

Self-Driving Lab for Solid-Phase Extraction Process Optimization and Application to Nucleic Acid Purification

Sebastian Putz, Jonathan Döttling, Tim Ballweg, Andre Tschöpe, Vitaly Biniyaminov, and Matthias Franzreb*

Sorptive bioprocesses are the basis for numerous biotechnological applications such as enzyme immobilization, biosensors, controlled drug delivery, water treatment, and molecular purification. Yet due to the complexity of these processes, their optimization is still time, labor, and cost-intensive. This research presents a flexible self-driving laboratory (SDL) designed for the accelerated development and optimization of solid-phase extraction processes. As a use case, the SDL was used to optimize a DNA purification process using silica magnetic beads. Through the integration of robotics, machine learning, and data-driven experimentation, the SDL demonstrates a highly accelerated process optimization with minimal human intervention. In the multistep purification approach, the system is able to optimize buffer compositions for DNA extraction from complex samples, demonstrating effectiveness in both conventional chaotropic salt-based methods and innovative chaotropic salt-free buffers. The study highlights the SDL's capability to autonomously refine process parameters, achieving significant enhancements in yield and purity of the product. This blueprint for future self-driving optimization of bioprocess parameters showcases the potential of autonomous systems to revolutionize biochemical process development, offering insights into scalable, environmentally sustainable, and cost-effective solutions.

and indispensable (bio-)chemical applications such as enzyme immobilization,^[1,2] biosensors,^[3,4] controlled drug delivery,^[5,6] water treatment,^[7,8] and molecular purification processes.^[9–11] For example, column chromatography has become indispensable in many industries, such as fine chemical and pharmaceutical production.^[12,13]

To optimize the binding capacity for target molecules, it is crucial to maximize the sorbent's surface area. Because of their extraordinarily high surface-to-volume ratio, nano- and microparticles are widely employed in adsorptive biochemical processes.^[14,15] Magnetic beads incorporating a superparamagnetic core of iron-oxides are a subclass of nano- and microparticles which have gained particular attention. Due to their superparamagnetic properties, magnetic beads are versatile tools, enabling simple and effective isolation and concentration of (bio)molecules by applying a magnetic field.^[16,17] In the simplest case, the only technical equipment needed for the separation process is a permanent magnet.


Due to the involvement of two phases and multiple process steps, there is a multitude of parameters that have to be optimized to achieve an efficient process. Simultaneous optimization for multiple targets, such as yield and purity in bioseparation processes, and the inference of the different process steps additionally increase the complexity of the task. Moreover, for biochemical processes, the variance of biological raw material can pose an additional challenge.^[18] Thus, the optimization of adsorptive biochemical processes remains highly time-, labor- and resource-intensive.^[19,20] The combination of automated high-throughput experimentation with data-driven and mechanistic modeling is currently widely used to accelerate the development of adsorptive bioprocesses such as chromatographical purification.^[21,22]

A very prominent biochemical adsorptive process is the purification of nucleic acids, DNA and/or RNA, the carriers of genetic information. Nucleic acid separation and purification are of immense importance across various fields, such as clinical diagnostics,^[23] therapeutics,^[24] nanotechnology,^[25] and forensics.^[26] Presently, solid-phase extraction (SPE) using silica as the adsorbent medium is the predominant method employed.^[16] Most commonly, the Boom method is used, in

1. Introduction

Adsorption is an interfacial process in which substances dissolved in a fluid adhere to the surface of a solid material, the sorbent. In the reversed process, called desorption, the adsorbed substances, called adsorbates, are released from the surface of the sorbent into the surrounding fluid phase or vacuum. These sorption processes are the basis for numerous widespread

S. Putz, J. Döttling, T. Ballweg, A. Tschöpe, V. Biniyaminov, M. Franzreb
Department for Bioengineering and Biosystems
Institute of Functional Interfaces (IFG)
Karlsruhe Institute of Technology (KIT)
76344 Eggenstein-Leopoldshafen, Germany
E-mail: matthias.franzreb@kit.edu

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/aisy.202400564>.

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/aisy.202400564

which DNA binds to silica at slightly acidic conditions (pH < 6.5) and high concentrations, up to 10 M, of chaotropic salts such as guanidine hydrochloride or guanidine isothiocyanate.^[27] For the binding step, the complex sample, containing the nucleic acids and various contaminants, such as proteins, carbohydrates, lipids, and cell debris is mixed with the binding buffer and exposed to the silica material, which is either in the form of magnetic beads or spin columns. After the binding step, multiple wash steps with high concentrations of alcohol are performed to remove unspecifically bound contaminants. Finally, the bound nucleic acids are released from the silica surface in the elution step. Although the process has been employed since the early 1990s, the exact adsorption mechanism of DNA/RNA to silica surfaces continues to be a topic of scientific debate.^[28]

Despite their widespread use, the DNA binding buffers containing high concentrations of chaotropic salts present significant drawbacks. First, residual chaotropic salts interfere with many downstream operations, such as polymerase chain reaction (PCR) due to their strongly denaturing effects.^[29] Furthermore, most chaotropic salts, such as guanidium thiocyanate and guanidium hydrochloride are classified as hazardous materials. Their toxicity poses a safety risk to lab personnel, necessitating protective equipment and careful handling. In response to these severe drawbacks, recent research has focused on exploring alternative binding buffer components to minimize or replace or chaotropic salts in nucleic acid binding buffers. For example, acetic acid (AA),^[28] potassium chloride (KCl),^[28] magnesium chloride (MgCl₂),^[30] polyethylene glycol (PEG),^[31] and various amino acids^[32,33] show great promise as binding buffer components. Binding simulations have suggested that the use of multiple binding agents can increase nucleic acid binding efficiency.^[33] However, it was not yet investigated, how these components perform for isolating nucleic acids from complex samples and whether synergistic effects between these components can be achieved by optimal concentration ratios for maximizing binding efficiency.

To solve complex optimization problems for adsorptive bio-processes such as nucleic acid purification, with minimal time, effort, and labor, so-called self-driving laboratories (SDLs) can be employed.^[34] SDLs combine modular lab automation via robotics and data-driven experiment planning via machine learning (ML) to create a closed-loop feedback system.^[35–37] These intelligent experimental platforms systematically execute cycles of experiments with parameters selected by ML algorithms to solve a task with a predefined target by the end-user. Thus, minimal human intervention is required, as the system autonomously performs all experimental and data analysis tasks. This autonomous operational framework minimizes human intervention by independently managing all experimental procedures and data analyses. Unlike traditional high-throughput screening (HTPS) systems that primarily focus on volume, SDLs enhance the efficiency of exploring experimental parameter spaces, thereby drastically conserving time and resources.^[38] The main advantages of the SDL approach over conventional experimental methods are presented in **Table 1**. While advantages 4 to 8 apply to automated experimental platforms in general, such as HTPS systems, points 1 to 3 are unique to SDLs.

SDLs have already proven to be extremely effective tools for the accelerated discovery, optimization, and synthesis in the field of complex organic compounds,^[38–42] nanomaterials,^[43–48] thin films^[49–52] as well as biomolecules and biosystems,^[53–59] among others.

In this research, we present our versatile SDL designed for process optimization of biochemical SPE methods, applicable for both magnetic and nonmagnetic solid-phase materials. As a proof-of-concept, this SDL was employed to optimize the buffer compositions in a multistep purification process of DNA from a complex sample. Initially, this involved using conventional binding buffers with chaotropic salts and silica-coated magnetic beads as the adsorbent. Subsequently, our system successfully optimized a nucleic acid binding buffer devoid of chaotropic salts, offering a promising alternative to the buffers commonly in use today.

Table 1. Advantages of SDLs over conventional experimental methods.

Advantage	Description
Efficient experiment planning	SDLs leverage ML methods to plan and optimize experiments efficiently, quickly identifying the most promising research paths and accelerating scientific discoveries. This ML-driven approach allows for adaptive experimentation, where the system learns from previous results to refine and focus subsequent experiments, speeding up the discovery process significantly. Compared to classical "brute-force" HTPS, this approach drastically reduces the amount of experiments.
Innovative discovery	By automating the exploration of experimental parameters, SDLs can uncover unexpected patterns or results that might not be apparent through traditional methods like design-of-experiments.
Resource efficiency	Over time, SDLs can reduce labor costs and minimize the use of reagents by optimizing experiments, making research more cost-effective. Due to lower number of experiments, the approach is much more resource-efficient than classical HTPS.
High throughput	Automation of experimentation and data processing enables more experiments to be conducted simultaneously and continuously, increasing overall throughput.
Improved reproducibility	Automation reduces human error and variation in experimental setup, leading to more reproducible results.
Remote operation	SDLs can be operated remotely, facilitating research continuity during disruptions (e.g., like pandemics) and enabling collaboration across different locations.
Data integration	SDLs integrate seamlessly with electronic lab notebooks and other data management systems, improving data traceability and accessibility for further analysis.
Safety	Automating hazardous processes can reduce the risk of accidents, enhancing safety in the laboratory environment.

2. Results and Discussion

2.1. SDL Platform Architecture and Workflow Overview

We developed a SDL platform specifically for the autonomous optimization of SPE processes. This system integrates eight laboratory devices on a custom-built workbench, as depicted schematically in **Figure 1a**. A photograph of the platform can be found in the Supporting Information (Figure S1, Supporting Information). The platform includes a liquid handling station (OT-2, Opentrons, USA) that offers pipetting, heating, shaking, and magnetic separation functionalities, facilitating various (bio-)chemical reactions and assays in well-plate format. A vacuum pump (MD 4 VARIO select, Vacuubrand, Germany), in conjunction with a custom vacuum manifold, enables the separation of nonmagnetic particles and solid materials from liquid phases. Magnetic particles with a minimum size of 100 nm can be separated using the automated magnetic separation module, which incorporates neodymium permanent magnets. Nonmagnetic particles are separated using commercially available 96-well filter plates, which feature pore sizes as small as 200 nm. The system's custom workbench houses storage, sorting, and stacking capabilities for well-plates, pipette tips, and liquid reservoirs. Additionally, a six-jointed robotic arm (UR5e, Universal Robots, Denmark) equipped with an adaptive gripper (Hand-E, Robotiq, Canada) and custom 3D-printed arms manages the transport and stacking of these items. For spectroscopic analysis, the platform includes a multimode plate-reader (Spark, Tecan, Switzerland) capable of conducting UV–vis spectroscopy, fluorescence, and luminescence measurements.

The well-plate-based part of the platform is connected to flow-based microcapillary system using a capillary positioning system (RotAXYS, Cetoni, Germany) to aspirate liquids from and dispense them into well-plates. Fluid transport in the capillary system is facilitated by precision syringe pumps (nemeSYS S, Cetoni, Germany) and a flow-selection valve module (6-port Qmix V Valve, Cetoni, Germany). Particle characterization is automated through a dynamic light scattering (DLS) and zeta potential analysis device (Zetasizer Nano ZS, Malvern Panalytical, UK), which is connected to the capillary system via a flow-through absorption cuvette (178.712-QS, Hellma, Germany). For the automated DNA purification process demonstrated in this work, only the liquid handling station, the robot arm, and the plate reader were utilized in the workflow. By additionally using the vacuum pump, the system was recently successfully employed for the automated determination of adsorption isotherms in batch incubation chromatography experiments.^[60] The modular design of the platform allows for easy expansion with additional devices, such as microreactors or other measurement modules, to accommodate specific application requirements.

Through the integration of a self-developed ML-reinforced control software the platform not only allows for automated but autonomous experimentation through intelligent planning of experimental cycles. The software utilizes a Python-based layered modular design, where each device and its functionalities are encapsulated into separate modules. This structure promotes easy updates and scalability through independent modifications. Each device is represented by a class in a separate module with

methods for the device functionalities. For example, initializing an object of the Tecan Spark class established the connection to the device. Device functionalities such as opening and closing the device or starting a measurement according to a configuration file can then be called with the methods of the class. The UR5e robot arm additionally uses lab-setup-specific modules that contain the exact joint positions and compound movements for transfer between the devices. New devices can be integrated with minimal configurational changes. These device classes for control and data exchange lay the foundation for automated experimentation. The software architecture is illustrated schematically in **Figure 1b**. In **Table 2**, the utilized application programming interface (API) and the communication protocol for the different devices are shown.

Every specific process, i.e., each type of experiment, is orchestrated by a superordinate main workflow script which coordinates the functionalities of the devices with the process-specific parameters in sequential order. For each experiment type templates of the main script and Opentrons protocols were created. Next to the device modules for automation, the main script also calls the modules for script generation, data processing, and optimization to enable autonomous experimentation. In each experimental cycle, the script generator first dynamically creates scripts for the liquid handling station and the plate reader based on the parameter values suggested by the optimization algorithm on the basis of process-specific templates. The data processing module contains functions for computation of target values and plot generation from raw data tailored to the measurement device, the sample number, and the experiment type. The optimization module contains functions to create an initial set of parameter combinations for the first experimental cycle and to generate new parameter combinations for following experimental cycles. Different initialization procedures and optimization algorithms (genetic, model-based, or hybrid; see Experimental Section) are included in this module. The database connector module automatically saves the generated data into a MySQL database on a local server.

Autonomous process optimization begins with the user defining the experiment. This definition includes selecting the experiment type, setting both fixed and variable process parameters, establishing variable limits, determining the objective and its value, setting configurations for the measurement devices, choosing the type of optimization algorithm, specifying algorithm parameters, and deciding on a maximum number of experimental cycles. Users navigate this setup via a custom-developed graphical user interface (GUI). Following this initial setup, the optimization algorithm generates the first set of parameter combinations for the variables. This triggers the automatic generation of a process workflow script based on a template of the main script for the according experiment type, which orchestrates the automated lab. The system then executes the first fully automated experimental cycle of the SPE process using these initial parameters. As the experiment proceeds, the measurement devices generate raw data, which the system automatically processes to perform calculations and generate plots. These data are stored in a MySQL database and analyzed by the optimization algorithm to determine optimized parameter values for subsequent experimental cycles. The included algorithms are either based on fitting an empirical model and

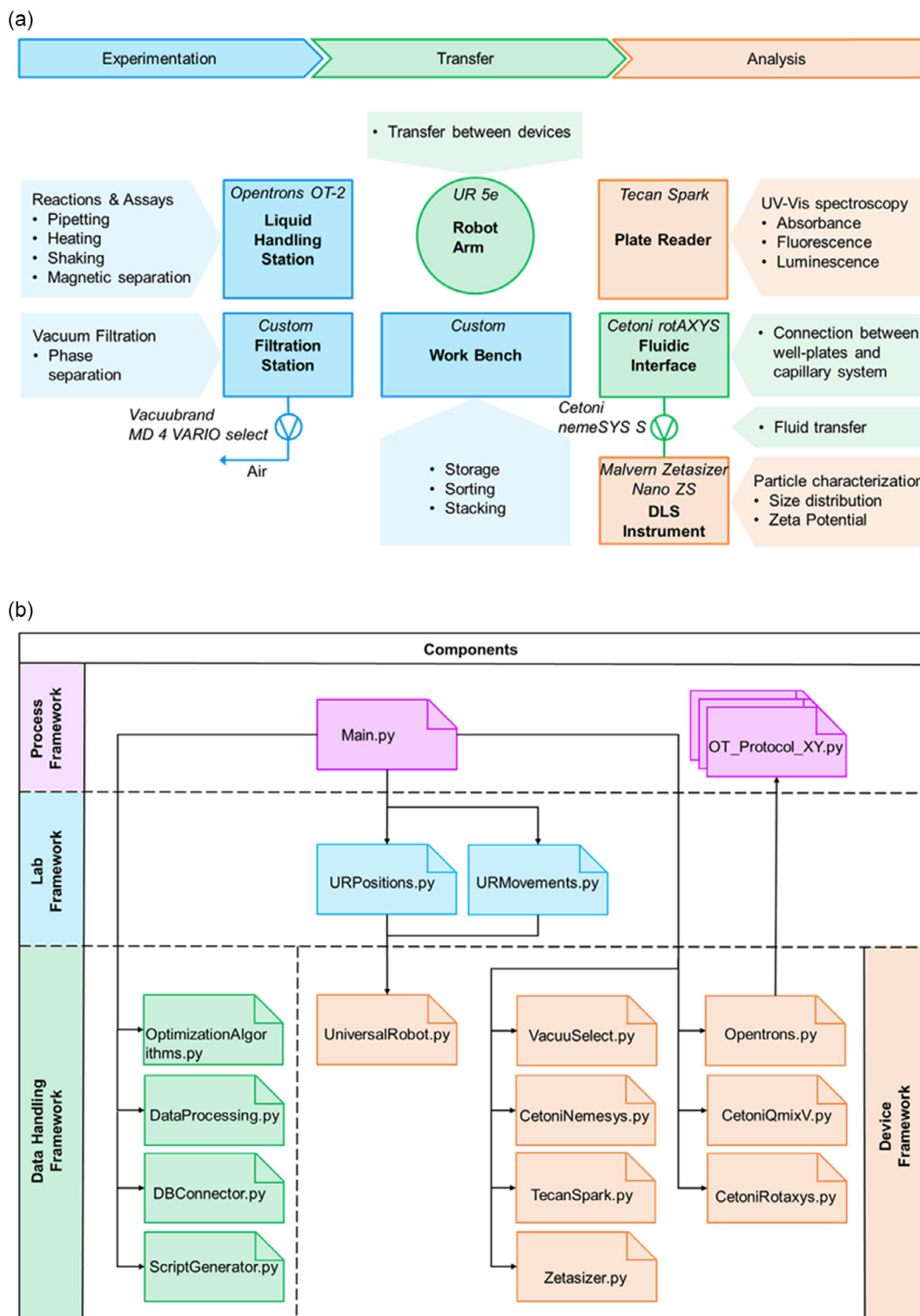


Figure 1. a) Hardware and b) software schematic of the developed SDL for SPE process optimization.

sampling points around the predicted optimum, an evolutionary principle as in a genetic algorithm (GA), or a hybrid of both (see Experimental Section). The system continues to repeat experimental cycles, adjusting parameters based on the

optimization algorithm's predictions, until it achieves the predefined objective or reaches the maximum number of cycles allowed. Upon completion, the system undertakes a final evaluation of all collected data, performs statistical model-fitting, and

Table 2. Devices integrated into the SDL with the API and communication protocol used for automation.

Device type	Manufacturer model	API	Communication protocol
Liquid handling station	Opentrons OT-2	HTTP ^[72]	TCP/IP
6-Joint Robot Arm	Universal Robots UR5e	Primary Interface ^[73] (Port 30001)	TCP/IP
Vacuum pump	VacuuSelect MD Vario 4	Modbus ^[74]	TCP/IP
UV-vis spectroscopy multimode plate reader	Tecan Spark	SiLA 2 ^[75]	gRPC
Positioning system for well-plate sampling	Cetoni RotAXYS	Cetoni SDK ^[76] (Python Integration)	CANopen
Syringe pumps	Cetoni nemeSYS	Cetoni SDK (Python Integration)	CANopen
Valve module	Cetoni Qmix V	Cetoni SDK (Python Integration)	CANopen
DLS instrument	Malvern Zetasizer Nano ZS	Custom GUI	TCP/IP

generates plots. All plots shown in this article were automatically generated by the data processing software module. A flowchart illustrating the autonomous process optimization is shown in **Figure 2**.

2.2. Autonomous Buffer Optimization for DNA Purification Processes

For a first proof-of-concept, the system was given the task to optimize the composition of the binding, wash, and elution buffer of a DNA purification process from a complex sample by varying the concentrations of Gu-HCl, EtOH, and Tris. In the first set of experiments, a GA (see Experimental Section) was used to iteratively optimize the buffer compositions.

The fitness function was defined as the product of loading and purity, with the purity being defined as the ratio of absorbance at 260 nm and 280 nm wavelength (A_{260}/A_{280}) in the purified sample (Equation (4)). The objective for the process optimization was a fitness value of $4 \mu\text{g mg}^{-1}$. In **Figure 3a**, the loading of the beads with DNA (calculated from the eluate), purity levels, and the corresponding value of the target function are shown for the best parameter combination of each experimental cycle. In the first generation, which was created randomly by the algorithm, a maximal fitness value of 2.35 was achieved with a loading of $1.20 \pm 0.14 \mu\text{g mg}^{-1}$ and a purity of 1.98 ± 0.09 . Throughout the experimental cycles, there was a consistent increase in maximal DNA loading, while the purity remained relatively stable, oscillating between 1.9 and 2.0. By the sixth experimental cycle, the objective of the optimization was met. The best buffer composition achieved a fitness value of 4.10, with a DNA loading of $1.98 \pm 0.10 \mu\text{g mg}^{-1}$ and a purity of 2.07 ± 0.07 . The normalized parameter values of the best combinations per generation alongside their respective loading are shown in **Figure 3b**. The best parameter combination in the initial generation was 5.20 M Gu-HCl, 65.0% EtOH, and 8.0 mM Tris. The overall best parameter combination in the sixth generation was 5.89 M Gu-HCl, 95.7% EtOH, and 24.4 mM Tris. There was no clear

trend of the Gu-HCl and Tris concentration throughout the experiment cycles, while the EtOH concentration of the best parameter combination continuously increased. In the Supporting Information, the data for all investigated parameter combinations in each experimental cycle are included (**Figure S2**, Supporting Information).

In this study, after achieving the specified optimization objective, a statistical model was automatically fitted to all collected data points to enhance process knowledge. A quadratic interaction model was selected to elucidate the dependence of DNA loading onto the beads on the concentrations of the buffer components. As illustrated in **Figure 3c**, the model accurately predicts the binding capacity, yielding an R^2 value of 0.90. Based on the fitted model, surface plots (**Figure 3d–f**) are generated to visually illustrate the influence of each parameter on the target. The surface plots revealed that optimal DNA loading on the beads is achieved with a high concentration of Gu-HCl in the binding buffer, a high concentration of EtOH in the wash buffer, and a low concentration of Tris in the elution buffer. The model predicts a maximal loading of $2.47 \mu\text{g mg}^{-1}$ at buffer concentrations of 6.0 M Gu-HCl, 100.0% EtOH, and 5.0 mM Tris.

The results demonstrate that the system successfully autonomously optimized the DNA purification process until the predefined objective was reached. Notably, the optimized buffer compositions closely mirror those used in the well-established Boom method, supporting the validity of both the optimized conditions and the statistical model predictions in line with prior research and established scientific theories. In the Boom method,^[27] DNA binds to silica at high concentration of chaotropic salts and at slightly acidic pH. Under high concentrations of chaotropic salts, adsorption is hypothesized to be driven by the hydrophobic effect due to dehydration of nucleic acids and the silica surface as well as salt bridging.^[61–64] The slightly acidic environment ($\text{pH} < 6.5$) decreases the electrostatic repulsion between DNA and silica, which are both negatively charged and thus facilitates adsorption.^[65] Following the binding step, multiple wash steps with high concentrations of alcohols, typically ethanol or isopropanol are performed. These steps are

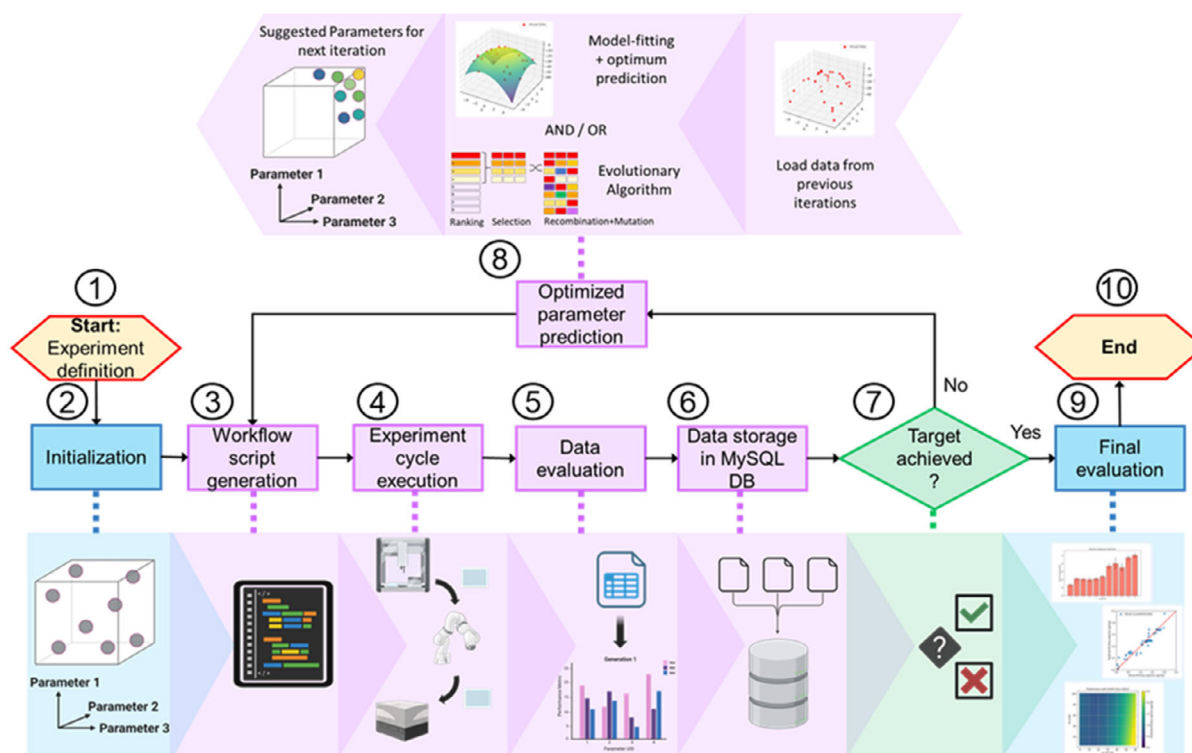


Figure 2. Flowchart of the autonomous SPE process optimization workflow. (1) The workflow starts with the experiment definition, where the experiment type, the optimization algorithm, the fixed parameters, the variable process parameters subject to optimization, and the optimization target are defined. (2) Based on the experiment definition a first set of parameter combinations is generated either random or according to an experimental design. (3) A dedicated software module subsequently generates the configuration files for each device according to the parameter combinations and the fixed parameters according to templates. (4) These are called from a subordinate workflow script for process orchestration in the adjacent execution of the experimental cycle. (5) The raw data generated during the experimental cycle is automatically evaluated and the target values for each parameter combination are calculated. (6) All generated data are then stored in a MySQL database on a local server. (7) The optimization algorithm then checks if the optimization target is reached, i.e., at least one parameter combination reaches the target fitness value. (8) If the target is not reached, the algorithm predicts optimized parameter combinations for an adjacent experiment cycle and repeats steps (3)–(8) until either the target is achieved or the maximum number of experimental cycles is reached. (9) Once the target is reached, a final data evaluation and plot generation are performed and the workflow is finished (10).

critical to remove unspecifically bound contaminants and excess salt from the binding buffer while precipitating the DNA onto the silica surface. High alcohol concentrations prevent the DNA from being eluted from the beads. Finally, the bound nucleic acids are released from the silica surface in the elution step into a low ionic strength buffer with slightly alkaline pH, typically 5 or 10 mM Tris or Tris-EDTA buffer with pH 8.5. The combination of low ionic strength and the slightly alkaline pH increase the electrostatic repulsion between DNA and silica and aids the elution of DNA from the beads.^[65,66]

The optimization algorithm employs elitism and prohibits mutations in the best parameter combinations during the recombination process, ensuring that each experimental cycle either improves or maintains the highest fitness value. The effectiveness of these features is evidenced by the continuous improvement in DNA loading and, consequently, the fitness function values throughout the experiments. The consistently high purity observed across the experimental cycles can be attributed to the selective DNA binding properties of the beads used, which appear to be largely independent of the buffer compositions.

This suggests that the beads effectively differentiate between DNA and other molecular species, ensuring a high purity of the extracted DNA regardless of variations in the buffer conditions.

The GA demonstrated its capability to optimize the buffer composition without requiring mechanistic input. This feature makes it particularly useful for optimizing processes where key influencing factors are known, but the details of their relationships with the target or their interactions with each other are not. However, the inherent trial-and-error nature of the algorithm means that the optimization process is not directed. Consequently, potentially favorable values of a parameter might be overlooked or discarded if altering another parameter leads to an overall increase in the fitness function value. For instance, in this series of experiments, the concentrations of Gu-HCl and Tris in the optimal combinations did not show a consistent pattern across the experimental cycles; rather, they exhibited fluctuations, illustrating the nonlinear and somewhat unpredictable behavior typical of GA optimizations.^[67]

To address the limitations of the GA and potentially accelerate the optimization process, we tested an alternative algorithm that

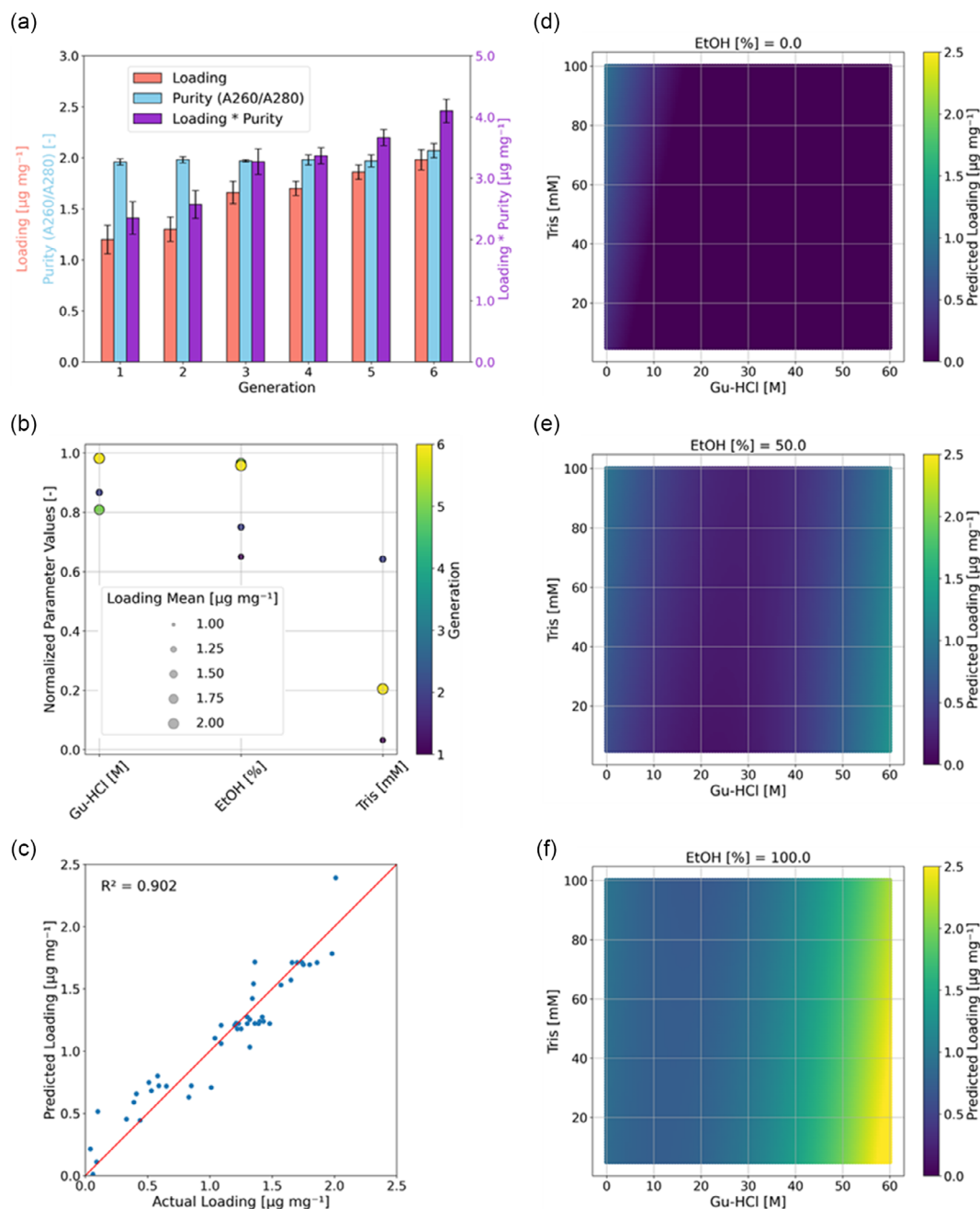


Figure 3. Results of the autonomous DNA purification process optimization using conventional buffer components and a GA. a) Loading, purity, and fitness function (loading*purity) value of the best parameter combination for each experimental cycle. b) Normalized parameter values and achieved loading of the best parameter combinations for each experimental cycle. Parameter values are normalized by the specified upper concentration limit. c) Experimentally determined loading versus loading predicted by the fitted quadratic interaction model. d–f) Surface plots of predicted loading at varying Gu-HCl and Tris concentration at fixed (0%, 50%, 100%) EtOH concentrations.

dynamically fits a statistical model to all available data points to predict the optimal parameter values for subsequent experimental cycles. We employed a quadratic interaction model for its adaptability across a variety of processes and its demonstrated accuracy with the data from the GA. To ensure comparability, the process optimization using the model-based algorithm

(MBA) commenced with the same initial parameter combinations as those used with the GA. The MBA achieved convergence by the second experimental cycle, effectively predicting the optimum from just eight data points gathered in the initial cycle. As shown in **Figure 4a**, the maximal fitness value in the first generation was 2.33 with a loading of $1.19 \pm 0.17 \mu\text{g mg}^{-1}$ and a purity

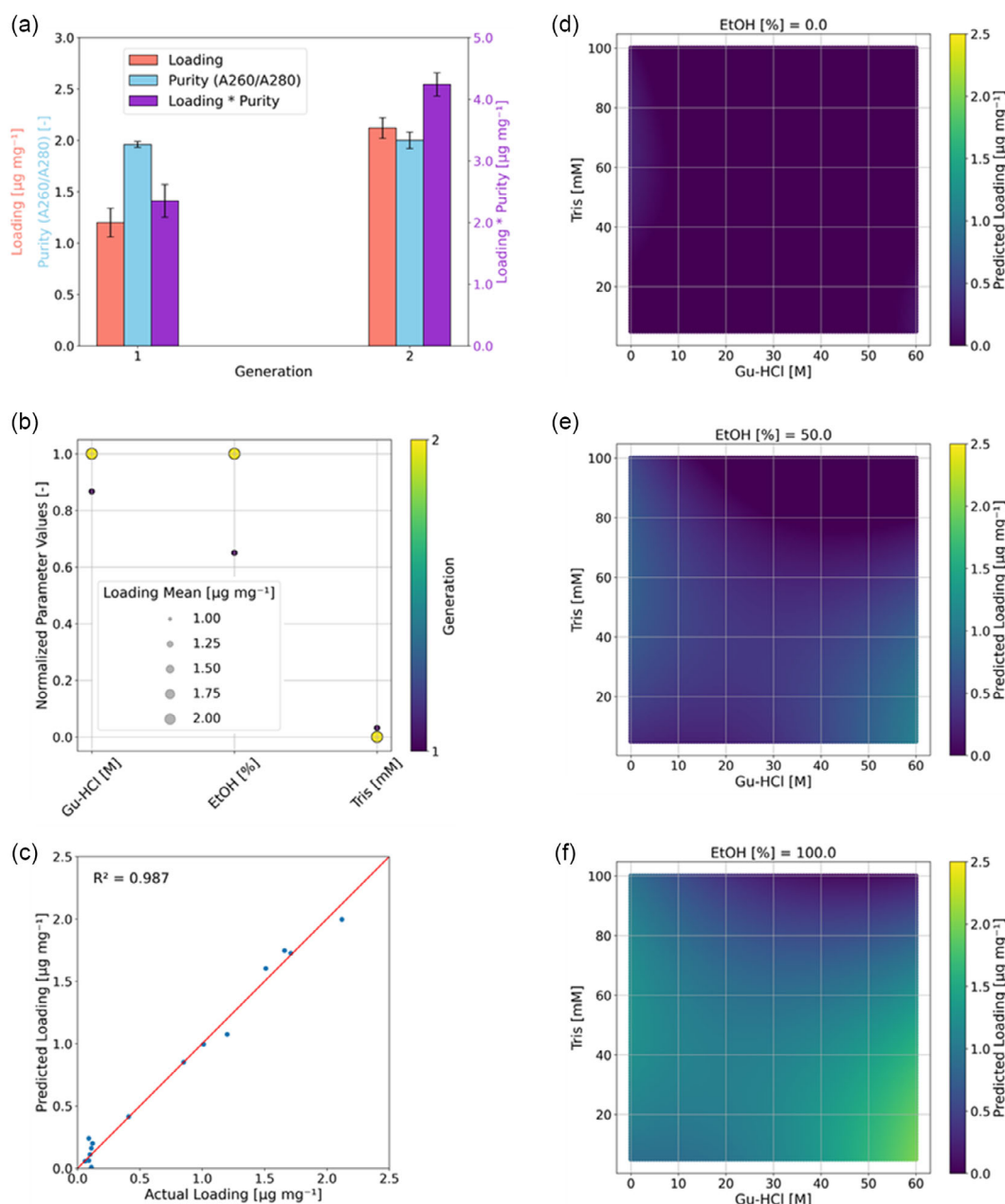


Figure 4. Results of the autonomous DNA purification process optimization using conventional buffer components and a GA. a) Loading, purity, and fitness function (loading*purity) value of the best parameter combination for each experimental cycle. b) Normalized parameter values and achieved loading of the best parameter combinations for each experimental cycle. Parameter values are normalized by the specified upper concentration limit. c) Experimentally determined loading versus predicted loading by the fitted quadratic interaction model. d–f) Surface plots of predicted loading at varying Gu-HCl and Tris concentration at fixed (0%, 50%, 100%) EtOH concentrations.

of 1.96 ± 0.03 . By the second generation, the objective was reached with a maximal fitness value of 4.24, corresponding to a loading of $2.12 \pm 0.10 \mu\text{g mg}^{-1}$ and a purity of 2.00 ± 0.08 . The best parameter combination for the first generation remained consistent with the GA, featuring 5.20 M Gu-HCl, 65.0% EtOH, and 8.0 mM Tris, as shown in Figure 4b. The overall best parameter combination in the second generation that reached the target fitness value comprised 6.00 M Gu-HCl,

100% EtOH, and 5 mM Tris. The data for all tested parameter combinations are shown in Figure S3 (Supporting Information). To verify the reproducibility of the experiments performed by the robotic setup, the initial experimental cycle was repeated a third time. The results shown in Supporting Information Figure S4 reveal a high reproducibility of the automated experiments, notably the standard deviation of the determined loading is $\leq 0.1 \mu\text{g mg}^{-1}$ between the same parameter

combinations in replicated cycles. The final quadratic-interaction model, following the second generation, maintained an excellent fit with an $R^2 = 0.987$, as seen in Figure 4c. The corresponding surface plots, illustrating the DNA loading at variable concentrations of Gu-HCl, Tris, and EtOH, are displayed in Figure 4d–f. Consistent with the previous results, predictions indicate that DNA loading increases with higher concentrations of Gu-HCl and EtOH and decreases with higher Tris concentrations. The maximal predicted loading, achieved with 6.00 M Gu-HCl, 100% EtOH, and 5 mM Tris, was 19.0% less than that predicted by the model fitted to data from the GA. However, the actual maximal loading reached with the best parameter combination using the MBA was 7.0% higher. This improvement can be attributed to slightly higher concentrations of Gu-HCl and EtOH, and a lower concentration of Tris compared to the optimal combination identified by the GA, leading to enhanced DNA binding, reduced wash loss, and increased elution yield.

In conclusion, the MBA effectively addresses the process optimization problem involving three variables, achieving convergence in just two experimental cycles—significantly fewer than the six required by the GA. Moreover, it identified a more optimal parameter combination that yielded a higher fitness value and enhanced DNA loading. The final models constructed from the generated data consistently demonstrated similar qualitative influences of the parameters. To maximize the yield of purified DNA, the concentrations of Gu-HCl and EtOH should be maximized, while keeping the Tris concentration minimized, aligning with findings from previous studies. Thus, the proposed quadratic-interaction model proved to be a simple but accurate model for describing the concentration influences on the yield of purified DNA. This suggests its potential as a viable algorithm in autonomous process optimization tasks, particularly when dealing with a smaller number of process variables. This is important because the number of coefficients in the model increases quadratically with the number of variables, which could lead to overfitting when data points are scarce. The maximally achieved loading is 26.5% of the binding capacity specified by the manufacturer. This could be attributed to the occupation of binding sites with unspecifically bound contaminants from the sample mixture. Especially bovine serum albumin (BSA), which is present in a 100-fold higher concentration than DNA in the sample, is known to have a strong interaction with silica surfaces, especially at low pH and high ionic strength.^[68,69] In addition, the DNA loading in this study is determined from the amount of eluted DNA. If the loading is determined using the DNA concentration in the supernatant from the binding step, a higher loading might be observed due to losses in the subsequent wash steps and only partial desorption of DNA during elution.

2.3. Autonomous Optimization of Alternative Binding Buffer

To demonstrate the capability of the autonomous system to accelerate scientific innovation, it was employed to develop an alternative binding buffer free from chaotropic salts. Chaotropic salts were excluded due to their drawbacks, including high chemical consumption, toxicity, and inhibition of downstream applications, among others. Instead, a mixture of alternative

compounds, namely, acetate, L-histidine, L-arginine, potassium chloride, magnesium chloride, and PEG-8000 was utilized to facilitate DNA adsorption to the silica magnetic beads. The system was allowed to arbitrarily mix these compounds within specified concentration limits (see Experimental Section) in 20 mM Tris-HCl buffer at pH 5 to limit the total chemical consumption. For the optimization process, a novel hybrid algorithm combining genetic and model-based elements was developed. This method leverages the exploratory capabilities of the GA alongside the predictive accuracy of the model-based approach to iteratively refine the buffer compositions. The wash and elution buffers were maintained at their previously optimized compositions to ensure consistency and to isolate the effects of the novel binding buffer components on the purification process. The objective for this optimization was defined as achieving a minimum product of loading and purity of $2 \mu\text{g mg}^{-1}$.

This target was successfully reached by the 11th experimental cycle, as shown in Figure 5a. The progression from the initial experimental cycle, which began with randomly selected parameters, to the final cycle demonstrated a significant improvement in performance. In the first experimental cycle the maximal loading was $0.27 \pm 0.02 \mu\text{g mg}^{-1}$. From cycle 2 to 6, the maximal loading remained relatively constant between 0.40 ± 0.02 and $0.48 \pm 0.02 \mu\text{g mg}^{-1}$ and then sharply increased to $0.74 \pm 0.06 \mu\text{g mg}^{-1}$ in the 7th cycle. Then the loading continuously increased until the target was reached in the 11th cycle with a loading of $1.09 \pm 0.06 \mu\text{g mg}^{-1}$ and a corresponding fitness value of 2.09. The only exception is cycle 9 with a small drop in the maximal loading, which however is not statistically significant ($p = 0.366$ in t-test). Throughout these cycles, the purity of the best parameter combinations remained consistent, fluctuating between 1.79 and 2.00. In Figure 5b, the normalized parameter values of the best parameter combinations in each generation are shown. The highest loadings were achieved when only the two amino acids L-histidine and L-arginine were present in the buffer. It was found that optimal results were obtained when both amino acids were present at concentrations below their maximum allowable limits. The optimal binding buffer composition was determined to be 65.8 mM L-histidine and 41.0 mM L-arginine in 20 mM Tris-HCl at pH 5. In the Supporting Information, the data for all investigated parameter combinations is shown in Figure S5.1 and S5.2 (Supporting Information). The final quadratic-interaction model has an $R^2 = 0.593$, indicating a moderate level of explanatory power (Figure 5c). Thus, the commonly used quadratic interaction model appears unable to accurately capture the underlying relationship between the parameters and the loading. Some effects may not be correctly represented by the parameter influences included in the model. Moreover, the quadratic interaction model assumes a continuous input-output relationship, which may not be appropriate for this problem, as we suspect discontinuities with certain combinations of compounds in the binding buffer. These discontinuities could potentially arise from factors such as synergistic or antagonistic interactions between compounds, as well as the crossing of solubility boundaries. In contrast, nonparametric models, such as Random Forest Regressors, Gaussian Processes, or Decision Trees, may be better suited. Indeed, a Random Forest Regression model achieved the best performance, with $R^2 = 0.933$ when fitted to all data and

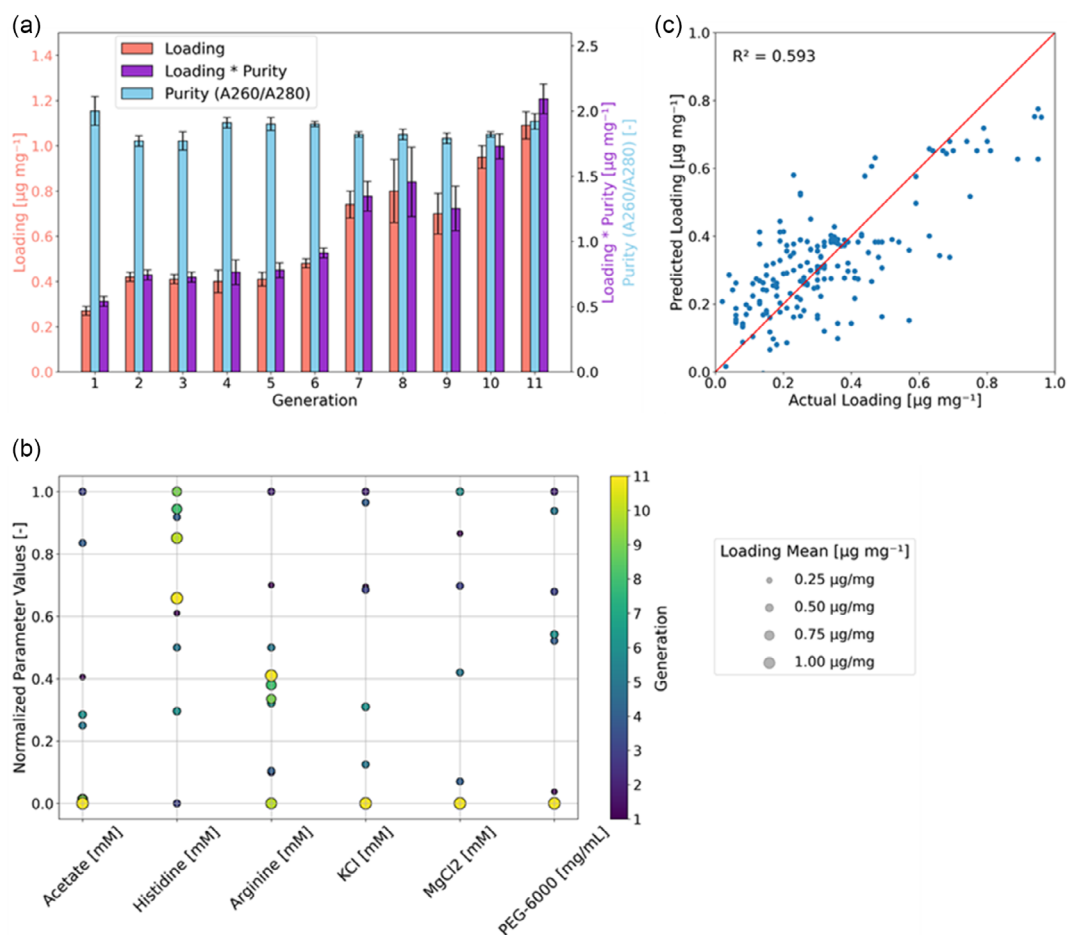


Figure 5. Results of the autonomous chaotropic salt-free DNA binding buffer optimization using a hybrid algorithm. a) Loading, purity, and fitness function (loading*purity) value of the best parameter combination for each experimental cycle. b) Normalized parameter values and achieved loading of the best parameter combinations for each experimental cycle. Parameter values are normalized by the specified upper concentration limit. c) Experimentally determined loading versus predicted loading by the fitted quadratic interaction model.

$R^2 = 0.629$ on the test set when using an 80/20 training/test split. For comparison, the quadratic interaction model yielded $R^2 = 0.593$ on all data and $R^2 = 0.348$ on the test set. A comparison of the model predictions is shown in Supporting Information in Figure S6.

In summary, the development and optimization of an alternative binding buffer were demonstrated successfully. The buffer achieved 51.4% of the loading capacity compared to the best chaotrope-based buffer while using only 3.6% of the chemicals by weight and maintaining equivalent purity levels. It could thus be considered a more economically and ecologically sustainable alternative to conventional chaotrope-based buffers, particularly in applications where maximal recovery is not crucial, such as when subsequent processes include PCR amplification. Previous studies have already suggested that buffers containing positively charged and polar neutral amino acids, particularly L-arginine and L-histidine, show promise as alternatives to chaotropic salt-based methods for solid-phase DNA extraction.^[32] However, these studies did not explore the performance of these buffers with complex samples containing impurities or

investigate potential synergistic effects between different amino acids. Bag et al. conducted a study precisely simulating the DNA adsorption to silica and unveiling the underlying cooperative adsorption mechanism.^[33] It was ultimately concluded that using multiple binding agents could enhance DNA binding affinity to silica surfaces. Our study extends these findings by demonstrating that amino acid buffers can selectively bind DNA from complex samples with high purity. Moreover, we observed synergistic effects between L-histidine and L-arginine that maximize DNA loading at their optimal concentrations, providing a compelling case for their combined use in practical applications. A possible explanation is a favorable combination of different binding mechanisms leading to more stable interactions. While both amino acids facilitate cooperative adsorption serving as positively charged mediator between the negatively charged DNA and silica under binding conditions, L-arginine additionally exhibits a strong interaction to the guanidine bases of DNA, especially through hydrogen bonding.^[70] Further studies focusing on simulating the interaction between DNA and silica under these conditions would be required for a definitive explanation.

2.4. Comparison of Conventional and Alternative Binding Buffer

For additional analysis of the DNA purified using both the conventional and alternative binding buffers, the conductivity of the wash buffer and elution supernatants was assessed with a conductivity cell (Table 2). Conductivity levels in the wash and elution supernatants were notably higher when the conventional binding buffer was used, implying a carry-over of chaotropic salts from the initial binding step. For instance, the conductivity of the eluate was $865.0 \pm 18.1 \mu\text{S}$ when using the conventional binding buffer and $285.9 \pm 10.0 \mu\text{S}$ when using the alternative binding buffer. Using a calibration curve (Figure S7, Supporting Information), the residual concentration of Gu-HCl in the elution buffer was estimated assuming the increased conductivity is solely caused by salt carry-over. The remaining Gu-HCl concentration in the eluate is $\approx 6 \text{ mM}$ which is 0.15% of the initial concentration during the binding step. In contrast, in the alternative process, the increased conductivity in the eluate compared to the pure elution buffer is likely attributable to the presence of amino acids from the binding buffer. Chaotropic salts, such as the utilized Gu-HCl are known to be PCR inhibitors due their strong denaturing effect and thus may be problematic for downstream applications, e.g., in diagnostics, forensics, or biosensing. PCR inhibition with increased Ct-values in qPCR is observable at Gu-HCl concentrations exceeding 100 mM.^[71] However, the extent of chaotropic salts carry-over can vary when using other protocols, strongly increasing when more supernatant remains in each separation step due to for example error-prone manual liquid handling, unfavorable setup geometry, or nonoptimized automated protocols. In addition, many protocols incorporate chaotropic salts in the wash buffer to reduce DNA loss during wash steps, which further increases salt carry-over into the eluate. These risks can be completely mitigated using the alternative amino acid binding buffer. Moreover, the alternative buffer produces 96.7% less waste of chemicals and uses no hazardous compounds, while having 89.4% lower cost. The main disadvantage is a 48.6% lower DNA yield, which however in applications involving downstream amplification techniques may not be problematic. To conclusively validate the efficacy of this novel buffer for DNA purification, it is required to extend testing to actual biological samples, incorporating an initial cell lysis step, as the present study employed a synthetic complex sample.

3. Conclusion

In conclusion, this study successfully demonstrated a SDL designed to optimize SPE processes for bioseparations. The developed autonomous system seamlessly integrates robotics, advanced ML algorithms, and an efficient laboratory workflow through a modular layered software framework. Specifically, in the presented use case of silica magnetic bead-based DNA purification, the system has proven to be a powerful tool for rapid and effective process optimization. By employing this system, we demonstrated that substantial improvements in process efficiency can be achieved with minimal human intervention.

The system was able to find optimal buffer compositions for all process steps in a multivariate process optimization aimed at maximizing yield and purity of the extracted DNA from a

complex sample. Both a GA and a model-fitting algorithm based on a quadratic interaction model have proven to be very effective tools in this use case. Based on these two algorithms, we developed a hybrid algorithm. This algorithm combines the exploratory strength of GAs with the predictive accuracy of dynamically updated models fitted to all existing data to guide parameter adjustments. Employing the hybrid algorithm, a binding buffer free from chaotropic salts was optimized by the system to avoid the drawbacks of these compounds. The final optimized buffer is composed of 65.8 mM L-histidine and 41.0 mM L-arginine in 20 mM Tris-HCl at pH 5. With this buffer 51.4% yield, equal purity measured by A_{260}/A_{280} and a lower salt contamination compared to the conventional binding buffer were accomplished. At the same time, the novel binding buffer achieves 96.7% less chemical consumption and 89.4% lower cost while avoiding toxic and potentially inhibitory compounds. These results highlight the capability of the SDL for the highly accelerated development of efficient, environmentally safe, and cost-effective processes.

Most importantly the system is highly flexible and can be employed to optimize SPEs across a diverse range of adsorbents and adsorbates. For example, the system can be used to optimize buffer compositions for chromatographic processes with batch adsorption experiments or to optimize enzyme immobilization processes. To further expand the capabilities of the autonomous laboratory, additional devices for experimentation and analysis, such as a robotic centrifuge or a HPLC system could be integrated into the modular platform. Future work will also focus on refining the optimization algorithms and developing suitable models, especially for processes with a large number of parameters.

Our research offers a blueprint for incorporating autonomous platforms into diverse bioprocessing applications, catalyzing major improvements in process development. The extremely potent synergy of artificial intelligence and automation is transforming whole industries already today. Through continued progress in artificial intelligence, robotic automation, and high-throughput instrumentation, intelligent autonomous systems will become even more powerful and indispensable tools in scientific discovery and industrial production.

4. Experimental Section

Materials: All chemicals and reagents were used as received without additional purification. Ethanol (EtOH) (Ph. Eur., >99.5%), guanidium hydrochloride (Gu-HCl), L-arginine (Arg), acetic acid (AA), magnesium chloride (MgCl_2), potassium chloride (KCl), polyethylene glycol 8000 (PEG-8000), Tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), and tris(hydroxymethyl)aminomethane (Tris) were purchased from VWR Chemicals (Germany). L-histidine hydrochloride monohydrate (His) was obtained from PanReac AppliChem (Germany). Sodium hydroxide (NaOH, 1.0 N), hydrochloric acid (HCl, 36.5–38%), DNA sodium salt from salmon testes, BSA, and cholesterol were obtained from Sigma Aldrich (Germany). SeraSil-Mag 700 silica magnetic beads ($d = 700 \text{ nm}$) were purchased from Cytiva (USA).

Automated DNA Purification and Quantification: For automated purification of DNA using magnetic beads, an optimized protocol for the robotic liquid handling workstation (OT-2, Opentrons, USA) was developed. Prior to each experimental cycle, the user exchanges nonautomatable labware and fills up the required reagent reservoirs for the buffer

solutions, the sample, and the magnetic beads. Once the user confirms correct deck setup, the workflow continues with the execution of the liquid handling protocol.

The protocol initiates by calculating the required mixture of different buffers based on the input concentrations specified for the experiment. These concentrations c_i are then transformed into the exact volumes of stock solutions, $V_{\text{stock},i}$, and diluents V_{diluent} needed for the procedure, using the following equations Equation (1) and (2) and the buffers are mixed in a designated well-plate.

$$V_{\text{stock},i} = \frac{c_i}{c_{\text{stock},i}} V_{\text{total}} \quad (1)$$

$$V_{\text{diluent}} = V_{\text{total}} - \sum_{i=1}^{i=n} V_{\text{stock},i} \quad (2)$$

Here, $c_{\text{stock},i}$ is the concentration of the stock solution of substance i and V_{total} is the total volume of the buffer.

After buffer mixing, 100 μL each of the binding buffer was transferred from the buffer mixing plate into a 96-well PCR plate positioned on the magnet separator module of the OT-2. Subsequently, the magnetic bead solution was homogenized in the reagent reservoir by pipetting and then 20 μL of the particle suspension was immediately added to the binding buffer in the PCR plate to ensure precise and consistent dosing. Lastly, 30 μL of the DNA-containing sample solution was added and the mixture was incubated for 10 min in the binding step. During the binding step, the solution was agitated every 2 min by pipetting to ensure homogeneity and enhance mass transfer. After the incubation, the magnet separator was activated to separate the magnetic beads from the solution. Once the beads were fully separated, the supernatant was carefully removed by slow pipetting (5 $\mu\text{L s}^{-1}$) and transferred to the liquid waste. A separation time of 5 min was determined to be optimal for all separation steps. Following the binding step, the system performs two identical wash steps to remove excess salt and unspecifically bound contaminants. For this purpose, the magnetic module was deactivated to release the magnetic beads, and 100 μL of the wash buffer was added. The beads were then resuspended by thorough pipetting. After 2 min of incubation in the wash buffer, the magnetic module was reactivated again to separate the beads and the supernatant was transferred to the liquid waste. Lastly, an elution step was performed to release the bound DNA. After deactivating the magnetic module, 150 μL of elution buffer was added and the beads were resuspended by thorough pipetting. The beads were incubated for 10 min with intermittent agitation every 2 min through pipetting. Finally, the beads were separated by the magnetic module and 75 μL of the eluate from each well was transferred to a 96-well UV-plate for UV-vis spectroscopy analysis. To procure 150 μL per well for the absorption measurement, eluates from two wells belonging to the same parameter combination were pooled.

The UV-plate containing the eluate was transferred into the UV-vis plate-reader (Spark, Tecan, Switzerland) by the robotic arm (UR 5e, Universal Robots, Denmark). The absorbance was then measured at 260 and 280 nm wavelength. The concentration of DNA in the elution buffer is quantified using a calibration curve established from readings at 260 nm. This calibration curve was generated in a 5 mM Tris buffer at pH 8.5, which is also employed as the blank for accurate measurements. From the DNA concentration in the elution buffer, the DNA loading of the magnetic beads was calculated using Equation (3):

$$q = \frac{c_{\text{DNA}}}{c_{\text{beads}}} \quad (3)$$

Here, c_{DNA} is the concentration of DNA and c_{beads} is the concentration of magnetic beads in the elution buffer.

To determine the purity P of the extracted DNA, the ratio of the blank-corrected absorbance between 260 and 280 nm (A_{260}/A_{280}) is calculated according to Equation (4):

$$\text{Purity} = P = A_{260}/A_{280} = \frac{A_{260 \text{ nm, sample}} - A_{260 \text{ nm, blank}}}{A_{280 \text{ nm, sample}} - A_{280 \text{ nm, blank}}} \quad (4)$$

Autonomous Process Optimization of DNA Purification with Silica Magnetic Beads: As the first proof-of-concept for capability of autonomous process optimization, the system was tasked with optimizing the binding, wash, and elution buffer in the automated DNA purification process. Key variables selected for optimization included the concentration of Gu-HCl in the binding buffer, EtOH in the wash buffer, and Tris in the elution buffer. The concentrations were freely adjustable in the limits of 0–6 M Gu-HCl, 0–100% (v/v) EtOH, and 5 to 100 mM Tris. To facilitate the mixing of arbitrary concentrations of the buffer components for optimization experiments, stock and diluent solutions for each the binding, wash, and elution and buffer were prepared according to **Table 3**.

To simulate a complex biological sample, a mixture containing salmon sperm DNA (0.1 mg mL⁻¹), BSA (10 mg mL⁻¹), glucose (1 mg mL⁻¹), and cholesterol (0.1 mg mL⁻¹) was prepared in UPW containing 1% EtOH to solubilize cholesterol by magnetic stirring overnight. The magnetic beads were first washed twice with UPW and then resuspended in 20 mM Tris-HCl at pH 5 by thorough vortexing. The magnetic beads were loaded into the reagent reservoir at a particle concentration of 15 mg mL⁻¹, establishing a final concentration of 2 mg mL⁻¹ during the binding step.

The concentrations of the buffer components were optimized using both a GA and a MBA in two separate sets of experiments, each composed of multiple experimental cycles. During each experimental cycle, 8 different parameter combinations were evaluated in the automated DNA purification process. A parameter combination was defined as a tuple composed of a Gu-HCl, EtOH, and Tris concentration and was given a unique identifier (UID), which is a sequential integer. For the initial experimental cycle, 8 parameter combinations were randomly generated by the algorithm. Subsequent to each experimental cycle, the resulting absorbance data were automatically evaluated and plotted (s. Supporting Information Figure S2 and S3) by the software. From the raw absorbance data, the loading of the magnetic beads as well as the purity of DNA in the eluate were calculated according to Equation (3) and (4).

To determine the parameter combinations for subsequent generations, the GA first computes the value of the fitness function for each parameter combination (=individual) as outlined in Equation (5):

$$\text{Fitness} = \text{Loading} \cdot \text{Purity} = q \cdot P \quad (5)$$

In Equation (5), q is the loading of the magnetic beads in $\mu\text{g mg}^{-1}$ and P is the dimensionless purity. The unit of the fitness function is thus $\mu\text{g mg}^{-1}$. Both loading and purity values are mean values derived from triplicated experiments to ensure reproducibility.

The parameter combinations are subsequently ranked by their fitness function value. The top four combinations are selected as parent candidates to generate the parameter set for the next generation. The top two parameter combinations are passed to the next generation. The remaining six parameter combinations are created by recombination. During this process, each parameter value (Gu-HCl, EtOH, and Tris concentration) in the offspring is randomly chosen from the values of the parent parameter combinations. After recombination, the individuals

Table 3. Conductivity of wash and elution supernatants from the DNA purification process using the conventional and the alternative binding buffer. For reference, the conductivity of the pure wash buffer (100% EtOH) and elution buffer (5 mM Tris at pH 8.5) are included. All values represent the mean and standard deviation of replicated experiments ($n = 3$).

	Wash 1 [μS]	Wash 2 [μS]	Eluate [μS]
Conventional binding buffer	3166.7 \pm 13.7	538.0 \pm 5.8	865.0 \pm 18.1
Alternative binding buffer	182.2 \pm 8.4	21.8 \pm 2.6	285.9 \pm 10.0
Reference	0.2 \pm 0.0	0.2 \pm 0.0	238.5 \pm 0.9

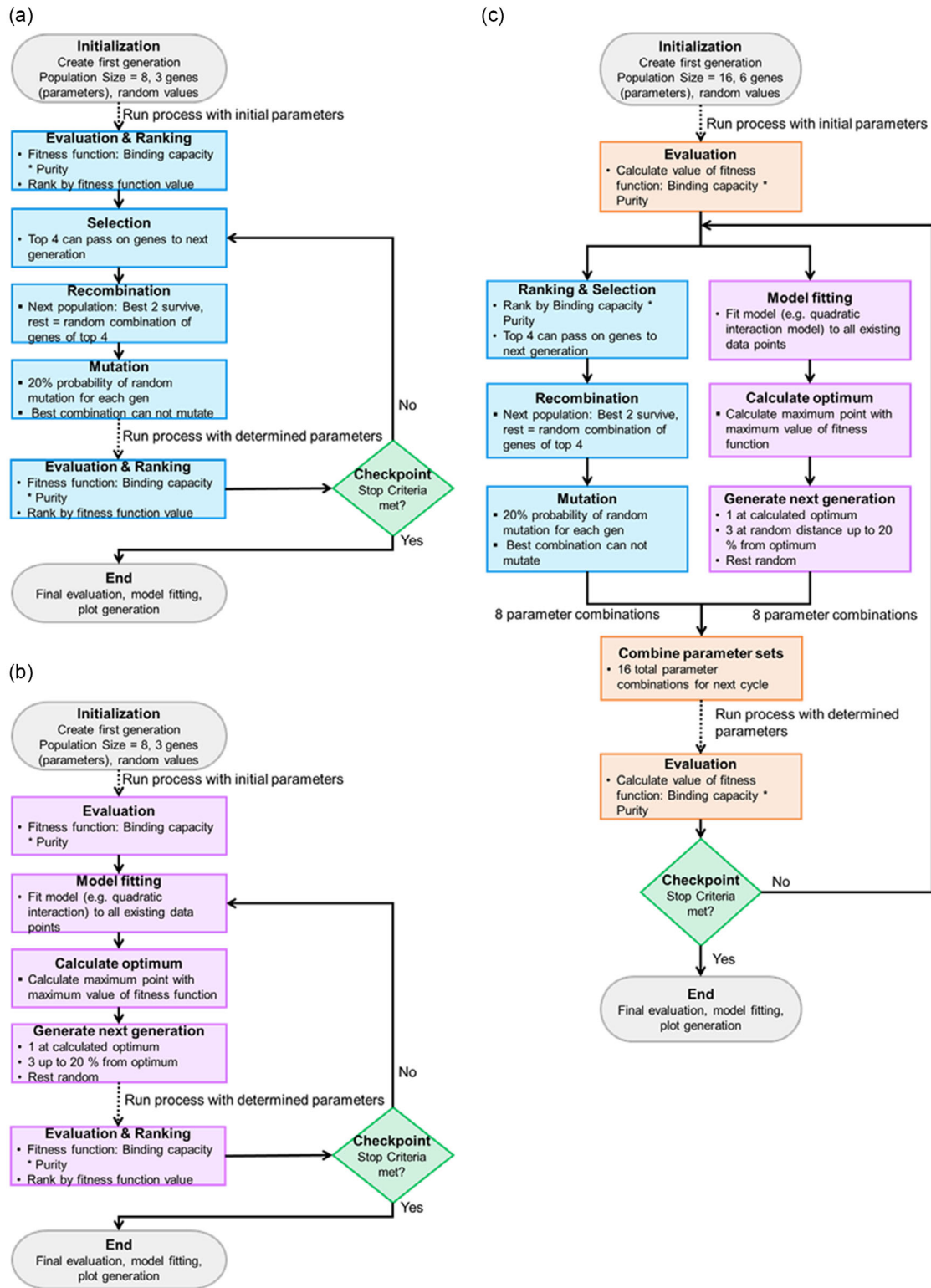


Figure 6. Flowcharts of the employed optimization algorithms for autonomous DNA purification process optimization. a) GA, b) MBA, and c) hybrid algorithm.

undergo a mutation step, in which the value of each parameter is altered to a random value within the predefined limits with a probability of 20%. Only the individual represented by the best parameter combination from the previous cycle is not allowed to mutate, so that each experimental cycle improves or at least maintains the highest achieved performance. The GA is visually summarized in a flow chart shown in **Figure 6a**.

The MBA employed in this study dynamically fits a quadratic interaction model to all accumulated data points to optimize the DNA purification process. This model predicts the fitness function, $q \cdot P$, as a function of the concentrations of the variable buffer components c_1, \dots, c_n . The model incorporates linear, quadratic, and interaction terms as delineated by Equation (6):

$$\text{Fitness} = q \cdot P = \beta_0 + \sum_{i=1}^n \beta_i c_i + \sum_{i=1}^n \beta_{ii} c_i^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \beta_{ij} c_i c_j \quad (6)$$

In Equation (6), β_0 is the intercept and β_i, β_{ii} , and β_{ij} are the coefficients for linear, quadratic, and interaction terms, respectively. Using the fitted equation, the algorithm identifies the parameter combination that maximizes the predicted fitness value by applying the L-BFGS-B optimization algorithm to minimize the negative of the fitness function. The parameter combinations for the next experimental cycle are determined as follows: One parameter combination is situated exactly at the calculated optimum. Three additional combinations are selected within a random distance of up to 20% from the optimum. This strategy tests the robustness of the model near the predicted optimum and explores slight variations that could potentially enhance performance. The remaining four parameter combinations are chosen randomly to introduce diversity into the dataset and expand the modeled parameter space, aiding in the robustness and generalizability of the model. In Figure 6b, the MBA is illustrated in a flowchart.

With either algorithm, the system runs experimental cycles until the defined target is met or a maximum number of experimental cycles is reached. For the DNA purification process optimization, a fitness value of $q \cdot P \geq 4$ was set as objective and the maximum cycle number was 20. Upon finishing the optimization, the system automatically proceeds to data analysis. This includes generating plots that compare maximum fitness values across generations and illustrate the parameter values of the best combinations from each cycle. Moreover, a quadratic-interaction model is refitted to all data collected through the experimental runs to validate and refine the predictions. The system is capable to fit other statistical and computational models, such as multiple linear regression and artificial neural networks. A plot comparing the actual versus predicted values is generated, additionally displaying the model's goodness of fit, represented by the coefficient of determination R^2 . Additionally, surface plots are created to visually represent the influence of each buffer component concentration on the DNA loading, thereby enhancing the understanding of the process and informing future optimizations.

Autonomous Optimization of Chaotropic Salt-Free Binding Buffer: In an effort to develop a chaotropic salt-free binding buffer, six alternative substances were evaluated for their potential to replace Gu-HCl. These substances were tested in arbitrary combinations and concentrations within the constraints outlined in **Table 4**. Each substance's stock solution was prepared at the specified concentration in 20 mM Tris, with the pH adjusted to 5 using 1 M HCl and 1 M NaOH. For dilution of the binding buffer to the final volume, 20 mM Tris at pH 5 was used (**Table 5**).

For all experiments, a wash buffer consisting of 100% EtOH and an elution buffer consisting of 5 mM Tris at pH 8.5 were utilized. The sample and bead preparation were conducted analog to the process optimization in the previous paragraph, maintaining a constant bead concentration of 2 mg mL⁻¹.

A hybrid algorithm, combining the previously described GA and MBA, was employed to optimize the composition of the binding buffer for maximal loading and purity. This approach is visualized in the flowchart in Figure 6c. After each cycle, measurement data were analyzed, and the fitness function value, based on Equation (5), was recalculated. Eight

Table 4. Buffer stock and diluent solutions composition for autonomous DNA purification process optimization experiments.

Buffer type	Stock solution	Diluent solution
Binding	6 M Gu-HCl	0 M Gu-HCl
	20 mM Tris-HCl	20 mM Tris-HCl
	pH 5	pH 5
Wash	100% (v/v) EtOH 0% UPW	0% (v/v) EtOH 100% UPW
Elution	100 mM Tris pH 8.5	5 mM Tris pH 8.5

Table 5. Alternative buffer compounds with concentration limits and stock solution concentration for the autonomous development of chaotropic salt-free binding buffer.

Substance	Lower concentration limit	Upper concentration limit	Stock solution concentration
Acetic acid	0 mM	200 mM	1000 mM
L-histidine	0 mM	100 mM	500 mM
L-arginine	0 mM	100 mM	500 mM
KCl	0 mM	400 mM	4000 mM
MgCl ₂	0 mM	400 mM	4000 mM
PEG-8000	0 g L ⁻¹	50 g L ⁻¹	400 g L ⁻¹

adjusted parameter combinations were then selected by each of the algorithms for further testing, resulting in a total 16 parameter combinations per experimental cycle. The objective for the optimization was set at a fitness value of 2 μg mg⁻¹ with a maximal cycle number of 20.

Comparison of Eluate Purity via Conductivity Analysis: To compare the purity of DNA purified using the conventional Gu-HCl binding buffer and the newly optimized chaotrope-free binding buffer, conductivity measurements were performed. The wash buffers and the eluates were analyzed using an ÄKTA conductivity cell (Cytiva, USA) connected to a conductometer (Konduktometer 703, Knick, Germany). To ensure adequate sample volume for these measurements, the automated DNA purification process was executed using both buffer systems across a complete PCR plate ($n = 96$). The supernatants from these processes were consolidated into three tubes to facilitate triplicate conductivity measurements.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work was funded by the KIT President's Strategy Fund under the title Auto.MAP within the Bio.CAR project.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

automation, DNA purification, magnetic beads, optimization algorithms, self-driving labs, silica, solid-phase extractions

Received: July 8, 2024

Revised: October 15, 2024

Published online:

- [1] T. Jesionowski, J. Zdarta, B. Krajewska, *Adsorption* **2014**, *20*, 801.
- [2] R. A. Sheldon, *Adv. Synth. Catal.* **2007**, *349*, 1289.
- [3] S. A. Bhakta, E. Evans, T. E. Benavidez, C. D. Garcia, *Anal. Chim. Acta* **2015**, *872*, 7.
- [4] T. D. Lazzara, I. Mey, C. Steinem, A. Janshoff, *Anal. Chem.* **2011**, *83*, 5624.
- [5] G. Ahuja, K. Pathak, *Indian J. Pharm. Sci.* **2009**, *71*, 599.
- [6] C. Pinholt, R. A. Hartvig, N. J. Medlicott, L. Jorgensen, *Expert Opin. Drug Delivery* **2011**, *8*, 949.
- [7] I. Ali, V. K. Gupta, *Nat. Protoc.* **2006**, *1*, 2661.
- [8] N. B. Singh, G. Nagpal, S. Agrawal, Rachna, *Environ. Technol. Innov.* **2018**, *11*, 187.
- [9] H. A. Chase, *Trends Biotechnol.* **1994**, *12*, 296.
- [10] M. Franzreb, M. Siemann-Herzberg, T. J. Hobley, O. R. T. Thomas, *Appl. Microbiol. Biotechnol.* **2006**, *70*, 505.
- [11] A. M. Ramos-de-la-Peña, J. González-Valdez, O. Aguilar, *J. Sep. Sci.* **2019**, *42*, 1816.
- [12] C. De Luca, S. Felletti, G. Lievore, T. Chenet, M. Morbidelli, M. Sponchioni, A. Cavazzini, M. Catani, *TrAC Trends Anal. Chem.* **2020**, *132*, 116051.
- [13] H. Lorenz, A. Seidel-Morgenstern, *Angew. Chem., Int. Ed.* **2014**, *53*, 1218.
- [14] D. S. Kohane, *Biotechnol. Bioeng.* **2007**, *96*, 203.
- [15] I. Safarik, M. Safarikova, *Chem. Pap.* **2009**, *63*, <https://doi.org/10.2478/s11696-009-0054-2>.
- [16] S. Berensmeier, *Appl. Microbiol. Biotechnol.* **2006**, *73*, 495.
- [17] C. Ruffert, *ECS Trans.* **2015**, *64*, 49.
- [18] C. Jiang, L. Flansburg, S. Ghose, P. Jorjorian, A. A. Shukla, *Biotechnol. Bioeng.* **2010**, *107*, 985.
- [19] M. Rabe, D. Verdes, S. Seeger, *Adv. Colloid Interface Sci.* **2011**, *162*, 87.
- [20] B. K. Nfor, T. Ahamed, G. W. Van Dedem, L. A. Van Der Wielen, E. J. Van De Sandt, M. H. Eppink, M. Ottens, *J. Chem. Technol. Biotechnol.* **2008**, *83*, 124.
- [21] E. Ramin, A. G. Cardillo, R. Liebers, J. Schmölder, E. Von Lieres, W. Van Molle, B. Niebel, L. Natalis, I. Meln, M. Perea-Vélez, D. Clénet, J. B. Jørgensen, B. Nilsson, D. G. Bracewell, K. V. Gernaey, *Curr. Opin. Chem. Eng.* **2024**, *43*, 100998.
- [22] D. Keulen, G. Geldhof, O. L. Bussy, M. Pabst, M. Ottens, *J. Chromatogr. A* **2022**, *1676*, 463195.
- [23] B. Kaltenboeck, C. Wang, *Advances in Clinical Chemistry*, Elsevier, Amsterdam **2005**, pp. 219–259.
- [24] S. D. Patil, D. G. Rhodes, D. J. Burgess, *AAPS J.* **2005**, *7*, E61.
- [25] N. C. Seeman, H. F. Sleiman, *Nat. Rev. Mater.* **2017**, *3*, 17068.
- [26] L. Roewer, *Invest. Genet.* **2013**, *4*, 22.
- [27] R. Boom, C. J. Sol, M. M. Salimans, C. L. Jansen, P. M. Wertheim-van Dillen, J. Van Der Nooraa, *J. Clin. Microbiol.* **1990**, *28*, 495.
- [28] P. E. Vandeventer, J. S. Lin, T. J. Zwang, A. Nadim, M. S. Johal, A. Niemz, *J. Phys. Chem. B* **2012**, *116*, 5661.
- [29] M. K. Hourfar, U. Michelsen, M. Schmidt, A. Berger, E. Seifried, W. K. Roth, *Clin. Chem.* **2005**, *51*, 1217.
- [30] J. R. Moll, *Nucleic Acids Res.* **2002**, *30*, 1240.
- [31] S. Vilanova, D. Alonso, P. Gramazio, M. Plazas, E. García-Fortea, P. Ferrante, M. Schmidt, M. J. Díez, B. Usadel, G. Giuliano, J. Prohens, *Plant Methods* **2020**, *16*, 110.
- [32] P. E. Vandeventer, J. Mejia, A. Nadim, M. S. Johal, A. Niemz, *J. Phys. Chem. B* **2013**, *117*, 10742.
- [33] S. Bag, S. Rauwolf, S. P. Schwaminger, W. Wenzel, S. Berensmeier, *Langmuir* **2021**, *37*, 5902.
- [34] M. Abolhasani, E. Kumacheva, *Nat. Synth.* **2023**, *2*, 483.
- [35] H. G. Martin, T. Radivojevic, J. Zucker, K. Bouchard, J. Sustarich, S. Peisert, D. Arnold, N. Hillson, G. Babnigg, J. M. Marti, C. J. Mungall, G. T. Beckham, L. Waldburger, J. Carothers, S. Sundaram, D. Agarwal, B. A. Simmons, T. Backman, D. Banerjee, D. Tanjore, L. Ramakrishnan, A. Singh, *Curr. Opin. Biotechnol.* **2023**, *79*, 102881.
- [36] F. Häse, L. M. Roch, A. Aspuru-Guzik, *Trends Chem.* **2019**, *1*, 282.
- [37] M. B. Rooney, B. P. MacLeod, R. Oldford, Z. J. Thompson, K. L. White, J. Tungjunyatham, B. J. Stankiewicz, C. P. Berlinguette, *Digital Discovery* **2022**, *1*, 382.
- [38] S. Steiner, J. Wolf, S. Glatzel, A. Andreou, J. M. Granda, G. Keenan, T. Hinkley, G. Aragon-Camarasa, P. J. Kitson, D. Angelone, L. Cronin, *Science* **2019**, *363*, eaav2211.
- [39] C. W. Coley, D. A. Thomas, J. A. M. Lummiss, J. N. Jaworski, C. P. Breen, V. Schultz, T. Hart, J. S. Fishman, L. Rogers, H. Gao, R. W. Hicklin, P. P. Plehiers, J. Byington, J. S. Piotti, W. H. Green, A. J. Hart, T. F. Jamison, K. F. Jensen, *Science* **2019**, *365*, eaax1566.
- [40] J. M. Granda, L. Donina, V. Dragone, D.-L. Long, L. Cronin, *Nature* **2018**, *559*, 377.
- [41] M. Christensen, L. P. E. Yunker, F. Adedeji, F. Häse, L. M. Roch, T. Gensch, G. Dos Passos Gomes, T. Zepel, M. S. Sigman, A. Aspuru-Guzik, J. E. Hein, *Commun. Chem.* **2021**, *4*, 112.
- [42] T. Ha, D. Lee, Y. Kwon, M. S. Park, S. Lee, J. Jang, B. Choi, H. Jeon, J. Kim, H. Choi, H.-T. Seo, W. Choi, W. Hong, Y. J. Park, J. Jang, J. Cho, B. Kim, H. Kwon, G. Kim, W. S. Oh, J. W. Kim, J. Choi, M. Min, A. Jeon, Y. Jung, E. Kim, H. Lee, Y.-S. Choi, *Sci. Adv.* **2023**, *9*, eadj0461.
- [43] R. W. Epps, M. S. Bowen, A. A. Volk, K. Abdel-Latif, S. Han, K. G. Reyes, A. Amassian, M. Abolhasani, *Adv. Mater.* **2020**, *32*, 2001626.
- [44] K. Abdel-Latif, R. W. Epps, F. Bateni, S. Han, K. G. Reyes, M. Abolhasani, *Adv. Intell. Syst.* **2021**, *3*, 2000245.
- [45] D. Salley, G. Keenan, J. Grizou, A. Sharma, S. Martín, L. Cronin, *Nat. Commun.* **2020**, *11*, 2771.
- [46] J. Li, J. Li, R. Liu, Y. Tu, Y. Li, J. Cheng, T. He, X. Zhu, *Nat. Commun.* **2020**, *11*, 2046.
- [47] J. Li, Y. Tu, R. Liu, Y. Lu, X. Zhu, *Adv. Sci.* **2020**, *7*, 1901957.
- [48] A. A. Volk, R. W. Epps, M. Abolhasani, *Adv. Mater.* **2021**, *33*, 2004495.
- [49] S. B. Harris, A. Biswas, S. J. Yun, K. M. Roccapiore, C. M. Rouleau, A. A. Poretzky, R. K. Vasudevan, D. B. Geohagan, K. Xiao, *Small Methods* **2024**, *8*, 2301763.
- [50] P. Nikolaev, D. Hooper, N. Perea-López, M. Terrones, B. Maruyama, *ACS Nano* **2014**, *8*, 10214.
- [51] B. P. MacLeod, F. G. L. Parlane, C. C. Rupnow, K. E. Dettelbach, M. S. Elliott, T. D. Morrissey, T. H. Haley, O. Proskurin, M. B. Rooney, N. Taherimakhosousi, D. J. Dvorak, H. N. Chiu, C. E. B. Waizenegger, K. Ocean, M. Mokhtari, C. P. Berlinguette, *Nat. Commun.* **2022**, *13*, 995.
- [52] B. P. MacLeod, F. G. L. Parlane, T. D. Morrissey, F. Häse, L. M. Roch, K. E. Dettelbach, R. Moreira, L. P. E. Yunker, M. B. Rooney, J. R. Deeth, V. Lai, G. J. Ng, H. Situ, R. H. Zhang, M. S. Elliott,

- T. H. Haley, D. J. Dvorak, A. Aspuru-Guzik, J. E. Hein, C. P. Berlinguette, *Sci. Adv.* **2020**, *6*, eaaz8867.
- [53] R. D. King, K. E. Whelan, F. M. Jones, P. G. K. Reiser, C. H. Bryant, S. H. Muggleton, D. B. Kell, S. G. Oliver, *Nature* **2004**, *427*, 247.
- [54] K. Williams, E. Bilsland, A. Sparkes, W. Aubrey, M. Young, L. N. Soldatova, K. De Grave, J. Ramon, M. De Clare, W. Sirawaraporn, S. G. Oliver, R. D. King, *J. R. Soc. Interface* **2015**, *12*, 20141289.
- [55] T. Si, R. Chao, Y. Min, Y. Wu, W. Ren, H. Zhao, *Nat. Commun.* **2017**, *8*, 15187.
- [56] M. Hamedirad, R. Chao, S. Weisberg, J. Lian, S. Sinha, H. Zhao, *Nat. Commun.* **2019**, *10*, 5150.
- [57] G. N. Kanda, T. Tsuzuki, M. Terada, N. Sakai, N. Motozawa, T. Masuda, M. Nishida, C. T. Watanabe, T. Higashi, S. A. Horiguchi, T. Kudo, M. Kamei, G. A. Sunagawa, K. Matsukuma, T. Sakurada, Y. Ozawa, M. Takahashi, K. Takahashi, T. Natsume, *eLife* **2022**, *11*, e77007.
- [58] P. Notin, N. Rollins, Y. Gal, C. Sander, D. Marks, *Nat. Biotechnol.* **2024**, *42*, 216.
- [59] J. T. Rapp, B. J. Bremer, P. A. Romero, *Nat. Chem. Eng.* **2024**, *1*, 97.
- [60] T. Ballweg, M. Liu, J. Grimm, E. Sedghamiz, W. Wenzel, M. Franzreb, *J. Chromatogr. A* **2024**, *1730*, 465089.
- [61] M. Cárdenas, K. Schillén, D. Pebalk, T. Nylander, B. Lindman, *Biomacromolecules* **2005**, *6*, 832.
- [62] V. Balladur, A. Theretz, B. Mandrand, *J. Colloid Interface Sci.* **1997**, *194*, 408.
- [63] T. H. Nguyen, M. Elimelech, *Biomacromolecules* **2007**, *8*, 24.
- [64] J. A. Libera, H. Cheng, M. Olvera De La Cruz, M. J. Bedzyk, *J. Phys. Chem. B* **2005**, *109*, 23001.
- [65] K. A. Melzak, C. S. Sherwood, R. F. B. Turner, C. A. Haynes, *J. Colloid Interface Sci.* **1996**, *181*, 635.
- [66] S. Ong, X. Zhao, K. B. Eisenthal, *Chem. Phys. Lett.* **1992**, *191*, 327.
- [67] S. Katoch, S. S. Chauhan, V. Kumar, *Multimed. Tools Appl.* **2021**, *80*, 8091.
- [68] K. Kubiak-Ossowska, K. Tokarczyk, B. Jachimska, P. A. Mulheran, *J. Phys. Chem. B* **2017**, *121*, 3975.
- [69] A. C. McUmber, T. W. Randolph, D. K. Schwartz, *J. Phys. Chem. Lett.* **2015**, *6*, 2583.
- [70] F. Sousa, C. Cruz, J. A. Queiroz, *J. Mol. Recognit.* **2010**, *23*, 505.
- [71] Effects of Low A260/A230 Ratios in RNA Preparations on Downstream Applications - QIAGEN Application Note, <https://www.qiagen.com/de/resources/resourcedetail?id=11226191-0a82-4a9b-ba4a-99800b6f8595&lang=en> (accessed: June 2024).
- [72] Opentrons Python API Version 2 Reference, https://docs.opentrons.com/v2/new_protocol_api.html (accessed: May 2024).
- [73] Universal Robots Remote Control Via TCP/IP Primary Interface, <https://www.universal-robots.com/articles/ur/interface-communication/remote-control-via-tcpip/> (accessed: May 2024).
- [74] Vacuu Select MD 4 Vario Manual, <https://shop.vacuubrand.com/de/vakuu-controller-vacu-select-20700000.html> (accessed: May 2024).
- [75] G. Hinkel, J. Kunert, J. Meredith, *SLAS Technol.* **2023**, *28*, 334.
- [76] CETONI SDK Documentation, https://cetoni.de/downloads/manuals/CETONI_SDK/index.html (accessed: May 2024).