Generating WESTPA templates of BioNetGen language models

Abstract

WEBNG is an open-source templating tool to simplify weighted ensemble simulations of rule-based models. WEBNG bridges the open source software packages WESTPA, for weighted ensemble path sampling, and BioNetGen, for rule-based modeeling. This package is designed to simplify enhanced rare-event sampling of biological rule-based models, allowing researchers to tackle more challenging models that contain harder to simulate processes within a well-defined framework. Depending on existing open-source packages that are proven and are in active development allows WEBNG to stay up-to-date while also being easy to install and maintain.

Introduction

Models of biochemical systems can get complicated, resulting in a large number of chemical species. Frequently these models result in combinatorial complexity when species have multiple states that they can be in. A rule-based modelling approach was developed to avoid manually enumerating every possible species and reactions. Rule-based models define rules which specify how species can interact and all possible reactions and species are generated off of these rules. BioNetGen (BNG) is an open-source rule-based modelling software package that allows for model construction in BioNetGen language (BNGL) and simulation of BNGL models.

Rare events in modelling are events that take a long time to occur compared to the time scale of the fastest proccess in the model. For some models, rare events are unavoidable because the fastest process in the system must be modelled for accuracy. This is frequently the case for biological models and while some modelling methods allow for simplification of the fastest process, it's generally ideal to model everything if possible. Models with rare events are also frequently hard to simulate because it takes many events to get a statistically robust estimate of any aspect of a biological process which means the simulations need to run for a long time to get enough rare events to occur.

Tackling this problem has been a focus of the computational modelling field for

a while and many rare event sampling methods have been developed. One of these methods is called weighted ensemble method. Weighted ensemble path sampling works by organizing multiple parallel trajectories, resampling them at a fixed time interval. Each trajectory is assigned a statistical weight that are tracked during WE resampling. This allows WE resampling scheme to keep model kinetics statistically rigorous while replicating trajectories making progress towards the rare event of interest and terminating trajectories that aren't.

WESTPA is an open-source, scalable and interoperable softare package that applies the WE strategy. WESTPA has been successfully used to study multiple challanging biological processes and is being actively maintained. WESTPA is interoperable and is designed to interface with any stochastic simulator, which makes it ideal to use with BNG stochastic simulation methods.

In bioinformatics modelling, researchers frequently avoid working with models that contains rare events because it is easier to coarse grain the fastest processes and keep the timescale separation in a model relatively small, reducing computational cost. However, this does come at the cost of accuracy and not being able to understand how the kinetics of the fast processes in a model affect the slow ones. Rare event sampling methods can alleviate this issue, allowing researchers to tackle more detailed, challenging models that contain rare events.

To this end we have developed a templating tool named WEBNG that makes it easier to setup WESTPA simulations of a BNGL model. The user needs to supply parameters for a WESTPA simulation along with a BNGL model and the tool creates a WESTPA simulation folder which can then be ran like any other WESTPA simulation. The tool also comes with some analyses built-in and is designed to tackle high dimensional WE simulations which is common for BNG models. WEBNG tries to lower the entry barrier for bioinformatics researchers to use weighted ensemble rare event sampling method and explore more models where rare events occur without having to worry about the computational cost.

How to use WEBNG

WEBNG is on python package index meaning that the installation can be done with a single command from the command line: `pip install webng`. This also installs a specific version of WESTPA that WEBNG is designed to work with as well as PyBioNetGen (link). This not only allows the user to avoid having to install the other two packages separately but also allows WEBNG to know where the packages are located automatically so no user input is needed as to where the other two packages are installed.

WEBNG comes with three subcommands to simplify the generation of the WESTPA simulation folder and analysis of the resulting simulation. First subcommand `template` allows you to point to the BNGL model file you want to simulate and this subcommand will generate a YAML file that's easy to read and
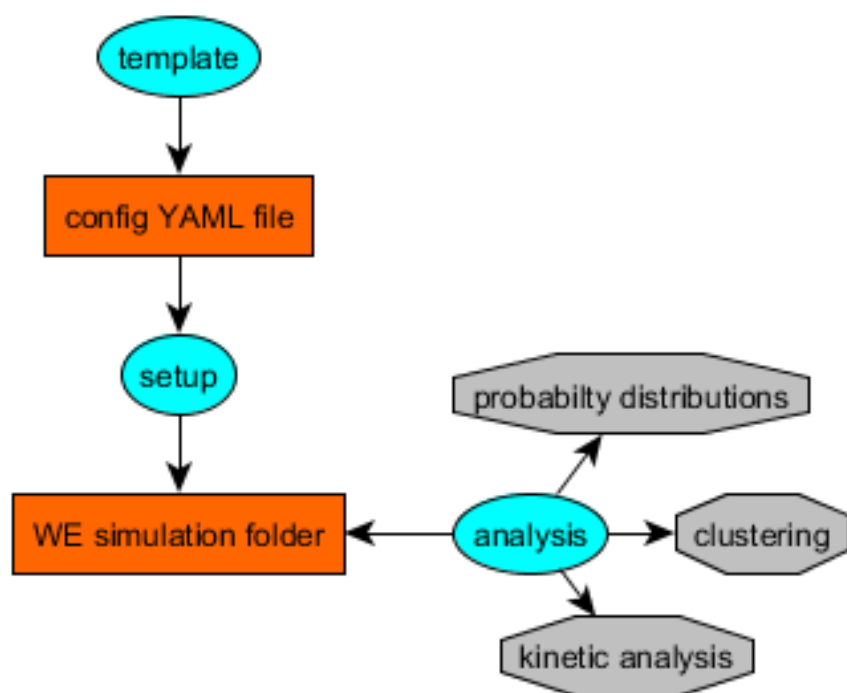
Figure 1: Figure1

change and is populated with sane default values. This YAML file contains information about where the packages are installed (which are automatically detected but can be changed here), WESTPA simulation parameters and parameters for the built-in analyses.

The next subcommand, `setup` takes in a single argument which is the YAML file generated (and adjusted by the user). This subcommand will use this information to setup a stadard WESTPA simulation folder that can be simulated. From this point the user can also adjust the folder further for more sophisticated options that aren't currently supported WEBNG. For information on what WEBNG currently supports, see WEBNG documentation (link). This workflow allows for a new user to immediately start simulating their BNG model while also allowing for an advanced WESTPA user to change simulation parameters and options freely.

Once the simulation is ran, the user can use the `analysis` subcommand, which also takes in a single YAML file. This subcommand will take in the parameters provided in the YAML file and run the build-in analyses on the simulation. The build-in analyses are (1) probability distributions of each progress coordinate over WE iterations (Fig1A), (2) 2D joint probability distributions of each progress coordinate against each other progress coordinate (Fig1B), (3) PCCA based clustering calculated from the transition matrix (this also generates a reusable transiton matrix file for other analyses) and calculation of rate constants between clusters (Fig1C) and (4) using the clusters to generate graph files that can be used to visualize clusters and transitions between them (Fig1D). These analyses can be more complicated for a new user to implement since the user needs to know how to program but also needs to learn how WESTPA files are structured. WEBNG allows a new user to start analyzing the WESTPA simulation immediately, analyses 1 and 2 are useful for tracking simulation progress and analyses 3 and 4 are hard to implement for new users.
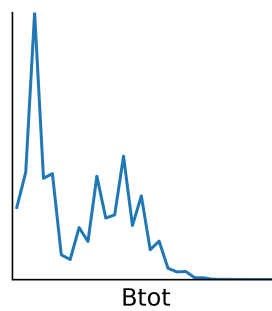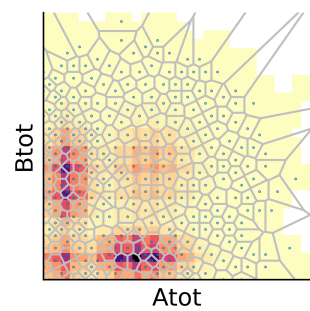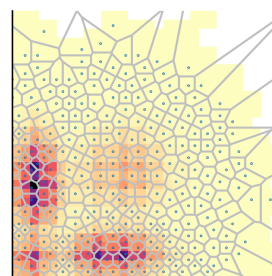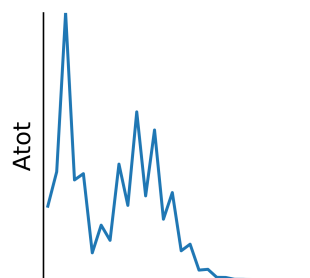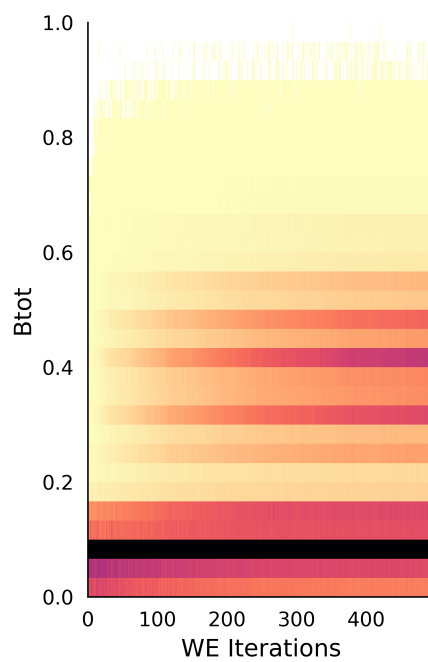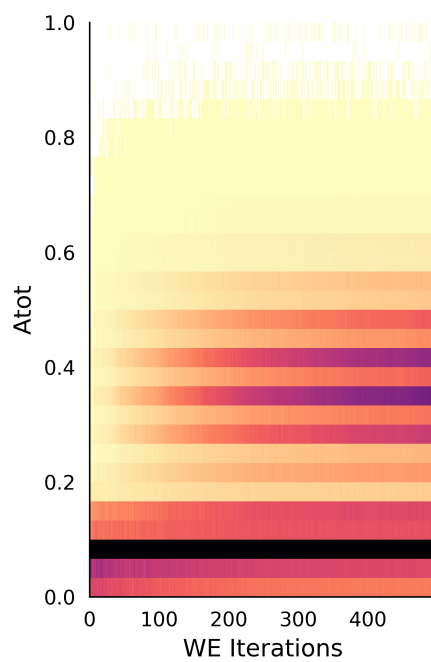
Transition matrix from simulation				
	lo/lo	hi/hi	lo/hi	hi/lo
lo/lo	9.05×10^{-1}	1.35×10^{-3}	4.85×10^{-2}	4.53×10^{-2}
hi/hi	1.82×10^{-5}	9.50×10^{-1}	2.37×10^{-2}	2.67×10^{-2}
lo/hi	2.41×10^{-2}	3.28×10^{-2}	9.42×10^{-1}	1.27×10^{-3}
hi/lo	3.12×10^{-2}	5.04×10^{-2}	1.31×10^{-3}	9.17×10^{-1}

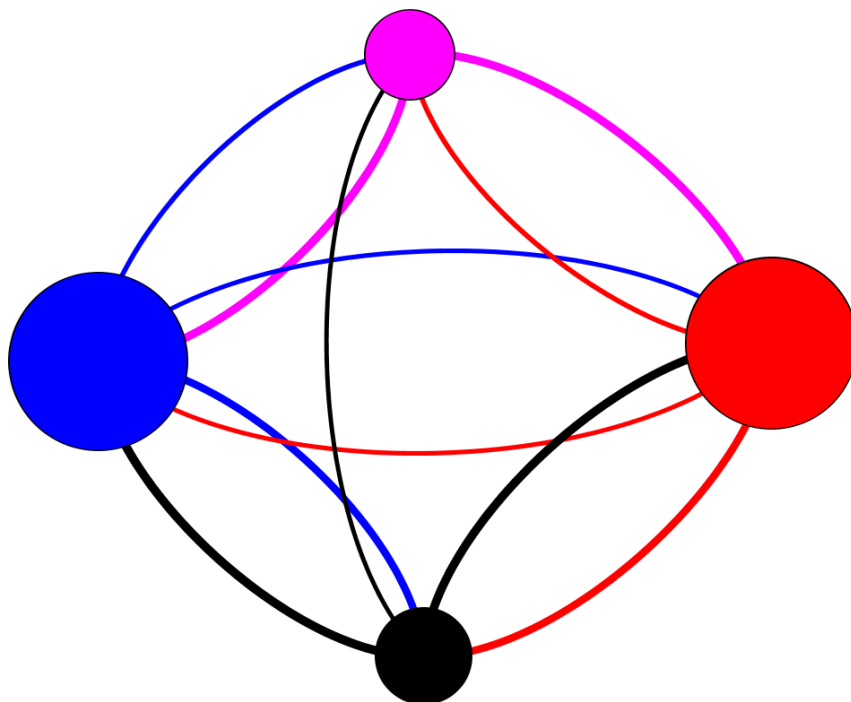| Stationary state probabilities (%)			
lo/lo	hi/hi	lo/hi	hi/lo
8.2	17.5	37.5	36.8

Figure2: Results of automated analyses on a simulation of X, replicated from Read study. (A) Probability distribution of each progress coordinate over WE iterations. Each distribution is calculated off of a running average over Y iterations. Lighter colors show less probable regions while darker colors show more probable regions. (B) 2D heatmaps of joint probabilty distributions of each progress coordinate against each other progress coordinate. Lighter colors show less probable regions while darker colors show more probable regions. The probabilty distributions are estimated over Z WE iterations.(C) Transition probabilities and rate constants between four clusters defined using PCCA clustering algorithm. Transitions between clusters HH and LL are rare events and the error bars show that the rate constants are well converted. (D) Network graph generated off of the clustering done in the previous clustering step. Thicker vertices mean higher probability (and therefore rate constant) transitions and node sizes scale with the probability of each state in steady state.

One advantage of a dedicated software package for templating these simulations is reproducibility. There are many studies where the model is implemented

in a language specific manner and contains many hardcoded variables which makes reproducing results challanging even when the code is provided with the published paper. Dedicated software packages like WEBNG allows for the same simulation setup to be generated and results analyzed in the same manner which will make results far more reproducible, assuming the simulations are ran until convergence.

Conclusion and future directions

WEBNG is a simple templating tool that bridges two open software packages, WESTPA and BioNetGen, and lowers the barrier to entry to weighted ensemble rare event sampling of rule-based models. WESTPA, BNG and WEBNG are all well documented which helps any potential researcher that wants to make BNG models that containt rare events to get a simple starting point. Using python package index as the distribution point not only makes installation simple but also allows for all three software packages to be kept up-to-date and in sync easily.

New analyses, better efficiency and simpler ways to generate templates are currently in development and as WESTPA and BNG packages get updated, WEBNG will be kept in sync with both software packages. More information on the documentation of WEBNG, WESTPA and BNG can be found in the supplementary information.