

SI2-SSE: GenApp - A Transformative Generalized Application Cyberinfrastructure

E. H. Brookes^a, J. E. Curtis^b, D. Fushman^c, S. Krueger^b and A. Savelyev^a

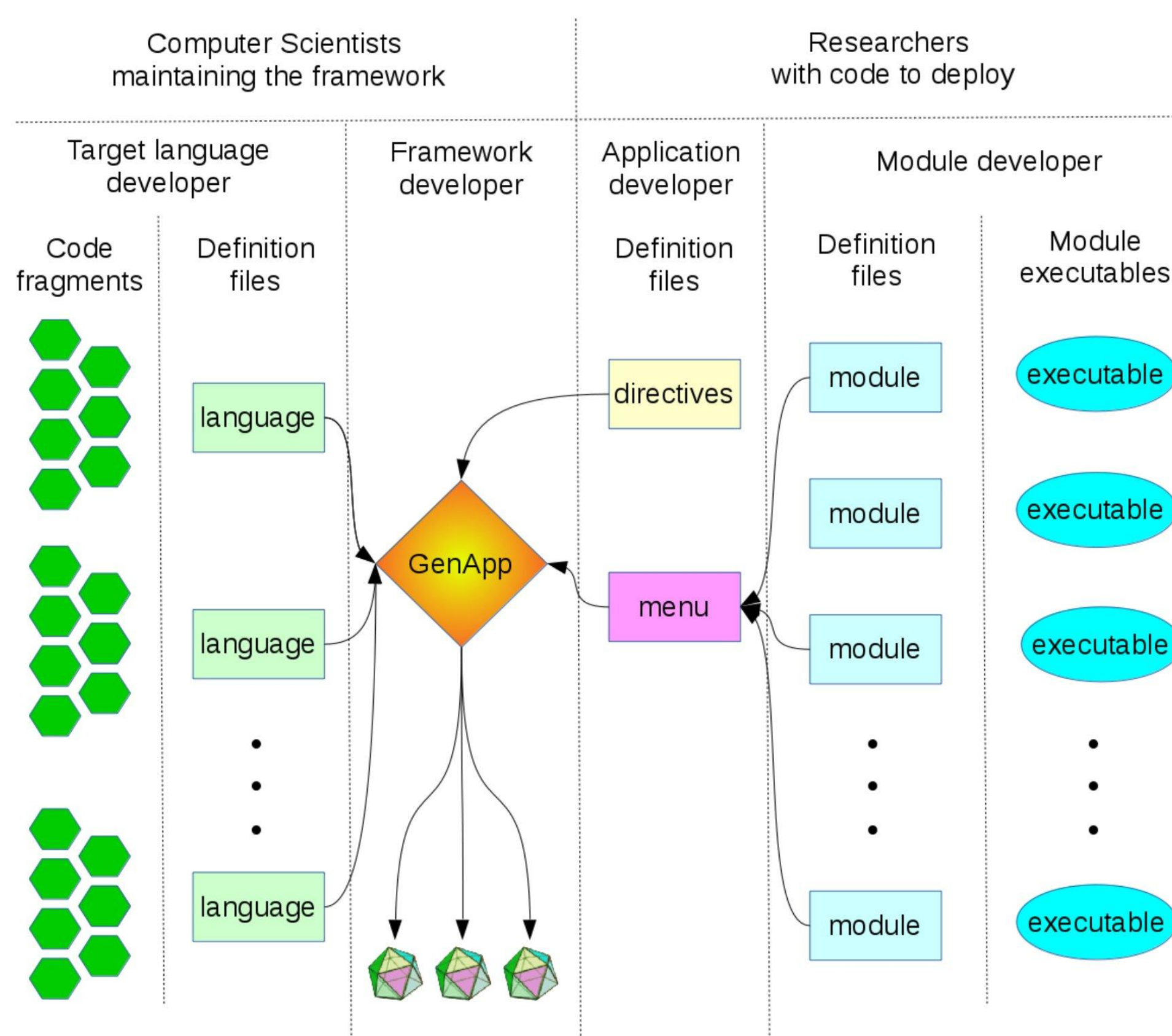
^aDepartment of Biochemistry & Structural Biology, University of Texas Health Science Center at San Antonio, San Antonio, Texas

^bNIST Center for Neutron Research, Gaithersburg, Maryland

^cDepartment of Chemistry & Biochemistry, University of Maryland, College Park, Maryland

OVERVIEW: GenApp, a [Generalized Application Framework](#) [1, <https://genapp.rocks>], is a framework for rapidly building and deploying applications to an extensible variety of targets including web-based science gateways. GenApp works by building targets from a collection of definition files. An existing program can be brought into GenApp by writing a definition file and wrapping or modifying the program to accept inputs and produce outputs as defined. GenApp was developed to support the needs of the SI2-CHE CCP-SAS [2] project to rapidly expose SAS programs to the community. Some notable features of generated gateways include: interactive 2D plots, 3D plots and atomic structures; user and job management with history and statistics; a “cloud” file system; integrated help and feedback; OAuth credential management; and support for a variety of methodologies for execution of programs such as direct, elastic (on clouds), and batch (on queue managed clusters or HPC resources). New features and capabilities are added as required by use cases. GenApp is successfully being used to deploy multiple software projects in the SAS domain, including **SASSIE** [3], **WillitFit** [4], **MULCh** [5], **QuaFit** [6], **Denfert** [7], **BayesApp** [8], **Pepsi-SAXS** [9] and **US-SOMO** [10]. GenApp is not limited to SAS software and is being used to deploy software in other areas of science, such as NMR and CFD. GenApp is an affiliate of NSF's Science Gateways Community institute and has an active engagement with the NSF's TrustedCI to improve cybersecurity of GenApp generated gateways.

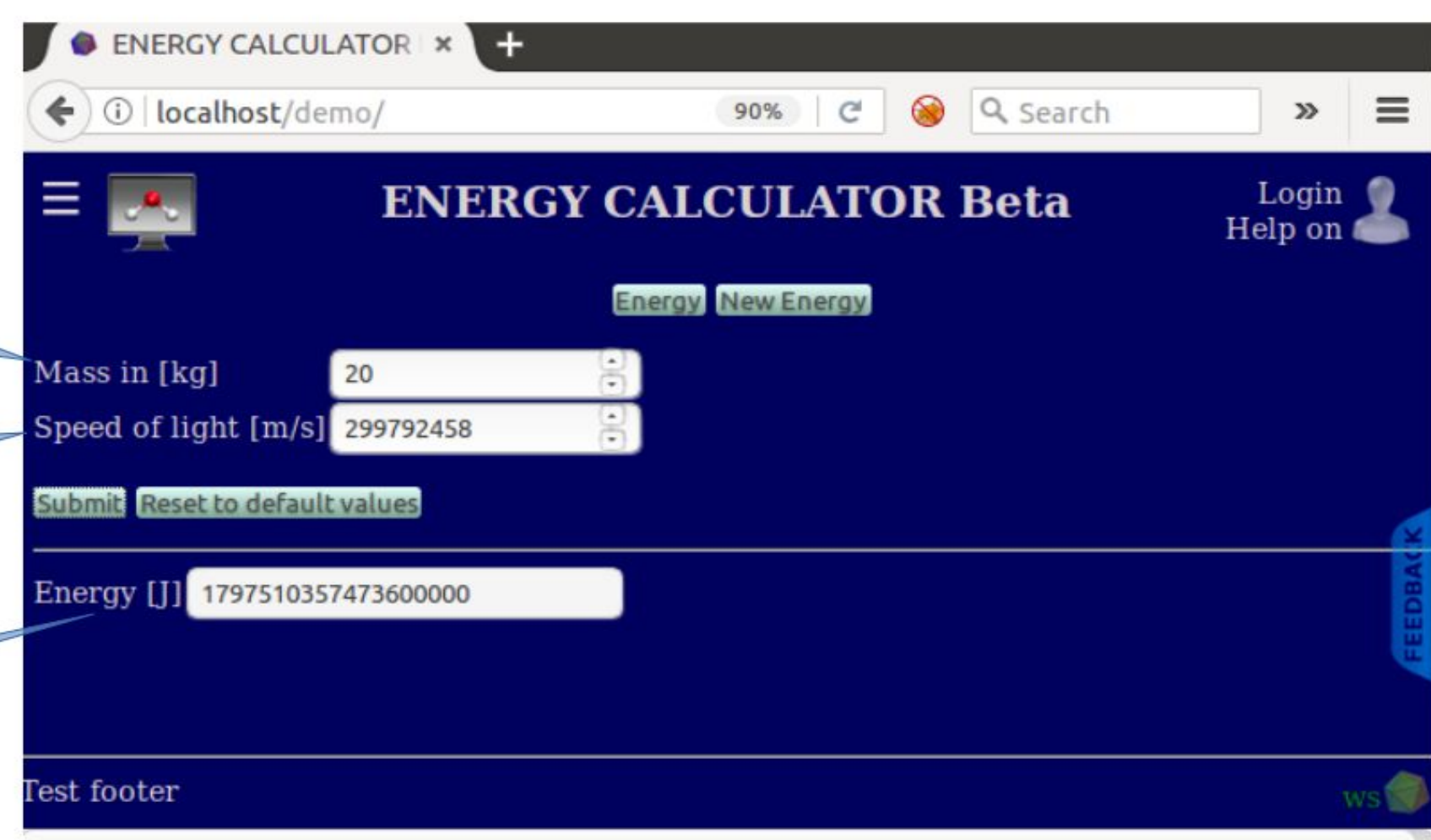
PHILOSOPHY: The [separation of powers](#) in GenApp. The researchers focus on their specific code defined as modules and are unconcerned about the technical details of the target language. The CS experts work on developing enhanced capabilities or new or variant target languages. This structure enables rapid adaptable dissemination of applications, minimizes researcher overhead and preserves hard-won scientific code in an ever-evolving software landscape.



TECHNICAL DETAILS: An example of the module definition files and the corresponding GenApp generated science gateway

```
def einstein(mass, speed of light): return mass*(speed_of_light ** 2.0)

{
  "moduleid": "penergy",
  "label": "Energy",
  "help": "Help for Energy",
  "executable": "penergy",
  "submitpolicy": "all",
  "uniqueid": "true",
  "fields": [
    {
      "role": "input",
      "id": "m",
      "label": "Mass in [kg]",
      "type": "float",
      "required": "true",
      "help": "Enter the mass in kilograms"
    },
    {
      "role": "input",
      "id": "c",
      "label": "Speed of light [m/s]",
      "type": "float",
      "default": "299792458",
      "required": "true",
      "help": "Enter speed of light in meters/second"
    },
    {
      "role": "output",
      "id": "e",
      "label": "Energy [J]",
      "type": "text"
    }
  ]
}
```

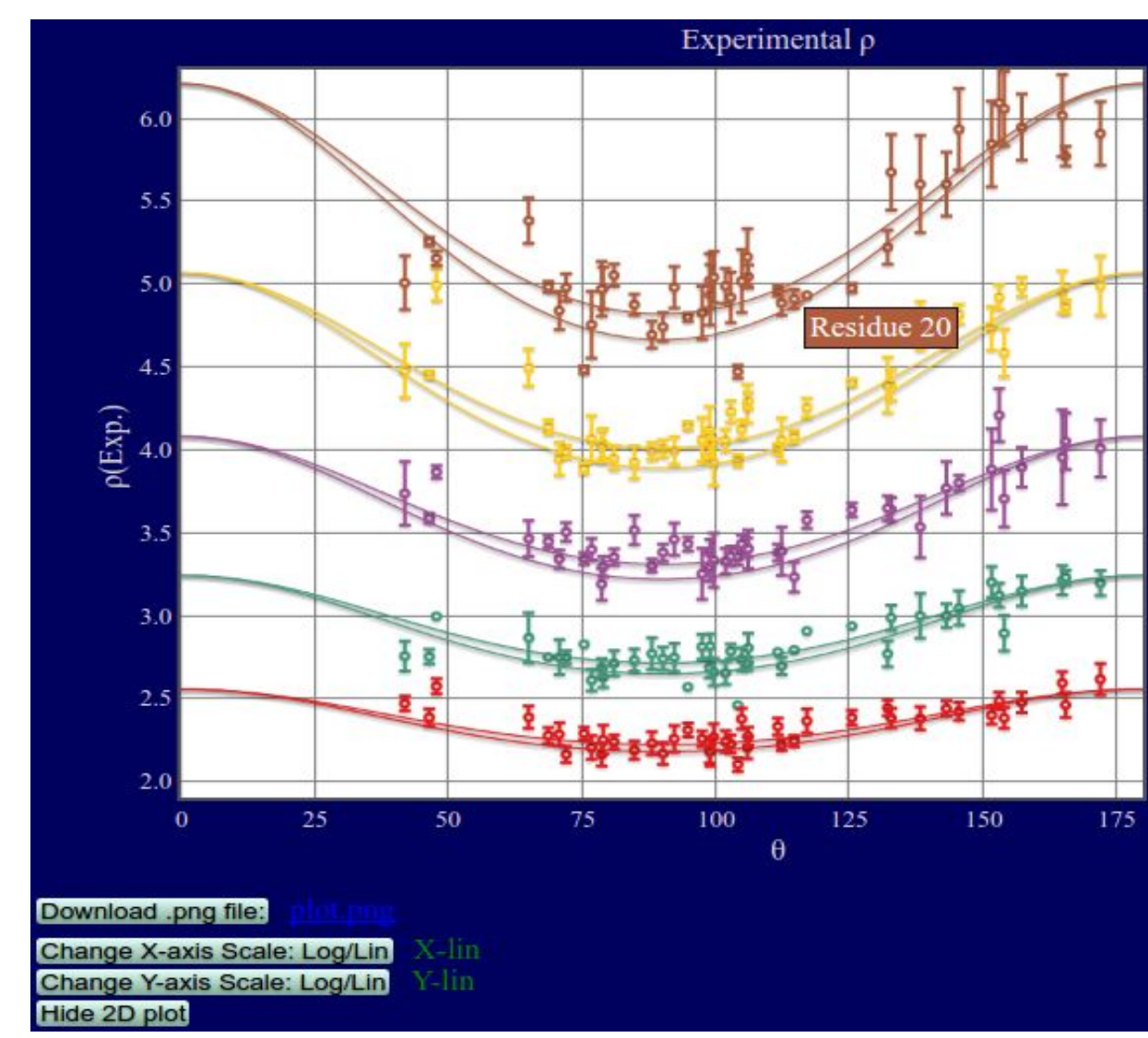
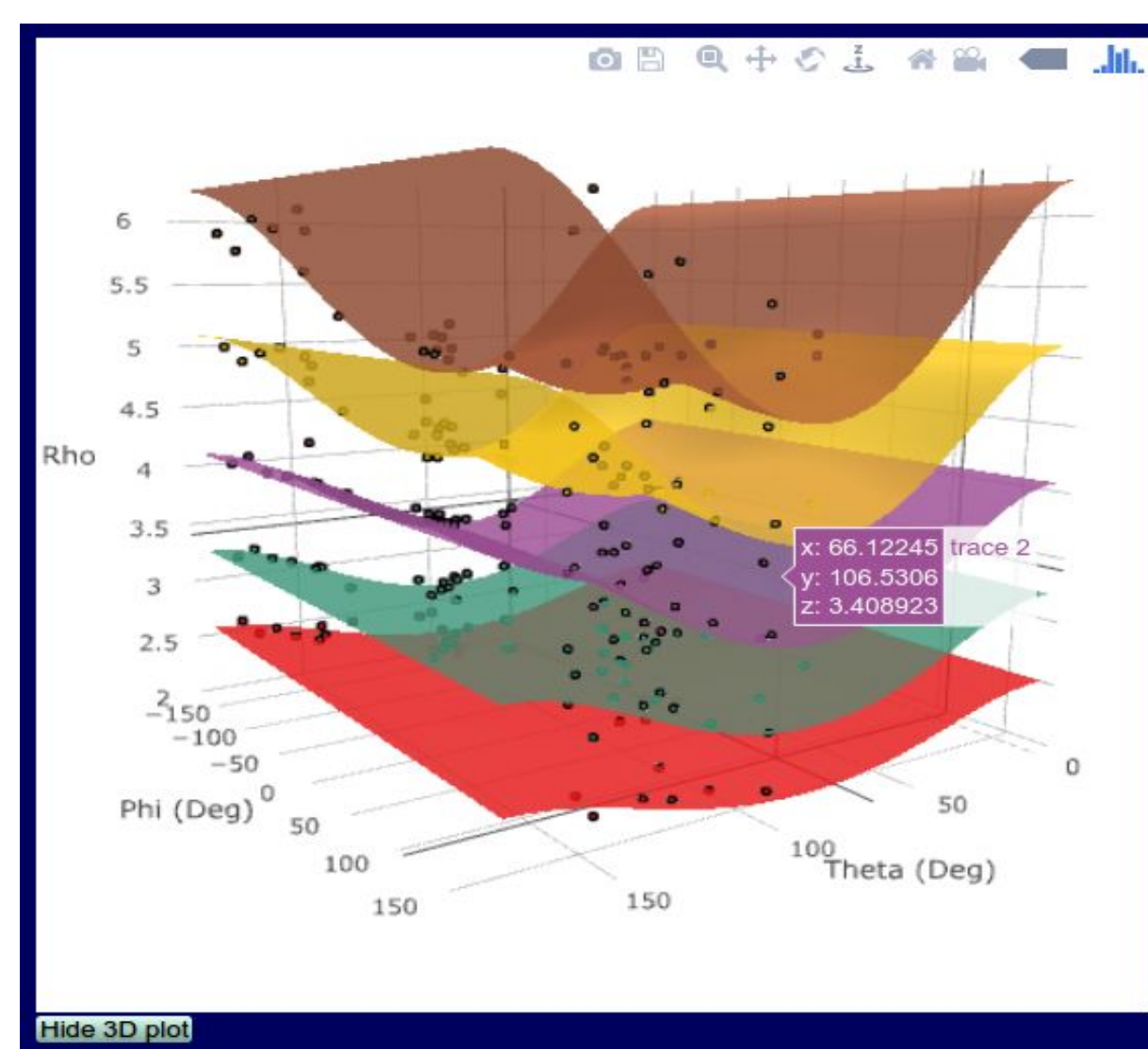
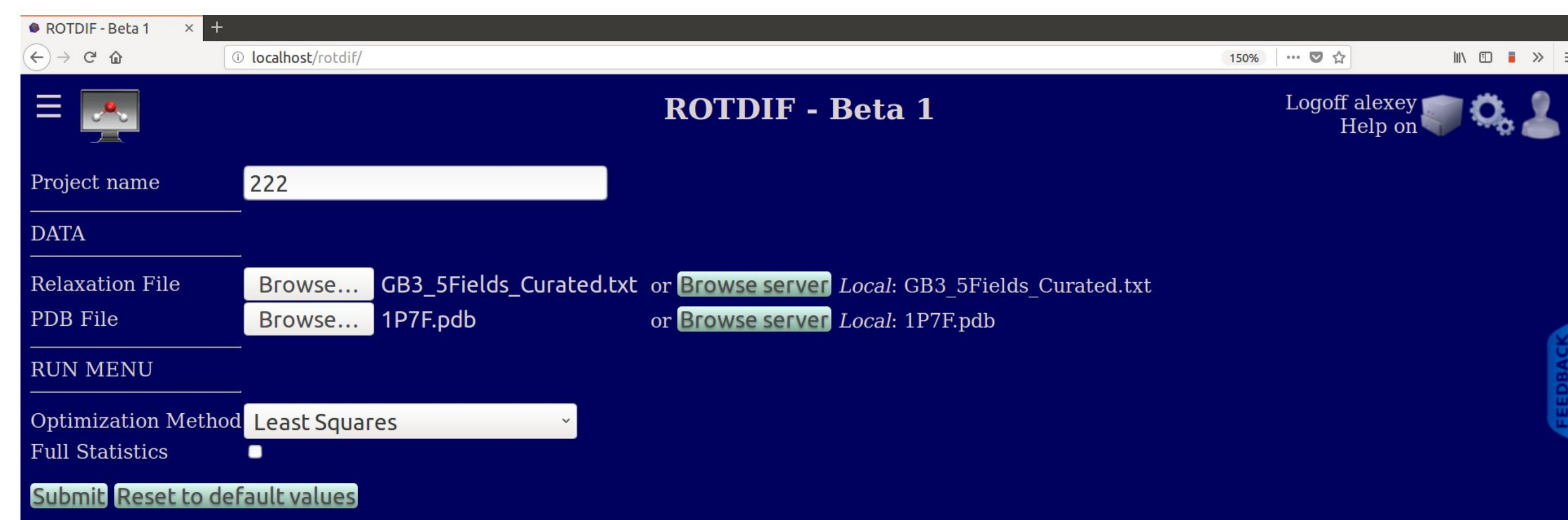


MAIN FEATURES:

- ability to generate applications in an extensible variety of target languages
 - HTML5/JavaScript/PHP, Java, Qt4, Qt5, Qt5/Android
- messaging, context sensitive help, user management and statistics
- multiple execution models: local to remote queue managed HPC and cloud-elastic
- job managements with multiple simultaneous reattachment to job running or run
- a “cloud” file system, caching of results; OAuth2 integration
- integrated feedback mechanism which links full failed job information for easy debugging.

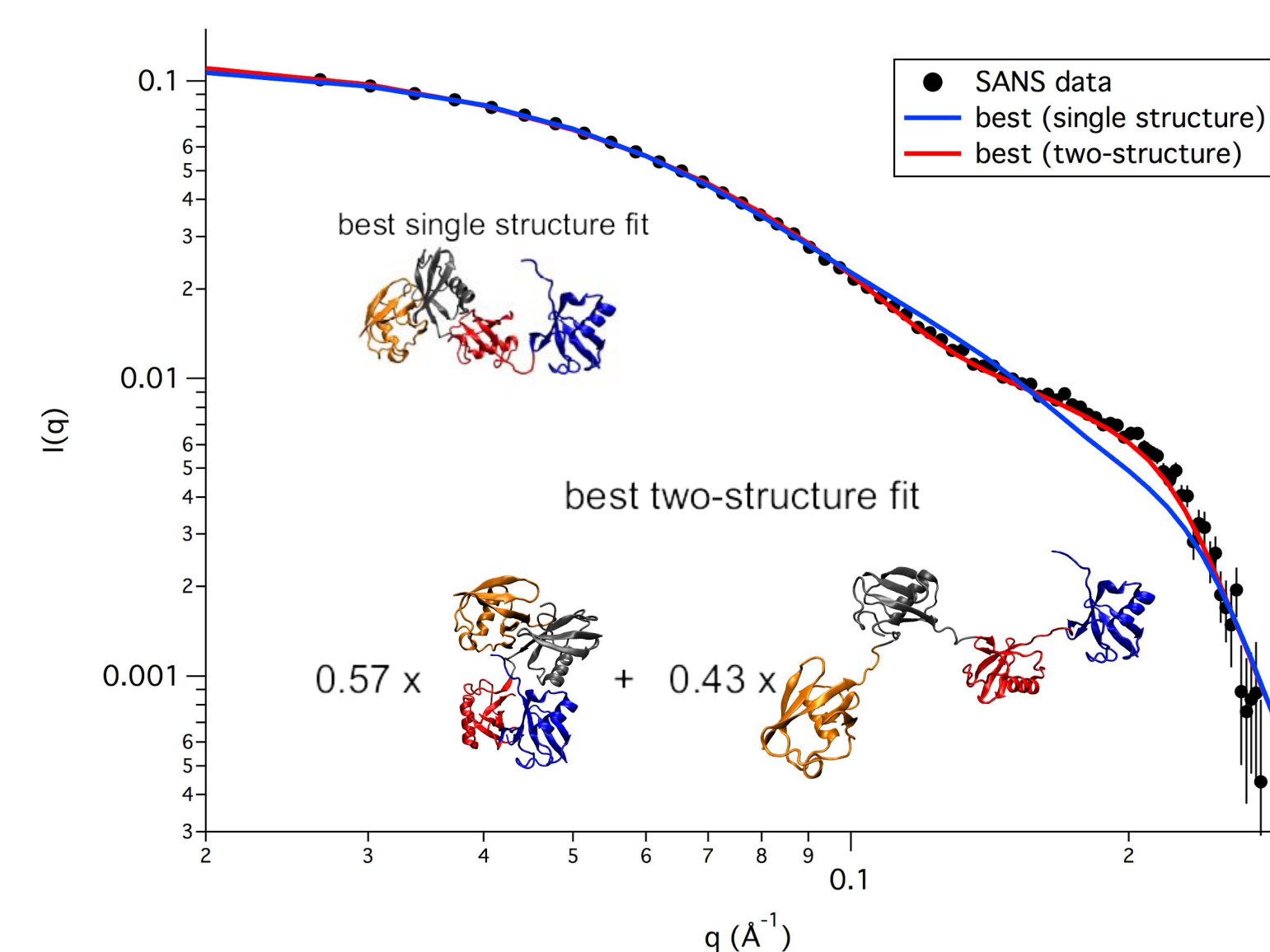
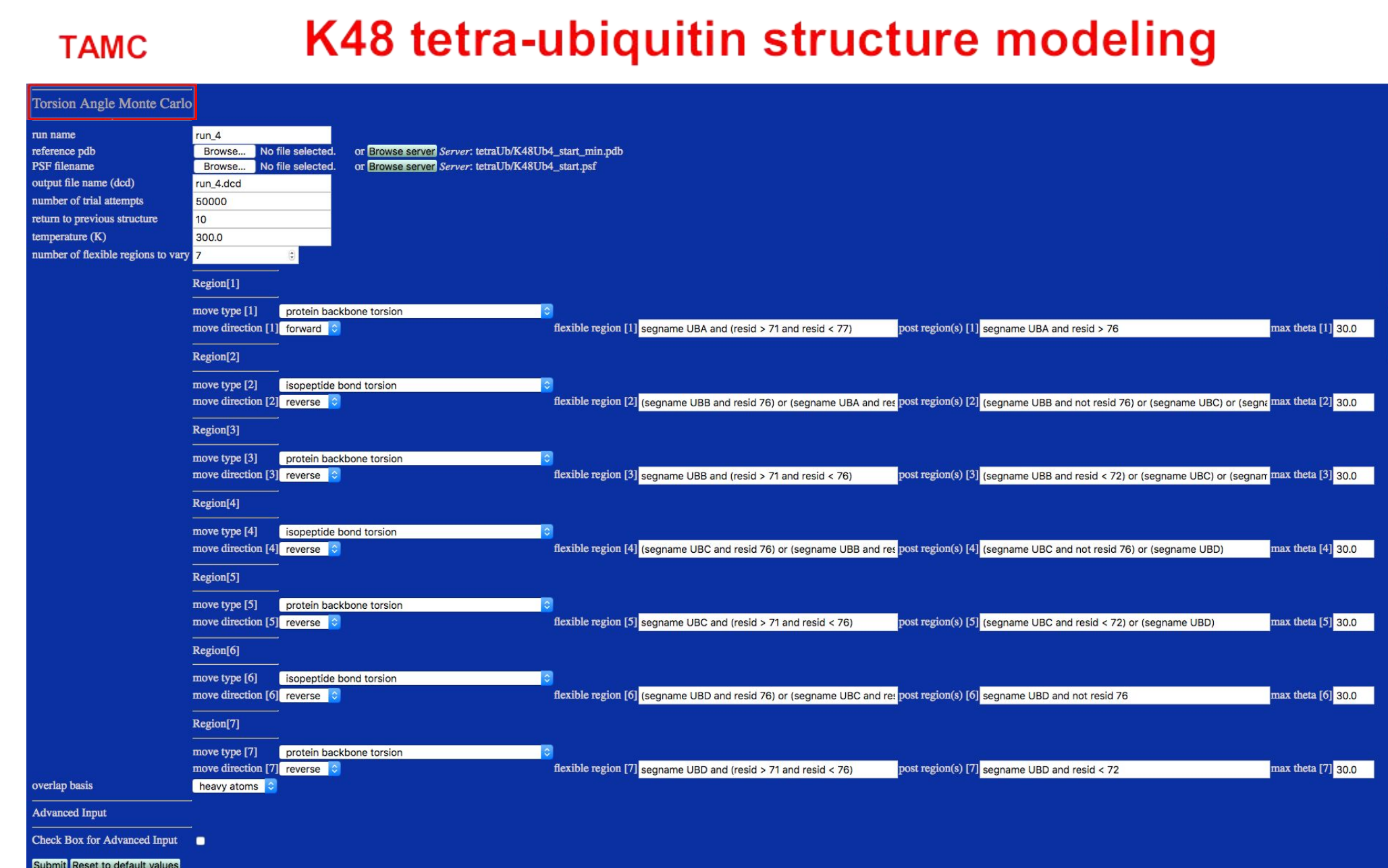
RECENT APPLICATION (NMR): ROTDIF

ROTDIF [11] is the first of several computer programs for analysis of various types of NMR data which has recently been turned into science gateway via GenApp. ROTDIF has several important capabilities. It can simultaneously analyze spin-relaxation data for ¹⁵N and ¹³C nuclei in proteins and/or nucleic acids and at several magnetic fields, to determine the overall rotational diffusion tensor and characterize the internal motions. The program also efficiently predicts a fully anisotropic rotational diffusion tensor of a macromolecule from atom coordinates [12], and performs rigid-body docking of molecular complexes guided by experimental rotational diffusion tensors [13].



APPLICATION (SAS): SASSIE-Web

SASSIE [3] is a suite of programs to enable molecular simulation of soft-matter systems to model X-ray and neutron scattering experimental data. The prototype of GenApp was first applied to SASSIE to create SASSIE-web. This has had a significant positive impact on our user community with over 500 registered users and 24000 jobs over the first 18 months since SASSIE-web became operational. Over 50 scientific publications have resulted from the use of SASSIE-web. GenApp provided the framework to handle a large user community and thus freeing our limited resources to work on other barriers facing our users.



A TAMC-generated ensemble of 34,655 K48Ub4 structures was analyzed to find the best match to small-angle neutron scattering (SANS) data. The best single structure does not match the data over the entire range. The best two-structure solution, consisting of an extended and compact structure, matches the data much better. The presence of the compact structure had been predicted by NMR.

References:

- [1] A. Savelyev and E. H. Brookes. Fut. Gen. Comp. Sys. (2017) In press.
- [2] S. J. Perkins *et al*, J. Appl. Cryst. 49 (2016) 1861-1875.
- [3] J. E. Curtis *et al*, Comp. Phys. Comm. 183 (2012) 382-389.
- [4] M. C. Pedersen *et al*, J. Appl. Cryst. 46 (2013) 1894-1898.
- [5] A. E. Whitten *et al*, J. Appl. Cryst. 41 (2008) 222-226.
- [6] Spinozzi *et al*, Biophys J 103 (2012) 511-521
- [7] A. Koutsoubas and J. Pérez, J. Appl. Cryst. 46 (2013) 1884-1888.
- [8] S. Hansen, J. Appl. Cryst. 33 (2000) 1415-1421.
- [9] S. Grudin *et al*, Acta Cryst. D. 73 (2017) 445-464.
- [10] E. H. Brookes and M. Rocco, Eur. Biophys. J. (2018) in press.
- [11] O. Walker *et al*, J Magn Reson 168 (2004) 336; K. Berlin *et al*, J Biomol NMR 57 (2013) 333.
- [12] Y. Ryabov *et al*, J Am Chem Soc 128 (2006) 15432.
- [13] Y. Ryabov *et al*, J Am Chem Soc 129 (2007) 7894; K. Berlin *et al*, Proteins 79 (2011) 2268.

Acknowledgements

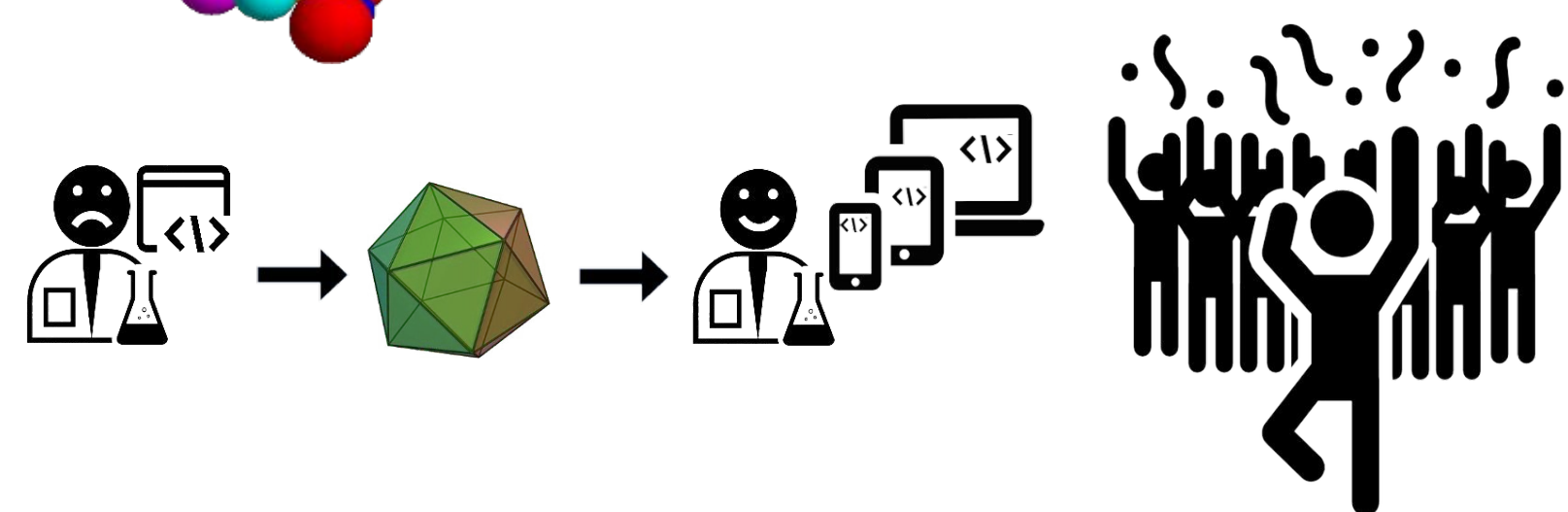
- Funding
 - NIH NIGMS K25GM090154 (E.B.)
 - NSF OCI-1032742 (to E.B.)
 - NSF CHE-1265817 (to E.B.)
 - NSF OAC-1740097 (to E.B.)
 - NSF OAC-1739549 (to D.F.)
- NSF XSEDE
- UTHSCSA BCF
- Google GSOC 2014, 2015
- CCP-SAS
- Apache Airavata

Funded by a grant from the National Institute of General Medical Sciences of the National Institutes of Health



GENAPP'S FUTURE:

- Security with Trusted CI engagement
- Improved UI with SGCI engagement
- Advances to SAS / SASSIE modules
- New NMR suite & modules
- API & Container generation
- Job staging
- Question response
- Stand alone GUI enhancements



<https://genapp.rocks>