# CHEM 436 / 630

## Molecular modelling of proteins

**Winter 2018 Term**

**Instructor:**
Guillaume Lamoureux
Concordia University, Montréal, Canada

# What to do with a force field ?

**Energy minimization :**

- The initial structure is relaxed to a configuration for which the forces on all atoms are zero.
- This relaxed structure is a single configuration out of the (usually) many distinct configurations that are thermodynamically accessible.

**Molecular dynamics (MD) :**

- The time course of all the atoms in the system is generated according to the Newton's equations ($\mathbf{a}_i = \mathbf{F}_i / m_i$).
- Thermodynamic and dynamic/kinetic information

**Langevin dynamics / Brownian dynamics :**

- The system is divided into *explicit* and *implicit* subsystems (e.g., explicit protein and implicit solvent) and the trajectories are calculated only for the atoms of the explicit subsystem. The influence of the implicit subsystem is described by a random force (representing molecular collisions).
- Thermodynamic information
- Dynamic/kinetic information is biased

**Monte Carlo (MC) :**

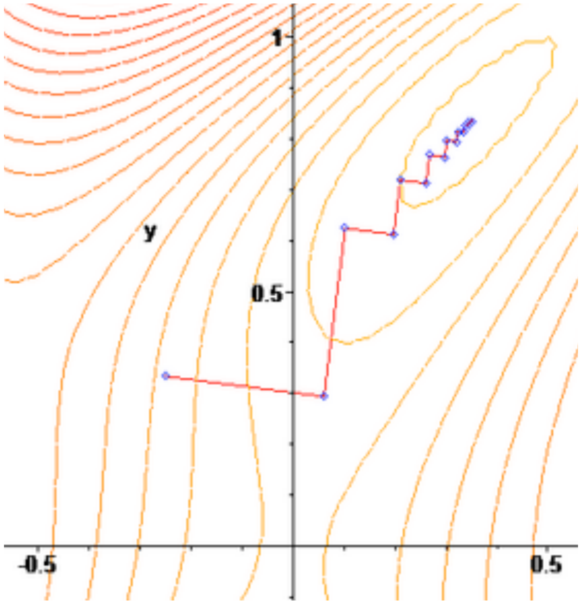
- A series of configurations characteristic of the *canonical ensemble* at temperature *T* is generated randomly.
- Thermodynamic information only

# Energy minimization

- Structure refinement (comparison/relaxation of sidechain rotamers, backbone, solvent, etc.)
- Preparing a structure for molecular dynamics

$$\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \ldots\}$$

(vector of 3*N* elements *R*$_i$)



## Steepest descent method

Requires the **forces** (first derivatives of *U*)

$$\mathbf{R}' = \mathbf{R} + \gamma \mathbf{F}$$

small constant

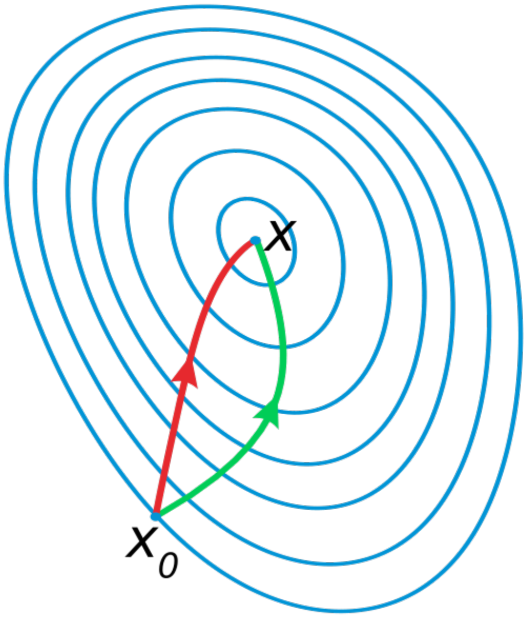
$$F_i = -\frac{\partial U}{\partial R_i}$$

## Newton–Raphson method

Requires the forces and the **Hessian matrix** (second derivatives of *U*)

$$\mathbf{R}' = \mathbf{R} + H^{-1} \mathbf{F}$$

$$H_{ij} = \frac{\partial^2 U}{\partial R_i \, \partial R_j}$$



Figures from :
Wikipedia: Gradient descent
Wikipedia: Newton's method in optimization

# What to do with a force field ?

## Energy minimization :

- The initial structure is relaxed to a configuration for which the forces on all atoms are zero.
- This relaxed structure is a single configuration out of the (usually) many distinct configurations that are thermodynamically accessible.

## Molecular dynamics (MD) :

- The time course of all the atoms in the system is generated according to the Newton's equations ($\mathbf{a}_i = \mathbf{F}_i / m_i$).
- Thermodynamic and dynamic/kinetic information

## Langevin dynamics / Brownian dynamics :

- The system is divided into *explicit* and *implicit* subsystems (e.g., explicit protein and implicit solvent) and the trajectories are calculated only for the atoms of the explicit subsystem. The influence of the implicit subsystem is described by a random force (representing molecular collisions).
- Thermodynamic information
- Dynamic/kinetic information is biased

## Monte Carlo (MC) :

- A series of configurations characteristic of the *canonical ensemble* at temperature $T$ is generated randomly.
- Thermodynamic information only

# Molecular dynamics

“Putting the thermal fluctuations back into a protein structure”

**Basic techniques :**
Simulation of a system undergoing thermal fluctuations

- Conformational sampling / searching
- Exploration of dynamics
- Debye–Waller factors (x-ray)
- Diffusion coefficients
- IR spectra
- NMR observables
- Raman spectra

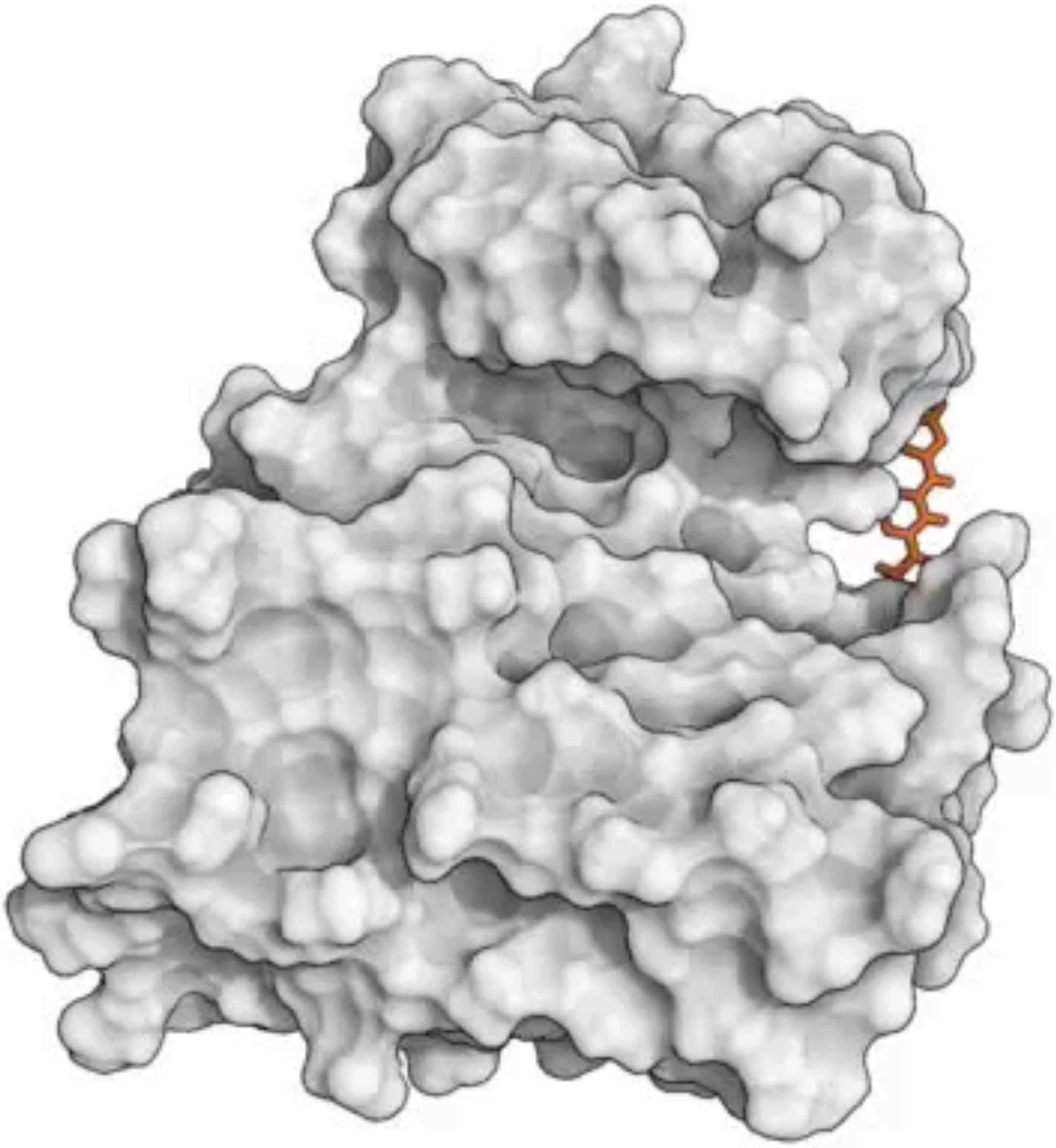
**Advanced techniques :**
Free energy calculations

- Reaction profiles (conformational change, substrate permeation, etc.)
- Binding free energies of ligands
- Partition constants (“log *P*”)
- Thermodynamic effects of a mutation
- $pK_a$ calculations
- Rate constants / kinetics

# Dasatinib binding to Src kinase

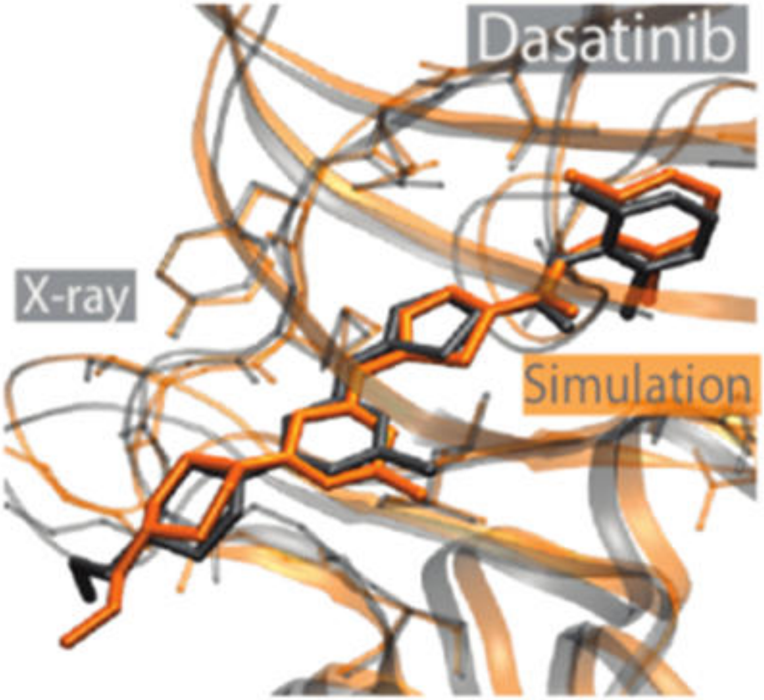
- Molecular dynamics simulations (total of 35 μs)
- Ligand "finds" the known position (PDB: 3G5D)

- Amber99SB ff for the protein
- TIP3P ff for water
- GAFF ff with AM1-BCC for the ligand



**Movie from:**
Y. Shan et al., *J. Am. Chem. Soc.* 2011, 133, 9181-9183.
http://dx.doi.org/10.1021/ja202726y

# Dasatinib binding to Src kinase



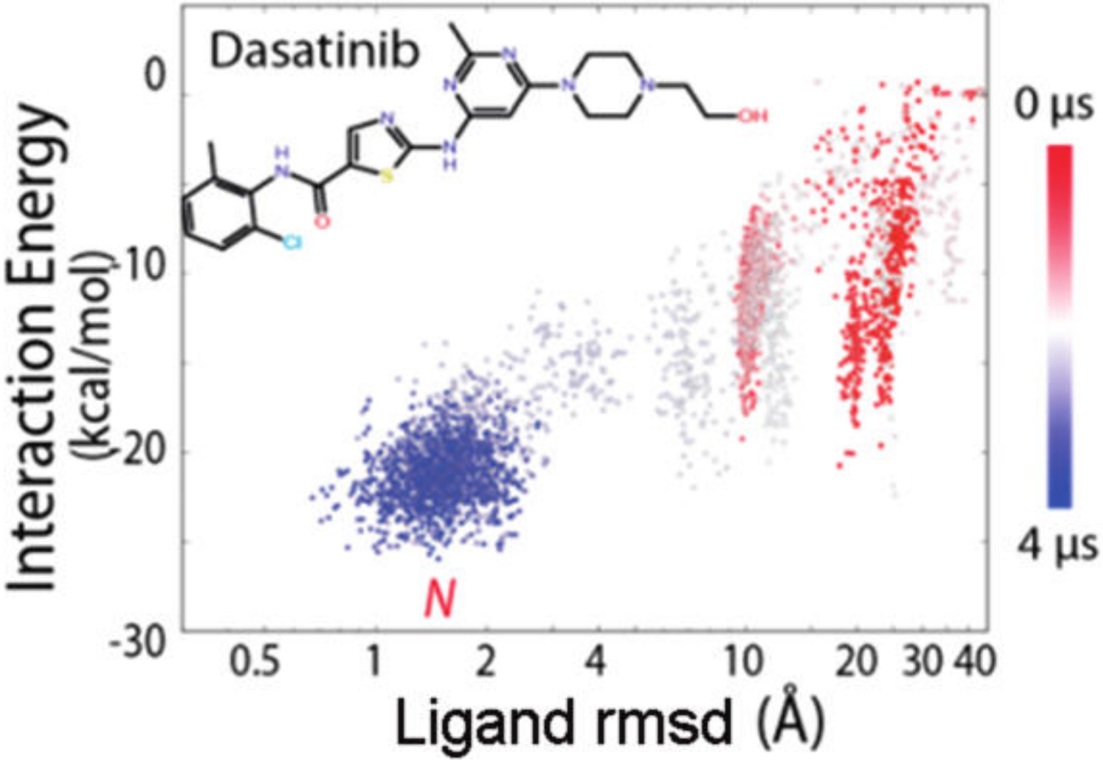


**Figure from:**
Y. Shan et al., *J. Am. Chem. Soc.* 2011, 133, 9181-9183.
http://dx.doi.org/10.1021/ja202726y

# Molecular dynamics

## Newton’s equations

acceleration of particle i (“effect”)

force on particle i (“cause”)

$$m_i\mathbf{a}_i(t) = \mathbf{F}_i(t) = -\frac{\partial U}{\partial \mathbf{r}_i}$$

$$\downarrow$$

$$\mathbf{v}_i(t)$$

$$\downarrow$$

$$\mathbf{r}_i(t)$$

### QM (“ab initio”)

$U =$ the result of a long electronic structure calculation

### MM (“empirical”)

$U =$ a simple function of the atomic coordinates

$\ldots$

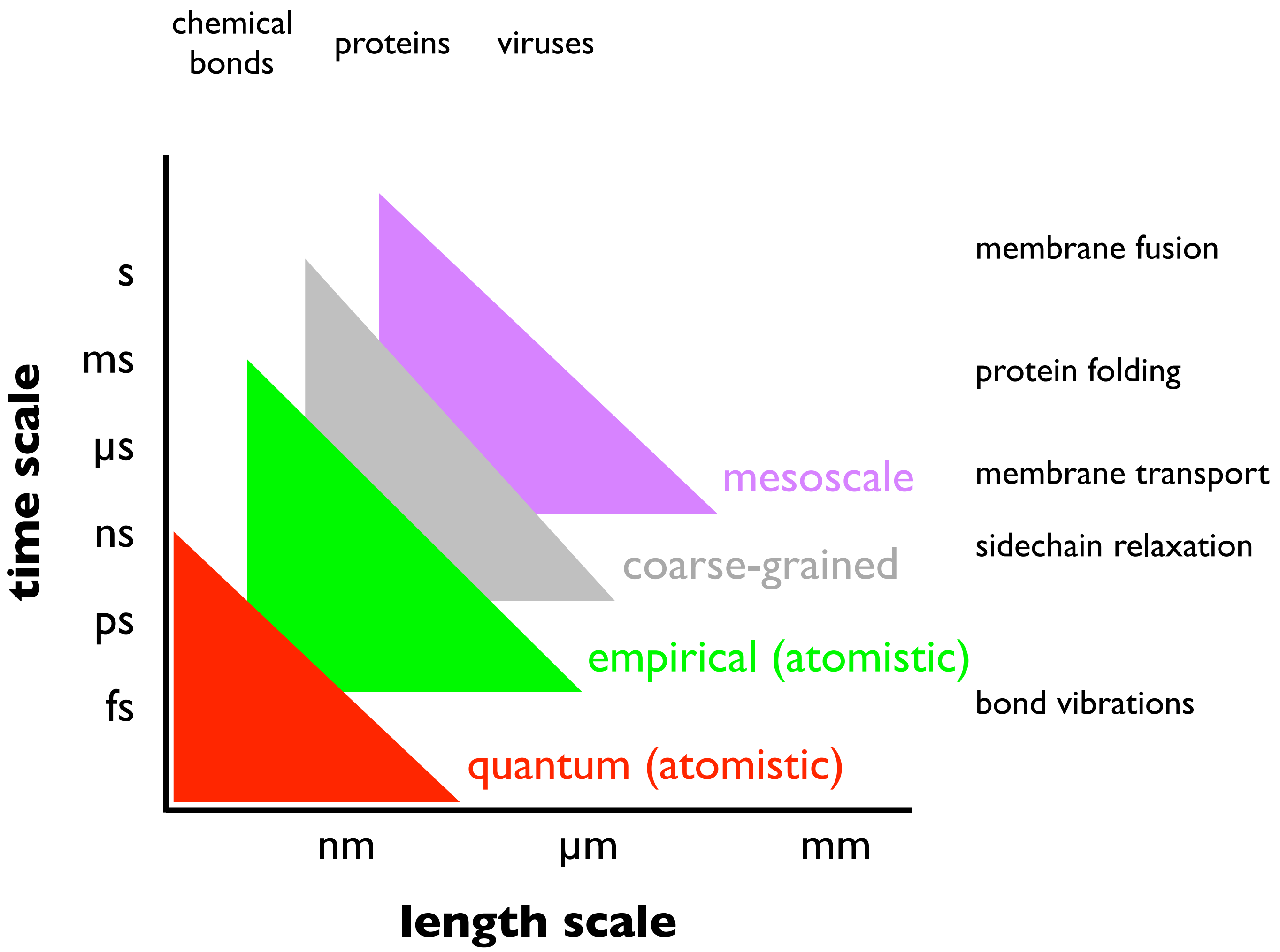
$\mathbf{r}(t)$

$\mathbf{r}(t+\delta t)$

$\mathbf{r}(t+2\delta t)$

$\ldots$

# Tell me how you calculate the forces…



(figure adapted from: Nielsen et al., *J. Phys.: Cond. Mat.* 2004, **16**, R481)

# **Potential energy functions** commonly used in molecular dynamics simulations of proteins

**CHARMM (proteins, nucleic acids, lipids)**
http://mackerell.umaryland.edu/charmm_ff.shtml

MacKerell *et al.*, " All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins", *J. Phys. Chem. B* 1998, **102**, 3586–3616.

Vanommeslaeghe *et al.*, "CHARMM General Force Field (CGenFF): A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields", *J. Comp. Chem.* 2010, **31**, 671-690.

**Amber (proteins, nucleic acids)**
http://ambermd.org/AmberModels.php

Cornell *et al.*, "A second generation force field for the simulation of proteins and nucleic acids", *J. Am. Chem. Soc.* 1995, **117**, 5179–5197.

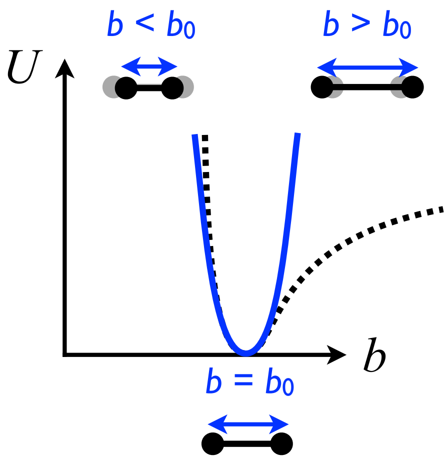
Ponder and Case, "Force fields for protein simulations", *Adv. Prot. Chem.* 2003, **66**, 27–85.

# Empirical force fields

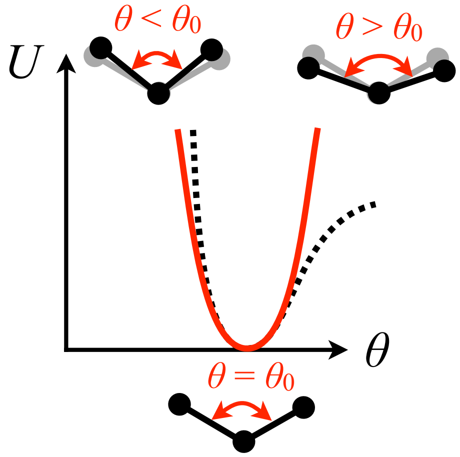
Energy calculated using a simple formula, in terms of the positions (*x*,*y*,*z*) of the individual atoms forming the molecule.

$$U = \sum_{\text{bonds}} k_b (b - b_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 \boxed{+ \ldots}$$

**bond strength**
(depends on the type of bond: single, double, atoms involved, etc.)



Bond lengths oscillate around their equilibrium values



Bond angles oscillate around their equilibrium values

**Note:**
These *harmonic* approximations are OK as long as we're interested only in thermal, quasi-classical vibrations (involving heavy atoms).

# par_all22_prot.inp (CHARMM)

```
BONDS
!
!V(bond) = Kb(b - b0)**2
!
!Kb: kcal/mole/A**2
!b0: A
!
!atom type Kb          b0
!
C    C    600.000     1.3350 ! ALLOW ARO HEM
                ! Heme vinyl substituent (KK, from propene (JCS))
CA   CA   305.000     1.3750 ! ALLOW   ARO
                ! benzene, JES 8/25/89
CE1  CE1  440.000     1.3400   !
         ! for butene; from propene, yin/adm jr., 12/95
CE1  CE2  500.000     1.3420   !
         ! for propene, yin/adm jr., 12/95
CE1  CT2  365.000     1.5020   !
         ! for butene; from propene, yin/adm jr., 12/95
CE1  CT3  383.000     1.5040   !
         ! for butene, yin/adm jr., 12/95
CE2  CE2  510.000     1.3300   !
         ! for ethene, yin/adm jr., 12/95
CP1  C    250.000     1.4900 ! ALLOW PRO
                ! 6-31g* AcProNH2, ProNH2, 6-31g*//3-21g AcProNHCH3 RLD 4/23/93
CP1  CC   250.000     1.4900 ! ALLOW PRO
                ! 6-31g* AcProNH2, ProNH2, 6-31g*//3-21g AcProNHCH3 RLD 4/23/93
CP1  CD   200.000     1.4900 ! ALLOW PRO
                ! 6-31g* AcProNH2, ProNH2, 6-31g*//3-21g AcProNHCH3 RLD 4/23/93
CP2  CP1  222.500     1.5270 ! ALLOW PRO
                ! 6-31g* AcProNH2, ProNH2, 6-31g*//3-21g AcProNHCH3 RLD 4/23/93
CP2  CP2  222.500     1.5370 ! ALLOW PRO
```

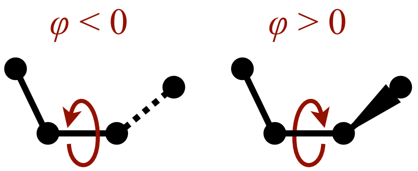
## par_all22_prot.inp (CHARMM)

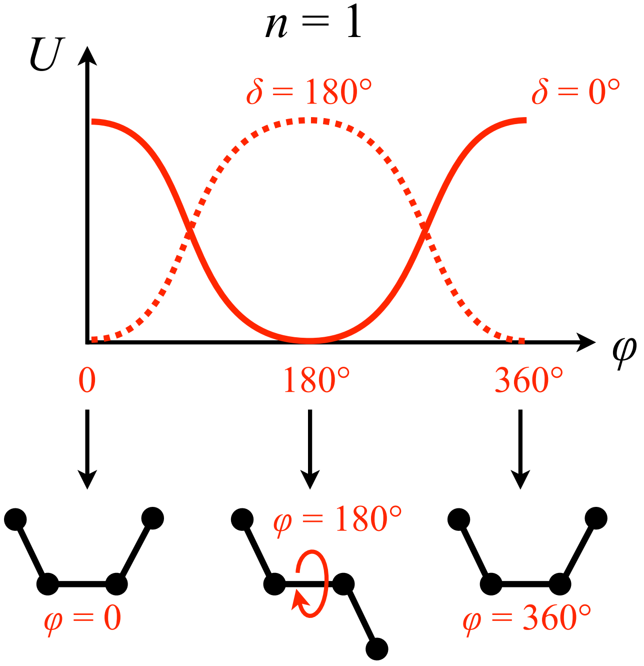
```
ANGLES
!
!V(angle) = Ktheta(Theta - Theta0)**2
!
!V(Urey-Bradley) = Kub(S - S0)**2
!
!Ktheta: kcal/mole/rad**2
!Theta0: degrees
!Kub: kcal/mole/A**2 (Urey-Bradley)
!S0: A
!
!atom types     Ktheta    Theta0   Kub     S0
!
CA   CA   CA     40.000    120.00   35.00   2.41620 ! ALLOW   ARO
                ! JES 8/25/89
CE1  CE1  CT2     48.00    123.50   !
                ! for 2-butene, yin/adm jr., 12/95
CE1  CE1  CT3     48.00    123.50   !
         ! for 2-butene, yin/adm jr., 12/95
CE1  CT2  CT3     32.00    112.20   !
         ! for 1-butene; from propene, yin/adm jr., 12/95
CE2  CE1  CT2     48.00    126.00   !
         ! for 1-butene; from propene, yin/adm jr., 12/95
CE2  CE1  CT3     47.00    125.20   !
         ! for propene, yin/adm jr., 12/95
CP1  N    C       60.000   117.0000 ! ALLOW PRO
                ! 6-31g* AcProNH2, ProNH2, 6-31g*//3-21g AcProNHCH3 RLD 4/23/93
CP2  CP1  C       52.000   112.3000 ! ALLOW PRO
                ! 6-31g* AcProNH2, ProNH2, 6-31g*//3-21g AcProNHCH3 RLD 4/23/93
CP2  CP1  CC      52.000   112.3000 ! ALLOW PRO
                ! 6-31g* AcProNH2, ProNH2, 6-31g*//3-21g AcProNHCH3 RLD 4/23/93
CP2  CP1  CD      50.000   112.3000 ! ALLOW PRO PEP
```

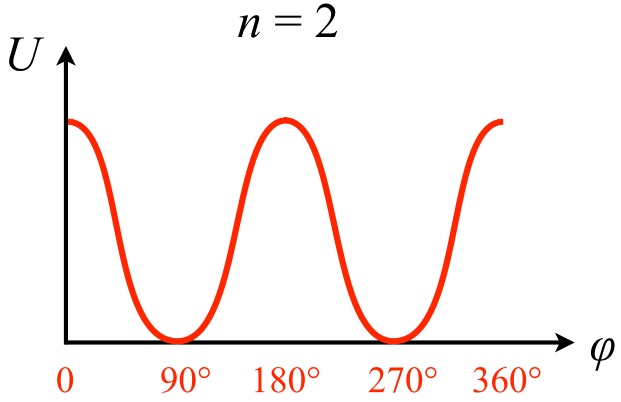
$$U = \ldots + \sum_{\text{dihedrals}} k_\phi [1 + \cos(n\phi - \delta)] \;\; + \ldots$$

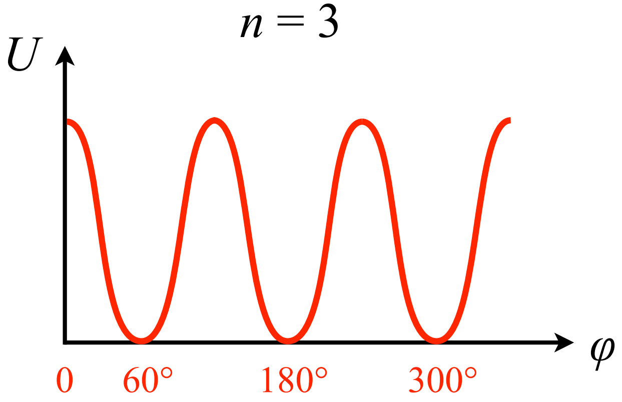
**multiplicity**
*n* = 3 for sp3-sp3 bonds (ex: ethane)
*n* = 2 for sp2-sp2 bonds (ex: ethylene)

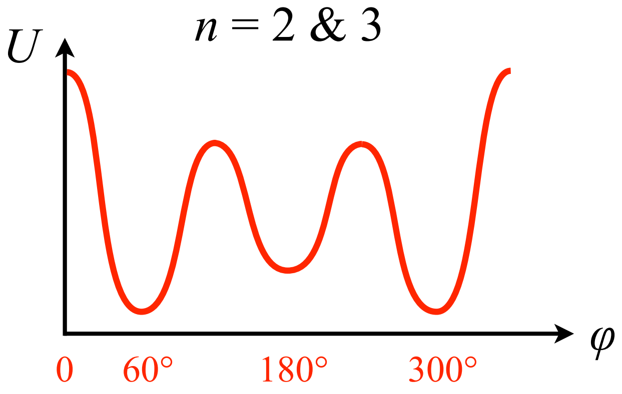
**phase**
δ = 0° if staggered is most stable
δ =180° if eclipsed in most stable







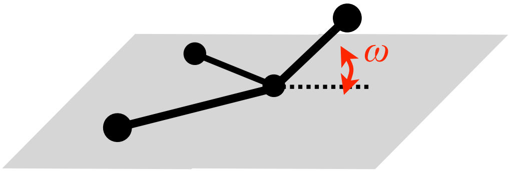




## par_all22_prot.inp (CHARMM)

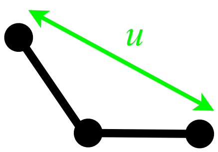
```
DIHEDRALS
!
!V(dihedral) = Kchi(1 + cos(n(chi) - delta))
!
!Kchi: kcal/mole
!n: multiplicity
!delta: degrees
!
!atom types             Kchi    n   delta
!
C    CT1  NH1  C        0.2000  1   180.00 ! ALLOW PEP
                ! ala dipeptide update for new C VDW Rmin, adm jr., 3/3/93c
C    CT2  NH1  C        0.2000  1   180.00 ! ALLOW PEP
                ! ala dipeptide update for new C VDW Rmin, adm jr., 3/3/93c
C    N    CP1  C        0.8000  3     0.00 ! ALLOW PRO PEP
                ! 6-31g* AcProNH2, ProNH2, 6-31g*//3-21g AcProNHCH3 RLD 4/23/93
CA   CA   CA   CA       3.1000  2   180.00 ! ALLOW   ARO
                ! JES 8/25/89
CA   CPT  CPT  CA       3.1000  2   180.00 ! ALLOW   ARO
                ! JWK 05/14/91 fit to indole
CA   CT2  CT1  C        0.0400  3     0.00 ! ALLOW   ARO
                ! 2.7 kcal/mole CH3 rot in ethylbenzene, adm jr, 3/7/92
CA   CY   CPT  CA       3.0000  2   180.00 ! ALLOW   ARO
                ! JWK 09/05/89
CA   NY   CPT  CA       3.0000  2   180.00 ! ALLOW   ARO
                ! JWK 05/14/91 fit to indole
CC   CP1  N    C        0.8000  3     0.00 ! ALLOW PRO PEP
                ! 6-31g* AcProNH2, ProNH2, 6-31g*//3-21g AcProNHCH3 RLD 4/23/93
CC   CT1  CT2  CA       0.0400  3     0.00 ! ALLOW   ARO
                ! 2.7 kcal/mole CH3 rot in ethylbenzene, adm jr, 3/7/92
CC   CT1  NH1  C        0.2000  1   180.00 ! ALLOW PEP POL
                ! ala dipeptide update for new C VDW Rmin, adm jr., 3/3/93c
```

$$U = \ldots + \sum_{\mathrm{impropers}} k_\omega(\omega - \omega_0)^2 + \sum_{\mathrm{Urey-Bradley}} k_u(u - u_0)^2$$



Used to keep “dangling” atom oscillating around their optimal orientation

**Examples:**
carbonyl oxygens,
hydrogens in aromatic groups,
etc.



Correction terms to reproduce some IR frequencies

**Example:**
between hydrogens of methyl groups

# par_all22_prot.inp (CHARMM)

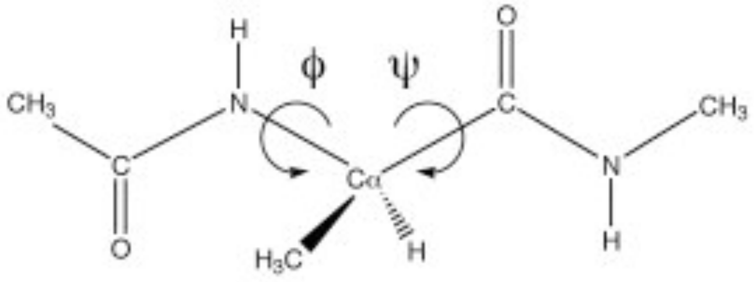
```
IMPROPER
!
!V(improper) = Kpsi(psi - psi0)**2
!
!Kpsi: kcal/mole/rad**2
!psi0: degrees
!note that the second column of numbers (0) is ignored
!
!atom types           Kpsi                   psi0
!
CPB  CPA  NPH  CPA    20.8000         0      0.0000 ! ALLOW HEM
                ! Heme (6-liganded): porphyrin macrocycle (KK 05/13/91)
CPB  X    X    CE1    90.0000         0      0.0000 ! ALLOW HEM
                ! Heme (6-liganded): substituents (KK 05/13/91)
CT2  X    X    CPB    90.0000         0      0.0000 ! ALLOW HEM
                ! Heme (6-liganded): substituents (KK 05/13/91)
CT3  X    X    CPB    90.0000         0      0.0000 ! ALLOW HEM
                ! Heme (6-liganded): substituents (KK 05/13/91)
!HA   C    C    HA     20.0000         0      0.0000 ! ALLOW   PEP POL ARO
                ! Heme vinyl substituent (KK, from propene (JCS))
HA   CPA  CPA  CPM    29.4000         0      0.0000 ! ALLOW HEM
                ! Heme (6-liganded): porphyrin macrocycle (KK 05/13/91)
!HA   CPB  C    C      20.0000         0      0.0000 ! ALLOW HEM ARO
                ! Heme (6-liganded): substituents (KK 05/13/91)
!HA   HA   C    C      20.0000         0    180.0000 ! ALLOW   PEP POL ARO
                ! Heme vinyl substituent (KK, from propene (JCS))
HE2  HE2  CE2  CE2     3.0            0      0.00   !
         ! for ethene, yin/adm jr., 12/95
HR1  NR1  NR2  CPH2    0.5000         0      0.0000 ! ALLOW ARO
                ! his, adm jr., 7/05/90
HR1  NR2  NR1  CPH2    0.5000         0      0.0000 ! ALLOW ARO
                ! his, adm jr., 7/05/90
```

$$U = \ldots + \sum_{\alpha\ \text{carbons}} \text{CMAP}(\phi, \psi)$$

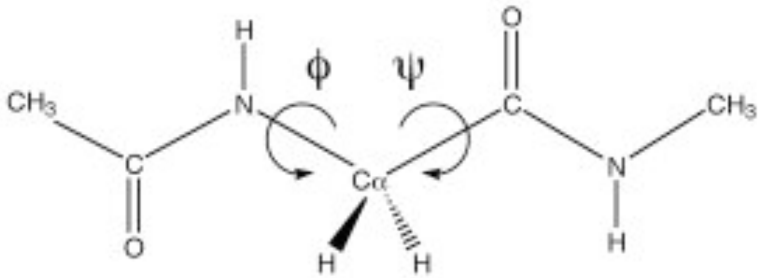
Correction terms to improve the φψ energy surfaces

(CHARMM force field only)

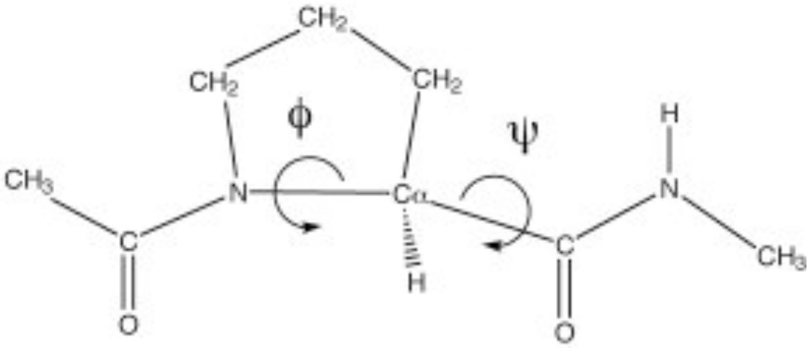
Alanine dipeptide



Glycine dipeptide



Proline dipeptide



**Reference:**
A.D. MacKerell Jr., M. Feig, C.L. Brooks, III, *J. Comp. Chem.* **25**, 1400-1415 (2004).
http://dx.doi.org/10.1002/jcc.20065

## par_all22_prot.inp (CHARMM)

```
CMAP
! 2D grid correction data.  The following surfaces are the correction
! to the CHARMM22 phi, psi alanine, proline and glycine dipeptide surfaces.
! Use of CMAP requires generation with the topology file containing the
! CMAP specifications along with version 31 or later of CHARMM.  Note that
! use of "skip CMAP" yields the charmm22 energy surfaces.
!
! references
!MacKerell, A.D., Jr., Feig, M., Brooks, C.L., III, Accurate Treatment of
!Protein Backbone Conformational Energetics in Empirical Force Fields, Submitted
!for publication.

!MacKerell, A.D., Jr., Feig, M., Brooks, C.L., III, Improved Treatment of the
!Protein Backbone in Empirical Force Fields, Journal of the American Chemical
!Society, In Press.


! alanine map
C    NH1  CT1  C    NH1  CT1  C    NH1   24

!-180
0.126790 0.768700 0.971260 1.250970 2.121010
2.695430 2.064440 1.764790 0.755870 -0.713470
0.976130 -2.475520 -5.455650 -5.096450 -5.305850
-3.975630 -3.088580 -2.784200 -2.677120 -2.646060
-2.335350 -2.010440 -1.608040 -0.482250

!-165
-0.802290 1.377090 1.577020 1.872290 2.398990
2.461630 2.333840 1.904070 1.061460 0.518400
-0.116320 -3.575440 -5.284480 -5.160310 -4.196010
-3.276210 -2.715340 -1.806200 -1.101780 -1.210320
```

# Lennard-Jones

$$U = \ldots + \sum_{\text{nonbonded}} \epsilon_{ij} \left[ \left( \frac{\text{Rmin}_{ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{\text{Rmin}_{ij}}{r_{ij}} \right)^{6} \right]$$

**$1/r^{12}$ term :**
Models the "exchange-repulsion-penetration" forces

Always **repulsive**

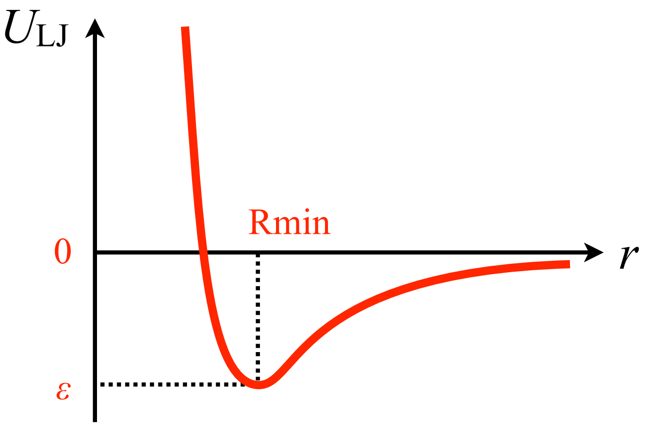
**$1/r^{6}$ term :**
Models the London dispersion forces (first term of the series $1/r^6$, $1/r^8$, $1/r^{10}$, etc.)

Always **attractive**

$\varepsilon_{ij}$ and $\text{Rmin}_{ij}$ are parameters defined for each pair of atoms, according to the following **mixing rules** :

$$\epsilon_{ij} \overset{\text{def}}{=} \sqrt{\epsilon_i \epsilon_j}$$

$$\text{Rmin}_{ij} \overset{\text{def}}{=} \tfrac{1}{2}\text{Rmin}_i + \tfrac{1}{2}\text{Rmin}_j$$



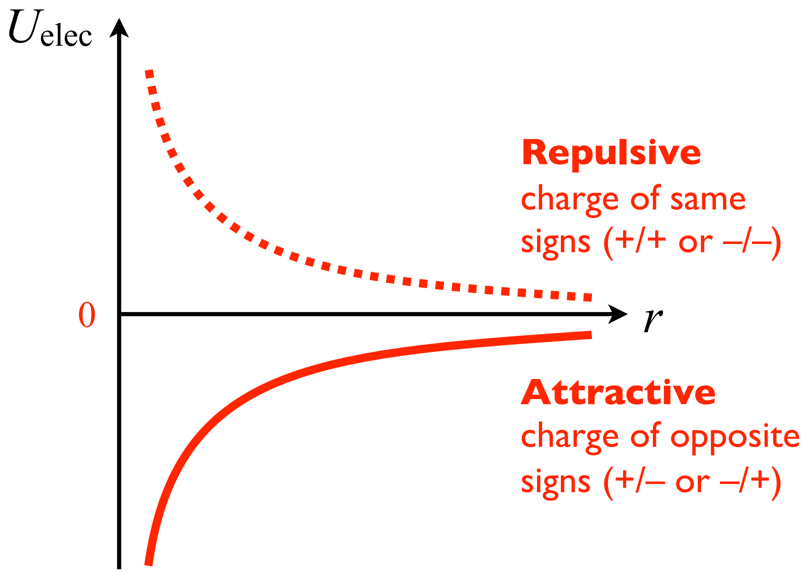
# Electrostatics

**atomic partial charges**

$$U = \ldots + \sum_{\text{nonbonded}} \frac{q_i q_j}{r_{ij}}$$

1/*r* term :
Models the Coulomb forces

**Attractive** or **repulsive**, depending on the charges



**Repulsive**
charge of same signs (+/+ or –/–)

**Attractive**
charge of opposite signs (+/– or –/+)

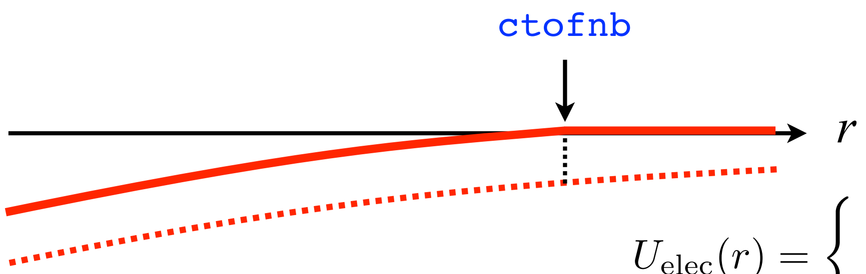
# par_all22_prot.inp (CHARMM)

```
NONBONDED nbxmod  5 atom cdiel shift vatom vdistance vswitch -
cutnb 14.0 ctofnb 12.0 ctonnb 10.0 eps 1.0 e14fac 1.0 wmin 1.5
                !adm jr., 5/08/91, suggested cutoff scheme
!
!V(Lennard-Jones) = Eps,i,j[(Rmin,i,j/ri,j)**12 - 2(Rmin,i,j/ri,j)**6]
!
!epsilon: kcal/mole, Eps,i,j = sqrt(eps,i * eps,j)
!Rmin/2: A, Rmin,i,j = Rmin/2,i + Rmin/2,j
!
!atom  ignored    epsilon      Rmin/2   ignored   eps,1-4       Rmin/2,1-4
!
!carbons
C      0.000000  -0.110000     2.000000 ! ALLOW   PEP POL ARO
                ! NMA pure solvent, adm jr., 3/3/93
CA     0.000000  -0.070000     1.992400 ! ALLOW   ARO
                ! benzene (JES)
CC     0.000000  -0.070000     2.000000 ! ALLOW   PEP POL ARO
                ! adm jr. 3/3/92, acetic acid heat of solvation
CD     0.000000  -0.070000     2.000000 ! ALLOW  POL
                ! adm jr. 3/19/92, acetate a.i. and dH of solvation
CE1    0.000000  -0.068000     2.090000 !
         ! for propene, yin/adm jr., 12/95
CE2    0.000000  -0.064000     2.080000 !
         ! for ethene, yin/adm jr., 12/95
CM     0.000000  -0.110000     2.100000 ! ALLOW HEM
                ! Heme (6-liganded): CO ligand carbon (KK 05/13/91)
CP1    0.000000  -0.020000     2.275000   0.000000  -0.010000     1.900000 ! ALLOW
                ! alkane update, adm jr., 3/2/92
CP2    0.000000  -0.055000     2.175000   0.000000  -0.010000     1.900000 ! ALLOW
                ! alkane update, adm jr., 3/2/92
CP3    0.000000  -0.055000     2.175000   0.000000  -0.010000     1.900000 ! ALLOW
                ! alkane update, adm jr., 3/2/92
```

# Cutoff scheme

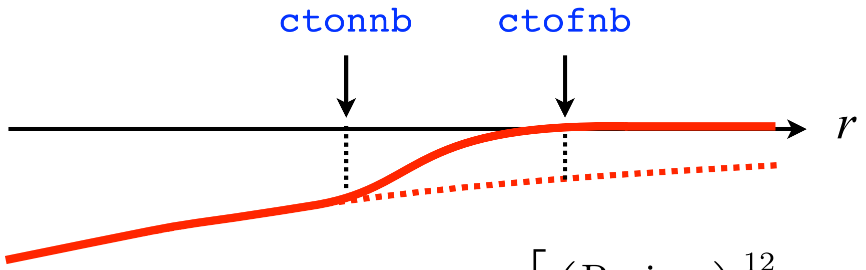
```
NONBONDED nbxmod 5 atom cdiel shift vatom vdistance vswitch -
cutnb 14.0 ctofnb 12.0 ctonnb 10.0 eps 1.0 e14fac 1.0 wmin 1.5
               !adm jr., 5/08/91, suggested cutoff scheme
```

“`shift`”
(for electrostatics)



$$U_{\text{elec}}(r) = \begin{cases} \frac{q_i q_j}{r}\left(1-\frac{r^2}{\text{ctofnb}^2}\right)^2 & \text{if } r < \text{ctofnb} \\ 0 & \text{otherwise} \end{cases}$$

“`vswitch`”
(for Lennard-Jones)



$$U_{\text{LJ}}(r) = \epsilon_{ij}\left[\left(\frac{\text{Rmin}_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{\text{Rmin}_{ij}}{r_{ij}}\right)^{6}\right] \times S(r)$$

$$S(r) = \begin{cases} 1 & \text{if } r < \text{ctonnb} \\ 0 & \text{if } r > \text{ctofnb} \\ \frac{\left(\text{ctofnb}^2 - r^2\right)^2\left[\left(\text{ctofnb}^2 - r^2\right) - 3\left(\text{ctonnb}^2 - r^2\right)\right]}{\left(\text{ctofnb}^2 - \text{ctonnb}^2\right)^3} & \text{otherwise} \end{cases}$$

# Force field parameterization

$$
\begin{aligned}
U = &\sum_{\text{bonds}} k_b(b-b_0)^2 + \sum_{\text{angles}} k_\theta(\theta-\theta_0)^2 \\
&+ \sum_{\text{dihedrals}} k_\phi[1+\cos(n\phi-\delta)] \\
&+ \sum_{\text{impropers}} k_\omega(\omega-\omega_0)^2 + \sum_{\text{Urey-Bradley}} k_u(u-u_0)^2 \\
&+ \sum_{\alpha\text{ carbons}} \text{CMAP}(\phi,\psi) \\
&+ \sum_{\text{nonbonded}} \epsilon_{ij}\left[\left(\frac{\text{Rmin}_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{\text{Rmin}_{ij}}{r_{ij}}\right)^{6}\right] \\
&+ \sum_{\text{nonbonded}} \frac{q_i q_j}{r_{ij}}
\end{aligned}
$$

| Term | Target data | Source |
|---|---|---|
| **Bonded** | | |
| Equilibrium terms ($b_0$, $\theta_0$, $n$, $\delta$) | Geometries | QM, electron diffraction, microwave, crystal surveys |
| Force constants ($K_b$, $K_\theta$, $K_\chi$) | Vibration spectra | QM, IR, Raman |
| **Nonbonded** | | |
| VdW terms ($\varepsilon_i$, $Rmin_i$) | Pure solvent properties ($\Delta_{vap}H$, molecular volume) | •Vapor pressure, calorimetry, densities |
| | Crystal properties ($\Delta_{sub}H$, lattice parameters, non-bonded distances) | •X-ray and neutron diffraction, vapor pressure, calorimetry |
| | •Interaction energies (dimers, interaction with rare gases, interaction with water) | •QM, microwave, mass spectrometry |
| Partial charges ($q_i$) | •Dipole moments | •QM, dielectric permittivity, Stark effect, microwave |
| | •Electrostatic potentials | •QM |
| | •Interactions energies (dimers, interaction with water) | •QM, microwave, mass spectrometry |
| | Aqueous solution ($\Delta_{solv}G$, $\Delta_{solv}H$, partial molar volume) | •Calorimetry, volume variations |

Reproduced from : MacKerell, "Atomistic Models and Force Fields" in *Computational Biochemistry and Biophysics* (Editors: Becker, MacKerell, Roux, Watanabe), 2001, p. 7–38.

## We need two things :

**Force field**
List of parameters associated with each atom type and their combinations (bonds, angles, dihedrals)

**+**

**Topology**
Description of all molecules and residues in term of their atoms and covalent bonds
+ partial atomic charges…

## top_all27_prot_lipid.rtf (CHARMM)

```
MASS     1 H       1.00800 H ! polar H
MASS     2 HC      1.00800 H ! N-ter H
MASS     3 HA      1.00800 H ! nonpolar H
MASS     4 HT      1.00800 H ! TIPS3P WATER HYDROGEN
MASS     5 HP      1.00800 H ! aromatic H
MASS     6 HB      1.00800 H ! backbone H
MASS     7 HR1     1.00800 H ! his he1, (+) his HG,HD2
MASS     8 HR2     1.00800 H ! (+) his HE1
MASS     9 HR3     1.00800 H ! neutral his HG, HD2
MASS    10 HS      1.00800 H ! thiol hydrogen
MASS    11 HE1     1.00800 H ! for alkene; RHC=CR
MASS    12 HE2     1.00800 H ! for alkene; H2C=CR
MASS    13 HA1     1.00800 H ! alkane, CH, new LJ params (see toppar_all22_prot_alipha
MASS    14 HA2     1.00800 H ! alkane, CH2, new LJ params (see toppar_all22_prot_aliph
MASS    15 HA3     1.00800 H ! alkane, CH3, new LJ params (see toppar_all22_prot_aliph
MASS    16 HF1     1.00800 H ! Aliphatic H on fluorinated C (see toppar_all22_prot_flu
MASS    17 HF2     1.00800 H ! Aliphatic H on fluorinated C (see toppar_all22_prot_flu
MASS    20 C      12.01100 C ! carbonyl C, peptide backbone
MASS    21 CA     12.01100 C ! aromatic C
MASS    22 CT1    12.01100 C ! aliphatic sp3 C for CH
MASS    23 CT2    12.01100 C ! aliphatic sp3 C for CH2
MASS    24 CT3    12.01100 C ! aliphatic sp3 C for CH3
MASS    25 CPH1   12.01100 C ! his CG and CD2 carbons
MASS    26 CPH2   12.01100 C ! his CE1 carbon
MASS    27 CPT    12.01100 C ! trp C between rings
MASS    28 CY     12.01100 C ! TRP C in pyrrole ring
MASS    29 CP1    12.01100 C ! tetrahedral C (proline CA)
MASS    30 CP2    12.01100 C ! tetrahedral C (proline CB/CG)
MASS    31 CP3    12.01100 C ! tetrahedral C (proline CD)
MASS    32 CC     12.01100 C ! carbonyl C, asn,asp,gln,glu,cter,ct2
MASS    33 CD     12.01100 C ! carbonyl C, pres aspp,glup,ct1
MASS    34 CPA    12.01100 C ! heme alpha-C
```

# top_all27_prot_lipid.rtf (CHARMM)

atom types vs atom names

```
RESI ALA          0.00
GROUP
ATOM N    NH1    -0.47  !     |
ATOM HN   H       0.31  !  HN-N
ATOM CA   CT1     0.07  !     |     HB1
ATOM HA   HB      0.09  !     |    /
GROUP                   !  HA-CA--CB-HB2
ATOM CB   CT3    -0.27  !     |    \
ATOM HB1  HA      0.09  !     |     HB3
ATOM HB2  HA      0.09  !   O=C
ATOM HB3  HA      0.09  !     |
GROUP                   !
ATOM C    C       0.51
ATOM O    O      -0.51
BOND CB CA  N  HN  N  CA
BOND C  CA  C  +N  CA HA  CB HB1  CB HB2  CB HB3
DOUBLE O  C
IMPR N -C CA HN  C CA +N O
CMAP -C  N  CA  C   N  CA  C  +N
DONOR HN N
ACCEPTOR O C
IC -C   CA   *N   HN    1.3551 126.4900  180.0000 115.4200  0.9996
IC -C   N    CA   C     1.3551 126.4900  180.0000 114.4400  1.5390
IC N    CA   C    +N    1.4592 114.4400  180.0000 116.8400  1.3558
IC +N   CA   *C   O     1.3558 116.8400  180.0000 122.5200  1.2297
IC CA   C    +N   +CA   1.5390 116.8400  180.0000 126.7700  1.4613
IC N    C    *CA  CB    1.4592 114.4400  123.2300 111.0900  1.5461
IC N    C    *CA  HA    1.4592 114.4400 -120.4500 106.3900  1.0840
IC C    CA   CB   HB1   1.5390 111.0900  177.2500 109.6000  1.1109
IC HB1  CA   *CB  HB2   1.1109 109.6000  119.1300 111.0500  1.1119
IC HB1  CA   *CB  HB3   1.1109 109.6000 -119.5800 111.6100  1.1114
```

# top_all27_prot_lipid.rtf (CHARMM)

```
RESI HSD          0.00  ! neutral HIS, proton on ND1
GROUP
ATOM N    NH1    -0.47  !     |         HD1    HE1
ATOM HN   H       0.31  !  HN-N           |      /
ATOM CA   CT1     0.07  !     |   HB1    ND1--CE1
ATOM HA   HB      0.09  !     |   |     /      ||
GROUP                   !  HA-CA--CB--CG       ||
ATOM CB   CT2    -0.09  !     |   |     \\     ||
ATOM HB1  HA      0.09  !     |   HB2    CD2--NE2
ATOM HB2  HA      0.09  !   O=C           |
ATOM ND1  NR1    -0.36  !     |         HD2
ATOM HD1  H       0.32
ATOM CG   CPH1   -0.05
GROUP
ATOM CE1  CPH2    0.25
ATOM HE1  HR1     0.13
ATOM NE2  NR2    -0.70
ATOM CD2  CPH1    0.22
ATOM HD2  HR3     0.10
GROUP
ATOM C    C       0.51
ATOM O    O      -0.51
BOND CB  CA   CG  CB   ND1 CG   CE1 ND1
BOND NE2 CD2  N    HN   N   CA
BOND C   CA   C   +N   CA  HA   CB  HB1
BOND CB  HB2  ND1 HD1  CD2 HD2  CE1 HE1
DOUBLE O  C   CG  CD2   CE1  NE2
IMPR ND1 CG CE1 HD1  CD2 CG  NE2 HD2  CE1 ND1 NE2 HE1
IMPR ND1 CE1 CG HD1  CD2 NE2 CG  HD2  CE1 NE2 ND1 HE1
IMPR N   -C  CA HN   C   CA  +N  O
CMAP -C  N  CA  C   N  CA  C  +N
DONOR HN N
```

# top_all27_prot_lipid.rtf (CHARMM)

```
RESI HSE          0.00  ! neutral His, proton on NE2
GROUP
ATOM N    NH1    -0.47  !     |                 HE1
ATOM HN   H       0.31  !  HN-N             __  /
ATOM CA   CT1     0.07  !     |   HB1    ND1--CE1
ATOM HA   HB      0.09  !     |   |     /      |
GROUP                   !  HA-CA--CB--CG       |
ATOM CB   CT2    -0.08  !     |   |     \\     |
ATOM HB1  HA      0.09  !     |   HB2    CD2--NE2
ATOM HB2  HA      0.09  !   O=C           |     \
ATOM ND1  NR2    -0.70  !     |          HD2    HE2
ATOM CG   CPH1    0.22
ATOM CE1  CPH2    0.25
ATOM HE1  HR1     0.13
GROUP
ATOM NE2  NR1    -0.36
ATOM HE2  H       0.32
ATOM CD2  CPH1   -0.05
ATOM HD2  HR3     0.09
GROUP
ATOM C    C       0.51
ATOM O    O      -0.51
BOND CB  CA   CG  CB   ND1 CG
BOND NE2 CD2  N   HN   N   CA
BOND C   CA   C   +N   NE2 CE1  CA  HA   CB HB1
BOND CB  HB2  NE2 HE2  CD2 HD2  CE1 HE1
DOUBLE   O   C   CD2 CG   CE1 ND1
IMPR NE2 CD2 CE1 HE2  CD2 CG  NE2 HD2  CE1 ND1 NE2 HE1
IMPR NE2 CE1 CD2 HE2  CD2 NE2 CG  HD2  CE1 NE2 ND1 HE1
IMPR N   -C  CA  HN   C   CA  +N  O
CMAP -C  N  CA  C   N  CA  C  +N
DONOR HN N
```

# top_all27_prot_lipid.rtf (CHARMM)

```
RESI TIP3         0.000 ! tip3p water model, generate using noangle nodihedral
GROUP
ATOM OH2  OT     -0.834
ATOM H1   HT      0.417
ATOM H2   HT      0.417
BOND OH2 H1 OH2 H2 H1 H2    ! the last bond is needed for shake
ANGLE H1 OH2 H2             ! required
ACCEPTOR OH2
PATCHING FIRS NONE LAST NONE

! Ion parameters from Benoit Roux and Coworkers
! As of 8/98 no NBFIX terms required
!
RESI SOD       1.00 ! Sodium Ion
GROUP
ATOM SOD  SOD  1.00
PATCHING FIRST NONE LAST NONE

RESI MG        2.00 ! Magnesium Ion
GROUP
ATOM MG   MG   2.00
PATCHING FIRST NONE LAST NONE

RESI POT       1.00 ! Potassium Ion
GROUP
ATOM POT   POT 1.00
PATCHING FIRST NONE LAST NONE

RESI CES       1.00 ! Cesium Ion
GROUP
ATOM CES  CES  1.00
PATCHING FIRST NONE LAST NONE
```

# **Warning :**
## Empirical force fields have limited transferability.

**They should be used in the context for which they were calibrated :**

- Amino acids around other amino acids (or around water)
- Ions in water
- Water around polar compounds

**Unusual combinations should be tested :**

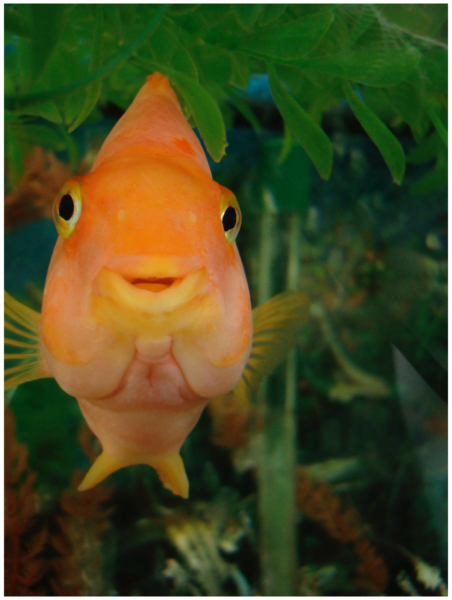
- ions next to nonpolar / aromatic residues
- water in small cavities
- metal binding

**For some of these special situations, <u>polarizable</u> force fields may be required.**

**But luckily :**

- "Like dissolves like" (usually)

**"Good force field"**



**"Bad force field"**