JOURNAL OF NANO- AND ELECTRONIC PHYSICS
Vol. **16** No 5, 05001(5pp) (2024)

ЖУРНАЛ НАНО- ТА ЕЛЕКТРОННОЇ ФІЗИКИ
Том **16** № 5, 05001(5cc) (2024)

OPEN ACCESS

REGULAR ARTICLE

Artificial Intelligence Analysis of Protein Compositions on Engineered Nanomaterials

P. William[1,*] ✉, N. Yogeesh[2], Lingaraju[2], R. Chetana[3], T.N. Vasanthakumar[2], V. Verma[4]

[1] *Department of Information Technology, Sanjivani College Engineering, Kopargaon, SPPU, Pune, India*
[2] *Department of Mathematics, Government First Grade College, Tumkur, Karnataka, India*
[3] *Department of Mathematics, Siddaganga Institute of Technology, Tumkur, Karnataka, India*
[4] *Department of Computer Science and Engineering, Bhilai Institute of Technology, Raipur, Chhattisgarh, India*

(Received 12 June 2024; revised manuscript received 14 October 2024; published online 30 October 2024)

Protein compositions applied on Engineered Nanomaterials (ENM) require the presence of nanoscale protein molecules for multiple biochemical uses. Potential toxicity hazards and the requirement for full safety evaluations caused by the complex interactions between nanoparticles and biological systems are issues. Effectively Fluorescamine approaches for predicting protein composition on synthetic nanomaterials ENM can clarify biochemical findings from ENMs that are in biological structures without needing long-term protein composition tests. The Polypeptide Chemical Reaction Optimized Resistant Logistic Regression Model (PCRO-RLRM) is an innovative Artificial Intelligence (AI) technology that would be utilized in this research. The protein composition is analyzed using the Z-score normalization technique. The key elements from the normalized data that are useful for studying proteins or amino acid areas are extracted using the Position-Specific Scoring Matrix, or PSSM. Applying Polypeptide Chemical Reaction Optimization (PCRO) to modify the algorithm's parameters improves the predicted performance of the RLRM method. The findings reveal that the PCRO-RLRM combination is superior to the analysis of Protein Composition algorithm in accuracy (96.57 %), sensitivity (94.5 %), and specificity (98.03 %). This novel approach has the potential to promote findings in biochemistry based on nanomaterials and to improve bioengineering techniques.

Keywords: Engineered nanomaterial (ENM), Protein compositions, Artificial Intelligence (AI), Polypeptide chemical reaction optimized resilient logistic regression model (PCRO-RLRM).

DOI: 10.21272/jnep.16(5).05001 PACS number: 07.05.Mh

1. INTRODUCTION

The field of nanotechnology has become innovative, allowing improvements in a wide range of scientific fields. Due to their unique properties, engineered nanomaterials (ENMs) were suitable for an array of fields, including electronics, ecological restoration and healthcare [1]. Understanding the method these types of nanomaterials affect protein becomes essential as they relate to organisms [2]. Knowledge of protein composition on ENMs that are rapidly growing field that craves to clarify the complex interactions between biology and nanomaterials. This broad research examines complex connections at the nanoscale, providing insight into potential impacts on environmental safety, human wellness, and the growth of new technologies [3]. The processes of proteins connect to ENMs. The importance of data regarding the security and bio-compatibility of nanomaterials could be acquired by analyzing the binding motion, conformational shifts, and post-interaction activities [4]. Researchers have the ability to understand the molecular mechanisms behind the protein-ENM interface by using a range of analytical techniques, including mass spectrum, fission resonance, and surface Plasmon resonance, to describe such interactions [5]. The larger effects on human wellness and surroundings must be taken seriously while analyzing protein composition on artificial nanomaterials [6]. To create and utilize nanotechnology, it would be necessary to address security issues, analyze complex biochemical pathways and depend on the functional results of protein-ENM interactions [7-8]. The main objective of the study is to analysis of protein compositions on engineered nanomaterials. To gain a full understanding and how proteins interact with nanomaterials.

2. RELATED WORK

A Study [9] showed the Bradford proteins test was a useful method for measuring the total quantity of protein absorbed after the alkaline hydrolysis of NP solutions based on the PLGA. A study [10] discovered the proso millet protein to be utilized in a barrier material encasing

* Correspondence e-mail: william160891@gmail.com

Cite this article as: P. William et al., *J. Nano- Electron. Phys.* 16 No 5, 05001 (2024) https://doi.org/10.21272/jnep.16(5).05001

curcumin, in which the protein was wet milled or extracted with 60% (v/v) ethanol in water. The Study [11] overviewed the effects of positive, negative and irrelevant, dendrimers as nonmetallic ENMs and metallic ENMs (heavy metals, metal oxides, semiconductors) in various biological systems. The study [12] discussed systemic absorption and toxicological properties of Nanomedicines and provided a framework for future clinical research on Nano particles-based medication delivery systems. It would highlight the behavior of NPs in biological media. The study [13] covered a broad spectrum of topics, such as the process enabling the formation of green nanomaterials, the parameters influencing the synthesis, including the harmful consequences of the resulting particles. The effects that have upon living things, such as good bacteria, individuals, pets and the environment in its entirety provide growing issues. Study [14] examined the nuclear corona protein in blood using mass chromatography (MS) based proteins and spectroscopic. Study [15] explored that the surface chemistry and nanoparticle dimensions affect protein corona growth, which in turn affects cell attachment, absorption, and transport. Consequently, charge polystyrene nanoparticles (PSNPs) measuring 50 and 200 nm and modified using sulfone or carbohydrates were injected into Caco-2 intestine cells at 9 nominal concentrations (15 – 250 μg/ml) during 10 – 120 minutes.

3. METHOD AND MATERIALS

This study suggests using a Polypeptide chemical reaction-optimized resilient logistic regression model (PCRO-RLRM) to identify protein compositions. We collected data from the ENM to identify Protein Compositions. Position-specific scoring matrix (PSSM) is used to extract the pertinent feature from the preprocessed data after Z-Score Normalization. Fig. 1 shows the flow of this study.

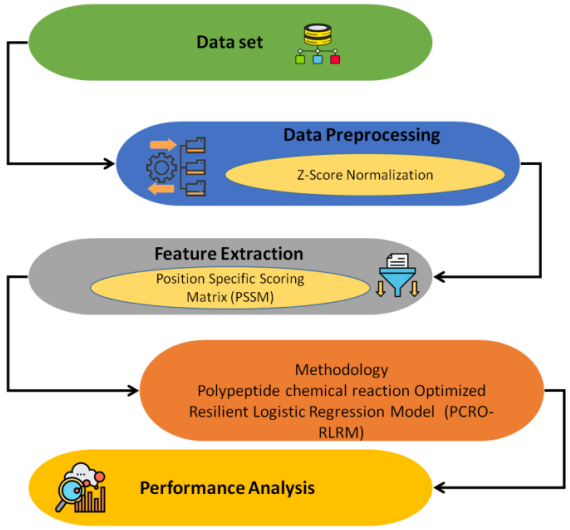

Fig. 1 – Overall Flow

3.1 Dataset

In this study, 962 distinct yeast proteins were identified as being enriched on Ag ENMs in the Ag ENM Pc database. Ag ENMs are widely employed in consumer products and the collection of data has a lot of proteins. The artificial intelligence algorithm was built using previously published information on the yeast amino acids enhancement on silver ENMs. [16]. Table 1 shows training factors.

Table 1 – Domains of chemical properties in the target and training datasets

Training factor	Range within dataset	Variable type
Features of proteins		
Weight of protein	6 to 559 kDa	Constant
Protein content	10-7.40-10-4.17	Constant
Positive amino acid	4.72–39.00	Constant
Negative amino acid	0–33.33	Constant
ENM characteristics		
ENM attributes	10 nm and 100 nm	Qualitative
ENM dimensions	Positive (+) and negative (–)	Qualitative

3.1.1. Creation of ENMs

Poor nutrition components show that the amino acids are richer in fluid & aren't included in the protein complex (non-PC), while positive nutrition components show that the amino acids have gone into the PC or enhanced on Ag ENMs. There are 3012 protein enrichment levels in total, listed as columns in the database; of these, 1208 protein component loops are classed as non-PC (40%) and 1810 amino particle packages as PC (60%) respectively. 1950 cells indicate that all proteins are made up of categories and periodic variables. Any protein connects with its corresponding type variable which indicates whether it fits into the Protein Class (PC) or not.

3.1.2. Metalized ENM

There are forty-four distinct nanoparticles (NPs) based on size, R_1 and R_2 reflexivity, and zeta potential, which is derived from the outermost intensity .This study evaluated the toxicity of 87 metallic and metal-oxide, dendrimer and polymer nanoparticles (NPs) on protein embryos using 6 different identifiers: concentration, functional groups, surfaces charge and zinc content. The minuscule particles were classified as dangerous by employing distinct threshold values for many endpoints, such as a CC_{50}, EC_{50}, IC_{50}, TC_{50}, and LC_{50}. Studies have also demonstrated their calculations of the cytotoxicity in NPs are more than 90% accurate. Two cubic nanoparticles of different materials (Al, Fe, Cu, Ag, Au, Pt) with a width of 0.6 – 0.3 nm are contained in the minerals.

3.1.3. Characteristics of Engineered Nanomaterials

For substances with the same qualities that must be produced commercially in a repeatable manner, characterization of the ENM is an essential phase. After that value or efficiency of an ENM in a particular application is compared, or its possible negative effects are evaluated. In multiple studies, that has been noted that poor characterization of ENM poses a significant challenge, limiting the replication and verification of investigations.

3.2 Data Preprocessing

Data preprocessing is an important phase in the artificial intelligence (AI) process for information analysis. It often included maintaining, transforming, and arranging basic information into a format that is suitable for additional research. The Z-score, which is presented in normal distributions, indicates how far certain protein compositions result from the mean. When a protein's Z-score is positive, it implies its composition is higher than the mean and when it is undesirable, it shows that it is lower than the average.

3.2.1. Z–Score Normalization

The Z–Score is commonly used as a score normalization method. It can be calculated by the typical variance and numerical mean. It is anticipated that this normalization method would function effectively if a normal score and mean score were well-known. Normalized score values exist by Eq. (1):

$$t_l' = \frac{t_l - \mu}{\sigma} \quad (1)$$

In this study, Z–Score normalization is helpful when comparing protein composition with possibly diverse scales or units.

3.3 Feature Extraction

Feature extraction is the procedure of identifying and modifying pertinent aspects from unprocessed data to provide a compact and informative version. A Position-Specific Scoring Matrix (PSSM) applied to informatics to illustrate the probability of each potential amino acid at each location in a sequence alignment is called PSSM. They are helpful in the investigation of proteins or amino acid regions.

3.3.1. Position-Specific Scoring Matrix (PSSM)

Utilizing the PSI-BLAST technique with three iterations and a cutoff E-value of 0.001, utilizing every amino acid structure as a starting point to represent biological data. The PSSM was built using several alignments of the highest-scoring hits compared to the first BLAST search. Using a dimension of K × 20 PSSM is a 20 amino acids matrix.

$$e(w) = \frac{1}{1+f^{-w}} \quad (2)$$

Where w is the original PSSM variable. We collected the Polypeptide chemical reaction (PCR) and amino acid composition (AAC) from the O_{PSSM} to convert a descriptor into a size-uniform matrix. The amino acid sample was symbolized by,

$$O_{PSSM} = \begin{bmatrix} o_{1,1} & o_{1,2} & o_{1,20} \\ o_{2,1} & o_{2,2} & o_{2,20} \\ o_{K,1} & o_{K,2} & o_{K,20} \end{bmatrix} \quad (3)$$

$$O = (o_1, o_2, \ldots\ldots o_i, \ldots\ldots, o_{20})^S (i = 1,2, \ldots\ldots,20) \quad (4)$$

$$o_i = \frac{1}{K}\sum_{j=1}^{K} o_{j,i} \quad (i = 1,2, \ldots\ldots,20) \quad (5)$$

Where o_i is the mean score of amino acid sequences in protein and a measure of base type i in PSSM.

$$O = (C_{1,1}, C_{1,2}, \ldots, C_{1,20}, C_{2,1}, C_{2,2}, \ldots. C_{2,20}, \ldots, C_{20,1}, C_{20,2}, \ldots\ldots, C_{20,20}) \quad (6)$$

$$C_{j,i} = \frac{1}{K-1}\sum_{l=1}^{K-1} o_{l,i} \times o_{l+1,i} \quad (1 \le j, i \le 20) \quad (7)$$

Where $C_{j,i}$ can be calculated by Eq. (7).

4. EXPERIMENTAL RESULT

The proposed solution is executed using Python version 3.10.1 on a Windows 10, provided with an Intel i7 core processor and 8GB of RAM. The suggested technique is (PCRO-RLRM) compared to existing methods such as Random forest (RF), K-Nearest neighbor (KNN), and Support vector machine (SVM).The performance of these methods is examined based on accuracy, sensitivity, and specificity. The study uses a novel technique called PCRO-RLRM with the goal of exploring the potential uses and collaborative effects of nanoparticles in protein engineering. The objective of the research is to explain the complex interactions between engineered nanomaterials and amino acids by applying novel approaches like PCRO and RLRM. The research findings have the potential to facilitate the creation of novel approaches for modifying protein structures and functions with the targeted integration of nanomaterials, therefore improving the scope of potential applications in diverse biochemical processes and biotechnology developments.

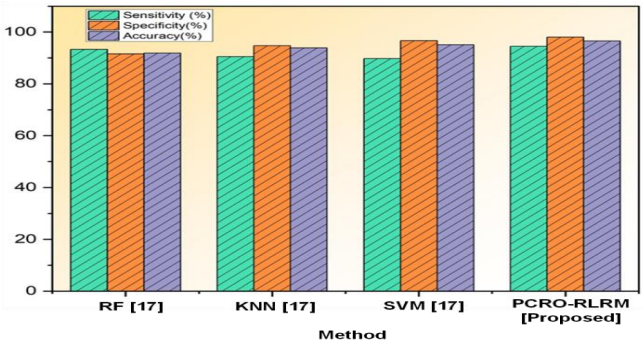

Fig. 2 – Overall Parameters

Table 2 – Values for Sensitivity, Specificity, Accuracy

Method	Sensitivity (%)	Specificity (%)	Accuracy (%)
RF [17]	93.3	91.58	91.86
KNN [17]	90.5	94.75	93.89
SVM [17]	89.79	96.66	95.14
PCRO-RLRM [Proposed]	94.5	98.03	96.57

The PCRO-RLRM system exhibits superior accuracy (96.57%) in early quality detection of protein compositions on Engineered Nanomaterials (ENMs). Outperforming existing methods such as RF (91.86%), KNN (93.89%), and SVM (95.14%), the proposed system demonstrates higher precision, enhancing reliability in assessing protein composition. Sensitivity in the design of PCRO-RLRM-based ENM on Protein Compositions. Figure 2 and Table 2 show the Sensitivity performance. The proposed PCRO-RLRM system achieves a Sensitivity of (94.5%). The existing system RF, KNN, and SVM, which are respectively 93.3%, 90.5%, and 89.79%. As a result, the proposed system is more precise than the existing approaches of early quality detection in the Protein Compositions.

Specificity measures the ability to exclude irrelevant data and avoid false positives. PCRO-RLRM system demonstrates high Specificity (98.03%), surpassing RF (91.58%), KNN (94.75%), and SVM (96.66%) in existing systems. Figure 3 depicts peptides with a significant concentration increase, crucial in biochemical reactions. Figure 4 showcases peptides with balanced positive and negative amino acids. Figure 5 highlights extreme acidity based on the isoelectric point. Accurate Fusion Protein (FC) identification is evident with high TP rates and low FP rates. Challenges in distinguishing FC and no FC arise with low TP rates and high FP rates.

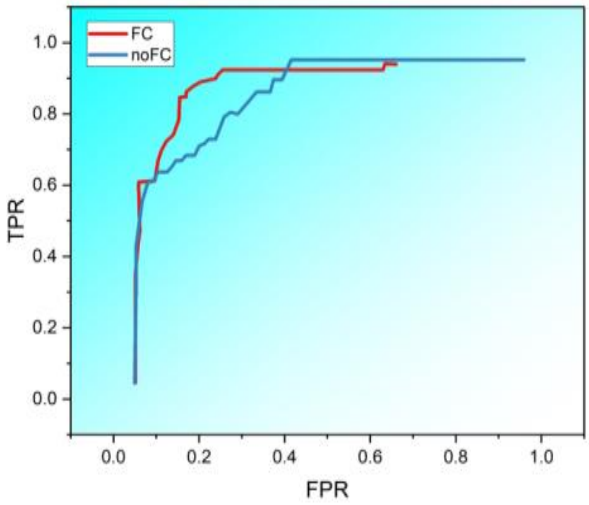

Fig. 3 – Peptides with abundant increase (AC) > 1

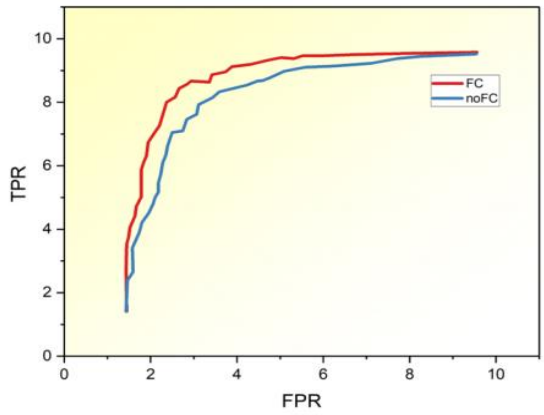

Fig. 4 – Peptides with AC within − 1 and 0

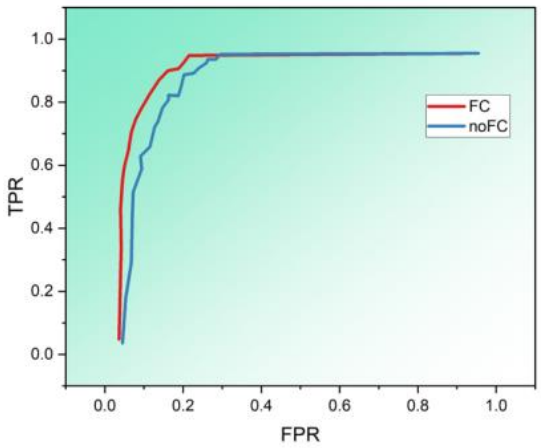

Fig. 5 – Protein with AC < − 1

Polypeptides are groups of amino acid sequences that are coupled to produce proteins that are necessary for a number of different metabolic activities. A rise in the false positive rate in biochemical tests could result in accurate polypeptide identification or characterization.

5. DISCUSSION

Frequently encountered disadvantages of RF may be extreme, when dealing with big datasets and many trees [17]. Using a lot of trees for training and prediction can use a lot of processing energy. KNN has some limitations; it is sensitive and noisy data. Outliers can significantly affect the calculation of distances and lead to less accurate predictions. Some of the falls in SVM that might hinder adaptation and ultimate efficiency are its reactivity to irrelevant or loud attributes, struggle with data inequalities, and requirement for proper tuning of parameters. [17].

6. CONCLUSION

The study concludes by highlighting the vital part that nanotechnology plays in several scientific fields and the importance of knowing the interactions that (ENMs) have with proteins. Investigating protein compositions on

synthetic nanomaterials, especially with the help of the suggested PCRO-RLRM model, offers light on intricate nanoscale interactions. Higher accuracy (96.57%), sensitivity (94.5%), and specificity (98.03%). The study statistics show that the PCRO-RLRM model exceeds other methods like RF, KNN, and SVM to detect protein compositions. Insufficient understanding of particular nanomaterials interact with proteins can prevent an examination of protein compositions, hence affecting accuracy of forecasts and analyses. The future holds opportunities of real-time dynamic interaction research, high quantities growth in technology and the combining of multi-omics for complete understanding of nanomaterial-protein structure in a variety of biochemical systems.

REFERENCES

1. M. Thakur, A. Sharma, M. Chandel, D. Pathania, *Green functionalized nanomaterials for environmental applications*, 259 (2022). Elsevier.
2. T.A. Saleh, *Environmental Technology & Innovation* **20**, 101067 (2020).
3. X. He, P. Fu, W.G. Aker, H.M. Hwang, *Journal of Environmental Science and Health*, Part C, **36** No 1, 21 (2018).
4. C. Liao, Y. Li, S.C. Tjong, *International journal of molecular sciences* **19** No 11, 3564 (2018).
5. S.S. Hinman, K.S. McKeating, Q. Cheng, *Analytical chemistry* **90** No 1, 19 (2018).
6. R. Feng, F. Yu, J. Xu, X. Hu, *Biomaterials* **266**, 120469 (2021).
7. A. Serra, G. Del Giudice, P.A.S. Kinaret, L.A. Saarimäki, S.S. Poulsen, V. Fortino, S. Halappanavar, U. Vogel, D. Greco, *Nanomaterials* **12** No 12, 2031 (2022).
8. N.A. Byzova, A.V. Zherdev, B.N. Khlebtsov, A.M. Burov, N.G. Khlebtsov, B.B. Dzantiev, *Sensors* **20** No 12, 3608 (2020).
9. K. Partikel, R. Korte, N.C. Stein, D. Mulac, F.C. Herrmann, H.U. Humpf, K. Langer, *European Journal of Pharmaceutics and Biopharmaceutics* **141**, 70 (2019).
10. L. Wang, P. Gulati, D. Santra, D. Rose, Y. Zhang, *Food Chemistry* **240**, 1039 (2018).
11. S.K. Verma, A.K. Das, M.K. Patel, A. Shah, V. Kumar, S. Gantait, *Science of the Total Environment* **630**, 1413 (2018).
12. T. Kopac, *International Journal of Biological Macromolecules* **169**, 290 (2021).
13. A. Rana, K. Yadav, S. Jagadevan, *Journal of Cleaner Production* **272**, 122880 (2020).
14. Z. Zhao, G. Li, Q.S. Liu, W. Liu, G. Qu, L. Hu, Y. Long, Z. Cai, X. Zhao, G. Jiang, *Journal of Hazardous Materials* **414**, 125582 (2021).
15. A. Abdelkhaliq, M. Van Der Zande, A. Punt, R. Helsdingen, S. Boeren, J.J. Vervoort, I.M. Rietjens, H. Bouwmeester, *Journal of Nanobiotechnology* **16**, 1 (2018).
16. M.R. Findlay, D.N. Freitas, M. Mobed-Miremadi, K.E. Wheeler, *Environmental Science: Nano* **5** No 1, 64 (2018).
17. Y. Wang, Q. Pan, X. Liu, Y. Ding, *Frontiers in Bioscience-Landmark* **27** No 1, 12 (2022).

Використання штучного інтелекту для аналізу білкових сполук на основі наноматеріалів

P. William[1], N. Yogeesh[2], Lingaraju[2], R. Chetana[3], T.N. Vasanthakumar[2], V. Verma[4]

[1] *Department of Information Technology, Sanjivani College Engineering, Kopargaon, SPPU, Pune, India*
[2] *Department of Mathematics, Government First Grade College, Tumkur, Karnataka, India*
[3] *Department of Mathematics, Siddaganga Institute of Technology, Tumkur, Karnataka, India*
[4] *Department of Computer Science and Engineering, Bhilai Institute of Technology, Raipur, Chhattisgarh, India*

Білкові композиції, що наносяться на розроблені наноматеріали (ENM), вимагають наявності нанорозмірних білкових молекул для багатьох біохімічних цілей. Потенційна небезпека токсичності та вимога повної оцінки безпеки, спричинена складною взаємодією між наночастинками та біологічними системами, є проблемними питаннями. Ефективні флуоресцамінові підходи для прогнозування білкового складу на синтетичних наноматеріалах ENM можуть уточнити біохімічні дані ENM, які знаходяться в біологічних структурах, без необхідності довгострокових тестів білкового складу. Модель стійкої логістичної регресії з оптимізованою поліпептидною хімічною реакцією (PCRO-RLRM) – це інноваційна технологія штучного інтелекту (ШІ), яка буде використана в цьому дослідженні. Білковий склад аналізують за допомогою методики нормалізації Z-score. Ключові елементи з нормалізованих даних, корисні для вивчення білків або амінокислотних ділянок, виділяються за допомогою позиційно-специфічної оціночної матриці, або PSSM. Застосування оптимізації поліпептидної хімічної реакції (PCRO) для зміни параметрів алгоритму покращує прогнозовану продуктивність методу RLRM. Результати показують, що комбінація PCRO-RLRM перевершує алгоритм аналізу складу білка за точністю (96,57 %), чутливістю (94,5 %) і специфічністю (98,03 %). Цей новий підхід має потенціал для сприяння відкриттям у біохімії на основі наноматеріалів і вдосконаленню методів біоінженерії.

Ключові слова: Штучно сформований наноматеріал (ENM), Білкові композиції, Штучний інтелект (AI), Оптимізована стійка логістична регресійна модель поліпептидної хімічної реакції (PCRO-RLRM).