DOI: 10.24996/ijs.2026.67.4.7

ISSN: 0067-2904

Characterization of Cysteine-Rich Peptides in Nodules from Different Legume Species

Aryan Mohamod Hamed[1], Najwa Ibrahim Khaleel Al-Barhawee[2]*

[1]*Biology Department, Pure Science and Education College, Al-Hamdaniya University, Mosul, Iraq*

[2]*Biology Department, Pure Science and Education College, Mosul University, Mosul, Iraq*

Received: 27/9/2024 Accepted: 9/4/2025 Published: 30/4/2026

Abstract

Nodules cystine-rich (NCR) peptides from legume nodules mediate bacteroid differentiation of *Rhizobium* sp. This study aimed to characterize NCR peptides of *Phaseolus vulgaris, Trifolium repens, Medicago sativa,* and *Vicia faba*, as relevant information on these legumes is still lacking. After exposing the root nodules of four leguminous plants to sonication, the peptides were separated according to their molecular weight by sodium dodecyl gel electrophoresis (SDS-PAGE). The NCR peptide was detected by high-performance liquid chromatography (HPLC), followed by the determination of the protein sequence and amino acid composition and their evolutionary history in terms of the phylogenetic tree by bioinformatics analysis. The results showed these four peptides had the same molecular weight of 60 kDa. The peaks of the curve varied with retention time depending on the legume from which the peptide was extracted. Bioinformatic analysis of the results showed that the amino acid substitutions varied in number and sequence for each species and that these peptides had three distinct wire, ribbon, and surface patterns within the structural modeling of their crystals. Side chains of amino acids in the loops and helices within the grid box may facilitate hydrophobic interactions via tyrosine, phenylalanine, or polar interaction sites via cysteine, histidine, and glutamine. These loops often serve as flexible regions to adapt to binding partners. In general, this study contributed to the characterization of NCR peptides that showed mismatches and differences in amino acid sequences and deletions of some amino acids at different locations depending on the type of legume from which the peptide was extracted.

Keywords: Bioinformatics, HPLC, NCR, Phylogenetic tree

توصيف الببتيدات الغنية بالسيستين في عقد أنواع مختلفة من البقوليات

اريان محمد حامد[1], نجوى إبراهيم خليل البرهاوي[2]*

[1]قسم علوم الحياة، كلية التربية للعلوم الصرفة ، جامعة الحمدانية ، الموصل ، العراق

[2]قسم علوم الحياة، كلية التربية للعلوم الصرفة ، جامعة الموصل ، الموصل ، العراق

الخلاصة :

تهدف هذه الدراسة إلى توصيف ببتيدات العقد الغنية بالسيستين التي تتوسط تمايز الرايزوبيوم الى البكتريويد . علما بان المعلومات ذات الصلة ببتيدات العقد البقولية في كل من *Phaseolus vulgaris* ، *Trifolium repens* ، *Medicago sativa* و Vicia faba، لا تزال غير متوفرة . بعد تعريض العقد

*Email: dr.najwa@uomosul.edu.iq

الجذرية لأربع نباتات بقولية للصوتنة، تم فصل الببتيدات وفقًا لوزنها الجزيئي بواسطة SDS–PAGE (الرحلان الكهربائي لهلام دوديسيل الصوديوم)، وتم الكشف عن الببتيد NCR بواسطة كروماتوغرافيا السائلة عالية الأداء (HPLC)، كما حدد تسلسل البروتين وتركيب الأحماض الأمينية، وتاريخها التطوري من حيث شجرة النشوء والتطور بواسطة تحليل البيانات بالمعلوماتية الحيوية. أظهرت النتائج أن هذه الببتيدات الأربعة كان لها نفس الوزن الجزيئي البالغ 60 كيلو دالتون ، بينما اختلفت قمم المنحنى مع زمن الاحتفاظ بحسب البقوليات التي استخرج منها الببتيد. في حين أظهرت النتائج بعد تحليل االبيانات بالمعلوماتية الحيوية أن بدائل الأحماض الأمينية اختلفت من حيث العدد والتسلسل لكل نوع، وأن هذه الببتيدات كان لها ثلاثة أنماط سلكية ، شريطية وسطحية متميزة ضمن النمذجة الهيكلية لبلوراتها. قد تسهّل السلاسل الجانبية للأحماض الأمينية في الحلقات والحلزونات داخل صندوق الشبكة التفاعلات الكارهة للماء، عبر التيروزين والفينيل ألانين، أو مواقع التفاعل القطبية، عبر السيستين والهيستيدين والغلوتامين. غالبًا ما تعمل هذه الحلقات كمناطق مرنة للتكيف مع شركاء الارتباط. وبشكل عام، أسهمت هذه الدراسة في توصيف ببتيدات NCR التي أظهرت عدم تطابق واختلافات في تسلسل الأحماض الأمينية وحذف بعض الأحماض الأمينية في مواقع مختلفة اعتمادًا على نوع النبات البقولي الذي استخرج منه الببتيد.

1. Introduction

Leguminous plants generate their root nodules through symbiosis with rhizobia. This soil-dwelling bacterium fixes nitrogen gas directly from atmospheric nitrogen and reduces it to ammonia as a food source. Inside the cells of the nodule, a symbiotic membrane formed from plants envelops the rhizobia, which separate into the nitrogen-fixing shape named as bacteroid, mediated by nodule cysteine-rich peptides (NCR) [1]. *Medicago truncatula* is an inverted repeat-lacking clade (IRLC) legume and its specialized symbiotic collaborator is *Sinorhizobium meliloti* [2]. Compared to other IRLC legumes, including those belonging to the genera of *Medicago, Vicia, Trifolium, Pisum,* etc., and *Aeschynomene* species*,* the nodules of this bean have a greater variety of NCR peptides including NCR169 and NCR211, which are essential for a successful symbiotic relationship [3], [4], [5]. Four or six cysteine residues are present in conserved places in NCR peptides. Their constituent amino acids are linked by disulfide linkages [6]. Only two NCR peptides have been studied structurally, and NCR044 is the only one with a three-dimensional structure. The disulfide linkages in NCR044 were located at C1 - C4 and C2 - C3, and in NCR247 at C1 - C2 and C3 - C421 (22). A key factor in bacterial differentiation is NCR169, a cationic peptide consisting of 38 amino acid residues including four cysteine amino acid residues and disulfide linkages [7].A recent study by Zorin *et al.,* investigated the gene family that codes for peptides rich in cysteines of root nodules in pea (*Pisum sativum* L.) and noted that cysteine-rich peptides have been extensively studied in the model legume *Medicago truncatula*, but relevant information is still fragmented in many other legumes [8].

As traditional extraction techniques lost their optimum performance, practitioners developed various methods to extract substances from different plant sources, and developments in the extraction process led to ultrasonic technology [9]. One of the main advantages of this technique is its low production costs and energy consumption, as well as its flexibility compared to other techniques [10]. Ultrasound-assisted extraction is one of the most promising techniques for extracting bioactive plant components. Because of its excellent performance with reduced solvent and time consumption, as well as its applicability to heat-sensitive chemicals, it is recognized as a green extraction approach [11]. The mechanism of ultrasound-assisted extraction (UAE) It involves applying the mechanical energy produced by ultrasonography to samples. As a result, the liquid develops tiny bubbles or cavities that explode into the solid sample, producing localized high pressures and

temperatures, resulting in the disruption of cell membranes and extraction of intracellular materials [12, 13].
S. Chuai, "Study on the mechanism of ultrasonic cavitation effect on the surface properties enhancement of TC17 titanium alloy", *Ultrasonics Sonochemistry*, vol.108, pp.106957, 2024.

The most commonly used technique for high-resolution analytical separation of protein mixtures is sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) according to size, structure, and molecular weight [14]. The procedure comprises first denaturing the proteins with an anionic detergent, which attaches to them and gives each one a negative charge according to its molecular mass, and followed by electrophoresis, the most commonly used technology for high-resolution analytical separation of protein mixtures is SDS-PAGE [15]. Thus, the assessment of the purity of protein samples, the evaluation of protein expression, and the quantification of proteins are methods that use SDS-PAGE [16, 17].

High-performance liquid chromatography (HPLC) has become the primary method for analyzing, purifying, and characterizing various starting peptides and proteins. It has contributed significantly to the recent progress of the biotechnological sciences [18, 19].This analytical technique has also been widely used to separate, identify and quantify compounds present in any sample; therefore, it has been widely used for quantitative and qualitative analysis of separated materials to determine their stability as a pharmaceutical product [20]. The sample is separated according to its molecular size by passing through a column that is filled with a substance whose pore size can be precisely controlled [21]. Recently, twelve free amino acids and nineteen bound amino acids, including L-glutamic acid and aspartic acid, were found by HPLC separation in the leaves of plants of the genus [22].

Bioinformatics has been defined as the application of computational techniques (informatics) to the analysis of biological data [23]More recently, it has been defined as one of the scientific disciplines related to genetics and genomics, which involves the use of computer technology to analyze biological data related to the sequences of DNA and amino acids and to provide a comprehensive interpretation of these sequences [24].

The growth of bioinformatics is a global project, as computer networks have made biological databases available to all researchers and allowed those interested in the field to develop software, databases and information for analysis quickly and easily [25], [18]. Bioinformatics uses computer programs for a variety of applications and analyses, particularly those related to determining the functions of genes and proteins and establishing evolutionary relationships between organisms [26].

The aim of this research was to characterize the amino acid sequences of cysteine-rich peptides, known as NCRs, extracted from the root nodules of four leguminous plants using several biological techniques, starting with sonication, followed by extraction of the peptide material by SDS-PAGE, detection by HPLC and the sequencing in the PDB files containing 3D structures data of four peptides were taken from different bioinformatics websites.

2. Material and Methods

2.1 Collecting root nodule samples

Four types of root nodules from four leguminous plants (*Phaseolus vulgaris, Trifolium repens, Medicago sativa,* and *Vicia faba*) were collected from local farms in the city of Mosul. According to Al-Barhawee and Ahmed [27], the nodules were washed with running

water to remove dirt, sterilized with 1% sodium hypochlorite solution, and washed several times with sterile distilled water to remove the effects of the sterilizer. They were dried on sterile filter paper for use in the sonication process for protein extraction.

Sonication process:

When preparing protein extracts, the cell is disrupted by sonication, where a titanium probe completely lyses the cells and extracts all DNA, RNA and protein contents from the cells.

Preparation of the sample Pre-sonication:

Sample preparation was sonicated as follows:

The nodules sample was washed with ice-cold phosphate-buffered saline (PBS), and the suspended nodules sample was placed in ice-cold PBS.

Protein extraction using the sonication procedure:

Samples were sonicated as follows: The sonicator probe was set to a frequency of 40 kHz and the nodule sample was placed in a 1.5 ml Eppendorf tube and suspended in 30-100 µl of RIPA buffer (50 mm Tris-HCl pH 7.4, 50 mm NaCl, 2 mm EDTA, 1% SDS, plus recently added the Phosphatase Inhibitor Cocktail (100X) is a proprietary mix of sodium fluoride, sodium pyrophosphate,β-glycerophosphate, and sodium orthovanadate).The microcentrifuge tube is then gently moved under the tip of the sonicator probe, causing the buffer to vibrate for 2-10 seconds and the cells may be sonicated 5 times, depending on the samples and sample viscosity. The cells were then incubated for 5 minutes on ice. Centrifuge at 10,000g for 20 minutes. The debris may contain undegraded cells, nuclei or organelles; the supernatants were collected and stored at -20 °C.

2.2Protein Purification:

The supernatant from the centrifugation was applied to a Sephadex G-25 gel filtration column and eluted with 5 mM Tris-Cl pH 8.0 and then applied to a Macro-Prep DEAE anion exchange column (Bio-Rad) for further separation. The column was eluted with 5 mM Tris-Cl pH 8.0 using a 50 mM NaCl gradient from 0-500 mM NaCl. Selected fractions were concentrated by centrifugation at 12,000 × g for 20 minutes at 4 °C. Salts were removed from the samples by repeating the concentration three times, each time washing the protein sample with fresh 5 mM Tris-Cl pH 8.0.

Protein gel extraction protocol:

The process involves localizing the protein of interest. Peptides can be separated according to their molecular weight using the SDS-PAGE (Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis) technology, removing the protein from the gel for subsequent experiments [28].

The detergent SDS is a component of the SDS-PAGE sample buffer. Peptide disulfide bonds are broken by SDS, which disrupts peptides' tertiary structure.

Prepare the gel:

All reagents except tetramethyl ethylene diamine (TEMED) are combined during gel preparation. When the gel is ready to be poured, TEMED is added with ammonium sulfate to act as a reaction catalyst. The gel is poured into the molding chamber in two layers: the resolving layer (bottom layer with 16% acrylamide and a high pH of about 8.8), a plastic comb is placed over the gel of this layer followed by the addition of the stacking gel solution, (top layer with a low acrylamide content of 5% and a pH of about 6.8), to separate the entire range from about 10 to 250 KDa. Note that each layer has different functions from the functions of the other layer. After removing the comb, 1 cm deep wells are formed into which protein samples are loaded.

Sample preparation:

50 µl 2-mercaptoethanol, 950 µl Laemmli loading buffer (106 mM Tris HCl, 141 mM Tris base, 0.51 mM EDTA, 0.22 mM SERVA Blue G-250, 0.175 mM phenol red, 2% LDS and 10% glycerol, at pH 8.5) from InvitrogenTM, at a final concentration of 355 mM, were added to the protein sample in a beaker, this mixture was transferred to microcentrifuge tubes and boiled at 95 degrees Celsius for less than five minutes to completely denature the proteins, then centrifuged at 16000 xg for five minutes and run against the protein marker by gel electrophoresis.

Electrophoresis:

In order to perform electrophoresis, take the gel cassette out of the casting stand and put it in the electrode assembly. Next, attach the electrode assembly to the clamp stand and pour one x electrophoresis buffer into it. (50 mM MOPS, 50 mM Tris Base, 1 mM EDTA, and 0.1% SDS, at pH 7.7) from Invitrogen^{TM} (trademark of Thermo Fisher Scientific Company), 30 milliliters of the denatured material should be pipetted into the wells of the gel to fill the aperture in the casting frame, in addition to filling the first well on the left with Protein Ladder 5-245 kDa (SMART SCIENCE CO., LTD), which was used as a standard for molecular weight. Cover the vessel with the lid and connect the unit to the power source, the voltage used and the run time of protein sample will vary depending on the percentage of resolving gel and stacking gel used and will not normally exceed 200V for 35 minutes and 100V for 40 minutes respectively. Place the gel in a plastic tray with Gel Fix solution. Place the tray on a rocking chair and leave for two hours. Add the Coomassie Blue protein staining solution and leave on the rocker for two to four hours. Wash the gel several times with distilled water, then add the destaining solution to the gel. Coomassie-stained protein gels can be imaged on the Odyssey imagers using NIR fluorescence.

2.3 Extraction of microproteins

After electrophoresis and staining with Coomassie blue dye, separation was carried out using a Protein Extraction Kit produced by Solarbio Life Sciences company, consisting of the following materials: 10 ml Buffer A, 60 ml Buffer B, 9.6 g dry powder, 40 mg DTT, filter column, combination tube.

2.4 Detection of NCR peptide by high-performance liquid chromatography (HPLC):

NCR peptide was detected by injecting 20 µl of the extracted protein into a high-performance liquid chromatograph (HPLC) using a C18 column and a mobile phase with methanol and water in a ratio of 80:20, respectively, at a flow rate of 1 ml/min at a wavelength of 365 nm [29].

2.5 Bioinformatic analysis

Construction of a Phylogenetic Tree with MEGA-X: MEGA-X (Molecular Evolutionary Genetic Analysis), employed by Kumar *et al.,* [30], was used to obtain sequences of the cysteine-rich peptides under study, and reliable databases such as UniProt were consulted and curated using FASTA software. Once the alignment work was complete, the "maximum likelihood" method of tree construction was chosen; this is suitable for making good estimates of evolutionary relationships. MEGA-X tools were then used to adjust the width of the tree to variable branch styles, collapse branches, add bootstrap values, etc., and then saved in the original MEGA format.

2.6 Bioinformatics to analyse the results

The PDB file was used to determine the atomic and morphological structure of the NCR peptides extracted from legume root nodules using the Pymol analysis program algorithms.

Far-ultraviolet light at 170-250 nm was used to determine the protein secondary structure content of alpha helices, beta sheets and loops.

3. Results and Discussion

Following NCR peptide electrophoresis of samples taken from the root nodules of four distinct leguminous plants (*Trifolium repens, Phaseolus vulgaris, Medicago sativa,* and *Vicia faba*) on polyacrylamide gel, the protein bands appeared on the gel. It was found that their molecular weight was equal and reached (60) KD as measured by the 245 kilodalton molecular weight marker, as shown in Figure 1.

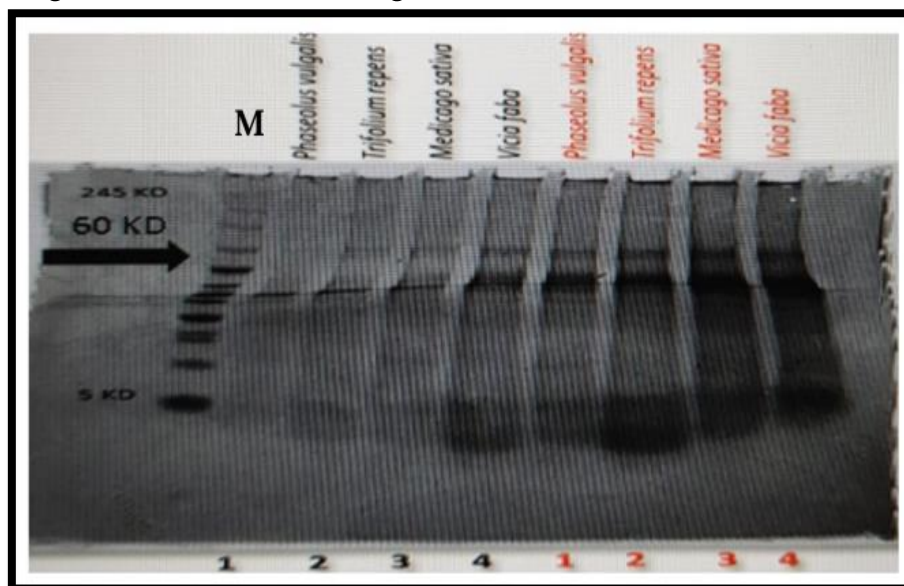

Figure 1: NCR protein bands in polyacrylamide gels electrophoresis.

M:Molecular weight marker.

A high-performance liquid chromatograph was used to separate and determine the concentration of each of the four peptides, and it was found that the highest peak of the curve was at retention times of 2.171, 2.121, 2.188, and 2.099 min in the non-infinitesimal stock of the peptides of *Phaseolus vulgaris*, *Trifolium repens*, *Medicago sativa*, and *Vicia faba* (Figure 5a-d), while the concentrations reached the following values: 50.218, 120.47, 80.2, and 120.613 mg/ml, respectively. Isozumi reported that the legume *Medicago truncatula,* which contains more than 600 cysteine-rich peptides, peaked at 17 and 19 minutes in a flow rate of 0.5 ml/min when analyzed by HPLC for two types of NCR 169 and at low concentration [6].

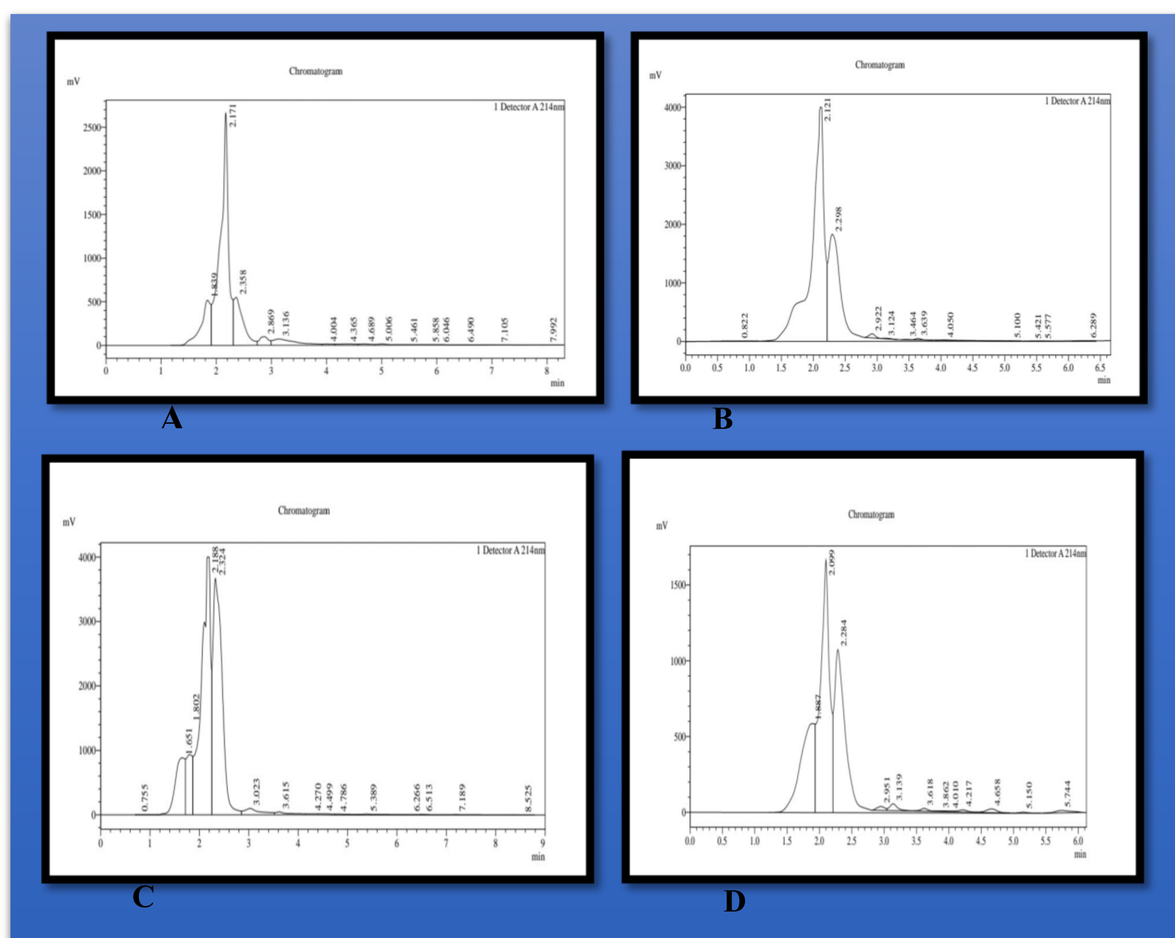

Figure 2: NCR peptide analysis using high-performance liquid chromatography (HPLC).
A- Phaseolus vulgaris, *B- Trifolium repens*, C- Medicago sativa, *and, D-* Vicia faba .

Sequence alignments were determined to highlight the similarities and differences between NCR proteins. Each colored block represents an amino acid; different colors for other amino acid residues according to an unspecified color coding system in the same image. Residues that are identical or highly conserved across species are placed in bins consistently throughout the alignment process. Such an alignment will be used to compare the sequence similarity and evolutionary relationships of NCR proteins between the four species studied, providing information on their functional roles and adaptations (Figure 3).

The alignment patterns obtained during bioinformatics analysis of *Rhizobium cremeum* sequences revealed distinct conserved regions across the studied isolates, particularly in motifs crucial for binding and interaction. Variations observed in non-conserved regions suggest potential functional differentiation. For instance, residues involved in hydrophobic interactions, such as tyrosine and phenylalanine, were predominantly located in loop regions, facilitating protein flexibility. These observations align with previously reported findings in similar proteins but also highlight unique features specific to *Rhizobium cremeum*.

This study shows the primary protein structure of the four NCR samples, defined as a series of amino acids linked linearly by peptide bonds [31]. They all begin at the N-terminus with the amino acid methionine, which has long been considered essential for the initiation of protein synthesis [32]. This amino acid has also been observed at the beginning of NCR peptide formation in both *P. sativum* and *M. truncatula* root nodules [8].Colour coding is a common convention used in many related studies, such as the sequencing of cysteine-rich peptides from *Aeschynomene* spp. [3]. Gaps, expressed as (-) in this figure, represent positions deleted from the sequence to improve the overall alignment between the sequences

[33]. Their differences in the type of amino acids they are composed of are probably due to differences in the genes responsible for encoding them, as pointed out by Zorin *et al.,* [8]. when explaining why the amino acid sequences of NCR peptides from alfalfa and pea plants differ. On the other hand, the paucity of studies characterising the NCR peptides of the legumes used in this study is probably due to the high interest in the NCR peptides of *Medicago truncatula* and the use of its results as a standard model for comparison, leaving the relevant information in many other legume species fragmented and incomplete. The native form of NCR peptides in legumes remains unknown because the amount found in roots is too small to be purified for analysis [6].

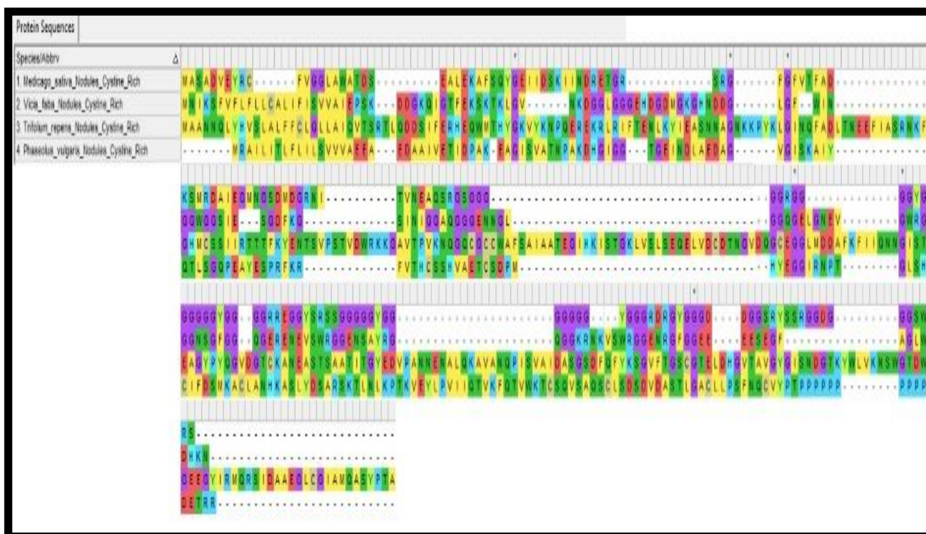

Figure 3:A multiple sequence alignment of four legume Nodules Cystine Rich (NCR) proteins of *Medicago sativa, Vicia faba, Trifolium repens,* and *Phaseolus vulgaris*.

As for the amino acid composition, both hydrophilic and hydrophobic, of the four peptides studied, it was observed that they are arranged in a line-by-line format with the amino acid names of each NCR type. The values in the columns represent the relative abundance of each amino acid present in these peptides. Within the columns, the ¨total¨ lines indicate the cumulative frequency of each species; the top row lines indicate the average frequency across the four species for each amino acid. In addition, the number of hydrophilic and hydrophobic amino acids was evenly distributed, as shown in Figure 4. A recent study by Alhhazmi confirmed that the proportion of hydrophobic amino acids did not exceed 50% in the NCR888 and NCR992 peptides from the legume *Medicago truncatula* [34]. The amino acid sequences of cysteine-rich peptides from *Aeschynomene* spp. were identical to the NCR peptides from *Medicago* spp. [3], but not to their sequences in the four peptides in this study.

	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu
Medicago sativa Nodules Cystine Rich	5.421686747	0.602409639	7.22891566	5.42168675	3.01204819	34.3373494	0	3.61445783	1.80722892	1.20481928
Vicia faba Nodules Cystine Rich	2.732240437	0.546448087	4.91803279	9.28961749	5.46448087	27.3224044	1.63934426	4.91803279	6.55737705	5.46448087
Trifolium repens Nodules Cystine Rich	9.037900875	2.915451895	4.95626822	6.12244898	4.08163265	9.32944606	1.74927114	6.12244898	6.12244898	5.5393586
Phaseolus vulgaris Nodules Cystine Rich	9.545454545	3.636363636	5.45454545	5.90909091	2.72727273	5.45454545	2.72727273	6.36363636	5	7.27272727
Avg.	7.236842105	2.192982456	5.48245614	6.57894737	3.8377193	16.5570175	1.64473684	5.48245614	5.15350877	5.15350877

	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr	Total
Medicago sativa Nodules Cystine Rich	2.40963855	2.40963855	0	1.20481928	9.63855422	10.2409639	2.40963855	2.40963855	1.204819277	5.421686747	166
Vicia faba Nodules Cystine Rich	1.09289617	7.65027322	0.546448087	2.73224044	3.82513661	6.55737705	1.09289617	3.82513661	3.278688525	0.546448087	183
Trifolium repens Nodules Cystine Rich	1.74927114	6.41399417	2.332361516	4.95626822	2.9154519	7.28862974	7.28862974	5.24781341	1.749271137	4.081632653	343
Phaseolus vulgaris Nodules Cystine Rich	1.36363636	2.72727273	9.090909091	3.18181818	3.18181818	9.09090909	7.27272727	6.81818182	0.454545455	2.727272727	220
Avg.	1.64473684	5.04385965	3.179824561	3.39912281	4.38596491	8.11403509	5.15350877	4.8245614	1.644736842	3.289473684	228

Figure 4: Amino acid composition for Nodules Cystine Rich (NCR) proteins from four leguminous plants: *Medicago sativa, Vicia faba, Trifolium repens*, and *Phaseolus vulgaris.*

The phylogenetic tree (Figure 5) describes the evolutionary history of various legumes with respect to their nodules and their cystine-rich content. It provides information on the following relationships; *Medicago sativa* nodules cystine rich and *Trifolium repens* nodules cystine rich are related as their common branch is supported by a bootstrap value of 61. These two are more distantly related to *Vicia faba* Nodules Cystine Rich, which is linked by a branch with a bootstrap value of 71. Finally, *Phaseolus vulgaris* Nodules Cystine Rich is the most distantly related of these four and branches separately. According to the results of evolutionary relationships, this difference can be attributed to the fact that plants have different tolerances when exposed to different selective pressures [31]. While the results of the study of nodule-specific evolutionary relationships in some legume species for cysteine-rich peptides (NCR) are arranged from closest to farthest in the following sequence: *Galega orientalis, Cicer arietinum*, and *Ononis spinosa*, followed by *Medicago, Trifolium, Pisum*, *and Vicia* [35].The results of amino acid sequencing of cysteine-rich peptides by alignment and phylogenetic analysis of the genes encoding NCR343, NCR341, NCR345 and NCR344 peptides in *Medicago truncatula* confirmed that sequence analysis of the NCR peptide is critical for its relationship to the differentiation of *Rhizobium* to Bacteroides and the successful fixation of atmospheric nitrogen to ammonia [36].

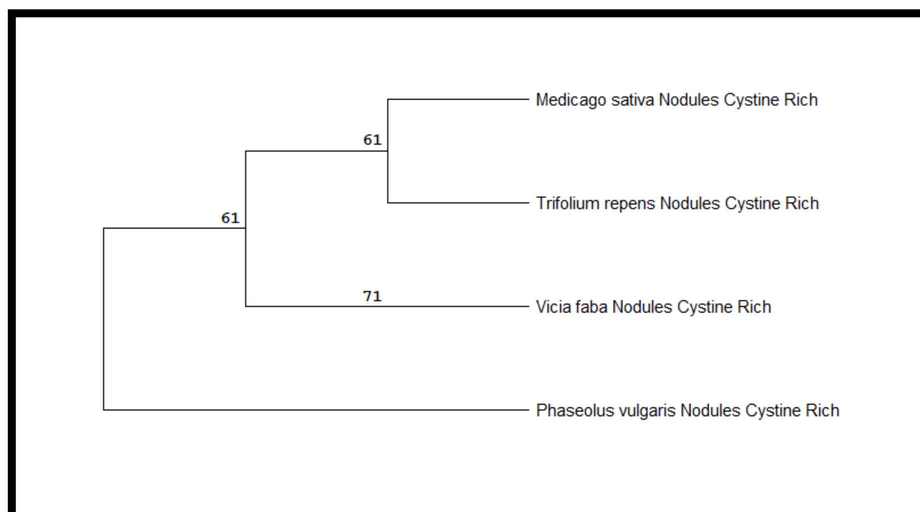

Figure 5: The phylogenetic tree shows the evolutionary history of nodules cysteine-rich (NCR) proteins in four different legume species.

The number of amino acid substitutions for each site between sequences is shown in Table (1). Analyzes were performed using the Poisson correction model noted by Zuckerkandl and Pauling [37]. This analysis included 4 amino acid sequences specific for each plant species, and for each pair of sequences all unclear place was eliminated using the pairwise deletion method. The final dataset includes 343 locations in total. Phylogenetic analyzes were performed using MEGA X software, described by Kumar *et al.,* [30].

Table 1: Estimates of evolutionary divergence between the amino acid sequences of cysteine-rich nodules.

Medicago sativa			
Vicia faba	1.32		
Trifolium repens	1.71	1.88	
Phaseolus vulgaris	2.16	2.21	2.69

Crystallization and structural modeling of cystine-rich nodule proteins extracted from root nodules of the legume *Medicago sativa*, *Trifolium repens*, *Phaseolus vulgaris* and *Vicia faba*, are analysed using bioinformatics tools and described in Figure 6.

The top image is of the NCR protein in a wireframe diagram (Figure 6, A), which shows the bonding between atoms, allowing the protein backbone to be seen and the position of the amino acids in 3D space to be determined. This diagram explains how the network of cross-links between amino acids forms the structure of a protein. It also provides insight into the complex network of bonds that determine the shape and molecular structure of this protein.
The second image shows the NCR protein in a ribbon diagram (Figure 6, B). This diagram shows the secondary structures of this protein, which are alpha-helices (coil-like parts) and beta-sheets (arrow-like parts). It provides a clear view of the folding pattern and structural elements of the protein that may be of functional importance.

The third, fourth and fifth images show the NCR protein in surface diagrams from three different directions (Figure 6, C). This diagram shows the 3D shape of the protein and the regions involved in molecular interaction. Green, red and blue generally represent carbon, oxygen and nitrogen atoms respectively. These color schemes are very useful for visualising the location of polar and non-polar regions, especially hydrophobic and hydrophilic regions on the protein surface. This is particularly important for the functional and interaction properties of a given protein. The three-dimensional structural modeling showed a combination of alpha-helices, beta-sheets, and flexible loop regions, indicative of a robust and adaptable structure. The surface diagram demonstrated distinct polar and hydrophobic regions, correlating with the protein's interaction capabilities. The binding interface, as revealed by the grid box positioning, emphasized residues critical for molecular recognition and stability. These findings provide insights into the structural mechanics of *Rhizobium cremeum* and its potential functional roles in symbiotic and environmental contexts.

In this context, bioinformatics computer programs have been used for a wide range of applications and analyses, especially regarding many gene and protein sequences, for use in revealing evolutionary relationships that have occurred over the years, facilitating their study, and predicting the three-dimensional shapes of cellular proteins with different functions [38]. A recent study [39], demonstrated the success of developing a new alignment that integrates seamlessly with protein bank data, allowing researchers in the field to align multiple protein sequences and compare them to a reference structure by providing specific alignment algorithms while designing them in 3D. On the other hand, these structures are an excellent means of understanding evolutionary relationships between and among proteins [40]. The folding phenomenon protects the NCR from degradation, and the disulfide bond leads to stabilization of the NCR peptides during
symbiosis [41].

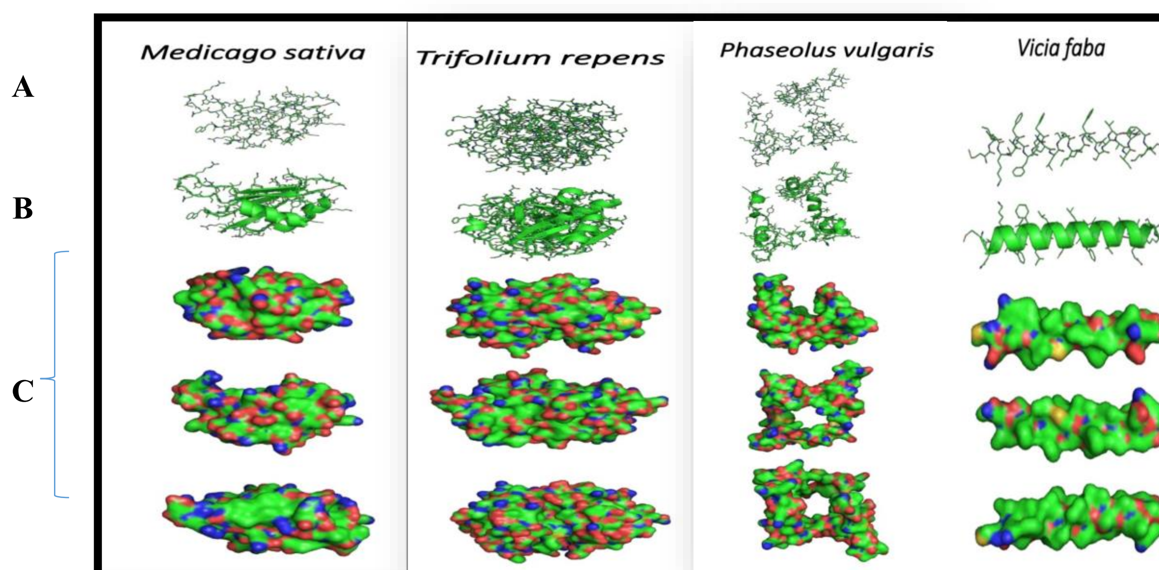

Figure 6: Crystallization and structural modeling of NCR proteins of *Medicago sativa, Phaseolus vulgaris, Trifolium repens,* and *Vicia faba*.
A: Wireframe diagram , B: Ribbon diagram, C: Surface diagram (Three Images)

As shown in Figure 7, within the interaction region, the middle part of the grid box probably represents the binding interface through which protein-protein interactions occur, and the structural arrangement here seems to provide a combination of surface accessibility through loops and helixes with stability through beta sheets, an emerging mechanism of such non-covalent peptide-surface interaction involves the addition of a beta-strand in the ligand to a beta sheet in the receptor [42]. Because the non-contributing forces that stabilize protein structures are not fully understood, one way to address this problem is to study the interactions between unfolded states and α-helices in peptides [43].

On the one hand, the amino acid side chains in the rings and helices within the grid box are likely to be hot spots that can facilitate hydrophobic interactions via tyrosine and phenylalanine, or polar interaction sites via cysteine, histidine, and glutamine, and these same amino acids have been observed in several types of NCR peptides from the legume *Medicago truncatula* [44].These loops often serve as flexible regions to adapt to binding partners. Therefore, interactions involving one or more amino acid side chains near the ends of protein helices stabilize the helical ends and shape the geometry of the adjacent loops, contributing significantly to the overall protein structure [45].The grid box indicates the active binding region for computational docking or interaction studies. Residues within this region are prime candidates for analysis in protein-protein docking, ligand binding, or catalytic activity [46]. Secondary structure results showed that two structures of the NCR169 peptides from the legume *Medicago truncatula* have a consensus C-terminal β-sheet attached to an extended N-terminal region with dissimilar features; one moves widely, whereas the other is relatively stapled [6].

Furthermore, structure comparison searches revealed that the structure of the four NCR peptides in this study did not fully match some of the structures reported in the PDB file, possibly due to the structural difference in the disulfide bonds connecting the loop to the β-sheet [47]. In this regard, Isozumi *et al.,* [6] indicated that NCR169-ox1 and NCR169-ox2 have a novel structure as they do not match any structure in the PDB, and they also differ in the position of the N-terminal region; in NCR169-ox1, the position of the extended N-

terminal region is non-specific, whereas the position of NCR169-ox2 is relatively stabilized by disulfide bonds and covers one side of the β-sheet.

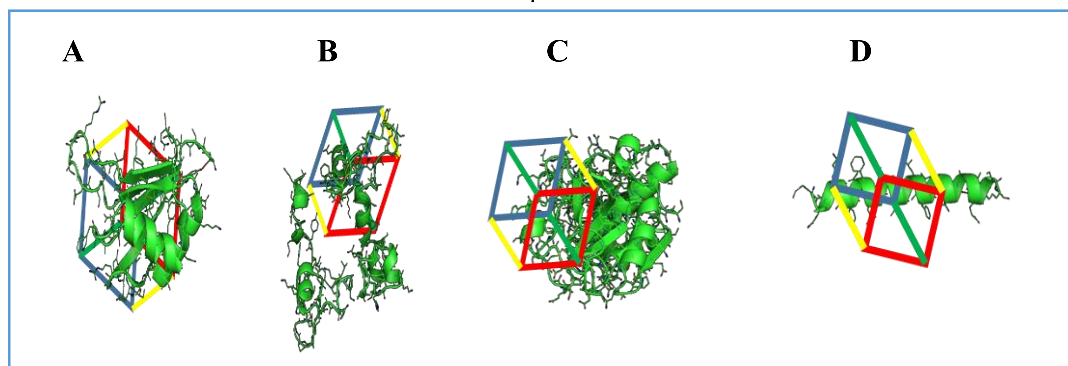

Figure 7: The grid box position wraps around most secondary structure elements, including alpha helices, beta sheets, and loops connecting these regions of the NCR in (A) *Medicago sativa*, (B) *Phaseolus vulgaris*, (C) *Trifolium repens*, and (D) *Vicia faba*.

4. Conclusions

Ultrasonic extraction at around 40 kHz is effective in extracting peptides from legume nodules. In addition to the success and accuracy of the SDS-PAGE tools in separating the peptides, as well as being a simple and inexpensive method, all the techniques used revealed significant differences between the four species in all their properties studied. In addition, their structures do not fully match the data in the PDB file, raising the possibility that these are new peptides discovered for the first time in bioinformatics programs.This demonstrates the speed and accuracy of the interpretation of complex scientific results and the importance of the convergence of biology and technology under the concept of bioinformatics. On the other hand, the fact that the plants used in this study belong to the same family (Fabaceae) does not mean that their NCR peptides are identical in terms of amino acid sequence or primary and secondary structure. Finally, the goal of identifying and characterizing NCR peptides was fully achieved in this study.

Acknowledgements

We would like to thank the University of Mosul, represented by its President, the Dean of the College of Education for Pure Sciences, and the Head of the Department of Biology, for the facilities they provided to carry out this research.
I want to acknowledge that this project is entirely the property of the researchers and no one else.

Conflict of Interest: the authors declare that have no conflict of interest

References:

[1] P.Mergaert,"Differentiation of symbiotic nodule cells and their rhizobium endosymbionts", *Advances in Botanical Research,* vol.94, pp.149-180, 2020.

[2] B. Horváth, B. Güngör, M. Tóth, Á. Domonkos, F. Ayaydin, F. Saifi, and P. Kaló, "The *Medicago truncatula* nodule-specific cysteine-rich peptides, NCR343 and NCR-new35 are required for the maintenance of rhizobia in nitrogen-fixing nodules", *New Phytologist,* vol.239, no.5, pp.1974-1988,2023.

[3] P. Czernic, "Convergent evolution of endosymbiont differentiation in dalbergioid and inverted repeat-lacking clade legumes mediated by nodule-specific cysteine-rich peptides", *Plant physiology*, vol.169,no.211, pp.1254-1265, 2015.

[4] B. Alunni and B. Gourion, " "Terminal bacteroid differentiation in the legume– rhizobium symbiosis: nodule-specific cysteine-rich peptides and beyond", *New Phytologist*, vol.211, no.2, pp.411-417, 2016.

[5] H. Pan, "Nodule cysteine-rich peptides maintain a working balance during nitrogen-fixing symbiosis", *Nature Plants*, vol.3, no.5, pp.1-6, 2017.

[6] N. Isozumi, "Structure and antimicrobial activity of NCR169, a nodule-specific cysteine-rich peptide of Medicago truncatula", *Scientific Reports*, vol. 11, no.1, pp.9923, 2021.

[7] B. Horváth, "Loss of the nodule-specific cysteine rich peptide, NCR169, abolishes symbiotic nitrogen fixation in the Medicago truncatula dnf7 mutant", *Proceedings of the National Academy of Sciences*, vol.112, no.49, pp. 15232-15237, 2015.

[8] E. A. Zorin, "A variable gene family encoding nodule-specific cysteine-rich peptides in pea (*Pisum sativum* L.) ", *Frontiers in plant science*, vol.13, pp.884726, 2022.

[9] M. M. A. Ranjha, "Sonication, a potential technique for extraction of phytoconstituents: A systematic review", *Processes*, vol. 9, no.8, pp.1406, 2021..

[10] M. R. Alhajhoj, "Importance and applications of ultrasonic technology to improve food quality", *Food processing*, vol.9, pp.144, 2019.

[11] I. M. Yusoff, "A review of ultrasound-assisted extraction for plant bioactive compounds: Phenolics, flavonoids, thymols, saponins and proteins", *Food research international*, vol.157, pp.111268, 2022.

[12] P. Geada, "Algal proteins: Production strategies and nutritional and functional properties", *Bioresource Technology*, vol.332, pp.125, 2021.

[13] S. Chuai, "Study on the mechanism of ultrasonic cavitation effect on the surface properties enhancement of TC17 titanium alloy", *Ultrasonics Sonochemistry*, vol.108, pp.106957, 2024.

[14] V. B. Gore, "Sodium Dodecyl Sulphate-polyacrylamide Gel Electrophoresis (SDS-PAGE) Characterization of Protein Fractions in Ten Peanut Bunch Type Varieties", *International Journal of Biochemistry Reserh &Reviw*, vol.33,no.6, pp.60-70, 2024.

[15] A. Q. Muhammed, "Efficacy of gel electrophoresis for proteins and biotechnological products–an overview", *Al Mustansiriyah Journal of Pharmaceutical Sciences*, vol.20, no.4, pp.45-56, 2020.

[16] D. W. Lim, "Improved Separation in Horizontal Protein SDS-PAGE with Double-Deck Flat Electrodes and a Field Inversion Gel Electrophoresis Module", *Methods and Protocols*, vol.6, no.6, pp.106, 2023.

[17] A. Q. Muhammed, "Efficacy of gel electrophoresis for proteins and biotechnological products–an overview", *Al Mustansiriyah Journal of Pharmaceutical Sciences*, vol.20, no.4, pp.45-56, 2020.

[18] M. I. Aguilar, "HPLC of peptides and proteins: basic theory and methodology", *HPLC of peptides and proteins: methods and protocols*, pp.3-8, 2004.

[19] A. S. Alhendi, "A Review: Protein Identification by LC-MS: Principles, Instrumentation, and Applications ", *Iraqi Journal of Science*, vol. 61, no.10, pp. 2448-2466, 2020.

[20] S. K. Bhardwaj, "A review: HPLC method development and validation", *International Journal of Analytical and Bioanalytical Chemistry*, vol.5, no.4, pp. 76-81, 2015.

[21] J. M. Belanger, "High performance liquid chromatography (HPLC)", *Techniques and instrumentation in analytical chemistry principles and applications*, vol. 18, pp. 37-59, 1997.

[22] L. Budniak, "Determination of amino acids of plants from *Angelica* L. genus by HPLC method", *Pharmacia*, vol.69, pp.437-446, 2022.

[23] A. Lesk, "Introduction to bioinformatics", *Oxford university press*, 2019.

[24] M. H. Medema, "The year 2020 in natural product bioinformatics: an overview of the latest tools and databases", *Natural Product Reports*, vol.38, no2, pp. 301-306, 2021.

[25] A. Bayat, "Science, medicine, and the future: Bioinformatics", *BMJ: British Medical Journal*, vol.324, no.7344, pp.1018, 2002.

[26] Z. A. Hussein and A. A. Al-Kazaz, "Bioinformatics evaluation of CRISP2 gene SNPs and their impacts on protein", *Iraqi Journal of Agricultural Sciences*, vol.54, no.2, pp. 369-377, 2023..

[27] N. I. K. Al-Barhawee, " Using Sequencing Technique for Diagnostic Different Species of Genus Rhizobium Which Isolated from Legume Plants ", *Iraqi Journal of Science*, vol.63. no.10, pp. 4213-4224, 2022.

[28] A. B. Nowakowski, "Native SDS-PAGE: high resolution electrophoretic separation of proteins with retention of native properties including bound metal ions", *Metallomics*, vol.6,no5, pp.1068-1078, 2014.
[29] N. I. K. Al-Barhawee, "Phytochemical analysis of Annona atemoya seed extract by HPLC and their ability to inhibit the growth of *Agrobacterium tumefaciens*", *Biodiversitas: Journal of Biological Diversity*, vol.24, no.1, pp.603-608, 2023.
[30] S. Kumar,"MEGA X: molecular evolutionary genetics analysis across computing platforms", *Molecular biology and evolution*, vol.35, no.6, pp.1547-1549, 2018.
[31] J. Durairaj, "Uncovering new families and folds in the natural protein universe", *Nature*, vol.622, no.7983, pp. 646-653, 2023.
[32] R. J. Savino, "The potential of a protein model synthesized absent of methionine", *Molecules*, vol.27, no.12, pp.3679, 2022.
[33] Z. Ghanbarzadeh, "The Roadmap of Plant Antimicrobial Peptides Under Environmental Stress From Farm to Bedside", *Probiotics and Antimicrobial Proteins*, vol.16, no.6, pp.2269-2304, 2024.
[34] A. A. Alhhazmi, "Novel antimicrobial peptides identified in legume plant, *Medicago truncatula*", *Microbiology Spectrum*, vol.12, no.2, pp. 18-23, 2024.
[35] X. Liu, " Expansion and evolutionary patterns of cysteine-rich peptides in plants" ,*BMC genomics*, vol.18,pp.1-14, 2017.
[36] J. Montiel, "Terminal bacteroid differentiation is associated with variable morphological changes in legume species belonging to the inverted repeat-lacking clade", *Molecular Plant-Microbe interactions*, vol.29,no.3, pp.210-219, 2016.
[37] L. Pauling, "Evolutionary divergence and convergence in proteins", *Academic Press* , pp. 97-166, 1965.
[38] D. B. Singh and R. K. Pathak, "Bioinformatics: methods and applications", *Academic Press*, 2021.
[39] S. Bittrich, "RCSB protein data bank: Exploring protein 3D similarities via comprehensive structural alignments", *Bioinformatics*, pp. 370, 2024.
[40] M. Varadi, "AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences", *Nucleic acids research*, vol.52, no.D1, pp.D368-375, 2024.
[41] M. Shabab, " Disulfide cross-linking influences symbiotic activities of nodule peptide NCR247" ,*Proceedings of the National Academy of Sciences*, vol.113, no.36, pp.10157-10162, 2016.
[42] G. Waksman, "Protein–protein interaction through β-strand addition", *Trends in biochemical sciences*, vol.31, no.8, pp. 436-444, 2006.
[43] E. G. Baker, "Local and macroscopic electrostatic interactions in single α-helices",*Nature chemical biology*, vol.11, no.3, pp. 221-228, 2015.
[44] M. Lima, "Unexplored arsenals of legume peptides with potential for their applications in medicine and agriculture", *Frontiers in Microbiology*, vol. 11, pp.1307, 2020.
[45] N. E. Newell, "Mapping side chain interactions at protein helix termini", *BMC bioinformatics*, vol.16, pp.1-21, 2015.
[46] D. González-Esparragoza , "In Silico Analysis of Protein–Protein Interactions of Putative Endoplasmic Reticulum Metallopeptidase 1 in *Schizosaccharomyces pombe*" ,*Current Issues in Molecular Biology*, vol.46,no.5, pp.4609-4629, 2024.
[47] S. J. Godwin, "Chickpea NCR13 disulfide cross-linking variants exhibit profound differences in antifungal activity and modes of action", *PLoS pathogens*, vol.20, no.12, pp. e1012745, 2024.