Ian Braun, Homework 1, BCB 4

1. Historically, what two factors have held back scientists from constructing high-quality whole-cell models?

The first factor is that historically not enough has been known about the list of which molecules and macromolecules were actually in the cell and the list of what interactions were taking place. In order to develop a whole-cell model, you first need to know enough of what the individual components are including the proteome, genome, transcriptome, metabolome, etc, and the tools to generate the datasets representing this knowledge are recent occurrences.

The second factor is that computationally accounting for the true nature of the complexity of how different interactions produce different phenotypes is simply not possible. To specify these reactions and their results in an unsimplified way would require an unrealistic amount of computing power, and simplifying the problem in a way that requires realistic resources but still accurately describes many true biological outcomes is very difficult.

2. What general strategy was used to construct the model in this paper?

The general strategy in making this whole-cell model was to break the cell down into submodules. The assumption used was that these submodules were independent over short periods of time, but dependent over longer periods of time. Therefore, the processes and components within them could be modelled in isolation for short runs before sharing the resulting information with all the submodules and getting a new combined state for all the parameters across the board.

3. After it was constructed, how was the model validated experimentally? Is there enough data to validate all the parameters defined in the model?

Data sets, containing values for the model parameters, which were not used to build or design the model itself were used to test the performance of the model. In other words, a set of starting conditions were provided the parameters of the model, a simulation run, and the resulting parameterized values and final predicted states were observed, and compared to the true biologically recorded final states given by the datasets. For many parameters, the model predictions were accurate to the actual experimental values from the datasets, but not enough data was present to validate all the parameters. As seen in Figure 2F, some predicted values have no experimental counterpart from the experimental datasets so we cannot gauge how the model performed on these parameters.

4. Give an example of how this model was used to for the purposes of predictive biology and explain and interpret the results.

One example of the type of predictive biology that this whole-cell model was used for was predicting the activity of DNA-binding proteins in the cell over time in the cell cycle. Their model predicted both the overall quantity of proteins interacting with chromosomal DNA at any given time point, as well as the identify of the individual reactions. They used this information to extract biological facts, such as that 50% of the chromosome is bound at least once in the first 6 minutes of the cycle, and 90% at least once in the first 20 minutes of the cell cycle. They also extracted what collisions were taking place between proteins attempting to bind the same DNA regions, finding that the most common collisions were between RNA polymerase taking the place of structural maintenance proteins. This prediction is interesting, and is exactly the kind of information that is useful to extract from a whole-cell model, because it is an emergent property that also makes biological sense, and biologists can now try and form hypotheses and expand on how maintenance is paused or abandoned when a region is transcribed, a research avenue that might have not been considered prior to the model.