Journal of Molecular Graphics and Modelling 129 (2024) 108718

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/jmgm

Structural and functional analysis of Cyanovirin-N homologs: Carbohydrate binding affinities and antiviral potential of cyanobacterial peptides

Gabriel Xavier *, Alenna Crystiene Lima Farias de Sousa, Larissa Queiroz dos Santos, Délia Aguiar, Evonnildo Gonçalves, Andrei Santos Siqueira

Biomolecular Technology Laboratory/Institute of Biological Sciences, Federal University of Pará, Belém-PA, Brazil

ARTICLE INFO

Keywords:
Cyanobacteria
Lectins
Bioinformatics
Molecular Biology
Biotechnology

ABSTRACT

Cyanobacteria, a group of photosynthetic prokaryotes, can sinthesize several substances due to their secondary metabolism, with notable properties, such as Cyanovirin-N(CVN), a carbohydrate-binding lectin, that exhibits antiviral activity against several pathogens, due to its ability to bind viral surface carbohydrates such as mannose, thus interfering with the viral entry on the cell. CVN has been described in several cyanobacterial strains and shows biotechnological potential for the development of drugs of pharmaceutical interest. This study focuses on the genomic exploration and characterization of Cyanovirin-N homologs to assess the conservation of carbohydrate-binding affinity within the group. The analysis of their antiviral properties was carried out using bioinformatics tools to study protein models through an in silico pipeline, following the steps of genomic prospection on public databases, homology modeling, docking, molecular dynamics and energetic analysis. Mannose served as the reference ligand, and the lectins' binding affinity with mannose was assessed across Cyanovirin-N homologs. Genomic mining identified 33 cyanobacterial lectin sequences, which underwent structural and functional characterization. The results obtained from this work indicate strong carbohydrate affinity on several homologs, pointing to the conservation of antiviral properties alongside the group. However, this affinity was not uniformly distributed among sequences, exhibiting significant heterogeneity in binding site residues, suggesting potential multi-ligand binding capabilities on the Cyanovirin-N homologs group. Studies focused on the properties involved in these molecules and the investigation of the genetic diversity of Cyanovirin-N homologs could provide valuable insights into the discovery of new drug candidates, harvesting the potential of bioinformatics for large-scale functional and structural analysis.

1. Introduction

Cyanobacteria are a group of gram-negative prokaryotes notable for their ability to carry out photosynthesis using water, sunlight, and carbon dioxide, resulting in the production of carbohydrates and oxygen, more efficiently than plants and with more than half the resources [1]. These organisms are often referred to as biofactories due to their secondary metabolism [2] by which a variety of bioactive compounds can be synthesized, such as anticancer, antibacterial, and antiviral substances [2,3]. Among these, lectins stand out as a group of proteins capable of reversible binding to oligosaccharides without altering the ligand structure, a property that gives them potential to be explored on the scope of rational drug design.

Due to their ability of carbohydrate binding, lectins can interact with oligosaccharides rich in mannose, present on the cell surface of several viruses, including Human Immunodeficiency Virus (HIV), Simian Immunodeficiency Virus (SIV), Rhinovirus, Herpes Virus, Influenza, Zaire Ebola Virus, and even SARS-CoV-2 [4–6]. The mechanism of viral inhibition occurs by preventing the pathogen from entering the host cell, interfering with viral replication. Besides, viral envelope carbohydrates exhibit significant homology among viruses, turning molecules that interact with them into potential targets for drug development.

One of the most notable enzymes on the group of cyanobacterial metabolites is Cyanovirin-N, a lectin that has demonstrated its efficacy in preventing vaginal transmission of HIV-1 and HIV-2, acting as a topical microbicide on in vivo studies [7]. Additionally, Cyanovirin-N has shown the potential to protect cells by inhibiting the formation of syncytium between healthy $CD4^+$ T cells and HIV-1 infected T cells, thereby preventing viral fusion with the host cell [8]. This molecule was initially discovered as a metabolite produced by the cyanobacterial

* Corresponding author.
E-mail address: gabriel.xavier@icb.ufpa.br (G. Xavier).

https://doi.org/10.1016/j.jmgm.2024.108718
Received 22 October 2023; Received in revised form 23 January 2024; Accepted 23 January 2024
Available online 7 February 2024

species *Nostoc ellipsosporum* [9].

The protein structure of Cyanovirin-N (PDB code: 1IIY) consists of a peptide chain of 101 amino acids, with molecular mass of approximately 10 kDa and containing four cysteine residues that form two disulfide bonds, without which it loses its function. The structural composition of the carbohydrate-binding domain includes two pseudo-symmetrical domains, demarcated by amino acids 1–38 and 90–101 (Domain A), forming the low-affinity domain, with a disulfide bond between residues Cys-58 and Cys-73; and amino acids 38–89 (Domain B), the high-affinity domain, with a disulfide bond between Cys-8 and Cys-22. The disulfide bonds are essential for stabilizing the molecule and they may also impair the oligosaccharide binding. These domains are linked by a sequence of 4 amino acids: Gln-Pro-Ser-Asn [10].

Each domain has two helical turns and five β-strands. Domain A has a carbon and nitrogen terminal, which are lacking in Domain B, making Domain A possibly more flexible and influencing its binding properties [12]. Consequently, regions of higher and lower affinity can be found within the domains, which may generate variable affinities at the binding sites. Fig. 1 provides a visualization of the Cyanovirin-N binding site and the main residues involved [11]. The antiviral activity of lectins is based on their affinity for mannose-rich glycans, such as the glycoproteins gp120 [13] and gp41 found on the surface of the HIV envelope, which play a crucial role in viral evasion and cell entry.

Cyanovirin-N interacts with these carbohydrates to inhibit viral fusion and block the connection between the pathogen and the host cell. It has already shown clinical potential as a topical microbicide [14], exhibiting interesting pharmaceutical properties such as low toxicity to the human body and resistance to denaturing processes, organic solvents, detergents, and temperature variations [10]. Heterologous expression systems have been developed to produce Cyanovirin-N in various organisms, including fungi, *Escherichia coli*, and even plants [15]. Its application in plants has gained prominence due to the use of genetically engineered soybeans to produce recombinant CVN, which has also shown antiviral properties, indicating that homologous sequences could also play a similar role [16].

Given the biotechnological potential of cyanobacterial lectins, this study focuses on evaluating the pharmacological potential of cyanobacterial metabolites that share the homologous domain of Cyanovirin-N (CVNH, Pfam: PF08881) through carbohydrate affinity analysis. Bioinformatic tools were employed to analyze the structural and functional aspects of these homologous proteins, as they could potentially exhibit antiviral properties based on their degree of similarity and conservation of the three-dimensional structure, which may offer advantages such as lower toxicity and more cost-effective production rates for broad-scale applications. With that in mind, this study aims to evaluate the existence of antiviral activity, by researching the difference between the binding specificities displayed by different members of the CVNH family.

2. Results

2.1. Genomic prospecting

Upon genomic prospection, 33 sequences from cyanobacterial strains were obtained, distributed amog several genera, with higher frequency of *Nostoc, Scytonema, Tolypothrix, Microcystis, and Calothrix*. The models showed variable size, ranging from 52 to 230 amino acids, which illustrates the wide distribution and diversity of Cyanovirin-N Homologs. The sequences were subjected to a filtering process for quality assessment and size separation. The initial characterization of the sequences is described in Table S1(See supplementary material), pointing structural features like protein length, identity with the template (PDB Code: 1IIY), and presence of signal peptide.

2.2. Molecular modeling

The 33 Cyanovirin-N homologous sequences proceeded to the molecular modeling step, with the construction of tridimensional models of the sequences. The Root Mean Square Deviation (RMSD) indices showed low structural assimetry, with RMSD values ranging from 0.10 to 0.20, which means the tridimensional structures are closely related to the template. In this work, five validation parameters were used to assess the models' quality and verify their stability, shown in Table S2.

An alignment on Espript3 (Fig. 2) was conducted to verify the conservation of essential elements in the lectin, such as the disulfide bonds – formed between 4 cysteines, without which the lectin loses its function. In Cyanovirin-N, they are found at positions Cys-8, 22, 58, and 73. The mannose-binding site is represented by the residues Lys-3, Gln-6, Thr-7, Glu-23, Thr-25, Glu-41, Asn-42, Asp-44, Ser-52, Asn-53, Glu-56, Thr-57, Lys-74, Arg-76, Gln-78.

2.3. Docking

After validating the models, they were subjected to docking. The sequences obtained favorable results in terms of potential carbohydrate-binding affinity. The docking step served a dual purpose: predicting the optimal coordinates for ligand anchoring and screening the sequences to identify the most promising candidates for further Molecular Dynamics (MD) simulations. The Cyanovirin structure and its homologs possess two binding sites that are symmetrically opposite, with the reference ligand being the disaccharide Man-α (1–2)Man. Although the constructed models were structurally related to the template, there was some degree of variation in the location of the mannose docking site residues. Fig. S1 provides a visualization of the two di-mannoses docked onto the Cyanovirin template (1IIY). Docking results of the sequences are displayed on Table S3, ranked by Mol Dock Score. The best result on the docking step was *Microcystis novacekii* (Identifier: TRU87446.1). Table S4 highlights an Energy Overview Parameters for this model and

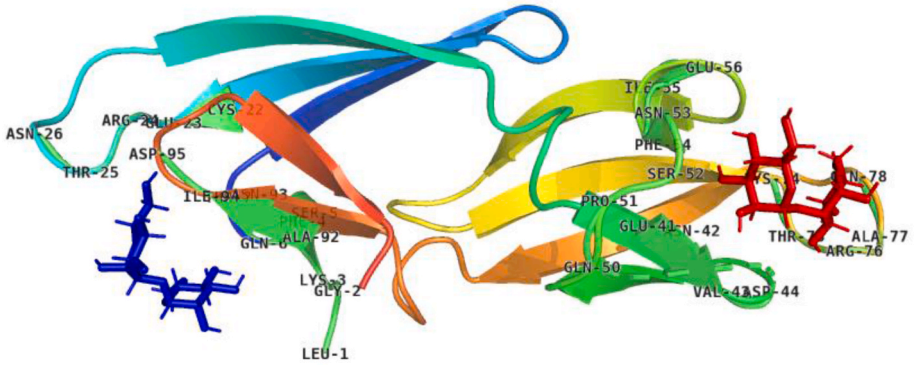

Fig. 1. Cyanovirin-N active sites A and B, representing the residues that interact directly with the Di-Mannoses.

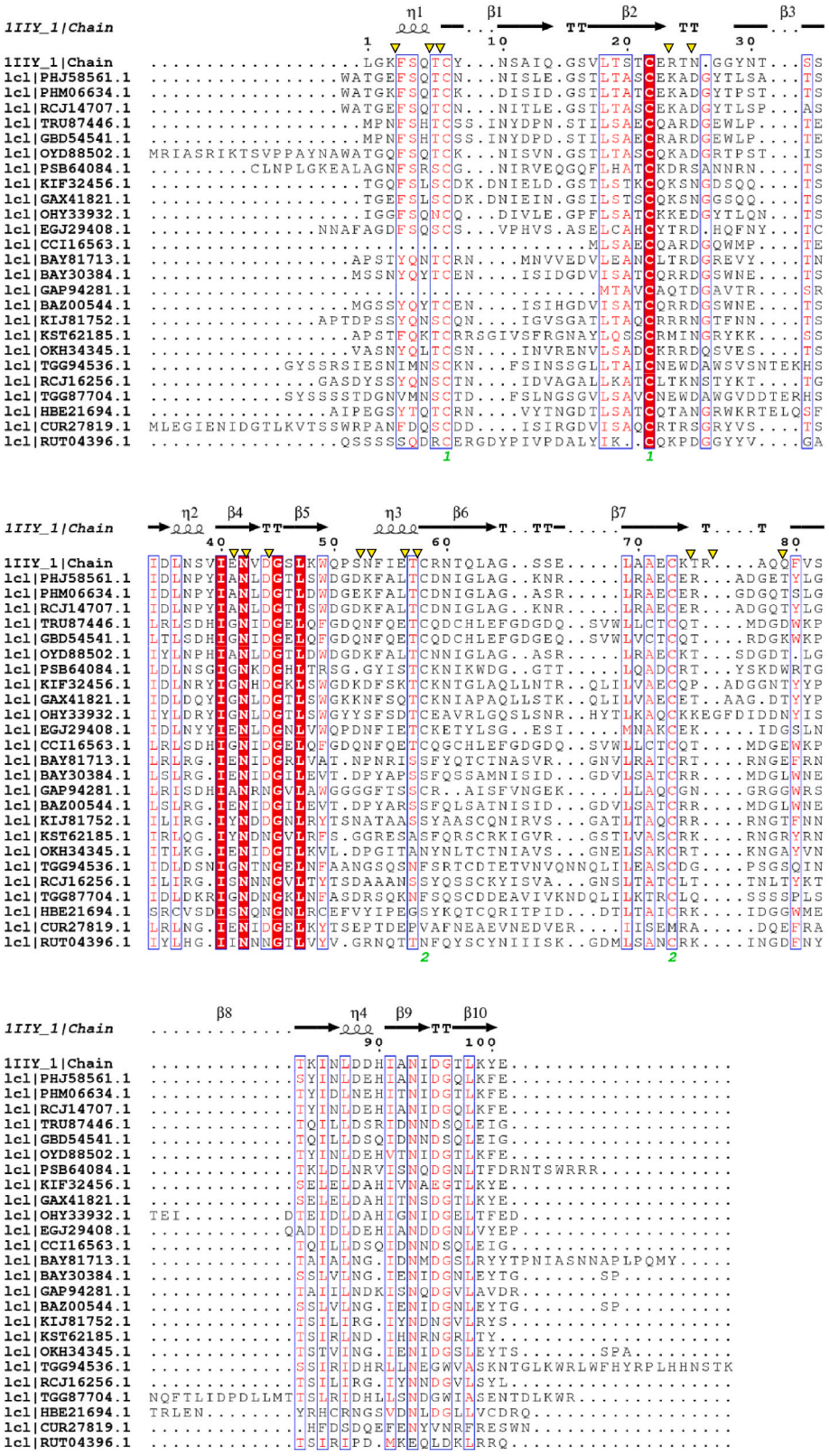

Fig. 2. Alignment of Cyanovirin-N (1IIY) and the homolog sequences on Espript3 software. The binding site residues of Cyanovirin are marked with yellow triangles. The red and blue boxes represent the sequences identity and similarity, respectively.

the ligand interactions for the best complexes at the docking step are shown in Figs. S2–S4.

2.4. Molecular dynamics

MD simulations involved producing 100ns for the models generated during the docking step. The stability of the models was evaluated using

RMSD, RMSF and energetic evaluations. In this study, it was utilized to calculate the average deviation between corresponding atoms of the protein and the ligand, thereby measuring their fluctuation. It also enabled examination of the carbohydrate-binding configuration to the protein model, providing insights into its stability over time to investigate the dynamics and stability of the protein-ligand complexes throughout the simulation and capturing the interaction of the enzymes with both binding sites of the complex.

A cross-reference analysis from the results of the three fields was implemented to evaluate and select the most promising models in terms of stability and ligand affinity through the 100ns of simulation at a temperature 300 K. The results of the binding free energy analysis were used as a parameter to filter the models with higher stability, since the lower the binding free energy resulting from the complex at the end of molecular dynamics, the higher the rate of well-formed and stable intermolecular bonds. The MD simulations highlighted significant differences in carbohydrate affinity among the binding sites, with values varying between −10 and −28 kcal/mol for Mannose Binding Site A and −2 to −26 kcal/mol for Site B. At the end of simulations, 6 systems were selected, through binding free energy and structural analysis, displaying as promising lectin models that could share antiviral properties of Cyanovirin-N with even higher carbohydrate affinity and less cytotoxicity. The Binding Free Energy results of the top systems are shown at Table 1, with the graph presenting being the average values obtained from the systems in triplicate. The RMSD graphs of the 3 best models are shown in Figs. 3–5 and RMSF graphs of the top 2 are shown in Figs. 7–8.

Most of the Dimmanose interactions that hold it together after lectin attachment are composed of hydrogen bonds that lies in the usual range of 1.5–2.5 Å, as shown on Figs. S5–S6, that present a picture of the ligands position and distances to the receptor after MD simulations. One of the parameters evaluated for a successful simulation was the correct attachment of the receptor and ligand since a receptor with low affinity might not hold it tightly. On Fig. S7 it is presented an illustration of *Nostoc calcicola* after simulation, with the ligands binded.

2.5. Energy decomposition

In addition to the Binding Free Energy calculation, the energetic analysis also encompasses the evaluation of individual residue contributions to binding of the complex. This analysis is crucial because certain residues may hinder interaction with the ligand. This study aims to identify the key residues within the complex that play a significant role and those that have minimal impact. Such information can guide further investigations, including in silico point mutations to remove or enhance specific residues. To examine whether the energetic contribution of individual residues in the binding site could potentially interfere with carbohydrate attachment and destabilize the systems, sequences representing the extreme ends of the Binding Free Energy results were selected. The Energy Decomposition analysis, depicted in Figs. 6–9 provides insights into the individual energetic contributions of residues towards mannose binding among models.

3. Discussion

3.1. Structural characterization

Cyanovirin-N (CVN), the first antiviral lectin identified in cyanobacteria, demonstrates considerable diversity within its taxonomic group, that can be seen by the variation between size and binding site residues among its homologs. However, the dissulfide bonds that maintain the lectin integrity remain conserved in between the group, and also the catalytic residues involved in Man-α (1–2) Man attachment. Alongside that, the CVN also shares the conformation of a higher affinity site and a lower affinity, present in various of the models tested. Most systems present a slightly greater binding affinity for ligand site A, in accordance with literature findings over cyanobacterial lectins. However, some of homologs analyzed here stand out for a different conformation, like *Microcystis aeruginosa NIES-298*(Identifier: GBD54541.1) and *Nostoc calcicola* (Identifier: OKH34345.1) that have high indices of ligand affinity on both binding sites, suggesting a greater efficiency at carbohydrate attachment than the original cyanovirin-N and, therefore, potential for higher antiviral affinity with less citotoxicity.

Besides, although binding site A is usually referred to as of higher affinity, some of the models displayed strong results for Mannose B ligand site instead, such as *Planktothrix rubescens* (Identifier: CUR27819.1) and *Microcystis aeruginosa PCC 9807*(Identifier: CCI16563.1), which is an example of the diversity among the group of CVN homologs and an indicator of multiple ligand Affinity, that has yet to be studied in order to unveil its possibilities.

Even among homologous sequences, a certain degree of variability is anticipated. Properties like molecular weight and adherence to the template exhibit heterogeneity. Nonetheless, all surveyed sequences were found to harbor a signal peptide responsible for tagging proteins for extracellular transport. The prevalence of this signal peptide in most sequences suggests potential involvement in extracellular processes, such as pathogen recognition or cell-to-cell interactions in cyanobacteria. These interactions are likely associated with colony formation, a prevalent feature in cyanobacteria [19]. Supporting evidence comes

Table 1
Binding free energy results of MM-GBSA and MM-PBSA calculations of higher affinity cyanovirin homologs considering both ligand sites.

Binding Site A

		MM-GBSA			MM-PBSA		
Organism	Identifier	Average	Std. Dev.	Err. of Mean	Average	Std. Dev.	Err. of Mean
Microcystis aeruginosa NIES-298	GBD54541.1	−34.81	3.47	0.22	−33.56	3.68	0.22
Nostoc calcicola	OKH34345.1	−26.48	2.76	0.27	−25.37	3.48	0.35
Nostoc sp. *membranacea cyanobiont*	OYD88502.1	−21.99	3.77	0.23	−17.39	4.45	0.32
Cyanobacteria bacterium	HBE21694.1	−21.14	3.93	0.26	−17.56	4.30	0.29
Planktothrix rubescens	CUR27819.1	−14.65	1.66	0.06	−14.25	1.82	0.07
Microcystis aeruginosa PCC 9807	CCI16563.1	−13.74	3.95	0.22	−13.97	3.83	0.22

Binding Site B

		MM-GBSA			**MM-PBSA**		
Organism	**Identifier**	**Average**	**Std. Dev.**	**Err. of Mean**	**Average**	**Std. Dev.**	**Err. of Mean**
Microcystis aeruginosa NIES-298	GBD54541.1	−31.18	3.16	0.10	−30.24	3.56	0.11
Nostoc calcicola	OKH34345.1	−27.53	2.66	0.27	−25.80	3.31	0.33
Nostoc sp. *membranacea cyanobiont*	OYD88502.1	−8.06	2.66	0.16	−8.64	2.73	0.17
Cyanobacteria bacterium	HBE21694.1	−5.89	1.75	0.07	−5.39	2.33	0.08
Planktothrix rubescens	CUR27819.1	−29.89	3.43	0.23	−27.35	3.52	0.24
Microcystis aeruginosa PCC 9807	CCI16563.1	−25.01	3.18	0.18	−23.27	3.60	0.21

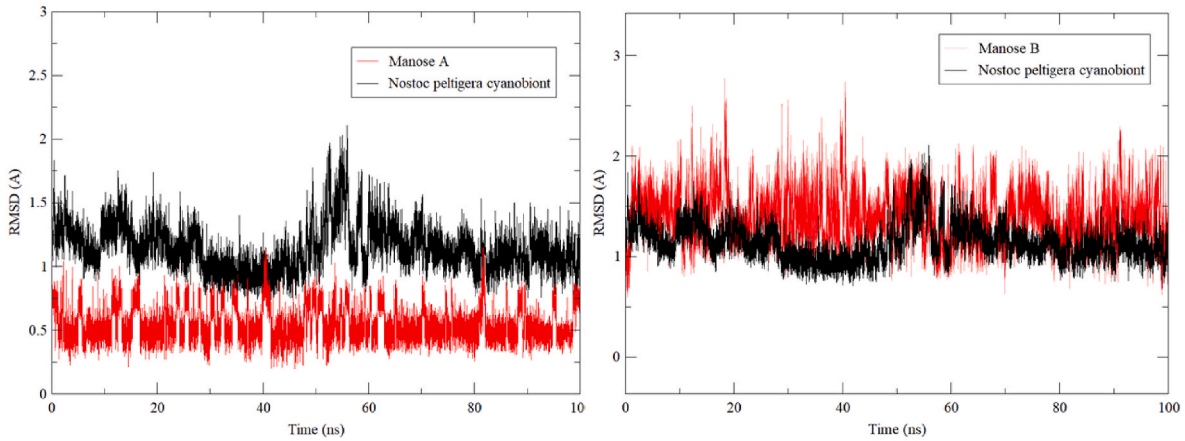

Fig. 3. RMSD graph of the interaction of *Nostoc* sp. *Peltigera membranacea cyanobiont* (Identifier: OYD88502.1) with ligands A and B, over the 100 ns of molecular dynamics simulation.

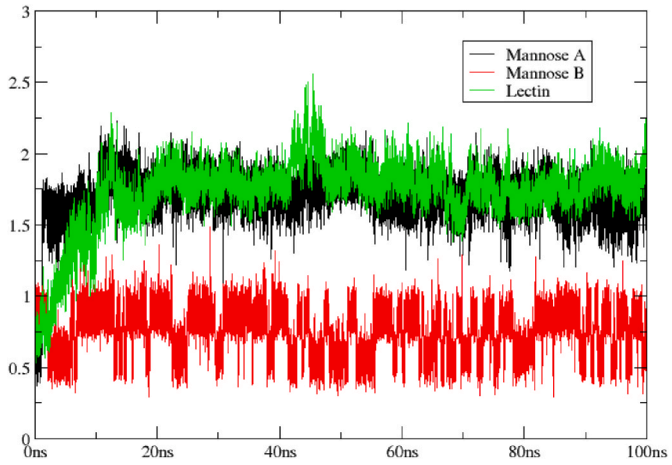

Fig. 4. RMSD graph of the interaction of *Microcystis aeruginosa NIES-298* with ligands A and B, over the 100 ns of molecular dynamics simulation.

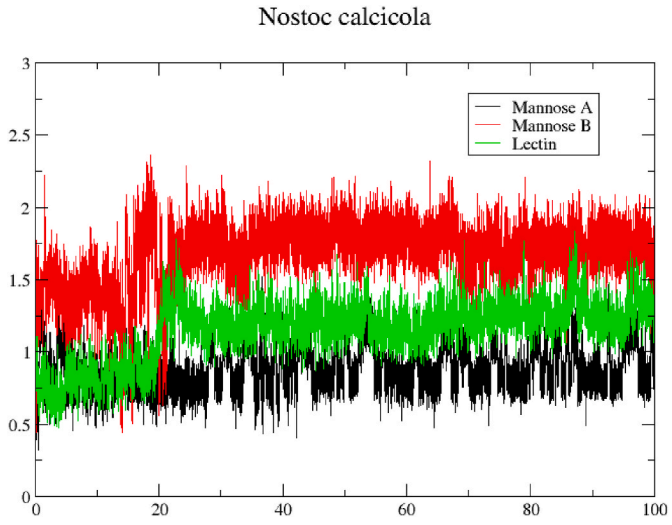

Fig. 5. RMSD graph of the interaction of *Nostoc calcicola* with ligands A and B, over the 100 ns of molecular dynamics simulation.

from other lectins, such as microvirin, which exhibits similar molecular weight and shares 33 % identity. Microvirin has been demonstrated to specifically recognize carbohydrates that differentiate and mediate cell attachment within colonies [20]. Although the original physiological role of lectins in cyanobacteria's survival and metabolism remains incompletely understood, it is conceivable that this function has been conserved within the family, offering new insights into lectin-carbohydrate binding on a macromolecular scale.

To assess the functionality of cyanovirin homologs, a structural analysis was conducted, generating three-dimensional models of CVNH (Cyanovirin Homologs) sequences complexed with their ligands. These models were constructed through a comparative modeling approach based on homology. Sequence alignments revealed positive identity values (Id), with the majority exceeding 30 %, surpassing the minimum expected rate for homologous sequences. Other modeling assessment parameters demonstrated favorable values, with most Qmean values falling between −2 and 0, indicating satisfactory local model quality. The Ramachandran analysis, measuring stereochemical quality, yielded positive results, with over 90 % of residues located within acceptable regions. Additionally, the sequences displayed low MolProbity [17,18] scores.

Upon analyzing the three-dimensional structure of the models, it becomes apparent that most folds are conserved and visually resemble cyanovirin. However, there is significant variability in mannose affinity among them, a feature noted in previous studies [21]. Molecular modeling facilitated the evaluation of crucial parameters, including the conservation of essential components for structural stability. Notably, the alignment using Espript3 revealed the conservation of some crucial cysteine residues that form disulfide bridges, vital for maintaining molecule stability.

The RMSD graphs indicate positive outcome from the systems, with variation of 1 Å at the most, and so does the RMSF Figures, that point to slight fluctuations on the protein structure over the 100ns of simulation. This suggests the possibility that CVN homologous sequences may exhibit affinity to other carbohydrates beyond mannose, either preferentially or simultaneously.

3.2. *Differences in carbohydrate affinity*

The assessment of Binding Free Energy provided valuable insights into the distribution of optimal models between ligands A and B within the family of CVN homologs. This distribution reveals varying affinities to glycans among homologs, suggesting the potential for in silico refinement of the complex interaction. Decomposition graphics identify candidate residues for in silico improvement, highlighting specific residues amenable to direct point mutations for enhanced carbohydrate

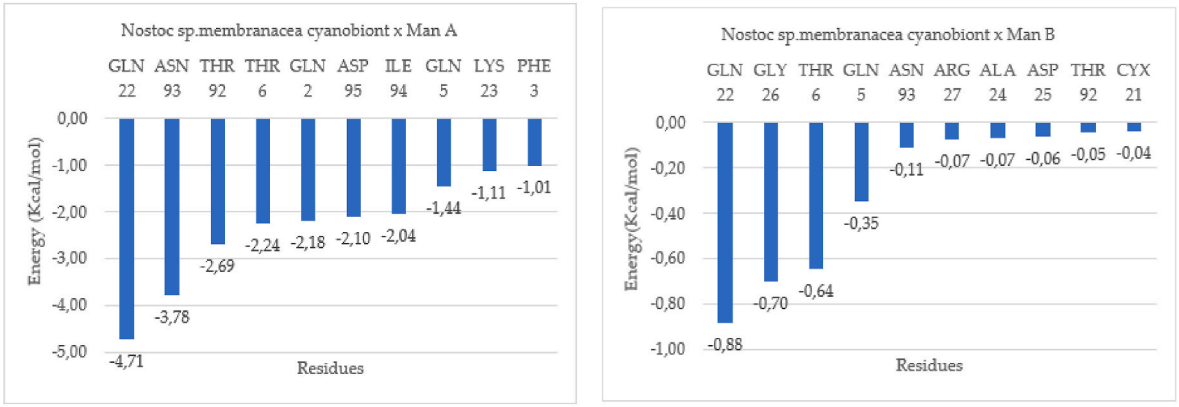

Fig. 6. Individual Residue Energy Decomposition from *Nostoc* sp. *membranacea cyanobiont 213* (Identifier: OYD88502.1) with the two ligands, separately.

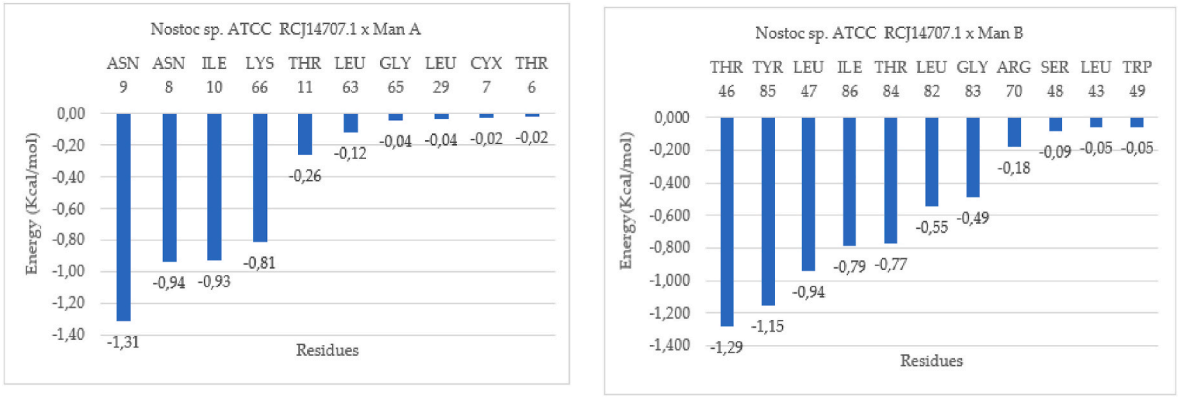

Fig. 7. Individual Residue Energy Decomposition from *Nostoc* sp. *ATCC RCJ14707.1 with* the two ligands.

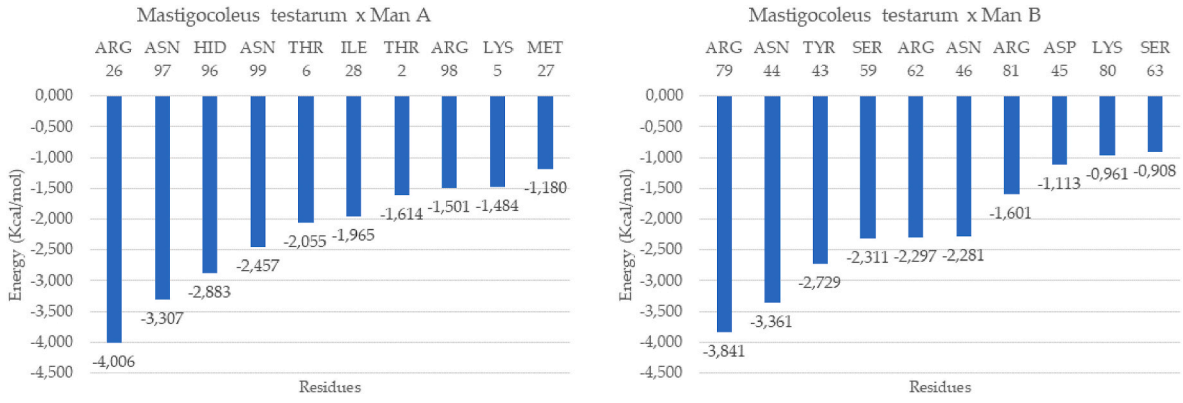

Fig. 8. Individual Residue Energy Decomposition from *Mastigocoleus testarum* with the two ligands, respectively.

affinity in the lectin, thereby improving viral inhibition. It is noteworthy that mutations in binding site residues of cyanovirin-like sequences typically compromise ligand affinity, indicating an evolutionarily selected optimal configuration [21]. Domain A of Cyanovirin-N typically possesses the low-affinity binding site, resulting in less efficient binding rates [22].

However, in Table 1 it is possible to see that several sequences from both domains A and B achieved good results, and domain A obtained better outcomes on average, when analyzing both free energy and RMSD results.

Differential binding affinity rates among both domains could offer advantages by incorporating low-affinity binding sites, granting supplementary binding specificity to the protein. This flexibility allows the identification and attachment of different ligands with diverse affinities, potentially useful given the rapid evolution or alterations in viral molecules. Therefore, low-affinity binding sites may serve as a protective mechanism, enabling the protein to adjust its binding interaction effectively. The results of Binding Free Energy exhibited a significant correlation with the deviation indices of the structures. RMSD graphs demonstrated overall stability in most systems during MD simulations, with some local fluctuations observed among ligands within the same systems. Sequences with poor binding energy correlated with higher RMSD, indicating lower affinity to the ligand, impacting the stability of the entire complex.

CVN homologs demonstrated differential affinity for the two carbohydrate binding sites, with binding site B displaying superior results.

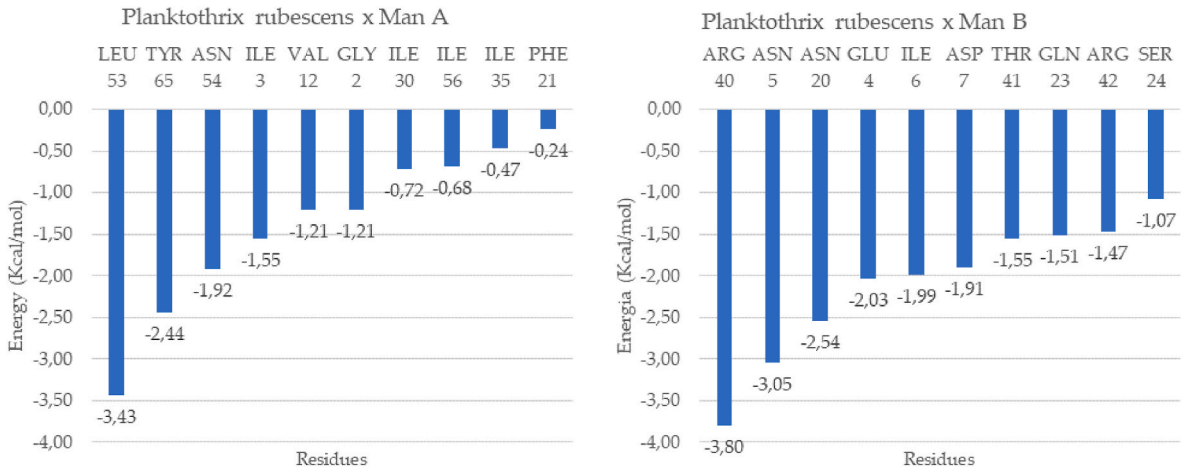

Fig. 9. Individual Residue Energy Decomposition for *Planktothrix rubescens* and the two ligands.

Individual residue decomposition graphics identified residues potentially impairing ligand attachment, supporting the hypothesis that lectin carbohydrate binding could be enhanced by introducing point mutations to remove residues negatively affecting the protein-receptor connection. Despite observed variability, several sequences exhibited affinity for dimannose, forming stable complexes during the 100ns MD simulation. Carbohydrate affinity appears conserved among homologous sequences of Cyanovirin-N, albeit with divergences not only between species but also between binding site domains. The lectin's pseudo-domain B consistently produced the best outcomes, indicating that while domain A has lower affinity, it likely serves a function adapted for multiple ligand binding.

Binding Free Energy values varied, with most enzyme-substrate complexes yielding results in the range of −20 to −30 kcal/mol, indicative of greater complex stability. Based on energetic analysis results and considering satisfactory outcomes for both ligands and stability during RMSD and RMSF calculations, the models from *Microcystis aeruginosa NIES-298* and *Nostoc calcícola* presented as the most promising lectins, standing out as a promising alternative to Cyanovirin-N in the field of antiviral inhibitors.

4. Materials and methods

4.1. Genomic prospection

This work used an *in-silico* approach, through bioinformatics tools, to identify and characterize cyanobacterial lectins' biotechnological potential. Initially, a genomic prospection was conducted by applying a search strategy based on domain conservation. This was made by searching for lectin coding sequences in cyanobacterial genomes available in public databases, like the National Center for Biotechnology Information (NCBI) [19]. The search was performed using a conserved domain identification approach. First, the amino acid sequences annotated as lectins and whose structure had been already experimentally resolved were obtained from NCBI Ref Seq [20]. After that, the CD-HIT tool [21] was used to group sequences with more than 90 % identity, creating a non-redundant dataset. This dataset was submitted to NCBI's Conserved Domain Database [22] to identify all the domains present. The result was used for lectin identification. All the coding sequences (CDS) annotated in cyanobacterial genomes used in this study were extracted and translated with Geneious R9 software [23], using the bacterial genetic code. Next, for each genome, the translated CDS were subjected to CDD. A script developed in Perl analyzed the result of each genome in the CDD and returned only the sequences with the same domains present in the NCBI dataset. Those sequences formed the initial dataset, that proceeded to the molecular modeling step.

4.2. Molecular modeling

For the construction of the tridimensional models, the chosen approach was homology modeling. The chosen template was the original Cyanovirin-N (PDB Code: 1IIY). Multiple sequence alignments were performed on Promals3D program [24] and ClustalW [25]. The sequences that showed reasonable identity values with the template deposited in the Protein Data Bank [26] were modeled using the Modeller 10.4 [27] software. One hundred models were generated for each sequence, considering the different possible spatial conformations, and choosing the best model. The models were evaluated on Modeller according to DOPE score, satisfying spatial constraints such as bond lengths, angles, and non-bonded interactions, in addition to the use of automatic loop refinement. After building the models, the stereochemical quality was evaluated using the Ramachandran, generated by the MolProbity server [17,18]. The folding quality of the model was analyzed using the Verify3D program [28] and the local quality of the model was measured using Qmean [29].

4.3. Docking

The Molecular Docking step is focused on predicting the optimal orientation of the Mannose ligands binding to the Lectins and is responsible to anchor the receptor and ligand structures, determining the interaction coordinates and evaluating the formed complex energetically, to verify the stability of the interaction. Molecular Docking was made on the Molegro Virtual Docking 5.5 software [30], by assessing potential ligand binding configurations to predict the most energetically favorable arrangement. For the docking parameters, it was used the grid resolution value of 0.30. The assessment of the docking runs was performed using the MolDock scoring function [31] through a piecewise linear potential (PLP) that considers the directionality of the grid. The best solutions were then re-evaluated to increase the efficiency of the process using the Re-Rank scoring function, which considers a sp2-sp2 twist term and a Lennard-Jones potential [32]. The ligand-protein interaction graphs were made using LigPlot + v.2.2.8 [33]. The following steps were the energy minimization and optimization of hydrogen bonds. Each procedure had 10 rounds of calculation, with a population of 100 and maximum iterations of 2000. The best complexes, selected through MolDock and ReRank, were selected for the MD simulatons.

4.4. Molecular dynamics

Unlike Docking, molecular dynamics simulations are used to model the dynamic behavior of atoms and molecules over time, exploring the

temporal evolution of molecular systems and analysing structural aspects, such as protein folding and conformational changes on a non-static system, aiming for a comprehensive understanding of the molecular interactions involved in the lectins' carbohydrate interactions, considering important factors such as temperature, force fields, chemical bonds, solvation, among others. The simulations were carried out to validate the models generated, as well as to evaluate the affinity of lectins that had their ligands predicted, on a dynamic process, to mimic a biological system in all their complexity.

For the assembling, preparation, solvation of the model systems, as well as the energy minimization, heating, and MD steps, the Amber 23 software [34,35] package tools were used, using the force fields Glycam06_j-1 [36] for the carbohydrates; Gaff [37] for the ions; and ff14SB [38] for the proteins. In order to proceed to MD, the models underwent stages of preparation, starting by protonation, that was made with H++ server [39,40] considering pH value of 7.0 and salinity of 0.15, alongside default parameters for protein protonation. The objective of this phase was to assess the molecular chemical surroundings. This assessment ensured the accurate determination of protein residue protonation states that are crucial for precisely calculating both intra and intermolecular interactions. Additionally, it aimed to simulate the pH influence within the system.

Ligand preparation and system assembly were performed using Amber modules Antechamber, Parmchk2 and Tleap, respectively. Solvation was performed using a 10 Å-sized TIP3P water box model in each direction from the tip of the protein. Na + or Cl-ions were added to the system to neutralize the excess charges to stabilize the simulation. After this, topology and coordinate files of the complex were generated and brought together to obtain the final complex file. After the system assembly, the lectin models proceeded to the energy minimization step using the Sander software, a component of the Amber 23 package tools. When water is added to a system in a vacuum, previously non-existing interactions are created that could destabilize the protein. For this reason, the energy minimization step was used to decrease the total energy of the solvated system, causing water and protein to stabilize their interactions. The minimization occurred in 5 steps, where in the initial 4 steps the heavy atoms had their movement restricted by a harmonic potential of 1000 kcal/mol*Å2 while in the final step all atoms were free. Altogether, 3000 steepest descent cycles and 5000 conjugate gradient cycles were performed during the first 4 stages, while in the last stage, 5000 steepest descent cycles and 20,000 conjugate gradient cycles were performed.

For the heating phase, the same principle of energy minimization was used, but in a greater number of steps, totalizing 14, with a lower harmonic potential restricting the heavy atoms, 25 kcal/mol*Å2, which was reduced to 0 at the end. In the first 13 steps, the temperature was gradually increased until reaching 300 K in an NVT ensemble, for a total of 650 ps. In the last step, 2ns of simulation were produced in an NPT ensemble to balance the system. The SHAKE algorithm was used to constrain the bond length of all hydrogen atoms. The Particle Mesh Ewald method was used to calculate the electrostatic interactions using a cutoff value of 10.0 Å for unbound interactions. Each system underwent 3 runs of 100 nanosseconds of MD simulation at 300 K in order to eliminate bias variations. structural analysis of the proteins with their respective ligands was made through RMSD (Root Mean Square Deviation) and RMSF (Root Mean Square Fluctuation) calculated from the N, CA, and C atoms of the main chain by the cpptraj module and for the free energy calculations.

4.5. Binding free energy

To evaluate the behavior and stability of the biological systems processed during the MD simulations, an energetic evaluation approach was employed, following 2 methodologies: Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) and Molecular Mechanics-Poisson-Boltzmann Surface Area (MM-PBSA), to calculate the Binding Free Energy, a thermodynamic approach to assess the affinity between molecules. These methods are considered "end-state" techniques, assessing the energy difference between the two final states of the system: bound (on) and unbound (off). The MM-PBSA, on the other hand, a more accurate approach to describe solute-solvent interaction. It solves the Poisson-Boltzmann equation to calculate the electrostatic potential around the solute.

Binding Free Energy for each frame of the simulation was measured, and its variation was calculated to indicate the temporal evolution of the free energy using the last 10 ns of the simulation trajectory (5000 frames) of MD simulations. This specific timeframe was selected to ensure the systems reached a stable state. Both methods involve calculating the molecular mechanics energy, which includes the binding energy, torsional energy, van der Waals energy, and electrostatic energy. The MM-GBSA is used to account for solvent effects, as a simplified approach to consider solvation, taking into account atom polarity and accessible surface area. This analysis was performed through Amber 23 MMPBSA.py script adapted for dual ligand evaluation.

Additionally, an assessment of individual residue energy decomposition was conducted, to investigate the residues' contribution for the ligand attachment. The residue decomposition graphs can help to identify residues that may be negatively interfering with the interaction, establishing targets for site-directed changes. Such mutations could be useful for improving the lectins' viral activity through point mutations in sílico.

5. Conclusions

One of the greatest challenges of bioinformatic approaches is the efficiency in mimetizing the complexity of a biological system in silico, capturing all molecular interactions and events that would happen in vivo to accurately make a biological inference. Several elements could bias the efficiency of this type of prediction, like the lenght of MD simulations, that are limited by the available computational resources. Also, interpretation of the energetic analyses may sometimes be subjective, what could impact the influence of conclusions. Given this framework, the results of this study altogether support the thesis that antiviral properties could be conserved among homolog sequences of Cyanovirin-N, what could be of great of importance in guiding future research, perhaps applying it in vitro and open doors to the possibility of enhancing lectin's affinity through point mutations and produce it by heterologous expression. Structural and functional prediction of biological molecules is a growing field of scientific research that has yet to be further explored. Improving in silico prospection by creating an efficient pipeline for harvesting biologically active molecules from cyanobacteria in large scale can be of great value for future research. This study aimed to investigate whether different homologous sequences of CVN share carbohydrate affinity and possess antiviral activity. The findings revealed significant variation in ligand affinity, which are most likely attributed to structural changes in the binding site across the sequences. These results provide insights into new research possibilities in the viral inhibition mechanism and to harnessing the genetic diversity of cyanobacterial strains to enhance HIV therapies. Based on the results obtained, the study highlights the genetic variability among Cyanovirin-N homologs, as evidenced by variations in size, molecular behavior, and carbohydrate affinity. This variability can pose both challenges, such as homologs with low affinity to the ligand, as well as advantages, given that the existence of numerous homologs enables a search for proteins with greater potential in rational drug design. Biotechnological products are currently gaining increasing prominence, driven by the abundance of species in nature and bioactive compounds with pharmacologically relevant antimicrobial properties. Cyanobacteria serve as an excellent source for discovering new compounds with biotechnological potential, highlighting their high capacity for producing biopharmaceuticals. Molecular dynamic simulations, along with screening and molecular docking methodologies, have proved to be

valuable assets in the research and development of novel bioproducts. Cyanovirin-N, which exhibits significant diversity within the cyanobacterial group, possesses binding sites that show substantial variation in conservation, indicating a potential affinity for different carbohydrates. This supports the hypothesis that it may have multiple ligands within the cyanobacterial group. Previous research has already highlighted the antiviral activity of Cyanovirin-N, and the analysis of its homologous proteins can serve as a crucial element in discovering new potential drugs with reduced toxicity, improved cost-benefit ratio, and sustainability in their acquisition and utilization.

Funding

This research was funded by Fundação Amazônia de Amparo a Estudos e Pesquisas do Pará (ICAAF 099/2014) | Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), grant number 322686/2015-0 (ECG)

CRediT authorship contribution statement

Gabriel Xavier: Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Alenna Crystiene Lima Farias de Sousa:** Validation. **Larissa Queiroz dos Santos:** Validation. **Délia Aguiar:** Project administration. **Evonnildo Gonçalves:** Methodology. **Andrei Santos Siqueira:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Gabriel Albuquerque Xavier reports financial support was provided by CNPQ. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

Molecular graphics and analyses were performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmgm.2024.108718.

References

[1] R.M.M. Abed, S. Dobretsov, K. Sudesh, Applications of cyanobacteria in biotechnology, J. Appl. Microbiol. 106 (2009) 1–12, https://doi.org/10.1111/J.1365-2672.2008.03918.X.

[2] S. Mazard, A. Penesyan, M. Ostrowski, I.T. Paulsen, S. Egan, Tiny microbes with a big impact: the role of cyanobacteria and their metabolites in shaping our future, Mar. Drugs 14 (2016), https://doi.org/10.3390/MD14050097.

[3] R.B. Dixit, M.R. Suseela, Cyanobacteria: potential candidates for drug discovery, Antonie Leeuwenhoek 103 (2013) 947–961, https://doi.org/10.1007/S10482-013-9898-0, 2013 103:5.

[4] A.R. Garrison, B.G. Giomarelli, C.M. Lear-Rooney, C.J. Saucedo, S. Yellayi, L.R. H. Krumpe, et al., The cyanobacterial lectin scytovirin displays potent in vitro and in vivo activity against Zaire Ebola virus, Antivir. Res. 112 (2014) 1–7, https://doi.org/10.1016/J.ANTIVIRAL.2014.09.012.

[5] Gao X, Chen W, Guo C, Qian C, Liu G, Ge F, et al. Soluble cytoplasmic expression, rapid purification, and characterization of cyanovirin-N as a His-SUMO fusion. APPLIED GENETICS AND MOLECULAR BIOTECHNOLOGY n.d. https://doi.org/10.1007/s00253-009-2078-5.

[6] A.S. Siqueira, A.R.J. Lima, D.C.F. Aguiar, A.S. Santos, JL. da SG. Vianez Júnior, E. C. Gonçalves, Genomic screening of new putative antiviral lectins from Amazonian cyanobacteria based on a bioinformatics approach, Proteins: Struct., Funct., Bioinf. 86 (2018) 1047–1054, https://doi.org/10.1002/prot.25577.

[7] H. Lotfi, R. Sheervalilou, N. Zarghami, An update of the recombinant protein expression systems of Cyanovirin-N and challenges of preclinical development, Bioimpacts 8 (2018) 139–151, https://doi.org/10.15171/bi.2018.16.

[8] H. Lotfi, M.A. Hejazi, M.K. Heshmati, S.A. Mohammadi, N. Zarghami, Cellular and Molecular Biology Optimizing Expression of Antiviral Cyanovirin-N Homology Gene Using Response Surface Methodology and Protein Structure Prediction, 2017, https://doi.org/10.14715/cmb/2017.63.9.17.

[9] M.R. Boyd, K.R. Gustafson, J.B. McMahon, R.H. Shoemaker, B.R. O'Keefe, T. Mori, et al., Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development, Antimicrob. Agents Chemother. 41 (1997) 1521–1530, https://doi.org/10.1128/AAC.41.7.1521.

[10] H. Mazur-Marzec, M. Cegłowska, R. Konkel, K.P. Pyrć, Antiviral Cyanometabolites-A Review, 2021, https://doi.org/10.3390/biom11030474.

[11] E. Matei, R. Basu, W. Furey, J. Shi, C. Calnan, C. Aiken, et al., Structure and glycan binding of a new cyanovirin-N homolog, J. Biol. Chem. 291 (2016) 18967–18976, https://doi.org/10.1074/jbc.M116.740415.

[12] R. Percudani, B. Montanini, S. Ottonello, The anti-HIV cyanovirin-N domain is evolutionarily conserved and occurs as a protein module in eukaryotes, Proteins 60 (2005) 670–678, https://doi.org/10.1002/PROT.20543.

[13] I. Maier, R.H. Schiestl, G. Kontaxis, Cyanovirin-N binds viral envelope proteins at the low-affinity carbohydrate binding site without direct virus neutralization ability, Molecules 26 (2021), https://doi.org/10.3390/molecules26123621.

[14] D.M. Colleluori, D. Tien, F. Kang, T. Pagliei, R. Kuss, T. McCormick, et al., Expression, purification, and characterization of recombinant cyanovirin-N for vaginal anti-HIV microbicide development, Protein Expr. Purif. 39 (2005) 229–236, https://doi.org/10.1016/J.PEP.2004.10.009.

[15] A. Sexton, P.M. Drake, N. Mahmood, S.J. Harman, R.J. Shattock, J.K. Ma, Transgenic plant production of Cyanovirin-N, an HIV microbicide, FASEB (Fed. Am. Soc. Exp. Biol.) J. 20 (2006) 356–358, https://doi.org/10.1096/FJ.05-4742FJE. Official Publication of the Federation of American Societies for Experimental Biology.

[16] B.R. O'Keefe, A.M. Murad, G.R. Vianna, K. Ramessar, C.J. Saucedo, J. Wilson, et al., Engineering soya bean seeds as a scalable platform to produce cyanovirin-N, a non-ARV microbicide against HIV, Plant Biotechnol. J. 13 (2015) 884, https://doi.org/10.1111/PBI.12309.

[17] I.W. Davis, A. Leaver-Fay, V.B. Chen, J.N. Block, G.J. Kapral, X. Wang, et al., MolProbity: all-atom contacts and structure validation for proteins and nucleic acids, Nucleic Acids Res. 35 (2007) W375, https://doi.org/10.1093/NAR/GKM216.

[18] V.B. Chen, W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, et al., MolProbity: all-atom structure validation for macromolecular crystallography, Acta Crystallogr D Biol Crystallogr 66 (2010) 12–21, https://doi.org/10.1107/S0907444909042073.

[19] National Center for Biotechnology Information n.d. https://www.ncbi.nlm.nih.gov/(accessed December 17, 2023).

[20] K.D. Pruitt, T. Tatusova, D.R. Maglott, NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, Nucleic Acids Res. 35 (2007) D61–D65, https://doi.org/10.1093/nar/gkl842.

[21] W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences, Bioinformatics 22 (2006) 1658–1659, https://doi.org/10.1093/BIOINFORMATICS/BTL158.

[22] A. Marchler-Bauer, M.K. Derbyshire, N.R. Gonzales, S. Lu, F. Chitsaz, L.Y. Geer, et al., CDD: NCBI's conserved domain database, Nucleic Acids Res. 43 (2015) D222–D226, https://doi.org/10.1093/NAR/GKU1221.

[23] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, et al., Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, Bioinformatics 28 (2012) 1647, https://doi.org/10.1093/BIOINFORMATICS/BTS199.

[24] J. Pei, M. Tang, N.V. Grishin, PROMALS3D web server for accurate multiple protein sequence and structure alignments, Nucleic Acids Res. 36 (2008) W30, https://doi.org/10.1093/NAR/GKN322.

[25] J.D. Thompson, D.G. Higgins, T.J. Gibson, W. Clustal, Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673, https://doi.org/10.1093/NAR/22.22.4673.

[26] P. Rose, The RCSB protein data bank: integrative view of protein, gene and 3D structural information, Nucleic Acids Res. 45 (2017), https://doi.org/10.1093/NAR/GKW1000.

[27] B. Webb, A. Sali, Protein structure modeling with MODELLER, Methods Mol. Biol. 2199 (2021) 239–255, https://doi.org/10.1007/978-1-0716-0892-0_14.

[28] D. Eisenberg, R. Lüthy, Bowie Ju, VERIFY3D: assessment of protein models with three-dimensional profiles, Methods Enzymol. 277 (1997) 396–404, https://doi.org/10.1016/s0076-6879(97)77022-8.

[29] P. Benkert, S.C.E. Tosatto, D. Schomburg, QMEAN: a comprehensive scoring function for model quality assessment, Proteins 71 (2008) 261–277, https://doi.org/10.1002/PROT.21715.

[30] G. Bitencourt-Ferreira, W.F. de Azevedo, Molegro virtual docker for docking, Methods Mol. Biol. 2053 (2019) 149–167, https://doi.org/10.1007/978-1-4939-9752-7_10.
[31] R. Thomsen, M.H. Christensen, MolDock: a new technique for high-accuracy molecular docking, J. Med. Chem. 49 (2006) 3315–3321, https://doi.org/10.1021/JM051197E/SUPPL_FILE/JM051197ESI20060314_081922.PDF.
[32] J. Li, A. Fu, L. Zhang, An overview of scoring functions used for protein–ligand interactions in molecular docking, Interdiscipl. Sci. Comput. Life Sci. 11 (2019) 320–328, https://doi.org/10.1007/s12539-019-00327-w.
[33] A.C. Wallace, R.A. Laskowski, J.M. Thornton, LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions, Protein Eng. 8 (1995) 127–134, https://doi.org/10.1093/protein/8.2.127.
[34] D.A. Case, H.M. Aktulga, K. Belfon, D.S. Cerutti, G.A. Cisneros, V.W.D. Cruzeiro, et al., AmberTools. J Chem Inf Model 63 (2023) 6183–6191, https://doi.org/10.1021/acs.jcim.3c01153.
[35] R. Salomon-Ferrer, D.A. Case, R.C. Walker, An overview of the Amber biomolecular simulation package, WIREs Comput Mol Sci 3 (2013) 198–210, https://doi.org/10.1002/wcms.1121.
[36] M.B. Tessier, M.L. DeMarco, A.B. Yongye, R.J. Woods, Extension of the GLYCAM06 biomolecular force field to lipids, lipid bilayers and glycolipids, Mol. Simulat. 34 (2008) 349, https://doi.org/10.1080/08927020701710890.
[37] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J. Comput. Chem. 25 (2004) 1157–1174, https://doi.org/10.1002/JCC.20035.
[38] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB, J. Chem. Theor. Comput. 11 (2015) 3696–3713, https://doi.org/10.1021/ACS.JCTC.5B00255.
[39] H++ (web-based computational prediction of protonation states and pK of ionizable groups in macromolecules) n.d. http://newbiophysics.cs.vt.edu/H++/index.php (accessed December 17, 2023).
[40] J.C. Gordon, J.B. Myers, T. Folta, V. Shoja, L.S. Heath, A. Onufriev, H++: a server for estimating pKas and adding missing hydrogens to macromolecules, Nucleic Acids Res. 33 (2005) W368–W371, https://doi.org/10.1093/nar/gki464.