Glycobiology, 2024, **34**, cwae037
https://doi.org/10.1093/glycob/cwae037
Advance access publication date 20 May 2024
Review

The three-sided right-handed β-helix is a versatile fold for glycan interactions

Audrey A. Burnim, Keith Dufault-Thompson, Xiaofang Jiang*

National Library of Medicine, National Institutes of Health, Building 38A, Room 6N607, 8600 Rockville Pike, Bethesda, MD 20894 United States
*Corresponding author: National Library of Medicine, National Institutes of Health, Building 38A, Room 6N607, 8600 Rockville Pike, Bethesda, MD 20894 United States. Email: xiaofang.jiang@nih.gov

Interactions between proteins and glycans are critical to various biological processes. With databases of carbohydrate-interacting proteins and increasing amounts of structural data, the three-sided right-handed β-helix (RHBH) has emerged as a significant structural fold for glycan interactions. In this review, we provide an overview of the sequence, mechanistic, and structural features that enable the RHBH to interact with glycans. The RHBH is a prevalent fold that exists in eukaryotes, prokaryotes, and viruses associated with adhesin and carbohydrate-active enzyme (CAZyme) functions. An evolutionary trajectory analysis on structurally characterized RHBH-containing proteins shows that they likely evolved from carbohydrate-binding proteins with their carbohydrate-degrading activities evolving later. By examining three polysaccharide lyase and three glycoside hydrolase structures, we provide a detailed view of the modes of glycan binding in RHBH proteins. The 3-dimensional shape of the RHBH creates an electrostatically and spatially favorable glycan binding surface that allows for extensive hydrogen bonding interactions, leading to favorable and stable glycan binding. The RHBH is observed to be an adaptable domain capable of being modified with loop insertions and charge inversions to accommodate heterogeneous and flexible glycans and diverse reaction mechanisms. Understanding this prevalent protein fold can advance our knowledge of glycan binding in biological systems and help guide the efficient design and utilization of RHBH-containing proteins in glycobiology research.

Key words: adhesin; beta-helix; carbohydrate-active enzymes; glycoside hydrolase; polysaccharide lyase.

Introduction

Glycan-interacting proteins play pivotal roles in numerous biological processes, from cellular communication to pathogen invasion (Varki 2017; Lee et al. 2022). The diversity and complexity of glycans pose unique challenges to understanding these interactions (Davies and Henrissat 1995; DeMarco and Woods 2008; Gabius et al. 2011). Data on carbohydrate-interacting proteins, such as the information collected in the carbohydrate-active enzyme database (CAZy) (Drula et al. 2022), UniLectin (Bonnardel et al. 2021), and GlycoGene databases (Togayachi et al. 2008), have significantly advanced the field by providing functional annotations for a variety of carbohydrate-active enzymes and carbohydrate-binding proteins. These resources facilitate the identification and characterization of glycan-interacting proteins, underlining the importance of databases and manual curation in propelling glycobiology research forward.

The three-sided right-handed β-helix (RHBH) has emerged as an important structural fold to glycobiology in both carbohydrate-active enzymes (CAZymes) and lectin-like proteins. Detailed studies have provided analyses of specific three-sided right-handed β-helix protein families with biological significance, including the first identified RHBH family, pectate lyases (Yoder et al. 1993a, 1993b), enzymes secreted by plant pathogens to degrade plant cell walls. In the gut microbiome, polysaccharide degradation is performed by *Bacteroides thetaiotaomicron* rhamnogalacturonan lyase to similarly digest pectin from plant cell walls (Luis et al. 2018). Another seminal RHBH study described the *Bordetella pertussis* virulence factor, an RHBH that is able to attach to mammalian cells and is used in whooping cough vaccines (Emsley et al. 1996). Other structural studies have specified the molecular determinants of lectin-like interactions, including cellulose-binding proteins implicated in biomass digestion (Close et al. 2014) and Type V secretion system proteins important for bacterial pathogenesis (Kajava and Steven 2006a). Additionally, the detection of antibiotic-resistant pathogens and degradation of biofilms by RHBH-containing bacteriophage tailspike proteins is an area of active research (Lee et al. 2020; Rice et al. 2021; Abdelkader et al. 2022). The versatility of the RHBH fold, being involved in crucial biological processes from pathogenesis to carbohydrate degradation, underscores the importance of understanding its properties for future therapeutic and biotechnological applications.

Systematic studies of β-solenoids, the family of proteins which includes the RHBH, have been performed, yet these studies often focus on the overall protein structure rather than on the details of the fold's function and mechanism (Jenkins et al. 1998; Jenkins and Pickersgill 2001; Iengar et al. 2006; Kajava and Steven 2006b; Roche et al. 2018). The broader structural superfamily of β-solenoids consists of 15 different cross-sectional shapes, including the three-sided left-handed, two-sided right-handed, and four-sided right-handed β-helices among other β-repeat folds. Here, we focus our structural review on the L-type β-solenoid, which is the three-sided right-handed β-sheet (Mesdaghi et al. 2023). The RHBH fold is characterized by three distinct sides of repeating parallel β-helix sheets interspaced with loops that form a larger solenoid structure, distinguishing it from other β-solenoids.

Received: March 13, 2024. **Revised:** May 13, 2024. **Accepted:** May 17, 2024
Published by Oxford University Press 2024.
This work is written by US Government employees and is in the public domain in the US.

Downloaded from https://academic.oup.com/glycob/article/34/7/cwae037/7676630 by NIH - Health Services Research Library user on 24 January 2025

In addition, traditional sequence analysis methods, such as hidden Markov model searches, do not effectively capture the diverse sequences associated with β-sheet repeats and solenoid domains, leading to the omission of distinct families like the phage tailspike proteins from many databases (Ciccarelli et al. 2002). By bridging the gap between sequence information and structural insights, we can provide a foundation for further advancements in the characterization of β-solenoids and their role in biological processes.

In this review, we present a comprehensive overview of three-sided RHBH proteins in the Protein Data Bank (PDB), focusing on their structural characteristics, prevalence, and functional roles in glycan interaction. We detail a unified overview showing that experimentally observed three-sided right-handed parallel β-helix structures are nearly exclusively associated with glycan interaction and adhesin functions, highlighting their functional specificity. The evolutionary trajectories of RHBH sequences exhibit distinct functional clustering, likely diversifying from a glycan-binding ancestral protein. Examination of RHBH structures reveals the features that allow RHBH folds to interact with glycans, including the shape, electrostatics, and hydrogen bonding propensity of the binding sites. The detailed examination of RHBH proteins in this review highlights their role as key players in glycobiology, illuminating their potential in unraveling complex biological processes and fostering the development of innovative therapeutic strategies targeting glycan-related diseases.

Right-handed β-helix proteins are associated with glycan binding functions

The RHBH is a single protein chain with β-sheet and loop repeats (Fig. 1A, CATH: 2.160.20 (Sillitoe et al. 2021)). These structural repeats fold in a circular organization such that there are three sides (Fig. 1A, PB1-3) interspaced by three turns (Fig. 1A, T1-3). The β-sheets stack in a parallel manner along a perpendicular axis to the sheet direction, causing the β-sheets to process along the major axis, forming a helical structure. Along the major axis, this helical structure forms a cleft with two neighboring loops or turns (Fig. 1A, T1 and T3), adding depth to the cleft formed by the β-sheet (PB1). The RHBH described in this work is distinct from other three-sided β-solenoid folds (CATH 2.160), such as the β-roll (CATH 2.30), left-handed β-helix (CATH 2.160.10), and the triple-stranded β-helix (e.g. PDB 4jj2). At the time of writing, seven of the 43 polysaccharide lyase families and 13 of the 189 glycoside hydrolase families in the CAZy database consist of parallel β-helix structures.

Utilizing the CATH and ECOD (Cheng et al. 2014) databases, we compiled an initial dataset of right-handed β-helix (RHBH) proteins from the Protein Data Bank (PDB), supplemented by manual searches and curation to remove duplicate chains and highly homologous sequences (within 95% identity), resulting in 151 representative structures from 258 entries (Table 1, Supplementary Table 1). All of the RHBH structures except for one (PDB 7o17) are identified as carbohydrate-active enzymes (CAZymes) (81/151), adhesins (35), or bacteriophage tailspike proteins (34). The RHBH sequences from this dataset are predominantly encoded by bacteria (88) and phages (34) (Table 1, Supplementary Table 1). RHBH sequences from eukaryotic organisms were also observed, specifically from plants and fungi (e.g. pectin degrading enzymes PDB 7b8b from *Arabidopsis thaliana* and PDB 4c2l from *Aspergillus tubingensis*).

Within structurally characterized RHBHs, 34 out of 81 CAZymes are classified as glycoside hydrolases (GH), which utilize general acid catalysis and water hydrolysis for the cleavage of glycosidic bonds (Fig. S1) (Davies and Henrissat 1995). Meanwhile, 37 out of 81 are identified as polysaccharide lyases (PL), enzymes that specifically target and cleave polysaccharides containing uronic acid (Fig. S1) (Garron and Cygler 2014; Drula et al. 2022). The less represented categories of CAZymes include carbohydrate esterases (Fig. 1F, CE8) and carbohydrate epimerases (Table 1, no CAZyme category), and are also observed to interact and isomerize pectin (Fries et al. 2007) and alginate isomers (Rozeboom et al. 2008; Wolfram et al. 2014) respectively. These diverse categories highlight the versatility of these folds, suggesting the RHBH can perform a range of carbohydrate-modifying functions.

Tailspike proteins are a functional category of RHBH that can act as both PL and GH enzymes. Previous work with phage tailspike proteins has demonstrated that the RHBH domain facilitates strong and specific glycan binding, is the determinant of substrate specificity, and facilitates the depolymerization of glycans (Simpson et al. 2015; Latka et al. 2017; Knecht et al. 2019). Most of the bacteriophage tailspike proteins in the PDB are not assigned to any CAZyme category (30/34). Tailspike proteins are receptor-binding proteins known for their ability to specifically recognize and degrade bacterial polysaccharides through their glycoside hydrolase and polysaccharide lyase activities (Knecht et al. 2019). Despite their structurally conserved RHBH domains, tailspike proteins are particularly diverse and cannot be easily categorized by sequence homology like the bacterial and eukaryotic CAZymes (Steinbacher et al. 1997; Timoshina et al. 2023). Four tailspike proteins, all acting on O-antigens (PDBs 2x6w, 2xc1, 3riq, and 2v5i), are categorized as glycoside hydrolases, and the latter three sequences are subclassified into GH90 (Drula et al. 2022). The remaining 30 tailspike proteins in the dataset consist of proteins that are observed to act on O-antigens (e.g. PDBs 5w6h, 6f7k), capsular polysaccharides (CPS, e.g. 7lzj), or exopolysaccharides (EPS, e.g. 5jsd) (Supplementary Table 1).

Most RHBH structures in membrane-containing organisms (that is, excluding viruses) feature a signal peptide sequence motif at the N-terminus, as highlighted by SignalP annotation in 79 of 115 cases (Almagro Armenteros et al. 2019). The presence of signal peptides in these proteins indicates that they are likely destined to be secreted out of the cell or integrated into a membrane (Pohlschröder et al. 1997; Nielsen 2017). A subset of these signal peptide-containing sequences are from known extracellular proteins, including adhesins and autotransporters (Fig. 1C, Table 1). Adhesin-like proteins represent a significant portion of the total RHBH structural dataset, mediating functions including biofilm formation (i.e. PDB 3syj), binding of glycoproteins like fibronectin (i.e. PDB 4mr0) and hemopexin (i.e. PDB 4rm6), cellulose binding (i.e. PDB 4wa0) and membrane disruption (i.e. PDB 5j44).

The association of a small subset of the RHBH-containing proteins with glycans is ambiguous. These proteins include F-box containing proteins (UniProt IDs Q86XK2 and Q9UK96), fibrocystin (UniProt IDs P08F94 and Q86WI1) (Fig. S2A and B), and the ATP-binding cassette transporter binding protein NosD (PDB 7o17) (Fig. S2C). While fibrocystin is crucial for kidney function and epithelial

Downloaded from https://academic.oup.com/glycob/article/34/7/cwae037/7676630 by NIH - Health Services Research Library user on 24 January 2025

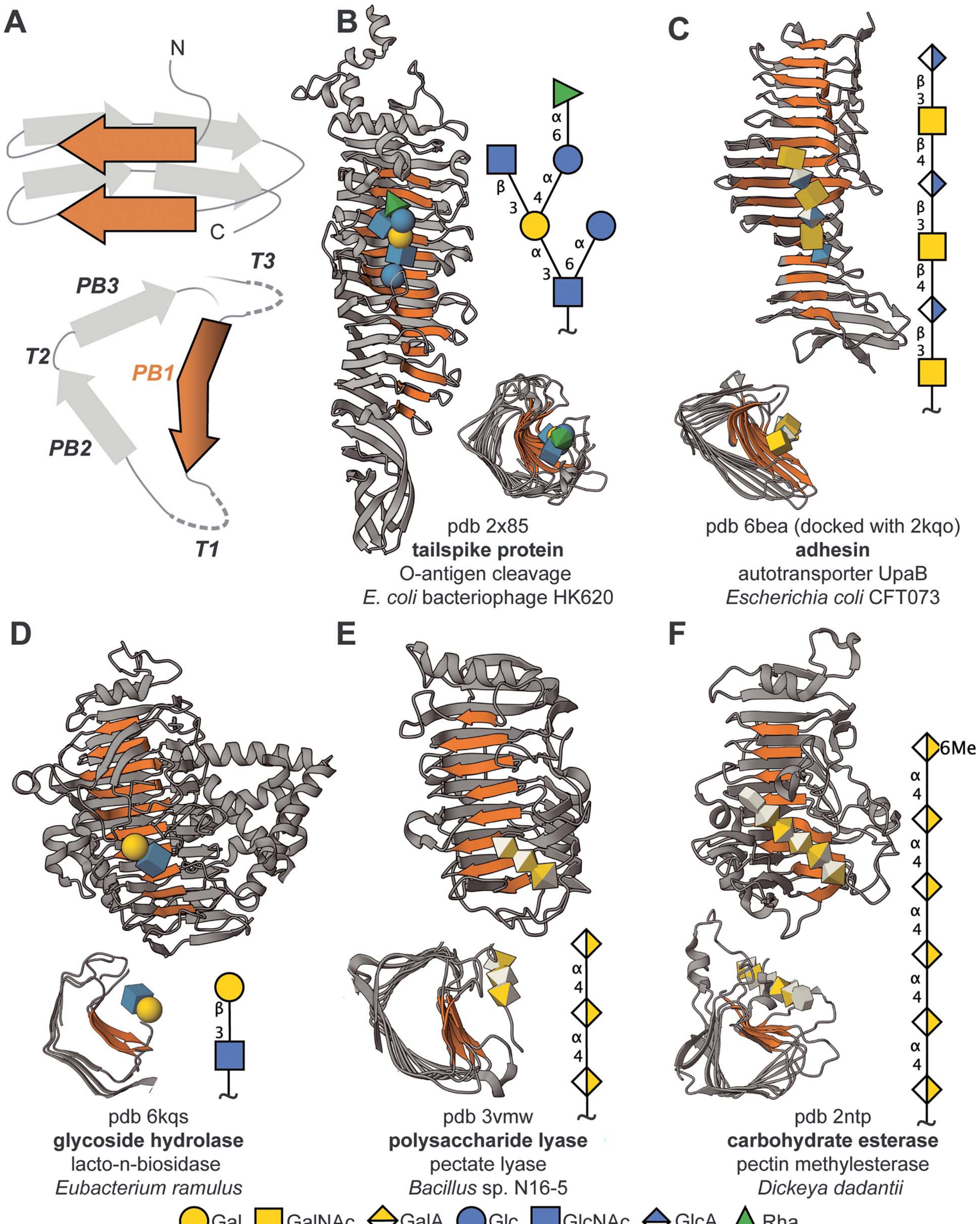

Fig. 1. The right-handed parallel β-helix fold and examples of carbohydrate binding. A) Cartoon of the basic building block of the right-handed β-helix fold, shown as two rungs. The folds are labeled with T1-T3 for loop turns that connect the β-helix, and PB1-3 for poly-β-helix. B–F) Examples of the main categories of β-helix proteins bound to carbohydrate substrates. In all panels, the top structure shows the N-terminus at the top and the C-terminus at the bottom. In all panels, the glycan structure is shown bound in 3D symbol nomenclature for glycan (SNFG) and in 2D SNFG representation to the right (Neelamegham et al. 2019). In all panels except for the adhesin (C), the glycans modeled are experimentally observed in their respective crystal structures. In (C) the adhesin shows PDB 2kqo, chondroitin, docked as described by (Paxman et al. 2019). The bottom shows a key describing SNFG identities: Gal, galactopyranose; GalNAc, N-acetyl-D-galactosamine; GalA, galactopyranuronic acid; Glc, glucopyranose; GlcNAc, N-acetylglucosamine; Rha, rhamnose. Protein-glycan complexes were visualized in Mol* viewer (Sehnal et al. 2021).

Downloaded from https://academic.oup.com/glycob/article/34/7/cwae037/7676630 by NIH - Health Services Research Library user on 24 January 2025

Table 1. Summary of β-helix specific functions. Overview of the function of β-helix structures with example PDB identifiers and substrates. The bottom shows a key describing symbol nomenclature for glycan (SNFG) identities of the glycans that β-helix proteins bind in these examples: Fuc, fucose; Xyl, xylose; GulA, guluronic acid; gal, galactose; GalA, galacturonic acid; man, mannose; ManA, mannuronic acid; Glc, glucose.

Functional classification	*Active site characteristics*	*CAZyme Family*	*Protein Data Bank deposition and substrate examples*
Polysaccharide Lyase	• Positively charged pocket along cleft • Calcium bound often	PL1, PL3, PL6, PL9, PL8, PL31, PL41	6a40: 3krg:
		Bacteriophage tailspike proteins (PL0)	7w1e:
Glycoside Hydrolase	• Negatively charged pocket along cleft • Aspartate and glutamate invariably involved in catalysis	GH28, GH49, GH55, GH82, GH87, GH90, GH9, GH110, GH118, GH120, GH136, GH141	3vsu: 3wwg:
		Bacteriophage tailspike proteins (GH0, GH90)	6f7k:
Carbohydrate Esterase	• Hydrophobic cleft for acyl group adjacent to negatively charged pocket • Two aspartates invariably involved in catalysis	CE8	2ntp:
Carbohydrate Epimerase	• Calcium bound	N/A	2pyh:
Cellular Interaction (Type V secretion, adhesins, F-box containing, fibrocystin-like)	• Hydrophobic interactions dominate	N/A	• Self-adhesin • Type V secretion system
Unclassified bacteriophage tailspike proteins	Not conclusive	Not conclusive	• Capsular polysaccharide degradation • Unknown glycan binding

Fuc Xyl GulA Gal GalA Man ManA Glc

cell adhesion (Hogan et al. 2003; Qiu et al. 2020; Ziegler et al. 2020), F-box proteins are involved in protein–protein interactions (Kipreos and Pagano 2000), with a subset of F-box proteins specific for glycoprotein interactions (Kumanomidou et al. 2015), and NosD is part of a copper-dependent complex essential for nitrogen and sulfur transport (Müller et al. 2022). The presence of RHBH domains in these proteins suggests the potential for glycan-binding functions or indicates an evolutionary adaptation towards novel, as yet unidentified, roles.

Right-handed β-helix proteins exhibit clustering by function and distinct evolutionary trajectories

The repetitive and diverse nature of RHBH sequences poses challenges for traditional alignment-based evolutionary analysis. While other studies have explored some of the sequence-level features of RHBH proteins (Jenkins and Pickersgill 2001; Kajava and Steven 2006b; Roche et al. 2018), they have typically been unable to investigate RHBH evolution at large scales due to extensive sequence diversity. We utilized the recently developed tool, evo-velocity (Hie et al. 2022), to overcome these obstacles, gain insights into the evolution of the RHBH, and facilitate a broader comparison of RHBH sequences. Evo-velocity analysis of the RHBH dataset revealed three main functional clusters: CAZymes, bacteriophage tailspike proteins, and adhesins (Fig. 2A). Most apparent from the evolutionary modeling are the distinct and relatively distant clusters of adhesin proteins (Fig. 2A, yellow) and tailspike proteins (Fig. 2A, green). Despite the tailspike proteins representing a mix of polysaccharide lyases and glycoside hydrolases, they formed a distinct cluster separate from other CAZymes (see Supplementary Table 1). Overlap, depicting primary sequence similarity and evolutionary relatedness, is particularly apparent within the bacterial and eukaryotic CAZyme categories (Fig. 2A blue, red, orange; Fig. 2C blue, orange), where polysaccharide lyase (PL) and glycoside hydrolase (GH) enzymes are interspaced closely with each other.

Through evolutionary pseudotime modeling, as a proxy for evolutionary trajectory, evo-velocity reveals that adhesin sequences are observed to be closest to the putative root of the RHBH dataset, with the earliest pseudotime (blue, Fig. 2B) and bacteriophage tailspike proteins and bacterial

Downloaded from https://academic.oup.com/glycob/article/34/7/cwae037/7676630 by NIH - Health Services Research Library user on 24 January 2025

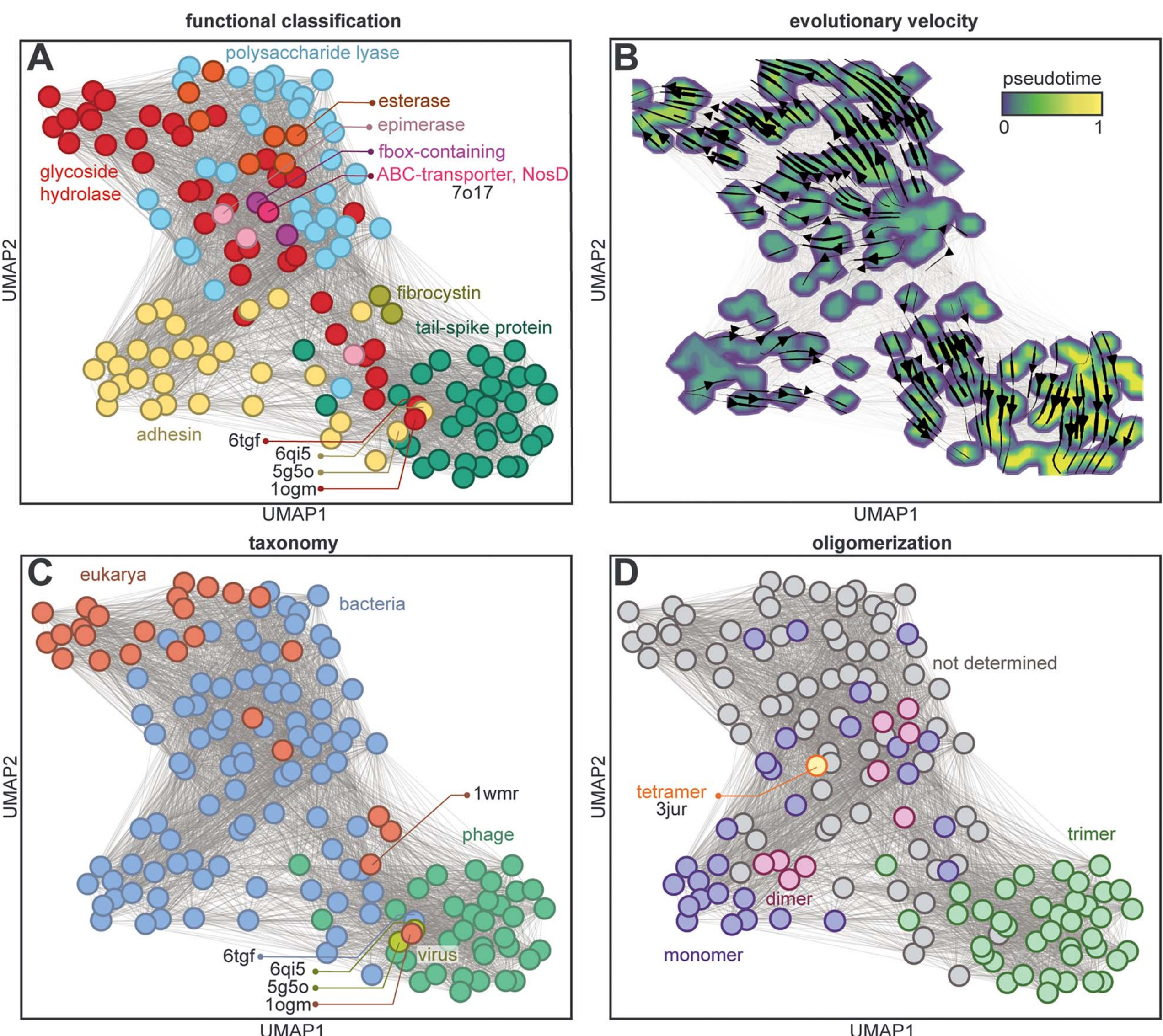

Fig. 2. Evolutionary velocity analysis of the diverse RHBH structural family. A) 151 protein structures mapped onto UMAP space representing an alignment-free evolutionary trajectory of the sequences analyzed by evo-velocity (Hie et al. 2022). Sequences are connected by gray lines representing network edges without any pseudotime directionality. Sequences are colored by the classification of function. B) Sequences as in (A, C, D) colored by pseudotime approximation corresponding to evolutionary trajectory by evo-velocity analysis. Sequences that are more likely to represent root nodes (pseudotime = 0) in the evolutionary trajectory are colored blue and sequences that are less likely to be root sequences, therefore more likely to have diverged in the sequence space are colored yellow (pseudotime = 1). C) Sequences as in (A, D) colored by superkingdom. D) Protein sequences as in (A, C) colored by known oligomeric state from source literature. Where the oligomeric state could not be inferred from the associated literature, the sequence is colored gray.

and eukaryotic CAZymes are most diverged in sequence space, farthest from the root of the RHBH dataset (yellow, Fig. 2B). This pseudotime modeling enables a hypothesis of the evolutionary trajectory of the divergent RHBH sequences (Haghverdi et al. 2016), suggesting the ancestral RHBH may have been able to only facilitate receptor binding or intramolecular interactions, with subsequent catalytic function conferred in some RHBHs, resulting in diverse CAZymes capable of degrading heterogeneous glycan substrates. While the currently described RHBH proteins provide evidence for evolution from a common ancestor, similarly suggested by others (Jenkins et al. 1998), comparable repeating solenoid folds have been hypothesized to have evolved independently multiple times (Todd et al. 2001; Chaudhuri et al. 2008). As more protein structures and sequences are characterized, new evolutionary patterns may become evident in this broadly distributed protein family.

Clustering by oligomeric state is also observed in the RHBH sequences, revealing a cluster of trimeric proteins associated with the bacteriophage tailspike proteins (Fig. 2D). This evolutionary constraint is especially pertinent considering a substantial number of these proteins function as multimers (i.e. tailspike proteins) or require oligomerization for function (i.e. adhesins) (Soto and Hultgren 1999; Klemm and Schembri 2000; Paxman et al. 2019). Approximately half of the RHBH protein sequences in the dataset have experimental evidence of oligomerization (78/151). Of these proteins with known oligomeric states, all sequences encoded by bacteriophages and the two minor coat protein sequences encoded by viruses are observed to be trimeric (Fig. 2D,

Downloaded from https://academic.oup.com/glycob/article/34/7/cwae037/7676630 by NIH - Health Services Research Library user on 24 January 2025

green circles). The monomeric RHBH proteins are localized throughout the adhesin and CAZyme clusters with a cluster associated near the adhesin proteins (Fig. 2D, blue circles). Some of these monomers are unique two-domain RHBH structures joined by a linker, fixing a dimer interaction (e.g. alginate lyases PDB 7dmk and 5gkq). One observable outlier oligomer is the tetramer PDB 3jur, a pectin-degrading polysaccharide lyase encoded by the extremophile *Thermotoga maritima* (Pijning et al. 2009). The evolutionary justification proposed for this oligomerization state by Pijning and coworkers was that it would allow the protein to be thermostable in the high-temperature environment that *T. maritima* inhabits. The functional clustering and evolutionary relatedness from the oligomeric states of RHBH proteins shed light on the structural diversity essential for varied functions.

Analysis of the evolutionary trajectory of RHBH sequences reveals interesting relationships between RHBH structures and bacteriophage tailspike proteins. The bacterial biofilm modifying glycoside hydrolase, PDB 6tgf, is found at the interface between bacterial CAZymes and bacteriophage tailspike proteins in the evo-velocity trajectory (Fig. 2A and C). The protein WceF was observed to be a trimer and shared considerable structural homology with bacteriophage tailspike proteins (Irmscher et al. 2021). Specifically, the neck domain between the N-terminus and β-helix domain of WceF has been reported to have significant structural homology to a domain in the bacteriophage CBA120 tailspike protein (Irmscher et al. 2021). Further experiments by Irmscher et al. showed that deletion of this N-terminus in the bacterial WceF prevented trimerization of the protein. Also observed at the interface between bacteriophages and CAZymes are two proteins from eukaryotic viruses, the LH3 protein from lizard atadenovirus LAdV-2 (PDB 6qi5) (Marabini et al. 2021) and the LH3 protein from snake atadenovirus SnAdV-1 (PDB 5g5o) (Menéndez-Conejero et al. 2017). These proteins both have a trimeric structure similar to bacteriophage tailspike proteins and are involved in stabilizing the viral capsid structures. Sequence similarity analysis showed that these adenovirus proteins were homologous to known CAZymes and phage tailspike proteins (Marabini et al. 2021). However, initial glycan-binding experiments on LH3 from lizard atadenovirus LAdV-2 did not show any binding or CAZyme activity (Menéndez-Conejero et al. 2017). The similarity between these proteins and bacteriophage tailspike proteins highlights the versatility of the RHBH structure to perform different functions, adapting through distinct evolutionary trajectories.

Right-handed β-helices exhibit spatial and electrostatic complementarity with glycans

The RHBH structures in our dataset that have bound glycans show, with a few exceptions, a glycan bound along the central cleft (PB1, Figs 1 and 3). The PB1 parallel β-sheet stack creates the main cleft of the right-handed β-helix, supporting an elongated and slightly twisted pocket that is complementary for an elongated glycan substrate. In our RHBH dataset of 151 proteins, most are identified as glycan-binding or having adhesin functions; however, only 30 proteins have experimentally confirmed glycans positioned within their structures. This low number is likely due to the technical challenges in synthesizing, purifying, and determining the structures of glycans, hindering a full understanding of glycan-protein interactions. The substrate is also observed to bind the interstitial space between two RHBH domains instead of or in addition to binding in the central cleft (PB1) (e.g. PDBs 2inv [2inu], 4urr, 3eqn, 7w1e). In each of these structures, the T1 and T3 loops are longer than a typical turn between β-sheets, even showing distinct secondary structures (e.g. 2inv) or insertions of protein domains (e.g. PDB 7w1e). A unique type of interstitial glycan-binding mode is also observed in β-1,3-glucanase PDBs 3eqn and 4tz1 (Figs 3E and S3C). These RHBH proteins consist of a single protein chain with two RHBH monomer domains that appear as a dimerization of two RHBH chains. The glycan substrate is bound and cleaved by residues on a long loop linking the two domains, stabilized by the backbone β-helix of both RHBH domains.

Loops adjacent to the main cleft (T1, T3, Fig. 1A) are the main drivers of the formation of the binding pocket shape and, in some cases, catalytic activity (Figs 3 and S3). The foundational form of the RHBH consisting of β-sheet and loop repeats facilitates structural complementarity to adapt to glycan binding (van Santen et al. 1999). The loops are flexible to accommodate diverse glycan structures and can also withstand insertions and deletions while maintaining the same RHBH fold (Linse et al. 2020). The evolutionary freedom of loop mutability highlights their role as a key reason the RHBH is able to conserve its overall structure while binding heterogeneous substrates, accommodating diverse glycan decorations such as acetyl or phosphate groups, and participating in multiple mechanisms that act on different glycosidic bonds.

The RHBH architecture can support positively and negatively charged active sites to accommodate versatile acid–base chemistry to act on diverse, highly charged glycans (Fig. 3). As described, the predominant carbohydrate-activating enzyme mechanisms of the RHBH scaffold are polysaccharide lyases (PL) and glycoside hydrolases (GH). GH invariably uses a negatively charged active site (Fig. 3A–C) consisting of aspartate and glutamate for general acid–base catalysis involving water, while PL invariably uses positively charged arginine and lysine (Fig. 3D–F) (Drula et al. 2022). PL catalysis is often supported by a calcium ion bound to a cluster of adjacent aspartate residues (e.g. PDB 1ofl and 4ew9 (Fig. 3A and C)). Glycans are widely regarded as some of the most hydrophilic, highly charged molecules in biology (e.g. the glycosaminoglycan heparin, the highest-charged polymer density in the animal kingdom (Nelson et al. 2008) (Figs 3A and S3A). Complementary electrostatics that are capable of favorably neutralizing charge for productive binding and favorable catalysis is a key feature of RHBHs that enable glycan interaction (Varki 1993, 2017). The oppositely charged chemistries of the two main mechanisms supported by the RHBH illustrate the diversity of glycan composition that can be complemented for polysaccharide cleavage or modification.

The hydrogen bonding capability of the RHBHs, mediated by loop and β-sheet repeats, is a key feature that facilitates versatile glycan interactions (Fig. S3). Due to their chemical properties, glycans are typically able to participate in hydrogen bonding, mediating strong bonding to carbohydrate-interacting enzymes (García-Hernández and Hernández-Arana 1999; Gabius et al. 2011). Formation of the β-sheet helix architecture requires parallel hydrogen bonding, often resulting in repeating residues that form stacks either of charged residues such as solvent-facing asparagine ladders or internal aromatic residues, or both (e.g. PDB 3vmv)

Downloaded from https://academic.oup.com/glycob/article/34/7/cwae037/7676630 by NIH - Health Services Research Library user on 24 January 2025

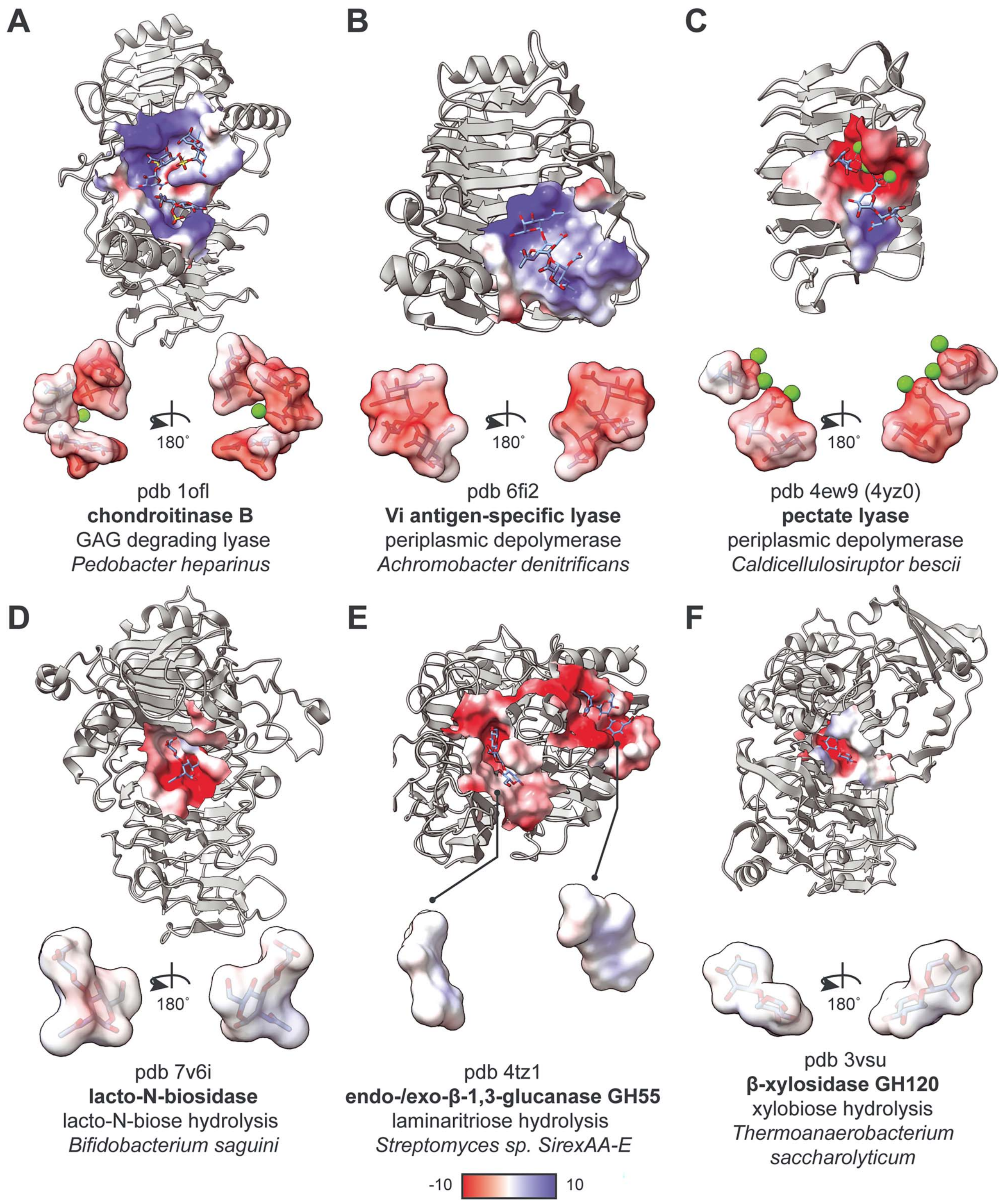

Fig. 3. Electrostatic complementarity of RHBH active sites and glycan substrates. A–C) Top, Coulombic surface potential of active site residues from representative polysaccharide lyase right-handed β-helix enzymes. Bottom, Coulombic surface potential of glycan substrates: (A) three molecules of dermatan sulfate, (B) O-acetylated vi antigen produced by *salmonella enterica* serovar Typhi, (C) trigalacturonic acid. In (A) and (C) calcium ions are represented as green spheres. D–F) Top, Coulombic surface potential of active site residues from representative glycoside hydrolase right-handed β-helix enzymes. Bottom, Coulombic surface potential of glycan substrates: (D) lacto-N-biose, (E) laminaritriose, (F) xylobiose. In all panels Coulombic surface potential is represented by red as -10 kcal/(mol·e) at 298 K and blue as 10 kcal/(mol·e) at 298 K. The structures were analyzed in ChimeraX (Pettersen et al. 2004). Coulombic electrostatic surface was calculated with default parameters as implemented in ChimeraX.

Downloaded from https://academic.oup.com/glycob/article/34/7/cwae037/7676630 by NIH - Health Services Research Library user on 24 January 2025

(Jenkins and Pickersgill 2001; Kajava and Steven 2006b). Ultimately, the RHBH represents a protein fold that is able to conserve a basic structural unit that facilitates multiple strategies favorable for glycan binding.

Challenges and opportunities in understanding RHBH-glycan interactions

While the increasing number of structurally characterized proteins has improved our understanding of RHBH-glycan interactions, there are still multiple open questions in this field. The vast amounts of structural data available can be used to provide information about substrate binding and reaction mechanisms, but many RHBH structures do not have any bound substrate, making it difficult to draw conclusions about the protein's function, especially when the substrate is poorly characterized or unknown. In addition, substrate specificity is one of the key features of RHBH-containing proteins that can be exploited in biotechnology applications, but the sequence and structural determinants of RHBH substrate specificity are still poorly understood. Lastly, this review focused on the RHBH, but various other β-solenoid domains exist across nature, for example the left-handed β-solenoids of Lamprey Variable Lymphocyte Receptors (PDB 6bxa and 3 g39) (Velikovsky et al. 2009; Gunn et al. 2018) and the β-propeller domain, which are both known to interact with glycans (Bonnardel et al. 2019). This highlights the broader biological significance of proteins with repeating β-solenoid folds, and further analyses of these families may provide insights into the general evolution and structural determinants of glycan-binding β-solenoids.

Understanding RHBH function at a structural and mechanistic level is an essential step in being able to take advantage of the RHBH in biomedical and biotechnology contexts. Work is already being done to develop bioassays with proteins that can be used to detect glycans from specific bacterial pathogens (Hsu and Mahal 2006; Geissner et al. 2014), and understanding phage tailspike specificity is a key step in the development of new phage therapies (Strathdee et al. 2023). Understanding the specific sequence and structural features that drive glycan binding in RHBHs can contribute significantly to these processes, leading to more efficient design and application of these assays and therapies. RHBH-containing proteins can also be used as the basis for new molecular glycobiology tools that can selectively recognize glycans and hydrolyze them at specific sites. Overall, an expanded understanding of the structural and mechanistic properties of this family of glycan-binding proteins will provide a foundation for multiple fields of research to build upon.

Acknowledgments

This work utilized the computational resources of the NIH HPC Biowulf cluster. (http://hpc.nih.gov).

Supplementary material

Supplementary material is available at *Glycobiology Journal* online.

Author contributions

Audrey A. Burnim (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Software [equal], Validation [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal]), Keith Dufault-Thompson (Data curation [equal], Methodology [equal], Supervision [equal], Writing—original draft [equal], Writing—review & editing [equal]), and Xiaofang Jiang (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Methodology [equal], Resources [equal], Supervision [equal], Validation [equal], Writing—original draft [equal], Writing—review & editing [equal]).

Funding

A.A.B., K.D., and X.J. are supported by the Intramural Research Program of the NIH, National Library of Medicine.

Conflict of interest statement

The authors declare that there are no conflicts of interest.

Data availability

The authors confirm that the data supporting the findings of this review are available within the article and its supplementary materials.

References

Abdelkader K, Gutiérrez D, Latka A, Boeckaerts D, Drulis-Kawa Z, Criel B, Gerstmans H, Safaan A, Khairalla AS, Gaber Y, et al. The specific capsule depolymerase of phage PMK34 sensitizes *Acinetobacter baumannii* to serum killing. *Antibiotics (Basel)*. 2022:**11**(5):677.

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol*. 2019:**37**(4):420–423.

Bonnardel F, Kumar A, Wimmerova M, Lahmann M, Perez S, Varrot A, Lisacek F, Imberty A. Architecture and evolution of blade assembly in β-propeller lectins. *Structure*. 2019:**27**(5): 764–775.e3.

Bonnardel F, Mariethoz J, Pérez S, Imberty A, Lisacek F. LectomeXplore, an update of UniLectin for the discovery of carbohydrate-binding proteins based on a new lectin classification, *Nucleic Acids Res*. 2021;**49**(D1):D1548-D1554.

Chaudhuri I, Söding J, Lupas AN. Evolution of the β-propeller fold. *Proteins*. 2008:**71**(2):795–803.

Cheng H, Schaeffer RD, Liao Y, Kinch LN, Pei J, Shi S, Kim B-H, Grishin NV. ECOD: an evolutionary classification of protein domains. *PLoS Comput Biol*. 2014:**10**(12):e1003926.

Ciccarelli FD, Copley RR, Doerks T, Russell RB, Bork P. CASH—a β-helix domain widespread among carbohydrate-binding proteins. *Trends Biochem Sci*. 2002:**27**(2):59–62.

Close DW, D'Angelo S, Bradbury ARM. A new family of β-helix proteins with similarities to the polysaccharide lyases. *Acta Crystallogr D Biol Crystallogr*. 2014:**70**(10):2583–2592.

Davies G, Henrissat B. Structures and mechanisms of glycosyl hydrolases. *Structure*. 1995:3(9):853–859.

DeMarco ML, Woods RJ. Structural glycobiology: a game of snakes and ladders. *Glycobiology*. 2008:**18**(6):426–440.

Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res*. 2022:**50**(D1):D571–D577.

Emsley P, Charles IG, Fairweather NF, Isaacs NW. Structure of *Bordetella pertussis* virulence factor P.69 pertactin. *Nature*. 1996:**381**(6577):90–92.

Fries M, Ihrig J, Brocklehurst K, Shevchik VE, Pickersgill RW. Molecular basis of the activity of the phytopathogen pectin methylesterase. *EMBO J*. 2007:**26**(17):3879–3887.

Downloaded from https://academic.oup.com/glycob/article/34/7/cwae037/7676630 by NIH - Health Services Research Library user on 24 January 2025

Gabius H-J, André S, Jiménez-Barbero J, Romero A, Solís D. From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem Sci*. 2011:**36**(6):298–313.

García-Hernández E, Hernández-Arana A. Structural bases of lectin-carbohydrate affinities: comparison with protein-folding energetics. *Protein Sci*. 1999:**8**(5):1075–1086.

Garron M-L, Cygler M. Uronic polysaccharide degrading enzymes. *Curr Opin Struct Biol*. 2014:**28**:87–95.

Geissner A, Anish C, Seeberger PH. Glycan arrays as tools for infectious disease research. *Curr Opin Chem Biol*. 2014:**18**:38–45.

Gunn RJ, Herrin BR, Acharya S, Cooper MD, Wilson IA. VLR recognition of TLR5 expands the molecular characterization of protein antigen binding by non-Ig-based antibodies. *J Mol Biol*. 2018:**430**(9):1350–1367.

Haghverdi L, Büttner M, Wolf FA, Buettner F, Theis FJ. Diffusion pseudotime robustly reconstructs lineage branching. *Nat Methods*. 2016:**13**(10):845–848.

Hie BL, Yang KK, Kim PS. Evolutionary velocity with protein language models predicts evolutionary dynamics of diverse proteins. *Cell Syst*. 2022:**13**(4):274–285.e6.

Hogan MC, Griffin MD, Rossetti S, Torres VE, Ward CJ, Harris PC. PKHDL1, a homolog of the autosomal recessive polycystic kidney disease gene, encodes a receptor with inducible T lymphocyte expression. *Hum Mol Genet*. 2003:**12**(6):685–698.

Hsu K-L, Mahal LK. A lectin microarray approach for the rapid analysis of bacterial glycans. *Nat Protoc*. 2006:**1**(2):543–549.

Iengar P, Joshi NV, Balaram P. Conformational and sequence signatures in β helix proteins. *Structure*. 2006:**14**(3):529–542.

Irmscher T, Roske Y, Gayk I, Dunsing V, Chiantia S, Heinemann U, Barbirz S. *Pantoea stewartii* WceF is a glycan biofilm-modifying enzyme with a bacteriophage tailspike-like fold. *J Biol Chem*. 2021:**296**:100286.

Jenkins J, Pickersgill R. The architecture of parallel β-helices and related folds. *Prog Biophys Mol Biol*. 2001:**77**(2):111–175.

Jenkins J, Mayans O, Pickersgill R. Structure and evolution of parallel β-helix proteins. *J Struct Biol*. 1998:**122**(1-2):236–246.

Kajava AV, Steven AC. The turn of the screw: variations of the abundant β-solenoid motif in passenger domains of type V secretory proteins. *J Struct Biol*. 2006a:**155**(2):306–315.

Kajava AV, Steven AC. β-rolls, β-helices, and other β-solenoid proteins. *Adv Protein Chem*. 2006b:**73**:55–96.

Kipreos ET, Pagano M. The F-box protein family. *Genome Biol***1**:REVIEWS3002. 2000:**1**(5):reviews3002.1.

Klemm P, Schembri MA. Bacterial adhesins: function and structure. *Int J Med Microbiol*. 2000:**290**(1):27–35.

Knecht LE, Veljkovic M, Fieseler L. Diversity and function of phage encoded depolymerases. *Front Microbiol*. 2019:**10**:2949.

Kumanomidou T, Nishio K, Takagi K, Nakagawa T, Suzuki A, Yamane T, Tokunaga F, Iwai K, Murakami A, Yoshida Y, et al. The structural differences between a glycoprotein specific F-box protein Fbs1 and its homologous protein FBG3. *PLoS One*. 2015:**10**(10): e0140366.

Latka A, Maciejewska B, Majkowska-Skrobek G, Briers Y, Drulis-Kawa Z. Bacteriophage-encoded virion-associated enzymes to overcome the carbohydrate barriers during the infection process. *Appl Microbiol Biotechnol*. 2017:**101**(8):3103–3119.

Lee I-M, Tu I-F, Yang F-L, Wu S-H. Bacteriophage tail-spike proteins enable detection of pseudaminic-acid-coated pathogenic bacteria and guide the development of antiglycan antibodies with cross-species antibacterial activity. *J Am Chem Soc*. 2020:**142**(46): 19446–19450.

Lee S, Inzerillo S, Lee GY, Bosire EM, Mahato SK, Song J. Glycan-mediated molecular interactions in bacterial pathogenesis. *Trends Microbiol*. 2022:**30**(3):254–267.

Linse S, Thulin E, Nilsson H, Stigler J. Benefits and constrains of covalency: the role of loop length in protein stability and ligand binding. *Sci Rep*. 2020:**10**(1):20108.

Luis AS, Briggs J, Zhang X, Farnell B, Ndeh D, Labourel A, Baslé A, Cartmell A, Terrapon N, Stott K, et al. Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic *Bacteroides*. *Nat Microbiol*. 2018:**3**(2):210–219.

Marabini R, Condezo GN, Krupovic M, Menéndez-Conejero R, Gómez-Blanco J, San Martín C. Near-atomic structure of an atadenovirus reveals a conserved capsid-binding motif and intergenera variations in cementing proteins. *Sci Adv*. 2021:**7**(14):eabe6008.

Menéndez-Conejero R, Nguyen TH, Singh AK, Condezo GN, Marschang RE, van Raaij MJ, San Martín C. Structure of a reptilian adenovirus reveals a phage tailspike fold stabilizing a vertebrate virus capsid. *Structure*. 2017:**25**(10):1562–1573.e5.

Mesdaghi S, Price RM, Madine J, Rigden DJ. Deep learning-based structure modelling illuminates structure and function in uncharted regions of β-solenoid fold space. *J Struct Biol*. 2023:**215**(3):108010.

Müller C, Zhang L, Zipfel S, Topitsch A, Lutz M, Eckert J, Prasser B, Chami M, Lü W, Du J, et al. Molecular interplay of an assembly machinery for nitrous oxide reductase. *Nature*. 2022:**608**(7923): 626–631.

Neelamegham S, Aoki-Kinoshita K, Bolton E, Frank M, Lisacek F, Lütteke T, O'Boyle N, Packer NH, Stanley P, Toukach P, et al. Updates to the symbol nomenclature for Glycans guidelines. *Glycobiology*. 2019:**29**(9):620–624.

Nelson DL, Lehninger AL, Cox MM. *Lehninger principles of biochemistry*. Macmillan: London, United Kingdom; 2008

Nielsen H. Predicting secretory proteins with SignalP. *Methods Mol Biol*. 2017:**1611**:59–73.

Paxman JJ, Lo AW, Sullivan MJ, Panjikar S, Kuiper M, Whitten AE, Wang G, Luan C-H, Moriel DG, Tan L, et al. Unique structural features of a bacterial autotransporter adhesin suggest mechanisms for interaction with host macromolecules. *Nat Commun*. 2019:**10**(1):1967.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004:**25**(13): 1605–1612.

Pijning T, van Pouderoyen G, Kluskens L, van der Oost J, Dijkstra BW. The crystal structure of a hyperthermoactive exopolygalacturonase from *Thermotoga maritima* reveals a unique tetramer. *FEBS Lett*. 2009:**583**(22):3665–3670.

Pohlschröder M, Prinz WA, Hartmann E, Beckwith J. Protein translocation in the three domains of life: variations on a theme. *Cell*. 1997:**91**(5):563–566.

Qiu L-R, Xu R-R, Tang J-H, Zhou J-H. Possible PKHD1 hot-spot mutations related to early kidney function failure or hepatofibrosis in Chinese children with ARPKD: a retrospective single center cohort study and literature review. *Curr Med Sci*. 2020:**40**(5): 835–844.

Rice CJ, Kelly SA, O'Brien SC, Melaugh EM, Ganacias JCB, Chai ZH, Gilmore BF, Skvortsov T. Novel phage-derived Depolymerase with activity against *Proteus mirabilis* biofilms. *Microorganisms*. 2021:**9**(10):2172.

Roche DB, Viet PD, Bakulina A, Hirsh L, Tosatto SCE, Kajava AV. Classification of β-hairpin repeat proteins. *J Struct Biol*. 2018:**201**(2): 130–138.

Rozeboom HJ, Bjerkan TM, Kalk KH, Ertesvåg H, Holtan S, Aachmann FL, Valla S, Dijkstra BW. Structural and mutational characterization of the catalytic A-module of the mannuronan C-5-epimerase AlgE4 from *Azotobacter vinelandii*. *J Biol Chem*. 2008:**283**(35): 23819–23828.

van Santen Y, Benen JA, Schröter KH, Kalk KH, Armand S, Visser J, Dijkstra BW. 1.68-Å crystal structure of endopolygalacturonase II from *aspergillus Niger* and identification of active site residues by site-directed mutagenesis. *J Biol Chem*. 1999:**274**(43): 30474–30480.

Sehnal D, Bittrich S, Deshpande M, Svobodová R, Berka K, Bazgier V, Velankar S, Burley SK, Koča J, Rose AS. Mol* viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res*. 2021:**49**(W1):W431–W437.

Sillitoe I, Bordin N, Dawson N, Waman VP, Ashford P, Scholes HM, Pang CSM, Woodridge L, Rauer C, Sen N, et al. CATH:

Downloaded from https://academic.oup.com/glycob/article/34/7/cwae037/7676630 by NIH - Health Services Research Library user on 24 January 2025

increased structural coverage of functional space. *Nucleic Acids Res*. 2021:**49**(D1):D266–D273.

Simpson DJ, Sacher JC, Szymanski CM. Exploring the interactions between bacteriophage-encoded glycan binding proteins and carbohydrates. *Curr Opin Struct Biol*. 2015:**34**:69–77.

Soto GE, Hultgren SJ. Bacterial adhesins: common themes and variations in architecture and assembly. *J Bacteriol*. 1999:**181**(4): 1059–1071.

Steinbacher S, Miller S, Baxa U, Budisa N, Weintraub A, Seckler R, Huber R. Phage P22 tailspike protein: crystal structure of the headbinding domain at 2.3 Å, fully refined structure of the endorhamnosidase at 1.56 Å resolution, and the molecular basis of O-antigen recognition and cleavage. *J Mol Biol*. 1997:**267**(4):865–880.

Strathdee SA, Hatfull GF, Mutalik VK, Schooley RT. Phage therapy: from biological mechanisms to future directions. *Cell*. 2023:**186**(1): 17–31.

Timoshina OY, Kasimova AA, Shneider MM, Matyuta IO, Nikolaeva AY, Evseev PV, Arbatsky NP, Shashkov AS, Chizhov AO, Shelenkov AA, et al. Friunavirus phage-encoded depolymerases specific to different capsular types of *Acinetobacter baumannii*. *Int J Mol Sci*. 2023:**24**(10):9100.

Todd AE, Orengo CA, Thornton JM. Evolution of function in protein superfamilies, from a structural perspective. *J Mol Biol*. 2001:**307**(4):1113–1143.

Togayachi A, Dae K-Y, Shikanai T, Narimatsu H. A database system for glycogenes (GGDB). In: Taniguchi N, Suzuki A, Ito Y, Narimatsu H, Kawasaki T, Hase S editors. Experimental glycoscience Japan Springer, 2008. p.423–425.

Varki A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*. 1993:**3**(2):97–130.

Varki A. Biological roles of glycans. *Glycobiology*. 2017:**27**(1): 3–49.

Velikovsky CA, Deng L, Tasumi S, Iyer LM, Kerzic MC, Aravind L, Pancer Z, Mariuzza RA. Structure of a lamprey variable lymphocyte receptor in complex with a protein antigen. *Nat Struct Mol Biol*. 2009:**16**(7):725–730.

Wolfram F, Kitova EN, Robinson H, Walvoort MTC, Codée JDC, Klassen JS, Howell PL. Catalytic mechanism and mode of action of the periplasmic alginate epimerase AlgG. *J Biol Chem*. 2014:**289**(9): 6006–6019.

Yoder MD, Keen NT, Jurnak F. New domain motif: the structure of pectate lyase C, a secreted plant virulence factor. *Science*. 1993a:**260**(5113):1503–1507.

Yoder MD, Lietzke SE, Jurnak F. Unusual structural features in the parallel β-helix in pectate lyases. *Structure*. 1993b:**1**(4):241–251.

Ziegler WH, Soetje B, Marten LP, Wiese J, Burute M, Haffner D. Fibrocystin is essential to cellular control of adhesion and epithelial morphogenesis. *Int J Mol Sci*. 2020:**21**(14):5140.

Downloaded from https://academic.oup.com/glycob/article/34/7/cwae037/7676630 by NIH - Health Services Research Library user on 24 January 2025