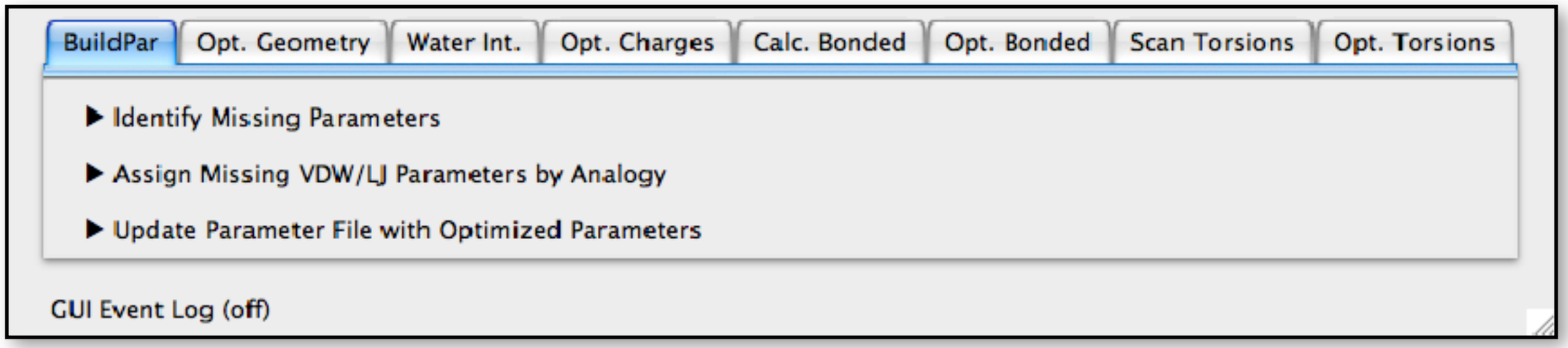
Parametrizing Small Molecules Using:

The Force Field Toolkit (ffTK)



Christopher G. Mayne, Emad Tajkhorshid
Beckman Institute for Advanced Science and Technology
University of Illinois, Urbana-Champaign

Klaus Schulten
University of Illinois, Urbana-Champaign

James C. (JC) Gumbart, Anna Pavlova
Georgia Institute of Technology

Computational Biophysics Workshop | DICP | July 11 2018

MD Simulations of Biological Systems

Molecular Mechanics Force Fields

$$U = \underbrace{U_{bonds} + U_{angles} + U_{dihedrals}}_{\text{bonded}} + \underbrace{U_{vdW} + U_{coulombic}}_{\text{non-bonded}}$$

The CHARMM Force Field

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

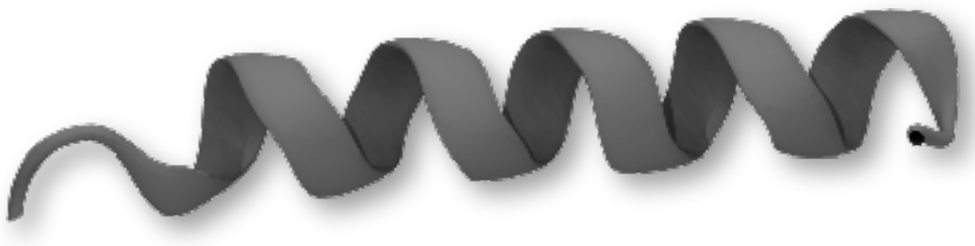
$$\sum_{dihedrals} k_i^{dihedral}[1 + \cos(n_i\phi_i + \delta_i)] +$$

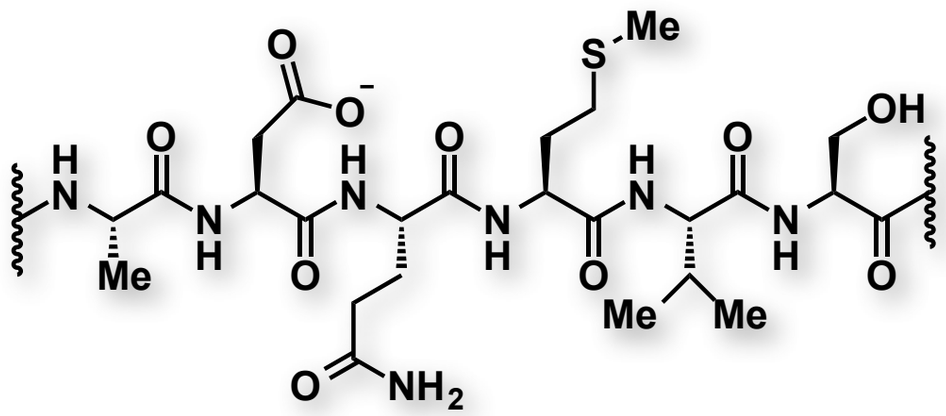
$$\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{r_{ij}}$$

Parameter Transferability In Biopolymers

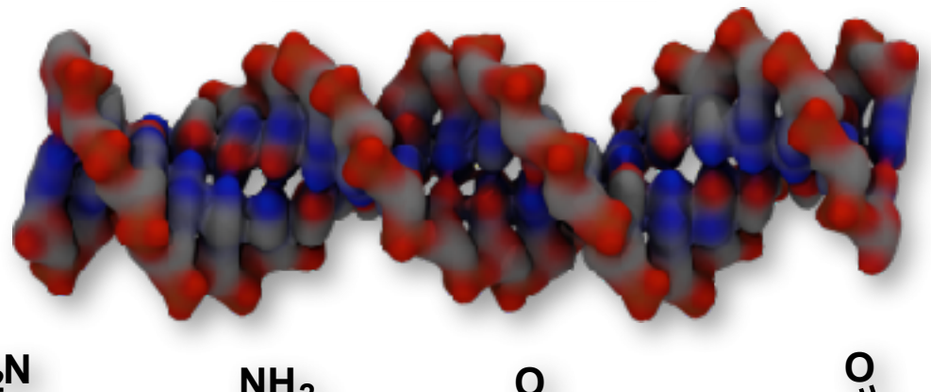
Parameter set describes molecular behavior in varied chemical (connectivity) and spatial (conformation) contexts

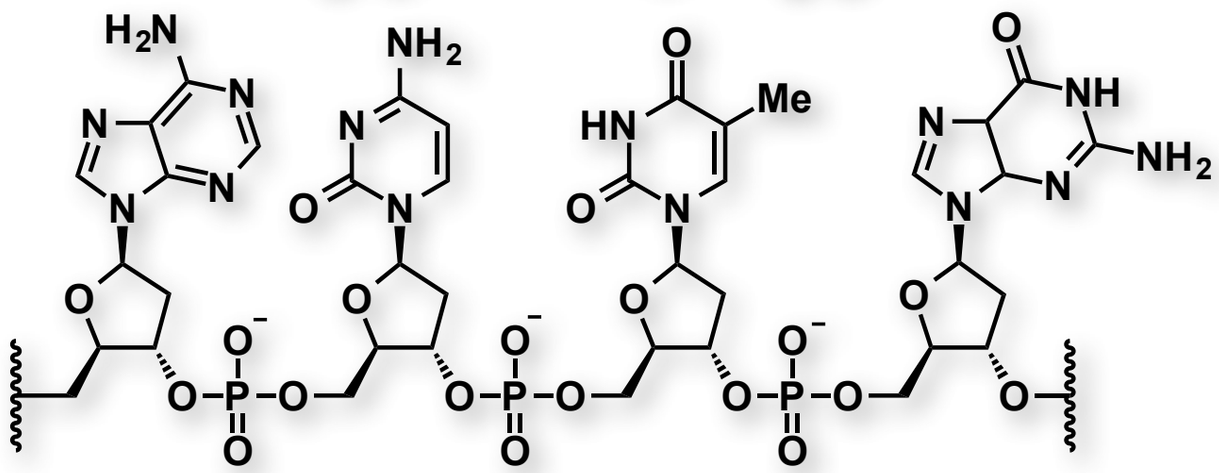
Peptides and Proteins



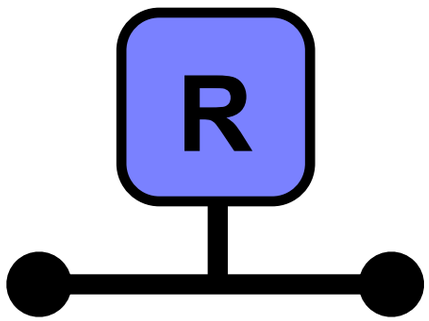


Nucleic Acids





Key Features:

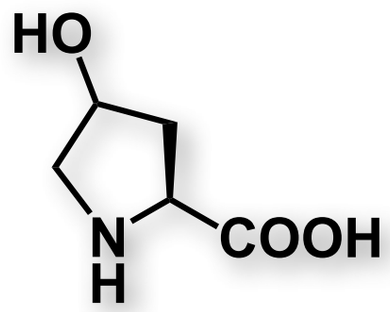


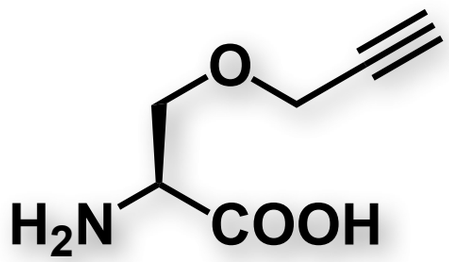
limited set of isolated building blocks

repetitive backbone unit

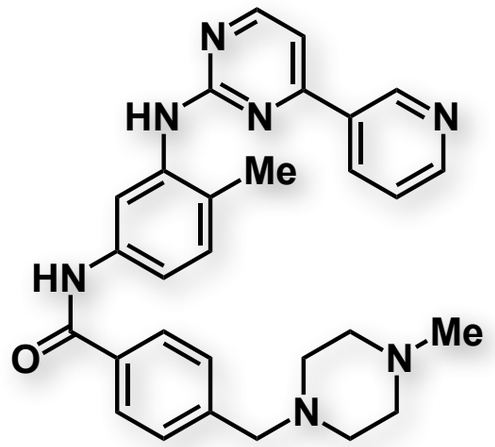
Parametrization as an Impasse

non-standard or engineered amino acids

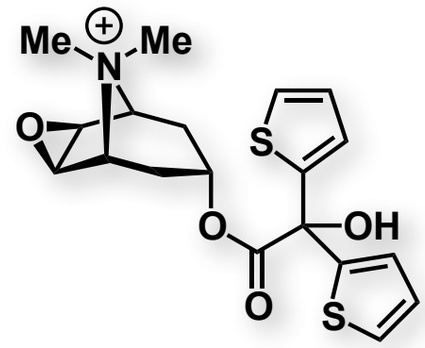




small molecule ligands

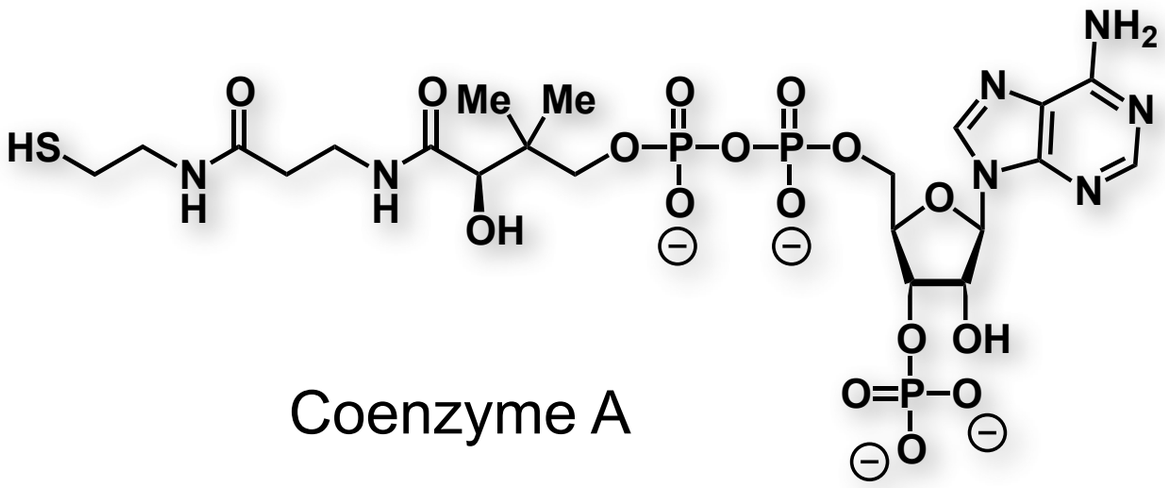


Imatinib (Gleevec)



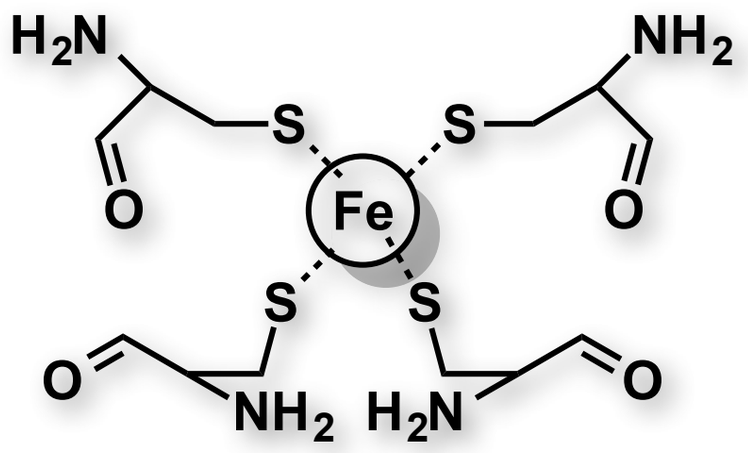
Tiotropium (Spiriva)

cofactors

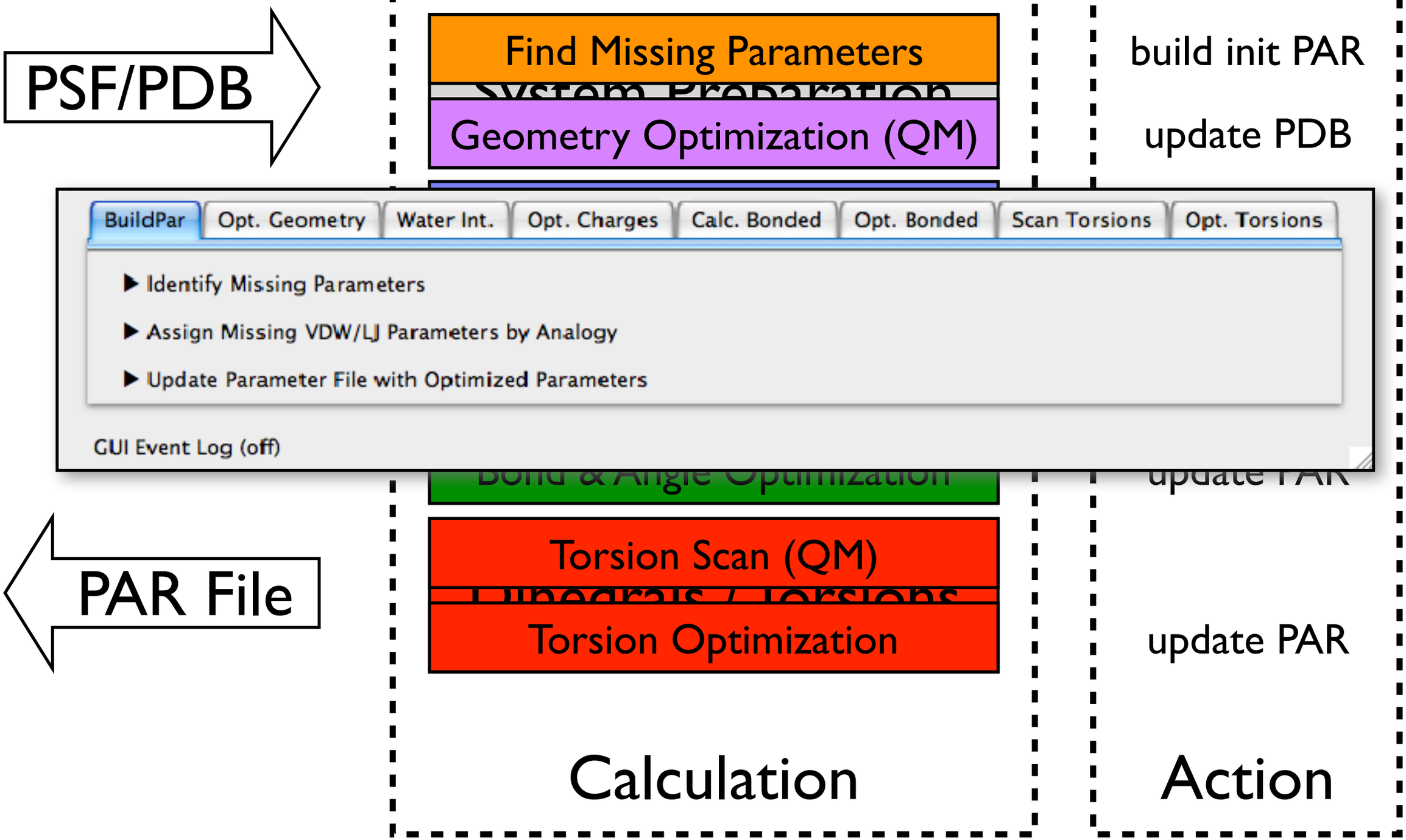


Coenzyme A

metal centers

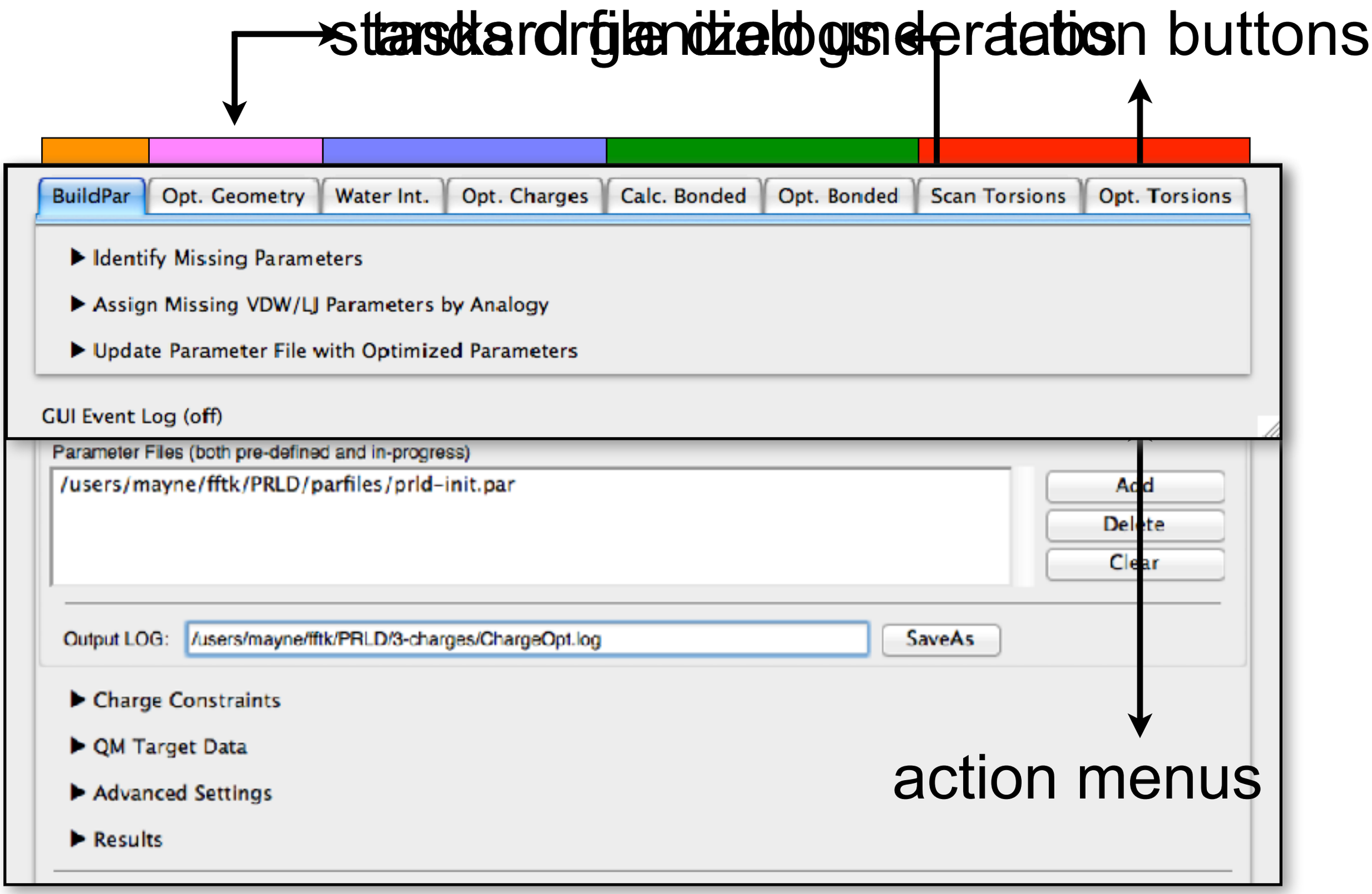


CGenFF Parametrization Workflow



K. Vanommeslaeghe *et al.*, *J. Comput. Chem.* **2010**, *31*, 671–690.
C. G. Mayne *et al.*, *J. Comput. Chem.* **2013**, *34*, 2757–2770.

ffTK Interface



Functionality Provided by ffTK

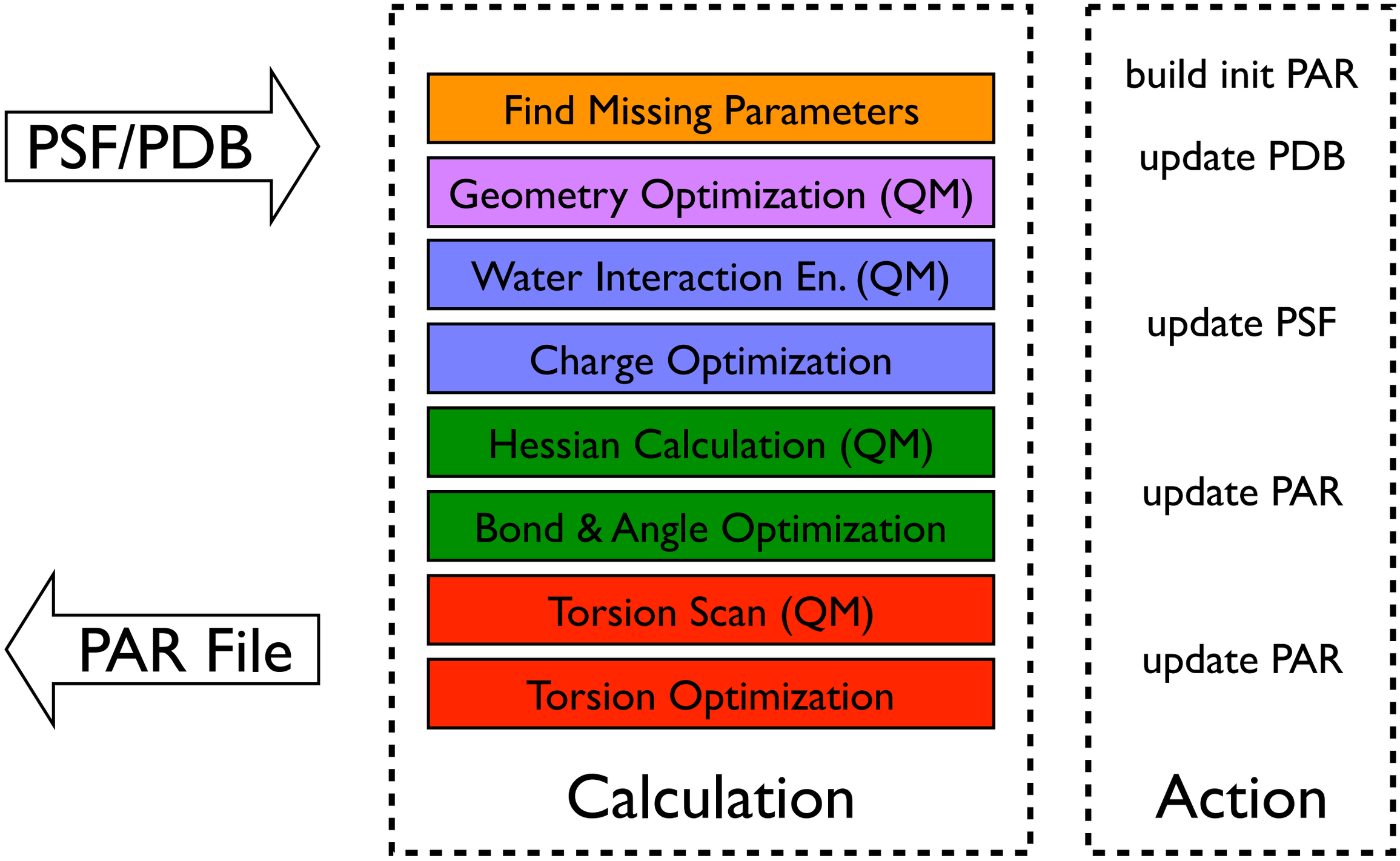
Core Functions

- Setup & Perform Multi-dimensional Optimizations
- Abstraction of Gaussian I/O (QM)
- Assess Performance of Parameters by Visualizing Optimization Data

Support Functions

- Auto-detect Water Interaction Sites
- Auto-detect Charge Groups
- Auto-detect Non-redundant Torsions
- Build & Update Parameter Files
- Browse Existing Parameter Sets
- Write Updated Charges to PSF
- Reset Opt. Input from Output
- Visualize Target Data in VMD
- Create Graphic Objects in VMD
- Label Atoms in VMD
- Read Input Parameters from File
- Read/Write Data From Opt. Logs
- Export Plot Data to File
- Monitor Optimization Progress

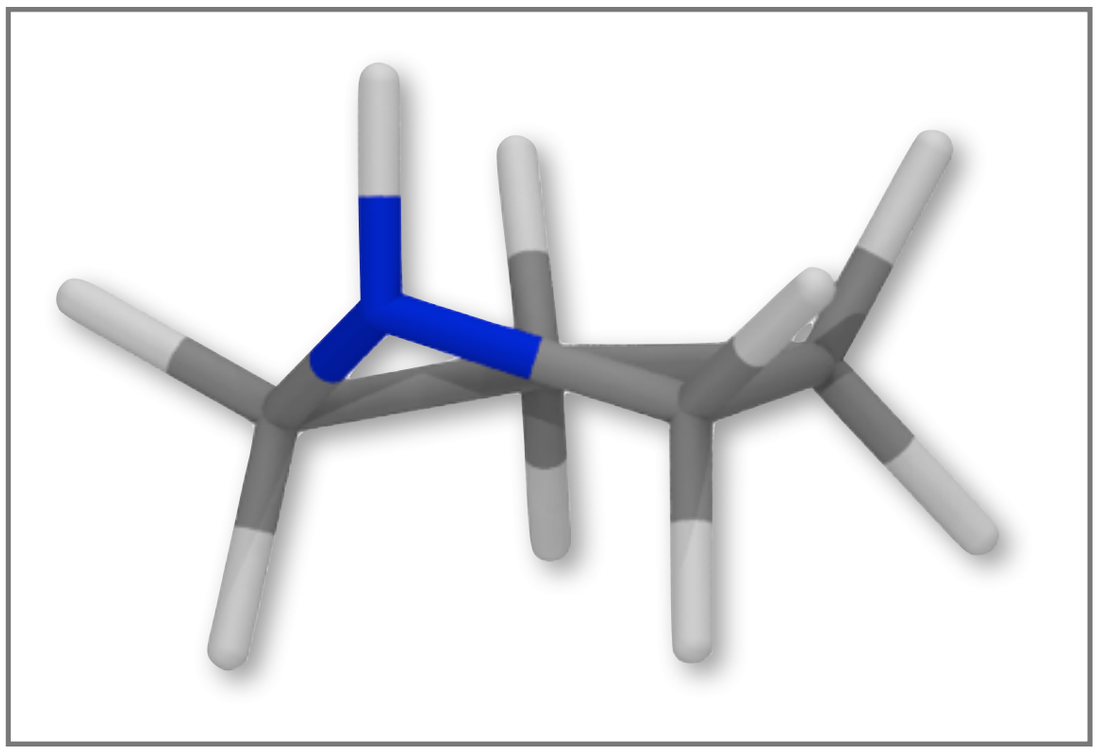
ffTK Exemplified by Charge Optimization



Generating Charge Optimization Target Data

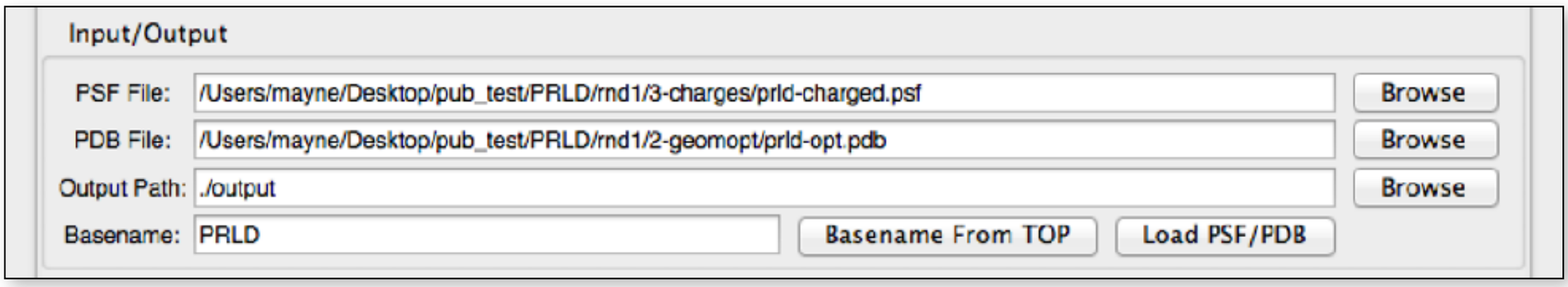
Load QM optimized geometry | Auto-detect interaction sites | Genera

pyrrolidine



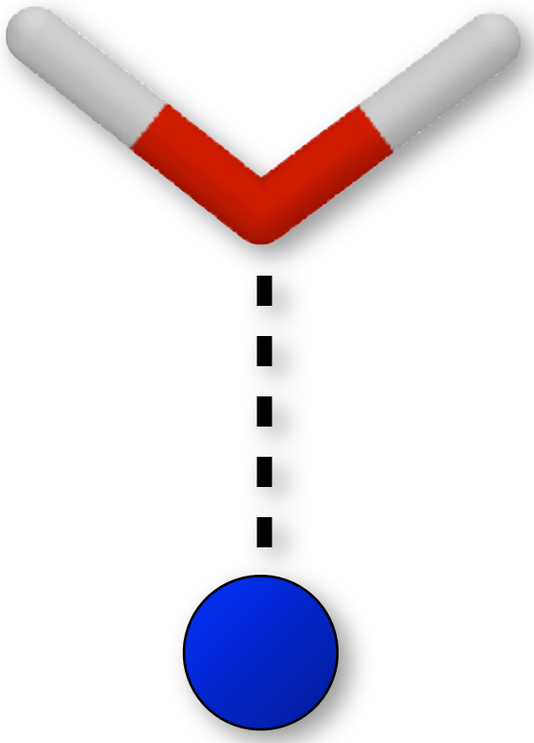
VMD main window

ffTK GUI

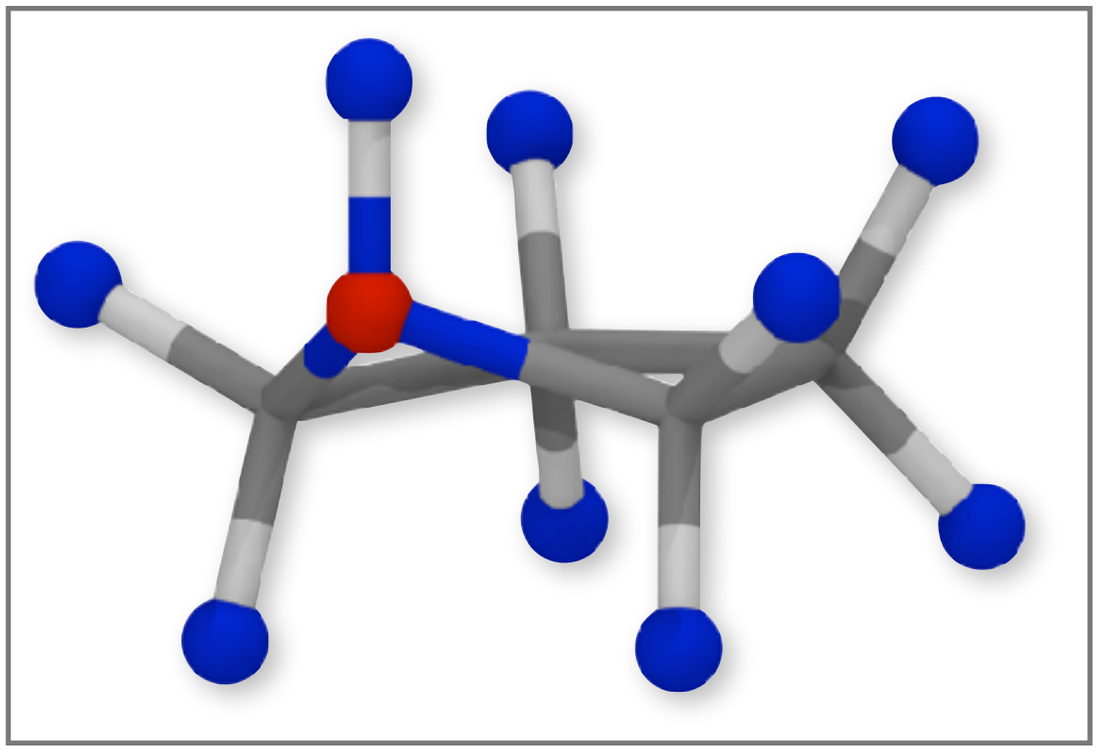


Generating Charge Optimization Target Data

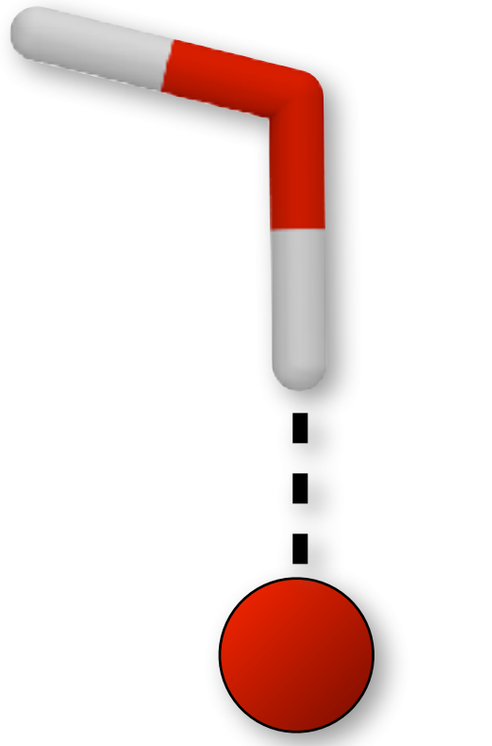
geometry | Auto-detect interaction sites | Generate Gaussian Input Files | Run



Donor

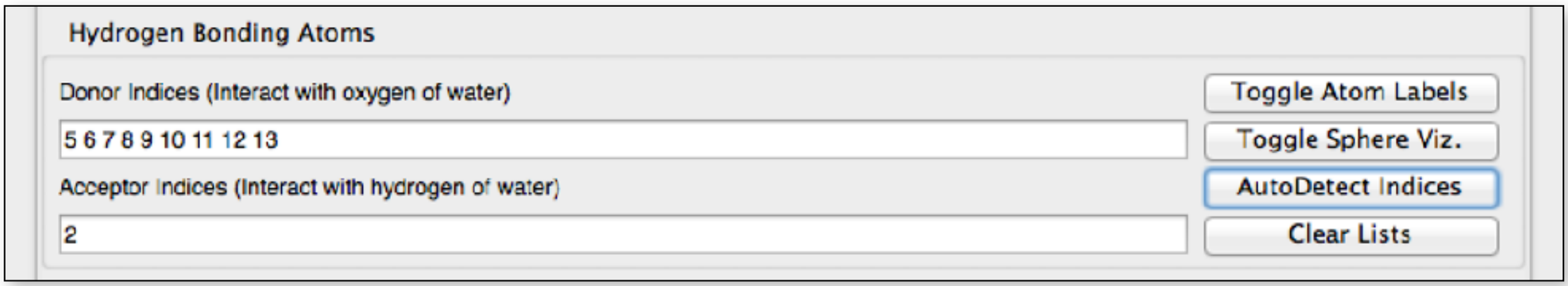


VMD main window



Acceptor

ffTK GUI

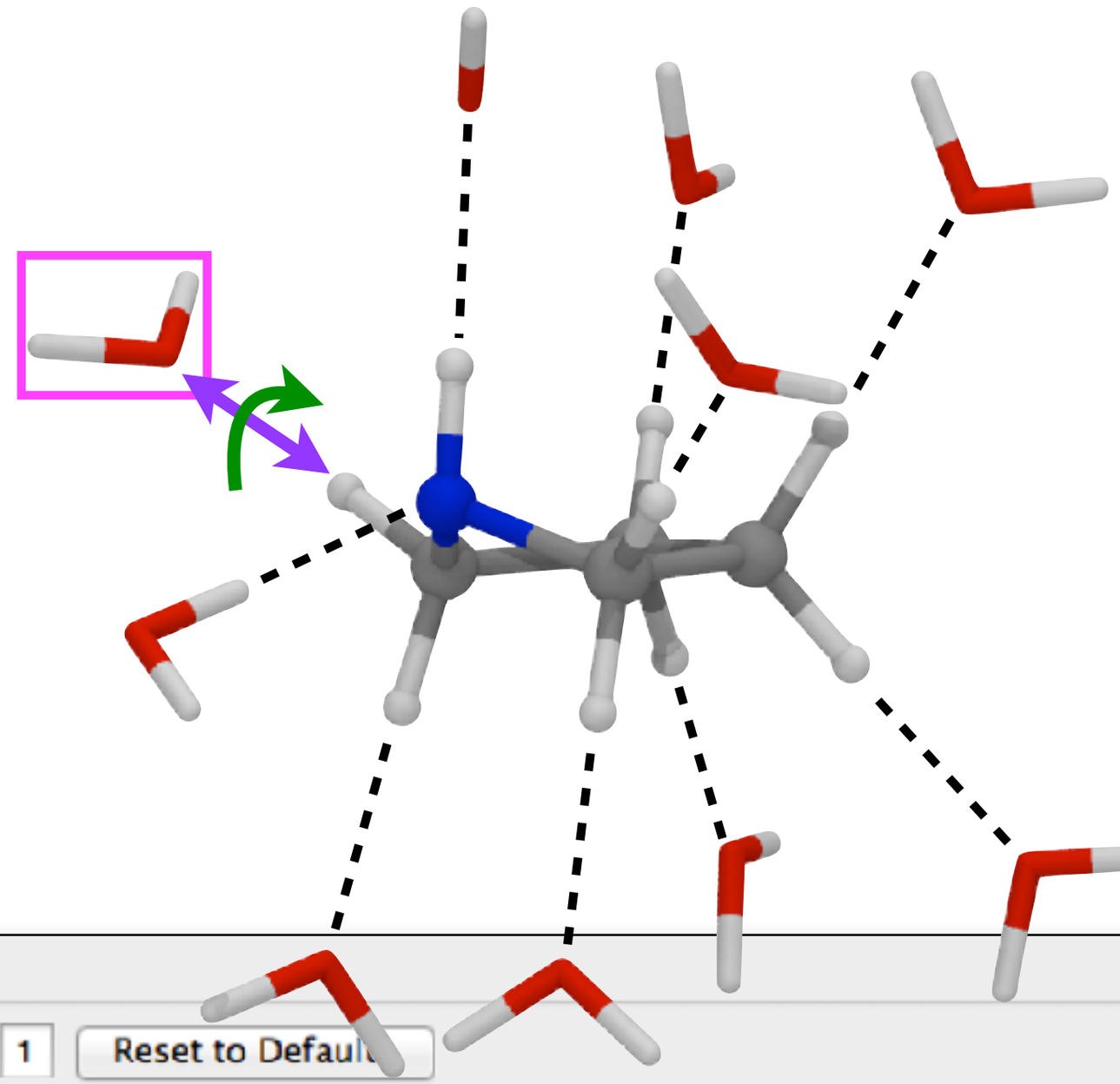


Generating Charge Optimization Target Data

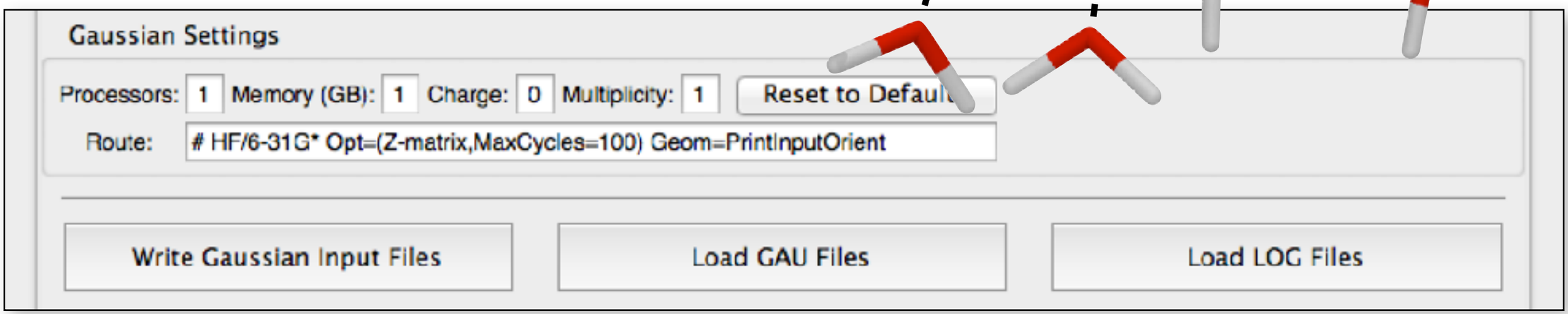
on sites | Generate Gaussian Input Files | Run QM | Inspect water optimization

Compute water position

Optimize
distance & rotation



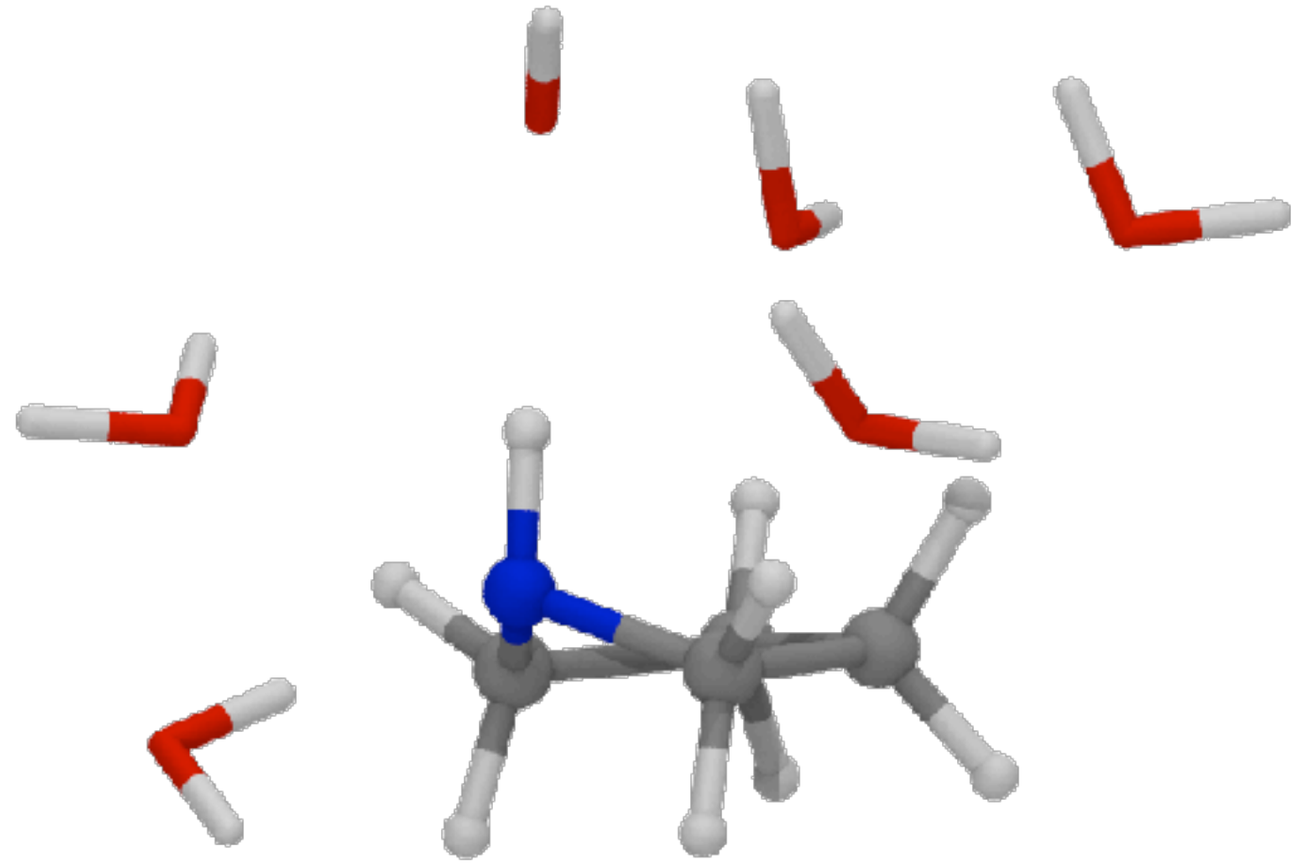
ffTK GUI



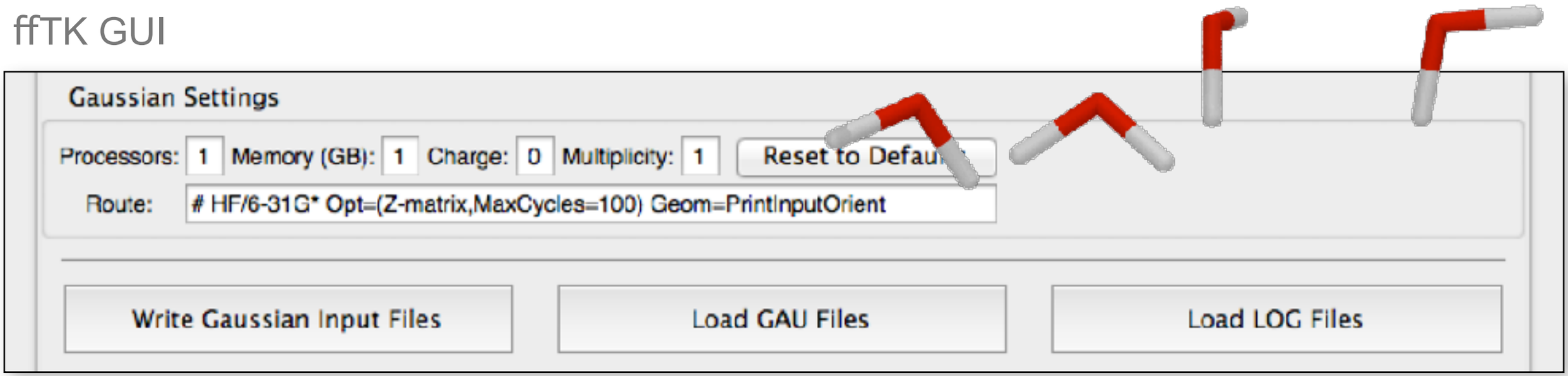
Generating Charge Optimization Target Data

| Run QM | Inspect water optimization

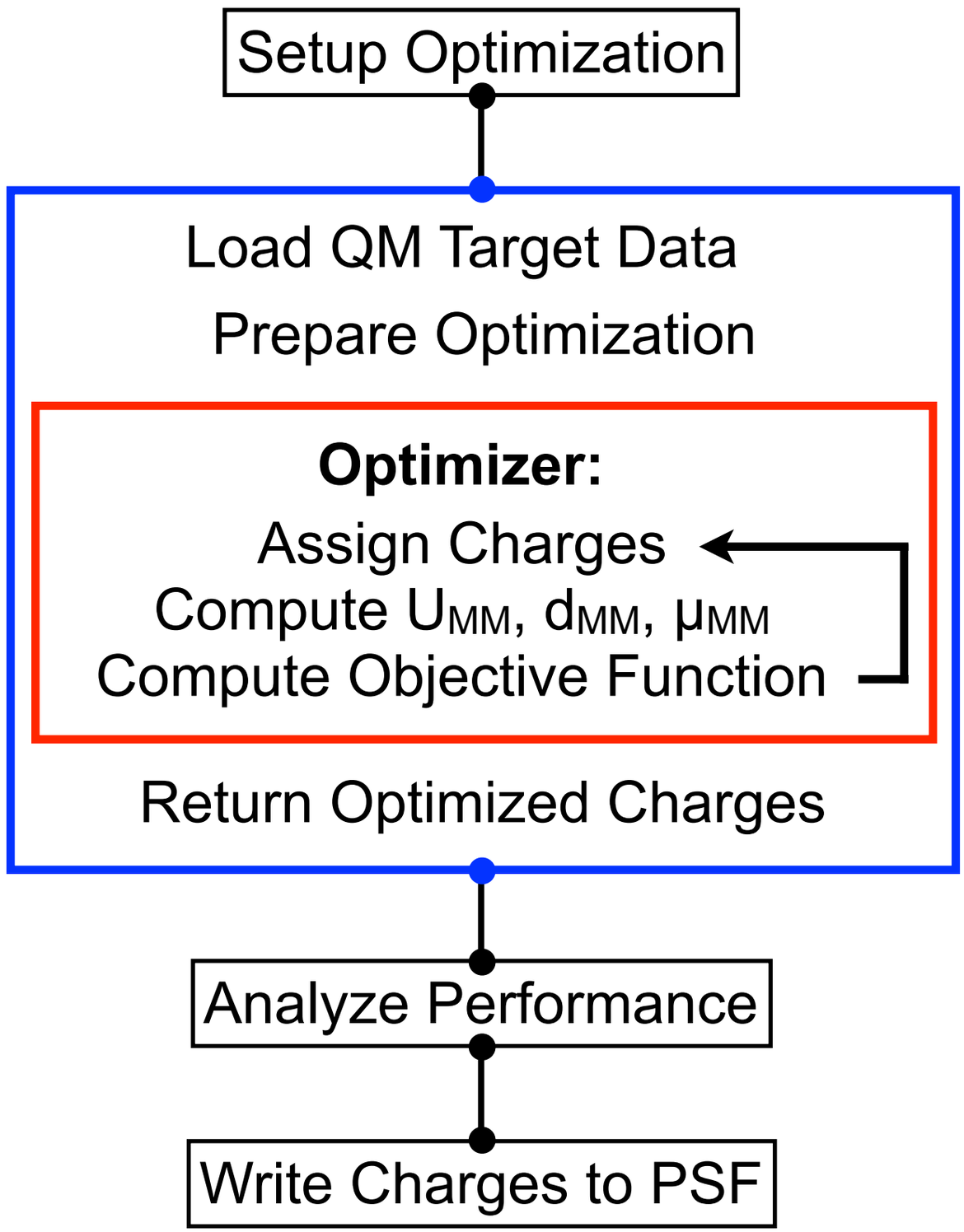
Visually assess QM-optimized water position(s)



ffTK GUI



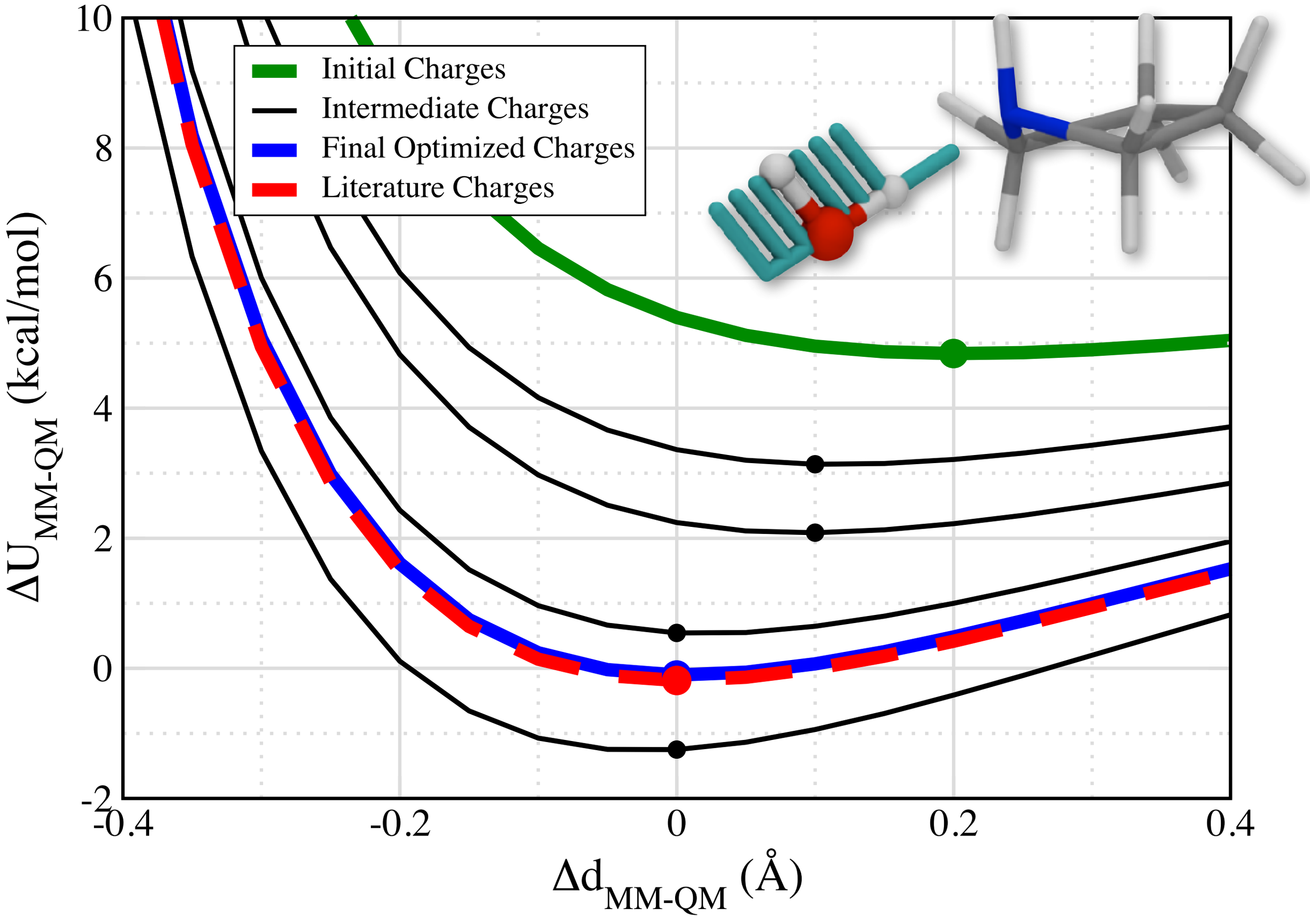
Charge Optimization



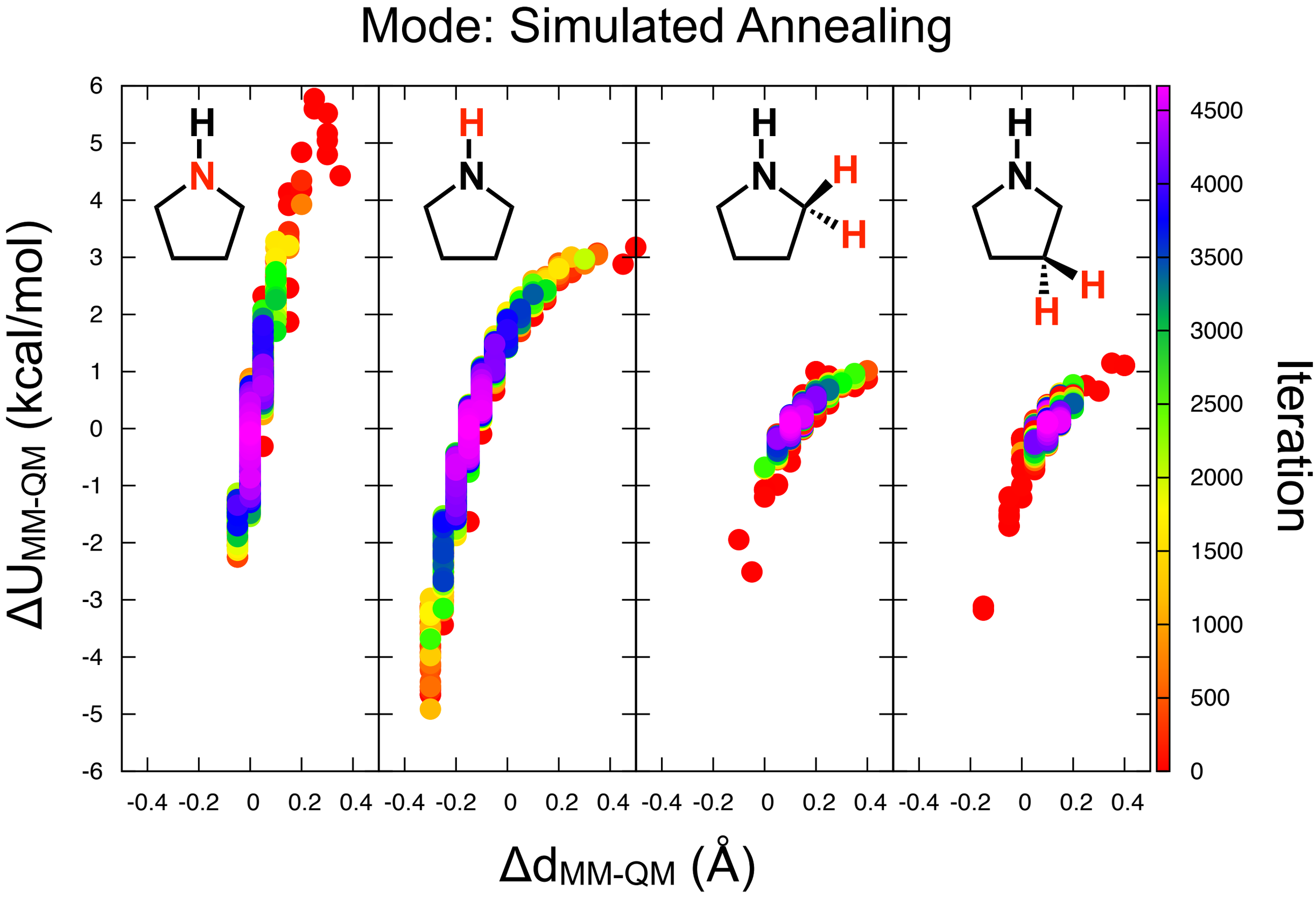
Objective Function

$$\sum_{\text{wat. int.}} f(U_{MM}\text{-}U_{QM}) + \sum_{\text{wat. int.}} f(d_{MM}\text{-}d_{QM}) + f(\mu_{MM}\text{-}\mu_{QM})$$

Assessing MM Water-Interaction Profiles



Sampling MM Water-Interaction Profiles



Tuning the Optimization

Objective Function

$$\sum_{\text{wat. int.}} w_i \left(\frac{U_{MMmin} - U_{QMmin}}{U_{tol}} \right)^2 +$$

$$w_d \sum_{\text{wat. int.}} w_i \left(\frac{d_{MMmin} - d_{QMmin}}{d_{tol}} \right)^2 +$$

$$n\, w_\mu \left[\left(\frac{\mu_{MM} - \mu_{QM}}{\mu_{tol}} \right)^2 + \left(\frac{\theta}{\theta_{tol}} \right)^2 \right]$$

Atom Name	Weight
N1	1.0
H1	1.0
H21	1.0
H22	1.0
H31	1.0
H32	1.0
H41	1.0
H42	1.0
H51	1.0
H52	1.0

w_i

w_d w_μ

Optimize Settings

Tolerance: 0.0001 Distance Weight: 1.0 Dipole Weight: 1.0

Mode: downhill

In practice, it is impossible to fit all of these perfectly! Often we decrease w_d and w_μ to improve the fit to the **energies**

Plotting Charge Optimization Data

LOG File: /Users/mayne/Desktop/pub_test/PRLD/rnd1/3-charges/ChargeOpt-SA.log

File

export data

Energy

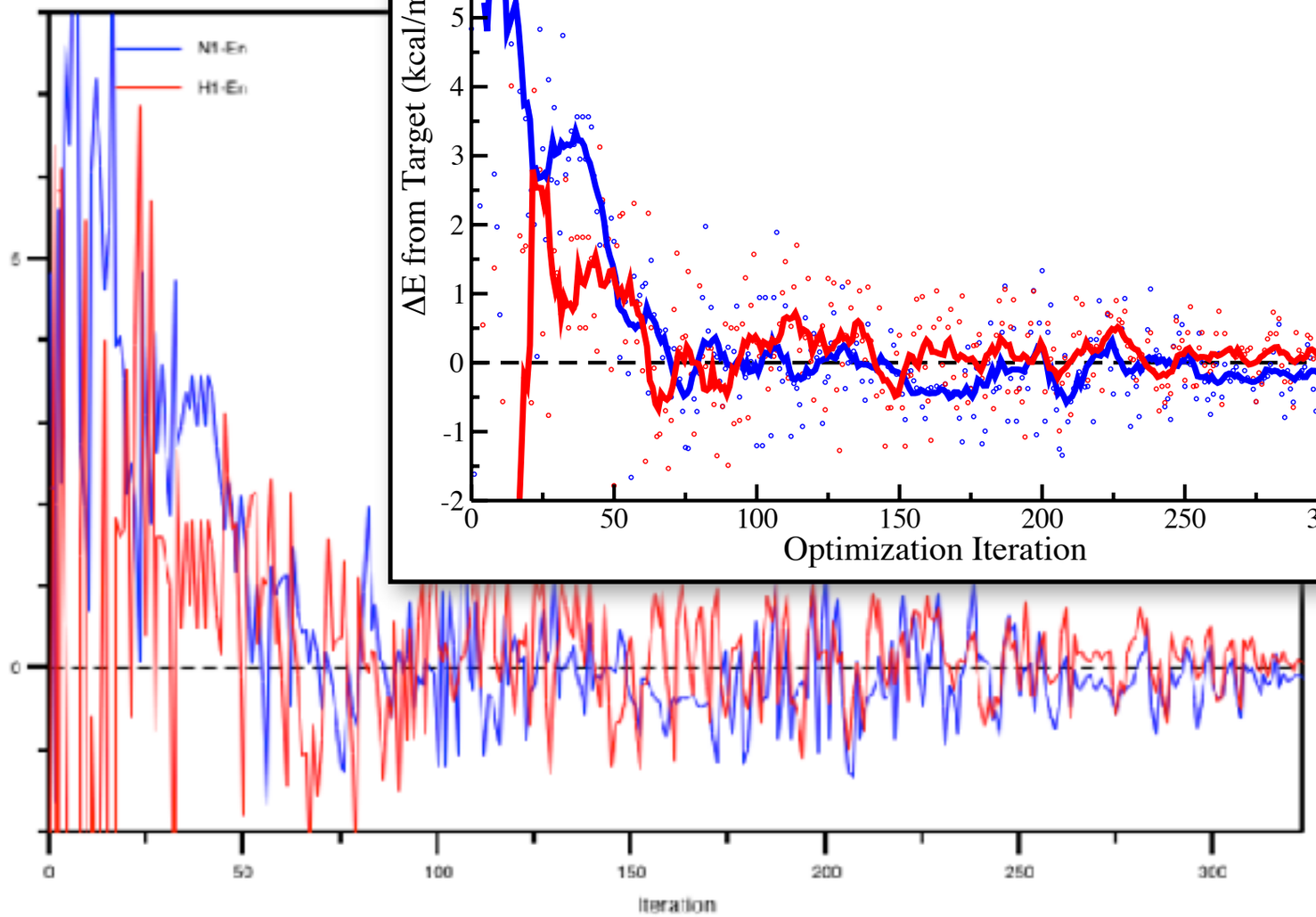
atom	color
N1	blue
H1	red
H21	black
H22	black
H31	black
H32	black
H41	black
H42	black
H51	black
H52	black

Distance

atom	color
N1	black
H1	black
H21	black
H22	black
H31	black
H32	black
H41	black
H42	black
H51	black
H52	black

Objective

type	color
total	black
energy	red
distance	blue
dipole	green



x-min y-min

x-max y-max

x-min auto x-max auto y-min -2 y-max 8 Set Axis

Axis Autoscaling

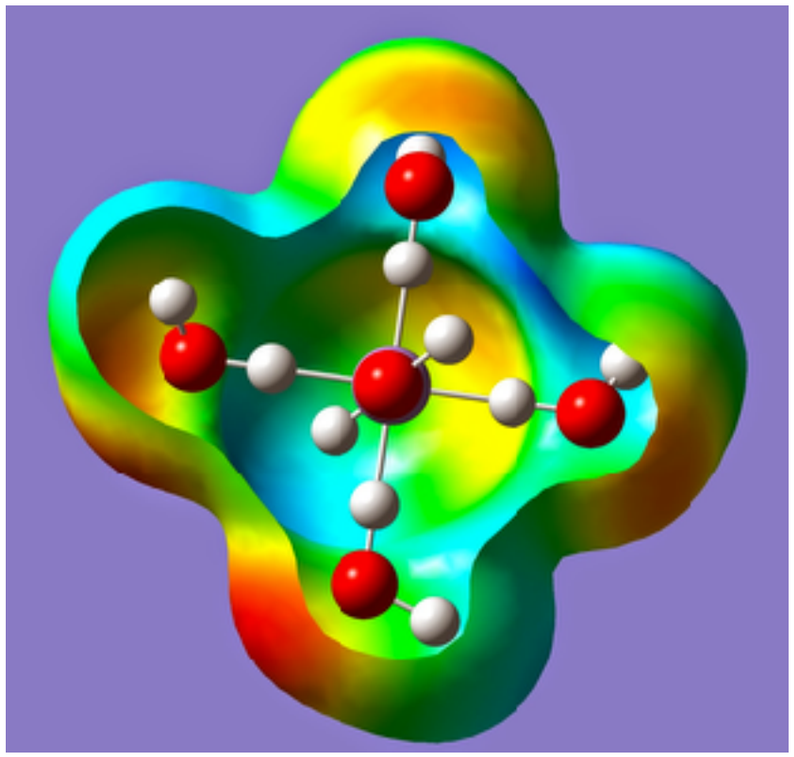
Set Plot Color: red

Restrained Electrostatic Potential (RESP) fitting

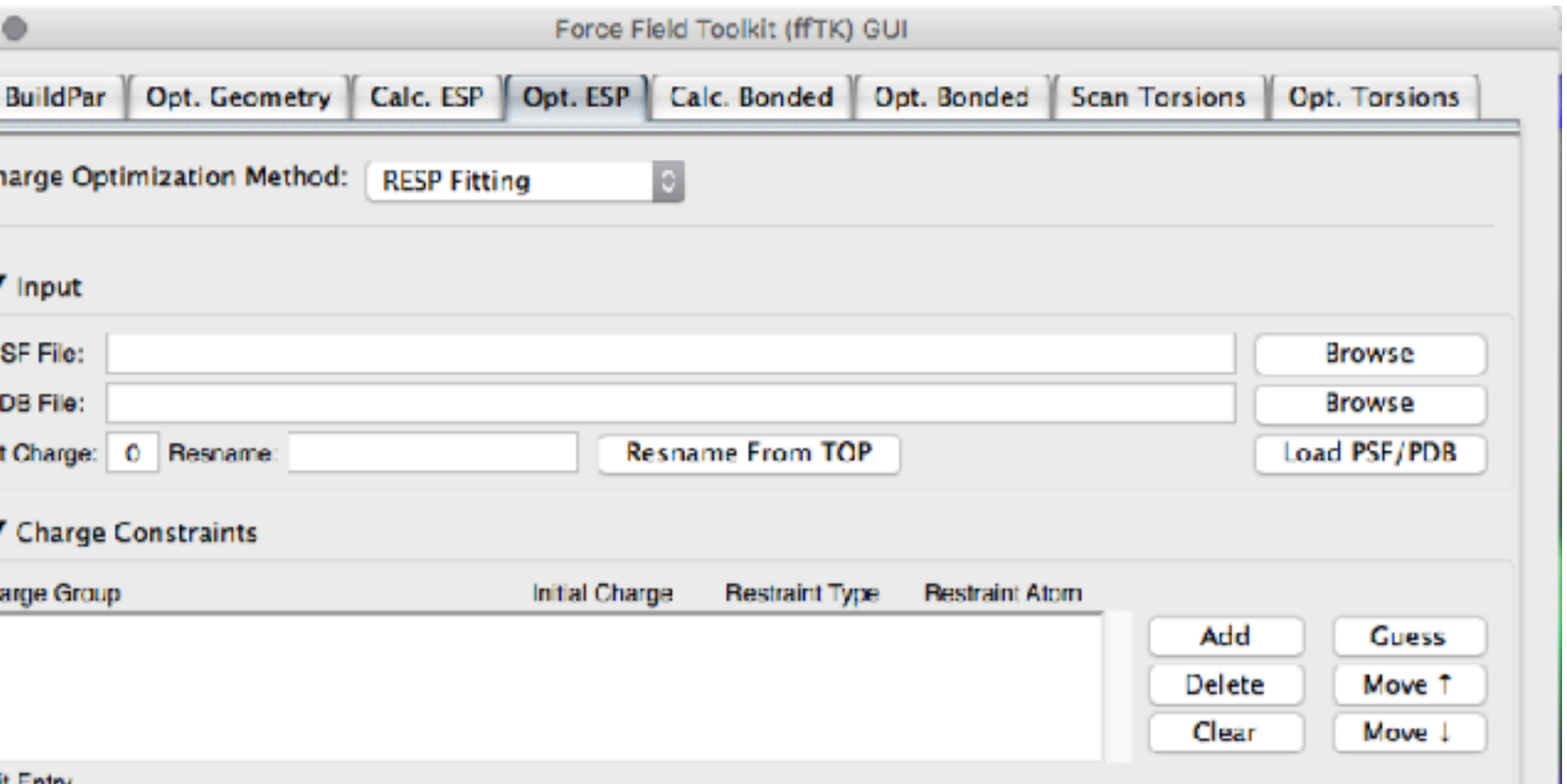
An alternative to water interactions for charges, commonly used in Amber

The QM electrostatic potential is calculated and then fit by optimizing the MM charges

Has problems with buried atoms, which may not noticeably affect the ESP



https://studynights.blogspot.com/2015/03/the-single-point-energy-of-mnh2o6-and.html

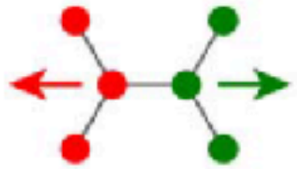


RESP fitting is supported by FFTK, requires downloading the **resp** program as part of AmberTools (free)

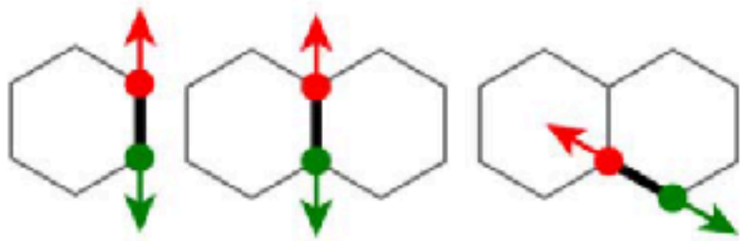
Fitting of Bonds and Angles

Bonds

1) nonredundant bond



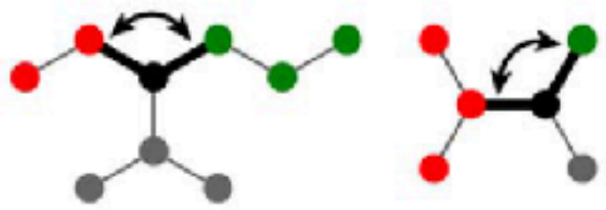
2) bond in a ring



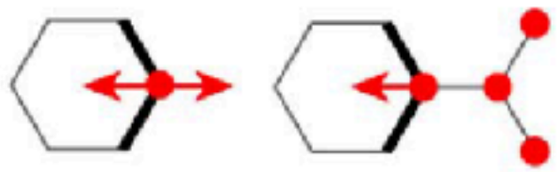
Angles

1) nonredundant angle

2) redundant angle; not part of a ring

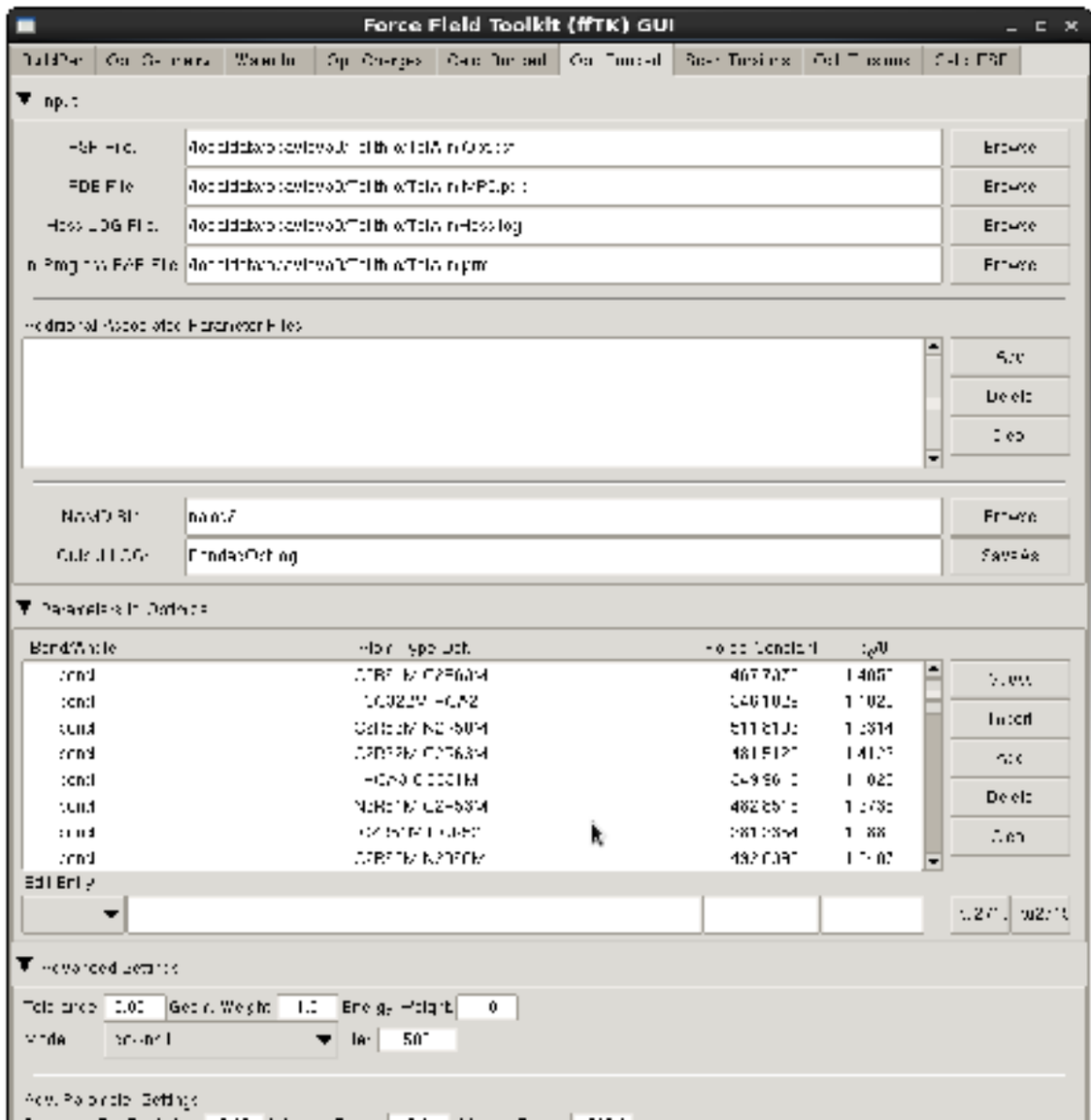


3) redundant angle; part of a ring

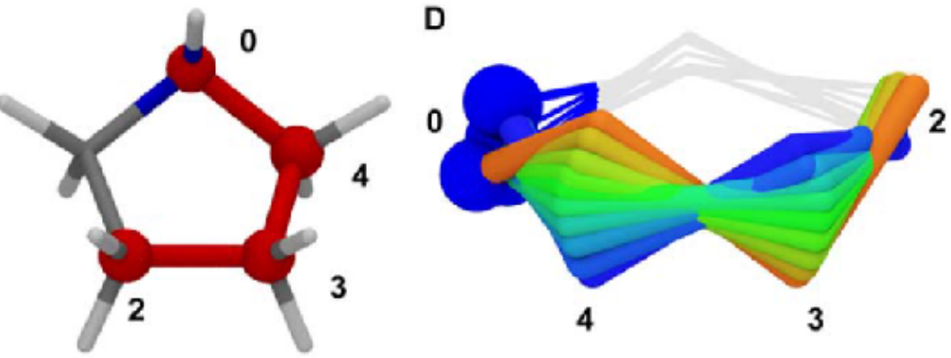


$$\sum_{\text{bonds}} k_i^{\text{bond}} (r_i - r_0)^2 + \sum_{\text{angles}} k_i^{\text{angle}} (\theta_i - \theta_0)^2$$

Bond and angle are fit by creating a small distortion of the bond/angle and calculating the QM energy and the MM energy, then choosing the force constants to match



Fitting of Dihedrals



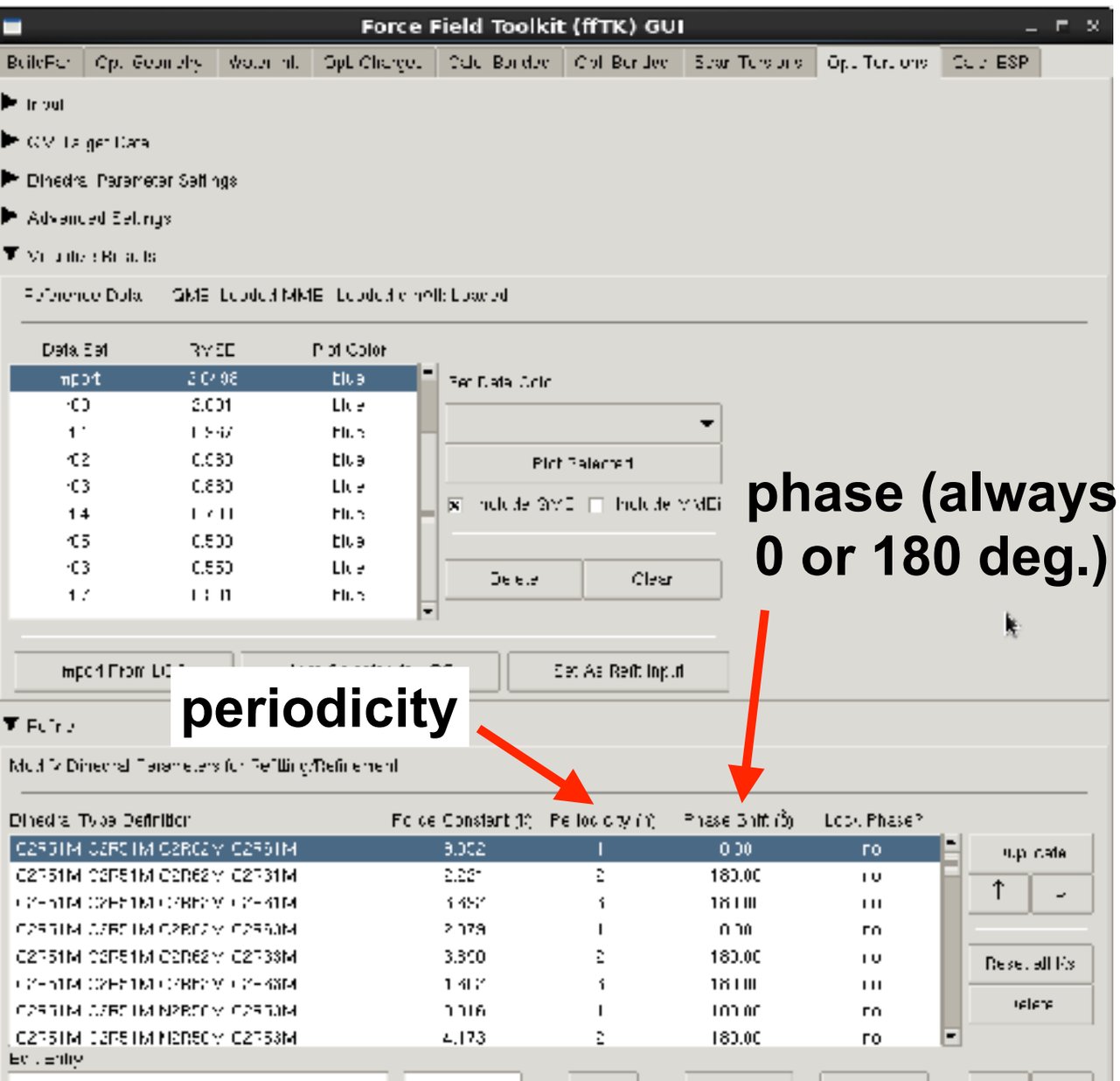
Dihedrals are scanned in QM in 10-15 deg. increments

Energies of each conformation are fit in MM according to the pre-determined dihedral terms included

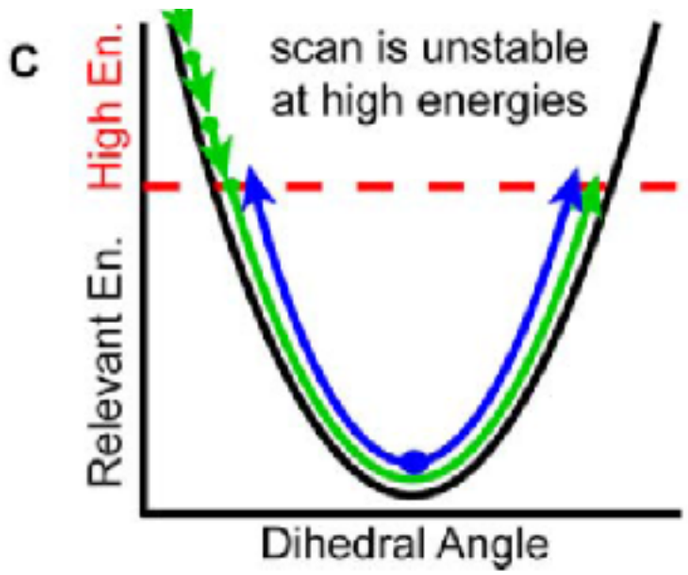
$$\sum_{\text{dihedrals}} k_i^{\text{dihedral}} [1 + \cos(n_i \phi_i + \delta_i)]$$

periodicity (1-6 possible)

phase (always 0 or 180 deg.)



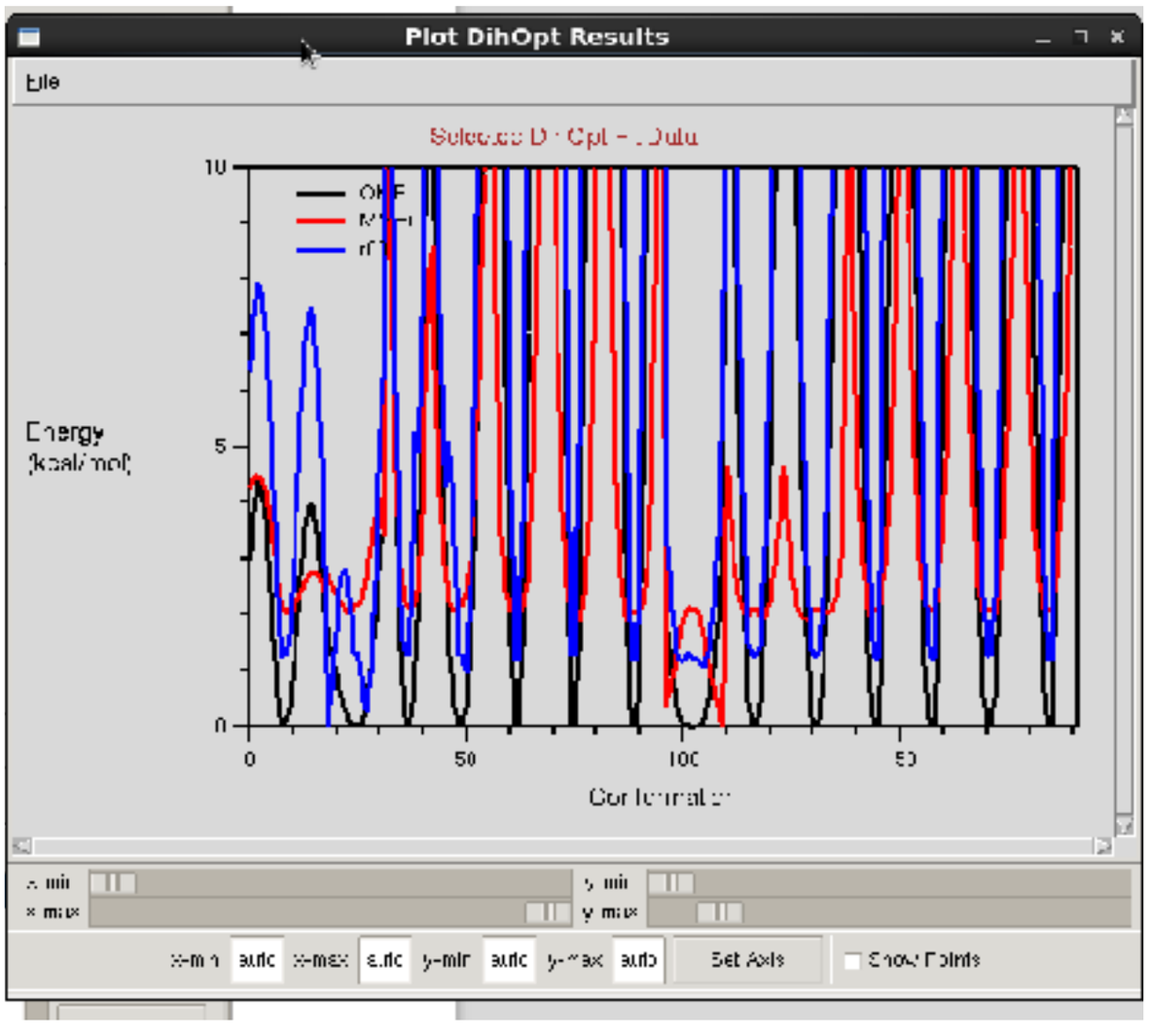
Energies above a threshold (e.g., 8-10 kcal/mol) are ignored

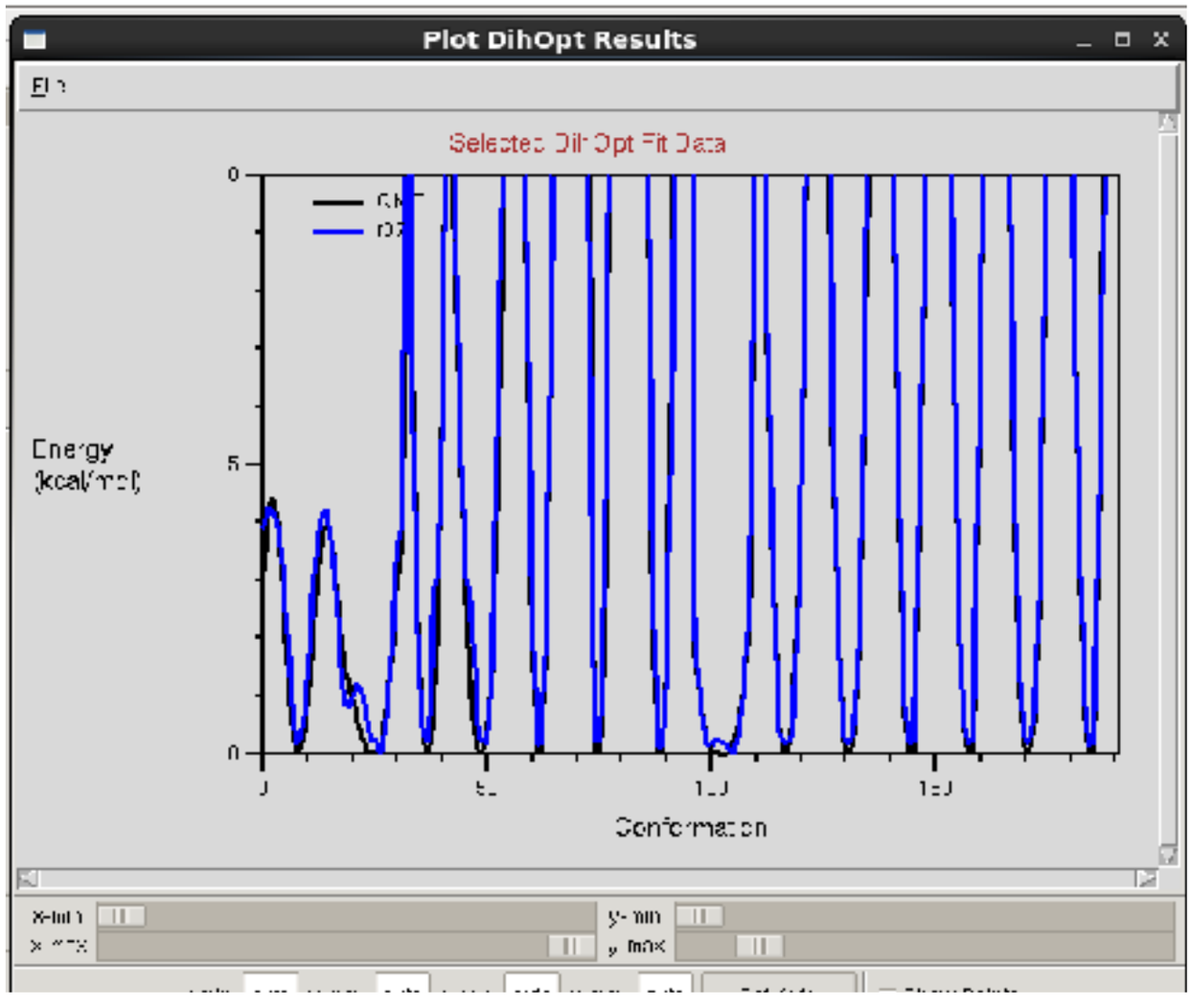


Fitting of Dihedrals

QM PES: black
initial: red
First fit: blue

After 7 rounds of simulated annealing, the fit is much better





Two Approaches to Fitting the Dihedrals

Several multiplicities and free phase

Pro: very good fit of QM PES
Cons: possible incorrect behavior multiple sets of force constants

One multiplicity and locked phase

Pros: limits incorrect behavior, sets of force constants
Cons: fit to QM PES not always possible

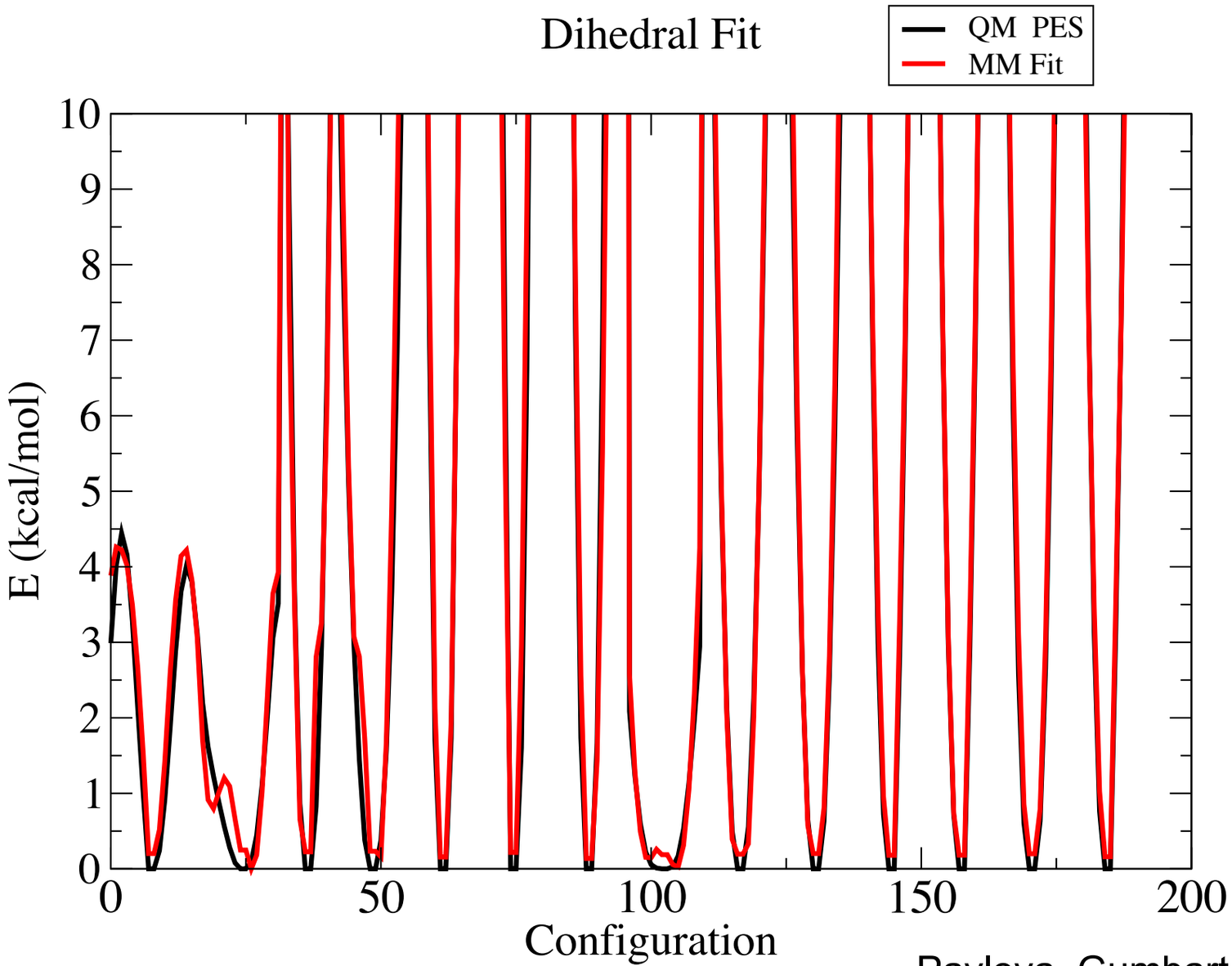
In practice a tradeoff is needed!

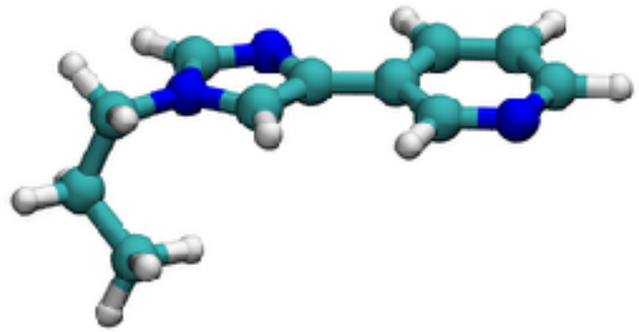
Example of Overfitting

imidazole-pyrridine moiety of antibiotic telithromycin

3 multiplicities, free phases for each dihedrals

$$k_1[1+\cos(\phi+\delta_1)]+k_2[1+\cos(2\phi+\delta_2)]+k_3[1+\cos(3\phi+\delta_3)]$$



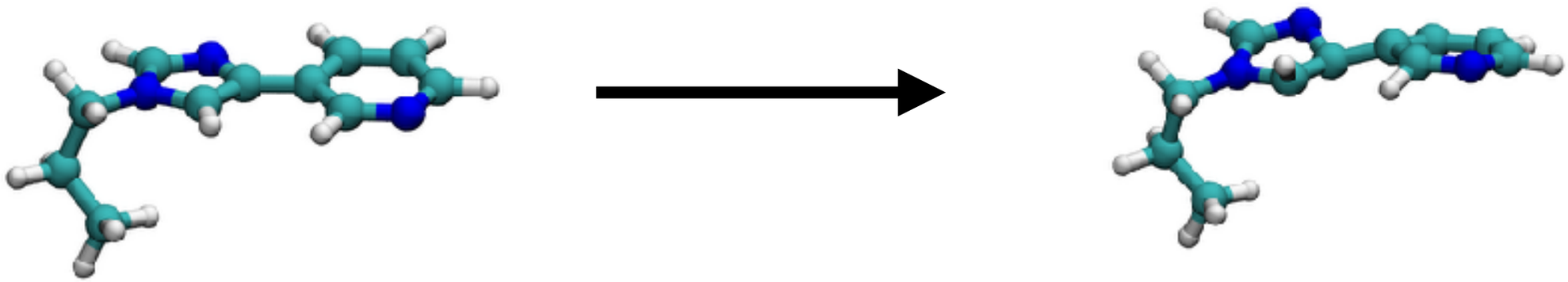


multiple dihedrals scanned; plotted simultaneously to reference a **common global minimum**

Pavlova, Gumbart. Parametrization of macrolide antibiotics using the force field toolkit. *J. Comp. Chem*. 2015, **36**, 2052−2063.

Example of Overfitting

imidazole-pyrridine moiety of antibiotic telithromycin



Using too many dihedral multiplicities
can leads to distortion of a planar molecule!

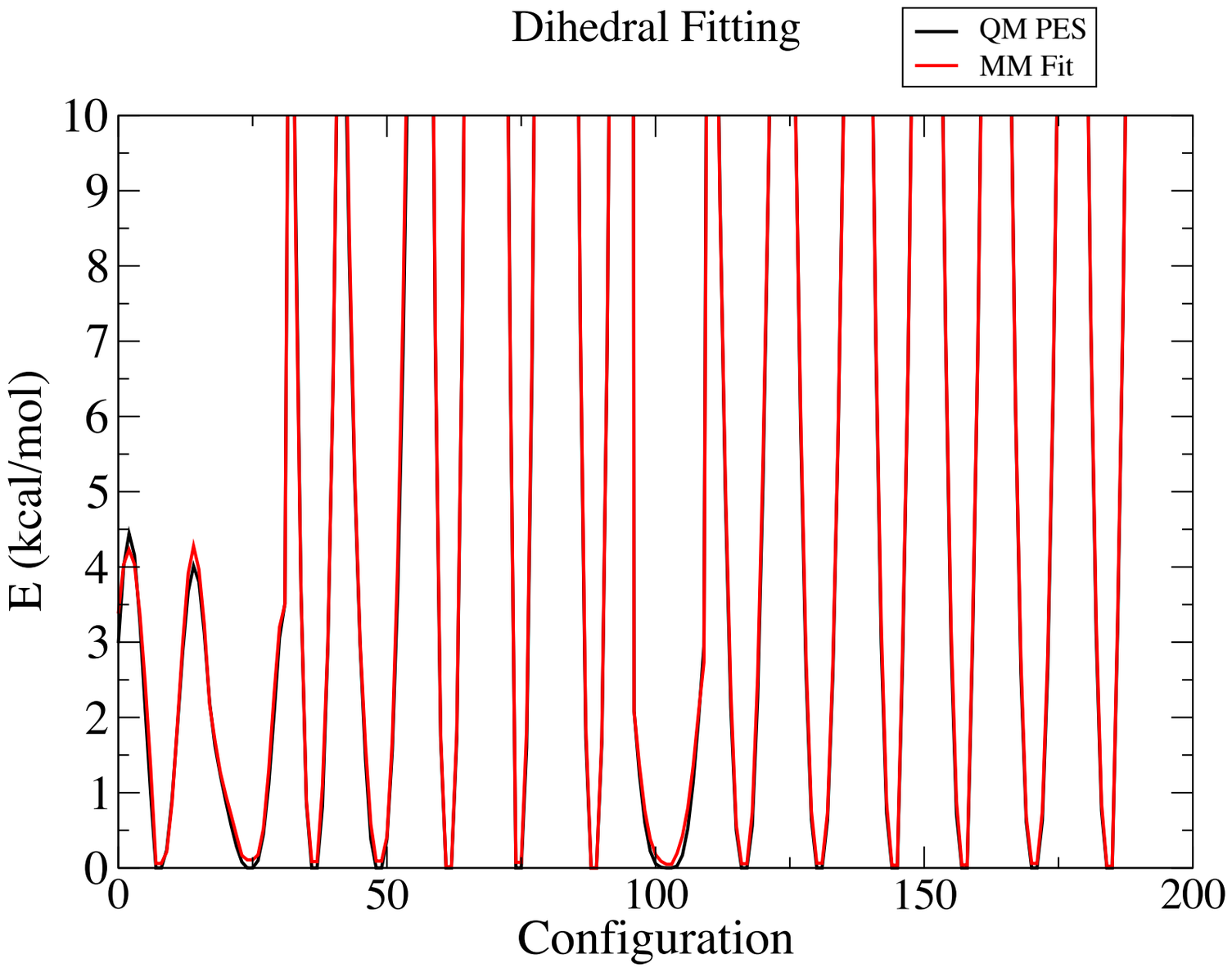
Pavlova, Gumbart. Parametrization of macrolide antibiotics using the force field toolkit. *J. Comp. Chem*. 2015, **36**, 2052−2063.

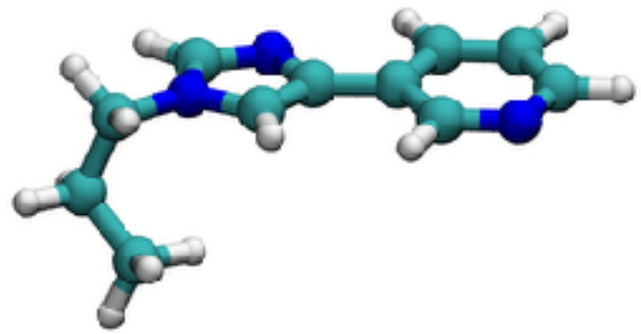
Example of Overfitting

imidazole-pyrridine moiety of antibiotic telithromycin

planar dihedrals have multiplicity 2 and phase 180 deg.

$$k_2[1 + \cos(2\phi + \pi)]$$





The fit looks (surprisingly) better despite using fewer terms (why?)

Fitting a lot of parameters simultaneously cannot always find the best fit!

Pavlova, Gumbart. Parametrization of macrolide antibiotics using the force field toolkit. *J. Comp. Chem*. 2015, **36**, 2052−2063.

Example of Overfitting

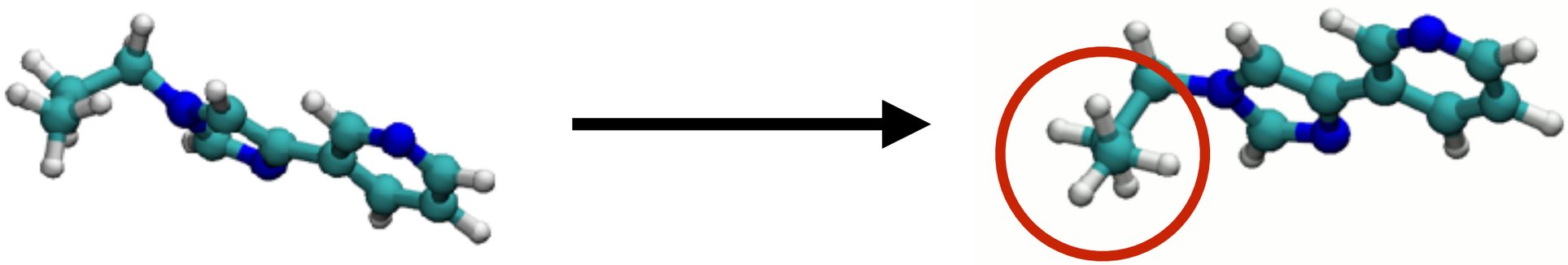
imidazole-pyrridine moiety of antibiotic telithromycin

planar dihedrals have multiplicity 2 and phase 180 deg.

$$k_2[1 + \cos(2\phi + \pi)]$$

Planarity is maintained!

Problems persist! Eclipsed conformation of the alkane



Restraining phase of CH dihedrals to 0 prevents eclipsed conformations

Pavlova, Gumbart. Parametrization of macrolide antibiotics using the force field toolkit. *J. Comp. Chem*. 2015, **36**, 2052−2063.

How to know what terms to include?

Back to CGenFF main page

CGenFF Frequently Asked Questions

Contents

How can dihedral contributions cancel out?

Why does a wind turbine or a propeller with 2 or 3 blades not wobble?

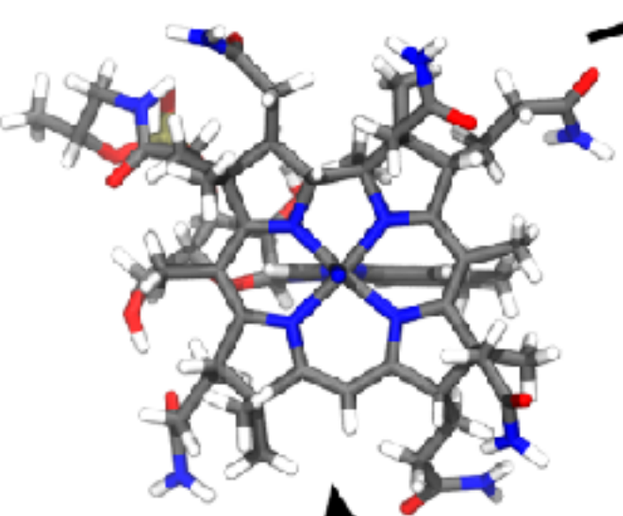
There are a number of common cases in which dihedral contributions cancel out:

- If a dihedral involves an sp^2 center as one of the inner atoms, and both substituents on this center are identical or have identical dihedrals, all terms with an odd multiplicity will cancel out.
- If a dihedral involves an sp^3 center as one of the inner atoms, and all three substituents on this center are identical or have identical dihedrals, all terms with a multiplicity that is not a multiple of 3 (ie. all terms other than 3-fold and 6-fold) will cancel out.
- Consequence 1: if both of the above conditions are satisfied (example: methyl rotation in toluele), only the 6-fold term will *not* cancel out.
- Consequence 2: if the substituents are not identical but some of their dihedral terms have the same phase and multiplicity, there may be partial cancellation and some of the terms can be omitted. (In practice, one would typically choose to omit H-X-X-H terms because these are more prone to transferability problems.)

All these observations follow from the expression for the dihedral energy:

https://mackerell.umaryland.edu/~kenno/cgenff/faq.php

Example: Parametrization of Cobalamins



ffTK

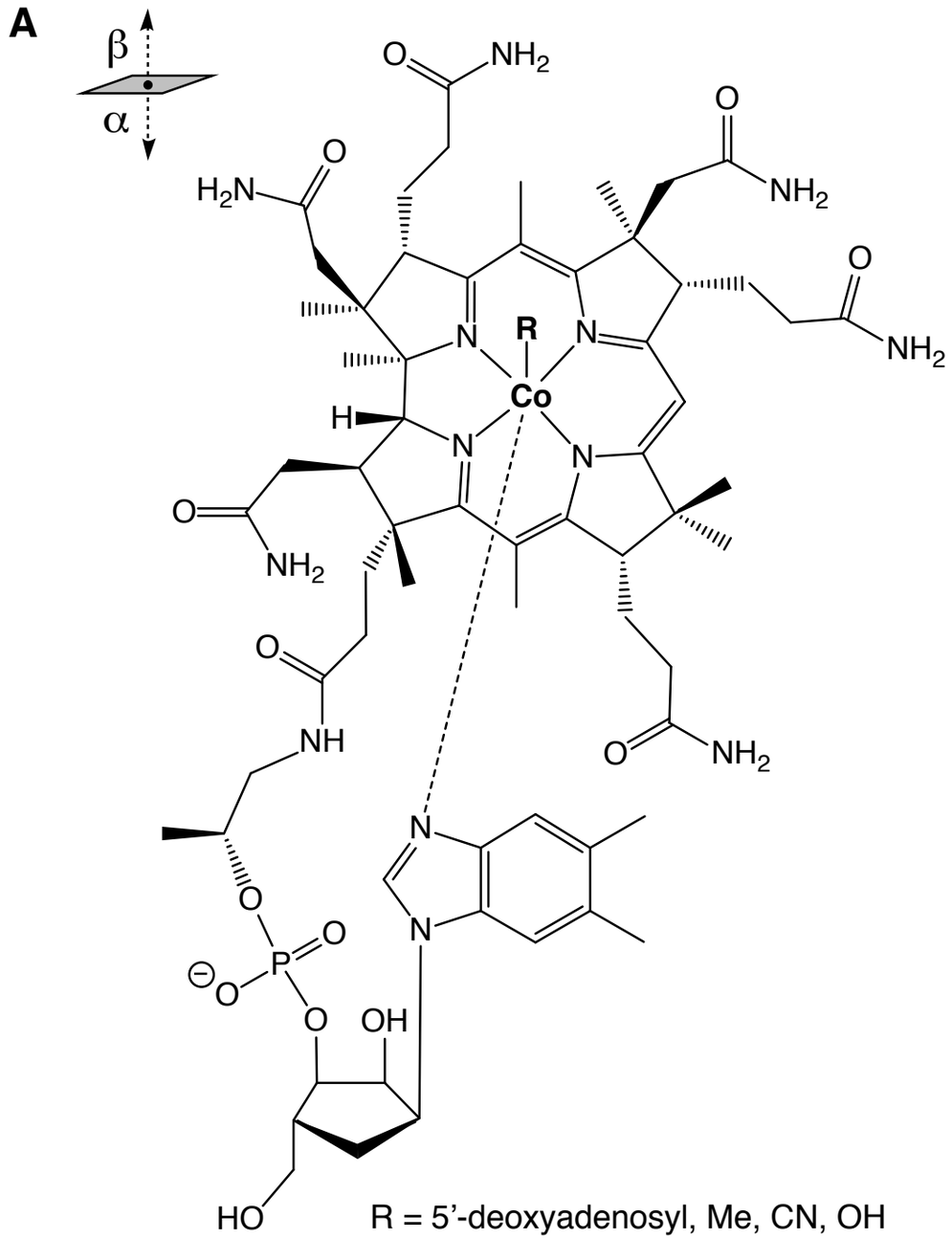
```
RESI  C03    0.00  ! cyano-cobalamin, Co(III)
GROUP
ATOM   CO  C06    0.234 !
ATOM  N21 NC06A  -0.387 !          H33A-N33-H33B
ATOM  N22 NC06B  -0.425 !                |
ATOM  N23 NC06C  -0.425 !          O34==C32-C31
```

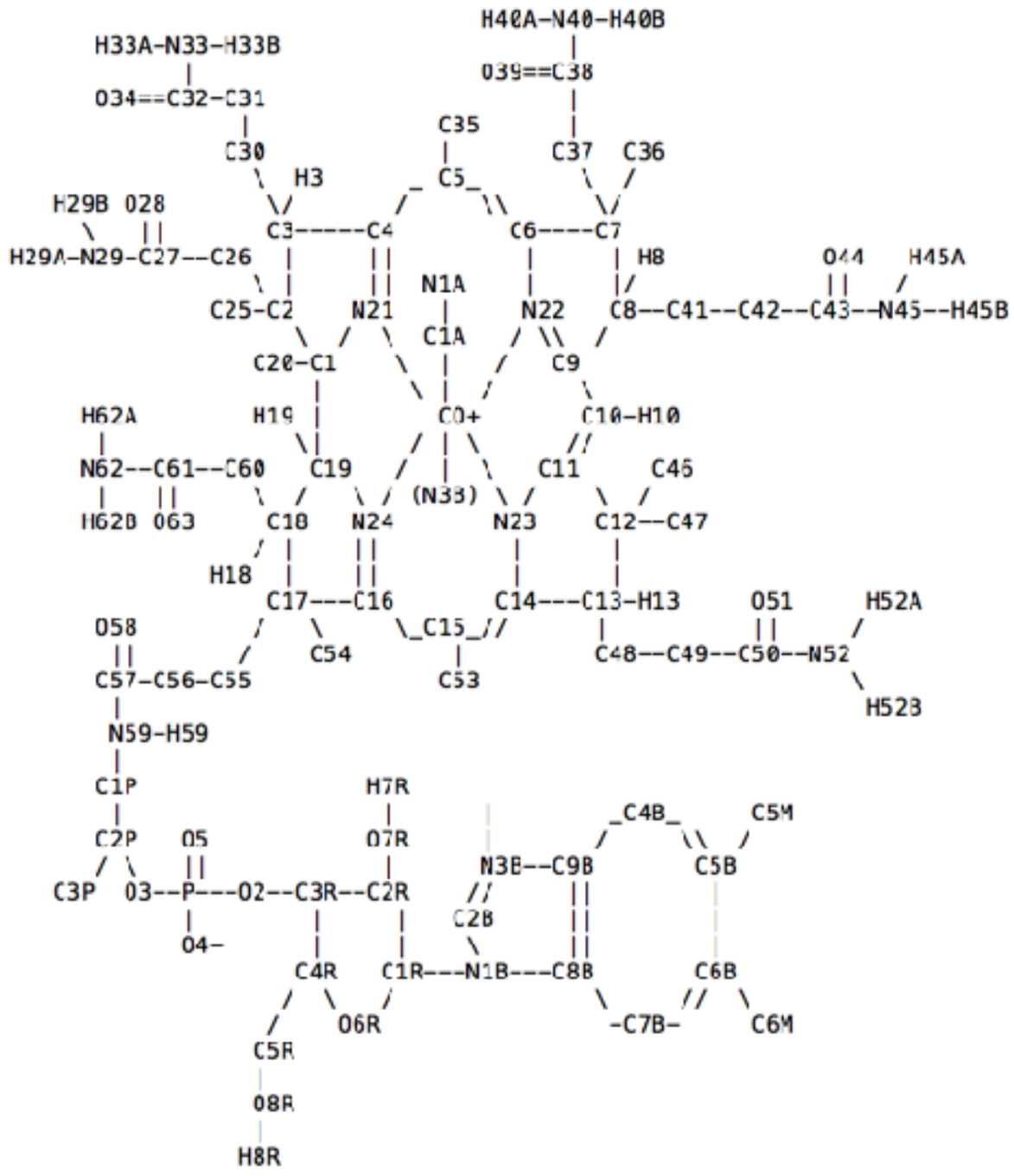
cobalamin is also known as vitamin B_{12}, is a large, cobalt-containing compound; inability to absorb vitamin B_{12} causes pernicious anemia

its large size and metal center make it particularly challenging for simulation

Pavlova, Parks, Gumbart. Development of CHARMM-Compatible Force-Field Parameters for Cobalamin and Related Cofactors from Quantum Mechanical Calculations. *J. Chem. Theory Comput*. 2018, **14**, 784−798.

Parametrization of Cobalamins





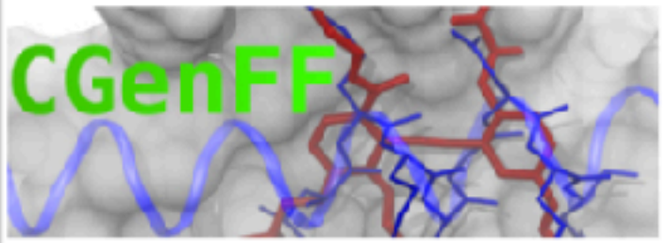
The first (tedious) step is to assign unique names to all the atoms!

(and make a nice ASCII schematic if you are so inclined!)

Pavlova, Parks, Gumbart. *JCTC*. 2018, **14**, 784−798.

Identify existing parameters w/CGenFF

Don’t reinvent the wheel! CGenFF webpage gives you estimated parameters derived from CHARMM General FF



New User? Register /Login

My Account | Upload molecule | More Info & Tools | About CGenFF

Welcome to CGenFF

The CGenFF program is a product of the ParamChem project. Other exciting technologies such as the lsfitpar program for robust fitting of bonded parameters are listed on ParamChem's technology page. Future directions for the CGenFF program itself can be found on our future prospects page.

The CHARMM General Force Field (CGenFF) program performs atom typing and assignment of parameters and charges by analogy in a fully automated fashion. Atom typing is done by a deterministic programmable decision tree. Assignment of bonded parameters is based on substituting atom types in the definition of the desired parameter. A penalty is associated with every substitution and the existing parameter with the lowest total penalty is chosen as an approximation for the desired parameter; the "penalty score" is returned to the user as a measure for the accuracy of the approximation. Charges are assigned using an extended bond-charge increment scheme that is able to capture short- and medium-range inductive and mesomeric effects.

CGenFF *program* links

- Usage information.
- Summary of output data and its utilization (**required** reading).
- FAQ (read this before contacting us with questions).
- How to cite / references.

CGenFF *force field* links

- Latest CGenFF version (**required** for using the output of the CGenFF program).
- Introduction.
- FAQ.
- Parameter optimization tutorial

https://cgenff.paramchem.org/

Identify existing parameters w/CGenFF

8EX_11_v2.str

```
* Toppar stream file generated by
* CHARMM General Force Field (CGenFF) program version 1.0.0
* For use with CGenFF version 3.0.1
*

read rtf card append
* Topologies generated by
* CHARMM General Force Field (CGenFF) program version 1.
*
36 1

! "penalty" is the highest penalty score of the associated parameters.
! 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.

RESI 8EX          0.000 ! param penalty= 271.500 ; charge penalty= 142.747
GROUP                   ! CHARGE   CH_PENALTY
ATOM C16     CG331  -0.272 !    2.380
ATOM H16     HGA3    0.090 !    0.325
ATOM H84     HGA3    0.090 !    0.325
ATOM H85     HGA3    0.090 !    0.325
ATOM C14     CG311  -0.103 !    8.307
ATOM H83     HGA1    0.090 !    0.954
ATOM C13     CG321  -0.179 !    8.719
ATOM H13     HGA2    0.090 !    0.808
ATOM H14     HGA2    0.090 !    0.808
ATOM C11     CG3RC1  0.056 !   27.453
ATOM H29     HGA1    0.090 !    2.560
ATOM O17     OG3C51 -0.194 !   23.768
ATOM C18     CG2R53  0.569 !   28.843
ATOM O19     OG2D1  -0.403 !    2.530
ATOM C20     CG25C1 -0.326 !  142.747
```

The output is a combined topology and parameter file using CGenFF **atomtypes**

Pay attention to the penalties! Low penalty charges/parameters can be kept; high ones need to be optimized

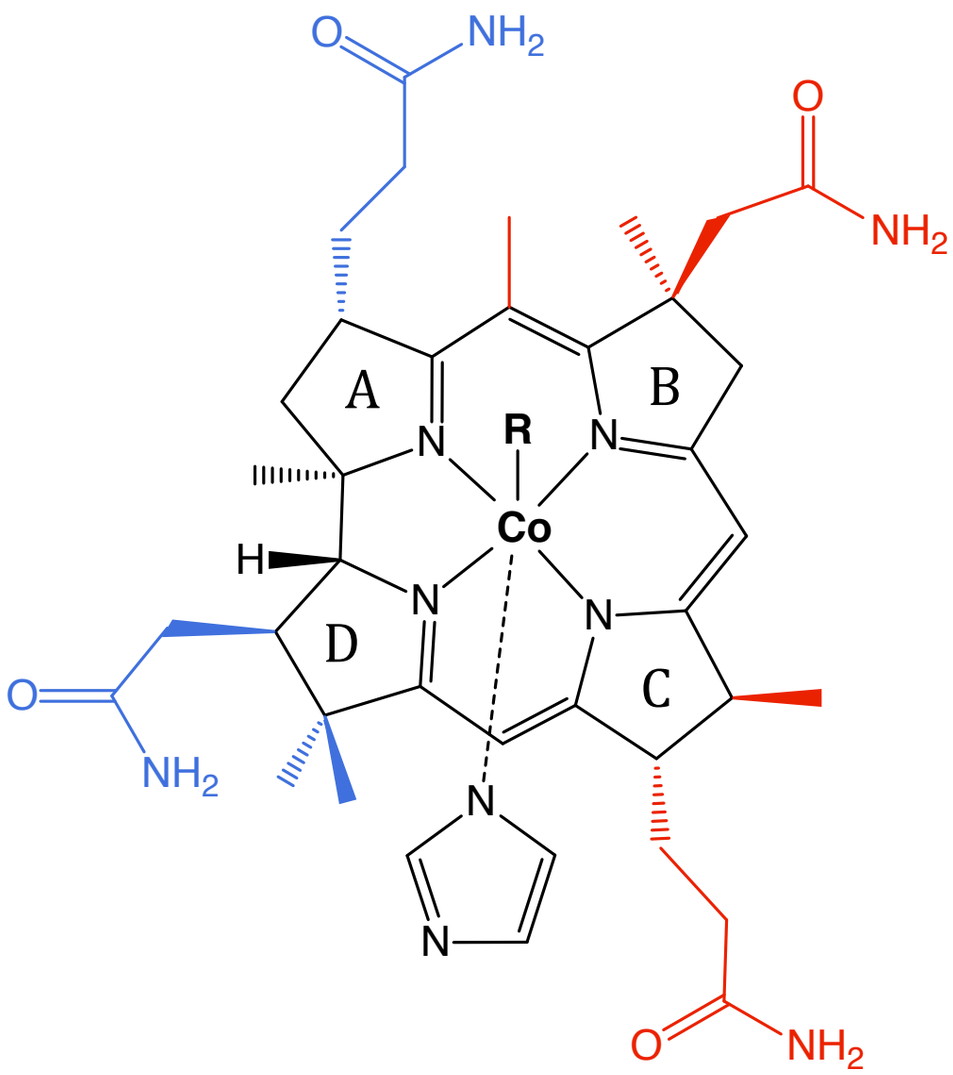
< 10 - keep

> 10 - optimize with FFTK

https://cgenff.paramchem.org/

Parametrization of Cobalamins

A



R = 5’-deoxyribose, Me, CN, OH

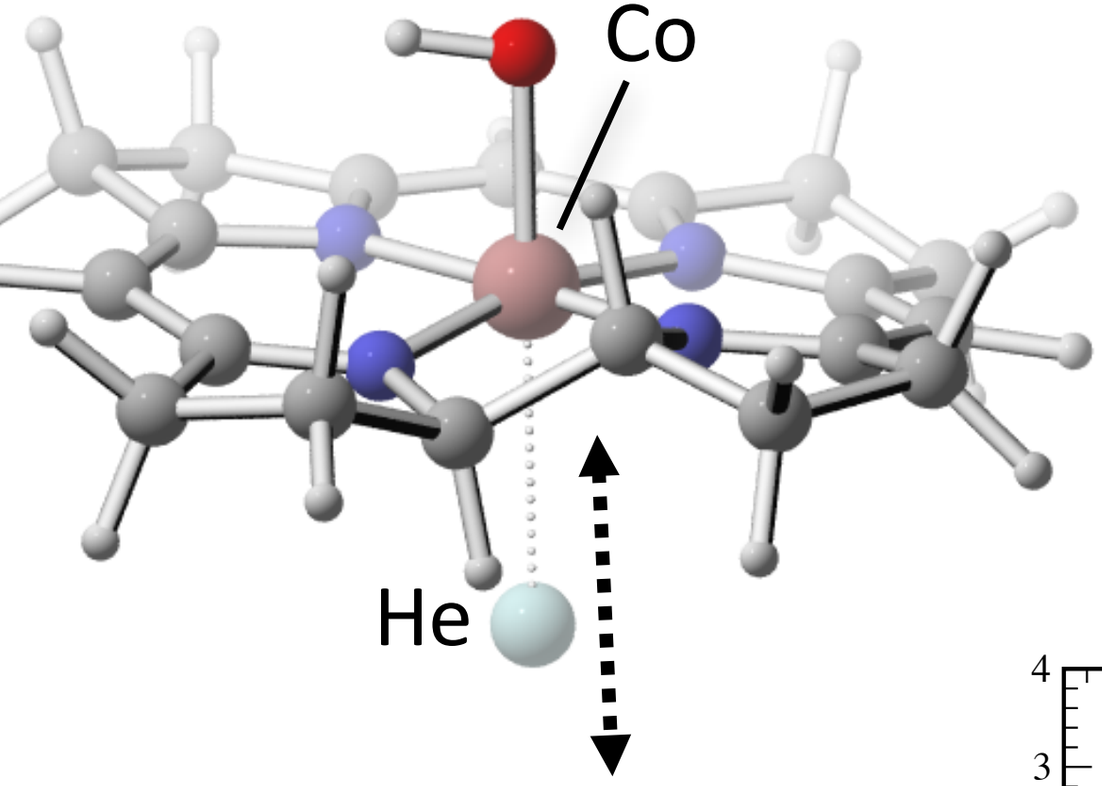
Often the molecule of interest is too large (> 50 atoms) and/or flexible for direct application of QM optimization

We create one or more molecule fragments for independent parametrization, combining them all at the end (possibly needing to create fragments for linker regions)

*Black is the corrin ring; w/**blue** and **red** side chains were separately used as well*

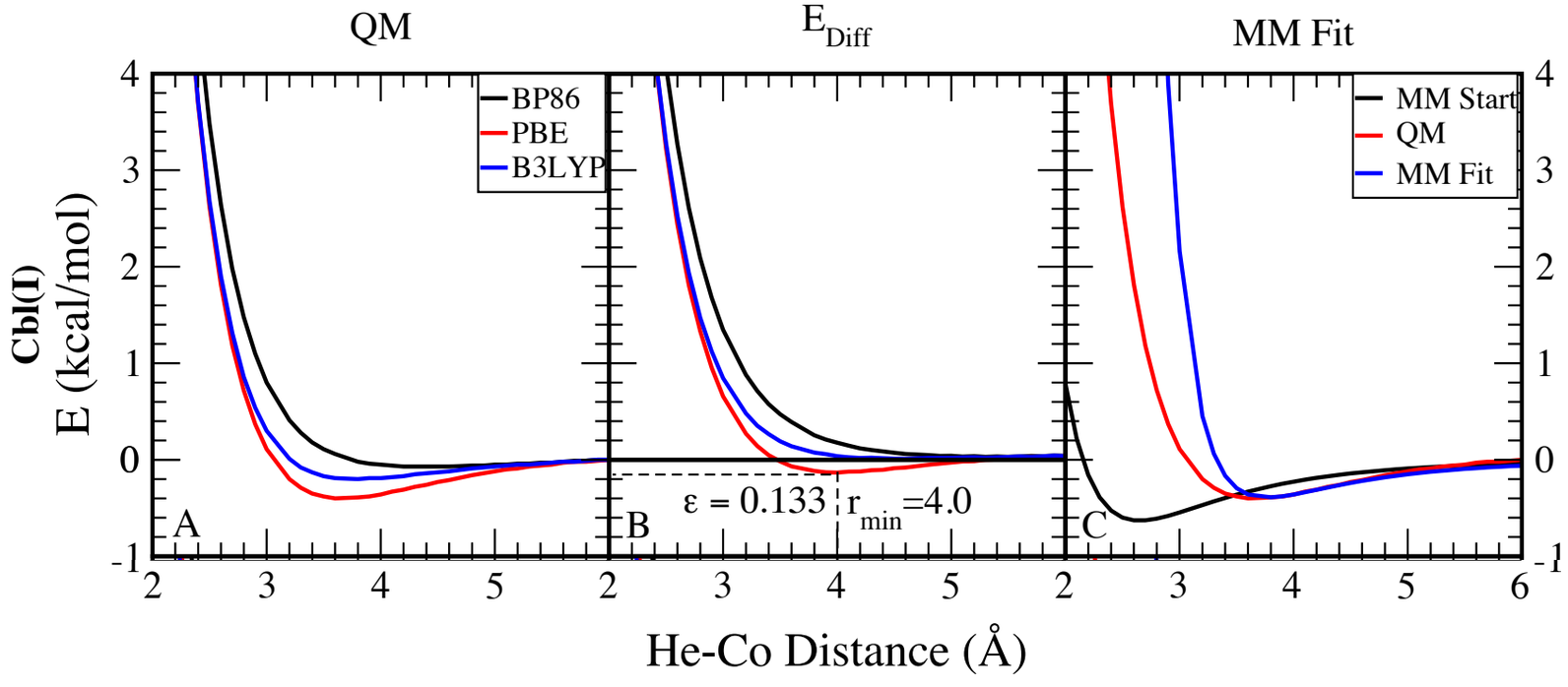
Pavlova, Parks, Gumbart. *JCTC*. 2018, **14**, 784−798.

Parametrization of Cobalamins: vdW



Optimizing van der Waals parameters is especially challenging, but rarely necessary as existing atom types are almost always appropriate

tried three different DFT methods (typically use MP2 or HF for CHARMM)

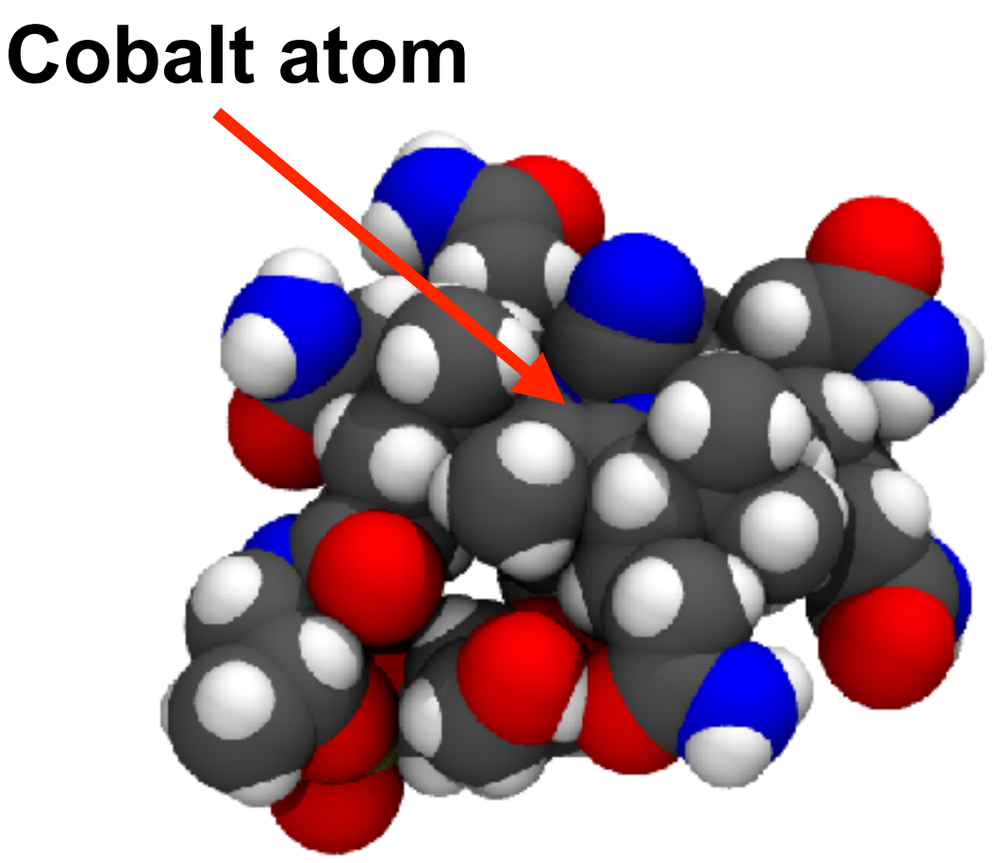


Used a Helium probe (no charge!) to fit interaction energies in QM and MM

Pavlova, Parks, Gumbart. *JCTC*. 2018, **14**, 784−798.

*Yin, D.; MacKerell, A. D., Jr. Combined ab initio/empirical approach for optimization of Lennard-Jones parameters. *J. Comput. Chem*. 1998, **19**, 334−348.
*Chen, I. J.; Yin, D.; MacKerell, A. D., Jr. Combined ab initio/empirical approach for optimization of Lennard-Jones parameters for polar-neutral compounds. *J. Comput. Chem*. 2002, **23**, 199−213.

Parametrization of Cobalamins: Charges



CHARMM focuses on interaction with waters - but where to place them for buried atoms?

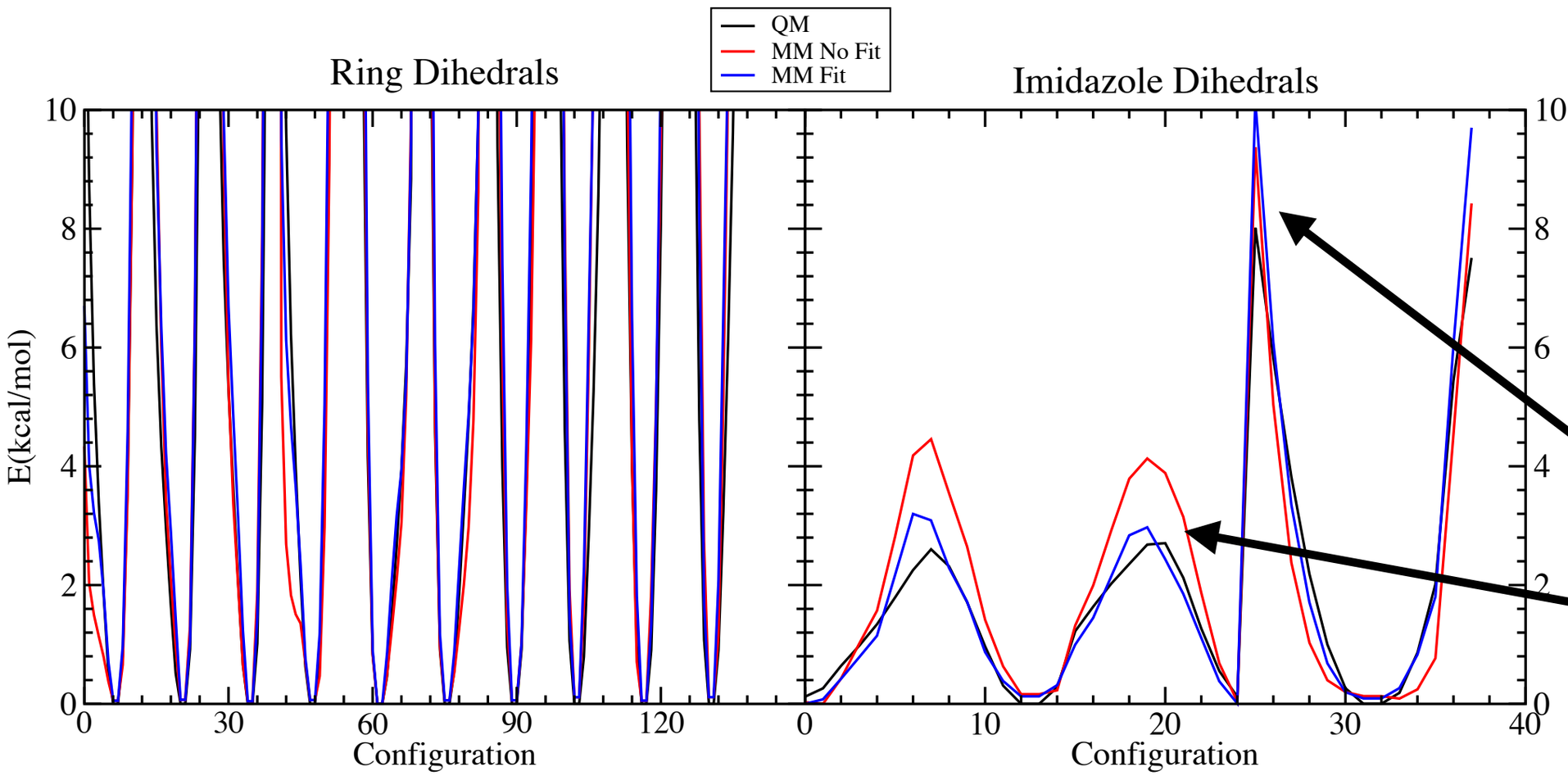
RESP approach was tried, but gave unphysical charges due to problems with buried atoms as well

Instead, a hybrid approach was used: Natural Population Analysis (NPA) for buried Co and N atoms; RESP for all others

Recommended approaches can and will fail!
Don't be afraid to experiment!

Pavlova, Parks, Gumbart. *JCTC*. 2018, **14**, 784−798.

Parametrization of Cobalamins: Dihedrals

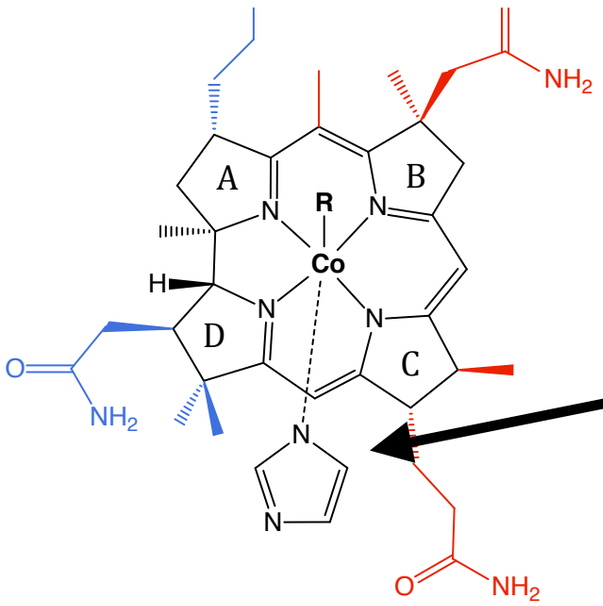


QM

MM (no dih.)

MM fit

fit to low energies is more important than fit to high (> 8 kcal/mol) energies



R = 5'-deoxyribose, Me, CN, OH

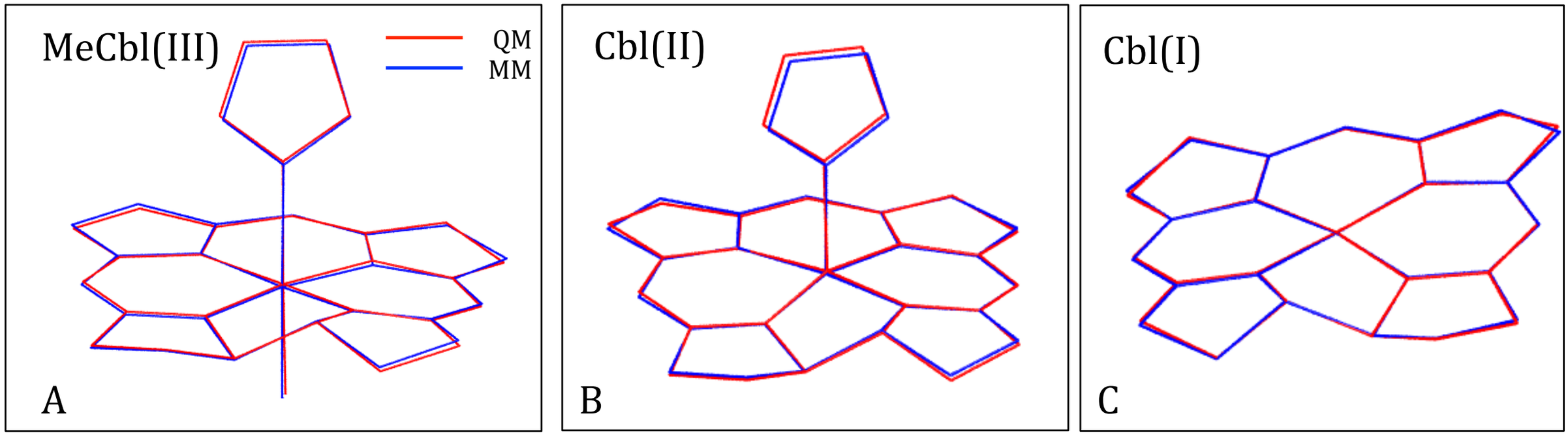
Many (but not all!) dihedrals within the corrin ring set to 0 (avoid overfitting)

Only *n* = 2 and *n* = 4 dihedrals for the imidazole were fit

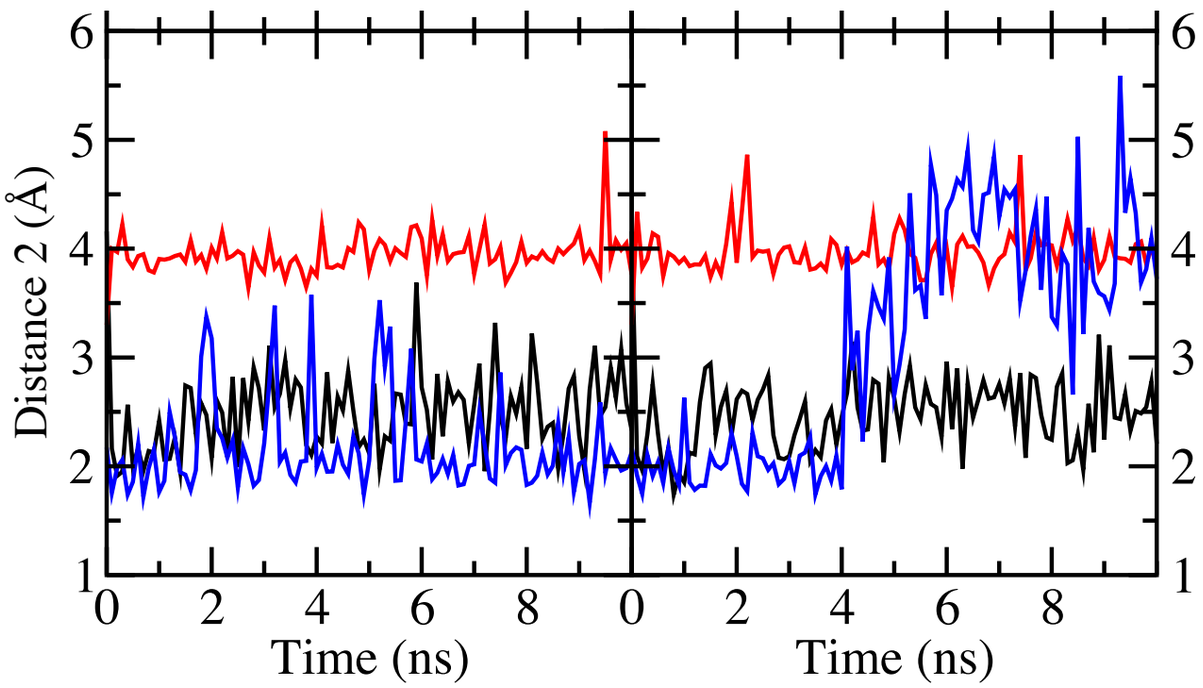
Multiple rounds of fitting and tested were required!

Pavlova, Parks, Gumbart. *JCTC*. 2018, **14**, 784−798.

Parametrization of Cobalamins: Validation



Minimization in NAMD using final parameters produced excellent agreement with QM minimized geometry (BP86/Def2-SVP)



Also ran simulations of Cbl bound to proteins, monitoring various interactions over time (each run **twice**)

Pavlova, Parks, Gumbart. *JCTC*. 2018, **14**, 784−798.

Conclusions

Find Missing Parameters

Geometry Optimization (QM)

Water Interaction En. (QM)

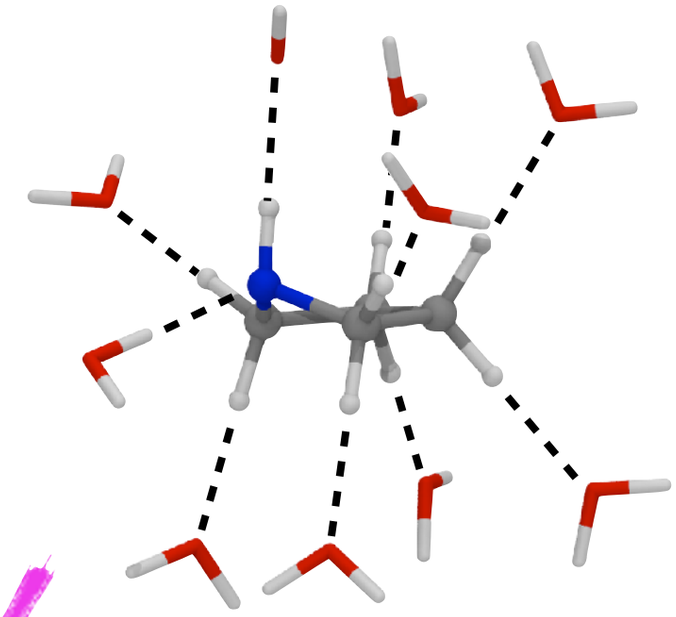
Charge Optimization

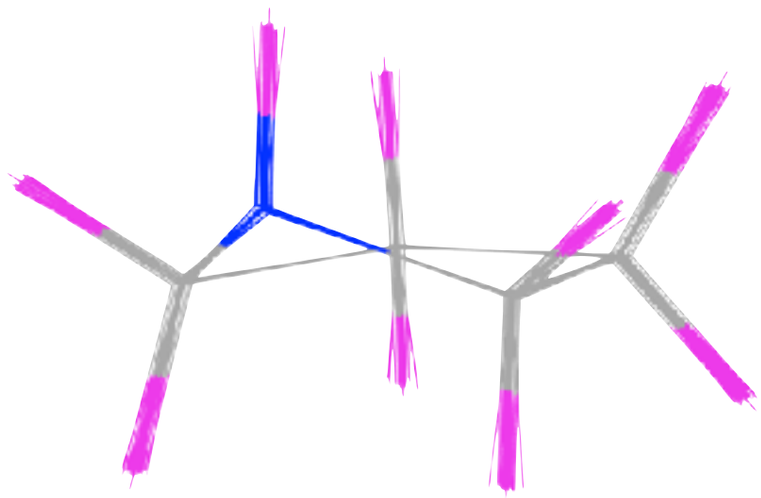
Hessian Calculation (QM)

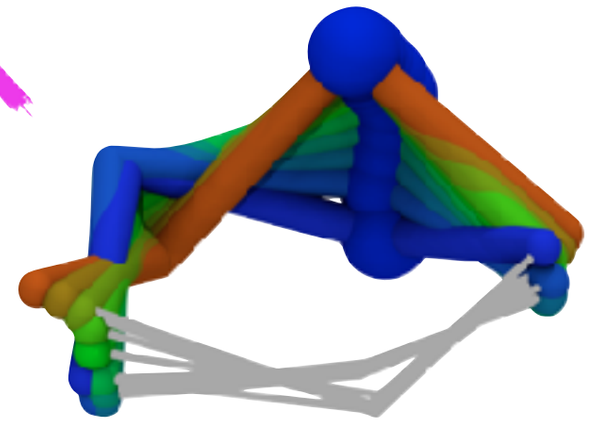
Bond & Angle Optimization

Torsion Scan (QM)

Torsion Optimization







ffTK:

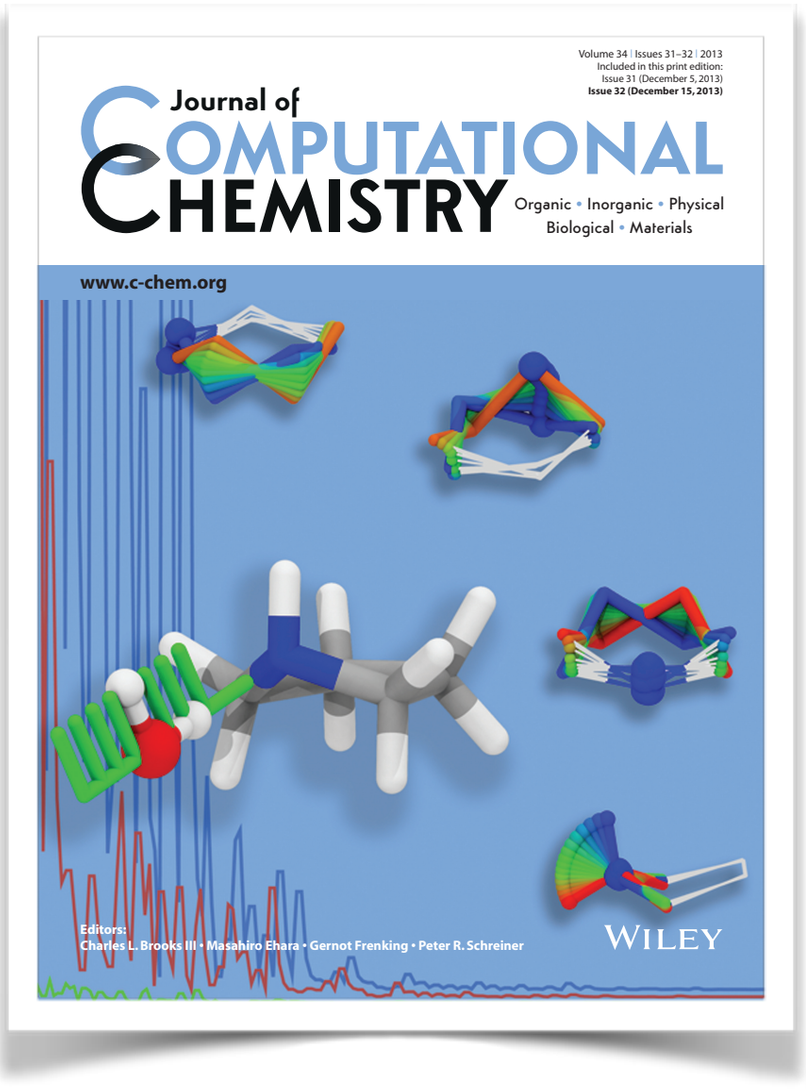
- Simplifies the parameterization workflow
- Offers opportunity for extensive customization
- Provides analytical tools to assess parameter performance

www.ks.uiuc.edu/Research/vmd/plugins/fftk

ffTK Mayne *et al.; J. Comp. Chem.* **2013,** 34, pp. 2757-2770

ffTK is available as a VMD Plugin (1.9.1 or newer)

http://www.ks.uiuc.edu/Research/vmd/plugins/fftk



Questions?