Principles and Algorithms for MultiScale Coarse-Graining (MS-CG)

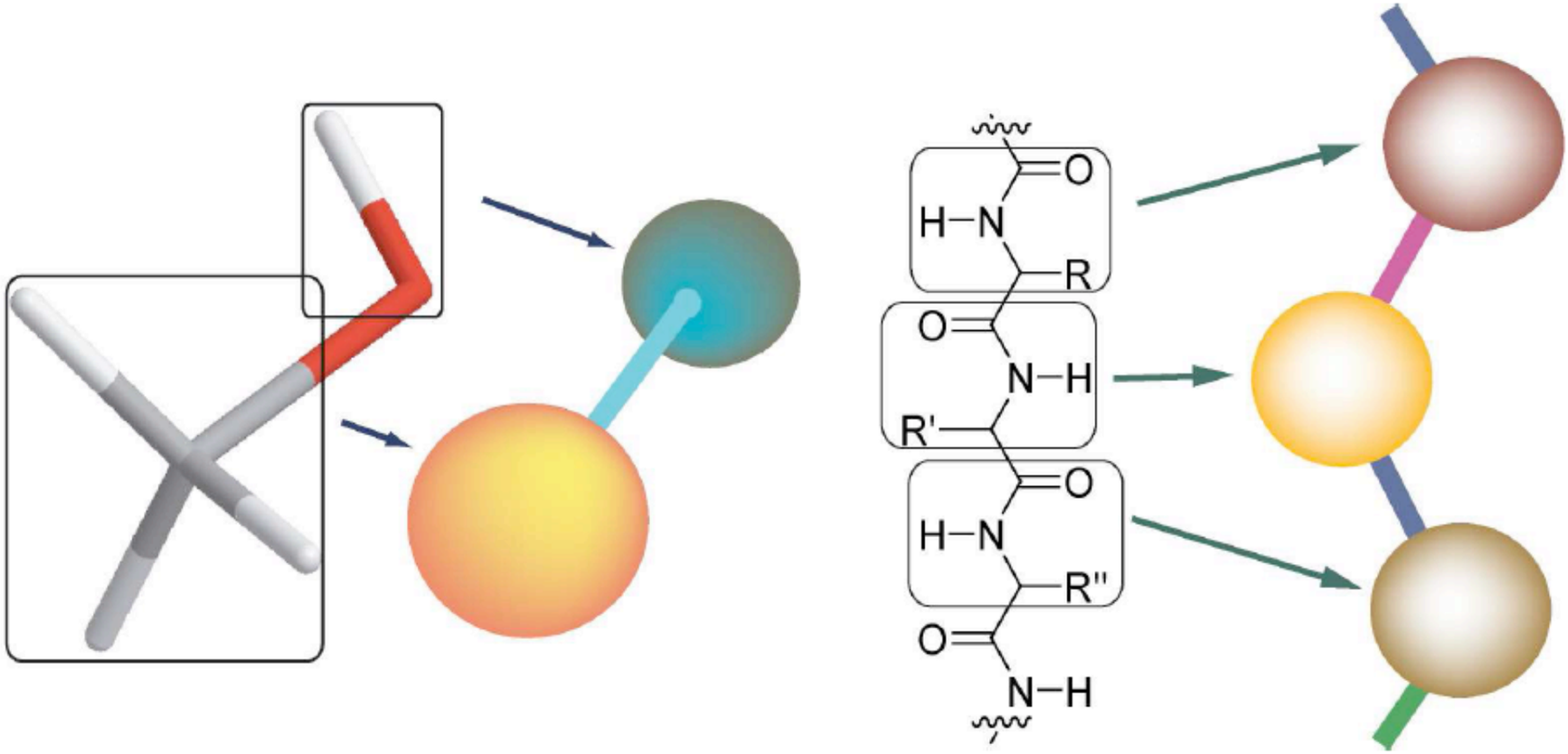
Hans C. Andersen
Department of Chemistry
Stanford University

MS-CG was developed by Greg Voth and his group starting around 2004. He and his group have applied it to a large variety of molecular systems, primarily organic liquids and biomolecular materials. I started collaborating with him a few years later, focussing on the theoretical basis of the method. I tried to understand why it worked as well as it did and why it failed when it failed. I also wanted to develop algorithms that get around the limitations of the method so that it could be applied in a wider variety of situations.

MS-CG method

Izvekov and Voth, JCP (2005)

- For each molecule in the system of interest, choose a set of CG sites designed to give a lower resolution model of the molecule.


O
H–N
R
O
N–H
R'
O
H–N
R''
O
N–H

MS-CG variational principle

- The potential of mean force of the CG sites (for fixed N,V,T) satisfies an exact variational principle [Noid *et al.*, JCP (2008)]
- The variational principle can be expressed in terms of static correlation functions for the forces on the atoms that define each site.
- Everything needed to construct the variational principle can be obtained straightforwardly from simulations of the atomistic system with all its degrees of freedom. The simulations should sample a canonical distribution.

THEOREM [Noid *et al.*, (2008)]

- Suppose the CG potential obtained this way is used as the potential energy function in an MD simulation in which
  - the sites are regarded as mass points
  - the site positions are the only degrees of freedom.
- Static equilibrium distributions and correlation functions of the site positions of the CG system calculated using this CG potential will be ***exactly*** the same as for the original atomistic system with all degrees of freedom ...

... IF

- the atomistic simulations give the necessary equilibrium structural correlation functions ***with no sampling error***
- the basis set for the variational calculation is ***complete***.

There is no question that the MS-CG method is in principle correct. The real questions are whether we have a complete enough set of basis functions, enough simulation data to have a small enough sampling error, and algorithms that minimize the effect of remaining sampling error on the final answer.

There are three important types of sampling error in the atomistic calculations. We have various strategies for dealing with each of them, and it is important to have such strategies in place.

Three types of sampling error

x

- "Shot noise" - atomistic simulations give information about the forces acting on the sites but only for the specific intersite distances observed in the atomistic configurations.
- *must avoid use of basis functions that fit the shot noise*

x

- Failure to explore all of the **thermally accessible** parts of CG configuration space.
- *must make sure the atomistic simulations used to construct the variational function are extensive enough to allow the molecules to explore all thermally accessible conformations.*

- We have no numerical force information for the **thermally inaccessible** parts of CG configuration space.
- *must make sure that the trial function used in the variational calculation is large and positive in the thermally inaccessible parts of CG configuration space. In particular, this means that the thermally inaccessible parts of configuration space, if any, must be identified.*

Choice of basis functions

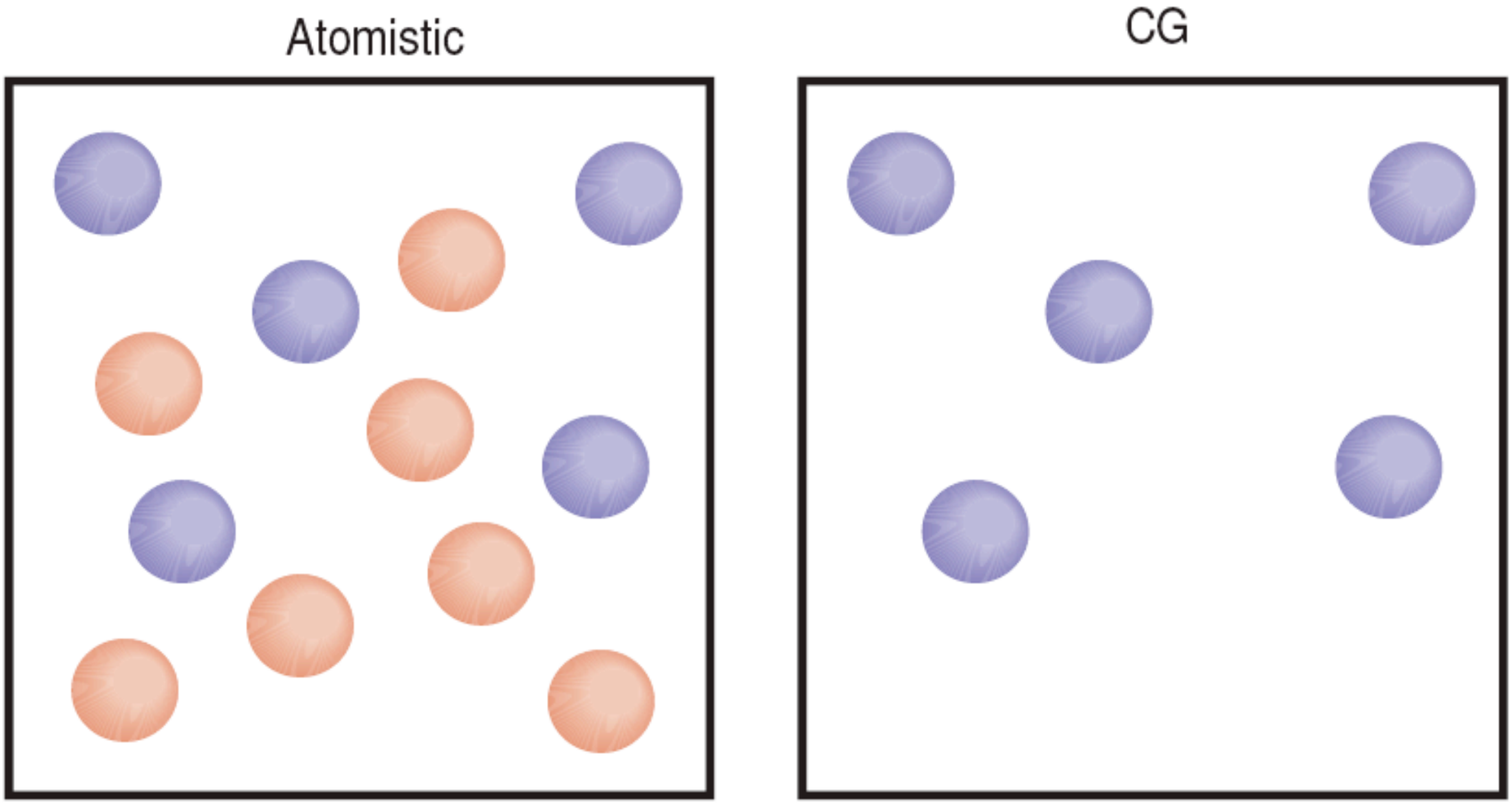
- Choose the types of interactions to be included in the CG potential.
- The typical choice is
  - pairwise additive nonbonded potentials
  - bond stretching potentials,
  - bond angle bending potentials
  - dihedral angle potential
- We will also discuss use of three body potentials.

- Construct basis functions for each type.
- Express the CG potential as a linear combination of basis functions.
- The coefficients are varied to minimize the variational function.
- The coefficients at the minimum give the CG potential.

STUDIES OF SIMPLE MODEL SYSTEMS

first model system

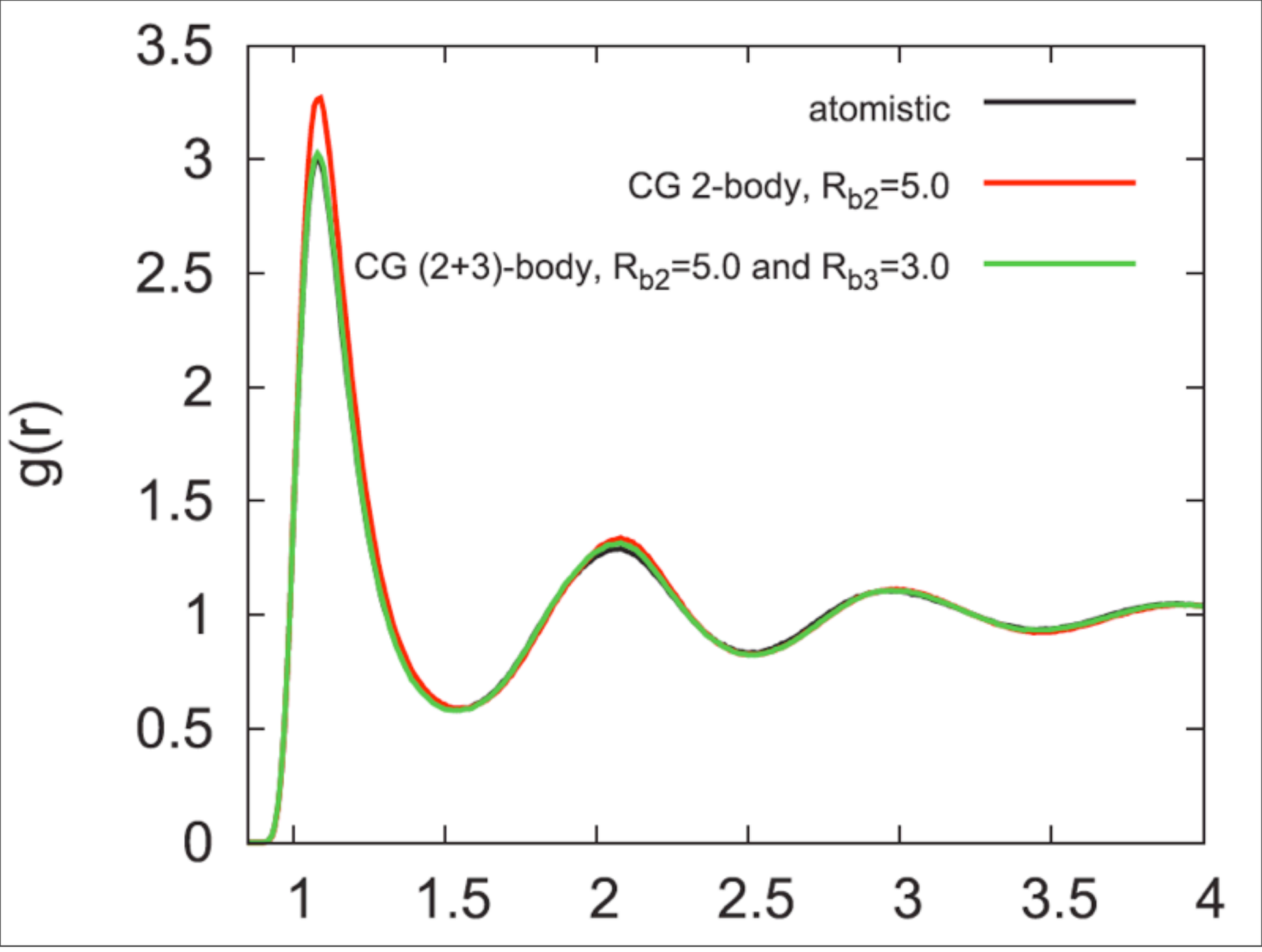
a two component Lennard-Jones mixture



- Atomistic system: binary solution of atomic particles (blue solutes and red solvents)

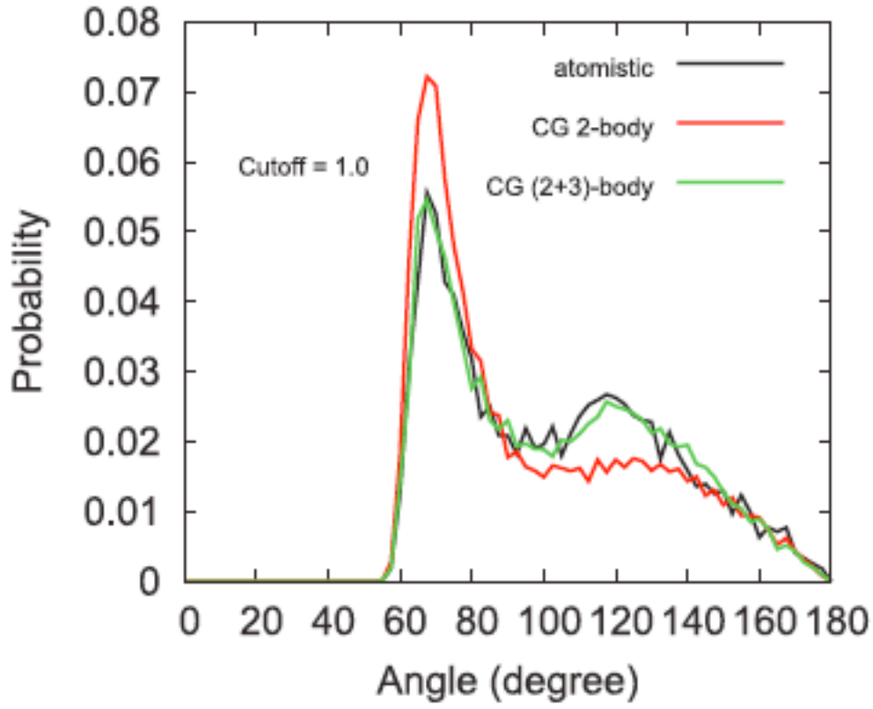
- CG system: one component system containing only solutes (a solvent-free model of the solution)

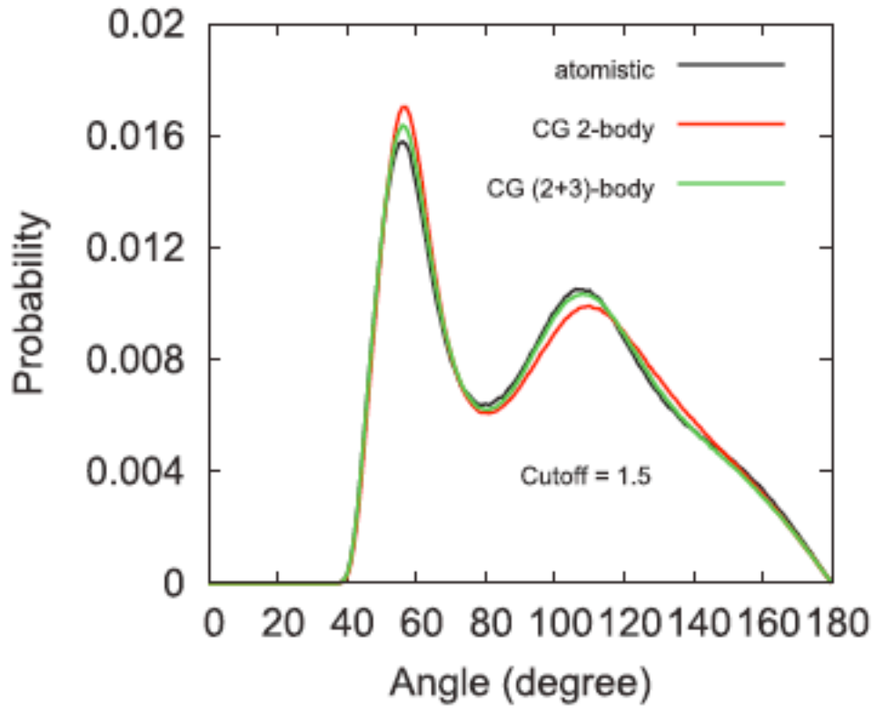
- Use basis functions that describe **two-body interactions** between CG sites.
- Then use basis functions that describe **two-body and three-body** interactions between CG sites.

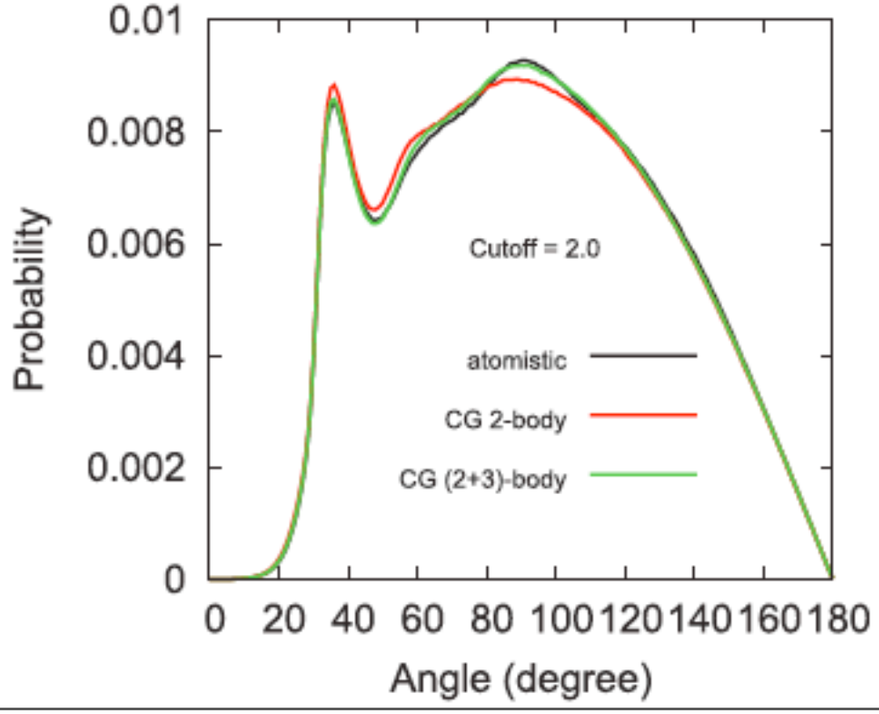


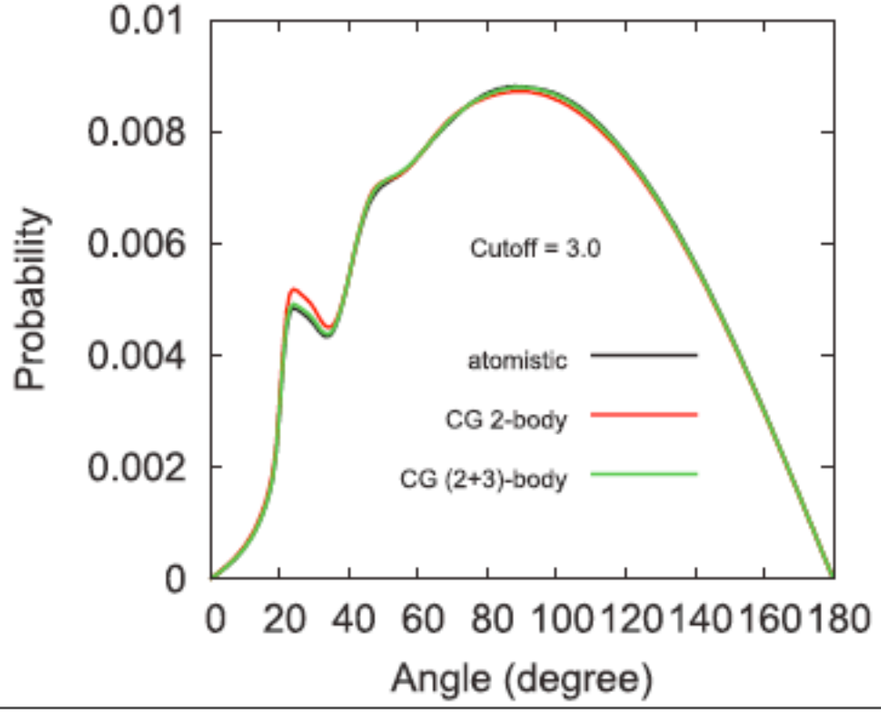
The two body result is better than typical two body MS-CG results, probably because for this model the sites are actually spherical and because we used a very good basis set that extended out to large distances. Moreover, we had very small statistical error in our atomistic data.

More typically, results are reasonable but not this good, due primarily to:
nonsphericity of the set of atoms associated with a site,
restriction of basis functions to those that can represent only a short ranged CG potential.


Probability
Angle (degree)
Cutoff = 1.0
atomistic
CG 2-body
CG (2+3)-body

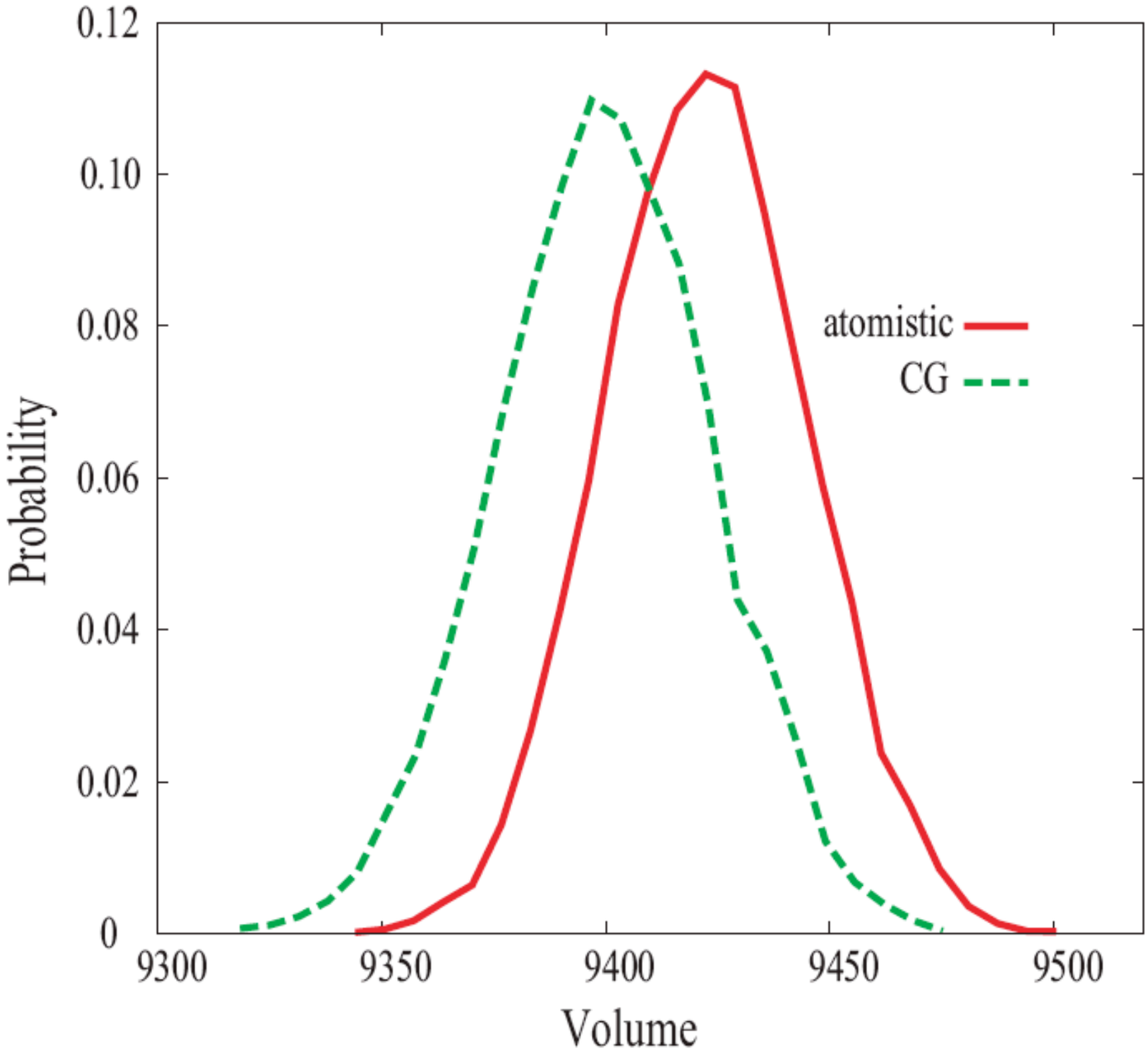

Probability
Angle (degree)
Cutoff = 1.5
atomistic
CG 2-body
CG (2+3)-body


Probability
Angle (degree)
Cutoff = 2.0
atomistic
CG 2-body
CG (2+3)-body


Probability
Angle (degree)
Cutoff = 3.0
atomistic
CG 2-body
CG (2+3)-body

constant pressure MS-CG

- The MS-CG method has been extended to the isothermal-isobaric ensemble, allowing MS-CG simulations at constant pressure.
- Volume fluctuations in the CG system are very similar to those of the atomistic system.



This shows that MS-CG calculations can be done for the isothermal-isobaric ensemble. The CG intermolecular interactions are the same as would be obtained for the canonical ensemble at the same volume. There is an additional volume dependent term in the CG potential that is required to describe the volume fluctuations correctly.

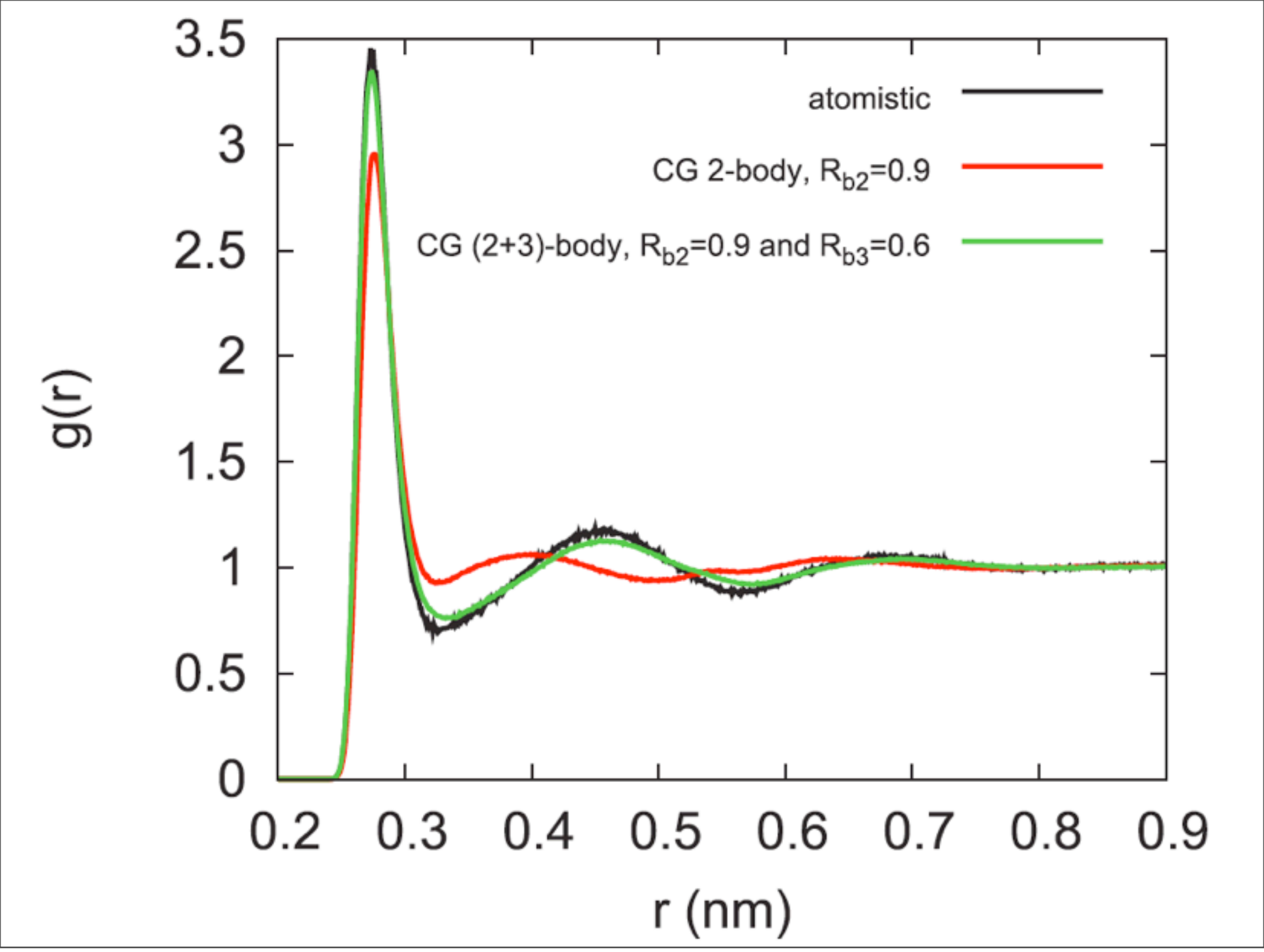
These calculations were performed using a CG potential that included only two body interactions and the volume dependent potential.

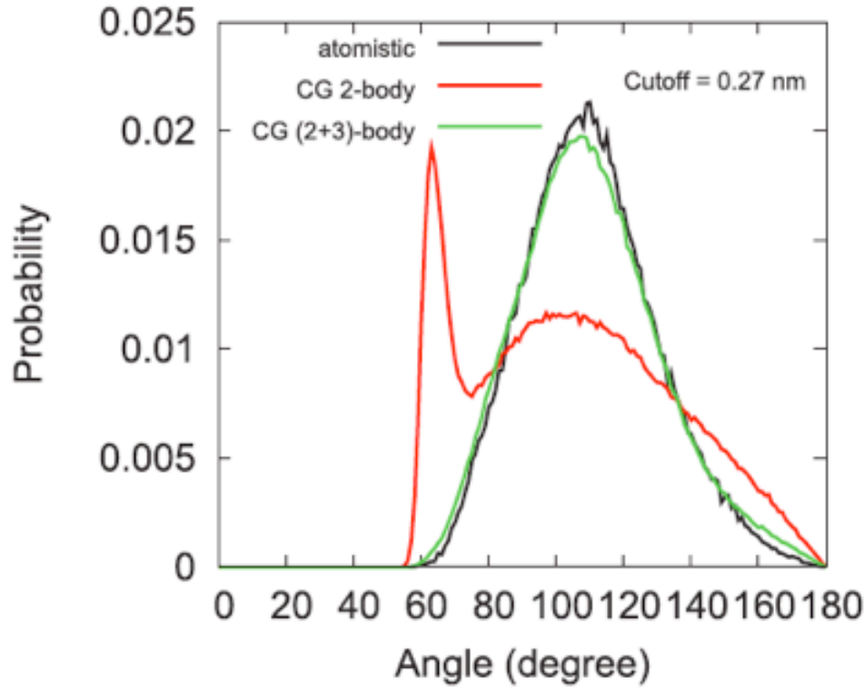
second model system

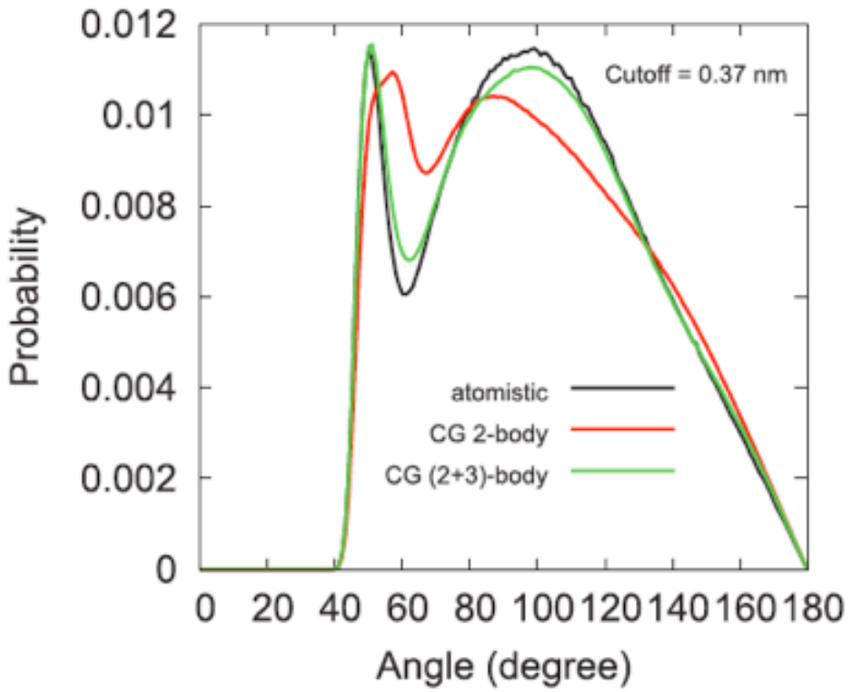
liquid water

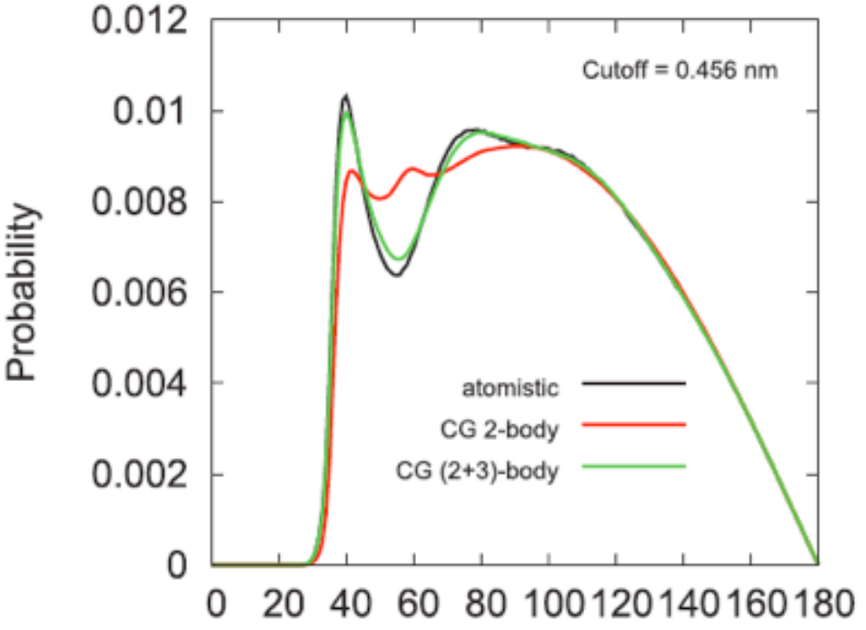
- The atomistic system - SPC/E water *with electrostatic interactions and Ewald summation*
- The CG system
  - a one site model of the water molecule
  - the site is placed at the center of mass of the molecule
  - the CG model is a mass point *with short ranged interactions*

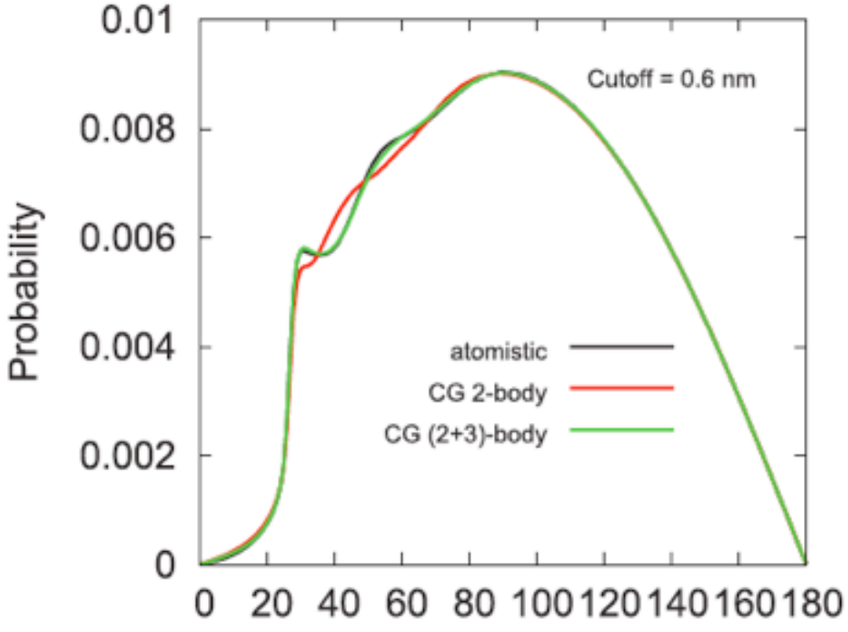
This CG model clearly can't describe the long ranged electrostatic effects of the interactions involving water. It is not at all clear that this is a useful model, but it might be in some circumstances.


3.5
3
2.5
2
1.5
1
0.5
0
g(r)
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
r (nm)
atomistic
CG 2-body, R_{b2}=0.9
CG (2+3)-body, R_{b2}=0.9 and R_{b3}=0.6


atomistic
CG 2-body
CG (2+3)-body
Cutoff = 0.27 nm
Probability
Angle (degree)
0.025
0.02
0.015
0.01
0.005
0
0 20 40 60 80 100 120 140 160 180


Cutoff = 0.37 nm
atomistic
CG 2-body
CG (2+3)-body
Probability
Angle (degree)
0.012
0.01
0.008
0.006
0.004
0.002
0
0 20 40 60 80 100 120 140 160 180


Cutoff = 0.456 nm
atomistic
CG 2-body
CG (2+3)-body
Probability
0.012
0.01
0.008
0.006
0.004
0.002
0
0 20 40 60 80 100 120 140 160 180


Cutoff = 0.6 nm
atomistic
CG 2-body
CG (2+3)-body
Probability
0.01
0.008
0.006
0.004
0.002
0
0 20 40 60 80 100 120 140 160 180

third model system

hexane in the gas phase and liquid phase

J. Chem. Theory Comput. **2009,** *5,* 3211–3223

JCTC Journal of Chemical Theory and Computation

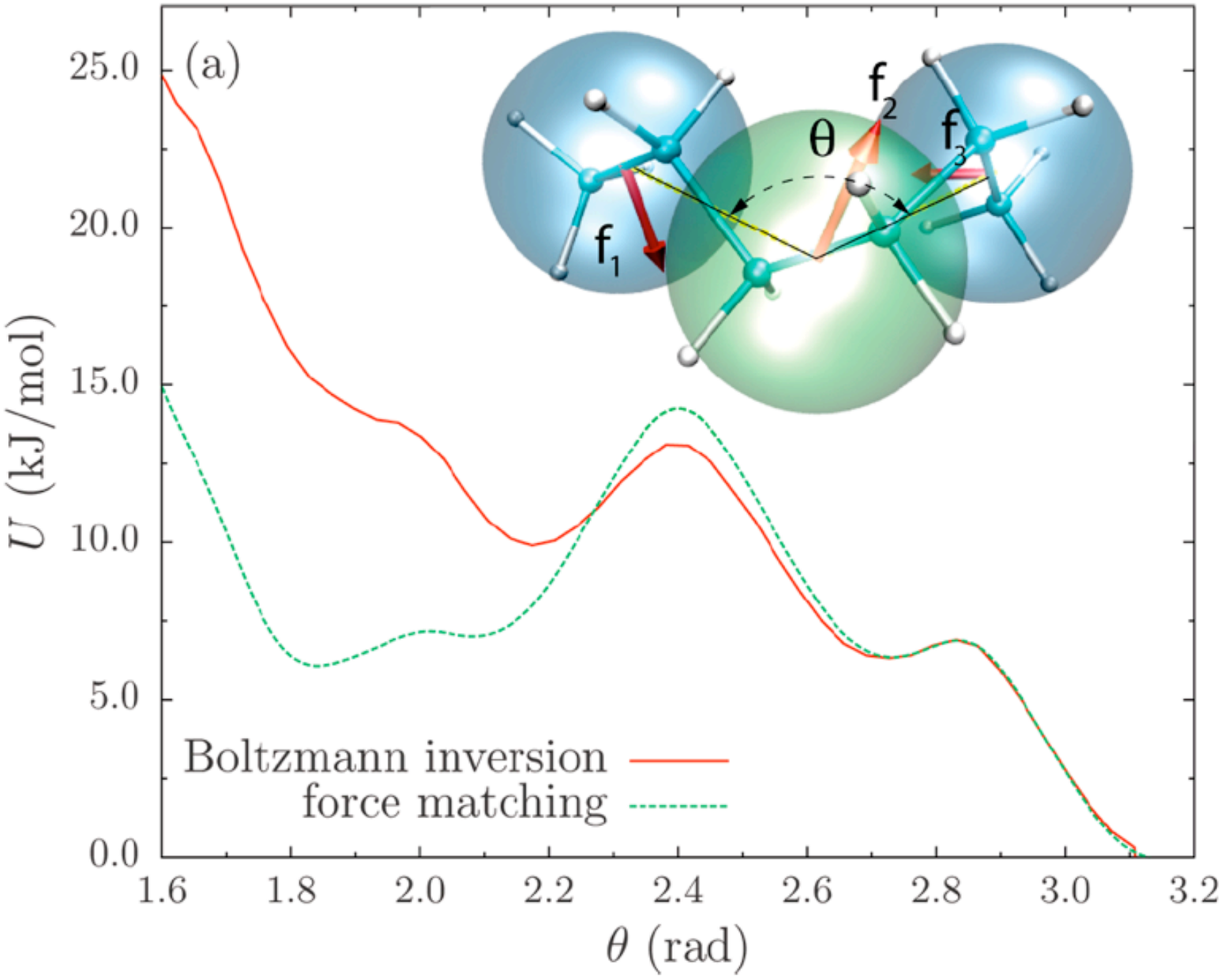
Versatile Object-Oriented Toolkit for Coarse-Graining Applications

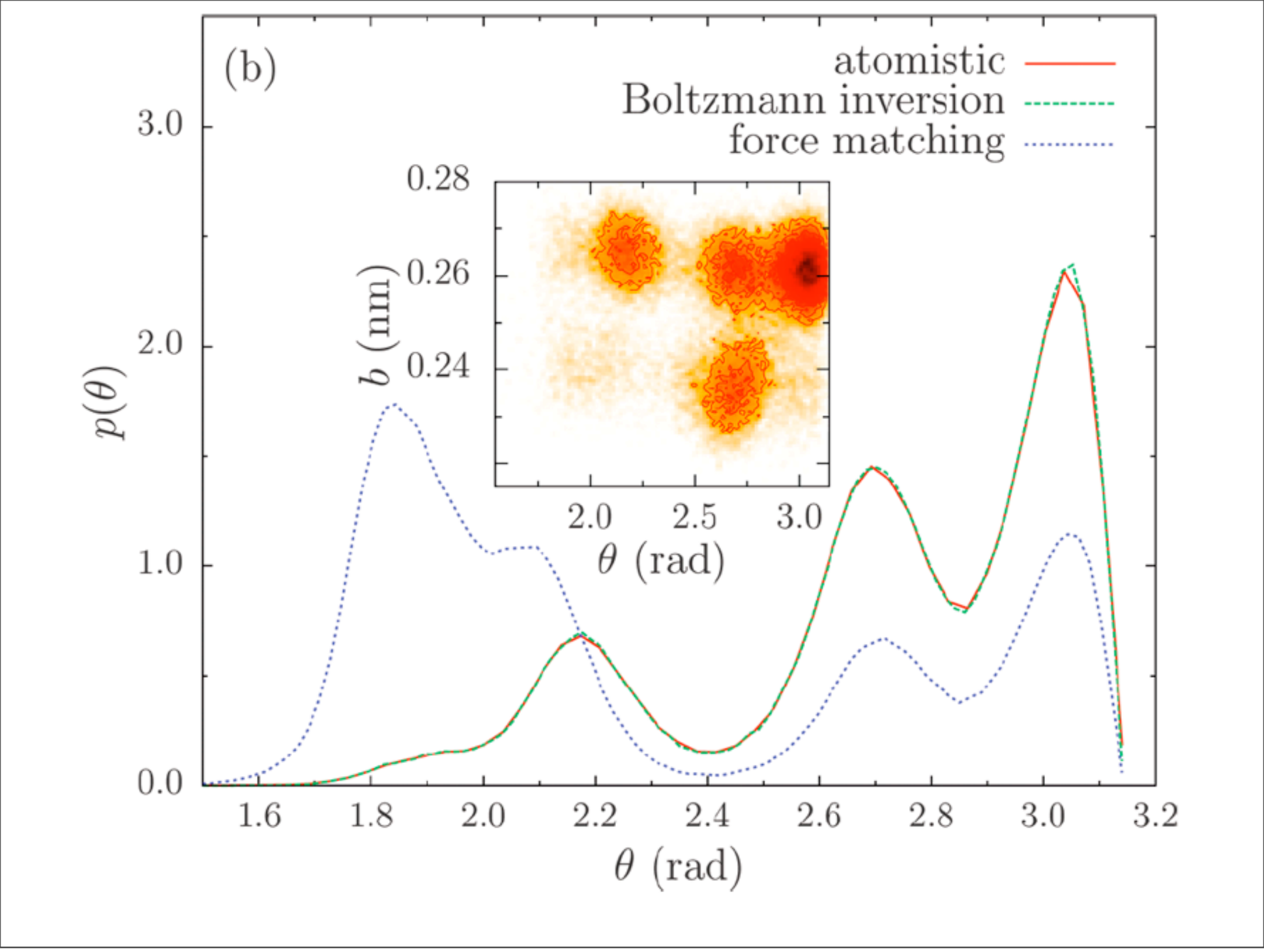
Victor Rühle, Christoph Junghans, Alexander Lukyanov, Kurt Kremer, and Denis Andrienko*

Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany

Received July 17, 2009

Abstract: Coarse-graining is a systematic way of reducing the number of degrees of freedom representing a system of interest. Several coarse-graining techniques have so far been developed, such as iterative Boltzmann inversion, force-matching, and inverse Monte Carlo. However, there is no unified framework that implements these methods and that allows their direct comparison. We present a versatile object-oriented toolkit for coarse-graining applications (VOTCA) that implements these techniques and that provides a flexible modular platform for the further development of coarse-graining techniques. All methods are illustrated and compared by coarse-graining the SPC/E water model, liquid methanol, liquid propane, and a single molecule of hexane.


(a)
25.0
20.0
15.0
10.0
5.0
0.0
U (kJ/mol)
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0
3.2
θ (rad)
Boltzmann inversion
force matching
f1
f2
f3
θ



The probability density in CG space is quite complicated, consisting of four regions. This is a reflection of the existence of dihedral angle rotations and of gauche and trans configurations of the molecule. This makes only certain combinations of bond angles and bond lengths possible. A simple Boltzmann factor for an angle potential and two Boltzmann factors for the two bond vibration potentials can not possible describe this distribution.

Recent advances in MS-CG algorithms

- Multiresolution basis functions - for representing CG potential functions (similar to wavelets)
- Elastic net method - for solving the numerical matrix problem associated with the variational calculation in MS-CG (provides automatic selection of basis functions from an extremely large set)
- Use of these two to calculate three-body CG potentials
- Method to deal with the inaccessible regions of configuration space

A. Das *et al.*, J. Chem. Phys. (2012) (3 papers)

It is these methods that have enabled the calculations I have presented today. I am hopeful that they will significantly expand the types of problems that can be successfully addressed using MS-CG.

Acknowledgments

- Avisek Das (constant pressure CG method, hierarchical basis functions, elastic net method, two-body and three-body CG potentials for Lennard-Jonesium and water, some of the hexane work)
- Greg Voth and the many members of the Voth group, especially Will Noid and Lanyuan Lu (some of the hexane work)
- NSF - CRC program