

Supplementary Information

GlycanInsight: an open platform for carbohydrate-binding pocket prediction and characterization

Qinyu Chu,^{†abcd} Xinheng He,^{†cd} Xinyi Tan,^{†c}, Zhiyong Gu,^{acd} Yin Luo,^c Zifu Huang,^c Mingyue Zheng,^{*acd} and Xi Cheng^{*bcd}

^a School of Pharmaceutical Science and Technology, Hangzhou Institute of Advanced Study, Hangzhou 330106, China

^b Medicinal Chemistry and Bioinformatics Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

^c State Key Laboratory of Drug Research and Information Management Office, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China

^d University of Chinese Academy of Sciences, Beijing 100049, China

† These three authors contributed equally to this work.

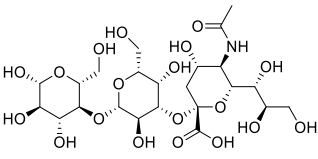
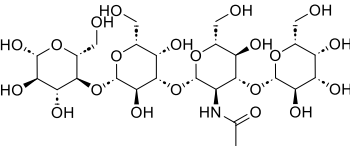
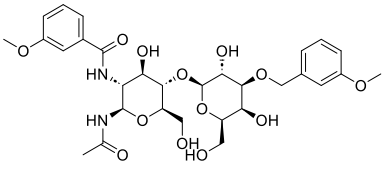
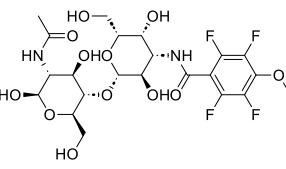
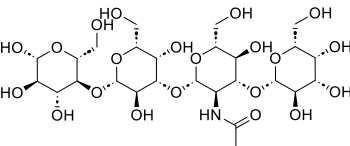
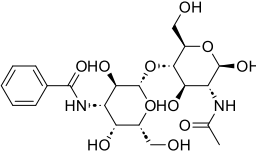
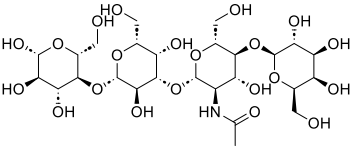
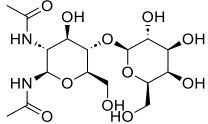
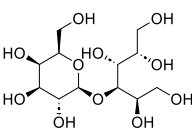
* To whom correspondence should be addressed.

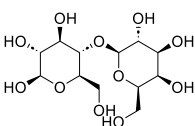
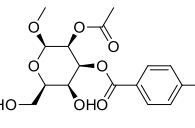
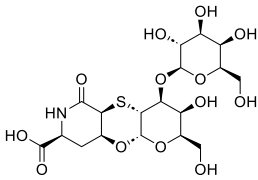
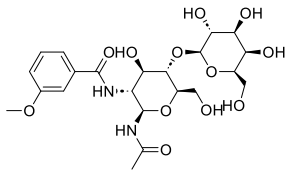
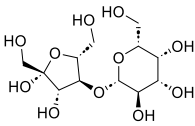
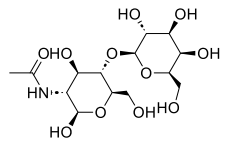
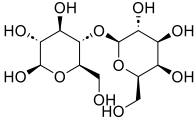
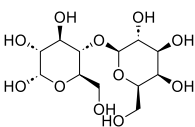
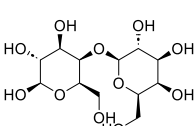
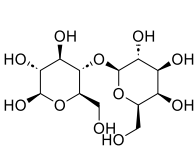
Emails: xicheng@simm.ac.cn; myzheng@simm.ac.cn

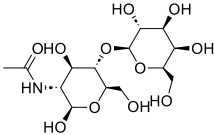
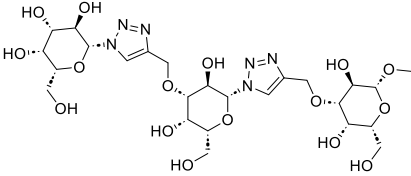
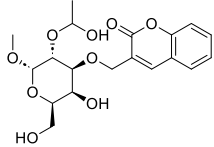
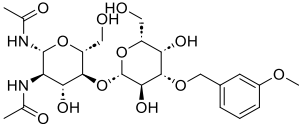
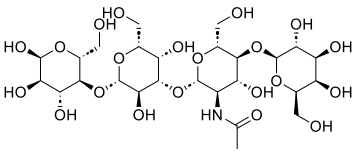
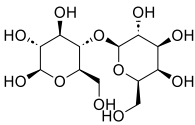
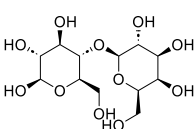
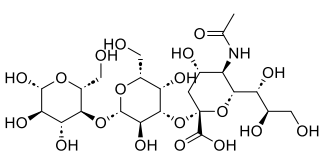
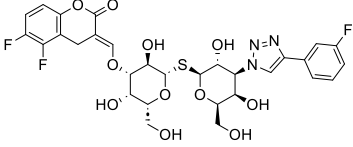
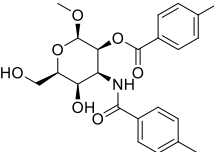
Table S1. GlycanInsight-predicted carbohydrate-binding pockets of pig pancreatic α -amylase (UniProt ID: P00690; PDB ID: 1ua3).

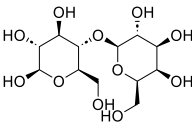
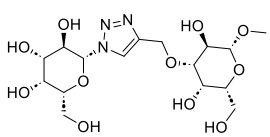
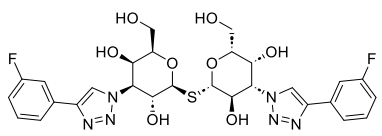
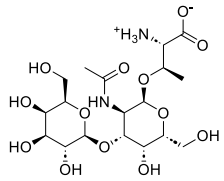
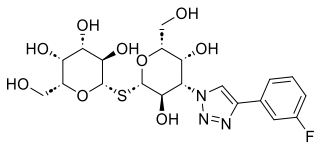
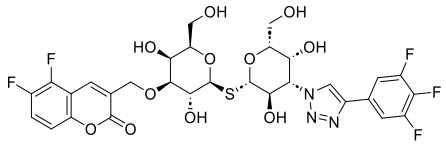
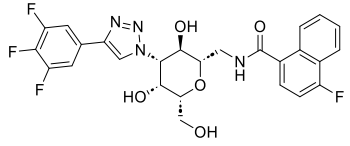
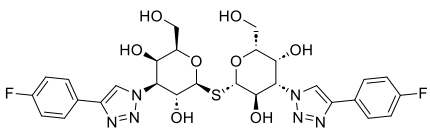
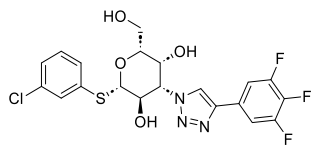
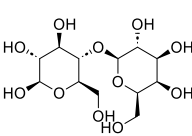
Pocket ID	Residue	Probability (0-1)	Conservation (0-1)	Comments
1	W58	0.715	0.807	catalytic pocket (active site)
	W59	0.757	0.624	
	E60	0.683	0.194	
	Y62	0.765	0.935	
	Q63	0.750	0.740	
	V98	0.593	0.772	
	H101	0.745	0.865	
	Y151	0.767	0.442	
	L162	0.726	0.606	
	V163	0.702	0.211	
	G164	0.531	0.566	
	L165	0.745	0.695	
	R195	0.751	0.897	
	D197	0.734	0.926	
	A198	0.762	0.75	
	K200	0.764	0.774	
	H201	0.766	0.754	
	E233	0.736	0.803	
	I235	0.745	0.416	
	L237	0.591	0.000	
	E240	0.740	0.412	
	H299	0.657	0.000	
	D300	0.771	0.000	
	H305	0.664	0.000	
	G306	0.662	0.000	
	A307	0.706	0.000	
	G309	0.507	0.000	
	E352	0.509	0.000	
2	P130	0.576	0.000	secondary pocket
	Y131	0.586	0.476	
	S132	0.581	0.219	
	W134	0.559	0.083	
	D135	0.576	0.458	
	K140	0.664	0.000	
	E171	0.709	0.305	
	K172	0.540	0.256	
	Y174	0.595	0.206	
	R176	0.542	0.570	
3	W203	0.513	0.200	secondary pocket
	G205	0.539	0.203	
4	E272	0.504	0.000	secondary pocket
	Y276	0.614	0.000	
5	N279	0.521	0.000	secondary pocket
	W284	0.617	0.000	
6	A318	0.635	0.000	secondary pocket
	K322	0.596	0.000	
	W388	0.56	0.000	
	E390	0.595	0.000	
7	T376	0.532	0.000	secondary pocket
	T377	0.582	0.000	

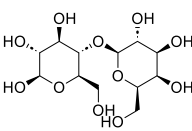
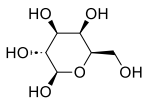
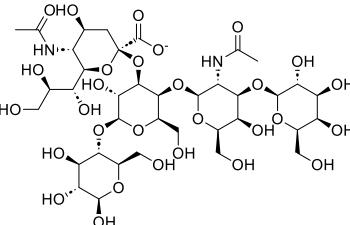
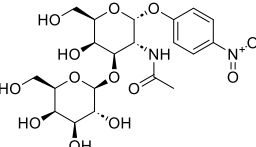
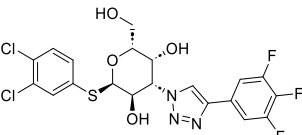
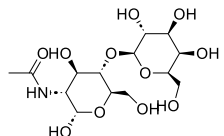
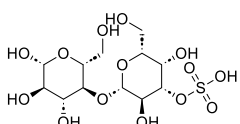
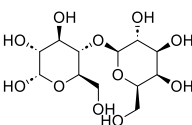
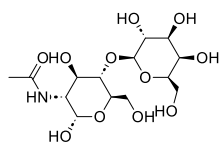
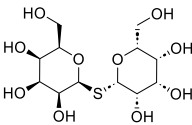
Table S2. GlycanInsight-suggested ligand information of human Galectin-3.

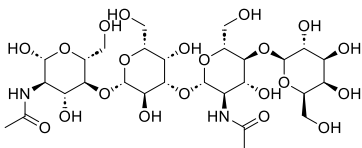
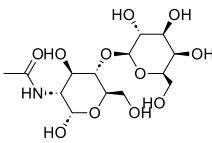
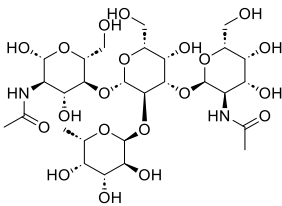
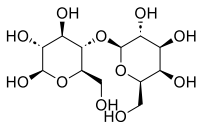
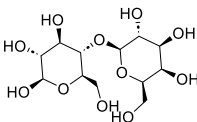
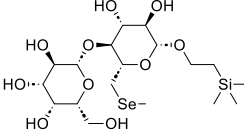
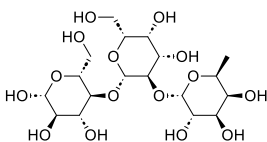
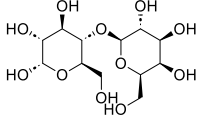
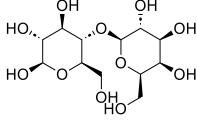
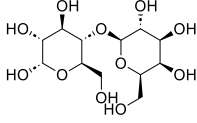
ID	PDB ID	Protein name	Ligand	Alignment score
1	4lbo	Galectin-3		9
2	4lbj	Galectin-3		8.94
3	5nfb	Galectin-3		8.9
4	1kjr	Galectin-3		8.88
5	4lbm	Galectin-3		8.87
6	2xg3	Galectin-3		8.86
7	4lbk	Galectin-3		8.85
8	5nf7	Galectin-3		8.84
9	6fk2	Galectin-3		8.82

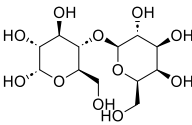
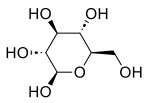
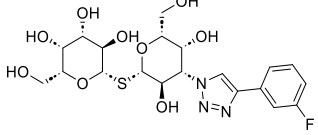
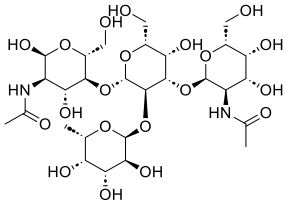
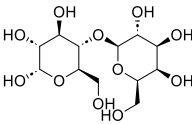
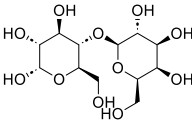
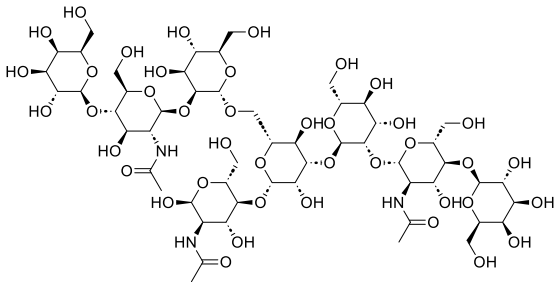
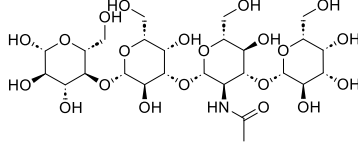
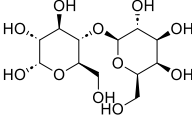
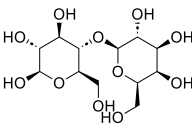
10	2nn8	Galectin-3		8.78
11	3t1l	Galectin-3		8.76
12	6g0v	Galectin-3		8.74
13	5nf9	Galectin-3		8.74
14	6b8k	Galectin-3		8.72
15	1kjl	Galectin-3		8.71
16	6y4c	Galectin-3		8.71
17	6h64	Galectin-3		8.7
18	6f6y	Galectin-3		8.68
19	6eym	Galectin-3		8.65

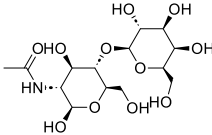
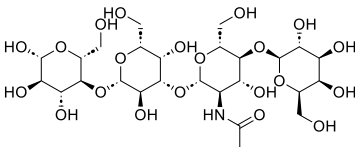
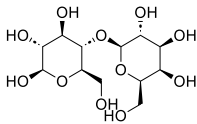
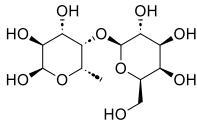
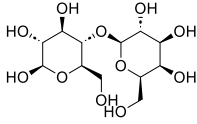
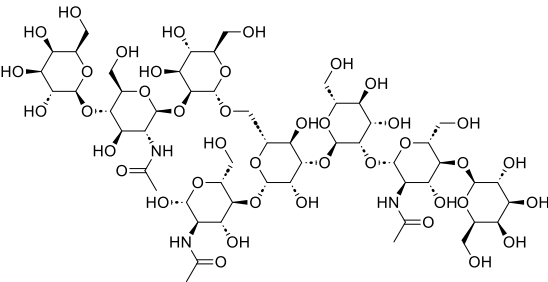
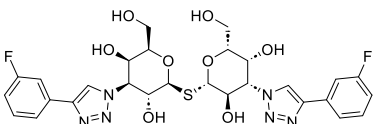
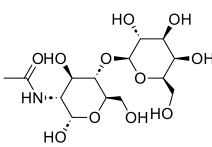
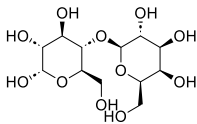
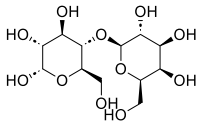
20	1a3k	Galectin-3		8.65
21	6q0q	Galectin-3		8.64
22	5exo	Galectin-3		8.62
23	5nfa	Galectin-3		8.59
24	4lbn	Galectin-3		8.57
25	6fof	Galectin-3		8.55
26	2nmn	Galectin-3		8.55
27	4lbl	Galectin-3		8.54
28	5oax	Galectin-3		8.52
29	3t1m	Galectin-3		8.52

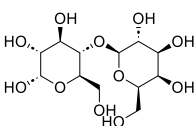
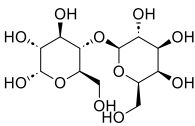
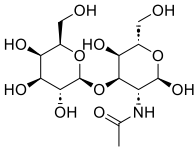
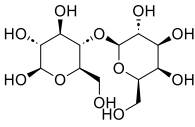
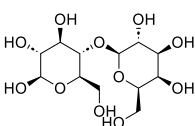
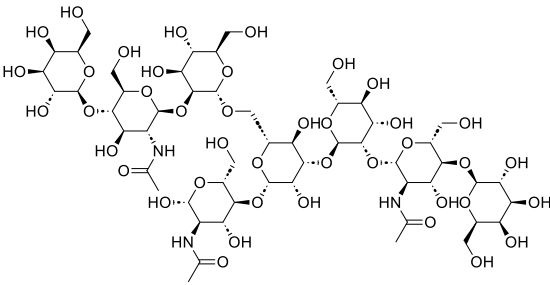
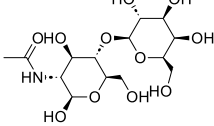
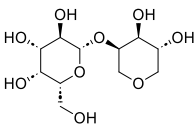
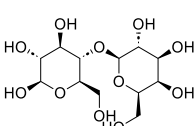
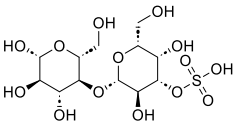
30	3zsj	Galectin-3		8.51
31	6q17	Galectin-3		8.51
32	5h9p	Galectin-3		8.51
33	3aya	Galectin-3		8.39
34	5h9r	Galectin-3		8.32
35	5ody	Galectin-3		8.31
36	6tf7	Galectin-3		8.3
37	5e8a	Galectin-3		8.29
38	6eog	Galectin-3		8.28
39	3wv6	Galectin-9		8.27

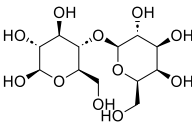
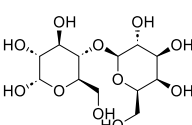
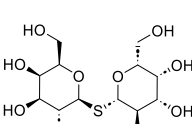
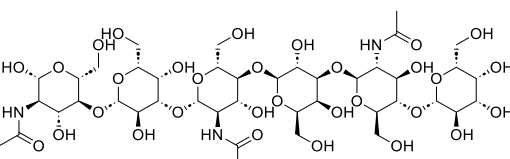
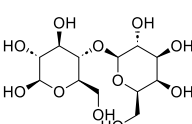
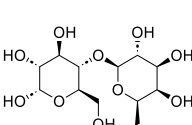
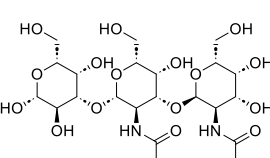
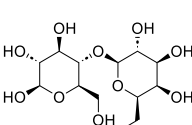
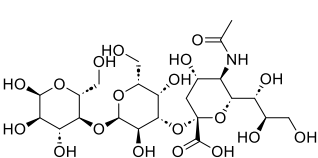
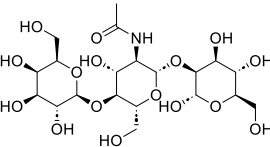
40	6y78	Galectin-3		8.24
41	2gal	Galectin-7		8.22
42	3ayc	Galectin-3		8.21
43	3ayd	Galectin-3		8.21
44	6eol	Galectin-3		8.19
45	6e20	Galectin-1		8.12
46	5duw	Galectin-4		8.10
47	2wsv	Porcine adenovirus 4 fiber protein		8.05
48	3nv2	Galectin-9		8.04
49	4jck	Galectin-3		8.03

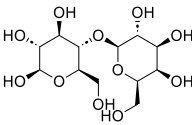
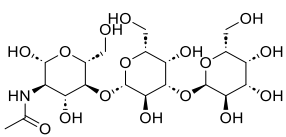
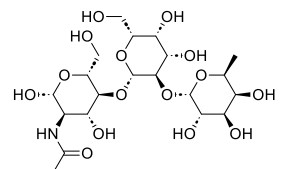
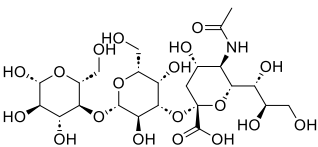
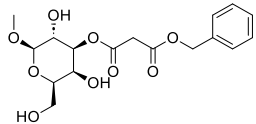
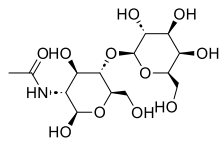
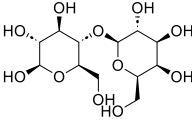
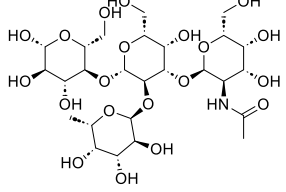
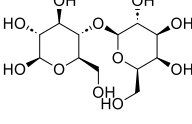
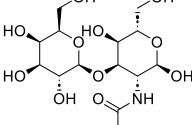
50	2d6o	Galectin-9		8.02
51	1slt	Galectin-1		8.02
52	3wg3	Cyclocybe cylindracea galectin		8.00
53	3vkl	Galectin-8		7.98
54	3vkm	Galectin-8		7.97
55	3wlu	Galectin-9		7.96
56	5dux	Galectin-4		7.95
57	5nle	Chicken GRIFIN		7.95
58	4gal	Galectin-7		7.94
59	2d6m	Galectin-9		7.94

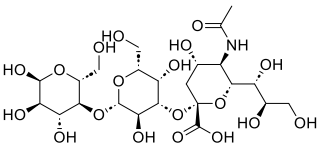
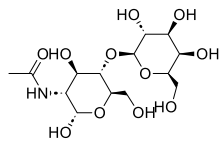
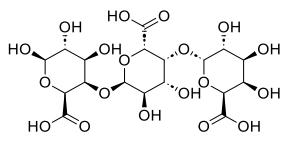
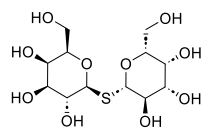
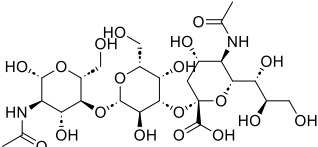
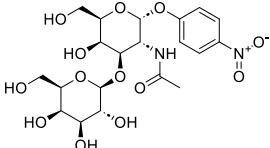
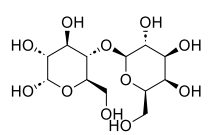
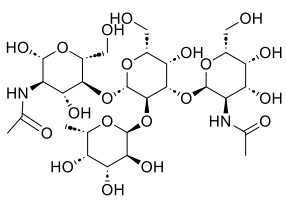
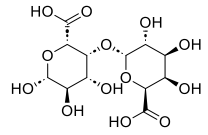
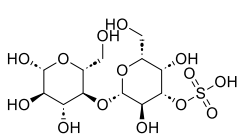
60	4ga9	Galectin-1		7.93
61	5glz	Tl-gal		7.93
62	5h9s	Galectin-7		7.93
63	5xfd	Cyclocybe cylindracea galectin		7.92
64	2eak	Galectin-9		7.9
65	5nld	Chicken GRIFIN		7.9
66	1slc	Galectin-1		7.88
67	5glt	Tl-gal		7.87
68	3wud	Xenopus laevis galectin-1b		7.81
69	5duv	Galectin-4		7.81

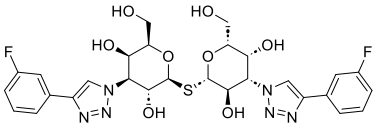
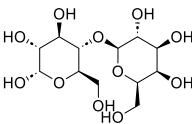
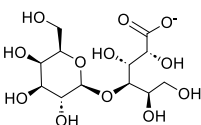
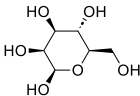
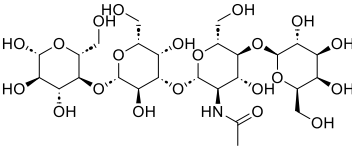
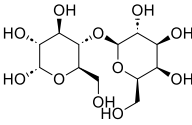
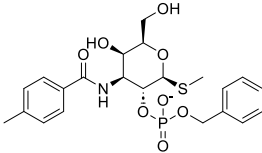
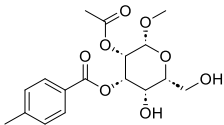
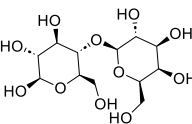
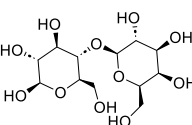
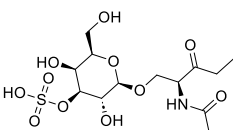
70	5glw	Tl-gal		7.8
71	2wt1	Porcine adenovirus 4 fiber protein		7.78
72	5ews	Galectin-2		7.77
73	3vv1	Caenorhabditis elegans galectin LEC-6		7.75
74	1wld	Congerin II		7.73
75	1slb	Galectin-1		7.72
76	5h9q	Galectin-7		7.7
77	5gal	Galectin-7		7.69
78	5nlh	Chicken GRIFIN		7.69
79	5jpg	Galectin-5		7.65

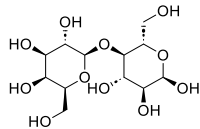
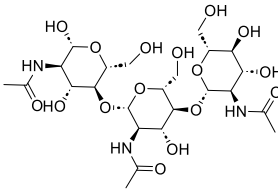
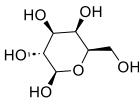
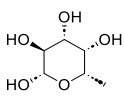
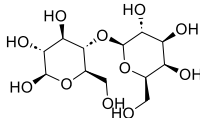
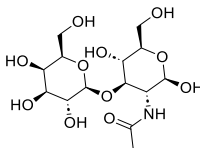
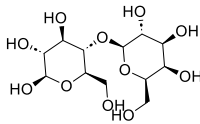
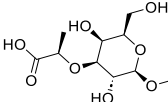
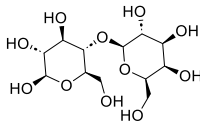
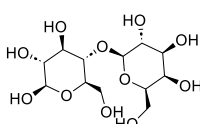
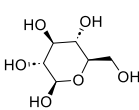
80	5nmj	Chicken GRIFIN		7.65
81	5nm1	Chicken GRIFIN		7.65
82	2d6p	Galectin-9		7.64
83	1c1l	Congerin I		7.64
84	1hlc	S-LAC Lectin		7.63
85	1sla	Galectin-1		7.63
86	2wt0	Porcine adenovirus 4 fiber protein		7.62
87	2wkk	Galectin-2		7.61
88	5dg1	Galectin-2		7.61
89	4ym2	Galectin-4		7.61

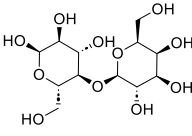
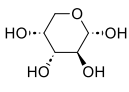
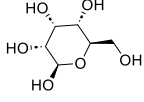
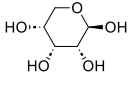
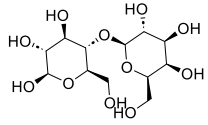
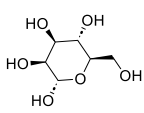
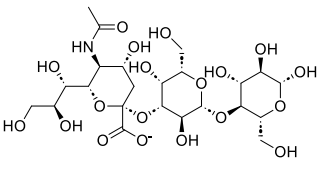
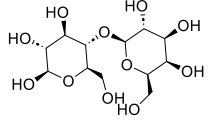
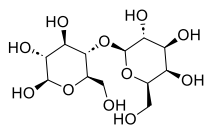
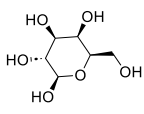
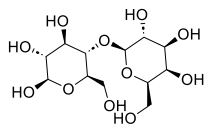
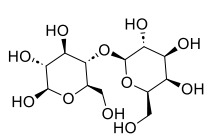
90	1is4	Congerin II		7.58
91	5dg2	Galectin-2		7.58
92	4jc1	Galectin-3		7.58
93	2wt2	Porcine adenovirus 4 fiber protein		7.58
94	1is3	Congerin II		7.57
95	1wlw	Congerin II		7.56
96	2eal	Galectin-9		7.55
97	6m5y	Galectin-1		7.54
98	1ww4	Cyclocybe cylindracea galectin		7.54
99	3nv3	Galectin-9		7.54

100	1ulc	Galectin-2		7.53
101	1ule	Galectin-2		7.53
102	1uld	Galectin-2		7.49
103	3nv4	Galectin-9		7.49
104	6w4z	Galectin-8		7.47
105	2d6n	Galectin-9		7.46
106	5gm0	TI-gal		7.46
107	1ulf	Galectin-2		7.46
108	1ww6	Cyclocybe cylindracea galectin		7.43
109	1ulg	Galectin-2		7.43

110	5glu	Tl-gal		7.42
111	1gan	Galectin-1		7.38
112	6kxb	Galectin-3		7.37
113	3oyw	Galectin-1		7.34
114	3vko	Galectin-8		7.33
115	3m3e	Cyclocybe aegerita galectin		7.29
116	3wuc	Xenopus laevis galectin-Va		7.27
117	3wg4	Cyclocybe cylindracea galectin		7.27
118	6kxa	Galectin-3		7.27
119	1ww5	Cyclocybe cylindracea galectin		7.25

120	4y24	Galectin-1		7.16
121	4bme	Galectin-8		7.13
122	3oy8	Galectin-1		7.01
123	6l6a	Galectin-10		7.00
124	5t7i	Galectin-8		6.99
125	5t7s	Galectin-8		6.96
126	3zxe	Galectin-7		6.9
127	3t2t	Galectin-1		6.84
128	6a63	Galectin-13		6.84
129	5gze	Galectin-8		6.83
130	6z6y	Galectin-8		6.83

131	3i8t	Galectin-4		6.82
132	2r0h	Fungal lectin CGL3		6.73
133	6l67	Galectin-10		6.71
134	6l6b	Galectin-10		6.69
135	4bmb	Galectin-8		6.64
136	5t7t	Galectin-8		6.63
137	6a62	Galectin-13		6.58
138	5vwg	Galectin-8		6.58
139	6a1t	Galectin-10		6.58
140	4fqz	Galectin-8		6.56
141	6l64	Galectin-10		6.56

142	3ap4	Galectin-8		6.56
143	6l6c	Galectin-10		6.55
144	6l68	Galectin-10		6.53
145	6qrn	Galectin-10		6.51
146	6a64	Galectin-13		6.42
147	1qkq	Galectin-10		6.40
148	3ap7	Galectin-8		6.39
149	3wg1	Cyclocybe cylindracea galectin		6.35
150	5gzf	Galectin-8		6.32
151	6n3r	Galectin-11		6.22
152	6ljr	Galectin-16		6.17
153	5gzg	Galectin-8		6.13

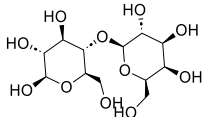
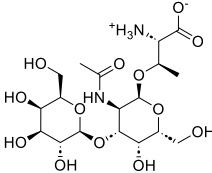
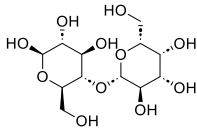
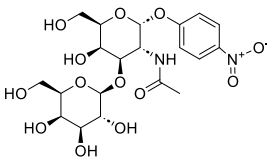
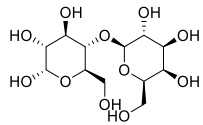
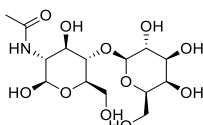
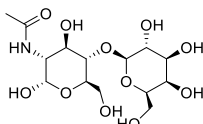
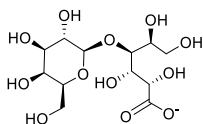
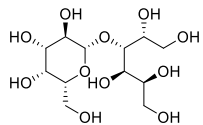
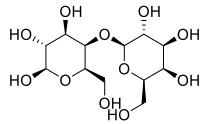
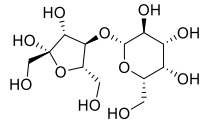
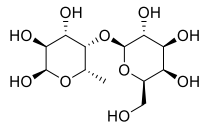
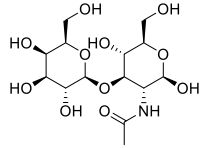
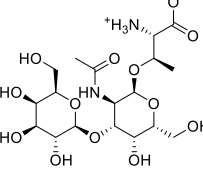
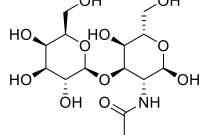
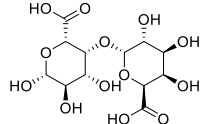
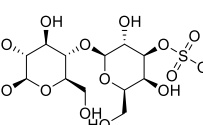
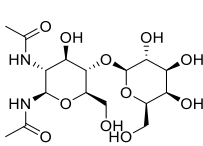
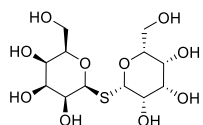
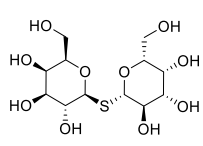
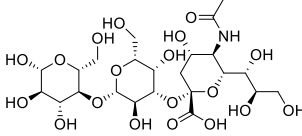
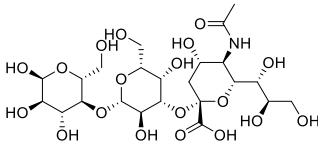
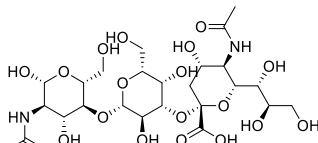
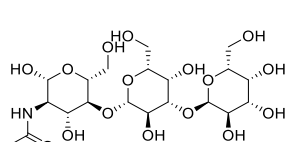
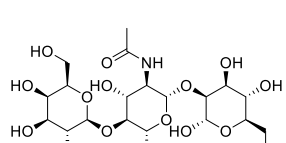
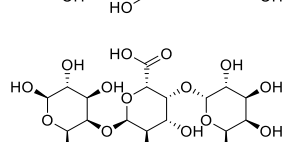
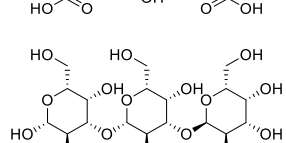
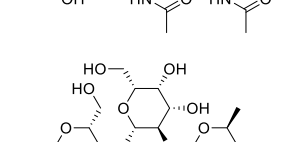
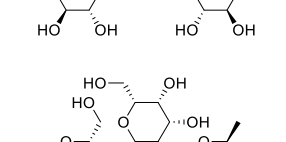
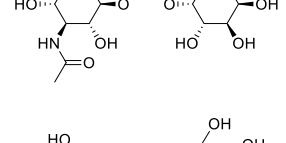
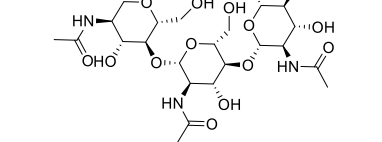
154	5gzd	Galectin-8		6.05
155	3afk	Cyclocybe cylindracea galectin		6
156	2zgm	Cyclocybe cylindracea galectin		5.94
157	3m3c	Cyclocybe cylindracea galectin		5.92

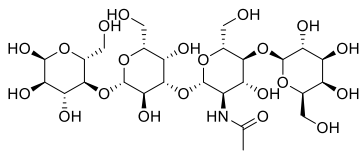
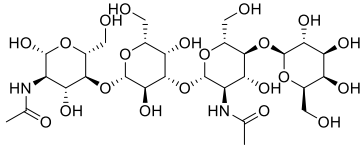
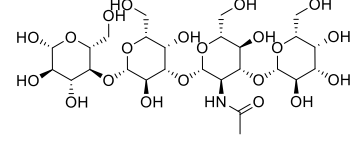
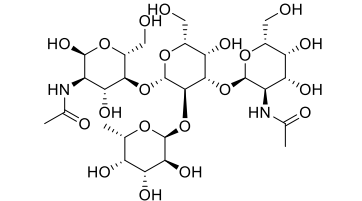
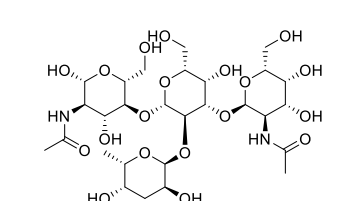
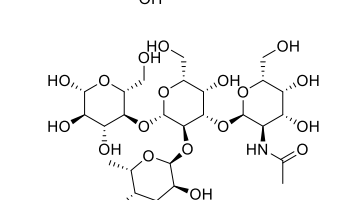
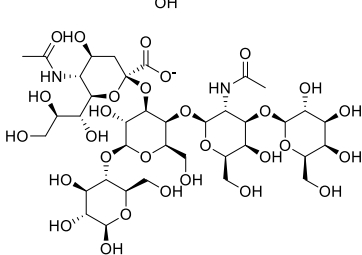
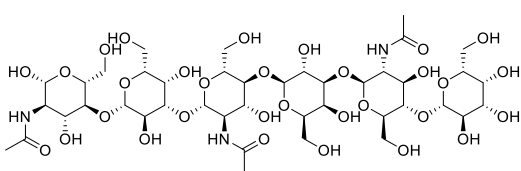
Table S3. GlycanInsight-suggested carbohydrate ligands that bind to human Galectin-3.

ID	GlyYouCan	IUPAC	Structure	Target
1		Gal		Galectin-7; Galectin-10; Galectin-11
2		Glc		Galectin-10; TI-gal;
3		Man		Galectin-10
4		All		Galectin-10
5		Fuc		Galectin-10
6		Ara		Galectin-10
7		RIP		Galectin-10
8	G84224TW	Gal(β1-4)Glc(β1-		Congerin I; Congerin II; Cyclocybe cylindracea galectin; Galectin-1; Galectin-2; Galectin-3; Galectin-4; Galectin-7; Galectin-8; Galectin-9; Galectin-10; Galectin-13; Galectin-16; S-LAC Lectin; TI-gal

9	G88362QR	Gal(β 1-4)Glc(α 1-		Chicken GRIFIN; Congerin II; Galectin-1; Galectin-2; Galectin-3; Galectin-5; Galectin-8; Galectin-9; Porcine adenovirus 4 fiber protein; Xenopus laevis galectin-Ib; Xenopus laevis galectin-Va Galectin-3; Galectin-7; Galectin-9; Porcine adenovirus 4 fiber protein; Tl-gal Galectin-1; Galectin-9
10	G00055MO	Gal(β 1-4)GlcNAc(β 1-		Galectin-3; Galectin-7; Galectin-9; Porcine adenovirus 4 fiber protein; Tl-gal
11	G79333MY	Gal(β 1-4)GlcNAc(α 1-		Galectin-1; Galectin-9
12	G18874TB	Gal(β 1-4)Glc- onic		Galectin-1
13	G26039ES	Gal(β 1-4)Glc-ol		Galectin-3
14	G08806VI	Gal(β 1-4)Gal(β 1-		Galectin-3
15	G57229GY	Gal(β 1-4)Fru(β 2-		Galectin-3

16	G61266PA	Gal(β 1-4)Fuc(α 1-		Caenorhabditis elegans galectin LEC-6; Galectin-2 Galectin-8
17	G00056MO	Gal(β 1-3)GlcNAc(β 1-		Galectin-8
18	G00031MO	Gal(β 1-3)GalNAc(α 1-		Cyclocybe aegerita galectin; Galectin-3
19	G01534TU	Gal(β 1-3)GalNAc(β 1-		Galectin-2; Galectin-9
20	G82552LT	GalA(α 1-4)GalA(β 1-		Galectin-3
21	G20044TD	Gal3S(β 1-4)Glc(β 1-		Cyclocybe cylindracea galectin; Galectin-4 Galectin-3
22	G06285TJ			Galectin-3
23	G45025VM			Galectin-3
24	G61532BD			Galectin-1; Galectin-3
25	G91237TK	Neu5Ac(α 2-3)Gal(β 1-4)Glc(β 1-		Cyclocybe cylindracea galectin; Galectin-3; Galectin-9; Galectin-8

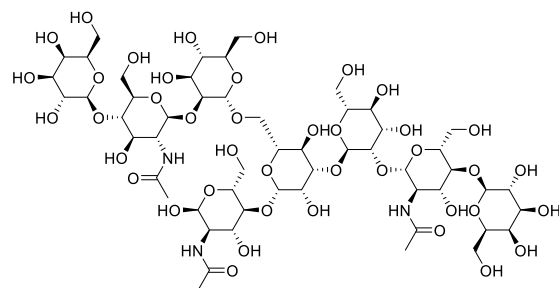
26	G92323ZH	Neu5Ac(α 2-3)Gal(β 1-4)Glc(α 1-		TI-gal
27	G00065MO	Neu5Ac(α 2-3)Gal(β 1-4)GlcNAc(β 1-		Galectin-8
28	G03499BP	Gal(α 1-3)Gal(β 1-4)GlcNAc(β 1-		Galectin-2
29	G28052FT	Gal(β 1-4)GlcNAc(β 1-2)Man(α 1-		Galectin-9
30	G60509JN	GalA(α 1-4)GalA(α 1-4)GalA(β 1-		Galectin-3
31	G81206JQ	GalNAc(α 1-3)GalNAc(β 1-3)Gal(β 1-		Galectin-9
32	G61399NC	Fuc(α 1-2)Gal(β 1-4)Glc(β 1-		Galectin-4
33	G61459JW	Fuc(α 1-2)Gal(β 1-4)GlcNAc(β 1-		Galectin-2
34	G47362BJ	GlcNAc(β 1-4)GlcNAc(β 1-4)GlcNAc(β 1-		Fungal lectin CGL3
35	G58896AZ	Gal(β 1-4)GlcNAc(β 1-3)Gal(β 1-4)Glc(β 1-		Galectin-3; Galectin-8; Porcine adenovirus 4 fiber protein

36	G14632ZD	Gal(β 1-4)GlcNAc(β 1-3)Gal(β 1-4)Glc(α 1-		Galectin-3
37	G00058MO	Gal(β 1-4)GlcNAc(β 1-3)Gal(β 1-4)GlcNAc(β 1-		Galectin-9
38	G00049MO	Gal(β 1-3)GlcNAc(β 1-3)Gal(β 1-4)Glc(β 1-		Galectin-3; TI-gal
39	G74303MH	GalNAc(α 1-3)[Fuc(α 1-2)]Gal(β 1-4)GlcNAc(α 1-		Cyclocybe cylindracea galectin
40	G74353FF	GalNAc(α 1-3)[Fuc(α 1-2)]Gal(β 1-4)GlcNAc(β 1-		Cyclocybe cylindracea galectin
41	G19524AS	GalNAc(α 1-3)[Fuc(α 1-2)]Gal(β 1-4)Glc(β 1-		Galectin-2
42	G48558GR	Neu5Ac(α 2-3)[Gal(β 1-3)GalNAc(β 1-4)]Gal(β 1-4)Glc(β 1-		Galectin-3
43	G46055MA	Gal(β 1-4)GlcNAc(β 1-3)Gal(β 1-4)GlcNAc(β 1-3)Gal(β 1-4)GlcNAc(β 1-		Porcine adenovirus 4 fiber protein

44

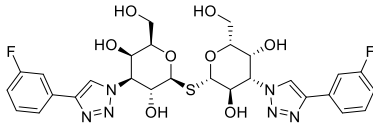
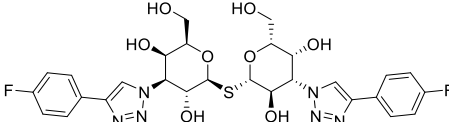
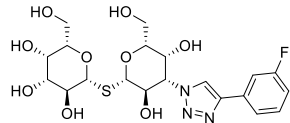
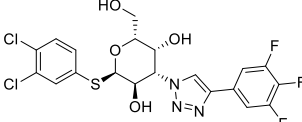
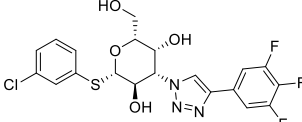
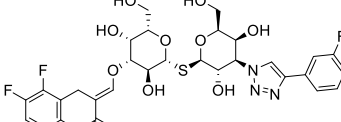
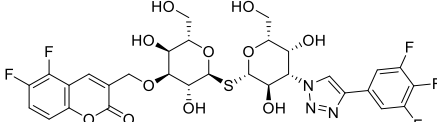
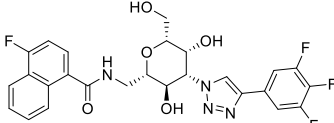
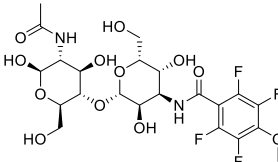
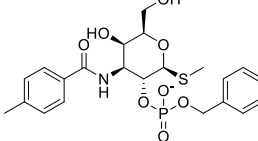
G41327HW

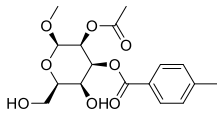
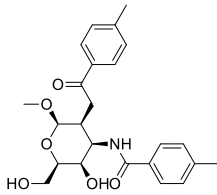
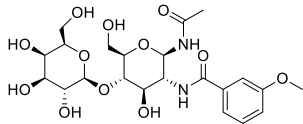
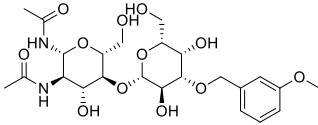
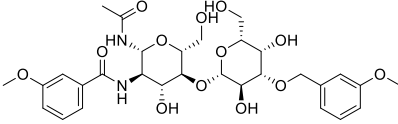
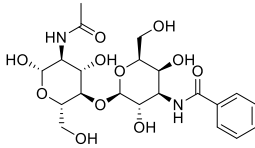
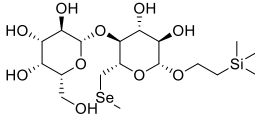
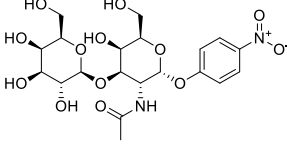
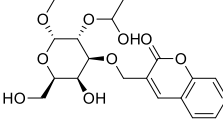
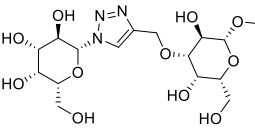
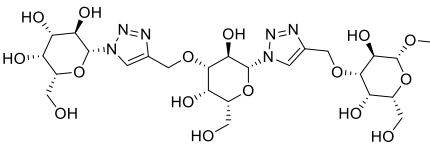
Gal(β 1-
4)GlcNAc(β 1-
2)Man(α 1-
3)[Gal(β 1-
4)GlcNAc(β 1-
2)Man(α 1-
6)]Man(β 1-
4)GlcNAc(β 1-

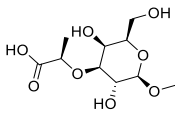
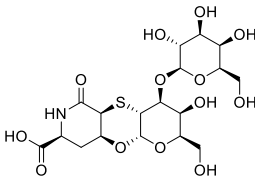
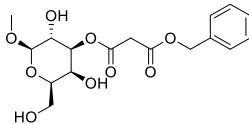
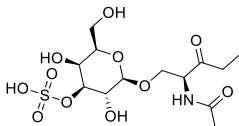


Galectin-1

Table S4. GlycanInsight-suggested carbohydrate-based ligands of human Galectin-3.

ID	PubChem	Name	Structure	Target
1	73774610	Olitigaltin; (GB0139)		Galectin-1; Galectin-3; Galectin-7
2	121493959	Compound 5		Galectin-3
3	121225588	TAZTDG		Galectin-3; Galectin-7
4	122443390	GB-1107		Galectin-3
5	134812702	Compound 2c		Galectin-3
6	133053961	Compound 36		Galectin-3
7	133107878	Compound 37		Galectin-3
8	154815575	Compound 19		Galectin-3
9	23585945	1kjr		Galectin-3
10	24776273	Compound 6		Galectin-7

11	44561021	Compound 1		Galectin-1; Galectin-3
12	54751685	Compound 2		Galectin-3
13	53392480	8VW		Galectin-3
14		Compound 3		Galectin-3
15	129009690	Compound 4		Galectin-3
16	72199541	Compound L2		Galectin-3
17	86223062	Selenolactose		Galectin-9
18	11168041	TFN		Galectin-3
19	122172824	Compound 12		Galectin-3
20	132535577	P8J		Galectin-3
21	146025984	P8G		Galectin-3

22	133107879	9QG		Galectin-8
23	138753182	EGZ		Galectin-3
24	154573933	SZS		Galectin-8
25	164575897	L4L		Galectin-8

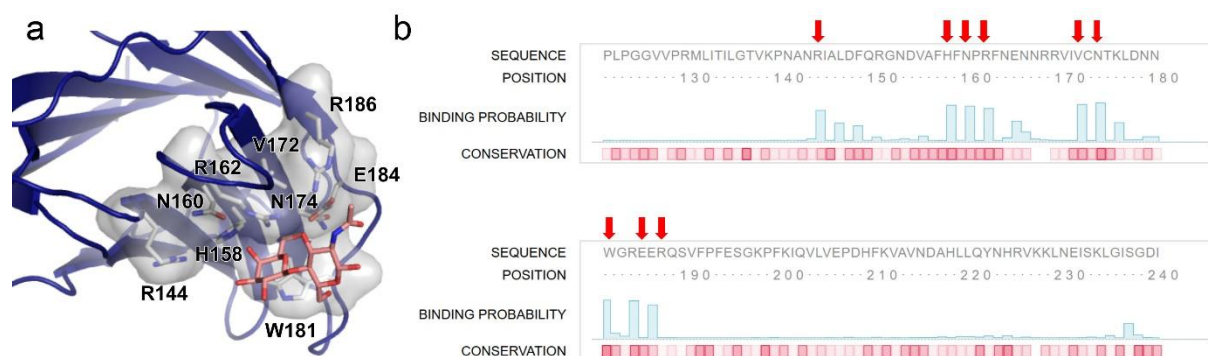


Fig. S1 GlycanInsight-predicted carbohydrate-binding pocket and experimentally identified carbohydrate-recognition domain of human Galectin-3. (a) The X-ray crystal structure of carbohydrate-bound human Galectin-3 (PDB ID: 1a3k). GlycanInsight-predicted carbohydrate-binding pocket is shown as a surface. Residues in the predicted pocket are shown as sticks. (b) Sequence annotations of the predicted pocket. A blue bar chart indicates the carbohydrate-binding probability for each residue. Conservation scores of residues are represented by transparency of red color; more conserved residues appear as deeper red. The experimentally identified carbohydrate-recognition domain of human Galectin-3 is indicated by red arrows.

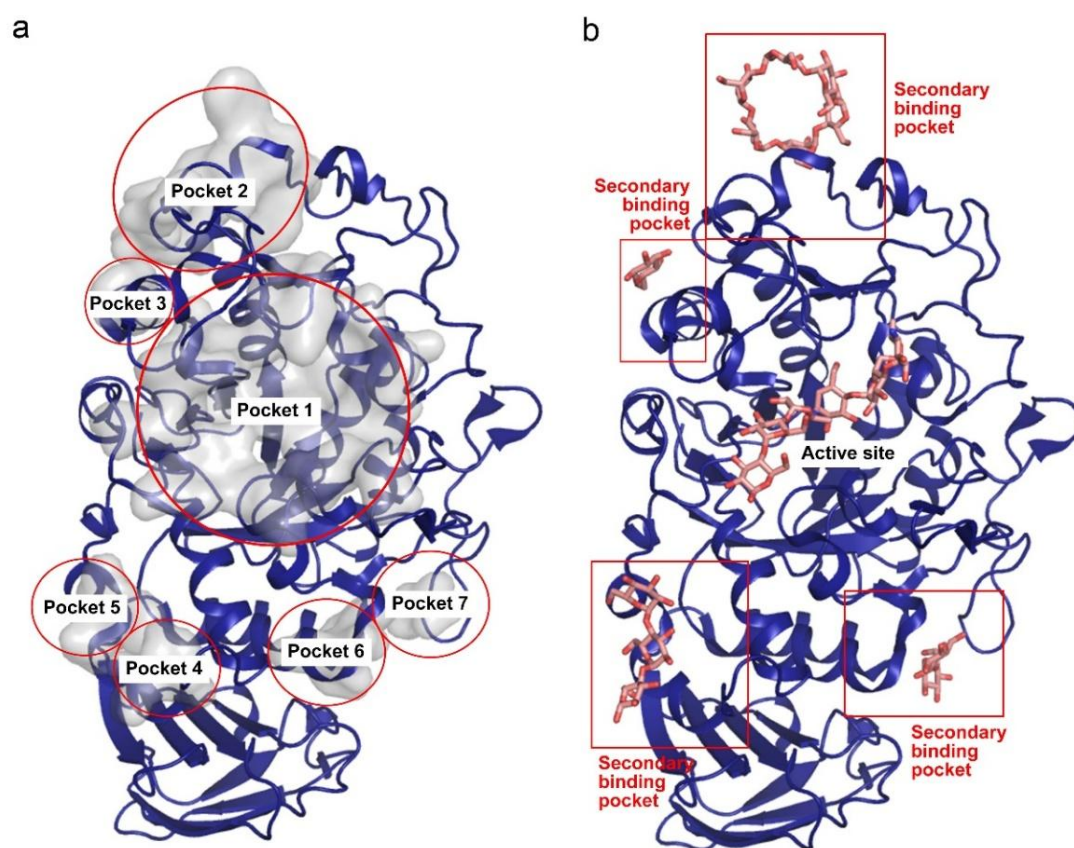


Fig. S2 GlycanInsight-predicted and experimentally identified carbohydrate-binding pockets of the pig pancreatic α-amylase. (a) Predicted carbohydrate-binding pockets of pig pancreatic α-amylase. Protein is shown in cartoon representation (UniProt ID: P00690; PDB ID: 1ua3). Predicted pockets are displayed in surface representation and marked with red circles. (b) Carbohydrates bound to the catalytic pocket (active site) and secondary pockets of pig pancreatic α-amylase. Superposition of three experimentally determined structures (PDB IDs: 3l2m, 1pig and 1ua3) indicates carbohydrate-binding pockets. Protein is depicted in cartoon representation (UniProt ID: P00690; PDB ID: 1ua3). Carbohydrates are shown as sticks.

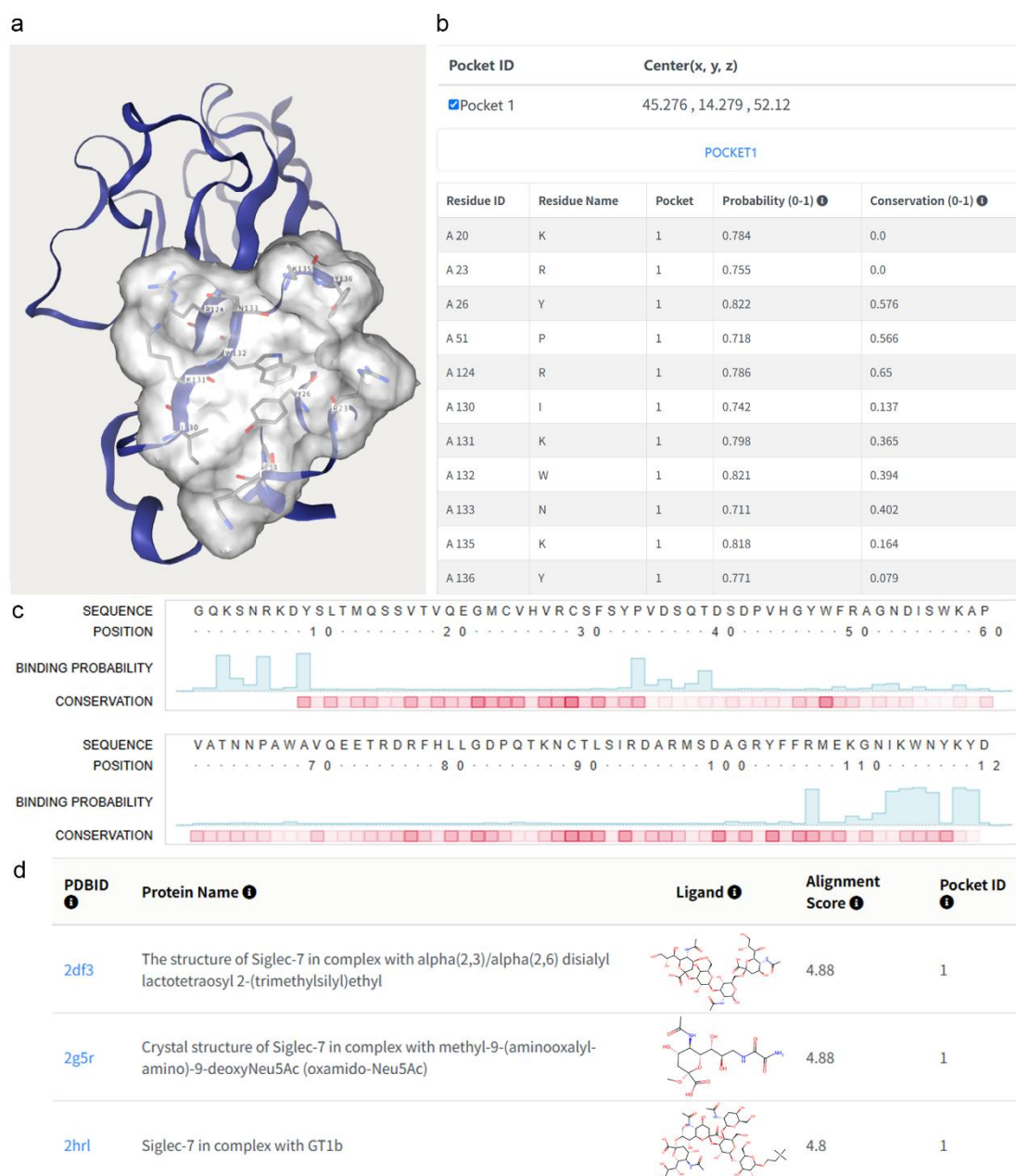


Fig. S3 Carbohydrate-binding pocket prediction and analysis of human Siglec-7 (PDB ID: 1o7v) by GlycanInsight. (a) An interactive molecular viewer of the predicted carbohydrate-binding pocket on the query protein. The predicted pocket is displayed as a surface and sticks. Users can rotate, zoom, and toggle individual residues. (b) Geometry and composition information of the predicted pocket. The center coordinates of the pocket are presented. Residues in the pocket are listed with carbohydrate-binding probabilities and conservation scores. (c) Sequence annotations of the predicted pocket. A blue bar chart indicates the carbohydrate-binding probability for each residue. Conservation scores of residues are represented by transparency of red color, with more conserved residues appearing as deeper red. (d) Putative binding ligands are suggested for the predicted pocket, which are presented with their chemical structures and scores.

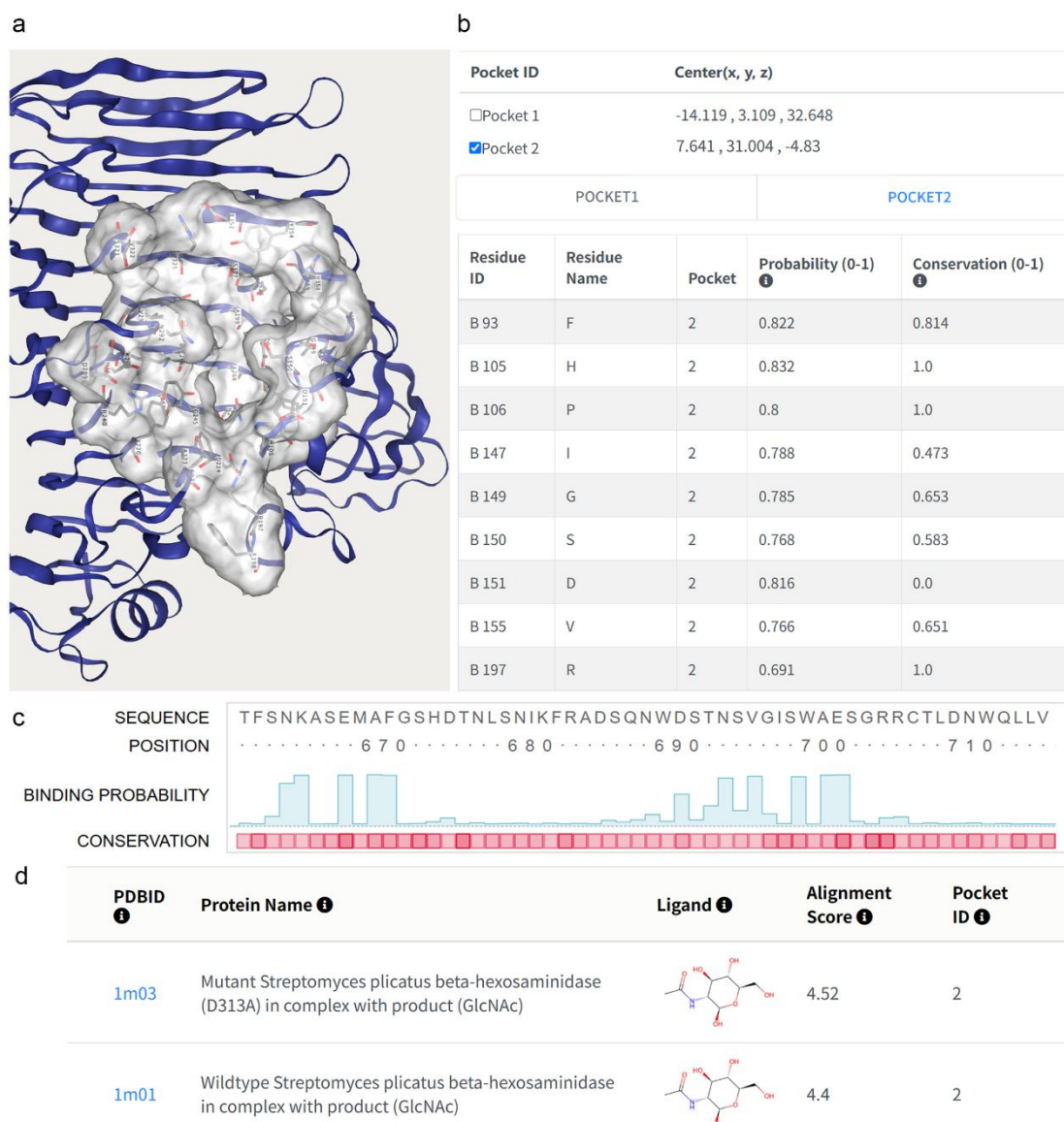


Fig. S4 Carbohydrate-binding pocket prediction and analysis of *Klebsiella* phage KP34p57 capsular depolymerase (PDB ID: 8bke) by GlycanInsight. (a) An interactive molecular viewer of the predicted carbohydrate-binding pocket on the query protein. The predicted pocket is displayed as a surface and sticks. Users can rotate, zoom, and toggle individual residues. (b) Geometry and composition information of the predicted pocket. The center coordinates of the pocket are presented. Residues in the pocket are listed with carbohydrate-binding probabilities and conservation scores. (c) Sequence annotations of the predicted pocket. A blue bar chart indicates the carbohydrate-binding probability for each residue. Conservation scores of residues are represented by transparency of red color, with more conserved residues appearing as deeper red. (d) Putative binding ligands are suggested for the predicted pocket, which are presented with their chemical structures and scores.

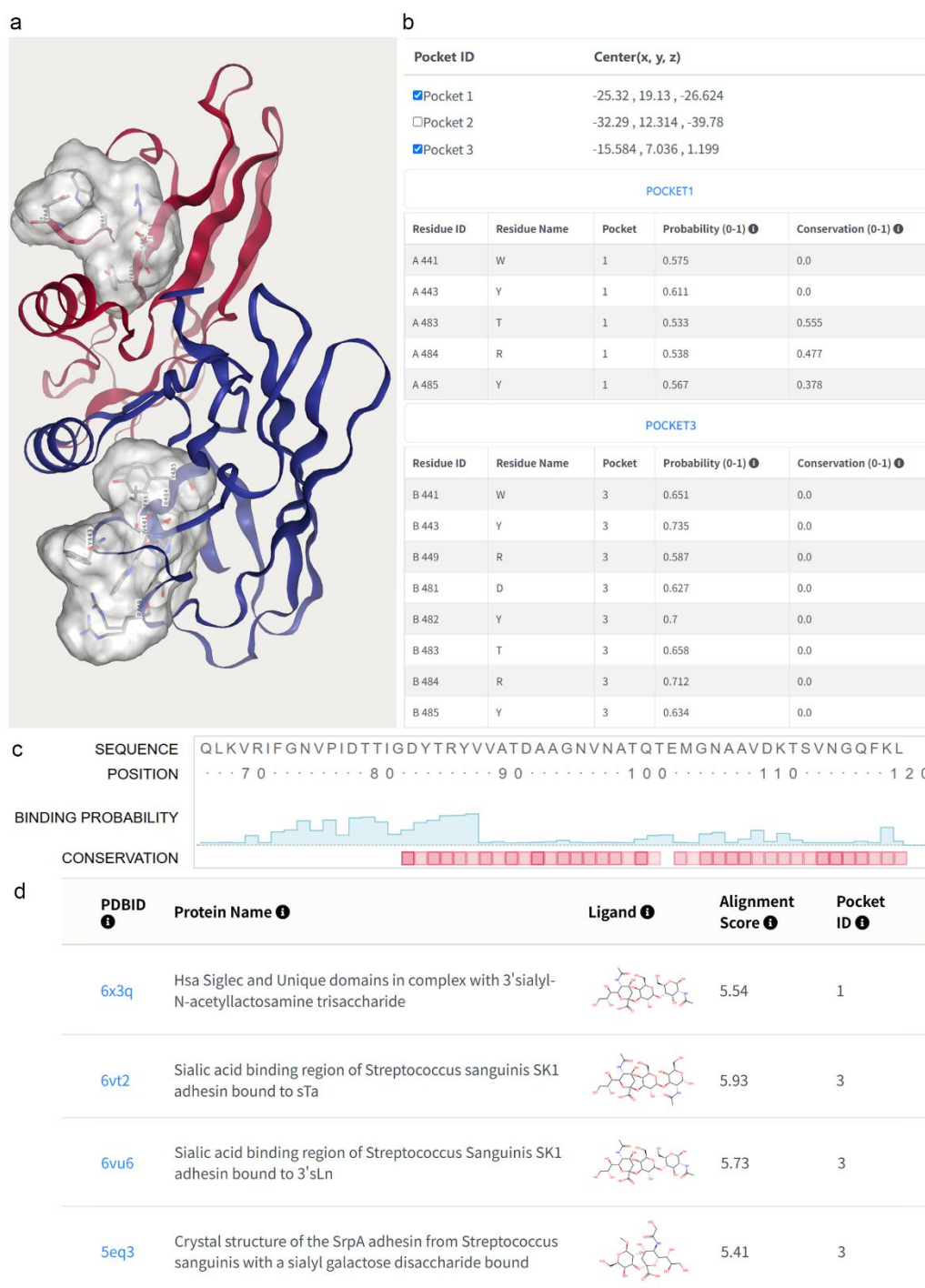


Fig. S5 Carbohydrate-binding pocket prediction and analysis of GspB (PDB ID: 5iuc) by GlycanInsight. (a) An interactive molecular viewer of the predicted carbohydrate-binding pocket on the query protein. The predicted pocket is displayed as a surface and sticks. Users can rotate, zoom, and toggle individual residues. (b) Geometry and composition information of the predicted pocket. The center coordinates of the pocket are presented. Residues in the pocket are listed with carbohydrate-binding probabilities and conservation scores. (c) Sequence annotations of the predicted pocket. A blue bar chart indicates the carbohydrate-binding probability for each residue. Conservation scores of residues are represented by transparency of red color, with more conserved residues appearing as deeper red. (d) Putative binding ligands are suggested for the predicted pocket, which are presented with their chemical structures and scores.