

## General planning.

### Avenues to explore in data.

Biology by the numbers in *E. coli*:

1. General overview/outline of proteome. Are the numbers compatible with prior expectations?
2. Things I had written down to explore: growth rate versus volume, Protein density
3. 'Building *E. coli*': make use of known fluxes, reaction rates, protein numbers, growth rates - do the numbers make sense? Also of interest to see how they vary with growth rate.
  - (a) replication: DNA polymerase; other proteins needed to find origin.
  - (b) transcription: RNA polymerase, sigma factors
  - (c) translation: ribosomal proteins, elongation factor, ...
  - (d) lipid synthesis
  - (e) nutrient uptake
  - (f) energy utilization and conversion

Additional insight from the data:

1. Knowing the unknown: One example here is the high variability in copy number for a large fraction of genes that are devoid of any regulatory annotation.
2. Observation that ratio of transcription factor copy number to (expected) DNA context is essentially constant across growth conditions.

Quantitative predictions:

1. Essential proteome: What fraction; how does it vary with growth rate?
2. Biological process network analysis. Can we understand how certain subsets/classes of the data are robust to changes in growth rate given knowledge of their biological processes?
3. Can proteome mass fraction occupancy comport with known kinetic properties? Are "slow" biological processes more highly represented in the proteome compared to the "fast" processes?
4. Can constraints on proteome occupancy be mapped to requirements for biological cofactors? Can get information about associated cofactors from the Uniprot databases.

### What other datasets might be valuable

There are additional data sets on protein copy number from Gene-Wei Li 2014, Taniguchi 2010. GC also had another proteomic dataset?. EcoCyc of course. What are other resources?

### Summary of what others have done already.

Papers:

1. 1.