CCP-SAS Year Three Report

The fundamental goal of the currently funded CCP-SAS project is to make the modeling capability required for solving the often complex macromolecular and supramolecular structural problems, primarily arising from small angle scattering (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 Monte Carlo (MC) and molecular dynamics (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 physics based 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.

The effort in Year One was heavily weighted towards organizational activities, setting up the necessary 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 at the end of which, at the halfway point of the project, we released a public beta of *SASSIE-web* with access to a fully functional and feature rich set of computational tools, running on a dedicated compute cluster backend, including over 20 modules supported by a substantial set of documentation available on line.

In Year Three the focus turned to completing the major computational functionality and responding to user feedback, culminating in the first full (non-beta) release of *SASSIE-web*, and to adjusting to the lessons learned from, and the realities of, the first two years. In particular, the reality identified last year that a four year project is not sufficient to build a fully self-sustaining community activity, has led us to start developing plans for interim and long-term strategies that include complementary and follow-on funding from various sources.

As part of that effort, the Tuesday session of the 4th CCP-SAS Joint Project Meeting held May 22-25, 2016, brought together a number of participants interested in a variety of SAS-related software issues, including some outside the scope of the current CCP-SAS effort, to begin a conversation on the common need to deal more globally with limited funding that too often translates into no resources for the long-term maintenance required for sustainability of projects. Topics ranged from exploring more proactive and efficient collaborative models to new funding paradigms.

It has also become clear that the technological needs of soft matter, as distinct from those of biological macromolecules, far exceeds the capability of the limited resources of this grant to provide any easily accessible user friendly tools. This was addressed on the first day (Sunday) at the May meeting, where the steering group (PIs) unanimously agreed that for the remainder of the current funding the focus should be squarely on delivering robust and user friendly software; ie, doing what we do now well, rather than trying to do too much poorly. This in fact also folds directly into the longer term support strategy by moving the bulk of soft matter innovations and aspirations to a future effort that in turn looks for its own supporting/follow on funding. Further, there was a significant interest in developing soft matter tools from several of the participants at Tuesday's session (discussed below) which also served as an opening to start discussing opportunities for soft matter SAS modeling. Of particular note was the keen interest from a UK industrial consortium that sees synergies in the CCP-SAS effort as a means to bridge between their own simulation efforts (mostly using DPD) and their (expensive) SAXS/SANS experiments.

On the other hand, the fact that a web front end with transparent access to HPC backend was provided through the development of a more general software infrastructure, *GenApp*, provided us the opportunity to rethink the goals of CCP-SAS to more broadly support analysis tools in the SAS community. Thus at the end of Year Three, the project is working with 6 other SAS projects, 4 of which already have a beta version of their software powered by CCP-SAS's *GenApp* and hosted on the CCP-SAS dedicated compute cluster. This is in keeping with the spirit of CCPs in general and forms the start of an attempt to consolidate long term support along the themes expressed at the Tuesday session: namely lowering long term costs by collaborating at multiple levels, not just on narrow identical projects.

Besides culminating in the full release of *SASSIE-web* and the highly successful 4th (3rd in the series of major open workshops) Joint Project Meeting, major accomplishments in Year Three include: the successful test deployment of *GenApp* and *SASSIE-web* on NSF's XSEDE Jetstream and Amazon Web Services (AWS); significant enhancements and additions to the *GenApp* and *SASSIE-web* codebase based on feedback from ~150 beta testers as well as on the original needs of the project; the logging of nearly 17,000 CPU hours on our dedicated compute cluster using the publically available beta version of *SASSIE-web* by 273 registered users; the implementation of the new free version of charmm into *SASSIE* leaving only one closed source module soon to be retired; the release of a general Contrast Calculator module in *SASSIE-web*; and the development of a new prototype module (PDB-Rx) that promises to automatically fix many problems in protein PDB files. Meanwhile community engagement at the *SASSIE* level included working with the Fushman group at the University of Maryland to bring their NMR constraints to *SASSIE* and making the CAPRIQORN code from the Hummer group at the Max Planck Institute for Biophysics in Frankfurt Germany available through *SASSIE-web*. Finally, the last day of the May meeting was our first developer hands-on tutorial attended by over a half dozen people learning how to get some computational code wrapped as a web application using *GenApp*.

Other developments in Year Three include the small (project team members only) interim Project Meeting held in conjunction with the triennial international small angle scattering

conference (an IUCr SAS commission sponsored event with ~500 participants) in Berlin in September 2015. The project took the opportunity of this major event to showcase the progress of the project through several talks, posters and a well-attended demonstration. Meanwhile, *SASSIE* continues to develop with modules graduating out of beta and a number of new modules in various stages of development such a SAS-DOCK and SASCALC-MD, while the already solid documentation continues to expand. The dedicated compute server at UTK was significantly upgraded this year, for example expanding the data space from 2TB to 24TB and implementing better backup protocols, to better serve the growing number of registered users. A number of papers and presentations along with tutorials were given as documented on the ccpsas.org website. Three summer undergraduate students worked as beta testers on projects using *SASSIE*, one of which was specifically aimed at better understanding soft matter needs, while one High School intern contributed to the code base. Reports of their efforts are available on the ccpsas.org website. Finally, testing of various sampling protocols and atomistic force field options for proteins was completed with further refinement of sampling protocols and force field options for nucleic acids and glycosylated proteins as well as evaluation of various atomistic implicit solvent models remains ongoing.

On the staffing front, the UTK postdoc took another position at short notice 4 months into Year Three while the UK Postdoc took another position 8 months in. Thus from the point of view of personnel directly supported under this grant, the project was effectively half-staffed during this period. Both UTK and the UK have now hired replacements and the UK side has asked and received a 3 month no-cost extension to mitigate the lost time, while UTK expects to request a 9 month to 1 year no-cost extension. Given the offset in starting dates between the US and UK funding this would mean that significant effort will remain available on both sides of the Atlantic for 5 months beyond that originally envisaged. Nonetheless the hiatus was unfortunate and only serves to demonstrate that talented software developers are difficult to retain in competitive economies.

Thus overall, despite the delays noted above, the project continues to meet and even exceed most of its very ambitious targets with the possible exception of a soft matter builder. More importantly the project has grown and matured in its scope and vision in a way that could embrace a much larger community and which positions it well to eventually transcend the lifetime of a particular grant. The unfortunate delay may actually allow us take advantage of a number of very recent simulation developments relevant to soft matter, as well as some new projects in the field, and thus provide significant soft matter capabilities after all. And, at the same time, provide a bit more of a bridge for transitioning to follow-on funding sources. This will be particularly important as that follow on funding is likely to not come in a monolithic project funding. Successfully managing a collective, collaborative effort with diverse funding sources, each with its independent timeline on deliverables, will be quite challenging and require time to properly organize.

With the grant beyond the half way point the project chose to focus on making as much progress as possible, taking advantage of opportunities as they arose, rather than spend too much time at this point in an exercise to refine and elaborate the Year Three goals. The original goals for Year Three set out in the proposal were:

The goals are to release the HPC enabled web-accessible interface to the general public, continue alpha/beta testing of non-bonding interactions for concentrated suspensions, and soft-matter builder. Continuing development tasks include the extension to use GPU, OpenGL canvas with multiple contrasts, and the development of fast-SAS calculators.

These were more specifically laid out as:

1) Release web-interface with the addition of a transparent HPC-layer to the general public. Expand usage metrics and HPC benchmarks to sample actual user activities. 2) Design and prototyping of a transportable HPC layer to international HPC centers. 3) Design and implement HPC layer to stand-alone version of code. 4) Release GPU capable code for beta testing and general public. Continue development and re-factoring to implement scalable fast-SAS calculators. 5) Continue the development of interactive SAS OpenGL capabilities and incorporate the ability to predict SAS profiles at varying contrast. 6) Release new non-bonding MC move-sets to alpha testing. Continue to develop parameterization by comparison to scattering data. 7) Release soft-matter builder by end of year 3. 8) Yearly Assessment Meeting (US/UK plus users from X-ray and neutron scattering) at NCNR. Review progress, re-address design goals, user feedback/wish-lists, set milestones. 9) Host a developer workshop to encourage expert users to contribute to the CCP-SAS code base. 10) Report upon academic and facility feedback and implement changes (usability, documentation, training) with consortium members.

It is interesting to note that nearly four years after putting down the objectives for Year Three, outside of calling out specific technologies that may no longer be appropriate and the stretch goal of a soft matter builder, they remain appropriate and relevant and that despite the staffing issues experienced to date, the project remains remarkably on target. The developments, activities and accomplishments for Year Three are summarized below:

- **Infrastructure developments and activities:**
    - Successful test deployment of *GenApp* and *SASSIE-web* on HPC: NSF XSEDE Jet-Stream and Amazon Web Services (AWS) after integrating Apache Airavata 0.15 into *GenApp* to allow submission to managed compute resources enabling applications to transparently run on those resources.
    - *SASSIE-web* released out of beta https://sasssie-web.chem.utk.edu/sassie2
    - Qt5 and JAVA target language support added to *GenApp* for eventual support of standalone deployments.
    - Numerous new UI tools added to *GenApp* and used to improve *SASSIE-web* GUI such as an optional Splash screen, Captcha and email verification for registration validation, integrated 3rd party license mechanism, administrator role with additional user statistics, user color management, background images, welcome pages, style controls, and more.
    - New *SASSIE-web* options implemented to easily access beta features and retired sections along with a registered developer only accessible alpha area to facilitate development cycle.

- *GenApp* API published on the GenApp Trac Wiki at http://scigenapp.org/wiki.
- Major additions to *SASSIE* docs & tutorials.
- Major storage upgrade with user data space increasing from 2 to 24 terabytes & backup protocols implemented on our primary dedicated hardware at The University of Tennessee Knoxville to support the needs of the registered users.
- Numerous hardening/bug fixes for *GenApp* and *SASSIE* based on user feedback.
- JavaScript additions to JSMol provided by High School Intern Sam Blackman.

- **Chemical physics developments and activities:**
  - *SASSIE* interface to the new freely available charmm implemented and released. This leaves only one closed source module in *SASSIE* which is slated for retirement soon.
  - Torsion angle molecular dynamics (TAMD) module using the new charmm, moved to full (non beta) release in *SASSIE-web*.
  - Generic Contrast Calculator module (for general chemical systems), released in *SASSIE-web*.
  - SASCALC scattering curve calculator module (neutron contrast and X-ray, infinite dilution), moved to beta section of *SASSIE-web*
  - TAMC prototype graduated to alpha module currently in *SASSIE-web* Alpha (developer access only)
  - PDB-Rx module to automate correcting starting PDB structure files and CAPRIQORN scattering curve calculator module including explicit solvent (X-ray only infinite dilution) in development in *SASSIE-web* alpha (developer access only).
  - Development of SASCALC-MD scattering curve calculator module, including periodic boundary conditions for concentrated systems, is ongoing
  - Work begun on ALTENS/ROT-DIF module implementing NMR constraints, SAS-DOCK module and a Normal Modes module.
  - Testing of various sampling protocols and atomistic force field options for proteins was completed while further refinement of sampling protocols and force field options for nucleic acids and glycosylated proteins is in progress.
  - Evaluation of various atomistic implicit solvent models and simulation protocols for proteins continues.
  - New multi-scale, coarse-grained protein modeling and sampling protocols for potential use in the CCP-SAS software suite are now under development.
- **Community engagement, interactions and dissemination activities:**
  - Organized a four day Project Meeting May 22-25, 2016 held at NIST in Gaithersburg, MD; USA. The first day was a closed project steering group meeting while the second day was the open meeting with roughly 50 attendees where the state of the project was presented to the community and feedback solicited.
  - Hosted a special ‘Future Collaborations’ Workshop on the third day of the pProject Meeting to discuss the potential for inter project collaborations, joint

funding opportunities, maintenance and support issues, deployment support issues etc. and generally begin a conversation about developing new models for long term support and sustainability.

- Held several workshops and tutorials including a hands-on Workshop on the fourth and last day of the Project Meeting attended by about 10 participants to learn how to use *GenApp* to create web applications.
- Started engaging the larger SAS community directly by making useful laboratory based codes available via assisted wrapping with *GenApp* followed by hosting on the CCP-SAS dedicated servers. Currently 6 projects are in various stages of deployment.
  - In beta:
    - *US-SOMO* (E. Brookes - U. Texas Health Science Center, San Antonio) https://somo.chem.utk.edu/somo
    - *Denfert* (J. Perez – Soleil Synchrotron) https://somo.chem.utk.edu/denfert
    - *Will it Fit* (L. Arleth – Copenhagen University.) https://somo.chem.utk.edu/willitfit
    - *QuaFit* (F. Spinozzi - Marche Polytechnic U. and ESRF) https://somo.chem.utk.edu/quafit
  - In Development:
    - *GenFit (F. Spinozzi - Marche Polytechnic U. and ESRF)*
    - *BioMolAnalysis Suite* (A. Savelyev - U. Texas Health Science Center, San Antonio)
- Engaging the community to incorporate appropriate contributions into *SASSIE* including two of the *SASSIE* modules under development described in the chemical physics section.
- Working with the Apache foundation "Airavata" project, secured two more funded Google Summer of Code [GSoC] 2015 summer students to help integrate Aiavata into *GenApp* for job submission to managed compute resources.
- Broader community impacts: *GenApp* used in XSEDE ECSS Vortex Shedding (A. Perlstein) https://gw165.iu.xsede.org/vortexsheddingalpha
- Work with local stakeholders to deploy *SASSIE-web* on SCARF (UK/STFC) and Titan (US/ORNL) is ongoing.
- Several tutorials and workshops were held as described in the outputs section below.
- CCP-SAS presentation and papers are being collected and documented on the website. These presentations are a key method for broadly disseminating the activities of the project and the capabilities of the software.

Outputs and other outcomes:

- Usage statistics for Year Three: *SASSIE-web* beta period: June 1, 2015 to May 20, 2016
  - ~ 150 beta testers & students (feedback using google groups)
  - 273 registered users on *SASSIE-web* as of May 2016
    - 941 projects running 14492 jobs

        - Total of >16765 CPU hours
        - Most jobs by Travis Teisemann from Thomas Jefferson University whose *Structure* paper was accepted in May 2016
- Outreach
    - 2 summer undergraduates from under-represented groups participated in *SASSIE* based projects at Illinois Institute of Technology (and BioCAT) in Chicago IL over the summer.
    - 1 summer undergraduate at NIST worked on a collaborative project between UK PI K. Edler and J. Curtis from NIST on scattering from nanodiscs, a testing project aimed at better understanding soft matter needs.
    - 1 summer High School intern working at NIST worked on scripting JSMol.
- Workshops and Tutorials
    - J. E. Curtis, S. Krueger, and H. Zhang *SASSIE tutorial* presented as part of the *Small Angle Scattering: Structural Biology and Soft Matter workshop*. ACA 2015, July 25, 2015; Philadelphia, PA, USA.
    - J. E. Curtis & D. Wright *Live demonstration of SASSIE-Web*. 16th International Conference on Small-Angle Scattering, September 14, 2015; Berlin, Germany.
    - J. E. Curtis, S. Krueger *SASSIE Studies of IDPs, Nucleic Acids and Complexes.* BioCat Advanced SAXS Training Course, November 3, 2015; Advanced Photon Source, Argonne National Labs, IL, USA.
    - E. H. Brookes *UltraScan-SOMO (US-SOMO): An Integrated Hydrodynamic and Small Angle Scattering Data Analysis Software Suite*. BioCat Advanced SAXS Training Course, November 3 2015; Advanced Photon Source, Argonne National Labs, IL, USA.
    - E.H. Brookes, M. Rocco, J.E. Curtis *US-SOMO Hydrodynamic Modeling and Small Angle Scattering and SASSIE2 Web Based Framework for Atomistic Modelling to Interpret Scattering Data.* Workshop at 22nd International Conference on Analytical Ultracentrafugation December 6-8, 2015; Melbourne, Australia.
    - E. H. Brookes *Wrapping your code for deployment in GenApp*. 4th Joint UK-US CCP-SAS Project Meeting, May 25, 2016; NIST, Gaithersburg, MD, USA.
- Presentations (full list including prior years available on website or previous reports):
    - Jianhan Chen. Multi-scale modeling of IDP structure and interaction. 251st American Chemical Society National Meeting, March 13-17, 2016. San Diego, CA, USA. **TALK**
    - S. Perkins. CCP-SAS – a community consortium for the atomistic modelling of scattering data: example applications. International Conference on Molecular Recognition. Feb 1-3, 2016. Zaragoza, Spain. **TALK**
    - K. Edler. Polymer stabilized phospholipid nanodiscs. In Symposium "Self-assembled Biofunctional Nanomaterials", Pacifichem, Dec 15-20, 2015, Honolulu, USA. **TALK**
    - K. W. Fung, R. Nan, M. Swann and S. J. Perkins. Dual polarization interferometry: characterisation of complement Factor H and C4b binding orientation to surfaces. Farfield and QCM-D User Meeting. Oct 26-28, 2015. Manchester University, UK. **TALK**
    - G. Hui, D. Wright & S. Perkins. CCP-SAS – a collaborative computational project for small angle X-ray and neutron scattering in application to human IgA1 antibody. 18th BRIC Dissemination Event. Oct 21-22, 2015. Manchester, UK. . **POSTER**

- J. E. Curtis. Neutron Scattering and Simulation for Structural Biology and Biotechnology. Oct 21, 2015. Kansas State University, USA. **TALK**
- Jianhan Chen. Multi scale Simulations of IDP Structure and Interaction. 12th New England Structure Symposium, Oct 10, 2015. Storrs, CT, USA. **TALK**
- E. H. Brookes, A. Kapoor, P. Patra, S. Marru, R. Singh, M. Pierce. *GSoC 2015 student contributions to GenApp and Airavata*, 10th Gateway Computing Environments Workshop, Sep 30, 2015, Boulder, Colorado, USA**. TALK**
- J. E. Curtis. Software for Atomistic Modeling and Analysis of X-ray and Neutron Scattering Data. Sep 21, 2015. Max Planck Institute of Biophysics, Frankfurt am Main, Germany. **TALK**
- J. E. Curtis. CCP-SAS - A community consortium for the atomistic modelling of scattering data. 16th International Conference on Small-Angle Scattering, Sep 13-18, 2015. Berlin, Germany. **TALK**
- S. Perkins. The asymmetric solution structures of native and patient monomeric human IgA1 reveal new insights on IgA nephropathy. 16th International Conference on Small-Angle Scattering, Sep 13-18, 2015. Berlin, Germany. **TALK**
- Weihong Zhang, Andrew Heindel, David Wright, Joseph E. Curtis and Jianhan Chen. Advanced sampling and atomistic modeling for structural interpretation of small angle scattering. 16th International Small Angle Scattering Conference, Sep 13-18, 2015. Berlin, Germany. **TALK**
- Jianhan Chen. Towards Reliable Atomistic Simulations of Disordered Ensembles. CECAM workshop on "Intrinsically Disordered Proteins - Bringing together Physics, Computation and Biology," Aug 18-21, 2015. Zürich, Switzerland **TALK**
- H. Zhang, J. E. Curtis, E. Brookes. Massively Parallel Computation for Small Angle Scattering. 65th Annual Meeting of American Crystallography Association, Jul 28, 2015. Philadelphia, USA. **TALK**
- S. Krueger. SANS Contrast Variation Experiments on Protein Complexes with Disordered Subunits. 65th Annual Meeting of American Crystallography Association, Jul 26, 2015. Philadelphia, USA. **TALK**
- J. E. Curtis, S. Perkins, P. Butler, S. King, J. Chen, H. Zhang, D. Wright, E. Brookes. CCP-SAS: A Community Consortium for the Atomistic Modelling of Scattering Data. 65th Annual Meeting of American Crystallography Association, Jul 26, 2015. Philadelphia, USA. **TALK**

- Publications using CCP-SAS or components being integrated into CCP-SAS (full list including prior years available on website or previous reports):
  - J. Zhiguang, J. Chen. Necessity of High-Resolution for Coarse-Grained Modeling of Flexible Proteins. J. Comp. Chem., 37(18), 1725-1733, (2016).
  - E. H. Brookes, A. Kapoor, P. Patra, S. Marru, R. Singh, M. Pierce. GSoC 2015 Student contributions to GenApp and Airavata. Concurrency and Computation: Practice and Experience. 28(7), 1960, (2016).
  - M. Green, L. Hatter, E. Brookes, P. Soultanas, D. J. Scott. Defining the Intrinsically Disordered C-Terminal Domain of SSB Reveals DNA-Mediated Compaction. J. Mol. Biol., 428(2), 357–364, (2016).
  - H. Zahid, L. Miah, A.M. Lau, L. Brochard, D. Hati, T.T.T Bui, A.F. Drake, J. Gor, S.J. Perkins and L.C. McDermott. Zinc-induced oligomerisation of zinc a2 glycoprotein reveals multiple fatty acid binding sites. Biochem. J. 473, 43-54 (2016).
  - C. A. Castaneda, J. E. Curtis, S. Krueger, D. Fushman. Linkage-specific conformational ensembles of non-canonical polyubiquitin chains. Phys. Chem. Chem. Phys., 18, 5771-5788 (2016). Structure Preview article
  - K. H. Lee, J. Chen. Multiscale Enhanced Sampling of Intrinsically Disordered Protein Conformations. Journal of Computational Chemistry. 37, 550-557 (2016).
  - 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. 27(16), 4292, (2015)

- C. A. Castaneda, E. Dixon, O. Walker, A. Chaturvedi, M. A. Nakasone, J. E. Curtis, M. R. Reed, S. Krueger, T. A. Cropp, D. Fushman. Linkage via k27 bestows ubiquitin chains with unique properties among polyubiquitins. Structure 24, 1-14 (2015).
- N. R. Zaccai, C. W. Sandlin, J. T. Hoopes, J. E. Curtis, P. J. Fleming, K. G. Fleming, S. Krueger. Deuterium Labeling Together with Contrast Variation Small-Angle Neutron Scattering Suggests How Skp Captures and Releases Unfolded Outer Membrane Proteins. Methods in Enzymology. 566, 159-210, (2015).
- G. K. Hui, D. W. Wright, O. L. Vennard, L. E. Rayner, M. Pang, S. C. Yeo, J. Gor, K. Molyneux, J. Barratt & S. J. Perkins. The solution structures of native and patient monomeric human IgA1 reveal asymmetric extended structures: implications for function and IgAN disease. Biochem. J. 471, 167-185 (2015).
- B. J. Tarasevich, J. S. Philo, N. K. Maluf, S. Krueger, G. W. Buchko, G. Lin, W. J. Shaw. The leucine-rich amelogenin protein (LRAP) is primarily monomeric and unstructured in physiological solution. J. Struct. Biol. 190, 81–91 (2015).