BBA Advances 7 (2025) 100154

Contents lists available at ScienceDirect

BBA Advances

journal homepage: www.sciencedirect.com/journal/bba-advances

Tools for structural lectinomics: From structures to lectomes

Frédérique Lisacek [a,b,*], Boris Schnider [a,b], Anne Imberty [c,*]

[a] SIB Swiss Institute of Bioinformatics CH-1227 Geneva, Switzerland
[b] Computer Science Department, UniGe CH-1227 Geneva, Switzerland
[c] Univ. Grenoble Alpes, CNRS, CERMAV 38000 Grenoble, France

ARTICLE INFO

Keywords:
Structural biology
Carbohydrate-binding
Lectin
Bioinformatics
Modelling
Prediction method
AI

ABSTRACT

Lectins are ubiquitous proteins that interact with glycans in a variety of molecular processes and as such, also play a role in diseases, whether infectious, chronic or cancer-related. The systematic study of lectins is therefore essential, in particular for understanding cell-cell communication. Accumulated protein three-dimensional structural data in the past decades boosted advance in AI-based prediction and opened up new options to characterise lectins that are known to often be multimeric and multivalent. This article reviews the methods to obtain structures of lectins, the current data available for lectin 3D structures and their interactions, how this knowledge is used to classify these proteins and shows that the combination of an array of bioinformatics tools should make the prediction of binding specificity possible in a near future.

1. Introduction

The *glycocode*, built by the complexity of glycan structures and conformations at the surface of cells, is deciphered by specific proteins able to bind to glycans [1]. Among them, lectins recognize specific structures of carbohydrates, and are different from glyco-enzymes, transporters, antibodies or glycosaminoglycan-binding proteins [2]. Lectins are present in all kingdoms of life. They can be soluble or anchored in cell membranes, presented on bacterial appendices or on virus capsids [3].

Lectins play a role in various biological functions. Their interactions with specific glycan structures are involved in quality control during protein folding and sorting, cell adhesion and trafficking, innate immunity, mammalian fertilization and embryo development among others [4]. They also play a role in many diseases, being involved in recognition of altered glycan structures related to cancer or other chronic diseases. Furthermore, many microbes, such as viruses, bacteria, parasites, and fungi, use lectins for specific recognition of glycans and adhesion on target tissues during the first step of infection or of symbiosis [5].

The glycan-binding properties of lectins makes them relevant as tools for biotechnology, with research application in cell biology, immunology, nanotechnology and pharmacology. They can be used for structural analysis of glycoconjugates, labelling of tissues or detection of pathogens. Lectins can also be used for diagnosis or even treatment in some human diseases [6]. This fosters the development of recombinant lectins, with quality-controlled production as well as engineered lectins with modified properties of affinity and specificity for specialised applications.

Lectins often exist in the form of well-defined modules or carbohydrate recognition domains (CRD). CRDs occurs as single protein, repetitive domain in tandem proteins, or as domains within larger proteins containing other modules/segments [7,8]. When a lectin is simply a CRD, it is generally functional upon oligomerization, resulting in the multivalent presentation of several CRDs, with high avidity for multiple copies of glycans presented on complex glycoconjugates or cell surfaces [9].

In this context, structural knowledge of lectins is of primary importance at several levels. Atomic details of the interaction between glycan ligands and lectin binding sites help in rationalizing the lectin function as well as related properties in terms of affinity and specificity. This allows for knowledge-based design of glycomimetics, i.e. high affinity synthetic compounds able to compete with natural interaction of lectins with human tissues [10].

"Functional lectinomics" was proposed a long time ago as a generic term to cover the global study of lectins [11]. In reference to this terminology but with a different angle, this review surveys the methods at hand to determine the three-dimensional structural properties of lectins and briefly covers the structural aspect of their interaction with glycans. It then details how structural features can support the classification of

* Corresponding authors.
E-mail addresses: Frederique.lisacek@sib.swiss (F. Lisacek), anne.imberty@cermav.cnrs.fr (A. Imberty).

https://doi.org/10.1016/j.bbadva.2025.100154
Received 31 December 2024; Received in revised form 24 February 2025; Accepted 5 March 2025
Available online 6 March 2025

lectins and in turn, how this classification sets the basis of automatic methods predicting species-specific lectomes and further supports the design of AI-driven specificity assignment.

2. Tools for structural determination of lectins

Lectin three dimensional structures can be retrieved from different sources, including the Protein Databank (PDB) [12], but also from more specialized databases, such as UniLectin3D [13] that is described below. The vast majority, i.e., 97 %, of the 2643 structures recorded in UniLectin3D, were solved by classical X-ray crystallography, as shown in Table 1, in which the origin of data is detailed.

2.1. X-ray crystallography

The 2566 X-ray crystallography structures are in most cases, refined at high resolution, except for 600 with resolution lower than 2.3 Å. Complexes between lectins and carbohydrate ligands can be obtained by co-crystallization or soaking, but most structures are solved with short oligosaccharides. If the carbohydrate chains were longer, it would create conformational disorders that is an obstacle to crystallisation. Crystals are in general collected at very low temperature to protect the protein from radiation damage. Room temperature data collection is sometimes preferred to obtain results closer to native conformation of side chains and to avoid the use of glycerol, a cryoprotectant that may compete for glycans in binding sites.

2.2. Nuclear magnetic resonance

Nuclear Magnetic Resonance (NMR) is a very useful approach for the determination of structures of proteins and glycan–protein complexes in solution. NMR has been successfully applied, first to solve the structures of small lectins consisting of short peptides, such as hevein, or trefoil factors [14,15]. However, larger domains, such as Siglec 8, have also been investigated by NMR, deciphering the modes of interaction with sulfated and sialylated Lewis structures [16]. Altogether 45 structures from 31 different lectins have been determined by NMR in UniLectin3D. While complete structure determination requires uniform isotopic labelling, other structure-related information can be obtained through the vast panel of experimental parameters that can be measured by NMR (chemical shifts, couplings, Nuclear Overhauser Effects (NOEs), relaxation rates…). They reflect ligand conformation, dynamics, affinity and contact to the binding site [17].

2.3. Neutron crystallography

Neutron crystallography provides highly relevant features for structural biology and more specifically for characterising protein-carbohydrate interaction [18]. The localisation of hydrogen atoms, when exchanged to deuterium, is clearly revealed in neutron density maps providing valuable information on hydrogen bond geometries, protonation state, and orientation of water molecules. Since neutrons do not induce radiation damage to the crystal, diffraction data can be collected at room temperature, avoiding the artefact of very low temperature and cryoprotectant. However, two technical limitations, i.e. deuteration of the protein and the ligands, and the need for large crystals, currently restrain the number of lectin neutron structures. UniLectin3D only contains seven structures corresponding to four distinct proteins. Recently, the use of *E. coli* strains engineered to produce monosaccharides allowed for perdeuterated fucose [19] and the neutron structure of its complex with LecB from *Pseudomonas aeruginosa* [20]. These data help rationalizing how calcium ions induce unusual types of hydrogen bonds and an infrequent strong affinity for monosaccharide (Fig. 1A).

Table 1
Origin of UniLectin3D content.

Methods	Structures	Distinct lectins
X-ray crystallography	2566	689
NMR	45	31
Neutron crystallography	7	4
Cryo-EM	25	17
Total	2643	741

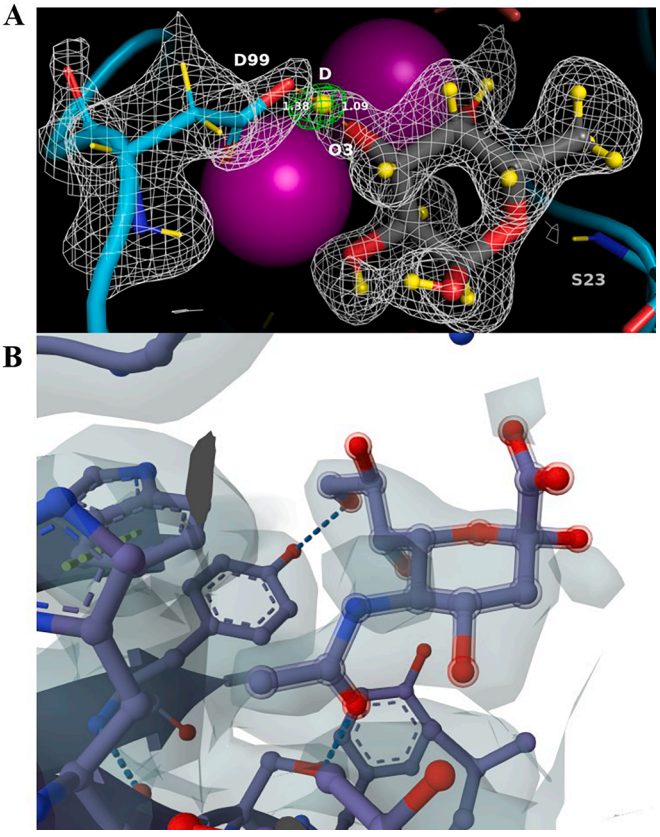

Fig. 1. Non-classical map of carbohydrates in lectin binding site. (A) neutron density maps around fucose in the binding site of LecB from neutron crystallography (PDB code 7PRG). (B) electron density map around sialic acid in the binding site of SARS-Cov2 spike protein from cryo EM (PDB code 7QUR)..

2.4. Cryo-electron microscopy

Cryo-Electron Microscopy (EM) is a very effective and fast developing technique. Direct electron detectors generate structural information while avoiding the constraints of labelling or crystallizing biological macromolecules. It is particularly suited to glycobiology, as samples with structural and conformational heterogeneity can be observed. Cryo-EM analyses of fully glycosylated spike protein of the feline alphacoronavirus led to model 27 out of the 33 identified N-glycans [21]. With progress in the quality of the resolution, it is now possible to also analyse carbohydrates in the binding site of large proteins that are difficult to crystallize. UniLectin3D currently includes 25 structures of lectins obtained by cryo-EM. Among them, the only structure of SARS-Cov2 spike glycoprotein in complex with host glycans was obtained by cryo-EM and revealed the presence of sialic acid in the binding site of the NTD domain [22] (Fig. 1B).

3. Lectins structures and basis of glycan recognition

3.1. Brief overview of lectin structures

The functional properties of lectins are structurally contained in their CRDs. Such domains have a wide variety of folds, indicating a broad evolutionary distribution [23]. The architecture of lectins generates another level of complexity: CRDs are found as single domains (merolectins), oligomers (hololectins), or associated with domains of different functions (chimerolectins) [24]. This large structural repertoire has been reviewed for plant [24,25], animal [26,27], or fungal [28] lectins. Lectins of microorganisms have also been structurally determined, with an emphasis on those placing an infectious burden on humans [5,29]. Even though CRDs adopt a large variety of folds, some structural families have been more studied, and a wide panel of structural data are available for C-type lectins [30], R-type lectins [31], galectins [32], I-type [33] and L-type [34] lectins. It should be noted that some canonical structures are shared in different branches of life, such as the L-type CRDs occurring as legume lectins in plant vacuoles and membranes [35,36], but also involved in eukaryotic cells for quality control of glycoproteins synthesis [37]. Lectin structures are not comprehensively reviewed in this article but databases for accessing the corresponding information are described in section 3.

3.2. Intermolecular interactions

Lectins have carbohydrate-binding sites that are generally shallow pockets and do not undergo significant conformational changes upon ligand-binding (Fig. 2). These proteins achieve specificity for carbohydrates through the recognition of stereospecific arrangement of glycan hydroxyl groups, which is dependent on ring shape, chiral center, and linkage sites [23]. Targeted modifications of hydroxyl by sulfation, phosphorylation or N-acetylation are also accommodated in lectin binding sites, thereby additionally contributing to specificity [16].

Since lectin sites are generally solvent-accessible, carbohydrate ligands have to compete with structured water molecules that occupy the same positions as the hydroxyl groups of sugars [38,39]. Upon binding, these hydroxyl groups are involved in hydrogen bonds with polar amino acids that favour the creation of bifurcated bonds, such as asparagine (Asn) and aspartate (Asp), glutamine (Gln) and glutamate (Glu), as well as arginine (Arg). Acidic groups of sialic acids also establish salt bridges with positively charged Arg and lysine (Lys). In some cases, water molecules play a bridging role between the carbohydrate and the binding site, establishing hydrogen bonds between both the sugar and the amino acids [40,41].

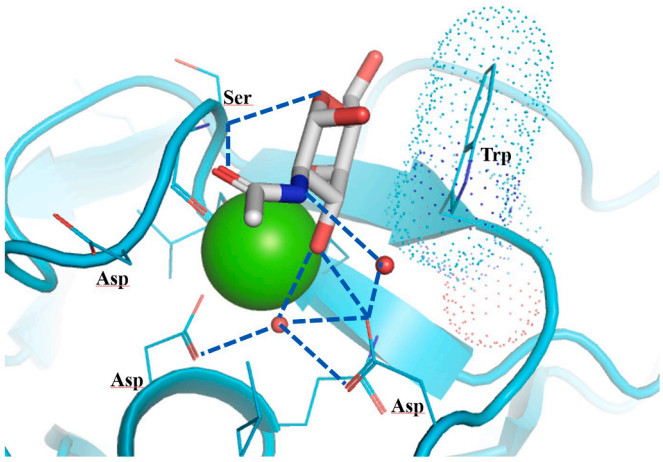

Fig. 2. Interactions between lectins and carbohydrates illustrated by binding of GalNAc to hemolytic lectin CEL-III (PDB code 2Z48). Interactions between GalNAc (represented by sticks) and lectin (represented by ribbon) include direct and water-bridged hydrogen bonds (represented as blue dotted lines), coordination of calcium ion (represented as green sphere) and π-stacking to a Trp residue.

The systematic analysis of amino acids in lectin binding sites demonstrated the predominant enrichment of Asn and Asp among the polar residues [42]. The same study however concludes in a much stronger enrichment of aromatic residues, with preference for tryptophan (Trp). Indeed, Trp, and to a lesser extent tyrosine (Tyr), establish CH-π interactions with the hydrophobic face of monosaccharides [43]. The binding energy resulting from CH-π interactions, depends on the stereochemistry of the monosaccharide. It is especially favoured for β-galactoside, but also observed for mannose, glucose and derivatives [44].

3.3. Metal ions and carbohydrate binding site: the role of calcium

Some lectins are known to require the presence of divalent ions in the assay medium for activity and their binding can be inhibited by chelators such as Ethylene Diamine Tetra Acetic acid (EDTA). Structural data confirm the presence of different ions in several families of lectins. For example, in the large family of legume lectins, that encompasses Concanavalin A, the binding site is conserved and requires the presence of two ions, calcium (Ca^{2+}) and manganese (Mn^{2+}) [35]. Legume lectins can be further classified in subgroups reflecting different binding specificities, ranging from mannose (Man), galactose (Gal) and N-acetyl galactosamine (GalNAc), N-acetyl glucosamine (GlcNAc), fucose (Fuc) or sialic acid (NeuAc) [36]. Nonetheless, the two Ca and Mn ions occur in all crystal structures. They do not interact directly with the bound carbohydrate, but play a structural role by stabilizing a rare cis-peptide linkage and the conformation of the adjacent loop [45].

Other families of lectins from animals [46], fungi [47] and bacterial [48,49] lectins have calcium ions intimately involved in the glycan binding site, directly bridging the sugar with (mostly) acidic side chains of the protein. The glycan is then directly involved in the coordination sphere of calcium, occupying two sites through vicinal hydroxyl groups. The cation therefore contributes to the specificity but also to the affinity, both through electrostatic interactions and through the release of previously calcium-bound water molecules.

The first crystal structure of a calcium-dependent lectin was rat mannose-binding protein complexed with an oligomannose [50] and this animal lectin family is still referred to as C-type lectin [51]. This lectin class now spans more than 50 different lectins with known structures, and specificity varies from Gal, N-acetyl galactosamine (GalNAc), GlcNAc or fucose (Fuc), depending on neighbouring amino acids.

Several other calcium-dependent lectin families have been identified. They are characterised with different folds and binding sites (Fig. 3). Calcium-containing binding appears to have evolved independently in several branches of life, to converge to a very efficient way of binding sugars. In addition to classical coordination of two vicinal glycan hydroxyl groups by one calcium ion, LecB from *Pseudomonas aeruginosa* and related lectins from Pseudomonota are characterized by

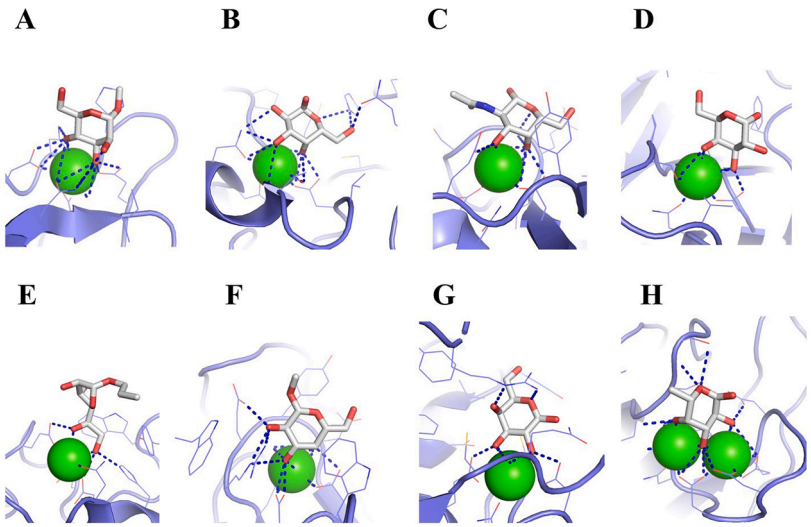

Fig. 3. Graphical representation of eight different calcium-dependent carbohydrate-binding sites found in crystal structures of lectins. (A) Human MPB-A complexed with mannoside (1KWU), (B) *Pseudomonas aeruginosa* LecA complexed with galactose (1OKO), (C) sea cucumber CEL-III complexed with GalNAc (2Z48), (D) oyster mushroom lectin complexed with galactose (6T1D), (E) human intelectin-1 complexed with galactofuranoside (4WMY), (F) *Candida glabrata* adhesin (4A3X) and (G) sea anemone lectin (6A56) complexed with galactose and (H) *P. aeruginosa* LecB complexed with fucose (1GZT).

the occurrence of two calcium ions in the carbohydrate binding site. The two calcium ions are spatially very close, resulting in unique coordination of the three hydroxyl groups of fucose or mannose [49] and energy-rich low barrier hydrogen bond [20].

3.4. Glycan conformation and lectin binding

Lectins are generally rather rigid proteins, with pre-organized binding sites that do not undergo major conformational change upon binding, although allostery has been observed within the C-type family lectins [52]. In contrast, glycans have complex conformational behaviour. While monosaccharides are generally considered as rigid blocks, the glycosidic linkage is semi-flexible. In turn, disaccharides can adopt different conformations, visible in "conformational maps" similar to the Ramachandran peptide maps. As a consequence of the flexibility of each linkage, large oligosaccharides, such as biantennary N-glycans can adopt a variety of shapes as demonstrated by molecular dynamics [53]. Some branched oligosaccharides, with vicinal linkages, such as Lewis oligosaccharides, may however show a more rigid conformational behaviour [54].

Upon binding, lectins select one conformation of the glycan. The subsequent loss of flexibility participates in the entropy cost estimated by analysing the thermodynamics of binding [55]. The bound conformations often correspond to the lowest energy observed in solution but in some cases, the geometry differs from that expected at energy minimum. Lectins, unlike enzymes, do not induce a high energy excited state, but "secondary" conformations, a few kcal above lowest energy, are often observed. This is the case, for example, for Lewis X (Gal(β1–4)[Fuc (α1–3)]GlcNAc). This branched trisaccharide adopts mainly a single "closed" conformation in solution due to steric hindrance, rings stacking and unconventional C-H…O bond [56,57]. However, its binding to β-propeller lectins results in a very different "open" conformation, as displayed for RSL, the lectin of *Ralstonia solanacearum* in Fig. 4 [58]. This conformation in the binding site is therefore either selected by the lectin from very low concentration in solution or more likely induced by sliding contact with a Trp residue in the binding site.

4. Structural databases and structure – based classification

While the CAZy database has been collecting glycoactive enzymes for more than 20 years [59], no such tools were available for lectins until recently. At the present time, the lectin section of GlyCosmos [60] based on the LectinFrontier collection [61], contains more than 5000 entries cross-referencing UniProt and PDB. This database provides raw data from different sources, and emphasizes the need for a classification. One that distributes lectins in 48 families based on their folds was proposed [62].

More recently, the development of UniLectin3D promoted the lectin

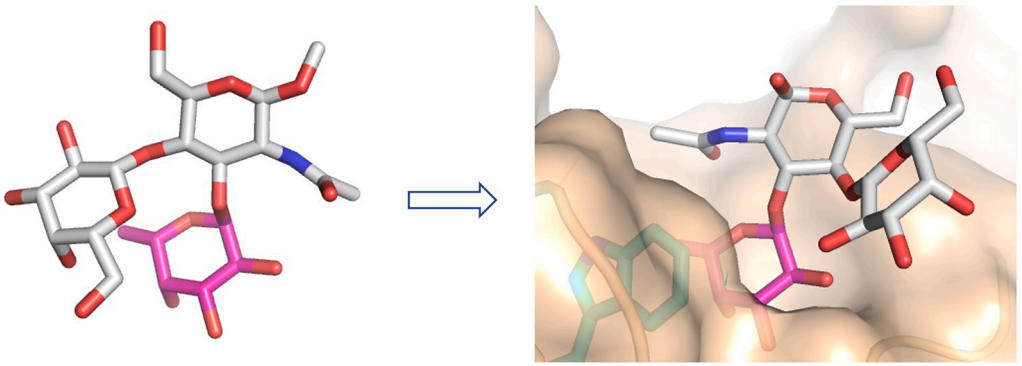

Fig. 4. Change of conformation for Lewis x in solution and crystal structure and in the binding site of RSL (PDB code 5AJB). For comparison sake, the fucose (coloured in pink) is shown with same orientation in the two panels to highlight the conformation flip of GlcNAc at the glycosidic linkage.

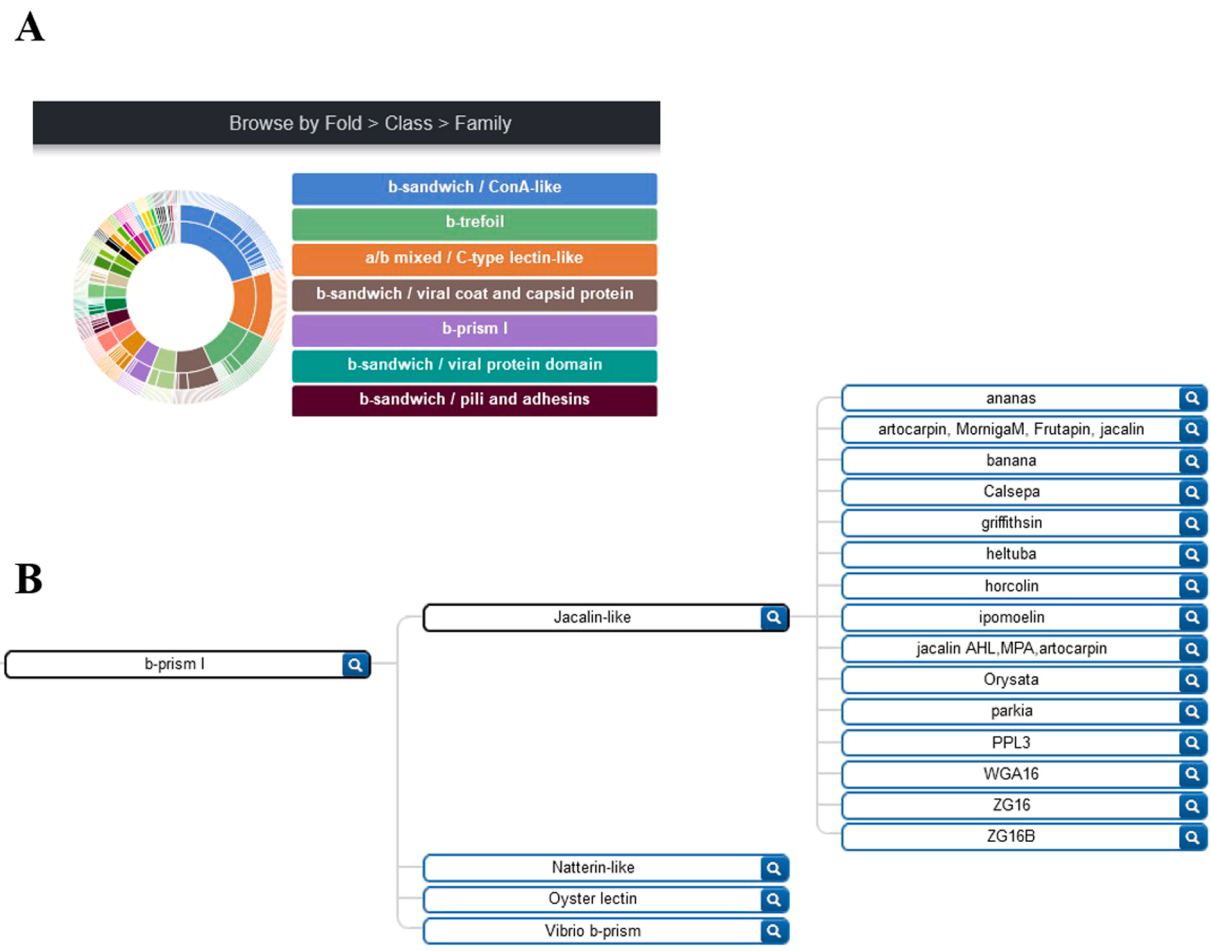

Fig. 5. (A) Sunburst of lectin folds in UniLectin3D, classified by most abundant entries (only the top seven ones are listed). (B) Example of classification for lectins with β-prism 1fold, with three levels from left to right: fold, class and family.

3D structures as the foundation for their classification [13]. In this database, lectins are hierarchically organized in three levels starting with 36 folds, then 109 classes based on amino acid sequence similarity, and lastly, 449 families reflecting species information. The sunburst representation of UniLectin3D content displayed in Fig. 5A, shows that β-sandwich, corresponding to an assembly of two β-sheets, is the top most frequent fold, as it is observed in well-studied legume lectins and galectins among others. The second is β-trefoil that lacks a regular secondary structure but possesses an internal 3-fold symmetry due to tandem repeats. This structure is reported in many studies on a variety of toxins [63]. The ranking shown in Fig. 5A tends to reflect research trends rather than biological facts, as in many other life science databases. Fig. 5B illustrates the hierarchy of different classes of lectins adopting the β-prism fold, and the next level of families for the jacalin-like class.

UniLectin3D is part of the UniLectin portal (Fig. 6A) that contains several databases (some are presented in the next section) and other glycobioinformatics tools related to lectins [64]. There are reciprocal links between UniProt and UniLectin3D. Recently, another group released the DIONYSUS database [65] built on protein-carbohydrate interactions filtered out from PDB, thereby including both lectins and glyco-active enzymes. A geometry-based clustering of the identified carbohydrate binding sites allows for the identification of clusters that can be used for comparative analysis. Note that Glyco@Expasy [66] and GlycoPedia (https://glycopedia.eu/) are two informative websites where the resources cited above and below are put into context. The former suggests a classification of databases and tools to support their reasoned use, while the latter provides background information organised in chapters.

5. Predicting structure from sequence: LectomeXplore and the human lectome

The second level of the UniLectin3D hierarchical classification (classes) was used to set the basis of a method predicting a putative lectin function from an amino acid sequence. A signature for each class was defined as a Hidden Markov Model (HMM) and each was systematically applied to screen popular protein sequence databases, namely, NCBI-nr [67] and UniProt [68]. The predicted lectin sequences are stored in LectomeXplore [69] that contains almost 1.5 million sequences of putative lectins across all kingdoms. LectomeXplore can be searched by organism and at each level of the classification. A score has been defined for selecting sequences with higher similarity with known lectins, and applying a value of 0.5 results in more than 170 000 sequences predicted to have lectin activity. Some classes, such as the fibrinogen fold of ficolins, are over-represented as displayed in Fig. 6B. Each entry provides information at the protein sequence level, indicating conserved amino acids in the binding site, and at genomic level, with adjacent domains identified. Structural information is available as an AlphaFold model when a UniProt accession number is available.

UniLectin also includes a complete catalogue of lectins in human tissues named HumanLectome [70]. This database was created from checking available literature [71,72] and using the LectomeXplore engine to process the entire human proteome (translated genome). This led to collect 215 putative lectins spread over 19 classes, the most populated ones being the C-type lectin-like, and the ficolin-like. About half of the content corresponds to curated entries, for which the lectin activity has been demonstrated. Low and very low evidence lectins, i.e. with no

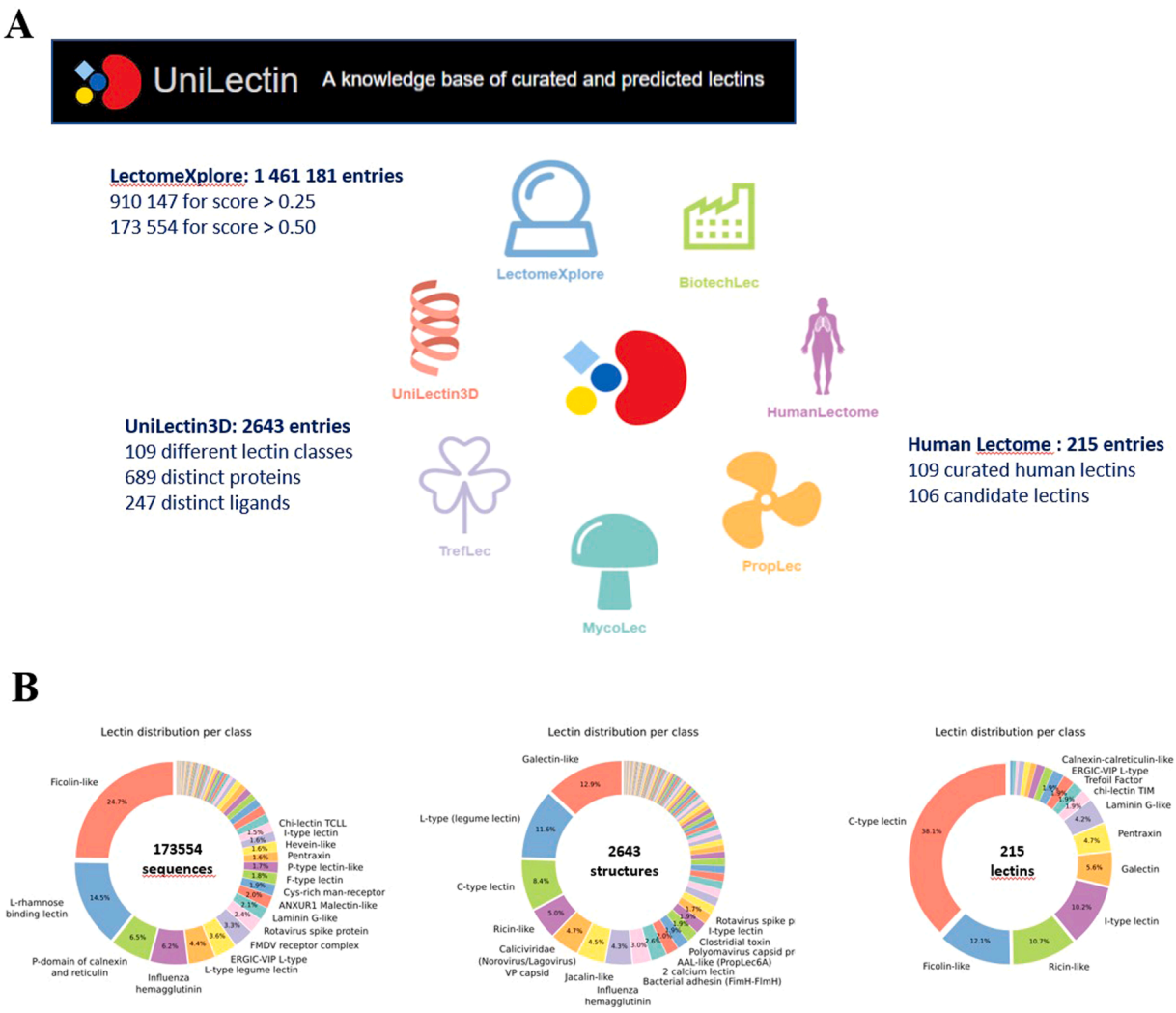

Fig. 6. The UniLectin Portal. (A) Schematic view of the different databases with content information. (B) Number of entries per lectin class for the different databases, with number and names indicated for the most populated ones.

experimental evidence for carbohydrate binding activity, have also been included since some of them may warrant more attention in the future. The information for each entry includes sequence, tissue localisation, AlphaFold model of the whole protein, and structure of the carbohydrate binding regions from PDB or built by homology using Modeller [73]. Additional information about the role of adjacent domains is also available.

6. Towards prediction of structure and specificity through machine learning approaches

Artificial Intelligence (AI) is now expanding fast in glycoscience, despite the sparsity and heterogeneity of published data [74]. In recent years, the interest for developing neural network models to predict glycan functional features and recognition has significantly raised and prompted several groups to design new AI-based approaches. This new trend follows the paradigm shift created by the spreading usage of the AlphaFold software that predicts the 3D structure of a protein from its amino acid sequence using large language models (LLMs) [75].

6.1. *The imprint of AlphaFold*

Within a few years, the AlphaFold output has become a standard for modelling the 3D structure of proteins in such a remarkable fashion that the work was crowned by a Nobel Prize in 2024. The high reliability achieved by AlphaFold has led some authors to attempt the mapping of no less than the protein universe [76] while others have systematically clustered proteins by fold to ease 3D structure search [77]. The occurrence and co-occurrence of folds for grouping proteins has guided the early protein structure classifications SCOP [78] and CATH [79] and these have evolved with AlphaFold towards a greater coverage of possible domains and domain combinations [80].

AI-powered classification is based on data available. As mentioned earlier, in the vast majority of cases, database content(s) used to train models is shaped by research trends as well as by the availability and limitations of technologies. In the case of proteins, PDB, due to its half-century longevity, contains much more proteins that easily crystallise leaving behind membrane proteins notoriously resistant to forming crystals. This bias does not reflect a biological but a methodological fact. Since UniLectin collects information from the PDB, it is also subjected to this bias, as shown in Table 1. Furthermore, immediate concerns for serious health threats produce an extensive coverage of some proteins such as Sars-Cov2 spike or cholera toxin, and are therefore another source of lopsided datasets that create knowledge gaps. These gaps are not always identifiable and may hinder or skew automated knowledge acquisition. This situation is revealed when using AlphaFold and related software for first glimpses at structures of lectins lacking experimental evidence. There is for example, no structural information on most algae

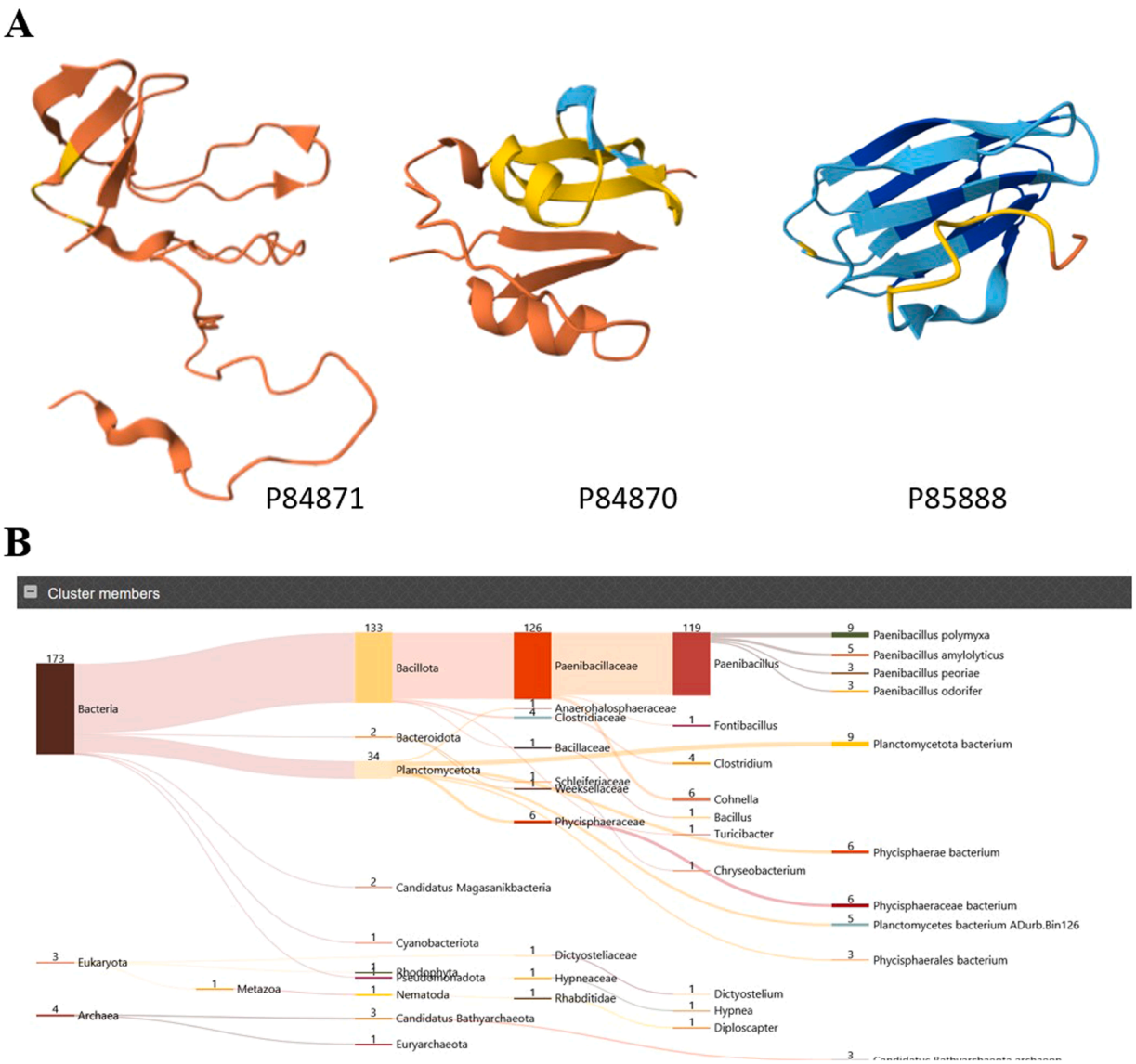

Fig. 7. Examples of use of AlphaFold in lectin structure prediction. (A) Predicted structures taken from the AlphaFold Structure Database for three lectins from red algae, HML (UniProt P84871), HCA (P84870) and hypnin-3 (P85888). AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. The colour code indicates the model confidence decreasing from blue (very high: pLDDT > 90) to orange (very low: pLDDT < 50). (B) Sankey diagram of the AlphaFold Clusters database (FoldSeek Clusters) showing species distribution of the cluster members to which hypnin-3 (P85888) is assigned.

lectins despite their promising applications in biotechnology. Among them, two related mucin-binding lectins, named HML and HCA, were purified two decades ago from the red marine algae *Hypnea musciformis* and *Hypnea cervicornis*, respectively [81] and unrelated hypnin-3 from *Hypnea japonica* was shown to be specific for core-fucosylation on N-glycans [82]. As shown in Fig. 7A, the AlphaFold Structure Database [83] fails to produce a reliable model for HCA (P84871) and outputs a low confidence one for HML (P84870) that somehow predicts a tandem repeat structure and disulfide bridges. Fig. 7A also confirms that AlphaFold provides a higher quality model for hypnin-3 (P85888). Despite the obvious advantage of relying on predicted structures, these examples illustrate the mitigated success met with AI method for truly novel sequences and the subsequent need for further checks. At the time of writing, hypnin-3 is still unclassified and therefore not included in UniLectin3D. This is a reflection of real cases regularly occurring when species are poorly studied. In this situation, it is possible to rely on FoldSeek [77] in the AlphaFold Clusters database [84] and assess the fold cluster in which it was assigned. FoldSeek is the result of clustering all protein folds collected in resolved and predicted (by AlphaFold) 3D structures in order to group proteins automatically by similar folds. In each cluster, species distribution is displayed as a Sankey diagram and Fig. 7B shows this diagram for the cluster where hypnin-3 is found. It highlights that hypnin-3 is the unique algal representative in this cluster dominated by Paenibacillus. Moreover, all proteins in the cluster are annotated as uncharacterized, opening the possibility of novel class of lectins. This example illustrates how the careful use of several bioinformatics resources can support shaping new assumptions.

The databases and other web resources related to lectins and cited in the text are listed and referenced in Table 2.

6.2. *The prediction of lectin specificity*

The last step of structural lectinomics, following 3D structure prediction, consists in inferring specificity from a carbohydrate binding site architecture. Determining specificity, affinity and ligand binding mode has become the next frontier. The prediction of specificity and affinity turns out to be much more complicated than prediction of structure. Many glycans are similar in terms of stereochemistry, but in contrast, a

Table 2
Summary of online resources cited in the text.

Name	Purpose	URL	Ref
CAZy	Glyco-enzyme knowledge portal	http://www.cazy.org	[59]
UniLectin	Lectin knowledge portal	http://unilectin.unige.ch	[64]
GlyCosmos	Japanese glycoinformatics portal	http://www.glycosmos.org	[60]
PDB	Repository of macromolecule 3D structures	https://www.rcsb.org/	[85]
UniProt	Database of protein sequences	http://www.uniprot.org	[68]
Rosetta	Software suite for modelling, design and analysis of macomolecules	http://rosettacommons.org/software	[86]
AlphaFold Structure database	Protein structure database	https://alphafold.ebi.ac.uk/	[83]
Modeller	Software for homology modelling of proteins	https://salilab.org/modeller/	[73]
DIONYSUS	database of protein–carbohydrate interfaces	http://www.dsimb.inserm.fr/DIONYSUS	[65]
FoldSeek clusters	Clusters from AlphaFold database	https://cluster.foldseek.com	[84]
Glyco@Expasy	Interactive website to access glycoinformatics resources	https://www.expasy.org/resources/glyco-expasy	[66]
GlycoPedia	Glycoscience online resource	https://glycopedia.eu/	

Table 3
Binding specificity predictive methods.

Name	Comments	URL	ref
GlycanDock	Molecular docking	https://docs.rosettacommons.org/docs/latest/application_documentation/carbohydrates/GlycanDock	[98]
PLIP	Protein-Ligand Interaction Profiler	https://github.com/demattox/plip	[90]
LectinOracle	Specificity prediction	https://github.com/BojarLab/LectinOracle	[93]
GlyNet	Specificity prediction	https://github.co/derdalab/GlyNet	[95]
GlyBERT	Specificity prediction	https://github.com/demattox/glyBERT	[92]
PeSTo-Carbs	Molecular docking	https://pesto.epfl.ch https://github.com/LBMEPFLPeSTo-Carbs	[99]
DeepGlycanSite	Molecular docking	https://github.com/xichengeva/DeepGlycanSite	[96]
CAPSIF:G	Binding site identifier	https://github.com/Graylab/CAPSIF	[97]

minor variation in a peptide sequence can result in different specificities. Consequently, lectins in the same class can present a large spectrum of specificities even when sequence identity is high. In some lectin classes, the abundance of crystal structures supports a rational understanding of specificity when examining amino acid variation in the binding site, as for the Man vs Gal preference in C-type lectins [87], or the Man/Fuc selectivity of 2-calcium bacterial lectins [88]. Similarly, in legume lectins, an explicit correlation was established between the lengths of the loops around the binding site and their carbohydrate specificities [89].

A handful of software tools that predicts binding specificity has been published, and each method hinges on different types of data (Table 3). In 2020, Mattox & Bailey-Kellogg [90] proposed a knowledge discovery approach through processing an extensive description of glycan binding pockets and outputting a selection of characteristic features. The same authors went on adapting SweetNet [91] a graph convolutional neural network representation of glycan structures to leverage the accumulation of glycan array data and predict a structure-function relationship [92] but this effort seems to have been dropped before publication. In parallel, two other groups followed a similar path to release methods with slightly different approaches and goals. To predict lectin specificity, LectinOracle [93] uniquely combines SweetNet with a transformer-based representation of proteins using ESM-1b [94]. GlyNet [95] also relies on a deep neural network trained on protein-glycan interaction data to quantitatively predict glycan affinity given a protein sample.

More influenced by structural biology, DeepGlycanSite [96] and CArbohydrate-Protein interaction Site IdentiFier (CAPSIF) [97] generate a model of binding from PDB data. The former extracts atomic coordinates of carbohydrate-protein complexes while the latter remains at the residue level. The Rosetta package, renowned for its excellence in protein structure prediction includes GlycanDock now defines a complete protein–glycoligand docking pipeline for building protein-glycan complexes [98]. Lastly, PeSTo-Carbs [99], an extension of the Protein Structure Transformer (PeSTo) engineered to predict protein-carbohydrate binding interfaces but limited to predicting interfaces involving cyclodextrins that are peculiar sugars unrelated to known ligands of lectins.

7. Conclusion

The significant progress achieved in recent years in modelling protein 3D structure is a springboard for many omics fields, particularly interactomics that is the next focal point in the improvement of 3D molecular modelling techniques. In this context, the study of protein-glycan interactions is a self-consistent problem that requires a systematic and stepwise approach presented in this review, as structural lectinomics. The latter also encompasses advancing the prediction of lectin specificity that remains challenging due to current knowledge gaps. Among these, the unknown extent of the lectin repertoire and the intricacy of multivalent binding. Nonetheless, solving these issues is essential to the elucidation of the glycocode and AI-based bioinformatics tools provide critical support to this end.

CRediT authorship contribution statement

Frédérique Lisacek: Writing – original draft, Visualization, Supervision, Resources, Data curation, Conceptualization. **Boris Schnider:** Writing – original draft, Visualization, Formal analysis, Data curation. **Anne Imberty:** Writing – original draft, Visualization, Supervision, Project administration, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

A.I. acknowledges Glyco@Alps (ANR-15-IDEX-0002) and Labex Arcane/CBH-EUR-GS (ANR- 17-EURE-0003).

Data availability

No data was used for the research described in the article.

References

[1] M.E. Taylor, K. Drickamer, A. Imberty, Y. van Kooyk, R.L. Schnaar, M.E. Etzler, A. Varki, Discovery and classification of glycan-binding proteins, in: P.H. Seeberger (Ed.), Essentials of Glycobiology, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2022, pp. 375–386.

[2] I.J. Goldstein, R.C. Hughes, M. Monsigny, T. Osawa, N. Sharon, What should be called a lectin? Nature 285 (1980) 66.

[3] H. Lis, N. Sharon, Lectins: carbohydrate-specific proteins that mediate cellular recognition, Chem. Rev. 98 (1998) 637–674.

[4] N. Sharon, H. Lis, Lectins as cell recognition molecules, Science 246 (1989) 227–234.

[5] A. Imberty, A. Varrot, Microbial recognition of human cell surface glycoconjugates, Curr. Opin. Struct. Biol. 18 (2008) 567–576.

[6] H. Ideo, A. Tsuchida, Y. Takada, Lectin-based approaches to analyze the role of glycans and their clinical application in disease, Int. J. Mol. Sci. 25 (2024).

[7] S. Notova, F. Bonnardel, F. Lisacek, A. Varrot, A. Imberty, Structure and engineering of tandem repeat lectins, Curr. Opin. Struct. Biol. 62 (2020) 39–47.

[8] F.H. Olvera-Lucio, H. Riveros-Rosas, A. Quintero-Martinez, A. Hernandez-Santoyo, Tandem-repeat lectins: structural and functional insights, Glycobiology 34 (2024).

[9] B.E. Collins, J.C. Paulson, Cell surface biology mediated by low affinity multivalent protein-glycan interactions, Curr. Opin. Chem. Biol. 8 (2004) 617–625.

[10] S. Leusmann, P. Menova, E. Shanin, A. Titz, C. Rademacher, Glycomimetics for the inhibition and modulation of lectins, Chem. Soc. Rev. 52 (2023) 3663–3740.

[11] H.J. Gabius, S. Andre, H. Kaltner, H.C. Siebert, The sugar code: functional lectinomics, Biochim. Biophys. Acta 1572 (2002) 165–177.

[12] H.M. Berman, T.N. Bhat, P.E. Bourne, Z. Feng, G. Gilliland, H. Weissig, J. Westbrook, The Protein Data Bank and the challenge of structural genomics, Nat. Struct. Biol. 7 (Suppl) (2000) 957–959.

[13] F. Bonnardel, J. Mariethoz, S. Salentin, X. Robin, M. Schroeder, S. Pérez, F. Lisacek, A. Imberty, UniLectin3D, a database of carbohydrate binding proteins with curated information on 3D structures and interacting ligands, Nucleic. Acids. Res. 47 (2019) D1236–D1244.

[14] G. Colombo, M. Meli, J. Canada, J.L. Asensio, J. Jimenez-Barbero, A dynamic perspective on the molecular recognition of chitooligosaccharide ligands by hevein domains, Carbohydr. Res. 340 (2005) 1039–1049.

[15] M.A. Williams, B.R. Westley, F.E. May, J. Feeney, The solution structure of the disulphide-linked homodimer of the human trefoil protein TFF1, FEBS Lett. 493 (2001) 70–74.

[16] J.M. Propster, F. Yang, S. Rabbani, B. Ernst, F.H. Allain, M. Schubert, Structural basis for sulfation-dependent self-glycan recognition by the human immune-inhibitory receptor Siglec-8, Proc. Natl. Acad. Sci. U S. A 113 (2016) E4170–E4179.

[17] J. Angulo, A. Arda, a. Bertuzzi, A. Canales, J. Ereno-Orbea, A. Gimeno, M. Gomez-Redondo, J.C. Munoz-García, P. Oquist, S. Monaco, A. Poveda, L. Unione, J. Jimenez-Barbero, NMR investigations of glycan conformation, dynamics, and interactions, Prog. Nucl. Magn. Reson. Spectrosc. 144 (2024) 97–152.

[18] J.R. Helliwell, Fundamentals of neutron crystallography in structural biology, Methods Enzymol. 634 (2020) 1–19.

[19] L. Gajdos, V.T. Forsyth, M.P. Blakeley, M. Haertlein, A. Imberty, E. Samain, J. M. Devos, Production of perdeuterated fucose from glyco-engineered bacteria, Glycobiology 31 (2021) 151–158.

[20] L. Gajdos, M.P. Blakeley, M. Haertlein, V.T. Forsyth, J.M. Devos, A. Imberty, Neutron crystallography reveals mechanisms used by Pseudomonas aeruginosa for host-cell binding, Nat. Commun. 13 (2022) 194.

[21] T.J. Yang, Y.C. Chang, T.P. Ko, P. Draczkowski, Y.C. Chien, Y.C. Chang, K.P. Wu, K. H. Khoo, H.W. Chang, S.D. Hsu, Cryo-EM analysis of a feline coronavirus spike protein reveals a unique structure and camouflaging glycans, Proc. Natl. Acad. Sci. U S. A 117 (2020) 1438–1446.

[22] C.J. Buchanan, B. Gaunt, P.J. Harrison, Y. Yang, J. Liu, A. Khan, A.M. Giltrap, A. Le Bas, P.N. Ward, K. Gupta, M. Dumoux, T.K. Tan, L. Schimaski, S. Daga, N. Picchiotti, M. Baldassarri, E. Benetti, C. Fallerini, F. Fava, A. Giliberti, P. I. Koukos, M.J. Davy, A. Lakshminarayanan, X. Xue, G. Papadakis, L.P. Deimel, V. Casablancas-Antras, T.D.W. Claridge, A. Bonvin, Q.J. Sattentau, S. Furini, M. Gori, J. Huo, R.J. Owens, C. Schaffitzel, I. Berger, A. Renieri, G.C.M. Study, J. H. Naismith, A.J. Baldwin, B.G. Davis, Pathogen-sugar interactions revealed by universal saturation transfer analysis, Science 377 (2022) eabm3125.

[23] J. Angulo, J. Zimmer, A. Imberty, J. Prestergard, Structural biology of glycan recognition, in: A. Varki, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger, R. D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer (Eds.), Essentials of Glycobiology, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2022, pp. 403–418.

[24] W.J. Peumans, E.J. Van Damme, A. Barre, P. Rougé, Classification of plant lectins in families of structurally and evolutionary related proteins, Adv. Exp. Med. Biol. 491 (2001) 27–54.

[25] J. Bouckaert, T. Hamelryck, L. Wyns, R. Loris, Novel structures of plant lectins and their complexes with carbohydrates, Curr. Opin. Struct. Biol. 9 (1999) 572–577.

[26] K. Drickamer, Increasing diversity of animal lectin structures, Curr. Opin. Struct. Biol. 5 (1995) 612–616.

[27] H. Kaltner, H.J. Gabius, Animal lectins: from initial description to elaborated structural and functional classification, Adv. Exp. Med. Biol. 491 (2001) 79–94.

[28] A. Varrot, S.M. Basheer, A. Imberty, Fungal lectins: structure, function and potential applications, Curr. Opin. Struct. Biol. 23 (2013) 678–685.

[29] A.L. Lewis, J.J. Kohler, M. Aebi, A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G. W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R. L. Schnaar, Microbial lectins: hemagglutinins, adhesins, and toxins, in: P. H. Seeberger (Ed.), Essentials of Glycobiology, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2022, pp. 505–516.

[30] K. Drickamer, M.E. Taylor, Recent insights into structures and functions of C-type lectins in the immune system, Curr. Opin. Struct. Biol. 34 (2015) 26–34.

[31] R.D. Cummings, R.L. Schnaar, Y. Ozeki, R-Type lectins, in: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N. H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger (Eds.) Essentials of Glycobiology, Cold Spring Harbor (NY), 2022, 419–430.

[32] G.R. Vasta, Galectins as pattern recognition receptors: structure, function, and evolution, Adv. Exp. Med. Biol. 946 (2012) 21–36.

[33] T. Angata, E. Brinkman-Van der Linden, I-type lectins, Biochim. Biophys. Acta 1572 (2002) 294–316.

[34] R.D. Cummings, M.E. Etzler, T.N.C. Ramya, K. Kato, G.A. Rabinovich, A. Surolia, L-type lectins, in: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P. H. Seeberger (Eds.), Essentials of Glycobiology, Cold Spring Harbor (NY), 2022, pp. 431–442.

[35] R. Loris, T. Hamelryck, J. Bouckaert, L. Wyns, Legume lectin structure, Biochim. Biophys. Acta 1383 (1998) 9–36.

[36] V.J.S. Osterne, G. De Sloover, E.J.M. Van Damme, Revisiting legume lectins: structural organization and carbohydrate-binding properties, Carbohydr. Res. 544 (2024) 109241.

[37] J.D. Schrag, J.J. Bergeron, Y. Li, S. Borisova, M. Hahn, D.Y. Thomas, M. Cygler, The structure of calnexin, an ER chaperone involved in quality control of protein folding, Mol. Cell 8 (2001) 633–644.

[38] R. Loris, Principles of structures of animal and plant lectins, Biochim. Biophys. Acta 1572 (2002) 198–208.

[39] J.M. Rini, Lectin structure, Annu. Rev. Biophys. Biomol. Struct. 24 (1995) 551–577.

[40] D.F. Gauto, S. Di Lella, C.M. Guardia, D.A. Estrin, M.A. Marti, Carbohydrate-binding proteins: dissecting ligand structures through solvent environment occupancy, J. Phys. Chem. B 113 (2009) 8717–8724.

[41] R. Loris, P.P. Stas, L. Wyns, Conserved waters in legume lectin crystal structures. The importance of bound water for the sequence-structure relationship within the legume lectin family, J. Biol. Chem. 269 (1994) 26722–26733.

[42] L.L. Kiessling, R.C. Diehl, CH-pi interactions in glycan recognition, ACS. Chem. Biol. 16 (2021) 1884–1893.

[43] C.H. Hsu, S. Park, D.E. Mortenson, B.L. Foley, X. Wang, R.J. Woods, D.A. Case, E. T. Powers, C.H. Wong, H.J. Dyson, J.W. Kelly, The dependence of carbohydrate-aromatic interaction strengths on the structure of the carbohydrate, J. Am. Chem. Soc. 138 (2016) 7636–7648.

[44] J.L. Asensio, A. Arda, F.J. Canada, J. Jimenez-Barbero, Carbohydrate-aromatic interactions, Acc. Chem. Res. 46 (2013) 946–954.

[45] J. Bouckaert, Y. Dewallef, F. Poortmans, L. Wyns, R. Loris, The structural features of concanavalin a governing non-proline peptide isomerization, J. Biol. Chem. 275 (2000) 19778–19787.

[46] T. Hatakeyama, H. Unno, Y. Kouzuma, T. Uchida, S. Eto, H. Hidemura, N. Kato, M. Yonekura, M. Kusunoki, C-type lectin-like carbohydrate recognition of the hemolytic lectin CEL-III containing ricin-type -trefoil folds, J. Biol. Chem. 282 (2007) 37826–37835.

[47] M. Perduca, L. Destefanis, M. Bovi, M. Galliano, F. Munari, M. Assfalg, F. Ferrari, H. L. Monaco, S. Capaldi, Structure and properties of the oyster mushroom (Pleurotus ostreatus) lectin, Glycobiology 30 (2020) 550–562.

[48] G. Cioci, E.P. Mitchell, C. Gautier, M. Wimmerova, D. Sudakevitz, S. Pérez, N. Gilboa-Garber, A. Imberty, Structural basis of calcium and galactose recognition by the lectin PA-IL of *Pseudomonas aeruginosa*, FEBS Lett. 555 (2003) 297–301.

[49] E. Mitchell, C. Houles, D. Sudakevitz, M. Wimmerova, C. Gautier, S. Pérez, A. M. Wu, N. Gilboa-Garber, A. Imberty, Structural basis for oligosaccharide-mediated adhesion of *Pseudomonas aeruginosa* in the lungs of cystic fibrosis patients, Nat. Struct. Biol. 9 (2002) 918–921.

[50] W.I. Weis, K. Drickamer, W.A. Hendrickson, Structure of a C-type mannose-binding protein complexed with an oligosaccharide, Nature 360 (1992) 127–134.

[51] R.D. Cummings, E. Chiffoleau, Y. van Kooyk, R.P. McEver, C-Type lectins, in: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger (Eds.) Essentials of Glycobiology, Cold Spring Harbor (NY), 2022, 455–474.

[52] B.G. Keller, C. Rademacher, Allostery in C-type lectins, Curr. Opin. Struct. Biol. 62 (2020) 31–38.

[53] A.M. Harbison, L.P. Brosnan, K. Fenlon, E. Fadda, Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: a molecular dynamics study, Glycobiology 29 (2019) 94–103.

[54] K.E. Miller, C. Mukhopadhyay, P. Cagas, C.A. Bush, Solution structure of the Lewis x oligosaccharide determined by NMR spectroscopy and molecular dynamics simulations, Biochemistry 31 (1992) 6703–6709.

[55] A. Audfray, A. Varrot, A. Imberty, Bacteria love our sugars: interaction between soluble lectins and human fucosylated glycans, structures, thermodynamics and design of competing glycocompounds, C. R. Chimie 16 (2013) 482–490.

[56] S. Pérez, N. Mouhous-Riou, N.E. Nifant'ev, Y.E. Tsvetkov, B. Bachet, A. Imberty, Crystal and molecular structure of a histo-blood group antigen involved in cell adhesion: the Lewis x trisaccharide, Glycobiology 6 (1996) 537–542.

[57] M. Zierke, M. Smiesko, S. Rabbani, T. Aeschbacher, B. Cutting, F.H. Allain, M. Schubert, B. Ernst, Stabilization of branched oligosaccharides: Lewis(x) benefits from a nonconventional C-H...O hydrogen bond, J. Am. Chem. Soc. 135 (2013) 13464–13472.

[58] J. Topin, M. Lelimousin, J. Arnaud, A. Audfray, S. Pérez, A. Varrot, A. Imberty, The hidden conformation of Lewis x, a human histo-blood group antigen, is a

determinant for recognition by pathogen lectins, ACS. Chem. Biol. 11 (2016) 2011–2020.
[59] E. Drula, M.L. Garron, S. Dogan, V. Lombard, B. Henrissat, N. Terrapon, The carbohydrate-active enzyme database: functions and literature, Nucleic. Acids. Res. 50 (2022) D571–D577.
[60] I. Yamada, M. Shiota, D. Shinmachi, T. Ono, S. Tsuchiya, M. Hosoda, A. Fujita, N. P. Aoki, Y. Watanabe, N. Fujita, K. Angata, H. Kaji, H. Narimatsu, S. Okuda, K. F. Aoki-Kinoshita, The GlyCosmos Portal: a unified and comprehensive web resource for the glycosciences, Nat. Methods 17 (2020) 649–650.
[61] J. Hirabayashi, H. Tateno, T. Shikanai, K.F. Aoki-Kinoshita, H. Narimatsu, The lectin frontier database (LfDB), and data generation based on frontal affinity chromatography, Molecules. 20 (2015) 951–973.
[62] Z. Fujimoto, H. Tateno, J. Hirabayashi, Lectin structures: classification based on the 3-D structures, Methods Mol. Biol. 1200 (2014) 579–606.
[63] S. Notova, F. Bonnardel, F. Rosato, L. Siukstaite, J. Schwaiger, N. Bovin, A. Varrot, W. Römer, F. Lisacek, A. Imberty, The choanoflagellate pore-forming lectin SaroL-1 punches holes in cancer cells by targeting tumor-related glycosphingolipid Gb3, Comm. Biol. 5 (2022) 594.
[64] A. Imberty, F. Bonnardel, F. Lisacek, UniLectin, a one-stop-shop to explore and study carbohydrate-binding proteins, Curr. Protoc. 1 (2021) e305.
[65] A. Gheeraert, T. Bailly, Y. Ren, A. Hamraoui, J. Te, Y. Vander Meersche, G. Cretin, R. Leon Foun Lin, J.C. Gelly, S. Perez, F. Guyon, T. Galochkina, DIONYSUS: a database of protein-carbohydrate interfaces, Nucleic. Acids. Res. 53 (2025) D387–D395.
[66] J. Mariethoz, D. Alocci, A. Gastaldello, O. Horlacher, E. Gasteiger, M. Rojas-Macias, N.G. Karlsson, N. Packer, F. Lisacek, Glycomics@ExPASy: bridging the gap, Mol. Cell Proteomics. (2018).
[67] E.W. Sayers, J. Beck, J.R. Brister, E.E. Bolton, K. Canese, D.C. Comeau, K. Funk, A. Ketter, S. Kim, A. Kimchi, P.A. Kitts, A. Kuznetsov, S. Lathrop, Z. Lu, K. McGarvey, T.L. Madden, T.D. Murphy, N. O'Leary, L. Phan, V.A. Schneider, F. Thibaud-Nissen, B.W. Trawick, K.D. Pruitt, J. Ostell, Database resources of the national center for biotechnology information, Nucleic. Acids. Res. 48 (2020) D9–D16.
[68] C. UniProt, UniProt: the universal protein knowledgebase in 2021, Nucleic. Acids. Res. 49 (2021) D480–D489.
[69] F. Bonnardel, J. Mariethoz, S. Perez, A. Imberty, F. Lisacek, LectomeXplore, an update of UniLectin for the discovery of carbohydrate-binding proteins based on a new lectin classification, Nucleic. Acids. Res. 49 (2021) D1548–D1554.
[70] B. Schnider, Y. M'Rad, J. el Ahmadie, A.G. de Brevern, A. Imberty, F. Lisacek, HumanLectome, an update of UniLectin for the annotation and prediction of human lectins, Nucleic Acid Res. 52 (2024) D1683–D1693.
[71] A.L. Peiffer, A.E. Dugan, L.L. Kiessling, Soluble Human lectins at the host-microbe interface, Annu. Rev. Biochem. 93 (2024) 565–601.
[72] C.D. Raposo, A.B. Canelas, M.T. Barros, Human lectins, their carbohydrate affinities and where to find them, Biomolecules. 11 (2021).
[73] B. Webb, A. Sali, Comparative protein structure modeling using MODELLER, Curr. Protoc. Protein Sci. 86 (2016), 2 9 1-2 9 37.
[74] D. Bojar, F. Lisacek, Glycoinformatics in the artificial intelligence era, Chem. Rev. 122 (2022) 15971–15988.
[75] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S.A. A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature 596 (2021) 583–589.
[76] J. Durairaj, A.M. Waterhouse, T. Mets, T. Brodiazhenko, M. Abdullah, G. Studer, G. Tauriello, M. Akdel, A. Andreeva, A. Bateman, T. Tenson, V. Hauryliuk, T. Schwede, J. Pereira, Uncovering new families and folds in the natural protein universe, Nature 622 (2023) 646–653.
[77] M. van Kempen, S.S. Kim, C. Tumescheit, M. Mirdita, J. Lee, C.L.M. Gilchrist, J. Soding, M. Steinegger, Fast and accurate protein structure search with Foldseek, Nat. Biotechnol. 42 (2024) 243–246.
[78] A. Andreeva, E. Kulesha, J. Gough, A.G. Murzin, The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures, Nucleic. Acids. Res. 48 (2020) D376–D382.
[79] I. Sillitoe, T.E. Lewis, A. Cuff, S. Das, P. Ashford, N.L. Dawson, N. Furnham, R. A. Laskowski, D. Lee, J.G. Lees, S. Lehtinen, R.A. Studer, J. Thornton, C.A. Orengo, CATH: comprehensive structural and functional annotations for genome sequences, Nucleic. Acids. Res. 43 (2015) D376–D381.
[80] A.M. Lau, N. Bordin, S.M. Kandathil, I. Sillitoe, V.P. Waman, J. Wells, C.A. Orengo, D.T. Jones, Exploring structural diversity across the protein universe with the encyclopedia of domains, Science 386 (2024) eadq4946.
[81] C.S. Nagano, H. Debray, K.S. Nascimento, V.P.T. Pinto, B.S. Cavada, S. Saker-Sampaio, W.R.L. Farias, A.H. Sampaio, J.J. Calvete, HCA and HML isolated from the red marine algae Hypnea cervicornis and Hypnea musciformis define a novel lectin family, Protein Sci. 14 (2005) 2167–2176.
[82] S. Okuyama, S. Nakamura-Tsuruta, H. Tateno, J. Hirabayashi, K. Matsubara, K. Hori, Strict binding specificity of small-sized lectins from the red alga Hypnea japonica for core (alpha1-6) fucosylated N-glycans, Biosci. Biotechnol. Biochem. 73 (2009) 912–920.
[83] M. Varadi, D. Bertoni, P. Magana, U. Paramval, I. Pidruchna, M. Radhakrishnan, M. Tsenkov, S. Nair, M. Mirdita, J. Yeo, O. Kovalevskiy, K. Tunyasuvunakool, A. Laydon, A. Zidek, H. Tomlinson, D. Hariharan, J. Abrahamson, T. Green, J. Jumper, E. Birney, M. Steinegger, D. Hassabis, S. Velankar, AlphaFold protein structure database in 2024: providing structure coverage for over 214 million protein sequences, Nucleic. Acids. Res. 52 (2024) D368–D375.
[84] I. Barrio-Hernandez, J. Yeo, J. Janes, M. Mirdita, C.L.M. Gilchrist, T. Wein, M. Varadi, S. Velankar, P. Beltrao, M. Steinegger, Clustering predicted structures at the scale of the known protein universe, Nature 622 (2023) 637–645.
[85] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I. N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic. Acids. Res. 28 (2000) 235–242.
[86] J.K. Leman, B.D. Weitzner, S.M. Lewis, J. Adolf-Bryfogle, N. Alam, R.F. Alford, M. Aprahamian, D. Baker, K.A. Barlow, P. Barth, B. Basanta, B.J. Bender, K. Blacklock, J. Bonet, S.E. Boyken, P. Bradley, C. Bystroff, P. Conway, S. Cooper, B.E. Correia, B. Coventry, R. Das, R.M. De Jong, F. DiMaio, L. Dsilva, R. Dunbrack, A.S. Ford, B. Frenz, D.Y. Fu, C. Geniesse, L. Goldschmidt, R. Gowthaman, J.J. Gray, D. Gront, S. Guffy, S. Horowitz, P.S. Huang, T. Huber, T.M. Jacobs, J.R. Jeliazkov, D.K. Johnson, K. Kappel, J. Karanicolas, H. Khakzad, K.R. Khar, S.D. Khare, F. Khatib, A. Khramushin, I.C. King, R. Kleffner, B. Koepnick, T. Kortemme, G. Kuenze, B. Kuhlman, D. Kuroda, J.W. Labonte, J.K. Lai, G. Lapidoth, A. Leaver-Fay, S. Lindert, T. Linsky, N. London, J.H. Lubin, S. Lyskov, J. Maguire, L. Malmstrom, E. Marcos, O. Marcu, N.A. Marze, J. Meiler, R. Moretti, V. K. Mulligan, S. Nerli, C. Norn, S. O'Conchuir, N. Ollikainen, S. Ovchinnikov, M. S. Pacella, X. Pan, H. Park, R.E. Pavlovicz, M. Pethe, B.G. Pierce, K.B. Pilla, B. Raveh, P.D. Renfrew, S.S.R. Burman, A. Rubenstein, M.F. Sauer, A. Scheck, W. Schief, O. Schueler-Furman, Y. Sedan, A.M. Sevy, N.G. Sgourakis, L. Shi, J. B. Siegel, D.A. Silva, S. Smith, Y. Song, A. Stein, M. Szegedy, F.D. Teets, S. B. Thyme, R.Y. Wang, A. Watkins, L. Zimmerman, R. Bonneau, Macromolecular modeling and design in Rosetta: recent methods and frameworks, Nat. Methods 17 (2020) 665–680.
[87] K. Drickamer, Ca(2+)-dependent sugar recognition by animal lectins, Biochem. Soc. Trans. 24 (1996) 146–150.
[88] J. Adam, M. Pokorna, C. Sabin, E.P. Mitchell, A. Imberty, M. Wimmerova, Engineering of PA-IIL lectin from *Pseudomonas aeruginosa* - unravelling the role of the specificity loop for sugar preference, BMC. Struct. Biol. 7 (2007) 36.
[89] V. Sharma, A. Surolia, Analyses of carbohydrate recognition by legume lectins: size of the combining site loops and their primary specificity, J. Mol. Biol. 267 (1997) 433–445.
[90] D.E. Mattox, C. Bailey-Kellogg, Comprehensive analysis of lectin-glycan interactions reveals determinants of lectin specificity, PLoS. Comput. Biol. 17 (2021) e1009470.
[91] R. Burkholz, J. Quackenbush, D. Bojar, Using graph convolutional neural networks to learn a representation for glycans, Cell Rep. 35 (2021) 109251.
[92] B. Dai, D.E. Mattox, C. Bailey-Kellogg, bioRxiv. (2021), https://doi.org/10.1101/2021.1110.1115.464532.
[93] J. Lundstrom, E. Korhonen, F. Lisacek, D. Bojar, LectinOracle: A generalizable deep learning model for lectin-glycan binding prediction, Adv. Sci. 9 (2022) e2103807.
[94] A. Rives, J. Meier, T. Sercu, S. Goyal, Z. Lin, J. Liu, D. Guo, M. Ott, C.L. Zitnick, J. Ma, R. Fergus, Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences, Proc. Natl. Acad. Sci. U S. A 118 (2021).
[95] E.J. Carpenter, S. Seth, N. Yue, R. Greiner, R. Derda, GlyNet: a multi-task neural network for predicting protein-glycan interactions, Chem. Sci. 13 (2022) 6669–6686.
[96] X. He, L. Zhao, Y. Tian, R. Li, Q. Chu, Z. Gu, M. Zheng, Y. Wang, S. Li, H. Jiang, Y. Jiang, L. Wen, D. Wang, X. Cheng, Highly accurate carbohydrate-binding site prediction with DeepGlycanSite, Nat. Commun. 15 (2024) 5163.
[97] S.W. Canner, S. Shanker, J.J. Gray, Structure-based neural network protein-carbohydrate interaction predictions at the residue level, Front. Bioinform. 3 (2023) 1186531.
[98] M.L. Nance, J.W. Labonte, J. Adolf-Bryfogle, J.J. Gray, Development and evaluation of GlycanDock: a protein-glycoligand docking refinement algorithm in rosetta, j. phys. chem. b (2021).
[99] P. Bibekar, L. Krapp, M.D. Peraro, PeSTo-carbs: geometric deep learning for prediction of protein-carbohydrate binding interfaces, J. Chem. Theory. Comput. 20 (2024) 2985–2991.