Comparative Approaches for Quantification of Product Yield in the Production of Recombinant Green Fluorescent Protein (GFPuv) Expressed in *E. coli*

By Gabrielle Rusch[1,2], Junhyeong Wang[3], Keith Breau[4], Katie Kilgour[5], Gary Gilleskie[6], Jeff Keele[6], Kurt Selle[6], Scott T. Magness[1], Stefano Menegatti[5,6]*, and Michael Daniele[1,2]*

[1] Joint Department of Biomedical Engineering
North Carolina State University and University of North Carolina, Chapel Hill
911 Oval Dr., Raleigh, NC 27695 (USA)

[2] Comparative Medicine Institute
North Carolina State University
1060 William Moore Dr., Raleigh, NC 27606 (USA)

[3] Department of Electrical & Computer Engineering
North Carolina State University
890 Oval Dr., Raleigh NC, 27695 (USA)

[4] Department of Cell Biology and Physiology
University of North Carolina, Chapel Hill
111 Mason Farm Road, Chapel Hill, NC 27599 (USA)

[5] Department of Chemical Engineering
North Carolina State University
911 Partners Way, Raleigh, NC 27695 (USA)

[6] Golden LEAF Biomanufacturing Training and Education Center (BTEC)
NC State University, Raleigh, NC 27695 (USA)

[*] Correspondence: mdaniel6@ncsu.edu, smenega@ncsu.edu

ABSTRACT

Process Analytical Technologies (PAT) are critical for efficient and automated bioprocessing; moreover, sensor and assay-driven automation will be necessary for the realization of Industry 4.0. Herein, we demonstrate methods for analyzing product yield from a pilot-scale bioreactor (300L) and conduct cross-method comparisons. We used a system of recombinant green fluorescent protein (GFPuv) expressed in *Escherichia coli* (*E. coli*), which is a simple, cost-effective model for evaluation at both laboratory and pilot scales. By comparing inline bioreactor measurements, plate reader assays, and a novel image analysis pipeline, we identify optimal harvest timelines and demonstrate the strengths and limitations of each technique. Results indicate peak cell density and GFPuv expression between 12.5h and 18h post-inoculation, with declining viability thereafter. Comparisons across techniques suggest that imaging methods may be more effective in capturing adverse outcomes, such as increased membrane permeability or cell death, which typically occur at later stages of operation. This integrated analysis offers actionable insights for optimizing biomanufacturing workflows and advances the development of scalable PAT applications, bridging the gap between laboratory research and industrial implementation.

KEYWORDS:

pilot-scale bioreactor; green fluorescent protein; GFPuv; process analytical technologies; fluorescence microscopy; image analysis

1. INTRODUCTION

As bioprocessing technologies advance and continue to scale from laboratory settings to industrial operations, the need for robust analytical methods to monitor and control yield becomes increasingly critical. Inline, at-line, and online Process Analytical Technologies (PAT) have emerged as essential tools for real-time monitoring, enabling enhanced process understanding and control [1]. However, the diversity of available PAT methods—spanning spectroscopic, chromatographic, and imaging-based techniques—presents a challenge in determining the most effective approaches for specific applications. Comparative assessments of these technologies are vital to identify their strengths, limitations, and suitability for diverse bioprocessing scenarios, particularly in the context of industrial demands for precision and scalability. Such comparative studies are not only fundamental for optimizing current biomanufacturing workflows but also for laying the groundwork for transformative innovations in Industry 4.0, including digital twins and predictive modeling [2-5].

For example, digital twins, as virtual representations of physical processes, are fundamentally dependent on the continuous flow of accurate and dynamic data, which in turn requires the application of precise and reliable PAT. In order to effectively inform and drive these digital twin systems, it is imperative to develop a robust understanding of how various PAT techniques perform under diverse operational conditions. Such knowledge enables the thoughtful weighting and integration of their outputs into digital twin frameworks and predictive models, ensuring their relevance and efficacy in optimizing biomanufacturing processes. For instance, as the industry progresses toward more sophisticated digital twin systems and predictive modeling, imaging datasets—such as those proposed in this study—will increasingly play a pivotal role, offering critical insights that bridge the gap between real-time process monitoring and advanced computational modeling [6-9].

The use of green fluorescent protein-expressing *Escherichia coli* (GFPuv-*E.coli*) represents a versatile and powerful model system for studying bioprocesses and recombinant protein biomanufacturing [10-17]. *E. coli* is widely utilized in the commercial production of a diverse array of proteins, including industrial enzymes, diagnostic reagents, and therapeutic molecules such as insulin, growth factors, and interferons [18, 19]. The inherent scalability and ease of genetic manipulation of *E. coli* solidify its status as a preferred platform for high-yield expression systems, facilitating the efficient biomanufacturing of complex recombinant proteins for both pharmaceutical and biotechnological applications [20]. Induction of GFPuv expression in *E. coli* can be realized by the introduction of isopropyl β-D-1-thiogalactopyranoside (IPTG) or temperature change, which accelerates fluorescent marker expression and proliferation simultaneously; nevertheless, while IPTG and temperature change are popular induction methods, it should be noted that induction conditions are highly dependent on expression system [23-25]. This dual-functionality makes GFPuv-*E.coli* an exemplary model for optimizing various biomanufacturing parameters, spanning from induction strategies and nutrient optimization to evaluating the influence of environmental conditions on yield [26-28]. In total, GFPuv-*E.coli* offers a close approximation to key aspects of industrial recombinant protein production, such as transcriptional and translational efficiency, metabolic burden, and stress responses. By genetically fusing GFP or GFPuv to the protein of interest or co-expressing it in parallel, researchers can dynamically track both the temporal and spatial aspects of protein production throughout the bioprocess continuum. In the GFPuv-*E. coli* model, GFPuv provides a direct, quantitative measure of protein yield, facilitating rapid assessment of culture conditions. As a highly quantifiable reporter protein, GFP enables real-time monitoring of expression kinetics, cellular proliferation, and metabolic activity, offering critical insights into bioprocess optimization and control. [21, 22, 29-31].

Although GFP is easily measured and GFPuv-*E.coli* is widely acknowledged as a reliable model for assessing bioprocess performance, the differences in sensitivity, resolution, and scalability across various analytical techniques for quantifying product yield can lead to discrepancies in the measurement of protein expression levels, complicating efforts to optimize bioprocess parameters. Spectrophotometric techniques, such as inline bioreactor probes, plate reader assays, and microscopy-based measurements, each offer distinct advantages and limitations. Inline probes allow for real-time monitoring but may lack the sensitivity to capture population or spatial heterogeneity [15, 32]. Plate readers facilitate higher-throughput analysis, yet are limited to bulk measurements, potentially missing localized variations within the bioreactor. On the other hand, advanced microscopy techniques offer single-cell resolution and can capture detailed spatial distributions, though scalability for continuous monitoring in industrial processes remains a concern [33]. These differences in methodology can lead to discrepancies in GFP expression levels, underscoring the need to carefully select analytical techniques based on the specific goals and scale of the bioprocess. Addressing these variations is crucial for ensuring reliable and accurate yield assessments, thereby optimizing both research and industrial applications. Combining and comparing these two perspectives of single cell and batch level analysis is important for accurate analysis of the product, while traditional methods of PAT typically focus on one method of analysis [34-38].

This study systematically compares and contrasts multiple approaches for quantifying cell proliferation and GFPuv expression over time, as illustrated in **Figure 1**. Using GFPuv-*E.coli* cultivated under industrial conditions, we assessed commonly employed analytical techniques in biomanufacturing. While GFP and these PAT techniques are well-documented in research, their continued evaluation and comparison hold utility for enhancing bioprocessing performance, but more

notably for identifying the degree of variation and limitations, which are critical yet underreported needs for realizing and implementing data driven control and prediction innovations, such as digital twins and continuous production paradigms. By comparing and refining quantification methods for GFP yield, this work contributes to the development of more precise, data-driven biomanufacturing systems, supporting their readiness for the future of industrial biotechnology and aligning with the evolving goals of Industry 4.0.

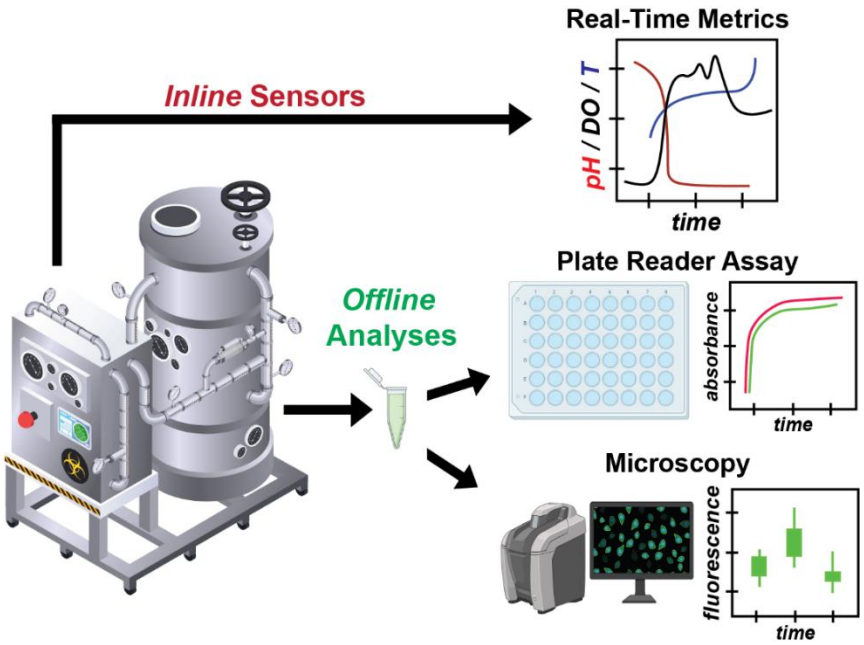

Fig 1. *Overview of inline and offline analysis for this bioprocess. Data was collected from several in line sensor modules to be compared against offline analytics including OD_{600} measurements from a plate reader and confocal image analysis to better understand the bioreactor performance.*

2. MATERIALS & METHODS

2.1 Production of GFPuv-*E.coli* and Sample Collection

A 300 L fed-batch bioreactor used for all processes. Following basic sterilization and preparatory steps, at 0 h, the bioreactor was inoculated with GFPuv-expressing *E.coli* (strain = BL21 (DE3)::pET17b::gfpuv). At 6 h, GFPuv expression was induced by introducing isopropyl β-D-1-thiogalactopyranoside (IPTG) into the culture medium. Culture continued for 24h post inoculation and was terminated upon harvest. Inline monitoring was performed throughout the 24-hour bioreactor run to measure several critical parameters, including optical density (OD_{600}), glucose concentration, pH, percentage of dissolved oxygen (DO%), temperature, and system weight. The culture was continuously agitated, with mixing speed (rotations per minute, rpm) and air sparging rate (standard liters per minute, slpm) recorded to ensure consistent aeration and homogeneity within the reactor. Two separate bioreactor runs were conducted. For both bioreactor runs, 2 x 1 mL samples were collected every hour between inoculation and induction. Post-induction, 2 x 1 mL samples were collected every 30 min. Samples were stored at -80ºC prior to analysis. The timeline for sample collection is presented in **Figure 3D**.

2.2 Plate Reader Assays

OD_{600} and fluorescence intensity were measured on a Cytiva XYB98 plate reader. Fluorescence intensity was measured at λ_{ex} = 395 nm and λ_{em} = 508 nm. Each sample underwent thawing and 100 µL of each sample was dispensed into a black 96-well plate with a clear bottom for analysis. Technical triplicates were prepared and analyzed for each sample from both bioreactor runs. Following sample preparation and initial OD_{600} and fluorescence measurements were recorded, the plates were centrifuged at 4000 xg for 3 min. Post-centrifugation, the supernatant was aspirated and plated alongside the pellet. OD_{600} and fluorescence measurements were then recorded.

2.3 Confocal Microscopy and Image Analysis

GFPuv-*E.coli* cells were also analyzed using confocal microscopy as a secondary offline analytical technique. Laser scanning confocal microscope (Leica Stellaris8, water immersion 63x) was used for all imaging. Prior to imaging, samples were thawed and 500 µL of each sample was removed, stained with CellBrite 640, fixed using 4% paraformaldehyde (PFA), and washed with phosphate buffered saline (PBS) following standard protocols. After fixing, staining, and washing, each sample was resuspended in 5 mL of PBS and further diluted 1:200 for optimal imaging density. However, sample S1 at the

0 h collection contained a low cell count for both runs, necessitating a 1:20 dilution instead of the standard 1:200 dilution. For ease of imaging multiple samples, basic flow cells were fabricated using polydimethylsiloxane (PDMS) sealed to glass slides as shown in **Figure 2**. Each well of the fabricated flow cell was filled with 25 µL of the prepared and diluted sample. GFP images were collected at λ_{ex} = 488 nm; λ_{em} = 507 nm set at 50.3% gain and 1.5% intensity, while the CellBrite 640 nm membrane stain images were collected at λ_{ex} = 642 nm; λ_{em} = 662 nm adjusted between 16-25% gain and 1.5% intensity for image acquisition. The membrane stain facilitated the identification of individual *E. coli* cells regardless of GFP expression to construct an imaging mask. When collecting CellBrite stain images, the gain and intensity were modified slightly from sample to sample, however the GFP images were collected at the same gain and intensity to allow for appropriate comparison of relative fluorescence per GFPuv-*E.coli*. The zoom was set to 1.0 and the pinhole was set to 1.00 AU for consistency in image capture. Each image was analyzed using a custom Python script (see Supplemental Material) to quantify the number of cells and fluorescence per cell using ImageJ. Post image acquisition, all images were analyzed using Image J. The Python script used details of the general guideline to follow for repeatable measurements for both fluorescence intensity per GFPuv-*E.coli* and cell count per image. The code allowed for a repeatable image analysis for an accurate comparison of relative GFP expression per GFPuv-*E.coli* and total cell count.

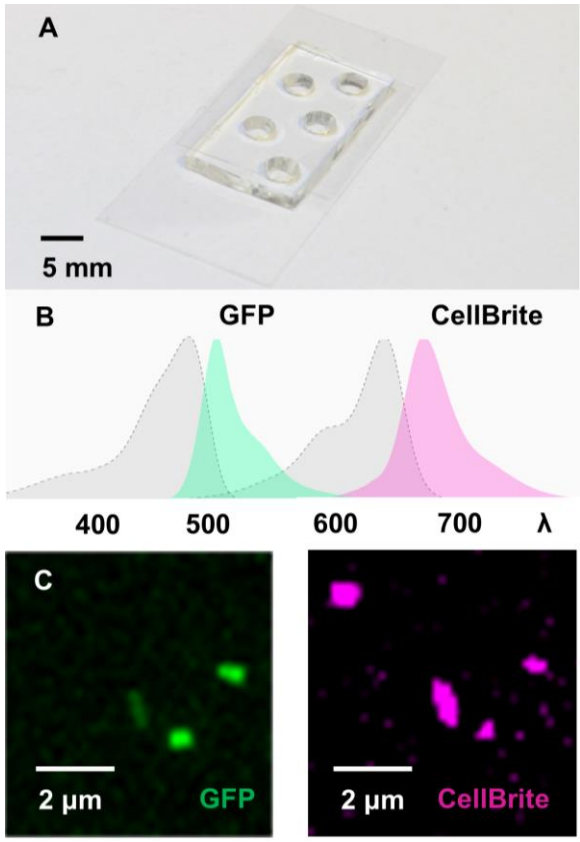

Fig 2. *Photograph of microdevice used for confocal imaging of E.coli samples.* ***A)*** *The device was manufactured using PDMS bonded to a glass slide. Wells for samples have a diameter of 3 mm.* ***B)*** *The peak excitation and emission wavelengths used for this study were* $\lambda_{ex,GFP}$ *= 495 nm,* $\lambda_{em,GFP}$ *= 509 nm,* $\lambda_{ex,CellBrite}$ *= 640 nm, and* $\lambda_{em,CellBrite}$ *= 662 nm.* ***C)*** *The GFPuv image compared with the CellBrite stained image.*

2.4 Statistical Analysis

All data points were collected in technical triplicates at various time points over 24 h throughout two independent 300L bioreactor runs. Technical triplicates were compared using a repeated measures ANOVA as well as a Tukey Means Comparison test (α = .05) to determine if there was significant difference between the time point samples. Due to identical bioreactor set up and time points of sample collection, data from both bioreactor runs were averaged and analyzed. This analysis method was used to find the statistically significant highest point for each metric. To assess the strength of correlation between different methods, Pearson's correlation coefficient was calculated. This analysis compared trendlines across various fluorescence intensity measurement techniques as well as different OD_{600} data collection methods.

3. RESULTS & DISCUSSION

3.1 Bioreactor Metric Analysis

During the bioreactor run, several inline (pH, DO, Temperature) and at-line (Glucose and OD_{600}) parameters were measured throughout the 24 h process, which can be seen in **Figure 3**. The measured variables included optical density (OD), which increased with GFPuv-*E.coli* proliferation and glucose which decreased as GFPuv-*E.coli* proliferated and consumed. The weight of the bioreactor content was measured using an internal calibrated scale to ensure it was unchanged and maintained within a specific range (275 kg+/- 10 kg), although some variability in this measurement existed due to removal of samples throughout the process and standard measuring error. Agitation began in response to changes in DO and the rpm was recorded. The dissolved oxygen content (DO) indicates oxygen available for consumption. Aeration (slpm) is used to promote oxygenation, and agitation (rpm) is introduced to promote both mixing and uniform oxygenation in the culture medium. The pH was measured between 6-7. Temperature and pressure were regulated between 29.8-30.1 °C and 4.95-5.05 psi, respectively. It should be noted that aeration, temperature, and pressure show some variation due to the scale of the presented values; however, there was statistically negligible variation in these parameters. **Figure 3** shows the change in these metrics over time for Run 2.

Results in **Figure 3A** show a correlation between glucose decline and an increase in optical density and weight (measured by settling of cells on the bottom of the bioreactor). This is expected, as glucose is consumed by cells during the proliferation period. **Figure 3B** depicts changes in dissolved oxygen (DO) as a percentage compared to agitation rpm which is inversely correlated. This relationship can be explained as when agitation and air sparging are introduced to the system, more dissolved air and therefore oxygen is added to the solution. These metrics are important to capture to ensure adequate functionality of the bioreactor and to give insight into the biological processes occurring during inoculation, GFP induction and thereafter.

Glucose levels correlate with cell proliferation because as cells are expanding, they are consuming glucose and therefore levels will decrease as cells proliferate. Optical Density (OD) also relates to *E. coli* proliferation as while cells are proliferating the optical density will increase due to a larger number of cells in the culture. Temperature monitoring is important for ensuring that the environment is suitable for cell proliferation. Other variables such as pressure, pH and DO % must all be tightly monitored and maintained within a certain range for optimal *E. coli* health, protein expression and proliferation. As cells are expanding and consuming glucose and oxygen, the glucose and DO% levels will drop.

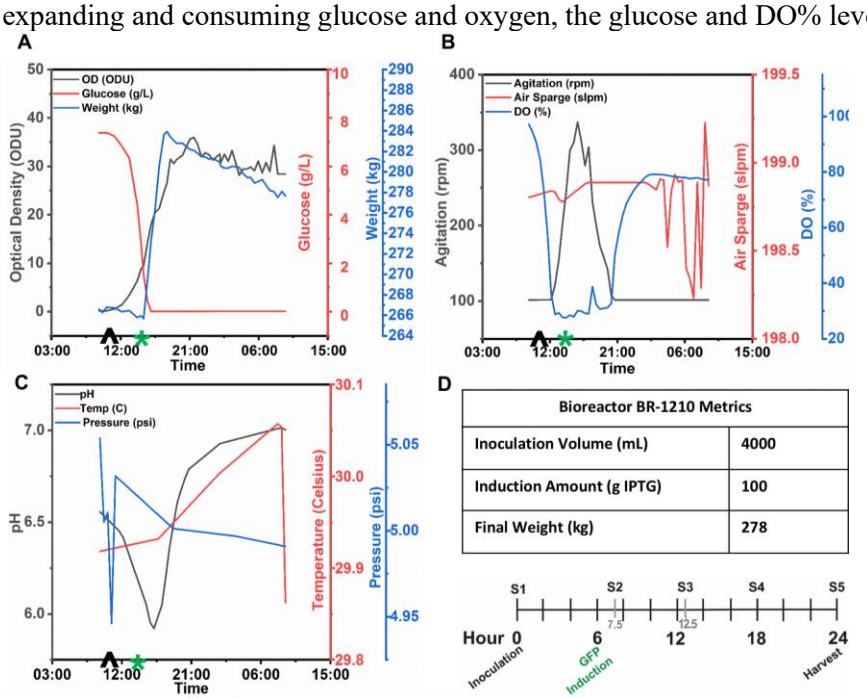

Fig 3. *Parameters measured during the bioreactor run, highlighting changes during inoculation at 09:08 (^) and induction at 15:08 (*).* ***A)*** *Optical density increased drastically after induction and levelled out afterwards. Glucose levels decreased with growth, as expected. Weight remained within 275 kg +/- 10 kg range. Fluctuations in weight are due to standard variability within measurement.* ***B)*** *Agitation, Air Sparge and Dissolved Oxygen were all monitored throughout the bioreactor run. The sharp decrease in DO% after inoculation levels out after induction which then increases after an increase in agitation. Air sparge remains at a steady rate and begins to vary towards the end of the process before final harvest.* ***C)*** *Changes in pH, temperature and pressure were all tightly monitored and maintained throughout the bioreactor run but do show a decline in pH and pressure after inoculation. Pressure begins to rise and level out immediately after inoculation while pH slowly declines until after induction and then begins to rise and level out until final harvest. The temperature remains within the strict 29.8 and 30.1 Celsius range.* ***D)*** *Table 4.D gives a brief overview of other important*

parameters measured for this run. The timeline in 4.D shows inoculation began at hour 0, GFP induction at hour 6, and final harvest at hour 24. Time points for the five samples analyzed included hour 0, hour 7.5, hour 12.5, hour 18 and hour 24.

3.2 Plate Reader Assays

Results from the plate reader for bioreactor runs 1 and 2 are shown in **Figure 4** and are considered offline analysis. OD_{600} and fluorescence intensity of *E.coli* suspended in the broth media were measured. Subsequently, the suspension was centrifuged to isolate the supernatant, which was then plated and measured separately. Technical triplicates were measured for each sample. The OD_{600} values relate to cell density while fluorescence intensity relates to GFP expression. The isolation and measurement of the supernatant fluorescence intensity is essential to infer potential GFP leakage from the *E.coli* into the broth media. It is well-established that bacterial membrane and cell wall become more permeable as they undergo cellular death, leading to "leakage" in their surrounding media [39]. Therefore, measuring fluorescence intensity in the supernatant can indicate the health of the cell population.

The OD_{600} values of the cells suspended in broth, indicative of cell density, show comparable changes in density profiles for both bioreactor runs. There was no significant change in cell density values of the supernatant for both runs, indicating that the supernatant was successfully isolated. The OD_{600} values of the suspended cells indicate the samples taken at hours 12.5 and 18 had the highest ($p<.0001$) cell density for run 1 and hours 12.5, 18 and 24 had the highest ($p<.0001$) OD_{600} values for run 2 based on a repeated measures ANOVA and Tukey Means comparison. Run 1 shows no significant difference between OD_{600} values for hours 12.5 and 18 and run 2 shows no significant difference in OD_{600} values between 12.5, 18 and 24 hours. In run 1, there was a significant decline in cell density after hour 18 ($p<.0001$), suggesting potential cell death occurred. Fluorescence intensity values for bioreactor run 1 indicate a significantly higher ($p<.0001$) overall fluorescence at hour 18 for suspended cells and hour 24 for supernatant. While GFP expression in bioreactor 2 peaks significantly at hour 12.5 for suspended cells and hour 24 for supernatant.

The OD_{600} and fluorescence intensity data obtained from the plate reader are part of the offline analysis of these bioreactor runs. The observed increase in fluorescence intensity of the supernatant across both bioreactor runs could indicate cell death or an increased membrane permeability, which is the highest at 24 hours (but not significantly higher than hour 18). The fluorescence intensity value of the suspended cells increases and peaks at hour 18 but thereafter does not significantly change at hour 24. These trends underscore the potential benefits of optimizing the duration of future bioreactor runs to enhance total product yield and streamline resource utilization in this bioprocess.

Statistical analysis, based on a Repeated Measures ANOVA with Tukey Means Comparison, was conducted on the technical triplicates from two separate biological runs collected for each time point via the plate reader. Comparison of each time point facilitated the identification of peak values for optical density (correlated with cell density) and fluorescence (associated with GFP expression).

3.3 Microscopy and Image Analysis Results

Image analysis was performed as a secondary method for quantification of relative fluorescence intensity, and cell density at each time point. Summary **Figure 5** highlights this offline analysis pipeline from confocal image capture through image analysis for samples collected from both bioreactor runs. Raw images were first collected for every sample using both λ_{ex} = 488 nm; λ_{em} = 507 nm and λ_{ex} = 642 nm; λ_{em} = 662 nm which produced two separate images for each region. Individual cells were identified based on the CellBrite 640 membrane stain. This membrane stain was used as a marker for the *E.coli*. Images were collected until a minimum of 200-300 individual GFPuv-*E.coli* bacterium were counted in each sample group, where later point samples realized individual counts of ~1000 for similar image numbers.

Once individual cells were identified, these identified cells were made into a cell mask. This mask was then overlayed on top of the EGFP image of the same region. Each individual cell, marked with the mask and overlayed on the EGFP image was then measured for fluorescence intensity. The code described earlier allows for repeatable fluorescence intensity measurements for each individual image for all samples. This method for measuring fluorescence allows for a higher resolution offline analysis as compared to the plate reader fluorescence intensity data. The confocal image analysis of fluorescence intensity per cell can be used to measure the distribution of fluorescence per cell as opposed to the overall fluorescence intensity per well measured in the plate reader. Multiple images were collected for each sample to accumulate hundreds of cells for fluorescence intensity analysis.

Image analysis of fluorescence intensity for two separate bioreactor runs can be seen in **Figure 6**. Fluorescent intensity data from both bioreactor runs follows a similar trend. Specifically, the highest mean fluorescence intensity can be found at hour 12.5 ($p<.0001$). The statistical analysis was performed via repeated measure ANOVA and Tukey Means Comparison with an alpha of .05.

Overall, the fluorescence intensity per cell area measured at five different time points (hour 0, 7.5, 12.5, 18, and 24) revealed significant variations (Repeated Measures ANOVA and Tukey Means Comparison, $p < 0.0001$) between all-time points for the two runs. Fluorescence intensity data shows a distinct peak in fluorescence intensity at the 12.5-hour time

point. After fluorescence reached its maximum levels at hour 12.5, the fluorescence intensity begins significantly declining in hours 18 and 24.

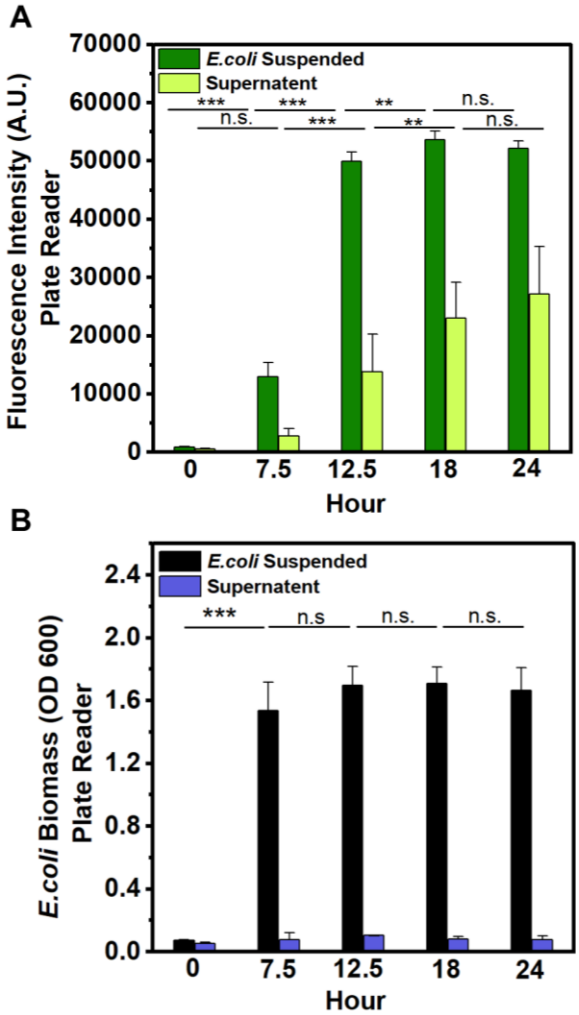

Fig 4. *Bar graph representations of fluorescence intensity and OD_{600} for two different bioreactor runs collected via plate reader.* ***A)*** *E. coli fluorescence values for 5 samples from bioreactor run 1 and 2 shows a significantly high (p<.001) relative fluorescence intensity of the suspended cell sample at hour 18 which is not significantly different than at hour 24 and a significant (p<.001) increase in fluorescence intensity of the supernatant at hour 18 which is not significantly greater than hour 24.* ***B)*** *E. coli OD_{600} values for 5 samples from bioreactor run 1 and 2 shows a significantly high (p<.0001) increase in OD_{600} value of the suspended cell sample at hour 7.5 which is not significantly different than hours 12.5, 18 or 24 indicating that cells proliferated and plateaued in growth after hour 7.5.*

These results indicate that peak GFPuv expression per *E.coli* is around the 12.5 hour period post inoculation. This peak in fluorescence intensity per cell area is different than the peak in overall fluorescence measured with the plate reader, which showed a peak at hours 18 and 24. This difference can be attributed to the fact that the plate reader takes a singular reading of fluorescence intensity of an entire sample and is much lower resolution than the confocal fluorescence intensity measurement per cell area. Although the overall fluorescence intensity may increase during hours 18 and 24, this does not account for any cellular membrane leakage or levels of proliferation that could impact the final measurement. Overall, the repeatable imaging pipeline and analytics presented in this study are key for multimodal analysis. Additionally, on the single cell analysis level, the imaging technique can be performed on a small number of either fixed or live cells which would allow for retrospective analysis. This imaging pipeline offers an alternative to flow cytometry which is a commonly used PAT for single cell level of analysis [29, 37, 40].

To calculate cell density per sample, a standardized number of images were used. Initially, 56 individual images were captured, and the total cell number was calculated per image using the cell mask process previously described. This cell density quantification method provides a total cell count per image as opposed to an OD_{600} value per well measured in the plate reader. This offline analysis method was used to both give higher resolution and a means of comparison for the data captured through plate reader analysis. Samples from hour 0 were diluted 10-fold less than all other samples due to lack of cell density and therefore cell counts per image were divided by 10.

After images were captured and analyzed, a sample of 50 randomly selected images were compiled to determine the distribution of cell count per image. Additionally, a sample of the remaining 6 images were compiled to determine the

distribution of cell count per image. The results of these two sample sizes were compared. The results from the 50-image sample size were more normally distributed than the results from the 6-image sample size and show more significance in the difference between the groups as seen in **Figure 7**.

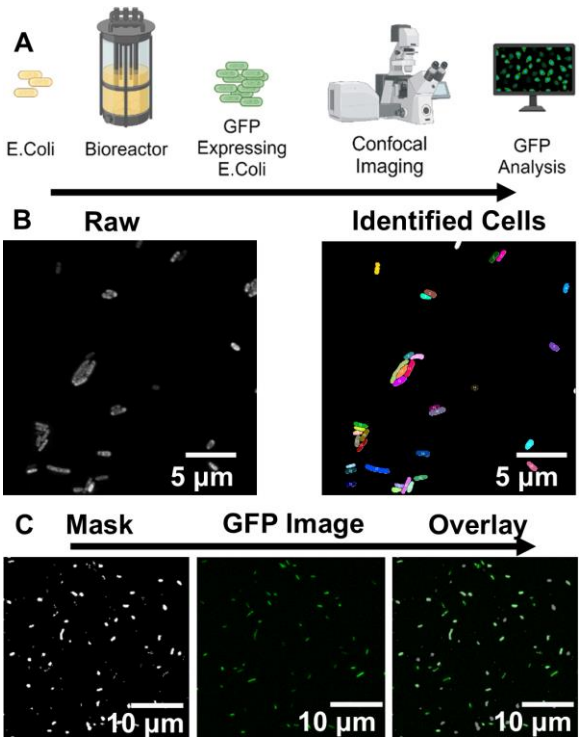

Fig 5. *Offline image analytics process for calculating fluorescence intensity per cell and cell count.* ***A)*** *Samples are collected from the bioreactor at various time points and analyzed using confocal microscopy.* ***B)*** *Raw images collected are processed using Python code in Image J to identify each individual E. coli cell based on the CellBrite 640 membrane stain.* ***C)*** *The identified cells then are used as a "mask" which when overlayed with the fluorescence image will measure the relative mean fluorescence intensity per E. coli.*

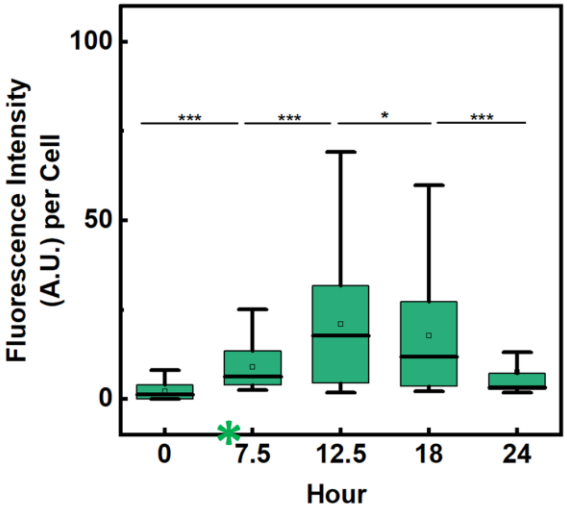

Fig 6. *Fluorescence intensity per cell as determined by confocal image analysis for two separate bioreactor runs. The peak of fluorescence intensity (A.U.) per E. coli for both runs is at hour 12.5 (p<.0001) which then begins to significantly decline at hour 18 (p<.05).*

Results from the 50-image sample size show the highest cell count per image to be at 12.5h for both run 1 (Repeated Measures ANOVA and Tukey Means Comparison, p<.0001) and run 2 (p <.0001). Analysis of 6 image sample size for run 2 shows the highest cell count per image to be at hour 12.5 (p<.0001). However, when analyzing the 6-image sample size for run 1, there is no significant difference between hours 7.5, 12.5, 18 and 24. Therefore, in order to determine the highest cell count per image, more than 6 images are necessary. These results of cell count per image follow a similar trend to the OD_{600} data from the plate reader analysis which both correlate to cell density. This information further supports the claim

that there may be a point of diminishing returns during this bioreactor where cell density begins to decrease due to increased cell death. **Figure 7** highlights the cell fragments at hour 24 that begin to appear in Run 2 and can impact the cell mask and subsequent count.

3.4 Comparison of Analytical Methods

Single-cell and batch-level analytical methods for the quantification of GFPuv yield were demonstrated and compared. While previous studies have focused on single cell level analysis of GFP expression, utilizing flow cytometry, or have focused on batch-level protein quantification via optical density OD_{600} and fluorescence assays, this study bridges the gap between these two analytical levels [17, 42]. **Figure 8** shows the representative comparative metrics, as illustrated by analysis of Run 2. Analysis of plate reader data, confocal image analysis, and inline bioreactor measurements suggest that the peak *E. coli* density and GFP expression are between hours 12.5-18h. A strong correlation was observed between inline bioreactor and plate reader methods for OD_{600} measurement (Pearson's correlation coefficient of 0.87). Microscopic image analysis and plate reader methods for quantification of GFPuv yield were also comparable (Pearson's correlation coefficient of 0.96) for samples collected during the initial 18h of the bioreactor runs. Peak GFPuv expression was expected at 18h.

A surprising outcome was realized after 18h. A steep decline and divergence of outputs from both plate reader assay and microscopic methods was observed, suggesting that as the cells are proliferating, there may be a point where they become overpopulated or underfed and start not only decreasing their rate of proliferation but also begin to die. This idea can be supported by the supernatant fluorescence intensity data, which shows a steady increase in GFPuv expression over time. This increased supernatant fluorescence intensity could indicate the *E. coli* have increased permeability, and GFPuv is leaking into the surrounding media. Apoptosis would also release GFPuv into the media [39, 41]. This effect was most apparent in the microscopy-based method, which when understood in conjunction with when used with batch-level analysis like plate reader measurements, provides crucial additional insights. This analytical pipeline not only identifies key time windows for peak protein expression but also offers insight into cell leakage and viability, as detected through supernatant fluorescence. This data suggests it might be optimal to shorten this bioprocessing timeline after 18h to improve GFP expression and *E.coli* viability; however, more data collected from future runs will be necessary to validate this claim. The limitations of this study include sample size, which was constrained by only two biological runs.

The multifaceted analysis and comparison of several metrics gives a higher resolution view of this bioprocess, which can be used for further optimization and enhancement in the future. This short study combines two methods for fluorescence intensity quantification which correlates with GFP expression which could indicate optimal bioreactor settings with further investigation. The microscopy-based method, unlike flow cytometry, requires fewer cells, does not demand specialized equipment or technical expertise, and provides differentiation of debris from target and a rapid inference of lysed cells in the population. [37, 40].

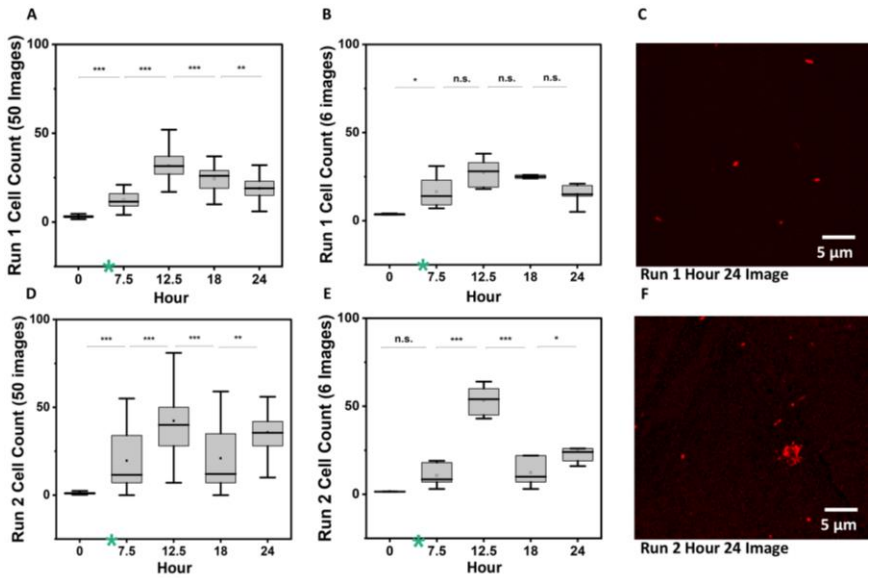

Fig 7. *Cell count per image analyzed at different sample sizes for two bioreactor runs.* ***A)*** *In run 1, with a 50-image sample size, the peak in cell number per image is observed at hour 12.5 (p<.0001).* ***B)*** *In run 1, based on a 6-image sample size, determination of the peak in cell number is inconclusive due to a non-significant difference in cell number when comparing hours 7.5, 12.5, 18, and 24.* ***C)*** *Hour 24 image of E.coli from run 1 showing intact E.coli which are fairly sparse.* ***D)*** *In run 2, also with a 50-image sample size, the peak in cell number per image is observed at hour 12.5 (p<.0001).* ***E)*** *Conversely, in run 2, based on a 6-image sample size, a peak is observed at hour 12.5 for cell number per image (p < .0001).* ***F)*** *Hour 24 image of E.coli from run 2 showing cells potentially beginning to die via apoptosis showing multiple cell fragments which can impact the cell mask and cell count.*

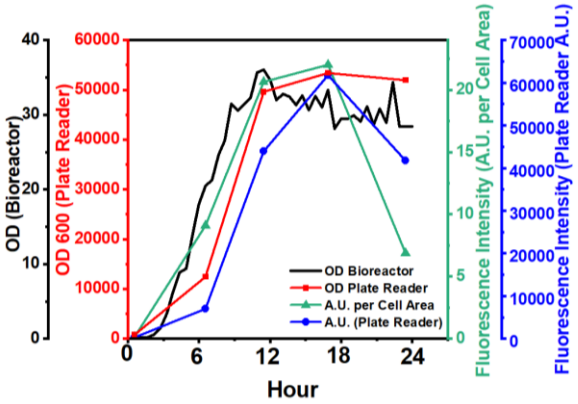

Fig 8. *Results from bioreactor Run 2 including both inline and offline quantification of GFPuv yield.* ***OD_{600}*** *data was collected inline from the bioreactor. OD_{600} and bulk fluorescence intensity data were collected using a plate reader, and fluorescence intensity per cell was computed via confocal image analysis. Peak yield was between 12.5-18h.*

4. CONCLUSION

This study underscores the importance of multimodal analytical approaches in bioprocess monitoring to advance Industry 4.0 applications, where scalable and automated process analytics are vital for high-throughput operations. For example, analyzing data at both the batch and cellular levels during the process development stage can identify divergent responses in your process and provide crucial information for accurately setting operational parameters—such as fermenter harvest time—before transitioning to manufacturing. This comparative approach supports more precise process optimization and smoother scale-up to industrial production.

In our study, we highlight this by comparing batch-level analysis with single-cell, semi-automated imaging analysis to measure the GFPuv yield across the entire bioreactor run. This study also offers a rare perspective on pilot-scale production (300L) with step-by-step data collection and analysis methods. A clear point of diminishing returns was identified with offline techniques, while image analysis revealed insights into factors contributing to productivity loss. Batch-level analysis, based on OD_{600} and fluorescence intensity measurements of the supernatant and cell suspension, highlighted protein leakage and changes in overall fluorescence expression. Inline metrics, including OD_{600}, pH, and glucose levels, provided real-time insights into the environmental conditions within the bioreactor. All three methods indicate that an earlier harvest could potentially achieve a more efficient product yield.

In conclusion, while inline and at-line metrics offer valuable real-time insights, offline techniques like image analysis provide high-resolution data that are crucial for evaluating cellular health and product distribution—key factors in assessing yield decline, which may signal lysis or cell death. These findings are instrumental for bioprocess optimization, particularly in identifying optimal harvest times, as evidenced by the strong correlations observed between the different analytical methods. While this study was limited by the scheduling constraints of bioreactor runs, future research should incorporate biological triplicates to enhance the robustness and generalizability of the results.

5. ACKNOWLEDGEMENTS

G.R was supported by the Chemistry of Life Pre-Doctoral Training Grant at NC State University (NIH Grant 5T32GM141887-03). This work was performed in part at the North Carolina State University BTEC Facility. The authors wish to acknowledge the support provided by the Novo Nordisk Foundation (AIM-Bio Grant NNF19SA0035474), National Science Foundation (Grant EBMS- 2033997), the Triangle Universities Center for Advanced Studies Inc. (TUCASI), and the North Carolina Viral Vector Initiative in Research and Learning (NC VVIRAL) at NC State University.

6. STATEMENTS AND DECLARATIONS

Conflict of Interest: The authors declare no conflict of interest.

Ethical Approval: Neither ethical approval nor informed consent was required for this study.

REFERENCES

1. Clegg I. (2020) Process analytical technology. *Specif. Drug Subst. Prod.* pp. 149-173. https://doi.org/10.1016/B978-0-08-102824-7.00007-5
2. Maischberger T. (2019) Optimized process and bioreactor characterization. *Chem. Ing. Tech.* 91(12):1719-1723. https://doi.org/10.1002/cite.201900134
3. Javaid M., Haleem A., Vaishya R., et al. (2020) Industry 4.0 technologies and their applications in fighting COVID-19 pandemic. *Diabetes Metab. Syndr.* 14(4):419-422. https://doi.org/10.1016/j.dsx.2020.04.032
4. Khanal O., Lenhoff A.M. (2021) Developments and opportunities in continuous biopharmaceutical manufacturing. *mAbs* 13(1):1903664. https://doi.org/10.1080/19420862.2021.1903664
5. Xu J., Xu X., Huang C., et al. (2020) Biomanufacturing evolution from conventional to intensified processes for productivity improvement: a case study. *mAbs* 12(1):1770669. https://doi.org/10.1080/19420862.2020.1770669
6. Gerlach I., Hass V.C., Brüning S., et al. (2013) Virtual bioreactor cultivation for operator training and simulation: application to ethanol and protein production. *J. Chem. Technol. Biotechnol.* 88(12):2159-2168. https://doi.org/10.1016/j.btre.2015.03.002
7. Stephenson M., Grayson W. (2018) Recent advances in bioreactors for cell-based therapies. *F1000Res.* 7:517. https://doi.org/10.12688/f1000research.12533.1
8. Pereira R.F.S., De Carvalho C.C.C.R. (2021) Optimization of multiparameters for increased yields of cytochrome B5 in bioreactors. *Molecules* 26(14):4148. https://doi.org/10.3390/molecules26144148
9. Gargalo C.L., De Las Heras S.C., Jones M.N., et al. (2020. Towards the development of digital twins for the bio-manufacturing industry. *Springer Int. Publ.* pp. 1-34. https://doi.org/10.1007/10_2020_142
10. Kensy F., Engelbrecht C., Büchs J. (2009) Scale-up from microtiter plate to laboratory fermenter: evaluation by online monitoring techniques of growth and protein expression in *E. coli* and *H. polymorpha* fermentations. *Microb. Cell Fact.* 8(1):68. https://doi.org/10.1186/1475-2859-8-68
11. Han S., Delvigne F., Brognaux A., et al. (2013) Design of growth-dependent biosensors based on destabilized GFP for the detection of physiological behavior of *Escherichia coli* in heterogeneous bioreactors. *Biotechnol. Prog.* 29(2):553-563. https://doi.org/10.1002/btpr.1694
12. Gustavsson R., Mandenius C.-F. (2013) Soft sensor control of metabolic fluxes in a recombinant *E. coli* fed-batch cultivation producing green fluorescence protein. *Bioprocess Biosyst. Eng.* 36(10):1375-1384. https://doi.org/10.1007/s00449-012-0840-z
13. Gorochowski T.E., Van Den Berg E., Kerkman R., et al. (2014) Using synthetic biological parts and microbioreactors to explore the protein expression characteristics of *E. coli*. *ACS Synth. Biol.* 3(3):129-139. https://doi.org/10.1021/sb4001245
14. Frank A., Brognaux N., Sorensen S., et al. (2011) Applicability of GFP microbial whole-cell biosensors to bioreactor operations - Mathematical modeling and related experimental tools. *Biosens. Emerging Mater. Appl.*. https://doi.org/ 10.5772/20088
15. Delisa M., Chae H., Weigand W., et al. (2001) Generic model control of induced protein expression in high cell density cultivation of *E. coli* using on-line GFP-fusion monitoring. *Bioprocess Biosyst. Eng.* 24(2):83-91. https://doi.org/10.1002/1097-
16. Bower D.M., Lee K.S., Ram R.J., Prather K.L.J. (2012) Fed-batch microbioreactor platform for scale-down and analysis of a plasmid DNA production process. *Biotechnol. Bioeng.* 109(8):1976-1986. https://doi.org/10.1002/bit.24498
17. Aucoin M.G., McMurray-Beaulieu V., Poulin F., et al. (2006) Identifying conditions for inducible protein production in *E. coli*: combining a fed-batch and multiple induction approach. *Microb. Cell Fact.* 5(1):27. https://doi.org/10.1186/1475-2859-5-27
18. Lara A.R., Leal L., Flores N., et al. (2006) Transcriptional and metabolic response of recombinant *E. coli* to spatial dissolved oxygen tension gradients simulated in a scale-down system. *Biotechnol. Bioeng.* 93(2):372-385. https://doi.org/10.1002/bit.20704
19. Wurm D.J., Marschall L., Sagmeister P., et al. (2017) Simple monitoring of cell leakiness and viability in *E. coli* bioprocesses—A case study. *Eng. Life Sci.* 17(6):598-604. https://doi.org/10.1002/elsc.201600204
20. Jamaledin R., Sartorius R., Di Natale C., et al. (2020) Recombinant filamentous bacteriophages encapsulated in biodegradable polymeric microparticles for stimulation of innate and adaptive immune responses. *Microorganisms* 8(5). https://doi.org/10.3390/microorganisms8050650
21. Jones J.J., Bridges A.M., Fosberry A.P., et al. (2004) Potential of real-time measurement of GFP-fusion proteins. *J. Biotechnol.* 109(1-2):201-211. https://doi.org/10.1016/j.jbiotec.2003.10.039

22. Feilmeier B.J., Iseminger G., Schroeder D., et al. (2000) Green fluorescent protein functions as a reporter for protein localization in *E. coli*. *J. Bacteriol.* 182(14):4068-4076. https://doi.org/10.1128/jb.182.14.4068-4076.2000
23. Savchuk O.A., Silvestre O.F., Adão R.M.R., Nieder J.B. (2019) GFP fluorescence peak fraction analysis-based nanothermometer for the assessment of exothermal mitochondria activity in live cells. *Sci. Rep.* 9(1). https://doi.org/10.1038/s41598-019-44023-7
24. Rosano G.N.L., Ceccarelli E.A. 2014. Recombinant protein expression in *E. coli*: advances and challenges. *Front. Microbiol.* 5. https://doi.org/10.3389/fmicb.2014.00172
25. Faust G., Stand A., Weuster-Botz D. (2015) IPTG can replace lactose in auto-induction media to enhance protein expression in batch-culture *E. coli*. *Eng. Life Sci.* 15(8):824-829. https://doi.org/10.1002/elsc.201500011
26. Garcia J.R., Cha H.J., Rao G., et al. (2009) Microbial nar-GFP cell sensors reveal oxygen limitations in highly agitated and aerated laboratory-scale fermentors. *Microb. Cell Fact.* 8(1):6. https://doi.org/10.1186/1475-2859-8-6
27. Zulkifly N.A.H., Selas Castiñeiras T., Overton T.W. (2023) Optimization of recombinant TNFα production in *E. coli* using GFP fusions and flow cytometry. *Front. Bioeng. Biotechnol.* 11. https://doi.org/10.3389/fbioe.2023.1171823
28. Shokri A.S.G.L.A. (2002) Growth rate-dependent changes in *E. coli* membrane structure and protein leakage. *Appl. Microbiol. Biotechnol.* 58(3):386-392. https://doi.org/10.1007/s00253-001-0889-0
29. Broger T., Odermatt R.P., Huber P., Sonnleitner B. (2011) Real-time on-line flow cytometry for bioprocess monitoring. *J. Biotechnol.* 154(4):240-247. https://doi.org/10.1016/j.jbiotec.2011.05.003
30. Hisiger S., Jolicoeur M. (2005) A multiwavelength fluorescence probe: is one probe capable for on-line monitoring of recombinant protein production and biomass activity? *J. Biotechnol.* 117(4):325-336. https://doi.org/10.1016/j.jbiotec.2005.03.004
31. Reischer H., Schotola I., Striedner G., et al. (2004) Evaluation of the GFP signal and its aptitude for novel on-line monitoring strategies of recombinant fermentation processes. *J. Biotechnol.* 108(2):115-125. https://doi.org/10.1016/j.jbiotec.2003.11.007
32. Delisa M.P., Li J., Rao G., et al. (1999) Monitoring GFP-operon fusion protein expression during high cell density cultivation of *E. coli* using an on-line optical sensor. *Biotechnol. Bioeng.* 65(1):54-64. https://doi.org/10.1002/(SICI)1097-0290(19991005)65:1<54::AID-BIT7>3.0.CO;2-R
33. Kavara A., Sokolowski D., Collins M., Schofield M. (2020) Recent advances in continuous downstream processing of antibodies and related products. *Approaches Purif. Anal. Charact. Antibody-Based Ther.* pp. 81-103. https://doi.org/10.3390/ijms23158663
34. Knabben I., Regestein L., Marquering F., et al. (2010) High cell-density processes in batch mode of a genetically engineered *E. coli* strain with minimized overflow metabolism using a pressurized bioreactor. *J. Biotechnol.* 150(1):73-79. https://doi.org/10.1016/j.jbiotec.2010.07.006
35. Binder D., Probst C., Grünberger A., et al. (2016) Comparative single-cell analysis of different *E. coli* expression systems during microfluidic cultivation. *PLoS ONE* 11(8):e0160711. https://doi.org/10.1371/journal.pone.0160711
36. Gomes L.C., Carvalho D., Briandet R., Mergulhão F.J. (2016) Temporal variation of recombinant protein expression in *E. coli* biofilms analysed at single-cell level. *Process Biochem.* 51(9):1155-1161. https://doi.org/10.1016/j.procbio.2016.05.016
37. Heins A.L., Hoang M.D., Weuster-Botz D. (2022) Advances in automated real-time flow cytometry for monitoring of bioreactor processes. *Eng. Life Sci.* 22(3-4):260-278. https://doi.org/10.1002/elsc.202100082
38. Girard P., Meissner P., Jordan M., et al. Small-scale bioreactor system for process development and optimization. *Springer Netherlands* pp. 323-327. https://doi.org/10.1016/S1369-703X(00)00110-8
39. Lowder M., Unge A., Maraha N., et al. (2000) Effect of starvation and the viable-but-nonculturable state on GFP fluorescence in GFP-tagged *Pseudomonas fluorescens* A506. *Appl. Environ. Microbiol.* 66(8):3160-3165. https://doi.org/10.1128/AEM.66.8.3160-3165.2000
40. Brognaux A., Han S., Sørensen S.J., et al. (2013) A low-cost, multiplexable, automated flow cytometry procedure for the characterization of microbial stress dynamics in bioreactors. *Microb. Cell Fact.* 12(1):100. https://doi.org/10.1186/1475-2859-12-100
41. Webb J.S., Barratt S.R., Sabev H., et al. (2001) Green Fluorescent Protein as a Novel Indicator of Antimicrobial Susceptibility in *Aureobasidium pullulans*. *Appl. Environ. Microbiol.* 67(12):5614-5620. https://doi.org/10.1128/AEM.67.12.5614-5620.2001
42. Brognaux A., Thonart P., Delvigne F., et al. (2013) Direct and indirect use of GFP whole-cell biosensors for the assessment of bioprocess performances: Design of milliliter scale-down bioreactors. *Biotechnol. Prog.* 29(1):48-59. https://doi.org/10.1002/btpr.1660

Supplementary Information

Comparative Approaches for Quantification of Product Yield in the Production of Recombinant Green Fluorescent Protein (GFPuv) Expressed in E. coli

By Gabrielle Rusch[1,2], Junhyeong Wang[3], Keith Breau[4], Katie Kilgour[5], Gary Gilleskie[6], Jeff Keele[6], Kurt Selle[6], Scott T. Magness[1], Stefano Menegatti[5,6]*, and Michael Daniele[1,2]*

[1] Joint Department of Biomedical Engineering
North Carolina State University and University of North Carolina, Chapel Hill
911 Oval Dr., Raleigh, NC 27695 (USA)

[2] Comparative Medicine Institute
North Carolina State University
1060 William Moore Dr., Raleigh, NC 27606 (USA)

[3] Department of Electrical & Computer Engineering
North Carolina State University
890 Oval Dr., Raleigh NC, 27695 (USA)

[4] Department of Cell Biology and Physiology
University of North Carolina, Chapel Hill
111 Mason Farm Road, Chapel Hill, NC 27599 (USA)

[5] Department of Chemical Engineering
North Carolina State University
911 Partners Way, Raleigh, NC 27695 (USA)

[6] Golden LEAF Biomanufacturing Training and Education Center (BTEC)
NC State University, Raleigh, NC 27695 (USA)

[*] Correspondence: mdaniel6@ncsu.edu; smenega@ncsu.edu

Code for Image Analyses: The code used for image analysis is compatible in both Python and Macros languages and should be used with FIJI Image J image analysis software. The description of each steps operation is denoted below with "//."

```
open(".lif");
//open up original image which should include both GFP and Cell Mask images
selectWindow("A-S1-IM3 - C=0");
//Select GFP Image (*DO NOT CHANGE BRIGHTNESS OR CONTRAST)
run("Z Project...", "projection=[Max Intensity]");
selectWindow("A-S1-IM3 - C=1");
//Select Cell Mask Image (CellBrite 640)
run("Z Project...", "projection=[Max Intensity]");
run("Show Info...");
run("Brightness/Contrast...")
// Adjust brightness and contrast to reduce background noise and enhance cell mask as needed.
run("Apply LUT");
run("16-bit");
setOption("ScaleConversions", true);
run("8-bit");
setAutoThreshold("Default no-reset");
run("Threshold...");
setThreshold(47, 255);
// Optimize the threshold to ensure no background noise is included
setOption("BlackBackground", true);
run("Convert to Mask");
// This sets the mask in order to define what is a cell and what is not
run("Close");
run("Watershed");
run("Analyze Particles...", "size=15-Infinity pixel display exclude clear summarize overlay");
saveAs("Results", "Results.csv");
```

```
selectWindow("A-S1-IM3 - C=1");
run("Set Measurements...", "area mean min add redirect=[ A-S1-IM3 - C=0] decimal=3");
// Overlay the cell mask onto the GFP image in order to quantify relative mean GFP per E.coli
run("Analyze Particles...", "size=15-Infinity pixel display exclude clear summarize overlay");
saveAs("Results", "S1Results.csv");
```