Structural Database for Lectins and the UniLectin Web Platform

François Bonnardel, Serge Pérez, Frederique Lisacek, Anne Imberty

Structural Database for Lectins and the UniLectin Web Platform

François Bonnardel [1,2,3], Serge Perez [1] , Frédérique Lisacek [2,3,4]*, and Anne Imberty [1]*

1. Univ. Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France.
2. Swiss Institute of Bioinformatics, CH-1227 Geneva, Switzerland.
3. Computer Science Department, UniGe, CH-1227 Geneva, Switzerland.
4. Section of Biology, UniGe, CH-1205 Geneva, Switzerland.

* To whom correspondence should be addressed. Anne Imberty (anne.imberty@cermav.cnrs.fr, Tel: +33 476 03 76 40, Twitter: @AnneImberty) Frédérique Lisacek (frederique.lisacek@sib.swiss, Tel: +4122 379 01 95)

Abstract

The search for new molecules requires a clear understanding of biosynthesis and degradation pathways. This view applies to most metabolites as well as other molecule types such as glycans whose repertoire is still poorly characterized. Lectins are proteins which recognize specifically and interacts non-covalently with glycans. This particular class of proteins is considered as playing a major role in biology. Glycan-binding is based on multivalence, which gives lectins a unique capacity to interact with surface glycans and significantly contribute to cell-cell recognition and interactions. Lectins have been studied for many years using multiple technologies and part of the resulting information is available online in databases. Unfortunately, the connectivity of these databases with the most popular omics databases (genomics, proteomics and glycomics), remains limited. Moreover, lectin diversity is extended and requires setting out a flexible classification that remains compatible with new sequences and 3D structures that are continuously released. We have designed UniLectin as a new insight in the knowledge of lectins, their classification and their biological role. This platform encompasses UniLectin3D, a curated database of lectin 3D structures that follow a periodically updated classification , a set of comparative and visualizing tools and gradually released modules dedicated to specific lectins predicted in sequence databases. The second module is PropLec, focused on β-propeller lectin prediction in all species based on five distinct family profiles. This chapter describes how UniLectin can be used to explore the diversity of lectins, their 3D structures and associated functional information as well as to perform reliable predictions of β-propeller lectins.

Keywords

Lectin, Carbohydrate-binding protein, Database, Classification, Sequence, 3D structure, Profile-based prediction

1 Introduction

Lectins are oligomeric proteins that bind mono- and oligosaccharides reversibly and specifically while displaying no catalytic or immunological activity [1]. Those complex carbohydrates (also referred to as glycans) occur in the form of single molecules, or as part of glycoconjugates (glycoproteins and glycolipids). They constitute the most abundant class of biomolecules on Earth. Complex carbohydrates are built for high-density bio-coding, the information being carried and encoded in their 3D-structures and sometimes in their dynamics. Lectins are powerful macromolecular tools to decipher the high-density bio-encoding of complex carbohydrates. Along with a high-specificity, lectins exhibit diverse architectures and modes of multivalence relating to their function. Some lectins exhibit ‘‘architectural multivalence’’, consisting of one macromolecular structure with several, equivalent carbohydrate recognition domains. Other lectins, such as adhesins, are membrane-bound and include only one carbohydrate recognition domain, attached to fimbriae or flagella tethered to the cell surface. Several fimbrial structures clustered together at the cell surface also give a multivalent presentation of the lectins in the extracellular environment. All of these recognition processes play important roles in fertilization, embryogenesis, inflammation, metastasis, and parasite–symbiote recognition in microbes, invertebrates, plants, and vertebrates.

In 1984, Gallagher [2] made a first attempt to classify and establish a nomenclature of lectins. Then, an overview of seven families of plant lectins [3] and a classification of animal lectins [4] were proposed. Historically, these classifications were based on the discovery of new lectins supported by the resolution of their 3D structures. The accumulation of data revealed the occurrence of similar or closely similar protein motifs across several kingdoms. Consequently, a species-independent classification based on lectin sequence and 3D structure appeared more relevant. Such an approach was proposed through the association of 3D features and Pfam domains [5]. Unfortunately, the approximate definition of Pfam domains [6] that are rarely specific to lectin sequences not only reduces the accuracy of classes but also precludes the classification of newly discovered lectins. In the same vein, a classification of fucose-binding lectins based on Pfam domains was suggested by H. Makyio and R. Kato in 2016 [7]. The following examples illustrate the possible sources of ambiguity. Using “lectin” and “glycan binding” as keywords to search the UniProtKB database [8] returns roughly 159,000 and 230,000 distinct proteins, respectively (UniProt 2019-03 release). Searching the Protein Data Bank (PDB) through PDBe [9] with the same keywords returns 3634 and 121 structures, respectively (PDB 2019-04-24 release). A close examination of the results indicates that some of these proteins are in fact glycoproteins that bind glycan covalently, and others are enzymes that recognize and modify glycans. This highlights the need for a robust classification of lectins and their ligands based on a large panel of finely curated or filtered information extracted from genomic data as well as 3D structures and their inter-atomic features.

2 Existing Lectin Databases

Due to the crucial role of lectins in cell recognition and interactions, a large amount of experimental information has been published. They encompass peptide sequences of the lectin domain as well as full protein sequences, 3D structures derived from diffraction studies which possibly contain the interacting carbohydrate, and glycan arrays which reflect the specificity for different glycans and their target affinity. Such a large body of information needs to be integrated and organized in a public database that is interoperable with other relevant omics databases. Several initiatives that we now briefly summarize, have been launched to address this question.

Glyco3D [10] includes a family of databases covering the 3D features of monosaccharides, disaccharides, oligosaccharides, polysaccharides, glycosyltransferases, lectins, monoclonal antibodies and glycosaminoglycan binding proteins that have been developed with nonproprietary software and are freely available to the scientific community (http://glyco3d.cermav.cnrs.fr).

The ***Lectin Frontier Database*** (LfDB: https://acgg.asia/lfdb2/) [11] contains 400 lectins with curated information including lectin name, Pfam family, name of the interacting glycan, species, fold, PDB 3D structure, protein sequence and reference. It is now integrated in GlyCosmos, a newly developed portal (https://glycosmos.org/lectins/) for accessing and exploring knowledge in glycobiology.

LectinDB (http://proline.physics.iisc.ernet.in/lectindb/) [12] is an annotated database mainly of plant lectins.

It can be searched based on their species, accession number, lectin domain, fold, PDB code and interacting glycan; a protein sequence can be compared to the lectins in lectinDB; an overview of lectin available in each species is also available; each lectin is associated with ataxon and a UniProt entry.

SugarBindDB (http://sugarbind.expasy.org/) [13] provides integrated information on pathogen lectins and their corresponding glycan targets. This curated database describes either bacterial or viral proteins and details their binding specificity. Each ligand can be matched to full glycan structures of GlyConnect [14] and their associated glycoproteins to reveal potential glycan-mediated interactions between pathogen lectins and host glycoproteins. When 3D structures are available from PDB, they can be visualized with the LiteMol software [15]. It also includes affinity data when available. Despite this high level of data and tool integration SugarBindDB remains focused on pathogenic virus and bacterial species.

The databases of the Consortium for Functional Glycomics (CFG resources: http://www.functionalglycomics.org/static/consortium/resources.shtml) [16] include a repository of glycan array data used to measure the specificity of a lectin to multiples glycans. These databases are unfortunately no longer maintained. The latest related resource is ***GLAD*** (https://glycotoolkit.com/GLAD/) [17] a web-based tool designed to visualize and analyze glycan microarray data, providing a list of available glycan array for lectins.

Most of these lectin-dedicated databases lack interaction details with protein and protein family [18, 19], protein structure classification databases [20, 21], glycans or protein glycan interaction analysis tools [22, 23], and do not necessarily comply with glycan textual and visual representations now commonly accepted in the glycoscience community [24]. Such spread out information led us to design and implement the ***UniLectin*** platform (https://unilectin.eu) in order to address the data integration and classification issues. The present chapter describes the current content and the possible use of the UniLectin comparative, predictive and visualizing tools. The platform consists of modules the main one being ***UniLectin3D*** a curated database of 3-dimensional structures of lectins, as established mainly from crystallographic methods. It includes a classification of lectins, along with the knowledge of interacting glycans. Based on this information, we are gradually populating other modules dedicated to specific lectins predicted in sequence databases.The second module, PropLec, is focused on families of lectins that display a β-propeller structure.

3 UniLectin3D, Database of Curated Lectin 3D Structure and their Interacting Ligands

The UniLectin3D database includes structural information on lectins along with their interactions with carbohydrate ligands. A curated classification is proposed based on origin and fold in association with the literature and functional data such as known specificity. The content of UniLectin3D is centered on 3-dimensional data, using PDB information, with an appropriate curation of the glycan topology. It provides a family-based classification and cross-links to specialized glyco-related databases. Finally, the 3D visualization of contacts between the lectin and the ligand, is visualized via the Protein-Ligand Interaction Profiler (PLIP) application. The introduction of such a feature is likely to meet the expectations of lectin specialists. The three-dimensional structures reported in UniLectin3D are those of lectins crystallized with or without their carbohydrates (glycans) ligands and non-carbohydrate ligands (see Background for detailed information). The current **2117** lectin structures were all manually curated; this corresponds to **534** different lectins (as of **2019-04-17**). Bibliographic entries cover **860** published articles describing at least one structure. The first classification level, referred to as “origins”, separates the lectins into seven different classes reflecting the main orders of the living kingdom. The second level orders the lectins according to the protein fold into **75** classes. The third level separates the lectins according to their species in **309** families.

Among the **2117** 3D structures, **1338** occur as complexed with glycans. The most commonly observed monosaccharides are as follows: Galactose (Gal) **31**%(**730**); N-Acetyl glucosamine (GlcNAc) **16**%(**377**), Glucose (Glc) **15**%(**345**), Mannose (Man) **12**%(**283**), Fucose (Fuc) **9**%(**211**), sialic acid (Neu5Ac) **9**% (**211**), N-Acetyl galactosamine (GalNAc) **7**% (**170**). Rarer sugars are also observed in complexes with lectins (Rhamnose, Arabinose …). The ligands occur as monosaccharides, but also as oligosaccharides or glycoconjugates. The set of distinct glycan ligands amounts to **222**.

The following options are available for searching by: (1) keywords, (2) kingdom order, (3) historical classification, (4) monosaccharide and associate IUPAC sequence, (5) fold of the binding site and (6) multiple criteria. Once selected, lectins can be explored (and their features pictured and downloaded) by sequence (with the UniProt AC) and structure (with the PDB ID). For each lectin a detailed page is available with 3D visualization, interactions and links to external databases.

3.1 Searching by Keywords

The search can be performed by entering keywords as textual input: i.e. human, PDB code, UniProt accession number, lectin name, type of domains, fragment of glycan sequence, or textual fragments of the title of a publication: .e : human, propeller, 1TL2 (PDB), Q47200 (UniProt), GalNAc, Lewis.

3.2 Searching by Kingdom Order, Carbohydrate Binding Site Class and Species Family

Different modes are available for browsing the database and visualizing data coverage with respect to taxonomy. Lectin structures can be explored using two interactive graphical representations: a sunburst and a tree. The inner circle of the sunburst corresponds to the kingdom orders along with their respective percentage of occurrence. For a given kingdom order, (for example animal) the class of the carbohydrate binding site is displayed in the central section (for example galectins) whereas the species families can be browsed on the outer section (for example galectin 3). The hierarchical taxonomic tree (Figure 1A) can be used to explore the classification with the tree leaves expanding by clicking on the blue node and at each level. An advanced search can be launched by clicking on the label of the corresponding level.

3.3 Searching by Monosaccharide and Associate IUPAC Aequence

Glycans are described following the encoded IUPAC condensed nomenclature (http://www.sbcs.qmul.ac.uk/iupac/2carb/38.html). Ex. Gal(b1-4)GlcNAc(b1-2)Man(a1-3)[GlcNAc(b1-2)Man(a1-6)]Man

The nature of the interacting glycans can be visualized on a sunburst (Figure 1B). The inner circle corresponds to the monosaccharide that composes the carbohydrate chains interacting with the lectins, together with their respective percentage of occurrence. The outer circle corresponds to the carbohydrate chains. For a given monosaccharide (D-Galp for example) the number of occurrences of glycans containing the selected monosaccharide is displayed (for example 29 in the case of Gal(b1-3)GalNAc). Clicking on a selected glycan (for example Gal(b1-3)GalNAc) lists the 3D structures of lectins complexed with this carbohydrate.

A distinct interface is accessible (Figure 1C) by clicking the glycan search button. It allows searching for glycans through a combination of monosaccharides. Only the carbohydrates present in at least one lectin 3D structure and containing the selected monosaccharides are displayed in the SNFG format (Figure 1D). By clicking on a carbohydrate name, an advanced search is launched based on it.

3.4 Searching by Fold

The lectin fold is defined by the relative spatial arrangement of secondary structure elements. It characterizes the multiplicity of glycan binding and it is directly linked to the ability of lectins to bind and cross-link glycan containing molecules in a multivalent fashion. A sunburst representation indicates the occurrence of the main fold of lectins. Clicking on a selected section of the circle displays the nature and the occurence of the folds as identified in the crystal structures. Upon a click, buttons prompt the information pertaining to the lectin displaying the particular fold.

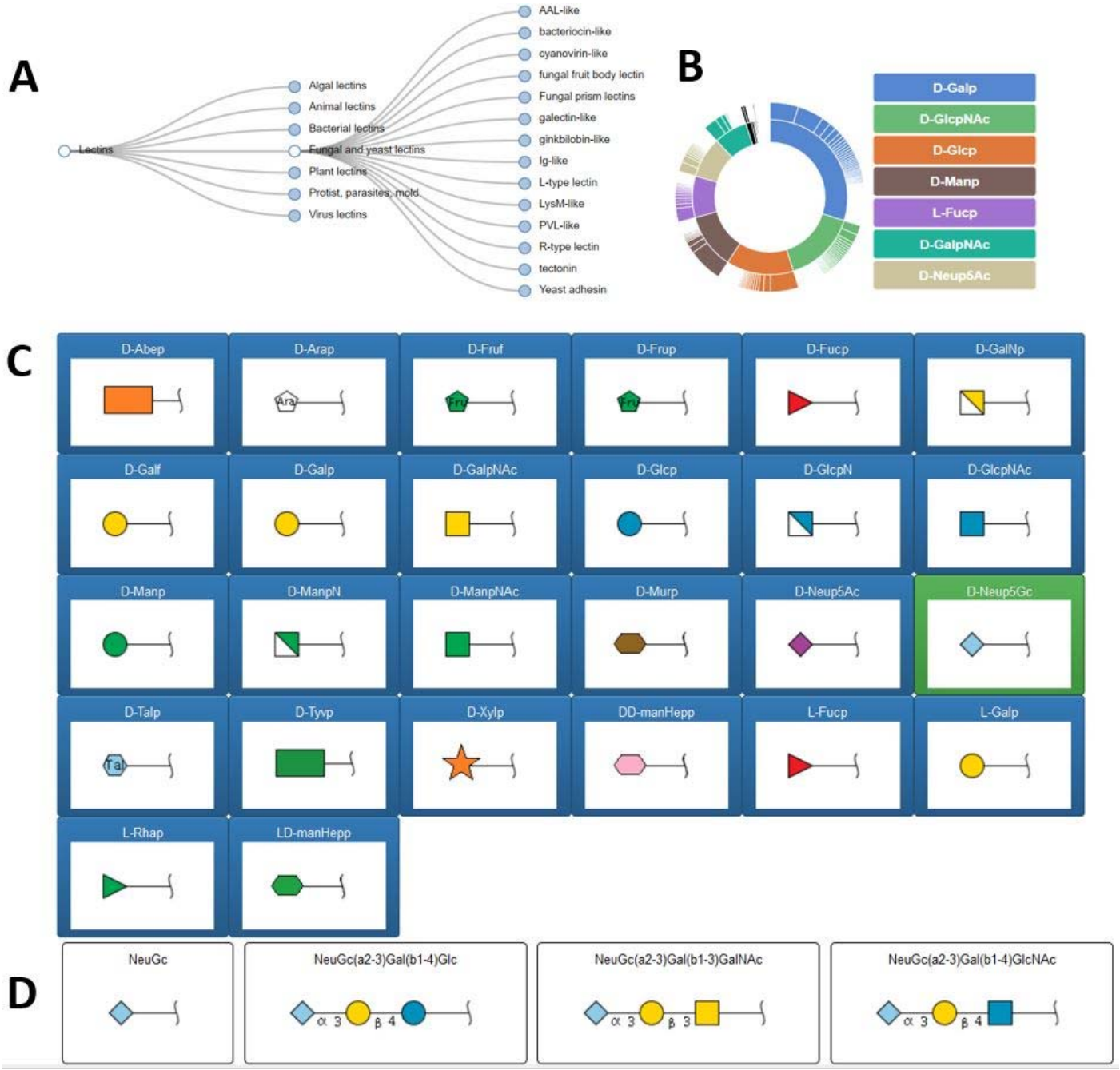

Figure 1. *Interfaces for the exploration of the lectin classification and associated carbohydrates ligands. A. Treeview of the classification highlighting the option of opening each node to display branching at the lower level. B. Sunburst as an overview of the glycans distribution. C. Glycan search dedicated page for selecting monosaccharides to be searched in all structures. D. Example of the oligosaccharides obtained by selecting D-Neup5Gc in the monosaccharide panel.*

3.5 Advanced Search, Searching by Multiple Criteria

This advanced search option offers a range of criteria to be selected in a combined fashion in order to search the whole database for specific lectins or structures. Practically, lectins can be searched by 3D structure or by sequence family with the support of drop-down lists (Figure 2A) and they can be filtered based on a large number of features (Figure 2B). (i) The classification of lectins (Origin, Class, and Family); (ii) the nature of the fold and taxonomic details of the lectin; (iii) Keywords from the title of a reference article; (iv) A unique feature is the search of fragments of glycan ligands, also called oligosaccharide motifs, that interact with the lectin. (v) Finally, a cutoff on the resolution (Å) of the X-ray structure can be used as a filter for selecting of high-quality data. UniLectin3D allows precise taxonomic search (vi) for all lectins that have been structurally characterized in a given organism.

The resolution criteria relate to the quality of the structural determination (High numeric values of resolution,

such as 4 Å, mean poor resolution, while low numeric values, such as 1.5 Å, indicate a good resolution. (The median resolution for X-ray crystallographic results in the Protein Data Bank is 2.05 Å). Whenever the resolution is set to 0, the structures are not filtered.

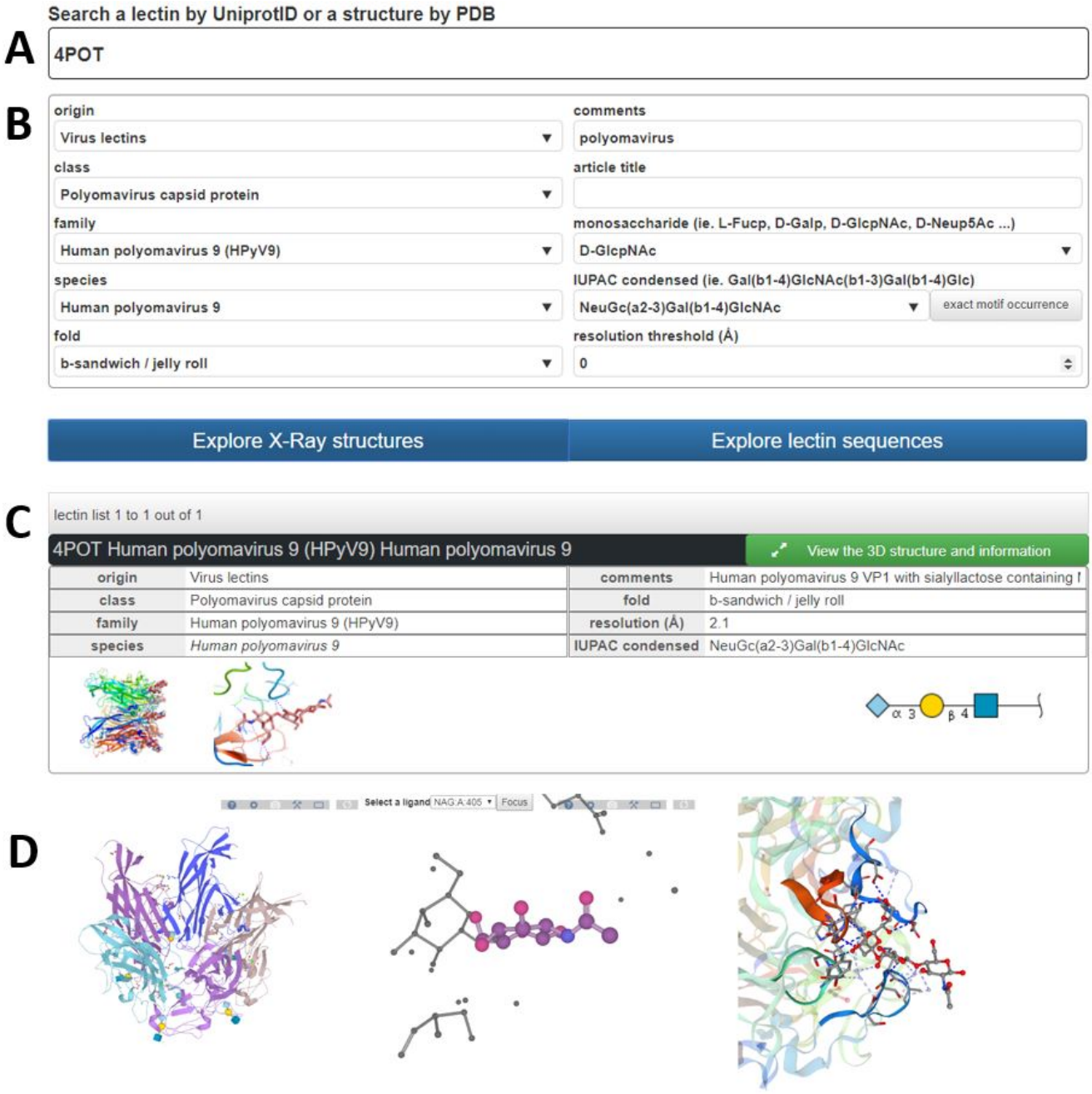

Figure 2. *Multiple criteria search of a lectin, with a detailed interface with information and 3D-visualization. A. Multicriterion window for advanced search. B. Example of results obtained by searching for an oligosaccharide. C. Detailed graphical information provided for the entry shown in B.*

3.6 Lectin Sequence Setailed Interface

The accession number (AC) which is assigned to each lectin sequence upon inclusion into UniProtKB can be used for searching. As one lectin sequence can be related to multiple PDB structures these are displayed in the results together with the ligand(s) shown in the SNFG representation. Some structures may have been published in multiple articles; these are listed together with the corresponding link to PubMed (http://www.ncbi.nlm.nih.gov/pubmed). When available, the structures are displayed on the main protein sequence in a 2D viewer provided by PDBe [9].

3.7 X-Ray Structure Detailed Interface

The PDB code which is assigned to each lectin structure is used to list available lectin structures and to display more detailed information (Figure 2C). Each structure is related to a protein with a UniProt AC. Other related PDB structures are listed together with their interacting glycans if any. The 3D X-ray structure of the lectin is visualized directly and interactively with the integration of the LiteMol [15] and NGL [25] viewers (Figure 2D).

Information about the interaction occurring between the glycan and the combining sites of the lectin can be obtained using the Protein-Ligand Interaction Profiler (PLIP) server [23]. The description of the glycan complies with representations and numerical descriptors that allow for cross-linking to other databases in glycoscience. The architecture and navigation tools are designed to extend the search to all organisms, as well as to search for all glycan epitopes complexed within specified binding sites. The NGL viewer adapted to SwissModel [26] displays the interactions as resulting from the PLIP application. This offers a detailed 3D visualization of the specific features of the interactions between the glycans and the surrounding amino-acid residues and possible metal ions. A complementary description of the 3D interactions between the lectin and glycan is given by the domain viewer of the PDBe.

4 PropLec, database of structure-based predicted β-propellers

The PropLec database includes predicted β-propeller lectins along with their features and conserved regions. In structural biology, a β-propeller is a particular type of beta-sheet protein architecture characterized by 4 to 8 highly symmetrical blade-shaped β-sheets arranged toroidally around a central axis. Together the β-sheets form a funnel-like active site. The blade consists of a small domain of less than 50 amino acids. The repeated blades hamper the identification of similar lectins when using common software based on pairwise sequence alignment such as BLAST [27]. However, the multiple alignment of blades manually adjusted with knowledge of 3D structures produces a unique conserved domain. This blade domain can then be used to compare all known β-propeller lectins and this systematic comparison led to the definition of five distinct families. To simplify the nomenclature, each family is named based on the number of blades, e.g. PropLec5A (Tachylectin 2 like), PropLec6A (RSL and AAL like), PropLec6B (Tectonin like), PropLec7A (PLL like) and PropLec7B (PVL like) (Figure 3A).

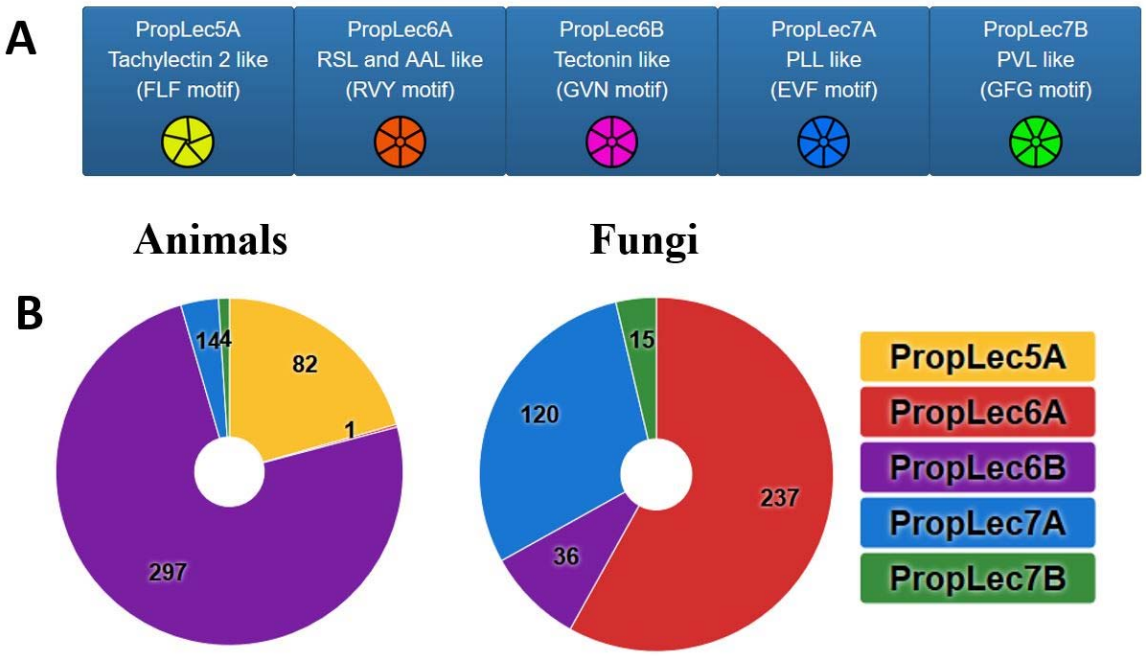

***Figure 3**. β-Propeller families. A. The five families of β-propellers used to build the database. B. Analysis of repartition of the different families exemplified by searching for animal lectins (metazoan) and fungal lectin*

The specific signature of each family has been used to predict with HMMER [28] possible β-propeller lectins from the UniProt sequence database [8] and are associated with RefSeq if available [29]. The results of this

prediction are stored and can be searched in the PropLec module of UniLectin, based on their family, species, taxonomy, number of blades, associated enzymatic functions and other additional features.

4.1 Searching by Keywords and by Family

The features of the predicted β-propeller lectins can be explored using multiple criteria from the homepage. A quick search can be performed by keywords: i.e. accession number, species name or protein name. The five families are then made accessible through buttons along with a pie chart depiction of the distribution of the number of predicted proteins in each family. Based on the two distinct sets Animal and Fungi, the distribution of the number of predicted β-propellers is represented (Figure 3B). As expected, a majority of the predicted animal β-propeller are PropLec6B and a majority of the predicted Fungi β-propellers are PropLec6A. Surprisingly, ProLec7A from *Photorhabdus luminescens* is predicted in both Animals and Fungi, and PropLec 7B from *Psathyrella velutina* is predicted in Animals.

4.2 Searching by Number of Blades in the Propeller

The number of blades in the predicted β-propeller lectins can be used to search for particular structures. The result is shown as a histogram representing the distribution of the predicted lectins relative to the number of blades identified in sequences. Clicking on any of the bars of the histogram prompts the details of the corresponding predicted lectins.

4.3 Searching by Phylum

The search can be performed by selecting in the Taxonomy sunburst either a superkingdom, a phylum or species. The distribution of predicted lectins across species can be explored at each level. As the sunburst is built as concentric circles, the inner circle represents the superkingdom, the next circle the kingdom, the third circle the phylum, the fourth circle the species group and the outer circle the species. Clicking on any of the circle sections prompts the details of the predicted lectins found in the selected taxonomic group.

4.4 Advanced Search

Predicted propeller lectins can be selected using a combination of criteria, and the corresponding lectins are then ordered by scores (highest to lowest). Possible combinations involve: (i) the prediction score threshold, which is set 0.25 by default; (ii) the identified propeller family; (iii) the number of blades identified; (iv) the maximum distance between blades; (v) keywords that exclude proteins based on their description, set by default set to “partial, synthetic and undefined”, (vi) taxonomy, (vii) Pfam domains, (viii) RefSeq AC, (ix) protein name and description; (x) UniProt AC. The “checkbox pathogen” button offers the possibility to restrict the search to a particular pathogen species, based on the NIH predefined list of pathogen species.

A graphic overview of the properties of predicted propeller lectins resulting from searching with a single criterion or a combination of criteria, is generated. It shows : (i) the distribution of the PropLec families; (ii) the distribution of the number of blades; (iii) a sunburst and a tree representation –as in the homepage. Clicking on the graphic sections displays further details. The predicted lectins matching the criteria are ordered by score with 20 items per page. For each predicted lectin, the following features are displayed: (i) protein name; (ii) UniProt AC and RefSeq AC which can be clicked to access the corresponding pages in the respective sequence databases; (iii) protein length, (iv) species, (v) domain identified in the Proplec families, (vi) number of blades; (vii) similarity score of the predicted protein blades to the reference blade; (viii) protein-coding gene list including chromosome number(s) and location.

For each predicted lectin, an in-house 2D sequence viewer indicates the localization of the predicted blades and possible Pfam domains. Zooming in the sequence is performed via a drag and drop button. Further details are available by clicking the more information button.

4.5 Detailed Results

For each lectin, a detailed panel and page are available with the NCBI gene viewer and a representation of the blade conservation compared to the reference. The protein features were described in the previous section. The protein-coding gene and its location on a chromosome is represented by the NCBI viewer, when the information is available [30]. The view can be scrolled back and moved by drag and drop to check the

surrounding genes. To compute the score, the predicted lectin blades have been aligned against the reference blades by the MUltiple Sequence Comparison by Log- Expectation (MUSCLE) [31]. To provide a more detailed view of the blade conservation, the resulting alignment is represented in two distinct bar charts. The first bar chart, on top, contains the amino acid conservation of the predicted blades. The second represents the amino acid conservation of the reference blade. As all blades are aligned, the bar chart position facilitates the comparison between sequences. The binding sites are represented, along with the amino acids known to interact with glycans either by a hydrogen bond or by hydrophobic interactions.

4.6 Searching by Other Pfam Functional Domains

Lectin domains can either bea whole lectin protein or only one part of a larger multi functional protein. Proteins with a predicted lectin β-propeller domain(s) and other functional Pfam domains can be explored to evaluate possible combination(s). Such complex architecture is of particular interest as it highlights the specificity of lectins towards a definite carbohydrate in combination with other functional domain(s) (ie. a glycan transferase). For each component of the protein architecture, a button prompts the list of predicted lectins displaying a similar pattern.

5 Conclusion and Discussion

The development of high quality glycomics databases counteracts the lack of precision reflected in the abundance of unreviewed and incorrect information regarding both glyconjugates and glycan-binding proteins in genome and protein databases. Here, we reviewed databases with information on lectins and their interacting glycans (mono, oligo, and polysaccharides). The recently released UniLectin platform provides a curated classification of lectins along with their reviewed interactions with glycans. Tools that facilitate lectin knowledge exploration were implemented. UniLectin has recently exceeded 2000 lectin structures. The platform also includes a tutorial that describes step by step the usage of simple and advanced search of the UniLectin databases covering lectin 3D structures and predicted β -propeller lectins .

We strive to ensure content accuracy and regular updates of the UniLectin platform as well as to provide a user-friendly tool collection (search, visualization, etc). Currently, UniLectin has a growing community whose feedback is key to driving further development. Based on UniLectin3D curated information and lectin classes, a high-quality prediction of lectins in genomes is a reachable short term goal but it still requires the revision of the current classification criteria, including amino acid sequence patterns.

Acknowledgments

The authors acknowledge support by the ANR PIA Glyco@Alps (ANR-15-IDEX-02) and the Alliance Campus Rhodanien Co-funds (http://campusrhodanien.unige-cofunds.ch).

References

1. Lis H, Sharon N (2002) Lectins: Carbohydrate-specific proteins that mediate cellular recognition Chem Rev 98:637-674.
2. Gallagher JT (1984) Carbohydrate-binding properties of lectins: A possible approach to lectin nomenclature and classification. Biosci Rep 4:621-632.
3. Peumans WJ, Van Damme EJ, Barre A et al. (2001) Classification of plant lectins in families of structurally and evolutionary related proteins. Adv Exp Med Biol 491:27-54.
4. Kaltner H, Gabius H-J (2011) Animal lectins: from initial description to elaborated structural and functional classification. *The Molecular Immunology of Complex Carbohydrates —2 Advances in Experimental Medicine and Biology, vol 491* ed Wu AM (Boston, MA, Springer), pp 79-94.
5. Fujimoto Z, Tateno H, Hirabayashi J (2014) Lectin structures: Classification based on the 3-D structures. Methods Mol Biol 1200:579-606.

6. Finn RD, Bateman A, Clements J et al. (2014) Pfam: The protein families database. Nucleic Acids Res 42:D222-D2230.

7. Makyio H, Kato R (2016) Classification and comparison of fucose-binding lectins based on their structures. Trends in Glycoscience and Glycotechnology 28:E25-E37.

8. Bateman A, Martin MJ, O'Donovan C et al. (2017) UniProt: The universal protein knowledgebase. Nucleic Acids Res 45:D158-D169.

9. Mir S, Alhroub Y, Anyango S et al. (2018) PDBe: Towards reusable data delivery infrastructure at protein data bank in Europe. Nucleic Acids Res 46:D486-D492.

10. Pérez S, Sarkar A, Rivet A et al. (2015) Glyco3d: A portal for structural glycosciences. Methods Mol Biol 1273:241-258.

11. Hirabayashi J, Tateno H, Shikanai T et al. (2015) The lectin frontier database (LfDB), and data generation based on frontal affinity chromatography. Molecules 20:951-973.

12. Chandra NR, Kumar N, Jeyakani J et al. (2006) Lectindb: A plant lectin database. Glycobiology 16:938-946.

13. Mariethoz J, Khatib K, Alocci D et al. (2016) SugarBindDB, a resource of glycan-mediated host-pathogen interactions. Nucleic Acids Res 44:D1243-D1250.

14. Alocci D, Mariethoz J, Gastaldello A et al. (2019) GlyConnect: Glycoproteomics goes visual, interactive, and analytical. Journal of Proteome Research 18:664-677.

15. Sehnal D, Deshpande M, Vařeková RS et al. (2017) LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data. Nat Methods 14:1121-1122.

16. Raman R, Venkataraman M, Ramakrishnan S et al. (2006) Advancing glycomics: Implementation strategies at the consortium for functional glycomics. Glycobiology 16.

17. Mehta AY, Cummings RD (2019) GLAD: GLycan Array Dashboard, a visual analytics tool for glycan microarrays. (in press) doi: 10.1093/bioinformatics/btz075

18. Bairoch A (2004) The Universal Protein Resource (UniProt). Nucleic Acids Res 33:D154-D159.

19. Jones P, Binns D, Chang HY et al. (2014) InterProScan 5: Genome-scale protein function classification. Bioinformatics 30:1236-1240.

20. Fox NK, Brenner SE, Chandonia JM (2014) SCOPe: Structural Classification of Proteins - Extended, integrating SCOP and ASTRAL data and classification of new structures. Nucleic Acids Res 42:D304-D309.

21. Sillitoe I, Lewis TE, Cuff A et al. (2015) CATH: Comprehensive structural and functional annotations for genome sequences. Nucleic Acids Res 43:D376-D381.

22. Lütteke T, von der Lieth CW (2004) pdb-care (PDB CArbohydrate REsidue check): A program to support annotation of complex carbohydrate structures in PDB files. BMC Bioinformatics 5:69.

23. Salentin S, Schreiber S, Haupt VJ et al. (2015) PLIP: Fully automated protein-ligand interaction profiler. Nucleic Acids Res 43:W443-W447.

24. Haltiwanger RS (2016) Symbol Nomenclature for Glycans (SNFG). Glycobiology 26:217-217.

25. Rose AS, Bradley AR, Valasatava Y et al. (2018) NGL viewer: Web-based molecular graphics for large complexes. Bioinformatics 34:3755-3758.

26. Bienert S, Waterhouse A, De Beer TAP et al. (2017) The SWISS-MODEL Repository-new features and functionality. Nucleic Acids Res 45:D313-D319.

27. Camacho C, Coulouris G, Avagyan V et al. (2009) BLAST+: architecture and applications. BMC bioinformatics 10:421.

28. Finn RD, Clements J, Arndt W et al. (2015) HMMER web server: 2015 Update. Nucleic Acids Res 43:W30-W38.

29. O'Leary NA, Wright MW, Brister JR et al. (2016) Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44:D733-D745.

30. Brown GR, Hem V, Katz KS et al. (2015) Gene: A gene-centered information resource at NCBI. Nucleic Acids Res 43:D36-D42.

31. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-1797.