Received: 27 July 2022 | Revised: 29 September 2022 | Accepted: 9 October 2022

DOI: 10.1002/bit.28261

ARTICLE

BIOTECHNOLOGY AND BIOENGINEERING WILEY

Explore or exploit? A model-based screening strategy for PETase secretion by *Corynebacterium glutamicum*

Laura M. Helleckes[1,2] | Carolin Müller[1,2] | Tim Griesbach[1] | Vera Waffenschmidt[1] | Matthias Moch[1] | Michael Osthege[1,2] | Wolfgang Wiechert[1,3] | Marco Oldiges[1,2]

[1]Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, Jülich, Germany

[2]Institute of Biotechnology, RWTH Aachen University, Aachen, Germany

[3]Computational Systems Biotechnology (AVT.CSB), RWTH Aachen University, Aachen, Germany

Correspondence

Marco Oldiges, Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße, 52428 Jülich, Germany.
Email: m.oldiges@fz-juelich.de

Funding information

Deutsche Forschungsgemeinschaft; CLIB-Kompetenzzentrum Biotechnologie (CKB) funded by the European Regional Development Fund (grant 34. EFRE-0300097); Microbial Bioprocess Lab – A Helmholtz Innovation Lab", part of the Enabling Spaces Program "Helmholtz Innovation Labs" of the German Helmholtz Association

Abstract

Extracellular production of target proteins simplifies downstream processing due to obsolete cell disruption. However, optimal combinations of a heterologous protein, suitable signal peptide, and secretion host can currently not be predicted, resulting in large strain libraries that need to be tested. On the experimental side, this challenge can be tackled by miniaturization, parallelization, and automation, which provide high-throughput screening data. These data need to be condensed into a candidate ranking for decision-making to focus bioprocess development on the most promising candidates. We screened for *Bacillus subtilis* signal peptides mediating Sec secretion of two polyethylene terephthalate degrading enzymes (PETases), leaf-branch compost cutinase (LCC) and polyester hydrolase mutants, by *Corynebacterium glutamicum*. We developed a fully automated screening process and constructed an accompanying Bayesian statistical modeling framework, which we applied in screenings for highest activity in 4-nitrophenyl palmitate degradation. In contrast to classical evaluation methods, batch effects and biological errors are taken into account and their uncertainty is quantified. Within only two rounds of screening, the most suitable signal peptide was identified for each PETase. Results from LCC secretion in microliter-scale cultivation were shown to be scalable to laboratory-scale bioreactors. This work demonstrates an experiment-modeling loop that can accelerate early-stage screening in a way that experimental capacities are focused to the most promising strain candidates. Combined with high-throughput cloning, this paves the way for using large strain libraries of several hundreds of strains in a Design–Build–Test–Learn approach.

KEYWORDS

Bayesian statistical modeling, experiment-modeling loop, high-throughput screening, polyethylene terephthalate hydrolase, secretion

Laura M. Helleckes and Carolin Müller contributed equally.

1 | INTRODUCTION

Plastics are widely used for packaging, building, and construction, among other applications. In the European Union, Norway, Switzerland, and the United Kingdom, 29 million tons of post-consumer plastic waste were collected in 2020. Of this waste, still 23.4% were sent to landfill, instead of recycling or at least energy recovery (PlasticsEurope, 2022). The discovery of a polyethylene terephthalate degrading enzyme (PETase) from the bacterium *Ideonella sakaiensis* in 2016 has strongly promoted research in the field of biotechnological plastics degradation (Yoshida et al., 2016). Reviews on this topic and subsequent plastic upcycling can be found elsewhere (Kawai et al., 2019; Taniguchi et al., 2019; Tiso et al., 2021). Other examples for PETases are the thermostable leaf-branch compost cutinase (LCC), whose mutant ICCG outperforms the *I. sakaiensis* PETase (Sulaiman et al., 2012; Tournier et al., 2020). The PETase polyester hydrolase (PE-H) from the marine mesophilic bacterium *Pseudomonas aestusnigri* in contrast shows higher activity at lower temperatures already. The rationally engineered PE-H mutant Y250S is also able to hydrolyze polyethylene terephthalate (PET) film substrates from PET bottles (Bollinger et al., 2020). Both PETases are presumably secreted by their native host mediated by a signal peptide (Bollinger et al., 2020; Sulaiman et al., 2012).

Since it is not possible to predict a suitable combination of a signal peptide and target protein for the selected host for heterologous protein secretion, a set of combinations have to be tested (Brockmeier et al., 2006; Degering et al., 2010; Müller et al., 2022). Such signal peptide screenings can be accelerated using automation, parallelization, and miniaturization by using a microbioreactor integrated into a liquid handling platform (Hemmerich et al., 2016). Müller et al. (2022) established automated workflows from strain construction to enable high-throughput screening for heterologous Sec-dependent secretion with *Corynebacterium glutamicum* mediated by *Bacillus subtilis* signal peptides. The Gram-positive *C. glutamicum* is a promising host for protein secretion because it has been used for decades for large-scale industrial production of amino acids and other small molecules and has low extracellular protease activity (Freudl, 2017; Liu et al., 2021; Vertès, 2013). Best-performing *B. subtilis* signal peptides for heterologous *Fusarium solani* f. sp. *pisi* cutinase secretion were identified using automated workflows (Müller et al., 2022). This cutinase is also able to degrade amorphous PET, albeit with lower activities than the LCC (Tournier et al., 2020; Vertommen et al., 2005).

We demonstrate in this study that the workflows for strain construction and, in particular, secretion screening for *B. subtilis* signal peptides are transferable to the LCC mutant ICCG and PE-H mutant Y250S as target proteins. To our knowledge, this is the first approach to secrete PETases with *C. glutamicum*. We have optimized and extended the workflow for comparing 24 different strains in one run. The aforementioned techniques pave the way for generating and testing hundreds of combinations; experimental design and efficient data analysis thus become increasingly important topics.

Notably, early-stage screening experiments need to reliably reduce the number of candidate strains before process optimization and validation in laboratory- and pilot-scale bioreactors follow (Hemmerich et al., 2018). In other words, small-scale screenings need to result in a candidate ranking regarding a key performance indicator (KPI) that is the decision criterion for later stages. These KPIs often cannot be observed directly, but rather need to be derived using models, for example, small regression models (Hemmerich et al., 2018; Neubauer et al., 2013). A second challenge is the batch-wise screening process, which demands to select, in the case of this study, 24 variants for each round of cultivation. However, with the number of potential candidates exceeding this number, many strategies of experimental design can be applied. For example, the same number of replicates for each strain can be selected and then randomly distributed over batches. Another possibility is to perform a prescreening with few replicates before investigating the top x-percent in more detail. Especially if experiments are prone to batch effects between runs, the experimental design decisions become more and more complex. Moreover, experimental capacity needs to be focused to the most promising candidate strains to obtain exact KPI estimates with low uncertainty, while low performers need to be immediately sorted out, not wasting experimental resources on them.

Generally, these challenges constitute a classical exploration–exploitation trade-off: On the one hand, a detailed examination of all variants could be aimed at, which corresponds to the goal of exploration (i.e., choosing new variants). On the other hand, expensive or time-intensive experiments may require maximizing the exploitation (i.e., choosing the best-known variants). For example, a greedy algorithm would choose the candidate with the so far highest mean KPI to make use of the obtained knowledge (Russo et al., 2018). For efficient screening, both objectives need to be balanced to make sure that novel variants are explored but no replicates are wasted for unpromising candidates that have little probability to outperform the others. A heuristic to balance this exploration–exploitation trade-off is Thompson sampling (Thompson, 1933), which recently gained popularity in various fields due to its simple implementation and good efficiency (Agrawal & Goyal, 2012; Chapelle & Li, 2011). The broad range of applications includes the exploration of chemical design spaces (Hernández-Lobato et al., 2017), mobile health applications (Tomkins et al., 2021), or real-time content selection for web pages (Hill et al., 2017).

In this paper, we present a model-based screening strategy to obtain the KPIs and apply probability-based decision-making. We make use of a Bayesian process model that accounts for various experimental effects such as pipetting errors or batch effects between different cultivations. The probability distributions for the KPIs are the basis for Thompson sampling, which we use to plan experiments for the next iteration of batch cultivation. Together with the aforementioned automation workflow, this study presents an efficient screening of signal peptides for secretion of two PETases by *C. glutamicum* and demonstrates how the interaction of experiment

and model in a Design–Build–Test–Learn (Carbonell et al., 2018; Pouvreau et al., 2018) approach can enhance early bioprocess development. This versatile approach can be universally applied to other secretory target proteins of interest and signal peptide libraries of larger sizes from various species.

2 | METHODS

2.1 | Cultivation media and strains

C. glutamicum ATCC 13032 (Kinoshita et al., 1957) was cultivated at 30°C in CGXII (Unthan et al., 2014) or brain heart infusion (BHI) medium. Cultivation in flasks was done in baffled flasks with a filling volume of 0.1× flask volume and a shaking diameter of 25 mm. For cultivation on plates, 10 gl^{-1} agar-agar were added to BHI medium with additional 91 gl^{-1} D-sorbitol. Media were supplemented with 30 µg ml^{-1} kanamycin. *E. coli* strains DH5*α* and TOP10 (both Thermo Fisher Scientific) were used for DNA cloning and cultivated in lysogeny broth (LB) with Miller's modifications (Miller, 1972) and 50 µg ml^{-1} kanamycin at 37°C. Solid LB was prepared with additional 15 gl^{-1} agar–agar.

2.2 | Preparation of PETase secretion strains

C. glutamicum strains with pCMEx8-based secretion of PE-H mutant Y250S or LCC mutant ICCG mediated by *B. subtilis* signal peptides were prepared as described elsewhere (Müller et al., 2022). All plasmids used in this study are listed in Supporting Information: Table S1. Briefly, the signal peptide sequence of *B. subtilis yncM* in pCMEx8 was exchanged by cassette mutagenesis (Wells et al., 1985). Oligonucleotides were purchased from Eurofins (Supporting Information: Table S2), hybridized and ligated with NdeI/EcoRI-digested pCMEx8 to generate plasmids pCMEx8-[SP_2] with SP_2 = 10 different *B. subtilis* signal peptide sequences. Synthetic PETase genes lcc^{ICCG} and $pe\text{-}h^{Y250S}$ were provided by Synbio Technologies in pUC57-based plasmids and cloned into plasmids pCMEx8-[SP_1] (Müller et al., 2022) and pCMEx8-[SP_2] by Golden Gate assembly. After sequence verification by Eurofins, *C. glutamicum* ATCC 13032 was transformed by electroporation as previously described (van der Rest et al., 1999).

2.3 | Determination of cell density

The optical density (OD) was measured at 600 nm. A total of 0.9% (w/v) NaCl solution was used as a blank and for sample dilution to fit OD 0.1–0.3 in the linear range. For cell dry weight determination, weighed 1.5 ml reaction tubes were dried at 90°C for 24 h and put into a desiccator. A volume of 1.5 ml cultivation samples was transferred into these reaction tubes and centrifuged at 21,500*g* and 4°C for 10 min. The supernatant was discarded, cells were washed with 0.9% (w/v) NaCl and dried at 90°C for 24 h before weight determination.

2.4 | Strain maintenance

Strain master cell banks were prepared in BHI medium inoculated from a single transformant. After overnight incubation at 30°C and 250 rpm, 750 µl culture were mixed with 750 µl 100% (w/v) glycerol in a cryogenic vial to a final glycerol concentration of 50% (w/v). For working cell banks, 3.5 ml CGXII per well of a Riplate® SW 10 ml 24 Wells plate (Ritter) were inoculated with 100 µl of a respective master cell bank and incubated overnight at 30°C and 900 rpm. After centrifugation for 5 min at 2000*g*, the supernatant was discarded and cells were washed with 1.5 ml fresh CGXII. Cells were resuspended in 3.5 ml fresh CGXII medium and incubated for 5 h at 30°C and 900 rpm. A volume of 100 µl culture was mixed with 100 µl 50% (w/v) glycerol in a 96-well microtiter plate (MTP). The MTP was sealed with aluminum foil and stored at −80°C. For bioreactor cultivations, additional working cell banks were prepared in flasks. A volume of 50 ml CGXII was inoculated with 250 µl of a master cell bank and incubated for around 15 h at 250–300 rpm at 30°C with online biomass monitoring (SFR vario, PreSense Precision Sensing). Cells were harvested in the late exponential phase by centrifuging with 9283*g* for 5 min. Sedimented cells were resuspended to an OD 60 with 0.9% (w/v) NaCl and 2× diluted with 50% (w/v) glycerol. 1 ml aliquots were stored in cryogenic vials at −80°C.

2.5 | High-throughput PETase secretion screening

High-throughput PETase secretion screenings were conducted using a Tecan Freedom EVO® robotic platform with washable fixed tips made of stainless steel and coated with polytetrafluoroethylene (PTFE). The robotic platform includes an integrated BioLector® Pro microcultivation device (Beckman Coulter), centrifuge (Rotanta 460 Robotic, Hettich), cooled sample storage (cooling carrier connected to a Microcool MC 600, Lauda), and microplate reader (Infinite® M Nano, Tecan). A 96-well MTP with cryo-conserved cultures and CGXII medium in a trough were placed on the deck of the robotic platform. For precultures, 760 µl CGXII were transferred to 10–24 wells of a FlowerPlate® with pH and DO optodes covered with a gas-permeable sealing foil with perforated silicone layer for automation (Beckman Coulter). A volume of 40 µl of each cryo-conserved culture was used to inoculate one of the preculture wells. The remaining cryo culture wells in the MTP not used for precultures were discarded.

Standard cultivation conditions were 30°C, 1400 rpm and ≥85% relative humidity with measurement of backscatter, pH and DO every 13 min. Data from BioLector online measurements was parsed during the cultivation using the bletl package (Osthege, Tenhaef, Helleckes, et al., 2022; Osthege, Tenhaef, Zyla, et al., 2022). For full automation of the process, backscatter-triggered inoculation and induction as

well as time-triggered sampling are managed by worklists written with the robotools package (Osthege & Helleckes, 2021) and executed by a device control system developed by M. Osthege and J. Hemmerich (Forschungszentrum Jülich GmbH, in preparation).

As soon as a preculture exceeded a certain device-dependent backscatter threshold in the exponential phase, wells for one or three main cultures were filled with 780 µl CGXII and inoculated with 20 µl of the respective preculture. Target protein expression in the main cultures was induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 500 µM triggered by a device-dependent backscatter signal in the early exponential phase. Cells were harvested 4 h after induction by centrifugation for 6 min at 3756*g* and 4°C. Supernatants were stored in a 1 ml deep well plate on a cooling carrier until all cultivations were finished for subsequent cutinase activity assay. A scheme for the screening workflow with 24 precultures and 24 main cultures is depicted in Figure 1.

For the second round of LCC screening, in which only two different signal peptide variants were chosen by the algorithm described in Section 2.10, the design was changed to the workflow described by Müller et al. (2022) to require less cryo cultures and increase the number of main cultures. Here, only the 12 wells in the outer columns were used for preculture cultivation. The remaining 36 wells were used for main culture cultivation in triplicates inoculated from respective precultures.

2.6 | Bioreactor cultivation

A preculture of 50 ml CGXII medium in a baffled flask was inoculated from a cryo-conserved *C. glutamicum* PETase secretion strain. Cells were incubated for 16 h at 30°C and 300 rpm with online biomass monitoring (SFR vario, PreSense Precision Sensing) until the late exponential growth phase. After centrifugation at 9283*g* for 5 min at 4°C, cells were resuspended in 0.9% (w/v) NaCl.

The main culture was carried out in four parallel 1.5 L DASGIP® bioreactors equipped with two Rushton-type impellers (six blades, 1 cm height and 3 cm distance). Included DASGIP® modules were TC4SC4 for temperature and agitation control, PH4PO4 for control of DO and pH, MF4 for mass flow controlled gassing, MP8 for control of feed flow rates, and GA4 exhaust analyzer. DO was measured with VisiFerm DO 225 optical probe and pH with 405-DPAS-SC-K8S electrodes. Deviating from the standard CGXII medium with 20 $g l^{-1}$ glucose, 3-(morpholin-4-yl)propane-1-sulfonic acid and urea were omitted in bioreactor CGXII medium and 50 µg ml^{-1} kanamycin, as well as an appropriate amount of Antifoam 204 (Sigma-Aldrich) were added. One liter CGXII in a bioreactor was inoculated to an OD of 0.3 from a respective preculture, and cells were incubated in batch mode at 30°C. The DO was kept ≥40% with a constant airflow rate q_{in} = 1 vvm by adjusting the agitation speed n = 400–1500 rpm. The pH was adjusted to pH 7 with 18% (w/w) NH_4OH and 30% (w/w) H_3PO_4.

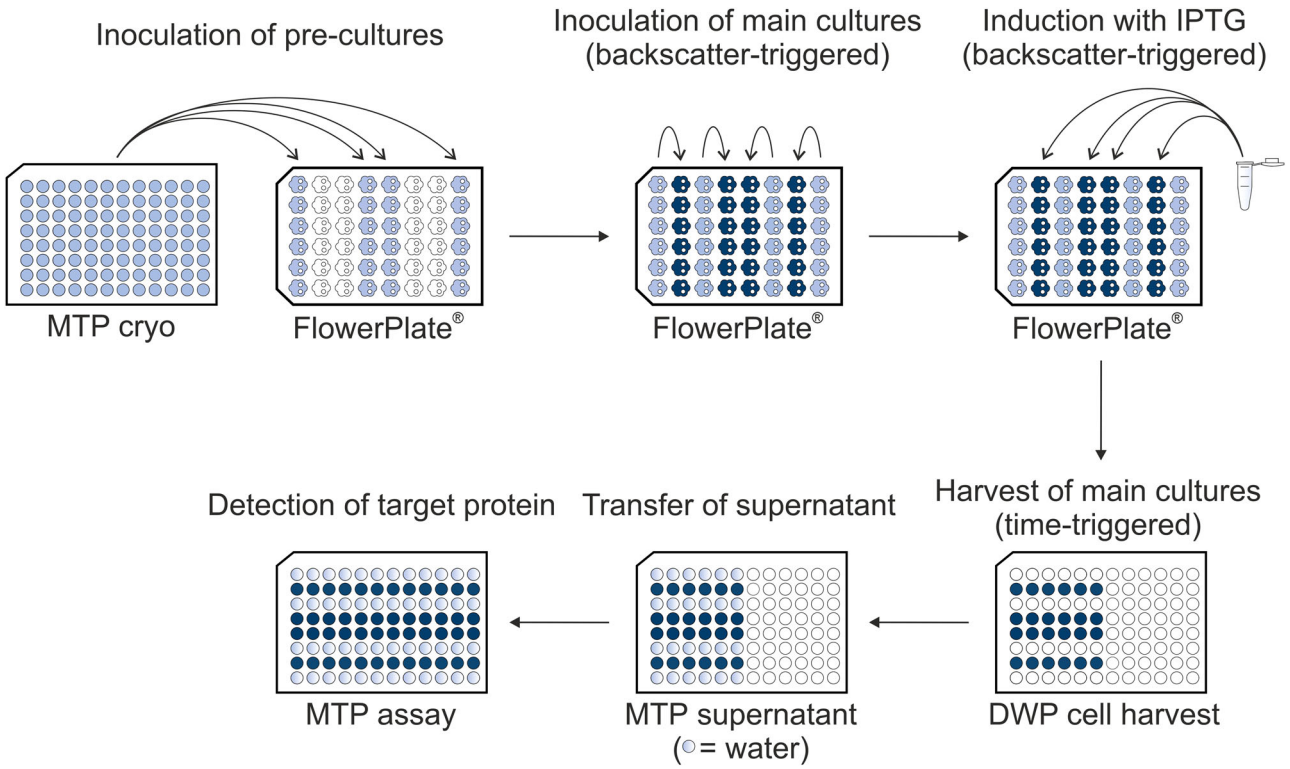

FIGURE 1 Screening workflow for 24 pre- and main cultures. Precultures are inoculated from cryo cultures frozen in MTPs. Upon reaching a certain backscatter level, 20 µl of the respective preculture were used to inoculate a main culture in the adjacent well. IPTG induction was triggered by backscatter in the early exponential phase. Harvesting of cells and centrifugation took place after 4 h. Samples were duplicated in an MTP for usage in the cutinase assay.

PETase secretion was induced around OD 1 by adding IPTG to a final concentration of 200 µM. If necessary, further sterile Antifoam 204 (Sigma-Aldrich) was added. Samples for measuring OD, cell dry weight, and activity assay were taken after inoculation as well as at six time points over the course of the fermentation (see Supporting Information).

2.7 | Cutinase activity assay

The activity of PETases in cultivation supernatant samples was determined spectrophotometrically by degradation of 4-nitrophenyl palmitate as a substrate analogue (Winkler & Stuckmann, 1979) as described by Müller et al. (2022). Briefly, nine parts reaction buffer (2.3 gl^{-1} Na-desoxycholate, 1.1 gl^{-1} gum arabic in 50 mM potassium phoshate buffer, pH 8) were mixed with one part 3 gl^{-1} 4-nitrophenyl palmitate in 2-propanol. A volume of 200 µl was filled into MTP wells and prewarmed to 37°C. Supernatant samples were appropriately diluted with 50 mM potassium phosphate buffer pH 8 (72× for BioLector samples and 12× for bioreactor samples). A volume of 40 µl diluted supernatant was pipetted into two wells filled with the reaction mix for technical duplicates. 4-nitrophenol formation was measured at 410 nm and 37°C for 40 min in a microplate reader.

A volume of 40 µl 0–2.3 mM 4-nitrophenol was mixed with 200 µl reaction mix in triplicates (final concentration 0–0.38 mM) and absorption was measured to convert absorption into product concentration.

In the traditional data analysis, enzymatic activities in U ml^{-1} were calculated in relation to the assay volume with the absorption A_{410} in a.u., slope of the 4-nitrophenol standard $m_{standard}$ in a.u. mM^{-1} and the unitless supernatant dilution factor DF (Equation 1). Ninety-five percent confidence intervals (95% CIs) were calculated with t-values from replicates.

$$EA = \Delta A_{410} \times \frac{1}{m_{standard}} \times DF \tag{1}$$

2.8 | Software

All analyses and plots presented in this study were performed with recent versions of Python 3.8, PyMC ≥4.0.0b2 (Wiecki et al., 2022), ArviZ ≥0.11.4 (Kumar et al., 2019), matplotlib ≥3.5 (Hunter, 2007), NumPy ≥1.20 (Harris et al., 2020), pandas ≥1.3 (McKinney, 2010; Reback, 2021), SciPy ≥1.7 (Virtanen et al., 2020) and related packages. For calibration models, the in-house developed, publicly available calibr8 package was used with versions ≥6.2.1 (Helleckes et al., 2022; Osthege & Helleckes, 2022). For parsing of BioLector data, the bletl package (Osthege, Tenhaef, Helleckes, et al., 2022; Osthege, Tenhaef, Zyla, et al., 2022) was applied. Photometric measurement were analysed using the in-house developed retl package (not published). The robotools Python package was used to facilitate multistep liquid-handling instructions on the robotic platform (Osthege & Helleckes, 2021). For a full list of dependencies, see the accompanying GitHub repository (Helleckes & Osthege, 2022).

2.9 | Process model

With the term process model, we refer to a model describing the enzymatic assay and the experimental effects influencing it, rather than a description of the unit operations in cultivation and analytics. The process model constructed in this paper is a Bayesian hierarchical model build with the PyMC Python package. A brief introduction to Bayesian process modeling is given here, while further details and mathematical theory can be found elsewhere (Gelman et al., 1995; Schmid & Brown, 2000; van de Schoot et al., 2021).

$$p_{\text{posterior}}(\text{rv} \mid \text{d}) = \frac{p_{\text{prior}}(\text{rv}) \times \mathcal{L}(\text{rv} \mid \text{d})}{\int p_{\text{prior}}(\text{rv}) \times \mathcal{L}(\text{rv} \mid \text{d})\ d\ \text{rv}} \tag{2}$$

According to Bayes' theorem (Equation 2), the probability of a random variable (rv) *given* some data (d), also called posterior probability, can be described in terms of the prior distribution $p_{\text{prior}}(\text{rv})$ and the likelihood $\mathcal{L}(\text{rv} \mid \text{d})$. The former term refers to the prior belief in the problem without seeing any data. In cases of biological application, this can for example include thermodynamically known constraints or knowledge about the biological system. The other term, the likelihood $\mathcal{L}$, is a purely data-driven expression quantifying how likely it is to make observations d given a fixed value of the random variable. It assumes that the random variable will follow a certain probability distribution when observed, for example, a normally distributed measurement noise.

In Bayesian inference, Bayes' theorem is used to estimate model parameters from data. Particularly in cases for which limited data is available, incorporating prior knowledge can facilitate parameter estimation. Moreover, the approach results in a probability distribution for each parameter instead of a point estimate, thus including uncertainty quantification. To estimate process parameters for this study, prior distributions were chosen for each parameter of the model accordingly. A graphical representation of the complete process model can be found in Figure 2.

Starting in the lower left, the measured product absorbance in the cutinase assay (Section 2.7) is evaluated against the modeled concentration using a customized likelihood. For this, a calibration model was used that was fitted with the calibr8 Python package (Osthege & Helleckes, 2022), assuming a linear relationship between product absorbance and concentration and a Student-*t*-distributed measurement noise. A plot of the fitted calibration model can be found in the Supporting Information: Figure S3. More details on the choice of likelihood distributions to describe measurement noise can be found elsewhere (Helleckes et al., 2022).

The product concentration during the assay was assumed to follow a first order mass action law, such that

$$P_t = S_0(1 - e^{(-k_{\text{assay}}t)}), \tag{3}$$

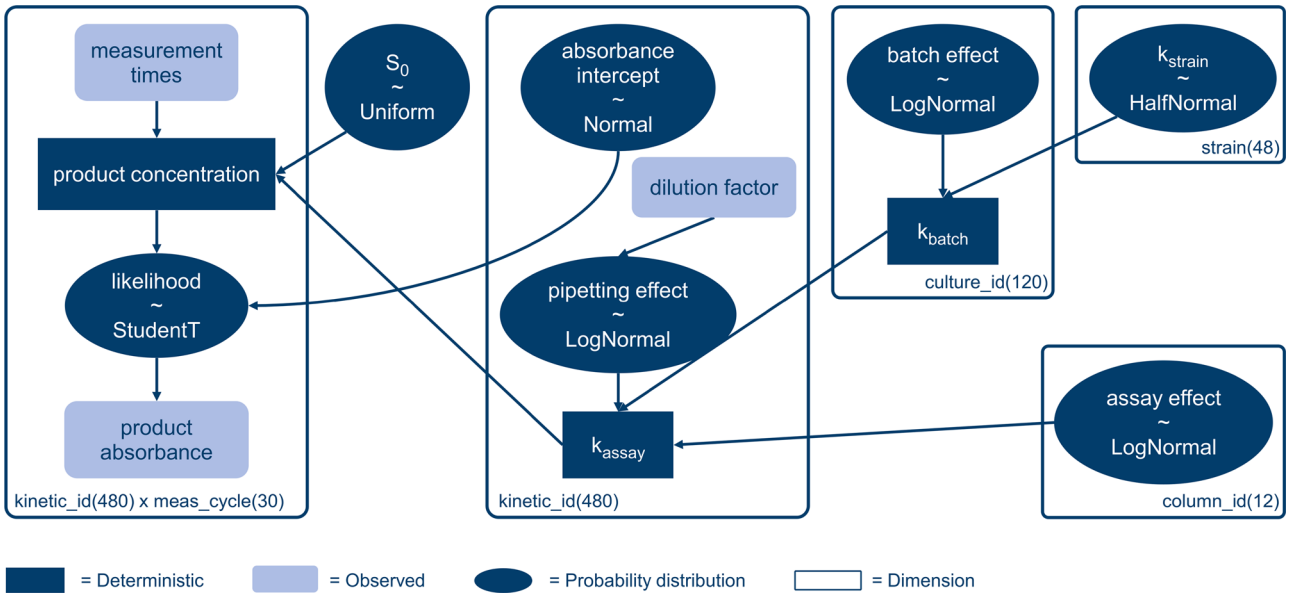

FIGURE 2 Graphical representation of the process model. The KPI used for ranking is the rate constant of each strain, k_{strain} (upper right).

with P_t as the product concentration at time t, S_0 as initial substrate concentration and k_{assay} as the rate constant.

An additional absorbance offset was observed at the beginning of each measurement, which could not be explained with the offset obtained from the calibration model. It was therefore modeled by the parameter `absorbance intercept`. While k_{assay} is the parameter that is observed in each reaction well in the MTP, the KPI of interest for the strain ranking is the actual rate k_{strain} of each variant. However, several biological and technical effects were found that influence this rate constant, for example, an effect of the position in the assay MTP (`assay effect`), which is further discussed in Section 3.1. Moreover, variance between batch runs, for example caused by small differences in viable cells in cryo cultures, as well as the pipetting error of robotic liquid handling are taken into account. In Figure 2, the route from k_{strain} (top right) down to k_{assay} (bottom middle) describes the assumptions on how the catalytic performance of each strain is perturbed by different sources of uncertainty before the apparent kinetic rate in one MTP well in the assay can be observed.

A mathematical notation of the model can be found in the Supporting Information. Additionally, the code for the process model can be found on the accompanying GitHub repository (Helleckes & Osthege, 2022). Posterior probabilities were obtained by Markov chain Monte Carlo sampling, using the No-U-Turn sampler (Hoffman & Gelman, 2014) in PyMC. Convergence checks and inspection of the traces were performed using ArviZ.

2.10 | Thompson sampling for experimental design

As described in Section 2.9, the reaction kinetics of PETases in the cutinase activity assay are modeled by a first order mass action, where k_{strain} is the rate constant of each strain in this reaction. The model inherently assumes that the reaction rate is proportional to the amount of secreted enzyme, thus k_{strain} was chosen as the KPI for a strain ranking. The posterior probabilities after each cultivation round were used as the basis for Thompson sampling. In this process, a batch of strain replicates for the following round was drawn by parallel Thompson sampling (Kandasamy et al., 2017) using the `sample_batch` function from the pyrff package (Osthege & Felton, 2020) and then randomly assigned to preculture wells in the FlowerPlate®. The obtained process data was then used to rerun the process model to obtain posterior probabilities of KPIs for the next round and so on.

3 | RESULTS AND DISCUSSION

3.1 | Model building and experimental learning

An essential part of the screening workflow is the Bayesian process model, which is described in more detail in Section 2.9. Before applying the process model to the LCC and PE-H secretion strain libraries, the influence of different experimental effects was investigated and an iterative cycle of model building and experimental improvements was performed. In the following, two findings of this workflow improvement are highlighted: first, the choice of the reaction kinetics is explained, second, a bias in the detected activities caused by the position of samples in the assay MTP is detailed.

In previous studies (Hemmerich et al., 2018; Müller et al., 2022), the slope of absorption in the cutinase activity assay has been used to calculate enzymatic activities. Therefore, a process model assuming linear reaction kinetics was constructed first (Figure 3a,c). Figure 3a shows the trajectory of absorbance value over the incubation time of

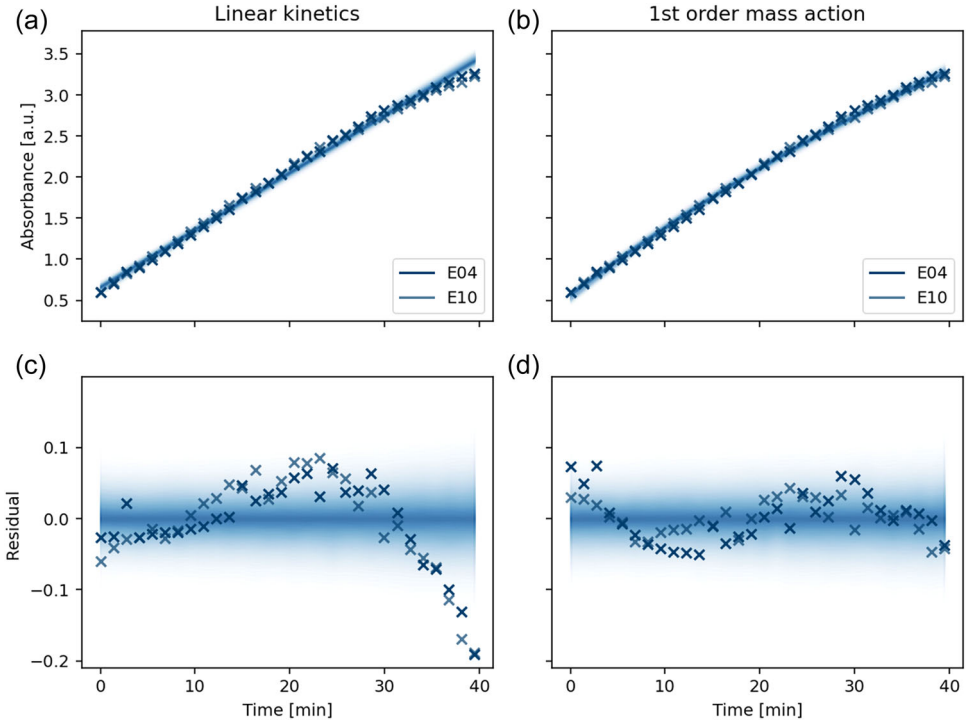

FIGURE 3 Comparison of linear (a, c) and exponential (b, d) model for reaction kinetics in wells E04 and E10 of PE-H screening in Round 1. Plots (c) and (d) show the residuals between data and model fit. The density bands visualize the distribution of predicted absorbance by the model.

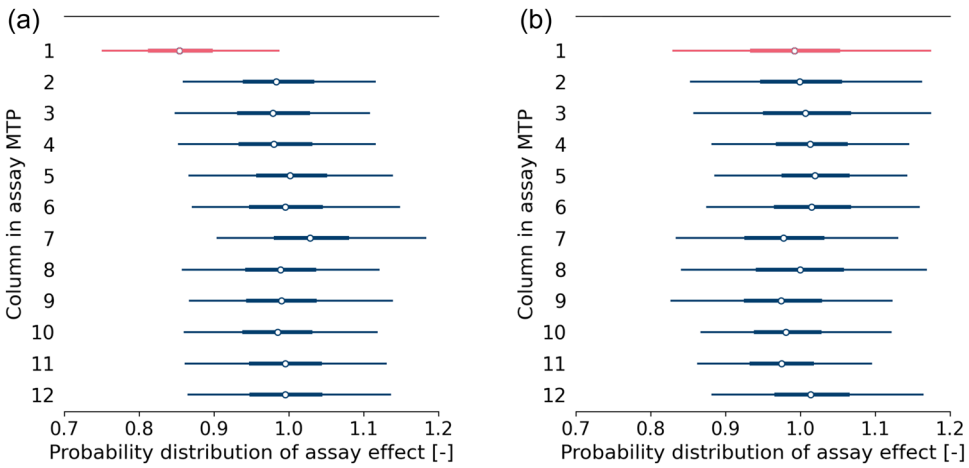

FIGURE 4 Influence of position in assay MTP before (a) and after (b) experimental optimization of the assay. The probability distributions of the assay effect for each column is shown. An assay effect of 1 indicates that the position has no influence while smaller and higher values indicate a negative and positive influence on the activity, respectively.

40 min for the LCC combined with LipB as a signal peptide, a variant with a comparably high activity (see Figure 8). After 35 min assay reaction time, the model significantly deviates from the measurements, which is also evident from the lack-of-fit indicated in the residual plot (Figure 3a,c). The trajectory indicates that the lower substrate concentration seems to limit the reaction rate towards the end of measurement. Therefore, the kinetics was changed to a first order mass action law (Figure 3b,d). This exponential trajectory describes the kinetics with a lower lack-of-fit (Figure 3d). The first order kinetics were thus chosen in the final process model.

As part of model building, the influence of pipetting errors, batch effects and positional effects was investigated within the different MTPs of the automated screening workflow (Section 2.5). Figure 4 shows one finding, namely a positional effect, where all 8 wells in the first column of the 96-well MTP show a systematically lower activity of the PETases (Section 2.9). The plot includes the median (white dot)

as well as intervals of 94% probability (whiskers) for the parameter `assay effect` in all 12 columns of the assay MTP. A value of 1 indicates that the respective column has no influence on the activity, values smaller and greater than 1 correspond to lower and higher activities caused by the position in the MTP. The probability distributions are deduced from replicate data of the same signal peptide which lie in different columns. Figure 4a shows the assay effect before experimental optimization. Here, it can be seen that wells in the first column show systematically lower activities with a median of around 0.85. However, the probability distribution interval of the first column is similar to all other columns, indicating that all wells of this column are affected in the same manner. With this feedback from the model, the experimental workflow of the cutinase activity assay was re-examined.

The robotic liquid handler uses fixed tips, which distribute the substrate solution column-wise. The substrate solution dispense step of the last column is followed by a washing step with water to clean the needles and to prevent carry-over of substrate. However, dispense of the substrate in the first column was done with needles that were not conditioned with the organic substrate solution. Since the pipettes were not wetted initially, the first column contained slightly less substrate. Wetting of the PTFE coating of the stainless steel tips with the organic substrate 4-nitrophenyl palmitate diminished the positional effect of the first column. The experimental procedure was thus changed to include a step of tip wetting before the substrate is distributed into the assay MTP. The improved result can be seen in Figure 4b.

After optimization, the positional effect of the first column was resolved and the posterior probabilities of the parameter `assay effect` are now similar for all columns. The prior for the assay effect is $f_{\text{assay effect}} \sim \text{LogNormal}(\mu = 0, \sigma = 0.1)$, which was chosen to allow for fluctuation of the activity around 20%. The highest density interval containing 94% of the prior probability mass is approximately $[0.83, 1.21]$. Overall, the distributions for all columns fluctuate in a similar range, with the median close to 1. This shows that the remaining uncertainty in the parameter `assay effect` is now reflecting random noise, rather than the systematic bias that was previously seen due to the positional effect. This shows that the assay effect was strongly reduced by the improvements of the pipetting operation and it can be neglected in the process model for future experiments. However, since the model can quantify and correct for the modulation of activity caused by the pipetting effect, it would still be possible to include data from such experiments in the analysis.

To conclude, the examples show that the Bayesian process parameters are easy to interpret, since they mimic experimental errors. The width of their probability distributions can give an insight on how well the respective effect is characterized by the observed data. The analysis can also point out experimental sources of error and targets for workflow improvements, as it was demonstrated for the parameter `assay effect`. It was thus shown how the process model can guide experimental developments in an iterative cycle, where experimental improvements can simplify the required process model. An additional influential factor, namely a batch effect between different cultivation rounds, is discussed in section Section 3.3.

3.2 | Screening of PE-H

After the successful strain construction of the *C. glutamicum* variants secreting the PETases PE-H mutant Y250S and LCC mutant ICCG mediated by 24 different *B. subtilis* signal peptides, secretion performance was examined. It has to be noted here that enzymatic activities and reaction kinetics are determined as 4-nitrophenyl palmitate hydrolytic activities rather than PET degradation. Nevertheless, the degradation of 4-nitrophenyl palmitate is a valid measure to compare the secretion performance between different signal peptides for the secretion of the same enzyme. However, secretion performance cannot be compared in absolute values regarding PET hydrolysis. This is due to the structural differences of the LCC and the PE-H (Bollinger et al., 2020; Sulaiman et al., 2014), which is why they can presumably convert the 4-nitrophenyl palmitate substrate to different extents. Moreover, for method development, which is the actual focus of this study, it is helpful to use an already established and automated activity assay for the detection of secreted target proteins.

First, the process model was applied to the strain library of 24 signal peptides combined with the PE-H mutant Y250S. To facilitate the process of model building and assess model parameters such as the batch effect, the first round of screening consisted of two batch cultivations as described in Section 2.5. In each batch cultivation, all 24 variants were cultivated, resulting in 24 main cultures which were automatically inoculated from the respective 24 precultures, induced with IPTG and harvested. The supernatant of each culture was subjected to the cutinase activity assay in duplicates (Section 2.7). Accordingly, the two batch cultivations led to four replicates for activity measurements per variant, which were simultaneously analyzed with the process model. In the following, the combined analysis of these two batch cultivations is always referred to as Round 1.

As a KPI for finding the most suitable signal peptide for PE-H secretion, the rate constant k of the first order reaction kinetics (Section 2.9) was chosen. The more active enzyme is present in the supernatant, the higher the rate constant k of the variant. The resulting probability of the KPI for Round 1 is depicted in Figure 5a.

The plot reveals that more than half of the signal peptides lead to a low rate constant, indicating that little active enzyme was present in the supernatant. This can be the consequence of either low overall secretion or misfolding and thus inactive enzyme. Strikingly, four technical replicates are enough to distinguish these unpromising candidates from the remaining top 6 *B. subtilis* signal peptides, namely YpjP, YolA, Pel, LipA, YoaW, and Mpr. Biological as well as technical errors such as the pipetting, which are taken into account by the model, cause uncertainty in the estimation, which is reflected by the broad distributions for k for the best signal peptides. For example, the 94% probability range (highest density interval) of the

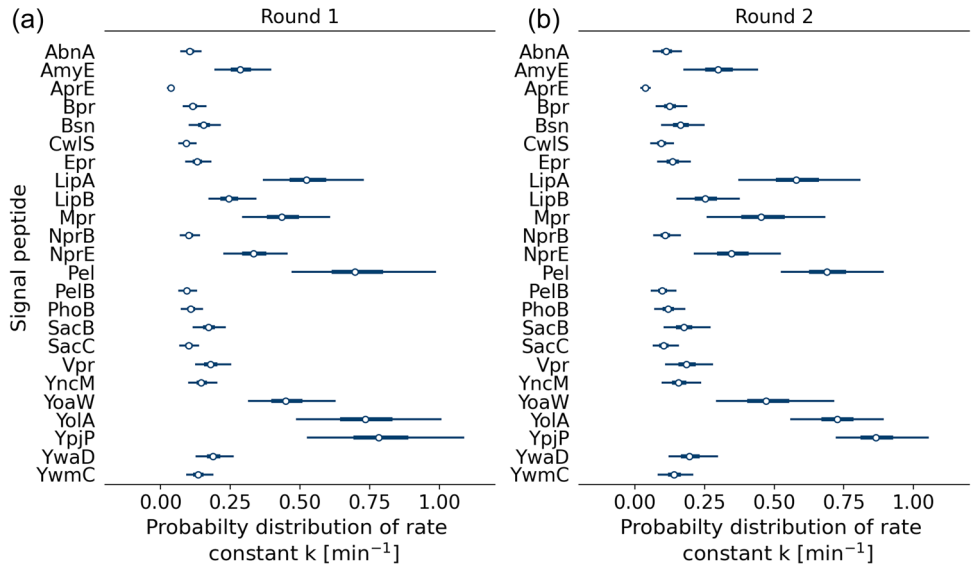

FIGURE 5 Posterior probability distributions of rate constant *k* for each signal peptide variant of PE-H after Round 1 (a) and Round 2 (b).

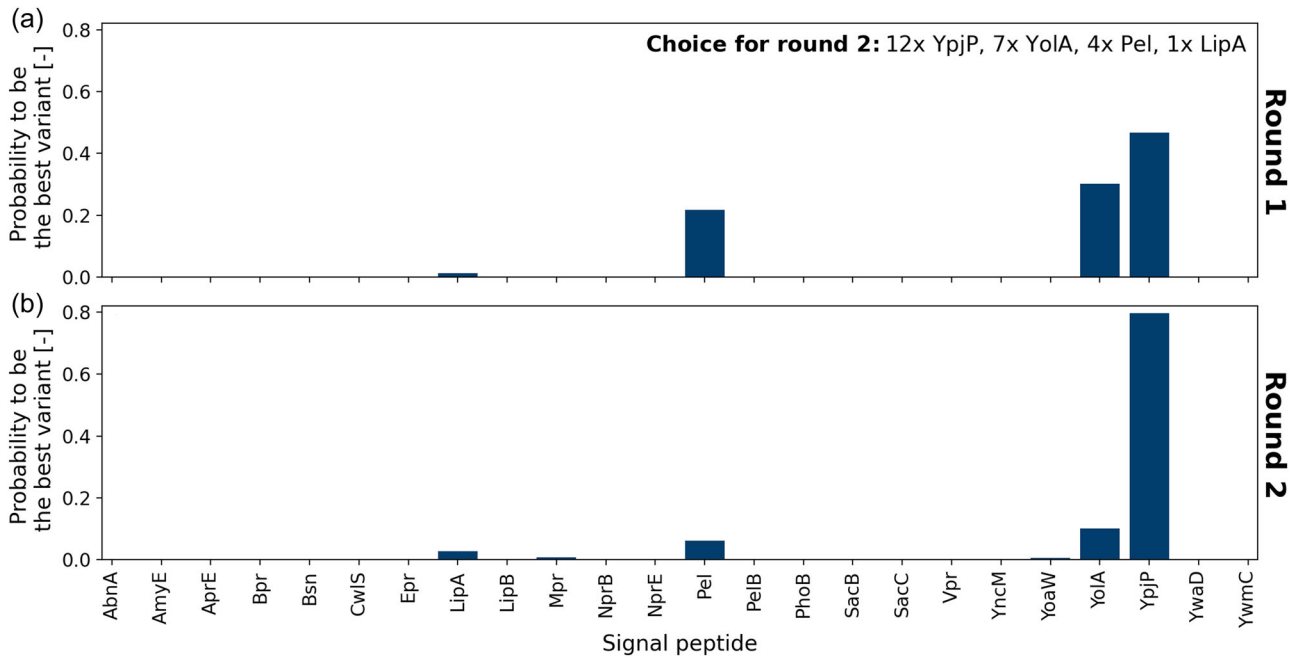

FIGURE 6 Probability for each signal peptide to be the best variant for PE-H secretion after (a) first and (b) second round of screening.

rate constant *k* for variant YpjP lies between [0.53, 1.09] min^{-1}. The distributions shown here are the basis for the Thompson sampling algorithm to draw variants for the next cultivation round. To understand this process, the probability for each variant to be the best signal peptide so far was calculated as described in Section 2.10. The results are shown in Figure 6a.

After the first round of two biological replicates per variant with four technical replicates of the assay, only four signal peptides have a probability significantly greater than zero to be the best performing one. This demonstrates the exploitation part of Thompson sampling. Since the process model can clearly distinguish low performing variants from the top ones, they have no chance to be chosen for the next rounds. However, exploration is still necessary to distinguish the top four variants. Within these, YpjP has the highest probability of being the best signal peptide at around 46%. It is followed by YolA with around 30%, Pel with around 22% and LipA with less than 2%. For comparison, using *C. glutamicum* and the same plasmid expression system for secretion of a heterologous cutinase resulted in very poor secretion performance for the *B. subtilis* signal peptides from Pel and YpjP (Müller et al., 2022). These results confirm that the selection

of an appropriate signal peptide is heavily dependent on the target protein.

As indicated by the text in subplot 6a, the choice for the next round reflects the probability of a signal peptide being the best, while small deviations in the ratio are caused by the small batch size of only 24 main culture wells. The whole screening workflow was conducted a second time with the respective replicates. Figure 6b shows the results after the second screening round. With overall 28 replicates from three batch cultivations, YpjP shows the highest probability of around 80% to be the best signal peptide, while YolA, Pel, LipA and Mpr show significantly lower probabilities smaller than 10%. The resulting probability distributions for the KPIs can be seen in Figure 5b. For comparison, in a strategy with equal numbers of replicates per signal peptide, it would have taken 14 rounds of cultivation and thus more than four times longer to obtain the same certainty about the KPI for the most suitable signal peptide YpjP for PE-H secretion. In another comparison, we only used three batch cultivations, corresponding to 72 main culture wells. With the same number of cultivation wells and rounds, each signal peptide could also be screened in a classical triplicates design ($24 \frac{\text{replicates}}{\text{round}} \times 3\,\text{rounds}/24\,\text{signal peptides} = 3\frac{\text{replicates}}{\text{signal peptide}}$). However, this would only allow to screen each signal peptide variant three times, resulting in six technical replicates, instead of the 28 replicates achieved for YpjP with the smarter experimental design. This shows the efficiency of combined process model and Thompson sampling algorithm.

3.3 | Screening of LCC

For signal peptide screening for secretion of the LCC mutant ICCG, the same strategy was used as for the successful screening with the secretory PE-H target protein. The first round comprised four technical replicates per variant accordingly, while the second was composed of suggestions from Thompson sampling. The resulting rankings after each round concerning the rate constant *k* are shown in Figure 7.

Again, compared to the heterologous secretion of the *F. solani* f. sp. *pisi* cutinase (Müller et al., 2022) or to the PE-H, two other signal peptide sequences, namely those of LipA and LipB, led to the highest extracellular LCC activities. All other signal peptides showed significantly less activity in the enzyme assay, more precisely less than half the rate constant. These results highlight the need to determine an appropriate signal peptide sequence for each target protein and thus the potential for combining automation and process models with Thompson sampling for accelerated screening. Moreover, it can be seen for Round 1 (Figure 7a) that the activities after secretion mediated with signal peptides LipA and LipB were very similar when taking measurement uncertainties into account, where the median (white dot) of LipB is slightly higher.

For Round 2 (Figure 7b), Thompson sampling selected the variants with LipA and LipB to an equal amount of replicates accordingly. With these data, the variants can be separated to a higher extend, where LipA shows a higher median in the probability distribution for *k*. Moreover, the higher number of replicates lead to narrower intervals of 95% probability (whiskers) compared to Round 1. Although the variants other than LipA and LipB were not measured in Round 2, their distributions are wider than before. This is caused by the the parameter `batch effect` of the process model (Section 2.9), which assumes an experimental error between different biological replicates in different rounds. Between Round 1 and Round 2, it could be observed that the activities increased for both LipA and LipB (compare Figure 10 in Section 3.5) due to an apparent batch effect. After only one round, no such effect could be assumed and the width of distributions is

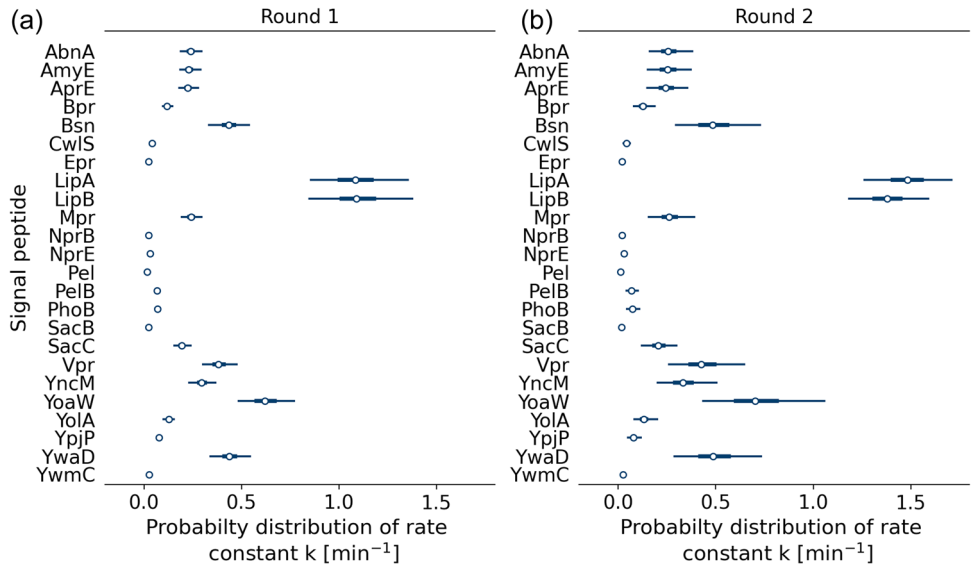

FIGURE 7 Probability distributions of rate constant *k* for each signal peptide variant of LCC after Round 1 (a) and Round 2 (b).

thus mainly influenced by deviation between technical replicates in the assay. However, with the information gained from LipA and LipB in both rounds, the process model can now account for the batch effect. Since the other variants were only measured in Round 1, but uncertainty is added by the batch effect, the posterior distributions in Figure 7b are wider. The batch effect will be further discussed in Section 3.5.

Taking the probability distributions for k from Figure 7, the question which signal peptide has the highest probability to be the best variant can be answered (Figure 8). As for the ranking, Figure 8a illustrates that LipA and LipB were very close after Round 1, with LipB having a slightly higher probability to be advantageous concerning secretion. In Thompson sampling, the candidates for the next round are drawn randomly according to their probability of being optimal based on the available data. Here, probabilities of LipA and LipB are close to 50%, which results in 18 main cultures of each variant being suggested for Round 2. Figure 8b shows the results after analysis of all 40 technical replicates of LipA and LipB after two rounds. Although the measurements were very similar for both variants (see classical statistical analysis in Section 3.5), the pairwise comparison of posterior samples results in a probability of around 75% that LipA is the best signal peptide. This shows the advantages of the process model, as no distinction between LipA and LipB could be seen in the classical statistical analysis. Importantly, the previously discussed positional effects and the batch effect between Round 1 and 2 are taken into account, assuring that the rate constant k of the secreted enzyme is the dominant factor in the ranking. To investigate whether the screening is representative for preselection of variants, both LipA and LipB, as well as the third-best variant YoaW, were chosen for a scale-up application.

3.4 | Comparison with batch fermentation in liter-scale bioreactors

To examine the scalability of the microcultivation process and to validate the results, *C. glutamicum* was cultivated in laboratory-scaled stirred tank reactors with 1 L working volume. LCC secretion mediated by the best *B. subtilis* signal peptides from microcultivation in the BioLector, namely LipA, LipB, and YoaW, were compared with laboratory-scale bioreactor cultivation data depicted in the Supporting Information (Figures S1 and S2). Maximum oxygen transfer rates of almost 100 mmol $L^{-1}h^{-1}$ were reached after approximately 12 h and the final cell dry weights were around 9.1–9.6 gl^{-1}. The probability distributions of the rate constant k between the microliter- and liter-scale are compared in Figure 9.

The laboratory-scale bioreactor data (red) show a similar ranking for the LipA and LipB signal peptides with YoaW approximately 40% behind. In comparison to the first BioLector run (blue), the laboratory-scale bioreactor data result in a very similar performance ranking. Thus, protein production with *C. glutamicum* could be transferred from the shaken cultivation in the BioLector in microliter-scale to the stirred tank reactor in the liter-scale, as demonstrated elsewhere for cutinase secretion by *C. glutamicum* (Rohe et al., 2012). However, uncertainties are higher for LipA and LipB in the liter-scale bioreactor despite the equal number of biological replicates (n = 2). Besides biological variance, the higher uncertainty is likely to be caused by the different scheme of sampling. While all BioLector experiments were sampled 4 h after induction, which corresponds to the early stationary phase, the fermentation was sampled at six different time points before and after induction. Due to different growth behaviors at the two scales, samples cannot be compared at the exact same process time. However, it is possible to take samples in identical

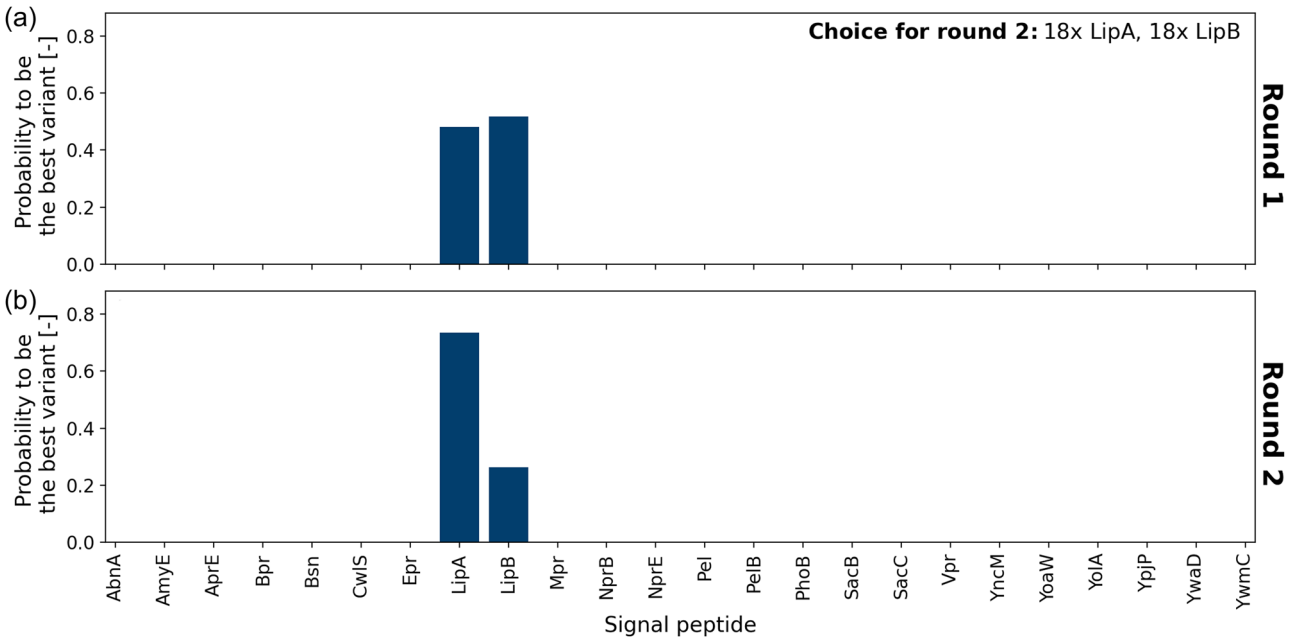

FIGURE 8 Probability for each signal peptide to be the best variant for LCC secretion after first (a) and second (b) round of screening.

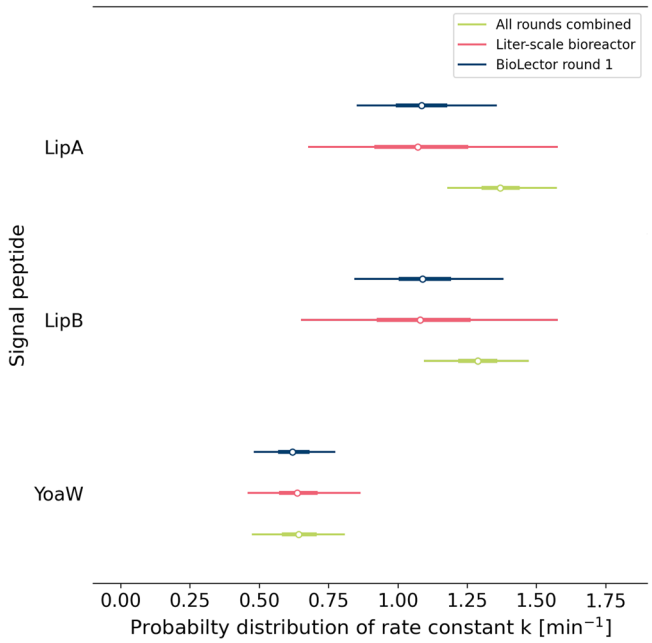

FIGURE 9 Comparison of LCC rankings after microliter-scale BioLector cultivation 1, fermentation in liter-scale bioreactors and the combined analysis of all BioLector and fermentation runs.

growth phases. Accordingly, all available time points in the stationary phase were used for analysis of liter-scale bioreactor data. In case of LipA and LipB, these correspond to 22, 25, and 26 h process time (see Supporting Information). The combination of three different sampling points with slightly different activities is likely to account for the higher uncertainty in the estimated rate constant *k* of LipA and LipB.

For the YoaW signal peptide, which was measured with four biological replicates in the bioreactor, the uncertainties are more similar to those of BioLector replicates. Moreover, only one time point lied within the stationary phase, which was after 23.5 h process time. This also suggests that the different time points have an influence on the uncertainty. In the future, a time-dependent analysis for the activity could be conducted for both scales. However, the obtained results for the rate constant *k* are in good accordance for different scales. This demonstrate the advantage of the high-throughput approach, which is representing the larger scale well.

The probability distributions in Figure 9 shown in green are the results of combining Round 1 of microliter-scale cultivation and the liter-scale bioreactor data with BioLector Round 2, where additional 18 biological replicates of LipA and LipB were measured. In contrast to the other probabilities, rate constants *k* for LipA and LipB can be better distinguished in the combined case. Moreover, a batch effect becomes evident, which can be detected from the higher overall median in the combined case. This is caused by overall higher activities in BioLector Round 2.

Overall, the scale-up shows that a large number of replicates is necessary to see differences of strain variants with a very similar secretion performance. This is often not feasible in laboratory-scale bioreactor cultivations due to the higher costs and efforts. For this purpose, microliter-scale cultivations are beneficial as they allow for a large number of replicates and thus a better resolution of very similar variants.

3.5 | Comparison to linear regression analysis

As a final validation of the method, a classical statistical analysis by linear regression of the absorbance measurements was conducted as described in Section 2.7 (Figure 10).

After the first round, LipA (dark blue) and LipB (light blue) cannot be distinguished, which can be seen by the overlapping 95% CIs between the variants. While the activities are still similar between the variants in the second round, the higher number of replicates leads to smaller CIs and indicates that LipA is the better signal peptide. However, the large batch effect between Round 1 and Round 2 for both variants makes it difficult to combine the results of both rounds in a purely statistical analysis. A main challenge is that technical and biological error could not be separated if all data are pooled for the linear regression, leading to large CIs that do not allow to distinguish the variants. This is important though to exclude that LipA is favored by positional or batch effect in the second round. Batch effects most likely result from the activity assay and not from the cultivation, since differences in cryo culture viability are compensated by backscatter-triggered individual inoculation of the main cultures and induction of target protein expression and secretion (Müller et al., 2022). In the activity assay, the reaction mix has a limited stability only (Hemmerich et al., 2018). Although directly prepared before each

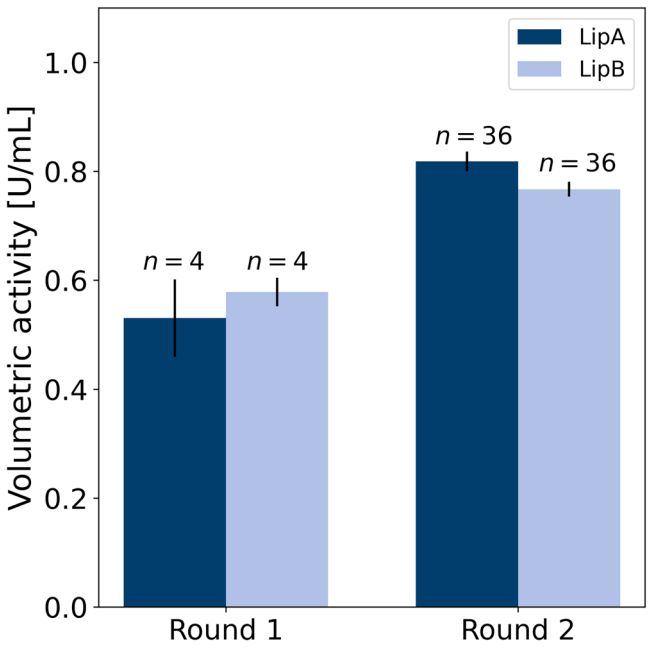

FIGURE 10 Results for LipA (dark blue) and LipB (light blue) for two rounds of screening with classical data analysis by linear regression. Error bars represent the 95% confidence intervals. The number of replicates refers to the technical replicates.

assay, the reaction mix could impact enzymatic 4-nitrophenyl palmitate degradation due to minor deviations in the composition and the poor solubility of the organic substrate. Here, the process model proves advantageous, since the uncertainty can be assigned to several experimental and biological effects and the data from different rounds can inform the probability of both, batch effects and pipetting errors. Using the process model we determined a probability of around 75% for LipA being the best signal peptide (see Figure 8). Overall, this demonstrates how the approach leads to more insight and faster screening and is thus promising for large libraries.

4 | CONCLUSION

To our knowledge, Sec-dependent secretion of the heterologous PETases LCC mutant ICCG and PE-H mutant Y250S could be shown for the first time using *C. glutamicum* as a secretion host. Utilizing our automated workflows for high-throughput screening and the accompanying modeling, the best suitable out of 24 *B. subtilis* Sec-type signal peptides was identified within two rounds of screening only. Even strain variants that were indistinguishable with classical methods due to batch effects could be resolved. The successful transfer from shaken cultivation in the BioLector to 1 L stirred tank bioreactors demonstrates the scalability of the secretion screening results.

Combined with the automated workflows for high-throughput cloning, this paves the way for screening of hundreds of strains or different process conditions in the Design–Build–Test–Learn cycle. Especially for application examples with greatly increased design space, where screening of all variants in multiple replicates becomes infeasible due to restrictions of resources in terms of both time and cost, model-based decision making as presented in this study is a valuable addition to the experimental high-throughput platforms.

In the future, the process model could be further extended by process conditions such as temperature. By solving these complex optimization problems such as strain and process conditions, our modeling approach can become an even more powerful tool to tackle the exploration–exploitation dilemma even in expanded design spaces in the future.

AUTHOR CONTRIBUTIONS

Laura M. Helleckes and Carolin Müller wrote the manuscript. *C. glutamicum* PETase secretion strains were prepared by Carolin Müller, Tim Griesbach and Vera Waffenschmidt. Laura M. Helleckes, Carolin Müller, Tim Griesbach, and Vera Waffenschmidt contributed to the automation of the screening workflows and Tim Griesbach screened for PETase secretion in microliter-scale. Matthias Moch did cultivations in laboratory-scale bioreactors and measured the PETase activities in the respective samples. Laura M. Helleckes build the process model with contributions from Michael Osthege. Laura M. Helleckes analyzed data and implemented Thompson sampling for experimental design. Marco Oldiges and Wolfgang Wiechert contributed to the study design, review and editing. Marco Oldiges was responsible for funding acquisition, project administration and supervision. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors thank the CLIB-Competence Center Biotechnology (CKB) funded by the European Regional Development Fund (ERDF), the Microbial Bioprocess Lab (MiBioLab), part of the Enabling Spaces Program "Helmholtz Innovation Labs" as well as the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for financial support. We thank Prof. Dr. Karl-Erich Jaeger from the Institute of Molecular Enzyme Technology (Heinrich-Heine University Düsseldorf) for his advice to use the PE-H enzyme and for provision of the gene. We thank Prof. Nick Wierckx from the Institute of Bio- and Geosciences, IBG-1: Biotechnology (Forschungszentrum Jülich) for providing the LCC gene. This study was funded by the European Regional Development Fund (grant 34. EFRE-0300097, CLIB-Kompetenzzentrum Biotechnologie [CKB]) and "Microbial Bioprocess Lab—A Helmholtz Innovation Lab," part of the Enabling Spaces Program "Helmholtz Innovation Labs" of the German Helmholtz Association. Open access publication was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—491111487. This work was performed as part of the Helmholtz School for Data Science in Life, Earth and Energy (HDS-LEE). Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at https://github.com/JuBiotech/petase-ts-paper.

ORCID

Laura M. Helleckes http://orcid.org/0000-0001-7825-7998
Carolin Müller http://orcid.org/0000-0002-6277-1009
Tim Griesbach http://orcid.org/0000-0003-0548-0667
Vera Waffenschmidt http://orcid.org/0000-0002-7737-1067
Michael Osthege http://orcid.org/0000-0002-2734-7624
Wolfgang Wiechert http://orcid.org/0000-0001-8501-0694
Marco Oldiges http://orcid.org/0000-0003-0704-5597

REFERENCES

Agrawal, S., & Goyal, N. (2012). Analysis of thompson sampling for the multi-armed bandit problem. In *Conference on Learning Theory* (pp. 39.1–39.6).

Bollinger, A., Thies, S., Knieps-Grünhagen, E., Gertzen, C., Kobus, S., Höppner, A., Ferrer, M., Gohlke, H., Smits, S. H. J., & Jaeger, K.-E. (2020). A novel polyester hydrolase from the marine bacterium *Pseudomonas aestusnigri*—Structural and functional insights. *Frontiers of Microbiology*, *11*, 114. https://doi.org/10.3389/fmicb.2020.00114

Brockmeier, U., Caspers, M., Freudl, R., Jockwer, A., Noll, T., & Eggert, T. (2006). Systematic screening of all signal peptides from *Bacillus subtilis*: A powerful strategy in optimizing heterologous protein

secretion in gram-positive bacteria. *J Mol Biol*, *362*(3), 393–402. https://doi.org/10.1016/j.jmb.2006.07.034

Carbonell, P., Jervis, A. J., Robinson, C. J., Yan, C., Dunstan, M., Swainston, N., Vinaixa, M., Hollywood, K. A., Currin, A., Rattray, N. J., Taylor, S., Spiess, R., Sung, R., Williams, A. R., Fellows, D., Stanford, N. J., Mulherin, P., Le Feuvre, R., Barran, P., … Scrutton, M.S. (2018). An automated Design-Build-Test-Learn pipeline for enhanced microbial production of fine chemicals. *Communications Biology*, *1*(1), 1–10. https://doi.org/10.1038/s42003-018-0076-9

Chapelle, O., & Li, L. (2011). An empirical evaluation of thompson sampling. *Advances in Neural Information Processing Systems*, *24*.

Degering, C., Eggert, T., Puls, M., Bongaerts, J., Evers, S., Maurer, K.-H., & Jaeger, K.-E. (2010). Optimization of protease secretion in *Bacillus subtilis* and *Bacillus licheniformis* by screening of homologous and heterologous signal peptides. *Applied and Environmental Microbiology*, *76*(19), 6370–6376. https://doi.org/10.1128/aem.01146-10

Freudl, R. (2017). Beyond amino acids: Use of the *Corynebacterium glutamicum* cell factory for the secretion of heterologous proteins. *Journal of Biotechnol*, *258*, 101–109. https://doi.org/10.1016/j.jbiotec.2017.02.023

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429258411

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Lerkwijk, M. H., Brett, M., Haldane, A., Rio, J. F., Weibe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Helleckes, L., & Osthege, M. (2022). *JuBiotech/petase-ts-paper: 2022-07-26* (Version 2022-07-26). Zenodo. https://doi.org/10.5281/zenodo.6908129

Helleckes, L. M., Osthege, M., Wiechert, W., vonLieres, E., & Oldiges, M. (2022). Bayesian calibration, process modeling and uncertainty quantification in biotechnology. *PLoS Computational Biology*, *18*(3), e1009223. https://doi.org/10.1371/journal.pcbi.1009223

Hemmerich, J., Rohe, P., Kleine, B., Jurischka, S., Wiechert, W., Freudl, R., & Oldiges, M. (2016). Use of a sec signal peptide library from *Bacillus subtilis* for the optimization of cutinase secretion in *Corynebacterium glutamicum*. *Microbial Cell Factories*, *15*(1), 208. https://doi.org/10.1186/s12934-016-0604-6

Hemmerich, J., Tenhaef, N., Steffens, C., Kappelmann, J., Weiske, M., Reich, S. J., Wiechert, W., Oldiges, M., & Noack, S. (2018). Less sacrifice, more insight: Repeated low-volume sampling of microbioreactor cultivations enables accelerated deep phenotyping of microbial strain libraries. *Biotechnol Journal*, *14*(9), 1800428.

Hernández-Lobato, J. M., Requeima, J., Pyzer-Knapp, E. O., & Aspuru-Guzik, A. (2017). Parallel and distributed thompson sampling for large-scale accelerated exploration of chemical space. In *International conference on machine learning* (pp. 1470–1479). PMLR.

Hill, D. N., Nassif, H., Liu, Y., Iyer, A., & Vishwanathan, S. (2017). An efficient bandit algorithm for realtime multivariate optimization. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1813–1821).

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*(1), 1593–1623.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/mcse.2007.55

Kandasamy, K., Dasarathy, G., Schneider, J., & Póczos, B. (2017). Multi-fidelity Bayesian optimisation with continuous approximations. In International conference on machine learning (pp. 1799–1808).

Kawai, F., Kawabata, T., & Oda, M. (2019). Current knowledge on enzymatic PET degradation and its possible application to waste stream management and other fields. *Applied Microbiology and Biotechnology*, *103*(11), 4253–4268. https://doi.org/10.1007/s00253-019-09717-y

Kinoshita, S., Udaka, S., & Shimono, M. (1957). Studies on the amino acid fermentation. *Journal of General and Applied Microbiology*, *3*(3), 193–205. https://doi.org/10.2323/jgam.3.193

Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. A. (2019). *ArviZ a unified library for exploratory analysis of Bayesian models in Python*. https://doi.org/10.21105/joss.01143

Liu, X.-X., Li, Y., & Bai, Z.-H. (2021). *Corynebacterium glutamicum* as a robust microbial factory for production of value-added proteins and small molecules: fundamentals and applications. In *Microbial cell factories engineering for production of biomolecules* (pp. 235–263). Elsevier. https://doi.org/10.1016/b978-0-12-821477-0.00006-4

McKinney, W. (2010). Data structures for statistical computing in Python, *Proceedings of the 9th python in science conference* (Vol. 445, pp. 51–56). https://doi.org/10.25080/majora-92bf1922-00a

Miller, J. (1972). *Experiments in molecular genetics*. Cold Spring Harbor Laboratory.

Müller, C., Bakkes, P. J., Lenz, P., Waffenschmidt, V., Helleckes, L. M., Jaeger, K.-E., Wiechert, W., Knapp, A., Freudl, R., & Oldiges, M. (2022). Accelerated strain construction and characterization of *C. glutamicum* protein secretion by laboratory automation. *Applied Microbiology and Biotechnology*, *106*, 4481–4497. https://doi.org/10.1007/s00253-022-12017-7

Neubauer, P., Cruz, N., Glauche, F., Junne, S., Knepper, A., & Raven, M. (2013). Consistent development of bioprocesses from microliter cultures to the industrial scale. *Engineering in Life Sciences*, *13*(3), 224–238. https://doi.org/10.1002/elsc.201200021

Osthege, M., & Felton, K. (2020). *Michaelosthege/pyrff: V2.0.1* (Version v2.0.1). Zenodo. https://doi.org/10.5281/zenodo.4317685

Osthege, M., & Helleckes, L. (2021). *Jubiotech/robotools: V1.3.0* (Version v1.3.0). https://doi.org/10.5281/zenodo.5745938

Osthege, M., & Helleckes, L. (2022). *Jubiotech/calibr8: Toolbox for non-linear calibration modeling*. https://doi.org/10.5281/zenodo.5961920

Osthege, M., Tenhaef, N., Helleckes, L., & Müller, C. (2022). *Jubiotech/bletl:V1.1.0* (Version v1.1.0). Zenodo. https://doi.org/10.5281/zenodo.6284777

Osthege, M., Tenhaef, N., Zyla, R., Müller, C., Hemmerich, J., Wiechert, W., Noack, S., & Oldiges, M. (2022). Bletl—A python package for integrating BioLector microcultivation devices in the Design-Build-Test-Learn cycle. *EngLife Science*, 22, 242–259. https://doi.org/10.1002/elsc.202100108

PlasticsEurope. (2022). *Plastics—The facts 2021: An analysis of european plastics production, demand and waste data*. Retrieved March 14, 2022, from https://plasticseurope.org/knowledge-hub/plastics-the-facts-2021/

Pouvreau, B., Vanhercke, T., & Singh, S. (2018). From plant metabolic engineering to plant synthetic biology: The evolution of the design/build/test/learn cycle. *Plant Science*, *273*, 3–12. https://doi.org/10.1016/j.plantsci.2018.03.035

Reback, J., jbrockmendel, McKinney, W., den Bossche, J. V., Augspurger, T., Cloud, P., Hawkins, S., gfyoung, Roeschke, M., Sinhrks, A., Petersen, T., Tratner, J., She, C., Ayd, W., Hoefler, P., Naveh, S., Garcia, M., Schendel, J., Hayden, A., Saxton, D., Darbyshire, J., & Seabolddd, S. (2021). *Pandas-dev/pandas: Pandas 1.3.4* (Version v1.3.4). Zenodo. https://doi.org/10.5281/zenodo.5574486

Rohe, P., Venkanna, D., Kleine, B., Freudl, R., & Oldiges, M. (2012). An automated workflow for enhancing microbial bioprocess optimization on a novel microbioreactor platform. *Microbial Cell Factories*, *11*(1), 144. https://doi.org/10.1186/1475-2859-11-144

Russo, D. J., Van Roy, B., Kazerouni, A., Osband, I., & Wen, Z. (2018). A tutorial on thompson sampling. *Foundations and Trends® in Machine Learning*, *11*(1), 1–96. https://doi.org/10.1561/2200000070

Schmid, C. H., & Brown, E. N. (2000). Bayesian hierarchical models. *Methods in Enzymology*, *321*, 305–330. https://doi.org/10.1016/S0076-

Sulaiman, S., Yamato, S., Kanaya, E., Kim, J.-J., Koga, Y., Takano, K., & Kanaya, S. (2012). Isolation of a novel cutinase homolog with polyethylene terephthalate-degrading activity from leaf-branch compost by using a metagenomic approach. *Applied and Environmental Microbiology*, *78*(5), 1556–1562. https://doi.org/10.1128/aem.06725-11

Sulaiman, S., You, D.-J., Kanaya, E., Koga, Y., & Kanaya, S. (2014). Crystal structure and thermodynamic and kinetic stability of metagenome-derived LC-cutinase. *Biochemistry*, *53*(11), 1858–1869. https://doi.org/10.1021/bi401561p

Taniguchi, I., Yoshida, S., Hiraga, K., Miyamoto, K., Kimura, Y., & Oda, K. (2019). Biodegradation of PET: Current status and application aspects. *ACS Catalysis*, *9*(5), 4089–4105. https://doi.org/10.1021/acscatal.8b05171

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, *25*(3–4), 285–294. https://doi.org/10.2307/2332286

Tiso, T., Winter, B., Wei, R., Hee, J., de Witt, J., Wierckx, N., Quicker, P., Bornscheuer, U. T., Bardow, A., Nogales, J., & Blank, L. M. (2021). The metabolic potential of plastics as biotechnological carbon sources—Review and targets for the future. *Metabolic Engineering*, *71*, 77–98. https://doi.org/10.1016/j.ymben.2021.12.006

Tomkins, S., Liao, P., Klasnja, P., & Murphy, S. (2021). IntelligentPooling: Practical thompson sampling for mHealth. *Machine learning*, *110*(9), 2685–2727. https://doi.org/10.1007/s10994-021-05995-8

Tournier, V., Topham, C. M., Gilles, A., David, B., Folgoas, C., Moya-Leclair, E., Kamionka, E., Desrousseaux, M.-L., Texier, H., Gavalda, S., Cot, M., Guémard, E., Dalibey, M., Nomme, J., Cioci, G., Barbe, S., Chateau, M., André, I., Duquesne, S., & Marty, A. (2020). An engineered PET depolymerase to break down and recycle plastic bottles. *Nature*, *580*(7802), 216–219. https://doi.org/10.1038/s41586-020-2149-4

Unthan, S., Grünberger, A., van Ooyen, J., Gätgens, J., Heinrich, J., Paczia, N., Wiechert, W., Kohlheyer, D., & Noack, S. (2014). Beyond growth rate 0.6: What drives *Corynebacterium glutamicum* to higher growth rates in defined medium. *Biotechnology Bioengineering*, *111*(2), 359–371. https://doi.org/10.1002/bit.25103

van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., & Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, *1*(1), 1–26. https://doi.org/10.1038/s43586-020-00001-2

van der Rest, M. E., Lange, C., & Molenaar, D. (1999). A heat shock following electroporation induces highly efficient transformation of *Corynebacterium glutamicum* with xenogeneic plasmid DNA. *Appl Microbiol Biotechnol*, *52*(4), 541–545. https://doi.org/10.1007/s002530051557

Vertès, A. A. (2013). Protein secretion systems of *Corynebacterium glutamicum*. In H. Yukawa, & M. Inui (Eds.), *Corynebacterium glutamicum: Biology and biotechnology* (pp. 351–389) Springer. https://doi.org/10.1007/978-3-642-29857-8_13

Vertommen, M. A. M. E., Nierstrasz, V. A., van der Veer, M., & Warmoeskerken, M. M. C. G. (2005). Enzymatic surface modification of poly(ethylene terephthalate). *Journal of Biotechnology*, *120*(4), 376–386. https://doi.org/10.1016/j.jbiotec.2005.06.015

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … van Mulbregt, P. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wells, J. A., Vasser, M., & Powers, D. B. (1985). Cassette mutagenesis: An efficient method for generation of multiple mutations at defined sites. *Gene*, *34*(2-3), 315–323. https://doi.org/10.1016/0378-1119(85)90140-4

Wiecki, T., Salvatier, J., Patil, A., Kochurov, M., Engels, B., Lao, J., Martin, C. O., Vieira, R., Willard, B. T., & Osthege, M. (2014). *Pymc-devs/pymc: v4.0.0b2*. (Version v4.0.0b2). https://doi.org/10.5281/zenodo.5850149

Winkler, U. K., & Stuckmann, M. (1979). Glycogen, hyaluronate, and some other polysaccharides greatly enhance the formation of exolipase by *Serratia marcescens*. *Journal of Bacteriology*, *138*3, 663–670. https://doi.org/10.1128/jb.138.3.663-670.1979

Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., Toyohara, K., Miyamoto, K., Kimura, Y., & Oda, K. (2016). A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science*, *351*(6278), 1196–1199. https://doi.org/10.1126/science.aad6359

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Helleckes, L. M., Müller, C., Griesbach, T., Waffenschmidt, V., Moch, M., Osthege, M., Wiechert, W., & Oldiges, M. (2023). Explore or exploit? A model-based screening strategy for PETase secretion by *Corynebacterium glutamicum*. *Biotechnology and Bioengineering*, 120, 139–153. https://doi.org/10.1002/bit.28261