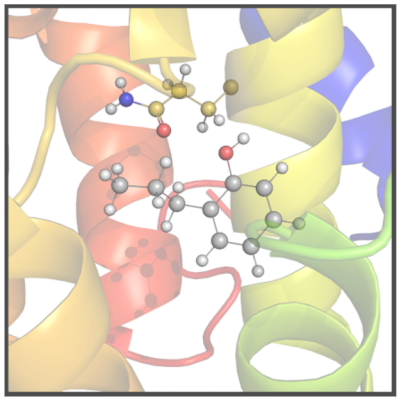
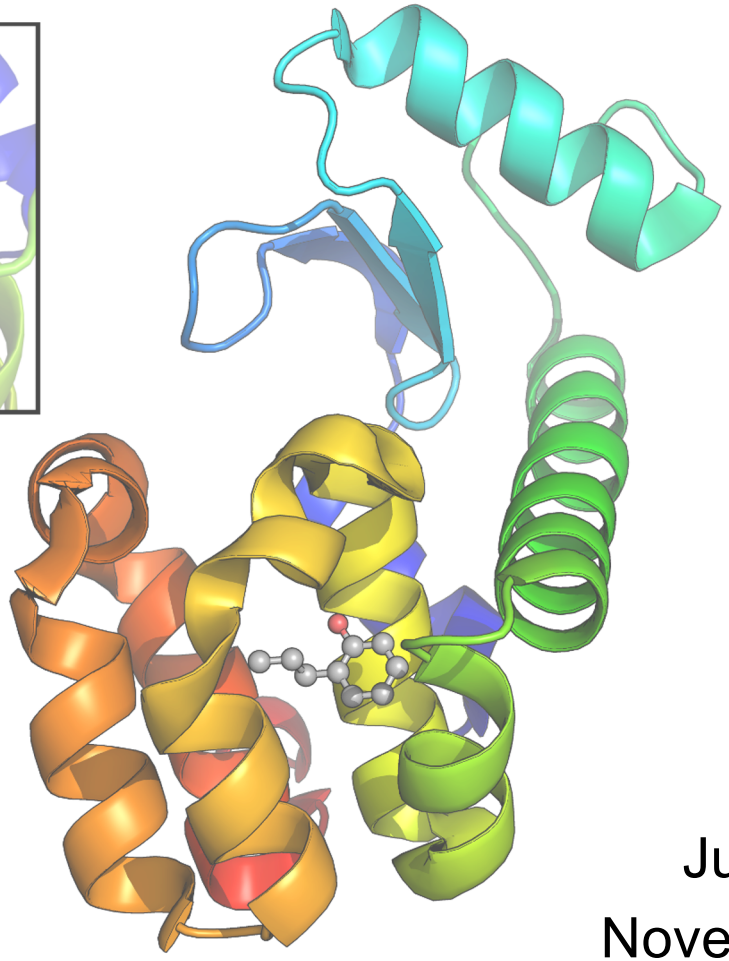
CHARMM Force Field Development History, Features, and Implementation in GROMACS



CHARMM

⇅

GROMACS



Justin A. Lemkul

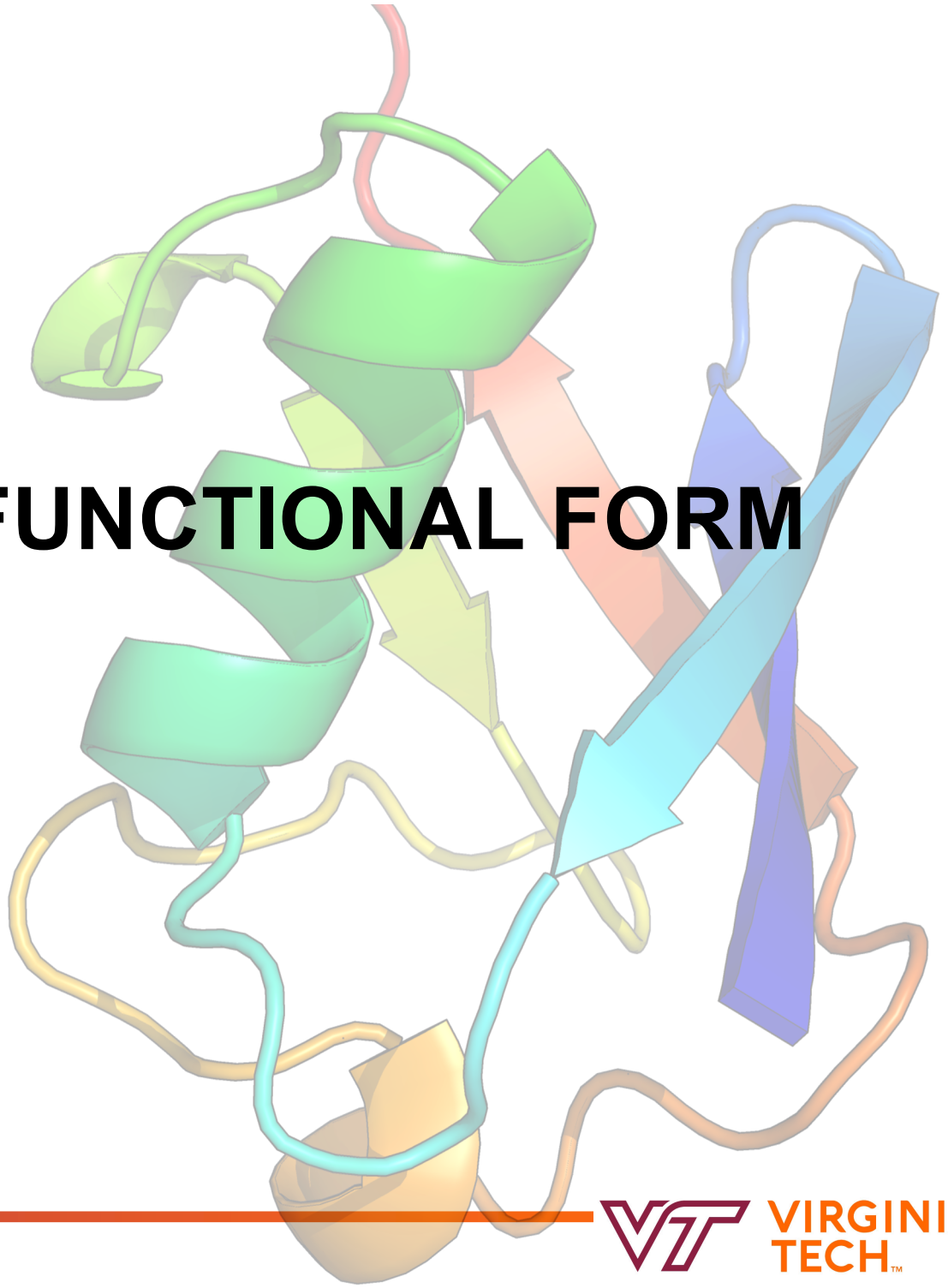
November 23, 2021

@JustinLemkulVT

jalemkul@vt.edu

VIRGINIA TECH™

I. CHARMM FUNCTIONAL FORM



The CHARMM Functional Form

$$
\begin{aligned}
& U(\mathbf{r}) \\
&= \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 \\
&+ \sum_{\text{U-B}} K_{UB}\left(S - S_0\right)^2 + \sum_{\text{dihedrals}} K_\phi[1 + \cos(n\phi - \delta)] \\
&+ \sum_{\text{impropers}} K_\omega(\omega - \omega_0)^2 + \sum_{\text{residues}} u_{CMAP}(\phi, \psi) \\
&+ \sum_{\text{NB pairs}} \varepsilon_{ij}\left[\left(\frac{r_{min}^{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{min}^{ij}}{r_{ij}}\right)^{6}\right] + \sum_{\text{NB pairs}} \frac{q_i q_j}{\varepsilon r_{ij}}
\end{aligned}
$$

The CHARMM Functional Form

$$U(\mathbf{r})$$

Bonded

$$= \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2$$

$$+ \sum_{\text{U-B}} K_{UB}\left(S - S_0\right)^2 + \sum_{\text{dihedrals}} K_\phi[1 + \cos(n\phi - \delta)]$$

$$+ \sum_{\text{impropers}} K_\omega(\omega - \omega_0)^2 + \sum_{\text{residues}} u_{CMAP}(\phi, \psi)$$

$$+ \sum_{\text{NB pairs}} \varepsilon_{ij}\left[\left(\frac{r_{min}^{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{min}^{ij}}{r_{ij}}\right)^{6}\right] + \sum_{\text{NB pairs}} \frac{q_i q_j}{\varepsilon r_{ij}}$$

The CHARMM Functional Form

$$
\begin{aligned}
& U(\mathbf{r}) \\
&= \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 \\
&+ \sum_{\text{U-B}} K_{UB}\left(S - S_0\right)^2 + \sum_{\text{dihedrals}} K_\phi[1 + \cos(n\phi - \delta)] \\
&+ \sum_{\text{impropers}} K_\omega(\omega - \omega_0)^2 + \sum_{\text{residues}} u_{CMAP}(\phi, \psi) \\
&+ \sum_{\text{NB pairs}} \varepsilon_{ij}\left[\left(\frac{r_{min}^{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{min}^{ij}}{r_{ij}}\right)^{6}\right] + \sum_{\text{NB pairs}} \frac{q_i q_j}{\varepsilon r_{ij}}
\end{aligned}
$$

Nonbonded

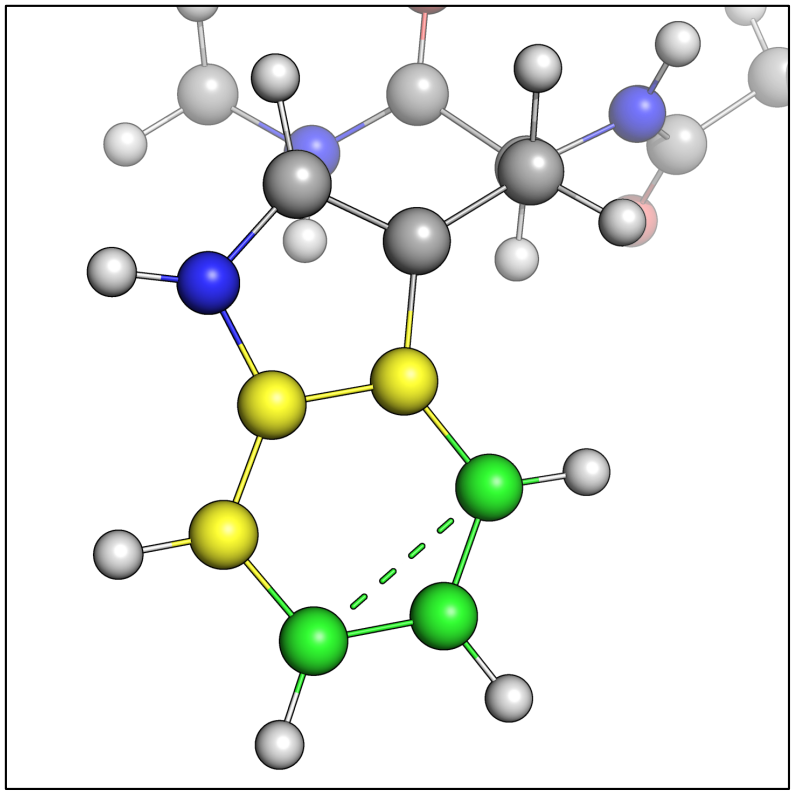
The CHARMM Functional Form

Bonded
Nonbonded

Urey-Bradley Angle Bending

```
[ angletypes ]
; i    j    k   func     theta0        ktheta          ub0          kub
CAI    CA   CA    5   120.000000   334.720000   0.24162000     29288.00
CPT    CPT  CAI   5   110.000000   418.400000   0.00000000         0.00
```

$$U(\mathbf{r}) = \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{U-B}} K_{UB}\left(S - S_0\right)^2$$



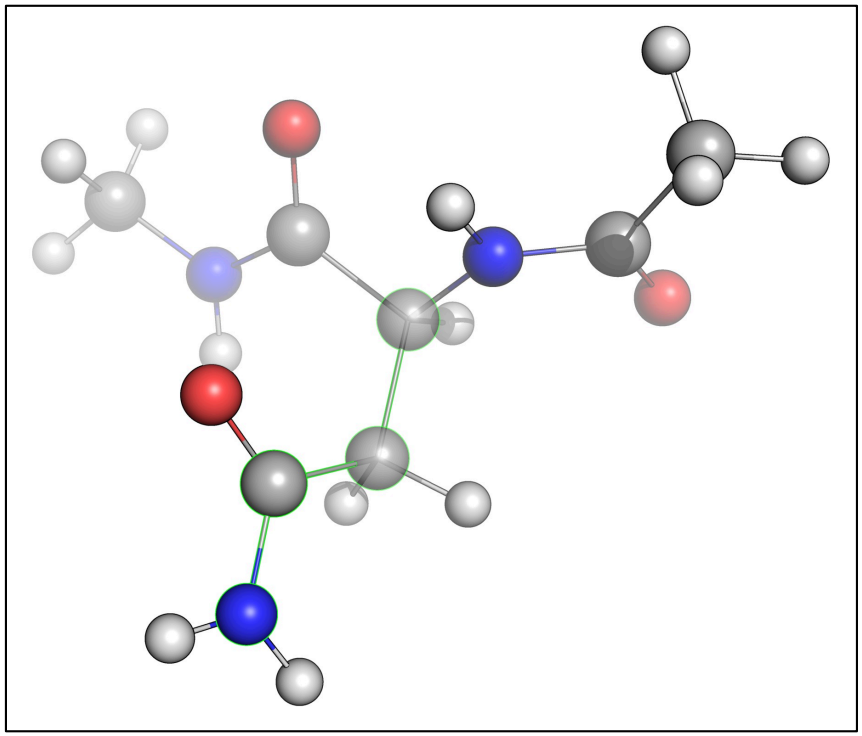
The CHARMM Functional Form

Bonded
Nonbonded

Dihedrals summed over multiplicities

```
[ dihedraltypes ]
;    i      j      k      l   func         phi0          kphi   mult
   CT1    CT2     CC    NH2      9   180.000000      2.594080      1
   CT1    CT2     CC    NH2      9   180.000000      2.761440      2
   CT1    CT2     CC    NH2      9   180.000000      3.012480      3
```

$$U(\mathbf{r}) = \sum_{\text{dihedrals}} K_\phi[1 + \cos(n\phi - \delta)]$$



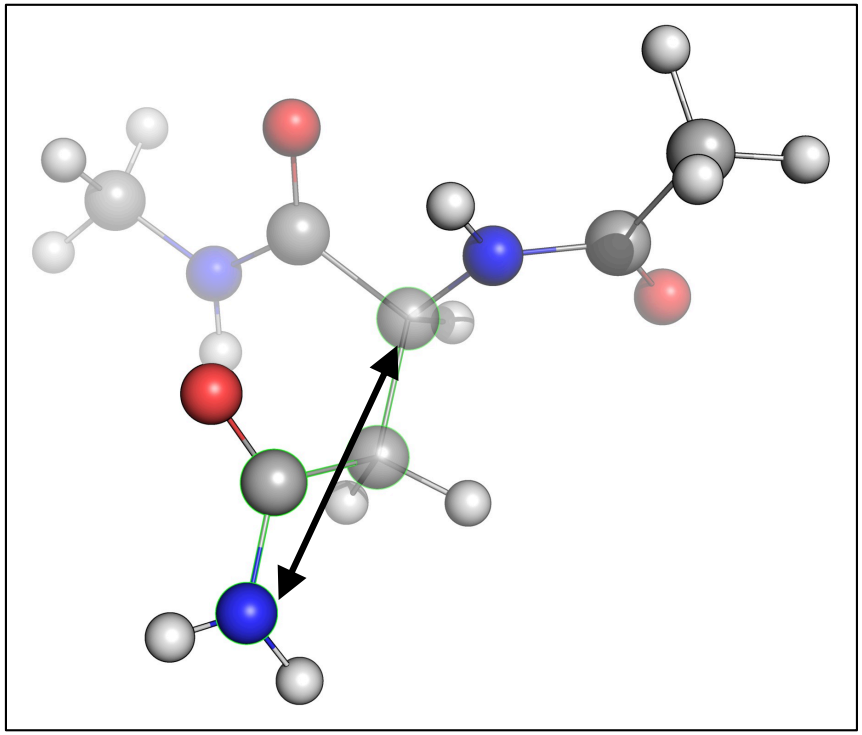
The CHARMM Functional Form

Bonded
Nonbonded

No scaling applied to 1-4 pairs

```
[ defaults ]
; nbfunc    comb-rule    gen-pairs    fudgeLJ fudgeQQ
      1            2          yes        1.0     1.0
```

Dihedral parameters and 1-4 scaling factors are directly linked!

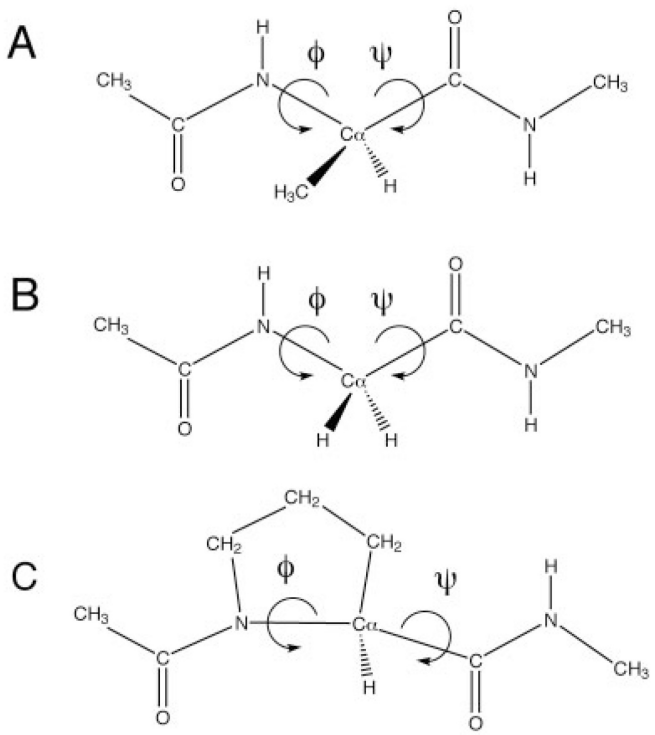


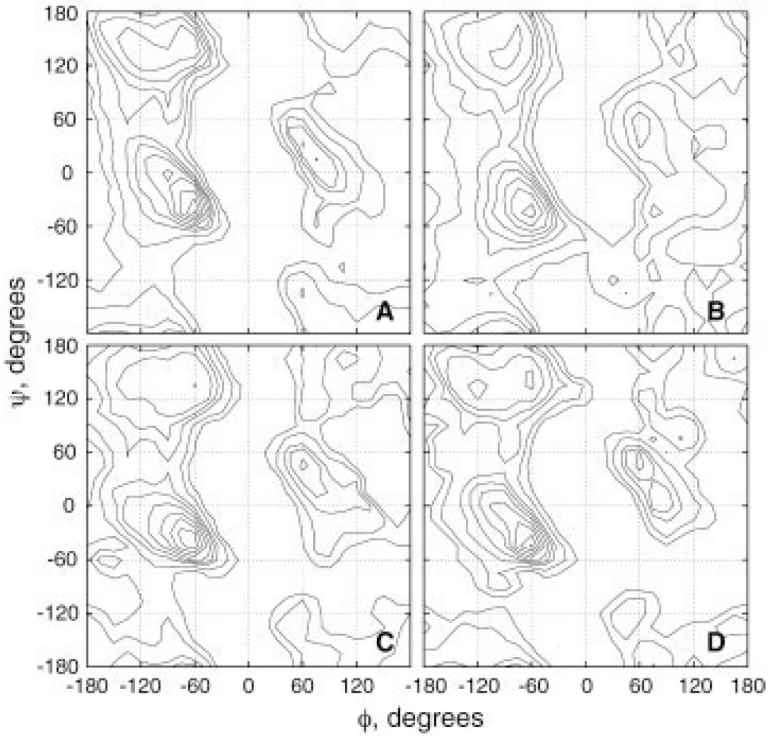
The CHARMM Functional Form

Bonded
Nonbonded

CMAP applied to (ϕ,ψ) pairs

```
[ cmaptypes ]
C NH1 CT1 C NH1 1 24 24\
0.54392000 3.22168000 4.05848000 5.23000000 8.87008000 11.38048000
8.74456000 7.48936000 3.26352000 -2.88696000\
```





A.D. MacKerell Jr., M. Feig, C.L. Brooks III (2004) *J. Comput. Chem.* 25: 1400-1415.

VIRGINIA TECH

The CHARMM Functional Form

Bonded
Nonbonded

$$F_{switch}^{LJ}(r_{ij}) = S(r_{ij})F^{LJ}(r_{ij})$$

$$S(r_{ij}) = \begin{cases} 1, & r_{ij} \leq r_{on} \\ \dfrac{(r_{off}^2 - r_{ij}^2)^2(r_{off}^2 + 2r_{ij}^2 - 3r_{on}^2)}{(r_{off}^2 - r_{on}^2)^3}, & r_{on} < r_{ij} \leq r_{off} \end{cases}$$

```
nbonds atom ewald pmewald kappa 0.32 -
fftx 32 ffty 32 fftz 32 order 6 -
vatom vfswitch -
cutnb 14.0 ctofnb 12.0 ctonnb 10.0
```

```
cutoff-scheme   = Verlet
vdwtype         = cutoff
vdw-modifier    = force-switch
rlist           = 1.2
rvdw            = 1.2
rvdw-switch     = 1.0
coulombtype     = PME
rcoulomb        = 1.2
dispcorr        = no
```

P.J. Steinbach and B.R. Brooks (1994) *J. Comput. Chem. 15*: 667-683.

VIRGINIA TECH™

II. PARAMETRIZATION STRATEGY

CHARMM Parametrization

QM Optimized Geometry

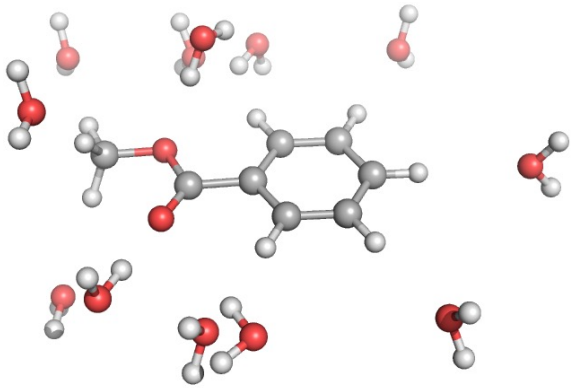
Initial Parameters by Analogy

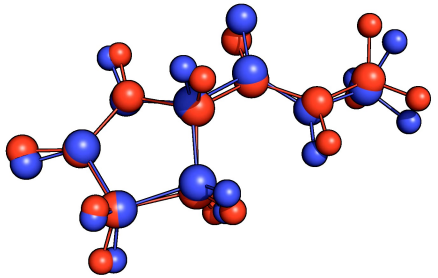
Bonded Optimization
Geometry
Vibrational Frequencies

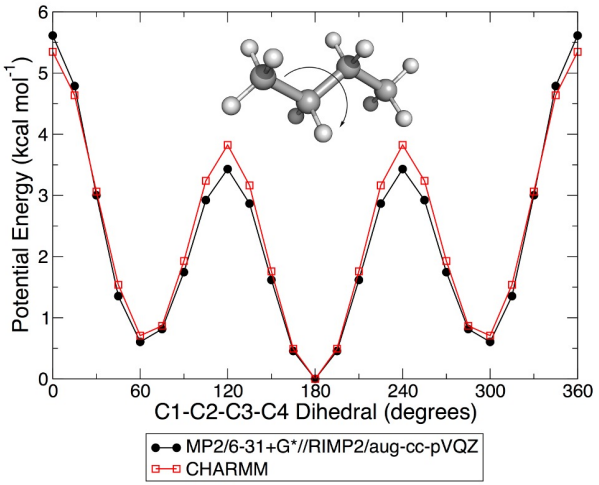
Electrostatic Optimization
Dipole moment
Water Interactions

Dihedral Parametrization
1D/2D surfaces

Empirical Validation







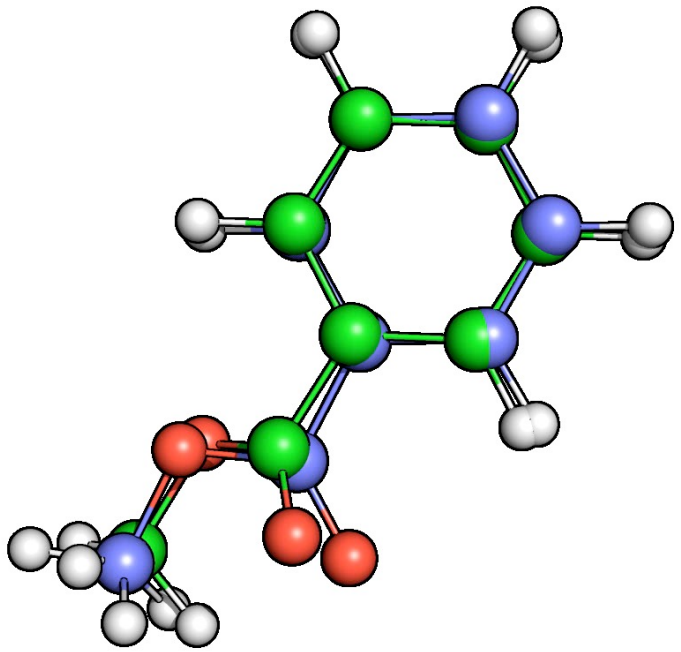
CHARMM Parametrization

MP2/6-31G* model chemistry

Optimized geometry

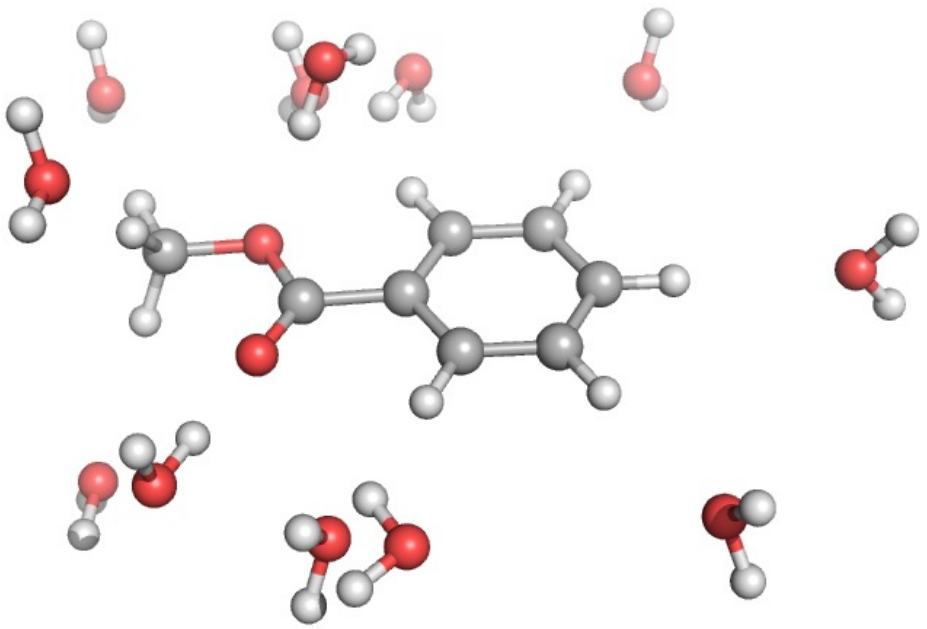
- Check bond lengths, valence angles, dihedral angles
- Vibrational frequency analysis

QM geometry used for initial steps of parametrization (check later with relaxed MM geometry)



QM optimized geometry
MM initial geometry

CHARMM Parametrization



Merz-Kollman charges may be used as a starting point

Target MP2/6-31G* dipole moment (overestimate ~20%)

MP2/6-31G* rigid scan of H-bond distance

- Fixed QM geometry of compound
- Water fixed in TIP3P geometry

HF/6-31G* single point interaction energy

- Shift QM minimum -0.2 Å
- Scale QM interaction energy x 1.16 (neutral compounds)

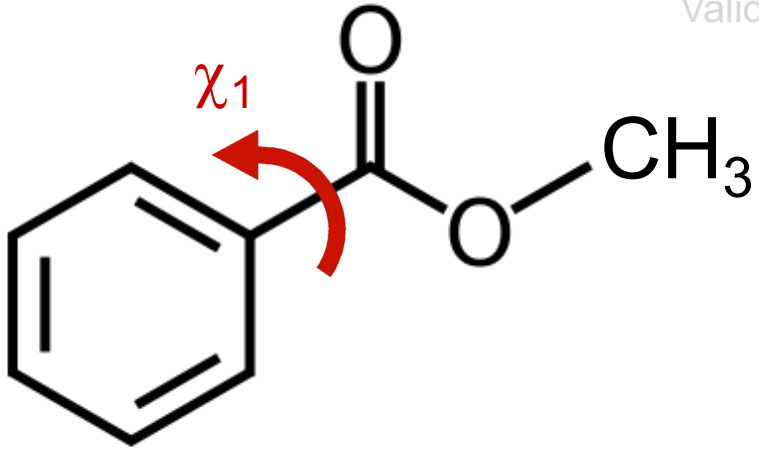
CHARMM Parametrization

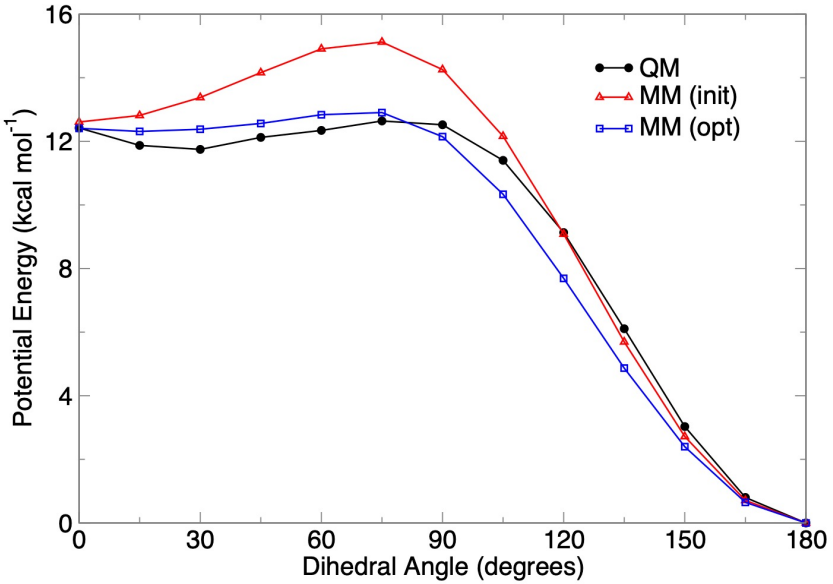
MP2/6-31G* model chemistry

Begin with QM optimized geometry

- Scan one degree of freedom (typically), 0 – 180° or 0 – 345° by 15°
- QM optimization, all other degrees of freedom can relax

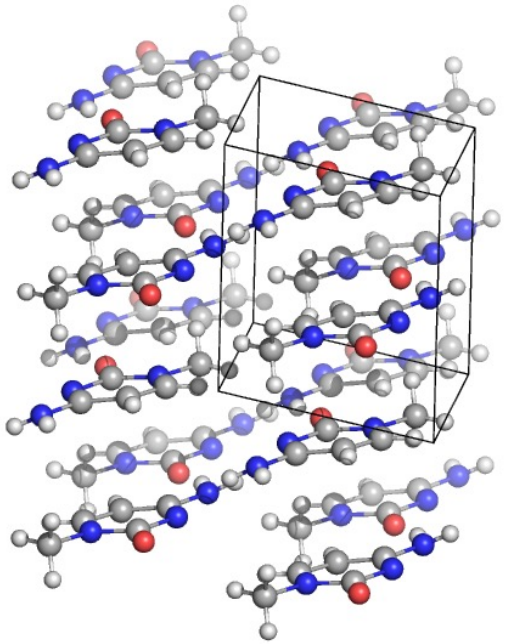
QM energy values targeted, ignore >12 kcal mol^{-1}, offset to global minimum



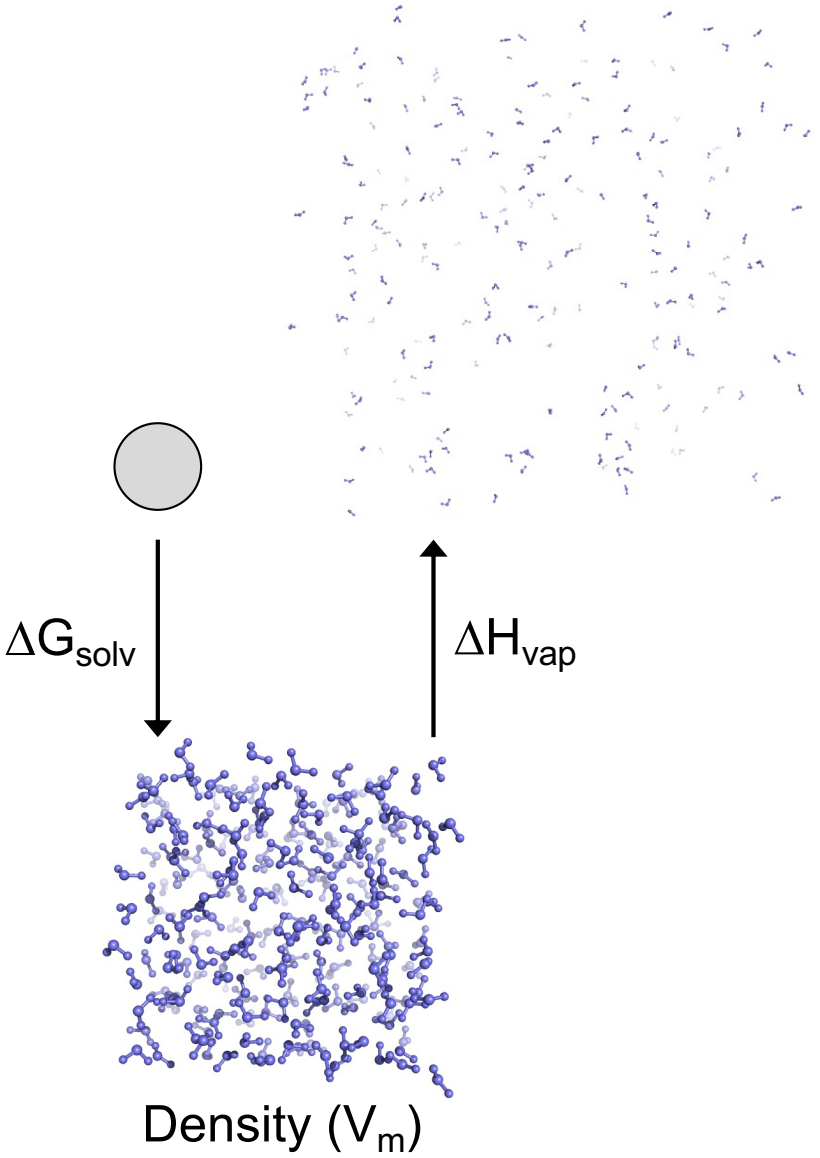



```
OG2D1  CG2O2  CG2R61  CG2R61  1.0250  2  180
OG302  CG2O2  CG2R61  CG2R61  0.8500  2  180
```

CHARMM Parametrization



Lattice parameters (A,B,C, α,β,γ)
Molecular volume (V_m)



CHARMM General Force Field (CGenFF)

CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields

K. VANOMMESLAEGHE, E. HATCHER, C. ACHARYA, S. KUNDU, S. ZHONG, J. SHIM, E. DARIAN, O. GUVENCH, P. LOPES, I. VOROBYOV, A. D. MACKERELL JR.
Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland 21201

JOURNAL OF CHEMICAL INFORMATION AND MODELING — Article — pubs.acs.org/jcim

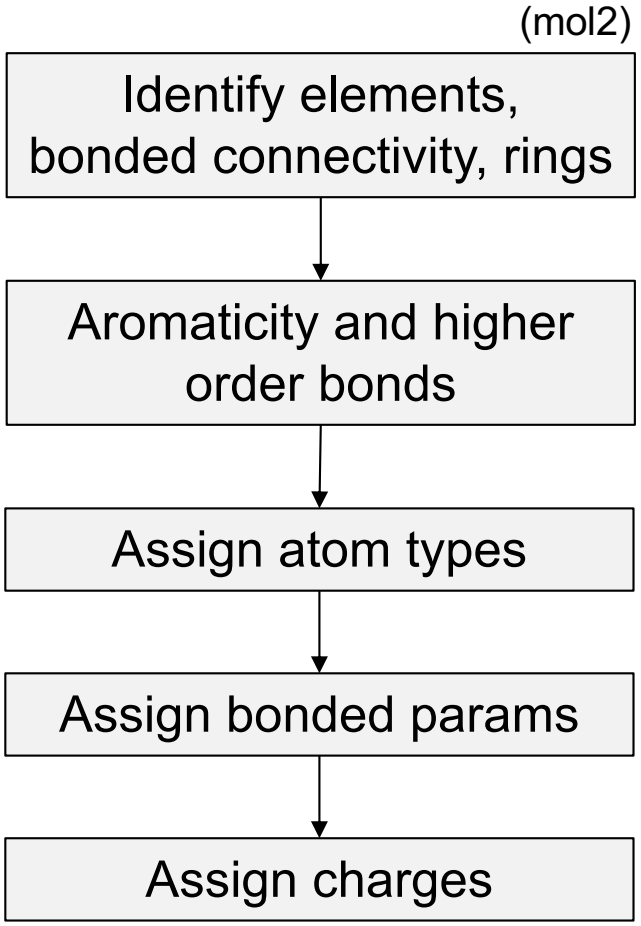
Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing

K. Vanommeslaeghe and A. D. MacKerell, Jr.*

JOURNAL OF CHEMICAL INFORMATION AND MODELING — Article — pubs.acs.org/jcim

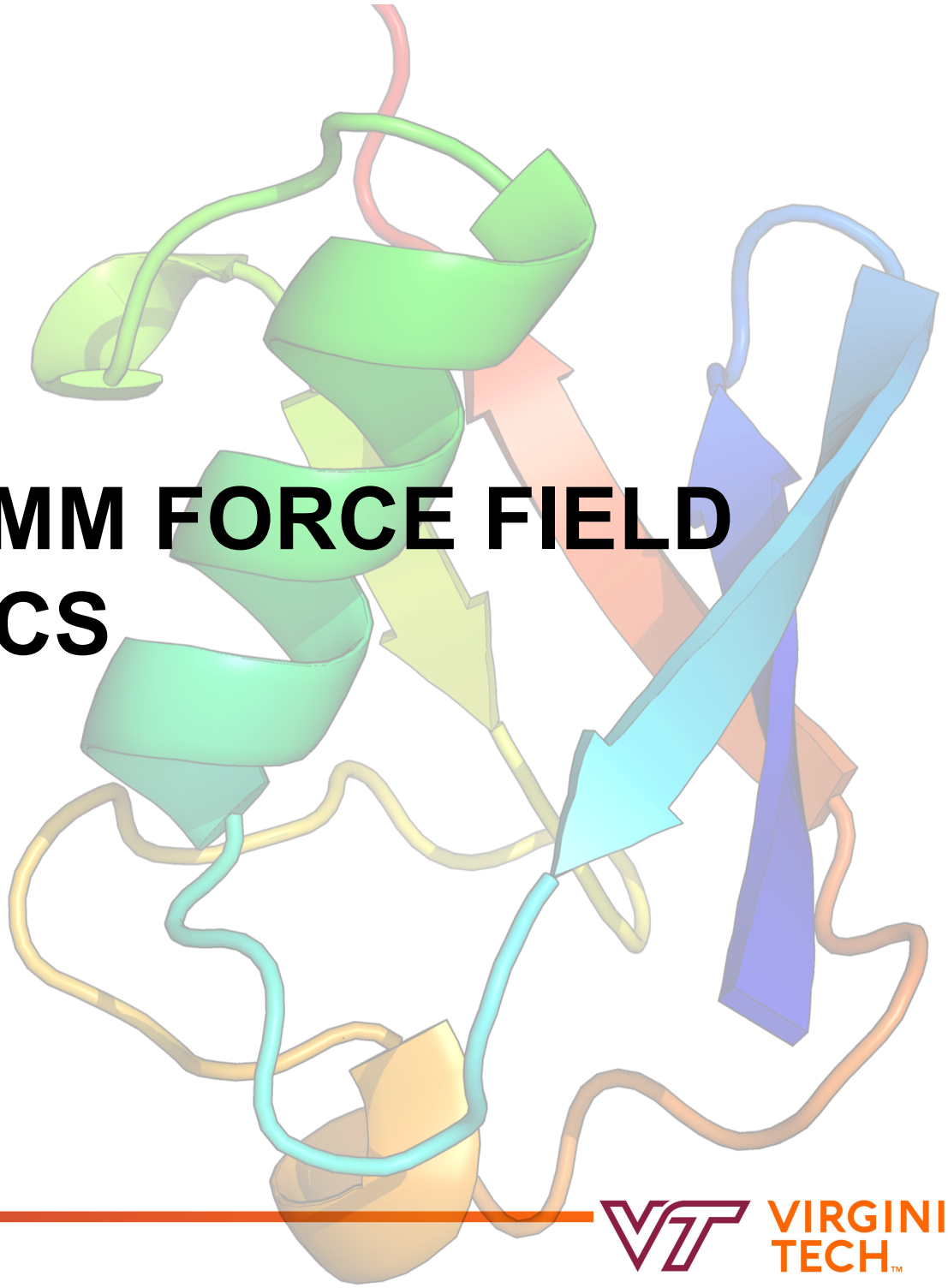
Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges

K. Vanommeslaeghe, E. Prabhu Raman, and A. D. MacKerell, Jr.*



https://cgenff.umaryland.edu/

III. THE CHARMM FORCE FIELD IN GROMACS



The CHARMM36 Port for GROMACS

http://mackerell.umaryland.edu/charmm_ff.shtml#gromacs

- Current release (July 2021): two versions
  - charmm36-jul2021.ff
  - charmm36_ljpme-jul2021.ff
- LJ-PME version supports new lipid topologies and parameters (Y. Yu et al. *JCTC* 2021, 17: 1581-1595)
- July 2021 version: some changes
  - New automated build system by András Wacha
  - Splitting of .rtp, .tdb, etc files based on molecule type rather than monolithic “merged” files
- CHARMM36m for proteins, can use modified TIP3P
  - ε = -0.1 kcal mol^{-1} for water-solute interactions

The CHARMM36 Port for GROMACS

- Validation is performed against a range of molecules (polypeptides including all amino acids, oligonucleotides, monosaccharides, lipids, water and salt solutions)

```
=== Protein ===
AA          CHARMM          GMX
ala       21.76441     21.76417
arg     -205.19292   -205.18738
asp       10.40257     10.40301
asn      -39.89309    -39.89149
cys       32.49317     32.49282
...

=== DNA ===
BASE        CHARMM           GMX
ade       46.23869      46.24115
cyt     -193.93475    -193.92638
gua     -273.53232    -273.52294
thy     -181.94695    -181.93905
```

```
=== LIPID ===
LIP         CHARMM           GMX
popc     11.72219      11.72342
pope      0.52813       0.52956
pops    -18.81789     -18.81596
popa    -40.71058     -40.70841
dppc      7.78802       7.78893

=== CARB ===
SUG         CHARMM           GMX
aglc    105.42133     105.41969
bglc    111.84464     111.84273
agal    106.43506     106.43331
bgal    111.38653     111.38479
```

Using CHARMM36 in GROMACS

J. Lee et al. (2016) *JCTC* 12: 405-413.

CHARMM

Coordinates
Topology
Subset of FF files

Internal coordinate builder
On-the-fly patching

.rtp entries
.hdb entries
.tdb modifications

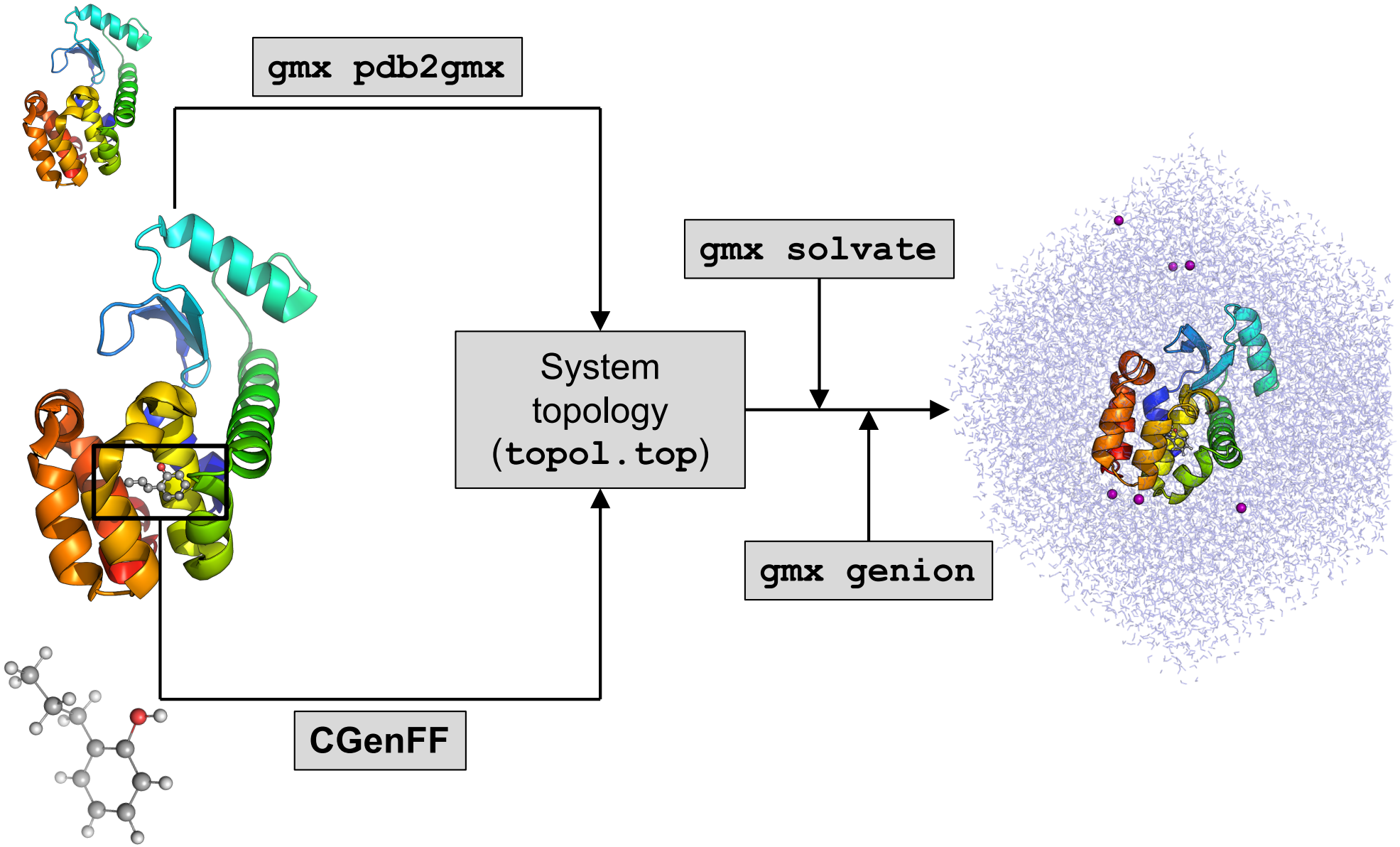
CHARMM36 port

Input
coordinate file

`gmx pdb2gmx`

Coordinates
Topology

Practical Example: Protein-Ligand Complex



http://www.mdtutorials.com/gmx/complex/index.html

Generating a CGenFF Topology

```
@<TRIPOS>MOLECULE
JZ4
 22 22 0 0 0
SMALL
GASTEIGER


@<TRIPOS>ATOM
      1 C4          24.2940  -24.1240   -0.0710 C.3     1  JZ4         -0.0650
      2 C7          21.5530  -27.2140   -4.1120 C.ar    1  JZ4         -0.0613
      3 C8          22.0680  -26.7470   -5.3310 C.ar    1  JZ4         -0.0583
      4 C9          22.6710  -25.5120   -5.4480 C.ar    1  JZ4         -0.0199
      5 C10         22.7690  -24.7300   -4.2950 C.ar    1  JZ4          0.1200
      6 C11         21.6930  -26.4590   -2.9540 C.ar    1  JZ4         -0.0551
      7 C12         22.2940  -25.1870   -3.0750 C.ar    1  JZ4         -0.0060
      8 C13         22.4630  -24.4140   -1.8080 C.3     1  JZ4         -0.0245
      9 C14         23.9250  -24.7040   -1.3940 C.3     1  JZ4         -0.0518
     10 OAB         23.4120  -23.5360   -4.3420 O.3     1  JZ4         -0.5065
     11 H           25.3133  -24.3619    0.1509 H       1  JZ4          0.0230
     12 H           23.6591  -24.5327    0.6872 H       1  JZ4          0.0230
     13 H           24.1744  -23.0611   -0.1016 H       1  JZ4          0.0230
     14 H           21.0673  -28.1238   -4.0754 H       1  JZ4          0.0618
     15 H           21.9931  -27.3472   -6.1672 H       1  JZ4          0.0619
     16 H           23.0361  -25.1783   -6.3537 H       1  JZ4          0.0654
     17 H           21.3701  -26.8143   -2.0405 H       1  JZ4          0.0621
     18 H           21.7794  -24.7551   -1.0588 H       1  JZ4          0.0314
     19 H           22.2659  -23.3694   -1.9301 H       1  JZ4          0.0314
     20 H           24.5755  -24.2929   -2.1375 H       1  JZ4          0.0266
     21 H           24.0241  -25.7662   -1.3110 H       1  JZ4          0.0266
     22 H           23.7394  -23.2120   -5.1580 H       1  JZ4          0.2921
@<TRIPOS>BOND
     1     4     3   ar
     2     4     5   ar
     3     3     2   ar
     4    10     5    1
     5     5     7   ar
     6     2     6   ar
     7     7     6   ar
     8     7     8    1
     9     8     9    1
...
```

Atom names

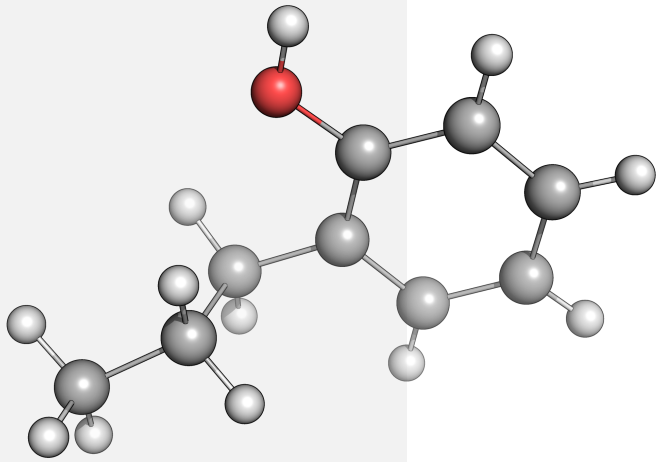
Rudimentary atom types

Bonded connectivity

Non-unique names will be reassigned

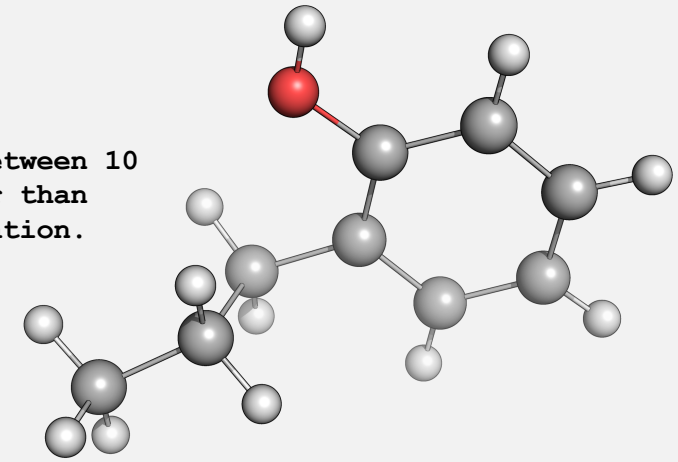
The CGenFF Topology

```
RESI JZ4            0.000 ! param penalty=   0.900 ; charge penalty=   0.342
GROUP              ! CHARGE   CH_PENALTY
ATOM C1     CG331  -0.271 !    0.285
ATOM C2     CG321  -0.183 !    0.045
ATOM C3     CG321  -0.178 !    0.342
ATOM C4     CG2R61 -0.007 !    0.218
ATOM C9     CG2R61  0.103 !    0.030
ATOM O10    OG311  -0.529 !    0.190
ATOM H22    HGP1    0.420 !    0.000
ATOM C8     CG2R61 -0.110 !    0.000
ATOM H21    HGR61   0.115 !    0.000
ATOM C5     CG2R61 -0.115 !    0.000
ATOM H18    HGR61   0.115 !    0.000
ATOM C6     CG2R61 -0.108 !    0.000
ATOM H19    HGR61   0.115 !    0.000
ATOM C7     CG2R61 -0.112 !    0.000
ATOM H20    HGR61   0.115 !    0.000
ATOM H1     HGA3    0.090 !    0.000
ATOM H2     HGA3    0.090 !    0.000
ATOM H3     HGA3    0.090 !    0.000
ATOM H4     HGA2    0.090 !    0.000
ATOM H5     HGA2    0.090 !    0.000
ATOM H6     HGA2    0.090 !    0.000
ATOM H7     HGA2    0.090 !    0.000


BOND C1   C2
BOND C2   C3
BOND C3   C4
BOND C4   C9
BOND C4   C5
...
```

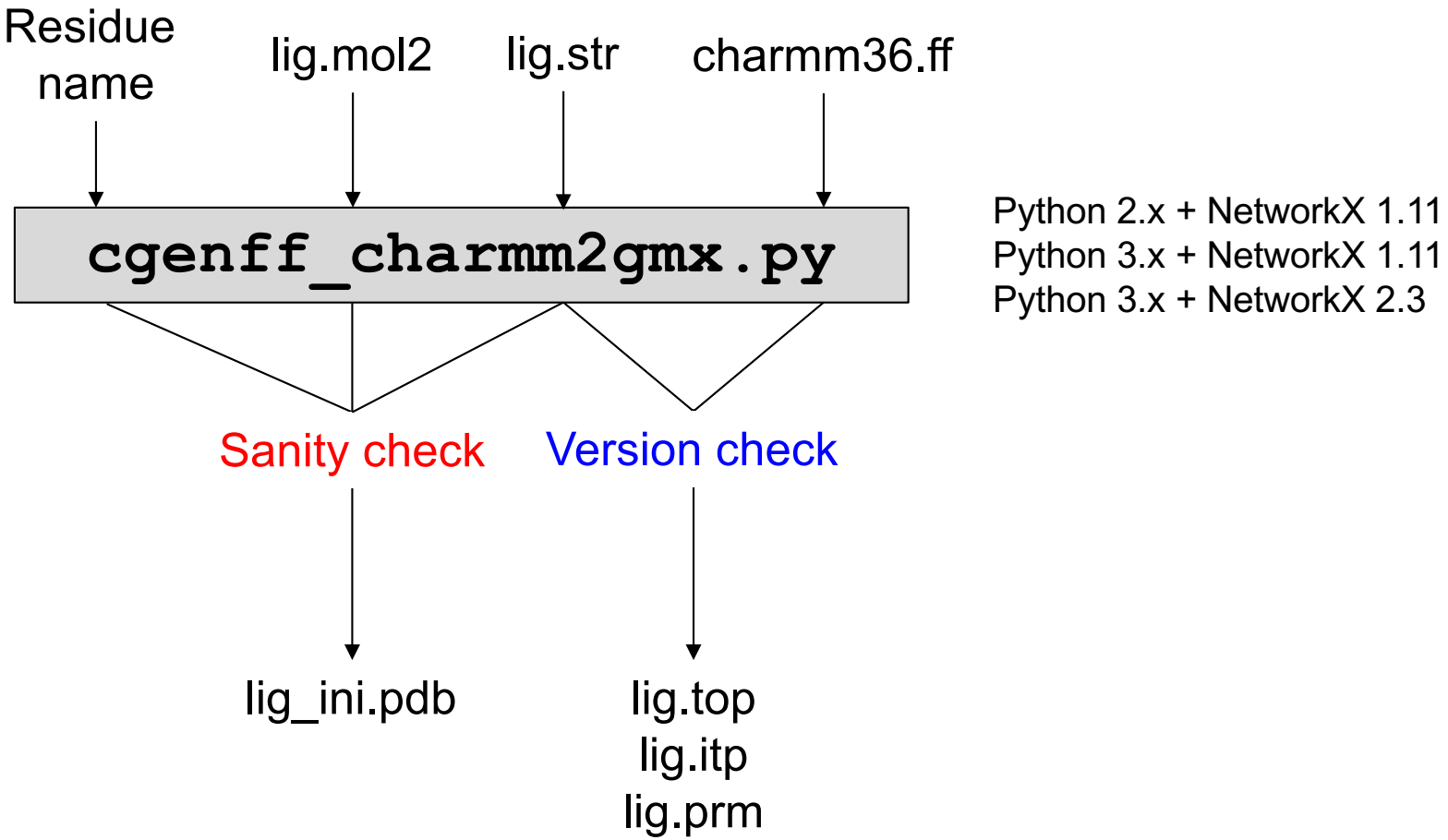


The CGenFF Topology

```
read param card flex append
* Parameters generated by analogy by
* CHARMM General Force Field (CGenFF) program version 1.0.0
*


! Penalties lower than 10 indicate the analogy is fair; penalties between 10
! and 50 mean some basic validation is recommended; penalties higher than
! 50 indicate poor analogy and mandate extensive validation/optimization.


BONDS


ANGLES


DIHEDRALS
CG321   CG2R61 CG2R61 OG311      2.4000  2   180.00 ! ***** , from CG311 CG2R61 CG2R61 OG311, penalty= 0.6
CG2R61 CG321   CG321   CG331      0.0400  3     0.00 ! ***** , from CG2R61 CG321 CG321 CG321, penalty= 0.9


IMPROPERS


END
RETURN
```



CHARMM → GROMACS Conversion



CHARMM → GROMACS Conversion

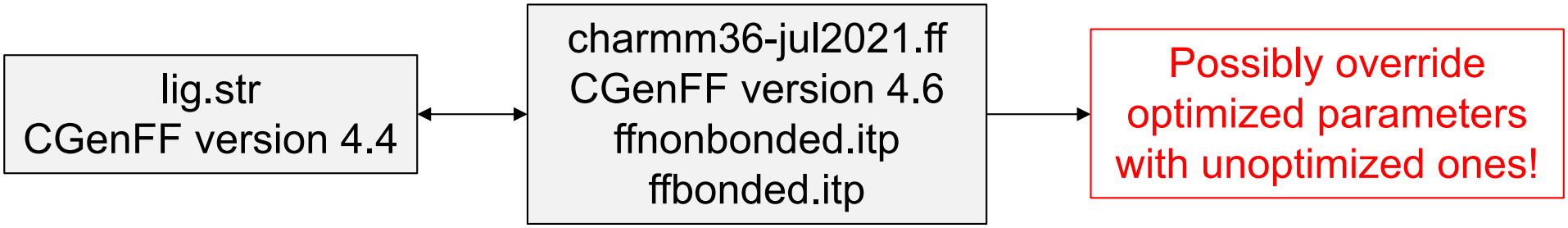
lig_ini.pdb	Coordinates from lig.mol2, atom names from lig.str; can be used in a complex or for a solution simulation.
lig.top	Standalone system topology for the small molecule; useful for preparing simulation in solution.
lig.itp	Contains the `[moleculetype]` definition of the ligand.
lig.prm	GROMACS topology file (.itp) that contains any additional parameters needed by the small molecule.

CHARMM → GROMACS Conversion

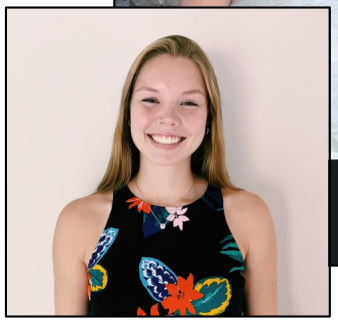
```
#include "charmm36-jul2021.ff/forcefield.itp"

#include "lig.prm"

[ moleculetype ]
; name            nrexcl
Protein           3
```

Location of this #include statement is key!

- Why do we perform a CGenFF version check during the conversion?
  - A force field is a self-consistent entity!



Acknowledgments



UNIVERSITY of MARYLAND
SCHOOL OF PHARMACY

NIH National Institutes of Health
Turning Discovery Into Health

F32GM109632
R35GM133754

Health Resources in Action
Advancing Public Health and Medical Research

Thomas F. and Kate Miller Jeffress Memorial Trust

USDA

C36 Port: András F. Wacha

VIRGINIA TECH™