CCP-SAS Year Two Report

The fundamental goal of the CCP-SAS project is to make the modeling capability required for solving the often complex macromolecular and supramolecular structural problems, primarily arising from SAS techniques, accessible to the bench scientist who needs it. While many classes of problems can be addressed by fitting data directly to analytical models, a case addressed by previous NSF funding leading to *SasView*, the present project addresses the problem of structures lacking any symmetry and/or exploring configurational space in a way that precludes an analytical solution. These must be modeled in real space and require tools and methods such as MC and MD, to reduce the ensemble of potential structures to test to something tractable. Thus this class of problems requires not only access to atomistic or coarse grained MC or MD techniques but also to transparent HPC resources.

The grant was conceived as an international project funded jointly by EPSRC in the UK and NSF in the US bringing together three researchers developing overlapping packages using similar approaches and philosophies, *SASSIE*, *US-SOMO*, and *SCT/SCTPL*, as well as scientists with strengths in small angle scattering and molecular dynamics and four major international scattering centers, one x-ray and one neutron each in the US and the UK. The aim is to provide a solid accessible tool for real space modeling of complex macromolecular structures by building a collaborative rather than competitive environment that harnesses this pool of complementary and overlapping skills and resources in a way that could not be achieved by the individual members of the grant alone. Long-term sustainability will be achieved by providing a suite of tools and a software framework which can be maintained by a core group of international facilities and developed through facility staff contributions, power user contributions and, for major new functionality and innovations, through new funding sources which will be able to leverage and build upon the existing stable base rather than start again from scratch.

While the effort in year one was heavily weighted towards organizational activities, setting up the necessary structures and infrastructures, developing detailed plans and starting work on a number of software tools, year two focused on ramping up the project as rapidly as possible. Thus at the halfway point of the project we have a public beta web front end, released May 19, 2015, with access to a fully functional and feature rich backend of computational tools, including over 20 modules supported by a substantial set of documentation available on line, a series of on-going projects using the software along with a number of papers already out, and engagement with some of the major sources in Europe and the US with an agreement to dedicate 10% of the TITAN supercomputer at ORNL to running the CCP-SAS developed software for use by SNS and HFIR staff and users. While some years away from user activity yet the ESS has expressed strong interest in our project and asked to be kept abreast of developments and that they be invited to any future meetings and tutorials, as has the ESRF.

Other developments in year two include the second project meeting in October 2014 held at the Diamond Light Source in the UK. This three day meeting gathered not only key project members, but also a number of interested researchers from the UK and Europe along with representatives from several of the CCPs, the ESRF and the ESS. The project also hosted several webinars during the course of the year, bringing the geographically disperse team together virtually, including talks by the UK software sustainability institute, the MANTID team's plans

for integrating HPC resources into their workflow, and a two part training webinar on the use of the new web interface for the Project members including a homework assignment. In May 2015, in tandem with the public beta release, CCP-SAS was introduced in a lecture at the highly regarded EMBO course on biological scattering in Grenoble France, and was followed by a three day hands-on tutorial. The testing team and collaborators, including three summer undergraduate students, continue to help refine the software while providing a growing number of papers and a significant number of presentations either about the software or using it in a project. On the software development side, major work has gone into creating and building GenApp, the underlying framework used to develop the web application and integrate eventually with HPC resources, allowing for the timely deployment of the web application. Significant work has also gone into preparing for the integration of HPC backends including some prototyping. Finally good progress continues to be made on the chemical physics work with the full integration of Torsion Angle MD (TAMD) and the CHARM force field into the web release and completion of prototypes of the new Torsion Angle Monte Carlo (TAMC) and SASCalc. Testing and evaluation of various sampling and simulation protocols for a variety of structures continues along with the evaluation of implicit solvent models.

Thus the project has met most of its very ambitious targets and has all the major elements in place. While significant effort remains of course on making HPC access more transparent and robust as well as on the core modules and web front end, the project is now ideally positioned to focus on building a strong user community while expanding the tools and capabilities to a much broader range of chemical systems and start integrating more constraint from other techniques such as NMR, EM, etc. That said, one of the very strong feedback messages we received from the other CCP projects invited to our second annual meeting was that 4 years is not at all sufficient to build a sustainable community of practice. The consensus feedback was that achieving a sustainable level requires 8 to 10 years of strong support. Thus one of the critical activities in the next 2 years, mostly for the executive and dissemination teams, will be to build and execute a strategy for locating appropriate funding/resource opportunities needed for the project to reach a fully sustainable level while working hard to ensure a stable minimal base resource to protect the project from collapse in the event of a hiatus in support.

A summary of the goals and achievements against those goals for each specialty team outlined in the grant application is outlined below:

Core software team

- **Summary of detailed goals set out in team document:**
    - Working with the dissemination team, host workshop for developers and users
    - Web version released for general use to public
    - HPC backend prototype implemented and alpha testing by grant members
    - API for HPC configuration and implementation drafted
    - Soft Matter builder design and early prototyping
- **Accomplishments:**
    - Several tutorials were held both for collaborators (webinar May 5 and 12) and for the wider community (half day at ACNS in Knoxville in June 2014 and 3 day course post EMBO in Grenoble; May 2015)
    - Official public release (beta) of Web front end for SASSIE May 19, 2015

- Major advancements to the computational framework, GenApp, including: Project, file and job management; integration with online documentation and user feedback; and support for large file trees in the file manager.
- HPC Alpha design completed aided in part by support from Google Summer of Code. GenApp web-applications currently running on Indiana University HPC (Xsede)
- Core libraries written to link Python / C++ and GPU code for parallelization tasks in preparation for year 3 & 4.

Chemical Physics team

- **Summary of detailed goals set out in team document:**
  - Finalize the optimal sampling protocols and force field options for proteins and porting the final protocols to SASSIE
  - Further testing of available force fields for nucleic acids followed by (limited) optimization and SASSIE implementation of the best option identified
  - Explore possible atomistic and coarse-grained models for large concentrated solutions of proteins and other chemicals.
  - Identify best modelling approaches for new chemical systems which may include synthetic polymers, micelles, bicelles, surfactants, among others
- **Accomplishments:**
  - Implementation of CHARMM and Torsion Angle MD (TAMD) into beta release of web application.
  - New Torsion Angle Monte Carlo (TAMC) prototype completed. Allows simulation of far more complex biological systems included double-stranded DNA, carbohydrates, proteins and their complexes.
  - SASCalc prototype completed and ported to GPU architecture.
  - SASCalc-MD prototype completed, in collaboration with Juergen Koefinger @ Max Planck, for SAS calculations from fully atomistic MD simulations with explicit solvent.
  - Evaluation and testing of various implicit solvent models and of simulation and sampling protocols and force field options for a variety of bio-chemical structures
  - Evaluation and development of new multi scale sampling techniques.

Dissemination team

- **Summary of detailed goals set out in team document:**
  - Start training the trainers (CCP-SAS collaborator & stakeholders, Facility scientists).
  - Start staging CCP-SAS Tutorial Workshops for the wider Community.
  - Begin implementing mechanisms for capturing CCP-SAS User Community feedback.
  - Begin implementing mechanisms for capturing CCP-SAS usage & publication metrics.
  - Continue engaging with European and North American facilities.
- **Accomplishments:**

- A UCL CCP-SAS project was the winner of best oral presentation award at the XXVth International Complement Workshop, Rio de Janeiro, Brazil, Sept 14-18, 2014.
- Held successful Year 2 project meeting in UK (October 2014) with interested European scientists along with representatives from CCPs, the ESS and the ESRF
- Visited ESS, ILL and ESRF and gave presentations about CCP-SAS and discussed possible common interests. The ESS has asked to be invited to all future meetings, workshops and tutorials while the ESRF is discussing contributing a module or two to SASSIE.
- Engaged with ORNL and have agreement to deploy on TITAN for SNS staff and users.
- Several tutorials were held both for collaborators (webinar May 5 and 12, 2015) and for the wider community (half day at ACNS in Knoxville June 1, 2014 and 3 day course post EMBO in Grenoble; May26-28, 2015)
- Numerous CCP-SAS related presentation and papers are being collected and documented on the website including a recent PNAS article published in collaboration with NIST. These presentations are a key method for broadly disseminating the activities of the project and the capabilities of the software.

Testing team

- **Summary of standing goals set out in team document:**
  - Identify potential projects to test the software and provide feedback as well as serve as examples of use to the larger community once completed and published.
  - Track which projects align when with the current status of the evolving functionality of the software.
  - Ensure communication between project lead and software development and training teams as necessary.
  - Maintain list of projects and their status.
  - Identify and help resolve bottlenecks in projects.
- **Accomplishments:**
  - A distinction between formal testing projects and general collaborative ones was made at the second annual CCP-SAS project meeting held at the Diamond Light Source in October 2014.
  - A Protocol for defining a Testing Project has been written and a master list of testing projects is being maintained. Currently the first UK testing project has been submitted for publication and 5 more are identified as currently active. A protocol for better identifying US testing projects as opposed to collaborative ones is still needed.
  - US and UK projects were presented and discussed at the joint CCP-SAS meeting in Oct 2014 at the Diamond Light Source.
  - Six abstracts were submitted to the major Berlin SAS meeting in Sept 2015 including a general CCP-SAS talk as well as several testing project presentations.

Outputs and other outcomes:

- Webinars
  - D. W. Wright. Training the Trainers - Part II. 12-May-15

- J. E. Curtis. Training the Trainers - Part I. 05-May-15
- E. H.Brookes. Gateways Application Generation with GenApp. XSEDE Weekly Gateways Seminar. 01-May-15
- N. Draper. Mantid HPC Challenges: The Challenges of delivering flexible HPC for novice end users. 12-Nov-14
- N. Hong. Introduction to the Software Sustainability Institute. 13-Aug-14
- R. Rambo. BioISIS. 16-Jul-14
- J. E. Curtis/E. H. Brookes. The current state of CCP-SAS development. 18-Jun-14

- Outreach
  - 3 summer undergraduates, two from under-represented groups, participated in testing projects over the summer.
- Workshops
  - J. E. Curtis, E. H. Brookes, D. W. Wright, H. Zhang *CCP-SAS 2015: Atomistic Modeling for Small-Angle Scattering: A Short Course*, May 26-28, 2015, Institut Laue-Langevin, Grenoble, France.
  - 2nd Joint US-UK CCP-SAS Project Meeting, Oct 6-7, 2014. Diamond Light Source, UK.
  - J. E. Curtis, H. Zhang. *CCP-SAS tutorial*. ACNS 2014, June 1, 2014. Knoxville, USA.
- Presentations (prior years available on website or previous reports):
  - J. E. Curtis. *SASSIE: A framework for multi-scale modeling of scattering data*, EMBO Practical Course, Small angle neutron and X-ray scattering from proteins in solution, May 19, 2015. Grenoble France. **TALK**
  - E. H Brookes. *An Open Extensible Application Generation Tool for Simple Rapid Deployment of Multi-Scale Scientific Codes*. University of Texas at San Antonio, Computer Science Department, 2015. San Antonio, Texas. **TALK**
  - E. H Brookes. *Gateways Application Generation with GenApp*. XSEDE Weekly Gateways Seminar. Online. (https://www.youtube.com/watch?v=AEyWr7_8qBE) 2015. **TALK**
  - K. H. Lee & J. Chen, *Refining Multi-scale Enhanced Sampling for Simulating Disordered Protein Conformations*, 59th Biophysical Society Annual Meeting, Feb 7-11, 2015, Baltimore, USA. **POSTER**
  - E. H. Brookes. *An Open Extensible Multi-Target Application Generation Tool for Simple Rapid Deployment of Multi-Scale Scientific Codes*. 2014 Annual Conference on Extreme Science and Engineering Discovery Environment, 2014. Atlanta, Georgia. **TALK**
  - E. H. Brookes, N. Anjum, J. E. Curtis, S. Marru, R. Singh, M. Pierce. *GenApp module execution and Airavata integration*. 9th Gateway Computing Environments Workshop, 2014. New Orleans, Louisiana. **TALK**
  - E. H. Brookes, M. Rocco. *UltraScan-SOMO (US-SOMO): An Integrated Hydrodynamic Calculation and Small Angle Scattering Data Analysis Software Suite*. 3rd S4SAS User Meeting and workshop, 2014. Diamond Light Source, Oxfordshire, United Kingdom. **TALK**
  - D. W. Wright, R. Nan, G.-K. Hui, J. E. Curtis, E. H. Brookes, S. J. Perkins. *CCP-SAS - Novel approaches for the atomistic modelling of small angle scattering data in biology.* Biophysical Meeting. Feb 7-11, 2015. Baltimore MD, USA. **TALK**
  - D. W. Wright, R. Nan, G.-K. Hui, J. E. Curtis, E. H. Brookes, S. J. Perkins. *CCP-SAS - Novel approaches for the atomistic modelling of small angle scattering data in biology*. CCP-Biosim Meeting. Jan 7-9, 2015. Leeds, UK. **TALK**
  - J. Chen, *Orderly Chaos of Proteins*. University of Kansas Medical Center, Biochemistry and Molecular Biology departmental seminar, Dec 19, 2014, Kansas City, USA. **TALK**
  - P. D. Butler. *CCP-SAS – a new Collaborative Computational Project for the atomistic and coarse grained modelling of SANS and SAXS data*. Joint PSB/COLLEGE 8 Seminar, Thursday, October 16th, 2014. Grenoble, France. **TALK**
  - J. Chen, *Multi-Scale Enhanced Sampling of Protein Structure and Interaction*. Shanghai Jiao Tong

University, Department of Physics seminar, Sept 22, 2014, Shanghai, China. **TALK**
- J. Chen, *Multi-Scale Enhanced Sampling of Protein Structure and Interaction*. Xiamen University, Department of Physics seminar, Sept 14, 2014, Xiamen, China. **TALK**
- E. Rodriguez, R. Nan, K. Li, J. Gor and S. J. Perkins *A revised mechanism for the activation of complement C3 to C3b and a molecular explanation of the difference between their C3S and C3F polymorphic forms*. XXV International Complement Workshop, Rio de Janeiro, Sep 14-18, 2014 **TALK - BEST PRESENTATION AWARD**
- S. J. Perkins, G.-K. Hui and L. Rayner. *Atomistic modelling of SANS and SAXS data in applications to diverse antibody classes important in biotechnology and disease*. Neutron Characterisation in Fundamental and Applied Biotechnology (NCFAB). Abingdon, UK. 3rd-5th September 2014. **TALK**
- J. Chen. *Accelerate sampling in atomistic energy landscapes using coarse-grained models for implicit solvent optimization*, 248th ACS National Meeting, Aug 10-14, 2014, San Francisco, USA. **TALK**
- S. J.Perkins. *CCP-SAS – a new community consortium for the atomistic modelling of SANS and SAXS data*. British Biophysical Society Biennial Meeting 2014, Jul 9-11, 2014. Warwick, UK. **TALK**
- R. Nan, C. M. Furze, D. W. Wright, R. Wallis, and S. J. Perkins. *The solution structures of MASP (mannose-binding lectin-associated serine protease) and MBL (mannose-binding lectin) provides insight on the activation of the lectin pathway of complement*. British Biophysical Society 2014 Biennial Meeting, Warwick University, 9-11th July 2014. **TALK**
- E. H. Brookes. *Optimizing utilizable information from small angle solution scattering of biological macromolecules*. British Biophysical Society Biennial Meeting 2014, Jul 9-11, 2014. Warwick, UK. **TALK**
- J. Chen. *Accelerate sampling in atomistic energy landscapes using topology-based coarse-grained models*, Telluride Workshop on Advances in Enhanced Sampling Algorithms, June 30 – July 4, 2014, Telluride, USA. **TALK**
- J. E. Curtis & H. Zhang. *SASCALC -- A fast and accurate small-angle scattering calculator for atomistic ensembles*. ACA 2014, May 24-28, 2014, Albuquerque, NM,. **TALK**
- P. Butler & J. E. Curtis. *CCP-SAS: A Collaborative Computational Project for Small Angle Scattering Modeling.* ACNS 2014, June 1-5, 2014. Knoxville, USA. **POSTER**

- Publications using CCP-SAS or components being integrated into CCP-SAS (prior years available on website or previous reports):
  - K. H. Lee and J. Chen. Multiscale Enhanced Sampling of Intrinsically Disordered Protein Conformations. *Journal of Computational Chemistry*. In Press (2015).
  - W. Zhang and J. Chen (2014). Replica exchange with guided annealing for accelerated sampling of disordered protein conformations. *Journal of Computational Chemistry*. **35** (23), 1682 (2014). DOI: 10.1002/jcc.23675
  - E. H. Brookes, N. Anjum, J. E. Curtis, S. Marru, R. Singh, M. Pierce. The GenApp framework integrated with Airavata for managed compute resource submissions. *Concurrency and Computation: Practice and Experience*. (2015). DOI: 10.1002/cpe.3519.
  - L. E. Rayner, G. K. Hui, J. Gor, R. K. Heenan, P. A. Dalby, S. J. Perkins. The solution structures of two human IgG1 antibodies show conformational stability and accommodate their C1q and FcγR ligands. J. Biol. Chem. 290, 8420-8438 (2015) doi:10.1074/jbc.M114.631002.
  - H. Zhao, R. Ghirlando, C. Alfonso, F. Arisaka, I. Attali, D. L. Bain, M. M. Bakhtina, D. F. Becker, G. J. Bedwell, A. Bekdemir, T. M. D. Besong, C. Birck, C. A. Brautigam, W. Brennerman, O. Byron, A. Bzowska, J. B. Chaires, C. T. Chaton, H. Cölfen, K. D. Connaghan, K. A. Crowley, U. Curth, T. Daviter, W. L. Dean, A. I. Díez, C. Ebel, D. M. Eckert, L. E. Eisele, E. Eisenstein, P. England, C. Escalante, J. A. Fagan, R. Fairman, R. M. Finn, W. Fischle, J. García de la Torre, J. Gor, H. Gustafsson, D. Hall, S. E. Harding, J. G. Hernández-Cifre, A. B. Herr, E. E. Howell, R. S. Isaac, S.-C Jao, D. Jose, S.-J Kim, B. Kokona, J. A. Kornblatt, D. Kosek, E. Krayukhina, D. Krzizike, E. A. Kusznir, H. Kwon, A. Larson, T. M. Laue, A. Le Roy, A. P. Leech, H. Lilie, K. Luger, J. R. Luque-Ortega, J. Ma, C. A. May, E. L. Maynard, A. Modrak-Wojcik, Y.-F Mok, N. Mücke, L. Nagel-Steger, G. J. Narlikar, M. Noda, A. Nourse, T. Obsil, C. K. Park, J.-K. Park, P. D. Pawelek, E.E. Perdue, S. J. Perkins, M. A. Perugini, C. L. Peterson, M. G. Peverelli, G. Piszczek,

G. Prag, P. E. Prevelige, B. D. E. Raynal, L. Rezabkova, K. Richter, A. E. Ringel, R. Rosenberg, A. J. Rowe, A. C. Rufer, D. J. Scott, J. G. Seravalli, A. S. Solovyova, R. Song, D. Staunton, C. Stoddard, K. Stott, H. M. Strauss, W. W. Streicher, J. P. Sumida, S. G. Swygert, R. H. Szczepanowski, L. Tessmer, R. T. Toth, IV, A. Tripathy, S. Uchiyama, S. F. W. Uebel, S. Unzai, A. Vitlin-Gruber, P. H. von Hippel, C. Wandrey, S.-H Wang, S. E. Weitzel, B. Wielgus-Kutrowska, C. Wolberger, M. Wolff, E. Wright, Y.-S Wu, J. M. Wubben, & P. A. Schuck. Multilaboratory comparison of calibration accuracy and the performance of external references in analytical ultracentrifugation. PLOS One 10: e0126420 (2015) doi:10.1371/journal.pone.0126420

- D. W. Wright, S.J. Perkins. SCT: A suite of programs for comparing atomistic models to small angle scattering data. J. Appl. Crystallography 48, in press (2015). doi:10.1107/S1600576715007062.
- E. Rodriguez, R. Nan, K. Li, J. Gor, S. J. Perkins. A revised mechanism for the activation of complement C3 to C3b: a molecular explanation of a disease-associated polymorphism. J. Biol. Chem. 290, 2334-2350 (2015). doi: 10.1074/jbc.M114.605691
- Y. Peng, J. E. Curtis, X. Fang, S. A. Woodson. Structural model of an mRNA in complex with the bacterial chaperone Hfq. Proc. Nat. Acad. Sci. 111, 17134–17139 (2014).doi: 10.1073/pnas.1410114111
- J. E. Curtis, H. Zhang, H. Nanda. SLDMOL: Structural characterization of thermally disordered membrane proteins. Comp. Phys. Comm. 185, 3010-3015 (2014). doi:10.1016/j.cpc.2014.07.006
- S. Krueger, J. H. Shin, J. E. Curtis, K. A. Rubinson, Z. Kelman. The solution structure of full-length dodecameric MCM by SANS and molecular modeling. Proteins: Structure, Function, and Bioinformatics 82, 2364-2374 (2014). doi: 10.1002/prot.24598
- Z. Zhang, J.-M.Y. Carrillo, S.-K. Ahn, B. Wu, K. Hong, G.S. Smith, C. Do. Atomistic Structure of Bottlebrush Polymers: Simulations and Neutron Scattering Studies. Macromolecules 47, 5808 (2014). DOI: 10.1021/ma500613c
- L.E. Rayner, G.K Hui, J. Gor, R.K. Heenan, P.A. Dalby, S.J. Perkins. The Fab conformations in the solution structure of human IgG4 restricts access to its Fc region: implications for functional activity. J. Biol. Chem. 289, 20740-20756 (2014).doi:10.1074/jbc.M114.572404
- N. J. Clark, M. Raththagala, N. T. Wright, E. A. Buenger, J. F. Schildbach, S. Krueger, J. E. Curtis. Structures of TraI in solution. J. Mol. Model. 20, 2308 (2014). doi: 10.1007/s00894-014-2308-3
- M. C. Watson and J. E. Curtis. Probing the average local structure of biomolecules using small-angle scattering and scaling laws. Biophys. J. 106 (11), 2474-2482 (2014).doi: 10.1016/j.bpj.2014.03.050