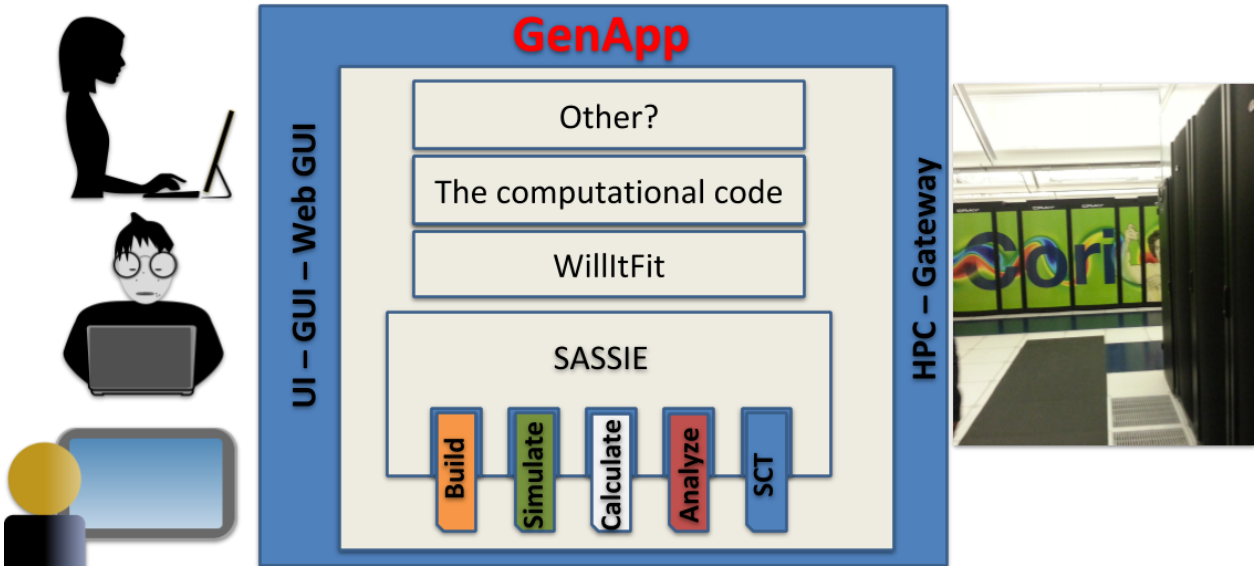
NSF OUTCOMES REPORT

Small Angle Neutron and X-ray Scattering (SANS and SAXS respectively) are powerful tools for the non-destructive probing of nano to micron scale structures, providing details on sizes, shapes, orientations and interaction in materials. To make these tools available to the scientific community there have been huge public investments in the building of the neutron and synchrotron X-ray sources needed to make such measurements which have led to advances in nearly all areas of science. However, as the techniques mature and the range and complexity of problems grows, the analysis techniques required to extract the full information content of the data are rapidly exceeding the capabilities of most experimentalists. In particular, the emerging strategy coupling measurement with advanced Molecular Dynamics (MD) and Monte Carlo (MC) simulation methods, is particularly useful for direct structural probes such as SANS and SAXS.

CCP-SAS (www.ccpsas.org), a Collaborative Computational Project, was established as an international collaboration between the US and UK (funded jointly with the EPSRC in the UK) in recognition of this bottleneck to reaping the full return on investment in instrumentation and materials. In its first five years, funded by this infrastructure grant, CCP-SAS focused on developing the necessary infrastructures and on integrating high end MD and MC methods into the analysis workflows, targeting first biomolecular sciences with an eye towards expansion to more general soft matter problems. The project brought together several groups working in parallel on similar software solutions, along with scattering facilities and other domain experts, to join forces to build a common supportable and sustainable platform moving forward that is accessible to the typical experimentalist via a well-documented, easy to use, web-based interface.

To achieve these goals, the infrastructure was designed and built in a modular and layered fashion (Fig. 1). First, a completely new deployment framework, GenApp (https://genapp.rocks/), for rapidly building and deploying graphical front ends while also transparently handling the backend interface to computational hardware was developed by the UTHSCSA portion of the grant. This outer layer allows not



only for clean separation of UI from the analysis codes, it allows for easily exposing so called “dark code”: the large amount of “home built” code (e.g. built by long gone grad students in individual labs) lacking the resources to make it easily available. Moreover, it provides a path for rapidly deploying alternative workflows that may be incompatible with the initial one. This is followed by the web based deployment of a plugin workflow framework, SASSIE-web (https://sassie-web.chem.utk.edu/sassie2/), using GenApp. This framework allows for connecting an extensible set of analysis modules into custom workflows. The innermost layer is the individual computational modules. These plugins can be simple wrappers around existing codes or written from scratch if appropriate codes do not already exist. Finally, partly as a demonstration exercise, several pieces of community developed codes were wrapped and made available on the project server at UTK. Aside from validating the concept of exposing code, it shows how the CCP-SAS umbrella can quickly start “supporting” the more general SAS community’s computational needs.

Aside from coordinating efforts and hosting the project computational server, the primary responsibility of the UTK portion of the effort, in collaboration with NIST partners, was to build and deploy the SASSIE-web framework, along with an initial suite of modules. The modules are collected into workflow categories for building starting structures, manipulating them, running simulations, calculating scattering patterns from the structures and analyzing the results against the data (Fig. 2). The design is nonrestrictive, allowing for a choice of algorithms and/or popping in and out of the workflow at any point. Thus, if algorithms not in the workflow are preferred, the outputs of those codes can be used as inputs at the appropriate place in the SASSIE-web workflow and the results exported to standard formats for further manipulation outside the workflow. There are currently 23 modules with more under development. In the last year of the grant there were nearly 200 active users running 16,000 jobs using 22,000 compute hours.

As CCP-SAS graduates from its initial funding phase, the project has successfully built a fully functional web based, modular and extensible workflow infrastructure populated with a rich set of modules particularly useful for biological macromolecules. The application is actively being deployed at various institutions in the UK and US, with the deployment on the UTK server already freely accessible to the community, while development is continuing through some new projects. Most importantly is the ongoing international collaboration that has been fostered which is seeking new partnerships and new projects to extend the capabilities.