APEX: Automated Protein EXpression in *Escherichia coli*

Martyna Kasprzyk[1], Michael A. Herrera[1,2], and Giovanni Stracquadanio[1,*]

[1]School of Biological Sciences, The University of Edinburgh, Edinburgh EH9 3BF, United Kingdom

[2]Current address: School of Chemistry, The University of Edinburgh, Edinburgh EH9 3FJ, United Kingdom

[*]Corresponding author. Email: giovanni.stracquadanio@ed.ac.uk.

Abstract. Heterologous protein expression is an indispensable strategy to generate significant amounts of recombinant proteins. To this end, *Escherichia coli* (*E. coli*) is one the most used microbial host for recombinant protein production due to its rapid growth, well-characterised genetics, and ability to produce recombinant proteins in high yields using modern recombinant DNA technology. However, while there is a plethora of robust protein expression protocols for *E. coli*, these methods are often unsuitable for high-throughput screening due to their significant resource and time consumption; these protocols are also susceptible to operator error and inconsistency.

To address these challenges, we have developed APEX, a robust and automated protocol for recombinant protein production in *E. coli*. APEX leverages the accessible, open-source Opentrons OT-2 platform to automate microbial handling and protein expression with high precision and reproducibility. APEX can be configured to perform heat shock transformation, colony selection, colony sampling, inoculation, subculturing and protein expression using a low-cost, minimal OT-2 hardware setup. We further demonstrate the efficacy of our automated transformation workflows using a variety of plasmids (2.7-17.7 kb), and exemplify the automated heterologous expression of a diverse array of proteins (27-222 kDa). Designed with customization, modularity and user-friendliness in mind, APEX can be easily adapted to the operator's needs without requiring any coding expertise.
APEX is available at https://github.com/stracquadaniolab/apex-nf under the AGPL3 license.

Introduction

Recombinant protein production is a widespread and important technique for the study, screening and characterisation of proteins. Microbial cell factories are commonly used as recombinant expression systems, particularly *Escherichia coli* (*E. coli*), due to rapid growth, straightforward culturing conditions, well-understood physiology, easy transformation, and low costs [1]. Recombinant proteins have significant applications, including the production of high-value chemicals [2] and therapeutic proteins

[3]. However, when processing numerous protein targets, such as in protein engineering where rapid identification of desirable properties is essential, these experiments can become labor-intensive and time-inefficient, leading to significant bottlenecks in research workflows [4]. Automation can transform these processes into high-throughput operations, significantly accelerating the processing of numerous protein targets.

Many workflows in biological laboratories have been successfully automated, streamlining processes such as DNA assembly [5], high-throughput screening (HTS) [6], and sample preparation for next-generation sequencing (NGS) [7]. Such adaptation of automation has played a significant role across various fields, including drug discovery [8], clinical diagnostics [9], genomics [10], and proteomics [11]. Automation enhances efficiency by streamlining repetitive tasks [12], while improving reproducibility by mitigating against commonly observed operator error, such as inconsistent pipetting [13]. By standardizing processes, automation allows for consistent performance, regardless of the operators' experience, increasing both experimental success rates and the reliability of results.

Opentrons robots have been used in various research applications, including Golden Gate [14] and BASIC assembly [15] methods, RNA extraction and RT-PCR diagnostics for SARS-CoV-2 [16, 17], and proteomics workflows for isobaric tag sample preparation [18] and shotgun proteomics [19].

Here we developed an end-to-end automation pipeline for recombinant protein expression in *Escherichia coli* (*E. coli*) called Automated Protein EXpression (APEX). This modular pipeline includes protocols for heat shock transformations, spotting, colony sampling, and protein expression. It used spreadsheets to guide the execution of protocols, allowing for customization without the need for programming experience. Protocols are generated via a computational pipeline that outputs a Python script (which can be directly uploaded to the Opentrons App) and a PDF file containing detailed protocol instructions and labware layouts for the reagents. To enhance reproducibility and streamline protocols for protein production in *E. coli*, we implemented the pipeline in Nextflow [20] and packaged all dependencies in a Docker container [21]. This allows the pipeline to be run on any computer system.

We validated our APEX platform and assessed its usability through an extensive series of experiments. Initially, we miniaturized the transformation volumes in DH5α to minimise reagent use. Subsequently, we transformed using plasmids ranging from 2.7 to 17.7 kb and compared the performance of APEX with that of a human operator. We evaluated the pipeline's ability to sample colonies, microculturing, and protein expression induction in BL21(DE3) cells using super folder Green Fluorescent Protein (sfGFP) as a benchmark. Finally, we executed the full pipeline with a variety of targets spanning molecular weights from 29 to 222 kDa, and analyzed the outcomes for protein production.

Collectively, our results demonstrate that our platform is highly robust, ensuring high throughput while minimising reagents use and maximising efficiency of protein production setup.

Results

The APEX workflow and implementation

The Automated Protein EXpression (APEX) workflow is designed to equip experimental scientists with a validated and optimised protein production protocol that leverages the power of affordable lab automation, while requiring little to no computational experience to onboard (see Figure 1). To this end, we designed our protocol to work on the Opentrons OT-2 liquid handling robot, the most widely used low-cost, open-source system, which provides sufficient capabilities to carry out protein expression. Importantly, to minimise hardware cost, we developed APEX to work on a minimal OT-2 setup, which consists only of the thermocycler module and any combinations of single and multi-channel pipettes.

The APEX protocol implements the typical four-step protein expression process in *E. coli*: (1) heat shock transformation, (2) colony selection on agar plates, (3) colony picking, and (4) protein expression induction. It is structured into four modular, self-contained OT-2 protocol steps, offering flexibility to execute complete workflows or individual steps as needed (see Figure 1B). Each module requires a JavaScript Object Notation (JSON) file for OT-2 hardware/labware setup and a Comma Separated Value (CSV) file for liquid handling details like source wells and volumes. These files are integrated into a master Python protocol (OT-2 APIs v2.15), which loads directly into the Opentrons app without additional coding. To assist with offline preparatory work such as loading DNA or media, APEX provides a PDF with instructions and visual representations of plate contents made with R scripts for user reference (see Figure 1C). The APEX workflow is released as a Nextflow [20] pipeline, where all dependencies are packaged as a Docker/Singularity [21] image, thus ensuring robustness and reproducibility of our software. APEX and its detailed documentation are available at: https://github.com/stracquadaniolab/apex-nf and Supplementary File 1.

Protocol 1 - Heat shock transformations

The most widely method used for delivering exogenous DNA into *E. coli* is the heat shock transformation. In brief, the cells and the exogenous DNA are pre-incubated together in ice-cold calcium chloride. A sudden, rapid shift in temperature enables the uptake of the exogenous DNA. The cells are subsequently recovered in a rich recovery medium [22]. The necessary aspect of this process is

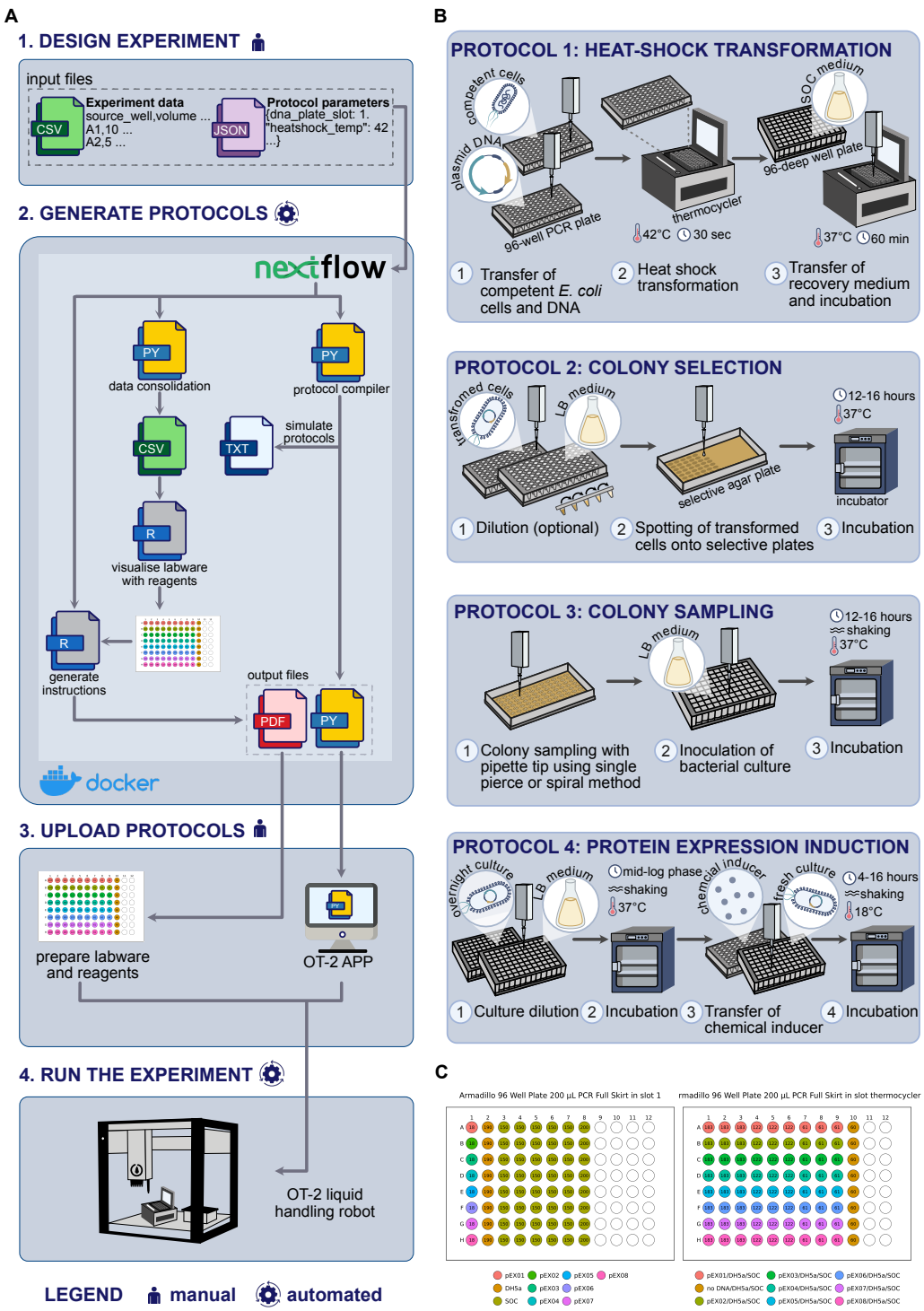

Figure 1: Automated Protein EXpression (APEX) workflow. A. The pipeline comprises several steps: (1) Users design experiments using a JSON file to configure hardware and labware setup, along with a CSV file that details experimental data such as wells containing reagents and their volumes. (2) These input files are integrated into a master Python protocol templates, generating ready-to-use Python protocols without any need for coding. Concurrently, another Python script consolidates data from the JSON and CSV files to form the CSV input for an R script, which visualizes the labware and plots each well's contents. An additional R script produces a PDF containing detailed instructions for users' reference. The workflow is released as a Nextflow pipeline and containerized in Docker to ensure efficient dependency management and reproducibility. (3) Users upload the Python protocols to the OT-2 app and prepare the robot deck. (4) The robot executes the experiment, completing the workflow. **B.** Example of APEX protocols usage: The workflow is designed in a modular fashion to mirror the structure of a protein expression workflow in *E. coli*, consisting of four main protocols: Protocol 1: Heat shock transformation to deliver DNA into the cells. Protocol 2: Colony selection via spotting transformed cells on agar plates supplemented with antibiotics. Protocol 3: Colony sampling and cell culturing. Protocol 4: Induction of protein expression in cultured cells. **C.** Example of visualized labware included in the instructions for user guidance.

ensuring rapid temperature changes. While the OT-2 temperature module can reach required transformation temperatures, its warm-up and cool-down rates are too slow [23], and require additional manual handling, like adding plate covers to prevent evaporation. Therefore, APEX uses a thermocycler module that enables rapid temperature gradients. By default, DNA is incubated at $4\,^{\circ}\mathrm{C}$ for 30 minutes, heat shocked at $42\,^{\circ}\mathrm{C}$ for 30 seconds, and recovered with SOC medium at $37\,^{\circ}\mathrm{C}$ for 1 hour. Users can easily modify these parameters in the configuration files to suit their specific requirements.

Protocol 2 - Colony selection

Following transformation, cells are plated on a selective solid medium, typically agar supplemented with antibiotic. This ensures that only cells carrying the DNA payload will survive. This time-consuming step can be efficiently replaced by liquid handling robots.

Protocol 2 enables the OT-2, which lacks sensing capabilities, to dispense transformed cells onto agar plates by calculating agar height from the plate base area, weight, and agar density [24], and includes an adjustable parameter for customising spotting height relative to the calculated agar height. APEX supports rectangular Thermo Scientific Nunc OmniTray plates for up to 96 transformations in an 8x12 grid and standard round 90mm Petri dishes with a 3D-printed adapter for up to 32 transformations in an 8x4 grid.

The density of agar was experimentally determined using a Jupyter notebook pre-installed on the OT-2, with an interactive script to measure agar height via a pipette tip. Given that environmental factors such as agar temperature while pouring can affect its solidification, we standardized these conditions. Agar density calculations revealed variations in plate types: $0.911\,\mathrm{g/cm^3}$ for rectangular plates and $0.954\,\mathrm{g/cm^3}$ for round plates (Supplementary Tables 5 and 6). The Jupyter notebook was further developed to enable users to verify and adjust agar density and spotting height, accommodating variations in plate preparation and ensuring correct plate depths in case of discrepancies caused by inaccurately specified plate dimensions or individual robot calibrations.

Protocol 3 - Colony sampling

Selecting single colonies for downstream culturing and protein expression is usually a tedious operation, which can be automated through the use of specialised equipment, such as the QPix system, which have detection capabilities. However, when the objective is to express a protein, the requirement for single colonies can be relaxed, and non isogenic cultures can be used. In this scenario, we were able to automate this process by using pipette with attached tips to sample colonies from agar plates.

Given that the success rate of this operation depends on the abundance and distribution of colonies across the spot, we developed 2 strategies (Figure 2C): (i) a single pierce sampling method for high cell densities, where the tip directly targets the centre of the initial cell dispension, and (ii) a spiral sampling method, where the pipette traces a spiral motion from the same centre coordinates outward for low or uneven cell densities. Agar height is calculated as outlined in Protocol 2, which also allows for adjusting the sampling height to ensure contact with the agar. The spiral method uses slow, smooth movements to minimise disturbance to the agar, and allows parameter adjustments to tailor the spiral shape based on colony distribution. Although slower, this method improves the likelihood of successful sampling compared to the quicker single pierce method. After sampling, cells are inoculated into fresh LB-antibiotic by mixing media in a 96 deep well plate.

Protocol 4 - Protein Expression Induction

After successful transformation and overnight culturing, cells are primed for protein expression induction. This stage enables the translation of the target gene into the corresponding protein. The induction method varies depending on the plasmid vector and expression strain used; this typically involves the addition of a chemical inducer like isopropyl-β-D-thiogalactoside (IPTG) or L-arabinose. The APEX protocol propagates cells in fresh LB-antibiotic media using an aliquot of the overnight culture. The cells are then manually transferred to a shaking incubator to reach mid-log phase. Following this, the robot dispenses the chemical inducer, and the cells are further cultured to produce protein.

Optimisation of transformation volume

When adapting laboratory protocols for automation, volumes are typically miniaturised to enhance efficiency and reduce costs. To evaluate the effects of volume reduction on transformation efficiency, standard volumes recommended by the competent cell supplier were adjusted to fit a $200\,\mu$l PCR plate suitable for a thermocycler module. We tested three transformation conditions: High Volume (HV), Medium Volume (MV) and Low Volume (LV) as per Table 1. Using Protocol 1, we performed automated heat shock transformations on DH5α cells at the specified volumes, using plasmids pEX05 (pJKR-H-araC encoding sfGFP, 1.5×10^{-4} pmol) and pEX01 (pUC19 as the positive control, 1.5×10^{-5} pmol) (see Figure 3A).

For plasmid pEX05, the highest transformation efficiency was observed using the LV configuration, which achieved 88% and 63% higher efficiencies compared to the MV and HV configurations, respectively (see Figure 3B, Supplementary Table 2). Conversely, for plasmid pEX01, the highest

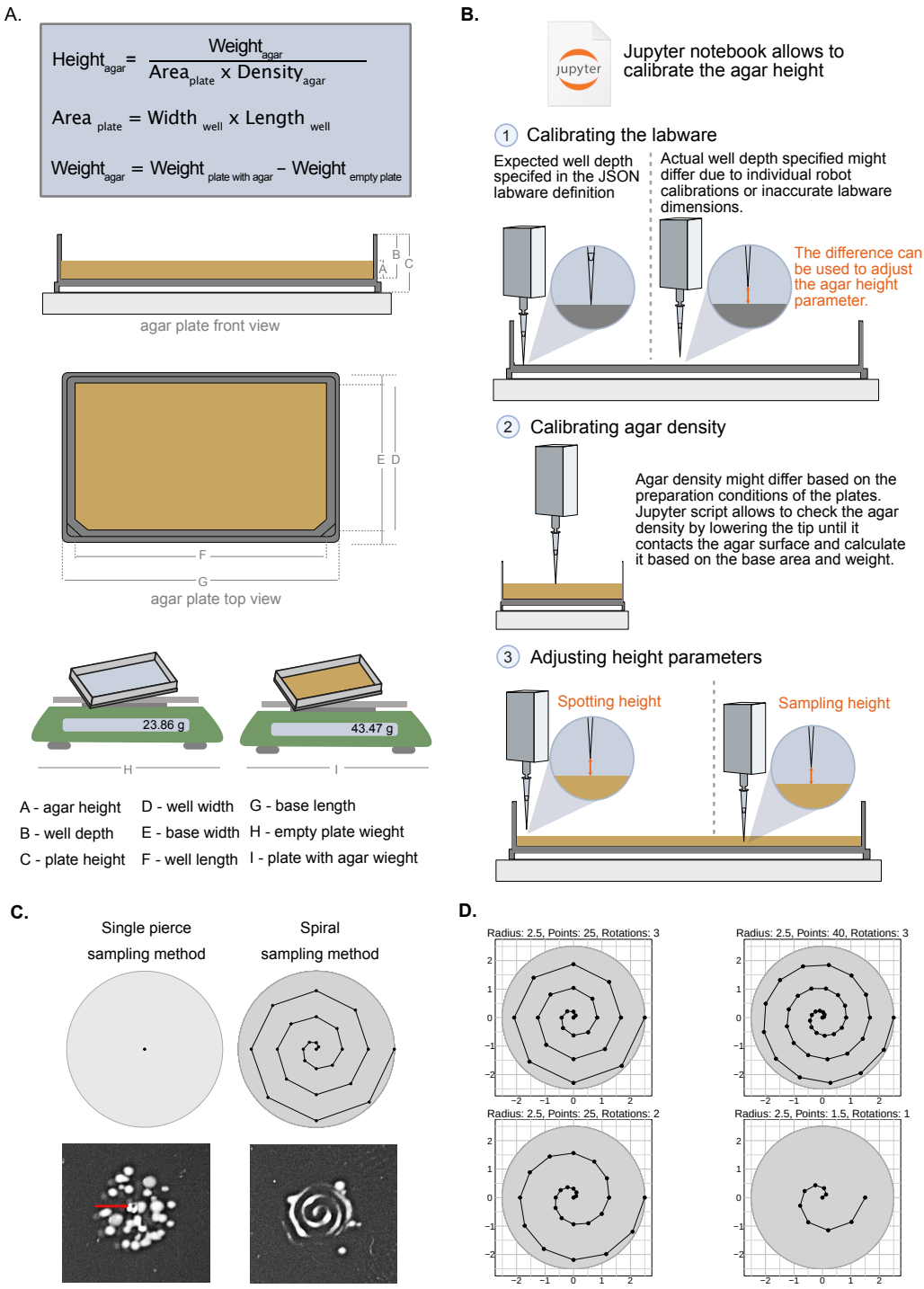

Figure 2: Developing methods of agar plate spotting and colony sampling. A. To determine the agar height for spotting, the pipeline uses parameters such as plate base area, agar density, empty plate weight, and filled plate weight are inputted for Protocols 2 and 3. The script determines the weight of the agar by subtracting the empty plate's weight from the filled plate's weight and divides the result by the product of the base area and agar density. **B.** To adjust for any discrepancies that might arise due to individual robot calibrations or inaccuracies in plate measurements and the environmental conditions that might affect how the agar solidifies, we developed a Jupyter notebook (Supplementary File 2) that allows to verify the agar heights. **C.** Two colony sampling methods were developed: a single pierce method, where the tip descends directly into the center of the spotted cells at the precise coordinates of initial cell dispersion, and a spiral method, where the pipette follows a spiral motion outward from the same central coordinates. **D.** Radius, number of points, and rotations can be adjusted to control the path and shape of the spiral. Increasing the number of rotations extends the spiral, adding more points creates a smoother spiral, and expanding the radius extends the spiral outward.

Table 1: Configurations of miniaturised transformation volumes. Volume of DNA, cells, and SOC medium used for High Volume (HV), Medium Volume (MV), and Low Volume (LV) configurations.

Configuration	DNA volume (µl)	Cells volume (µl)	SOC volume (µl)	Total volume (µL)
HV	3	30	150	183
ML	2	20	100	122
LV	1	10	50	61

efficiency occurred in the HV configuration, with efficiency 77% greater than MV, but only 21% higher than LV configuration. Given these results, the LV configuration was selected for all subsequent experiments. This approach not only reduces reagent consumption, but also enhances throughput by facilitating quicker processing of smaller volumes without sacrificing transformation efficiency.

Comparison of automated and manual transformation efficiency at different plasmid sizes

To assess the versatility of APEX for heat shock transformations using the OT-2 thermocycler, we tested a spectrum of plasmids varying in size, given that transformation efficiency generally decreases with increasing plasmid size [22]. We transformed the following plasmids (1.5×10^{-4} pmol)into DH5α cells: pEX01 ($2.7\,\mathrm{kb}$), pEX02 ($3.9\,\mathrm{kb}$), pEX03 ($4.3\,\mathrm{kb}$), pEX04 ($4.7\,\mathrm{kb}$), pEX05 ($4.9\,\mathrm{kb}$), pEX06 ($7.1\,\mathrm{kb}$), pEX07 ($11.6\,\mathrm{kb}$), and pEX08 ($17.7\,\mathrm{kb}$). Automated transformations were performed under LV conditions specified by Protocol 1, while manual transformations followed the supplier's protocol. After transformation, SOC-recovered outgrowths were handled as per Protocol 2. While all transformations were directly spotted onto selective agar plates, pEX01 was diluted 1:4 and pEX02 was diluted 1:2 to adjust for high transformation efficiency. Transformation Efficiency (TE) was calculated as described in Methods section. Results demonstrated successful colony formation across all plasmids (see Figure 4A, Supplementary Table 3).

A per-plasmid comparison showed no difference between automated and manual transformation (Supplementary Table 4) and, as expected, a consistent inverse relationship between plasmid size and TE. Taken together, these data exemplify the advantage of time-saving, high-throughput automation without compromising TE.

Colony sampling, microculturing and protein expression

To evaluate the end-to-end capability of the APEX workflow - including colony sampling, microculturing, and protein expression - we initially selected sfGFP as our recombinant protein target. We performed 32 transformations of the pEX05 plasmid in BL21(DE3) cells using Protocol 1, followed

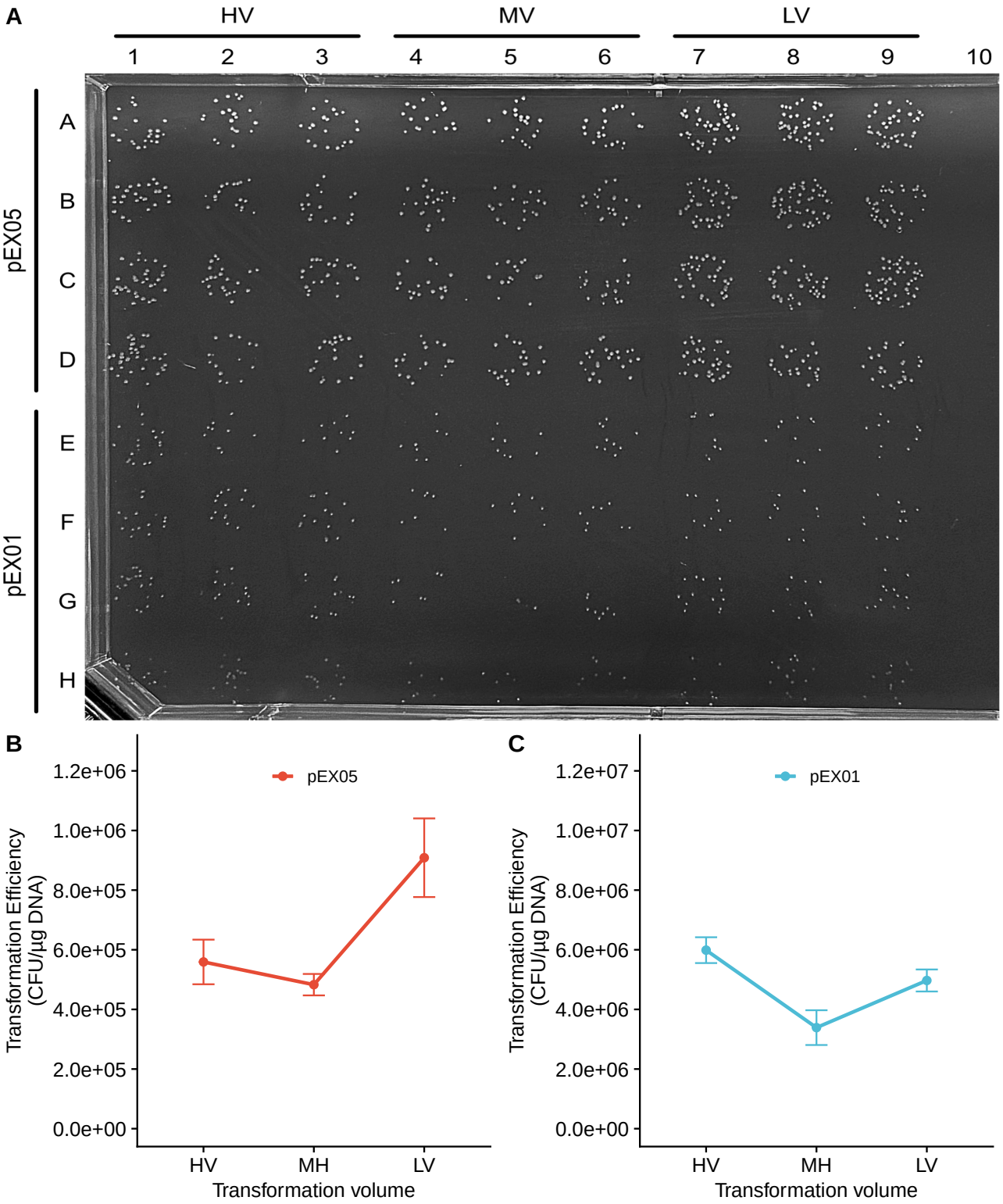

Figure 3: Transformation efficiency of *E. coli* DH5α at reduced volumes. A. DH5α transformations with pEX05 (Rows A-D) and pEX01 (Rows E-H, positive control). Columns 1-3 show High Volume (HV) transformations, columns 4-7 Medium Volume (MV), and columns 8-11 Low Volume (LV), with each row representing a technical replicate. Row 10 is a negative control. Panels **B.** and **C.** show transformation efficiencies (CFU/µg) of pEX05 and pEX01, respectively. Error bars represent standard deviation from the mean value (n=4).

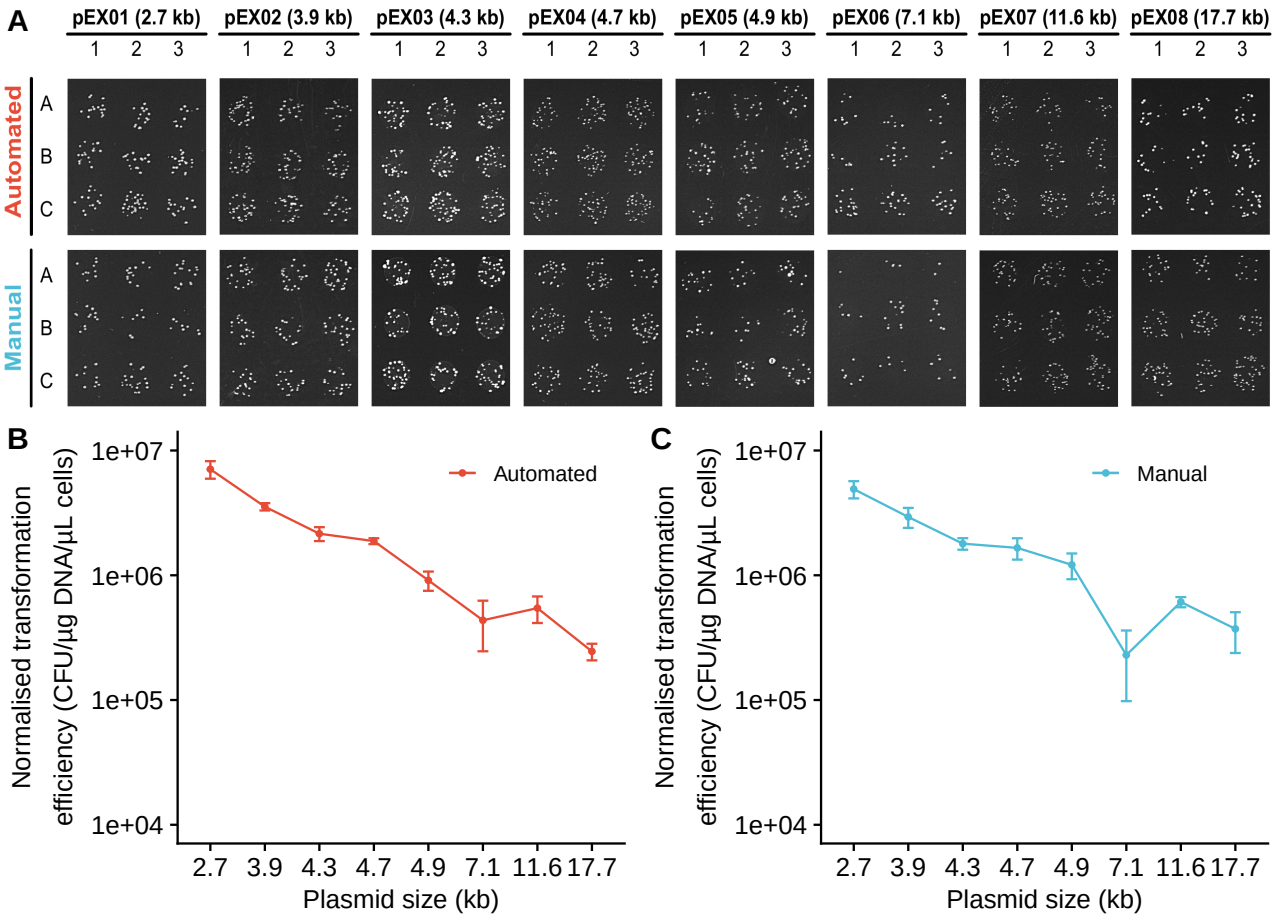

Figure 4: Comparative analysis of automated DH5α transfromations against manual method across various plasmid sizes. A. Transformation of DH5αcells with plasmids pEX01 (2.7 kb), pEX02 (3.9 kb), pEX03 (4.3 kb), pEX04 pACYC-Duet (4.7 kb), pEX05 (4.9 kb), pEX06 (7.1 kb), pEX07 (11.6 kb), and pEX08 (17.7 kb). Each transformation was carried out with 1 µl of DNA at 1.5×10^{-4} pmol. The automated method used miniaturised volumes with 10 µl of cells and 50 µl of SOC media as per Protocol 2 with Low Volume (LV) configuration. The manual method used 50 µl of cells and 250 µl of SOC media, adhering to the manufacturer's protocol. Experiments were conducted in biological replicates (Rows A-C) and technical replicates (Columns 1-3). For each transformation, 5 µl of culture was spotted on LB plates supplemented with antibiotics; plasmids pEX01 and pEX02 were diluted 4-fold and 2-fold, respectively, and incubated overnight. **B** and **C** show the normalized transformation efficiency for automated and manual methods, respectively. Error bars represent standard deviation from the mean value (n=3).

by colony selection by spotting the outgrowth directly and with a 1:5 dilution, as per Protocol 2. This setup created variable colony densities in order to assess our two colony sampling strategies: (i) high density for the single pierce method and (ii) low density for the spiral method. All transformations yielded colonies with the intended cell densities in accordance with the dilution strategy (see Figure 5A). OD_{600} measurements from overnight cultures confirmed successful cell propagation.

The single pierce method could sample and cultivate cells in 28 out of 32 biological replicates; significant challenges were presented when the distribution of colonies was non-central, even in instances where overall cell density was high. Conversely, the spiral method sampled cells across all biological replicates, demonstrating the efficacy of this sampling strategy. Subcultures were induced with L-arabinose to express sfGFP (Protocol 4), with fluorescence measurements verifying protein production in all cultures (see Figure 5C). These results showcase the APEX workflow as a robust pipeline for automating colony sampling, microculturing, and protein expression in *E. coli*.

Analysis of the end-to-end protein production using different expression plasmids

To demonstrate the versatility of the APEX pipeline for heterologous protein production, we selected three target enzymes from different prokaryotic organisms, covering a range of molecular weights from small to large: *Bacillus subtilis* 4'-phosphopantetheinyl transferase Sfp ($29\,\mathrm{kDa}$, encoded by pEX04); *Chromobacterium violaceum* ω-transaminase CVTA ($49\,\mathrm{kDa}$, pEX09); and an engineered polyketide synthase module derived from *Mycobacterium tuberculosis* ($222\,\mathrm{kDa}$, pEX07). We applied all four APEX protocols for automated transformation, selection, sampling with microculturing, and IPTG-inducible protein expression using BL21(DE3) cells. APEX protocol 4 was further configured to perform expression tests using different microculturing volumes ($0.5\,\mathrm{mL}$ and $1\,\mathrm{mL}$) in a 96 deep-well plate. Although a slight growth deficit was observed in the $1\,\mathrm{mL}$ microcultures (see Figure 6A), possibly due to less effective oxygen transfer, all three target proteins were well-expressed in soluble form at both the $0.5\,\mathrm{mL}$ and $1\,\mathrm{mL}$ scales (see Figure 6B). This demonstration highlights the APEX platform's versatility in automating heterologous protein expression in *E. coli*.

Conclusion

Conventional microbiological techniques and methodologies are often not amenable to high-throughput automation. To address this bottleneck, we developed an open-source, highly-customisable suite of protocols for automated microbial handling and recombinant protein expression, named APEX. APEX interfaces with the affordable Opentrons OT-2 liquid handling robot, using a minimal hardware setup.

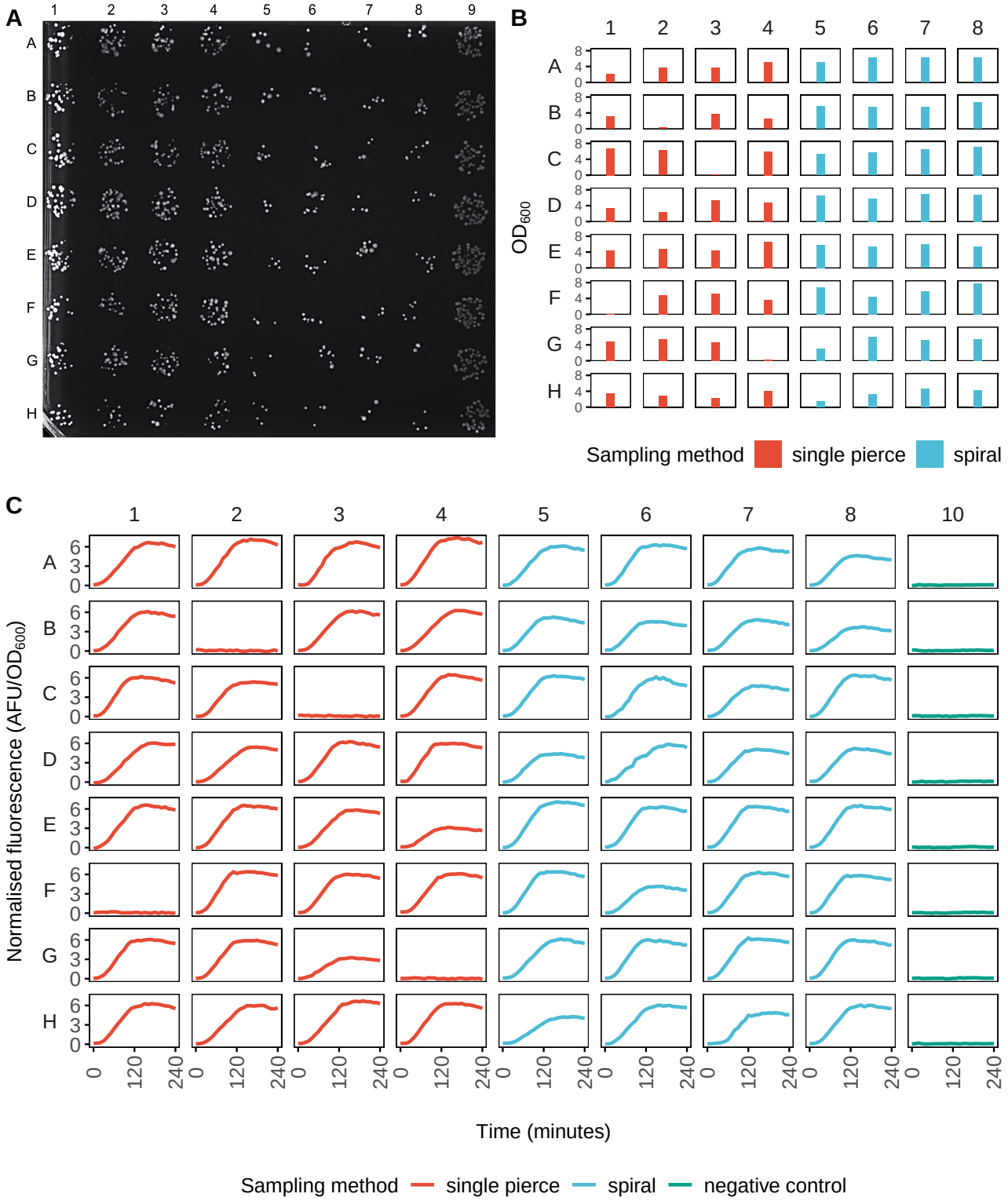

Figure 5: sfGFP expression in *E. coli* BL21(DE3) using automated sampling methods. A. *E. coli* BL21(DE3) cells transformed with plasmids pEX05 encoding sfGFP (Columns 1-8), and pEX01 (Column 9, positive control), following Protocol 1. Columns 1-4 show direct spotting onto selective antibiotic agar, while cells in Columns 5-8 were diluted 1:5 before spotting to simulate high- and low-density colonies respectively (Protocol 2). **B.** Final OD_{600} values of *E. coli* BL21(DE3) cells harbouring pEX05 and pEX01 were sampled using two sampling methods outlined in Protocol 3: the single pierce method for Columns 1-4 and the spiral method for Columns 5-8. **C.** Normalised fluorescence (AFU/OD_{600}) showing sfGFP expression under Protocol 4. L-arabinose (0.2% w/v) was added at time zero, with fluorescence monitored over 240 minutes. No inducer was added to Column 9.

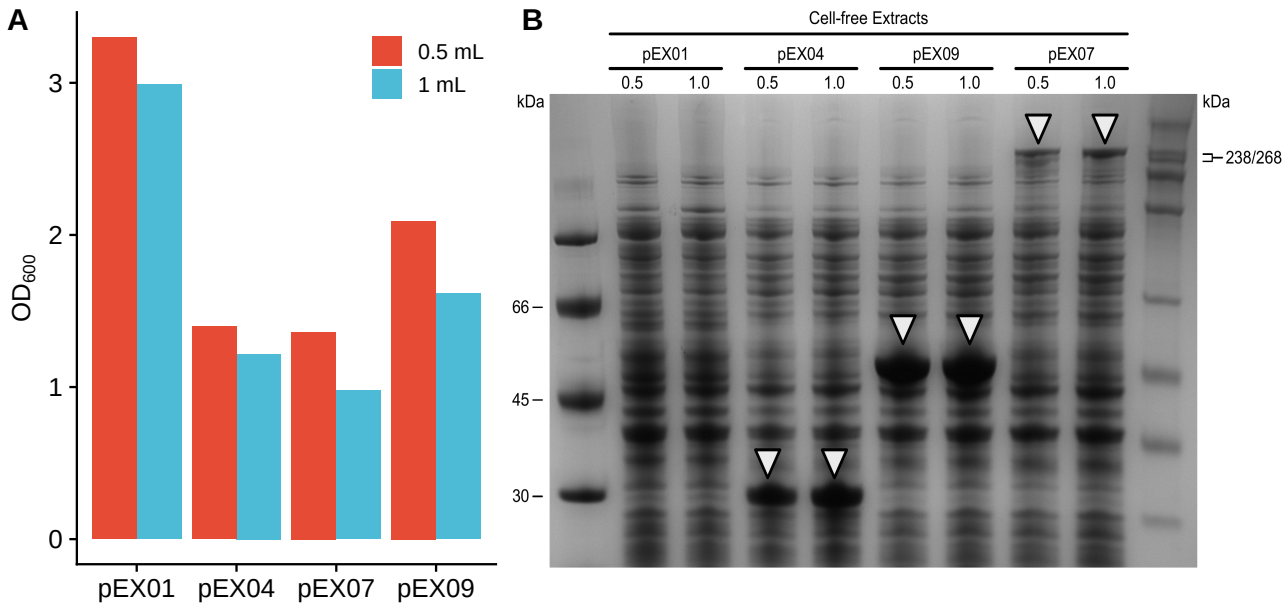

Figure 6: Analysis of the APEX end-to-end protein production pipeline across various molecular weights and culture volumes. A. Final OD_{600} values of *E. coli* BL21(DE3) after 16-hour expression at 18 °C. Cells harboring plasmids pEX01, pEX04, pEX09, and pEX07 were cultured in 96 deep V-well plate with 0.5 mL or 1 mL volumes. **B.** SDS-PAGE analysis of cell-free extracts from induced *E. coli* BL21(DE3) microcultures (0.5 mL and 1 mL). Cells harboring plasmids pEX01 (negative control), pEX04 (29 kDa), pEX09 (49 kDa), and pEX07(222 kDa), were induced with 0.1 mM IPTG. Proteins are marked by white triangles, with molecular weights verified using an Amersham low molecular weight calibration kit (leftmost lane) and a HiMark pre-stained protein standard (rightmost lane).

APEX can account for different hardware configurations and requirements as specified by the operator, and can be challenged to transform different *E. coli* cell lines with a variety of expression plasmids, encoding a diverse array of protein targets. We demonstrate the time and resource advantages of APEX-programmed automation and microculturing without compromising TE or soluble protein expression. Overall, APEX offers a viable solution for automated *E. coli* cultivation and recombinant protein production; we anticipate that the versatility of APEX will extend the accessibility of robust, high-throughput automation to small and medium-sized laboratories.

Materials and methods

Reagents, Strains and Plasmids

Two *E. coli* strains were used: DH5α and BL21(DE3), both sourced from Thermo Fisher Scientific. An overview of plasmids used in this study can be found in Table 2. pUC19 was purchased from New England Biolabs and pJKR-H-araC [25] was a gift from George Church (Addgene plasmid # 62563). Plasmids pEX04, pEX06, pEX07, pEX08, pEX09 were kindly gifted by Dominic Campopiano at the University of Edinburgh. Plasmids were prepared using the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) following the supplier's protocol. Antibiotics were sourced as follows: carbenicillin

from Thermo BioReagents, kanamycin from Gibco, zeocin and chloramphenicol from Thermo Fisher Scientific, and chloramphenicol, used at concentrations of $100\,\mu g\,ml^{-1}$, $50\,\mu g\,ml^{-1}$, $25\,\mu g\,ml^{-1}$, and $25\,\mu g\,ml^{-1}$ respectively.

ID	Size (bp)	Vector	Gene Product	Origin	Resistance
pEX01	2686	pUC19	Empty	N/A	AmpR
pEX02	3943	pJUMP27-1A(sfGFP)	N-acetyl-L-glutamate Kinase	Escherichia coli	KanR
pEX03	4287	pGAPZalpha B	α-Galactosidase	Homo sapiens	BleoR
pEX04	4695	pACYC-Duet	4'-phosphopantetheinyl Transferase	Bacillus subtilis	CmR
pEX05	4897	pJKR-H-araC	Superfolder Green Fluorescent Protein	Aequorea victoria	AmpR
pEX06	7109	pET26b	Fatty Acyl-ACP Ligase	Corynebacterium glutamicum	KanR
pEX07	11610	pET26b	Engineered Polyketide Synthase Module	Mycobacterium tuberculosis	KanR
pEX08	17658	pET26b	Polyketide Synthase Bimodule	Mycobacterium tuberculosis	KanR
pEX09	6568	pET28a	ω-Transaminase	Chromobacterium violaceum	KanR

Table 2: List of plasmids used for APEX experimental validation.

Lysogeny Broth (LB) was prepared with 5 g/L yeast extract, 10 g/L tryptone, and 10 g/L NaCl, all obtained from Merck Millipore. Super Optimal Broth with Catabolite Repression (SOC)was prepared with 2.5 g/L yeast extract, 10 g/L tryptone, and 10 mmol/L NaCl, 10 mmol/L KCl and 20 mmol/L glucose (both from Thermo Fisher Scientific), and 10 mmol/L $MgSO_4$ and 10 mmol/L $MgCl_2$ (both from Sigma-Aldrich). LB agar was prepared with 5 g/L yeast extract, 10 g/L tryptone, 10 g/L NaCl, and 1.5% agar (Oxoid). For labware and hardware used in this study see Supplementary Table 1.

Competent cells preparation

A single colony of *E. coli* strain was inoculated into $5\,mL$ of LB medium and incubated overnight at $37\,°C$. The overnight culture was 100-fold diluted into $50\,mL$ of fresh LB medium to grow at $37\,°C$ and $300\,rpm$ until it reached $OD_{600} = 0.4$. The cells were incubated on ice for 20 minutes and harvested by centrifugation at $4000\times g$ and $4\,°C$ for 10 minutes. Cells were prepared as calcium-competent using the Hanahan method [26], and the pellets were resuspended in $2\,mL$ of $0.1\,M$ $CaCl_2$ with 15% (v/v) glycerol solution and stored at $-80\,°C$.

Calculation of transformation efficiency

Transformation Efficiency (TE) is quantified as the number of colony-forming units (CFUs) per μg of plasmid DNA. To ensure consistency, the following parameters were standardized: resuspension volume at $2\,mL$ from $50\,mL$ bacterial cultures, cell density at OD = 0.4, use of the same bacterial stock within each experiemnt, same heat shock conditions, and plating $5\,\mu L$ of transformation. Images of the agar plates were taken using the Syngene™ NuGenius Gel Documentation System and CFUs was counted using ImageJ [27]. TE was calculated as follows:

$$\text{TE} = \left(\frac{\text{CFUs}}{\text{DNA}\,(\mu g)}\right) \times \left(\frac{\text{total transformation volume}\,(\mu L)}{\text{plated volume}\,(\mu L)}\right) \times \text{dilution factor} \tag{1}$$

Moreover, to account for variations in aliquots of competent cells and the cells-to-DNA ratio, TE was normalized as:

$$\text{Normalised TE} = \frac{\text{CFUs}}{\text{DNA}\,(\mu g) \times \text{cells volume}\,(\mu L)} \times \left(\frac{\text{total transformation volume}\,(\mu L)}{\text{plated volume}\,(\mu L)}\right) \times \text{dilution factor} \tag{2}$$

super folder Green Fluorescent Protein (sfGFP) experiment

Overnight cultures of *E. coli* BL21(DE3) transformed with the pJKR-H-araC plasmid were diluted and grown to an OD_{600} of approximately 0.2, at which point they were induced with $1.33\,mM$ L-arabinose. $150\,\mu L$ of cells were transferred to a clear-bottom 96-well black plate (Thermo Fisher Scientific) and incubated with shaking at $37\,^{\circ}C$ and $960\,rpm$ high speed in a Varioskan LUX multimode microplate reader (Thermo Fisher Scientific). OD_{600} and sfGFP fluorescence (excitation $485\,nm$, emission $528\,nm$) were measured every 10 minutes for 4 hours, and fluorescence (AFU) normalised to OD_{600} was reported.

End-to-end recombinant protein production

BL21(DE3) transformants harbouring plasmids pEX01, pEX04, pEX07, or pEX09 were selected and propagated in antibiotic media as per APEX protocols 1–3. APEX protocol 4 was configured to back-dilute overnight cultures in fresh LB media (OD_{600} = 0.1, $0.5\,mL$ or $1\,mL$) in a 96 V-well 2 mL plate (Greiner). The cells were grown to mid-log phase at $37\,^{\circ}C$, after which protein expression was induced by the addition of $0.1\,mM$ IPTG as per APEX protocol 4. The microcultures were sealed using a Breathe-Easy sealing membrane (Diversified Biotech) and incubated overnight at $18\,^{\circ}C$ with

rigorous shaking. Approximately 2.4×10^8 cells were sampled from each microculture and harvested by microcentrifugation (13 300 rpm, 10 minutes). The spent media was discarded, and the harvested cells were lysed in BugBuster Mastermix (25 µL, Merck) at room temperature for 15 minutes. The lysate was clarified by microcentrifugation (13 300 rpm, 10 minutes). A sample of the cell-free extract (20 µL) was diluted 2-fold in standard 2X SDS-PAGE sample buffer and boiled for 6 minutes. The samples were resolved on a NuPAGE 4–12% Bis-Tris precast gel (Invitrogen) using 1X NuPAGE MES SDS running buffer (200 V, 35 minutes). The gel was recovered and stained using InstantBlue Coomassie protein stain (Abcam).

Contributions

G.S. conceived the study. M.K developed the APEX protcols with help from M.A.H. and supervised by G.S. M.K. and M.A.H. performed lab experiments. All experimental work was supervised by G.S. G.S., M.K. and M.A.H. wrote the manuscript.

Acknowledgments

This work was supported by the UKRI EPSRC Fellowship (EP/V033794/1), and the IBioIC feasibility funding (FF-2022-04) to G.S, and the UKRI Biotechnology and Biological Sciences Research Council (BBSRC) grant BB/T00875X/1 for M.K.
We also thank the Edinburgh Genome Foundry (EGF) for the support and useful discussions.

References

[1] G. L. Rosano and E. A. Ceccarelli. “Recombinant protein expression in Escherichia coli: advances and challenges”. In: *Frontiers in Microbiology* 5 (Apr. 2014). ISSN: 1664-302X.

[2] E. N. Kissman et al. “Expanding chemistry through in vitro and in vivo biocatalysis”. en. In: *Nature* 631.8019 (July 2024), pp. 37–48. ISSN: 0028-0836, 1476-4687.

[3] R. Assenberg et al. “Advances in recombinant protein expression for use in pharmaceutical research”. en. In: *Current Opinion in Structural Biology* 23.3 (June 2013), pp. 393–402. ISSN: 0959440X.

[4] B. Jia and C. O. Jeon. “High-throughput recombinant protein expression in *Escherichia coli* : current status and future perspectives”. en. In: *Open Biology* 6.8 (Aug. 2016), p. 160196. ISSN: 2046-2441.

[5] D. H. Kang et al. “RoboMoClo: A Robotics-Assisted Modular Cloning Framework for Multiple Gene Assembly in Biofoundry”. en. In: *ACS Synthetic Biology* 11.3 (Mar. 2022), pp. 1336–1348. ISSN: 2161-5063, 2161-5063.

[6] E. Guo et al. “Robotic Construction and Screening of Lanthipeptide Variant Libraries in *Escherichia coli*”. en. In: *ACS Synthetic Biology* 11.12 (Dec. 2022), pp. 3900–3911. ISSN: 2161-5063, 2161-5063.

[7] E. Meijers et al. “Automating the Illumina DNA library preparation kit for whole genome sequencing applications on the flowbot ONE liquid handler robot”. en. In: *Scientific Reports* 14.1 (Apr. 2024), p. 8159. ISSN: 2045-2322.

[8] G. Schneider. “Automating drug discovery”. en. In: *Nature Reviews Drug Discovery* 17.2 (Feb. 2018), pp. 97–113. ISSN: 1474-1776, 1474-1784.

[9] K. Antonios, A. Croxatto, and K. Culbreath. “Current State of Laboratory Automation in Clinical Microbiology Laboratory”. en. In: *Clinical Chemistry* 68.1 (Dec. 2021), pp. 99–114. ISSN: 0009-9147, 1530-8561.

[10] H. Tegally et al. “Unlocking the efficiency of genomics laboratories with robotic liquid-handling”. en. In: *BMC Genomics* 21.1 (Dec. 2020), p. 729. ISSN: 1471-2164.

[11] Q. Fu et al. “Automated proteomic sample preparation: The key component for high throughput and quantitative mass spectrometry analysis”. en. In: *Mass Spectrometry Reviews* 42.2 (Mar. 2023), pp. 873–886. ISSN: 0277-7037, 1098-2787.

[12] I. Holland and J. A. Davies. “Automation in the Life Science Research Laboratory”. In: *Frontiers in Bioengineering and Biotechnology* 8 (Nov. 2020), p. 571777. ISSN: 2296-4185.

[13] G. Lippi et al. “Estimating the intra- and inter-individual imprecision of manual pipetting”. In: *Clinical Chemistry and Laboratory Medicine (CCLM)* 55.7 (Jan. 2017). ISSN: 1437-4331, 1434-6621.

[14] J. A. Bryant et al. “AssemblyTron: flexible automation of DNA assembly with Opentrons OT-2 lab robots”. en. In: *Synthetic Biology* 8.1 (Jan. 2023), ysac032. ISSN: 2397-7000.

[15] M. Storch, M. C. Haines, and G. S. Baldwin. “DNA-BOT: a low-cost, automated DNA assembly platform for synthetic biology”. In: *Synthetic Biology* 5.1 (Jan. 2020), ysaa010. ISSN: 2397-7000.

[16] F. Lázaro-Perona et al. “Evaluation of two automated low-cost RNA extraction protocols for SARS-CoV-2 detection”. en. In: *PLOS ONE* 16.2 (Feb. 2021). Ed. by A. M. Abd El-Aty, e0246302. ISSN: 1932-6203.

[17] P. Camacho-Martínez et al. “Optimización y adecuación de la capacidad diagnóstica para la realización de grandes volúmenes de RT-PCR de SARS-CoV-2”. es. In: *Enfermedades Infecciosas y Microbiología Clínica* 40.8 (Oct. 2022), pp. 441–444. ISSN: 0213005X.

[18] X. Liu, S. P. Gygi, and J. A. Paulo. “A Semiautomated Paramagnetic Bead-Based Platform for Isobaric Tag Sample Preparation”. en. In: *Journal of the American Society for Mass Spectrometry* 32.6 (June 2021), pp. 1519–1529. ISSN: 1044-0305, 1879-1123.

[19] Y. Han et al. “Shotgun Proteomics Sample Processing Automated by an Open-Source Lab Robot”. en. In: *Journal of Visualized Experiments* 176 (Oct. 2021), p. 63092. ISSN: 1940-087X.

[20] P. Di Tommaso et al. “Nextflow enables reproducible computational workflows”. en. In: *Nature Biotechnology* 35.4 (Apr. 2017), pp. 316–319. ISSN: 1087-0156, 1546-1696.

[21] D. Merkel. “Docker: lightweight linux containers for consistent development and deployment”. In: *Linux journal* 2014.239 (2014), p. 2.

[22] D. Hanahan. “Studies on transformation of Escherichia coli with plasmids”. en. In: *Journal of Molecular Biology* 166.4 (June 1983), pp. 557–580. ISSN: 00222836.

[23] Opentrons. *Temperature Module for Heating and Cooling*. Apr. 2022.

[24] S. Taguchi et al. “Automation of yeast spot assays using an affordable liquid handling robot”. en. In: *SLAS Technology* 28.2 (Apr. 2023), pp. 55–62. ISSN: 24726303.

[25] J. K. Rogers et al. “Synthetic biosensors for precise gene control and real-time monitoring of metabolites”. In: *Nucleic Acids Research* 43.15 (Sept. 2015), pp. 7648–7660. ISSN: 0305-1048.

[26] D. Hanahan, J. Jessee, and F. R. Bloom. “[4] Plasmid transformation of Escherichia coli and other bacteria”. en. In: *Methods in Enzymology*. Vol. 204. Elsevier, 1991, pp. 63–113. ISBN: 978-0-12-182105-0.

[27] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri. “NIH Image to ImageJ: 25 years of image analysis”. en. In: *Nature Methods* 9.7 (July 2012), pp. 671–675. ISSN: 1548-7091, 1548-7105.