CCP-SAS – a community consortium for the atomistic modeling of scattering data

Joseph E. Curtis
NIST Center for Neutron Research
Scatters & Simulators

Develop software with a community of users and developers to advance atomistic modeling of soft-condensed matter.
ccp-sas: nsf/epsrc 2013-2017

Software Infrastructure for Sustained Innovation (SI2) - Grand Challenges in the Chemical Sciences

USA
NIST
UTK
U. Texas SA
KSU
APS

UK
UCL
Kings College London
University of Bath
University of Nottingham
ISIS
Diamond

Teams

Core Software -- Chair: Curtis, Brookes(UT), Draper(ISIS), Perkins(UCL)

Chemical physics -- Chair: Chen(KSU), Curtis, Butler, Edler(UB), King(ISIS), Heenan(ISIS)

Dissemination -- Chair: King(ISIS), Terrill(Diamond), Butler, Irving(APS)

Testing and Applications --
Co-Chair: Perkins(UCL), Barlow(KCL), Edler(UB), Scott(UN), Terrill(Diamond)
Co-Chair: Butler, Krueger, Fushman(UMD), Liu(UD), Schildbach(JHU), Van Duyne (UP), Wright(JMU)

~4 developers

kickoff meeting: October 24-27, 2013 @ NCNR
deliverables by year

[1] web prototype and alpha testing

[2] web released, hpc backend prototype alpha testing

[3] disseminate hpc beta to international centers, incorporate contributed community code

[4] advance gpu implementation, soft-matter builder, polish, hpc release
types of systems

proteins
nucleic acids
lipids
polymers
sans/saxs
reflectivity
dynamics

Community software is more than ‘free’

Peer-improvement

Code encapsulation and re-use
So many options . . .

Bio-simulation (All-atom MD packages):

<table>
<thead>
<tr>
<th>Amber*</th>
<th>CHARMM*</th>
<th>Amber*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHARMM*</td>
<td>CVFF</td>
<td>CVFF</td>
</tr>
<tr>
<td>NAMD*</td>
<td>COSMOS-NMR</td>
<td>COSMOS-NMR</td>
</tr>
<tr>
<td>GROMACS*</td>
<td>GROMACS*</td>
<td>GROMACS*</td>
</tr>
<tr>
<td>GROMOS</td>
<td>GROMOS</td>
<td>GROMOS</td>
</tr>
<tr>
<td>LAMPPS*</td>
<td>OPLS*</td>
<td>OPLS*</td>
</tr>
<tr>
<td>PINYMD*</td>
<td>ENZMIX</td>
<td>ENZMIX</td>
</tr>
<tr>
<td>HIPPO</td>
<td>ECEPP/2</td>
<td>ECEPP/2</td>
</tr>
<tr>
<td>GPIUTMD</td>
<td>QCFF/PI</td>
<td>QCFF/PI</td>
</tr>
<tr>
<td>DL_POLY*</td>
<td>UFF</td>
<td>UFF</td>
</tr>
<tr>
<td>ESPResSo</td>
<td>CFF*</td>
<td>CFF*</td>
</tr>
<tr>
<td>MacroModel*</td>
<td>MMFF</td>
<td>MMFF</td>
</tr>
<tr>
<td>MACSIMUS</td>
<td>MM2, MM3, MM4*</td>
<td>MM2, MM3, MM4*</td>
</tr>
<tr>
<td>MOLDY</td>
<td>XPOL</td>
<td>XPOL</td>
</tr>
<tr>
<td>MOSCITO</td>
<td>SIBFA</td>
<td>SIBFA</td>
</tr>
<tr>
<td>ProtoMol</td>
<td>AMOEBA</td>
<td>AMOEBA</td>
</tr>
<tr>
<td>TINKER*</td>
<td>VALBOND</td>
<td>VALBOND</td>
</tr>
<tr>
<td>MDGrape</td>
<td>DRF90</td>
<td>DRF90</td>
</tr>
<tr>
<td>Materials Studio (InsightII)*</td>
<td>CG MD*</td>
<td>CG MD*</td>
</tr>
</tbody>
</table>

Classical force fields:

<table>
<thead>
<tr>
<th>Amber*</th>
<th>CHARMM*</th>
<th>Amber*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVFF</td>
<td>COSMOS-NMR</td>
<td>CVFF</td>
</tr>
<tr>
<td>OPLS*</td>
<td>ENZMIX</td>
<td>OPLS*</td>
</tr>
<tr>
<td>ENZMIX</td>
<td>ECEPP/2</td>
<td>ENZMIX</td>
</tr>
<tr>
<td>QCFF/PI</td>
<td>UFF</td>
<td>QCFF/PI</td>
</tr>
<tr>
<td>CFF*</td>
<td>MMFF</td>
<td>CFF*</td>
</tr>
<tr>
<td>MM2, MM3, MM4*</td>
<td>XPOL</td>
<td>MM2, MM3, MM4*</td>
</tr>
<tr>
<td>XPOL</td>
<td>SIBFA</td>
<td>XPOL</td>
</tr>
<tr>
<td>AMOEBA</td>
<td>VALBOND</td>
<td>AMOEBA</td>
</tr>
<tr>
<td>DRF90</td>
<td>CG MD*</td>
<td>DRF90</td>
</tr>
</tbody>
</table>

Methods & details*:

<table>
<thead>
<tr>
<th>Thermostats</th>
<th>Barostats</th>
<th>Thermostats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barostats</td>
<td>Electrostatics (PBC)</td>
<td>Barostats</td>
</tr>
<tr>
<td>Polarizability</td>
<td>Implicit Solvent</td>
<td>Polarizability</td>
</tr>
<tr>
<td>Langevin Dynamics</td>
<td>Parallel Tempering</td>
<td>Langevin Dynamics</td>
</tr>
<tr>
<td>Replica Exchange</td>
<td>Steered MD</td>
<td>Replica Exchange</td>
</tr>
<tr>
<td>Free-Energy Calculations</td>
<td>Umbrella Sampling</td>
<td>Free-Energy Calculations</td>
</tr>
<tr>
<td>VMD</td>
<td>Normal Mode Analysis</td>
<td>VMD</td>
</tr>
<tr>
<td>Pymol</td>
<td>Umbrella Sampling</td>
<td>Pymol</td>
</tr>
<tr>
<td>Chimera, O</td>
<td>Normal Mode Analysis</td>
<td>Chimera, O</td>
</tr>
<tr>
<td>APBS</td>
<td>VMD</td>
<td>APBS</td>
</tr>
<tr>
<td>Gaussian, Gamess, CPMD</td>
<td>Free-Energy Calculations</td>
<td>Gaussian, Gamess, CPMD</td>
</tr>
<tr>
<td>Beowulf, GPU Clusters</td>
<td>Umbrella Sampling</td>
<td>Beowulf, GPU Clusters</td>
</tr>
<tr>
<td>Rosetta</td>
<td>Parallel Tempering</td>
<td>Rosetta</td>
</tr>
<tr>
<td>PHYRE</td>
<td>Steered MD</td>
<td>PHYRE</td>
</tr>
<tr>
<td>Folding @ Home</td>
<td>Free-Energy Calculations</td>
<td>Folding @ Home</td>
</tr>
<tr>
<td>TIP3, TIP3P, TIP4P, SPC, ST2</td>
<td>Umbrella Sampling</td>
<td>TIP3, TIP3P, TIP4P, SPC, ST2</td>
</tr>
</tbody>
</table>

...
Technical details:

Command line, force-field, basic PDB repair, input files, approximations, equilibration, mining trajectories, HPC proposals, HPC queue systems ...

If a simple answer is going to take too long users do something else

large configuration space

50E6 atoms : 100s of $\mu$s
MD / MC & SAS

- Reduce degrees of freedom
- Rapid configuration space coverage
- Ask/Answer “What if” questions
- Prepare representative ensembles for further simulation and analysis

Figure 2. Comparison of configuration space coverage of HIV-1 integrase by molecular dynamics simulation (TOP) – compared to the more extensive conformations (orange) from Monte Carlo simulation (BOTTOM).

100 ns to > ms
A solution: SASSIE (part-time boot-strap N=0.2)

Developed to enable NCNR user community to efficiently develop molecular models for the neutron/X-ray scattering/reflectivity experiments.

Support collaborations with an eye towards code-reusability and modularity.

> 20 manuscripts (http://www.smallangles.net/sassie/SASSIE/Results.html)

Prototype | Beta | Refactoring | CPC release
--- | --- | --- | ---
2004 | 2009 |  | 2011

www.smallangles.net/sassie

Simulation for SAS Community

Directed tools to build structures & topology (FF)

Easy access to MC / MD programs AND hardware

Workflow connection from structures to SAS

Modular: in and out to use what you want

Handle long jobs … re-attach / restart

Constraints (NMR, AUC, etc.)

Open source: SAS users and developers
**GenApp Framework**

**Philosophy**

Simplify broad deployment and insure preservation of scientific codes in an ever-evolving software environment landscape.

Lower entry barrier for implementation of new ideas and new codes.

Open code without onerous licensing.

Community governance.

---

**Scientist**

“has code”

“writes text file”

to define GUI/link
code to GenApp

Then “compiles”

Done

has web-app, HPC links, and

GUI

---


“application generator” : web, QT-GUI, Java, Andriod/iOS
**GenApp Framework**

**Philosophy**

- Simplify broad deployment and insure preservation of scientific codes in an ever-evolving software environment landscape.
- Lower entry barrier for implementation of new ideas and new codes.
- Open code without onerous licensing.
- Community governance.

Let's build a community

http://genapp.rocks

ds-DNA : Howell/Chu (GWU)
SasCalc-MD : Koefinger (Max Planck)
MULCH : Trewhella (U. Sydney)
GenFit : Spinozzi (UPDM)
Quafit : Spinozzi (UPDM)
NMR Constraints : Fushman (UMD)
Denfert : Perez (Soleil)
Vortex Shedding : Pearlstein (UCIC)
ccp-sas: core-software team

Extensible application framework

Web UI

Support browser
User access benefits

HPC (core)
Or
HPC (gateway)

Server we control “instant legacy”
Add HPC gateway based on demand
Code is run on server not user’s laptop/tablet

Modularity: use what you need (in & out)
HPC is relative

Enable advanced MD & APBS
XSEDE (IU) & Langan & Smith (ORNL)
**Monomer Monte Carlo:** single chain protein or ss-NA (think IDP / ID-NA)

**Complex Monte Carlo:** multiple single chain protein or ss-NA (think IDP / ID-NA complexes)

**Energy Minimization:** anything FF supports (NAMD)

**Torsion Angle MD:** protein/NA/Carbs. FF supports (CHARMM)

**Docking:** anything FF supports

**Two-Body Grid:** anything FF supports
Much more robust than Cartesian MD (no high-frequency bond/angle vibrations)

Larger MD time steps

Sample the most relevant degrees of freedom (dihedrals)

Convenient classification of rigid and flexible regions

Can handle internal loops

Bio & soft-matter* systems

Requires extensive energy minimization to start

Run MC then TAMD: multi-scale sampling → use SAS as a guide

Jianhan Chen’s poster (was Monday)

J. Chen, W. Im, C. Brookes, JCC 26, 1565 (2005), Zhang et al. (in preparation)
TAMD: combine MC w/ fast MD

MC (50000)

Spatial Clusters (~100)

Protein, DNA, RNA, carbohydrates ... and complexes

Implicit solvent effects

Zhang et al. (in preparation) Constrained Rg

EFF1, SASA, GB, ACE
Multiple, multi-chain species: ds-DNA, proteins, RNA, carbohydrates, etc.

User defined torsion sampling

backbone, concerted motions, side-chains, carbohydrates, iso-peptide bonds, etc.

Group based rotations ... can handle odd topologies (mAb) natively.

Implicit solvent options

Not all chains need to be flexible but can simultaneously sample all DOF

Howell et al. (submitted) & Curtis et al. (in preparation)
Calculate

Build -- evaluate and clean up structure files

- **PDB Scan** Generates a report that characterizes the user supplied PDB file.
- **PDB Rx** Attempts to correct mistakes in user supplied PDB file. **ALPHA**
- **CG Builder** Tools to assist the generation of coarse-grain structures. **ALPHA**

Input a TRAJECTORY and get an *ensemble* of scattering profiles

Calculate -- generate theoretical scattering data from structures

- **SasCalc** Calculates neutron and X-ray scattering profiles from input structures. **ALPHA**
- **SasCalc-MD** Calculates neutron and X-ray scattering profiles from MD trajectories with explicit water. **ALPHA**
- **Xtal2sas** Calculates neutron scattering profiles from input structures.
- **SCT Calculate** Calculates neutron and X-ray scattering profiles from input structures.
- **SLD MOL** Calculates neutron and X-ray reflectivity scattering length density from user supplied structures. Utilities for experimental planning and isotopic labeling and optimization of ensemble populations are supplied.
- **EM to SANS** Calculates neutron scattering profile from user supplied electron density map.
Both $X^2 \sim 1$

Free Energy differs by $> 400$ kCal/mol

General Soft Matter

MD & Force-fields most advanced for proteins, NA

General builder for all soft-matter systems is not ready for non-specialist … Martini is well documented!

Many investigators have trajectories … can use SasCalc (alpha) … all-atom or CG … neutrons and/or X-ray

→ Hailiang Zhang’s poster today

For community: documentation and training for FF development and coarse-grain and TAMD simulations

→ Karen Edler’s poster today

General CG builders and links to simulation engines in a few years

all contrasts (n & X) at once for ensemble
Create new and enable existing open-source simulation tools & HPC to model scattering data to dramatically improve accessibility by non-experts

Adapt further physical constraints to problem (both experimentally and from advanced simulation methods)

Disseminate software to scattering centers

Build sustainable community

ccpsas.org/impact.html

DOCS -> sassie training ...
thank you SAS-2015 Berlin and . . .

- Susan Krueger susan.krueger@nist.gov
- **Emre Brookes**
  emre@biochem.uthscsa.edu
- Hailiang Zhang hailiang.zhang@nist.gov
- Dave Wright dave.wright@ucl.ac.uk
- Katie Sarachan (NRC now @ RNA Institute SUNY)
- Max Watson (NRC now @ Data Analytics)
- Nicholas Clark (Post-Doc now @ Regeneron)
- Sindhu Raghunandan (SURF now @ UCSD)
- Harry Bullen (SURF now @ Google)

CCP-SAS Co-I's, Collaborators & Advisors:

- Paul Butler (NIST/UTK)
- Stephen Perkins (UCL)
- Jianhan Chen (Kansas State University)
- Tom Irving (IIT/APS)
- Dave Barlow (King’s College London)
- Karen Edler (Bath University)
- Richard Heenan (ISIS Neutron & Muon Source)
- Steve King (ISIS Neutron & Muon Source)
- Dave Scott (Nottingham University / Research Complex at Harwell)
- Nick Terrill (Diamond Light Source)
- Nick Draper (Tessella Ltd)
- Cameron Neylon (PLoS)

Stakeholder Collaborators:

- Alun Ashton (Diamond Light Source)
- Rob Rambo (Diamond Light Source)
- Luke Clifton (ISIS Neutron & Muon Source)
- David Fushman (University of Maryland)
- Richard Gillilan (Cornell High Energy Synchrotron Source)
- Arwel Hughes (ISIS Neutron & Muon Source)
- Michael Kent (Sandia National Laboratory)
- Jeremy Lakey (Newcastle University)
- Daniel Myatt (ISIS Neutron & Muon Source)
- Alison Paul (Cardiff University)
- Joel Schildbach (Johns Hopkins University)
- Gregory Van Duyne (University of Pennsylvania)
- David Willock (Cardiff University)
- Nathan Wright (James Madison University)