactions with TIP3 water

van der Waals interaction

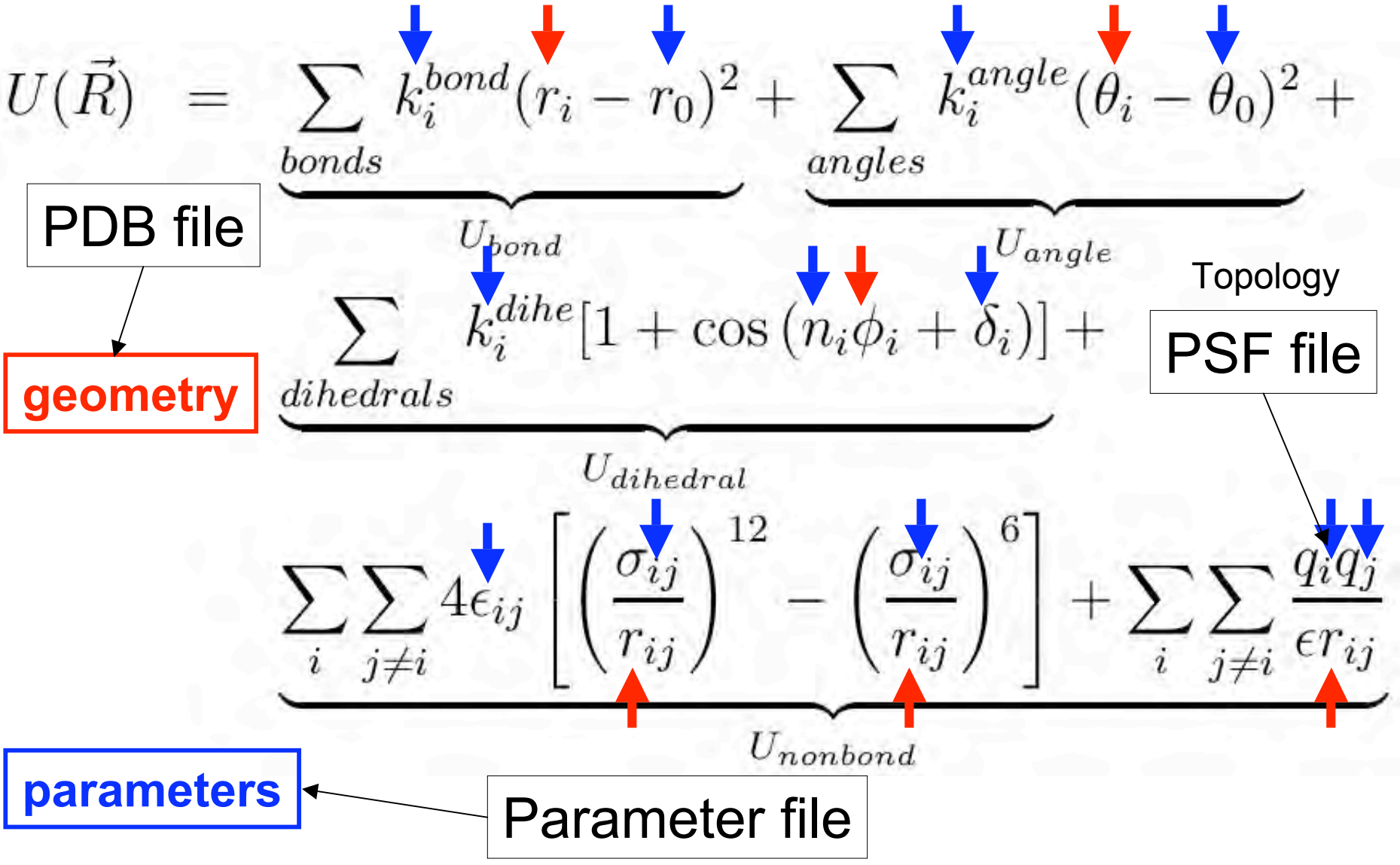


$$\varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]$$

Short range

From MacKerell

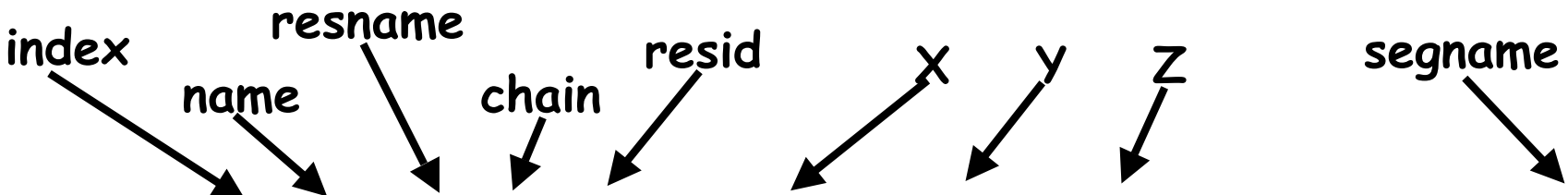
CHARMM Potential Function



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

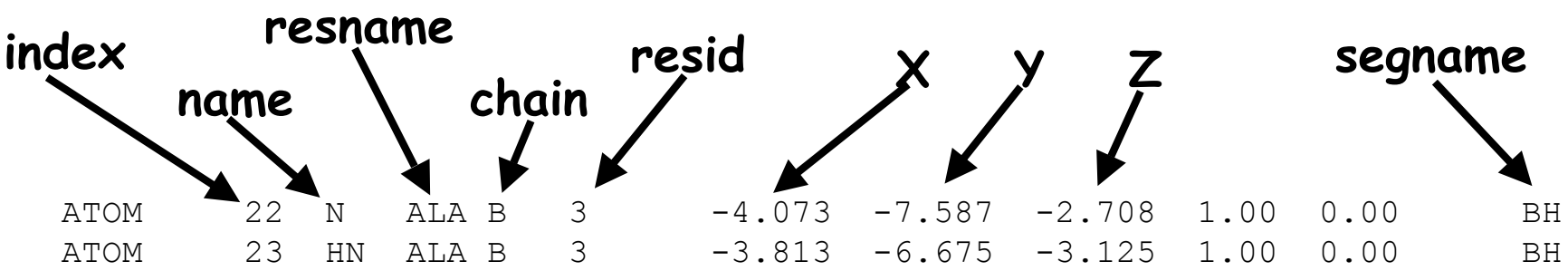



```
ATOM     22  N   ALA B   3      -4.073  -7.587  -2.708  1.00  0.00      BH
ATOM     23  HN  ALA B   3      -3.813  -6.675  -3.125  1.00  0.00      BH
ATOM     24  CA  ALA B   3      -4.615  -7.557  -1.309  1.00  0.00      BH
ATOM     25  HA  ALA B   3      -4.323  -8.453  -0.704  1.00  0.00      BH
ATOM     26  CB  ALA B   3      -4.137  -6.277  -0.676  1.00  0.00      BH
ATOM     27  HB1 ALA B   3      -3.128  -5.950  -0.907  1.00  0.00      BH
ATOM     28  HB2 ALA B   3      -4.724  -5.439  -1.015  1.00  0.00      BH
ATOM     29  HB3 ALA B   3      -4.360  -6.338   0.393  1.00  0.00      BH
ATOM     30  C   ALA B   3      -6.187  -7.538  -1.357  1.00  0.00      BH
ATOM     31  O   ALA B   3      -6.854  -6.553  -1.264  1.00  0.00      BH
ATOM     32  N   ALA B   4      -6.697  -8.715  -1.643  1.00  0.00      BH
ATOM     33  HN  ALA B   4      -6.023  -9.463  -1.751  1.00  0.00      BH
ATOM     34  CA  ALA B   4      -8.105  -9.096  -1.934  1.00  0.00      BH
ATOM     35  HA  ALA B   4      -8.287  -8.878  -3.003  1.00  0.00      BH
ATOM     36  CB  ALA B   4      -8.214 -10.604  -1.704  1.00  0.00      BH
ATOM     37  HB1 ALA B   4      -7.493 -11.205  -2.379  1.00  0.00      BH
ATOM     38  HB2 ALA B   4      -8.016 -10.861  -0.665  1.00  0.00      BH
ATOM     39  HB3 ALA B   4      -9.245 -10.914  -1.986  1.00  0.00      BH
ATOM     40  C   ALA B   4      -9.226  -8.438  -1.091  1.00  0.00      BH
ATOM     41  O   ALA B   4     -10.207  -7.958  -1.667  1.00  0.00      BH
0000000000000000000000000000000000000000000000000000000000000000000000000000
        10        20        30        40        50        60        70
```

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

"No connectivity information"

VMD Atom Selection Commands



(name CA CB) and (resid 1 to 4) and (segname BH)

protein and resname LYS ARG GLU ASP

water and within 5 of (protein and resid 62 and name CA)

water and within 3 of (protein and name O and z < 10)

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 - dihedrals??

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

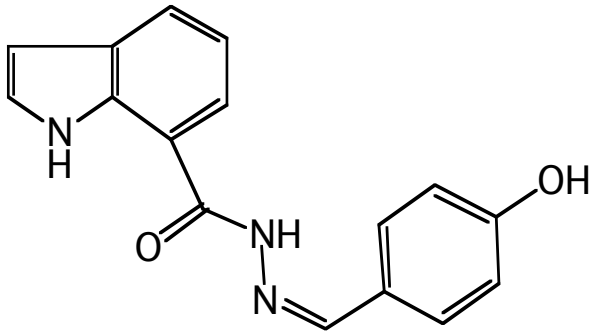
CHARMM topology (parameter files)

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)

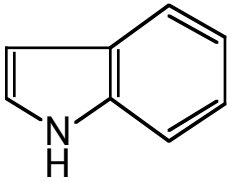
NA and lipid force fields have new LJ parameters for the alkanes, representing increased optimization of the protein alkane parameters. Tests have shown that these are compatible (e.g. in protein-nucleic acid simulations). For new systems is suggested that the new LJ parameters be used. Note that only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

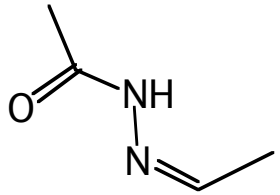


A



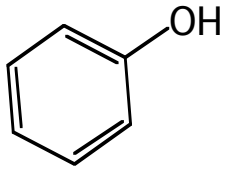
Indole

B



Hydrazine

C



Phenol

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

```
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     0.110
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
```

From MacKerell

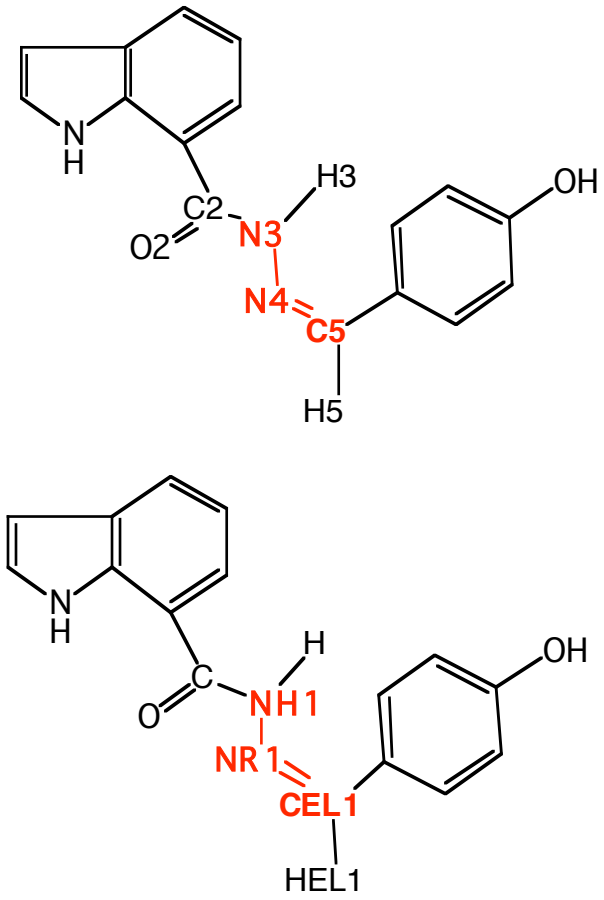
Top_all22_model.inp contains all protein model compounds. Lipid, nucleic acid and carbohydate model compounds are in the full topology files.

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

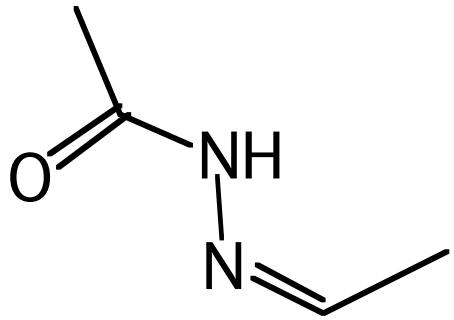
Comparison of atom names (upper) and atom types (lower)



From MacKerell

Creation of topology for central model compound

```
RESI  Mod1 ! Model compound 1
Group
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 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

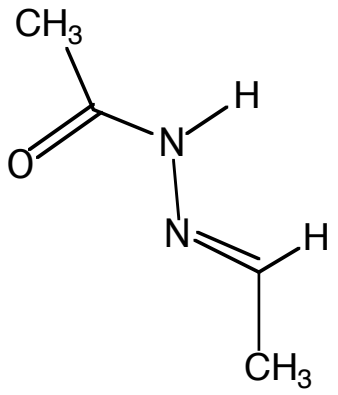
Method Dependent Choices

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.

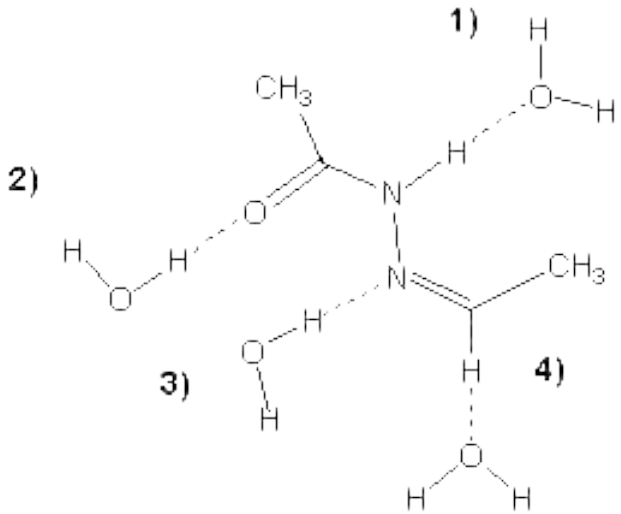
From MacKerell



Starting charges??
Mulliken population analysis
Analogy comparison

peptide bond
methyl
imidazole (N-N=C)?

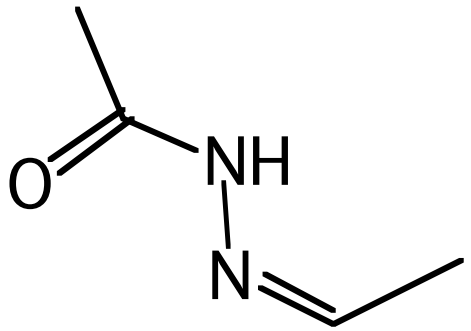
Final charges (methyl, vary q_C to maintain integer charge, $q_H = 0.09$)
interactions with water (HF/6-31G*, monohydrates!)



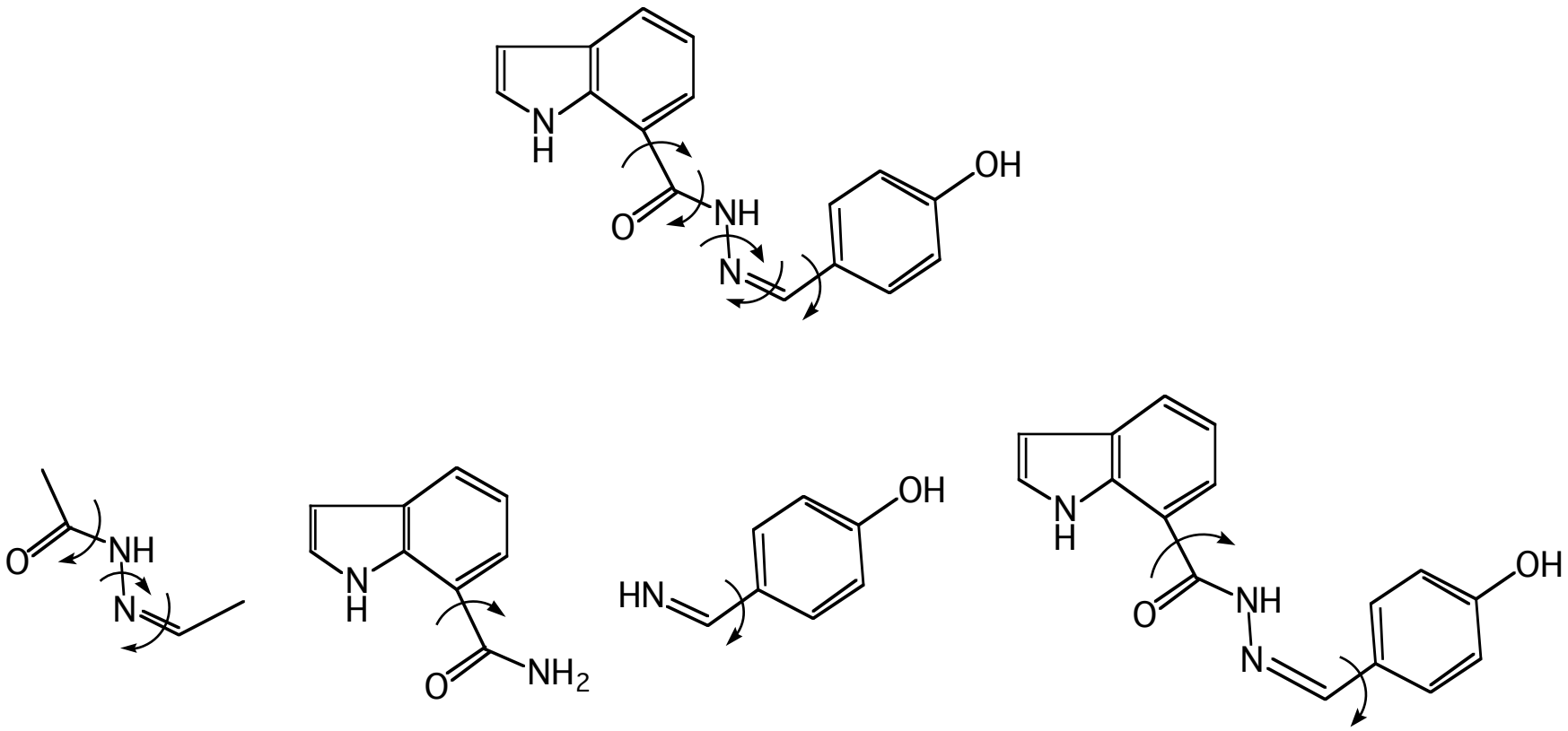
From MacKerell

Comparison of analogy and optimized charges

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

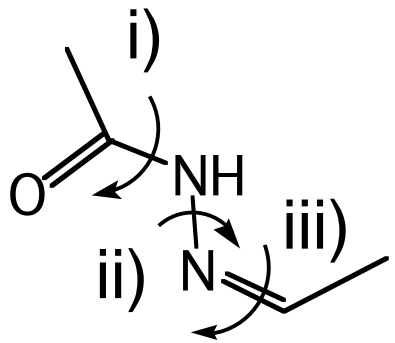


Dihedral optimization based on QM potential energy surfaces (HF/6-31G* or MP2/6-31G*).



From MacKerell

Potential energy surfaces on compounds with multiple rotatable bonds



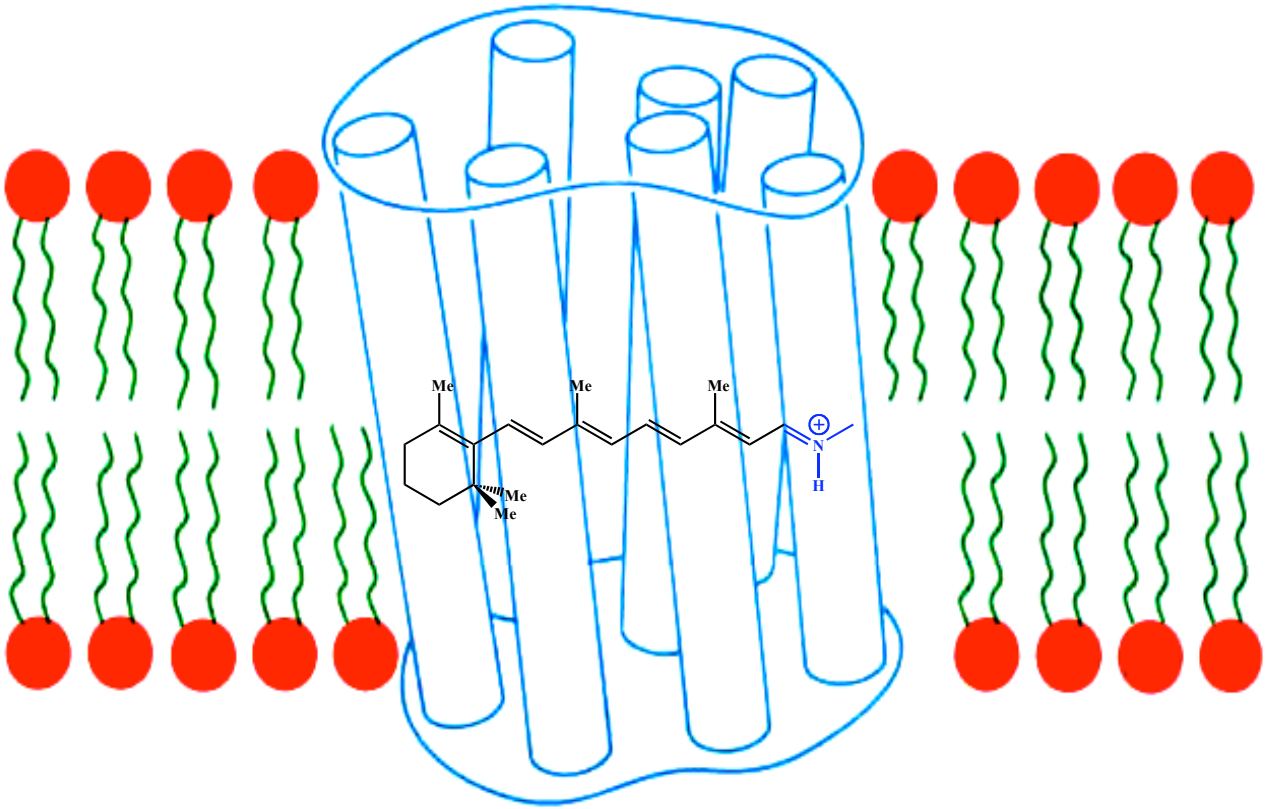
1) Full geometry optimization
2) Constrain n-1 dihedrals to minimum energy values or trans conformation
3) Sample selected dihedral surface
4) Repeat for all rotatable bonds dihedrals
5) Repeat 2-5 using alternate minima if deemed appropriate

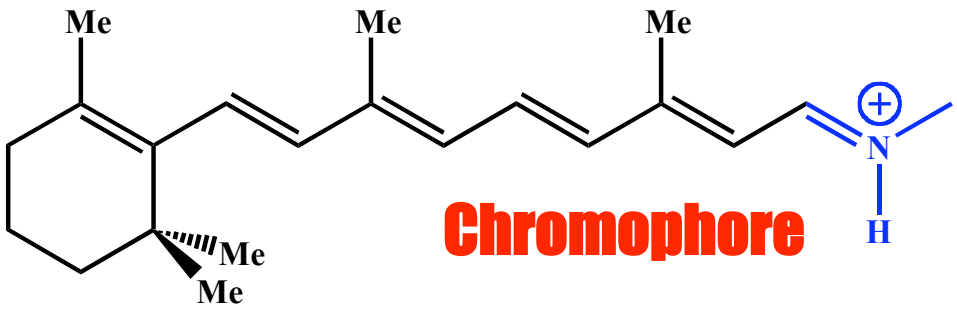
From MacKerell

QM development of force field parameters for retinal

Used for rhodopsin and bacteriorhodopsin simulations

Retinal Proteins -- Rhodopsins

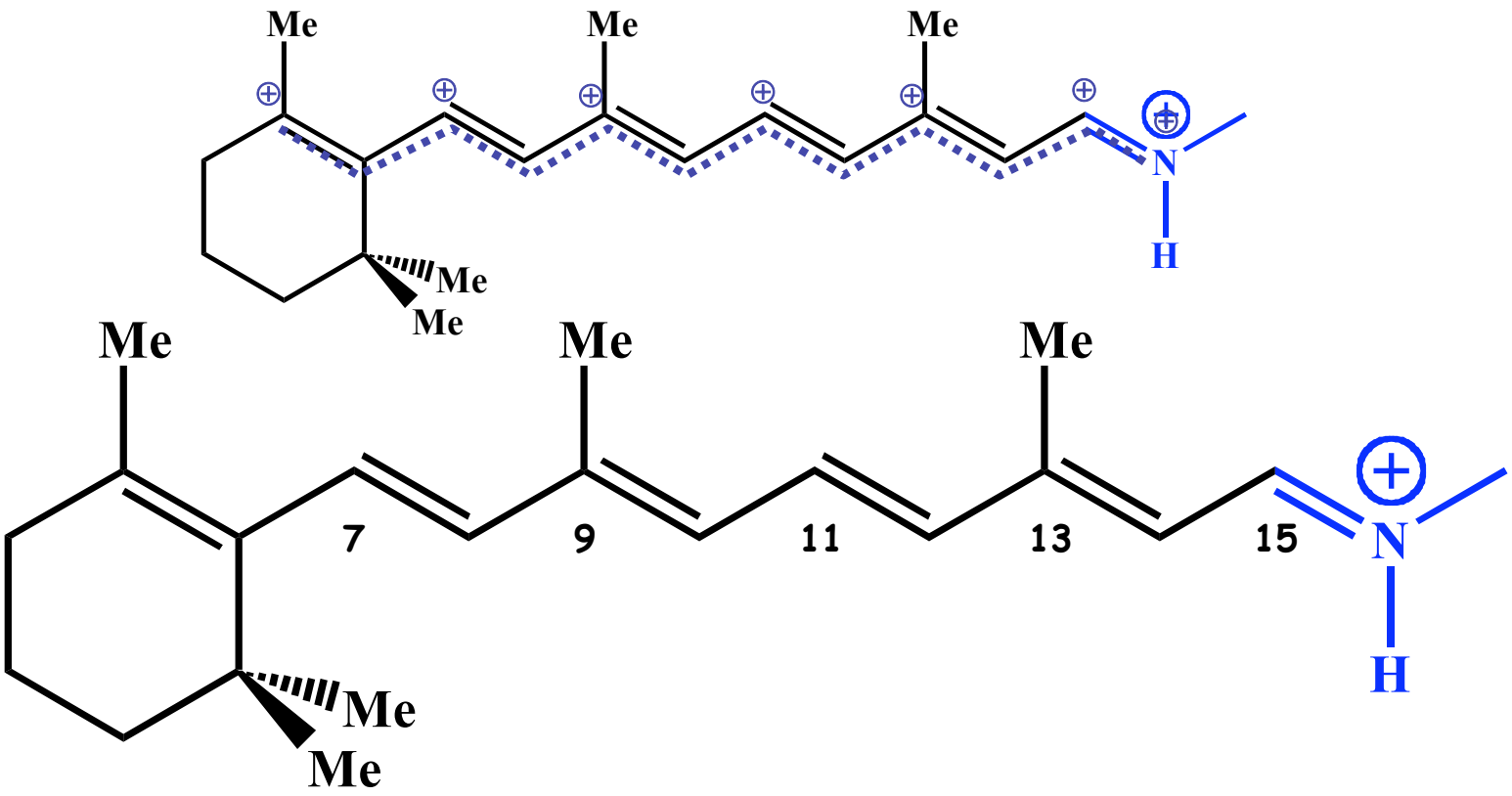




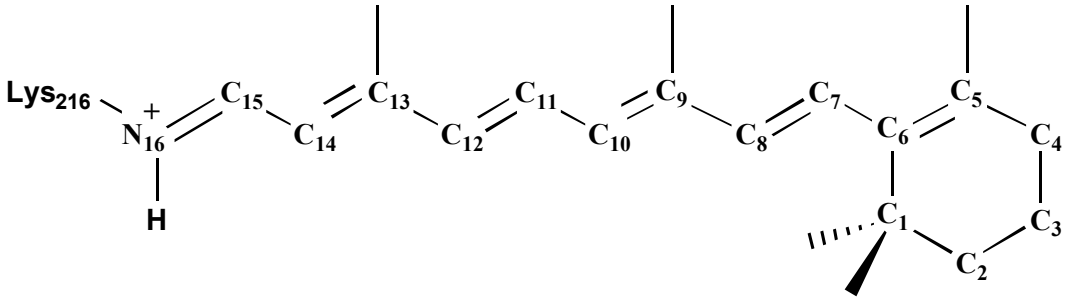
Chromophore

- Covalently linked to a lysine
- Usually protonated Schiff base
- all-*trans* and 11-*cis* isomers

Unconventional chemistry



Isomerization Barriers in retinal



DFT/6-31G**

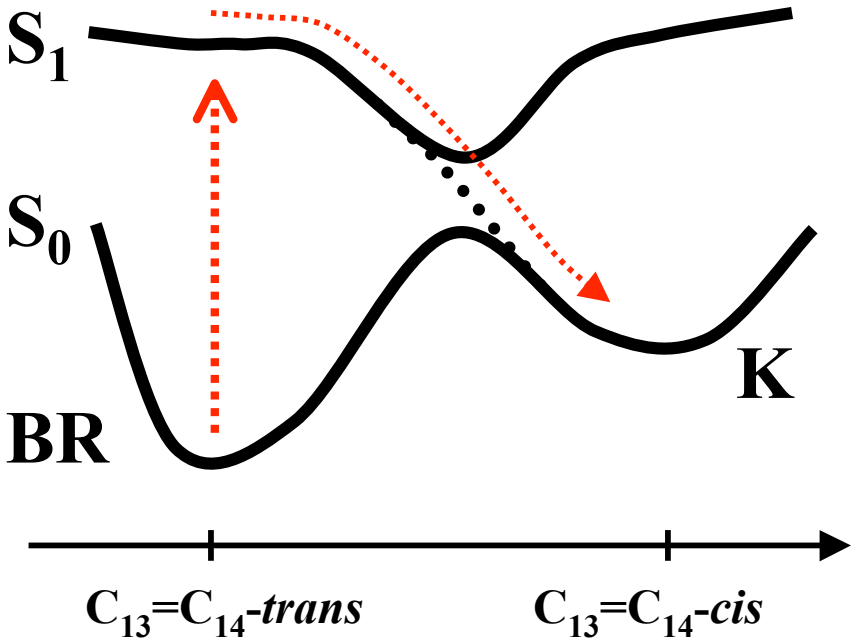
TABLE 2 The parameter set B used for the torsional potentials of the main polyene chain of the retinal Schiff base

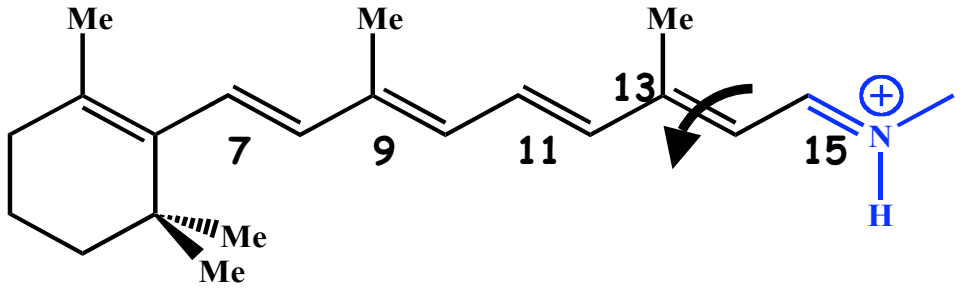
ϕ_i	k_i (kcal/mol)*	n_i	δ_i (deg)
$C_5{=}C_6{-}C_7{=}C_8$	11.24	2.0	180.00
$C_6{-}C_7{=}C_8{-}C_9$	39.98	2.0	180.00
$C_7{=}C_8{-}C_9{=}C_{10}$	17.03	2.0	180.00
$C_8{-}C_9{=}C_{10}{-}C_{11}$	37.28	2.0	180.00
$C_9{=}C_{10}{-}C_{11}{=}C_{12}$	22.50	2.0	180.00
$C_{10}{-}C_{11}{=}C_{12}{-}C_{13}$	35.08	2.0	180.00
$C_{11}{=}C_{12}{-}C_{13}{=}C_{14}$	28.30	2.0	180.00
$C_{12}{-}C_{13}{=}C_{14}{-}C_{15}$	29.46	2.0	180.00
$C_{13}{=}C_{14}{-}C_{15}{=}N_{16}$	30.43	2.0	180.00
$C_{14}{-}C_{15}{=}N_{16}{-}C_{\varepsilon}$	28.76	2.0	180.00

Tajkhorshid et al., 1999.

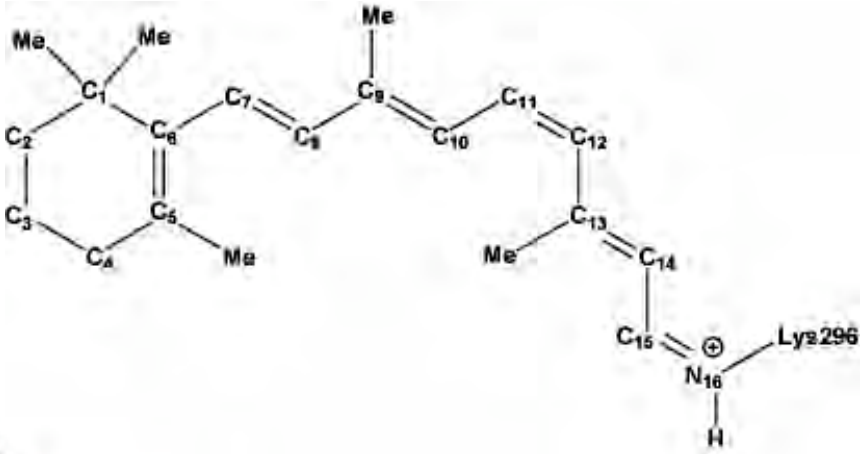
*$E_i^{\text{dihedral}} = (1/2)k_i[1 + \cos(n_i\varphi_i - \delta_i)]$.

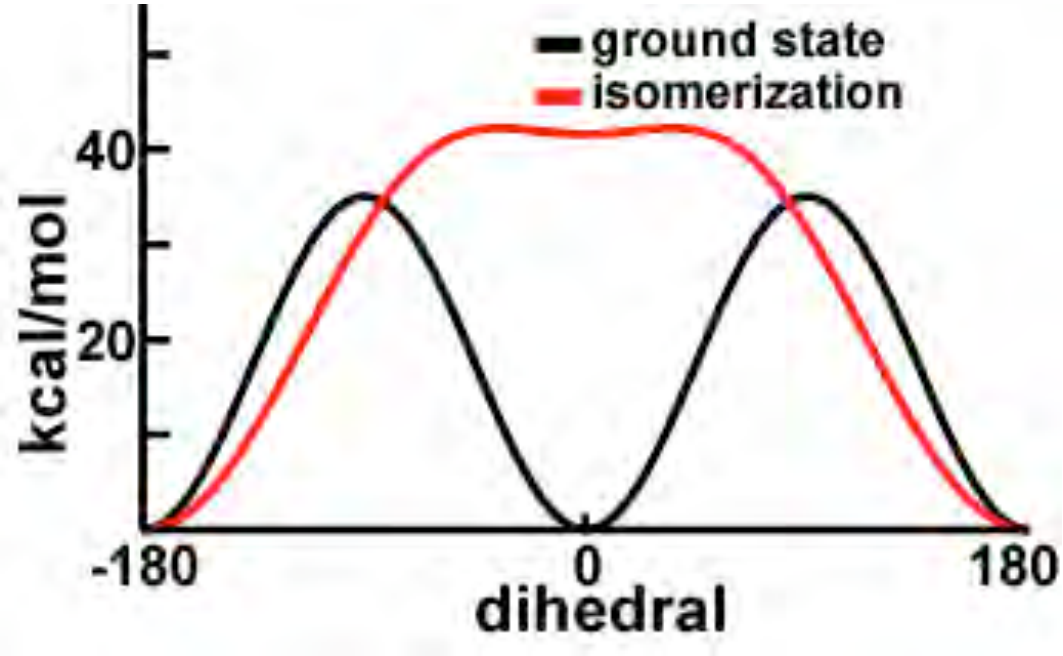
Coupling of electronic excitation and conformational change in bR





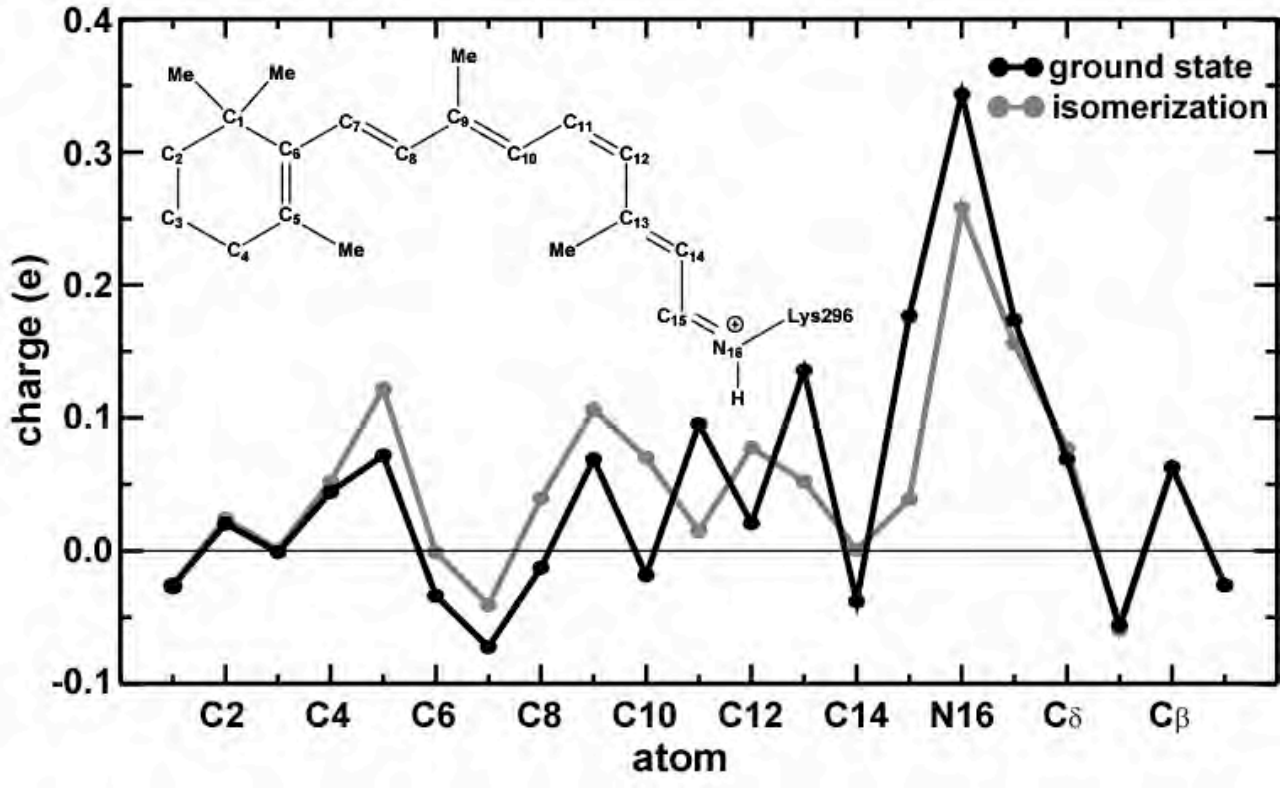
Inducing isomerization





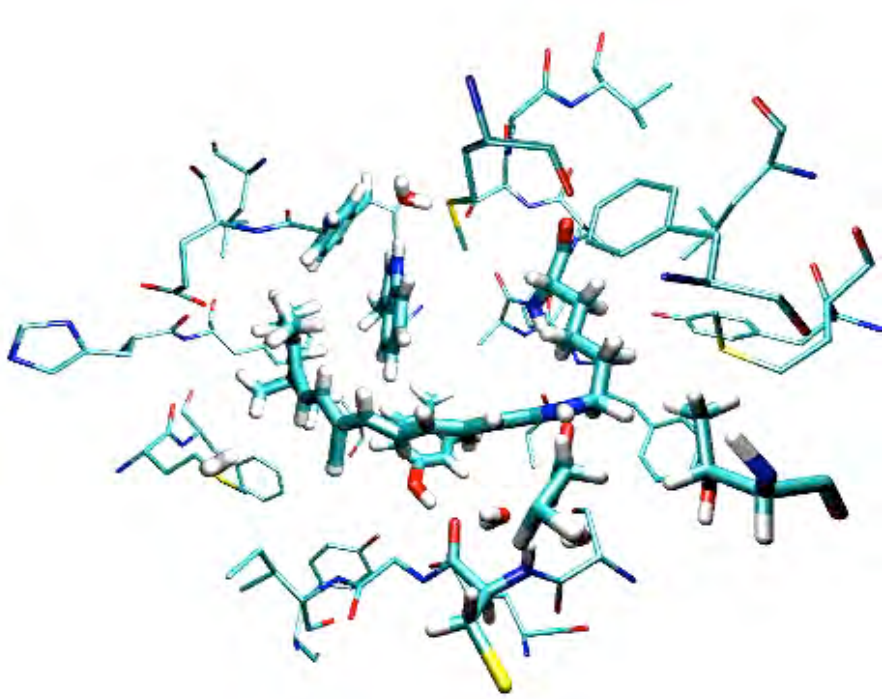
500 nm
~50 kcal/mole

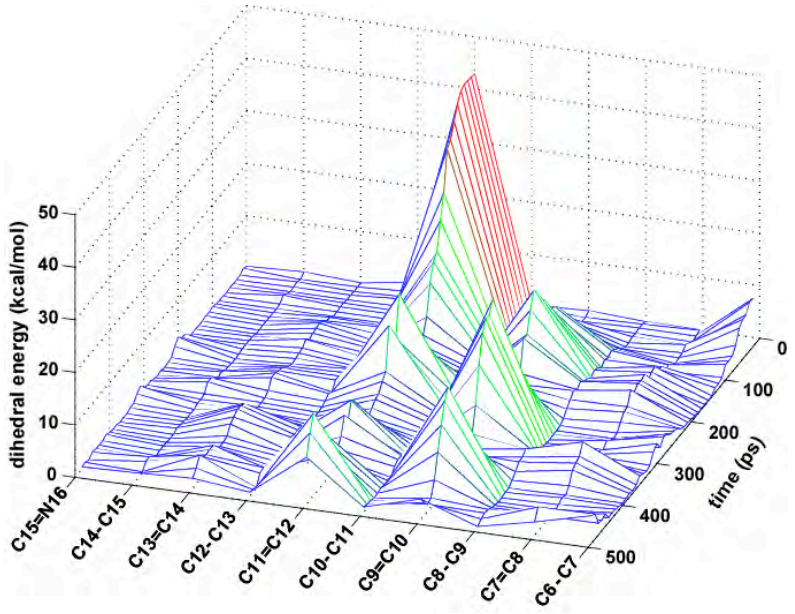
Retinal Charge Distribution



QM/MM derived partial atomic charges

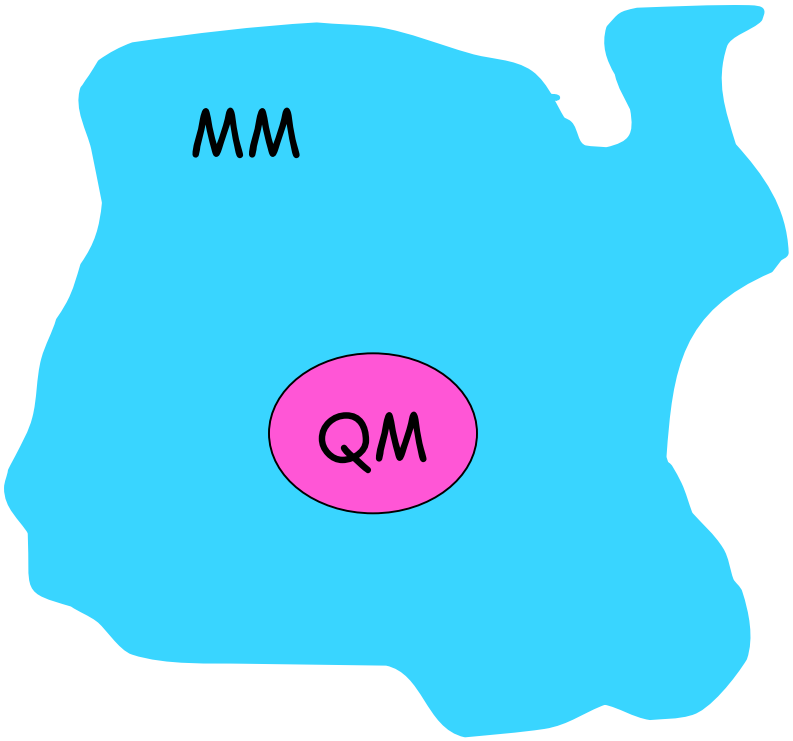
Classical Retinal Isomerization in Rhodopsin

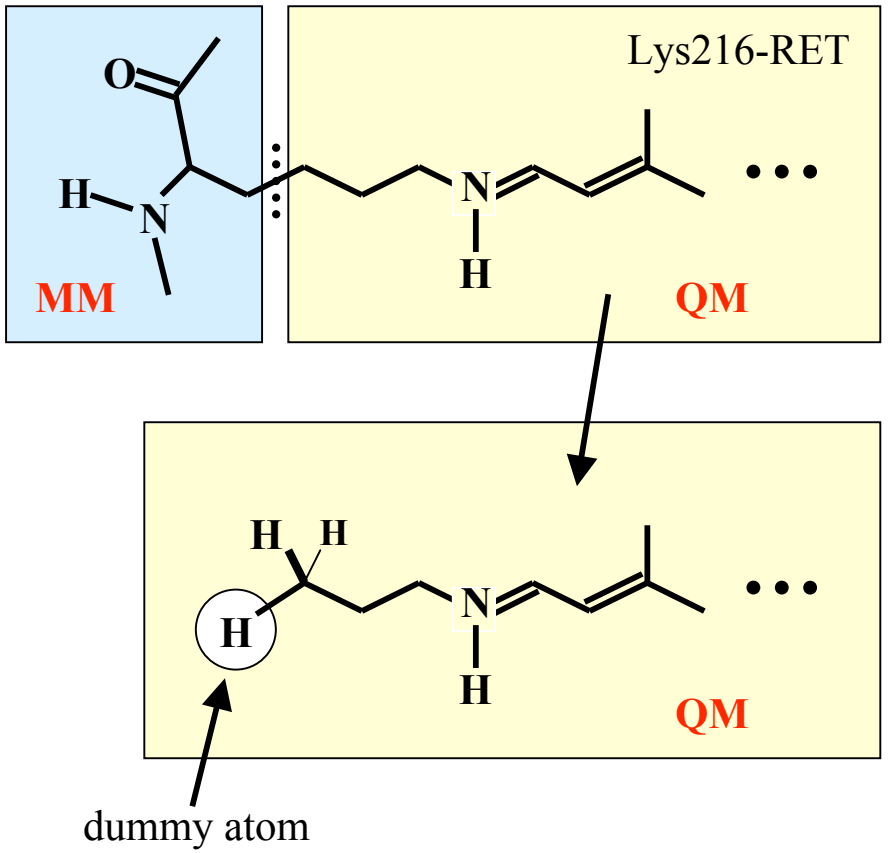




Twist Propagation

QM/MM calculations

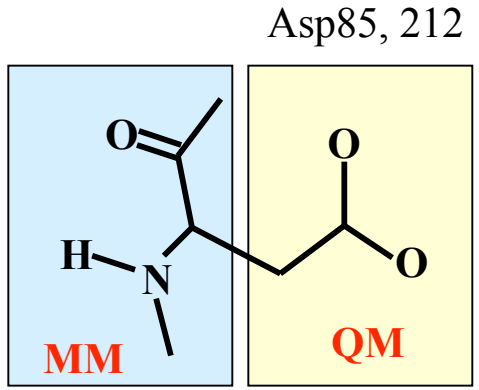




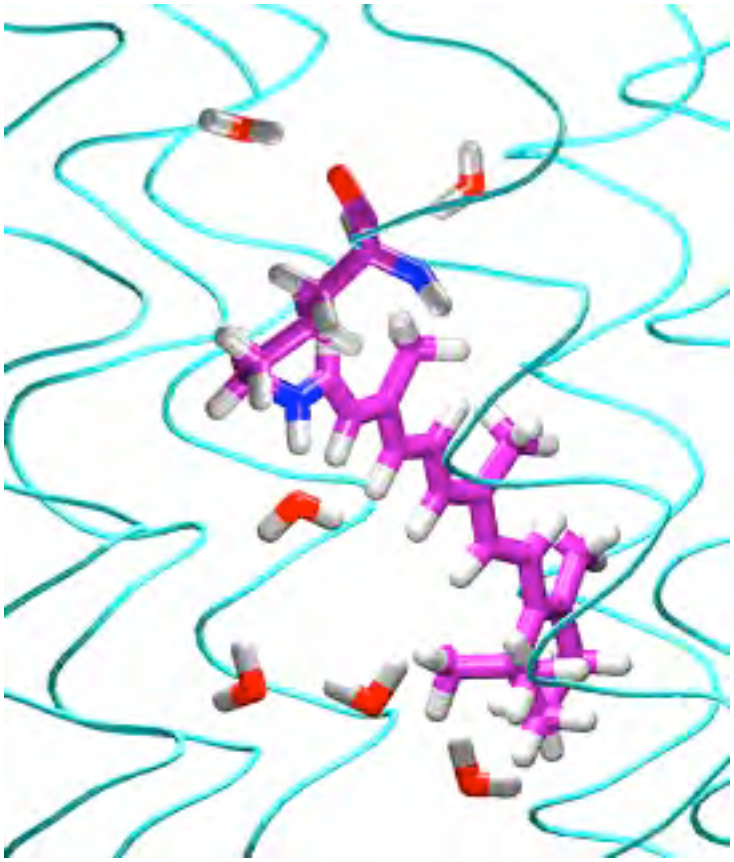
$$\hat{H} = \sum_i \frac{1}{2} p_i^2 + \sum_i \sum_A \frac{Z_A}{r_{iA}} + \sum_{i>j} \frac{1}{r_{ij}} + \sum_{A>B} \frac{Z_A Z_B}{r_{AB}}$$

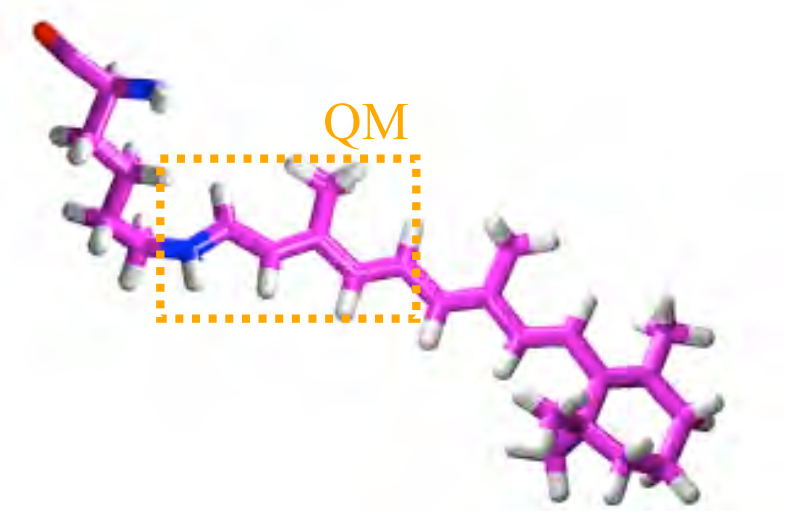
$$+ \sum_i \sum_p \frac{q_p}{r_{ip}} + \sum_A \sum_p \frac{Z_A q_p}{r_{Ap}}$$

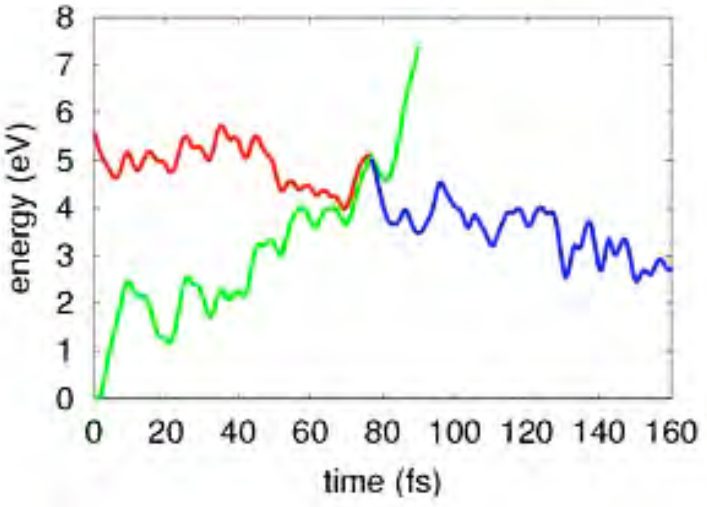
$$+ V_{QM-MM}^{MM} + V_{MM}^{MM}$$



Ab Initio QM/MM Excited State MD Simulation

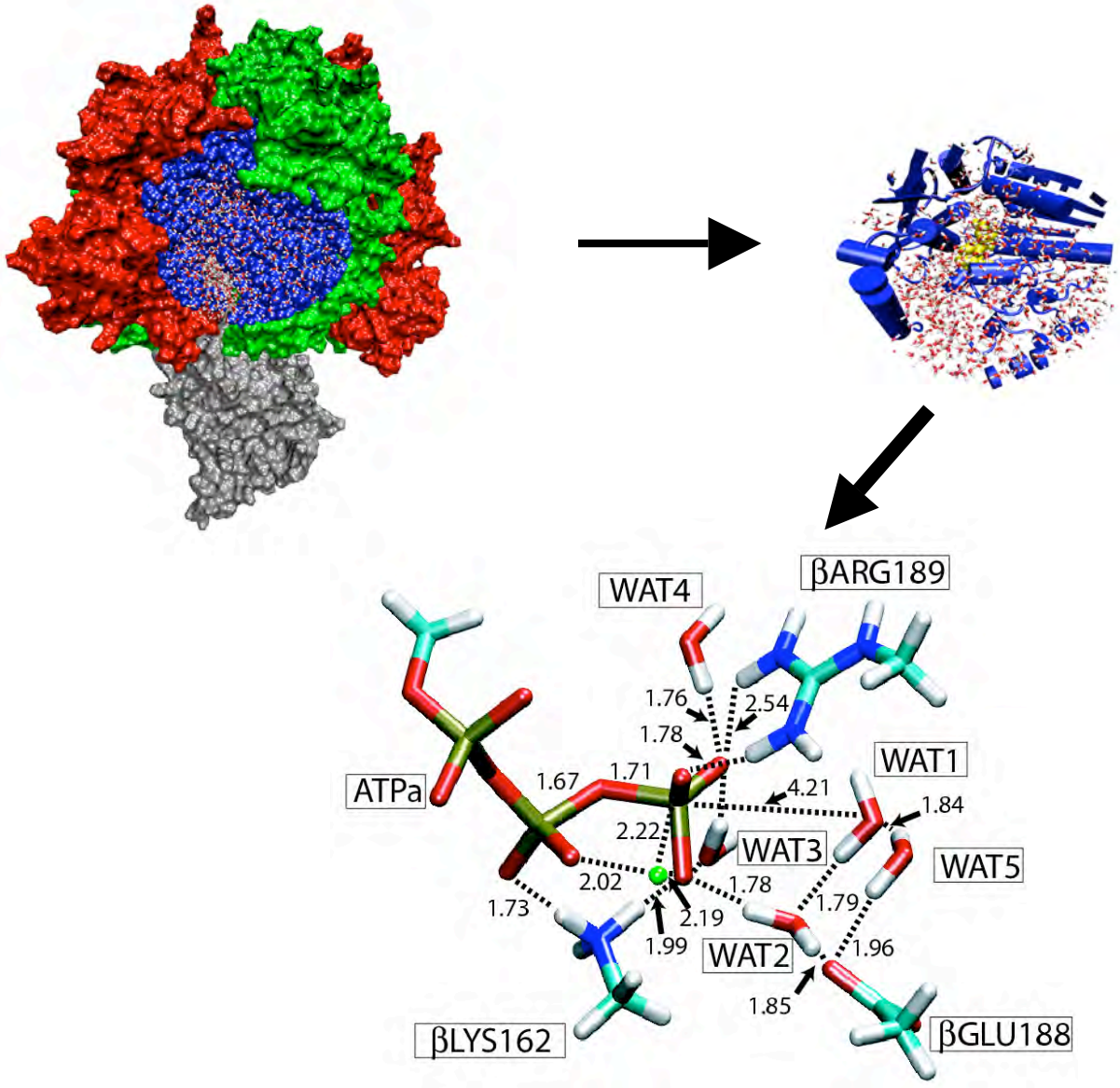


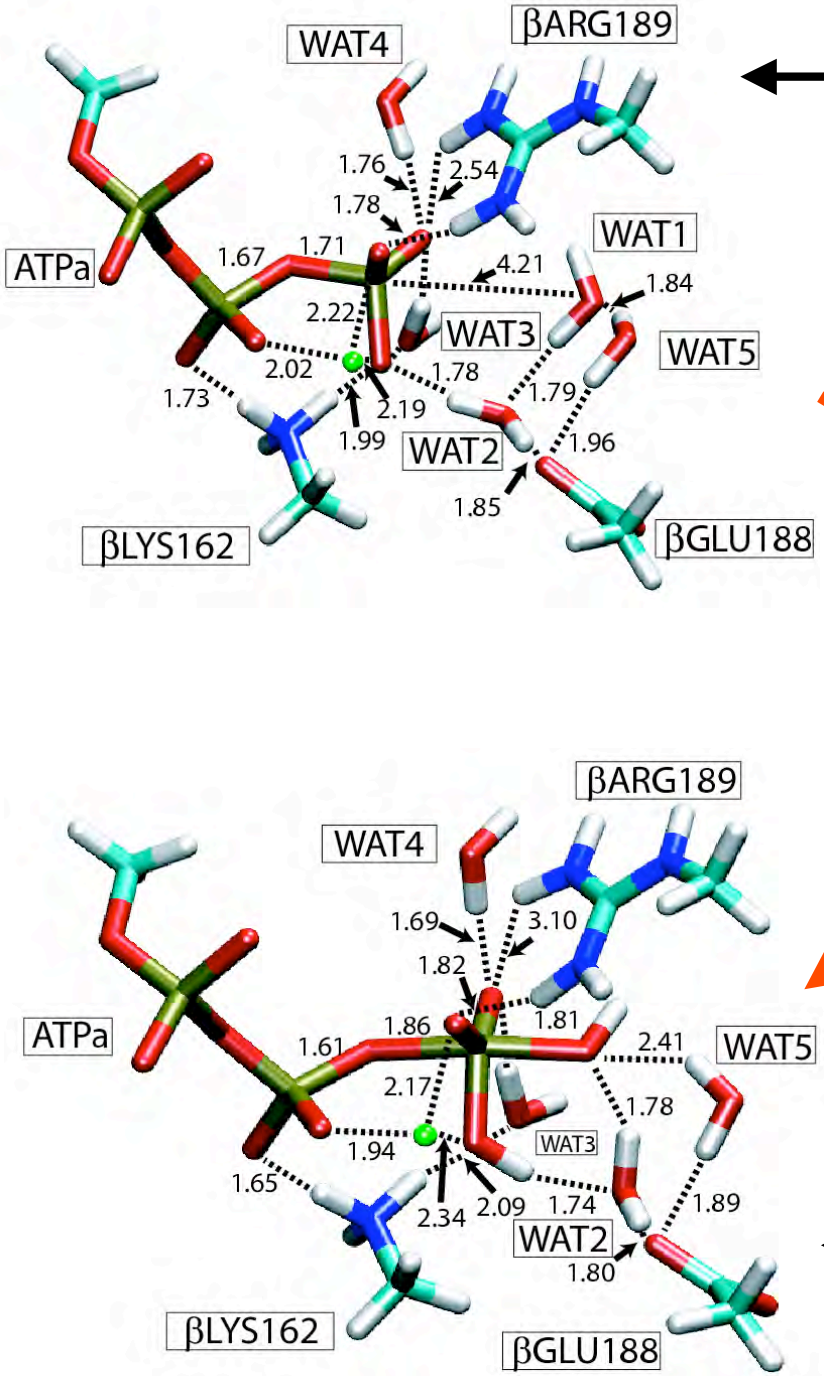


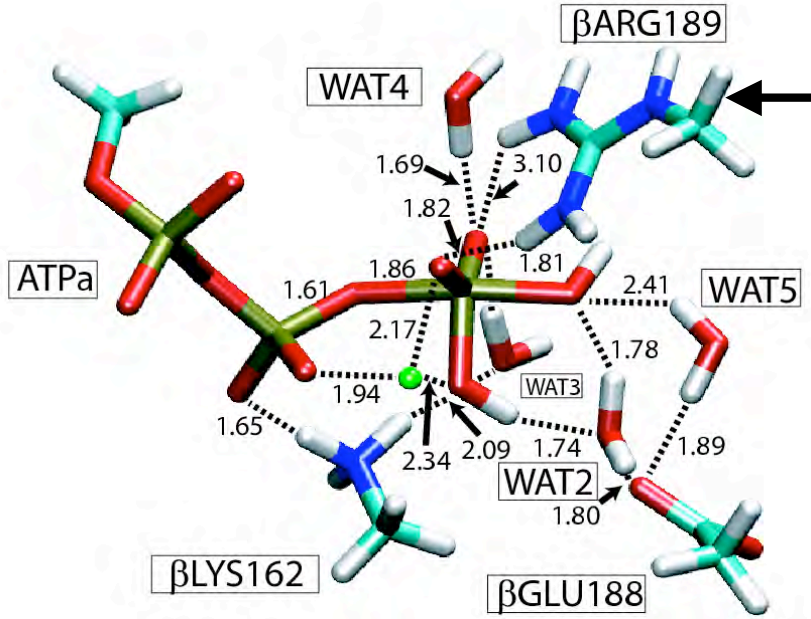


Quantum mechanical (QM) treatment of the chromophore, and force field (MM) treatment of the embedding protein

QM/MM calculation of ATP hydrolysis

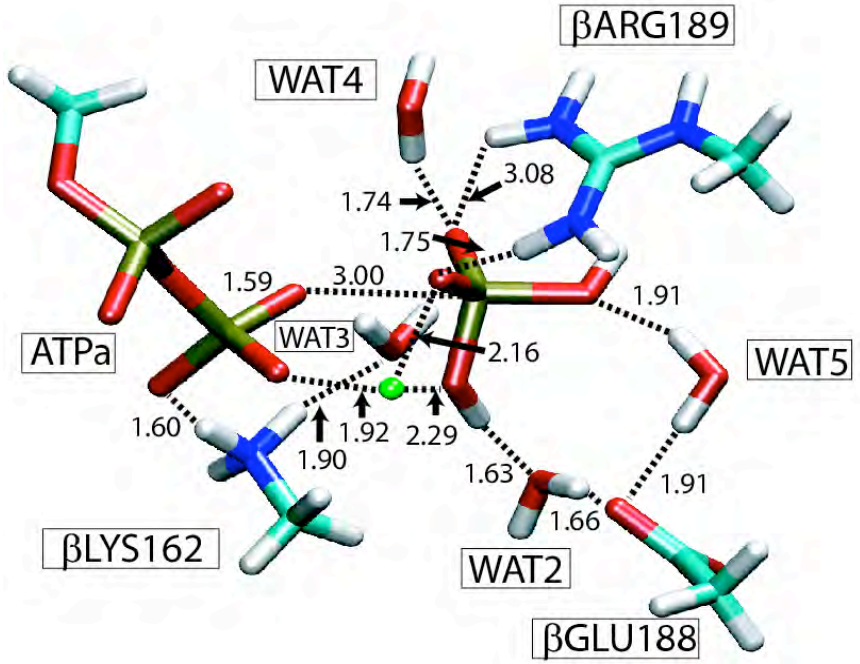



WAT4
βARG189
1.76
2.54
1.78
WAT1
ATPa
1.67
1.71
4.21
1.84
2.22
WAT3
WAT5
2.02
1.78
1.73
2.19
1.79
1.99
WAT2
1.96
1.85
βLYS162
βGLU188
Initial configuration
Transition state
βARG189
WAT4
1.68
3.11
1.77
ATPa
1.61
1.88
1.90
2.41
WAT5
2.87
WAT3
1.53
1.95
1.37
1.64
2.09
2.22
WAT2
1.91
1.56
βLYS162
βGLU188
βARG189
WAT4
1.69
3.10
1.82
ATPa
1.86
1.81
1.61
2.41
WAT5
2.17
1.78
1.94
WAT3
1.65
2.34
2.09
1.74
1.89
WAT2
1.80
βLYS162
βGLU188
Intermediate structure

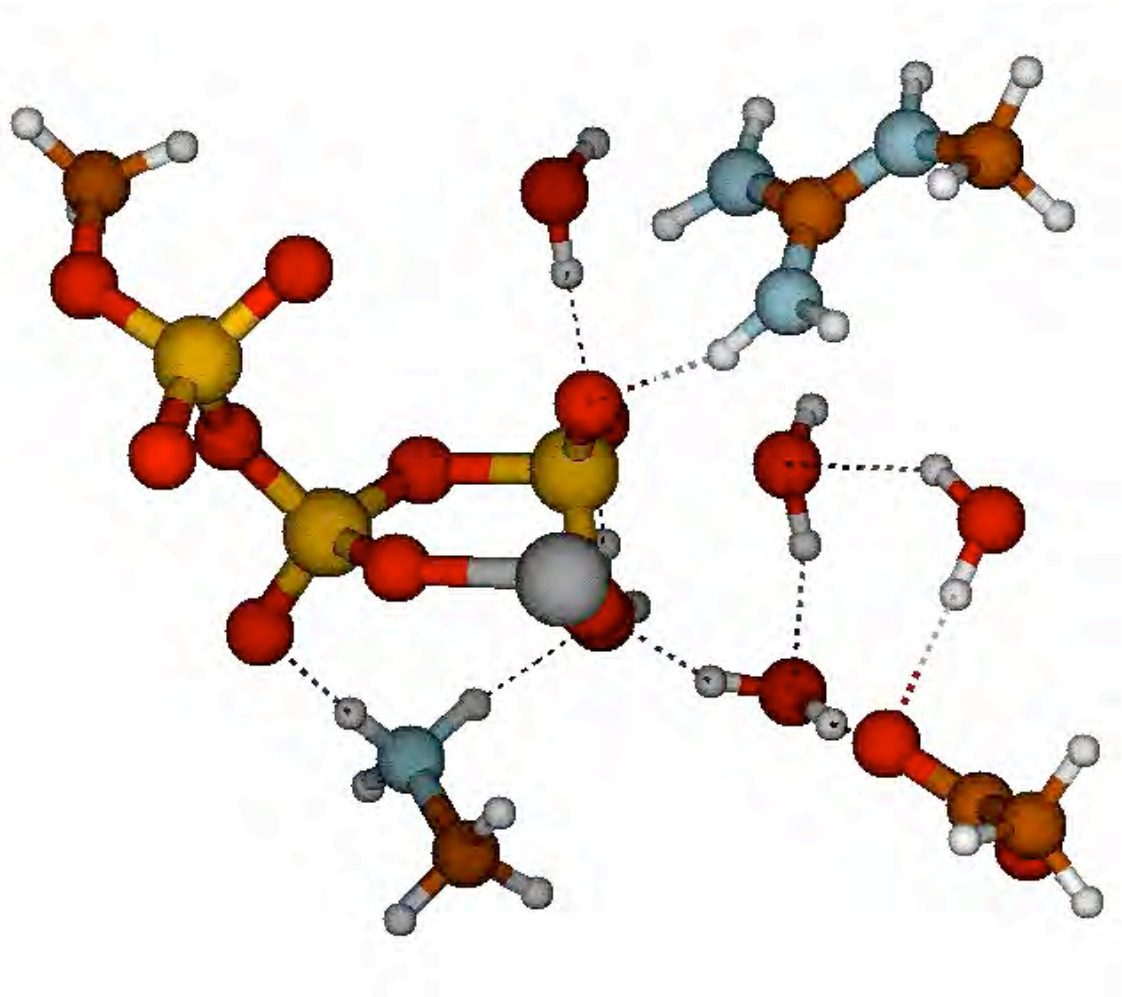


Intermediate structure

Product



ATP hydrolysis in β_{TP}



Coarse grain modeling of lipids

150 particles

9 particles!