Supplemental Information

Best Practices for Making Reproducible

Biochemical Models

Veronica L. Porubsky, Arthur P. Goldberg, Anand K. Rampadarath, David P. Nickerson, Jonathan R. Karr, and Herbert M. Sauro

General-Purpose Reproducible Biochemical Modeling Checklist

Data aggregation:

☐ New data collection and experimental procedures are reproducible

    ☐ The experimental protocol is provided

    ☐ Conditions that deviate from the standard protocol are provided

    ☐ Measurement uncertainty is quantified

    ☐ A record of when the data was collected is provided

    ☐ A record of the individual or lab that produced the data is provided

☐ Data collected from databases and literature are curated and metadata are provided

    ☐ All manipulations applied to measurements are described and the original data is referenced (e.g. statistical analyses, normalization)

    ☐ Database and data source entries are provided

    ☐ Ownership of the original data is credited

Model construction:

☐ All species and parameter names are biologically-relevant and unambiguous

    ☐ Shortened identifiers used in the model description are linked to the full species or parameter name in a supplemental table

    ☐ The full species or parameter names follow standardized naming conventions

☐ All reactions are uniquely identified

☐ Model components are described in computer-readable tables and automatically imported by simulation study code

☐ If a model uses differential equations, then its biochemical reactions and differential equations are both reported

Parameter estimation:

☐ Software generated to perform parameter estimation is provided

☐ Confidence intervals are reported on estimated parameters

☐ Families of parameter values are reported for non-identifiable models, if there are a discrete number of possibilities

☐ Uncertainty quantification is performed

Simulation:

☐ All initial conditions and parameter values are provided for each published simulation experiment

    ☐ Well-documented tables of parameter values for each unique simulation experiment are provided with the publication

☐ All simulation experiments are fully defined (events listed, collection times and measurements specified, algorithms provided, simulator specified, etc.)

☐ Algorithms required for executing the simulation are reported

    ☐ Numerical integration algorithms are reported

- ☐ Stochastic algorithms are reported
- ☐ Random number generator algorithms are reported
- ☐ Other algorithms are reported (e.g. steady state methods)

☐ Stochastic simulations are repeated to generate a representative distribution of simulation results

- ☐ Seeds are recorded for each simulation result, a method for obtaining the seeds is provided, or a method for generating a statistically-similar distribution of results is provided

Results storage:

☐ All data is provided in supplemental tables or computer-readable spreadsheets (e.g. HDF file)

- ☐ Unprocessed results are provided

☐ All data is annotated for comprehension

☐ Source data and code to produce published figures is provided

Verification & validation:

☐ All custom modeling code is systematically verified

- ☐ A program is written which checks that all model components and numerical methods exhibit expected behavior

☐ The model and associated simulation experiments are validated to assess biological relevance

☐ The model and associated simulation experiments are tested in an independent computing environment to ensure that the software is portable and reproducible

Documentation:

☐ All data and software used to construct the model are documented

☐ Installation and usage instructions are recorded for the model and associated software (e.g. as a README file)

☐ Design decisions and assumptions are recorded

- ☐ The biological relevance of the modeling study, including mathematical descriptions, and any simplifying assumptions are justified
- ☐ The biological system context is described
- ☐ The environmental context is described

☐ The model and all associated programs for simulation, analysis, and verification and validation are explained through comments

☐ Figures visualize the model—all components and interactions—and the results of simulation experiments, include detailed figure legends

Packaging:

☐ All model artifacts are organized in a single archive

☐ The model archive contains sub-directories which are described, along with all model artifacts, in a manifest file

Dissemination:

☐ The modeling study is disseminated in a pre-print server (e.g. bioRxiv)

☐ The modeling study is published in a peer-reviewed journal

☐ The model and source code are available at a public repository or personal web site

- ☐ A permissive license to use repository materials is provided