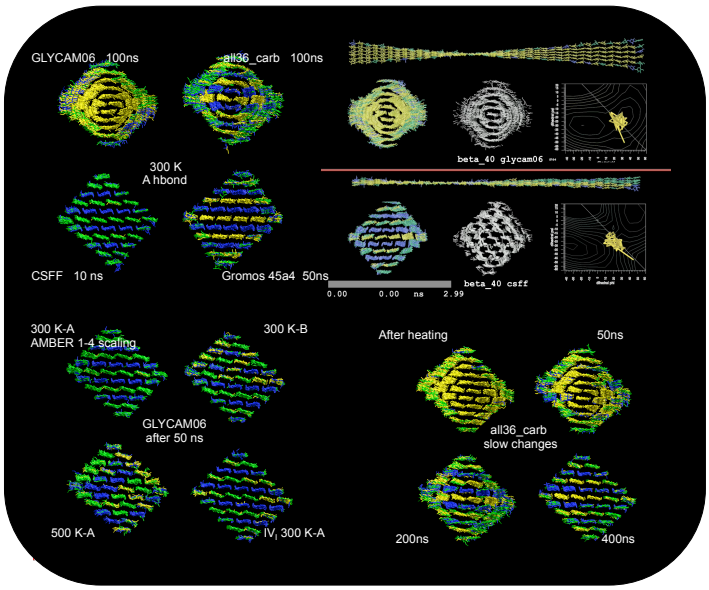
Cellobiohydrolase Processivity:

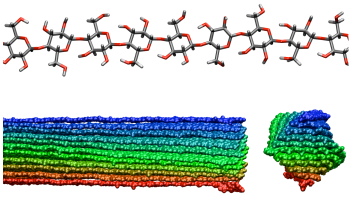
Improved enzymes for bioethanol

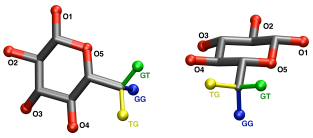
James Matthews, Eric Knoll, Antti-Pekka Hynninen, Mark Nimlos, Wes Jones, Michael Himmel, Michael Crowley • NREL, Golden, CO
Malin Bergenstrahle, John Brady • Cornell University, Ithaca, NY
Mike Garrahan, Charles Brooks III • University of Michigan, Ann Arbor, MI
Ross Walker, Mark Williamson, Andreas W. Goetz • San Diego Supercomputer Center, La Jolla, CA

Cellulose Modelling: Force Fields and Structure

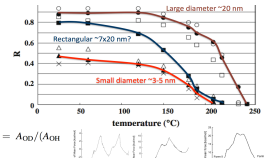
Matthews (NREL), Bergenstrahle,(Cornell), Brady(Cornell), Crowley(NREL)







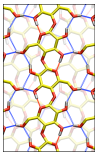
Primary alcohol conformations of glucose. Colors same in Fig at top.



$R = A_{OD}/(A_{OH}$

Free energy for phase transition in terephthalic acid. Figures from Beckham et al. JACS 2007, 129, 4715 Beckham et al. JPC-B 2008, 112, 7460

At high temperature, hydrogen/deuterium exchange proceeds to the interior of the crystals. This indicates a significant change in hydrogen bond pattern, most likely including a change in hydroxymethyl rotamer conformation. Figure from Horikawa et al. Cellulose 2009, 16, 1-8



Hydrogen bonds in celluose I. In proposed crystal structures there are no hydrogen bonds between layers, and all primary alcohol groups are TG. Elevated temperature makes every second layer GG, allowing for hydrogen bonding between layers.

CHAMBER: Support for CHARMM Force Fields within AMBER's MD engines

Williamson(SDSC), Walker (SDSC), Crowley(NREL)

CHAMBER:

- A tool to convert CHARMM protein structure file (PSF), coordinate file (COR), and the associated force field files into an AMBER topology file (prmtop) and the associated coordinate file (inpcrd)
- A true representation of the CHARMM force field in AMBER providing energies and forces that are the same to the limits of machine precision.

Benefits:

- Enables the simulation of CHARMM parameterised models using AMBER's PMEMD engine, which has improved serial and parallel efficiency over large numbers of CPUs.
- Software was released in the new 1.3 version of the free AMBERTools suite. DOI: 10.1002/qua.22372

CHARMM generated test system tri-alanine peptide:

Relative error in energy term:

ANGL	7.6e-16	ELEC	8.0e-15
CMAP	-5.7e-15	ENER	9.1e-16
DIHE	-6.2e-16	VDW	-5.5e-14
IMPR	1.4e-13		

RMSD of forces on all atoms: 1.9e-12

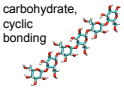
Relative error in energy between two gas phase single point energy evaluations in CHARMM c36a2 and SANDER running a CHAMBER converted psf file for the "tri-alanine peptide" test case. Energy units in Kcal/Mol.

Validation Suite for MD Codes and Force Fields

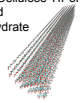
Garrahan(U. Michigan), Williamson(SDSC), Walker (SDSC)

- Charmm27 force field + carbohydrates
- 4 Systems test different features
- PSF, coordinates, topology, parameters
- Energies and gradient at double precision
- Comparison script for validation
- Implemented in CHARMM, Amber, LAMMPS
- Instructions provided for NAMD, Desmond, …
- More FF coming: Amber, GROMOS, Glycam

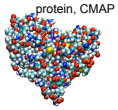
Cellohexose, 129 atoms carbohydrate, cyclic bonding



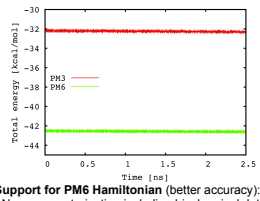
6-6-40 Cellulose TIP3P solvated carbohydrate 88,101 atoms, PME



DHFR, 2480 atoms protein, CMAP



QM Region Methods Needed in High Performance Codes



Support for PM6 Hamiltonian (better accuracy):

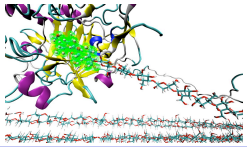
- New parameterization including biochemical data
- Explicit diatomic core-core repulsion parameters
- d-type orbitals

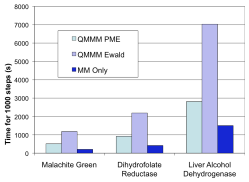
Support for PM3/ZnB Hamiltonian:

- Proper description of catalytic and structural zinc ions in protein environments

High performance QM/MM for cellulase studies. Designed to be a drop in replacement for classical simulations to enable the study of **reactive processes**.

- **Periodic boundaries** (PME)
- **Implicit solvent** (GB)
- **Parallel** support
- **Thermodynamic integration**
- **Complex restraints**
- **Accurate gradients** (NVE support)





Development of **external QM/MM library**:

- Simplify integration with other program packages

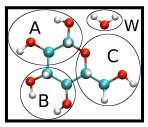
Under development:

- Integration with density functional theory
- Variable QM/MM solvent region

Coarse-graining of β-D Glucose using force-matching method

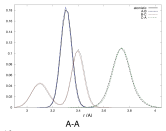
- Each atom belongs to a single coarse-grained (CG) bead
- Each glucose unit is represented with 3 beads
- Water is represented with a single bead
- CG bead is set at the center of mass of atoms
- CG force is equal to the sum of atomistic forces
- CG force-field consists of:
  - Bonded interactions within molecules
  - Non-bonded interactions between different molecules
- Force matching is done by minimizing the least-squares difference between CG forces and the CG force field

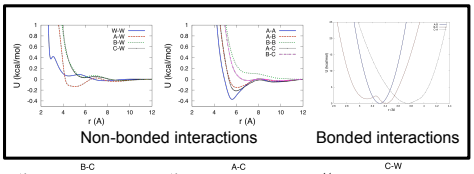
$$\left\langle \sum_{i=1}^{N} \left| \mathbf{f}_i - \mathbf{F}_i^{CG}(\phi) \right| \right\rangle$$



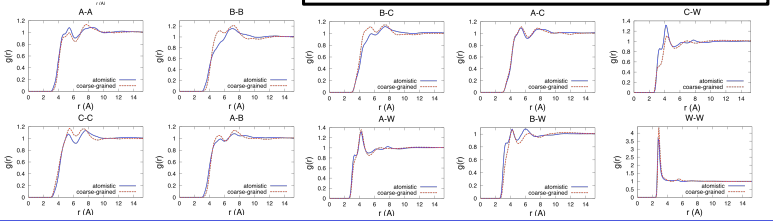
Glucose model with definition of the coarse-grained beads

Distribution functions





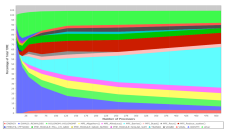
Non-bonded interactions Bonded interactions



Enhanced Codes and Modernization

Garrahan(U. Michigan), Knoll(NREL), Williamson(SDSC), Walker (SDSC), Crowley(NREL)

- Modernization of CHARMM from Fortran 77 to 95
  - Dynamic memory allocation
  - Conversion of common blocks into modules
  - Elimination of obsolete constructs
  - Update from fixed-form to free-form syntax
  - CruiseControl – a continuous integration tool
- Modernization allows easier use of tools
  - Used TAU (Tuning & Analysis Utilities), an algorithm profiling and tracing tool, to instrument CHARMM and other MD codes to identify parallel bottlenecks within the MPI communication layer
  - Valgrind memcheck to detect array boundary violation and uninitialized values

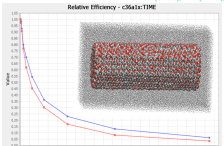


Profiling of CHARMM showing the percentage of total time for each subroutine or MPI call vs. the number of processors to identifiy parallel scaling problems in the code.

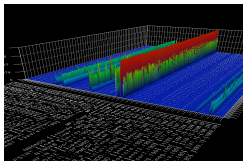
Current Benchmarks

Williamson(SDSC), Knoll(NREL), Walker (SDSC)

Current relative efficiency of the CHARMM code running the cellulose benchmark, over increasing numbers of processors. Inset: A 408,609 atom cellulose fiber in a water box. Bundle length ~200 Å; diameter ~70 Å. The simulation runs NVT for 500 steps using a 1fs time step, giving 0.5 ps of simulation length.



CHARMM c36a1x timing decompositions per MPI thread for the cellulose benchmark over 512 nodes, generated using TAU. The x axis is showing which subroutine is being evaluated, MPI thread is shown on the y axis and the z axis is time spend in the subroutine. Red indicates that more time is being spent in this subroutine.

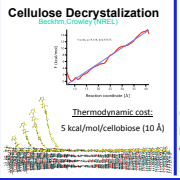


Simulations Enabled by CHAMBER and New Restraints

Cellulose and Force Fields

Cellulose Decrystalization

Beckhm,Crowley (NREL)

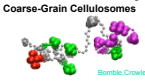


Thermodynamic cost:

5 kcal/mol/cellobiose (10 Å)

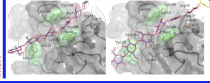
Cellulosomes and Cellulosomal Enzymes

Coarse-Grain Cellulosomes



Bomble,Crowley (NREL)

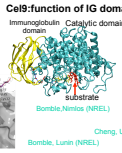
Structures of CBM4 from CbhA



Bomble, Lunin (NREL)

BESC BioEnergy Science Center

Cel9:function of IG domain



Bomble,Nimlos (NREL)

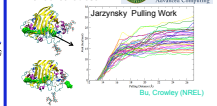
Conformational changes in CelS

Bomble (NREL)

Cheng, Uberbacher (ORNL)

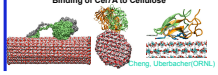
Processivity of Cel7A: Thermodynamics of exit of cellobiose

SciDAC Scientific Discovery through Advanced Computing



Bu, Crowley (NREL)

Binding of Cel7A to Cellulose



Cheng, Uberbacher(ORNL)

This work was supported by the DOE Office of Science SciDAC Program