AMRITA VISHWA VIDYAPEETHAM

Team D-14

AI-driven analysis of NRPS for Drug Development

INTEGRATED WITH DESIGN AND ANALYSIS OF ALGORITHMS AND INTELLIGENCE OF BIOLOGICAL SYSTEMS 2

MHOKESH
CB.SC.U4AIE23316

KOWSHIK
CB.SC.U4AIE23314

VIBHU SANCHANA
CB.SC.U4AIE23347

ASMI K
CB.SC.U4AIE23351

Introduction

What is NRPS?

- Nonribosomal Peptide Synthetases (NRPS) are enzymes that produce complex peptide-based drugs like antibiotics and anti-cancer agents.

Why is NRPS Important?

- Many life-saving drugs, such as penicillin, are derived from NRPS.
- Studying NRPS helps in discovering new medicines to fight diseases.

Role of AI in NRPS Analysis

- AI helps analyze vast biological data faster and more accurately.
- Machine learning models can predict new drug structures based on NRPS patterns.

Literature Review

S no.	Research papers	Challenge	Proposed Idea
1)	**Peptide-based drug discovery through artificial intelligence** Wang, Hui, et al. “Peptide-Based Drug Discovery through Artificial Intelligence: Towards a New Era of Therapeutic Peptide Design.” *Briefings in Bioinformatics*, vol. 25, no. 4, 2024, bbae275, Oxford University Press, https://academic.oup.com/bib/article/25/4/bbae275/7690345.	Peptide drugs often suffer from short half-lives, limited oral bioavailability, and susceptibility to plasma degradation, hindering their clinical application.	The study suggests employing AI-assisted peptide design and validation pipelines to enhance peptide stability and bioavailability, thereby expediting the development of novel therapeutic peptides
2)	**Synthetic-bioinformatic natural product-inspired peptides** Rogers, Ben J., et al. “Synthetic-Bioinformatic Natural Product-Inspired Peptides.” *Natural Product Reports*, 2025, Royal Society of Chemistry, https://pubs.rsc.org/en/content/articlehtml/2025/np/d4np00043a.	Accurately predicting the incorporation of fatty acids, modifications by tailoring enzymes, and peptide release mechanisms in NRPS-derived peptides remains difficult.	Advancements in bioinformatics techniques are recommended to improve the prediction accuracy of these modifications, facilitating the discovery of bioactive peptides that are currently inaccessible
3)	**Harnessing the Power of Artificial Intelligence in Drug Discovery** Patel, Aditi, et al. “Future Perspective: Harnessing the Power of Artificial Intelligence in Drug Discovery.” Biomolecules, vol. 14, no. 10, 2024, MDPI, https://www.mdpi.com/2218-273X/14/10/1303.	Integrating diverse data types to capture the complexity of peptide–drug interactions poses a significant challenge for AI models.	The paper emphasizes the need for high-quality, diverse training data and the integration of multiple data types to enhance the predictive power of AI in peptide drug discovery.

Literature Review

S no.	Research papers	Challenge	Proposed Idea
4)	**Structural, biochemical and bioinformatic analyses of NRPS** Carter, Thomas A., et al. “Structural, Biochemical, and Bioinformatic Analyses of Nonribosomal Peptide Synthetases.” Natural Product Reports, 2024, Royal Society of Chemistry, https://pubs.rsc.org/en/content/articlelanding/2024/np/d3np00064h	Understanding the substrate specificity of NRPS adenylation domains is complex due to their structural diversity	Combining structural, biochemical, and bioinformatic analyses can elucidate the substrate specificity of these domains, aiding in the rational design of NRPS for novel peptide synthesis
5)	**How Machines Learned to Discover Drugs** **Kolker, Robert.** “How Machines Learned to Discover Drugs.” The New Yorker, 9 Sept. 2024, https://www.newyorker.com/magazine/2024/09/09/how-machines-learned-to-discover-drugs.	Balancing traditional laboratory work with AI-driven methods in drug discovery is challenging	Integrating AI to process vast datasets can predict effective molecules against pathogens, complementing traditional lab research and enhancing drug discovery efficiency.

Literature Review

S no.	Research papers	Challenge	Proposed Idea
6)	**Wang, Yuxin, et al.** "AdenPredictor: Accurate Prediction of the Adenylation Domain Specificity in NRPS." Bioinformatics, vol. 39, Supplement_1, 2023, pp. i40-i48	Predicting the substrate specificity of adenylation domains in NRPS is challenging due to the diversity of possible substrates.	The paper introduces "AdenPredictor," a machine learning tool designed to accurately predict the specificity of adenylation domains, facilitating the discovery and engineering of novel nonribosomal peptides.
7)	**Schmitt, Elisa, et al.** "Evolution-Inspired Engineering of Nonribosomal Peptide Synthetases." Science, vol. 380, no. 6643, 2023, pp. 1096-1101	Engineering NRPS to produce novel peptides is complex due to the intricate nature of these enzymatic assembly lines	The study discusses utilizing evolution-inspired strategies to modify NRPS, enabling the biosynthesis of new peptides with potential therapeutic applications

Traditional Methods of Drug Discovery and Their Limitations

- Trial-and-Error Approach – Drug discovery has traditionally relied on experimental screening, which is **slow and expensive**.
- Time-Consuming Process – It takes years to identify potential drug candidates and bring them to market.
- High Costs – The development of new drugs requires significant financial investment, often in billions of dollars.
- Limited Understanding of NRPS – Since NRPS pathways are **complex**, manually identifying potential peptides for drug development is difficult.

Challenges in Analyzing NRPS Manually

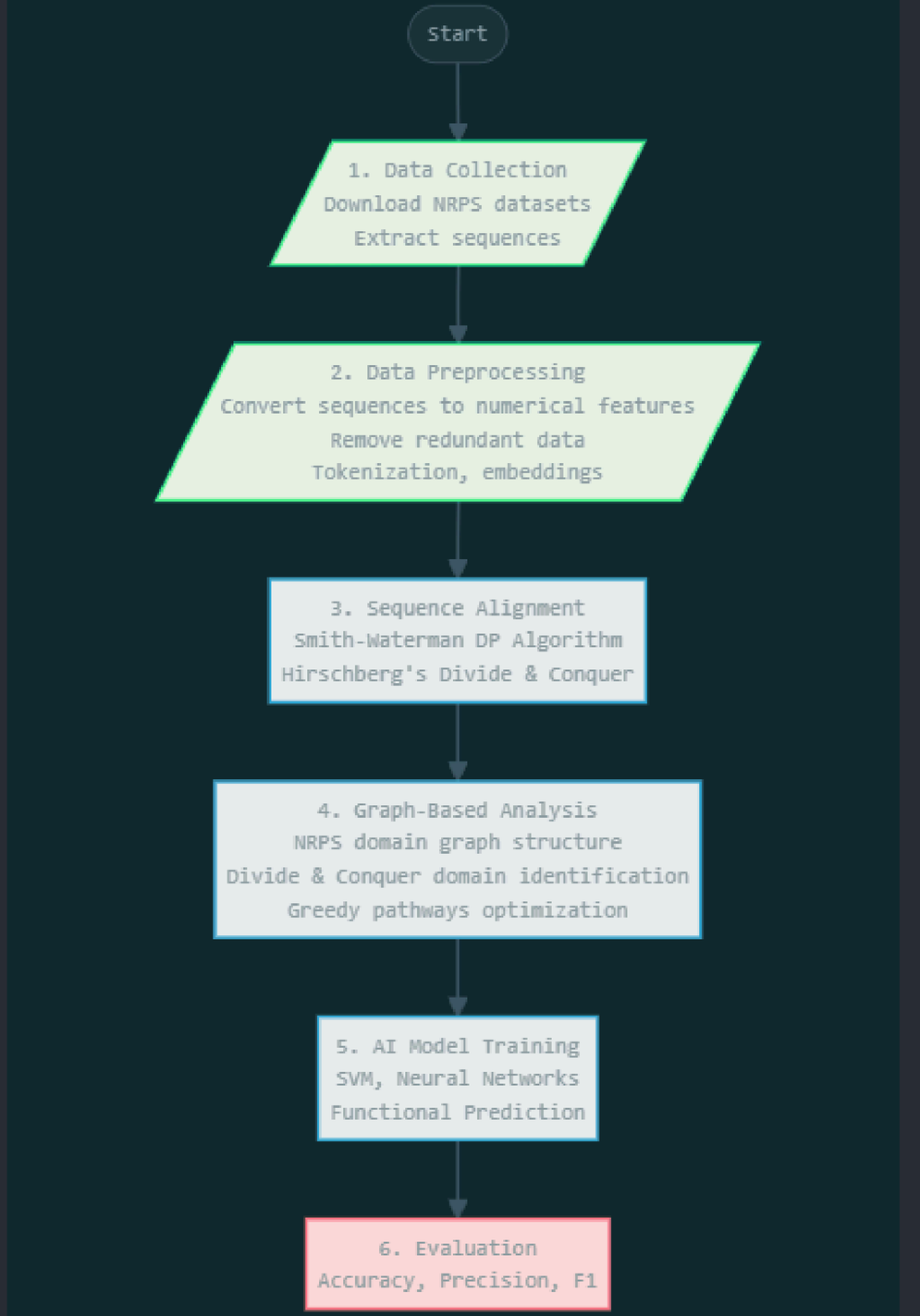
1. Complexity of NRPS Pathways – NRPS consists of **multiple enzymatic domains**, making it hard to predict peptide sequences.
2. Huge Data Volume – **Large datasets** need to be analyzed, which is beyond human capability.
3. Lack of Efficient Prediction Models – **Traditional computational** tools fail to accurately predict new NRPS-derived compounds.
4. Error-Prone Analysis – Manual methods are subject to **human error** and **inconsistencies**.

How AI and Algorithms Can Improve Efficiency

- Automating NRPS Analysis – AI can process **large datasets** rapidly, reducing human effort.
- Accurate Peptide Prediction – Machine Learning models can identify peptide synthesis patterns **more precisely** than manual methods.

1. Using dynamic programming approaches like **Smith-Waterman** for **finding functional NRPS domains.**
2. Hirschberg's Algorithm (**Divide and Conquer**) significantly **reduces memory usage** while maintaining accuracy. It **recursively** divides sequence alignment problems into **smaller subproblems**, solving them efficiently and optimizing the entire NRPS sequence

- Graph-Based Domain Analysis—AI can model NRPS as a shortest pathway using **graph structure** i.e. greedy algorithm, and **clustering** and D&C and optimized score help modify pathways.


Start
1. Data Collection
Download NRPS datasets
Extract sequences
2. Data Preprocessing
Convert sequences to numerical features
Remove redundant data
Tokenization, embeddings
3. Sequence Alignment
Smith-Waterman DP Algorithm
Hirschberg's Divide & Conquer
4. Graph-Based Analysis
NRPS domain graph structure
Divide & Conquer domain identification
Greedy pathways optimization
5. AI Model Training
SVM, Neural Networks
Functional Prediction
6. Evaluation
Accuracy, Precision, F1

Graph-Based Approach for NRPS Analysis

NRPS domains function as modular units, with each domain responsible for selecting and incorporating specific amino acids into peptides. This modularity makes graph representation an ideal choice:

- Nodes represent different functional domains.
- Edges represent interactions between these domains.
- Graph traversal enables a better understanding of NRPS synthesis pathways.

This representation allows us to efficiently analyze domain relationships, clustering, and pathway modifications for peptide discovery.

Applying Clustering Techniques

- Helps in identifying groups of domains that work together to synthesize specific peptide classes
- Useful for identifying relationships between NRPS sequences.
- Helps in recursively breaking down NRPS sequences into manageable subgroups.

Finding Shortest Modification Pathways (Using Greedy Algorithm)

After domain identification, we optimize the modification pathways using a Greedy Algorithm to identify the best sequence of modifications for synthesizing a desired peptide.

Greedy Algorithm Steps:

1. Initialize: Start from the root domain in the NRPS synthesis pathway.
2. Selection Step: At each step, choose the most optimal domain modification (e.g., most likely enzymatic reaction) based on probability scores.
3. Path Construction: Continue modifying until the shortest possible pathway to the target peptide is formed.
4. Optimization: Ensure that selected modifications produce a viable peptide with desired bioactivity.

Why Use a Greedy Approach?

- Ensures fast and efficient selection of NRPS modifications.
- Reduces unnecessary pathway exploration, focusing only on the most promising modifications.
- Improves computational efficiency compared to exhaustive search methods.

NRPS Functional Prediction

- In this step, we develop a Machine Learning (ML) model to predict the function of NonRibosomal Peptide Synthetases (NRPS). This AI-driven approach enhances the efficiency of identifying novel bioactive compounds by learning from existing NRPS datasets.
- We use two primary Machine Learning models for predicting NRPS functionality.

1. SVM
2. Neural Networks

Why SVM?

- Works well for high-dimensional datasets like protein sequences.
- Efficient for classification problems, such as predicting whether a given NRPS domain belongs to a specific functional class.

How SVM is used in our project?

1. Convert NRPS sequences into numerical feature vectors (e.g., using embeddings).
2. Train SVM on labeled NRPS datasets with functional annotations.
3. Use the trained model to classify unknown NRPS sequences into functional categories.

How Neural Networks are used in our project?

1. Convert NRPS sequences into numerical representations (embeddings, one-hot encoding).
2. Train a deep learning model (e.g., LSTM, CNN, or Transformer) on NRPS data.
3. Predict functional roles of NRPS domains and potential drug-like properties.

- SVM is used for efficient classification tasks and works well with complex, nonlinear data.
- Neural Networks can perform advanced predictions, such as drug-like property analysis.

Model Performance Metrics

Accuracy

- Measures the overall correctness of predictions.
- High accuracy indicates that the model correctly classifies NRPS domains into functional categories.

Precision (Positive Predictive Value)

- Measures how many of the predicted positive classes (functional NRPS domains) are actually correct.
- A high precision value ensures fewer false positives, reducing incorrect drug candidate selections.

F1-Score (Harmonic Mean of Precision & Recall)

- Balances Precision and Recall, especially useful when data is imbalanced.
- A high F1-score ensures strong model performance in classifying functional NRPS domains.

Output

The primary outcome of this project is an AI-driven analysis framework for NonRibosomal Peptide Synthetases (NRPS), which aids in drug discovery by efficiently identifying and predicting NRPS functions. The key outputs include:

✓ Accurate classification of NRPS domains based on functional roles.

✓ Optimized sequence alignment using Hirschberg's Divide & Conquer Algorithm to improve similarity matching.

✓ Graph-based NRPS domain identification using clustering and recursive Divide & Conquer techniques.

✓ Shortest modification pathway identification using the Greedy Algorithm, helping in bioengineering drug synthesis.

✓ Functional prediction of NRPS sequences using AI models (SVM, Neural Networks) with hyperparameter optimization

✓ Finding the evaluation metrics like precision, F1 score, Recall.

Thank You