



HAL
open science

UniLectin, A One-Stop-Shop to Explore and Study Carbohydrate-Binding Proteins

Anne Imberty, François Bonnardel, Frédérique Lisacek

► **To cite this version:**

Anne Imberty, François Bonnardel, Frédérique Lisacek. UniLectin, A One-Stop-Shop to Explore and Study Carbohydrate-Binding Proteins. *Current Protocols*, 2021, 1 (11), 10.1002/cpz1.305 . hal-03452596

HAL Id: hal-03452596

<https://hal.science/hal-03452596>

Submitted on 27 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UniLectin, A One-Stop-Shop to Explore and Study Carbohydrate-Binding Proteins

Anne Imberty,^{1,5}  François Bonnardel,^{1,2,3}  and Frédérique Lisacek^{2,3,4,5} 

¹Université Grenoble Alpes CNRS, CERMAV, Grenoble, France

²SIB Swiss Institute of Bioinformatics, Geneva, Switzerland

³Computer Science Department, UniGe, Geneva, Switzerland

⁴Section of Biology, UniGe, Geneva, Switzerland

⁵Corresponding authors: anne.imberty@cermav.cnrs.fr; frederique.lisacek@sib.swiss

All eukaryotic cells are covered with a dense layer of glycoconjugates, and the cell walls of bacteria are made of various polysaccharides, putting glycans in key locations for mediating protein-protein interactions at cell interfaces. Glycan function is therefore mainly defined as binding to other molecules, and lectins are proteins that specifically recognize and interact non-covalently with glycans. UniLectin was designed based on insight into the knowledge of lectins, their classification, and their biological role. This modular platform provides a curated and periodically updated classification of lectins along with a set of comparative and visualization tools, as well as structured results of screening comprehensive sequence datasets. UniLectin can be used to explore lectins, find precise information on glycan-protein interactions, and mine the results of predictive tools based on HMM profiles. This usage is illustrated here with two protocols. The first one highlights the fine-tuned role of the O blood group antigen in distinctive pathogen recognition, while the second compares the various bacterial lectin arsenals that clearly depend on living conditions of species even in the same genus. © 2021 The Authors. Current Protocols published by Wiley Periodicals LLC.

Basic Protocol 1: Searching for the structural details of lectins binding the O blood group antigen

Basic Protocol 2: Comparing the lectomes of related organisms in different environments

Keywords: carbohydrate • glycan-binding • lectin • structural biology

How to cite this article:

Imberty, A., Bonnardel, F., & Lisacek, F. (2021). UniLectin, a one-stop-shop to explore and study carbohydrate-binding proteins. *Current Protocols*, 1, e305. doi: 10.1002/cpz1.305

INTRODUCTION

The next frontier in understanding protein function is deciphering the “PTM code.” Glycosylation, as the most variable post-translational modification, complicates the picture with its own challenging glycodecode borne by eukaryotic surface glycoconjugates or microbial membrane polysaccharides. The ability of lectins to non-covalently bind glycans makes these proteins the key readers of information encoded by glycans (Gabiuis, Siebert, Andre, Jimenez-Barbero, & Rudiger, 2004). The roles of glycans and glycoconjugates

Imberty et al.

1 of 14

are manifold and revealed in various medical, biochemical, and biotechnological applications. Glycoscience is now reaching out to other omics and systems biology following recent technological progress in carbohydrate structure resolution and *ab initio* synthesis, as well as functional screening methods.

Lectins are diverse within and across all species. This particular class of proteins is considered as a major player in many biological processes taking place on the surface of cells. Lectins have been studied for many years using broad technologies ranging from X-ray crystallography to various screening methods, e.g., microarrays; nonetheless, they are poorly annotated in genomes and no reference database was available before the launching of the UniLectin platform in 2019 (<https://www.unilectin.eu>).

Structuring lectin knowledge has been challenged by the lack of a reliable classification system that is robust with respect to the continuous production of new sequences and the slow elucidation of corresponding 3D structures. Several criteria for grouping lectins have been proposed based on phylogeny or binding specificity, but none have proven satisfactory (Bonnardel, Perez, Lisacek, & Imberty, 2020). A key aspect is that glycan binding relies on multivalence, which gives lectins a unique capacity to interact with surface glycans. β -Propellers in which each propeller contains a glycan-binding site are examples of this situation.

The initial content of UniLectin consisted of manually curated lectin 3D structures together with their interacting glycans grouped in the UniLectin3D module (Bonnardel et al., 2019). Then, the potential for prediction from sequence data was validated with tandem repeat lectins. As a proof of concept, β -propellers were categorized by their propeller content, and sequence databases were screened with corresponding Hidden Markov Model (HMM) profiles. Prediction was proven reliable and experimentally tested (Bonnardel et al., 2019). These results were compiled in the PropLec module. In 2020, accumulated data set the basis for a robust hierarchical classification built on 35 protein folds and expanding into 109 classes. The latter were used to define 109 HMM profiles and extensively screen NCBI-nr and UniProt. This resulted in the LecomXplore module, which includes approximately 500,000 lectin predictions across all kingdoms (Bonnardel, Mariethoz, Perez, Imberty, & Lisacek, 2021). This analysis revealed that species distribution is wider than expected. For instance, calnexin, calreticulin, and malectin, involved in quality control during glycoprotein biosynthesis, are not limited to animal species, and cytolysin-like lectins are found elsewhere than

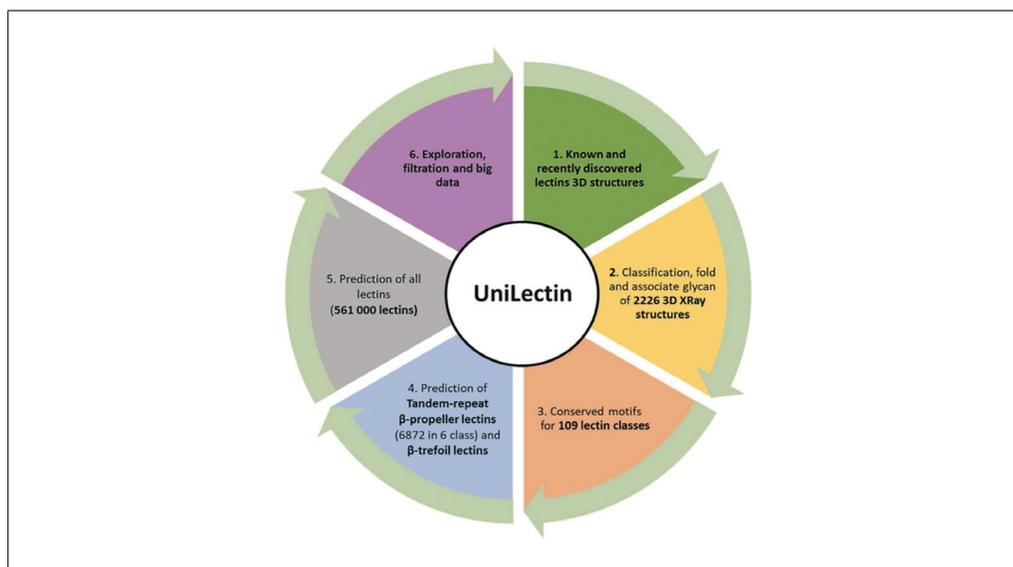


Figure 1 The different phases of development of the UniLectin platform.

fungi. Figure 1 visually describes the chronology of development of the UniLectin platform. UniLectin is available at: <https://www.unilectin.eu/>. The homepage displays the modules, each one being dedicated to a specific aspect of lectins. A simple cross-module search can be launched from the top window; otherwise, each module includes an advanced search tool. UniLectin relies on standard protein cross-references (e.g., UniProt and PDB) as well as on standard representations of glycans. The latter comply with both the IUPAC condensed and the SNFG (Standard Nomenclature For Glycans) simplified notation (Neelamegham et al., 2019). The SNFG symbolism is now popularized in the literature and in the main chemical compound resources. It is implemented in PubChem (Kim et al., 2019) since 2020, and expected in ChEBI in 2022 (Hastings et al., 2016).

In these protocols, usage of the UniLectin3D and LectomeXplore modules is presented.

SEARCHING FOR THE STRUCTURAL DETAILS OF LECTINS BINDING THE O BLOOD GROUP ANTIGEN

BASIC PROTOCOL 1

UniLectin3D is a curated database that classifies and describes 3D structures of lectins by origin and fold, with cross-links to literature and other databases in glycoscience. Among the 2330 entries (sept 2021), more than 60% are complexes between a protein and a carbohydrate ligand, bringing invaluable information on the molecular basis of specificity and affinity. Many lectins recognize glycans present on human tissues, playing a role in cancer, development, immunity, and infection. Structural variations of some of these glycans differentiate individuals, defining for example the basis of the ABO human blood group system (Marionneau et al., 2001). Some pathogens, such as *Vibrio cholerae* or noroviruses, infect sub-populations with given blood group phenotypes more efficiently, and the lectins involved in their recognition have been well studied (Heggelund, Varrot, Imberty, & Krenzel, 2017). This protocol, using bacterial lectins binding the blood group O oligosaccharide (also called the H epitope) as an example, will demonstrate the search function of UniLectin3D for the analysis of the structural details of this interaction.

Necessary Resources

Hardware

Computer with internet connection

Software

Up-to-date web browser such as Chrome or Firefox. Using Safari or Edge may not be as reliable.

Access to UniLectin3D and search interface

1. Go to the UniLectin platform webpage: <https://www.unilectin.eu/>.
2. Select the UniLectin3D module by clicking on the corresponding blue button on the homepage. This action opens <https://www.unilectin.eu/unilectin3D/>. The UniLectin3D landing page, shown in Figure 2, offers different search modes (keywords, fields, or glycan ligands). It also displays visual summaries of the module content in the form of sunbursts enabling data browsing. Finally, it shows a hierarchical chart that can be expanded to detail the underlying lectin classification built on folds branching to classes branching to families. In all cases, these charts are interactive and open new windows for accessing the lectin pages.

Searching UniLectin3D with a glycan-based query for H-type 1 epitope

3. Oligosaccharide structures can be searched through the “Glycan Search” option of the UniLectin3D interface. Clicking on this button opens https://www.unilectin.eu/unilectin3D/glycan_search and displays the table of monosaccharides shown in Figure 3.

Imberty et al.

3 of 14

Figure 2 UniLectin3D homepage. Left: the main searching options and sunbursts for browsing. Right: the clickable classification at the fold level.

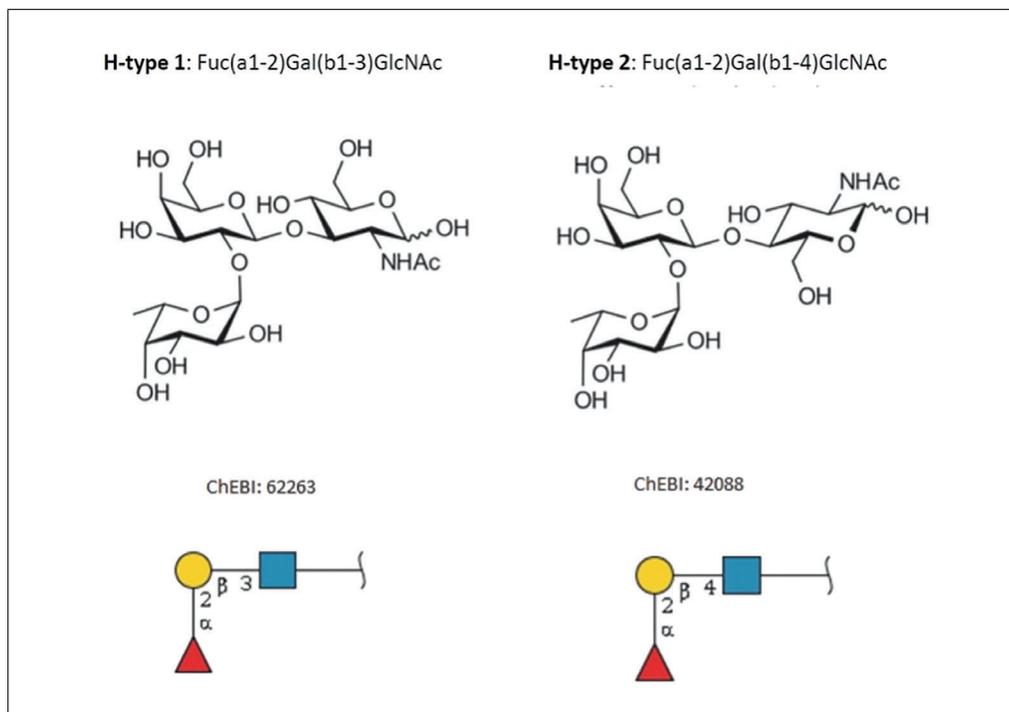


Figure 3 Representation of the two types of blood group O(H) antigens. From top: IUPAC condensed representation, schematic chemical drawing, corresponding ChEBI accession numbers, and SNFG representation. SNFG details: red triangle: fucose; yellow disk: galactose; and blue square: *N*-acetyl-glucosamine. Carbons are numbered from 1 to 6 to specify the linkages between monosaccharides.

4. To select the O(H) blood group antigen, the user needs to know it is a trisaccharide composed of one fucose (*L*-Fuc_p), one galactose (*D*-Gal_p), and one *N*-acetylglucosamine (*D*-GlcNAc_p). Two types of this molecule are known, corresponding to a subtle difference in the linkage between the galactose and the *N*-acetylglucosamine. These are described by IUPAC sequences along with the

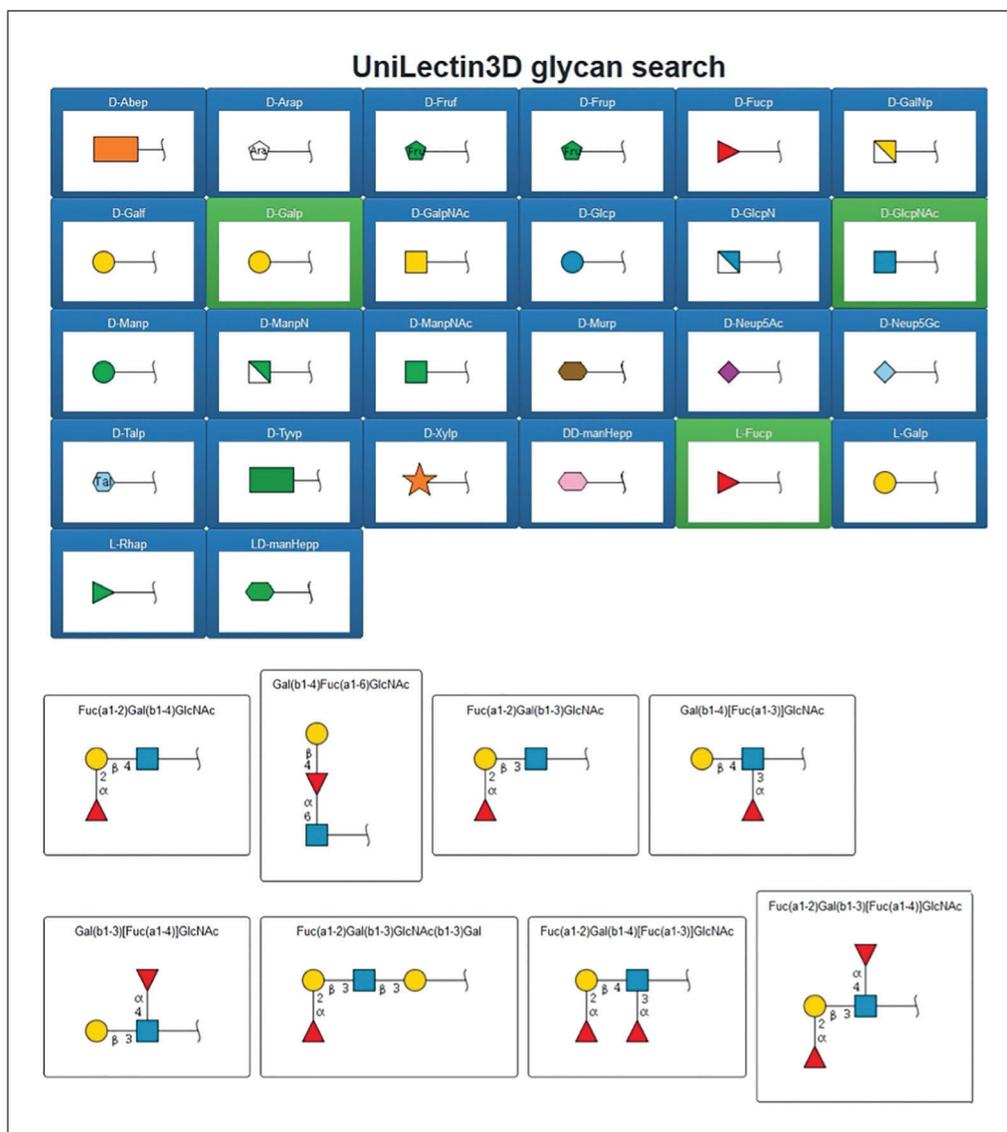


Figure 4 Top: Collection of monosaccharides available for search in the Glycan search option. Successive mouse clicks on the three boxes (L-Fucp, D-Galp and D-GlcNAcp) turns their background from blue to green. Bottom: Excerpt of the output selection of 30 oligosaccharides matching the query. H-Type 1 (Fuc(a1-2)Gal(b1-3)GlcNAc) and H-Type 2 (Fuc(a1-2)Gal(b1-4)GlcNAc) are shown.

matching 2D structures recorded in ChEBI and in the SNFG symbolic notation in Figure 4. With this information, it is then possible to click on either of the three monosaccharides that are part of the blood group O(H) epitope or to simultaneously select the three of them, as displayed in Figure 3 (green boxes). This results in outputting all oligosaccharides containing these three monosaccharides and present in binding sites of lectin structures in UniLectin3D.

- Click on the Fuc(a1-2)Gal(b1-3)GlcNAc motif to display all lectins for which a three-dimensional structure is available in complex with this ligand or with a longer oligosaccharide comprising this epitope. The result by default shows lectins in all species, resulting in a total of 11 entries, two of bacterial and nine of viral origin, mainly noroviruses that attach to blood group O in intestine during infection. The query can be limited to bacteria only by selecting “Bacterial lectins” in the “Origin” box. This search results in two structures from two different species of *Burkholderia*, as represented in Figure 5. These two lectins belong to different classes, one

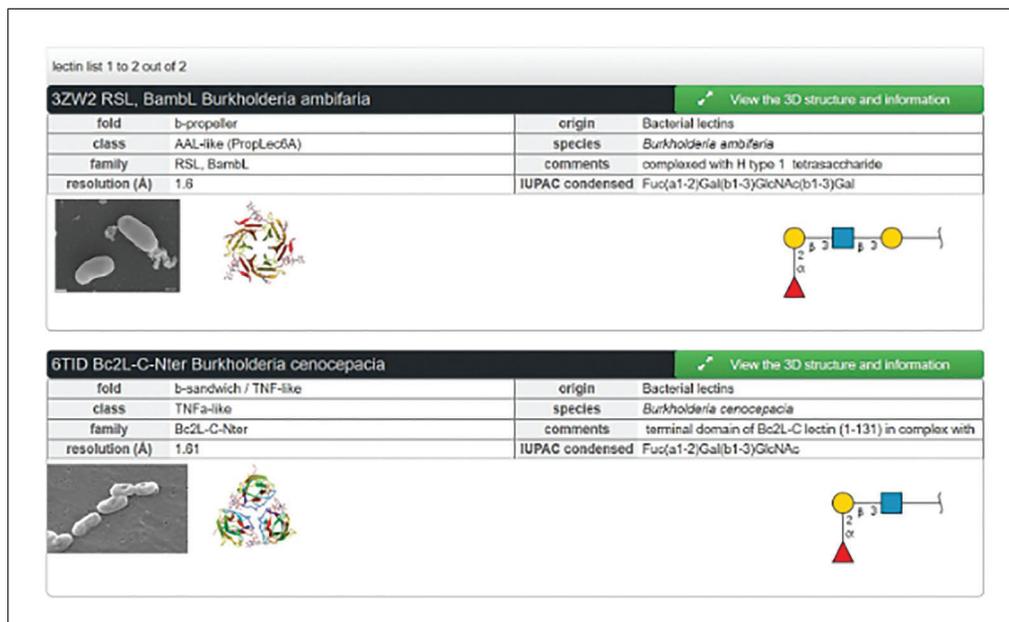


Figure 5 Results of the search based on H-type 1 oligosaccharide in bacterial lectins.

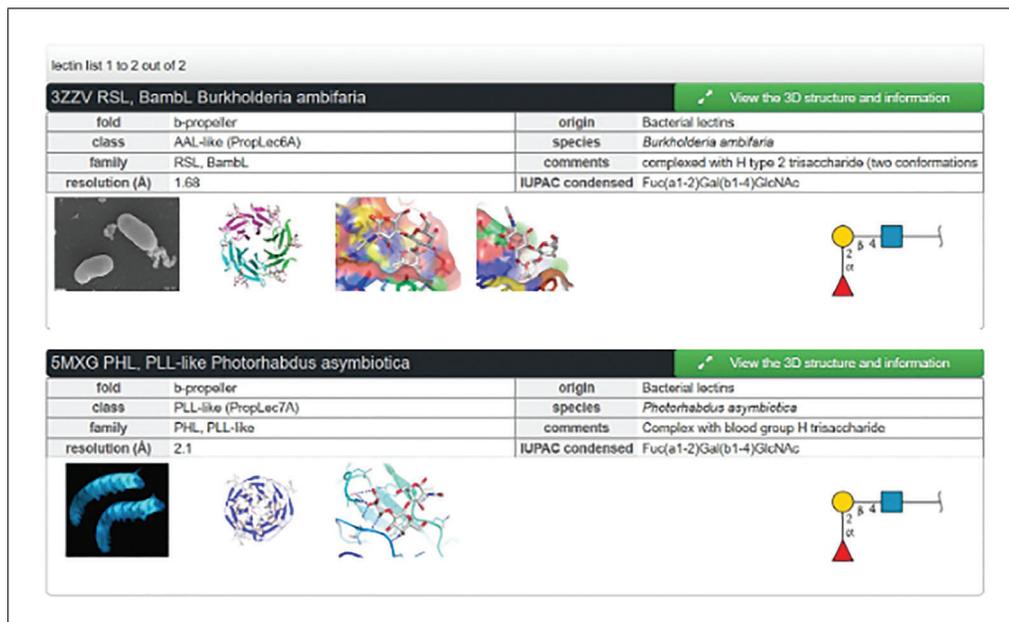


Figure 6 Results of the search based on H-type 2 oligosaccharide in bacterial lectins.

adopting a β -propeller fold (AAL-like class) and the other one a β -sandwich/TNF-like fold (TNF α -like fold). Both are opportunistic bacteria from the *Burkholderia cepacia* complex responsible for lung infection in immunocompromised or cystic fibrosis patients (Mahenthalingam, Baldwin, & Dowson, 2008).

Searching UniLectin3D by field search query for H-type 2 epitope

- For more expert users mastering the IUPAC condensed representation of glycans, a shortcut is exemplified here using the H-type 2 oligosaccharide. In this case, the “Fuc(a1-2)Gal(b1-4)GlcNAc” motif is entered in the “IUPAC condensed” box of the “Search by field” menu and can be directly associated with “Bacterial lectins” in the “Origin” box. As shown in Figure 6, this results in two lectins from two different bacteria, i.e., *Burkholderia ambifaria* (as seen in step 4) and *Phototrhhabdus asymbiotica*, an insect pathogen also able to infect humans (Waterfield, Ciche, & Clarke,

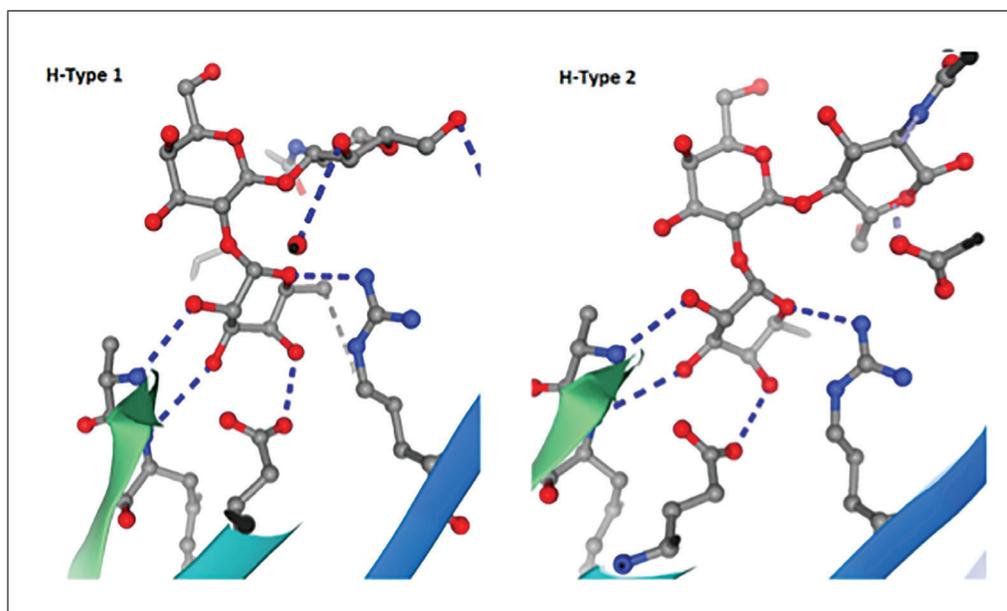


Figure 7 Screen captures of the PLIP window illustrating the basis for the recognition of H-type 1 and H-type 2 epitopes by BamBL.

2009). Both adopt a β -propeller fold, but they belong to different classes, i.e., AAL-like (6-bladed β -propeller) for BamBL and PLL-like (7-bladed β -propeller) for PHL.

Visualization of the structural basis for blood group O recognition by bacterial lectins

7. The previous search results show that the BamBL lectin of *Burkholderia ambifaria* binds both H-type 1 and H-type 2 oligosaccharides. Detailed information on BamBL lectin interacting with the Fuc(a1-2)Gal(b1-3)GlcNAc type 1 ligand can be obtained by clicking on the “View the 3D structure and information” button on the upper right corner of the listed item. This action will open https://unilectin.eu/unilectin3D/display_structure?pdb=3ZW2. Information on BamBL is accessible either by clicking on any green button that links to corresponding entries in UniProt (UniProt, 2021), PDB (European or American sites; Burley et al., 2021), and GlyConnect (Alocchi et al., 2018), or by exploring the structural information displayed on the page with two different viewers. Litemol software (Sehna et al., 2017) is integrated to show the lectin 3D structure, which can be manipulated by holding the mouse to turn the molecule around and, for example, bring the ligand to the fore. PLIP software (Salentin, Schreiber, Haupt, Adasme, & Schroeder, 2015) is integrated to detail atomic interactions between the lectin and the ligand. Mousing over one of the listed interactions on the left updates the view on the right to locate where that particular interaction is located in the structure. In the same way, details of the same BamBL interacting with the Fuc(a1-2)Gal(b1-4)GlcNAc type 2 ligand are found at https://unilectin.eu/unilectin3D/display_structure?pdb=3ZZV, and comparison of both ligands is displayed in Figure 7.

In the end, a user can shape new assumptions on blood group O recognition based on a clearer understanding of specificity. Moreover, structural data can be downloaded for further computational studies.

COMPARING THE LECTOMES OF RELATED ORGANISMS IN DIFFERENT ENVIRONMENTS

LectomeXplore is a module of UniLectin dedicated to the exploration of predicted lectins from each of the 109 classes of UniLectin3D. Translated genomes (proteomes) stored in UniProtKB and RefSeq sequence databases have been screened to generate all putative

lectins (lectome) for each available species. LectomeXplore can be searched by assessing the extent of a class of lectins in the kingdoms of life. In this protocol, we demonstrate how to analyze the lectome of a specified organism and how to compare it with other species of the same genus, but living in different environments. This is applied to investigating the lectomes of two *Pseudomonas* species that have been widely characterized, *P. syringae*, a plant pathogen, and *P. fluorescens*, a ubiquitous bacterium present in soil with much less pathogenic behavior. More specifically, *P. syringae* is a plant pathogen that causes damage in a wide range of plants, for example, with large impact on kiwi fruit production in New Zealand and in Italy (Donati et al., 2020). In contrast, *P. fluorescens* is an environmental bacterium that grows in soil and is in general non-pathogenic. The database also contains strains of *P. cichorii*, a bacterium first isolated on endives and known to cause green midrib rot of greenhouse-grown lettuces (Cottyn et al., 2009).

The protocol also shows how to investigate a specific lectin class common to *Pseudomonas* species. The OAA-like lectins were chosen. These organisms were structurally described in the cyanobacteria *Oscillatoria agardhii* and considered of pharmaceutical significance due to an antiviral activity related to their high affinity for mannose (Koharudin & Gronenborn, 2011), and later characterized in *Burkholderia* and *Pseudomonas* species.

Necessary Resources

Hardware

Computer with internet connection

Software

Up-to-date Web browser such as Chrome or Firefox. Using Safari or Edge may not be as reliable

Access and search LectomeXplore with a query

1. LectomeXplore is accessible via the UniLectin portal via its corresponding button (<https://www.unilectin.eu/predict/>). On the LectomeXplore homepage, click on the “search by field” button. The corresponding interface displays search fields spanning lectin characteristics (Fold, class) or phylogenetic properties (Kingdom, species ...) to be filled individually or in combination (Fig. 8). Note that a second bioinformatics-dedicated box offers the option to query LectomeXplore with protein or protein properties using accession numbers of different databases.

Data mining for species-specific genomes

2. Fill in the species field of the search window to launch the *P. syringae* lectome search, and then click on the “Load predicted lectin” button. This action results in showing that 65 available strains have been predicted and span six different classes of lectins. The same search is then performed replacing the species name by *P. fluorescens*, yielding 76 strains spread across 13 classes. These distinct distributions are shown in Figure 9. This cross-species comparison seems to indicate that in the narrower niche of plant pathogens, the variety of lectins is reduced. To strengthen this statement, a third *Pseudomonas* species is searched by typing *P. cichorii* in the species search field. Only 11 strains and two lectins in two distinct classes are found. In the end, the result of these three searches shows that the lectome of *Pseudomonas* species living in soil with a more saprophytic and opportunistic behavior, such as *P. fluorescens*, is more diverse. It could potentially be involved in adhesion to various surfaces.

Searching for a specific lectin in selected genomes

3. Search can be then focused on a lectin of interest, such as OAA-like lectin that was predicted in both *P. syringae* and *P. fluorescens* lectomes (see Fig. 9). In the search

LectomeXplore field search

Please select a score and optional other criteria

Minimum score (similarity to the reference) 0.25	Superkingdom kingdom (Metazoa;Fungi;Viridiplantae)	Genus
Fold	Phylum (Chordata;Arthropoda;Nematoda) (Actinobacteria;Bacteroidetes) (Firmicutes;Proteobacteria)	Species
Lectin class		exact keyword
Keyword to exclude (word1;word2)		

Bioinformatics detailed criteria Expand

Protein name	Chromosome	Multiple lectin classes
UniProt AC	Pathogen species	
NCBI AC	UniRef cluster representative	
PFAM	POB structures only	
PFAM name		

Figure 8 LectomeXplore query interface.

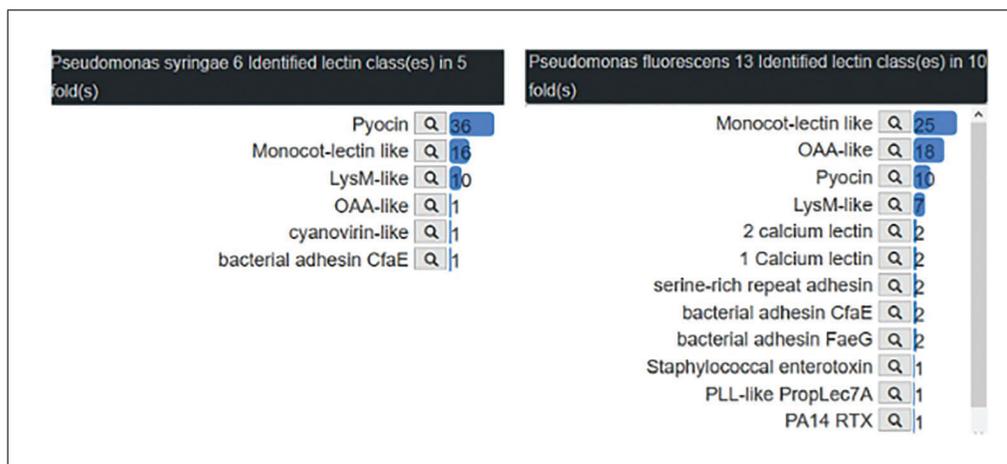


Figure 9 Lectome comparison obtained by entering *Pseudomonas syringae* (left) and *Pseudomonas fluorescens* (right) in the “Species” search box of LectomeXplore.

window, type OOA-like in the “class” field and *Pseudomonas* in the “genus” field. This results in the listing of 90 genes, in exactly 90 *Pseudomonas* genomes in the database, as shown in Figure 10. The monogenic character of this lectin is rather unusual. Interestingly, 20 of these 90 genes are not correctly annotated, and the lectin is not functionally identified in Pfam (Mistry et al., 2021).

Analyzing the quality of the prediction

- As mentioned above, 20 out of 90 lectins are poorly annotated. Clicking on the magnifying glass icon next to “uncharacterized protein” will display all corresponding entry summaries. Then, clicking on the top right corner of the listed *P. putida* entry opens a new page with further information on the alignment of this putative lectin with the reference of the class, here OAA-like. This “Amino acid conservation” also singles out the specific amino acids involved in carbohydrate binding. Figure 11 details this information and shows the match of five out of six binding residues. The sequence logo is another guide for assessing the quality of the alignment. As a matter

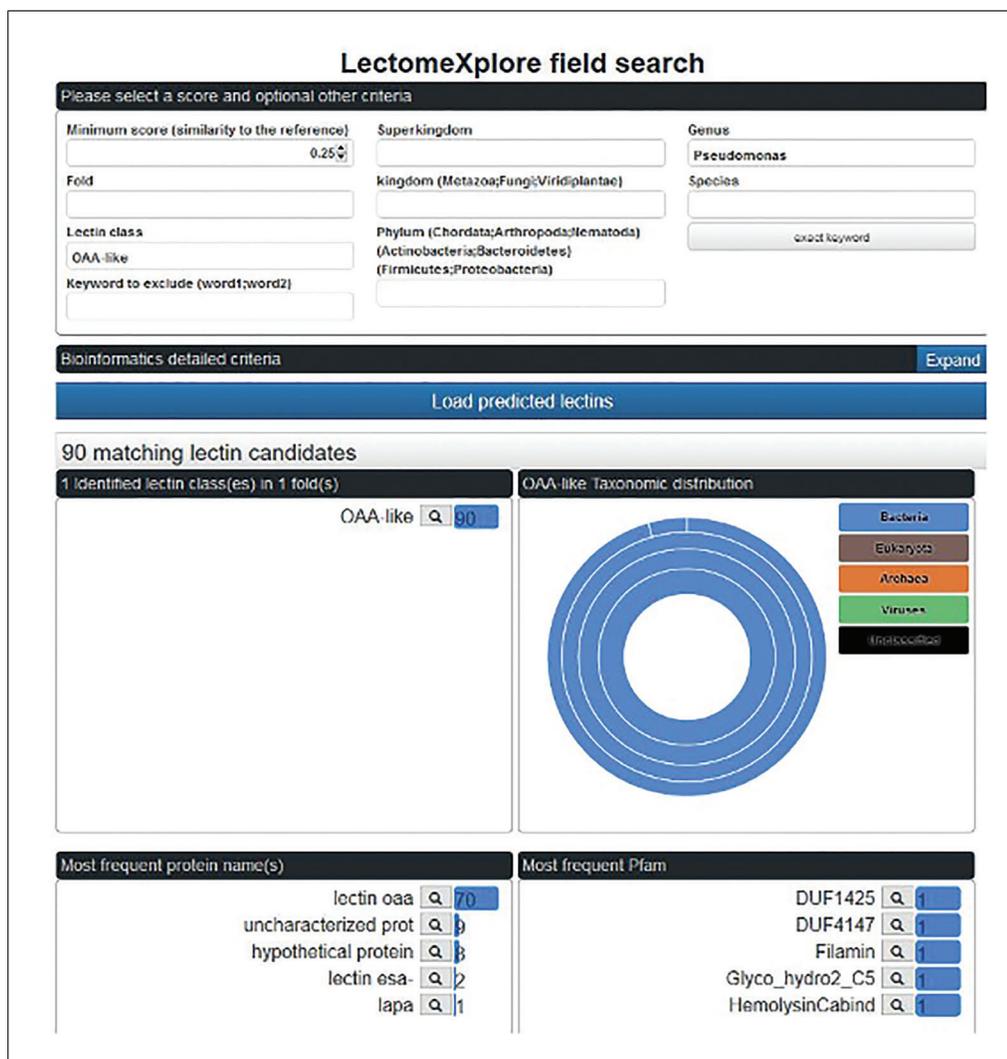


Figure 10 Results of searching for the occurrence of OAA-like lectins in the *Pseudomonas* genus.

of fact, many entries in LectomeXplore are named “uncharacterized,” “putative,” or “hypothetical” proteins (approximately 20% of lectins predicted with a score above 0.25). The present example illustrates how lectins of potential interest can be “mined” in genomes where they were not precisely identified.

GUIDELINES FOR UNDERSTANDING RESULTS

UniLectin was released for the first time in 2019, and unavoidably still faces teething problems in 2021. The content has grown substantially, and issues tend to be specific to each module gradually added to the platform.

The UniLectin3D module is a database of annotated three-dimensional structures of lectins, and, as such, strongly depends on information stored in the PDB. In this context, users should be aware that while crystallography is the main method for structure determination, oligosaccharides are not always treated with the same care as their protein counterparts. Frequent errors in ring shapes and glycosidic linkage conformation have been spotted (Agirre, Davies, Wilson, & Cowtan, 2015). However, cooperation between glycoscientists and the PDB has been fruitful, and recent structures have been subjected to more stringent controls.

The LectomeXplore module is destined to evolve rapidly. For the time being, lectome prediction is based on 109 classes reflecting currently available structural information.



Figure 11 Entry page of the OAA-like lectin predicted to occur in the *P. putida* genome.

Consequently, prediction misses potential motifs that are present in lectins that have not been structurally characterized. The number of classes should grow with the inclusion of new glycan-binding information that will undoubtedly be generated in the years to come. Furthermore, the identification of putative lectins is based on HMM models, and therefore on sequence similarity over the whole domain of interest, but the definition of this similarity may be hampered by the frequent occurrence of tandem repeats, i.e., repeats of more or less conserved stretches of 30–60 amino acids. This situation occurs in several families of lectins in which multi-valency is achieved by tandem repeats, probably through a gene duplication process (Notova, Bonnardel, Lisacek, Varrot, & Imberty, 2020). This problem was overcome by designing HMM motifs based on aligning individual repeats in the lectins of interest, and resulted in creating two modules: PropLec that covered β -propellers, the prediction of which supported the identification of a novel architecture (Bonnardel et al., 2019), and TrefLec dedicated to β -trefoil lectin prediction.

Finally, motif detection is run in parallel for each class and the highest score is kept. In this way, similarity reflected in high HMM scores does ensure that the predicted and target proteins share the same fold, but not necessarily the same function, i.e., glycan binding ability and specificity. This can result in the prediction of a false positive, especially in the case of ubiquitous folds occurring in some lectins and many other proteins, such

as the immunoglobulin domain. In such cases, the conservation of not only the whole domain sequence, but also of the amino acids involved in carbohydrate binding, should be carefully considered.

COMMENTARY

Background Information

The UniLectin platform was designed to address the data integration and classification issues encountered by researchers in glycobiology. Several other databases had been previously developed. The Lectin Frontier Database collecting 400 lectins with curated information (Hirabayashi, Tateno, Shikanai, Aoki-Kinoshita, & Narimatsu, 2015) is now integrated in GlyCosmos, a newly developed portal for accessing and exploring knowledge in glycobiology (Yamada et al., 2020). LectinDB is focused on plant lectins (Chandra et al., 2006). SugarBindDB describes pathogen lectins and their corresponding glycan targets (Mariethoz et al., 2016). Most recently, ProCarbDB was released with more than 5000 3D X-ray crystal structures of protein-carbohydrate complexes and annotated carbohydrate ligands (Copoiu, Torres, Ascher, Blundell, & Malhotra, 2020). Most of these lectin-dedicated databases only partially cover the topic, and, in general, are not based on a solid classification. Lastly, they do not necessarily comply with the latest standards for glycan nomenclature and visualization.

We recently demonstrated the efficiency of LectomeXplore for exploring the genomes of organisms in relation to biological questions. The analysis of the vaginal microbiome and its variations that impact women's health demonstrated that vaginal bacterial species associated with infection and inflammation produce a larger variety of lectins (Bonnardel et al., 2021). The size of the predicted lectome and the variety of putative specificities correlated with pathogenicity. The UniLectin portal also contains the MycoLec module, a publicly available and searchable database for fungal lectomes, based on genomes available in the JGI MycoCosm (Grigoriev et al., 2014) and NCBI (Sayers et al., 2020) databases. The lectome content of this module led to

correlation of phylogeny and the very diverse lifestyles of fungi, from saprophytes to symbionts and pathogens (Lebreton et al., 2021).

Curated and predicted lectin data are now available to the community and have already fed several machine-learning approaches to detect the structural basis of specificity or novel lectin function. In particular, the bulk data of UniLectin3D was recently used in a comprehensive analysis of lectin-glycan interactions aiming at revealing the determinants of lectin specificity (Mattox & Bailey-Kellogg). LectinOracle combined a powerful glycan-specific deep-learning engine (Burkholz, Quackenbush, & Bojar, 2021) with a transformer-based protein sequence language model to predict glycan-protein interactions (Lundström, Korhonen, Lisacek, & Bojar, 2021). The sequence language model was built from UniLectin3D and LectomeXplore.

Critical Parameters

Except for UniLectin3D, modules of UniLectin were created with HMM profiles for each specific class in each context. A critical parameter is therefore the prediction score, which involves the HMM p -value ($<10^{-2}$) and the evaluation of amino acid sequence alignment generated by HMMER during the search. More specifically, at each position of the alignment, a cumulative counter is incremented depending on amino acid similarity to set the basis of a normalized score (see Bonnardel et al., 2021, for details). Predictions above 0.5 are highly reliable. Nonetheless, predicted sequences with a score between 0.20 and 0.50 can reveal unknown lectins. Further checks are needed, such as the cautious examination of other potentially present domains and amino acid conservation at binding residue positions.

Table 1 Sources and Solutions to Potential Errors

Problem	Possible cause	Solution
Dysfunctional structure viewers (LiteMol or PLIP)	Desynchronized software updates	Please let us know; our helpdesk is responsive
Unresponsive cross-reference buttons	Desynchronized database releases	Please let us know; our helpdesk is responsive

Troubleshooting

Table 1 lists potential errors that may arise with the protocols in this article, along with their potential sources and solutions.

Acknowledgments

The authors acknowledge support by the ANR PIA Glyco@Alps (ANR-15-IDEX-02), the Alliance Campus Rhodanien Co-funds (<http://campusrhodanien.unige-cofunds.ch>), and the Labex Arcane/CBH-EUR-GS (ANR-17-EURE-0003).

Author Contributions

Anne Imberty: conceptualization, data curation, funding acquisition, investigation, methodology, resources, supervision, validation, writing original draft; **François Bonnardel**: methodology, resources, software, visualization; **Frédérique Lisacek**: conceptualization, formal analysis, funding acquisition, methodology, resources, supervision, validation, writing original draft.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

These data were derived from the following resources available in the public domain: <https://www.unilectin.eu/>.

Literature Cited

- Agirre, J., Davies, G., Wilson, K., & Cowtan, K. (2015). Carbohydrate anomalies in the PDB. *Nature Chemical Biology*, *11*, 303. doi: 10.1038/nchembio.1798.
- Alocchi, D., Ghraichy, M., Barletta, E., Gastaldello, A., Mariethoz, J., & Lisacek, F. (2018). Understanding the glycome: An interactive view of glycosylation from glyco compositions to glycoepitopes. *Glycobiology*, *28*, 349–362. doi: 10.1093/glycob/cwy019.
- Bonnardel, F., Haslam, S. M., Dell, A., Feizi, T., Liu, Y., Tajadura-Ortega, V., ... Imberty, A. (2021). Proteome-wide prediction of bacterial carbohydrate-binding proteins as a tool for understanding commensal and pathogen colonisation of the vaginal microbiome. *Biofilms and Microbiomes*, *7*, 49. doi: 10.1038/s41522-021-00220-9.
- Bonnardel, F., Kumar, A., Wimmerova, M., Lahmann, M., Perez, S., Varrot, A., ... Imberty, A. (2019). Architecture and evolution of blade assembly in β -propeller lectins. *Structure*, *27*, 764–775. doi: 10.1016/j.str.2019.02.002.
- Bonnardel, F., Mariethoz, J., Perez, S., Imberty, A., & Lisacek, F. (2021). LectomeXplore, an update of UniLectin for the discovery of carbohydrate-binding proteins based on a new lectin classification. *Nucleic Acids Research*, *49*, D1548–D54. doi: 10.1093/nar/gkaa1019.
- Bonnardel, F., Mariethoz, J., Salentin, S., Robin, X., Schroeder, M., Perez, S., ... Imberty, A. (2019). UniLectin3D, a database of carbohydrate binding proteins with curated information on 3D structures and interacting ligands. *Nucleic Acids Research*, *47*, D1236–D44. doi: 10.1093/nar/gky832.
- Bonnardel, F., Perez, S., Lisacek, F., & Imberty, A. (2020). Structural database for lectins and the UniLectin web platform. *Methods in Molecular Biology*, *2132*, 1–14. doi: 10.1007/978-1-0716-0430-4_1.
- Burkholz, R., Quackenbush, J., & Bojar, D. (2021). Using graph convolutional neural networks to learn a representation for glycans. *Cell Reports*, *35*, 109251. doi: 10.1016/j.celrep.2021.109251.
- Burley, S. K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G. V., ... Zhuravleva, M. (2021). RCSB Protein Data Bank: Powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Research*, *49*, D437–D51. doi: 10.1093/nar/gkaa1038.
- Chandra, N. R., Kumar, N., Jeyakani, J., Singh, D. D., Gowda, S. B., & Prathima, M. N. (2006). Lectindb: A plant lectin database. *Glycobiology*, *16*, 938–946. doi: 10.1093/glycob/cwl012.
- Copoiu, L., Torres, P. H. M., Ascher, D. B., Blundell, T. L., & Malhotra, S. (2020). ProCarbDB: A database of carbohydrate-binding proteins. *Nucleic Acids Research*, *48*, D368–D75. doi: 10.1093/nar/gkz860.
- Cottyn, B., Heylen, K., Heyrman, J., Vanhouteghem, K., Pauwelyn, E., Bleyaert, P., ... Maes, M. (2009). *Pseudomonas cichorii* as the causal agent of midrib rot, an emerging disease of greenhouse-grown butterhead lettuce in Flanders. *Systematic and Applied Microbiology*, *32*, 211–225.
- Donati, I., Cellini, A., Sangiorgio, D., Vanneste, J. L., Scortichini, M., Balestra, G. M., & Spinelli, F. (2020). *Pseudomonas syringae* pv. actinidiae: Ecology, infection dynamics and disease epidemiology. *Microbial Ecology*, *80*, 81–102. doi: 10.1007/s00248-019-01459-8.
- Gabius, H. J., Siebert, H. C., Andre, S., Jimenez-Barbero, J., & Rudiger, H. (2004). Chemical biology of the sugar code. *Chembiochem*, *5*, 740–764. doi: 10.1002/cbic.200300753.
- Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otilar, R., ... Shabalov, I. (2014). MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Research*, *42*, D699–704. doi: 10.1093/nar/gkt1183.
- Hastings, J., Owen, G., Dekker, A., Ennis, M., Kale, N., Muthukrishnan, V., ... Steinbeck, C. (2016). ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Research*, *44*, D1214–9. doi: 10.1093/nar/gkv1031.
- Heggelund, J. E., Varrot, A., Imberty, A., & Krenzel, U. (2017). Histo-blood group antigens as

- mediators of infections. *Current Opinion in Structural Biology*, *44*, 190–200. doi: 10.1016/j.sbi.2017.04.001.
- Hirabayashi, J., Tatenno, H., Shikanai, T., Aoki-Kinoshita, K. F., & Narimatsu, H. (2015). The Lectin Frontier Database (LfDB), and data generation based on frontal affinity chromatography. *Molecules*, *20*, 951–973. doi: 10.3390/molecules20010951.
- Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., ... Bolton, E. E. (2019). PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Research*, *47*, D1102–D09. doi: 10.1093/nar/gky1033.
- Koharudin, L. M., & Gronenborn, A. M. (2011). Structural basis of the anti-HIV activity of the cyanobacterial *Oscillatoria agardhii* agglutinin. *Structure*, *19*, 1170–1181. doi: 10.1016/j.str.2011.05.010.
- Lebreton, A., Bonnardel, F., Dai, Y.-C., Imbert, A., Martin, F. M., & Lisacek, F. (2021). Comprehensive phylogenetic and bioinformatics survey of lectins in the fungal kingdom. *Journal of Fungi*, *7*, 453. doi: 10.3390/jof7060453.
- Lundstrøm, J., Korhonen, E., Lisacek, F., & Bojar, D. (2021). LectinOracle—A generalizable deep learning model for lectin-glycan binding prediction. *BioRxiv*, doi: 10.1101/2021.08.30.458147.
- Mahenthiralingam, E., Baldwin, A., & Dowson, C. G. (2008). *Burkholderia cepacia* complex bacteria: Opportunistic pathogens with important natural biology. *Journal of Applied Microbiology*, *104*, 1539–1551. doi: 10.1111/j.1365-2672.2007.03706.x.
- Mariethoz, J., Khatib, K., Alocci, D., Campbell, M. P., Karlsson, N. G., Packer, N. H., ... Lisacek, F. (2016). SugarBindDB, a resource of glycan-mediated host-pathogen interactions. *Nucleic Acids Research*, *44*, D1243–50. doi: 10.1093/nar/gkv1247.
- Marionneau, S., Cailleau-Thomas, A., Rocher, J., le Moullac-Vaidye, B., Ruvoen, N., Clement, M., & le Pendu, J. (2001). ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. *Biochimie*, *83*, 565–573. doi: 10.1016/s0300-9084(01)01321-9.
- Mattox, D. E., & Bailey-Kellogg, C. Comprehensive analysis of lectin-glycan interactions reveals determinants of lectin specificity. *PLoS Computational Biology*, *17*(10), e1009470. doi: 10.1371/journal.pcbi.1009470.
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., ... Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research*, *49*, D412–D19. doi: 10.1093/nar/gkaa913.
- Neelamegham, S., Aoki-Kinoshita, K., Bolton, E., Frank, M., Lisacek, F., Lutteke, T., ... Group, S. D. (2019). Updates to the symbol nomenclature for glycans guidelines. *Glycobiology*, *29*, 620–624. doi: 10.1093/glycob/cwz045.
- Notova, S., Bonnardel, F., Lisacek, F., Varrot, A., & Imbert, A. (2020). Structure and engineering of tandem repeat lectins. *Current Opinion in Structural Biology*, *62*, 39–47. doi: 10.1016/j.sbi.2019.11.006.
- Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F., & Schroeder, M. (2015). PLIP: Fully automated protein-ligand interaction profiler. *Nucleic Acids Research*, *43*, W443–7. doi: 10.1093/nar/gkv315.
- Sayers, E. W., Beck, J., Brister, J. R., Bolton, E. E., Canese, K., Comeau, D. C., ... Ostell, J. (2020). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *48*, D9–D16. doi: 10.1093/nar/gkz899.
- Sehnal, D., Deshpande, M., Varekova, R. S., Mir, S., Berka, K., Midlik, A., ... Koca, J. (2017). LiteMol suite: Interactive web-based visualization of large-scale macromolecular structure data. *Nature Methods*, *14*, 1121–1122. doi: 10.1038/nmeth.4499.
- UniProt, C. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*, D480–D89. doi: 10.1093/nar/gkaa1100.
- Waterfield, N. R., Ciche, T., & Clarke, D. (2009). Photorhabdus and a host of hosts. *Annual Review of Microbiology*, *63*, 557–574. doi: 10.1146/annurev.micro.091208.073507.
- Yamada, I., Shiota, M., Shinmachi, D., Ono, T., Tsuchiya, S., Hosoda, M., ... Aoki-Kinoshita, K. F. (2020). The GlyCosmos Portal: A unified and comprehensive web resource for the glycosciences. *Nature Methods*, *17*, 649–650. doi: 10.1038/s41592-020-0879-8.