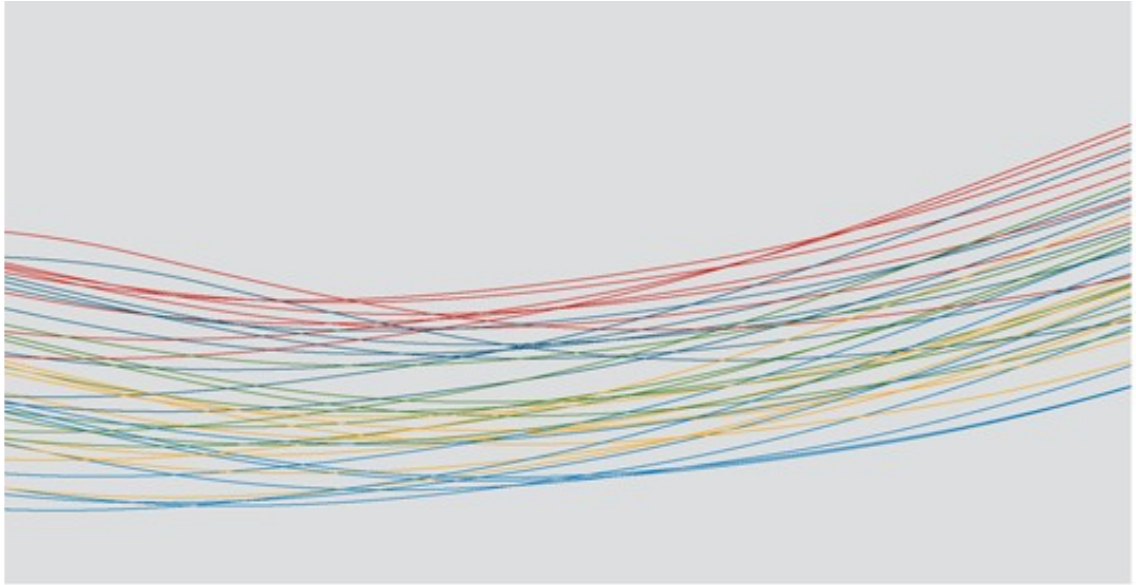
Fraunhofer SCAI

Connecting Quantum Mechanics to Coarse-Grained Simulations: the Role of Force Field Development



Karl N. Kirschner, Ph.D.
Computational **Ch**emistry
Engineering Group (CoChE)

Synopsis of SCAI and the “CoChE” Group

Fraunhofer SCAI

Part of “Institutszentrum Birlinghoven” - One of the largest computer science research centers in Germany

SCAI: ~90 Employees + 40 PhD Students; highly interdisciplinary

2008 Annual budget : 8.6 Mio. Euro

40% of the SCAI budget from industry

CoChE founded in 2007: 7 senior staff (PhDs), interdisciplinary 2 PhD-Students and 3-5 Undergraduates

Universität zu Köln – Lehrstuhl Angewandte Mathematik / Wissenschaftliches Rechnen

Bonn Aachen International Center for Information Technology
Life Science Informatik (Prof. Hofmann-Apitius)

DLR-SISTEC – Simulations- und Softwaretechnik
(Leitung: Prof. U. Trottenberg, R. Hempel)

Verteilte Systeme und Komponentensoftware
Software QS und eingebettete Systeme

Talk Overview

Fraunhofer SCAI

Force Fields: an old topic, but remains important

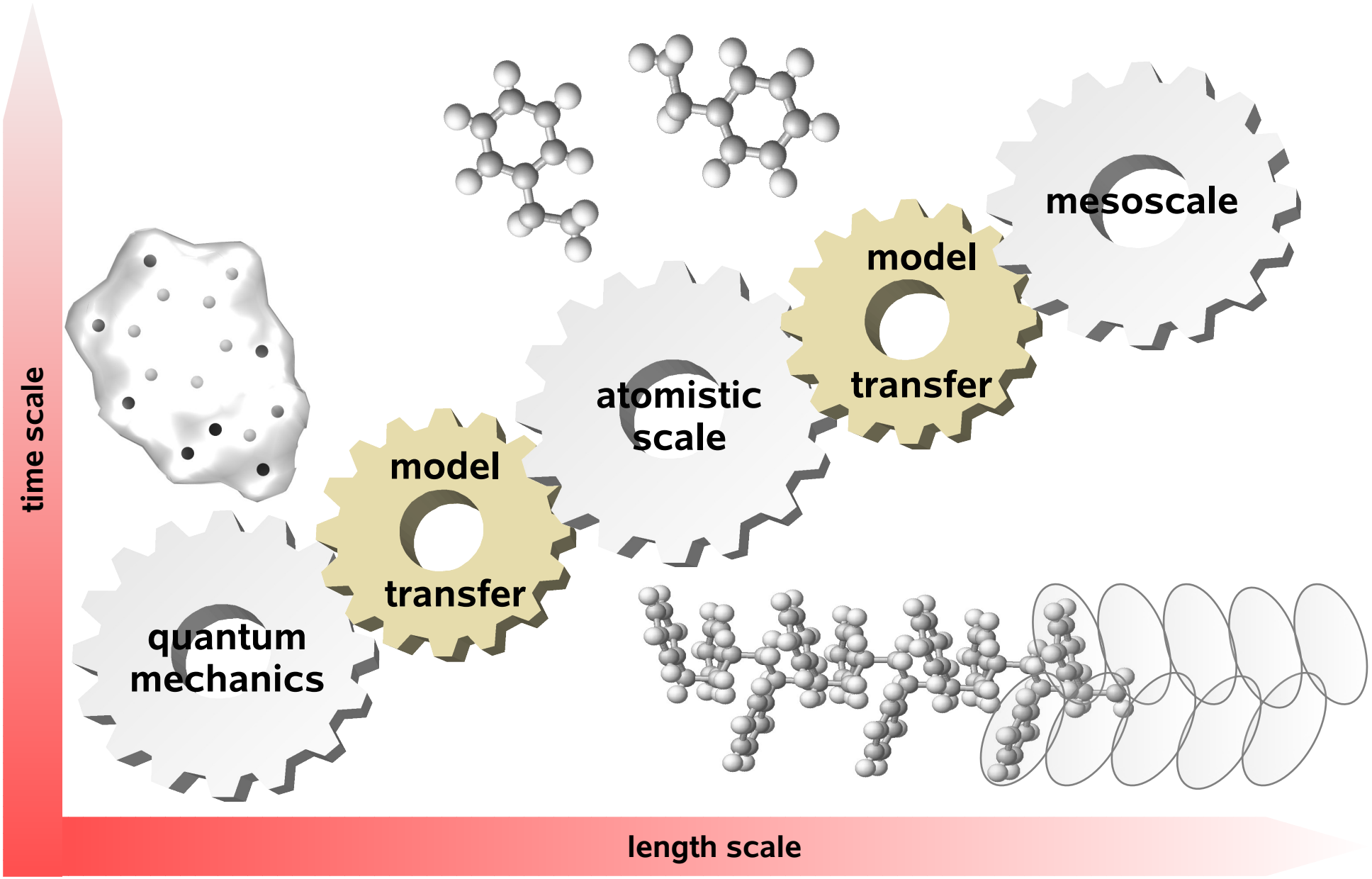
Different routes / methods / philosophies exists for their development --- “The art of parameterization”

Challenge: From our experiences and the data in the literature, can we begin to formulate more standard routes towards developing and optimizing a force field in a timely and reliable fashion?

Why are we thinking about this? Industry applications & contribute to the scientific community.

Soft matter - Multi-scaling modeling

Fraunhofer SCAI



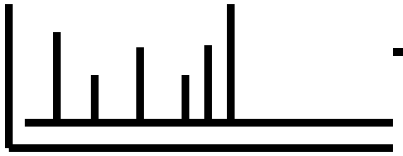
Fraunhofer SCAI

What is the Current State?

Basic atomistic force field equations appear to be established

$$V = \sum_{\text{bonds}} K_r (r - r_{eq})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}\left[1 + \cos(n\phi - \delta_n)\right] + \sum_{i<j}\left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\epsilon r_{ij}}\right]$$

Types:



→ **Type I**

$H\Psi = E\Psi$ → **Type II**

FF + MD → **Type III**

Molecular Classes:

a) Functional/Small molecule (transferable) → Hydrocarbon, Alcohol, Esters...

b) Specialized (nontransferable) → Solvent, Molecule specific

c) Biological (something in between??) → Protein, DNA, Carbo.

Force Fields. What is the Current State? Cont.

Fraunhofer SCAI

Current research towards how to including additional quantum effects (eg polarization)

Different routes exist towards developing new parameters

- Fully manual method
- Cross terms
- L.-J. Parameters
- QM – which levels
- Nearly full auto. method (target exp value)
- Truncation of Fourier Series
- Partial atomic charges (QM or empirical)
- Keep parameters close to chemical intuition vs completely adjustable
- Weighting specific molecular conformations
- How is experimental data used? a) training b) validation
- What exp observables are used -geom, energy, ΔH°_{vap}
- Generic or Explicit torsion terms (eg X-C_{sp3}-C_{sp3}-X vs C_{sp3}-C_{sp3}-C_{sp3}-C_{sp3})

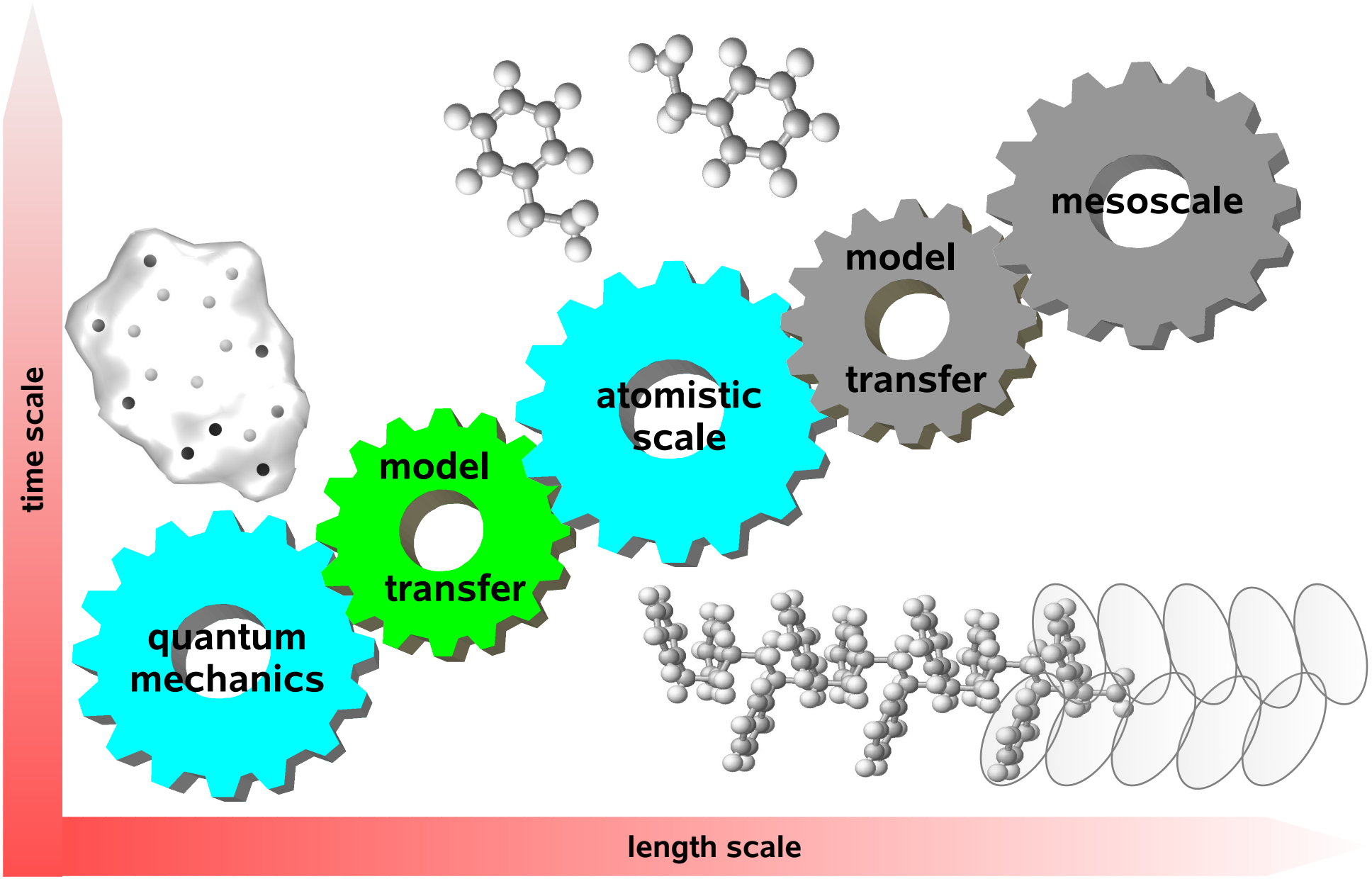
Different programs use different force fields (eg MM3, Parm99SB, GROMOS, and ports)

Internal default settings effect the force field performance (eg scaling factors)

Challenge: Develop a time-efficient and systematic route for new reliable parameters.

Soft matter - Multi-scaling modelling

Fraunhofer SCAI



MM3, Glycam06 + Literature

Fraunhofer SCAI

Current and future Goals via the lessons learned so far (atomistic)

- Have a *Development* and *Validation* suite of molecules
- Isolated chemical families for parameterization

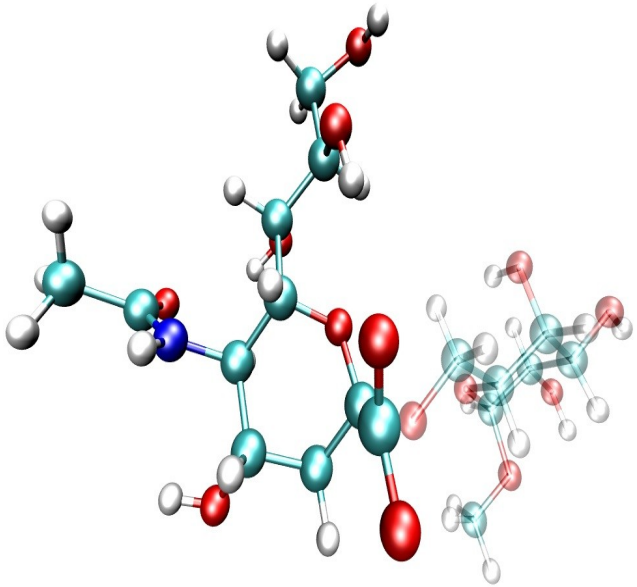
- Use a basic form of the FF Equation, *initially* (eg no cross terms)
- Keep the number of atom types as low as possible
- Parameters should be transferable & program independent

- Remove 1-4 nonbonded & electrostatic scaling factors
- Use QM to develop the force field parameters (at least for version 1)
  HF/6-31G(d)//HF/6-31G(d) → *Geometries*
  MP2/ or B3LYP/6-31++G(2d,2p)//HF/6-31G(d) → *Energies*
  B3LYP/ or HF/aug-cc-pVTZ//HF/6-31G(d) → *Electrostatics*

- Partial Atomic Charge Sets → Ensemble averaged or Boltzmann weighted (via RESP)
- Explicit parameters (C_{sp3}-C_{sp3}-C_{sp3}-C_{sp3} vs ~~X-C_{sp3}-C_{sp3}-X~~)

- Final test of parameters is the comparison between theoretical and experimental results

Provide a firm foundation for a) future parameter refinement, b) adding e^- lone pairs, and c) adding polarizability

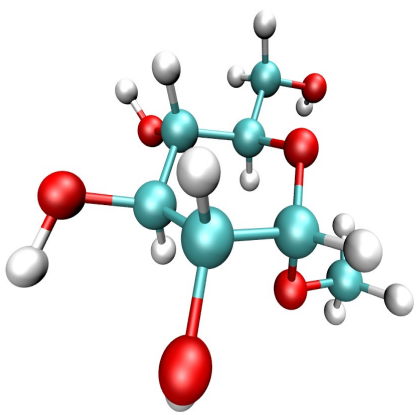


Chemical Functional Groups:

Hydrocarbons	Ether Amides
Alcohols	Alcohol Amides
Ethers	Ether Alcohols
Amides	Ether Carboxylates
Carboxylates	

Develop the parameters in a systematic step-wise fashion → allows the physics to be isolated to each parameter

1-4 Scaling Factors (AMBER)



Why scaling factors:
it was justified "given the known overestimation of short-range repulsion by a 6-12 form of potential."†

Why it is a problem:
1-5 interaction ≡ a 1-4 interaction (O•••O equal dist)
captured by torsions parameters

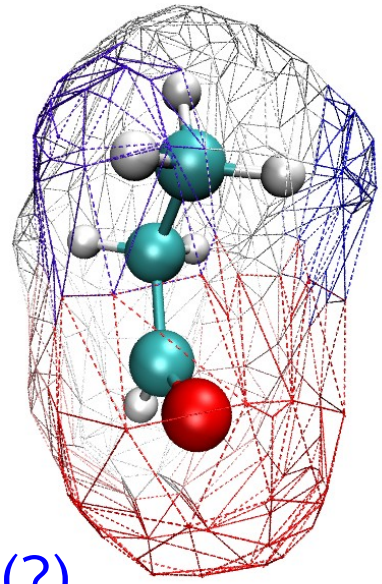
† Cornell et al. *J. Am. Chem. Soc.*. 115 (1993) 9631 & 115 (1995) 5179.
‡ Kirschner et al. *J. Comput. Chem.* 29 (2008) 662.

Partial Atomic Charges (P.A.C.)

Fraunhofer SCAI

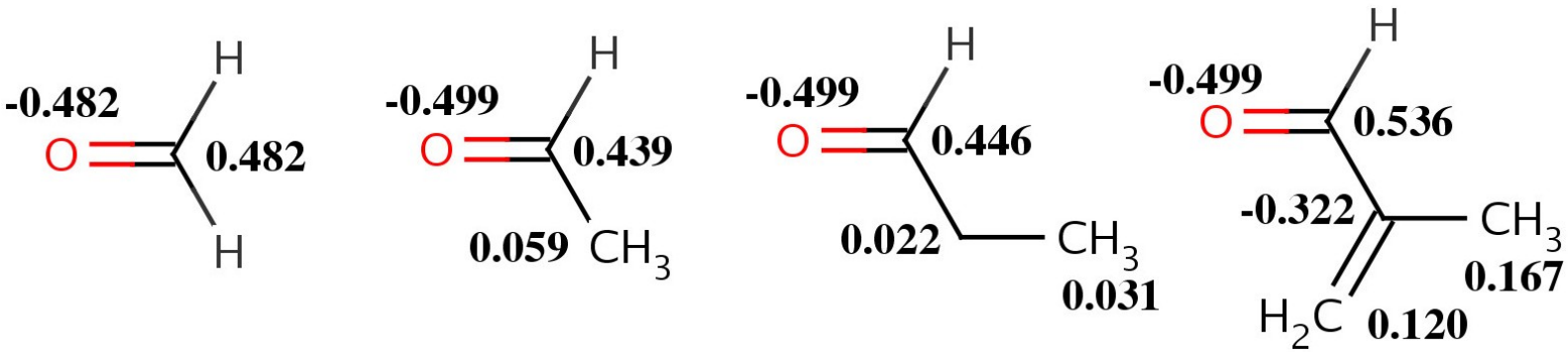
RESP Model†

- Input: QM ESP (CHELPG & large basis set)
- Uses a hyperbolic restraint (varied by a weighting factor)
  → purpose is to keep the P.A.C lower magnitude



- Helps to correct the high dependency of conformation in P.A.C. due to buried atoms

- Consistent P.A.C. in molecular *fragments*
  → allows for better generation of torsion parameters & L-J (?)



- Boltzmann Weighting vs Ensemble Averaging
  - Boltzmann Weighting: *gas-phase* QM relative energies
  - Ensemble Averaging: 1st approximationg to capturing *solution-phase* effects & conformations

† Kollman et al J Phys Chem 97 (1993) 10269.

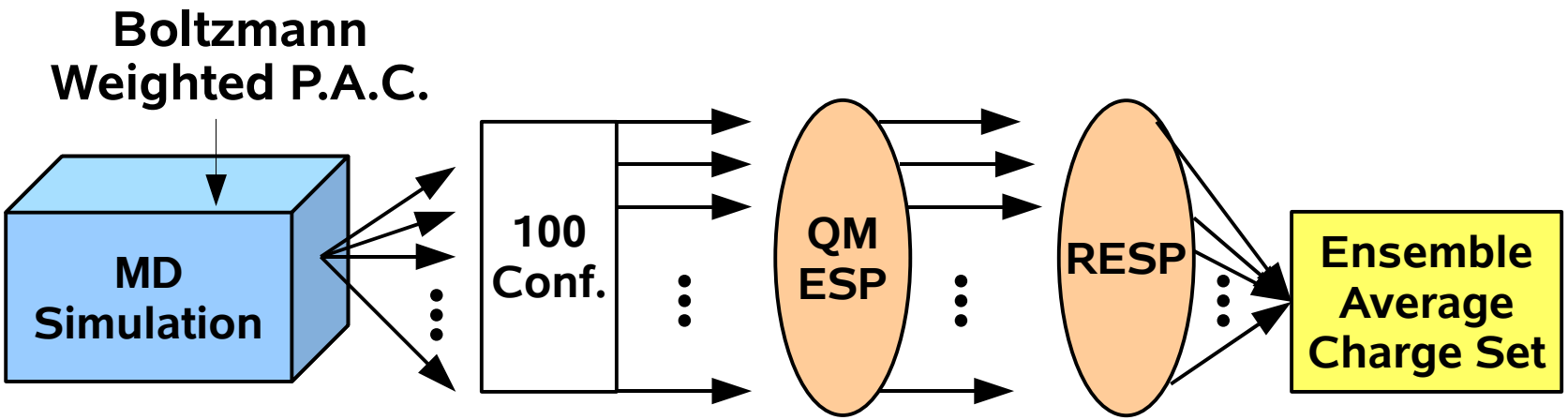
Ensemble Charge Set

Fraunhofer SCAI

Do not need every atom to obtain a good RESP fitting to the ESP Charge are more chemically intuitive†

For carbohydrates, removal of aliphatic hydrogens result in optimal number for fitting‡ ... this holds true for other compounds as well

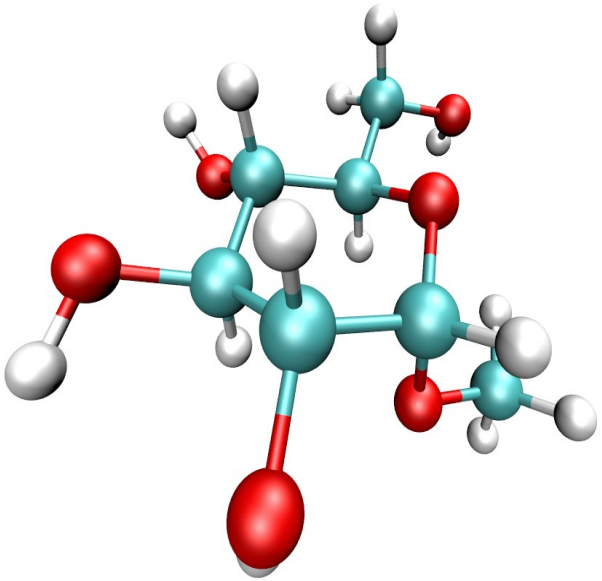
Methodology for final charge set (eg 100+ carbohydrate residues)

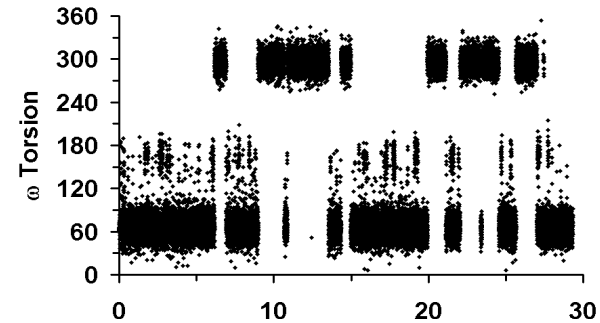


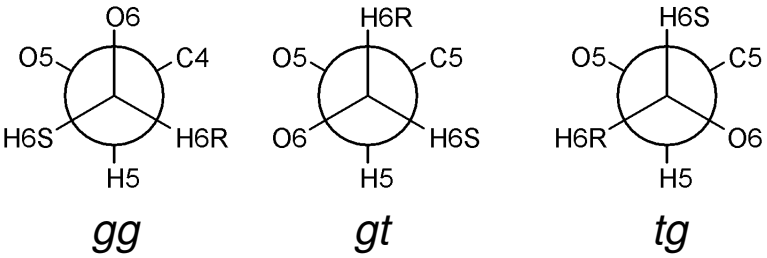
Take home point → a charge set that is less conformationally biased and reflexs the solution phase

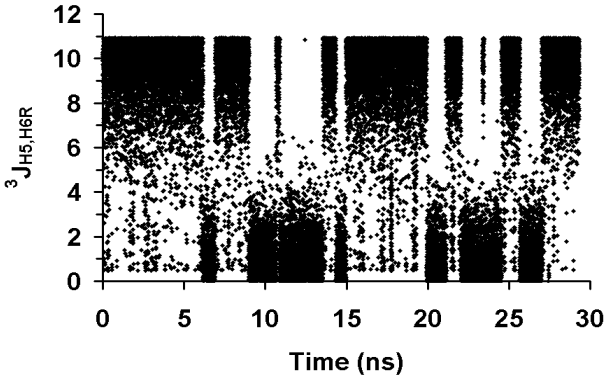
† Francl et al. *J. Comput. Chem.* 17 (1996) 367 & *Glycoconjugate J.* 14 (1997) 501.
‡ Basma et al. *J. Comput. Chem.* 27 (2001) 1125.
‡ Kirschner et al. *J. Comput. Chem.* 29 (2008) 662.

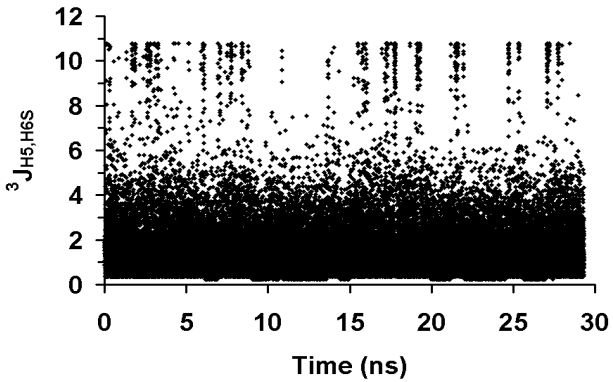
Glycam06



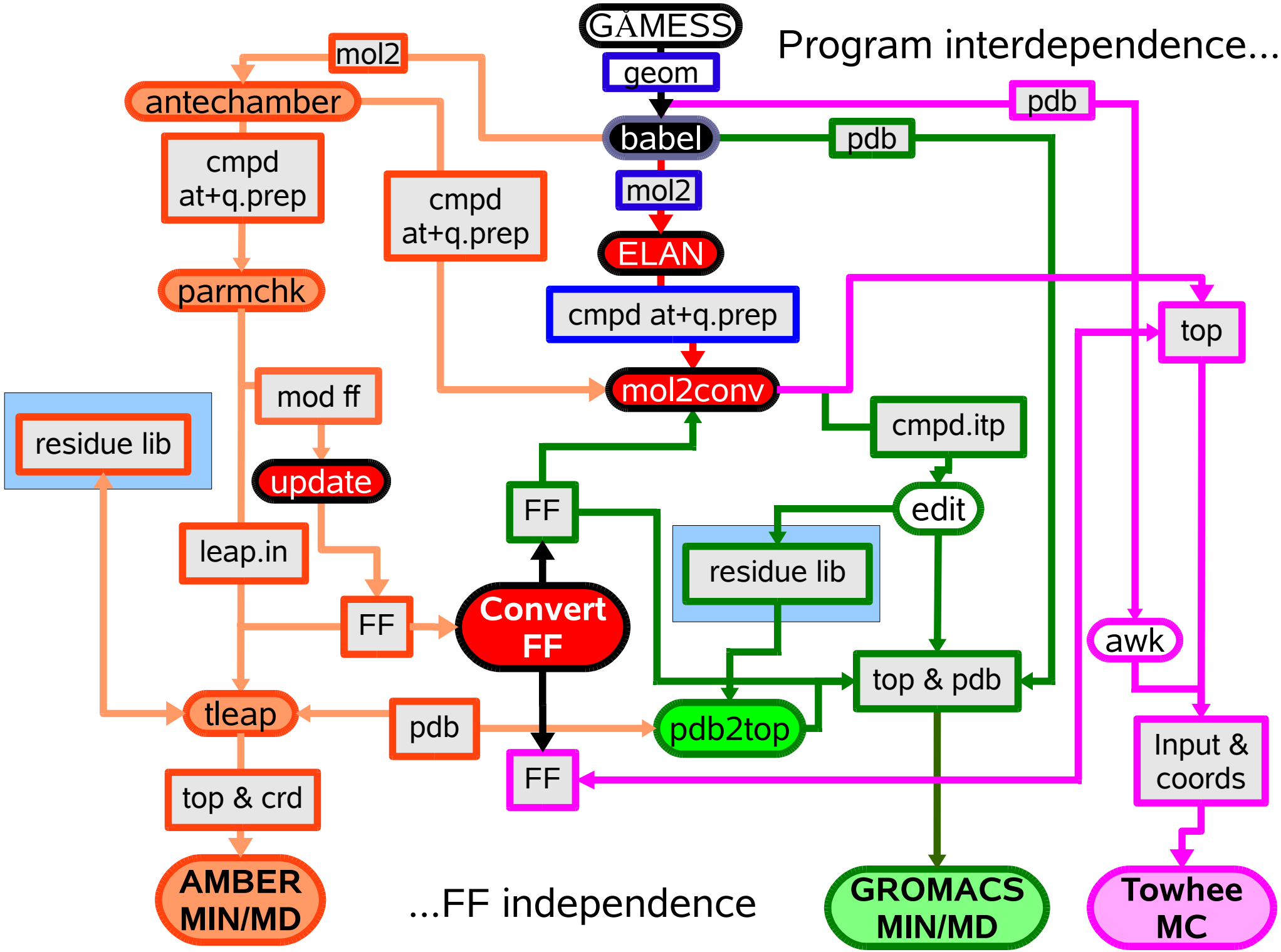








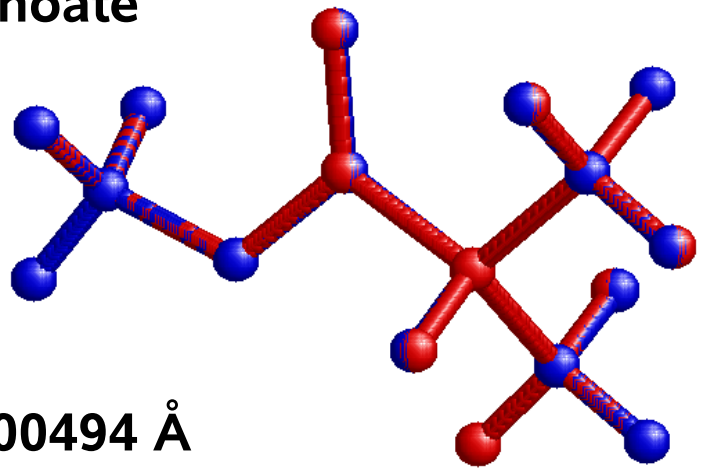
α-D-Glc*p*OMe	$^3J_{H5,H6R}$	$^3J_{H5,H6S}$
NMR (avg)	5.4	2.3
Glycam06	5.4±1.7	2.9±2.0
gt	9.5±1.4	1.8±1.2
tg	2.3±1.6	9.0±1.9
gg	1.0±1.0	1.6±1.1


Program interdependence...
GAMESS
geom
mol2
antechamber
babel
pdb
pdb
cmpd
at+q.prep
cmpd
at+q.prep
mol2
ELAN
parmchk
cmpd at+q.prep
top
mol2conv
mod ff
residue lib
cmpd.itp
update
FF
edit
leap.in
residue lib
FF
Convert
FF
awk
top & pdb
tleap
pdb
pdb2top
Input &
coords
FF
top & crd
AMBER
MIN/MD
...FF independence
GROMACS
MIN/MD
Towhee
MC

Amber & Gromacs

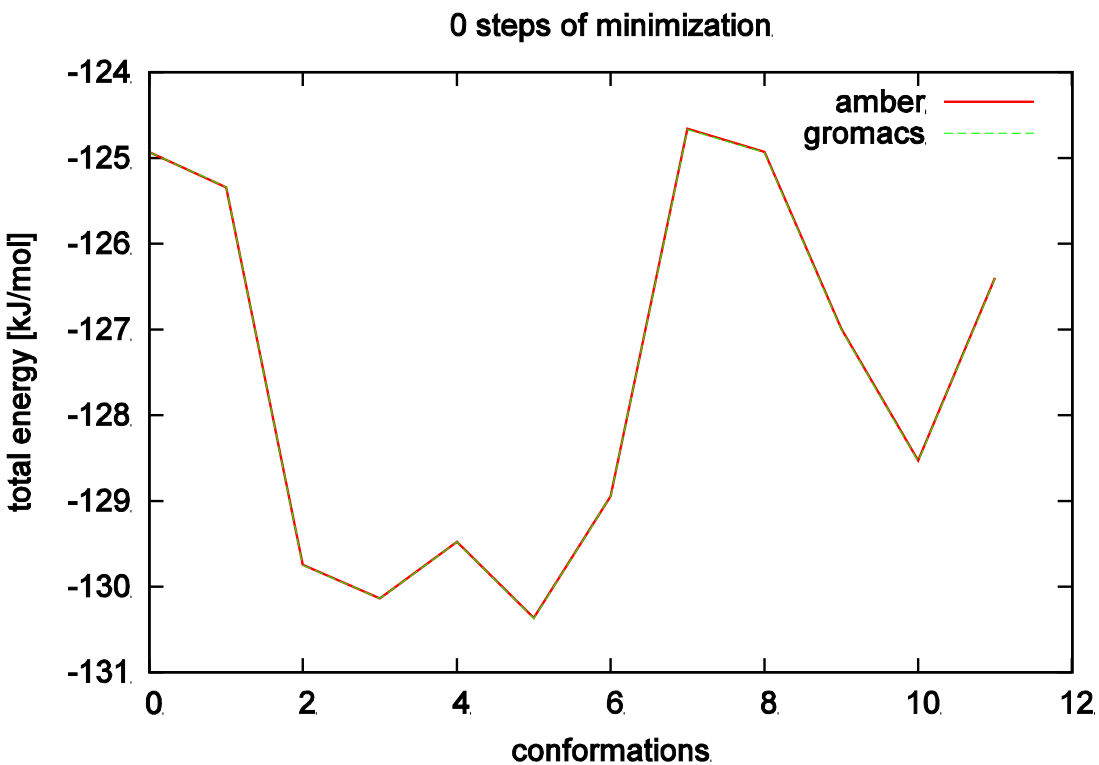
Fraunhofer SCAI

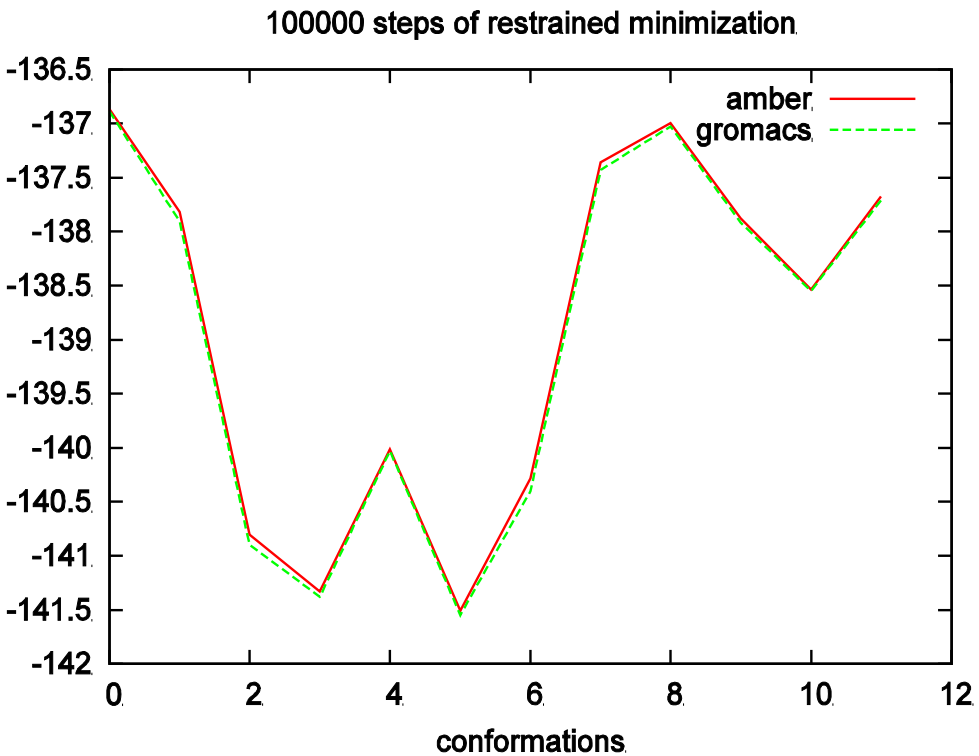
Methyl 2-methylpropanoate

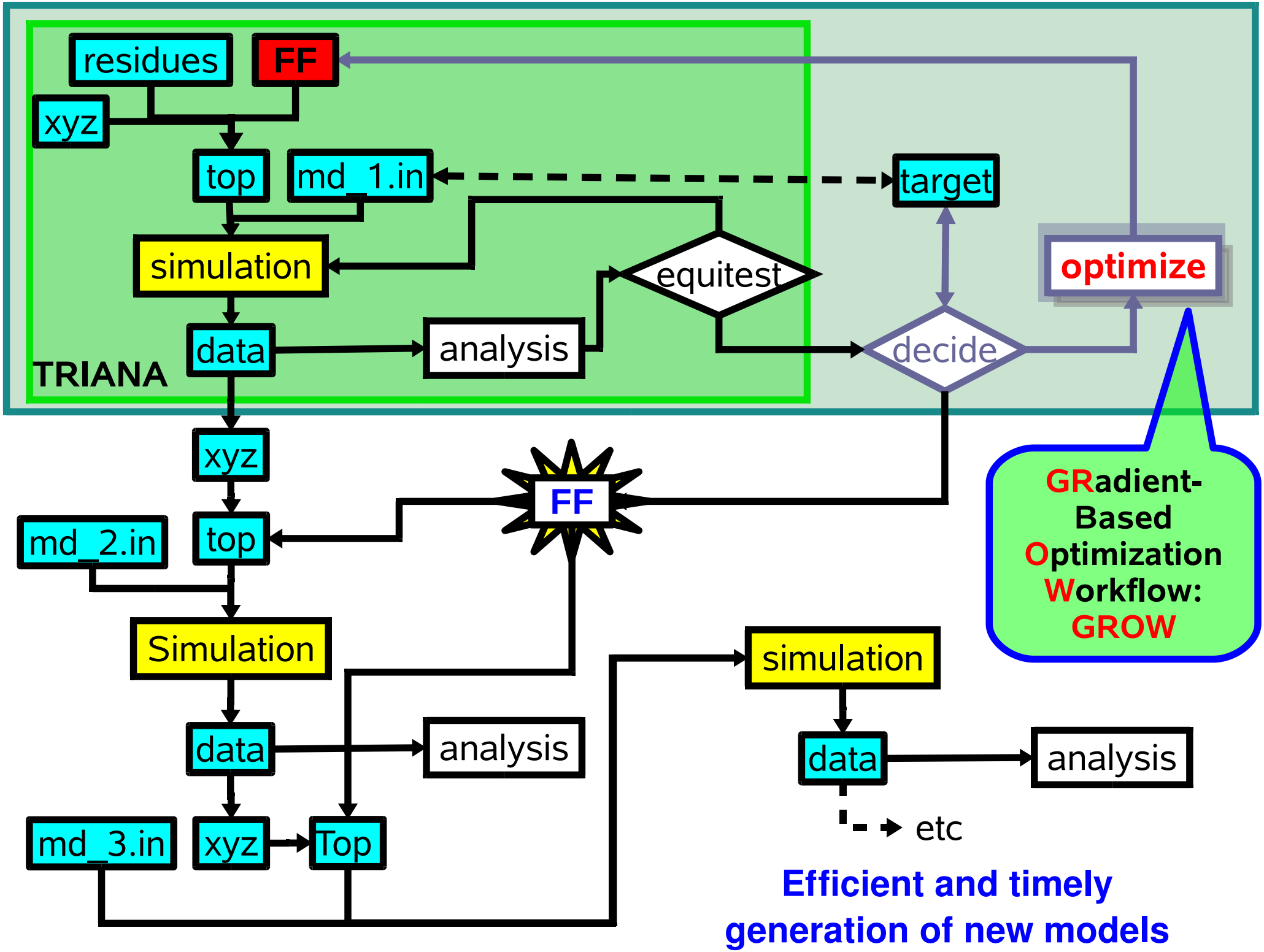


max. rms = 0.00494 Å

max. rms = 0.0196 Å

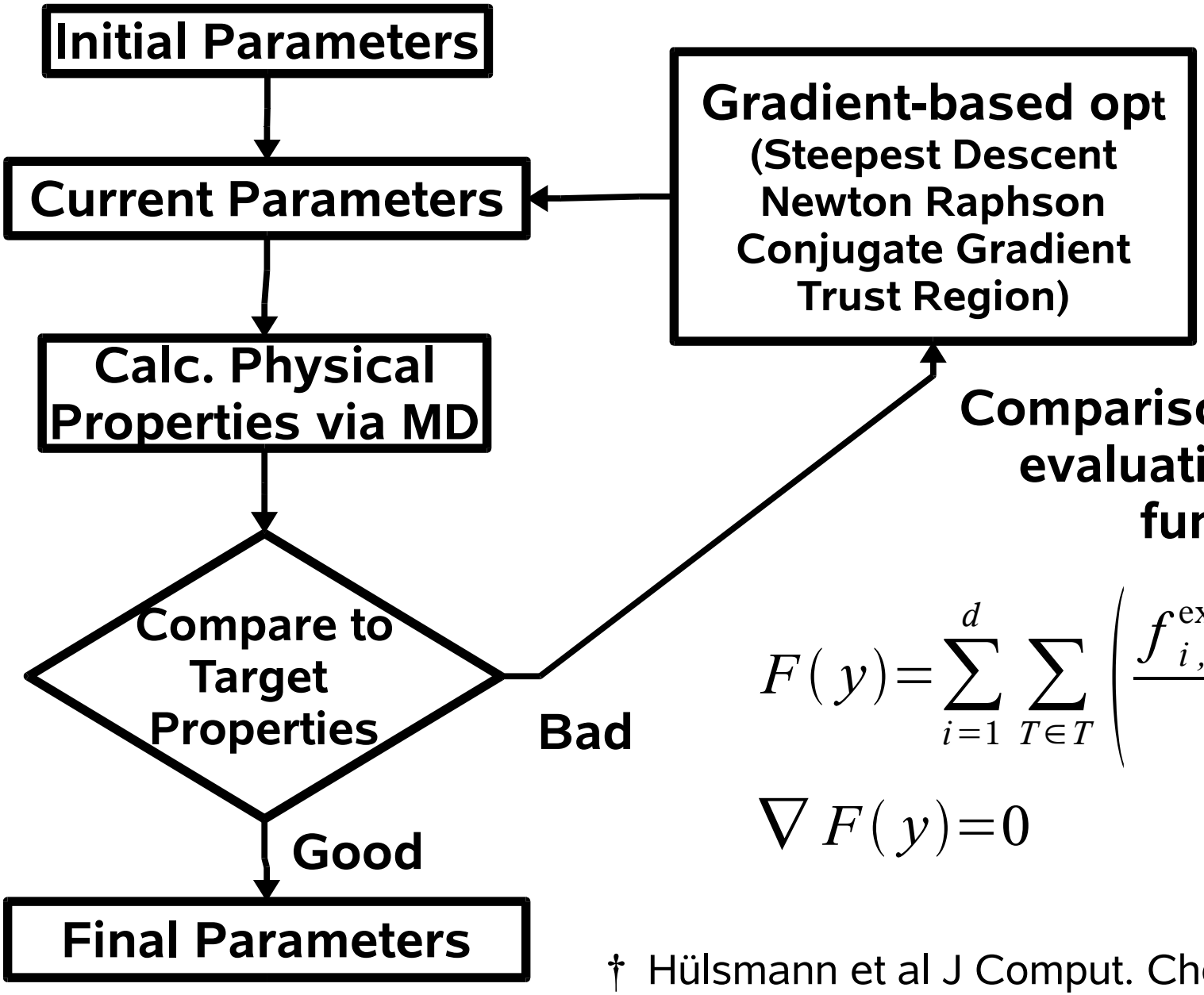





residues
FF
xyz
top
md_1.in
target
simulation
equitest
optimize
data
analysis
decide
TRIANA
xyz
FF
GRadient-Based Optimization Workflow: GROW
md_2.in
top
Simulation
simulation
data
analysis
data
analysis
md_3.in
xyz
Top
etc
Efficient and timely generation of new models

GROW† (opt of L-J parameters)

Fraunhofer SCAI



Comparison is done by evaluating the loss function

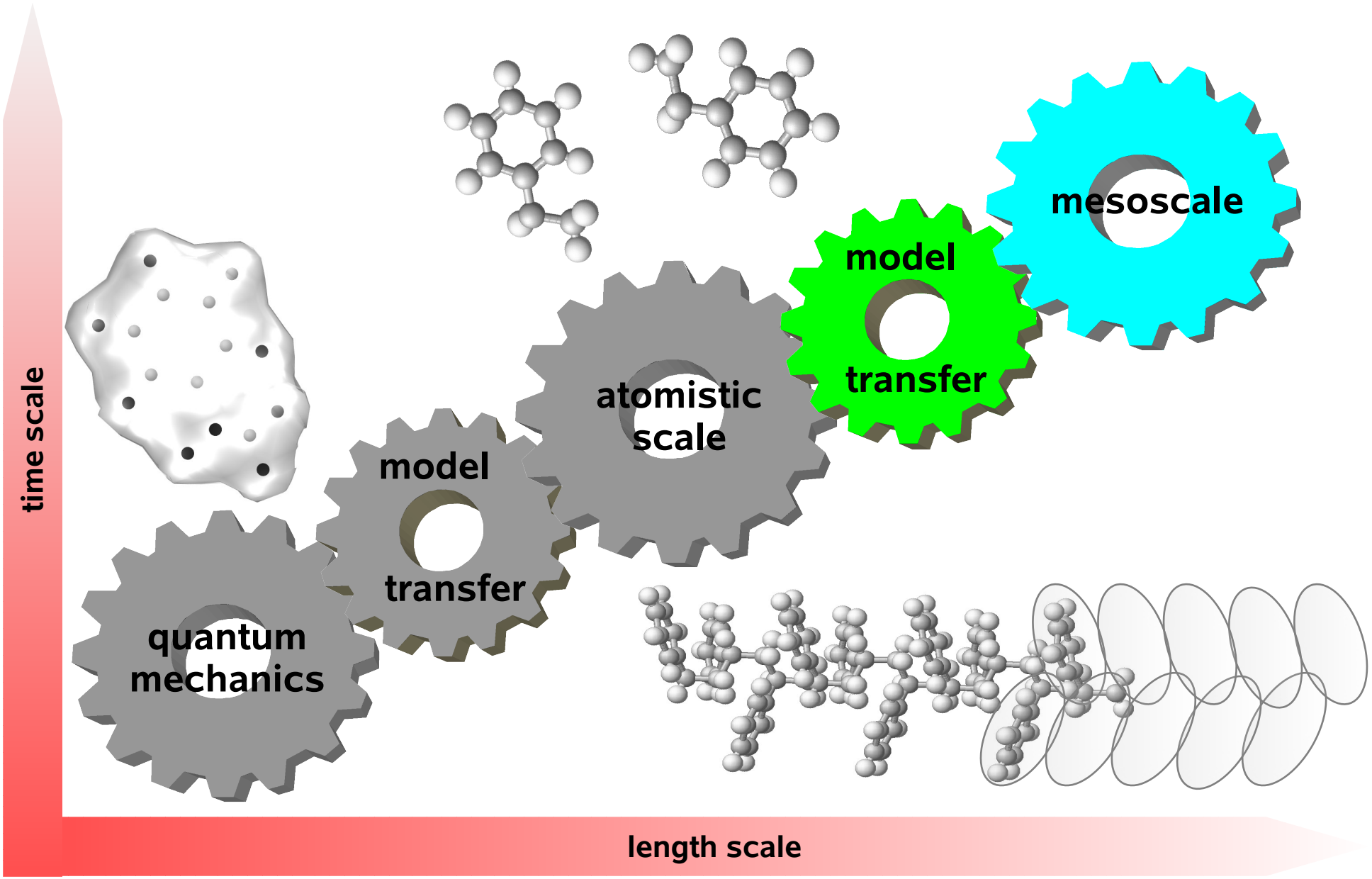
$$F(y)=\sum_{i=1}^{d}\sum_{T\in T}\left(\frac{f_{i,T}^{\exp}-F_{i,T}^{sim}(y)}{f_{i,T}^{\exp}}\right)^2$$

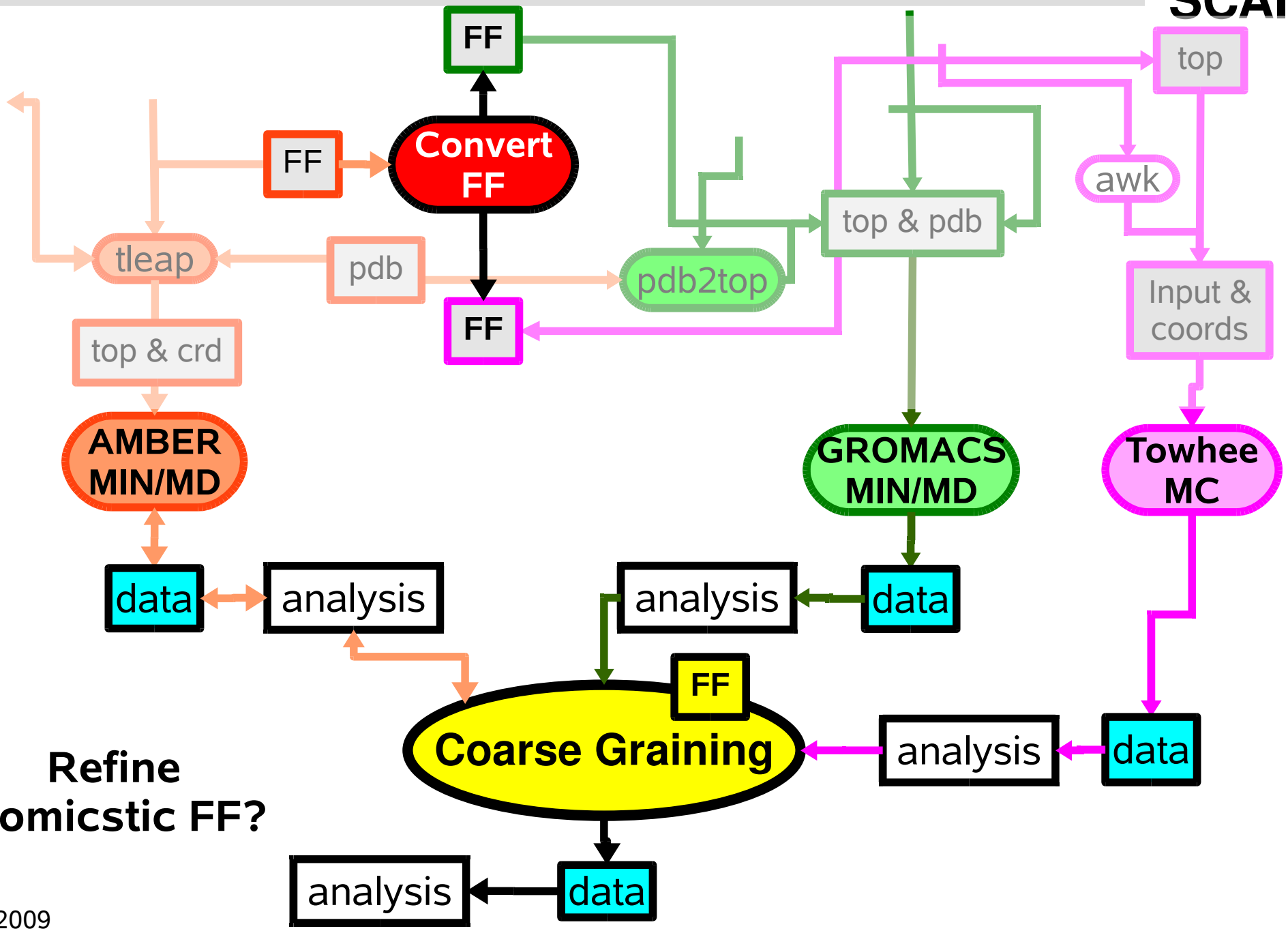
$$\nabla F(y)=0$$

† Hülsmann et al J Comput. Chem Submitted & Hülsmann et al Comp Phys Comm submitted

Soft matter - Multi-scaling modeling

Fraunhofer SCAI





Refine atomicstic FF?

Fraunhofer SCAI

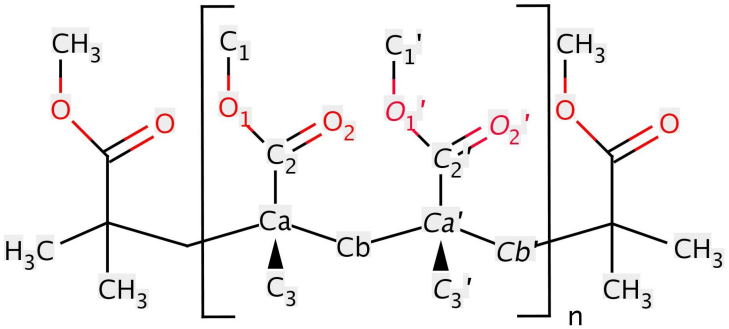
Atomistic model

Number of polymers: 59

Number of monomers per molecule: 8

Number of atoms: 7375

Temperature: 450 K

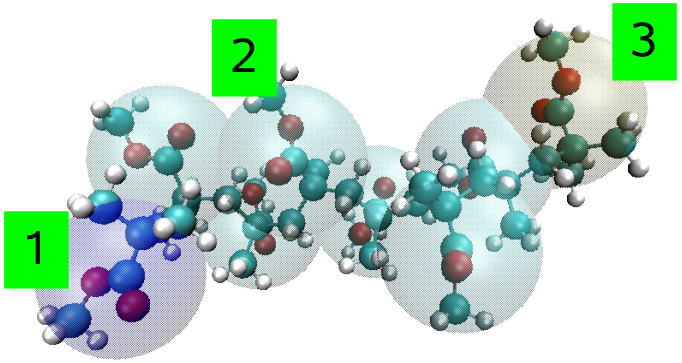


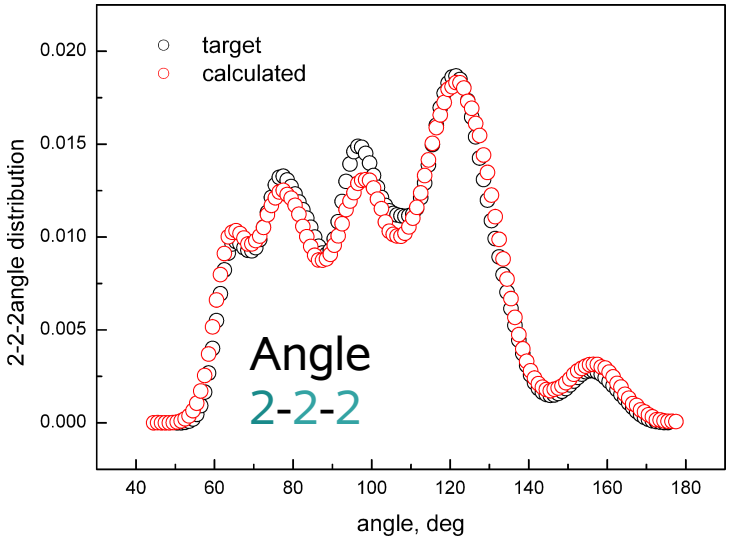
CG model

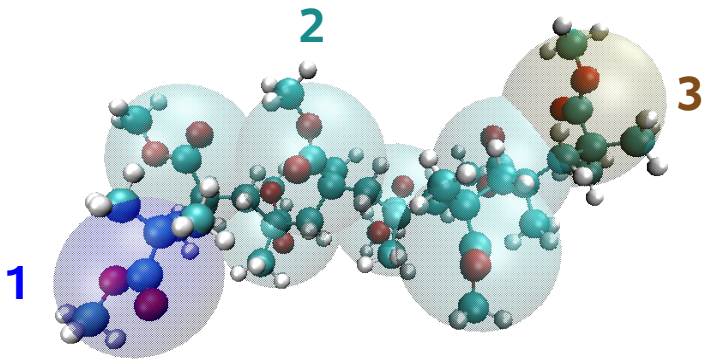
Number of beads: 472

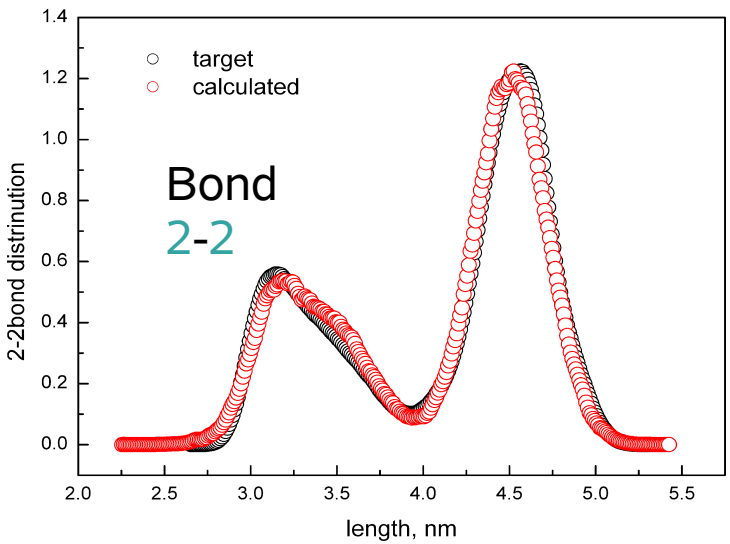
Number of different type of the beads: 3

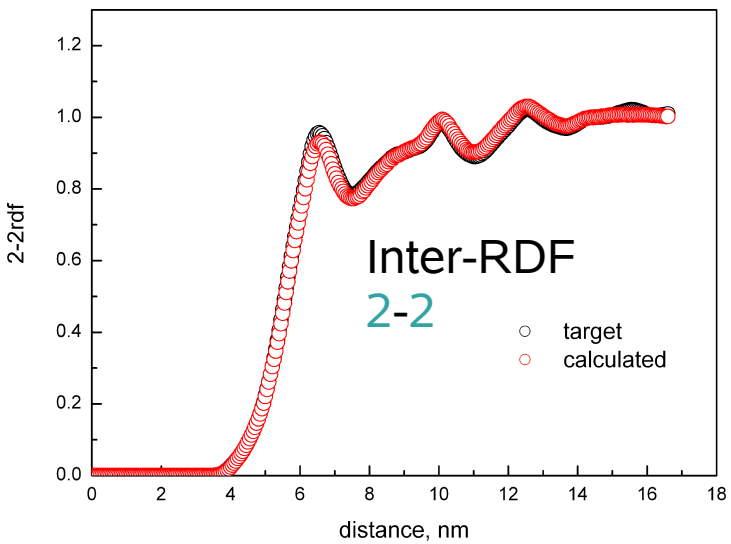
The C_2 position has been chosen as center of the beads











We reproduce atomistic MD target distributions

for inter- and intramolecular interactions

ESPResSO: Extensible **S**imulation **P**ackage for **Res**earch on **So**ft matter, MPI for polymer research (MPIP), Mainz

Open source, users in about 100 institutes all over the world

ESPResSo++: collaborative project of MPIP and Fraunhofer SCAI

Integration scheme:

- Standard MD (velocity-Verlet, constant energy)
- MD with constant T (eg Langevin) or constant P

Interactions

- Lennard-Jones, Morse, Gay-Berne (ellipsoids), tabulated potential
- Harmonic or nonlinear expander bonds, angle potentials, dihedrals

Spatial constraints: walls, spheres, cylinders

Electrostatics; Hydrodynamics via Lattice-Boltzmann; Analysis

Summary

Fraunhofer SCAI

FF Generation + Opt

a) Isolate chemical functionalities
b) Explicit parameters
c) No nb or el 1-4 scaling
d) RESP: Boltzmann, Ensemble Averaging
e) In-house scripts

Hydrocarbon (un- saturated), Alcohols, Ester, Ethers, (Amides)

?

Coarse Graining

a) Collaboration
b) ESPResSo++

Program Independent

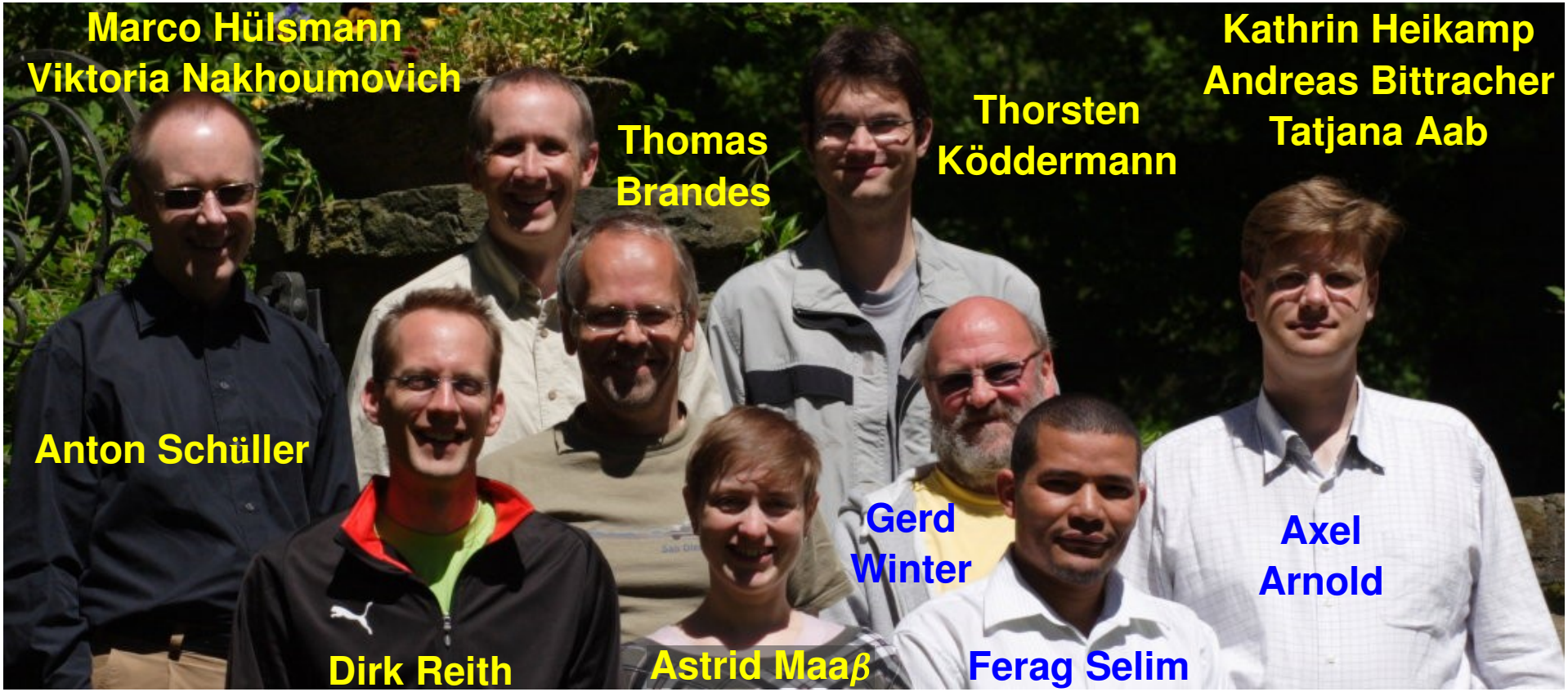
a) AMBER, GROMACS, Towee, etc.
b) In-house script converters for FF, residue lib., etc.
c) Must test FF performance – min & MD

FF Refinement

a) TRIANA Workflow to automate job cycle
b) GROW: parameter optimization towards target goal

Acknowledgments & Coworkers

Fraunhofer SCAI



Glycam06
Robert Woods (UGA)
Sarah Tschampel (UGA)
Austin Yongye (UGA)

Coarse Graining
Florian Müller-Plathe (TU Darmstadt)
Elena Algaer (TU Darmstadt)
Ali Karimi (TU Darmstadt)

Funding: NIH, NSF, PNNL