Readme for Ecoli_geneExpression_TASEP

This README will help guide you through running and using the Ecoli_geneExpression _TASEP GUI.

Running simulations:

First, the necessary rate parameters are specified (Fig. 1A).

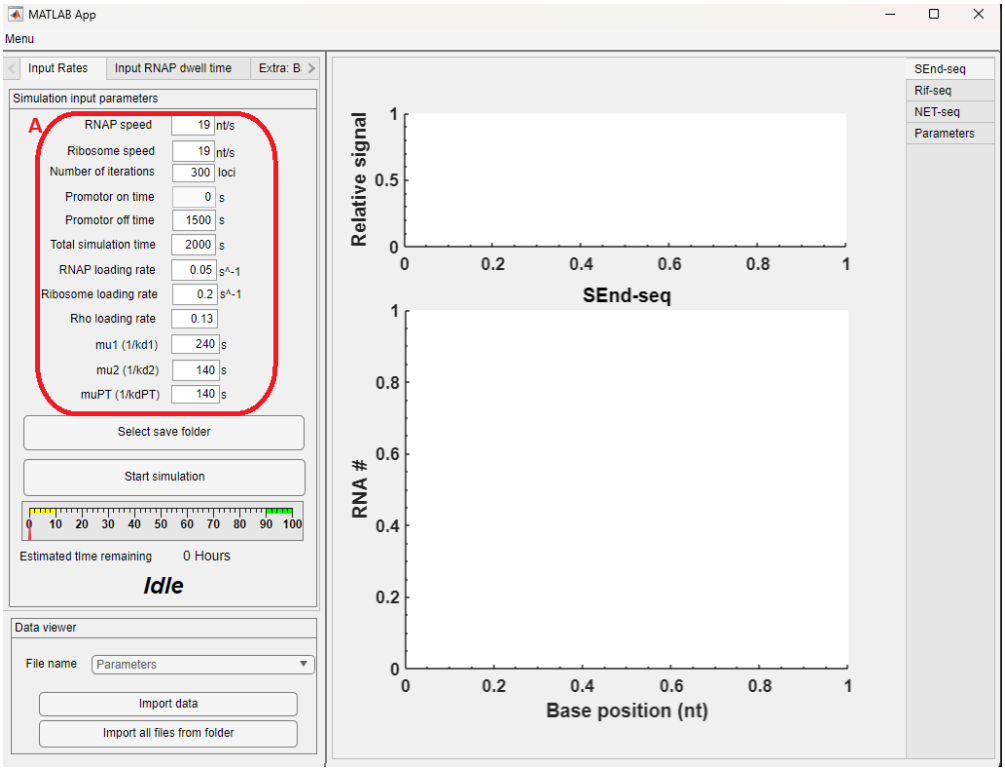

Figure 1

- RNAP and ribosome speeds are the average speed (gene length/total travel time).
- Number of iterations can be considered as the number of DNA loci to be averaged for the final output.
- Promoter on time is when the first RNAP loads (just like t = 0 in the IPTG induction experiment).
- Promoter off time is when the transcription initiation stops (like when glucose or rifampicin are added in experiments).
- Total simulation time is the last time processes are calculated.
- RNAP and ribosome loading rates are the rates at which these molecules are loaded at x = 0 position.
- Rho loading rate is for the probability of Rho loading. This input number is multiplied by the free RNA space (between the RNAP and the leading ribosome)

at a given moment and divided by the total gene length to determine the probability of Rho loading at that moment. If a random number (generated from a uniform distribution [0,1]) is lower than the probability, Rho loads at a random location in the free space. The Rho will translocate from 5' to 3' direction at a speed 5 times greater than the RNAP speed.

- Mu values are the mean lifetimes of different RNA species. They are inverse of the decay rates mentioned in the paper.

Second, RNAP pause profile for the simulation is specified (Fig. 2).
In the RNAP dwelltime tab, users provide a gene length and a profile of RNAP dwelltimes along the gene. The default setting is a uniform profile where RNAP spends a constant average dwell time (gene length/RNAP speed) at every nucleotide. A dwelltime distribution can be used by clicking the Import dwelltime distribution button (Fig. 2A). Selecting a CSV(e.g: NETSEQ_lacZ.csv) file with an intended dwell time distribution. Once the file is successfully imported, the RNAP profile will be displayed (Fig. 2B).

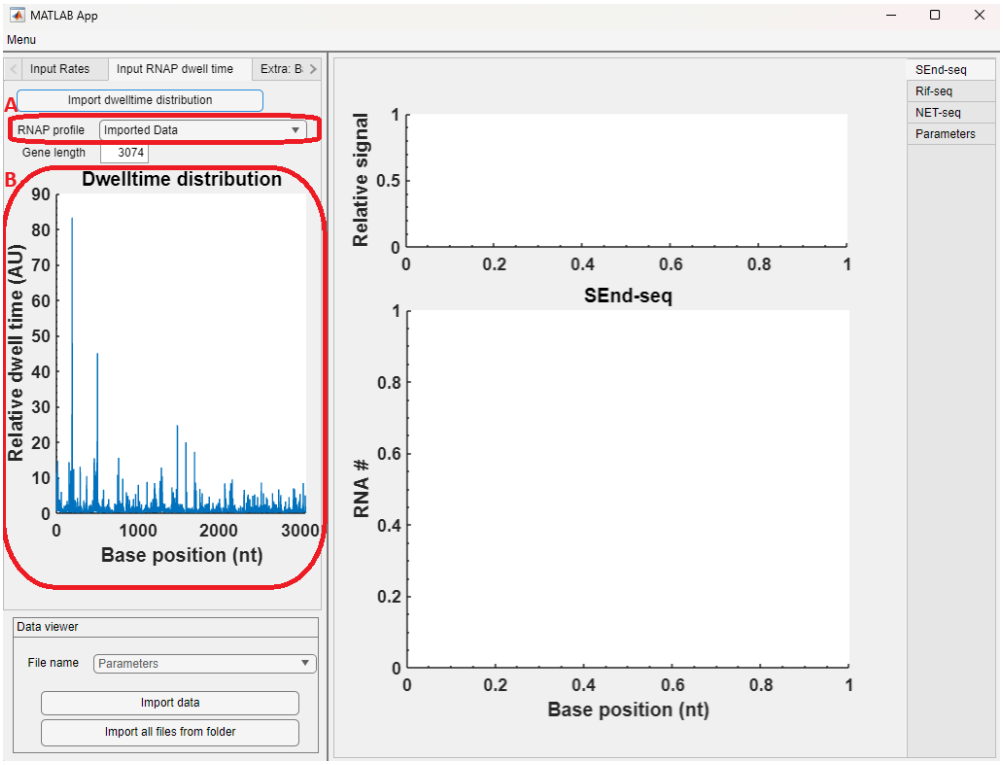

Figure 2

Note: Please make sure the file name of the dwelltime distribution is formatted as "NETSEQ_InsertgeneName.csv". We have included the lacZ NETSEQ profile, which can be used as a sample.

Lastly, select the save folder in the Input Rates tab (Fig. 3A) to determine where to save the simulation result.

To start the simulation, click the start simulation button (Fig. 3B). The Idle label will change to ”simulation running”. Once the simulation finishes running, this label should change to “Simulation done”. Simulation times can vary based on the parameters used.

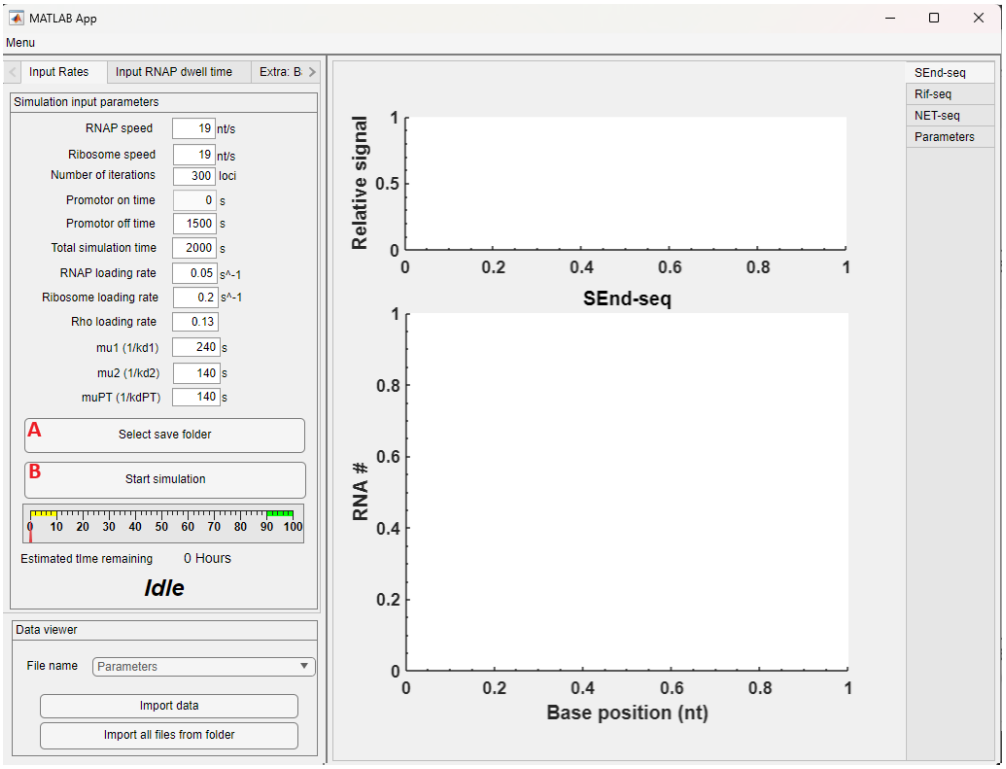

Figure 3

Extra: batch simulations:

In the **“**Extra: batch simulations” tab, users can set up multiple simulations to occur sequentially. Select a .csv file containing values for all the simulation parameters by clicking on the Import parameter file button (Fig. 4A). An example csv file has been provided which can be used as a template (\batchSimParameters.csv). Select the folder (\NETSEQ RNAP profile\) which contains the required dwelltime distributions to be simulated through the Select gene folder button (Fig. 4B). The table (Fig. 4C) should now be updated with the intended simulation runs. Click Start batch simulation to start the simulations.

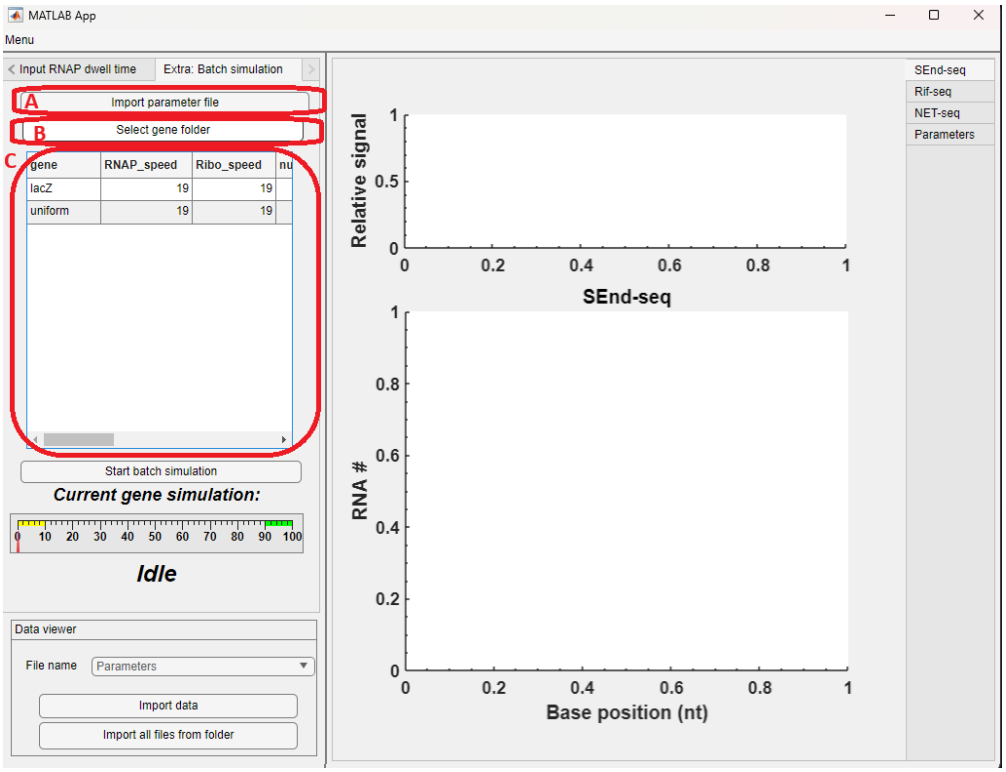

Figure4

Plotting simulated data:

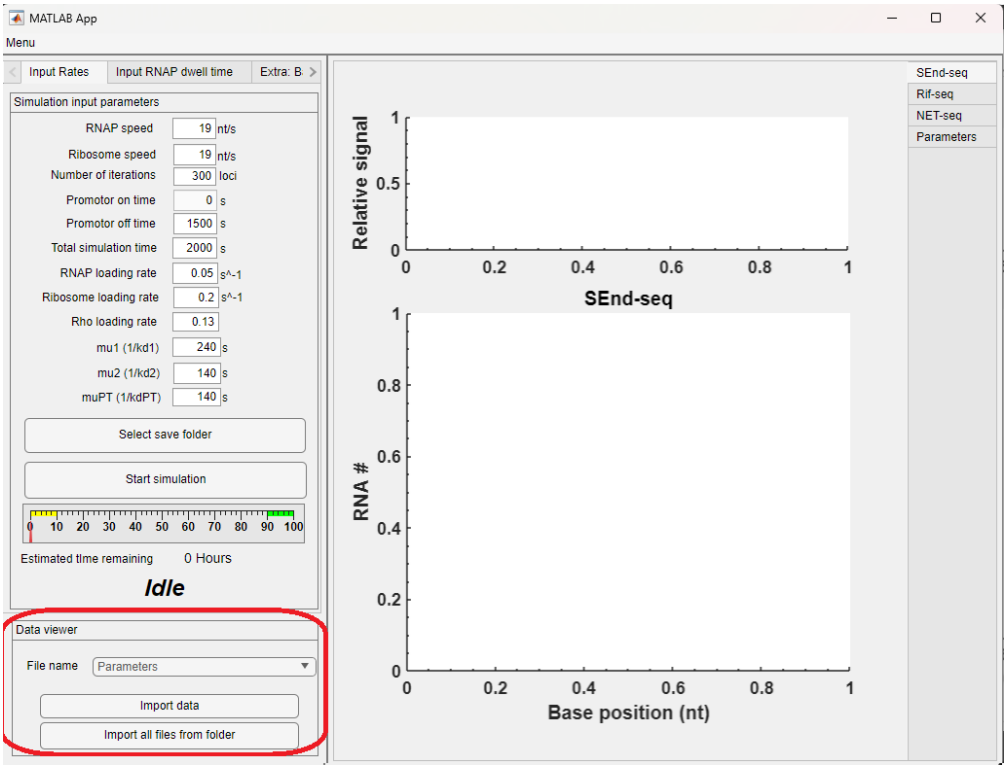

Figure 5

To run the analysis on previously simulated data, you can click on the ‘Import data’ button (Fig. 5) to select the simulation file. Alternatively, “Import all files from folder” is to import all simulation .mat files from a folder.

After the selection, a drop-down menu will show files that can be selected for data viewer (Fig. 6A).

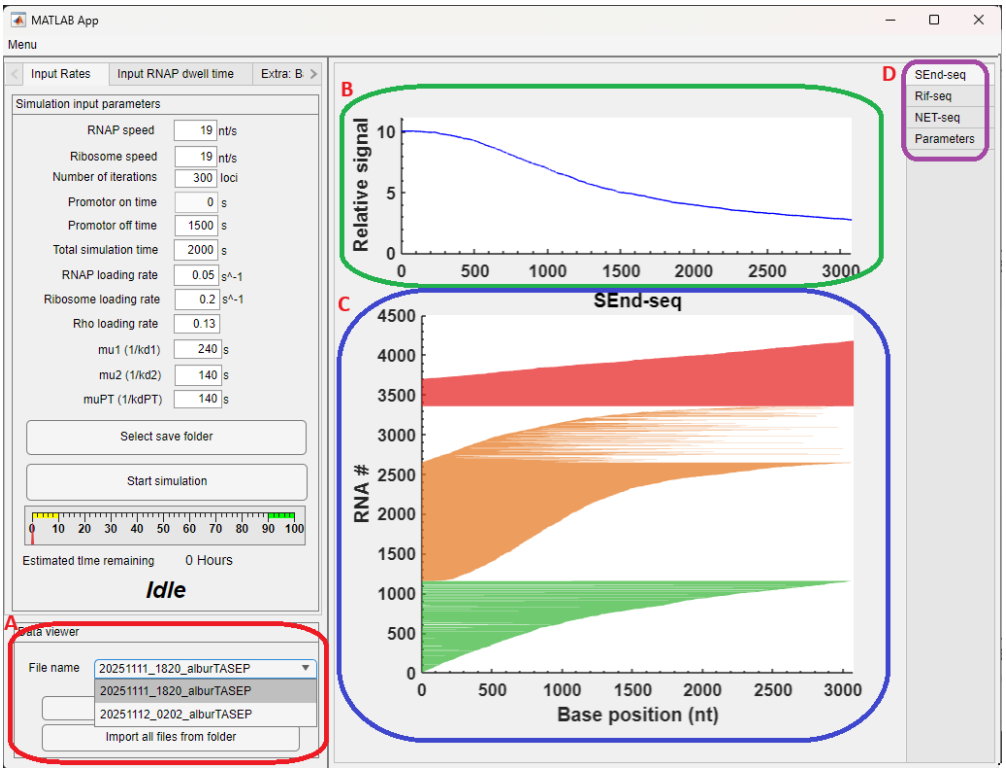

Figure 6

SEnd-seq results are shown with two graphs. The top graph (Fig. 6B) corresponds to the RNA coverage for the dataset, with the x-axis indicating base position.
The bottom graph (Fig. 6C) is individual RNAs as would be identified by SEnd-seq experiments. Green horizontal lines correspond to nascent RNA, orange lines correspond to RNA that have undergone premature termination, and red lines correspond to RNA that has finished full-length transcription.
We can navigate different types of analysis on the dataset by using the tabs on the right side (Fig. 6D).

Note: To view some of the SEnd-seq results shown in Supplementary Figure 1. We can choose the following datasets provided:
“\Simulated datasets\lacZ datasets\2_lacZ_alburTASEP” for weak RBS with lacZ dwelltime distribution.

“\Simulated datasets\uniform datasets\20251111_1820_alburTASEP” for weak RBS with uniform RNAP dwelltime distribution.
(Due to Github space limitation, we could only provide these data sets.)

Plotting multiple Rif-seq plots:
The Rif-seq tab displays the simulated Rif-seq graph for the current dataset (Fig. 7). To compare multiple datasets, click to check the hold graph checkbox in the blue box shown. Use the drop-down field in the red box to select the next dataset. The GUI should plot the new dataset on top of the current dataset.

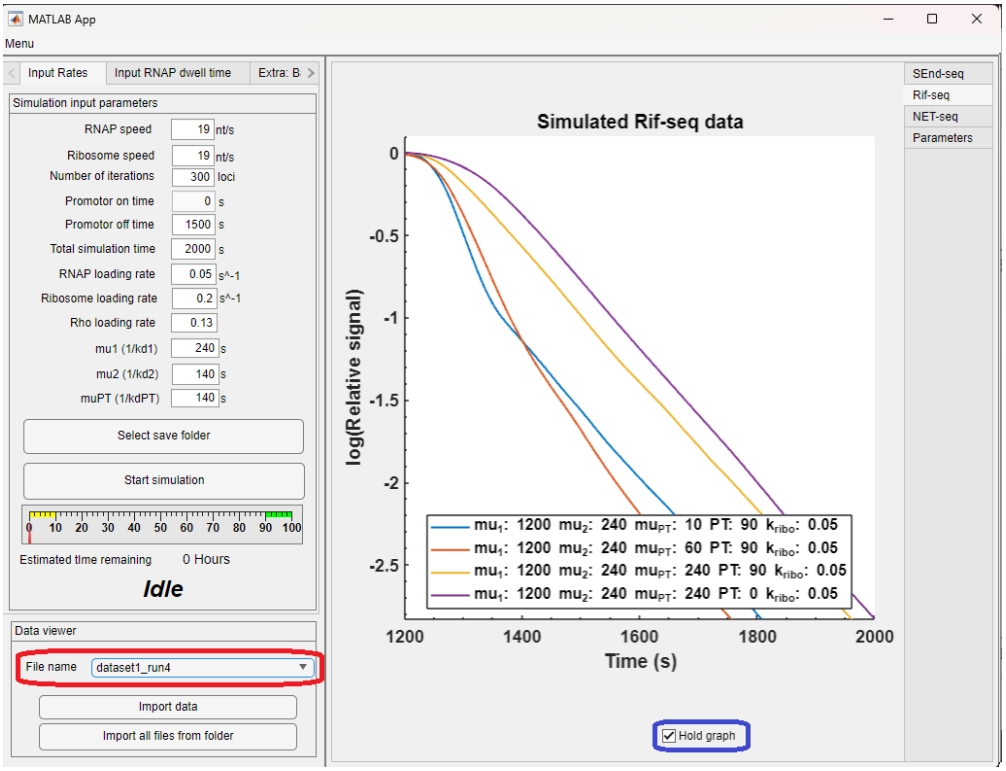

Figure 7

Note: To view Rif-seq results shown in Extended Data Figure 7h, please choose the following datasets provided:
“\Simulated datasets\Rif-seq datasets\dataset1_run1” for PT =0.9, kdPT = 1 min^{-1}.
“\Simulated datasets\Rif-seq datasetss\dataset1_run2” for PT =0.9, kdPT = 6 min^{-1}.
“\Simulated datasets\Rif-seq datasets\dataset1_run3” for PT =0.9, kdPT = 0.25 min^{-1}.