

B.Tech. Project Report - *Phase I*

Modelling the Reactive Extraction Process of Beta-Alanine using Machine Learning (ML) Models



Submitted in partial fulfilment of requirements for the award of the degree of Bachelor of Technology from IIT (Indian Institute of Technology) Guwahati

Under the supervision of
Dr. Senthilkumar Sivaprakasam

Submitted by
Sohail-ur-Rehman
200106069

November 2023
Department of Biosciences and Bioengineering
Indian Institute of Technology Guwahati,
Guwahati 781039, Assam, INDIA

Certificate

This is to certify that the work presented in the report entitled “**Modelling the Reactive Extraction Process of Beta-Alanine using Machine Learning (ML) Models** ” by **Sohail-ur-Rehman**, in fulfilment of the requirements of the degree of **Bachelors of Technology in Biosciences and Bioengineering**, is a Bonafide record of the work carried out by him under the guidance of **Dr. Senthilkumar Sivaprakasam** at the Department of Biosciences and Bioengineering, Indian Institute of Technology, Guwahati. This study has not been submitted elsewhere for a degree.

Signature of student:

Date: 11/11/2023

Place: IIT Guwahati



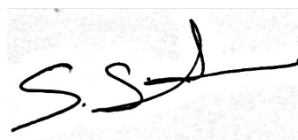
Sohail-ur-Rehman

200106069

Signature of supervisor:

Date: 13/11/2023

Place: IIT Guwahati



Dr. Senthilkumar Sivaprakasam

Professor, Department of
Biosciences and Bioengineering,
Indian Institute of Technology,
Guwahati

Signature of Head of Department:

Date:

Place:



Dr. Rakhi Chaturvedi

Professor & Head, Department of
Biosciences and Bioengineering,
Indian Institute of Technology,
Guwahati

ACKNOWLEDGEMENT

I would like to express my gratitude to the Indian Institute of Technology, Guwahati, for granting me the opportunity to pursue a Bachelor of Technology in Biosciences and Bioengineering. I want to specifically thank my project supervisor, Dr. Senthilkumar Sivaprakasam, a Professor at IIT Guwahati, for allowing me to undertake my bachelor's thesis project in the BioPAT laboratory. Throughout the project, his unwavering support, guidance, and supervision were invaluable.

I am also thankful to the Department of Biosciences and Bioengineering at IIT Guwahati for providing the necessary facilities for my research. Special thanks go to Ms. Sandhya, a Ph.D. scholar at BioPAT, for her mentoring. Additionally, I want to express my sincere appreciation to my lab-mates, whose motivation and valuable feedback played a crucial role in the successful completion of my work. I extend my thanks and regards to my family for their constant support, well-wishes, and encouragement, which I consider priceless.

In conclusion, my heartfelt thanks go out to everyone who has been a part of my academic and research endeavours at IIT Guwahati. I am confident that the skills and knowledge gained during my undergraduate studies will serve as a solid foundation for my future endeavours in the field of biosciences and bioengineering.

Date: 11/11/2023

Sohail-ur-Rehman
200106069

TABLE OF CONTENTS

| | |
|--|-----------|
| ABSTRACT | 5 |
| 1. INTRODUCTION | 5 |
| 1.1 β Alanine | 5 |
| 1.2 β Alanine applications | 6 |
| 1.3 Reactive extraction | 6 |
| 1.4 Responsive Phase Methodology | 7 |
| 1.5 Machine Learning (ML) Models | 8 |
| 2. OBJECTIVES | 8 |
| 3. LITERATURE REVIEW | 9 |
| 4. MATERIALS AND METHODS | 10 |
| 5. RESULTS AND CONCLUSIONS | 11 |
| 6. FUTURE WORK | 13 |
| 7. REFERENCES | 13 |

ABSTRACT

In recent years, the field of Biochemical Engineering has witnessed a growing interest in optimising processes through the utilisation of artificial intelligence (AI) techniques. This research aims to contribute to this evolving landscape by focusing on the reactive extraction process of beta-alanine, a compound of significance in pharmaceutical and industrial applications.

The primary objective of this study is to employ machine learning models to enhance the efficiency of the reactive extraction of beta-alanine. To achieve this, we adopted the responsive phase methodology, a technique designed to systematically explore the parameter space and generate comprehensive data sets for training and evaluating machine learning models.

The responsive phase methodology allows for a detailed understanding of the system's behaviour under varying conditions, providing a solid foundation for the subsequent application of machine learning models. In this project, we trained several machine learning models with the acquired data, exploring different architectures and configurations to identify the most effective model for predicting and optimising the reactive extraction process.

This introduction outlines the motivation behind the research, the methodology employed, and the overarching goal of optimising the reactive extraction of beta-alanine using machine learning models. Subsequent sections will delve into the experimental setup, data collection, and the results obtained, providing a comprehensive analysis of the contributions and implications of this study.

1. Introduction

1.1 β Alanine:

Beta-alanine is a non-essential amino acid, meaning the body can produce it naturally. It is a crucial component in the synthesis of carnosine, a dipeptide found predominantly in skeletal muscle tissues. Carnosine plays a vital role in buffering acid produced during high-intensity exercise, thereby helping to delay the onset of muscle fatigue and improve overall exercise performance.

Unlike many amino acids that are incorporated into proteins, beta-alanine primarily exists free in the body. Its supplementation has gained attention in the fields of sports nutrition and exercise physiology due to its potential ergogenic effects. By increasing carnosine levels,

beta-alanine supplementation is believed to enhance exercise capacity and performance, particularly in activities involving short bursts of high-intensity effort.

Moreover, beta-alanine has applications beyond the realm of sports nutrition. Its significance extends to various industrial and pharmaceutical processes, making it a compound of interest for research and optimization.

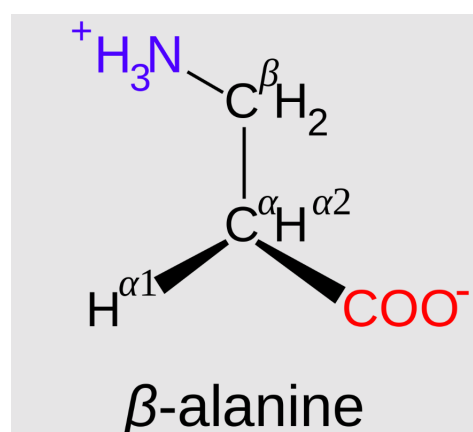


Fig 1.1: Structure of β alanine

1.2 β Alanine applications:

Beta-alanine holds diverse applications across sports nutrition, medicine, and industrial processes. In the realm of sports nutrition, it serves as a popular supplement, enhancing exercise performance by increasing carnosine levels in muscles and delaying the onset of fatigue during high-intensity activities. Beyond its role in sports, beta-alanine has implications in exercise physiology research, potentially improving endurance and power output. In the pharmaceutical domain, the antioxidant and anti-inflammatory properties of carnosine, synthesised with beta-alanine, make it an area of interest for therapeutic applications. Industrial processes, including reactive extraction, leverage beta-alanine's properties, showcasing its significance in chemical engineering and optimization. Additionally, beta-alanine finds utility in food additives, cosmetics, and skincare formulations, underscoring its multifaceted role in various fields.

1.3 Reactive extraction:

Liquid-liquid extraction, a fundamental separation process, involves the transfer of a specific solute from the aqueous phase to the organic phase. In the context of this research, reactive extraction is a specialised form of liquid-liquid extraction tailored for solutes featuring carboxylic acid groups, such as beta-alanine. Unlike conventional liquid-liquid extraction,

reactive extraction involves a distinctive aspect—the occurrence of a chemical reaction between the extractant and solute at the interface.

This process holds economic advantages, as both the diluent and extractant components utilised in the reactive extraction of beta-alanine can be recycled. This recyclability contributes to enhanced cost-effectiveness and a sustainable approach to the extraction process. Notably, the ability to reuse components is a critical factor known to facilitate superior product recovery, ensuring a final product with high purity.

Several factors intricately influence the efficiency of beta-alanine extraction through reactive methods. These factors include the concentration of **D2EHPA** (Di(2-ethylhexyl) phosphoric acid), **incubation time**, **pH levels**, **solute concentration**, **reaction temperature**, **agitation rate**, and **the choice of diluent extraction system**. The optimization of these parameters is pivotal for achieving an efficient and selective extraction of beta-alanine, a compound of significance in pharmaceutical or industrial applications. The intricate interplay of these factors will be explored in detail throughout the subsequent sections of this report, shedding light on their individual and collective impacts on the overall effectiveness of the reactive extraction process.

1.4 Responsive Phase Methodology:

In the pursuit of optimising the reactive extraction process for beta-alanine, the Responsive Phase Methodology was employed to systematically acquire data and comprehensively understand the intricate interplay of key factors. This methodological approach involves the deliberate exploration of the experimental parameter space, enabling the generation of a robust dataset. The data obtained through the Responsive Phase Methodology serves as the foundation for training and evaluating artificial neural networks (ANNs) in the subsequent phases of the research.

The predictors identified for this study encompass crucial parameters known to influence the efficiency of beta-alanine extraction. These include the concentration of **D2EHPA** (Di(2-ethylhexyl) phosphoric acid), **pH levels**, **temperature**, **solute concentration**, **incubation time**, and the **phase ratio**. Each of these predictors plays a distinctive role in shaping the outcome of the extraction process, and their systematic variation during the Responsive Phase allows for a comprehensive examination of their individual and collective impacts on extraction efficiency.

The response variable, **extraction efficiency**, is a key metric that reflects the effectiveness of the reactive extraction process. This variable serves as the basis for evaluating the performance of different neural network models during the training and optimization phases. The utilisation of extraction efficiency as a response variable enables the identification of optimal conditions for beta-alanine extraction, facilitating the overarching goal of enhancing the efficiency of the reactive extraction process.

1.5 Machine Learning (ML) Models:

In the quest to optimise the reactive extraction process for beta-alanine, the utilisation of machine learning (ML) algorithms, particularly machine learning models, emerges as a powerful tool to model and predict the intricate relationships within the dataset obtained through the Responsive Phase Methodology.

Machine learning models are computational approaches that can capture complex patterns and non-linear relationships within datasets, making them particularly advantageous for applications in chemical engineering and process optimization.

In this study, machine learning models are employed to discern the underlying patterns between the predictors (D2EHPA concentration, pH, temperature, solute concentration, incubation time, and phase ratio) and the response variable (extraction efficiency). The training of machine learning models involves exposing the algorithm to the dataset, allowing it to learn and adapt to the intricate relationships present in the experimental data. Once trained, the machine learning model can then make predictions and optimise the extraction process by identifying the combination of parameters that yields the highest extraction efficiency.

This approach offers a data-driven and adaptive solution to the complex task of optimising the reactive extraction of beta-alanine. By harnessing the power of machine learning models, this study aims to develop a predictive model that not only accurately represents the experimental data but also provides valuable insights for enhancing the efficiency of the reactive extraction process.

2. OBJECTIVES

The primary goal of this study is to identify a robust algorithm capable of accurately fitting and predicting the system dynamics in the reactive extraction process for beta-alanine:

To achieve this, advanced machine learning techniques, specifically Artificial Neural Networks (ANNs), are employed. The adaptability and pattern-recognition capabilities of ANNs are crucial for understanding the relationships between key parameters and extraction efficiency. Subsequent sections of this report will detail the methodology and outcomes, emphasising the selection of a robust algorithm for effective modelling and prediction in the context of beta-alanine extraction.

3. LITERATURE REVIEW

A method of reactive extraction aimed at isolating amino acids from a mixture. This process utilises four acidic extractants (DNNSA, D2EHPA(S), D2EHPA, and Versatic 10) dissolved in toluene to extract a range of amino acids (arginine, phenylalanine, alanine, glycine, and aspartic acid) from a solution. The extractive efficiency of the organic phase displays a defined hierarchy for amino acids, and separation factors are detailed for various amino acid pairs. Furthermore, the extractive and loading capabilities of the extractants exhibit a specific diminishing order, with DNNSA identified as a particularly promising extractant for both amino acid extraction and fractionation [1].

The study using di(2-ethyl-hexyl) phosphoric acid in aqueous solutions found that amino acid extraction is pH-dependent and influenced by the amino acids' hydrophobic nature. Higher pH values enhance extraction efficiency, especially for amino acids with longer carbon chains or cyclic structures. Conversely, those with additional polar groups exhibit lower extraction efficiency. Furthermore, successful amino acid extraction from cottonseed meal hydrolysate was achieved through microbial fermentation and HCl hydrolysis, with optimal results at an initial feed pH of 6.5. The use of D2EHPA reactive extraction was emphasised for its ability to simplify the process, avoiding multiple purification stages and yielding purer back-extracts. This approach streamlines the separation and purification of amino acids effectively [2].

The extraction equilibria of L-phenylalanine using liquid ion-exchange extractants, specifically Aliquat 336 and di-2-(ethylhexyl) phosphoric acid (DEHPA), in the diluent methyl isobutyl ketone. The results compare the extraction efficiency of each single extractant against a mixed extractant system combining equimolar amounts of Aliquat 336 and DEHPA. While the mixed system shows lower loading values than the single-extractant systems, it maintains uptake capacity across a broader pH range (1.8 to 9.2) [3].

The use of di-(2-ethylhexyl) phosphoric acid (D2EHPA) in reactive extraction for separating amino acids, revealing the potential for selective separation based on pH and the acidic or basic nature of each amino acid. Through multistage extraction, distinct groups of amino acids (neutral, basic, acidic) are successfully separated at specific pH ranges [4].

The optimization of the culture medium played a pivotal role in significantly enhancing the production of human interferon alpha 2b (huIFN α 2b) in glycoengineered *P. pastoris*. The application of a Design of Experiments (DoE)-based medium optimization strategy not only improved efficiency but also demonstrated its effectiveness at both shake flask and bioreactor levels. This successful scale-up underscores the practical applicability of the optimised conditions in larger-scale industrial settings. Furthermore, the confirmed N-glycosylation and biological activity of the purified huIFN α 2b highlight the importance of optimization in achieving a structurally characterised and functionally potent protein. The study's findings underscore the critical role of optimization in maximising productivity and ensuring the potential therapeutic relevance of huIFN α 2b, particularly in the context of cancer therapy [5]. This study addressed a critical challenge in hyaluronic acid (HA) production related to the highly viscous nature of the fermentation broth, making real-time monitoring with physical sensors challenging. The objective was to non-invasively track the state variables involved in

HA production and deduce critical process parameters using a hybrid model. By leveraging online bioreactor data, including pH, dissolved oxygen percentage (DO%), %CO₂ evolved, feed rate, and agitation rate, the hybrid model enabled real-time tracking of HA bioprocess dynamics. The study utilised a comprehensive dataset comprising historical batches and newly performed fed-batch runs at various specific growth rate set-points (μ_{sp}). The hybrid model, trained on this dataset, demonstrated accurate predictions of biomass and HA concentration for test runs, with a mean squared error of prediction ranging from 0.018 to 0.049 (g/L)². Additionally, recurrent neural networks were evaluated for forecasting specific growth rate (μ) and HA productivity rate (qHA) to observe desired process trajectories. This research contributes to the field by showcasing the applicability of a hybrid model-based soft-sensor in predicting the trend of process parameters in HA fermentation, offering valuable insights for optimising and controlling this complex bioprocess [6].

The initiation of a synthetic chemist's training involves conducting experiments based on traditional recipes found in textbooks and manuscripts, fostering the development of practical skills and chemical intuition through successful reactions. This approach often persists throughout a researcher's career, with new recipes evolving from established protocols and deviations guided by experiential learning from failed experiments. However, this paper underscores that when delving into the understanding of chemical systems, model-based, algorithmic, and miniaturised high-throughput techniques surpass human chemical intuition. They achieve reaction optimization with superior efficiency in terms of both time and materials, a topic explored comprehensively within this review. Notably, many synthetic chemists lack exposure to these advanced techniques during undergraduate education, leading to a significant number of scientists desiring reaction optimization without the knowledge or utilisation of these methodologies. This literature review illuminates both the fundamentals and cutting-edge aspects of modern chemical reaction optimization, highlighting its implications for process scale-up. By providing a comprehensive reference, it aims to empower and inform scientists, serving as a valuable resource for those inspired to explore and apply these techniques in their research endeavours [7].

4. MATERIALS AND METHODS

Extraction:

For the extraction of beta-alanine, the reactive extraction technique was employed. This method involves the simultaneous extraction and chemical reaction of the target component from the liquid phase into an immiscible solvent phase. The choice of reactive extraction aimed to optimise the process for beta-alanine extraction.

Dataset Generation:

To generate the dataset for this study, a combination of the Responsive Phase Methodology and experimental data obtained in the laboratory was employed. The Responsive Phase Methodology systematically explores the experimental parameter space, allowing for a comprehensive understanding of the system's behaviour. This hybrid approach, incorporating both methodologies, facilitated the collection of data crucial for training and evaluating the subsequent Machine Learning (ML) Models.

Data Analysis:

The collected data were then analysed using various Python libraries, including NumPy, Pandas, Scikit-learn, TensorFlow, and Keras. TensorFlow, a widely used machine learning library, facilitated the implementation and training of artificial neural networks. The analysis was conducted within a Jupyter notebook, providing a collaborative and interactive environment for exploring and interpreting the results. The integration of these diverse libraries allowed for a comprehensive examination of the data and the application of machine learning techniques to the reactive extraction process of beta-alanine.

5. RESULTS AND CONCLUSIONS

In evaluating the performance of the machine learning models trained on the generated data, the Mean Squared Error (MSE) and the R-squared value were utilised as evaluation metrics. The MSE serves as a measure of the average squared difference between predicted and actual values, with lower values indicating a more accurate model. Simultaneously, the R-squared value gauges the proportion of variance in the dependent variable that is predictable from the independent variable. A high R-squared value, approaching 1, signifies a model that effectively captures the variability in the data.

To mitigate the risk of overfitting and ascertain optimal hyperparameter values, a careful strategy was employed in dataset partitioning. Typically, datasets are separated into three subsets: training, validation, and testing, to ensure robust model evaluation. However, due to the limited number of observations in this study, a pragmatic decision was made to split the data into only training and testing sets. The partitioning ratio was chosen as 4:1, maintaining a balance between obtaining reliable training results and reserving a sufficient dataset for robust testing. This approach aims to strike a balance between model complexity and generalizability, avoiding the pitfalls of overfitting that may arise with a more complex partitioning strategy.

Linear Regression Model:

In the initial phase of model exploration, simple multiple linear regression models were employed to establish a baseline understanding of the relationship between predictor variables and the extraction efficiency of beta-alanine.

The results obtained from these initial linear regression models are as follows:

- Training Mean Squared Error (MSE): 25.43
- Training R-squared: 0.91
- Testing Mean Squared Error (MSE): 25.75
- Testing R-squared: 0.92

The coefficients and intercept for the linear regression model are as follows:

- D2EHPA Concentration: 0.774
- pH: 4.200
- Temperature: -0.572
- Solute Concentration: -0.005
- Incubation time: 0.019
- Phase Ratio: -3.299
- Intercept: 12.699

These results underscore the efficacy of linear regression in providing a reasonably good fit to the data, capturing a significant portion of the variance with high R-squared values. However, recognizing the potential existence of non-linear relationships within the dataset, the investigation now turns towards exploring the capabilities of more complex models, specifically Artificial Neural Networks (ANNs). The transition to ANNs aims to uncover and harness the complexities inherent in the reactive extraction process of beta-alanine, potentially improving the model's predictive accuracy by accounting for non-linear dependencies.

Feed Forward ANN (Artificial Neural Network):

In the subsequent phase of model refinement, a Feedforward Artificial Neural Network (ANN) was employed to explore the complex relationships within the dataset, aiming to capture nonlinear dependencies. To optimise the performance of the ANN, an exhaustive grid search was conducted, considering different combinations of hyperparameters. The hyperparameters explored included epochs (25, 50, 75, 100) and batch size (8, 16, 32, 64). The optimal hyperparameters identified through this process were:

- Best Parameters: {'batch_size': 8, 'epochs': 100}

The performance metrics for the Feedforward ANN, utilising the optimal hyperparameters, are as follows:

- Training Mean Squared Error (MSE): 28.49
- Training R-squared: 0.89
- Testing Mean Squared Error (MSE): 22.18
- Testing R-squared: 0.93

These results highlight the effectiveness of the Feedforward ANN in capturing the complexities of the reactive extraction process for beta-alanine. The model demonstrates notable improvement, as evidenced by the enhanced R-squared values on both the training and testing datasets. The subsequent sections will delve deeper into the analysis of the ANN's architecture and explore additional considerations to further enhance its predictive accuracy.

The code for the models can be found [here](#)

6. FUTURE WORK

The successful development of a robust machine learning model, capable of accurately predicting the efficiency of the beta-alanine reactive extraction process, marks a significant milestone in this project. Moving forward, the second phase will focus on leveraging this established model to enhance the efficiency of the extraction process.

In this subsequent phase, we intend to employ the trained machine learning model to predict extraction efficiency under varying conditions. Specifically, we aim to optimise the input features such as pH, D2EHPA concentration, and temperature to maximise the efficiency of the extraction process. By systematically adjusting these parameters based on the model's predictions, we anticipate achieving an enhanced and optimised extraction efficiency.

7. REFERENCES

- [1] Kelly, N.A., Lukhezo, M., Reuben, B.G., Dunne, L.J., & Verrall, M.S. (1998). Reactive solvent extraction of amino acids with cationic extractants. *J. Chem. Technol. Biotechnol.*, 72(4), 347-355. [See original paper here](#)
- [2] Zhang, J., Zhang, S., Yang, X., Qiu, L., Gao, B., Li, R., & Chen, J. (2016). Reactive extraction of amino acids mixture in hydrolysate from cottonseed meal with di(2-ethylhexyl) phosphoric acid. *J. Chem. Technol. Biotechnol.*, 91(2), 483-489. [See original paper here](#)
- [3] Cameron, T., & Husson, S. M. (2001). Recovery of amino acids from aqueous solution by reversible complexation with mixed organic extractants. *Separation Science and Technology*, 36(5-6), 849-866. [See original paper here](#)

- [4] Cascaval, D., Oniscu, C., & Galaction, A. I. (March 2001). Selective separation of amino acids by reactive extraction with D2EHPA. Science Direct Working Paper, No. S1574-0331(04)70486-5. [See original paper here](#)
- [5] Katla S, Karmakar B, Tadi SRR, Mohan N, Anand B, Pal U, Sivaprakasam S. High level extracellular production of recombinant human interferon alpha 2b in glycoengineered Pichia pastoris: culture medium optimization, high cell density cultivation and biological characterization. J Appl Microbiol. 2019 May;126(5):1438-1453. doi: 10.1111/jam.14227. Epub 2019 Mar 19. PMID: 30776176. [See original paper here](#)
- [6] Rathinavelu, S., Pavan, S.S. & Sivaprakasam, S. Hybrid Model-based Framework for Soft Sensing and Forecasting Key Process Variables in the Production of Hyaluronic Acid by Streptococcus zooepidemicus. Biotechnol Bioproc E 28, 203–214 (2023). [See original paper here](#)
- [7] Taylor, C. J.; Pomberger, A.; Felton, K. C.; Grainger, R.; Barecka, M.; Chamberlain, T. W.; Bourne, R. A.; Johnson, C. N.; Lapkin, A. A. Chem. Rev. 2023, 123, 6, 3089–3126. [See original paper here](#)