International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Exploring the Molecular Diversity of Lectins: Leveraging Bioinformatics for Gene Characterization and Phylogenetic Insights

Nivedita Patil[1], Dr. Soma Chaki[2]

[1]PhD Scholar, Faculty of Life and Allied Health Sciences, M. S Ramaiah University of Applied Sciences, Bengaluru, Karnataka, India
Email: nivedpatil5[at]gmail.com)

[2]Associate Dean, Faculty of Life and Allied Health Sciences, M. S Ramaiah University of Applied Sciences, Bengaluru , Karnataka, India
(somachaki14[at]gmail.com)

Abstract: ***Lectins are sugar - binding protein molecules, highly specific for the carbohydrate groups on glycoproteins, glycolipids, or polysaccharides and present in bacteria, viruses, plants and animals. They exhibit a high degree of affinity towards sugar moieties (1). Lectins are known to be heterogenous in their occurrence among different species, organs, and tissues, in addition to having structural and functional diversity. Most plant lectins are present as secretory proteins and are deposited in the vacuoles of a cell. For many years now, to understand plant lectins' functional specificity, one of the best - studied assayswas to agglutinate erythrocytes. Cancer treatments, biomedical applications, drug discovery are few areas where plant lectins have been found useful abundantly and are being constantly exploited in other fields of research as well. However, one of the less explored applications of plant lectins is the understanding at the molecular level. Although much of the literature information is available on the physiochemical aspect of the lectin protein, the gene structure and protein motifs studies are yet to be understood. An answer to this is the use of bioinformatics in establishing the phylogenetic territory of plant lectins, which will help demonstrate the enhanced role of lectins in each organism and further help widen the research application. The present review involves an understanding of bioinformatics tools' use to characterize lectin genes and build an association among the known lectins of plant, animal, and bacterial origin.***

Keywords: Plant lectins, Bioinformatics, Lectin Genes, Glyco - informatics

1. Introduction

Plants contain lectins in various parts, including seeds, leaves, bark, roots, tubers and fruits [2]. Many plant lectins linked to cell motility, cell - cell interaction, embryogenesis and organ formation have been identified as secretory proteins, meaning they enter the secretory system and accumulate in vacuoles, cell walls, or intercellular spaces [3 - 6]. Plants use lectins to protect themselves from insects and fungi and transport and store sugar [7]. One of the most studied protein groups is lectins due to their unique biological roles such as biological and medical applications, including the isolation of glycoconjugates from cells, microorganism recognition, monitoring of changes in carbohydrate expression on living cells, mitogenic simulation, anti - proliferative effects, anti - tumour, and drug targeting to the gastrointestinal tract [8 - 11].

The term “lectin” (derived from the Latin legere, which means to select, to pick out, to choose) was first introduced by Boyd and Shapleigh in 1954 to refer to the fact that lectins can recognize and bind specific carbohydrate structures [12]. However, this type of selectivity is not a general characteristic of plant lectins as reported in the literature over the past 100 years of research, and a steady progress has been made in isolating many different plant lectins. Most of the research is focused on legume lectins since they were primarily found in legume seeds and consequently, concanavalin A was the first lectin to be purified from jack bean seeds (*Canavalia ensiformis*) [15]. With different approaches and techniques, lectinology studies were extended from seeds to other plant tissues and have furthermore led to the discovery of new lectins owing to progress in purification techniques and a better understanding of carbohydrate - binding properties and a variety of molecular structures.

Plant lectin research revolved around seeds or vegetative storage tissues for a long time as the practical approach in extracting these lectins was known but in the past two decades, the research focus area has shifted to non - storage tissues such as leaves, roots and flowers. The change of research interestis due to the availability of advanced technologies and the discovery of novel lectins from various plants leading to a better understanding of protein interaction with sugar molecules.

The field of glycobiology provides insight into plant lectins' contributions, which are considered significant tools for understanding glycoconjugates on cell surfaces and in solution. Consequently, lectins are exciting tools to study protein - carbohydrate interactions playing an essential role in, e. g., host - pathogen interaction (s), development, cell–cell communication, and cell signalling. For example, concanavalin A is now one of the most used lectins for the characterization and purification of high - mannose N – glycan containing glycoconjugates and detecting these carbohydrate structures on biomolecules and cells [19 - 20].

Chemical and biological data on lectins, as in the case of any family of proteins, when interpreted through their tertiary structures, provide the most remarkable insight into their function and role in the biological system [21 - 22]. The vast amount of data on lectins, as in many other classes of proteins, makes it difficult and time - consuming to

simultaneously comprehend all the various advances, mainly because the sources of these various pieces of data are highly diverse as also the areas of specialization they emerge from. To effectively use all the data toward understanding the function and for any possible application, organising these seemingly independent data into a common framework is essential. It can be made possible by collating and integrating these different data using an appropriate framework. Carbohydrate - binding motifs sequence studies available through bioinformatics tools reveal that these are widespread in different organisms, including plants, animals, fungi, and bacteria. But the same database also reveals that few other motif sequences are restricted to few plant families only.

Bioinformatics allows us to use several programs to understand the gene structure and explore its applications. The current paper highlights few of the standard bioinformatics tools used to uncover a detailed assessment of plant lectin protein structure and functions.

Plant Lectin Classification:

Lectins are a group of heterogenous proteins was established quite early in the lectinology research [23, 24], and the focus area of classifying them was mainly based on their carbohydrate - binding nature such as mannose - binding lectins, glucose binding lectins, galactose binding lectins and so on. Although this classification was useful for short period, difficulties were faced to categorize lectins based on molecular and taxonomical relationships among plant species. More sequence - based lectins were isolated and classified accordingly with the advances in techniques available. This not only helps in better understanding of lectin's sugar - binding ability but also helps establish a evolutionary - based comprehensive analysis of lectins.

Relevant data from plant genome and transcriptome analyses show that plant lectins can be categorized into 12 families depending on their carbohydrate recognition domains (CRDs) and the lectin polypeptide sequence. In principle, each lectin family comprises all proteins with a domain that is evolutionarily related to sequence similarity to the characteristic CRD. Although the classification system proves to be useful, it needs to be updated periodically depending on the novel lectins detected in newer plant species.

Table 1:
Overview of the 12 plant lectin domains [25 - 27]

Lectin domain	Protein characteristics/ 3D structure	Specificity	Localization in plant cell
Agaricus bisporus homolog	Homodimer β-Sandwich	T antigen *N*-glycans	Nucleus, cytoplasm
Amaranthin domain	Homodimer β-Trefoil	GalNAc T antigen	Nucleus, cytoplasm
Homolog of class V chitinases	Homodimer TIM barrel	High-mannose *N*-glycans	Vacuole
Cyanovirin domain	Homodimer Triple-stranded β-sheet and a β-hairpin	High-mannose *N*-glycans	Vacuole
Euonymus europaeus lectin domain	Homodimer Structure unknown	Galactosides, high-mannose *N*-glycans	Nucleus, cytoplasm Vacuole?
Galanthus nivalis agglutinin domain	Different oligomerization states β-Barrel	Mannose, oligomannosides High-mannose *N*-glycans, Complex *N*-glycans	Vacuole Nucleus, cytoplasm
Hevein domain	Different oligomerization states Hevein domain	GlcNAc $(GlcNAc)_n$	Vacuole, cell wall
Jacalin domain	Different oligomerization states β-Prism	Mannose-specific subgroup Galactose-specific subgroup	Nucleus, cytoplasm Vacuole
Legume domain	Different oligomerization states β-Sandwich	Man/Glc, Gal/GalNAc, $(GlcNAc)_n$, fucose, Siaα2,3Gal/GalNAc, complex *N*-glycans	Vacuole Cytoplasm?
Lys M domain	Different oligomerization states β-α-α-β-Structure	$(GlcNAc)_n$	Vacuole Nucleus, cytoplasm
Nictaba-like domain	Homodimer Structure unknown	$(GlcNAc)_n$ High-mannose *N*-glycans Complex *N*-glycans	Nucleus, cytoplasm
Ricin-B domain	Different oligomerization states β-Trefoil	Gal/GalNAc Siaα2,6Gal/GalNAc	Vacuole Nucleus, cytoplasm

Glyco - Bioinformatics:
Study of glycans and its associated research belong to a field called Glycoscience and helps in understanding transport, catalysis, recognition, biological mechanism of synthesis of glycans. These glycans can be made up of either one or multiple monosaccharides in the form of chains or branched complex structure. Various proteins in the form of enzymes, receptors or transporters interact with glycans, forming a conjugate expressed by cells in all organisms. Although laboratory research gives first - hand information on the chemical nature and molecular functioning of these conjugates, biochemists do need certain databases online to access complete structure analysis; hence, the field of glycol - bioinformatics is a fast - growing domain.

The projected idea of glyco - Bioinformatics is to organise information reading glycan and glycoconjugate to generate new in - silico knowledge with the help of interconnection of available resources. The in - silico technique along with analytical methods helps in determining the composition of glycan which largely helps characterise glycans and conjugates. In Bioinformatics, a lot of understanding and analysis is observed by integrating genomics, transcriptomics, and proteomics. Glycomics and metabolomics, although have immense potential, are isolated due to a lack of knowledge and their complex nature.

Domains of Glyco informatics:

The different domain of Glyco informatics (Davide Alocci, Frederique Lisacek 2019)

Lectin Prediction:
A classic bioinformatics tool such as BLAST and HMMER can be the first step in identification of new lectin occurring in a species or scan all possible lectins in an organism. It has been shown that F - type lectins, first identified in the eel, are present in fish and participate in the adhesion of pathogenic bacteria to fucosylated glycoconjugates (Vasta et al.2012), (Vasta et al.2017). Similarly, a parallel analysis focused on the different domains of plant lectins identified in the genomes of five representative central angiosperms (Arabidopsis thaliana, Glycine max, Cucumis sativus, Oryza sativa ssp. japonica and Oryza sativa ssp. indica) (Van Holle et al.2017).

The approach is much more complex if a prediction is required on the entire lectome which is the information regarding all lectins present in a species. In such a case, a search on lectin sequences must be initiated using databases such as Pfam and CATH as the quality of annotation is insufficient using the basic tool search. PLecDom is an exclusive tool developed to identifyplant lectin genomes (Shridhar et al.2009). Althoughseveral tools and databases have been developed, the classification of lectins plays a crucial role in using a specific database.

Lectin prediction: new methods and databases
A LectomeXplore database on UniLectin portal has been immensely used to predict potential lectin candidates across kingdoms after the new lectin classification was implemented. The 9.3.1 Prediction method and LectomeXplore module was able to generate different conserved patterns of classes of lectins in addition to combining 2D protein sequence and 3D structure information. A tool MUSCLE contributes in obtaining conserved patterns of lectins and helps in aligning distantly related lectin sequences which in turn benefits in constructing a phylogenetic tree.

UniLectin3D was created using the glyco3D - lectin3D database. The new database has been expanded in terms of both the quantity of entries (2200 in December 2020 versus 1500 in Lectin3D) and the information provided. Each lectin entry now includes information about its origin, fold, class, and origin, as well as a UniProt AC that contains comparable proteins. Graphical information is available on the glycan linkage site thanks to a collaboration with the PLIP tool developers and SwissModel. External linkages are also provided to the RCSB, PDBe, PDB - CARE, GlyTouCan, GlyConnect, SugarBind, PubMed, and CFG glycan networks. The UniLectin 3D database is updated directly from the PDB's weekly output (300 to 500 structures with 2 to 20 readings per week) enabling the addition of new 3D lectin structures. After filtering out all carbohydrate - active enzymes, the existence of a carbohydrate residue with no covalent attachment to the protein is an efficient selection criterion for identifying lectins in the PDB. The accompanying publication can provide information on the lectin glycan - binding activity when there is no interacting glycan. Unilectin3D should be accessible to a broader audience of glycobiologists and glycochemists rather than only lectin specialists.

The NCBI - nr and UniProtKB databases offer the advantage of covering all kingdoms (without having to manually get the proteome of each species of interest), but the disadvantage of being largely redundant. The LectomeXplore online database makes use of a SQL database containing predicted lectin functions and communicates with the PHP server to build website pages based on user requests. For all of the information displayed in the web database, the LectomeXplore database contains numerous tables. The Protein table includes the NCBI and UniProt AC attributes, the protein name and length, the taxonomy of the linked species, the genetic source, and the identified Pfam name. The taxonomy is included in the species table.

2. Conclusion

Lectins areubiquitous in nature, as they are found in all kinds of organisms, from viruses to humans (Sharon 2008). These lectins are of immense interest in biotechnology and used as catalytic agents for studying glycoconjugates either

in cell or solution. This study helps in cell characterization, defining the glycan structure, and understanding lectin biological function such as immune response, inflammation, or lectin involvement in infection process. The information covering lectin knowledge such as sugar binding, agglutination, or biochemical properties available in online databases is sparse. Tools such as UniProt, RefSeq, Pfam and CATH do not provide a comparative criterion and lack information pertaining to curated and automatic lectin annotations.

Because of the simplicity of a single downloadable protein file containing all available species, rather than other platforms such as GenBank, and due to the limitations of the server available for the analysis, lectin prediction requires protein datasets to be analysed, with both UniProt and NCBI - nr protein sequence datasets used. Regrettably, such databases contain protein from both verified and unverified genetic origins. As a result, protein derived from incomplete and fragmented genomes contains many sequencing mistakes and related information, necessitating a filtering step. There are currently no concrete criteria in LectomeXplore to ensure the quality of a predicted lectin source genome. The solution could be to allow the user to select reference genomes and the quality of projected lectins. Many 3D structures of lectins are described each year and provide new information on glycan recognition and binding methods. To provide the scientific community with annotations and classification of 3D lectin structures, new databases and tools such as LectomeXplore were developed. This tool facilitates and maintains a simple and efficient lectin database which is the need of the hour.

References

[1] URS Rutishauser and Leo Sachs (May 1, 1975). "Cell - to - Cell Binding Induced by Different Lectins". Journal of Cell Biology.65 (2): 247–257. doi: 10.1083/jcb.65.2.247. PMC 2109424. PMID 805150.
[2] Nareddy, P. K., Bobbili, K. B., Swamy, M. J., Purification, physico - chemical characterization and thermodynamics of chito - oligosaccharide binding to cucumber (Cucumis sativus) phloem lectin, International Journal of Biological Macromolecules, 95 (2017), 910 - 919. https: //doi. org/10.1016/j. ijbiomac.2016.10.078.
[3] Ogawa, H., Date, K., The “white kidney bean incident” in Japan, Lectins, Springer, (2014), 39 - 45.
[4] Vasconcelos, I. M., Oliveira, J. T. A., Antinutritional properties of plant lectins, Toxicon, 44 (2004), 385 - 403. https: //doi. org/10.1016/j. toxicon.2004.05.005.
[5] Van Holle, S., Van Damme, E. J., Signaling through plant lectins: modulation of plant immunity and beyond, Biochemical Society Transactions, 46 (2018), 217 - 233. https: //doi. org/10.1042/BST20170371.
[6] Brown, G. D., Willment, J. A., Whitehead, L., C - type lectins in immunity and homeostasis, Nature Reviews Immunology, 18 (2018), 374 - 389. https: //doi. org/10.1038/s41577 - 018 - 0004 - 8
[7] Rüdiger, H., Gabius, H. - J., Plant lectins: occurrence, biochemistry, functions and applications, Glycoconjugate journal, 18 (2001), 589 - 613. https: //doi. org/10.1023/A: 1020687518999
[8] Lehr, C. - M., Bioadhesion technologies for the delivery of peptide and protein drugs to the gastrointestinal tract, Critical Reviews In Therapeutic Drug Carrier Systems, 11 (1994), 119 - 160.
[9] Zhang, J., Shi, J., Ilic, S., Xue, Jun., Kakuda, Y., Biological properties and characterization of lectin from red kidney bean (Phaseolus vulgaris), Food Reviews International, 25 (2008), 12 - 27. https: //doi. org/10.1080/87559120802458115
[10] Gabius, H. - J., André, S., Jiménez - Barbero, J., Romero, A., Solís, D., From lectin structure to functional glycomics: principles of the sugar code, Trends in biochemical sciences, 36 (2011), 298 - 313. https: //doi. org/10.1016/j. tibs.2011.01.005
[11] De Hoff, P. L., Brill, L. M., Hirsch, A. M., Plant lectins: the ties that bind in root symbiosis and plant defense, Molecular Genetics and Genomics, 282 (2009), 1 - 15. https: //doi. org/10.1007/s00438 - 009 - 0460 - 8
[12] Boyd WC, Shapleigh E (1954) Specifi c precipitating activity of plant agglutinins (lectins). Science 119: 419
[13] Sharon N, Lis H (1990) Legume lectins – a large family of homologous proteins. FASEB J 4: 3198–3208
[14] Strosberg AD, Buffard D, Lauwereys M, Foriers A (1986) Legume lectins: a large family of homologous proteins. In: Liener IE, Sharon N, Goldstein IJ (eds) The lectins, properties, functions, and applications in biology and medicine. Academic, Orlando, FL, pp 249–264
[15] Sumner JB, Howell SF (1936) The identification of the hemagglutinin of the Jack bean with concanavalin A. J Bacteriol 32: 227–237.
[16] Lannoo N, Van Damme EJM (2010) Nucleocytoplasmic plant lectins. BiochimBiophys Acta Gen Subj 1800: 190–201.
[17] 854. ISBN ISBN: 978 - 1 - 4419 - 7876 - 9 30. Smith DF, Song X, Cummings RD (2010) Use of glycan microarrays to explore specifi city of glycan - binding proteins. Methods Enzymol 480: 417–444
[18] Van Damme EJM, Barre A, Rougé P, Peumans WJ (2004) Cytoplasmic/nuclear plant lectins: a new story. Trends Plant Sci 9: 484–489
[19] Paulson JC, Blixt O, Collins BE (2006) Sweet spots in functional glycomics. Nat Chem Biol 2: 238–248
[20] Fry S, Afrough B, Leathem A, Dwek M (2012) Lectin array - based strategies for identifying metastasis - associated changes in glycosylation. Methods Mol Biol 878: 267–272
[21] Vijayan, M. and Chandra, N. (1999) Lectins. Curr. Opin. Struct. Biol., 9, 707–714
[22] Chandra, N. R., Prabu, M. M., Suguna, K., and Vijayan, M. (2001) Structural similarity and functional diversity in proteins containing the legume lectin fold. Protein Eng., 14, 857–866.
[23] Sharon N, Lis H (1990) Legume lectins – a large family of homologous proteins. FASEB J 4: 3198–3208.
[24] Strosberg AD, Buffard D, Lauwereys M, Foriers A (1986) Legume lectins: a large family of homologous proteins. In: Liener IE, Sharon N, Goldstein IJ (eds) The lectins, properties, functions, and applications in

biology and medicine. Academic, Orlando, FL, pp 249–264.
[25] Hirabayashi J (2008) Concept, strategy and realization of lectin - based glycan profiling. J Biochem 144: 139–147
[26] Blixt O, Head S, Mondala T, Scanlan C, Hufl ejt ME, Alvarez R, Bryan MC, Fazio F, Calarese D, Stevens J, Razi N, Stevens DJ, Skehel JJ, van Die I, Burton DR, Wilson IA, Cummings R, Bovin N, Wong C - H, Paulson JC (2004) Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. Proc Natl Acad Sci U S A 101: 17033–17038 Els J. M. van Damme.
[27] Taylor ME, Drickamer K (2009) Structural insights into what glycan arrays tell us about how glycan - binding proteins interact with their ligands. Glycobiology 19: 1155–1162
[28] François Bonnardel. Bioinformatics study of lectins: new classification and prediction in genomes. Structural Biology [q - bio. BM]. Université Grenoble Alpes [2020]; Université de Genève, 2021.
[29] Abhinav K V., Sharma A, Vijayan M (2013) Identification of mycobacterial lectins from genomic data. Proteins Structure FunctionBioinforma 81: 644–657. doi: 10.1002/prot.24219
[30] Armougom F, Moretti S, Poirot O, et al (2006) Expresso: Automatic incorporation of structural information in multiple sequence alignments using 3D - Coffee. Nucleic Acids Res. doi: 10.1093/nar/gkl092
[31] Bauters L, Naalden D, Gheysen G (2017) The distribution of lectins across the phylum nematoda: A genome - wide search. Int J Mol Sci 18: doi: 10.3390/ijms18010091
[32] Dang L, Van Damme EJM (2016) Genome - wide identification and domain organization of lectin domains in cucumber. Plant Physiol Biochem 108: 165–176. doi: 10.1016/j. plaphy.2016.07.009
[33] Shridhar S, Chattopadhyay D, Yadav G (2009) PLecDom: A program for identification and analysis of plant lectin domains. Nucleic Acids Res 37: doi: 10.1093/nar/gkp409
[34] Taherzadeh G, Zhou Y, Liew AWC, Yang Y (2016) Sequence - Based Prediction of ProteinCarbohydrate Binding Sites Using Support Vector Machines. J Chem Inf Model 56: 2115–2122. doi: 10.1021/acs. jcim.6b00320
[35] Taherzadeh G, Zhou Y, Liew AWC, Yang Y (2016) Sequence - Based Prediction of ProteinCarbohydrate Binding Sites Using Support Vector Machines. J Chem Inf Model 56: 2115–2122. doi: 10.1021/acs. jcim.6b00320
[36] Vasta GR, Ahmed H, Bianchet MA, et al (2012) Diversity in recognition of glycans by F - type lectins and galectins: molecular, structural, and biophysical aspects. Ann. N. Y. Acad. Sci.1253: E14– E26