Journal of Microbiological Methods 245 (2026) 107516

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

# Hybrid experimental–machine learning framework for media optimization enhances antioxidant production in *Micrococcus endophyticus* SS-1

Md Sourav Sarker [a,b], Asmamaw Abat Getu [a,b], Juvens Sugira Murekezi [a,b], Geng Chen [a,b], Jing Zhang [a,d], Yong Xiao [a,b,c,*]

[a] *State Key Laboratory of Regional and Urban Ecology, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen, Fujian 361021, PR China*
[b] *University of Chinese Academy of Sciences, Beijing 100049, China*
[c] *Xiamen Key Laboratory of Physical Environment, Institute of Urban Environment, Xiamen 361021, China*
[d] *Fuzhou University, Fuzhou, Fujian 350116, PR China*

ARTICLE INFO

*Keywords:*
Antioxidant activity
Media optimization
Machine learning
Neural network
Genetic algorithm
Orthogonal array

ABSTRACT

Culture medium composition is a key determinant of microbial antioxidant production, yet systematic optimization remains challenging due to nonlinear interactions among nutrients and the large experimental design space. In this study, a hybrid experimental–computational framework integrating orthogonal array design, neural-network surrogate modeling, and genetic algorithm optimization was developed for antioxidant production in *Micrococcus endophyticus* SS-1. A three-level L81 orthogonal array was used to evaluate ten nutritional and cultivation factors, generating a dataset for surrogate model training. The trained neural network was coupled with a genetic algorithm to explore the formulation space efficiently and to identify candidate media predicted to maximize antioxidant activity. Experimental validation confirmed that the genetic algorithm–optimized medium exhibited higher antioxidant activity than both the baseline medium and the best orthogonal-array formulation, as determined using the 2,2-diphenyl-1-picrylhydrazyl radical-scavenging assay. To account for the concentration-dependent nature of single-point scavenging measurements, dose–response experiments were performed and the half-maximal effective concentration values were determined. The experimentally validated genetic algorithm–optimized medium showed a reduced half-maximal effective concentration, indicating enhanced antioxidant potency rather than an assay-dependent increase in scavenging percentage.

## 1. Introduction

Reactive oxygen species (ROS) and free radicals are unavoidable byproducts of cellular metabolism, and their accumulation underlies oxidative stress, aging, and the onset of chronic diseases such as cancer, cardiovascular disorders, and neurodegeneration (Leyane et al., 2022; Valko et al., 2007). Antioxidants play a central role in neutralizing ROS, thereby preserving cellular homeostasis and mitigating oxidative damage (Karayat et al., 2025). Microorganisms are increasingly recognized as sustainable producers of antioxidant metabolites that can complement or even replace plant-derived compounds (Rani et al., 2021). Actinobacteria of the genus *Micrococcus* are notable in this regard, synthesizing carotenoids, phenolic derivatives, and other bioactive molecules with antioxidant, antimicrobial, and anticancer properties (Manivasagan et al., 2014; Mohana et al., 2013). For example, *Micrococcus luteus* has been reported to achieve radical scavenging activities up to ~79% (Akbar et al., 2014). *Micrococcus endophyticus*, first described as an endophyte from *Aquilaria sinensis* roots (Chen et al., 2009), remains largely unexplored with respect to antioxidant production, and its metabolic response to medium composition has not been systematically examined. Production of microbial secondary metabolites is highly sensitive to medium composition and culture parameters (Sanchez and Demain, 2002). Traditional one-factor-at-a-time (OFAT) approaches are inefficient and fail to capture interactions among medium components (Zhou et al., 2023). Statistical designs such as response surface methodology (RSM) have improved efficiency (Chauhan et al., 2021), but their reliance on simplified polynomial equations constrains their ability to capture the nonlinear, high-

* Corresponding author at: State Key Laboratory of Regional and Urban Ecology, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen, Fujian 361021, PR China.

*E-mail address:* yxiao@iue.ac.cn (Y. Xiao).

https://doi.org/10.1016/j.mimet.2026.107516
Received 31 January 2026; Received in revised form 4 April 2026; Accepted 15 April 2026
Available online 16 April 2026
0167-7012/

dimensional relationships typical of microbial systems (Hashizume and Ying, 2024). Although RSM has been successfully applied to optimize antioxidant extraction from plants (Begum et al., 2023), its scalability remains constrained. In microbial systems, factors such as carbon availability, phosphate buffering, nitrogen source quality, and pH are known to influence redox balance and precursor supply, indicating that antioxidant production may depend strongly on multidimensional nutrient interactions.

Machine learning (ML) approaches have emerged as powerful alternatives, offering the ability to capture nonlinear factor interactions and uncover predictive relationships in high-dimensional data (Hashizume and Ying, 2024; Zhou et al., 2023). Recent studies highlight the promise of combining ML surrogates with metaheuristic optimizers such as genetic algorithms (GA) or Bayesian optimization (BO) to efficiently explore large formulation spaces (Narayanan et al., 2025; Yoshida et al., 2023). Hybrid frameworks have already been demonstrated in microbial systems, including Plackett–Burman/Box–Behnken-based media optimization for *Bacillus velezensis* (Shi et al., 2024) and AI-guided fermentation regulation that rewires metabolic fluxes for enhanced antibiotic titers (Xu et al., 2025). Collectively, these advances show that ML-guided workflows can accelerate process optimization while reducing experimental burden, a trend increasingly recognized in microbial biotechnology and industrial applications (Nisha et al., 2025; Shah et al., 2024). However, despite these advances, hybrid ML–DOE frameworks have rarely been applied to antioxidant-producing actinobacteria, and multidimensional nutrient interactions governing antioxidant biosynthesis in *Micrococcus endophyticus* have not yet been systematically characterized.

In this study, antioxidant metabolite production in *M. endophyticus* SS-1 was optimized using an integrated framework that combines orthogonal-array design, data-driven factor selection, neural-network surrogate modeling, and GA-driven global search. Screening of fourteen nutrient media was first used to characterize baseline antioxidant activity and to identify influential medium components. These fourteen formulations were selected from widely used microbiological media for actinobacteria and encompass diverse nutrient environments, thereby providing a biologically grounded basis for identifying medium components relevant to antioxidant production. These factors were then evaluated using a ten-factor, three-level L81 orthogonal array, enabling comprehensive exploration of nutrient interactions. A neural-network surrogate model was then incorporated into a GA to propose candidate high-performance formulations.

This work contributes a compact and generalizable workflow for microbial media optimization and provides a systematic evaluation of how multidimensional nutrient interactions influence antioxidant metabolite production in *Micrococcus endophyticus* SS-1. By integrating structured experimentation with ML-guided modeling, the study advances both mechanistic understanding and practical strategies for accelerating bioprocess development in antioxidant-producing microorganisms.

## 2. Materials and methods

### 2.1. Microorganism and molecular identification

A bacterial isolate was obtained from the effluent of a wastewater treatment plant in Harbin, Heilongjiang Province, China, and preserved in the culture collection of the Institute of Urban Environment, Chinese Academy of Sciences. Wastewater effluent was selected as a sampling source because it represents a chemically diverse environment where microorganisms are frequently exposed to oxidative and chemical stresses, which may favor the presence of metabolically versatile or bioactive metabolite–producing bacteria. The isolate was obtained through serial dilution of the effluent sample followed by plating on nutrient agar and incubation under aerobic conditions. Distinct colonies were selected and purified through repeated streaking prior to molecular identification. Genomic DNA was extracted using the phenol–chloroform method. The 16S rRNA gene was amplified with universal primers 27F (5′-AGAGTTTGATCCTGGCTCAG-3′) and 1492R (5′-GGTTACCTTGTTACGACTT-3′) in a 25 μL PCR mixture containing 1× buffer, 1.5 mM $MgCl_2$, 200 μM dNTPs, 0.2 μM of each primer, 1 U Taq polymerase, and approximately 50 ng template DNA. PCR cycling conditions were: initial denaturation at 95 °C for 5 min; 35 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 90 s; followed by a final extension at 72 °C for 10 min. The ~1.5 kb amplicon was purified and sequenced (Biosune Co., Ltd., Xiamen, China). BLAST analysis of the consensus sequence showed 99.2% identity with *Micrococcus endophyticus*. The strain was designated *M. endophyticus* SS-1, and the 16S rRNA sequence has been deposited in GenBank under the accession number PX472913.

### 2.2. Preliminary media screening and rationale

To survey how nutrient environments influence antioxidant output, fourteen culture media were selected from the *Handbook of Microbiological Media* (Atlas and Atlas, 2004), chosen to represent (i) complex rich formulations (e.g., Peptone Glucose Yeast, Trypticase Soy, Brain–Heart Infusion), (ii) extract-based broths (e.g., Beef Extract Peptone, Nutrient Broth), (iii) enriched formulations containing yeast or organ extracts, and (iv) selective or stress-inducing media supplemented with antibiotics or alternative carbon sources (streptomycin, neomycin, lactate, pyridine). A summary of these formulations and their characteristic nutrient profiles is provided in Table 1. These media collectively

**Table 1**
Summary of the *fourteen-culture* media tested for *Micrococcus endophyticus* growth and DPPH radical-scavenging activity.

| ID | Medium Name | Key Nutritional Profile | Intended Use / Rationale |
|---|---|---|---|
| M1 | PGY (Peptone Glucose Yeast) | Glucose, peptone, yeast extract | Rich carbon and nitrogen; supports robust aerobic growth |
| M2 | Trypticase Soy Broth (TSB) | Casein digest, soy peptone | Balanced protein hydrolysates; commonly used for *Micrococcus* |
| M3 | Brain Heart Infusion (BHI) | Calf brain, beef heart, peptone | Highly nutritious; supports fastidious bacterial species |
| M4 | Nutrient Broth | Beef extract, peptone | General-purpose medium; baseline for bacterial growth |
| M5 | Yeast Extract Peptone Glucose (YEPG) | Yeast extract, peptone, glucose | High carbohydrate + vitamin supply; promotes biomass |
| M6 | Beef Extract Peptone | Beef extract, peptone | Simple extract-based formulation; widely used in antioxidant assays |
| M7 | Glucose Peptone Medium | Glucose, peptone | Minimal yet carbohydrate-enriched; suitable for oxidative metabolism |
| M8 | Lactate Medium | Sodium lactate, peptone | Selective carbon source; evaluates organic acid utilization |
| M9 | Pyridine-Containing Medium | Pyridine, peptone | Stress-inducing formulation; tests tolerance to heteroaromatic compounds |
| M10 | Neomycin Supplemented Medium | Peptone, yeast extract, neomycin | Selective; mimics antibiotic pressure and stress response |
| M11 | Streptomycin-Supplemented Medium | Peptone, yeast extract, streptomycin | Selective; probes resistance/ antioxidant linkage under antibiotic stress |
| M12 | Calf Brain Infusion | Infusion from calf brains | Specialized nutrient-rich extract; supports robust growth of *Micrococcus* |
| M13 | Pancreatic Digest of Gelatin | Gelatin hydrolysate | Provides amino acids/ peptides; supports protein metabolism |
| M14 | Soytone Medium | Soy peptone | Plant-derived nitrogen source; alternative to animal peptones |

span a broad physiological range used for culturing actinobacteria, ensuring that the initial screen sampled nutrient conditions relevant to both general growth and secondary metabolite production in *Micrococcus* spp.

Each medium was prepared according to its published formulation, dispensed as 50 mL aliquots into 150 mL Erlenmeyer flasks, inoculated with the same starting inoculum, and incubated with orbital shaking (150 rpm) for 24 h. All screening experiments were performed in biological triplicate to ensure reproducibility, and the complete screening dataset is provided in Supplementary Materials S1. Biological triplicates refer to independent cultures initiated from separately prepared inocula rather than splitting a single starter culture, with each replicate cultured in an independent flask under identical conditions. Supernatants were collected after incubation and used for subsequent 2,2-diphenyl-1-picrylhydrazyl (DPPH) radical-scavenging assays. Occasional negative scavenging values observed during screening were retained for completeness, as these reflected sample-specific absorbance interference rather than true pro-oxidant activity.

### *2.3. Identification of factors for systematic optimization*

The preliminary screening of fourteen culture media provided diverse nutrient environments and corresponding antioxidant responses of *Micrococcus endophyticus* SS-1. Because the dataset contained only fourteen formulations, factor identification was based on measurable variation and biological relevance rather than predictive modeling.

All medium components were first evaluated for non-zero variance and for detectable linear association with DPPH scavenging using Pearson correlation coefficients. Components were selected for systematic optimization if they showed (i) measurable variation across media, (ii) a detectable association with antioxidant activity, and (iii) biological relevance to microbial growth, redox balance, or secondary metabolite production. Components not meeting these conditions were excluded.

Ten factors were retained for optimization: tryptone, peptone, beef extract, yeast extract, glucose, $K_2HPO_4$, NaCl, streptomycin, temperature, and pH. For all subsequent analyses, the mean value of each biological triplicate was used as the response variable.

Although factors such as NaCl, temperature, and streptomycin exhibited weaker linear correlations with DPPH scavenging, they were intentionally retained because they showed non-zero variance and have documented physiological effects in *Micrococcus* spp. Their inclusion allowed the orthogonal array to capture potential nonlinear or interaction effects that simple correlation analysis cannot resolve.

### *2.4. Orthogonal array design for systematic medium optimization*

The ten factors identified through variance analysis and biological relevance criteria were evaluated systematically using a three-level L81 orthogonal array ($3^{10}$). This design was selected because it enables efficient exploration of nonlinear responses while requiring substantially fewer experiments than a full factorial approach (59,049 combinations). The low, medium, and high levels assigned to each factor—summarized in Table 2—were derived from the minimum, intermediate, and maximum concentrations observed across the fourteen preliminary media, ensuring that all formulations remained within physiologically suitable ranges for *M. endophyticus* SS-1. The streptomycin concentrations therefore reflect values present in the selected preliminary culture media and were retained in the orthogonal array design without modification. The selected temperature range (25–30 °C) reflects the incubation temperatures used during the preliminary media screening experiments.

The orthogonal matrix was generated algorithmically using a Galois-field–based construction to ensure mutual orthogonality and uniform level distribution across all columns. The resulting 81 unique medium compositions (MC1–MC81) provided a balanced experimental framework for quantifying the influence of each component independently of confounding effects. All formulations were prepared in triplicate to produce a dataset of 243 observations, which served as the foundation for surrogate modeling and optimization. Detailed formulations for each orthogonal combination and the generation script are provided in Supplementary Materials S2.

**Table 2**
Experimental factors and levels used in the L81 orthogonal array design.

| Component | Low | Medium | High |
|---|---|---|---|
| Tryptone (or Trypticase) (g/L) | 3 | 5 | 17 |
| Peptone (g/L) | 2 | 5 | 10 |
| Streptomycin (g/L) | 0.25 | 0.50 | 0.75 |
| NaCl (g/L) | 1 | 3 | 5 |
| Beef Extract (g/L) | 1 | 2 | 3 |
| Yeast Extract (g/L) | 1 | 2 | 4 |
| Glucose (g/L) | 0.5 | 1 | 2 |
| $K_2HPO_4$ (g/L) | 0.20 | 0.61 | 1.20 |
| Temperature (°C) | 25 | 28 | 30 |
| pH | 6 | 7 | 8 |

### *2.5. Determination of DPPH radical-scavenging activity*

Antioxidant activity across all media formulations (the fourteen-medium screening and the L81 orthogonal array) was initially assessed using the 2,2-diphenyl-1-picrylhydrazyl (DPPH) radical-scavenging assay, reported as percent inhibition. Culture supernatants were obtained after 24 h of incubation by centrifugation followed by filtration to remove cells and particulates. A methanolic DPPH solution was prepared freshly for each experimental batch by dissolving 2,2-diphenyl-1-picrylhydrazyl (0.0025 g per 100 mL) in methanol. For each medium formulation, a fixed volume of culture supernatant was reacted with DPPH solution and incubated in the dark at room temperature for 30 min. Absorbance was measured at 517 nm using a UV–Vis spectrophotometer. Solvent blanks (methanol only) and DPPH controls (methanol + DPPH) were included in each experimental batch. Ascorbic acid was used as a reference antioxidant as a positive control to verify the responsiveness of the DPPH assay under the same measurement conditions. All absorbance values were corrected by subtracting the methanol blank prior to further calculation. DPPH radical-scavenging activity was calculated relative to the DPPH control according to the following equation (Pianpumepong and Noomhorm, 2010):

$$\text{Scavenging activity } (\%) = \left[1 - \frac{A_{sample}}{A_{control}}\right] \times 100$$

where $A_{\text{sample}}$ is the absorbance of the reaction mixture containing culture supernatant and DPPH, and $A_{\text{control}}$ is the absorbance of the DPPH–methanol control. All measurements were performed in biological triplicate, and the mean scavenging percentage was used as the response variable for subsequent orthogonal-array analysis and surrogate modeling.

Although single-point DPPH scavenging measurements do not capture the full complexity of antioxidant mechanisms, this assay provides a robust, reproducible, and internally consistent proxy for comparative evaluation of medium-dependent antioxidant production under controlled process conditions. These percent-inhibition values were therefore used as the primary response metric for systematic screening and optimization, with concentration-normalized $EC_{50}$ values determined separately for validation purposes.

### *2.6. Determination of DPPH $EC_{50}$ values*

To obtain a concentration-normalized measure of antioxidant performance and to validate trends observed in single-point DPPH scavenging measurements, half-maximal effective concentration ($EC_{50}$)

values were determined for a representative subset of media formulations spanning low, intermediate, and high scavenging activities. Culture supernatants were prepared as described above and subjected to a six-point dilution series (3.125–100% *v*/v) in methanol.

Each dilution was mixed with an equal volume of freshly prepared DPPH solution and incubated in the dark at room temperature for 30 min under the same spectrophotometric conditions used for single-point measurements. Absorbance was measured at 517 nm. Solvent blanks (methanol only) and DPPH controls (methanol + DPPH) were included in each experimental batch to ensure consistency with the primary DPPH assay.

Scavenging activity at each dilution was calculated using the same equation described above for single-point measurements. For each formulation, the calculated scavenging percentages were plotted against the logarithm of the dilution fraction, and $EC_{50}$ values were estimated by nonlinear logistic regression to determine the dilution required to achieve 50% radical scavenging. $EC_{50}$ calculations were based on the mean values from biological triplicates at each dilution level.

Dilution series that exhibited irregular or non-monotonic responses, resulting in poor curve definition, were excluded from $EC_{50}$ estimation, and these limitations are noted explicitly in the Results section. $EC_{50}$ values are reported as the percentage (*v*/v) of culture supernatant required to reduce DPPH radical absorbance by 50%. These measurements were used exclusively as a validation metric to confirm that improvements observed in single-point scavenging activity reflected enhanced antioxidant potency rather than dilution-dependent assay artifacts.

### 2.7. Machine-learning surrogate Modeling and optimization

Machine-learning modeling was conducted exclusively on the 243 observations generated from the L81 orthogonal array, ensuring sufficient variation and preventing information leakage from the preliminary screening dataset. For each formulation, the mean scavenging activity from biological triplicates was used as the response variable.

The input variables consisted of ten factors in total, comprising eight medium components plus two cultivation parameters (temperature and pH). All variables were standardized using z-score normalization prior to modeling.

A feed-forward neural network (multilayer perceptron) was implemented in Python using scikit-learn and TensorFlow. The dataset was randomly partitioned into training (70%), validation (15%), and test (15%) subsets. The neural network employed two hidden layers with ReLU activation, L2 regularization, and the Adam optimizer. Hyperparameters—including learning rate (0.001), number of neurons per layer (32 and 16), batch size (32), and regularization strength—were tuned empirically based on validation performance. Early stopping with a patience of 20 epochs was used to minimize overfitting.

Model accuracy was evaluated using the coefficient of determination ($R^2$), root mean squared error (RMSE), and mean absolute error (MAE) on the held-out test set. Cross-validation results and parity plots demonstrated that the neural-network surrogate captured the multidimensional response surface with high fidelity.

A genetic algorithm (GA) was used to explore the continuous factor space within the bounds defined by the orthogonal array. The GA, implemented using the DEAP library, encoded each candidate medium as a real-valued vector. Crossover, mutation, and tournament selection operators iteratively generated new populations, and the trained neural network served as a fast surrogate to estimate fitness (predicted DPPH scavenging). After convergence, the highest-scoring GA-predicted formulations were experimentally evaluated to assess model generalization under conditions not included in training.

### 2.8. Statistical analysis

All experimental measurements are expressed as mean ± standard deviation (SD) from biological triplicates. One-way ANOVA followed by Tukey's HSD post hoc test ($p < 0.05$) was used to assess differences among media formulations. Statistical analyses were performed in Python using SciPy and statsmodels.

## 3. Results

### 3.1. Screening of nutrient media reveals variability in antioxidant activity

Antioxidant activity of *Micrococcus endophyticus* SS-1 varied widely across the fourteen-culture media tested, with DPPH radical-scavenging values spanning from apparent negative values to over 75% (Fig. 1). Media containing balanced carbon- and nitrogen-rich components (e.g., Beef Extract Peptone and PGY) tended to exhibit higher scavenging activities, whereas antibiotic-supplemented media exhibited the lowest values. Growth analysis under streptomycin-containing conditions showed that although biomass accumulation was reduced, OD600 values increased over time, indicating that the applied streptomycin concentration did not completely inhibit bacterial growth but instead imposed a sub-inhibitory stress condition (Supplementary Fig. S1).

Negative scavenging values were observed in a small number of cases. These resulted from sample absorbance exceeding the DPPH control and were retained without adjustment. Such negative values are known artifacts in DPPH assays when pigments or reducing sugars absorb at 517 nm, producing apparent increases in absorbance that do not reflect true pro-oxidant activity.

The substantial variability across the fourteen media confirmed that antioxidant production in *M. endophyticus* is sensitive to nutrient composition and provided the basis for selecting factors for systematic evaluation in the subsequent orthogonal-array design.

### 3.2. Effect of medium components on DPPH scavenging activity (orthogonal array design)

The L81 orthogonal array enabled systematic evaluation of ten medium components and their effects on DPPH radical-scavenging activity in *Micrococcus endophyticus* SS-1. Across the 81 formulations, antioxidant activity ranged from approximately 25% to over 80% (Fig. 2A),

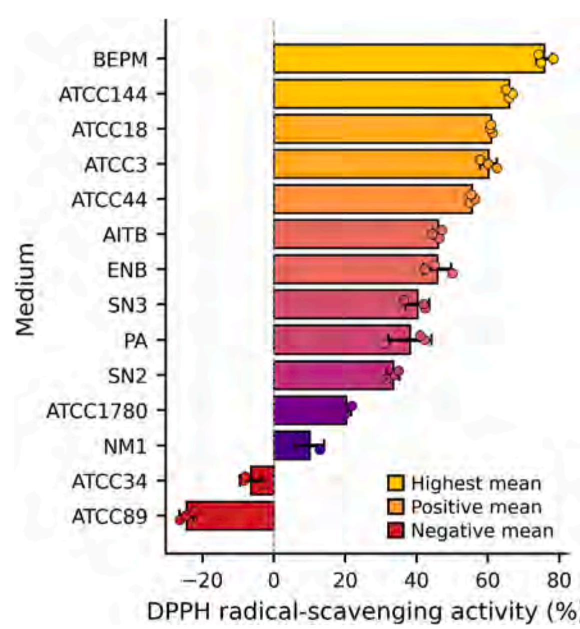


**Fig. 1.** Antioxidant activity of *Micrococcus endophyticus* SS-1 cultivated in 14 culture media. DPPH radical-scavenging activity (%) is shown as mean ± SD from biological triplicates, highlighting variability across nutrient environments.

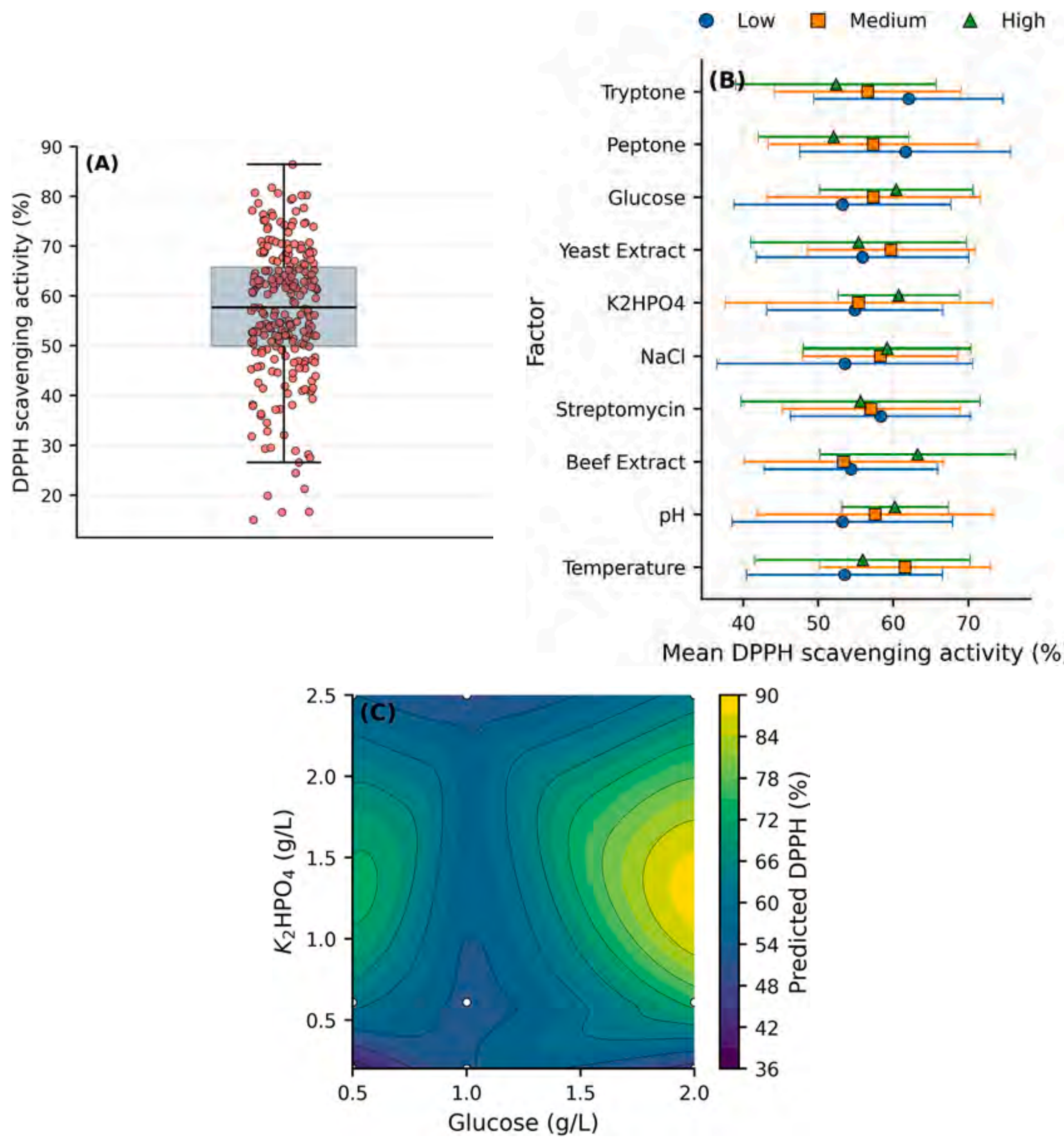


**Fig. 2.** Medium effects on DPPH scavenging activity.
(A) Distribution of DPPH radical-scavenging activity across the 81 formulations of the L81 orthogonal array.
(B) Main-effect plot showing mean DPPH activity at low, medium, and high levels of each factor.
(C) Two-factor contour plot showing predicted DPPH scavenging activity as a function of glucose and $K_2HPO_4$, with other variables fixed at mid-levels.

demonstrating substantial variation across the tested nutrient conditions.

Main-effect analysis based on mean DPPH values at each factor level (Fig. 2B) showed that glucose and $K_2HPO_4$ produced the largest changes across their tested concentration ranges. Beef extract and yeast extract produced moderate shifts, whereas pH showed a modest effect and temperature produced minimal changes in mean activity. Increasing streptomycin concentrations were associated with lower scavenging activity.

Interaction patterns between selected nutrients were examined using a two-factor response surface. The glucose × $K_2HPO_4$ surface (Fig. 2C) showed increased scavenging activity from low to intermediate concentrations of both components, followed by a plateau at higher levels. This surface represents the interaction between these two factors with all other variables held at their mid-levels.

Together, these results describe the individual and combined effects of nutrient components on DPPH scavenging activity within the L81 orthogonal array.

### *3.3. Neural network modeling for predicting DPPH scavenging activity*

A feed-forward multilayer perceptron was trained on the orthogonal array dataset to serve as a surrogate model linking medium composition to DPPH radical-scavenging activity. The input space comprised ten variables (eight medium components plus temperature and pH), and the response variable was the measured DPPH scavenging percentage for each formulation. The dataset was randomly partitioned into training, validation, and test subsets in a 70:15:15 ratio.

Model performance was first evaluated by five-fold cross-validation on the training data, yielding a mean $R^2$ of 0.94 ± 0.04 and RMSE values between 2.0 and 4.7 percentage points. Evaluation on the independent test set showed similarly high agreement between predicted and observed activities ($R^2 \approx 0.97$). The training history displayed a smooth, monotonic decrease of both training and validation loss until convergence under the early-stopping criterion (Fig. 3A). Predicted versus observed DPPH scavenging values for the test set clustered tightly around the parity line with no evident systematic bias (Fig. 3B), indicating that the neural network provided an accurate surrogate

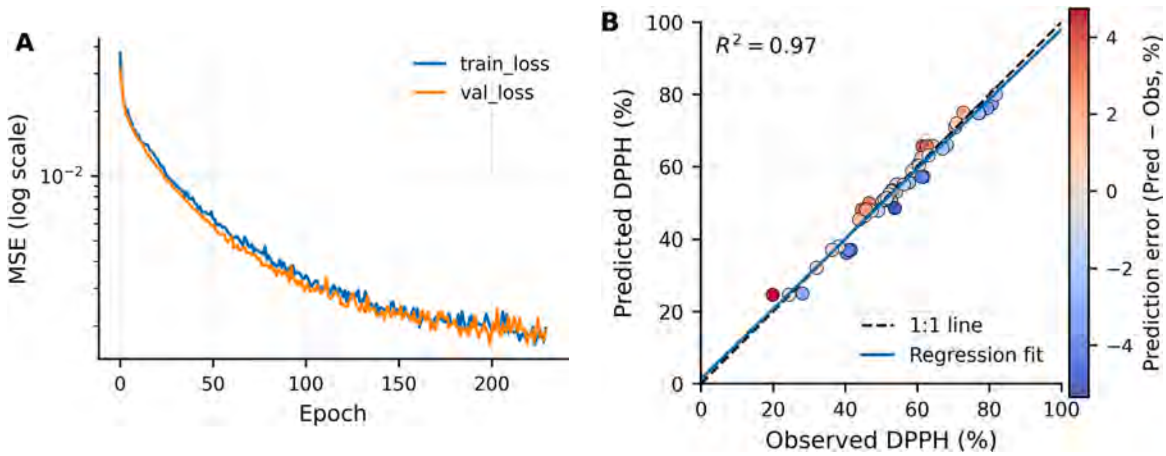


**Fig. 3.** Neural network modeling of DPPH scavenging activity.
(A) Training history showing convergence of training and validation loss during model optimization.
(B) Predicted versus observed DPPH radical-scavenging activity (%) for the test dataset, with the 1:1 reference line and regression fit.

representation of the experimental response surface suitable for subsequent optimization.

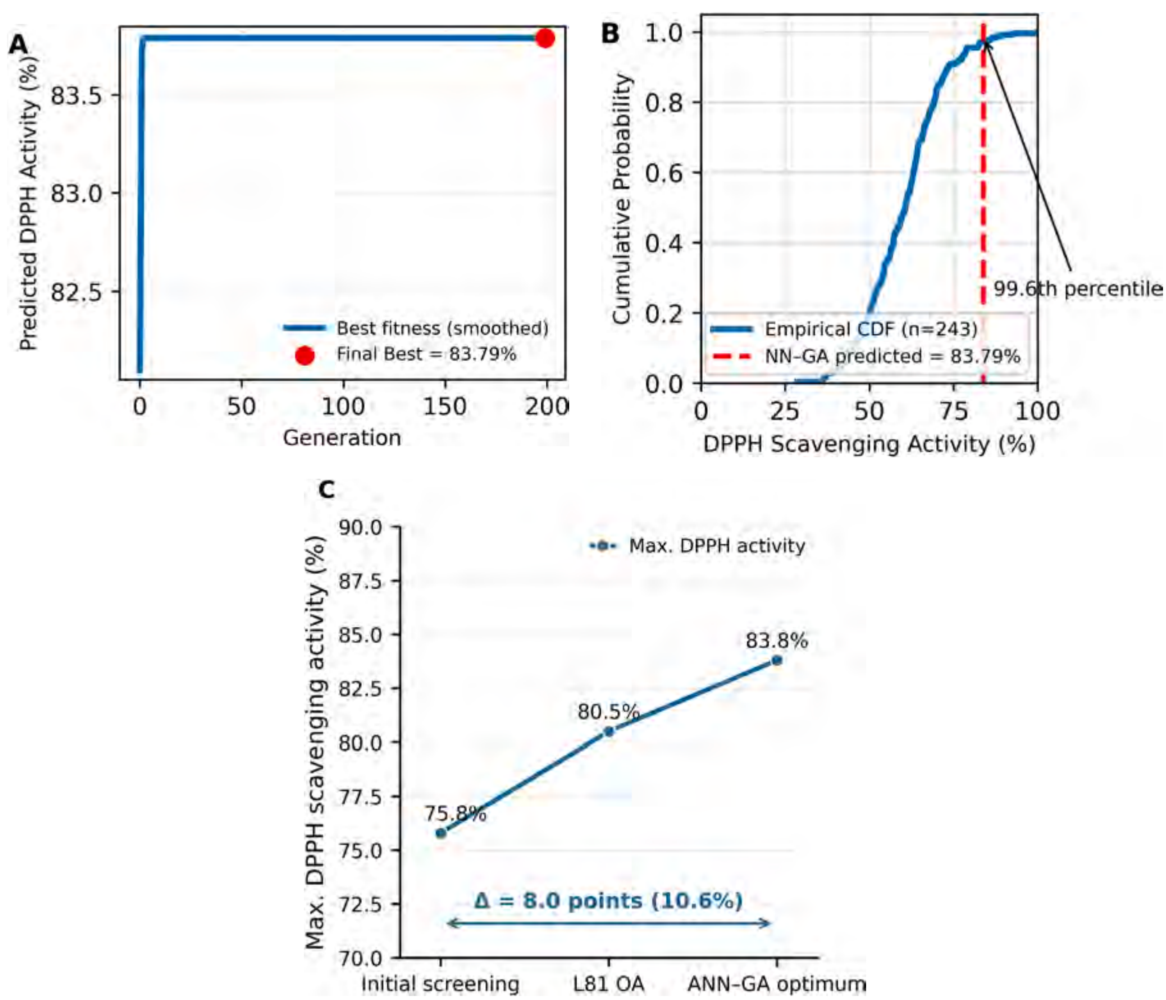


**Fig. 4.** Optimization of medium composition using the neural network–genetic algorithm framework.
(A) Convergence of predicted DPPH radical-scavenging activity during genetic algorithm optimization across generations.
(B) Empirical cumulative distribution function of experimentally measured DPPH scavenging activity values ($n$ = 243), with the neural network–genetic algorithm predicted optimum indicated.
(C) Comparison of the maximum DPPH scavenging activity obtained from the initial medium screening, the L81 orthogonal array, and the neural network–genetic algorithm–predicted formulation.

### 3.4. Optimization identifies a high-performance medium

The neural network surrogate model was integrated into a genetic algorithm (GA) to identify medium formulations predicted to maximize DPPH radical-scavenging activity. The GA explored the same concentration bounds defined in the orthogonal array and evaluated candidate solutions across 200 generations using a population size of 200, a crossover probability of 0.8, and a mutation rate of 0.1. As shown in Fig. 4A, the optimization trajectory displayed a rapid increase in predicted scavenging activity during the early generations, followed by stabilization once the population converged. A plateau was reached by approximately generation 25, indicating that the algorithm had identified a consistent optimum within the defined search space.

The empirical cumulative distribution function (ECDF) in Fig. 4B compares the GA-predicted optimum (83.79%) with all experimentally measured values ($n = 243$). The predicted optimum lies at the extreme right tail of the distribution, indicating that it represents the highest expected value relative to the observed experimental distribution.

To illustrate how performance improved across the optimization workflow, Fig. 4C compares the best-performing results from (i) the initial 14-medium screening, (ii) the L81 orthogonal array, and (iii) the ANN–GA prediction. This comparison highlights the progressive enhancement achievable through systematic experimental design and model-guided optimization.

The GA identified GA1 as the highest-scoring formulation in silico (predicted DPPH = 83.79%) within the explored bounds. Because surrogate-model predictions require experimental confirmation, representative GA-generated candidates were subsequently evaluated experimentally; the validation results and $EC_{50}$ analysis are presented in Section 3.5.

### 3.5. Experimental validation of GA-optimization and $EC_{50}$ determination

To minimize inter-batch variability inherent to DPPH assays, all experimental validation comparisons were performed within a single experimental batch under identical control and blank-correction conditions. At 100% (undiluted) culture supernatant, the GA-optimized medium exhibited higher DPPH radical-scavenging activity than both BEPM and MC72 (Fig. 5A), confirming that GA-driven optimization translated into experimentally verifiable differences in antioxidant activity.

To assess antioxidant potency, serial dilution assays were performed and $EC_{50}$ values were estimated by nonlinear logistic regression. As shown in Fig. 5B, the GA-optimized medium displayed a reduced $EC_{50}$ compared with BEPM and MC72, indicating enhanced antioxidant effectiveness at lower concentrations rather than a concentration-dependent increase in scavenging percentage alone.

## 4. Discussion

The composition of the culture medium plays a central role in regulating microbial metabolism and secondary metabolite biosynthesis, including antioxidant production. In this study, a hybrid framework integrating orthogonal experimental design with machine-learning-assisted optimization was developed to enhance antioxidant production by *Micrococcus endophyticus* SS-1. Rather than relying on exhaustive experimental screening alone, the proposed approach combines data-driven modeling with targeted experimental validation to efficiently explore a multidimensional formulation space.

Results from preliminary screening and orthogonal array experiments indicated that medium components associated with carbon availability, nitrogen sources, and phosphate concentration exerted a strong influence on DPPH radical-scavenging activity. These effects are consistent with established principles of microbial physiology, where balanced carbon–nitrogen metabolism supports both cellular growth and secondary metabolite biosynthesis. Members of the genus *Micrococcus* are known to produce antioxidant metabolites such as carotenoid pigments (e.g., sarcinaxanthin) and other redox-active compounds that contribute to oxidative stress protection. Carbon–nitrogen balance can influence the supply of metabolic precursors required for pigment biosynthesis, while phosphate availability is known to regulate secondary metabolism in many actinobacteria. In addition, environmental stresses may stimulate pigment production and other antioxidant defense mechanisms, which could contribute to the observed variation in radical-scavenging activity under different nutrient conditions. Complex nitrogen sources such as peptone and yeast extract may provide amino acids, vitamins, and growth factors that enhance metabolic flexibility, while excessive phosphate availability has been reported to repress secondary metabolism in some microbial systems.

The integration of machine-learning methods enabled the identification of influential medium parameters and the construction of a surrogate model capable of capturing nonlinear relationships between

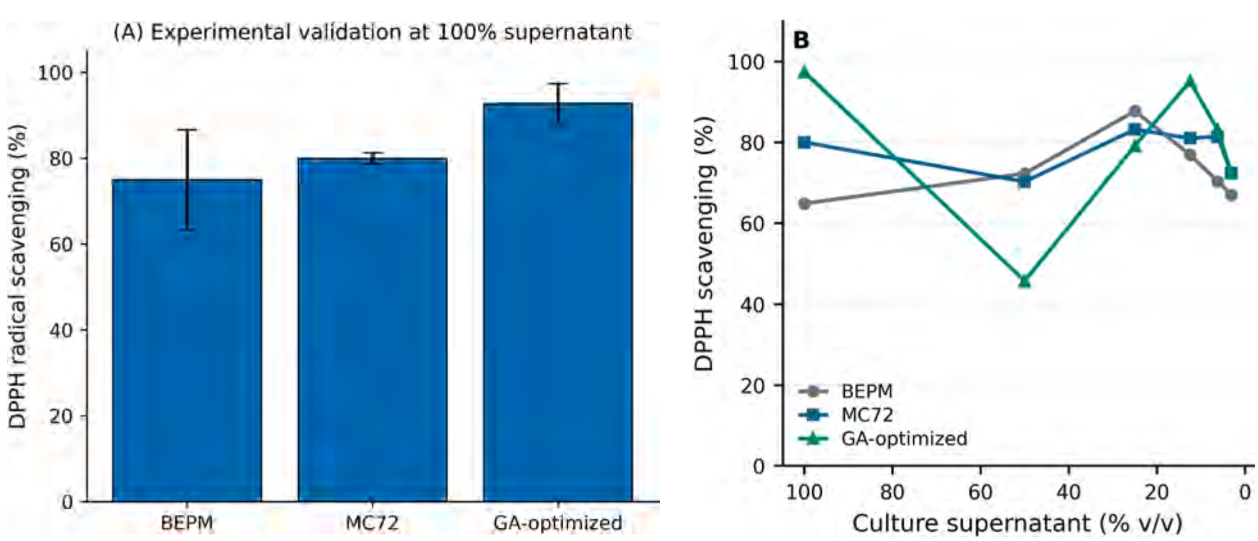


**Fig. 5.** Experimental validation and $EC_{50}$ analysis of the optimized medium.
(A) DPPH radical-scavenging activity (%) of undiluted (100% *v*/v) culture supernatants obtained from the baseline medium (BEPM), the best orthogonal-array formulation (MC72), and the experimentally validated genetic algorithm–optimized medium. (B) Dose–response curves of DPPH radical-scavenging activity across a six-point dilution series for the same media, with $EC_{50}$ values estimated by nonlinear logistic regression.

formulation variables and antioxidant response. In this framework, machine learning served as a decision-support tool rather than a substitute for experimental investigation, guiding the selection of promising formulations for experimental validation. Compared with conventional one-factor-at-a-time or response surface approaches, the surrogate-model-guided genetic algorithm provided an efficient strategy to navigate the high-dimensional design space while reducing experimental burden.

Experimental validation demonstrated that the GA-optimized medium achieved higher DPPH radical-scavenging activity than both the baseline medium and the best-performing orthogonal-array formulation. Importantly, dose–response analysis showed a reduction in $EC_{50}$ values for the GA-validated medium, indicating enhanced antioxidant potency rather than an assay-dependent increase in single-point scavenging percentages. This distinction is essential for evaluating functional improvements in antioxidant production and strengthens the process relevance of the optimization outcome.

From a bioprocess development perspective, the DOE–ML–GA framework reduced the experimental search space from tens of thousands of possible medium combinations to a manageable set of orthogonal experiments followed by targeted validation. Such an approach is particularly advantageous during early-stage process development, where experimental resources are limited and rapid identification of promising operating conditions is required.

Although the present study focused on DPPH-based evaluation of antioxidant activity, this assay was employed as a comparative proxy rather than a comprehensive characterization of antioxidant mechanisms. The antioxidant activity evaluated in this study is based on the DPPH radical-scavenging assay, which serves as a comparative proxy rather than direct chemical identification. The specific antioxidant metabolites responsible for the observed activity were not characterized in this study and remain to be identified in future work. Future work integrating complementary antioxidant assays (e.g., ABTS and FRAP), together with metabolite profiling, would provide deeper insight into the chemical nature of the bioactive compounds and the underlying metabolic pathways. Further investigation of physiological indicators such as housekeeping metabolic enzymes could also help clarify how changes in medium composition influence cellular metabolism and antioxidant biosynthesis. Systematic evaluation of cell morphology under different medium conditions may also help elucidate stress responses associated with nutrient composition. The inclusion of streptomycin in certain formulations originated from the preliminary media compositions used in the screening stage; although streptomycin-containing media influenced growth and antioxidant activity, the intrinsic resistance of *Micrococcus endophyticus* SS-1 to streptomycin was not experimentally evaluated in this study. Therefore, conclusions regarding antibiotic tolerance or resistance should be interpreted cautiously. In addition, extending validation to controlled bioreactor systems would further assess the scalability and robustness of the optimized medium under industrially relevant conditions.

## 5. Conclusion

This study establishes a hybrid experimental–machine learning framework for efficient medium optimization to enhance antioxidant production in *Micrococcus endophyticus* SS-1. Initial screening and L81 orthogonal-array experiments revealed strong medium-dependent variability in DPPH radical-scavenging activity and enabled systematic evaluation of ten nutritional and cultivation factors. A neural-network surrogate model trained on the orthogonal-array dataset captured nonlinear nutrient–response relationships and was integrated with a genetic algorithm to propose high-performance candidate formulations within the explored design space.

Importantly, experimental validation performed within a single experimental batch under identical control and blank-correction conditions confirmed that the GA-optimized medium improved antioxidant performance relative to both the baseline medium and the best orthogonal-array formulation. Dose–response analysis further demonstrated a reduced $EC_{50}$ for the GA-optimized medium, indicating enhanced antioxidant potency rather than a concentration-dependent assay artifact. Overall, the proposed DOE–ML–GA strategy provides a scalable and experimentally grounded approach for accelerating media optimization in antioxidant-producing microorganisms. Future work will focus on identifying the dominant antioxidant metabolites and expanding validation using complementary antioxidant assays (e.g., ABTS and FRAP) to strengthen mechanistic interpretation and application potential.

## CRediT authorship contribution statement

**Md Sourav Sarker:** Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Asmamaw Abat Getu:** Writing – review & editing. **Juvens Sugira Murekezi:** Writing – review & editing. **Geng Chen:** Writing – review & editing. **Jing Zhang:** Writing – review & editing. **Yong Xiao:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

## Funding sources

This work was supported by the National Natural Science Foundation of China (Grant No. 22276183); the Youth Innovation Promotion Association of the Chinese Academy of Sciences (Grant No. Y2022082); and ANSO Scholarship for Young Talents.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mimet.2026.107516.

## Data availability

All raw data, orthogonal array matrices, and analysis scripts are provided in the Supplementary Materials and at: https://github.com/MDSOURAVSARKER/DPPH-Scavenging-Activity.git

Additional datasets or analysis files generated during the study are available from the corresponding author upon reasonable request.

## References

Akbar, A., Sitara, U., Ali, I., Muhammad, N., Khan, S.A., 2014. Isolation and characterization of biotechnologically potent *Micrococcus luteus* strain from environment. Pak. J. Zool. 46, 967–973.

Atlas, R.M., 2004. Handbook of Microbiological Media. CRC Press, Boca Raton, FL. https://doi.org/10.1201/9781420039726.

Begum, S.N., Hossain, M., Adnan, M., Rahaman, C.H., Reza, A., 2023. Optimization and characterization of phenolic extraction conditions and antioxidant activity evaluation of *Adenanthera pavonina* l. bark. Plants 12, 3902. https://doi.org/10.3390/plants12223902.

Chauhan, V., Dhiman, V., Kanwar, S.S., 2021. Combination of classical and statistical approaches to enhance fermentation conditions and increase lipopeptide yield by *Pseudomonas* sp. OXDC12. Turk. J. Biol. 45, 695–710. https://doi.org/10.3906/biy-2106-59.

Chen, H.H., Zhao, G.Z., Park, D.J., Zhang, Y.Q., Xu, L.H., Lee, J.C., Kim, C.J., Li, W.J., 2009. *Micrococcus endophyticus* sp. nov., isolated from surface-sterilized *Aquilaria sinensis* roots. Int. J. Syst. Evol. Microbiol. 59, 1070–1075. https://doi.org/10.1099/ijs.0.006296-0.

Hashizume, T., Ying, B.W., 2024. Challenges in developing cell culture media using machine learning. Biotechnol. Adv. 70, 108293. https://doi.org/10.1016/j.biotechadv.2023.108293.

Karayat, M., Rahate, K., Singh, S., 2025. Update on the role of cellular redox system in health and illness. Curr. Drug Discov. Technol. 22. https://doi.org/10.2174/0115701638349511250121114323.

Leyane, T.S., Jere, S.W., Houreld, N.N., 2022. Oxidative stress in ageing and chronic degenerative pathologies. Int. J. Mol. Sci. 23, 7273. https://doi.org/10.3390/ijms23137273.

Manivasagan, P., Kang, K.-H., Sivakumar, K., Li-Chan, E.C.Y., Oh, H.-M., Kim, S.-K., 2014. Marine actinobacteria: an important source of bioactive natural products. Environ. Toxicol. Pharmacol. 38, 172–188. https://doi.org/10.1016/j.etap.2014.05.014.

Mohana, D., Thippeswamy, S., Abhishek, R., 2013. Antioxidant, antibacterial, and UV-protective properties of carotenoids isolated from *Micrococcus* spp. Radiat. Prot. Environ. 36, 168–172. https://doi.org/10.4103/0972-0464.142394.

Narayanan, H., Hinckley, J.A., Barry, R., Dang, B., Wolffe, L.A., Atari, A., Tseng, Y.Y., Love, J.C., 2025. Accelerating cell culture media development using Bayesian optimization-based iterative experimental design. Nat. Commun. 16, 6055. https://doi.org/10.1038/s41467-025-61113-5.

Nisha, S.J., Uma, G., Sathishkumar, R., Prakash, V.S.G., Isaac, R., Citarasu, T., 2025. Optimization and characterization of bioactive secondary metabolites from *Streptomyces* sp. CMSTAAHL-4. BMC Microbiol. 25, 57. https://doi.org/10.1186/s12866-025-03763-5.

Pianpumepong, P., Noomhorm, A., 2010. Isolation of probiotic bacteria from turmeric (*Curcuma longa* Linn.) and its application in enriched beverages. Int. J. Food Sci. Technol. 45, 2456–2462. https://doi.org/10.1111/j.1365-2621.2010.02337.x.

Rani, A., Saini, K., Bast, F., Mehariya, S., Bhatia, S., Lavecchia, R., Zuorro, A., 2021. Microorganisms: a potential source of bioactive molecules for antioxidant applications. Molecules 26, 1142. https://doi.org/10.3390/molecules26041142.

Sanchez, S., Demain, A.L., 2002. Metabolic regulation of fermentation processes. Enzyme Microb. Technol. 31, 895–906. https://doi.org/10.1016/S0141-0229(02)00172-2.

Shah, I., Uddin, Z., Hussain, M., Khalil, A.A.K., Amin, A., Hanif, F., Ali, L., Amirzada, M. I., Shah, T.A., Dawoud, T.M., Bourhia, M., Li, W.J., Sajjad, W., 2024. *Streptomyces* sp. from desert soil as a biofactory for antioxidants. BMC Microbiol. 24, 419. https://doi.org/10.1186/s12866-024-03586-w.

Shi, Y., Niu, X., Yang, H., Chu, M., Wang, N., Bao, H., Zhan, F., Yang, R., Lou, K., 2024. Optimization of fermentation media of *Bacillus Velezensis* BHZ-29 using Plackett–Burman design and response surface methodology. Front. Microbiol. 15, 1355369. https://doi.org/10.3389/fmicb.2024.1355369.

Valko, M., Leibfritz, D., Moncol, J., Cronin, M.T.D., Mazur, M., Telser, J., 2007. Free radicals and antioxidants in normal physiological functions and human disease. Int. J. Biochem. Cell Biol. 39, 44–84. https://doi.org/10.1016/j.biocel.2006.07.001.

Xu, F., Su, L., Wang, Y., Gao, H., Hu, K., Ben, R., Guo, Y., Li, X., Chu, J., Tian, X., 2025. AI-guided fermentation process regulation reveals metabolic rewiring in gentamicin c1a overproduction. Bioresour. Technol. 436, 132968. https://doi.org/10.1016/j.biortech.2025.132968.

Yoshida, K., Watanabe, K., Chiou, T.Y., Konishi, M., 2023. High-throughput optimization of medium composition for *Escherichia coli* protein expression using deep learning and Bayesian optimization. J. Biosci. Bioeng. 135, 127–133. https://doi.org/10.1016/j.jbiosc.2022.12.004.

Zhou, T., Reji, R., Kairon, R.S., Chiam, K.H., 2023. A review of algorithmic approaches for cell culture media optimization. Front. Bioeng. Biotechnol. 11, 1195294. https://doi.org/10.3389/fbioe.2023.1195294.