Estimating the errors in r_{cat}

Flux Variability Analysis

One of the main concerns in our approach for calculating the maximal catalytic rate of an enzyme, is how to get a representative flux value to plug in the equation:

$$r_{\mathrm{cat}} = \frac{v_i}{E_i} \tag{1}$$

Currently, we solve parsimonious FBA (pFBA), which minimizes the total sum of fluxes. It was shown that pFBA agrees with trnascriptional data from E. coli in 97% of the time, and is considered a reasonable approch from flux predictions among the FBA community.
However, the flux values we get for each reaction are random sample from a possible flux distribution and may not represent the "real" flux through a given reaction. To reduce biases in our r_{cat} estimations, we can consider parsimonious *FVA*, which returns a range of fluxes for each reaction. It works in the following manner:

(1) solve pFBA
(2) set all reaction boundaries accordingly
(3) for each reaction in the model:
  ($3a$) minimize the reaction flux (v_{min})
  ($3b$) maximize the reaction flux (v_{min})

The range of flux for reaction i is then:

$$v_{\mathrm{max},i} - v_{\mathrm{min},i} \tag{2}$$

$$\tag{3}$$

Taking the mean of the v range may be a better approach for r_{cat} estimations. Error bars (or more precisely, confidence intervals) can be described as:

$$\overline{r_{\mathrm{cat},i}} = \frac{v_{\mathrm{max},i}}{E_i} \tag{4}$$

$$\underline{r_{\mathrm{cat},i}} = \frac{v_{\mathrm{min},i}}{E_i} \tag{5}$$

$$\tag{6}$$

Most of our reactions seem to have a very small dynamic range of fluxes, resulting in almost zero flux-associated errors (figure 1). Usually, if a reactions have small flux variability it means that they are biomass related, thus must support a given amount of flux to allow growth. Nevertheless, this is not the

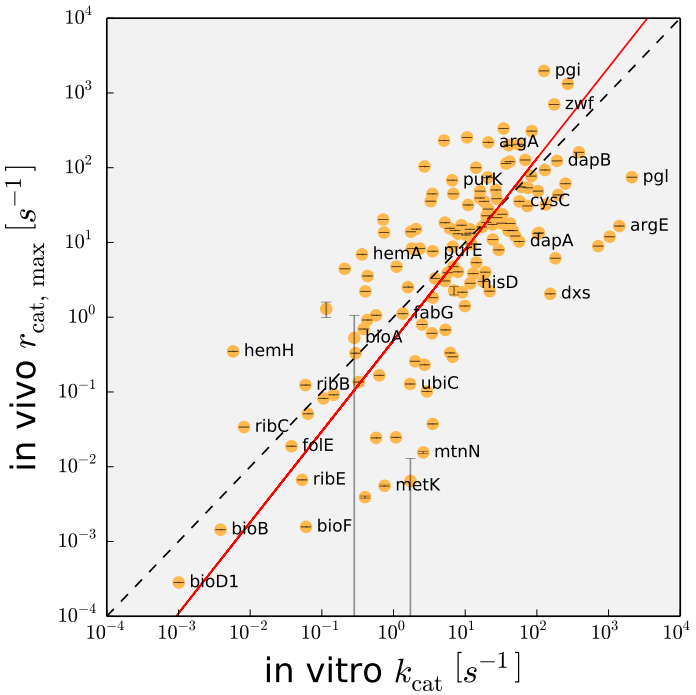
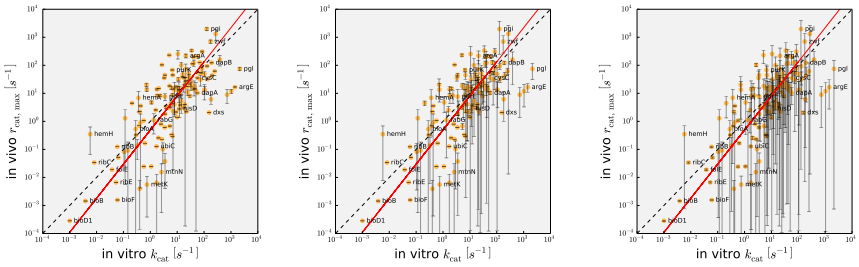


Figure 1: k_{cat} to $r_{\mathrm{cat}}^{\mathrm{max}}$ correlation using v_{mean} as flux value for $r_{\mathrm{cat}}^{\mathrm{max}}$ calculation. Dashed black line represents $y = x$ and red line represents best fit by orthogonal regression. correlation: $r^2 = 0.556$, $p_{val} < 10^{-24}$, rmse $= 0.6$



(a) relaxation = 0.1% (b) relaxation = 1% (c) relaxation = 10%

Figure 2

case for most of the 135 reactions we analyze (or at least I cannot see how they are biomass related). It is possible that the minimization of fluxes constrain, on top of the other constrains (growth rate and oxygen uptake rate) leave very little room for flux deviations. We can further relax the parsimonious constraint, that is, allow the sum of fluxes to be x% larger. A common relaxing factor is in the range of 0.1-5%. Figure 2

Standard deviation of the maximum

We can ask how reliable $r_{\text{cat}}^{\max}$ is. It may be that the maximal catalytic rate across all conditions. For example if the enzyme copy number in a particular condition was reported to be extremely small, $r_{\text{cat}}^{\max}$ will be disproportionally high. For this, we can calculate the standard deviation of the top three r_{cat} values:

$$\text{std}([r_{max}^{1st}, r_{max}^{2nd}, r_{max}^{3rd}]),$$

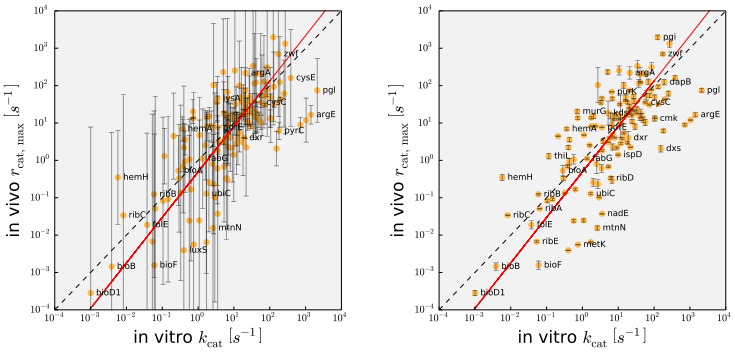
Figure 3a shows the resulting error bars. A reason for such large deviations is that the catalytic rate of many enzymes is log normal distributed, thus taking the std of the data results in huge errors. We can consider using:

$$\text{std}(\log([r_{\text{max}}^{1st}, r_{\text{max}}^{2nd}, r_{\text{max}}^{3rd}])),$$

yet then the errors are very small as we are only using three numbers (fig 3b).

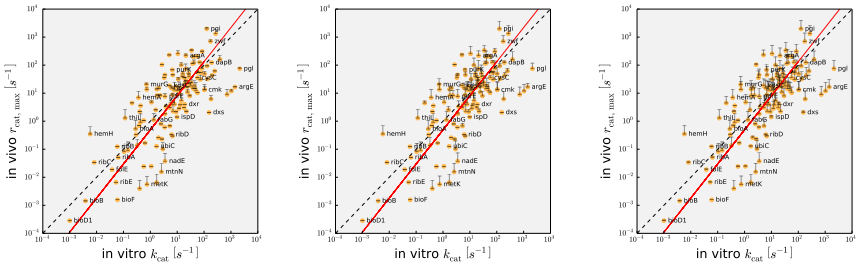
What Next?

What should we do? Maybe we can only show the positive error bars in figure 2, assuming that the minimal flux is not really biologically feasible, i.e., the algorithm may be using "irrelevant" metabolic pathways, allowing it to minimize the flux through the reactions significantly. It will look like this (fig 4):



(a) log of standard deviations (b) standard deviations of log

Figure 3



(a) relaxation = 0.1% (b) relaxation = 1% (c) relaxation = 10%

Figure 4