


## ARTICLE OPEN ACCESS

# Optimized Machine Learning for Autonomous Enzymatic Reaction Intensification in a Self-Driving Lab

Sebastian Putz<sup>1</sup> | Niklas Teetz<sup>2</sup> | Michael Abt<sup>2</sup> | Pascal Jerono<sup>3</sup> | Thomas Meurer<sup>3</sup> | Matthias Franzreb<sup>1</sup> 

<sup>1</sup>Department for Bioengineering and Biosystems, Karlsruhe Institute of Technology (KIT), Institute of Functional Interfaces (IFG), Eggenstein-Leopoldshafen, Germany | <sup>2</sup>Department for Electrobiotechnology, Karlsruhe Institute of Technology (KIT), Institute of Process Engineering in Life Sciences (BLT), Karlsruhe, Germany | <sup>3</sup>Department for Digital Process Engineering, Karlsruhe Institute of Technology (KIT), Institute of Mechanical Process Engineering (MVM), Karlsruhe, Germany

**Correspondence:** Matthias Franzreb ([matthias.franzreb@kit.edu](mailto:matthias.franzreb@kit.edu))

**Received:** 15 April 2025 | **Revised:** 25 June 2025 | **Accepted:** 27 July 2025

**Funding:** This study was funded from the KIT via the subproject BioCAR in the project Auto.MAP and the Helmholtz Program “Materials Systems Engineering” under Program No. 43.31.01.

**Keywords:** automation | autonomous | digitalization | enzymes | machine learning | optimization algorithm | self-driving labs

## ABSTRACT

Optimizing enzymatic catalysis is crucial for enhancing the efficiency and scalability of many bioprocesses such as biotransformations, pharmaceutical synthesis, and food processing, as well as for improving the performance of analytical applications, including assays and biosensors. However, optimizing these reactions is challenging due to the multitude of interacting parameters such as pH, temperature, and cosubstrate concentration that require precise adjustment for maximum enzyme activity. Current optimization methods are often labor-intensive and time-consuming, especially when accounting for complex parameter interactions in highly dimensional parameter spaces. To overcome these challenges, we present a machine learning-driven laboratory platform that enables rapid, data-informed optimization of enzymatic reaction conditions in a fully automated environment. By conducting over 10,000 simulated optimization campaigns on a surrogate model generated via linear interpolation of experimentally obtained data, we identified and fine-tuned the most efficient machine learning algorithm for optimizing enzymatic reactions. This allows the platform to autonomously determine optimal reaction conditions with minimal experimental effort and without human intervention. The effectiveness of our approach is demonstrated by the accelerated optimization of reaction conditions in a five-dimensional design space across multiple enzyme-substrate pairings. In conclusion, our self-driving lab platform, equipped with a tailored optimization algorithm, offers a novel and superior alternative to traditional optimization methods. Moreover, the methodology for selecting the most efficient problem-specific optimization algorithm can be extended to self-driving lab platforms with broader applications.

## 1 | Introduction

Enzymatic reactions are fundamental to various industries, including fine chemical production (Hara et al. 2014; Liese and Vilella Filho 1999; Thompson et al. 2019; Carrea and Riva 2000), pharmaceutical synthesis (Wu et al. 2021; Reetz et al. 2024; Kim et al. 2024; Simić et al. 2022; Chapman et al. 2018), food processing (Fernandes 2010; Yang et al. 2023;

Wang et al. 2024), and analytical applications such as assays (Bisswanger 2014; Gan and Patel 2013; Asensio et al. 2008) and biosensors (Katz et al. 2001; Kilic et al. 2023; Melo et al. 2024) amongst others. These biocatalysts offer unique advantages over inorganic catalysts, such as high activities, reaction selectivity, substrate specificity, and stereo-selectivity under mild reaction conditions (Rao et al. 2024). In addition, enzymatic processes are considered sustainable and environmentally friendly

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Biotechnology and Bioengineering* published by Wiley Periodicals LLC

(Jegannathan and Nielsen 2013). For instance, in pharmaceutical synthesis, enzyme cascades facilitate the production of complex molecules with high selectivity, often surpassing traditional chemical methods in both efficiency and environmental compatibility (Siedentop and Rosenthal 2022).

However, optimizing enzymatic reactions is a complex endeavor due to the multitude of interacting parameters—such as pH, temperature, and cosubstrate concentration—that must be precisely adjusted to achieve maximum enzyme activity. This high-dimensional design space, coupled with intricate parameter interactions, renders traditional optimization methods labor-intensive and time-consuming. For example, in enzyme cascade reactions, different optimal conditions for each enzyme involved can lead to challenges like side reactions and reduced overall efficiency (Siedentop et al. 2021; Onyeogaziri and Papanephytous 2019).

To overcome these challenges in complex optimization problems, the concept of a self-driving laboratory (SDL) has emerged. SDLs integrate automation with artificial intelligence (AI) to enhance experimental planning and execution (Canty et al. 2023; Abolhasani and Kumacheva 2023; Häse et al. 2019). SDLs can autonomously conduct experiments, analyze data, and iteratively refine conditions, thereby expediting the optimization process and reducing human intervention. This approach not only accelerates discovery but also improves reproducibility and efficiency in laboratory workflows (Bennett and Abolhasani 2022; Martin et al. 2023; Rooney et al. 2022; Soldatov et al. 2021). SDLs have already proven to be extremely effective tools for the accelerated discovery, optimization, and synthesis in the field of complex organic compounds (Steiner et al. 2019; Coley et al. 2019; Granda et al. 2018; Christensen et al. 2021; Ha et al. 2023), nanomaterials (Epps et al. 2020; Abdel-Latif et al. 2021; Salley et al. 2020; Li et al. 2020a, 2020b; Volk et al. 2021), thin films (Harris et al. 2024; Nikolaev et al. 2014; MacLeod et al. 2022, 2020) as well as biomolecules and biosystems (King et al. 2004; Williams et al. 2015; Si et al. 2017; Hamedirad et al. 2019; Kanda et al. 2022; Notin et al. 2024; Rapp et al. 2024), amongst others. Data-driven optimization algorithms used in SDLs have already outperformed human scientists in certain applications such as chemical reaction optimization (Shields et al. 2021). However, to enable SDLs to effectively accelerate experimental optimization and discovery, selecting the appropriate optimization algorithm is essential (Hickman et al. 2023a, 2023b; MacLeod et al. 2023).

In this study, we present an autonomous, machine learning-driven platform designed to rapidly optimize enzymatic reaction conditions within a fully integrated, self-driving laboratory environment. Additionally, we introduce a method to identify efficient optimization algorithms to fine-tune SDLs for specific tasks. This method includes an initial high-throughput screening to generate an exemplary data set for the intended application of the SDL, which is then used for the in-silico evaluation and optimization of multiple machine learning algorithms. For validation, the optimized machine learning algorithm is subsequently employed for optimization experiments on the SDL platform. After successful validation, the optimized algorithm can prospectively be used on the SDL to

solve the intended task in the most efficient way. We exemplarily carried out these steps, for an SDL tailored to enzymatic reaction condition optimization. The results demonstrate that the optimized algorithm, Bayesian Optimization (BO) with a specific kernel and acquisition function, is highly generalizable across various enzyme-substrate pairings. Using the fine-tuned BO, enzymatic reaction conditions of multiple enzyme-substrate pairings could be identified robustly and significantly accelerated compared to traditional methods.

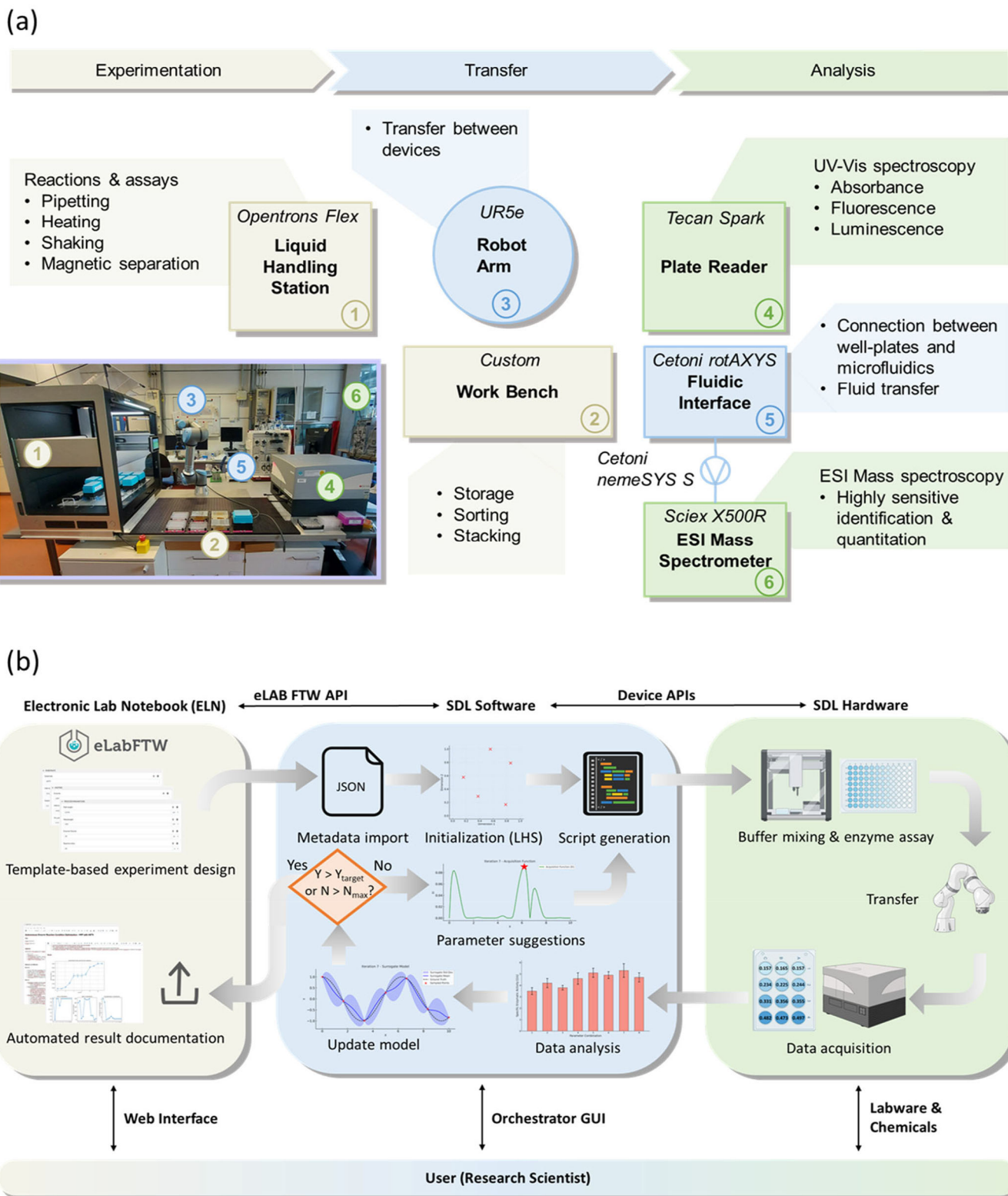
## 2 | Results and Discussion

### 2.1 | Self-Driving Lab Platform and Workflow Overview

We developed a self-driving lab platform tailored for the autonomous optimization of biochemical reactions and analyses. Building upon our previous SDL work in solid-phase extraction processes (Putz et al. 2024), this platform represents our first application to biocatalysis and bioanalytics, featuring substantial hardware and software advances. The first steps included setting up a hardware and software framework which allows for automated enzymatic reactions. The SDL currently integrates six laboratory devices and a custom-built workbench, as depicted schematically and by a photograph in Figure 1a. At the core of the platform is a liquid handling station (OT Flex, Opentrons, USA), which provides pipetting, heating, shaking, gripper, and magnetic separation functionalities, enabling a wide range of (bio-)chemical reactions and assays in well-plate format. The custom-designed workbench offers organized storage, sorting, and stacking for well-plates, pipette tips, and liquid reservoirs. A 6-DOF (six degrees of freedom) robotic arm (UR5e, Universal Robots, Denmark) equipped with an adaptive gripper (Hand-E, Robotiq, Canada) and custom 3D-printed fingers facilitates the automated transport and arrangement of labware and chemicals. Spectroscopic analysis is carried out using a multimode plate reader (Spark, Tecan, Switzerland), which supports UV-vis spectroscopy, fluorescence, and luminescence measurements.

Additionally, the platform's well-plate-based system connects to a flow-based microcapillary setup through a capillary positioning module (RotAXYS, Cetoni, Germany). Precision fluid transport is managed by high-accuracy syringe pumps (nemeSYS S, Cetoni, Germany) and a flow-selection valve module (6-port Qmix V Valve, Cetoni, Germany), while integration with an electrospray-ionization mass spectrometer (ESI-MS) (X500-R, Sciex, USA) coupled to an UPLC (Exion AD, Sciex, USA) enables highly sensitive detection and characterization of (bio-)analytes. In this study, the platform primarily utilized the liquid handling station, robotic arm, and plate reader for automated colorimetric enzymatic assays. The modular architecture of the SDL, however, ensures flexibility and expandability, enabling the addition of specialized devices such as microreactors or alternative measurement modules to meet diverse experimental needs and further enhance its capabilities.

The integration of diverse commercial platforms into our Python-based framework presented varying technical effort depending on vendor API support. Platforms with native



**FIGURE 1** | (a) Hardware schematic of the self-driving lab platform for biochemical reaction optimization with photograph as inset. (b) Flowchart of the autonomous optimization workflow exemplified for the optimization of enzymatic reaction conditions.

Python APIs (e.g., Opentrons, Universal Robots) enabled direct integration, while proprietary systems utilizing C# interfaces (e.g., Tecan Spark SiLA Server, Sciex OS Control API) required the development of custom Python wrappers. These wrapper implementations typically required 2–3 days of development per device interface. The prevalence of Python APIs in

laboratory instrumentation and the robust scientific computing ecosystem continue to support Python as the preferred framework for SDL backend development.

For a detailed explanation of the software architecture and APIs, we want to refer to our previous work (Putz et al. 2024)

and the *Supporting Information* (Figure S2). In brief, it is a Python-based modular framework, enabling high flexibility and easy integration of new devices and software modules. Key software advancements include the development of a Python wrapper for the Sciex OS API enabling automated analytical instrument control, and comprehensive end-to-end Electronic Laboratory Notebook (ELN) integration with eLabFTW using the eLabFTW Python API (API 2024). The ELN integration enables a seamless workflow from experimental design in the web-based notebook through automated metadata import, experiment execution, and final upload of all data and auto-generated reports back to the ELN for permanent documentation. Users can interact with the self-driving lab using a custom-programmed GUI (Supporting Information S1: Figure S1).

The workflow (Figure 1b) starts in the web interface of eLabFTW, where the user designs the experiment based on a predefined template. Each specific process, i.e., type of experiment, has its own template, where relevant metadata is entered into a structured input mask. Using the control software of the SDL, this metadata is imported to the local machine via the eLabFTW API. Following, the experiment is initialized: For autonomous iterative optimization processes an initial set of parameter combinations is generated using Latin Hypercube Sampling (LHS). For nonautonomous experiments, the parameter combinations are generated according to the experimental design specified in the metadata. Subsequently, the software generates the protocols to control the laboratory devices, namely the liquid handling station, the plate reader, and the ESI-MS by embedding the relevant parameters and metadata into experiment-specific templates. These templates are executed as part of a workflow script that orchestrates the operation of all connected laboratory devices during each experimental cycle.

For instance, in the optimization of enzymatic assays, the workflow starts with the preparation of buffers and assay mixtures by the liquid handler. Then the robot arm transfers the well-plate into the plate reader for absorbance measurements. After the measurement, the robot arm either returns the plate back into the liquid handler or moves it to the storage bench if its fully used. The robot arm further prepares the setup for the next experimental cycle by replacing labware and reagents as needed. At the end of each experimental cycle, raw data is analyzed to perform calculations and generate plots. In the case of autonomous experiments, this data is used by the machine learning algorithm for the suggestion of the process parameter combinations for the next experimental cycle. The employed algorithms are either model-based, such as BO, or model-free and evolutionary like a Genetic Algorithm (GA). Using the parameter suggestions from the algorithm, new device control protocols are generated and the next experimental cycle begins.

In certain cases, human intervention is required after several cycles—for example, when labile compounds need special handling outside the SDL platform, or when the labware in the storage is depleted. Once the optimization process is complete, that is, the convergence criterion of the algorithm is met, a predefined target is reached ( $Y > Y_{\text{target}}$ ) or a specified number of experimental cycles is executed ( $N > N_{\text{max}}$ ), a final data

evaluation is performed. This comprehensive, experiment-type-specific evaluation generates plots using data from all experimental cycles, highlighting improvements in the objective function and the influence of individual parameters. Finally, raw and processed data, including plots, are uploaded to the ELN using the API. The experimental documentation is completed by generating and uploading a written report, where placeholders in a predefined template are automatically replaced with the actual metadata and results.

## 2.2 | Evaluating Machine Learning Algorithms for Efficient Enzymatic Reaction Optimization

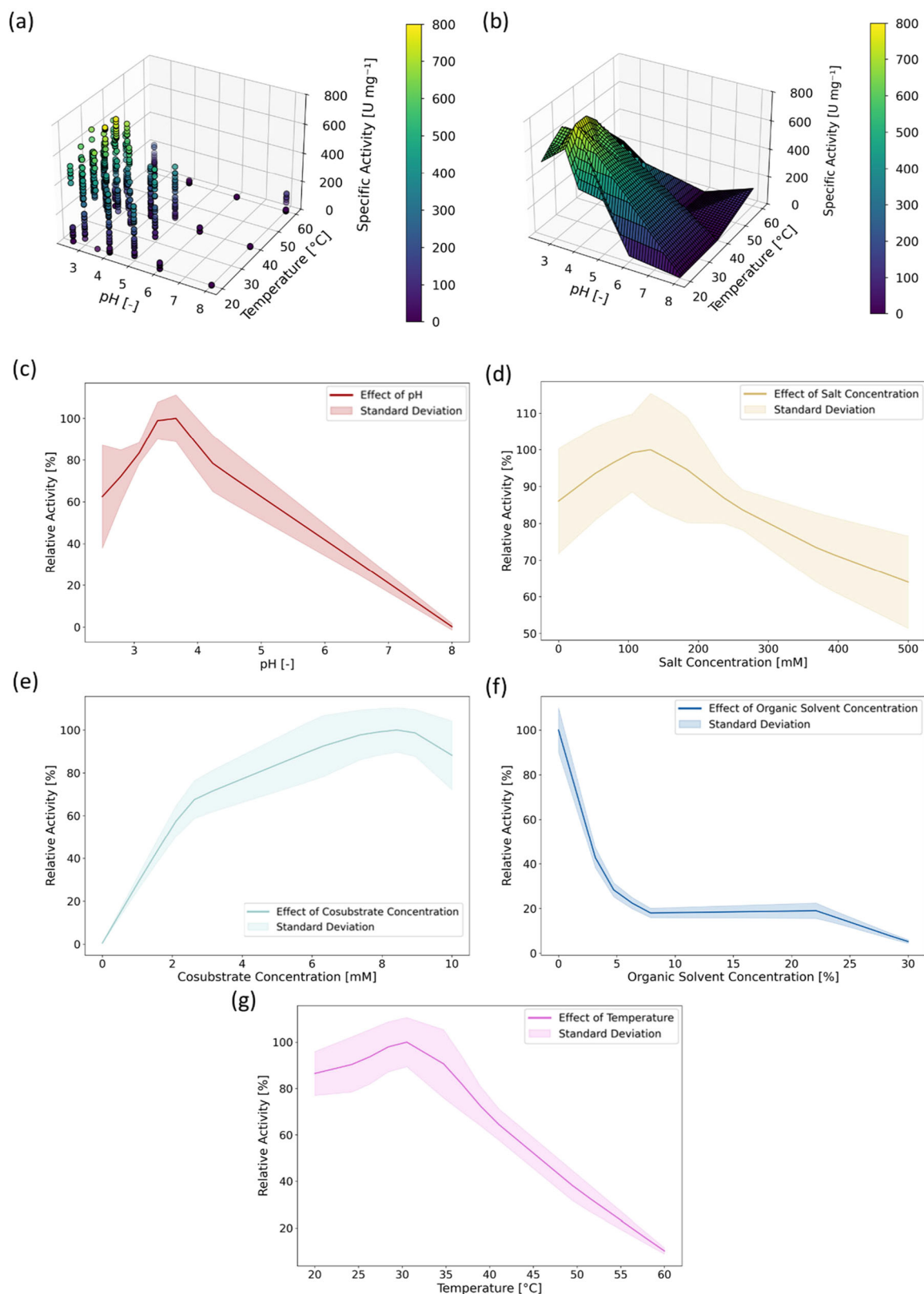
Selecting an appropriate machine learning algorithm is critical for efficiently optimizing enzymatic reaction conditions. To ensure robust and rapid optimization, multiple machine learning approaches were systematically evaluated, identifying the most effective algorithm for guiding our self-driving lab. To thoroughly assess different machine learning algorithms, first a comprehensive data set using high-throughput screening was created, capturing a wide range of enzymatic reaction conditions for an exemplary reaction. Next, in-silico experiments were conducted, simulating thousands of optimization campaigns to evaluate each algorithm's performance in navigating the complex parameter space.

### 2.2.1 | High-Throughput Screening (HTPS) of Initial Enzymatic Reaction Rates

To provide a comprehensive data set for testing in-silico different optimization algorithms in regard of their efficiency for process optimization of enzymatic reactions, a high-throughput screening of a typical model reaction was conducted. The initial reaction rates of an unspecific peroxygenase AaeUPO-PaDa-I with a Twin-Strep-Tag (Twin-Strep-UPO) with the substrate 2,2-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) were screened under variation of the five key parameters pH, temperature, salt concentration, organic solvent concentration and cosubstrate concentration. Sodium sulfate ( $\text{Na}_2\text{SO}_4$ ) was used as additional salt, acetonitrile (ACN) as additional organic solvent and hydrogen peroxide ( $\text{H}_2\text{O}_2$ ) as cosubstrate.

In Figure 2a, the mean values of the initial reaction rates are shown over pH and temperature for all investigated parameter combinations. The initial reaction rate presents a maximum at  $\text{pH} = 3.5$ ,  $T = 30^\circ\text{C}$ , with cosubstrate concentration  $c_{\text{H}_2\text{O}_2} = 8.75 \text{ mM}$ , salt concentration  $c_{\text{Na}_2\text{SO}_4} = 120 \text{ mM}$ , and organic solvent concentration  $c_{\text{ACN}} = 0\% \text{ (v/v)}$ . The peak activity at the maximum is  $790.4 \pm 80.6 \text{ U mg}^{-1}$ . Due to the iterative Design of Experiments (DoE)-based sampling, which employed three sequential 3-Level-Full-Factorial designs, the regions closer to the optimal conditions are sampled more densely than those farther away.

To ensure all possible parameter combinations within the design space were computationally accessible from the finite set of recorded data points, linear interpolation was applied for the in-silico experiments (see Section 2.3). The resulting interpolated response surface as a function of pH and temperature is



**FIGURE 2** | (a) Specific enzymatic activities of the AaeUPO PaDa-I with Twin-Strep-Tag as a function of pH and temperature from the high throughput screening under variation of pH, temperature, cosubstrate ( $\text{H}_2\text{O}_2$ ), salt ( $\text{Na}_2\text{SO}_4$ ) and organic solvent (ACN) concentration. The values are the mean of replicated experiments ( $n = 2$ ). (b) Response surface of the enzymatic activity as a function of pH and temperature obtained through linear interpolation of mean experimental data points. (c–g) One dimensional projection of the interpolated response surface for varying pH (c), salt concentration (d), cosubstrate concentration (e), organic solvent concentration (f) and temperature (g) while keeping all other parameters fixed at the optimal values. The fixed values are  $\text{pH} = 3.5$ ,  $T = 30^\circ\text{C}$ , with cosubstrate concentration  $c_{\text{H}_2\text{O}_2} = 8.75$  mM, salt concentration  $c_{\text{Na}_2\text{SO}_4} = 120$  mM, and organic solvent concentration  $c_{\text{ACN}} = 0\%$  (v/v).

illustrated in Figure 2b. A critical discussion of the interpolation model's confidence, associated uncertainty, and a demonstration of its suitability for algorithm benchmarking is provided in the *Supporting Information, Section 4. Linear interpolation surrogate model confidence and uncertainty quantification*.

To assess the influence of individual parameters, one dimensional projections of the interpolated response surface were generated (Figure 2c–g). These show the influence of the individual parameters while keeping the other four parameters constant at their optimal values. The initial enzymatic reaction rate is highest at pH = 3.5 and decreases to 61.1% at pH = 2.4% and 2.0% at pH = 8, the extremes of the design space (Figure 2c). The pH dependency of enzymatic activity typically exhibits a bell-shaped curve, with a steeper decline in activity as the pH increases beyond its optimal value. The approximately linear relationship between pH = 4 and pH = 8 observed in Figure 2c is a result of interpolation in sparsely populated regions of the 5-dimensional design space. Similarly, the temperature exhibits a single optimum at T = 30°C and the relative activity declines to 85.5% at T = 20°C and 10.2% at T = 60°C (Figure 2g). Increasing the concentration of the cosubstrate H<sub>2</sub>O<sub>2</sub> from 0 to 8.5 mM continually increases the reaction rate (Figure 2e) from 0.3% to 100%. Increasing H<sub>2</sub>O<sub>2</sub> concentration further to 10 mM reduces the reaction rate to 88.1% of the peak value. At 0 mM H<sub>2</sub>O<sub>2</sub>, the reaction rate approaches zero, as no cosubstrate for the reaction is available. As shown in Figure 2f, the reaction rate is maximal at 0% (v/v) ACN and decreases to 4.8% when increasing the organic solvent concentration to 30% (v/v). On average, adding additional 125 mM Na<sub>2</sub>SO<sub>4</sub>, yields the highest activity (Figure 2d). The relative activity is slightly decreased to 87.2% when no Na<sub>2</sub>SO<sub>4</sub> is added to the buffer and more strongly decreased to 64.8% when a high concentration of 500 mM Na<sub>2</sub>SO<sub>4</sub> is added.

The identified optimal reaction conditions for the substrate ABTS align well with previous studies on recombinant AaeUPO enzymes, which typically report pH optima between 3.0 and 4.0 and temperature optima around 30°C (Molina-Espeja et al. 2014, 2015; Yan et al. 2024). In this study, the enzyme was not pre-incubated at the target temperature; instead, the initial activity was measured immediately after the enzyme was added to the reaction mixture. Interestingly, higher temperatures exceeding 30°C still led to reduced activity, likely due to the rapid thermal deactivation kinetics of the enzyme. UPOs typically show a decreasing activity in ABTS assay with increasing organic solvent concentration (Martin-Diaz et al. 2021), which could also be observed in the conducted screening. An H<sub>2</sub>O<sub>2</sub> concentration of 8.75 mM was found to be optimal to maximize the initial enzymatic reaction rate. This concentration is substantially higher than 2 mM which is typically used for ABTS assays with UPOs (Dolz et al. 2023; Bormann et al. 2015). While elevated H<sub>2</sub>O<sub>2</sub> levels enhance the reaction rate, they can simultaneously promote enzyme inactivation (Hofrichter et al. 2015). As this study focused solely on the initial reaction rate within a 60-second timeframe, 8.75 mM appears to strike the most favorable balance between enzyme inactivation and reaction kinetics. Additionally, the screening suggests adding 120 mM Na<sub>2</sub>SO to the phosphate-citrate buffer maximizes the initial enzymatic reaction rate. It can be hypothesized that additional salt in the reaction buffer could favorably alter the electrostatic enzyme-enzyme and enzyme-substrate

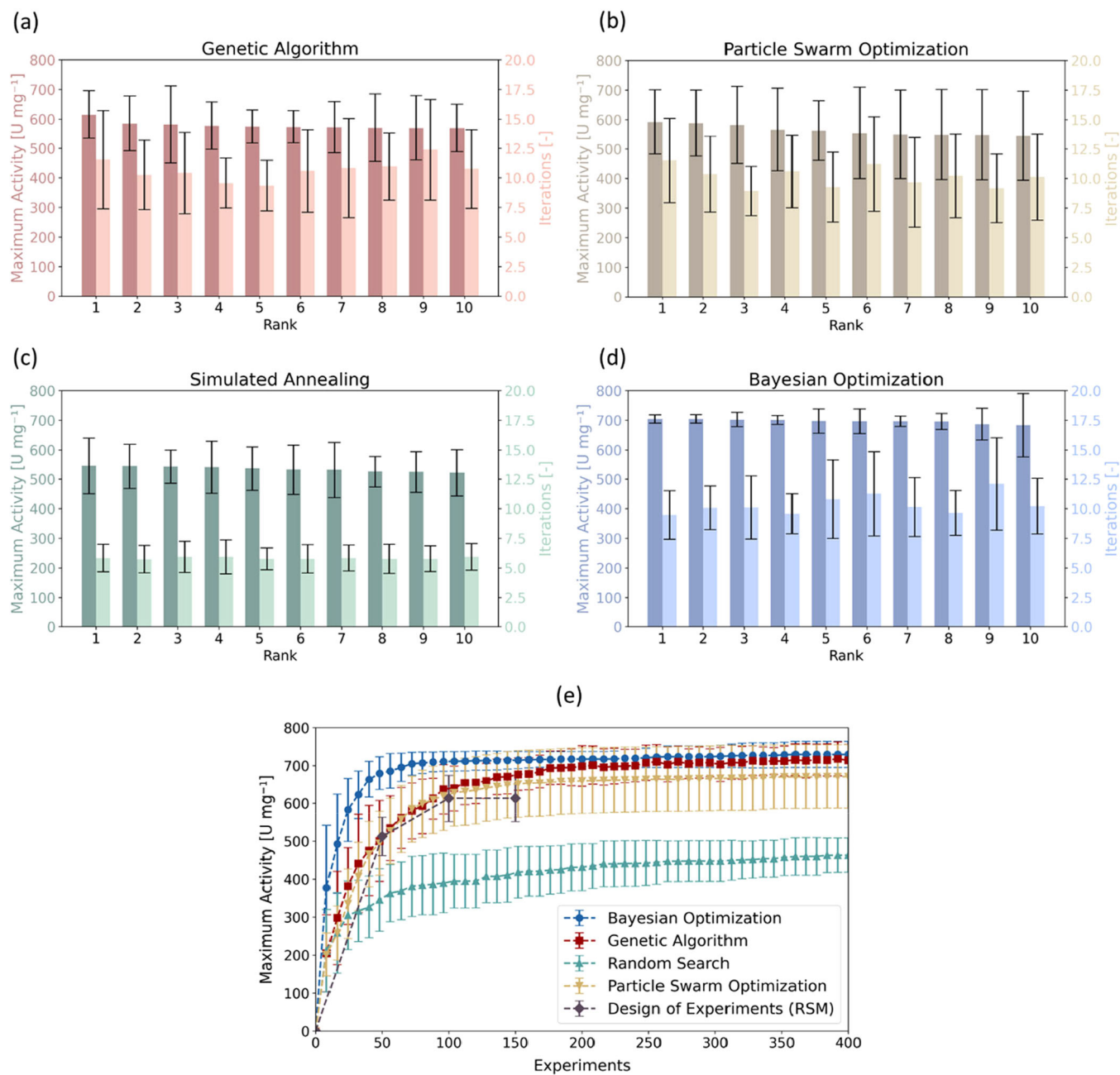
interactions to increase the enzymatic reaction rate. The average coefficient of variation (CV) for the ABTS assay was 13.2%, which is comparable to reported reference values of approximately 12% (Dolz et al. 2023).

## 2.2.2 | In-Silico Testing and Fine-Tuning of Optimization Algorithms

To identify the most effective algorithm for this optimization problem—maximizing the initial enzymatic reaction rate by varying reaction conditions—a screening and fine-tuning of several optimization algorithms was conducted. These algorithms were evaluated in simulated experimental optimization campaigns (see 4. *Experimental Section*), where both the number of iterations and the highest enzymatic activity achieved were tracked. Detailed descriptions of the utilized optimization algorithms, flow charts and explanations of the investigated hyperparameters are presented in the *Supporting Information*.

Figure 3a–d display the highest scores and corresponding iteration counts for the top 10 variants of the Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Simulated Annealing (SA), and Bayesian Optimization (BO). Among these, SA exhibited the lowest performance, achieving a maximum activity of 527.4 U mg<sup>-1</sup> within 5.83 iterations averaged over the top 10. PSO performed better, with a mean maximum activity of 563.8 U mg<sup>-1</sup> over an average of 10.11 iterations. GA ranked second, reaching an average best score of 578.6 U mg<sup>-1</sup> across 10.66 iterations. BO was the most effective algorithm, achieving a superior average maximum activity of 696.8 U mg<sup>-1</sup> in 10.33 iterations.

To gain insights into the hyperparameter tuning, in Tables 1–4, the hyperparameters of the top 5 performing algorithm variants are shown for the GA, PSO, BO and SA. For the GA, an efficient setup included migrating 1 to 2 top individuals into subsequent generations (referred to as elitism), a moderate mutation extent (0.3–0.6), random crossover during recombination, and rank-based selection. The mutation probability seems to have a weak effect, as all investigated values are present in the top 5 (Table 1). These settings balance exploration and exploitation, maintaining diversity while ensuring convergence. To construct an efficient PSO algorithm for the specific optimization problem, a high initial inertia weight  $w_0$  (0.9–1.1), a slow decay factor for the inertia  $c_d$  (0.85–0.95) and larger cognitive than social factor,  $c_1 \geq c_2$ , should be chosen (Table 2). This configuration emphasizes individual particle exploration, which seems beneficial in the investigated search space. BO performed best with a Matérn kernel and an acquisition function that is either EI, UCB or PI (Table 3). These choices facilitate a balance between exploring new areas and exploiting known promising regions. Briefly, the acquisition functions control how the algorithm navigates the trade-off between sampling regions with high predicted activity (exploitation) and exploring uncertain regions (exploration). The choice of kernel, such as Matérn, influences the flexibility and smoothness of the surrogate model, allowing it to capture varying levels of complexity in the underlying activity landscape (for detailed explanations, see *Supporting Information, Section 2. Utilized optimization algorithms*). For SA no clear trend could be observed, except



**FIGURE 3** | Comparison of optimization algorithms by simulated optimization campaigns. (a–d): Maximal enzymatic activity and number of iterations for the top 10 best performing algorithms for (a) Genetic Algorithm (GA), (b) Particle Swarm Optimization (PSO), (c) Simulated Annealing (SA) and (d) Bayesian Optimization (BO). All values represent the mean and standard deviation averaged over 30 optimization campaigns using the same 30 initial parameter combinations for all algorithms. (e) Comparison of the maximal activity for the best-performing BO, GA and PSO to Random Search (RS) and Response Surface Modeling (RSM) over 400 iterations. All values represent the mean and standard deviation averaged over 30 optimization campaigns using the same 30 initial parameter combinations for all algorithms. For RSM, the experimental design was fixed but the optimization was run 30 times, deviations arise from the noise in the simulated experiments.

that a large initial step size of 0.8 seems most effective for the optimization task (Table 4).

In Figure 3e the performances of the top GA, PSO, and BO algorithms are compared with Random Search (RS) and Response Surface Modeling (RSM) as benchmarks. Over 400 experiments (50 iterations), BO achieved the fastest convergence, reaching an average maximum activity of  $695.9 \pm 37.1$  U mg<sup>-1</sup> within 64 experiments (8 iterations). It plateaued at  $729.7 \pm 34.43$  U mg<sup>-1</sup> after

400 experiments. The best parameter combinations identified across all optimization runs are clustered closely around the single optimum in the parameter space with values of pH =  $3.5 \pm 0.0$ , T =  $20.9 \pm 2.5$ °C, c<sub>H<sub>2</sub>O<sub>2</sub></sub> =  $9.6 \pm 0.8$  mM, c<sub>Na<sub>2</sub>SO<sub>4</sub></sub> =  $211.7 \pm 132.3$  mM, and c<sub>ACN</sub> =  $0 \pm 0.0\%$  (v/v). In contrast, GA and PSO converged more slowly. After 64 experiments, their maximal activities were similar ( $560.9 \pm 80.0$  U mg<sup>-1</sup> for GA and  $558.1 \pm 88.2$  U mg<sup>-1</sup> for PSO), while both continued to improve gradually. GA eventually plateaued at  $714.0 \pm 42.4$ , while PSO

**TABLE 1** | Hyperparameters of the top five best performing genetic algorithm (GA) variants from the in-silico optimization algorithm testing.

Rank	Elitism	Mutation probability	Mutation extent	Cross-over	Selection
1	2	0.4	0.3	Random	rank
2	1	0.1	0.3	Random	rank
3	1	0.2	0.3	Random	rank
4	1	0.1	0.6	Random	rank
5	1	0.8	0.3	single point	rank

**TABLE 2** | Hyperparameters of the top 5 best performing particle swarm optimization (PSO) variants from the in-silico optimization algorithm testing.

Rank	$w_0$	$c_d$	$c_1$	$c_2$
1	0.9	0.95	2.5	1.0
2	0.9	0.95	1.5	1.0
3	1.1	0.95	1.5	1.5
4	0.9	0.95	2.0	1.0
5	1.1	0.85	2.5	1.5

**TABLE 3** | Hyperparameters of the top 5 best performing Bayesian optimization (BO) variants from the in-silico optimization algorithm testing.

Rank	Kernel Function	Acquisition function
1	Matérn 32	EI
2	Matérn 32	UCB
3	Matérn 12	EI
4	Matérn 12	UCB
5	Matérn 32	PI

**TABLE 4** | Hyperparameters of the top five best performing simulated annealing (SA) variants from the in-silico optimization algorithm testing.

Rank	$T_0$	$c_T$	$S_0$
1	380	0.7	0.8
2	150	0.6	0.8
3	150	0.7	0.8
4	380	0.9	0.8
5	75	0.5	0.8

stabilized at  $671.3 \pm 84.3$ . Thus, the GA and PSO can find optima in a similar range than the BO, but converge at a slower rate. Therefore, the lack-of-improvement convergence criterion (less than 5% improvement over 5 rounds) terminated the GA and PSO in the grid-search algorithm testing before reaching the maximum score. Additionally, the BO has the lowest standard deviation

averaged over the 30 initial generations indicating only little dependence of maximal activity on the initial conditions. In contrast, PSO has the largest standard deviation, indicating a stronger dependence on the initial parameter combinations.

In comparison, a 3-step Response Surface Modeling (RSM) identifies a maximal score of  $615.9 \pm 54.7$  in 150 experiments. Notably, the quadratic interaction models fitted at each round of the RSM procedure achieved coefficients of determination ( $R^2$ ) of  $0.762 \pm 0.005$  in the first round,  $0.826 \pm 0.007$  in the second round, and  $0.860 \pm 0.013$  in the final round. This progressive improvement in  $R^2$  demonstrates that as RSM iteratively narrows the search region, the local fit to the experimental data improves. However, the overall lower maximal activity achieved by RSM compared to model-free (GA, PSO, SA) or GP-based (BO) optimization algorithms highlights a key limitation: the quadratic model, despite good fits in the sampled region, does not fully capture the complex parameter dependencies present in the enzymatic system and can thus fail to efficiently guide the search toward the global optimum. Up to 100 experiments, the RSM performs comparably to the GA and PSO, which then outperform the RSM by converging to a higher maximal score at a slow rate. The BO plateaus at a similar number of experiments as the RSM, approximately 100, however reaching a higher maximal enzymatic activity. This can be explained by the rigid quadratic interaction model used in the RSM, which seems not to accurately cover the underlying relationship of the parameters on the enzymatic reaction rate. Unlike RSM, GA and PSO are model-free optimization methods that do not rely on fitting a predefined equation or model to the data. This allows these algorithms to explore the parameter space more flexibly. On the other hand, BO fits a model—specifically, a Gaussian process (GP)—that leverages all available data to predict the objective function. The GP model is highly flexible, capable of adapting to complex, nonlinear relationships in the parameter space. The comparison to RS shows that either method is more effective than random sampling within the design space. This demonstrates the importance of leveraging information-driven optimization techniques for effectively exploring and exploiting the design space.

Overall, these results indicate among all tested algorithms BO is most effective for optimizing enzymatic reaction conditions on the provided data set. This can be attributed to the inherent strength of the GP model, which provides a probabilistic framework for modeling the objective function. By quantifying uncertainty and incorporating prior knowledge, BO efficiently balances exploration and exploitation. The Gaussian process enables BO to focus experimental efforts on the most promising



regions of the parameter space, adapting dynamically as new data becomes available. This approach is particularly effective for capturing complex relationships in the enzymatic reaction rate, leading to faster convergence to an optimal solution compared to other algorithms. BO is among the most common experimental planning strategies for SDLs (Tom et al. 2024). It has proven to be a very effective optimization algorithm in SDLs for diverse applications including protein engineering (Rapp et al. 2024), adhesives (Rooney et al. 2022), nanoparticles (Vaddi et al. 2022), metal-organic frameworks (Xie et al. 2021), thin films (Ohkubo et al. 2021) and battery electrolytes (Dave et al. 2020).

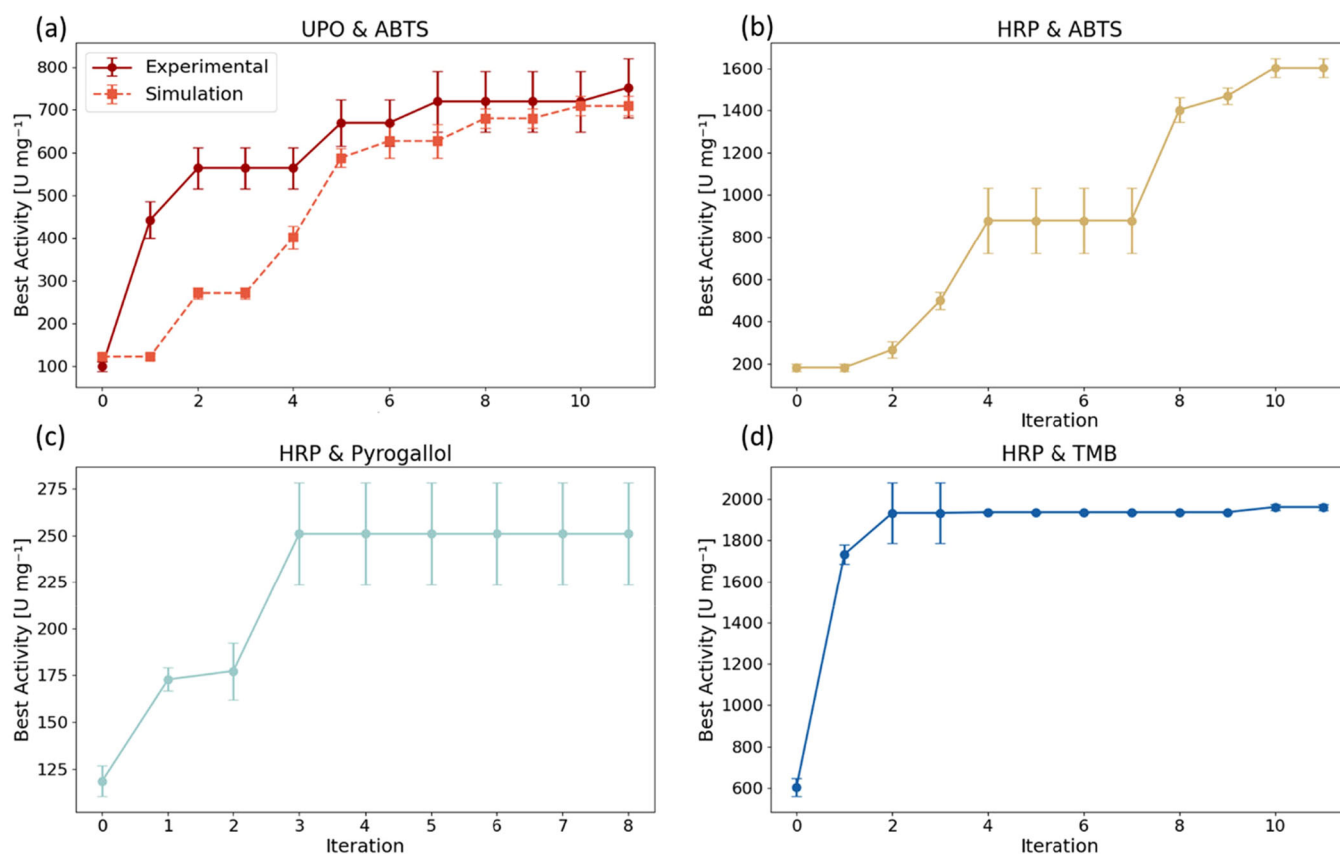
Given the similarity of optimization challenges across enzyme-substrate systems, we hypothesize that the fine-tuned BO algorithm can be broadly applied to optimize reaction conditions for various enzymes and substrates. To test this hypothesis, we conducted autonomous optimization campaigns on the SDL with different enzyme-substrate pairings (see Section 2.4).

### 2.3 | Autonomous Enzymatic Reaction Optimization With Fine-Tuned Algorithm

The best-performing algorithm from the in-silico testing, Bayesian Optimization (BO) with a Matérn 3/2 kernel and Expected Improvement (EI) acquisition function, was deployed on

the SDL to conduct real experiments for autonomously optimizing the reaction conditions of enzymes with different substrates. Each experimental campaign was initialized de novo without prior knowledge, using LHS to generate the initial parameter sets for each new enzyme-substrate combination. After each experimental cycle, the algorithm suggested eight new parameter combinations to test in the subsequent cycle.

To validate the results and compare simulated and experimental outcomes, the reaction conditions for the Twin-Strep-UPO with ABTS as the substrate were optimized first. Figure 4a displays the highest enzymatic reaction rate over the iterations for this UPO-ABTS pairing. In the experiment, the highest enzymatic activity increased from  $133.3 \pm 11.7 \text{ U mg}^{-1}$  in the initial cycle to  $708.2 \pm 69.7 \text{ U mg}^{-1}$  by iteration 11. For the simulated experiment with the same initial parameter combinations, the enzymatic activity increased from  $123.0 \pm 7.8$  to  $708.6 \pm 22.5 \text{ U mg}^{-1}$  during the optimization. Although in the experiments with the SDL the overall highest activity initially improved faster, experiment and simulation align closely from iteration 5 onwards. Comparing the optimal conditions identified by the SDL and the simulated experiments (Table 5), small differences can be found. While both identify an optimal pH of 3.4 and organic solvent concentration of 0% (v/v), the SDLs optimum has a lower temperature ( $20^\circ\text{C}$  vs.  $30^\circ\text{C}$  in the simulation), a lower salt concentration (0 mM vs. 125 mM) and a higher  $\text{H}_2\text{O}_2$



**FIGURE 4** | Autonomous optimization of enzymatic reaction conditions in real experiments on the SDL platform. (a): Overall highest enzymatic activity versus iterations for UPO - ABTS and comparison to simulated experiment. The same initial parameter combinations were used for experiment and simulation in iteration 0. (b-d): Overall highest enzymatic activity versus iterations for HRP - ABTS (b), HRP-pyrogallol (c) and HRP- TMB (d). All values represent the mean and standard deviation of replicated experiments ( $n = 3$ ). In each iteration 8 different parameter combinations were tested.

**TABLE 5** | Optimal enzymatic reaction conditions determined by the SDL platform in autonomous experiments.

Enzyme	Substrate	pH [-]	Temperature [°C]	Na <sub>2</sub> SO <sub>4</sub> [mM]	ACN [% (v/v)]	H <sub>2</sub> O <sub>2</sub> [mM]
UPO (Exp.)	ABTS	3.4	20	0	0	10.0
UPO (Sim.)	ABTS	3.4	30	125	0	8.0
HRP	ABTS	3.6	28	25	0	2.4
HRP	Pyrogallol	8	20	125	0	10.0
HRP	TMB	4.4	60	175	0	10.0

concentration (10.0 mM vs. 8.0 mM). These differences can be explained by experimental noise and inaccuracies of the simulated activities obtained from the interpolated response surface which influence the measured and simulated activities. This changes the different GP models and parameter suggestions, especially in the first iterations where most sampled points are far from the optimum and the simulated values are obtained from sparsely populated regions of the design space. With increasing number of iterations, the sampled points are closer to the optimum, where the predictions obtained from the response surface are more accurate. Thus, both the sampled parameter combinations and the enzymatic activities at those points align more closely between experiment and simulation. Due to the discretization of the design space with a step size of 0.2 for pH, the optimum identified in the HTPS with pH 3.5 could not be reached in the optimization. This is important to notice as pH has the strongest influence on the enzymatic activity for UPO-ABTS and most other tested enzyme-substrate pairings (Supporting Information S1: Figure S10). However, the optimal conditions are still very similar to those identified in the HTPS for both the experiment and simulation and yield comparable peak activities. In conclusion, the experiments validate the effectiveness of the fine-tuned BO for autonomous experimental optimization of enzymatic reaction conditions on the SDL.

To test the versatility of the SDL and the fine-tuned optimization algorithm, the enzymatic reaction conditions of another enzyme, horseradish peroxidase (HRP), with different substrates were autonomously optimized. For HRP-ABTS, the highest specific enzymatic activity over the iterations is shown in Figure 4b. The highest specific rate is  $181.2 \pm 18.3 \text{ U mg}^{-1}$  in the initial generation and increases to  $918.1 \pm 155.7 \text{ U mg}^{-1}$  by cycle four, where it plateaus for two cycles. In cycle seven the overall highest specific rate increases again and reaches an overall best of  $1603.9 \pm 45.4 \text{ U mg}^{-1}$  by cycle 10. The optimal reaction conditions for HRP-ABTS identified by the SDL are shown in Table 5. For HRP and ABTS the optimal reaction conditions align closely to those reported in previous studies. Gallati reported interfering effects between pH and ABTS and H<sub>2</sub>O<sub>2</sub> concentration (Gallati 1979). It was shown that decreasing the ABTS concentration lowers the pH optimum of HRP. In addition, H<sub>2</sub>O<sub>2</sub> and ABTS show competitive binding to the peroxidase. For 0.5 mM ABTS a pH optimum of 3.8 is reported by Gallati et. al and following the reported trend, for the 0.3 mM ABTS used in this study, the optimal pH should be lower. At 0.25 mM ABTS a  $K_m = 0.158 \text{ mM}$  for H<sub>2</sub>O<sub>2</sub> is reported, indicating that at 2.4 mM ABTS over 90% of the maximal turnover of H<sub>2</sub>O<sub>2</sub> is reached. The identified temperature optimum of 28°C is consistent with studies on HRP using phenols and guaiacol as

substrates reporting similar values (Lavery et al. 2010). To the best of our knowledge, no prior studies exist on the optimal salt and organic solvent concentrations for this enzyme-substrate pairing. The results suggest, similar to UPOs, that moderate salt concentrations and the absence of organic solvents maximize enzymatic activity of HRP with ABTS.

For HRP with pyrogallol, the maximal enzymatic activity increases from  $118.6 \pm 8.1 \text{ U mg}^{-1}$  in the initial cycle to  $250.9 \pm 27.3 \text{ U mg}^{-1}$  by iteration 3 and then plateaus (Figure 4c). In this study, an optimal pH of 8 for the oxidation of pyrogallol is found (Table 5), while previous studies reported an optimal pH of 7 for phenolic substrates (Lavery et al. 2010; Patel et al. 1997). This discrepancy may be attributed to the strongly increasing autoxidation of pyrogallol in alkaline solutions (Veselinović et al. 2012; Doona and Kustin 1993). The optimal pH identified by the SDL is thus the optimum for the superposition of both the enzymatic and nonenzymatic oxidation of pyrogallol in the investigated pH range. This highlights the importance of carefully selecting the parameter ranges for the optimization process.

Using 3,3,5,5-tetramethylbenzidine (TMB) as substrate for HRP, the overall highest activity plateaued after two iterations, with minor improvements in iteration 4 and iteration 10. Overall, the activity increased from  $603.1 \pm 43.9 \text{ U mg}^{-1}$  in the initial generation to  $1960.4 \pm 14.3 \text{ U mg}^{-1}$  in iteration 10. The optimal pH of 4.4 (Table 5) is in accordance with previous studies which report pH optima between 4.0 and 5.0 for HRP and TMB (Gallati and Pracht 1985). For H<sub>2</sub>O<sub>2</sub> concentration and temperature, lower values, 1 mM and 20°C–30°C are reported (Gallati and Pracht 1985). However, these values are based on longer incubation times of 30 min. In this study, the initial enzymatic rate in the first 30 s is studied, minimizing the effects of enzyme deactivation while covering the rate-enhancing effects.

For all enzyme-substrate pairings, convergence of the BO is indicated by a decreasing posterior variance of the model, inferring a decreasing uncertainty about the objective function (Supporting Information S1: Figure S9). For a CV of 10%–15% in the experiments, a posterior variance of 0.01–0.0225 indicates the model uncertainty equals the noise level, which is the minimum that can be achieved. Values in this range have been achieved for all autonomous optimization runs. A threshold for the posterior variance could thus serve as an important criterion to determine when to terminate the optimization process, indicating that further iterations may yield negligible improvements.

Each iteration in the autonomous optimization cycle, including reagent preparation, dispensing, incubation, measurement, computations and automated setup for the next cycle, takes approximately 75 min. In our current setup, every four cycles require manual replacement of reagent reservoirs to ensure stability of labile compounds, which is performed during regular working hours. As a result, a typical optimization campaign is completed in 2–3 days, despite the total experimental runtime being around 15 h for 12 iterations. Future improvements in automated reagent storage could enable fully continuous optimizations runs within a single day.

In summary, the SDL demonstrated the ability to autonomously and efficiently optimize the enzymatic reaction conditions for multiple enzyme-substrate pairings. Optimal reaction conditions within the five-dimensional design space are identified within 11 iterations, often also within less. For example, in the cases of HRP-pyrogallol and HRP-TMB, optimal or near-optimal conditions were reached within only three experimental cycles. This accelerated convergence can largely be attributed to more favorable initial samples generated by LHS, as the highest activities in these initial sets were already closer to the eventual optimum compared to UPO-ABTS and HRP-ABTS, effectively accelerating the optimization process. However, this additionally indicates that the rate of the optimization is still dependent on the initial parameter combinations. To further improve the system, prior knowledge, for example, from simulations or data from literature or databases could be used to initialize the first set of parameter combinations and further improve performance of the SDL.

The identified enzymatic reaction optima align well with existing studies, highlighting the SDLs reliability, versatility and efficiency for accelerating the optimization of enzymatic reactions. The fine-tuned BO algorithm, initially tested on a single enzyme-substrate pairing, proved highly generalizable, reinforcing its utility for similar optimization problems across diverse systems. This demonstrates the effectiveness of the proposed approach for tailoring experimental planning algorithms to improve SDL efficiency.

### 3 | Conclusion

In this study, we have demonstrated the development of an autonomous, machine learning-driven self-driving laboratory (SDL) platform for the rapid and precise optimization of biochemical reactions. By integrating fully automated experimental workflows, machine learning-driven optimization, and a seamless Electronic Lab Notebook (ELN)-connected workflow, the system accelerates traditionally labor-intensive optimization processes, while being highly user-friendly. This SDL was applied to optimize enzymatic reaction conditions in a five-dimensional design space. To further enhance the efficiency of the SDL, we tested and fine-tuned multiple machine learning algorithms *in silico*, using simulated optimization campaigns on an exemplary data set. Among the tested algorithms, Bayesian Optimization (BO) with a Gaussian Process (GP) model, a Matérn 3/2 kernel, and an Expected Improvement (EI) acquisition function emerged as the most efficient, achieving superior convergence rates and robust high performance across multiple initial parameter sets.

Real-world optimization campaigns using the fine-tuned BO algorithm on the SDL validated its efficiency and generalizability to different enzyme-substrate pairings. Using the BO, the SDL successfully identified optimal reaction conditions for diverse enzyme-substrate pairings with a small number of experiments. Importantly, the identified optima aligned well with prior literature, further validating reliability and generalizability of the approach. These results emphasize the platform's potential to accelerate the solving of complex, multi-parameter optimization challenges across biocatalytic applications. For the best-performing optimization algorithm identified in the presented work, we expect similar performance for other enzyme-substrate systems and, more generally, for similar biocatalytic optimization problems involving continuous-valued parameters in moderate to high-dimensional experimental spaces. The presented SDL platform, as implemented, thus has the potential to accelerate process development in diverse fields of biocatalysis including enzyme-driven biotransformations, pharmaceutical synthesis, and biosensors. Moreover, the proposed approach for fine-tuning SDLs with tailored optimization algorithms can be broadly generalized to different fields of application.

While the SDL platform demonstrated robust autonomous optimization capabilities, several challenges remain. Our current system is designed primarily for water-based enzymatic reactions, as many organic solvents and chemical transformations are not compatible with the current liquid handling hardware. To support a broader chemistry space, additional device modules and safety infrastructure, such as fume hoods, would be required. Manual intervention is still necessary for accurate preparation and placement of reagents and labware. Fully automating these steps through robotic sample preparation and handling would minimize hands-on labor and improve reproducibility. Handling and storage of labile or temperature-sensitive reagents currently require manual replacement of reservoirs after several cycles. Automated cooled storage solutions would enable longer, uninterrupted campaigns. Experiment time, reagent consumption, and safety constraints are managed via manual pre-experiment estimation, and campaign length and parameter boundaries are set accordingly. Future iterations could integrate automated tracking of reagent usage, time, and cost directly into the workflow and optimization logic, allowing for dynamic, constraint-aware experimentation. Computational scalability is another bottleneck, particularly for Bayesian Optimization as the data set and design space grow. While current computational times, maximally up to a few minutes, are not prohibitive, they are expected to become a limiting factor with larger datasets and higher dimensional design spaces. Algorithmic improvements, GPU acceleration, or distributed computing will be needed for large-scale applications. Lastly, while the workflow is modular, significant expertise is required for the implementation of new processes.

Looking ahead, integrating large language models (LLMs) into the system is a promising approach to address this challenge while enhancing the SDL's intelligence and versatility. The use of agentic LLMs, AI models equipped with scientific and experimental tools, offers exciting prospects for the field: enabling flexible control of laboratory automation through natural language, automated generation of experimental protocols and

device scripts, interactive data analysis and AI-lead, data-driven optimization campaigns. Integrating LLMs with the SDL framework has the potential to not only make experimental planning and execution more accessible to a broader user base, but also to accelerate discovery through intelligent hypothesis generation and adaptive experimentation.

In summary, the presented SDL platform establishes an efficient and extensible foundation for accelerated autonomous process optimization and discovery, paving the way for future advancements in biocatalysis and related fields. Ongoing efforts to address current limitations, alongside continued integration of advanced AI technologies, promise to further enhance the platform's efficiency, flexibility, and scientific impact.

## 4 | Experimental Section

### 4.1 | Materials

Hydrogen peroxide ( $\text{H}_2\text{O}_2$ , 3% (w/w)) and citric acid monohydrate (CA) were purchased from VWR Chemicals (Germany). Horseradish peroxidase Type VI (HRP), di-sodium hydrogen phosphate ( $\text{Na}_2\text{HPO}_4$ ), pyrogallol and 3,3',5,5'-tetramethylbenzidine dihydrochloride hydrate (TMB-d), sodium hydroxide (NaOH, 1.0 N) and hydrochloric acid (HCl, 36.5-38%) were obtained from Sigma Aldrich (Germany). Dimethyl sulfoxide (DMSO), acetonitrile (ACN), sodium sulfate anhydrous ( $\text{Na}_2\text{SO}_4$ ), sodium sulfite anhydrous ( $\text{NaSO}_3$ ) and sulfuric acid ( $\text{H}_2\text{SO}_4$ , 95-97%) were purchased from Merck (Germany). 2,2-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt (ABTS) and sodium azide 1% ( $\text{NaN}_3$ ) were procured from Thermo Fisher Scientific (Germany) and G-Biosciences (USA). All chemicals and reagents were used as received without additional purification. Buffers were prepared in UPW and pH adjusted using 1 M NaOH and 1 M HCl.

### 4.2 | Automated HTPs of Initial Enzymatic Reaction Rates

In the first step, the automated robotic setup of the SDL was employed to generate a comprehensive data set of enzymatic activities, designed for in-silico testing of various optimization strategies. The initial reaction rate of an unspecific peroxidase (UPO) with a streptavidin tag (generously provided by Niklas Teetz, BLT, KIT, Germany) was screened using an ABTS assay. This screening was conducted across a range of pH, temperature, and concentrations of salt, organic solvent, and cosubstrate. The colorimetric ABTS assay is based on the principle that the enzymatic oxidation of ABTS produces a green-colored radical cation ( $\text{ABTS}^{\bullet+}$ ), whose formation can be monitored spectrophotometrically at 420 nm. A phosphate-citrate (PC) buffer, consisting of varying ratios of 0.1 M CA and 0.2 M  $\text{Na}_2\text{HPO}_4$  to adjust pH, was used as buffer system.  $\text{Na}_2\text{SO}_4$  was used as additional salt, ACN as additional organic solvent, and  $\text{H}_2\text{O}_2$  as cosubstrate. The parameter ranges investigated during this study are summarized in Table 6.

To efficiently explore the entire design space with a limited number of experiments, an iterative approach using 3-Level Full-Factorial Designs was employed. In the first iteration, the

**TABLE 6** | Parameter ranges for the HTPs of initial enzymatic reaction rates of UPO in ABTS assay.

Parameter	Lower limit	Upper limit
pH [–]	2.5	8
Temperature [°C]	20	60
Salt concentration [mM]	0	500
Organic solvent concentration [% (v/v)]	0	30
Cosubstrate concentration [mM]	0	10

experimental bounds were defined by the predefined upper and lower parameter limits. For the second iteration, the parameter ranges were halved and centered around the best-performing conditions from the previous iteration—specifically, the experimental setup that achieved the highest initial enzymatic reaction rate. If centering around the optimum led to impractical values (e.g., negative or zero), the center point was repositioned within the new range to ensure feasible experimental conditions. For instance, if a parameter previously ranged from 0 to 100 and the optimal value was found at 0, the next iteration adjusted the range to 0 to 50 instead of extending into negative values (e.g., –50 to 0). The final third iteration was performed analogously again halving the parameter ranges and centering around the updated optimum from the second iteration. All experiments of the screening were conducted as duplicates.

Each iteration of the screening proceeds according to the following workflow. First, during the initialization step the experimental design is generated according to the specified bounds provided in the JSON metadata file. The design is then split into chunks ordered by pH and temperature. Each chunk is randomly shuffled to eliminate any potential bias associated with the timing of experiments. To assess reproducibility and potential instability of reagents, the center point of the design is included in every chunk. For each shuffled chunk of parameter combinations, representing an experimental cycle, control scripts for the pipetting robot (Flex, Opentrons, USA) and the plate reader (Spark, Tecan, Switzerland) are automatically generated before the experiments commence.

Every experimental cycle begins with buffer mixing in 96-well-plate format by the pipetting robot to procure the mixtures of the parameter combinations to be tested. The salt concentration is adjusted by mixing phosphate-citrate (PC) buffers with no additional salt and those with a high salt concentration, both prepared at the desired pH. In the screening, the pH value is adjusted by selecting reservoirs containing buffers at the required pH. The volumes of the cosubstrate,  $V_{\text{cosubstrate}}$ , and organic solvent,  $V_{\text{solvent}}$ , are dosed accordingly to adjust the correct concentrations,  $C_{\text{cosubstrate}}$  and  $C_{\text{solvent}}$ , from the provided stock solutions with the concentrations  $C_{\text{cosubstrate,stock}}$  and  $C_{\text{solvent,stock}}$ . The required volumes during this buffer mixing step are calculated according to the following Equations (1) – (4). The total assay volume  $V_{\text{total}}$  was fixed at 200  $\mu\text{L}$ . The final

concentration,  $c_{\text{substrate}}$ , and volume,  $V_{\text{substrate}}$ , of the substrate ABTS were fixed at 0.3 mM and 20  $\mu\text{L}$ , respectively. Stock concentrations of 2 M  $\text{Na}_2\text{SO}_4$ , and 100 mM  $\text{H}_2\text{O}_2$  in PC buffer at the according pH were used. ACN was supplied as pure solvent at 100% (v/v) in the reagent reservoir.

$$V_{\text{buffer,high salt}} = \frac{c_{\text{salt}}}{c_{\text{salt,stock}}} V_{\text{total}} \quad (1)$$

$$V_{\text{cosubstrate}} = \frac{c_{\text{cosubstrate}}}{c_{\text{cosubstrate,stock}}} V_{\text{total}} \quad (2)$$

$$V_{\text{solvent}} = \frac{c_{\text{solvent}}}{c_{\text{solvent,stock}}} V_{\text{total}} \quad (3)$$

$$V_{\text{buffer,low salt}} = V_{\text{total}} - V_{\text{buffer,high salt}} - V_{\text{cosubstrate}} - V_{\text{solvent}} - V_{\text{substrate}} \quad (4)$$

Once the correct volumes of low and high salt buffer, cosubstrate, and solvent are dispensed into the assay well-plate, the well-plate is transferred onto the heater-shaker-module of the pipetting robot using its internal gripper. The plate is then heated to the target temperature for the assay. After reaching the target temperature, the protocol delays for 5 min to ensure thermal equilibrium between the heating plate and the liquids in the wells. To initiate the enzymatic reaction, substrate and enzyme are dispensed column-wise into the plate using a multi-channel pipette. The reaction is allowed to proceed for exactly 60 s before being quenched by the addition of 20  $\mu\text{L}$  of 1% sodium azide. This process is repeated for each column of the well-plate being tested. After completing the assays, the robotic arm (UR5e, Universal Robots, Denmark), transfers the well-plate into the plate reader, where the absorbance at 420 nm is measured to quantify the formation of the  $\text{ABTS}^{\bullet+}$  radical.

Next, the enzymatic reaction rates are calculated according to Equations 5 and 6. First the concentration of the  $\text{ABTS}^{\bullet+}$  radical is calculated from the raw absorbance data following the Beer-Lambert law (Equation 5):

$$c_i = \frac{A_{\lambda,\text{sample}} - A_{\lambda,\text{blank}}}{\epsilon_{i,\lambda} d} \quad (5)$$

Herein,  $c_i$  is the concentration of the substance  $i$ ,  $A_{\lambda,\text{sample}}$  the absorbance of the sample at the wavelength  $\lambda$ ,  $A_{\lambda,\text{blank}}$  the absorbance of the blank,  $\epsilon_{i,\lambda}$  the extinction coefficient of the substance  $i$  at wavelength  $\lambda$  and  $d$  the path length. From dividing  $c_i$  by the reaction time  $t$  and the enzyme concentration  $c_{\text{enzyme}}$  the specific reaction rate  $v_{i,\text{spec}}$  is obtained (Equation 6). Using ABTS and TMB as substrates, the unit (U) of enzymatic activity is defined as the turnover of 1  $\mu\text{mole}$  of substrate per minute. For HRP-Pyrogallol 1 U oxidizes 1 mg of Pyrogallol in 20 s. In this study,  $v_{i,\text{spec}}$  is consequently always presented in  $\text{U mg}^{-1}$ .

$$v_{i,\text{spec}} = \frac{c_i}{t c_{\text{enzyme}}} \quad (6)$$

The data evaluation including calculation of enzymatic reaction rates concludes an experimental cycle and the station prepares

for the following cycle by replacing the used well-plate with a new one from the storage on the work-bench. Every third cycle, the robotic arm replaces the reagent reservoir in the liquid handler with a new one retrieved from the designated storage positions on the workbench. The system is capable of completing an entire iteration of screening—comprising a 3-Level Full-Factorial Design divided in 9 experimental cycles and a total of 513 individual assays—without any human intervention.

### 4.3 | Uncertainty Sources and Mitigation Strategies

The SDL system faces several potential sources of experimental uncertainty that must be systematically addressed to ensure reliable optimization results. The liquid handling system was equipped with two pipette types to optimize accuracy across different volume ranges: a Flex 1-Channel Pipette (1–50  $\mu\text{L}$ ) for smaller volumes and a Flex 1-Channel Pipette (5–1000  $\mu\text{L}$ ) for larger volumes (Opentrons Flex 1-Channel Pipette - Opentrons 2025). For volumes of 20  $\mu\text{L}$  or lower, the smaller volume range pipette was used to minimize error. This dual-pipette approach minimizes the impact of volume-dependent errors on the most sensitive assay components. Additionally, stock concentrations of reagents were strategically chosen, so that no pipetting volume was smaller than 5  $\mu\text{L}$  had to be dispensed.

Chemical stability can represent another source of variability in the autonomous system. Enzyme solutions were maintained on ice before addition to assays to prevent thermal deactivation, while liquid reservoirs were sealed with aluminum foil to prevent evaporation and photodegradation. Laboratory lighting and instrument internal lighting were turned off during operations to minimize light-induced degradation of photosensitive compounds, particularly important for the chromogenic substrates used in the assays. All stock solutions were freshly prepared for each optimization campaign to ensure consistent starting conditions.

Environmental control and temporal factors also contribute to system variability. Temperature and humidity in the laboratory were maintained constant throughout optimization campaigns. During batches, each temperature equilibration step required 4–5 min to ensure thermal stability before reaction initiation, with temperature conditions processed in ascending order to minimize thermal stress on the system. Each experimental cycle required approximately 75 min, during which environmental conditions remained stable. Buffer mixing was completed immediately before temperature equilibration to minimize evaporation and concentration drift that could affect final concentrations.

These mitigation strategies collectively resulted in an average coefficient of variation of 13.2% for the replicates in the HTPS for the ABTS assay, which is consistent with literature values of approximately 12% (Dolz et al. 2023).

### 4.4 | In-Silico Optimization Algorithm Testing in Simulated Experiments

The generated data set for enzymatic reaction rates was used to test and fine-tune various optimization algorithms in simulated

experiments. The identified optimal algorithm finds the optimal reaction conditions, i.e., those that yield the highest reaction rate with the smallest number of iterations, i.e., conducted experiments. Four different types of optimization algorithms were tested: Genetic Algorithm (GA) (Katoch et al. 2021; Vanneschi and Silva 2023), Particle Swarm Optimization (PSO) (Jain et al. 2018; Wang et al. 2018), Simulated Annealing (SA) (Rutenbar 1989; Aarts et al. 2005), and Bayesian Optimization (BO) (Frazier 2018; Shahriari et al. 2016). These algorithms have proven effective for the optimization of continuous variables in diverse biotechnological applications, such as enzymatic reactions (Tachibana et al. 2023; Siedentop et al. 2023; Elder et al. 2021; Muffler et al. 2007; Liu et al. 2008) and fermentation processes (Garlapati et al. 2010; Skolpap et al. 2008; Sarma et al. 2009; Subba Rao et al. 2008; Nagata and Chu 2003; Parkhey et al. 2017; Dhagat and Jujjavarapu 2021). Additionally, the algorithm performance was compared to RSM, the most established optimization method in the field of biocatalysis (Aktaş 2005; Lee et al. 2006; Mushtaq et al. 2015; Tacias-Pascacio et al. 2019). RS was used as a baseline to benchmark the performance of guided optimization algorithms against brute-force, unguided sampling of the design space.

For each algorithm, key hyperparameters were defined and systematically tested across a range of values. In this context, hyperparameters refer to the parameters affecting the algorithm itself, for example, the mutation rate in a GA. All possible permutations of these hyperparameter values (hereafter referred to as algorithm variants) were evaluated using a grid search approach.

All optimization algorithms were implemented in-house using Python, relying only on standard scientific libraries for data handling and modeling. For BO, Gaussian process regression was implemented using the *scikit-learn* library, with the kernel function selected from the available options, default zero mean prior, and the noise level set to 0.02 to reflect the average coefficient of variation (CV) of 13.2% observed in experimental measurements. Hyperparameters were optimized by log marginal likelihood with ten random restarts, and standard acquisition functions were employed with typical literature values. While this homoscedastic noise assumption simplified the modeling, we acknowledge that more advanced heteroscedastic noise modeling (using measured experimental uncertainties) could further improve predictive accuracy in future studies. Detailed descriptions of the optimization algorithms are available in the *Supporting Information, Section 2* and in the published repository (putzsebastian 2025).

The testing process was standardized to mimic the cyclic experimental workflow of a self-driving lab. Each algorithm was initialized with the same 30 unique sets of initial parameter combinations, with each set containing 8 parameter combinations generated via LHS. In the simulated experiments, the enzymatic reaction rate for each parameter combination was calculated using linear interpolation based on the data set. To simulate real-world experimental noise, a uniform noise in the range of 13.2%, corresponding to the average coefficient of variation (CV) observed during HTPS, was added to the calculated rates. After the enzymatic reaction rates are calculated for the initial generation, the algorithm variant suggests 8 new

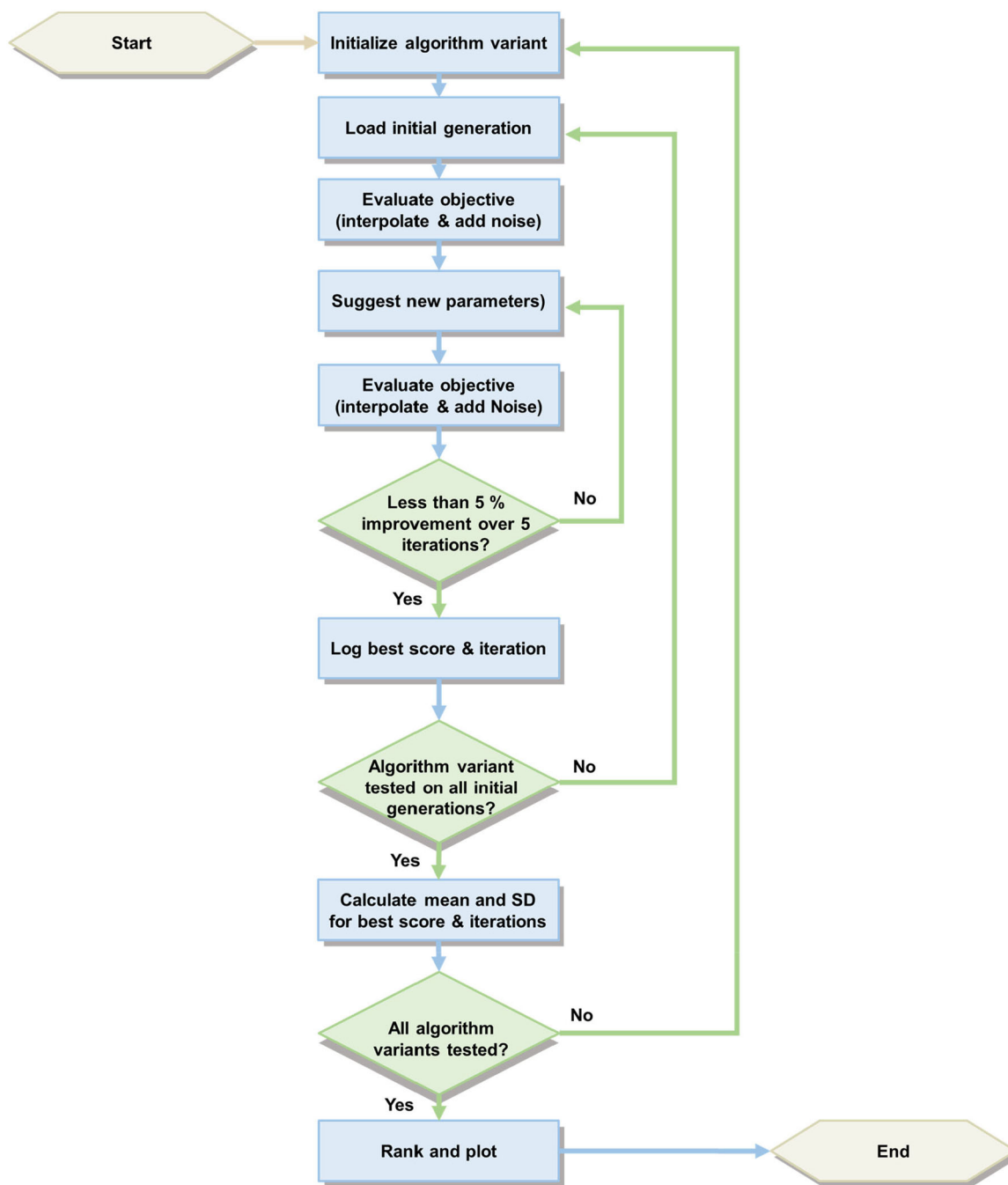
parameter combinations that are tested in the next round. The best target value, that is, the highest enzymatic reaction rate, in each iteration is tracked. A lack-of-improvement criterion was selected to determine convergence of the algorithm. If the best target value did not improve by at least 5% over the last 5 iterations, the optimization process was terminated. The highest enzymatic reaction rate identified and the number of iterations until convergence were recorded. This process was repeated for all 30 initial parameter sets for each algorithm variant. As metrics for comparing different algorithms, the mean and standard deviation of the highest enzymatic reaction rate and the number of iterations until convergence, were calculated across the 30 runs for each algorithm variant. The workflow for this screening of algorithm variants is illustrated as flow-chart in Figure 5.

#### 4.5 | Autonomous Optimization of Enzymatic Reaction Conditions in Real Experiments

To validate the best-performing in-silico algorithm in real experiments, autonomous optimization campaigns were conducted using the SDL. The platform was utilized to optimize the reaction conditions for the enzyme HRP using the substrates ABTS, Pyrogallol and TMB-d. Additionally, the reaction conditions for the UPO-ABTS pairing were optimized to compare them to a simulated optimization run. The optimization aimed to maximize the initial enzymatic reaction rate by varying pH, temperature, and the concentrations of salt, organic solvent, and cosubstrate.

Similar to the screening, a phosphate-citrate (PC) buffer system, consisting of varying ratios of 0.1 M CA and 0.2 M  $\text{Na}_2\text{HPO}_4$ , was used.  $\text{Na}_2\text{SO}_4$  was used as additional salt, ACN as additional organic solvent and  $\text{H}_2\text{O}_2$  as cosubstrate. The parameter ranges remained consistent with those previously outlined in Table 6. However, for practicality and computational efficiency, the design space was discretized into manageable increments, see Table 7. To accommodate this discretization, the BO was adapted to operate directly on the discrete parameter grid, with acquisition functions evaluated at each grid point rather than through continuous optimization followed by rounding. This approach ensures all proposed parameter combinations are experimentally feasible while avoiding potential artifacts from continuous-to-discrete parameter mapping.

To enable testing of different pH values and temperatures within the same experimental cycle, the protocol differed slightly from the automated HTPS experiments. For the first experimental cycle, 8 parameter combinations were generated using LHS. During each cycle, buffers were prepared according to the tested parameter combinations using Equations (1)–(4). However, to adjust pH and salt concentrations without premixed buffers, the buffer volumes  $V_{\text{buffer, high salt}}$  and  $V_{\text{buffer, low salt}}$  were further split into separate volumes of 0.1 M CA and 0.2 M  $\text{Na}_2\text{HPO}_4$  with no additional  $\text{Na}_2\text{SO}_4$  and 1 M additional  $\text{Na}_2\text{SO}_4$ , according to the mixing table for the PC buffer (McIlvaine 1921). After dosing the required volumes of CA, CA with added salt,  $\text{Na}_2\text{HPO}_4$ ,  $\text{Na}_2\text{HPO}_4$  with added salt, 100 mM  $\text{H}_2\text{O}_2$  in UPW and ACN (100% (v/v)) were dispensed at according volumes to adjust the correct concentrations of



**FIGURE 5** | Flow-chart of the in-silico optimization algorithm screening.

**TABLE 7** | Parameter ranges for the HTPS of initial enzymatic reaction rates of UPO enzyme in ABTS assay.

Parameter	Lower limit	Upper limit	Step size
pH [–]	2.4	8	0.2
Temperature [°C]	20	60	2
Salt concentration [mM]	0	500	25
Organic solvent concentration [% (v/v)]	0	30	2.5
Cosubstrate concentration [mM]	0	10	0.2

**TABLE 8** | Assay-specific parameters for ABTS, Pyrogallol and TMB-d assay with the enzymes HRP and UPO.

Enzyme Substrate	UPO ABTS	HRP ABTS	HRP Pyrogallol	HRP TMB-d
Substrate concentration	0.3 mM	0.3 mM	40 mM	1 mM
Substrate solvent	UPW	UPW	UPW	DMSO
Enzyme concentration [mg mL <sup>-1</sup> ]	5 × 10 <sup>-4</sup>	5 × 10 <sup>-4</sup>	2.5 × 10 <sup>-3</sup>	1 × 10 <sup>-4</sup>
Reaction time [s]	60	60	60	30
Stop reagent	Sodium azide (1%)	Sodium azide (1%)	Sodium azide (1%) + 1 M NaSO <sub>3</sub>	H <sub>2</sub> SO <sub>4</sub> (2 M)
Stop reagent volume [μL]	20	20	20	50
Measurement wavelength [nm]	420	420	420	450

cosubstrate and organic solvent. Then the buffer mixing was complete and the well-plate was heated to the required temperatures specified in the tested parameter combinations, processed in ascending order. Once the lowest target temperature was reached and stabilized, the pipetting robot dispensed 20 μL substrate and 20 μL enzyme solution (in UPW) into the wells corresponding to that temperature to initiate the reaction. The reaction was stopped at the specified reaction time for each assay. This process was repeated for all temperature levels until all assays for that experimental cycle were completed. The assay-specific differences are shown in Table 8.

After completing the assays, the robotic arm transferred the well-plate to the plate reader for absorbance measurements. The well-plate was then returned to the liquid-handling station. Every fourth cycle, the well-plate and reagent reservoirs were replaced with fresh ones from the storage on the workbench. Reaction rates were calculated at the end of the cycle using Equations 5 and 6.

The BO algorithm then determined the 8 parameter combinations that are tested in the subsequent experiment cycle. In brief, first, a GP model with a Matérn32 kernel as covariance function was fitted to all thus far generated data points. Using the EI acquisition function, the parameter combination with the highest EI was chosen as candidate for the next cycle. For batch-wise optimization with multiple parameter combinations per cycle, the Kriging-believer algorithm was employed to generate a batch of 8 parameter combinations. To avoid redundancy, a feature was implemented so that no parameter combination was tested more than once across all experimental cycles. Details on the utilized BO and all other tested optimization algorithms can be found in the *Supporting Information, Section 2*.

#### Author Contributions

**Sebastian Putz:** conceptualization, methodology, experimental work, data analysis, visualization, writing – original draft. **Niklas Teetz,** **Michael Abt:** conceptualization, resources (enzyme material), writing – review and editing. **Pascal Jerono,** **Thomas Meurer:** methodology, writing – review and editing. **Matthias Franzreb:** conceptualization, supervision, project administration, funding acquisition, writing – review and editing.

#### Acknowledgments

This study was funded from the KIT via the subproject BioCAR in the project Auto.MAP and the Helmholtz Program “Materials Systems Engineering” under Program No. 43.31.01. Open Access funding enabled and organized by Projekt DEAL.

#### Conflicts of Interest

The authors declare no conflicts of interest.

#### Data Availability Statement

The primary data set and codebase that support the findings of this study are openly available in GitHub under <https://github.com/putzsebastian/sdl-enzymes-optimization> (<https://doi.org/10.5281/zenodo.15730516>). Additional data and code are available from the corresponding author upon reasonable request.

#### References

- Aarts, E., J. Korst, and W. Michiels. 2005. *Search Methodologies*, edited by E. K. Burke and G. Kendall, 187–210. Springer US.
- Abdel-Latif, K., R. W. Epps, F. Bateni, S. Han, K. G. Reyes, and M. Abolhasani. 2021. “Self-Driven Multistep Quantum Dot Synthesis Enabled by Autonomous Robotic Experimentation in Flow.” *Advanced Intelligent Systems* 3: 2000245.
- Abolhasani, M., and E. Kumacheva. 2023. “The Rise of Self-Driving Labs in Chemical and Materials Sciences.” *Nature Synthesis* 2: 483–492.
- Aktaş, N. 2005. “Optimization of Biopolymerization Rate by Response Surface Methodology (RSM).” *Enzyme and Microbial Technology* 37: 441–447.
- API—elabftw 5.1.11 documentation. 2024. <https://docelabftw.net/api.html>.
- Asensio, L., I. González, T. García, and R. Martín. 2008. “Determination of Food Authenticity by Enzyme-Linked Immunosorbent Assay (ELISA).” *Food Control* 19, no. 1: 1–8.
- Bennett, J. A., and M. Abolhasani. 2022. “Autonomous Chemical Science and Engineering Enabled by Self-Driving Laboratories.” *Current Opinion in Chemical Engineering* 36: 100831.
- Bisswanger, H. 2014. “Enzyme Assays.” *Perspectives in Science* 1: 41–55.
- Bormann, S., A. Gomez Baraibar, Y. Ni, D. Holtmann, and F. Hollmann. 2015. “Specific Oxyfunctionalizations Catalysed by Peroxygenases: Opportunities, Challenges and Solutions.” *Catalysis Science & Technology* 5: 2038–2052.
- Canty, R. B., B. A. Koscher, M. A. McDonald, and K. F. Jensen. 2023. “Integrating Autonomy Into Automated Research Platforms.” *Digital Discovery* 2: 1259–1268.



- Carrea, G., and S. Riva. 2000. *Angewandte Chemie International Edition* 2226: 39.
- Chapman, J., A. E. Ismail, and C. Z. Dinu. 2018. "Industrial Applications of Enzymes: Recent Advances, Techniques, and Outlooks." *Catalysts* 8: 238.
- Christensen, M., L. P. E. Yunker, F. Adedeji, et al. 2021. "Data-Science Driven Autonomous Process Optimization." *Communications Chemistry* 4: 112.
- Coley, C. W., D. A. Thomas, J. A. M. Lummiss, et al. 2019. "A Robotic Platform for Flow Synthesis of Organic Compounds Informed by AI Planning." *Science* 365: eaax1566.
- Dave, A., J. Mitchell, K. Kandasamy, et al. 2020. "Autonomous Discovery of Battery Electrolytes With Robotic Experimentation and Machine Learning." *Cell Reports Physical Science* 1: 100264.
- Dhagat, S., and S. E. Jujjavarapu. 2021. "Simulated Annealing and Artificial Neural Network as Optimization Tools to Enhance Yields of Bioemulsifier and Exopolysaccharides by Thermophilic *Brevibacillus Borstelensis*." *Journal of Environmental Chemical Engineering* 9: 105499.
- Dolz, M., D. T. Monterrey, A. Beltrán-Nogal, et al. 2023. *Methods in Enzymology*, 73–109. Elsevier.
- Doona, C. J., and K. Kustin. 1993. "Kinetics and Mechanism of Pyrogallol Autoxidation: Calibration of the Dynamic Response of an Oxygen Electrode." *International Journal of Chemical Kinetics* 25: 239–247.
- Elder, S., C. Klumpp-Thomas, A. Yasgar, et al. 2021. "Cross-Platform Bayesian Optimization System for Autonomous Biological Assay Development." *SLAS Technology* 26: 579–590.
- Epps, R. W., M. S. Bowen, A. A. Volk, et al. 2020. *Advanced Materials* 32: 2001626.
- Fernandes, P. 2010. "Enzymes in Food Processing: A Condensed Overview on Strategies for Better Biocatalysts." *Enzyme Research* 2010: 1–19.
- Frazier, P. I. 2018. *Recent Advances in Optimization and Modeling of Contemporary Problems*, edited by E. Gel, L. Ntaimo, D. Shier, and H. J. Greenberg, 255–278. INFORMS.
- Gallati, H. 1979. "Peroxidase aus Meerrettich: Kinetische Studien Sowie Optimierung Der Aktivitätsbestimmung mit den Substraten H<sub>2</sub>O<sub>2</sub> und 2,2'-Azino-di-(3-ethyl-benzthiazolinsulfonsäure-(6)) (ABTS)." *Clinical Chemistry and Laboratory Medicine* 17, no. 1: 1–8.
- Gallati, H., and I. Pracht. 1985. "Bestimmung von Peroxidase mit H<sub>2</sub>O<sub>2</sub> und 3,3',5,5'-Tetramethylbenzidin." *Journal of Clinical Chemistry and Clinical Biochemistry* 23: 453–460. <https://doi.org/10.1515/cclm.1985.23.8.453>.
- Gan, S. D., and K. R. Patel. 2013. "Enzyme Immunoassay and Enzyme-Linked Immunosorbent Assay." *Journal of Investigative Dermatology* 133: 1–3.
- Garlapati, V. K., P. R. Vundavilli, and R. Banerjee. 2010. "Evaluation of Lipase Production by Genetic Algorithm and Particle Swarm Optimization and Their Comparative Study." *Applied Biochemistry and Biotechnology* 162: 1350–1361.
- Granda, J. M., L. Donina, V. Dragone, D.-L. Long, and L. Cronin. 2018. "Controlling an Organic Synthesis Robot With Machine Learning to Search for New Reactivity." *Nature* 559: 377–381.
- Ha, T., D. Lee, Y. Kwon, et al. 2023. "AI-Driven Robotic Chemist for Autonomous Synthesis of Organic Molecules." *Science Advances* 9: eadj0461.
- Hamedirad, M., R. Chao, S. Weisberg, J. Lian, S. Sinha, and H. Zhao. 2019. "Towards a Fully Automated Algorithm Driven Platform for Biosystems Design." *Nature Communications* 10: 5150.
- Hara, K. Y., M. Araki, N. Okai, S. Wakai, T. Hasunuma, and A. Kondo. 2014. "Development of Bio-Based Fine Chemical Production Through Synthetic Bioengineering." *Microbial Cell Factories* 13: 173.
- Harris, S. B., A. Biswas, S. J. Yun, et al. 2024. "Autonomous Synthesis of Thin Film Materials with Pulsed Laser Deposition Enabled by In Situ Spectroscopy and Automation." *Small Methods* 8, no. 9: 2301763. <https://onlinelibrary.wiley.com/doi/10.1002/smt.202301763>.
- Häse, F., L. M. Roch, and A. Aspuru-Guzik. 2019. "Next-Generation Experimentation With Self-Driving Laboratories." *Trends in Chemistry* 1: 282–291.
- Hickman, R., M. Sim, S. Pablo-García, et al. 2023b. "Atlas: A Brain for Self-driving Laboratories." *ChemRxiv*. <https://doi.org/10.26434/chemrxiv-2023-8nrxx>.
- Hickman, R. J., J. Ruža, H. Tribukait, L. M. Roch, and A. García-Durán. 2023a. "Equipping Data-Driven Experiment Planning for Self-driving Laboratories With Semantic Memory: Case Studies Of Transfer Learning in Chemical Reaction Optimization." *Reaction Chemistry Engineering* 8: 1606: 2284.
- Hofrichter, M., H. Kellner, M. J. Pecyna, and R. Ullrich. 2015. *Mono-oxygenase, Peroxidase and Peroxygenase Properties and Mechanisms of Cytochrome P450*, edited by E. G. Hrycay and S. M. Bandiera, 341–368. Springer International Publishing.
- Jain, N. K., U. Nangia, and J. Jain. 2018. "A Review of Particle Swarm Optimization." *Journal of The Institution of Engineers (India): Series B* 99: 407–411.
- Jegannathan, K. R., and P. H. Nielsen. 2013. "Environmental Assessment of Enzyme Use in Industrial Production – A Literature Review." *Journal of Cleaner Production* 42: 228–240.
- Kanda, G. N., T. Tsuzuki, M. Terada, et al. 2022. "Robotic Search for Optimal Cell Culture in Regenerative Medicine." *eLife* 11: e77007.
- Katoch, S., S. S. Chauhan, and V. Kumar. 2021. "A Review on Genetic Algorithm: Past, Present, and Future." *Multimedia Tools and Applications* 80: 8091–8126.
- Katz, E., A. F. Bückmann, and I. Willner. 2001. "Self-Powered Enzyme-Based Biosensors." *Journal of the American Chemical Society* 123: 10752–10753.
- Kilic, N. M., S. Singh, G. Keles, S. Cinti, S. Kurbanoglu, and D. Odaci. 2023. "Novel Approaches to Enzyme-Based Electrochemical Nanobiosensors." *Biosensors* 13: 622.
- Kim, S., S. Ga, H. Bae, et al. 2024. "Multidisciplinary Approaches for Enzyme Biocatalysis in Pharmaceuticals: Protein Engineering, Computational Biology, and Nanoarchitectonics." *EES Catalysis* 2: 14–48.
- King, R. D., K. E. Whelan, F. M. Jones, et al. 2004. "Functional Genomic Hypothesis Generation and Experimentation by a Robot Scientist." *Nature* 427: 247–252.
- Lavery, C. B., M. C. MacInnis, M. J. MacDonald, et al. 2010. "Purification of Peroxidase From Horseradish (*Armoracia rusticana*) Roots." *Journal of Agricultural and Food Chemistry* 58: 8471–8476.
- Lee, W. C., S. Yusof, N. S. A. Hamid, and B. S. Baharin. 2006. "Optimizing Conditions for Enzymatic Clarification of Banana Juice Using Response Surface Methodology (RSM)." *Journal of Food Engineering* 73: 55–63.
- Li, J., J. Li, R. Liu, et al. 2020a. *Nature Communications* 11: 2046.
- Li, J., Y. Tu, R. Liu, Y. Lu, and X. Zhu. 2020b. *Advanced Science* 7: 1901957.
- Liese, A., and M. Vilella Filho. 1999. "Production of Fine Chemicals Using Biocatalysis." *Current Opinion in Biotechnology* 10: 595–603.
- Liu, J., X. Guan, D. Zhu, and J. Sun. 2008. "Optimization of the Enzymatic Pretreatment in Oat Bran Protein Extraction by Particle Swarm Optimization Algorithms for Response Surface Modeling." *LWT - Food Science and Technology* 41: 1913–1918.
- MacLeod, B. P., F. G. L. Parlane, T. D. Morrissey, et al. 2020. "Self-Driving Laboratory for Accelerated Discovery of Thin-Film Materials." *Science Advances* 6: eaaz8867.

- MacLeod, B. P., F. G. L. Parlane, C. C. Rupnow, et al. 2022. "A Self-Driving Laboratory Advances the Pareto Front for Material Properties." *Nature Communications* 13: 995.
- MacLeod, B. P., F. G. L. Parlane, and C. P. Berlinguette. 2023. "How to Build an Effective Self-Driving Laboratory." *MRS Bulletin* 48: 173–178.
- Martin, H. G., T. Radivojevic, J. Zucker, et al. 2023. "Perspectives for Self-Driving Labs in Synthetic Biology." *Current Opinion in Biotechnology* 79: 102881.
- Martin-Diaz, J., P. Molina-Espeja, M. Hofrichter, F. Hollmann, and M. Alcalde. 2021. *Biotech & Bioengineering* 118: 3002.
- McIlvaine, T. C. 1921. "A Buffer Solution for Colorimetric Comparison." *Journal of Biological Chemistry* 49: 183–186.
- Melo, R. L. F., F. S. Neto, D. N. Dari, et al. 2024. *International Journal of Biological Macromolecules* 264: 130817.
- Molina-Espeja, P., S. Ma, D. M. Mate, R. Ludwig, and M. Alcalde. 2015. "Tandem-Yeast Expression System for Engineering and Producing Unspecific Peroxygenase." *Enzyme and Microbial Technology* 73–74: 29–33.
- Molina-Espeja, P., E. Garcia-Ruiz, D. Gonzalez-Perez, R. Ullrich, M. Hofrichter, and M. Alcalde. 2014. "Directed Evolution of Unspecific Peroxygenase From *Agroclybe Aegerita*." *Applied and Environmental Microbiology* 80: 3496–3507.
- Muffler, K., M. Retzlaff, K.-H. Van Pée, and R. Ulber. 2007. "Optimization of Halogenase Enzyme Activity by Application of a Genetic Algorithm." *Journal of Biotechnology* 127: 425–433.
- Mushtaq, M., B. Sultana, H. N. Bhatti, and M. Asghar. 2015. "Rsm Based Optimized Enzyme-Assisted Extraction of Antioxidant Phenolics From Underutilized Watermelon (*Citrullus lanatus* Thunb.) rind." *Journal of Food Science and Technology* 52: 5048–5056.
- Nagata, Y., and K. H. Chu. 2003. "Optimization of a Fermentation Medium Using Neural Networks and Genetic Algorithms." *Biotechnology Letters* 25: 1837–1842.
- Nikolaev, P., D. Hooper, N. Perea-López, M. Terrones, and B. Maruyama. 2014. "Discovery of Wall-Selective Carbon Nanotube Growth Conditions via Automated Experimentation." *ACS Nano* 8: 10214–10222.
- Notin, P., N. Rollins, Y. Gal, C. Sander, and D. Marks. 2024. "Machine Learning for Functional Protein Design." *Nature Biotechnology* 42: 216–228.
- Ohkubo, I., Z. Hou, J. N. Lee, et al. 2021. "Realization of Closed-Loop Optimization of Epitaxial Titanium Nitride Thin-Film Growth via Machine Learning." *Materials Today Physics* 16: 100296.
- Onyeogaziri, F. C., and C. Papanephytous. 2019. "A General Guide for the Optimization of Enzyme Assay Conditions Using the Design of Experiments Approach." *SLAS Discovery* 24: 587–596.
- Opentrons Flex 1-Channel Pipette - Opentrons. <https://opentrons.com/products/opentrons-flex-1-channel-pipette>, accessed: Jun., 2025.
- Parkhey, P., P. Gupta, and J. S. Eswari. 2017. "Optimization of Cellulase Production From Isolated Cellulolytic Bacterium: Comparison Between Genetic Algorithms, Simulated Annealing, and Response Surface Methodology." *Chemical Engineering Communications* 204: 28–38.
- Patel, P. K., M. S. Mondal, S. Modi, and D. V. Behere. 1997. "Kinetic Studies on the Oxidation of Phenols by the Horseradish Peroxidase Compound II." *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* 1339: 79–87.
- Putz, S., J. Dötting, T. Ballweg, A. Tschöpe, V. Biniyaminov, and M. Franzreb. 2024. "Self-Driving Lab for Solid-Phase Extraction Process Optimization and Application to Nucleic Acid Purification." *Advanced Intelligent Systems* 7, no. 1: 2400564. <https://doi.org/10.1002/aisy.202400564>.
- putzsebastian, 2025. <https://doi.org/10.5281/ZENODO.15730516>.
- Rao, K. H., S. Sahoo, and J. Gupta. 2024. *in* Recent Advances in Bioprocess Engineering and Bioreactor Design, edited by S. Dhagat, S. E. Jujjavarapu, N. S. Sampath Kumar, and C. Mahapatra, 65–73. Springer Nature.
- Rapp, J. T., B. J. Bremer, and P. A. Romero. 2024. "Self-Driving Laboratories to Autonomously Navigate the Protein Fitness Landscape." *Nature Chemical Engineering* 1: 97–107.
- Reetz, M. T., G. Qu, and Z. Sun. 2024. "Engineered Enzymes for the Synthesis of Pharmaceuticals and Other High-Value Products." *Nature Synthesis* 3: 19–32.
- Rooney, M. B., B. P. MacLeod, R. Oldford, et al. 2022. "A Self-Driving Laboratory Designed to Accelerate the Discovery of Adhesive Materials." *Digital Discovery* 1: 382–389.
- Rutenbar, R. A. 1989. "Simulated Annealing Algorithms: An Overview." *IEEE Circuits and Devices Magazine* 5: 19–26.
- Salley, D., G. Keenan, J. Grizou, A. Sharma, S. Martín, and L. Cronin. 2020. "A Nanomaterials Discovery Robot for the Darwinian Evolution of Shape Programmable Gold Nanoparticles." *Nature Communications* 11: 2771.
- Sarma, M. V. R. K., V. Sahai, and V. S. Bisaria. 2009. "Genetic Algorithm-Based Medium Optimization for Enhanced Production of Fluorescent *Pseudomonas* R81 and Siderophore." *Biochemical Engineering Journal* 47: 100–108.
- Shahriari, B., K. Swersky, Z. Wang, R. P. Adams, and N. De Freitas. 2016. "Taking the Human Out of the Loop: A Review of Bayesian Optimization." *Proceedings of the IEEE* 104: 148–175.
- Shields, B. J., J. Stevens, J. Li, et al. 2021. "Bayesian Reaction Optimization as a Tool for Chemical Synthesis." *Nature* 590: 89–96.
- Si, T., R. Chao, Y. Min, Y. Wu, W. Ren, and H. Zhao. 2017. "Automated Multiplex Genome-Scale Engineering in Yeast." *Nature Communications* 8: 15187.
- Siedentop, R., and K. Rosenthal. 2022. "Industrially Relevant Enzyme Cascades for Drug Synthesis and Their Ecological Assessment." *International Journal of Molecular Sciences* 23: 3605.
- Siedentop, R., M. Siska, N. Möller, et al. 2023. "Bayesian Optimization for An Atp-Regenerating In Vitro Enzyme Cascade." *Catalysts* 13: 468.
- Siedentop, R., C. Claaßen, D. Rother, S. Lütz, and K. Rosenthal. 2021. "Getting the Most Out of Enzyme Cascades: Strategies to Optimize In Vitro Multi-Enzymatic Reactions." *Catalysts* 11: 1183.
- Simić, S., E. Zukić, L. Schermund, K. Faber, C. K. Winkler, and W. Kroutil. 2022. "Shortening Synthetic Routes to Small Molecule Active Pharmaceutical Ingredients Employing Biocatalytic Methods." *Chemical Reviews* 122: 1052–1126.
- Skolpap, W., S. Nuchprayoon, J. M. Scharer, N. Grisdanurak, P. L. Douglas, and M. Moo-Young. 2008. "Fed-Batch Optimization of  $\alpha$ -amylase and Protease-Producing *Bacillus Subtilis* Using Genetic Algorithm and Particle Swarm Optimization." *Chemical Engineering Science* 63: 4090–4099.
- Soldatov, M. A., V. V. Butova, D. Pashkov, et al. 2021. "Self-Driving Laboratories for Development of New Functional Materials and Optimizing Known Reactions." *Nanomaterials* 11: 619.
- Steiner, S., J. Wolf, S. Glatzel, et al. 2019. "Organic Synthesis in a Modular Robotic System Driven by a Chemical Programming Language." *Science* 363: eaav2211.
- Subba Rao, C., T. Sathish, M. Mahalaxmi, G. Suvana Laxmi, R. Sreenivas Rao, and R. S. Prakasham. 2008. "Modelling and Optimization of Fermentation Factors for Enhancement of Alkaline Protease Production by Isolated *Bacillus Circulans* Using Feed-Forward Neural Network and Genetic Algorithm." *Journal of Applied Microbiology* 104: 889–898.

- Tachibana, R., K. Zhang, Z. Zou, S. Burgener, and T. R. Ward. 2023. "A Customized Bayesian Algorithm to Optimize Enzyme-Catalyzed Reactions." *ACS Sustainable Chemistry & Engineering* 11: 12336–12344.
- Tacias-Pascacio, V. G., B. Torrestiana-Sánchez, L. Dal Magro, et al. 2019. "Comparison of Acid, Basic and Enzymatic Catalysis on the Production of Biodiesel After RSM Optimization." *Renewable Energy* 135: 1–9.
- Thompson, M. P., I. Peñafiel, S. C. Cosgrove, and N. J. Turner. 2019. "Biocatalysis Using Immobilized Enzymes in Continuous Flow for the Synthesis of Fine Chemicals." *Organic Process Research & Development* 23: 9–18.
- Tom, G., S. P. Schmid, S. G. Baird, et al. 2024. *Chemical Reviews* 9633: 124.
- Vaddi, K., H. T. Chiang, and L. D. Pozzo. 2022. "Autonomous Retro-synthesis of Gold Nanoparticles via Spectral Shape Matching." *Digital Discovery* 1: 502–510.
- Vanneschi, L., and S. Silva. 2023. *Lectures on Intelligent Systems*, 45–103. Cham: Springer International Publishing.
- Veselinović, A., R. Nikolić, and G. Nikolić. 2012. "Application of Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) for Resolving Pyrogallol Autoxidation in Weakly Alkaline Aqueous Solutions." *Open Chemistry* 10: 1942–1948.
- Volk, A. A., R. W. Epps, and M. Abolhasani. 2021. "Accelerated Development of Colloidal Nanomaterials Enabled by Modular Microfluidic Reactors: Toward Autonomous Robotic Experimentation." *Advanced Materials* 33: 2004495.
- Wang, D., D. Tan, and L. Liu. 2018. "Particle Swarm Optimization Algorithm: An Overview." *Soft Computing* 22: 387–408.
- Wang, H., Z. Sun, Y. Qi, Y. Hu, Z. Ni, and C. Li. 2024. "Update Application of Enzyme in Food Processing, Preservation, and Detection." *Food Bioengineering* 3: 380–394.
- Williams, K., E. Bilsland, A. Sparkes, et al. 2015. "Cheaper Faster Drug Development Validated by the Repositioning of Drugs Against Neglected Tropical Diseases." *Journal of the Royal Society Interface* 12: 20141289.
- Wu, S., R. Snajdrova, J. C. Moore, K. Baldenius, and U. T. Bornscheuer. 2021. "Biocatalysis: Enzymatic Synthesis for Industrial Applications." *Angewandte Chemie International Edition* 60: 88–119.
- Xie, Y., C. Zhang, H. Deng, et al. 2021. "Accelerate Synthesis of Metal–Organic Frameworks by a Robotic Platform and Bayesian Optimization." *ACS Applied Materials & Interfaces* 13: 53485–53491.
- Yan, X., X. Zhang, H. Li, et al. 2024. "Engineering of Unspecific Peroxygenases Using a Superfolder-Green-Fluorescent-Protein-Mediated Secretion System in *Escherichia coli*." *JACS Au* 4: 1654–1663.
- Yang, W., F. Lu, and Y. Liu. 2023. "Recent Advances of Enzymes in the Food Industry." *Foods* 12: 4506.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section.

**Figure S1:** Workflow control panel in the GUI of the SDL software. The GUI was created using the package customtkinter in Python 3. **Figure S2:** Detailed overview of the SDL software showing software modules, repositories, devices, utilized and created files. **Figure S3:** Flow-chart for the utilized genetic algorithm (GA). **Figure S4:** Flow-chart for the utilized Particle Swarm Optimization (PSO) algorithm. **Figure S5:** Flow-chart of the utilized Bayesian Optimization (BO) algorithm. **Figure S6:** Flow-chart of the utilized Simulated Annealing (SA) algorithm. **Figure S7:** Flow-chart of the utilized random search (RS) algorithm. **Figure S8:** Flow-chart of the Response Surface Modelling (RSM). **Figure S9:** Posterior variance in the autonomous enzymatic reaction condition optimization experiments on the SDL for the enzyme-

substrate-pairings (a) UPO-ABTS, (b) HRP-ABTS, (c) HRP-Pyrogallol and (d) HRP-TMB. **Figure S10:** Parameter importances determined by Random Forest Regression in the autonomous enzymatic reaction condition optimization experiments on the SDL for the enzyme-substrate-pairings (a) UPO-ABTS, (b) HRP-ABTS, (c) HRP-Pyrogallol and (d) HRP-TMB. **Figure S11:** Visualization of the linear interpolation surrogate model and uncertainty proxies for the enzymatic activity landscape as a function of pH and temperature (other parameters fixed at optimal values:  $c\text{H}_2\text{O}_2 = 8.75 \text{ mM}$ ,  $c\text{Na}_2\text{SO}_4 = 120 \text{ mM}$ ,  $c_{\text{ACN}} = 0\% \text{ v/v}$ ). **Figure S12:** Robustness analysis of the interpolated landscape to experimental noise ( $\pm 1 \text{ SD}$ ). Six independent realizations (a)–(f) of the interpolated mean activity surface as a function of pH and temperature (other parameters fixed at optimal values:  $c\text{H}_2\text{O}_2 = 8.75 \text{ mM}$ ,  $c\text{Na}_2\text{SO}_4 = 120 \text{ mM}$ ,  $c_{\text{ACN}} = 0\% \text{ v/v}$ ) were generated by adding random noise within  $\pm 1$  standard deviation at each grid point. **Figure S13:** Robustness analysis of the interpolated landscape to increased experimental noise ( $\pm 2 \text{ SD}$ ). Six independent realizations (a)–(f) of the interpolated mean activity surface as a function of pH and temperature (other parameters fixed at optimal values:  $c\text{H}_2\text{O}_2 = 8.75 \text{ mM}$ ,  $c\text{Na}_2\text{SO}_4 = 120 \text{ mM}$ ,  $c_{\text{ACN}} = 0\% \text{ v/v}$ ) were generated by adding random noise within  $\pm 2$  standard deviation at each grid point. **Table S1:** Tested hyperparameters of the genetic algorithm (GA) in the in-silico algorithm screening. **Table S2:** Tested hyperparameters of the Particle Swarm Optimization (PSO) algorithm in the in-silico algorithm screening. **Table S3:** Tested hyperparameters of the Bayesian Optimization (BO) algorithm in the in-silico algorithm screening. **Table S4:** Tested hyperparameters of the Simulated Annealing (SA) algorithm in the in-silico algorithm screening.