

# WEBNG - a templating tool for weighted ensemble sampling of rule-based models

## Abstract

WEBNG is an open-source templating tool written in Python programming language that simplifies 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 modeling. 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 being written in python makes it easy to install and maintain.

## Introduction

### Rule-based modelling and BioNetGen

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. Rule-based modeling 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 modeling software package that allows for model construction in BioNetGen language (BNGL) and simulation of BNGL models.

### Rare events and the sampling problem

Rare events in modeling are events that take a long time to occur compared to the time scale of the fastest process in the model. For some models, rare events are unavoidable because the fastest process in the system must be modeled for accuracy. This is often the case for biological models and while some modeling

methods allow for simplification of the fastest process, it's generally ideal to model as much as possible. Models with rare events are hard to simulate because it takes many events to get a statistically robust estimate of any aspect of a biological process which means the model has to be simulated for a long time in order to sample enough rare events.

## **Enhanced sampling and weighted ensemble path sampling**

Tackling this problem has been a focus of the computational modeling field and many rare event sampling methods have been developed. One of these methods is called weighted ensemble path sampling method. Weighted ensemble path sampling works by organizing multiple parallel trajectories and resampling them at a fixed time interval. Resampling allows WE to replicate trajectories that are making progress towards the rare event of interest and terminating trajectories that aren't. Each trajectory is assigned a statistical weight which is tracked during resampling. This ensures that model kinetics aren't biased and kept statistically rigorous.

## **WESTPA**

WESTPA is an open-source, scalable and interoperable software package that applies the WE strategy. WESTPA has been successfully used to study multiple challenging biological processes and is being actively developed. The fact that WE doesn't bias the kinetics of the model makes it ideal to use with rule-based models and WESTPA being designed to interface with any stochastic simulator makes it ideal to use with BNG stochastic simulation methods.

## **Simplification of rare events at the cost of accuracy**

In bioinformatics modeling, researchers frequently avoid working with models that contains rare events either because the computational cost is too high or because it is relatively easy to simplify 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.

## **WEBNG**

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 to simulate and the tool creates a WESTPA simulation folder which can then be ran like any other WESTPA simulation. WEBNG 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.

## Features

### Simple installation

To lower the barrier to entry for this sampling method, we distribute WEBNG via the python package index. This simplifies the installation to a single command from the command line. 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 they are installed.

### Default templates to smoothen the learning curve for WE sampling (? bad but I'm unsure what else to write)

WEBNG comes with three subcommands that will setup the WESTPA simulation folder and analyze the resulting simulation in steps. First subcommand, **template**, allows the researcher to point to the BNGL model file they want to simulate and this subcommand will generate a YAML file that's easy to read and modify and is populated with reasonable default values, some of which are extracted from the model itself. 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.

### Straightforward WESTPA folder generation

The next subcommand, **setup**, takes in a single argument which is the YAML file generated in the previous step. This subcommand will use this information to generate the standard WESTPA simulation folder which is read to 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 options WEBNG currently supports, see WEBNG documentation ([link](#)). This

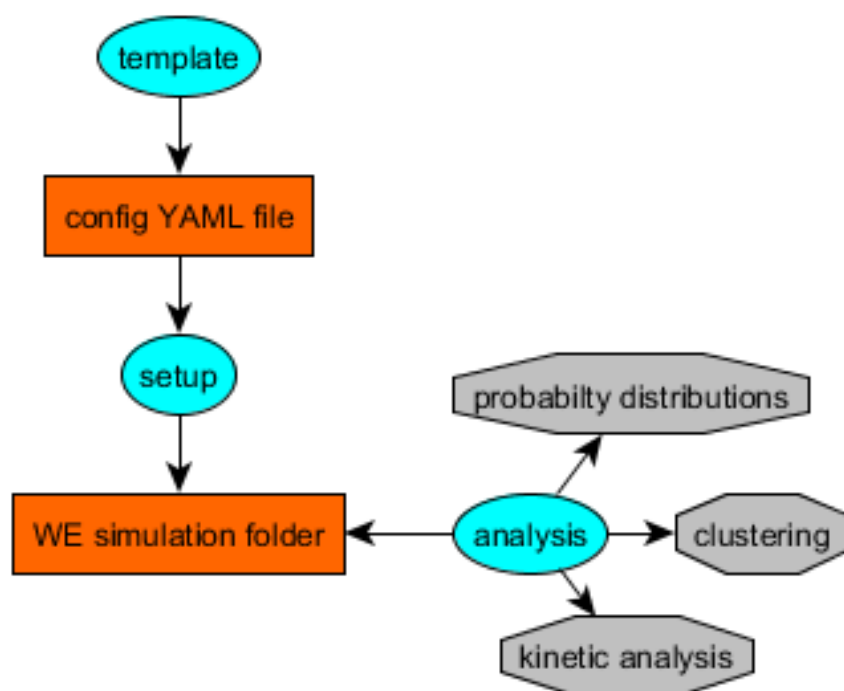


Figure 1: General workflow for WEBNG

workflow allows for a new user to quickly start simulating their BNG model with reasonable defaults while also allowing for an advanced WESTPA user to change simulation parameters and options freely.

## BNG model focused built-in analyses

Once the simulation is ran, the user can use the `analysis` subcommand, which also takes in the same YAML file generated in the `template` step. This subcommand will take in the parameters provided in the YAML file and run the enabled analyses on the simulation. The built-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) calculation of the transition matrix and PCCA+ clustering based on the transition matrix as well as calculation of rate constants between PCCA+ clusters (Fig1C) and (4) network graphs that visualize the calculated clusters and transitions between them (Fig1D). These analyses are complicated for a new user to implement since the user not only 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 identify important states in the progress coordinate space and calculate rate constants between them.

## Example published system

While designing WEBNG we used Tse et al. as a reference, aiming to be able to reproduce the results, quickly setup our simulation for the systems used in the paper and replicate some of the analyses done in a streamlined fashion. Tse et al. uses weighted ensemble to run stochastic simulations of multiple gene regulatory network models, identifies different phenotypes of the model and estimates mean first passage times between those phenotypes.

## ExMISA model

We have replicated the results of a two gene network motif that was studied in Tse et al. as an example and the results from the built-in analyses provided with WEBNG are shown in Fig1. This is a model of exclusive mutual inhibition and self-activation where there are two genes, generically named A and B, that encode transcription factors that activate their own transcription and repress the transcription of the other gene (Fig1A). The model includes the creation and degradation of the transcription factors as well as binding and unbinding of these transcriptions factors to the regulatory regions of both genes.

The analysis reveals 4 expected high probability regions (Fig1B), one region where both proteins are low in count (same notation as Tse et al., protein A count/protein B count, lo/lo), one where both proteins are high in count (hi/hi) and two regions where one protein is low and the other is high in count (hi/lo and lo/hi). The automated clustering correctly identifies each cluster correctly (Fig1C) and automated network visualization can visualize each state and transitions between them (Fig1D).

## **Helps reproducibility**

Another 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 (including Tse et al. study where the WE simulation is coded in Matlab) and contain hard-coded variables which makes reproducing results challenging 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 a consistent manner which makes results far more reproducible.

## **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 model processes that contain rare events using rule-based modeling 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 for WEBNG.

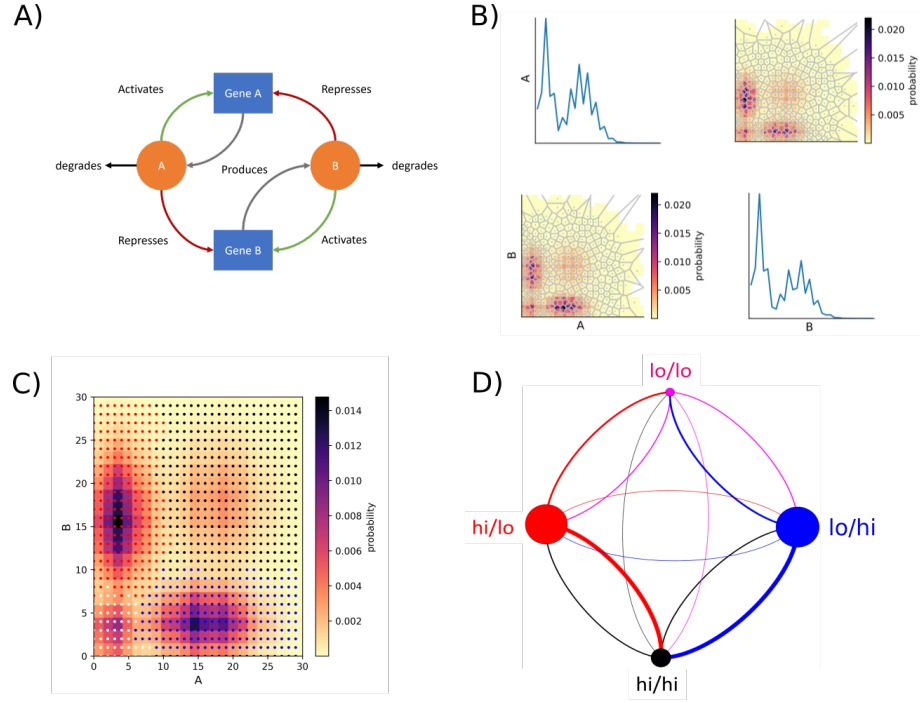


Figure 2: Results of automated analyses on a simulation of ExMISA that was ran for 500 WE iterations, replicated from Tse et al. study (A) Diagram of the ExMISA model (B) 2D heatmaps of joint probability distributions of protein A versus protein B counts. Lighter colors show less probable regions while darker colors show more probable regions. The probability distributions are estimated over the last 100 WE iterations. The grey lines indicate voronoi bins used in adaptive binning. (C) Automated clustering correctly identifies each high probability phenotype. Clusters are represented by colored dots overlaid on top of a 2D heatmap of joint probability distribution of protein A counts versus protein B counts. Each colored dot represents one discrete possible state the model is in and each color represents a different cluster. (D) Network graph generated off of the clustering done in the previous step. Thicker edges mean higher probability transitions between clusters and node sizes scale with the probability of each phenotype in steady state.