Automated Force Field Parameterization for Atomic Models Based on Ab Initio Target Data

Li-Ping XU

Supervisor: Prof. Wiest

Prof. Wu


第十二届全国量子化学会议
发展精确的蛋白质分子力场
吴云东
北京大学化学与分子
北京大学深圳研
2014年6月13-15日
全国量化会，太原

第十二届全国量子化学会议

UNIVERSITY OF MINNESOTA

Computational Methods for Complex Systems

Beyond QM/MM: Development of a Quantum Mechanical Force Field

第十二届全国量子化学会议
（The12th National Conference of Quantum Chemistry)
山西太原， China： June13-15, 2014

Jiali Gao （高加力）

理论化学计算国家重点实验室，吉林大学
and
University of Minnesota

Outline

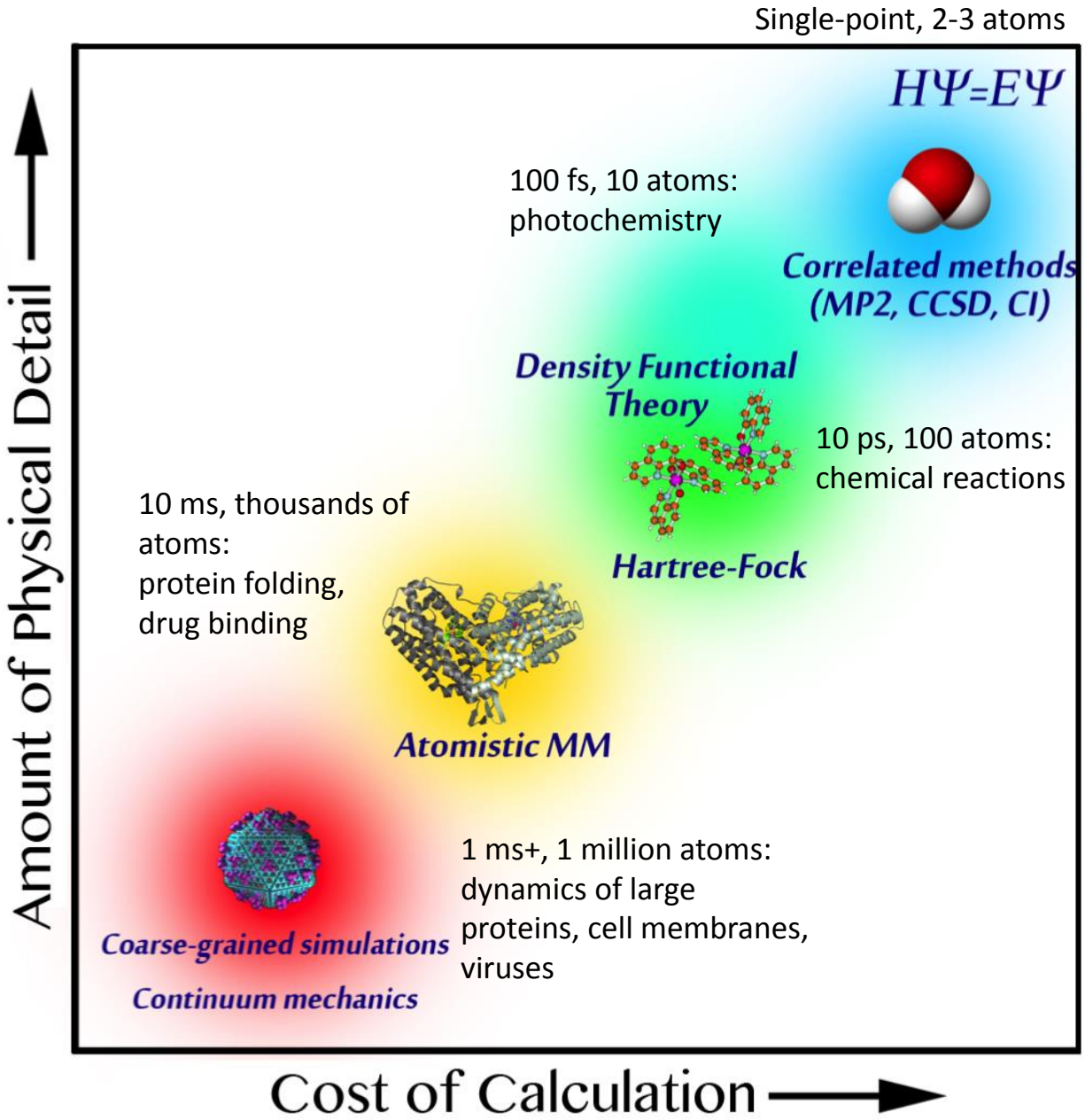
Introduction

- Force fields in molecular mechanics
- The ingredients of a force field
  - Functional form
  - Reference data
  - Optimization method

General Automated Atomic Model Parameterization

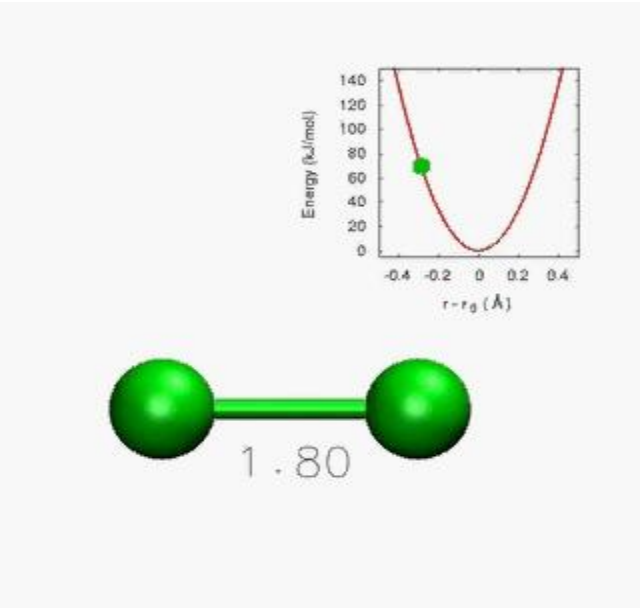
- Overview of method
- Results and discussion
- Limitation and possible improvements

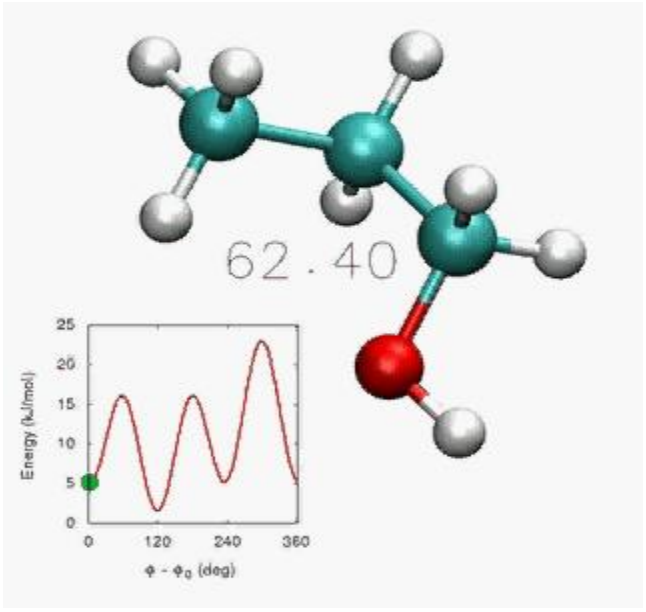
Introduction: A Wide Range of Simulation Domains

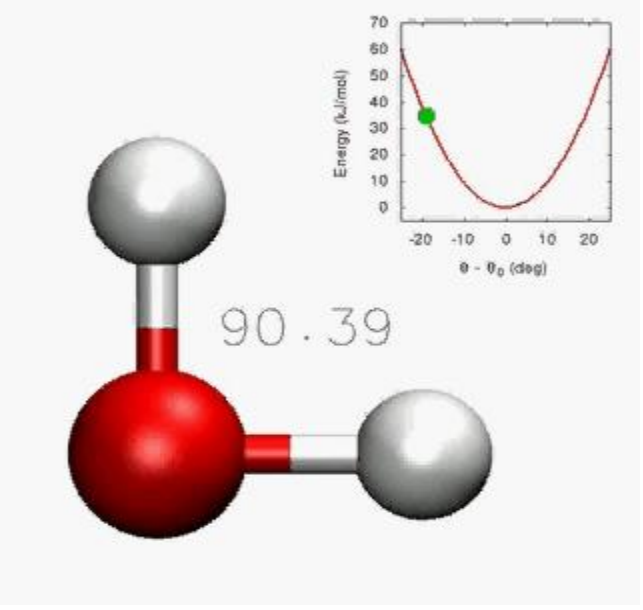


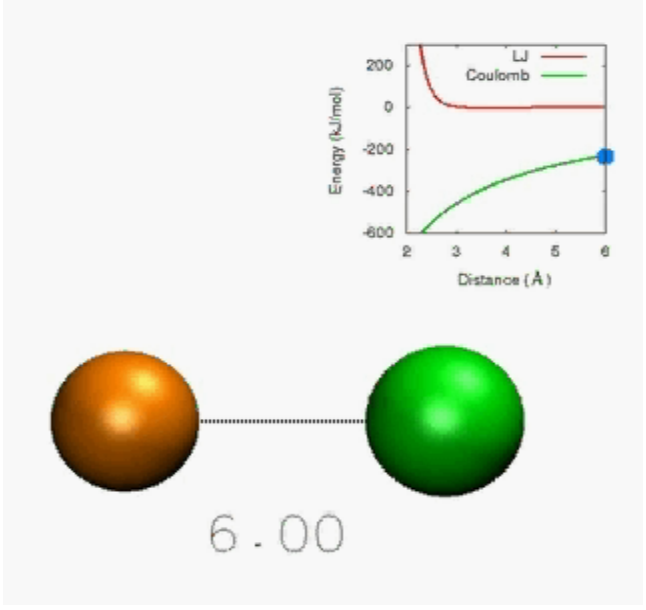
- Computer simulations of atoms and molecules span a vast range of detail
- More detailed theories can describe complex phenomena and offer higher accuracy
- Less detailed theories allow for simulation of larger systems/longer timescales
- In molecular mechanics simulation, the potential energy of molecules is represented using an empirical force field

Introduction: Force Fields









- **Force fields** are built from *functional forms* and empirical *parameters*
- Interactions include bonded pairwise, 3-body, and 4-body interactions…
- … as well as non-bonded pairwise interactions
- Simulation accuracy depends critically on choice of parameters

Introduction: Force Fields

The common paradigm for running simulations is to choose a force field from a large literature selection.

PROTEINS:

AMBER
"Assisted Model Building with Energy Refinement"

- Main series: ff94, ff96, ff99, ff03, ff10
- Dihedral modifications: ff99sb, ff99sb-ildn, ff99sb-nmr, ff99-phi
- GAFF (Generalized AMBER force field)

OPLS
"Optimized Potential for Liquid Simulation"

- OPLS-UA (united atom), OPLS-AA (all atom)
- OPLS-AA/L (revised torsions)
- OPLS-2001, OPLS-/2005 (improved solvation free energies)

CHARMM
"Chemistry at Harvard Molecular Mechanics"

- CHARMM19 (united atom), CHARMM27 (all atom)
- CHARMM36 (carbohydrates)
- CMAP (two-dimensional dihedral corrections)
- CGenFF (General CHARMM force field)

AMOEBA
"Atomic Multipole Optimized Energetics for Biomolecular Applications"

- Contains polarizable point dipoles

WATER:

TIP3P, TIP4P, TIP5P
"Transferable Intermolecular Potential"

- AMBER, OPLS, and CHARMM are "paired" with TIP3P
- TIP3P water melts at -146 ºC and boils at -90 ºC

SPC, SPC/E, SPC/Fw
"Simple Point Charge"

- Same functional form as TIP3P, different parameters

TIP4P/Ew, TIP4P/Ice, TIP4P/2005

- Reparameterization of TIP4P model
- Improved fits to experimental properties of water

Various polarizable models

- SWM4-DP, SWM4-NDP (contains Drude particle)
- AMOEBA (contains polarizable point dipoles)
- DPP, DPP2 (distributed point polarizable model)
- TTM2-F, TTM2-R, TTM3-F (Thole type model)
- TIP4P-FQ, SPC-FQ (Fluctuating charge model)

There are too many to choose from…

Can we create a force field that is best for our research project?

Creating a Force Field: Functional Form

Step 1: Choose a functional form to represent the potential energy surface, or design your own.

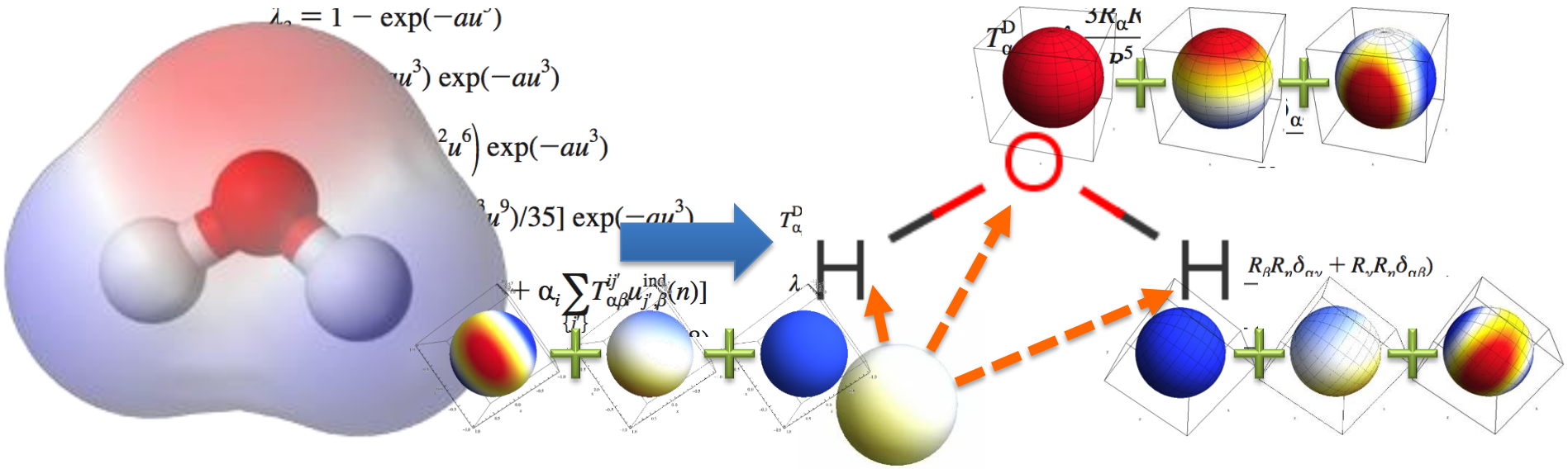
AMBER fixed-charge force field:

$$\sum_{i<j} \frac{q_i q_j}{r_{ij}}$$

- Point charge on each atom

AMOEBA polarizable force field:

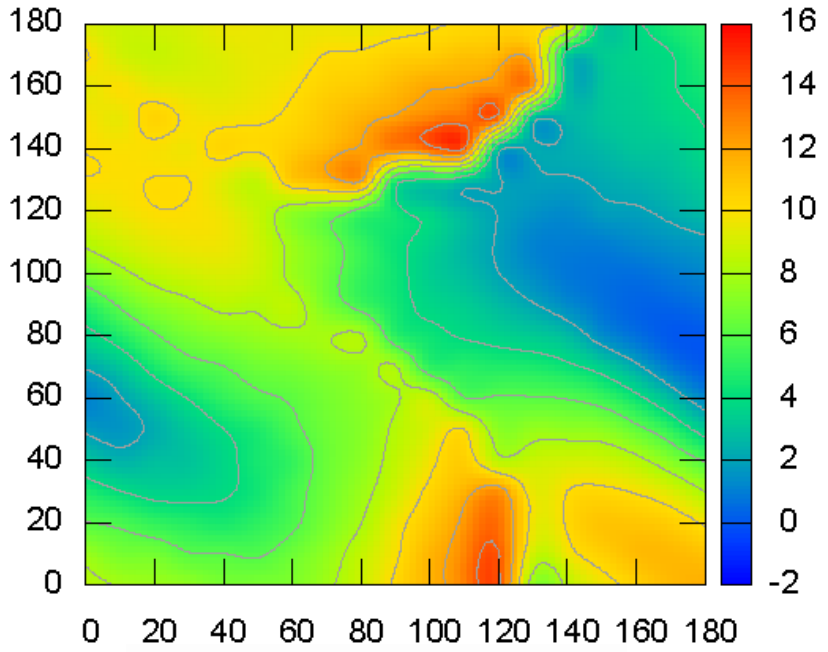
- Point charge, dipole, and quadrupole on each atom
- Polarizable point dipole on each atom with short-range damping



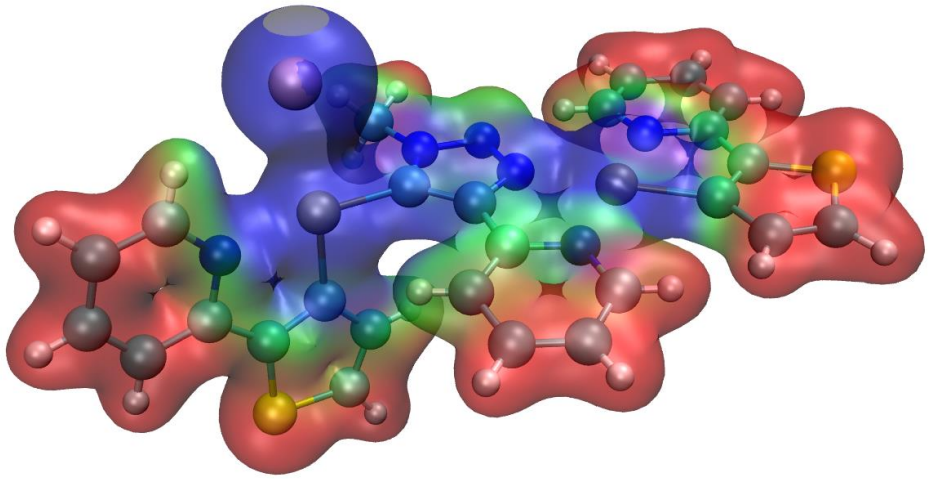
Creating a Force Field: Reference Data

Step 2: Create a reference data set from theoretical calculations or experimental measurements.

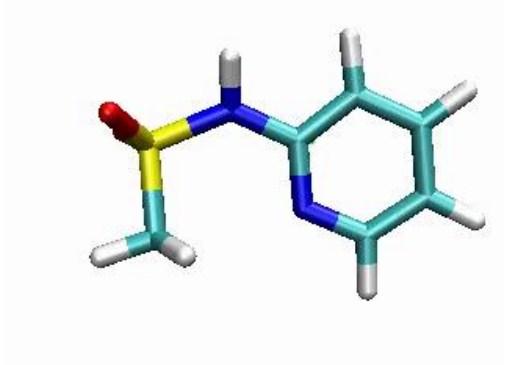
Energy scan across 2 dihedral angles

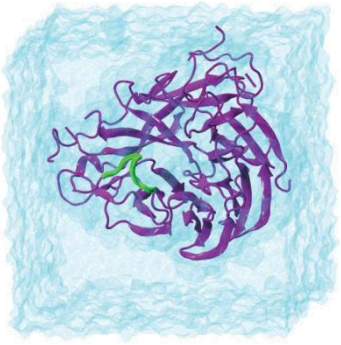


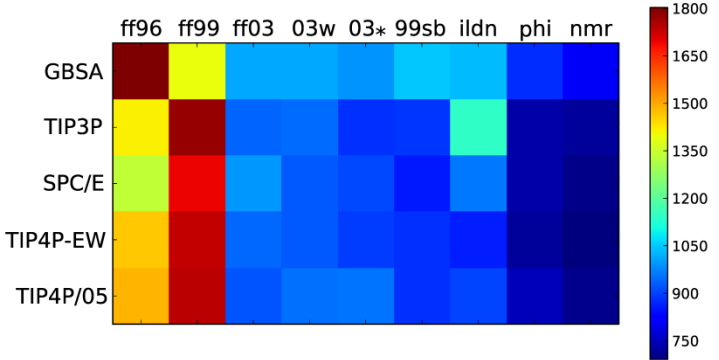
Electrostatic potential on a molecular surface (red = positive, blue = negative)



Simulated vs. experimental NMR chemical shifts for proteins (red = bad, blue = good)







Creating a Force Field: Optimization Method

Step 3: Construct an objective function and apply an optimization method to minimize it.

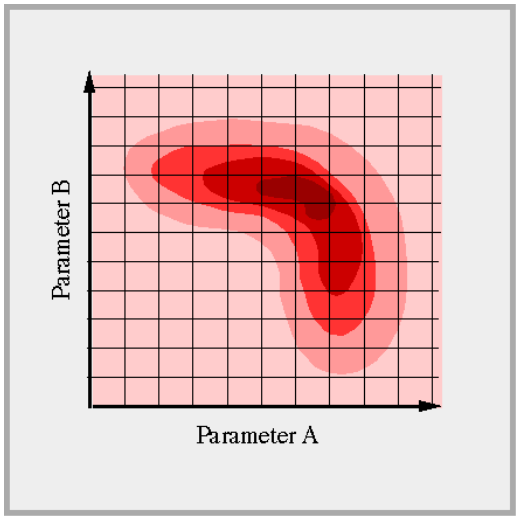
- The **objective function** measures the disagreement between the reference data and corresponding simulation result.
- An **optimization algorithm** searches for parameters that minimize the objective function.

R = Reference Data

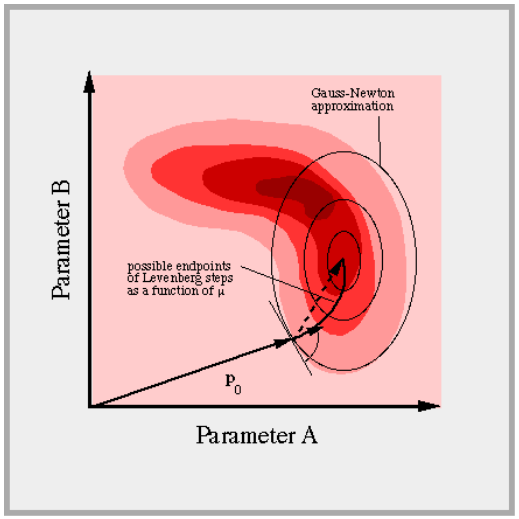
S = Simulation Result

$$\chi^2(\mathbf{k}) = \left(R - S(\mathbf{k})\right)^2$$

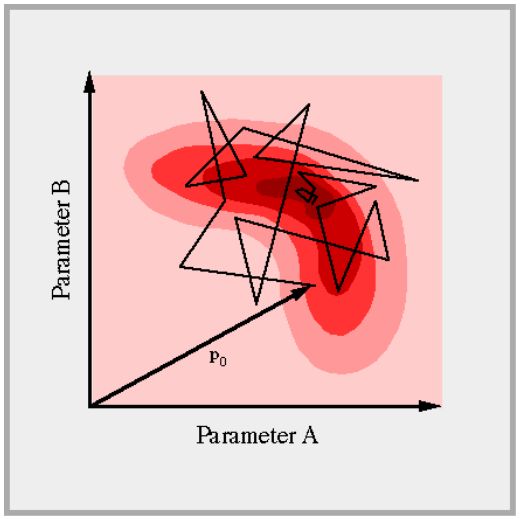
$$\mathbf{k}_{opt} = \min_{\mathbf{k}} \chi^2(\mathbf{k})$$



Grid Scan



Newton-Raphson



Simulated Annealing

Outline

Introduction

- Force fields in molecular mechanics
- The ingredients of a force field
  - Functional form
  - Reference data
  - Optimization method

General Automated Atomic Model Parameterization (GAAMP)

- Overview of method
- Results and discussion
- Limitation and possible improvements

Huang, L.; Roux, B. *J. Chem. Theory Comput.* **2013**, *9*, 3543-3556

Introducing GAAMP

There are some tools for parameterization of novel molecules:

- **Antechamber**: automatically parameterize small compounds in accord with general Amber force field (GAFF)
- **CGenFF**: provide CHARMM-consistent force field parameters for small compounds and drug-like molecules.

However, partial charges and dihedral parameters have limited transferabilities.
The accuracy is still a problem.

Introducing GAAMP

- General automated atomic model parameterization (**GAAMP**) aiming at achieving an automatic parameterization for small molecules using ab initio QM results as the primary target data.

A. Functional form (Nonpolarizable CHARMM force field)

$$
\begin{aligned}
E = & \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 \\
& + \sum_{\text{Urey-Bradley}} K_{UB}(r_{1,3} - r_{1,3;0})^2 \\
& + \sum_{\text{dihedrals}} K_\phi(1 + \cos(n\phi - \delta)) \\
& + \sum_{\substack{\text{improper} \\ \text{dihedrals}}} K_\varphi(1 + \cos(n\varphi - \varphi_0)) \\
& + \sum_{\text{nonbonded}} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]
\end{aligned}
$$

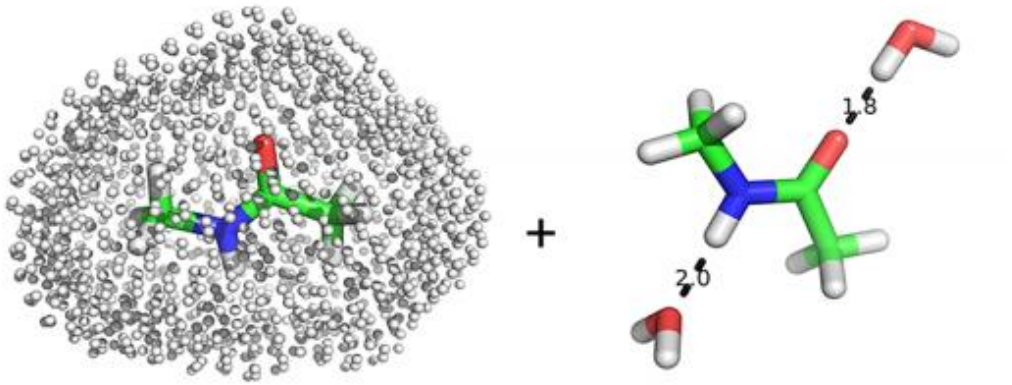
Introducing GAAMP

B. Parameterization method

1. Bond length and angle parameters: from GAFF, CGenFF geometry or QM calculation

2. Charge fitting: combination of ESP fitting and compound-water interaction fitting

$$\chi^2 = \chi_{\mathrm{ESP}}{}^2 + \chi_{\mathrm{wat\_int}}{}^2 + \chi_{\mathrm{CG}}{}^2$$

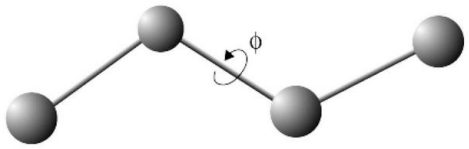


QM Electrostatic potential (ESP)

QM interactions with water, E_{int} and R_{int}

Introducing GAAMP

B. Parameterization method



3. Dihedral parameter fitting

Identification of all conformers (dihedrals rotate);

Clustering dihedrals and delete redundancies;

Dihedral scan first at MM level then at the QM level to find optimal structures;

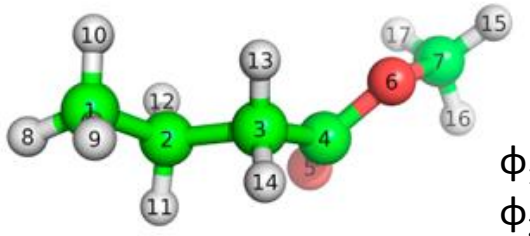
Fitting using scan information and conformer energy as reference.

4. Optimization algorithm: Augmented Lagrangian conjugated with L-BFGS algorithm

GAAMP: Result and Discussion

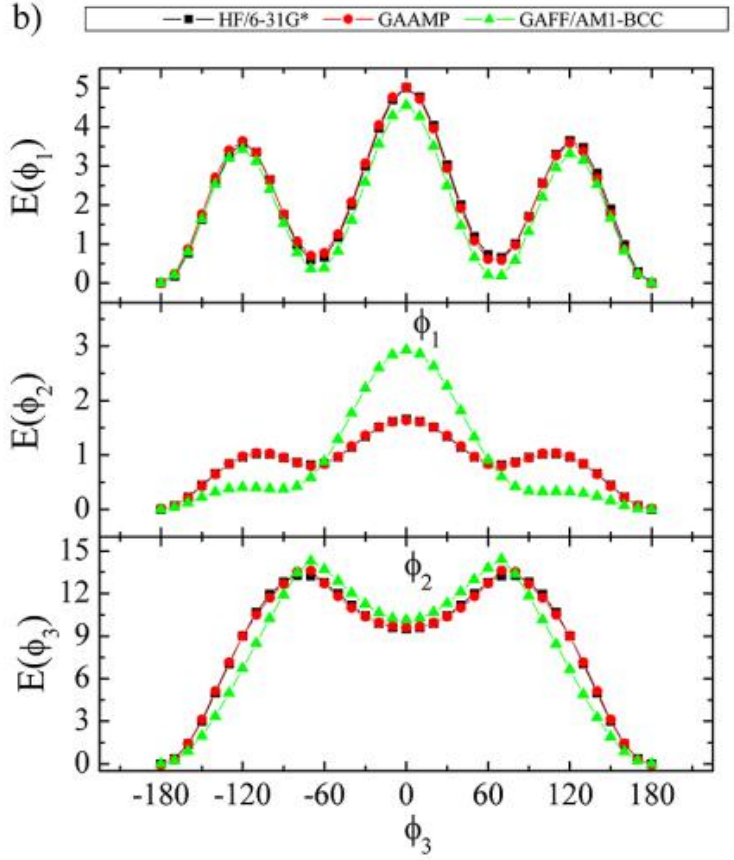
Results 1. Dihedral parameters

a)



ϕ_1 = 1-2-3-4
ϕ_2 = 2-3-4-6
ϕ_3 = 3-4-6-7

b)

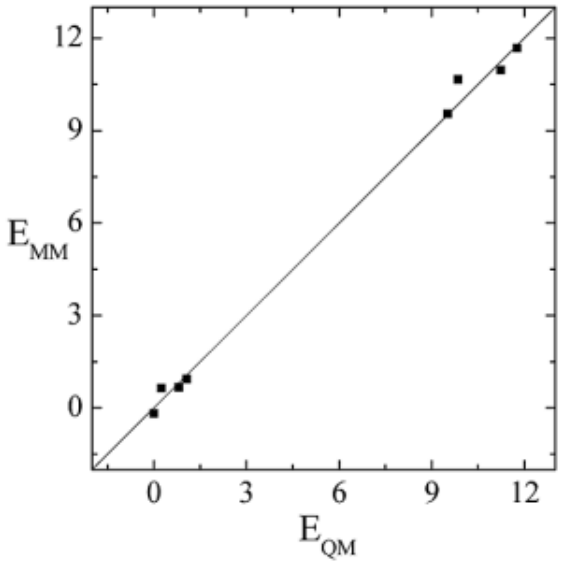


GAFF/AM1-BCC works reasonably for this molecule;

GAAMP perfectly matches QM results of the dihedral energy profiles;

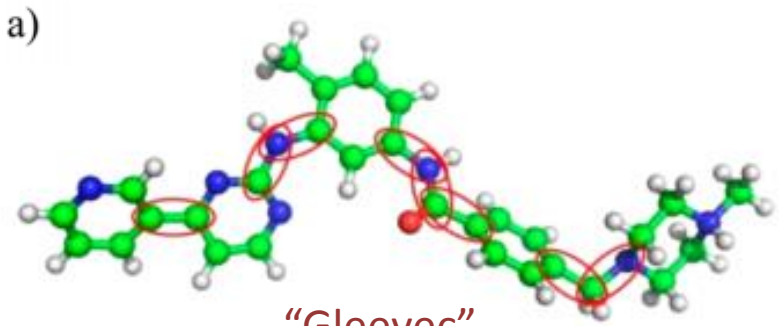
QM conformer energies also can be reproduced well.

c)

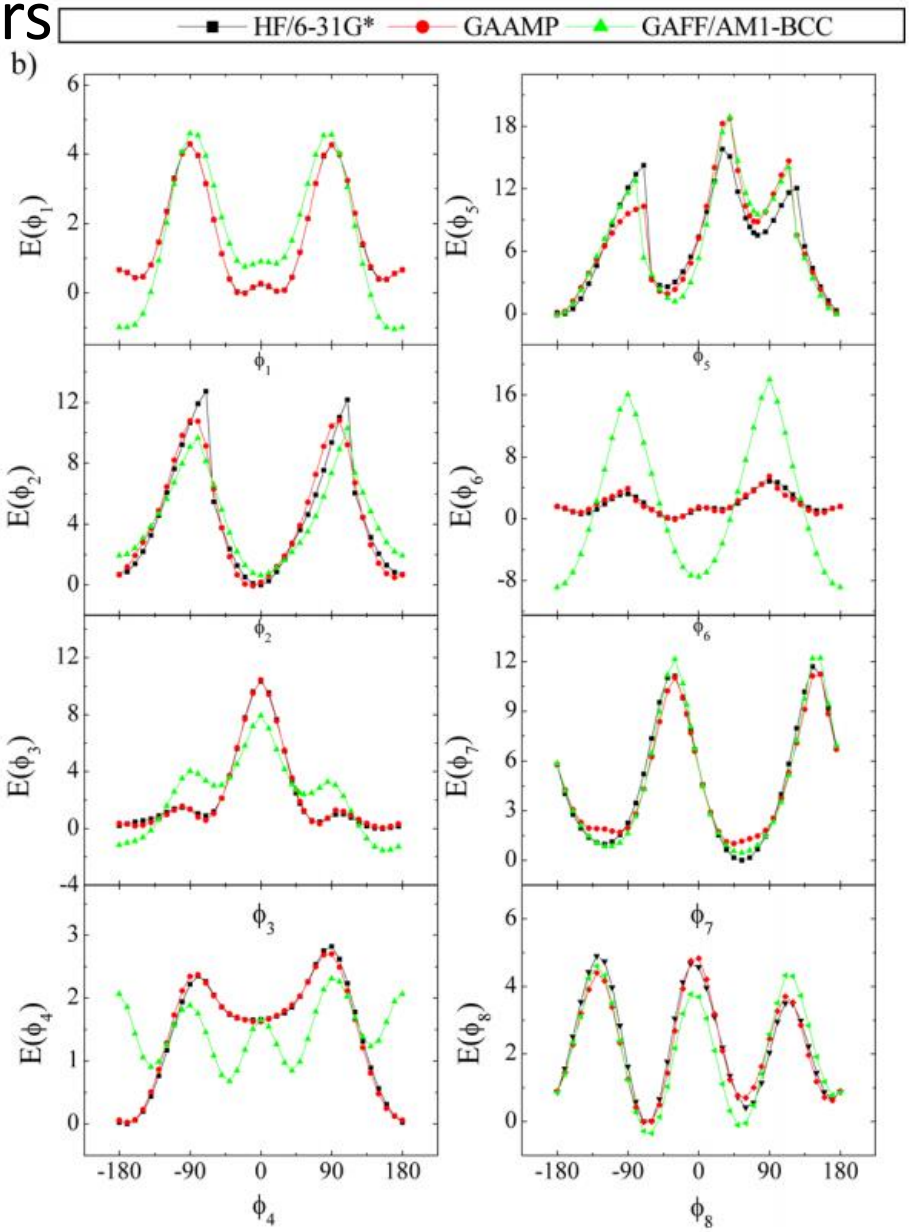


GAAMP: Result and Discussion

Results 1. Dihedral parameters



"Gleevec"



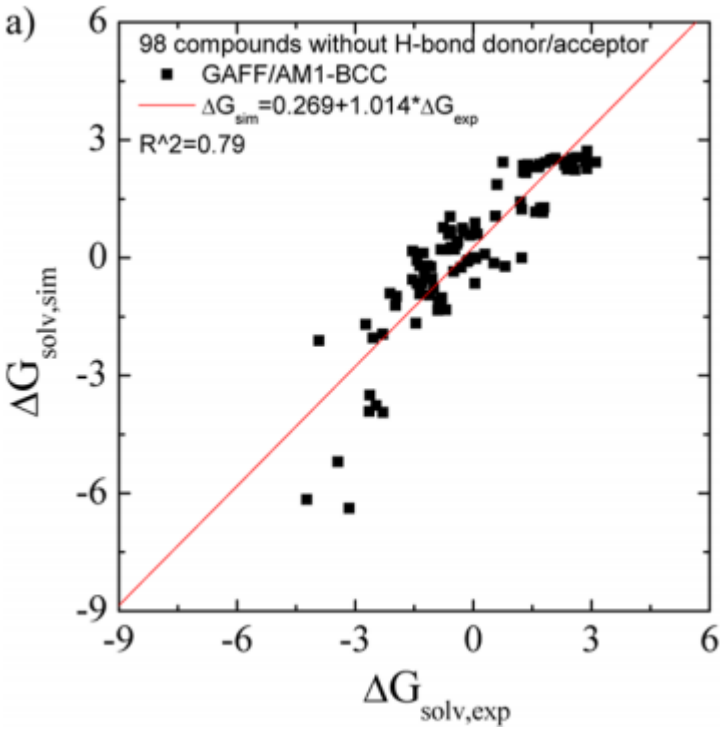
GAFF/AM1-BCC does not perform well For ϕ_4 and ϕ_6;

GAAMP can reproduce QM energy Profiles reasonably well for all dihedrals.

GAAMP: Result and Discussion

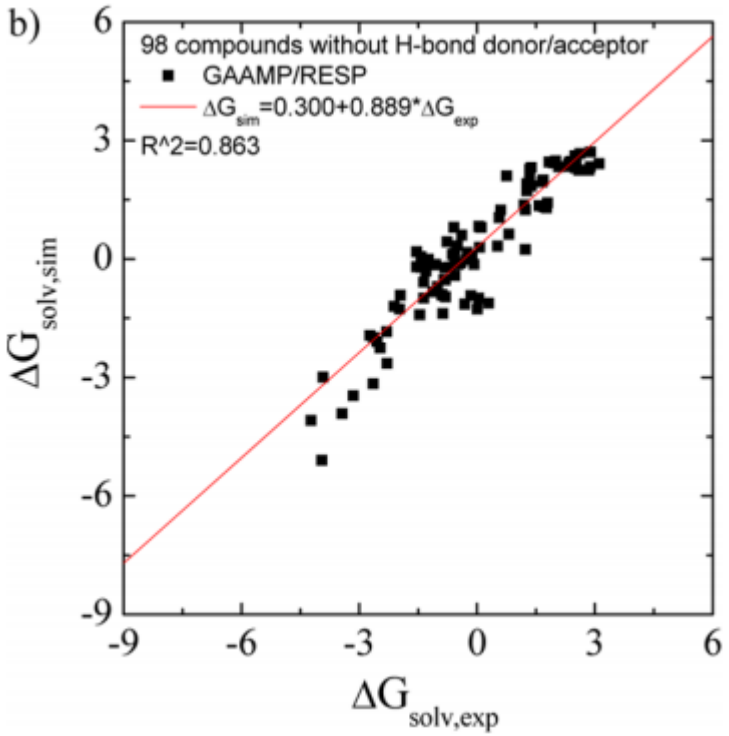
Results 2. Solvation free energies of 217 compounds

98 compounds without hydrogen-bond donor/acceptor



GAFF/AM1-BCC

AUE = 0.74 kcal/mol



GAAMP/RESP

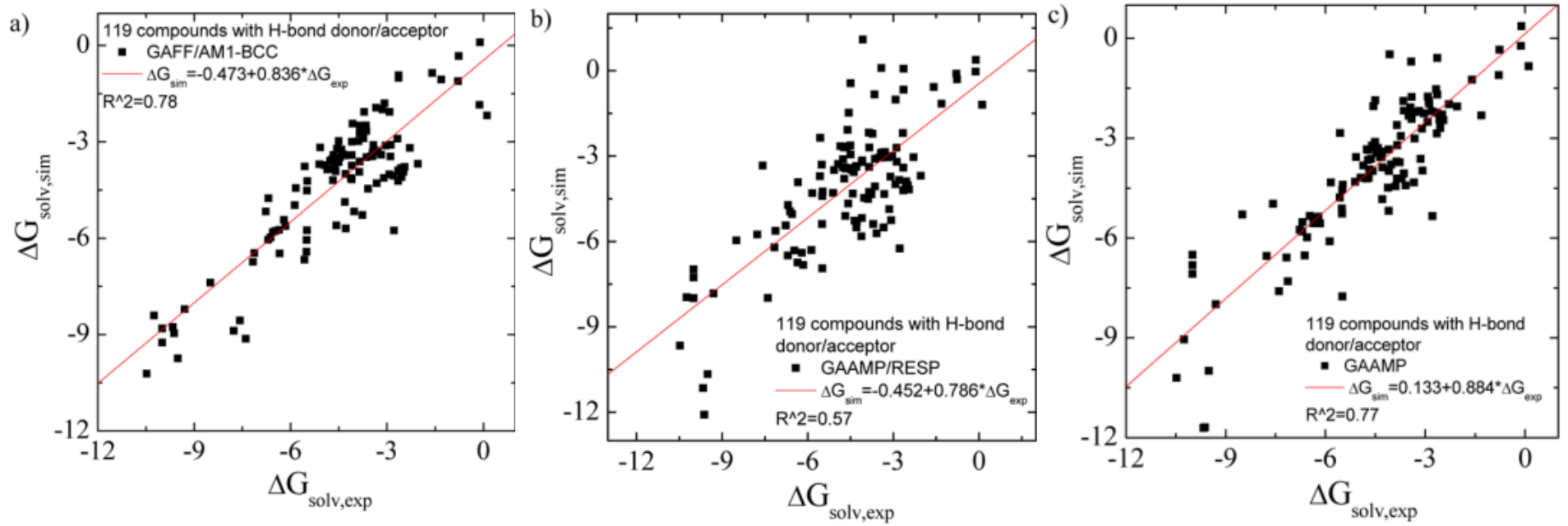
AUE = 0.58 kcal/mol

AUE: average unassigned error, the lower, the better.

GAAMP: Result and Discussion

Results 2. Solvation free energies of 217 compounds

119 compounds with hydrogen-bond donor/acceptor



GAFF/AM1-BCC

AUE = 0.94 kcal/mol

GAAMP/RESP

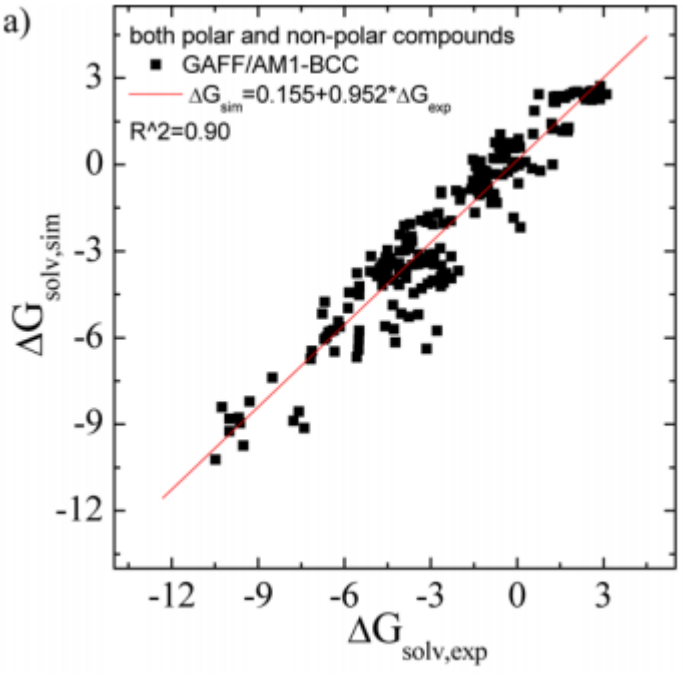
AUE = 1.35 kcal/mol

GAAMP (Fitting both RESP and compound-water interactions)

AUE = 1.00 kcal/mol

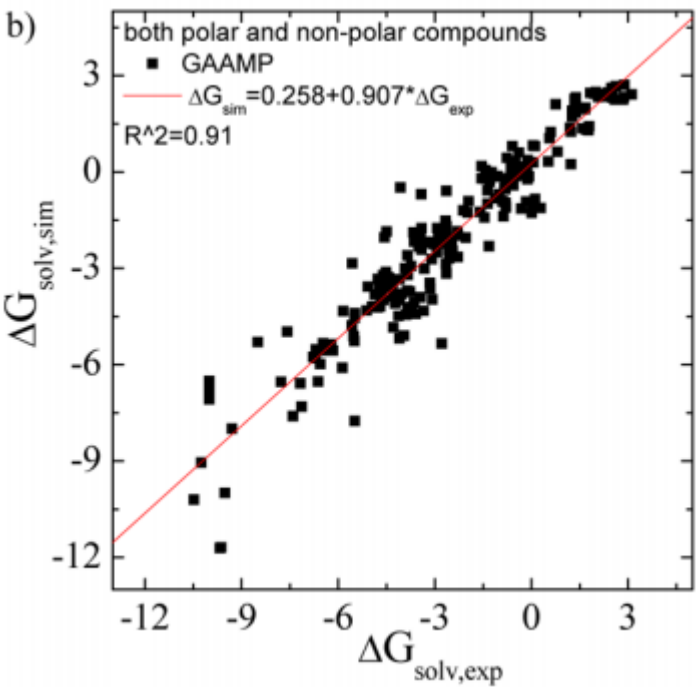
Results 2. Solvation free energies of 217 compounds

217 compounds including both polar and nonpolar molecules



GAFF/AM1-BCC

AUE = 0.85 kcal/mol



GAAMP (Fitting both RESP and compound-water interactions)

AUE = 0.81 kcal/mol

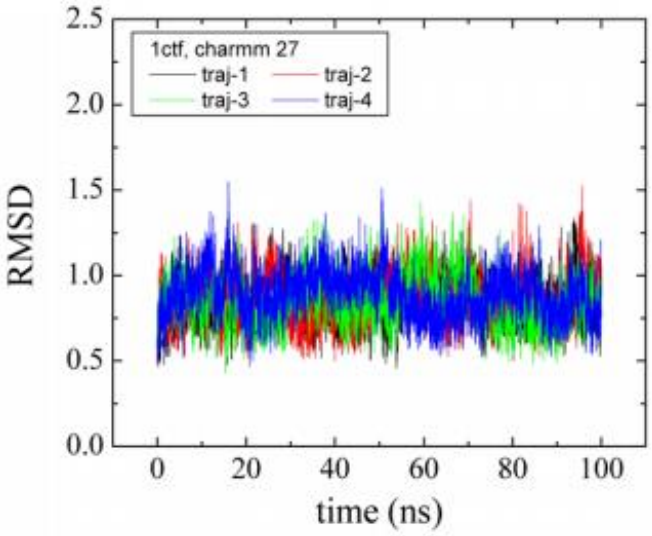
GAAMP: Result and Discussion

Results 3. GAAMP vs CHARMM27 in protein simulation

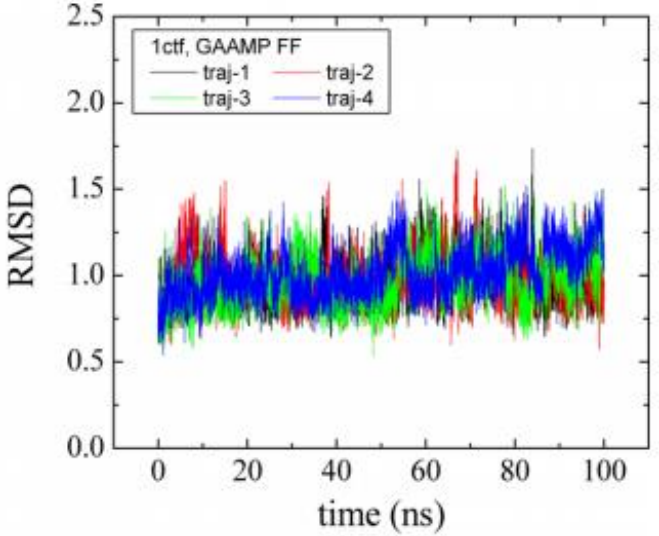
Four independent 100 ns simulation from crystal structure

1ctf

CHARMM27



GAAMP



The protein is stable both in CHARMM27 and GAAMP with conformational fluctuations.

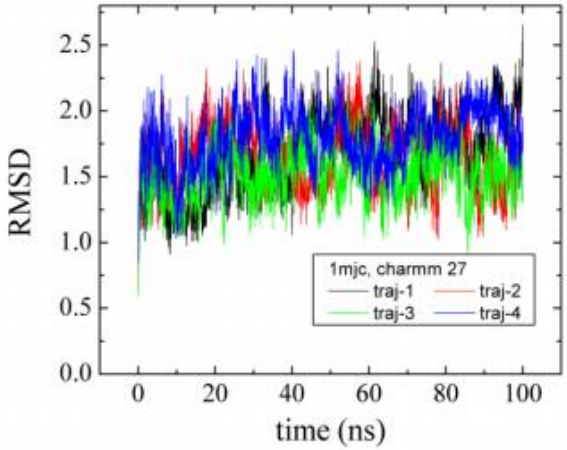
The parameters of amino acids generated by GAAMP are consistent with existing CHARMM27.

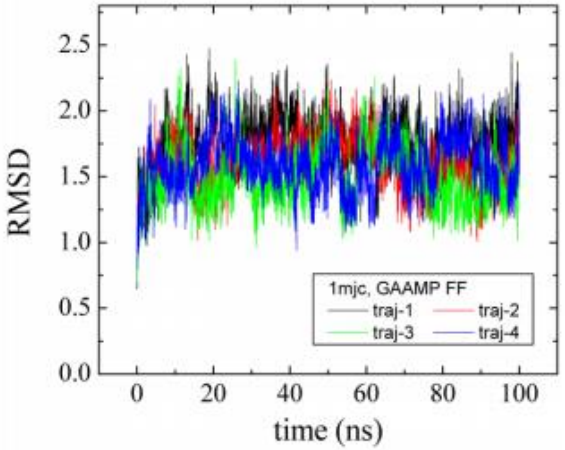
Results 3. GAAMP vs CHARMM27 in protein simulation

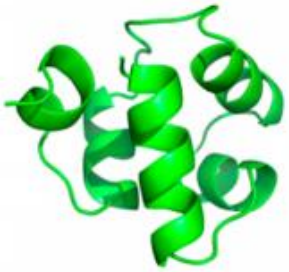
Four independent 100 ns simulation from crystal structure

CHARMM27 GAAMP

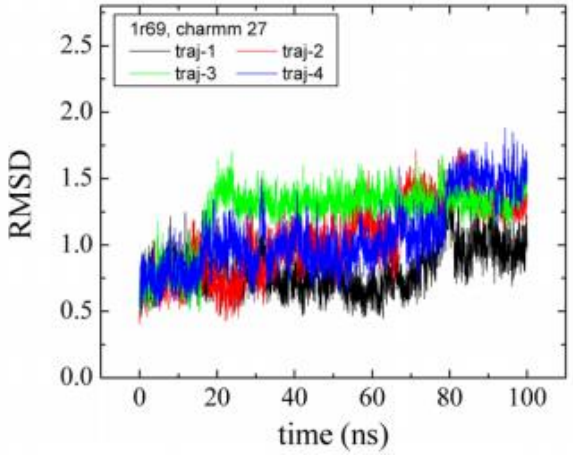
1mjc

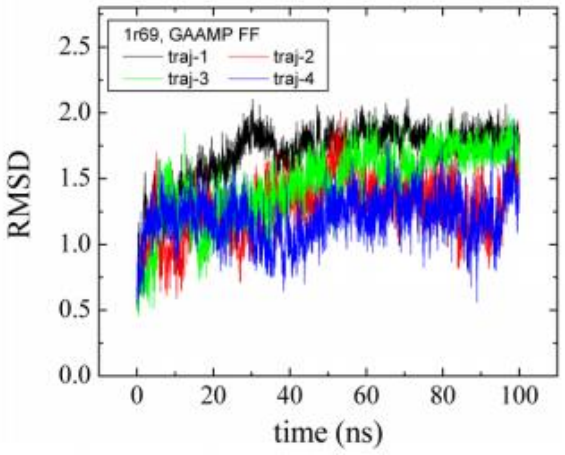






1r69





GAAMP: Limitations and Possible Improvements

- GAAMP targets ab initio calculations, which could be expensive.
- ✓ Cut large molecule into smaller fragments, parameterize them separately, then join them together.

- Geometry optimization is performed in a vacuum.
- ✓ QM geometry optimization with a continuum solvent method.

- For molecules inherently not supported by GAFF or CGenFF, hard to get initial parameters, hard to parameterize.
- ✓ Manually or use other force field development method, such as Q2MM, to prepare a reasonable initial FF

Thank You for Your Listening!