Parametrizing Small Molecules Using: The Force Field Toolkit (ffTK)

BuildPar	Opt. Geometry	Water Int.	Opt. Charges	Calc. Bonded	Opt. Bonded	Scan Torsions	Opt. Torsions	
► Identify Missing Parameters								
Assign Missing VDW/LJ Parameters by Analogy								
Update Parameter File with Optimized Parameters								
GUI Event Log (off)								

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



The CHARMM Force Field

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

$$\sum_{\text{dihedrals}} \mathcal{K}_{i}^{\text{dihedral}} \left[1 + \cos(n_{i} \varphi_{i} + \delta_{i}) \right] + \delta_{i}$$

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

Parameter Transferability In Biopolymers

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



Parametrization as an Impasse

non-standard or engineered amino acids



small molecule ligands





Imatinib (Gleevec)

Tiotropium (Spiriva)



metal centers



Geefal Parametrization Workflow



K. Vanorom Magneghet at. all, Compapit home 12020,10,43276742090.

ffTK Interface

->stansetardrfjændzielogssel	eratetissn buttons

	V					
BuildPar	Opt. Geometry	Water Int.	Opt. Charges	Calc. Bonded	Opt. Bonded	Scan Torsions Opt. Torsions
► Identify	Missing Parame	ters				
Assign	Missing VDW/LJ	Parameters b	by Analogy			
	rarameter File v		a Parameters			
Parameter Fil	g (off) es (both pre-define	d and in-progre	ss)			
/users/may	yne/fftk/PRLD/p	arfiles/prld-	init.par			Add
						Clear
Output LOG	: /users/mayne/fft	k/PRLD/3-char	ges/ChargeOpt.log			SaveAs
Charge	Constraints					
QM Tar	get Data				_	votion monu
Advance	ed Settings				C	
Results						

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



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



pyrrolidine

ffTK GUI

Input/Outp	put					
PSF File:	/Users/mayne/Desktop/pub_test/PRLD/rnd1/3-charges/prld-charged.psf					
PDB File:	/Users/mayne/Desktop/pub_test/PRLD/rnd1/2-geomopt/prld-opt.pdb					
Output Path:	./output	Browse				
Basename:	PRLD Basename From TOP Load PSF/PDB					





Charge Optimization

Assessing MM Water-Interaction Profiles

Sampling MM Water-Interaction Profiles

Mode: Simulated Annealing

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

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

•	Force Field	Toolkit (ffTK) GU	1		
BuildPar 🍸 Opt. Geometry 🖹 Calc. ESP	Opt. ESP Ca	lc. Bonded O	pt. Bonded	Scan Torsions	Opt. Torsions
arge Optimization Method: RESP Fitt	ing O				
Input					
SF File:					Browse
DB File:					Browse
Charge: 0 Resname:	Resna	me From TOP			Load PSF/PDB
Charge Constraints					
arge Group	Initial Charge	Restraint Type	Restraint Ato	m	
				Add	Guess
				Delet	e Move †
				Clea	/ Move↓

https://studynights.blogspot.com/2015/03/thesingle-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

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

		Force	e Field Toolki	t (ffTK) GU	l		- 5
uHPer Col Gen	анку — Жана	hi Op Castje	 Over the out 	Co. Turc-d	See Tesi ex (5:1 T ixinis	 CHEERE
np.t							
-SE -re.	400 tititataro	caviovale of this of t	dAnOxer				Erowse
FDE File	400 tititataro	caviovoQitto itti ort	tia nit-PE.pt :				Erowse
Hospie DG FLC.	Aos sidstaro	caviovo0;Tolith o/T	ala nHasalog				Erowse
n Prograw FAR Flo	e dor rittetaen		na repro				Friere
	e Haranetar H	les					
							≜ 5.0
							Dela
							I eo
							•
NAMD BIT	na osž						Friere
CURULOS:	E toda::Och	00					SaveAs
CUR 11/26-	f:Ocebar 3	ng					Sav=4x
Cutrul 1004 Derevelars In Oct	ErndasOch Inde	ng					Save 4a
Culsui (26) Derevelars in Oct Bond/White	ErndesOct hose	ng lor y	po Joh		-o co "Lenciari	البي:	59V=48
Cutrul 106- Maranelais In Oct BandWhale Cand	FindesOct hite	ng Hor y ATB1 M	ipe uch ∙⊂2≅6aM		- o do "Jeneio" 467 787	يني: 1 405"	5av=4a
Cute UT COM Devendens In Cont Bond Whole Cond Cond	FindesOct	ng Horis ATBR M 10023	poliust. • C2E6aM VIII-CA2		- 6 55 Londa 4 467 787 2461 025	تني: 1 4051 1 102.	5av=4a
Cute UT COM Devendens In Dort Bond Whole Cond Cond Cond	FindesOct	ng Horis ATBA M 10022 Ostooty	роция. - С256ан VIII-СА2 - N2-5004		- o do filonola 4 467 7871 1261025 511 8102	:20 14051 14021 12314	Faveés
Cutriul 1004 Parameters in Dort Bond/Whole Cond Cond Cond Cond Cond	E-mdaxOxt	ng History ATRA IS CO22 Calcom ARDAN	роция. - С256ан - Ч.С.А2 - К.С5094 - С.2.56394		- o do fuenda 1 467 7871 0467 7871 0467 1025 011 0102 181 5121	: 20 1 4051 1 1021 1 2314 1 4123	Saveés
Cutrul 1004 Parameters in Dort BondWhole Cond Cond Cond Cond Cond Cond Cond	EndesOct hos	ng History ATRA M CO22 Ostocoty ASD394 HO230	роция. • С256аня • НСА2 • N2-6004 • С256аня с 25211м		- 6 00 1.00012.1 467 7.371 0.461 0.25 611 61 0.2 481 51 21 0.49 56 0	:20 1 4051 1 102, 1 2314 1 41,23 1 025	Faveles
Cuts U1.064 Developed to Dort Bond/Whole ornd ocnd ocnd ocnd ocnd ocnd ocnd ocnd oc	EnnlasOct Ince	ng Hibring DOBCIN COLODA CARCAN HORAN NORCIN	роция. • С256/ди • НСА2 • N2-50/4 • С256/3/4 С256/16 • С256/16 • С256/16		- o do fuenda 1 467 7871 0461025 611 8105 481 5121 049 56 0 482 851 5	:20 1 4051 1 1022 1 2314 1 4123 1 022 1 2735	Faveles
Cute U1.064 Parameters In Def BendWhole anns annns anns anns anns annn	EndesOct Ince	ng Hori y DTRA M DC022 Ostoom DStoom HCP00 NDR010 NDR010 C2 S15	роция. • 62563-4 • 1-0,42 • 1-0,42 • 1-0,50-4 • 0,2563-4 • 0,2563-4 • 1-0,550 • 1-0,550		- o políticnola (1 467 747) 1467 747) 1467 1025 617 8102 481 5127 049 5610 482 6510 287 2854	:20 1 405 1 102 1 2314 1 412 1 232 1 233 1 233 1 233	Favelas
Cute U1.064 Parameters In Def BendWhole Cond	EnnlasOct Ince	ng Hibring Other M Code Osteom Code Noteom Noteom Noteom Codern Codern	рацов 10256304 МСА2 150-5004 10256304 0000114 10255304 21 50550 1525604	k	- o co lucado 1 467 7671 0461025 611 8102 481 5121 049 5610 482 8510 981 984 482 0351	:20 1 4051 1 1022 1 2314 1 4123 1 2735 1 2735 1 288 1 1107	FaveAs
Cuts U1.064 Parameters In Dert BerndWhole bend bend bend bend bend bend bend ben	EnnlasOct Ince	ng Hori y Diret M Colody Ostody Datady Horad Nataria Nataria Dataria	рацов 10256304 МСА2 15256304 10256304 Соосны 10256304 2150660 1525604	k	- 6 00 lende 1 467 767 646 1025 611 8105 481 5121 649 56 1 482 8510 681 984 482 6081	:20 1 405 1 102 1 2314 1 412 1 273 1 273 1 273 1 273 1 28 1 7:07	Favelas
Cuts U1.06* Parameters In Def Bend/Minole Cond Cond Cond Cond Cond Cond Cond Con	EndesOct hote	og Horry CTRALM CO22 Ostboth CSR52M HCR31M CSR51M CSR51M	рацот. 10256304 У НСА2 150-5004 150-5004 10056304 2006050 150-5000 150-5000	k	- 0 00 londo 11 467 767 0461025 611 8102 481 5121 049 56 10 482 8515 081 0854 482 0381	:g/0 1 4051 1 1021 1 2314 1 4123 1 4123 1 2735 1 2735 1 88 1 7:07	5 avs 4 a 5 avs 4 a 1 u cel 0 cel
Cute UT DOA Parameters in Dorf Bond/Whole Cond Con	FindesOct http://	ng Horry CCR21M CC022 CSR25M HCR30 NCR31M K245TM CSR21M	реция. 102563-4 9140,42 150363-4 160363-4 000014 162463-4 162463-4 162463-4 162463-4 162463-4 162463-4 162463-4	k	- 0 00 luncia 1 407 707 C401025 611 8102 481 5127 C49 50 0 482 8515 081 0394 492 0097	:20 1405 1102 12314 1412 1273 1273 1273 1273 1273 128 1707	 Fave 4s >.ev. b.col b.col b.col b.col b.col b.col b.col
Cute UT DOV Peranetare to Dort Bond Whole cond con	EndesOct http://www.	ng HDY I y DTRY M DC022 C2D23M HCP0 NCR21M K2P37M K2P37M C2D21M	реция. 172563-4 172563-4 172563-4 172563-4 172563-4 172563-4 172563-4 172563-4 172563-4 172563-4 172564	k	- 0 00 10001011 407 7071 0401025 611 8102 481 9121 049 96 0 462 8510 981 9354 492 0091	:20 1405 1405 15314 1412 1620 1573 1620 1573 168 1760	Fave 4s
Cute UT DOV Perenetaria In Dorf Bond/Whole cond co	EnndesOct	ng Hori y ATRA M LO223 Ostoch Ostoch Ostoch NSRCM HORO NSRCM CARATM	реция. • 62563-м • 6256-м • 6	k	- 6 00 10001011 467 7071 0201025 611 8100 481 5121 049 56 0 482 8510 081 0854 482 6010	: 270 1 405 1 102 1 2314 1 412 1 273 1 273 1 273 1 273 1 28 1 7407	Fave 48
Cute UT COM Parameters in Cont BendWhole cond	EnndesOct Inice : :	ng HDY S ATRA M LC022 Catalow HCP00 NDRCM CATRA CA	реция. • 6286634 • 6286634 • 6206334 • 6206114 • 6226334 • 6226134 • 6226334 • 622634 • 62264 • 622646 • 62264666666666666666666666666666666666	k	- 6 00 10001011 467 7071 0201025 011 8100 481 5121 049 56 0 482 8510 081 0854 482 0091	: 270 1 405 1 102 1 2314 1 412 1 273 1 273 1 273 1 273 1 28 1 7407	Fave 48

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

Force Field Toolkit (ffTK) GUI C I X SkileFor Op. Gountry Woten nil. Opt Charge. Odd Bundze Onl Bundze Spar Tansons Opt. Tans	$\sum_{dihedral} \left[1 + \cos(n_i \varphi_i + \delta_i) \right]$ periodicity (1-6 possible) phase (always 0 or 180 deg.)
Dela Eal RYEE Pol Color nEDt 2 (2000) Lke 11 1.947 Hub 12 C(3) Lke 13 C(3) Lke 14 1711 Hub 15 C(50) Lke 16 C(50) Lke 17 10.0 Hub 18 nuk de GYC Huk de YME 19 Delate Clear 11 Hub Delate Clear 11 Hub Hub Ext As Reft Input Ext As Reft Input 12 Hub Ext As Reft Input Ext As Reft Input	Energies above a threshold (e.g., 8-10 kcal/mol) are ignored
Directral Twos Definition Folde Constant (I) Pellodicity (I) Phase Brith (b) Look Phase? C2551M 02F61M 02R62M 02551M 8.002 1 0.00 ro 1 up inste C2551M 02F61M 02R62M 02551M 8.002 2 180.00 ro 1 up inste C2551M 02F61M 02R62M 02551M 8.201 2 180.00 ro 1	Elevant Ele

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!

imidazole-pyrridine moiety of antibiotic telithromycin

3 multiplicities, free phases for each dihedrals

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

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.

imidazole-pyrridine moiety of antibiotic telithromycin planar dihedrals have multiplicity 2 and phase 180 deg.

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

the force field toolkit. J. Comp. Chem. 2015, 36, 2052-2063.

imidazole-pyrridine moiety of antibiotic telithromycin

planar dihedrals have multiplicity 2 and phase 180 deg.

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

- Technical questions
 - <u>How do | compile CHARMM with CGenFF support?</u>
 What should I do if I get "LEVEL -4 WARNING FROM <RTFRDR> LIMIT EXCEEDED"?
 <u>What should I do if I get "LEVEL -3 WARNING FROM <PARRDR> Maximum no. of dihedrals reached"?</u>
- Using CGenFF

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 sp² 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 sp³ 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

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

C37

C11

C36

H8

C8-

C10-H10

C46

051

C5M

C6M

C5B

C6B

-C49--C50--N52

C12--C47

C48-

044

H45A

-N45--H45B

H52A

H52B

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

Welcome to CGenFF

The CGenFF program is a product of the ParamChem project. Other exciting technologies such as the lsfupar 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
                                                   The output is a combined
*
read rtf card append
                                                   topology and parameter file
* Topologies generated by
* CHARMM General Force Field (CGenFF) program version 1.
                                                   using CGenFF atomtypes
*
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
                       param penalty= 271.500 ; charge penalty= 142.747
                0.000 !
GROUP
               ! CHARGE
                         CH PENALTY
                                     Pay attention to the penalties! Low
ATOM C16
          CG331 -0.272 !
                            2.380
                  0.090 !
                            0.325
ATOM H16
          HGA3
          HGA3
                0.090 !
                            0.325
ATOM H84
                                     penalty charges/parameters can be
                            0.325
ATOM H85
          HGA3
                 0.090 !
ATOM C14
          CG311 -0.103 !
                            8.307
                                     kept; high ones need to be optimized
                            0.954
ATOM H83
          HGA1
                  0.090 !
          CG321
                -0.179 !
                            8,719
ATOM C13
ATOM H13
          HGA2
                  0.090 !
                            0.808
                                     < 10 - keep
ATOM H14
          HGA2
                  0.090 !
                            0.808
          CG3RC1 0.056 !
                           27,453
ATOM C11
ATOM H29
                            2.560
           HGA1
                  0.090 !
ATOM 017
          OG3C51 -0.194 !
                           23.768
ATOM C18
          CG2R53 0.569 !
                           28.843
                                     > 10 - optimize with FFTK
ATOM 019
          0G2D1 - 0.403
                            2.530
                          142.747
ATOM C20
          CG25C1 -0.326 !
```

https://cgenff.paramchem.org/

Parametrization of Cobalamins

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

Parametrization of Cobalamins: vdW

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!

Parametrization of Cobalamins: Dihedrals

R = 5'-deoxyribose, Me, CN

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**)

Conclusions

ff**TK:**

- 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?