Beyond fluorescein:

Using fluorescent protein calibrants for direct and absolute quantification of protein production in synthetic biology

Eszter Csibra and Guy-Bart Stan

Centre for Synthetic Biology (CSynB), Imperial College London

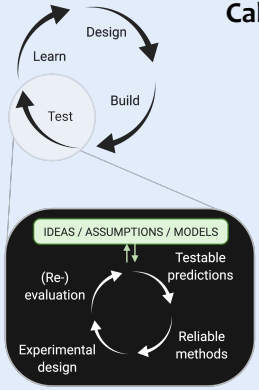
e.csibra@imperial.ac.uk

@eszter_cs

Imperial College London

UKRI Engineering and Physical Sciences Research Council

Royal Academy of Engineering



Calibration for quantitative analysis of fluorescence

While inter-lab calibration standards are approaching mainstream usage in synthetic biology (Beal et al., 2018; 2020), such calibrations are not in fact sufficient for absolute protein quantification required for modelling synthetic circuits. Fluorescein-based calibration of plate reader and flow cytometry instruments allows the measurement of green fluorescent protein (GFP) in synthetic cells to graduate from arbitrary units to calibrated units, but retains important caveats. Fluorescein is only a good calibrant for green FPs, leaving other FPs uncalibrated, and only provides conversions to units of brightness, not to molecule numbers.

Ideal assay calibrants in molecular biology consist of the same molecule as the one to be measured – in this case, a purified preparation of the appropriate fluorescent protein. Here we show that by using purified FP calibrants, all protein species in a synthetic circuit can be quantified in absolute terms using no advanced instrumentation. We develop a SEVA (Standardised European Vector Architecture)-based expression vector that allows the high-level production of soluble protein and describe a straightforward protocol for the purification of micrograms of FP, followed by a calibration that relates fluorescence activity to protein mass.

Calibration Type	NONE	FLUORESCEIN	FLUORESCENT PROTEIN
Protein Calibrants			
BFP	none	none	BFP
GFP (dim)	none	Fluorescein	GFP (dim)
GFP (bright)	none	Fluorescein	GFP (bright)
RFP (ex 590)	none	none	RFP (ex 590)
RFP (ex 620)	none	none	RFP (ex 620)
Comparison capability			

Why count proteins?

Mathematical modelling of synthetic circuits relies on our ability to quantitatively characterise gene expression. One of the biggest limitations is our inability to measure circuit components easily, affordably and accurately. While omics technologies allow us to exhaustively characterise a circuit or cell state at one point in time, their use is limited by cost and the need for continuous monitoring of dynamic processes with minimal manual intervention. Fluorescent protein monitoring is cheap, reliable and already part of most synthetic biologists' workflows.

Beyond synthetic circuits, characterising absolute protein number has broad applications. Fluorescent monitoring is easier, quicker and higher throughput than Western blotting, and working at the wrong order of magnitude (100 vs 100,000 molecules/cell) can have important implications for experimentation and engineering (Ceroni et al., 2015).

Discussion

While certainty about protein numbers on a single-cell level will ultimately require single-cell techniques, and cell-to-cell variation remains an important consideration, assumptions about the relative uniformity of bacterial cultures are valid for many circuits and can be verified by flow cytometry. The low cost and high throughput of batch techniques means that microplate assays remain an important screening platform, necessitating the development of methods for extracting informative numbers from such data.

We show here that simply by making appropriate calibrants, ballpark molecule-per-cell values are achievable without specialised instrumentation or expense, and that the information gained from such efforts can be important for: (1) easing experimentation by allowing comparison of experiments from different instruments, (2) giving us the tools to evaluate microplate protocols with quantitative information, and (3) enabling circuit debugging required for the Design Build Test cycle.

References

Wu, 2015. BioEssays. What is the total number of protein molecules per cell volume? A call to rethink some published values. 2015 Dec; 35(12): 1050–1055. DOI: 10.1002/bies.201300066

Ceroni et al., 2015. Quantifying cellular capacity identifies gene expression designs with reduced burden. 12(5): 415–418. DOI: 10.1038/NMETH.3339.

Beal et al., 2018. Quantification of bacterial fluorescence using independent calibrants. PLoS ONE 13(8): e0199432. DOI: 10.1371/journal.pone.0199432.

Beal et al., 2020. Robust estimation of bacterial cell count from optical density. Communications Biology. DOI: 10.1038/s42003-020-01127-5.

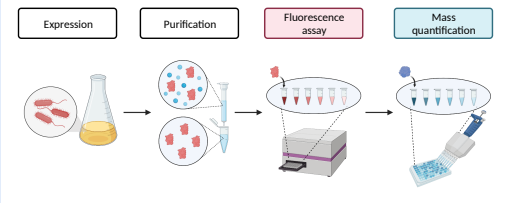
Fedorec, Robinson et al., 2020. FlopR: An Open Source Software Package for Calibration and Normalization of Plate Reader and Flow Cytometry Data. ACS Synth. Biol, 9, 2258–2266. DOI: 10.1021/acssynbio.0c00236.

Hecht et al., 2016. When Wavelengths Collide: Bias in Cell Abundance Measurements Due to Expressed Fluorescent Proteins. ACS Synth. Biol. 2016, 5, 1024–1027. DOI: 10.1021/acssynbio.6b00072.

Figures created with RStudio and BioRender.com. Fluorescent protein spectrum data from FPbase.

Fluorescence Protein Calibration Method

WET LAB

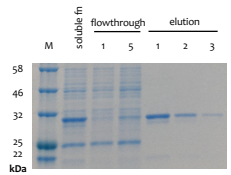


Protein purification need not be as complex as many newcomers to the technique think, particularly for stable proteins like FPs. This protocol uses standard techniques with commonly available reagents. A standardised expression vector was constructed from an arabinose-inducible His-tagged FP construct in a high copy SEVA vector. Protein is purified using a His-Cobalt matrix. No specialised equipment is needed to produce micrograms of protein, and all FPs trialled purify with high yield and purity.

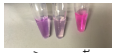
FPs are serially diluted for fluorescence measurements for all instruments and filter channels required, followed by quantification to determine protein concentration. Several protein quantification assays exist. Here, we use the bicinchoninic acid (BCA) assay for its sensitivity, ease of use and low protein-to-protein variability. Estimates from BCA assays are cross-validated using complementary techniques such as SDS-PAGE based Coomassie staining.

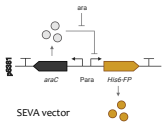
Typical fluorescent protein purification.

Purifications yield high % recovery and pure protein

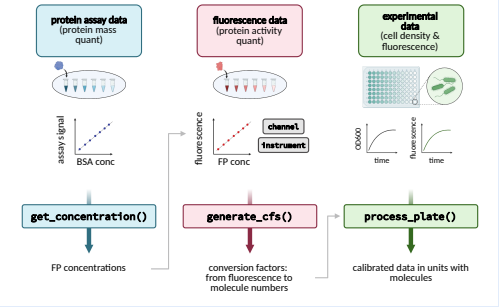


Protein observable in elutions



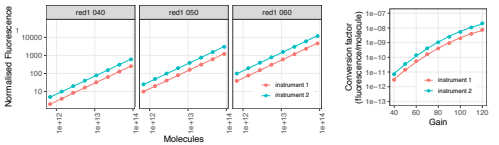


DRY LAB



Data analysis has been assembled into an R package consisting of automatable functions for the analysis of each data set. First, FP concentration can be determined from the protein assay data. This generates a csv file which can be incorporated into the fluorescence assay data to generate conversion factors from fluorescence to molecule number for each condition calibrated. Finally, these conversion factors, along with others for cell density, can be used to generate values in fluorescence protein molecules per cell from microplate data. The analysis methodology is based on the approach of the recent flopR package developed for fluorescein calibrations (Fedorec, Robinson *et al.*, 2020).

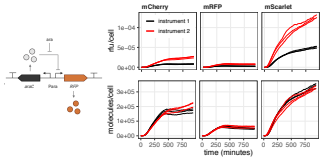
Conversion factors may be calculated by relating fluorescence to molecule number



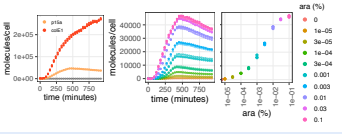
Method Validation

Quantification of RedFP circuits in molecules/cell can be used to compare between genetic variants, experimental conditions and instruments

Demonstration: Calibrations allow the comparison of three different RFPs across instruments despite different relative fluorescence values.



Characterisation of mCherry protein level with a SEVA expression vector shows protein levels per cell fall in the 10^2 to 10^3 range

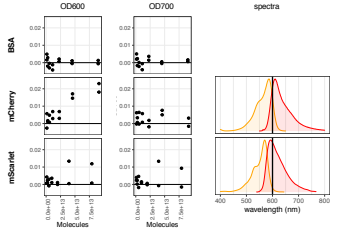


Evaluation of protocols

Calibration allows for a re-evaluation of assumptions about microplate assay quantification

Absorbance of RFPs at 600nm has been suggested to interfere with bacterial density estimation, typically assessed as the optical density (absorbance + scatter) of cultures at 600nm ('OD600') (Hecht *et al.*, 2016), however, it has never been quantified how many molecules of RFPs might be required for this effect to occur.

Using purified RFPs, it is possible to show that mCherry absorbance in one microplate reader approximates ~0.02 OD600 units for ~10^{14} molecules. For a typical cell density of $2x10^9$ cells per well (OD~0.45), an intermediate mCherry level of $5*10^5$ proteins/cell would be expected to increase apparent OD600 by only 0.0002 (0.04%). It would require overexpression to $5*10^6$ proteins/cell to increase the OD by a significant amount (0.02, or 4%), which would make mCherry the most abundant protein in the cell and ~25% of the proteome (Milo, 2013; http://book.bionumbers.org/).



Circuit Quantification

Calibration allows for a direct comparison of multiple protein species in the same circuit

Using a circuit consisting two fluorescently labelled components, calibration allows us to quantify the molecular levels of both, showing that the Ptac promoter results in 100s of molecules even in the absence of IPTG and suggesting that self degradation is one route to controlling transcriptional leakage.