Chapter 6: Glycosidic Anthracycline Production, Optimization, Purification, and Characterization from *Streptomyces fragilis*

6.1. Introduction

Anthracyclines represent one of the most important classes of natural product chemotherapeutic agents, commonly known for their potent anticancer activity against a wide spectrum of solid tumors and hematological malignancies (Martins-Teixeira & Carvalho, 2020). It is originally produced from Streptomyces bacteria, and the characteristic red color of glycosidic anthracyclines, such as doxorubicin and daunorubicin, is primarily due to the presence of the anthraquinone chromophore (WilliamáLown, 1993). The extensive conjugation and specific electronic structure of this core enable the absorption of light in the blue-green region, resulting in their red appearance. The glycosidic attachments (sugar moieties) of these compounds do not significantly affect the color, but they influence other pharmacological properties of the drug (Grebinyk, 2021; Penasse, 2022). Despite efficacy, anthracyclines are associated with significant toxicity, particularly cardiotoxicity, and to mitigate these toxicity effects, numerous semisynthetic derivatives have been developed (W Edwardson et al., 2015). Notably, modifications at the aminodeoxy sugar moiety (L-daunosamine) have led to some of the most significant advancements in reducing cardiotoxicity while maintaining therapeutic efficacy (Grynkiewicz & Szeja, 2016).

The production of anthracyclines or bioactive compounds through fermentation, coupled with the optimization of the fermentation process, significantly enhances the yield of these valuable drugs. This approach not only boosts production efficiency but also offers an environmentally friendly alternative to traditional chemical synthesis methods. By optimizing fermentation parameters, a higher amount of bioactive compounds can be produced by minimizing environmental impact. From this perspective, producing anthracyclines from *Streptomyces fragilis* ensures a safe and eco-friendly production of anticancer agents.

Screening and selection of key nutritional components and various physical parameters are essential for developing an effective and cost-efficient culture medium for the fermentation process (Reddy Tadi et al., 2017). To achieve this, various screening methods, such as the Taguchi orthogonal array and Plackett-Burman design, can be utilized to evaluate the impact of each factor on product formation (Sy Mohamad et al., 2020). To achieve the highest product formation, bioprocesses can be optimized using various methods, focusing on identifying crucial factors and their optimum levels (Mandenius & Brundin, 2008).

To optimize metabolite production in fermentation processes, several advanced techniques were employed, including Response Surface Methodology (RSM), Artificial Neural Networks (ANN), and machine learning (ML) models like Levenberg-Marquardt (LM), Support Vector Machine (SVM), General Regression Neural Network (GRNN), Broyden-Fletcher-Goldfarb-Shanno (BFGS), and Scaled Conjugate Gradient (SCG). RSM is a statistical method that identifies optimal conditions by exploring relationships between variables, making it essential for bioprocess optimization (Tirado-Kulieva et al., 2021). ANNs, inspired by the human brain, recognize complex data patterns and predict outcomes based on historical data, proving useful for nonlinear relationships in fermentation. It uses non-linear multivariate modeling to optimize process variables in several studies (Paturi et al., 2021). The ANN approach has a significant edge over response surface methodology (RSM) models, as ANN can approximate all types of non-linear functions, while RSM can only solve quadratic functions (Silva et al., 2024). ML models like LM and SVM provide efficient and accurate predictions and optimizations, with LM excelling in solving nonlinear problems and SVM being effective for classification and regression. GRNN offers smooth transitions for regression tasks, BFGS optimizes parameter settings to enhance yields, and SCG efficiently trains neural networks without costly line searches.

The present chapter describes the production process for glycosidic anthracycline from *Streptomyces fragilis* through submerged fermentation by selecting cost-effective and readily

available carbon and nitrogen sources. The concentration of the carbon and nitrogen ratio and several physical parameters like temperature, pH, and RPM were then optimized by the RSM approach, making it highly useful for producing secondary metabolites. In addition to RSM, machine learning (ML) models (LM, SVM, GRNN, BFGS, and SCG) were also utilized to predict metabolite production. The purification and structural characterization of isolated anthracycline was done by several techniques, including HPLC, FTIR, 1H, and 13C NMR, while its thermal characteristics were elucidated by XRD, DSC, and TGA.

6.2. Materials and methods

6.2.1. Microorganism and Fermentation

The microorganism used in this study was *Streptomyces fragilis*, which was isolated from sheep wool at IIT (BHU), India, and this strain has the NCBI GenBank accession number OL823085, which was determined via 16S rRNA analysis.

For initial growth, the seed culture medium comprised the following components: 1.5% glucose, 0.75% peptone, 0.75% yeast extract, 0.5% corn steep liquor, 0.5% sodium chloride, and 0.2% calcium carbonate. To produce glycosidic anthracycline, the production media used was M:3, with the following composition: 1% soluble starch, 1% glucose, 0.75% peptone, 0.75% malt extract, 0.3% sodium chloride, 0.1% magnesium sulfate, and TSS. The pH of the medium was adjusted to 7 before autoclaving. The culture was then incubated under 200 rpm rotary agitation at 30°C for 72 hours (Shanbhag et al., 2015a).

6.2.2. Microbial Growth and Metabolite Production

Streptomyces fragilis was cultivated in growth media at 30°C for three days to assess the growth of biomass and production of metabolites. Samples were taken every 3 hours over a 72-hour period, and the biomass was measured to determine the dry cell mass (g/L). To quantify the metabolite production, the absorbance of the filtrate was measured at 498 nm

using a spectrophotometer. Subsequently, the absorbance values were converted to (g/L) using a standard curve (Thakur et al., 2009).

6.2.3. Optimization process for bioactive metabolites production

6.2.3.1. Effect of different carbon sources on metabolite production

In fermentation processes, optimizing the production medium and process conditions is crucial for determining the microbe's growth and the quantity and quality of the product (V. Singh et al., 2017). Statistical tools were used to increase metabolite production. The primary factor influencing the process is selecting the substrate used in the media. Different carbon sources, such as sucrose, fructose, lactose, molasses, dextrose, tapioca powder, maltose, and corn powder, were assessed for their ability to produce a particular result. The one that produced the most notable outcomes was selected for additional refinement, keeping other variables such as pH, temperature, and agitation speed constant.

6.2.3.2. Effect of different nitrogen sources on metabolite production

In fermentation processes, the selection of nitrogen sources is crucial as it significantly influences microbial growth and the quality of the metabolite product (N. Srivastava et al., 2019). To evaluate the efficiency of different nitrogen sources, 1% concentrations of nitrogen content equivalent to ammonium sulfate, beef extract, casein peptone, and yeast extract were used to analyze which nitrogen source is most efficient in metabolite production. Once the most effective nitrogen source was identified, it was chosen for further optimization experiments, while the physical parameters such as pH, temperature, and agitation speed were kept constant during optimization. Statistical tools were employed to maximize metabolite production, ensuring that the fermentation conditions were optimal for the selected nitrogen source.

6.2.3.3. Different techniques for production process optimization

The process variables utilized for the optimization studies were pH (X1), Temperature (X2), Agitation Rate (X3), and Carbon:Nitrogen ratio (X4), as shown in Table 6.1. The RSM and ML analyses were performed using Python 3.12.4 and MATLAB R2017a software programs, respectively.

Table 6.1: Process parameters selected for fermentation of *Streptomyces fragilis* for the production of the metabolites.

Parameters	**Level**		
	-1	**0**	**+1**
(X1) pH	6.5	7	7.5
(X2) Temperature (°C)	26	30	34
(X3) Agitation Rate (RPM)	150	200	250
(X4) C:N	0.53	1.06	1.6

6.2.3.4 Production Optimization using RSM

The statistical technique of response surface methodology based on the Box-Behnken design (BBD) was used to explore the interaction between the desired criterion and improved metabolite production (Latha et al., 2017). The data was analyzed using the Python 3.12.4 program and analysis of variance (ANOVA) to determine the relationship between the determinants and responses.

The experiment began with the Box-Behnken design (BBD) comprising 27 runs to estimate the process variables' ranges and their combined impact on metabolite production. The variables were tested at different coded levels (-1, 0, +1) as indicated in Table 6.2. The data collected from the experimental runs were then used to create a quadratic model for metabolite production, expressing the relationship between the outcome and the variables through a regression equation.

$$\left[y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \beta_{ii} x_i^2 + \sum_i \sum_j \beta_{ij}\, x_i x_j\right] \quad \text{(Equation 1)}$$

Where y represents the response variable, β_0 represents the constant, n denotes the number of factors, β_i denote linear coefficient, β_{ii} represents the quadratic coefficient of the input factor x_i and β_{ij} shows the interaction coefficient of the input factor x_i and x_j.

Table 6.2: A matrix for the Box-Behnken experimental design is presented, depicting four factors along with their coded and experimental values. Additionally, the response is shown in terms of both the experimental and RSM-predicted values for the production of metabolites. (The experimental values were calculated as the mean of triplicate values).

Run	C:N	Agitation Rate (RPM)	Temp (°C)	pH	Experimental metabolites production (g/L)	Predicted metabolites production (g/L)
1	0.53(-1)	150(-1)	30(0)	7(0)	2.880	2.789
2	0.53(-1)	250(1)	30(0)	7(0)	3.170	3.093
3	1.6(1)	150(-1)	30(0)	7(0)	3.346	3.329
4	1.6(1)	250(1)	30(0)	7(0)	3.305	3.301
5	1.06(0)	200(0)	26(-1)	6.5(-1)	3.383	3.268
6	1.06(0)	200(0)	34(1)	6.5(-1)	3.224	3.172
7	1.06(0)	200(0)	26(-1)	7.5(1)	3.687	3.607
8	1.06(0)	200(0)	34(1)	7.5(1)	3.502	3.485
9	1.06(0)	200(0)	30(0)	6.5(-1)	3.646	3.795
10	1.06(0)	250(1)	30(0)	6.5(-1)	3.178	3.250
11	1.06(0)	150(-1)	30(0)	7.5(1)	3.483	3.436
12	1.06(0)	250(1)	30(0)	7.5(1)	3.380	3.405
13	0.53(-1)	200(0)	26(-1)	7(0)	2.952	3.034
14	1.6(1)	200(0)	26(-1)	7(0)	3.592	3.584
15	0.53(-1)	200(0)	34(1)	7(0)	3.030	3.100
16	1.6(1)	200(0)	34(1)	7(0)	3.316	3.297
17	1.06(0)	150(-1)	26(-1)	7(0)	2.955	3.058
18	1.06(0)	250(1)	26(-1)	7(0)	3.435	3.452
19	1.06(0)	150(-1)	34(1)	7(0)	3.152	3.204
20	1.06(0)	250(1)	34(1)	7(0)	3.122	3.088
21	0.53(-1)	200(0)	30(0)	6.5(-1)	3.214	3.179
22	1.6(1)	200(0)	30(0)	6.5(-1)	3.131	3.112
23	0.53(-1)	200(0)	30(0)	7.5(1)	3.017	3.069
24	1.6(1)	200(0)	30(0)	7.5(1)	3.816	3.883
25	1.06(0)	200(0)	30(0)	7(0)	4.497	4.477
26	1.06(0)	200(0)	30(0)	7(0)	4.469	4.477
27	1.06(0)	200(0)	30(0)	7(0)	4.540	4.477

The statistical analysis was conducted using an analysis of variance (ANOVA). Fisher's F-test was used to establish the overall significance of the model, accompanied by the associated probability p(F), correlation coefficient R, and determination coefficient R2, which assesses the accuracy of the regression model's fit. Additionally, the analysis involved the use of Student's t-value to assess the estimated coefficients and their corresponding probabilities p(t). Surface plots were used to represent the quadratic models for each variable.

The optimal combinations were determined based on the surface plots. The present study is centered on optimizing the production conditions for metabolites produced by *S. fragilis* through the application of statistical techniques. Response surface methodology (RSM) based on the Box-Behnken design (BBD) was an effective statistical approach for assessing the interaction between the four key factors and enhancing metabolite production.

To verify the interaction between the factors and responses, the data was analysed using the Python 3.12.4 software and an analysis of variance (ANOVA).

6.2.3.5 Process optimization using Machine Learning methodology

Researchers have introduced the uniform design (UD) method to obtain statistics suitable for real-world production situations. The UD approach has been successfully utilized in various optimization processes (Sibalija & Majstorovic, 2012). The alternative methods are artificial intelligence-based methodologies, including machine learning (ML) algorithms and genetic algorithms (GA). ML algorithms have been effectively utilized for media optimization and can simulate various aspects of biological information processing, providing an effective approach for studying nonlinear problems (Qiu et al., 2016)

ML algorithms are mathematical models that represent the connections within experimental data through iterative computation, as opposed to regression equations that are derived from a function of the experimental data. Due to their higher modeling accuracy and improved generalization skills, ML algorithms can simulate the bioprocess and forecast outcomes (Duong-Trung et al., 2023). ML algorithms can model all nonlinear multivariate functions, unlike standard statistical techniques that can only do so for quadratic functions (Kanevski et al., 2004).

According to other reports, ML algorithms are often more accurate than RSMs. UDs typically contain patterns that are representative and evenly spaced. ML algorithms can produce equally accurate models based on these high-quality patterns, using much less data than expected (Miriyala et al., 2022). A predictive machine learning model with a feed-forward architecture is a powerful tool for various applications, including optimizing metabolite production in microorganisms (Mondal et al., 2023). These models consist of an input layer, hidden layers, and an output layer. Each input gets a specific weight (W) and bias (b). The sum of these weighted inputs and bias is used as the input for the transfer functions.

6.2.3.6 BFGS (BFGS Quasi-Newton Algorithm)

The Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton algorithm is employed for training ML models due to its efficiency in handling large optimization problems. In the context of metabolite production, a neural network is created, trained, and evaluated using experimental data. The network consists of an input layer with several independent variables, a hidden layer, and an output layer predicting metabolite yield. BFGS adjusts weights and biases to minimize errors between predicted and actual outputs. Model performance is evaluated using metrics like mean squared error (MSE), root mean squared error (RMSE), and correlation coefficients (R, R^2) (Rafati & Marica, 2020).

6.2.3.7 GRNN (General Regression Neural Network)

GRNN is applied to solve function approximation challenges in complex systems. GRNNs consist of an input layer, pattern layer, summation layer, and output layer. The input layer receives the variables, which are then processed through the pattern layer, where each neuron corresponds to a training sample. The summation layer aggregates these inputs, and the output layer provides the predicted value. In metabolite production, GRNNs predict yields based on input conditions, ensuring smooth interpolation and high accuracy. The performance is assessed using similar metrics, ensuring the network effectively predicts the optimum production conditions (Kartal et al., 2018).

6.2.3.8 LM (Levenberg-Marquardt Algorithm)

The LM algorithm is another effective training method for ML models, particularly useful for small to medium-sized datasets. An ML model trained using the LM algorithm includes an input layer with variables, a hidden layer, and an output layer predicting metabolite yields. The LM algorithm combines gradient descent and the Gauss-Newton method, adjusting weights and biases to minimize prediction errors. This approach is efficient and provides accurate results, with model performance evaluated through MSE, RMSE, and correlation coefficients (Abubakar et al., 2016).

6.2.3.9 SCG (Scaled Conjugate Gradient Algorithm)

The SCG algorithm is advantageous for large-scale problems due to its memory efficiency and speed. In an ML model, the SCG algorithm is used to train a network with an input layer, a hidden layer, and an output layer for predicting metabolite production. SCG scales the conjugate gradient method, making it suitable for extensive datasets typical in bioprocess optimization. The algorithm adjusts the weights and biases to minimize prediction errors, and the model's performance is measured using MSE, RMSE, and correlation coefficients (Al-Arbo & Al-Kawaz, 2021).

6.2.3.10 SVM (Support Vector Machine)

SVM is an ML model used for classification and regression tasks. An SVM model created to predict metabolite production includes input variables such as microbial strain, nutrient levels, and environmental conditions. SVM finds the best hyperplane to separate data into classes or predict continuous values with minimal error. The model's performance is evaluated based on accuracy, precision, and recall, ensuring it effectively categorizes production conditions into high, medium, or low yield. This predictive capability aids in optimizing fermentation processes, enhancing the yield and quality of metabolites (Khaleghi et al., 2021).

A predictive machine learning model was created, trained, and evaluated using the BBD experimental set. The model employed a neural network with an input layer consisting of four independent variables, a hidden layer with 6 neurons, and an output layer with enzyme activity as the response variable. In the neural network model, each input in the input layer is assigned a weight (W) and a bias (b). The sum of the weighted inputs and the bias is used as the input to the transfer functions. The hidden layer uses the tangent sigmoidal (tansig) transfer function, while the output layer uses the linear (purelin) transfer function for optimal results. The data was split into three sets by the network in a 70:15:15 ratio, comprising a training set, a validation set, and a test set. Five different training algorithms were used to train the neural network to predict the optimal production of metabolites (g/L) under the given experimental conditions. The study evaluated five training algorithms: Levenberg-Marquardt (LM), scaled conjugate gradient (SCG), BFGS quasi-Newton algorithm (BFGS), General Regression Neural Network (GRNN), and Support Vector Machine (SVM). The results, depicted as a predictive bioprocess model using machine learning, were assessed by achieving a low mean squared error (MSE)

(equation 2), a low root mean squared error (RMSE) (equation 3), and high values for the correlation coefficient (R) and regression coefficient (R^2).

Mean Squared Error (MSE):

The MSE measures the average of the squares of the errors or deviations between the actual values and the predicted values. It is given by:

$$\left[\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2\right] \qquad \text{(Equation 2)}$$

Root Mean Squared Error (RMSE):

The RMSE is the square root of the average of squared differences between the predicted and actual observations. It quantifies the average magnitude of the error. It is given by:

$$\left[\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2}\right] \qquad \text{(Equation 3)}$$

Where Y_i and $\widehat{Y}_i$ are the experimental and predicted values, respectively

6.2.4 Validation of the experimentally derived model

Validation and comparison of the optimized process variable levels obtained from both the BBD and ML-based optimizers were carried out by conducting experimental runs in triplicate.

6.2.5 Downstream process and detection of Anthracycline Epirubicin

The downstream process of glycosidic anthracycline was performed using a protocol elucidated by Madduri et al., with some modifications (Madduri, Kennedy, Rivola, Inventi-Solari, Filippini, & Zanuso, G., ... & Hutchinson, 1998). After an incubation period of three days, the broth culture of *Streptomyces fragilis* was subjected to sonication for 30 minutes, alternating between 10 minutes of sonication and 5 minutes of rest cycles. The resulting mixture was centrifuged at 10,000 rpm for 20 minutes at 4 degrees Celsius. The supernatant obtained after centrifugation was adjusted to a pH of 8.2 and mixed with

the solution of dichloromethane and methanol (80:20) in a 1:1 ratio. The resulting solvent phase was collected and combined with 0.02 M hydrochloric acid in a ratio of 1:0.16 (solvent: 0.02 M HCl). The mixture was incubated for 10 minutes at a temperature of 30 degrees Celsius, and the aqueous phase was collected and adjusted to a pH of 5.

HPLC analysis was carried out to detect and isolate epirubicin (anthracycline) from the extract obtained from *Streptomyces fragilis.* It was performed using Shimadzu LC-20AD dual binary pumps and Shimadzu Prominence SPD-M20A PDA detector equipped with a C-18 reversed-phase column (Phenomenex, Gemini 5µ, 250 mm length × 4.6 mm internal diameter). Two different solvents, namely solvent A (water at pH 3 adjusted by orthophosphoric acid) and solvent B (a mixture of methanol and acetonitrile in a ratio of 20:80), were mixed in a 60:40 ratio, which was filtered through a 0.45 µm membrane filter and utilized as mobile phase (Madduri, Kennedy, Rivola, Inventi-Solari, Filippini, & Zanuso, G., ... & Hutchinson, 1998). The resulting solvents and extracts were subjected to degassing using an ultrasonic method before conducting HPLC. During the experiment, a 20 µL sample was taken as the injection volume, and the pump was set to a flow rate of 0.4 mL/min. Epirubicin detection was performed using a wavelength of 254 nm. To confirm the presence of epirubicin in the extract of *Streptomyces fragilis*, a comparative analysis was conducted by comparing the retention times and peak areas with those of pure epirubicin.

6.2.6 Structural Elucidation and Thermal Characteristics of Epirubicin

6.2.6.1. FTIR Spectral Analysis

Fourier Transform Infra-Red Spectroscopy was conducted for pure epirubicin, and epirubicin was separated by the HPLC technique of *Streptomyces fragilis*. The analysis was performed with a Nicolet Is5 model provided by the Central Instrument Facility at the Indian Institute of Technology, BHU. To prepare the sample, a 1:100 ratio of KBr was

mixed with the samples, and the resulting mixture was pulverized with a mortar and pestle. The resulting thin pellet was placed on a sample holder and scanned in the range of 400 to 4000 cm^{-1} wave number to obtain the characteristic absorption spectra of the sample (Sravan Kumar et al., 2015).

6.2.6.2. 1H and 13C Nuclear Magnetic Resonance

The peak fraction obtained from HPLC analysis was characterized using 1H NMR and 13C NMR techniques to determine the structure of epirubicin. The analysis was carried out utilizing AVH D 500 AVANCE III HD 500 MHz OneBay NMR Spectrometer (Bruker BioSpin International AG) operating at 500 MHz, by dissolving 20 mg epirubicin in 1 mL deuterated DMSO (Shukla et al., 2023).

6.2.6.3. High-Resolution Mass Spectroscopy

The high-resolution mass spectra of Epirubicin were determined using the Thermo Fischer Scientific Q-Exactive Plus Biopharma-High Resolution Spectrometer (HR-MS). Full-scan MS spectra were collected in the mass range of 200-2000 m/z, and the data were analyzed to identify the molecular ion peak and confirm the molecular formula (Liu et al., 2017).

6.2.6.4. X-ray diffraction (XRD) analysis

To determine the crystalline or amorphous structure of epirubicin, XRD analysis was conducted using a Rigaku Miniflex 600 Desktop X-Ray Diffraction System (RIGAKU Corporation) with Cu Kα radiation and nickel filter. The scanning range used was $2\theta = 5°$ to 90° with a scanning speed of 5°/min and a scan step of 0.02° at room temperature (Jangid et al., 2012).

6.2.6.5. Differential Scanning Calorimetry (DSC)

DSC was used to study the melting process of epirubicin using a differential scanning calorimeter (DSC-60 plus, M/s Shimadzu, Asia Pacific). The equipment was calibrated using indium, and an empty, sealed, but pierced aluminium pan was used as a reference. The experiment was conducted using 99.9% pure nitrogen gas with a flow rate of 100 mL/min. The heating rate was constant, typically from room temperature to 350°C (Cedeño et al., 2001; Colombo et al., 2010).

6.2.6.6. Thermogravimetric analysis (TGA)

TGA was employed to investigate the alterations in the physical and chemical properties of epirubicin when exposed to elevated temperatures. It was conducted using TGA-50 (M/s Shimadzu, Asia Pacific) with a sample weight of approximately 3.627 mg in pin-holed platinum crucibles. The experimental procedure involved subjecting the sample to a heating rate of 6 °C/min, starting from 25°C and reaching a maximum temperature of 600°C. Throughout the process, a continuous flow of nitrogen purging was maintained at a rate of 100 mL/min (Akram et al., 2023).

6.3. Results and Discussions

6.3.1. Microbial Growth and Metabolite Production

The growth and metabolite production curves for *S. fragilis* illustrated the progression of biomass accumulation and secondary metabolite, specifically anthracycline production, over a 72-hour period, as represented in Figure 6.1. The biomass growth curve was determined by measuring the dry cell mass every 3 hours, while metabolite production was assessed using a spectrophotometer at 498 nm. The biomass growth curve, indicated by the black line, exhibited a rapid increase in cell mass of *S. fragilis*, beginning at 6 hours and continuing to rise sharply until 24 hours. After this initial stage, the growth rate decreased, but biomass continued to rise steadily, reaching a peak of 7 g/L by the 72-hour mark. This pattern represented a typical bacterial growth curve, showing a log phase

followed by a stationary phase, indicating that the initial 24 hours provided optimal nutrient availability and growth conditions for *S. fragilis* (Ghaly et al., 2003)

The production of secondary metabolites (anthracyclines) from *S. fragilis* began to increase significantly after 24 hours, as represented by the red line. The increase in metabolite production continued until it reached 1.6 g/L at around 66 hours, after which the production rate plateaued. The lag in metabolite production relative to biomass growth suggested that secondary metabolite production was associated with the stationary phase of growth (Williams et al., 2012). The shape of the production curve indicated that once *S. fragilis* reached a certain biomass threshold and environmental conditions triggered secondary metabolism, it started accumulating secondary metabolites.

6.3.2. Production of secondary metabolites using different carbon sources

The *S. fragilis* was employed to produce bioactive secondary compounds through a batch fermentation process. To evaluate the effects of different carbon sources on metabolite production, different types of carbon sources were optimized in the fermentation media while keeping all other factors constant during the experiment.

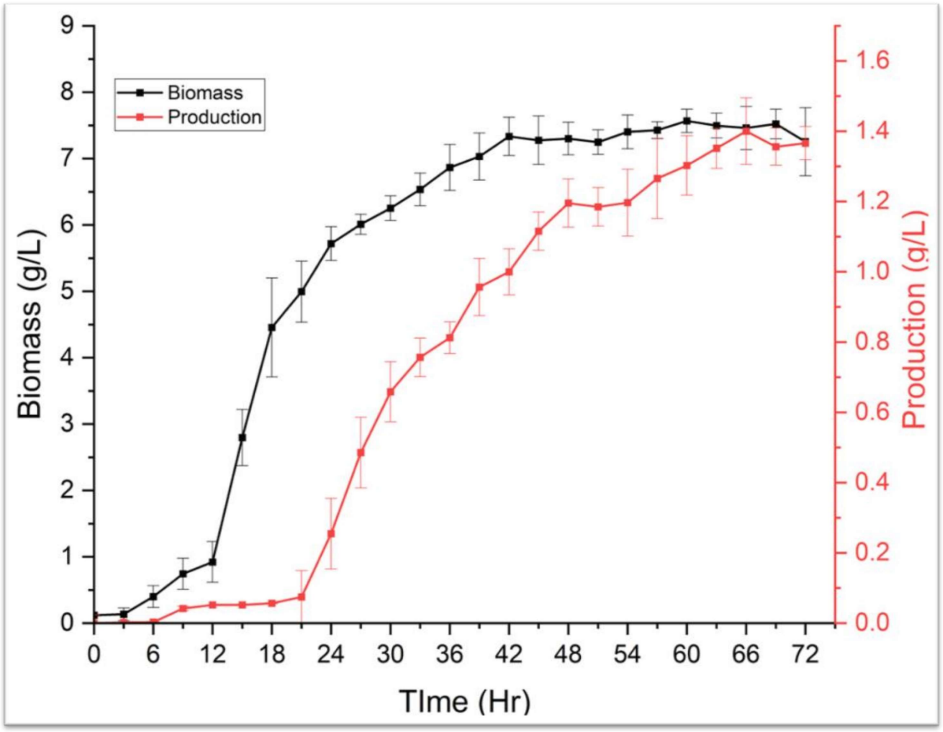

Figure 6.1: The growth curve (black line) and metabolite production curves (red line) for *S. fragilis* represent the biomass accumulation and secondary metabolite production over a 72-hour period.

During the fermentation process, the primary factor influencing cellular material and metabolite production is the carbon source. Microbes generate secondary metabolites in response to the challenging conditions encountered during fermentation. These metabolites act as a protective barrier against the acidity or alkalinity of the environment (Parekh et al., 2000). The influence of different carbon sources on metabolite production was observed and is depicted in Figure 6.2. The production medium was supplemented with sugarcane molasses, fructose, lactose, sucrose, maltose, dextrose, corn powder, and tapioca powder. Among these, molasses was the most effective, yielding 4.54 g/L, followed by lactose at 4.1 g/L. Cornflour exhibited the lowest production yield at 0.89 g/L. The utilization of sugarcane molasses, an agricultural by-product, can substantially reduce production costs. Molasses and tapioca powder possess carbon contents of about 24% and 40%, respectively, while corn flour has a carbon content of around 46%. Molasses is composed of 75% carbohydrates and 22% water, with no protein or fat. The

composition of molasses includes sucrose (29% carbohydrate), glucose (12%), and fructose (13%) (Shukla et al., 2022).

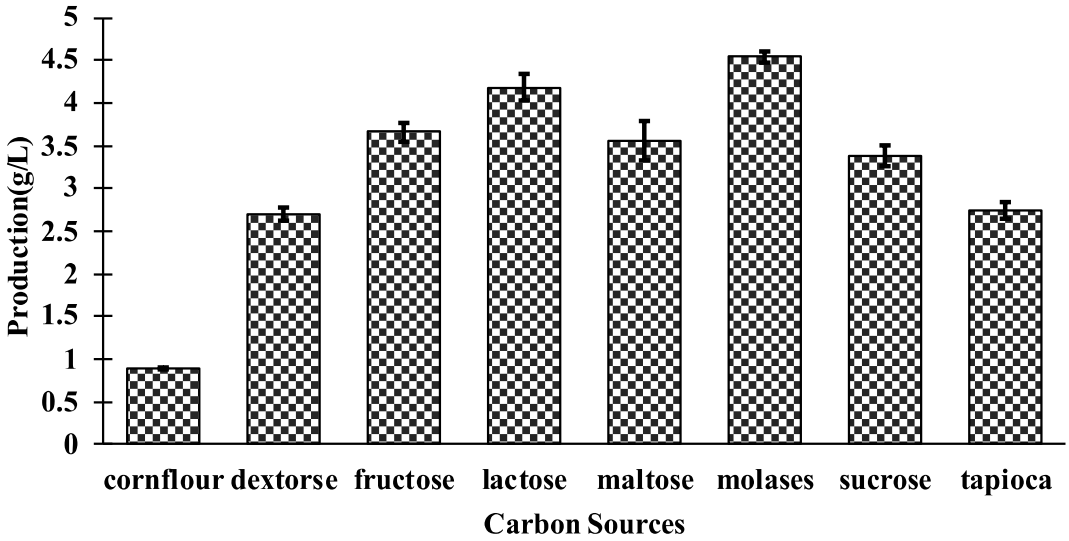

Figure 6.2: Effect of different carbon sources on metabolite production using *S. fragilis*, along with pH 7, agitation rate 200, and temperature 30℃.

6.3.3. Effect of Nitrogen on the production of secondary metabolites

The impact of various nitrogen sources on metabolite production was investigated using ammonium sulfate, peptone, casein, beef extract, and yeast extract, which were added at 1% (w/w) concentration. Figure 6.3 illustrates the effect of these nitrogen sources on secondary metabolite production, measured in grams per liter (g/L). The results reveal notable variations in metabolite production across the different nitrogen sources. Yeast extract was the most effective nitrogen source, resulting in the highest metabolite production of 2.7 g/L. Beef extract was the next most effective, with 2.5 g/L of metabolite production. Ammonium sulfate displayed moderate effectiveness, resulting in a production of 2.05 g/L of metabolites. Peptone and casein were the least effective, with peptone yielding 1.8 g/L and casein around 1.5 g/L. These findings align with previous results demonstrating that nitrogen sources like yeast extract can significantly enhance the production of bioactive metabolites in microbial fermentation (Chatterjee & Mazumder, 2023; S. N. Kumar et al., 2012).

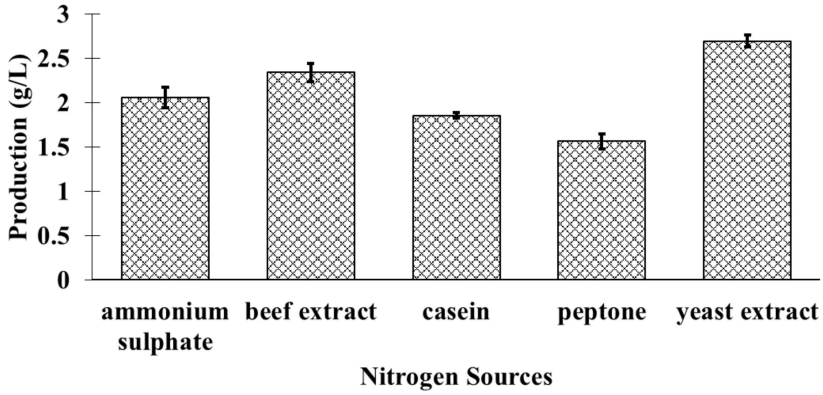

Figure 6.3: Effect of different nitrogen sources on metabolite production using *S. fragilis*, along with pH 7, agitation rate 200, and temperature 30℃.

6.3.4. Maximizing the production of metabolites using the RSM technique.

In the current analysis, the Box-Behnken Design (BBD) of a second-order polynomial equation was brought into work to escalate metabolite production. The boundaries of these preliminary frameworks were established based on an individual criterion selection trial. The computational analysis of metabolite production is as follows:

$$Y_{Production} = -127.1 + 0.109M + 0.0503A + 2.032T + 27.00P - 0.02619\ M*M - 0.000070A*A - 0.03486T*T - 1.995P*P - 0.000166M*A - 0.00442M*T + 0.0883M*P - 0.000319A*T - 0.00148A*P + 0.0214T*P$$

Where M represents Molasses percentage, A represents Agitation Rate, P represents pH, and T represents Temperature.

Table 6.3: ANOVA analysis of quadratic model

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	14	5.25	0.37	41.97	0.000
Linear	4	0.78	0.19	21.81	0.000
C:N	1	0.42	0.42	46.91	0.000
AgitationRate	1	0.05	0.05	5.37	0.039
Temp	1	0.04	0.04	4.13	0.065
pH	1	0.30	0.30	34.02	0.000
Square	4	4.36	1.09	122.06	0.000

C:N*C:N	1	2.33	2.33	261.44	0.000
AgitationRate*AgitationRate	1	2.45	2.45	274.60	0.000
Temp*Temp	1	1.83	1.83	205.19	0.000
pH*pH	1	1.33	1.33	148.98	0.000
2-Way Interaction	6	0.34	0.06	6.29	0.004
C:N*AgitationRate	1	0.03	0.03	3.08	0.105
C:N*Temp	1	0.03	0.03	3.50	0.086
C:N*pH	1	0.19	0.19	21.73	0.001
AgitationRate*Temp	1	0.07	0.07	7.30	0.019
AgitationRate*pH	1	0.02	0.02	2.08	0.175
Temp*pH	1	0.00	0.00	0.02	0.895
Error	12	0.11	0.01		
Lack-of-Fit	10	0.10	0.01	8.28	0.113
Pure Error	2	0.00	0.00		
Total	26	5.35			

Where DF- Degree of freedom, Adj. SS- Adjusted sum of squares, Adj. MS- adjusted mean of squares.

The results from the experimental design show that the C-N ratio (P-value < 0.0001) was the most significant factor influencing metabolite production. The linear and quadratic model terms of the variables were both highly significant, as seen from their P-values in Table 6.3. A Fisher F-test value of 41.97 indicates that the model was highly significant. The lack-of-fit (LOF) P-value of the model was 0.1125 (insignificant), suggesting that the regression model adequately explains the functional interdependence between the experimental factors and the response values. The RSM model had an R^2 value of 98.00%, an adjusted R^2 of 95.66 %, and a predicted R^2 value of 89.72%. Surface plots were created and analyzed to determine the optimal levels of physical process parameters affecting metabolite production (**Figure 6.4**).

Through the utilization of 3-D surface plots, the polynomial equation was employed to visually represent the variations in responses resulting from the combined influences of the factors. An investigation into the interplay of molasses, temperature, pH, and agitation

rate was conducted, and all the combined studies indicate that molasses may have a role in the production process. The ANOVA analysis of the RSM model is presented in Table 6.3. The ANOVA analysis showed that metabolite production was better than average. The highest predicted metabolite production was 4.65 g/L with a C-N ratio of 1.157, a temperature of 29.72°C, a pH of 7.1, and an agitation rate of 203 rpm. The experiments were conducted in triplicate to ensure accuracy, and the maximum metabolite production was approximately 4.59 ± 0.13 g/L.

The application of agricultural waste cane molasses was tested as a replacement for glucose as the carbon source in a predetermined medium for the production of bioactive compounds. Various production conditions, such as pH, temperature, and agitation rates, were examined individually. Additionally, different concentrations of molasses were tested to improve and increase production yield. After comparing productivity with molasses and glucose, the overall productivity saw an enhancement, resulting in a final yield of 4.59 ± 0.13 g/L. It has been noted that molasses contains various types of sugar, with sucrose being the primary sugar. Previous studies have suggested that molasses is a better carbon source than glucose for metabolite production (S. Zhang et al., 2021).

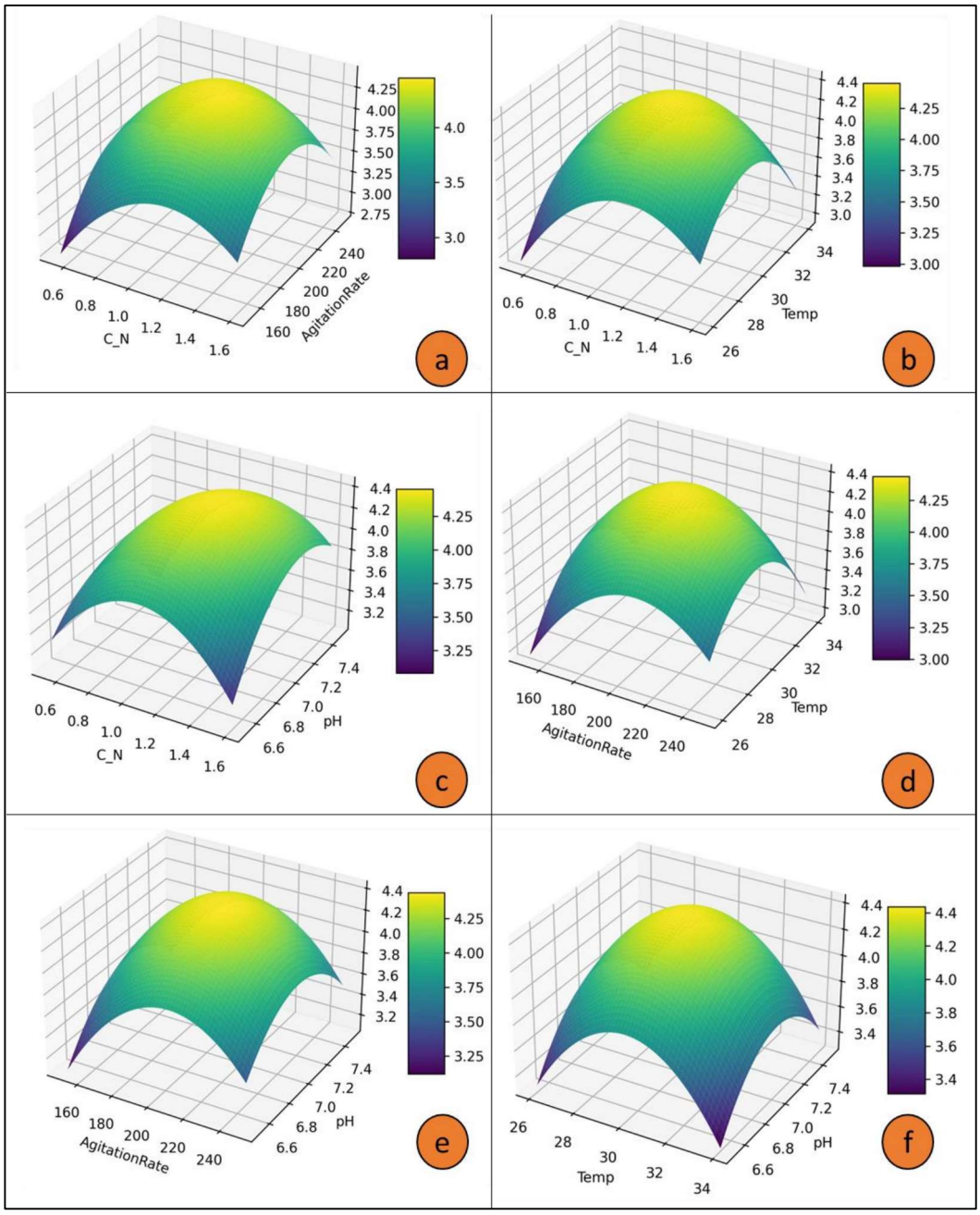

Figure 6.4: Response surface plots of the quadratic model depicting the interactive effects between the process parameters. Plots showing interaction between (a) Molasses (%) and Agitation Rate (RPM) (b) Agitation Rate (RPM) and temperature (°C) (c) Molasses (%) and pH (d) Agitation Rate (RPM) and temperature (°C) (e) Agitation Rate (%) and pH (f) temperature (°C) and pH.

6.3.5. Prediction of metabolite production using machine learning.

The ML modeling began with an architecture consisting of 4 inputs (mentioned process parameters) and utilized metabolite production (g/L) as the output to generate the predictive bioprocess model. The model was trained using five different algorithms:

General Regression Neural Network (GRNN), BFGS quasi-Newton (BFGS), Scaled Conjugate Gradient (SCG), Support Vector Machine (SVM), and Levenberg-Marquardt (LM).

The SVM model is widely used for classifying and regressing data in supervised learning tasks. The SCG method is an advanced approach to supervised learning that steers clear of the need for complex line search methods and is utilized through the trainscg function in neural networks. The LM technique, also known as a damped least squares method, finds extensive application in solving problems involving the sum of squared errors and is implemented through the trainlm function in neural networks (Suryawanshi et al., 2019). The GRNN serves as a type of neural network utilized for regression tasks, offering seamless transitions between data points. BFGS stands as a popular method for accelerating the convergence of neural networks, and it is applied through the train_bfgs function (Chaki, 2019).

Figure 6.5 represents the SVM model's performance, which shows a strong correlation between experimental and predicted values (section A). Section B and C illustrate that residuals are randomly distributed around zero, indicating minimal systematic errors. In this model, Performance metrics are Mean Squared Error: 14.0909, Root Mean Squared Error: 3.7538, Correlation Coefficient: 0.99825, and Coefficient of Determination: 0.9965, which highlights the model's high accuracy, predictive power, and consistency.

The performance of the GRNN model is depicted in Figure 6.6. It shows a high correlation coefficient (R) and coefficient of determination (R^2), along with a low MSE and RMSE, which indicate the GRNN model's exceptional predictive accuracy and consistency. Performance metrics include MSE: 14.2121, RSME: 3.7699, R: 0.99976, and R^2: 0.99953. The residual plots (Section B and C) further confirm the model's robustness by showing a random scatter of residuals around zero and minimal systematic errors.

Figure 6.7 represents a comprehensive performance evaluation of the BFGS model. Section A displays scatter plots that compare predicted values to actual target values for training (R=0.94743), validation (R=0.20153), test (R=0.97096), and overall datasets (R=0.94642). The training and test plots show strong correlations, while the validation plot indicates lower predictive accuracy. Section B presents MSE over 25 epochs, with the best validation performance at epoch 19 (MSE=0.016398). Section C compares predicted and experimental production values, demonstrating strong alignment. These data highlight the model's high accuracy in training and testing, but lower accuracy in validation.

Figure 6.8 provides a comprehensive performance evaluation of the SCG model, depicted through three sections: A, B, and C. In this, section A compares predicted and actual target values for training (R=0.94597), validation (R=0.97221), test (R=0.98057), and overall datasets (R=0.95177), showing strong correlations and high predictive accuracy. Section B represents the MSE plot over 28 epochs, with the best validation performance at epoch 22 (MSE=0.036337). Section C compares predicted metabolite production values to experimental values, confirming strong predictive accuracy. Thus, the SCG model demonstrates high accuracy across all datasets, as evidenced by the correlations in Section A and the optimal performance highlighted in Section B.

Figure 6.9 represents a performance evaluation of the LM model, in which section A shows scatter plots of predicted versus actual values for training (R=0.99994), validation (R=0.97396), test (R=0.99955), and overall datasets (R=0.99569), showing very high predictive accuracy. Section B tracks MSE over 3 epochs for training, validation, and testing datasets, with the best validation performance achieved at epoch 1 (MSE=0.011282). Section C compares predicted and experimental metabolite production values, showing a close alignment that confirms the model's strong predictive accuracy.

Overall, the LM model demonstrates exceptional performance across all datasets, with strong correlations and minimal errors.

In comparison, the LM model shows exceptional performance with nearly perfect predictive accuracy. Its MSE is impressively low at 0.011282, indicating optimal performance early in the training process. The SVM model and GRNN model also perform well but have higher MSE and RMSE values compared to the LM model, indicating slightly less precision. The SCG model shows high predictive accuracy but is slightly less accurate than the LM model. The BFGS model performs the worst, with significantly lower validation accuracy (R=0.20153), making it the least reliable. Based on these evaluations, the LM model was selected for further analysis due to its superior accuracy, minimal errors, and consistent performance across all datasets. This study is also consistent with previous research on process optimization using machine learning, in which different training algorithms were compared, and the Levenberg-Marquardt algorithm was found to yield the best optimization results (Suryawanshi et al., 2019).

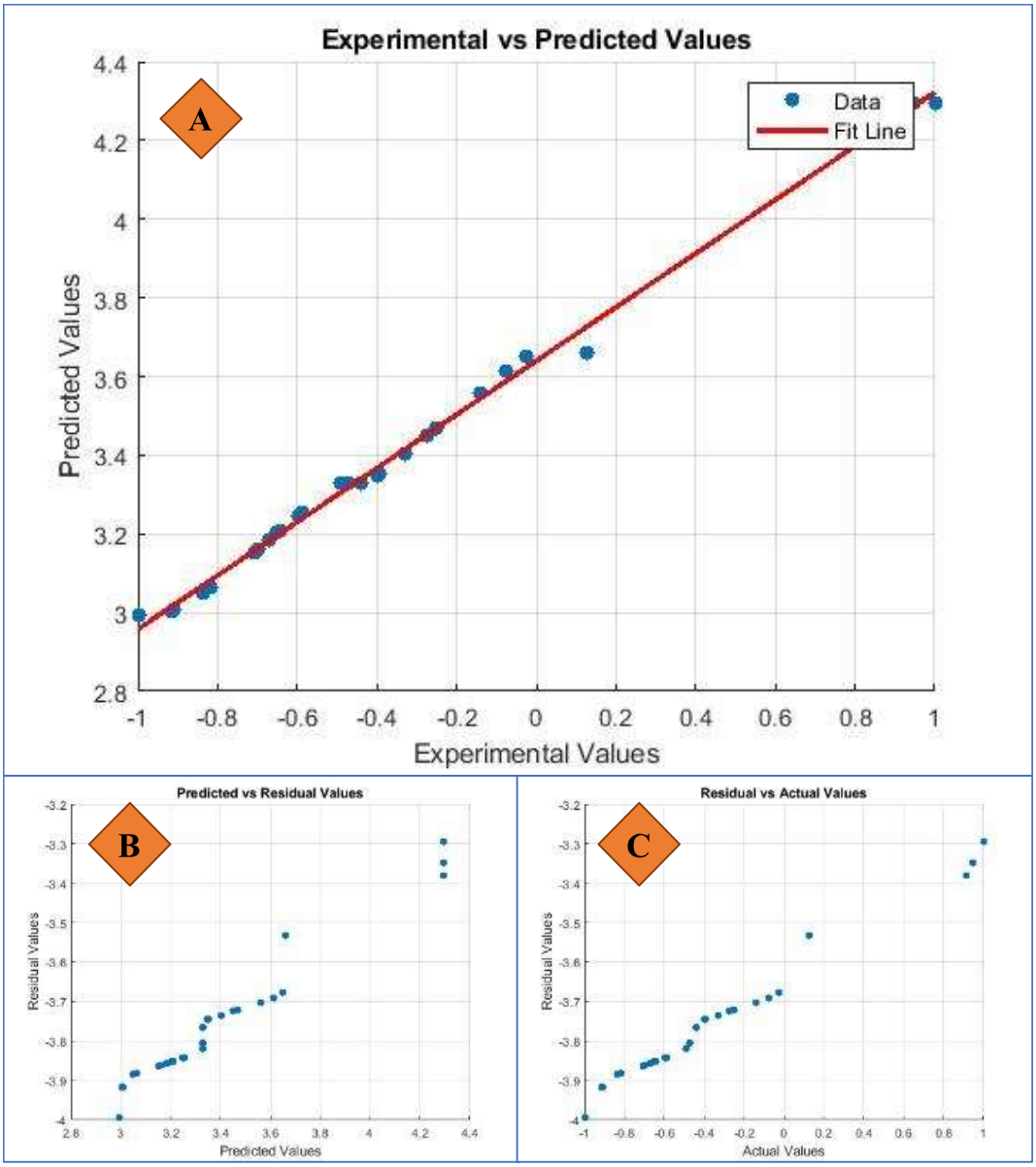

Figure 6.5: This figure presents a comprehensive performance evaluation of the SVM (Support Vector Machine) model. Section A displays a scatter plot of predicted versus experimental values with a fit line, demonstrating a strong alignment between predicted and actual values. Section B shows a plot of residuals versus predicted values, indicating the distribution of errors relative to the predicted values. Section C presents a plot of residuals versus actual values, further illustrating the error distribution across the range of actual values. These visualizations confirm the SVM model's strong predictive accuracy and consistency.

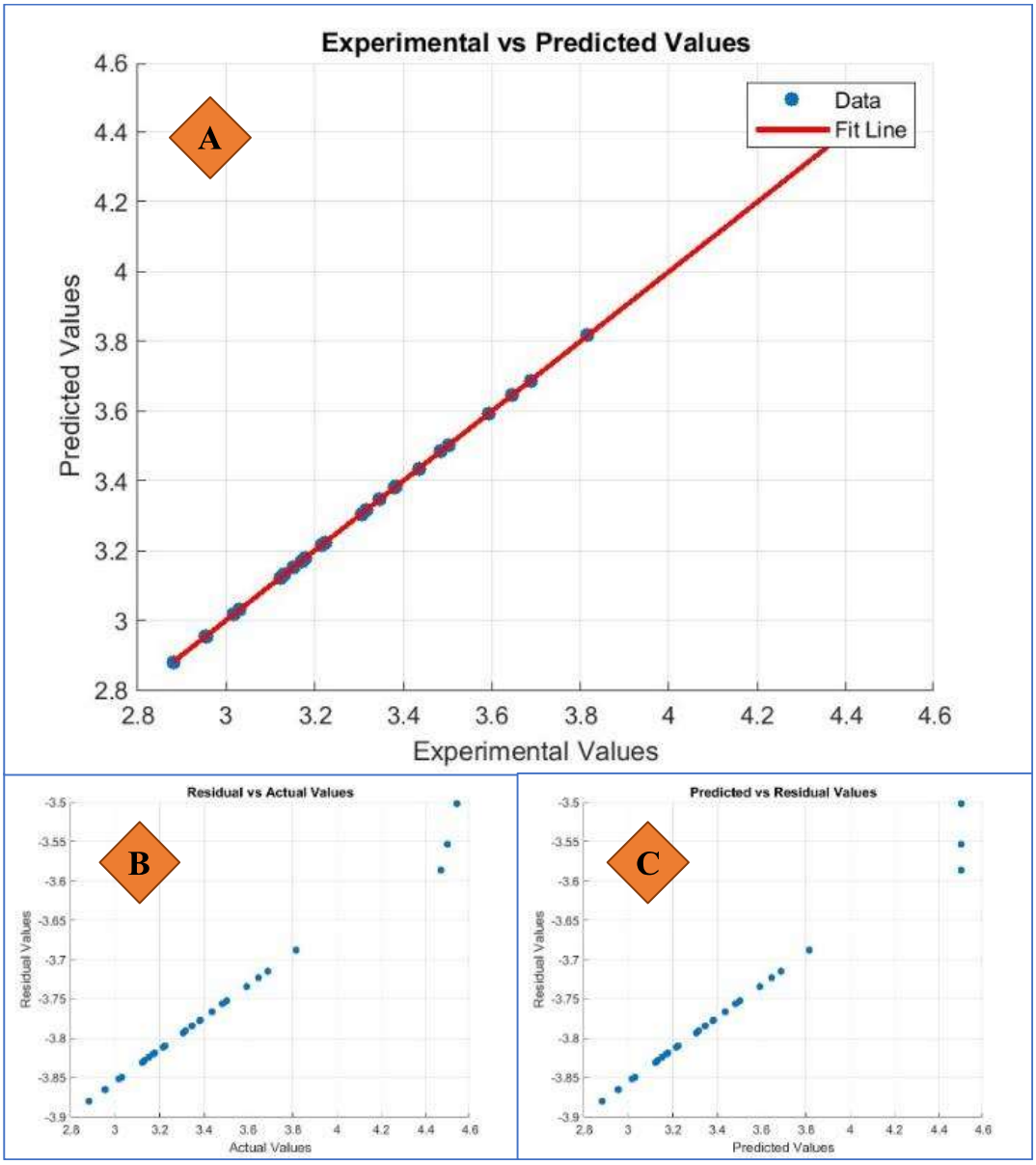

Figure 6.6: This figure presents a comprehensive performance evaluation of the GRNN (General Regression Neural Network) model. A display a scatter plot of predicted versus experimental values with a fit line, demonstrating a high degree of alignment between predicted and actual values. B shows a plot of residuals versus actual values, indicating the distribution of errors across the range of actual values. C presents a plot of residuals versus predicted values, further illustrating the error distribution relative to the predicted values. These visualizations confirm the GRNN model's strong predictive accuracy and consistency.

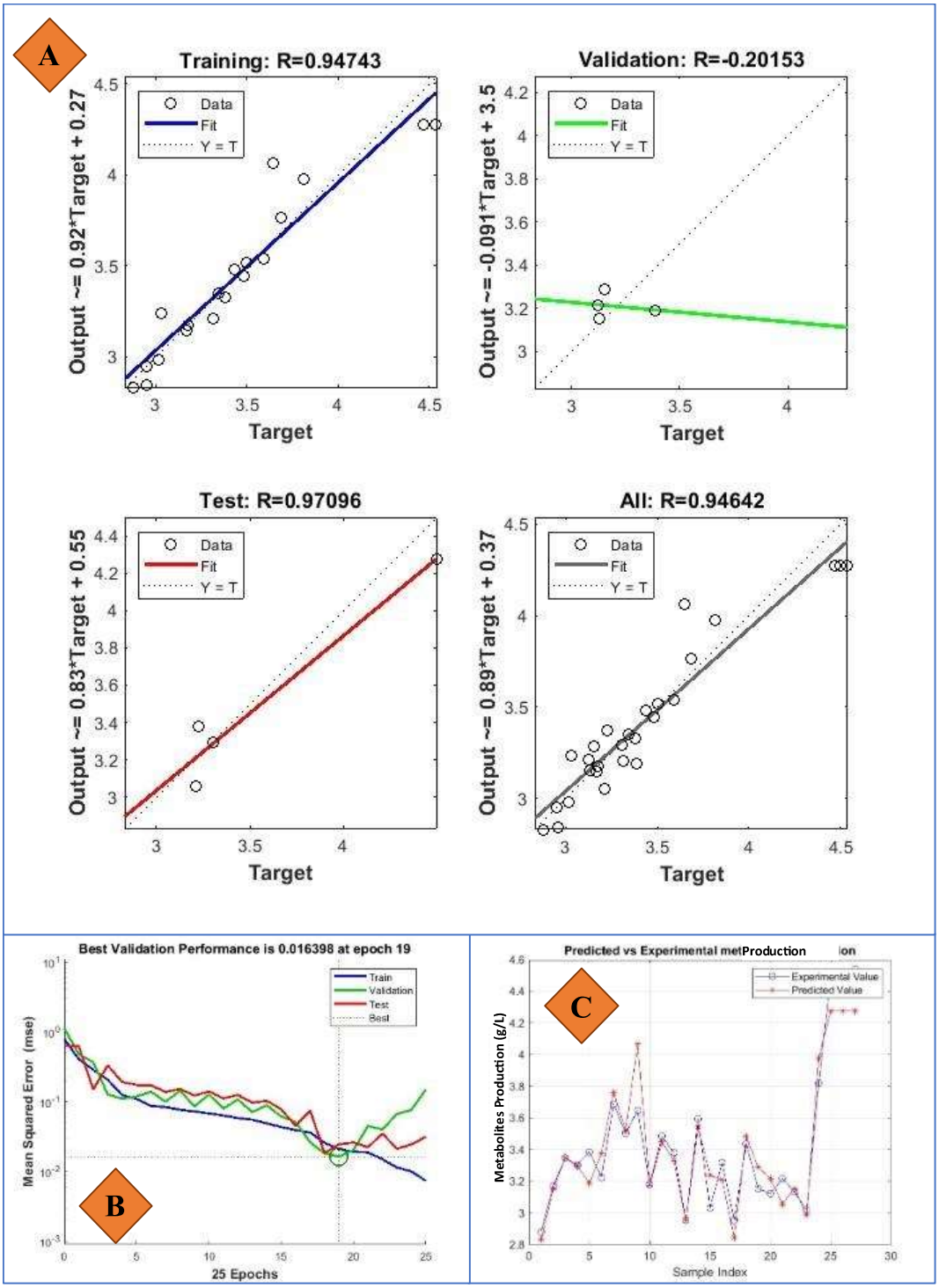

Figure 6.7: This figure presents a comprehensive performance evaluation of the BFGS (Broyden-Fletcher-Goldfarb-Shanno) model. A displays scatter plots of predicted versus actual target values for training (R=0.99779), validation (R=0.99494), test (R=0.88105), and overall datasets (R=0.97651), each with corresponding linear fit lines, indicating varying predictive accuracy across datasets. B shows the mean squared error (MSE) plot over 54 epochs, highlighting the best validation performance achieved at epoch 48 with a value of 0.010983. C illustrates the predicted versus experimental production values, further validating the model's predictive accuracy.

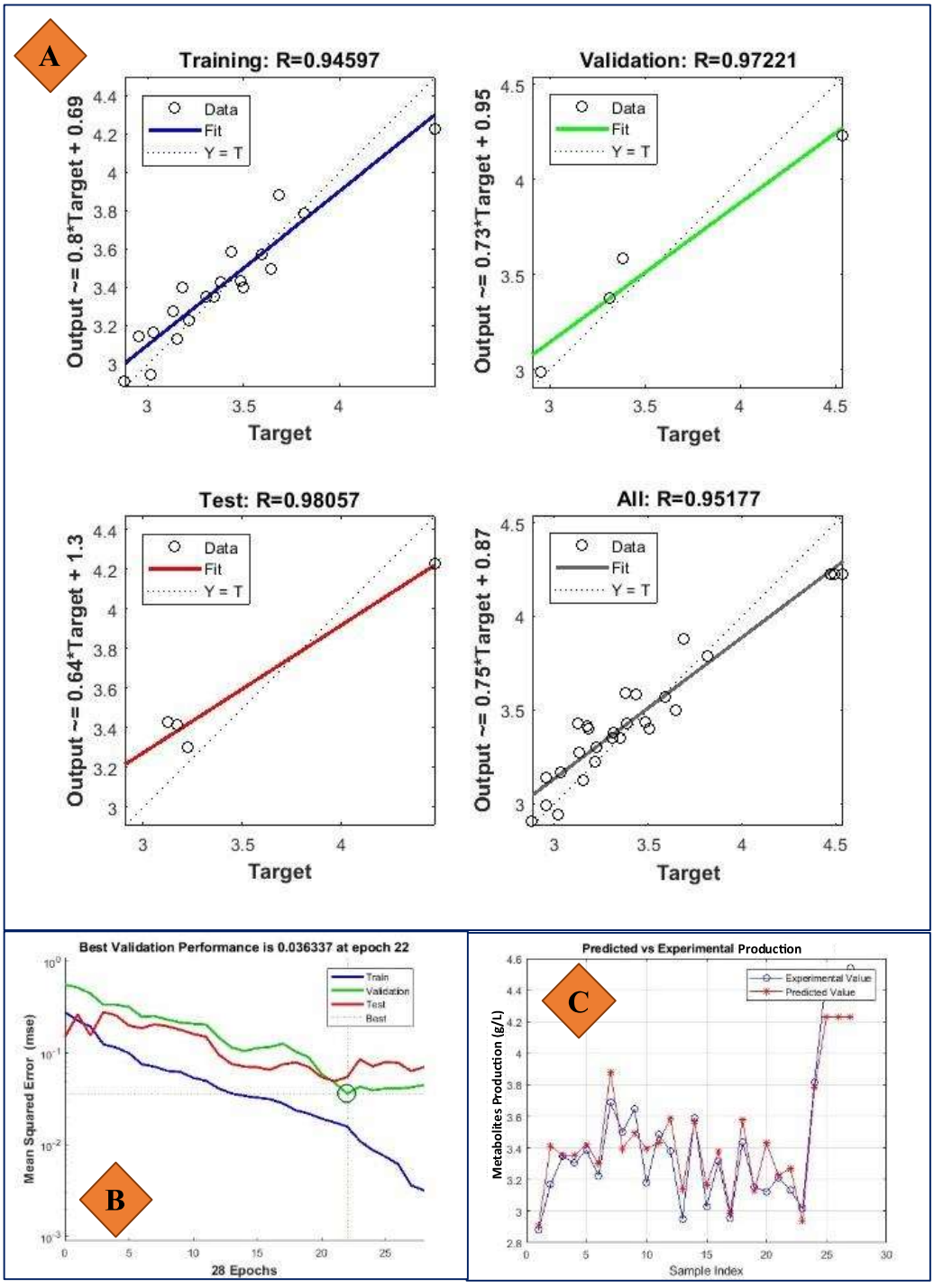

Figure 6.8: This figure presents a comprehensive performance evaluation of the SCG (Scaled Conjugate Gradient) model. A displays scatter plots of predicted versus actual target values for training (R=0.97417), validation (R=0.99642), test (R=0.74653), and overall datasets (R=0.97108), each with corresponding linear fit lines, indicating varying predictive accuracy across datasets. B shows the mean squared error (MSE) plot over 32 epochs, highlighting the best validation performance achieved at epoch 26 with a value of 0.005651. C illustrates the predicted versus experimental production values, further validating the model's predictive accuracy.

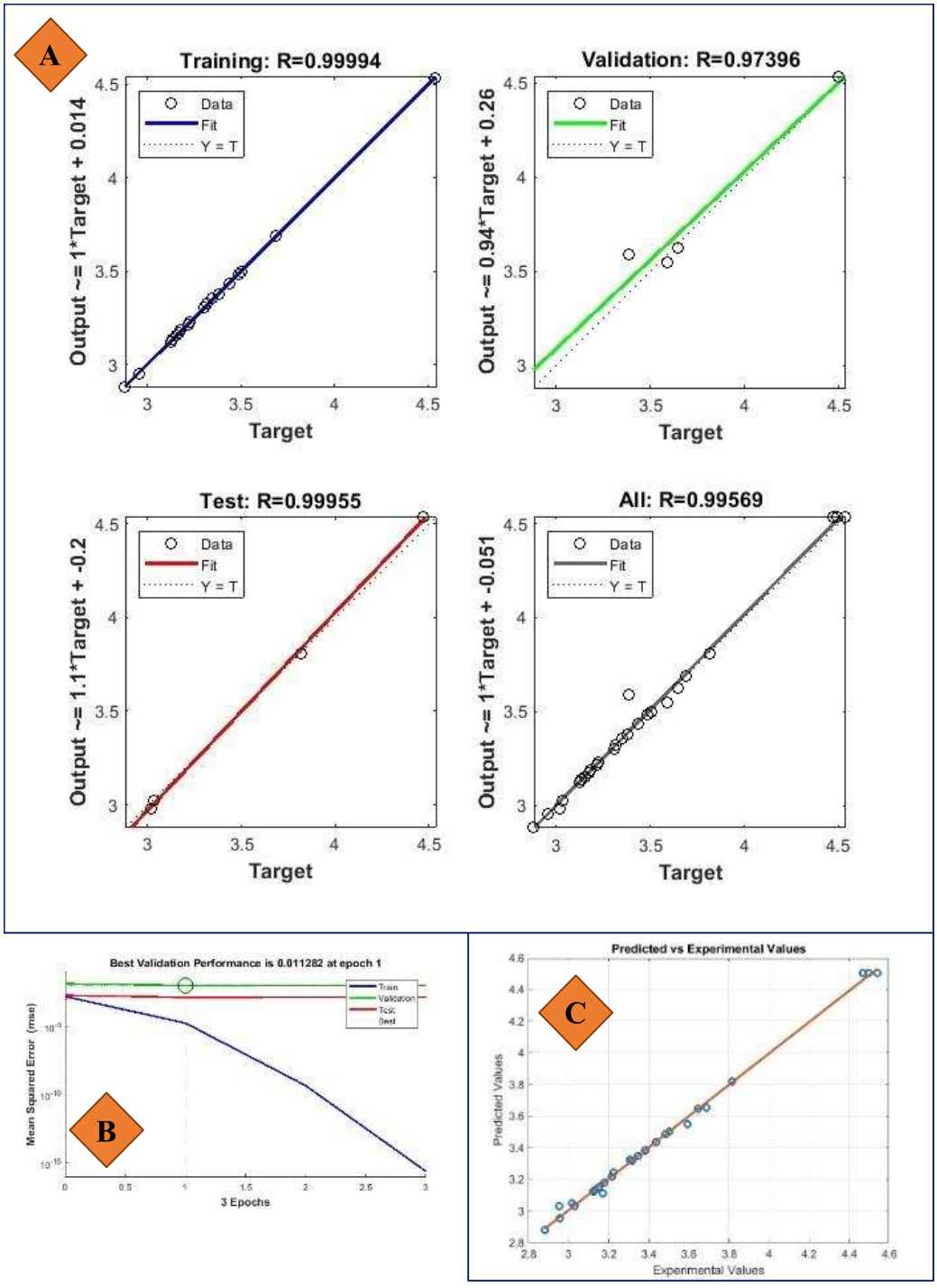

Figure 6.9: This figure presents a performance evaluation of the LM (Levenberg-Marquardt) model. A displays scatter plots of predicted versus actual target values for training (R=0.99994), validation (R=0.9997), test (R=0.99243), and overall datasets (R=0.99593), each with corresponding linear fit lines, indicating high predictive accuracy. B shows the mean squared error (MSE) plot over epochs, highlighting the best validation performance achieved at epoch 1 with a value of 0.00031272. C illustrates the predicted versus experimental values, further validating the model's predictive accuracy.

6.3.6. Validation of the model by experiments

Optimal process parameters, determined by the response optimizer tool of RSM in Python, were used to conduct experiments with the goal of maximizing metabolite production. The ML model, trained using the LM algorithm in MATLAB, was employed for this purpose. The response optimizer of RSM and the ML model predicted the production of 4.65 g/L and 4.58 g/L, respectively, at the optimized level of process parameters: pH(X_1) 7.09, temperature(X_2) of 29.7°C, agitation rate(X_3) of 203 RPM and C:N ratio (X_4) of 1.15. The experimental production of 4.59 g/L was observed, which was very close to the ML predicted value, in 72 hours of fermentation time.

6.3.7. Anthracycline detection and separation

After the downstream process of anthracycline, HPLC was conducted to detect and separate epirubicin. **Figure 6.10** illustrates the HPLC chromatogram of pure epirubicin and the downstream part of Streptomyces fragilis extract.

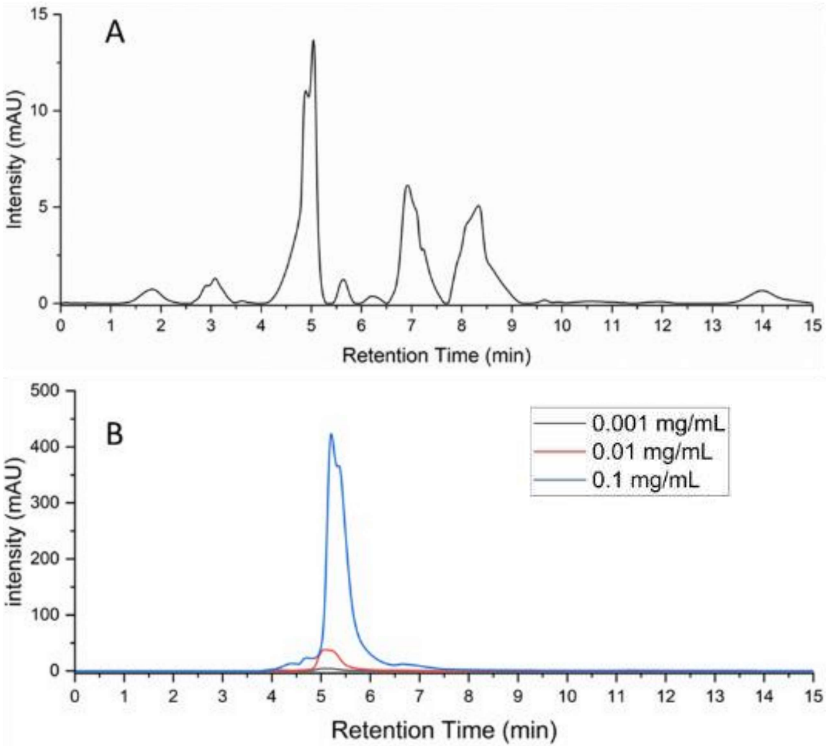

Figure 6.10: HPLC chromatogram of the downstream part of *Streptomyces fragilis* extract (Black; Figure. 6.1.A) and pure epirubicin drug (multicolor; Figure. 6.1.B). The peak at RT of ~5 min indicates the presence of epirubicin in the extract.

During the analysis of the pure sample, it was found that epirubicin was eluted with a retention time of 5 minutes. Upon analysis of the downstream portion, a peak was observed at around 5 minutes retention time, which approximately coincided with the retention time of the standard epirubicin chromatogram. It suggests the presence of epirubicin in the *Streptomyces fragilis* extract. Moreover, the shape and intensity of the peak observed in the extract were similar to those of the standard chromatogram of epirubicin. These observations strongly support the presence of epirubicin in the *Streptomyces fragilis* extract. The calibration curve depicted in **Figure 6.11** shows the total amount of 0.0014 mg epirubicin produced per batch.

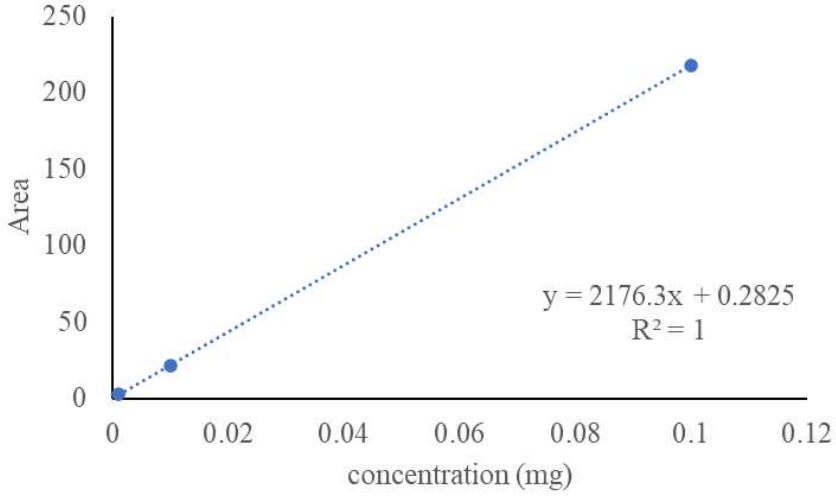

Figure 6.11: HPLC calibration curve of epirubicin.

6.3.8. Structural Elucidation and Thermal Characteristics of Epirubicin

6.3.8.1. FTIR Spectral Analysis

The FTIR technique was utilized to investigate the existence of epirubicin in the secondary metabolites produced by *Streptomyces fragilis*. The IR spectrum of the pure epirubicin and the extract of Streptomyces fragilis was analyzed and depicted in **Figure 6.3**, the broad peaks at 3426 cm^{-1} representing the O-H alcohol group stretching. Peaks at 3064 cm^{-1} and 2925 cm^{-1} corresponded to the stretching vibration of C=C alkene and

C-H alkane, respectively. A peak at 1718 cm^{-1} indicated the contribution of the C=O stretching mode, signifying the presence of carboxylic acid or aliphatic ketone. Another peak at 1621 cm^{-1} indicated the contribution of N-H bending and suggested the presence of an amine group. Two bands between 1500 and 1300 cm^{-1} at 1465 cm^{-1} and 1450 cm^{-1} were observed, signifying C−H bending and denoting the presence of methylene and methyl groups. The signal at 1400 cm^{-1} and 1369 cm^{-1} denoted the O-H bending and signified the presence of alcoholic and phenolic functional groups 25. The peak at 1284 cm^{-1} was assigned to the C–N bending mode and indicated the presence of an aromatic amine functional group. The peak at 1059 cm^{-1} represented the C-O bending and suggested the occurrence of a primary alcohol (Ragavendran et al., 2011). Comparison of the IR spectra of pure epirubicin and the extract of Streptomyces fragilis revealed similar peaks in the 3400–1000 cm−1 region, indicating the presence of epirubicin in the *Streptomyces fragilis* extract.

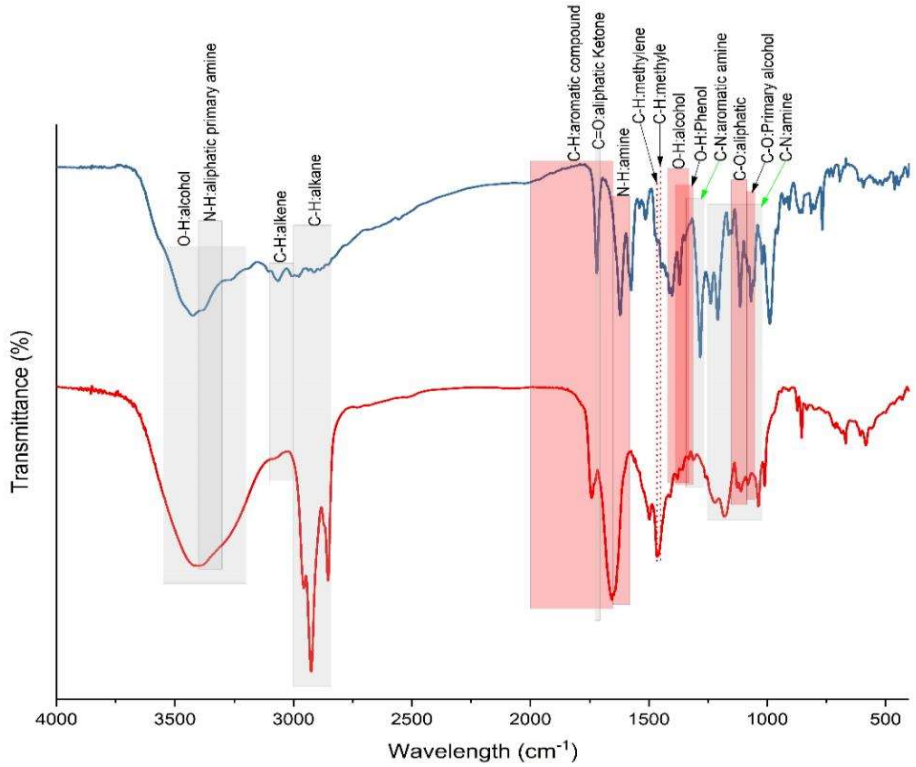

Figure 6.12: FT-IR spectra depicting a potential band of epirubicin (Blue) and the downstream part of *Streptomyces fragilis* extract (Red)

6.3.8.2. ^{1}H and ^{13}C Nuclear Magnetic Resonance

NMR is a highly adaptable and effective technique for resolving molecular structures and dynamics, facilitating a comprehensive understanding of the geometrical properties of chemical compounds (Toukach & Ananikov, 2013).

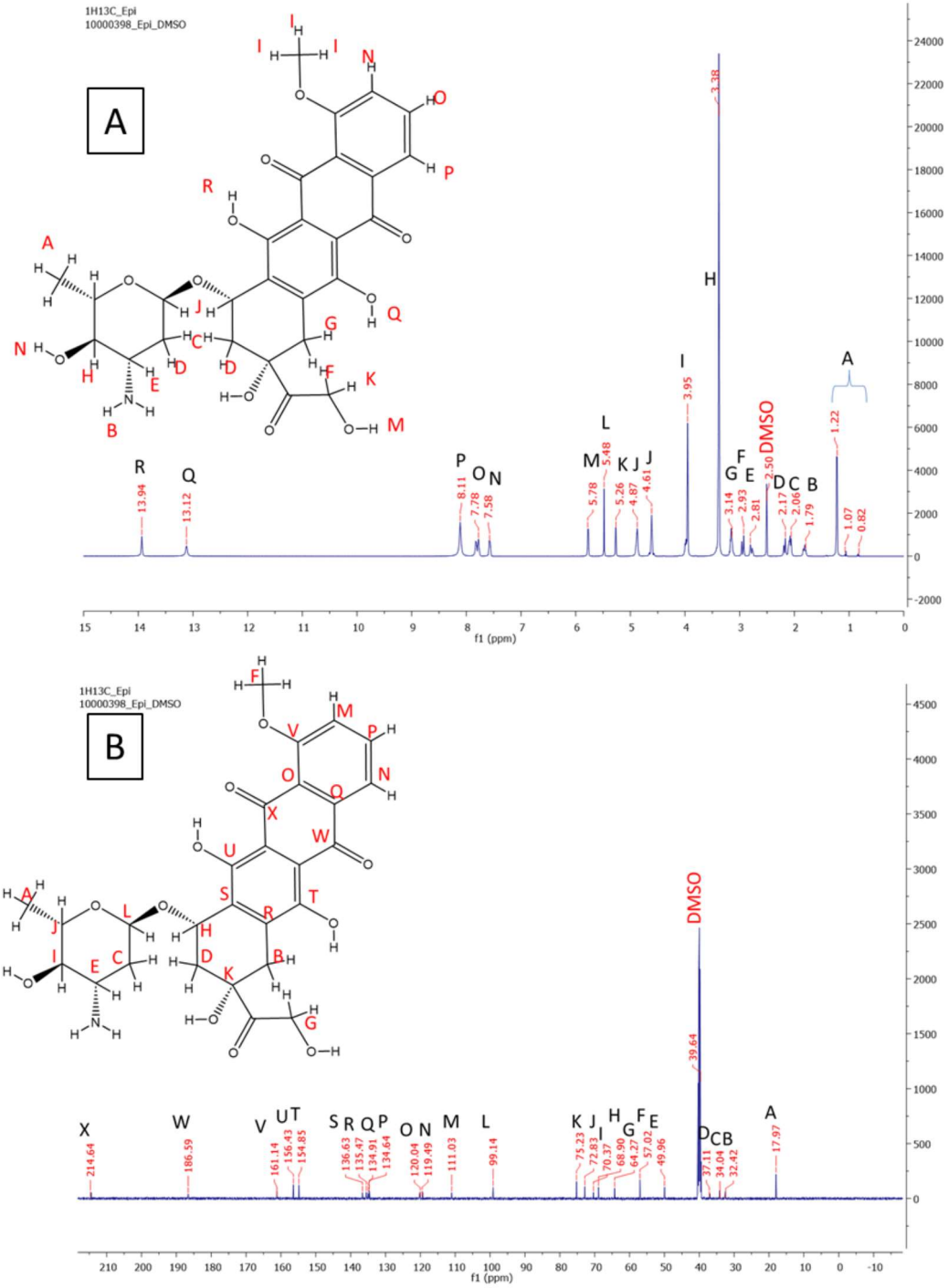

Figure 6.13: Structural elucidation of epirubicin by ^{1}H NMR (A) and ^{13}C NMR (B)

6.3.8.3. High-Resolution Mass Spectroscopy

Figure 6.14 displays the high-resolution mass spectrometry (HR-MS) spectrum of epirubicin produced by *S. fragilis*. The x-axis represented the mass-to-charge ratio (m/z) from 540 to 580, and the y-axis showed ion intensity. A prominent peak at m/z 566.1576, corresponding to the sodium adduct of epirubicin ([C27 H29 N O11] + Na)+, confirmed the identification of epirubicin (Kelly et al., 2010).

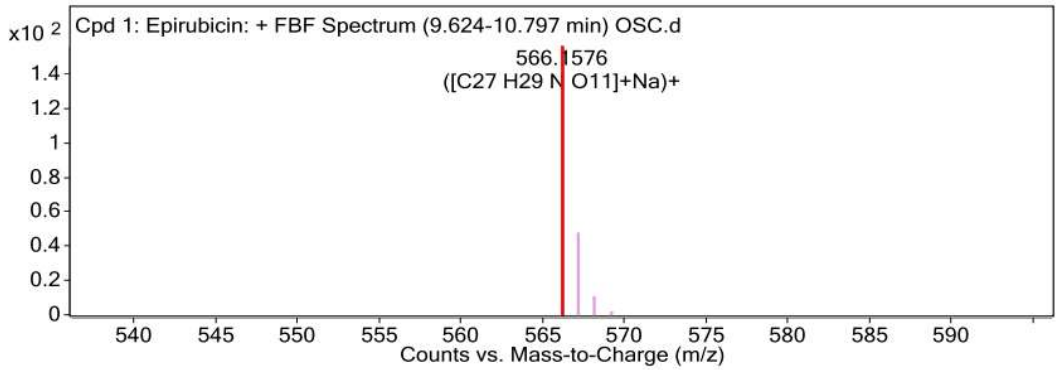

Figure 6.14: HR-MS spectra of epirubicin produced by *S. fragilis*

6.3.8.4. X-ray diffraction (XRD) analysis

XRD is a widely used analytical technique that provides essential information on the composition and crystal/amorphous structure of a sample (Shah et al., 2006). In this study, the XRD spectra of pure epirubicin showed distinct peaks at specific 2θ angles.

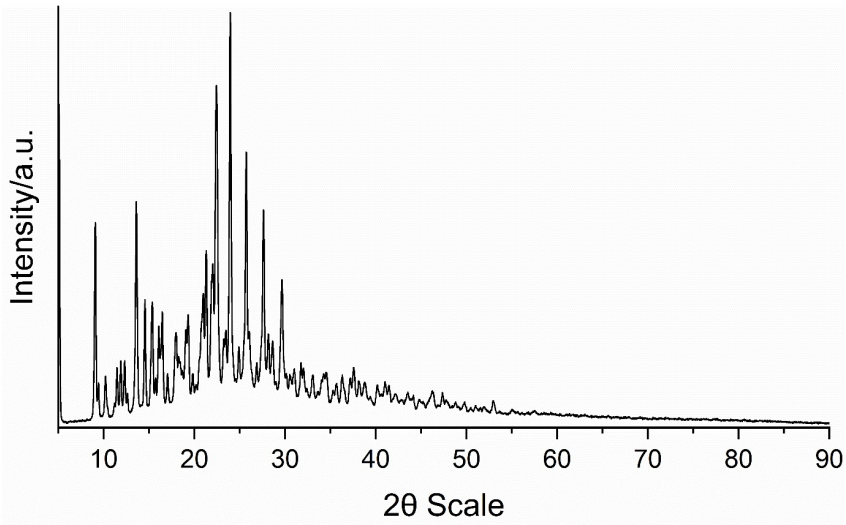

Figure 6.15: X-Ray diffraction spectra of epirubicin

The peaks observed in the XRD spectra were located at 2θ angles of 9.12, 13.64, 22.52, 23.88, 25.62, 27.56, and 29.7, as presented in **Figure 6.15**. XRD detected well-defined and sharp peaks, indicating a high degree of crystallinity in the epirubicin (Bhadra & Khastgir, 2008).

6.3.8.5. Differential Scanning Calorimetry (DSC)

The thermosensitivity of epirubicin was investigated through differential Scanning Calorimetry, which allowed for the determination of the temperature at which epirubicin undergoes a phase shift and enters a melted state (Posada-Murcia et al., 2022). **Figure 6.16** revealed the onset of a phase transition at approximately 186 ˚C, indicating the initiation of the melting process of epirubicin and a change in its physical state. Furthermore, a well-defined and sharp endothermic peak was observed at 206.6 ˚C, signifying complete melting and loss of the solid state of epirubicin.

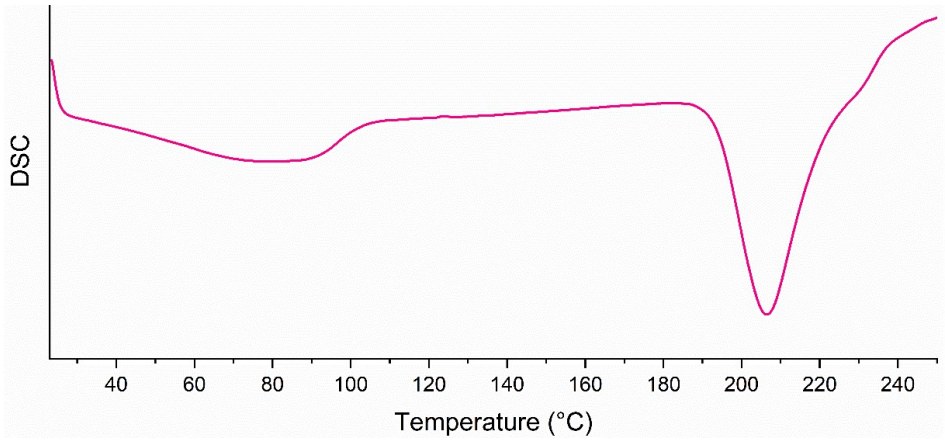

Figure 6.16: Differential scanning calorimetry thermogram analysis of epirubicin.

6.3.8.6. Thermogravimetric analysis (TGA)

The thermal stability and weight loss or gain of epirubicin were explored through the TGA technique. The experiments were conducted by subjecting pure epirubicin to temperatures varying from 25°C to 600°C under a constant flow of nitrogen gas. The change in the drug weight was continuously monitored in a controlled environment as a

function of temperature and time. **Figure 6.17** represents the TGA curves of pure epirubicin, indicating that epirubicin exhibited a certain level of stability throughout the tested temperature range, with a 50% weight loss occurring at a temperature of 600°C.

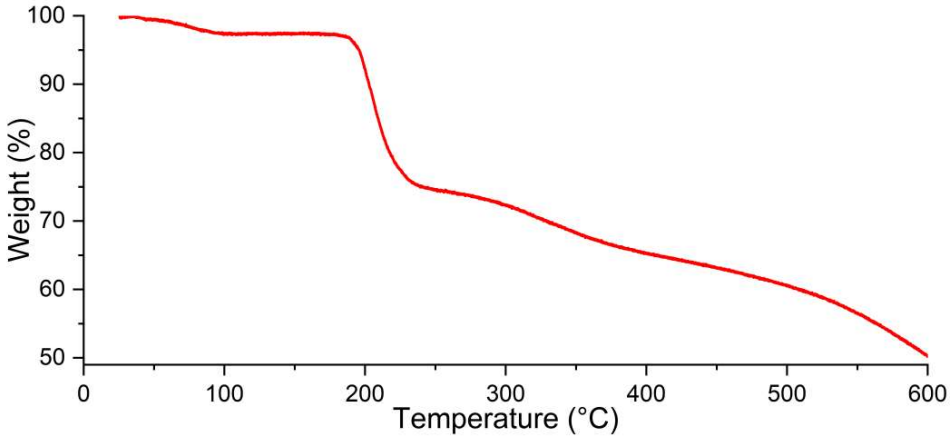

Figure 6.17: Thermogravimetric analysis of epirubicin at 25°C to 600°C with a nitrogen flow rate of 100 mL/min.

6.4. Conclusions

This chapter explored the optimization of bioactive secondary compound production by *Streptomyces fragilis* using various carbon sources and nitrogen sources in a batch fermentation process. Among the tested carbon sources, molasses was the most effective, yielding 4.54 g/L of metabolites, followed by lactose with a yield of 4.1 g/L. In comparison, other carbon sources like fructose and corn flour showed lower metabolite yields, with corn flour being the least effective, with a yield of 0.89 g/L. The cost-effectiveness of molasses, being an agricultural by-product, adds to its attractiveness as a carbon source for industrial applications. In addition to carbon sources, this chapter also evaluated the impact of different nitrogen sources on metabolite production. Yeast extract was identified as the most effective nitrogen source, resulting in the highest metabolite production of approximately 2.7 g/L. This effectiveness is likely due to yeast extract's comprehensive nutrient profile, which is rich in amino acids, peptides, vitamins, and minerals. Beef extract and ammonium sulphate also performed well, yielding 2.5 g/L and

2.05 g/L of metabolites, respectively. In contrast, peptone and casein were less effective, producing 1.8 g/L and 1.5 g/L, respectively.

To optimize the production process, Response Surface Methodology (RSM) was employed using the Box-Behnken Design (BBD). The optimization focused on key parameters such as agitation rate, pH, temperature, and C:N ratio. The RSM model indicated that the carbon-to-nitrogen ratio was the most critical factor influencing metabolite production, and the model predicted a maximum metabolite yield of 4.65 g/L, which closely matched the experimental yield of 4.59 g/L. In addition to RSM, machine learning (ML) models (LM, SVM, GRNN, BFGS, and SCG) were utilized to predict metabolite production. The LM algorithm emerged as the most effective, demonstrating exceptional predictive accuracy, while the BFGS model was the least reliable, with significantly lower validation accuracy. The RSM and ML models predicted metabolite productions of 4.65 g/L and 4.58 g/L, respectively, with the experimental yield being 4.59 g/L. This empirical validation demonstrates the effectiveness of integrating advanced modeling techniques in optimizing fermentation conditions. This chapter also encompasses downstream processing, isolation, and characterization of the secondary metabolite epirubicin, an anthracycline. The structural characterization of epirubicin (anthracycline) was conducted using various techniques such as HPLC, FTIR, 1H, and 13C NMR, while its thermal characteristics were elucidated by XRD, DSC, and TGA analysis. These comprehensive analyses verified the presence and structural integrity of epirubicin in the fermentation extract, underscoring the feasibility of producing high-value bioactive compounds through optimized fermentation processes. Overall, the integration of agricultural waste molasses as a carbon source, yeast extract as a nitrogen source, and the application of advanced optimization techniques significantly improved the efficiency and cost-effectiveness of metabolite production by *Streptomyces fragilis.*