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

Joseph E. Curtis

NIST Center for Neutron Research

SAS 2015 : Berlin : September 16, 2015

EPSRC
Engineering and Physical Sciences Research Council

CCP-SAS

NSF

ccpsas.org

Scatters & Simulators

Develop software with a community of users and developers to advance atomistic modeling of soft-condensed matter

CCP-SAS Project

About the project

home | about CCP-SAS | meetings/workshops | resources | people | impact

CCP-SAS - Collaborative Computational Project for advanced analyses of structural data in chemical biology and soft condensed matter

an SI2-CHE cyberinfrastructure project addressing Grand Challenges in the Chemical Sciences

A typical bench scientist purifies and characterizes samples, collects the SAXS/SANS/AUC data, and interprets the results using simplistic models. It is rare that the same individual also has the skills to use advanced atomistic simulation software.

The CCP-SAS project is focused on developing an easy-to-use open-source modeling package that enables users to generate physically accurate atomistic models, calculate scattering profiles and compare results to experimental scattering data sets in a single web-based software suite. This enables a broad range of scattering scientists to access often complicated simulation and scattering analysis methods seamlessly thus providing a significant acceleration to the discovery process.

The objectives of the project include providing a web-based GUI front-end with a high-performance back-end to increase the accessibility of advanced atomistic modeling of scattering data by novice users. Advanced analysis modules and new simulation methods, including implicit solvent models, are also being developed to increase the accuracy of scattering calculations and simulation protocols. The project includes both a testing team and a development team and is built around a continuous development and feedback model.

The project is funded through a joint EPSRC/NSF (UK/US) grant with participants from US and UK universities and X-ray and neutron scattering facilities. It aims to leverage a user and developer community that uses high-performance computing resources on a wide-range of cutting-edge chemical problems. The goal is to build a truly international effort that includes facilities and users from around the world which will be self sustaining long after the end of this grant. To that end we welcome all interested parties to discuss and join with us at whatever level is appropriate.

Components of CCP-SAS

Projects directly involved with this project:

- **GenApp** is a modular framework for multiscale science computations being developed at the University of Texas that enables web & gui applications to be generated from same source code.
- **SASSIE** is a NIST NCNR project to provide atomistic simulation of SAS data.
- **SCT/SCTPL/HYDRO** is a fortran suite out of Prof. Perkins' lab at UCL to compare molecular models of biological systems to SAS data.
- **US-SOMO** is part of the University of Texas UltraScan II suite for the analysis of AUC and SAS data.

Related Projects of Interest:

- The CCP steering panel brings together the chairs of the active UK CCP projects. Professor Stephen Perkins represents CCP-SAS on that panel.
- SasView is an open source collaborative SAS analysis software package based primarily on analytical models.
- The SAS portal is a project being developed by the canSAS group.

lower barriers

Open Source

diamond

Science & Technology Facilities Council ISIS

Argonne NATIONAL LABORATORY

NCNR

EPSRC Engineering and Physical Sciences Research Council

UCL

NIST National Institute of Standards and Technology Technology Administration, U.S. Department of Commerce

SCHOOL OF MEDICINE UT HEALTH SCIENCE CENTER DEPARTMENT OF BIOCHEMISTRY

THE UNIVERSITY OF TENNESSEE KNOXVILLE

NSF

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

ccp-sas: nsf/epsrc 2013-2017

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

deliverables by year

[1] web prototype and alpha testing

[2] web released, hpc backend prototype alpha testing

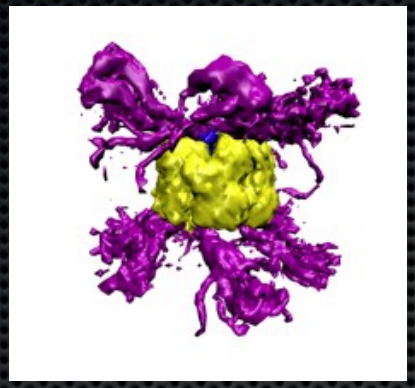
[3] disseminate hpc beta to international centers, **incorporate contributed community code**

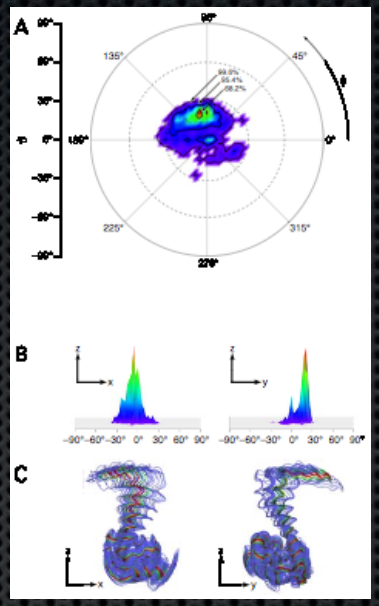
[4] advance gpu implementation, soft-matter builder, polish, hpc release

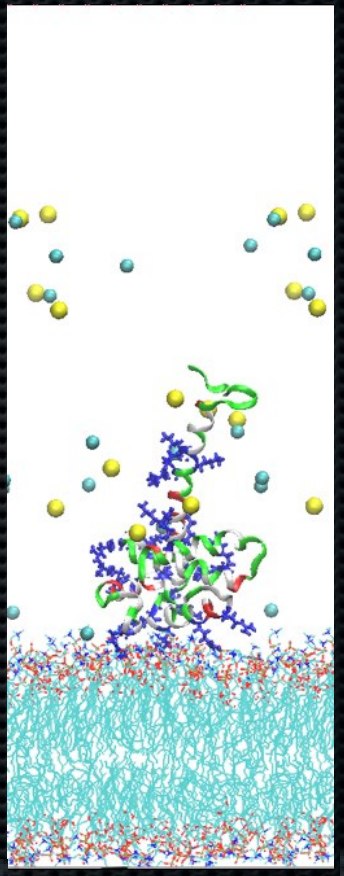
types of systems

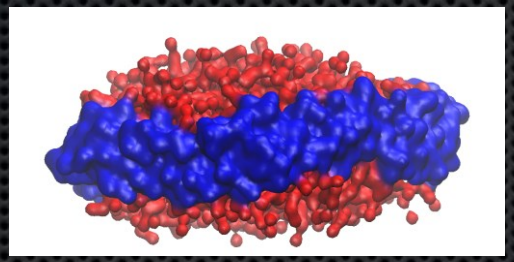
proteins
nucleic acids
lipids
polymers

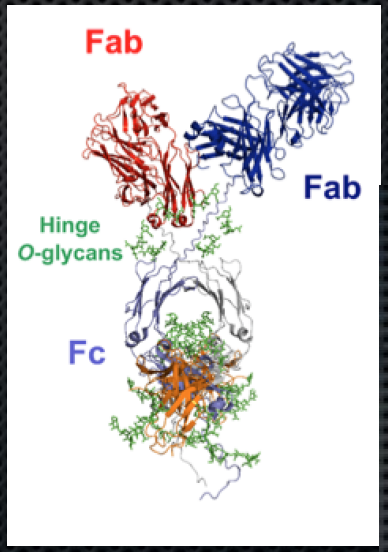
sans/saxs
reflectivity
dynamics

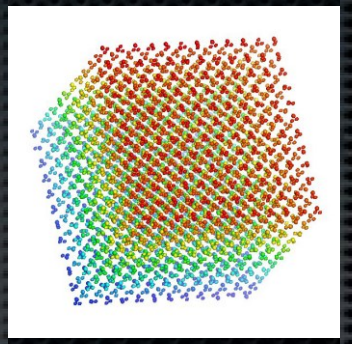


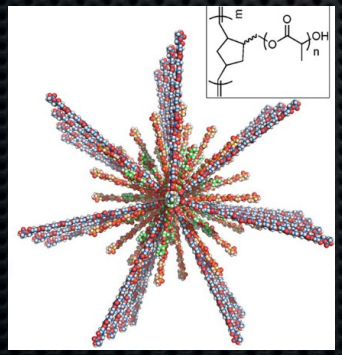


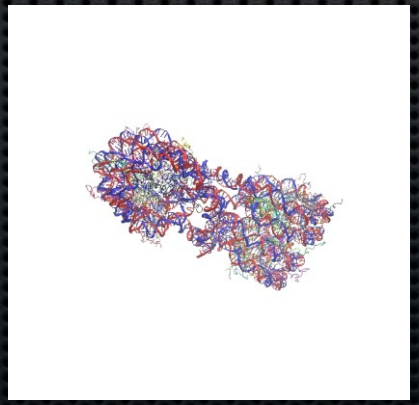


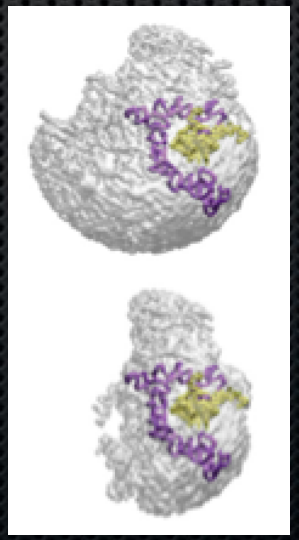


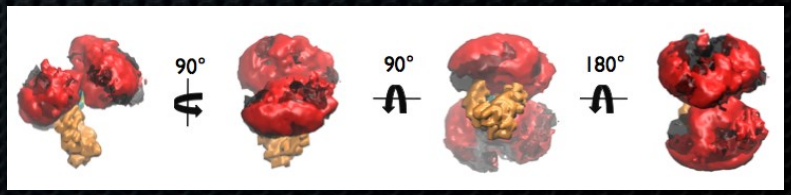












md community

MacKerell, Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr., R.L.; Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.T.K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.T.; Prodhom, B.; Reiher, III, W.E.; Roux, B.; Schlenkrich, M.; Smith, J.C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics Studies of proteins. Journal of Physical Chemistry B, 1998, 102, 3586-3616.

The Amber 8 authors are: D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, B. Wang, D.A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J.W. Caldwell, W.S. Ross, and P.A. Kollman. D.A. Case, T.E. Cheatham, III, T. Darden, H. Gohlke, R. Luo, K.M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R. Woods. The Amber biomolecular simulation programs. *J. Computat. Chem.* 26, 1668-1688 (2005).

Code encapsulation and re-use

Community software is more than ‘free’

Peer-improvement

So many options . . .

Bio-simulation (All-atom MD packages):

Amber*
CHARMM*
NAMD*
GROMACS*
GROMOS
LAMPPS*
PINYMD*
HIPPO
GPIUTMD
DL_POLY*
ESPResSo
MacroModel*
MACSIMUS
MOLDY
MOSCITO
ProtoMol
TINKER*
MDGrape
Materials Studio (InsightII)*
. . .

Classical force fields:

Amber*
CHARMM*
CVFF
COSMOS-NMR
GROMACS*
GROMOS
OPLS*
ENZMIX
ECEPP/2
QCFF/PI
UFF
CFF*
MMFF
MM2, MM3, MM4*
XPOL
SIBFA
AMOEBA
VALBOND
DRF90
CG MD*

Methods & details*:

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

not all MD / FF will work for your system

Technical details:

large configuration space

```
#!/bin/csh -f

cat > mdin <<EOF
  MD run, Temp=300K, start
 11 A water box, const eps=1.0,
 nscm=0, npscal=1, frameon=1 - p
 const pressure MD, seperate ter
 &cntrl
  irest=1, ibelly=0, imin=0,
  ipol=1,
  ntx=5, ntxo=1, ig=71277, tempi
  ntb=2,
  ntt=0, temp0=300.0, tautp=0.2,
  ntp=1,    pres0=1.0, comp=44.6
  nscm=0,
  nstlim=10,  t=1170.0, dt=0.001
  ntc=2,    tol=0.0000005,
  ntf=2,    nsnb=5,
  cut=9.0, dielc=1.0,
  ntpr=1, ntwx=5,  ntwv=5, ntwe=
  ioutfm=0,
  ntr=0,    ntave=100,  ntrx=1
 /
 &ewald
    ew_type = 0, verbose=0,
    vdwmeth=1, maxiter=10, indme
    frameon=1, irstdip=1, scaldi
 /
EOF

set
```

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

50E6 atoms :
100s of μs

MD / MC & SAS

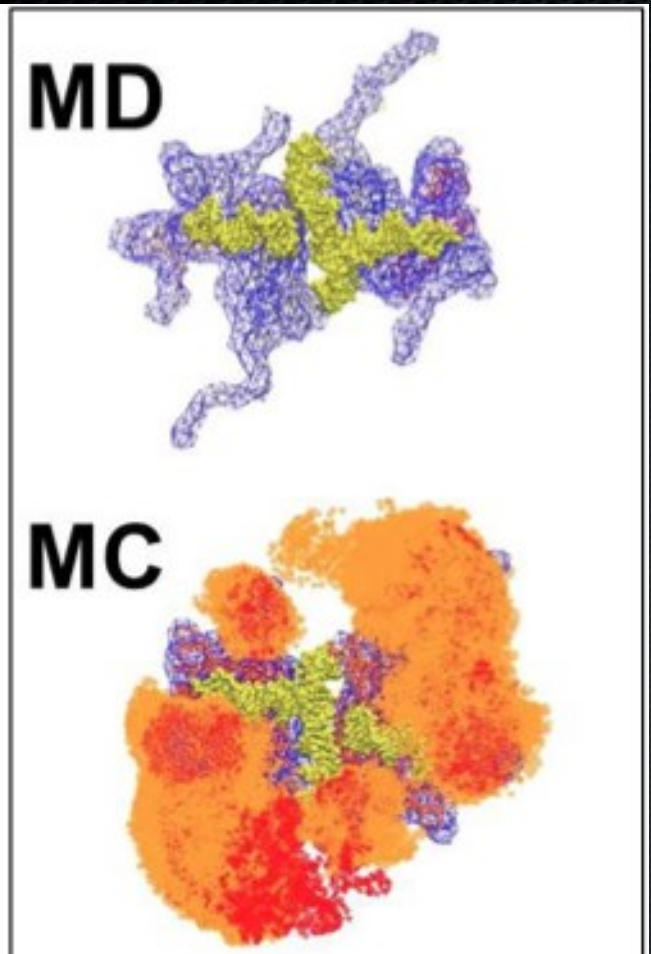


Figure 2. Comparison of configuration space coverage of HIV-1 integrase by molecular dynamics simulation (TOP) – compared to the more extensive conformations (orange) from Monte Carlo simulation (BOTTOM).

Reduce degrees of freedom

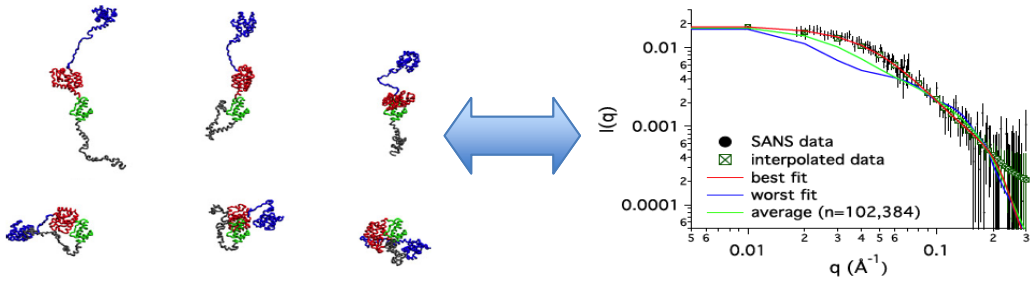
Rapid configuration space coverage

Ask/Answer “What if” questions

Prepare representative ensembles for further simulation and analysis

100 ns to > ms

A solution: SASSIE (part-time boot-strap N=0.2)



Structure & Dynamics

Neutron Experiments

- 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)

www.smallangles.net/sassie

Joseph E. Curtis *et al,* SASSIE: A program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints. *Computer Physics Communications*, 2012, 183(2), 382–389

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.

http://genapp.rocks

Scientist "has code"	"writes text file" to define GUI/link code to GenApp Then "compiles"	Done has web-app, HPC links, and GUI

1. E. H. Brookes, N. Anjum, J. E. Curtis, S. Marru, R. Singh, M. Pierce. GenApp module execution and airavata integration. Proceedings of the 9th Gateway Computing Environments Workshop, SC14, Nov 16-21, 2014. New Orleans, USA. IEEE Press, Piscataway, NJ, USA 2014. ISBN 978-1-4799-7030-8.

1. Brookes, E.H. 2014. An Open Extensible Multi-Target Application Generation Tool for Simple Rapid Deployment of Multi-Scale Scientific Codes. XSEDE '14. ACM DOI=10.1145/2616498.2616560

"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.

http://genapp.rocks

sassie-web

SCT

US-SOMO

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)

Google Summer of Code 2015

Let's build a community

ccp-sas: core-software team

Extensible application framework

HPC (core)
Or
HPC (gateway)

Support browser

User access benefits

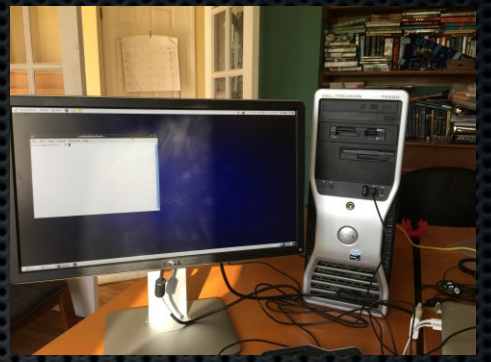
Server we control
"instant legacy"

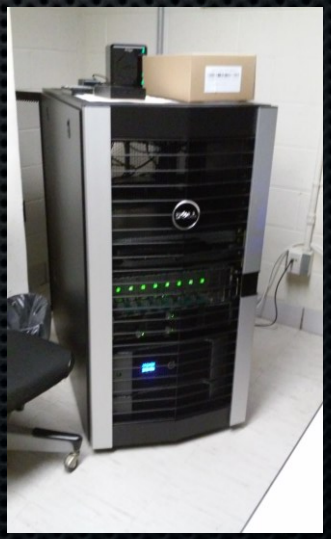
Add HPC gateway
based on demand

Code is run on server
not user's laptop/
tablet

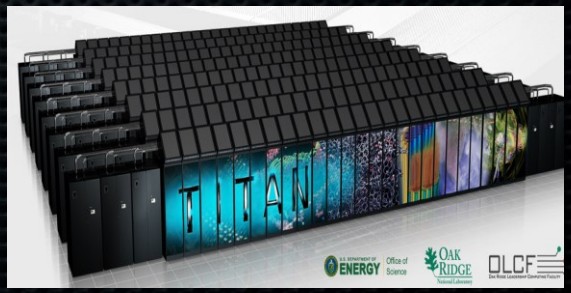
Modularity: use what you need (in & out)

HPC is relative





Enable advanced MD & APBS
XSEDE (IU) &
Langan & Smith (ORNL)



sassie-web -- Simulate

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

Rigid body

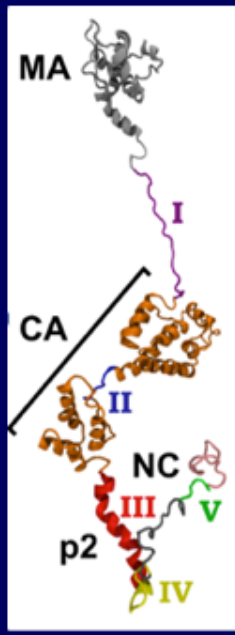
Monomer Monte Carlo

run name	run_0
reference pdb	Choose File hiv1_gag.pdb or Browse server *Local*: hiv1_gag.pdb
output file name (dcd)	hiv1_gag_monte_carlo.d
number of trial attempts	10000
return to previous structure	20
temperature (K)	300.0
molecule type	protein
number of flexible regions to vary	5
maximum angle sampled for each region	30.0,30.0,30.0,30.0,30.0
residue range for each flexible region	123-144,277-282,354-374,378-389,408-412
structure alignment: low residue	284
structure alignment: high residue	350
overlap basis	heavy atoms

Advanced Input

Check Box for Advanced Input

Submit Reset to default values



Domain	Flexible Region	Residues
MA		1 - 122
linker	I	123 - 144
CA		145 - 276
	II	277 - 282
		283 - 353
p2	III	354 - 377
linker	IV	378 - 389
NC		390 - 407
	V	408 - 412
		413 - 432

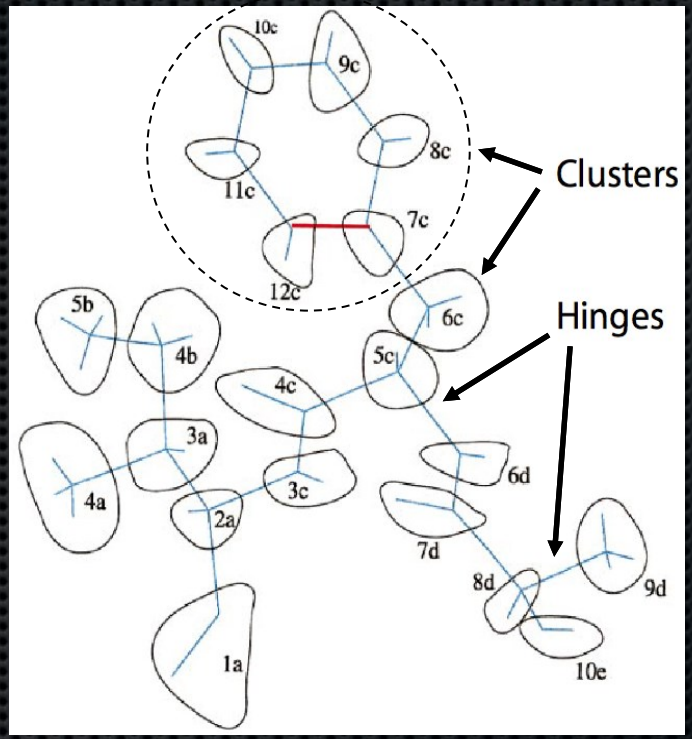
Generate Ensemble

Energy Minimize
&
Calculate I(q)

Compare to Exp.:
Chi-square Filter
Density Plot

MD

- 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

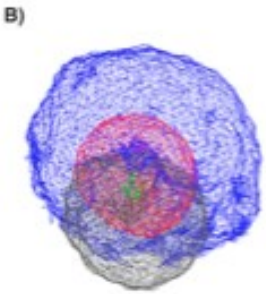
→ Jianhan Chen's poster (was Monday)

J. Chen, W. Im, C. Brookes, JCC 26, 1565 (2005), Zhang et al. (in preparation)

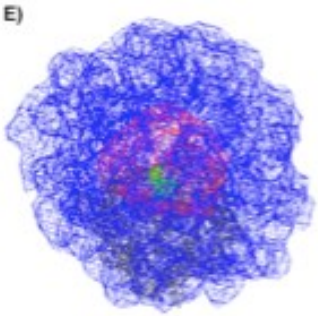
MD

TAMD : combine MC w/ fast MD

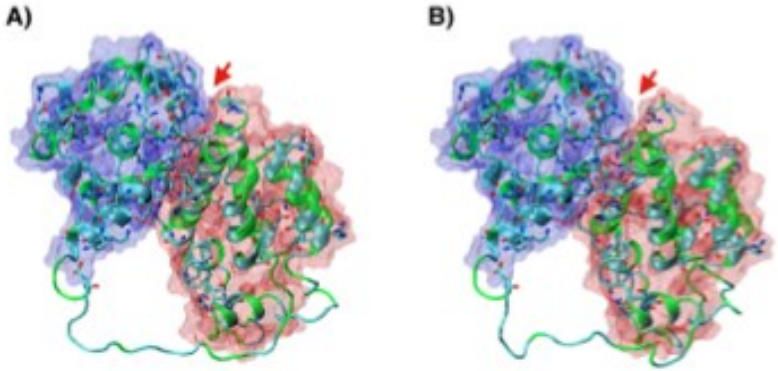
MC (50000)

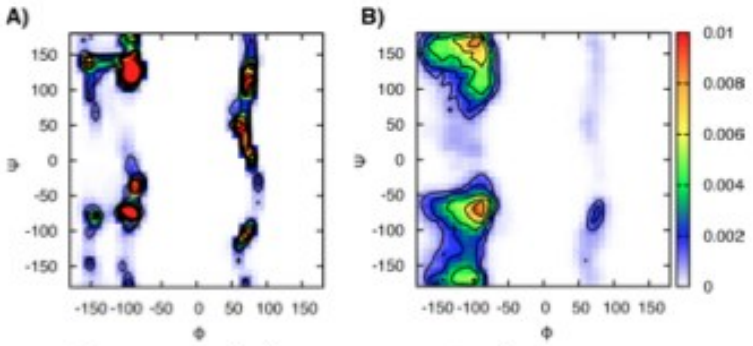


Spatial Clusters (~100)



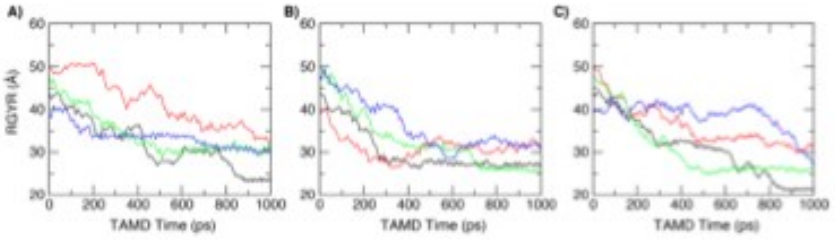
Protein, DNA, RNA, carbohydrates . . . and complexes





Zhang et al. (in preparation)

Implicit solvent effects



Constrained Rg

EFF1, SASA, GB, ACE

ALPHA -- Torsion Angle MC

Multiple, multi-chain species: ds-DNA, proteins, RNA, carbohydrates, etc.

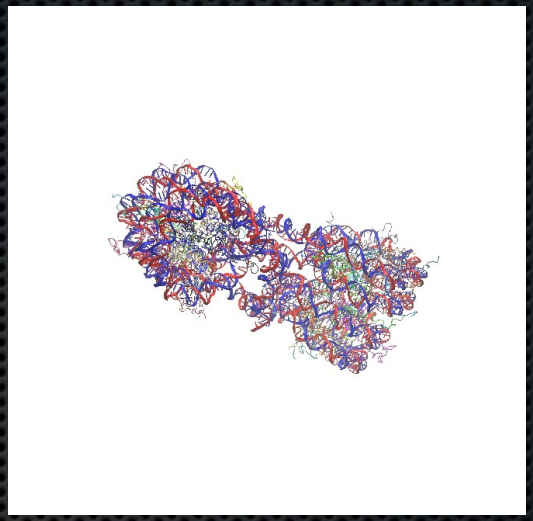
User defined torsion sampling

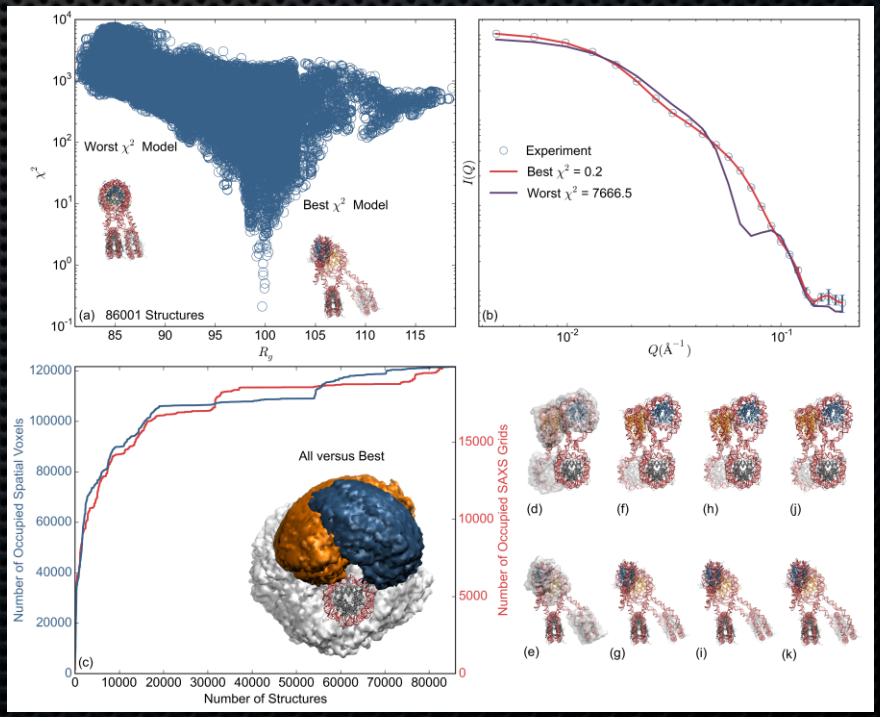
backbone, concerted motions, side-chains, carbohydrates, iso-peptide bonds, etc.

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.

sassie-web -- Analyze

Analyze -- compare theoretical data to experimental data and advanced analysis methods

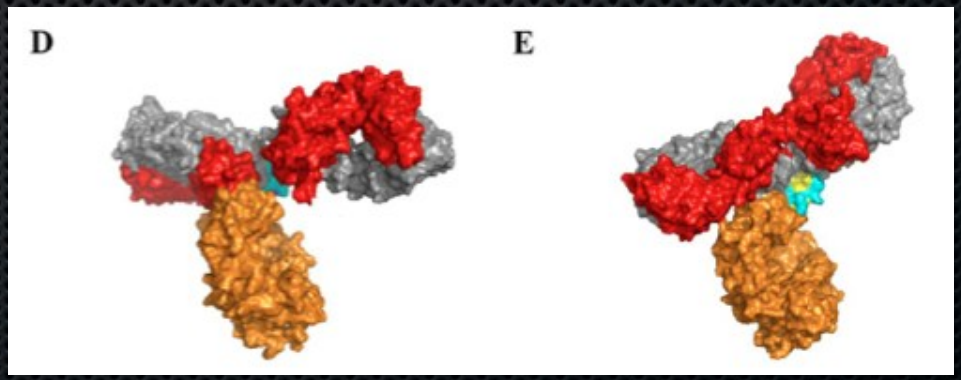
Chi-Square Filter Compares calculated scattering profiles to user supplied interpolated scattering profiles. No fitting.

SCT Analyze Compares calculated scattering profiles to user supplied interpolated scattering profiles. No fitting.

Density Plot Calculates volumetric gaussian cube files from user supplied structures. These files can be rendered in VMD and Pymol.

APBS Calculates non-polar solvation energy using adaptive Poisson-Boltzman solver from user supplied structures.

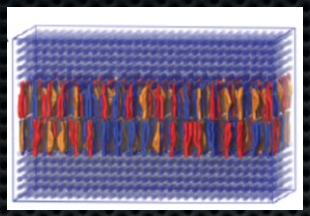
Both X^2 ~ 1

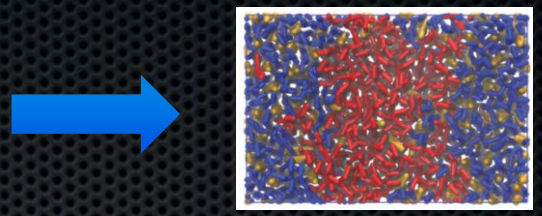


Free Energy differs by > 400 kCal/mol

Clark et al. J. Phys. Chem. B 2013, 117, 14029−14038, 2013

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

summary

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)
- Kushol Gupta (SURF now @ UCSD!)

~ 150 beta testers & students

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)

ccpsas.org