CCP-SAS Year One 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 3 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.

The primary objectives of year one were to establish the underlying organizational structure required to successfully manage such a diverse and geographically disperse collaboration, move the current and emerging tools into an accessible and coherent alpha framework with a web front end and start laying the foundation for the critical testing and dissemination activities. As initially outlined in the grant, an executive team was formed with overall coordination responsibilities, along with four specific teams covering the various aspects of the project: the core software team, the chemical physics team, the testing team and the dissemination team. Note that from the outset, the importance of long term sustainability is addressed by focusing half the teams on specific architecture and algorithm problems with the other half on testing, engagement and dissemination activities.

As the end of year one approaches, the project is on track to meet its goals. Each team has prepared a document outlining the scope and goals of the team. The international executive team holds regular meetings, the UK group holds quarterly face to face meetings, and the teams and individuals have established regular communication channels. A branded web presence was established and the highly successful kick-off meeting aided the early engagement with other key facilities. The small dedicated testing cluster was purchased and installed at UTK in Tennessee and the alpha web framework is nearly complete and installed on that cluster, while a server was purchased and installed at UCL in the UK. Significant progress towards expanding existing

capabilities beyond protein molecules has been made. A number of new capabilities have been added through integration of existing packages or development of new methods and packages and new algorithms and best approaches to a number of key new features have been researched. Aside from a number of ongoing external collaborations, eight testing projects, internal to the grant, have been identified and some are underway. Several presentations, tutorials and training sessions have already been given or are planned and a number of papers acknowledging the project have been published or submitted. Finally the project has attracted the attention of some potential sources of new complementary funding. A more detailed summary of goals and accomplishments against those goals is outlined below:

Start-up organization of the project

- **Primary Goals:** Have first grant wide kickoff meeting in US. Get organizational structure up and running. Refine and elaborate Year 1 goals. Hire two full-time postdocs, one at UCL and one at UTK, a half-time postdoc at UTHSCSA, and a Graduate Student at KSU.
- **Accomplishments:**
    - The executive team was established with the UK and US PI plus Joseph Curtis from NIST in US and Steve King from ISIS in the UK. Regular web conference meetings are held roughly fortnightly. Also, the UK and US PIs have exchanged visits outside of kickoff.
    - Internal mailing lists and the four specialty teams were established and team documents with milestones were developed. As provided in the grant each team elaborated the goals and milestones provided in the initial grant application. This forms the basis against which progress is monitored.
    - Quarterly face to face meetings brings together the UK members of the grant, while teams and individuals have established regular communication channels.
    - A highly successful kickoff meeting was held at NIST Feb 6-8, 2014.
    - The UCL hire is in place as is the KSU graduate student.
    - The 3 summer interns for start of year 2 (two at BioCat one at NIST) have been identified.

Each specialty team outlined in the grant application produced a document outlining the scope of the activities overseen by the team along with a specific set of goals and milestones for year 1. Goals and milestones for future years were included at a coarse grain level as a guide for the future but with the clear intent that at the beginning of each year that year's goals and milestones will be revisited and elaborated.

Core software team

- **Summary of detailed goals set out in team document:**
    - SCT/SCTPL refactored and ready to be a module for CCP-SAS. SASSIE modules installed in UK. Identify US-SOMO modules to be ported to CCP-SAS web framework. Release API for SASMOL package.
    - Purchase and install cluster at UTK and develop and deploy web prototype application. Install a server at UCL. Begin alpha testing with grant members.

Explore mechanism for access to UK HPC resources and global applicability of gateway solutions.
    - Investigate requirements to extend to more general chemical system problems including soft matter builder.
- **Accomplishments:**
    - SCT/SCTPL refactored and provided as open source in Feb 2014. SASSIE modules installed in UK PI lab. New Golden-Vector method implemented.
    - Plugin Framework concept, GenApp, developed and first prototype successfully used to wrap SASSIE modules with web front end.
    - 128 CPU and 20,000GPU system installed at UTK and web prototype using GenApp framework deployed. UCL server also installed.
    - Hybrid atomistic/affine modeling and coarse grain sampling methods developed towards supporting more general chemical problems.

Chemical Physics team

- **Summary of detailed goals set out in team document:**
    - Implement CCP-SAS interface to CHARMM and port TAMD (torsional angle molecular dynamics) to CCP-SAS.
    - Test various atomistic implicit solvent models and simulation protocols for proteins. Initiate testing of sampling protocols and force field options for nucleic acids and glycosylated proteins.
    - Identify computationally tractable (non-bio) chemical problems.
- **Accomplishments:**
    - CHARMM and TAMD now available to CCP-SAS framework.
    - Initial testing of sampling protocols and force field options for nucleic acids and glycosylated proteins completed and testing of solvent models is underway.
    - Several new tools have been implemented such as contrast calculator and HPLC-SAXS.
    - Lipid problems (particularly nanodiscs and vesicles) were identified as a first step to bridging between the needs of biological macromolecular problems and more general colloidal chemical problems.

Dissemination team

- **Summary of detailed goals set out in team document:**
    - Establish a branded web presence linked from appropriate other sites, including a project web page, user mailing lists, rss feeds etc. as appropriate.
    - Identify and begin to engage other facilities as potential partners.
    - Start creating Project level materials to be used as needed by any member of the project including first project highlight suitable for circulation to funding agencies.
    - Identify/explore technologies for dissemination activities such as video demos/course, capturing outputs from testing teams, capturing feedback from eventual external users.
    - Identify meetings to target and create living list of presentations.
- **Accomplishments:**

- CCP-SAS branded website established, *www.ccpsas.org*, and linked from several other sites.
- SNS, ESS, ILL, and other sectors of APS engaged. SNS, ILL, and APS representatives attended kickoff meeting. ESS invited to 2nd annual meeting in UK.
- Project acknowledgment slide and Poster was created. On target to prepare 1st highlight by July 1 – mid-point between the UK and US start date anniversaries.
- Technology tools exploration is in progress.
- Training courses are planned in the UK for UK participants and a tutorial is scheduled for ACNS in US June 1. Several other tutorials and lectures in planning stages for year 2. An invited presentation at Neutrons in Biology and Biotechnology (NIBB) in Grenoble France, introduced project. A project wide list of presentations and potential meetings to target was started.

Testing team

- **Summary of detailed goals set out in team document:**
  - Define expectations of formal testing projects vs. general collaborative ones.
  - Construct first list of formal testing projects and have at least one started.
  - Identify technologies to use in tracking projects, getting feedback capturing documentation outputs, and providing forums for sharing experiences.
- **Accomplishments:**
  - Expectations of outputs are defined in testing document and 8 specific testing team projects were identified. Several projects have begun.
  - Discussions are ongoing to identify appropriate useful technologies.

Outputs and other outcomes:

- Invited to join the CCP steering committee in UK - CCP-SAS now a member
- Attracting complementary funding
  - Invited to submit new EPSRC grant application to fund computational projects allied to a current CCP project.
  - The community project "Apache Airavata" expressed great interest in the framework and is mentoring a Google funded "Google Summer of Code" student to integrate Airavata to the framework.
- Code committed to open repository
  - SCT/SCTPL now available on github at https://github.com/dww100/sct
- Workshops
  - J. Perez, M. Rocco, E. Brookes. BioSAXS Data Analysis within US-SOMO. Full session tutorial. ACA 2014, May 24-28 2014.
- Presentations:
  - S. Perkins, D. Wright, J. Curtis. *CCP-SAS – a new community consortium for the atomistic modelling of SANS and SAXS data*. Neutrons in Biology and Biotechnology, February 19-21, 2014. Grenoble, France **TALK**
  - P. Butler. *CCP-SAS: A Collaborative Computational Project for Small Angle Scattering Modeling*. SI2 PI workshop, Feb 24-25, 2014. Arlington VA. **POSTER**
  - E. Brookes. *Progress in Parsimonious Spatial Modeling of Biological SAS Experimental Data.* ACA 2014, May 24-28, 2014. **TALK**
  - P. Vachette, J. Perez, M. Rocco, E. Brookes. *New developments in the UltraScan SOlution MOdeler (US-SOMO) HPLC-SAXS data analysis module*. ACA 2014, May 24-28, 2014. **TALK**

- Publications using CCP-SAS or components being integrated into CCP-SAS
    - E. J. Yearly, P. D. Godfrin, T. Perevozchikova, H. Zhang, P. Falus, L. Porcar, M. Nagao, J. E.Curtis, P. Gawande, R. Taing, I. E. Zarraga, N. J. Wagner, Y. Liu. Observation of small cluster formation in concentrated monoclonal antibody solutions and its implications to solution viscosity. *Biophys. J.* 106, 1763-1770 (2014). **highlighted in C&E News.**
    - S. J. Perkins, K.-W. Fung, S. Khan. Molecular interactions between complement factor H and its heparin and heparan sulphate ligands. *Frontiers in Immunology.* **5**, 126, 1-14 (2014).
    - L. E. Rayner, N. Kadkhodayi-Kholghi, R. K. Heenan, J. Gor, P. A. Dalby, S. J. Perkins. The solution structure of rabbit IgG accounts for its interactions with the Fc receptor and complement C1q and its conformational stability. *J. Mol. Biol.* **425**, 506-523 (2013).
    - S. Khan, K.-W. Fung, E. Rodriguez, R. Patel, J. Gor, B. Mulloy, S. J. Perkins. The solution structure of heparan sulphate differs from that of heparin: implications for function. *J. Biol. Chem.* **288**, 27737-27751 (2013).
    - E. Brookes, J. Pérez, B. Cardinali, A. Profumo, P. Vachette and M. Rocco. Fibrinogen species as resolved by HPLC-SAXS data processing within the UltraScan Solution Modeler (US-SOMO) enhanced SAS module. *J Appl Cryst*. 46 , 1823(2013).
    - N. J. Clark, H. Zhang, S. Krueger, H. J. Lee, R. R. Ketchem, B. A. Kerwin, S. R. Kanapuram, M. J. Treuheit, A. McAuley, J. E. Curtis. Small-angle neutron scattering study of a monoclonal antibody using free-energy constraints. *J. Phys. Chem. B* 117, 14029-14038 (2013).
    - K.L. Sarachan, J. E. Curtis, S. Krueger. Small-angle scattering contrast calculator for protein and nucleic acid complexes in solution. *J. Appl. Cryst*. 46 1889-1893 (2013).
    - H. Zhang, S. Khodadadi, S. L. Fiedler, J. E. Curtis. Role of water and ions on the dynamical transition of RNA. *J. Phys. Chem. Letters* 4, 3325-3329 (2013).
    - M.C. Watson and J. E. Curtis. Rapid and accurate calculation of small-angle scattering profiles using the golden ratio. *J. Appl. Cryst.* 46, 1171-1177 (2013).
    - Several others accepted for publication.

Challenges encountered and mitigation strategy implemented

There were surprisingly few unanticipated challenges for such a large project. The first such challenge was the fact that the two month gap between the startup dates of the US-NSF (June 1) and UK-EPSRC (Aug 1) sides of the grant, coupled with the temporary shutdown of the US government leading to a 4 month delay of the kick-off meeting, put that meeting in month 8 of the US-NSF side of the grant. That delay was addressed by using GoToMeeting heavily and proactively to start the collaboration, even before the UK grant officially started, and by having the UK PI and postdoc visit the US PI lab independently of the kick-off meeting. Further the extra time was effectively used to produce the best possible meeting and helped achieve the early positive engagement with other facilities. The second challenge was the inevitable startup hiring issue. While this is not applicable to UK-EPSRC side whose start date was based on their hire date, it was particularly acute for the UTK portion of the grant where an ideal candidate was identified but was then unable to start before June 9, 2014 (start of year 2 for the NSF part of the grant). To help offset this delay the NIST partner took on board more effort than expected upfront. The delay in hiring at UTHSCSA on the other hand is the result of careful consideration and a decision that hiring a full time person sometime in year two would be more productive overall than the half time person for 4 years originally envisaged.