BioSimulations: A unified, extensible, repository of models, simulations, and visualizations

Bilal Shaikh[1], Lucian P. Smith[2], Micheal L. Blinov[3], Herbert P. Sauro[2], Ion I. Moraru[3] and Jonathan R, Karr[1]

Abstract— Comprehensive, predictive mechanistic models have great potential for biology, bioengineering, and medicine. These models will likely require collaboration between investigators that can easily reuse and combine each other's simulations. Several community resources, including languages, ontologies, and repositories have been developed to facilitate reuse. Unfortunately, many gaps still remain in model reproducibility, including the complexity of methods, tools, and formats, the siloing of domains across repositories, and limited quality controls. To facilitate the share and reuse of simulations, we developed BioSimulations, a central repository for models, simulations, and visualization of their results, which we anticipate will be a driving force toward more comprehensive and more predictive models.

Index Terms— simulations, dynamical modeling, reproducibility, standards, algorithms, pathways, COMBINE, SBML, CellML, SED-ML, BNGL, KiSAO

I. MOTIVATION

More comprehensive and predictive mechanistic models have great potential for advancing science, bioengineering, and medicine. A prerequisite for successful modeling of large-scale systems, such as creating models of whole cells, is the ability to combine models of various subsystems. To fully capture the complexity of biology, collaboration between diverse teams of modelers and experimentalists will be required, with each team focusing on the submodels within their area of expertise. These submodels will likely use various frameworks, methodologies, and algorithms applied across the various scales of biological systems. To be effective, investigators will need to be able to share and reuse each other's models and quickly iterate on simulations.

Several community resources, such as modeling languages, ontologies, formats, and repositories have been developed to enhance interoperability and facilitate model reuse. Languages such as SBML, CellML, and BNGL, provide standardized descriptions of models, SED-ML specifies simulation experiments, and ontologies such as KiSAO provide information about various algorithms and their parameters. However, the large variety and number of these resources, as well as the unequal level of support by different software tools can make it challenging to employ them properly, and utilize the appropriate tools for each model. Additionally, models are scattered across various repositories, siloed by simulation method, modeling language, or simulation tool, and often have little to no curation and quality controls. BioSimulations is a unified repository for models, simulations, and visualizations of their results to enable investigators to find, quality control, share, modify and execute entire simulation studies across a range of simulation methods, frameworks, and model languages.

II. IMPLEMENTATION

BioSimulations is built upon existing community resources and integrates a wide range of model languages, formats, ontologies, tools, and algorithms into a comprehensive set of standards that enable interoperability and reproducibility across a range of biological systems and scales. Where gaps exist in the existing standards, BioSimulations includes guidelines and conventions to bridge together these resources. This required refining the Simulation Experiment Description Markup Language, a standard for describing simulations; expanding the Systems Biology Ontology, used for categorizing modeling frameworks; expanding the Kinetic Simulation Algorithm Ontology to specify simulation methods; creating new formats for simulation results, logs of simulations, data visualizations of simulation results, and capabilities of simulation tools; developing BioSimulators, a registry of simulation tools; providing utilities for quality controlling simulations and simulation tools; and integrating these resources together into a single platform. Furthermore, BioSimulations is designed to be extensible, allowing the community to expand the range of supported simulation methodologies and model formats, and contribute additional tools. BioSimulations has already incorporated nearly 1000 projects from community repositories and continues to expand.

III. AVAILABILITY

BioSimulations is freely available at biosimulations.org. Extensive user documentation, developer guides, and tutorials are available at docs.biosimulations.org

The source code for the web application, REST API, and simulation service is available under the MIT license in our GitHub repository, at https://github.com/biosimulations/biosimulations.

[1]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. Emails, respectively: bilalshaikh42@gmail.com and karr@mssm.edu

[2]Department of Bioengineering, University of Washington, Seattle, WA, USA Emails, respectively: lucianoelsmitho@gmail.com and hsauro@uw.edu

[3]Center of Cell Analysis and Modeling, University of Connecticut School of Medicine, Farmington, CT, USA Emails respectively: blinov@uchc.edu and moraru@uchc.edu