SUPPORTING INFORMATION

Binding Sites for Oligosaccharide Repeats from Lactic Acid Bacteria Exopolysaccharides on Bovine β-Lactoglobulin Identified by NMR Spectroscopy

Johnny Birch,[†] Sanaullah Khan,[†] Mikkel Madsen,[†] Christian Kjeldsen,[‡] Marie Sofie Møller,[†] Emil G. P. Stender,[†] Günther H. J. Peters,[∥] Jens Ø. Duus,[‡] Birthe B. Kragelund,[§*] and Birte Svensson[†*]

[†]Enzyme and Protein Chemistry, Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads, Building 224, DK-2800 Kgs. Lyngby, Denmark
[‡]NMR Spectroscopy, Department of Chemistry, Technical University of Denmark, Kemitorvet 207, DK-2800 Kgs. Lyngby, Denmark
[§]Structural Biology and NMR Laboratory, Department of Biology, University of Copenhagen, Ole Maaloes Vej 5, DK-2200 Copenhagen N, Denmark
[∥]Biophysical and Biomedicinal Chemistry, Department of Chemistry, Technical University of Denmark, Kemitorvet 207, DK-2800 Kgs. Lyngby, Denmark.

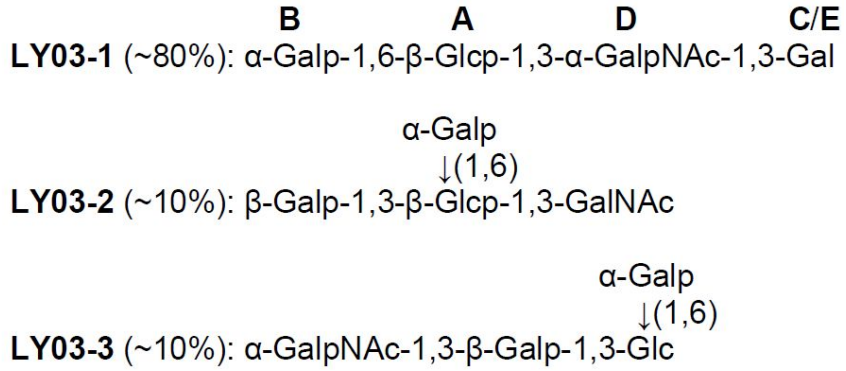
*Corresponding authors: bis@bio.dtu.dk; phone +45 4525 2740 and bbk@bio.ku.dk; phone +45 3532 2081

Content

Structure Elucidation of LY03-OS and 1187-OS

LY03-OS

In the NMR structure elucidation of the oligosaccharides, three different tetrasaccharides of the repeating unit were identified (Figure 2B, also schematized below), labelled as LY03-OS1, LY03-OS2 and LY03-OS3 and present in approximate ratios of 80%, 10% and 10%, respectively. A full assignment of the main tetrasaccharide LY03-OS1 can be found in Table S1 and all HMBC and ROESY signals from the anomeric positions are listed in Table S2.



From the ^{1}H, ^{13}C HSQC spectrum (Figure S2), it is evident that there are five distinct anomeric signals of similar intensity, as well as a few lower intensity ones. These five were labelled **A**-**E** in order of descending ^{13}C chemical shift, and the tetrasaccharide was dubbed LY03-OS1. Additionally, the sample contained quite a lot of glycerol from the purification process and 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) as internal standard.

The first of the five distinct anomeric positions **A** was determined to be from a 6-substituted β-glucose. The second anomeric position **B** was determined to be from an unsubstituted α-galactose, while the third anomeric position **C** was determined to be from a 3-substituted reducing end β-galactose. The fourth anomeric position **D** was determined to be from a 3-substituted *N*-acetylated α-galactosamine (GalNAc), and finally, the fifth of the anomeric signals, **E**, was from a 3-substituted reducing α-galactose. The substitution pattern of each monosaccharide unit was determined both by comparing to chemical shifts of the corresponding methyl glycoside in the case for non-reducing units and the corresponding glycoside for reducing ends, as well as using HMBC and ROESY data. The full chemical shift assignment of LY03-OS1 can be seen in Table S1.

The major carbohydrate component of the mixture is a single repeating unit cleaved between the β-galactose, which corresponds to the reducing ends **C** and **E**, and the disubstituted β-glucose, which corresponds to **A**, schematized in Figure 2. The order and substitution of the carbohydrates were confirmed using HMBC and ROESY (Figure S5 and S6), and a summary of the anomeric HMBC and ROESY correlations can be seen in Table

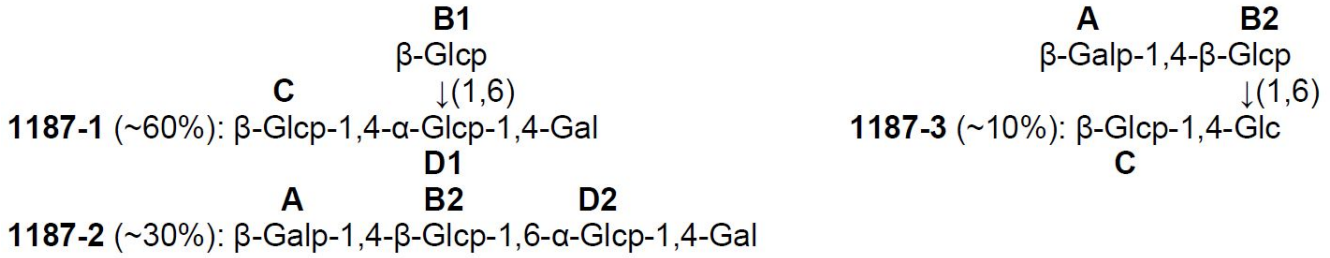
S2. The α-galactose **B** was confirmed to have a 1,6-linkage to the β-glucose **A**, which had a 1,3-linkage to the α-GalNAc **D**, which in turn had a 1,3-linkage to the reducing ends **C** and **E**.

However, as evident from the HSQC (Figure S2), there are some signals of lesser intensity originating from other carbohydrates than LY03-OS1. In total it was determined that two other tetrasaccharides were present, LY03-OS2 and LY03-OS3 (Figure 2B), and they were both in similar concentration of around 10% of the sample's total carbohydrate content. LY03-OS2 was hydrolyzed between the α-GalNAc and the β-galactose of the repeating unit while LY03-OS3 was hydrolyzed between the β-glucose and the α-GalNAc of the repeating unit, both are schematized in Figure 2B.

LY03-OS2 and LY03-OS3 have not been assigned as they are in too low concentration, but the structures were determined from the anomeric chemical shifts that differed from the main fragment. The two additional anomeric signals close the anomeric position $\mathbf{A}_1$ would correspond to the β-Gal anomeric position of the terminal and 3-substituted β-galactoside present in LY03-OS2 and LY03-OS3, respectively, and the additional signal in the α-anomeric region, close to the α-GalNAc $\mathbf{D}_1$ would correspond to the terminal α-GalNAc present in LY03-3. The reducing ends for LY03-OS2 and LY03-OS3 have very low signal intensity and are hard to distinguish from the noise.

1187-OS

The NMR structure analysis showed three different tetrasaccharides corresponding to different hydrolysis patterns of the repeating unit. This meant that it was rather inhomogeneous, even though many of the signals were quite similar. In the anomeric region of the HSQC spectrum six large signals were identified, alongside a few smaller ones. Two of the large signals and two of the smaller signals corresponded to two separate reducing ends, and they were found to be in a ratio of approximately 9:1. These different anomeric signals all varied in intensity, only supporting the idea of inhomogeneity. Furthermore, the sample contained glycerol, present as a result of the purification process, and DSS as internal standard.



Amongst the other signals, three corresponded to β-glycosides and one corresponded to an α-glycoside. The four non-reducing saccharide units were given the labels **A**-**D**. The first anomeric signal at 4.443 ppm and 103.80 ppm in ^{1}H and ^{13}C chemical shift, respectively, was designated as **A** and was found to be from a β-galactopyranoside. As all of its chemical shifts, see Table S4, was found to be similar to that of methylated β-galactopyranoside[1], it was concluded to be non-substituted.

The second signal in the anomeric region was found to be a mixture of two different β-glucopyranosides. One of them, designated as **B**1, had the chemical shifts of 4.504 ppm and 103.15 ppm for ^{1}H and ^{13}C, respectively, while the other one, **B2**, had chemical shifts of 4.518 and 103.025 for ^{1}H and ^{13}C, respectively. The only difference between these two glucosides were found to be a 4-substition on **B**2 while **B**1 was unsubstituted. The third anomeric signal at 4.577 ppm and 103.08 ppm ^{1}H and ^{13}C chemical shift, respectively, was also found to be an unsubstituted β-glucopyranoside and was called **C**. The last non-reducing carbohydrate anomeric signal was also found to originate from a pair of glucopyranosides, but of α-configuration, and the difference between these two was determined to be a 4,6-substitution on one of them and a 6-substitution on the other and they were called **D**1 and **D**2, respectively. The anomeric positions of these two overlapped almost completely with an anomeric signal at 4.911 ppm for ^{1}H and 100.58 ppm for ^{13}C. Finally, there are the different reducing end anomeric signals, and there were two large ones and two quite small ones. The large ones corresponded to a 4-substituted galactose while the small ones were determined to be from a glucose based on its 1- and 2-position chemical shifts. Due to its small presence in the sample it was not possible to fully assign the glucose reducing end.

The next step, after identifying all the monosaccharide units, was to determine how they were attached to one another. This was done using HMBC and ROESY experiments (shown in Figure S7 and S8), and the results are summarized in Table S5 and discussed here: The anomeric position on A, the β-galactopyranoside, had correlations through HMBC and ROESY to the 4-position of the β-glucopyranoside **B**2. The anomeric signals of the two units **B**1 and **B**2 had very similar correlations to the 6-positions of **D**1 and **D**2, respectively, whereas the anomeric signals of **D**1 and **D**2 had identical correlations to the 4-position of the reducing galactose. The last major signal to account for is the β-glucopyranoside **C** which was found to be attached to the 4-position of the α-glucose **D**1. This fits quite nicely with the repeating unit structure and suggests that the bond hydrolyzed is the β-galactose-1,4- bond in the backbone. However, when looking at the different intensities of the signals it is evident that the sample mixture is far from a single homogeneous saccharide. As the ^{1}H NMR overlap was quite extensive HSQC data was used to approximate the ratios between the different units. The most intense anomeric signal is the one corresponding to **B**1 and **B**2 and it was referenced as 1, whereas the other peaks integrals relative to **B** were as shown in Table S3. Based on the data in Table S3 it was possible to obtain an approximate ratio between the following three tetrasachharides labeled as 1181-OS1, 1187-OS2 and 1187-OS3 (Figure 2D):

All of the chemical shifts of the different monosaccharide units can be found in Table S4 and when compared to the relevant values in Faber et al. 2001[2] they fit quite nicely. As this was a mixture of three different species in different concentrations some of the signals overlap quite extensively and all chemical shifts are based on HSQC. The chemical shifts for the two glucosides **B**1 and **C** barely differ, as the only difference is whether they are attached to a 4- or a 6-position of the α-glucoside.

Table S1. Chemical shift table for LY03-OS1.

Unit		Position 1	Position 2	Position 3	Position 4	Position 5	Position 6
α-Gal **B**	^{1}H	4.981	3.833	3.826	3.976	3.91	3.73
	^{13}C	97.31	68.69	67.48	68.28	70.17	60.16
β-Glc **A**	^{1}H	4.564	3.292	3.466	3.48	3.618	3.932/3.749
	^{13}C	103.50	71.94	74.84	68.65	73.39	64.95
α-GalNAc **D**	^{1}H	5.08	4.403	4.168	4.202	4.197	3.73
	^{13}C	92.97	47.38	76.41	67.76	69.81	60.16
reducing α-Gal **E**	^{1}H	5.292	3.933	3.941	4.172	4.043	3.73
	^{13}C	91.36	65.83	72.47	64.34	69.45	60.16
reducing β-Gal **C**	^{1}H	4.632	3.594	3.746	4.118	3.654	3.73
	^{13}C	95.50	69.49	75.8	63.74	74.23	60.16

Table S2. HMBC and ROESY correlations for the anomeric positions in LY03-OS1.

Anomeric position	HMBC correlations	ROESY correlations
α-Gal - $\mathbf{B}_1$	$\mathbf{B}_2$, $\mathbf{B}_5$, $\mathbf{A}_6$	$\mathbf{A}_6$, $\mathbf{B}_2$
β-Glc - $\mathbf{A}_1$	$\mathbf{A}_2$, $\mathbf{D}_3$	$\mathbf{A}_3$, $\mathbf{A}_5$, $\mathbf{D}_3$
α-GalNAc - $\mathbf{D}_1$	$\mathbf{C}_3$, $\mathbf{D}_3$, $\mathbf{D}_5$	$\mathbf{C}_3$, $\mathbf{C}_4$, $\mathbf{E}_3$, $\mathbf{E}_4$/$\mathbf{D}_3$[a]
reducing α-Gal - $\mathbf{E}_1$	$\mathbf{E}_5$	$\mathbf{E}_2$
reducing β-Gal - $\mathbf{C}_1$	$\mathbf{C}_2$	$\mathbf{C}_3$, $\mathbf{C}_5$

a: overlapping signals

Table S3. Relative integrals measured in HSQC for 1187-OS.

Unit	**A**	**B**	**C**	**D**	Reducing Gal	Reducing Glc
Relative integral	0.43	1	0.76	0.91	0.89	0.14

Table S4. Chemical shifts of the units of the three 1187-OS DP4.

Unit		Position 1	Position 2	Position 3	Position 4	Position 5	Position 6
β-Gal **A**	^{1}H	4.443	3.541	3.658	3.922	3.720	3.774/3.774
	^{13}C	103.80	71.66	73.22	69.2	76.04	61.66
β-Glc **B**1	^{1}H	4.504	3.314	3.499	3.400	3.461	3.731/3.906
	^{13}C	103.15	73.70	76.28	70.16	76.60	61.31
β-Glc **B**2	^{1}H	4.518	3.362	3.594	3.661	3.655	3.814/3.979
	^{13}C	103.03	73.38	75.31	79.01	74.83	60.67
β-Glc **C**	^{1}H	4.577	3.318	3.499	3.400	3.461	3.731/3.906
	^{13}C	103.09	73.81	76.28	70.16	76.60	61.31
α-Glc **D**1	^{1}H	4.911	3.582	3.849	3.785	4.393	3.982/4.157
	^{13}C	100.58	72.26	71.77	78.69	70.00	67.75
α-Glc **D**2	^{1}H	4.905	3.459	3.726	3.543	4.291	3.815/4.046
	^{13}C	100.58	70.95	72.92	69.84	71.61	68.18
Reducing β-Gal	^{1}H	4.634	3.531	3.729	4.001	3.757	3.758/3.892
	^{13}C	97.36	72.26	72.89	78.05	75.8	60.42
Reducing α-Gal	^{1}H	5.293	3.849	3.922	4.059	4.141	3.814/3.814
	^{13}C	93.01	71.77	69.2	79.34	71.45	60.67
Reducing β-Glc	^{1}H	4.666	3.288	3.618	nd	nd	nd
	^{13}C	96.65	nd	nd	nd	nd	nd
Reducing α-Glc	^{1}H	5.857	3.572	3.853	3.592	nd	nd
	^{13}C	92.06	Nd	nd	nd	nd	nd

Table S5. HMBC and ROESY correlations from the anomeric protons in 1187-OS.

Anomeric position	HMBC correlations	ROESY correlations
β-Gal **A**$_1$	**B**2$_4$	**B**2$_4$, **A**$_3$, **A**$_5$
β-Glc **B**1$_1$	**D**1$_6$	**D**1$_{6a}$, **D**1$_{6b}$, **B**1$_3$, **B**1$_5$
β-Glc **B**2$_1$	**D**2$_6$	**D**2$_{6a}$, **B**2$_5$, **B**2$_3$
β-Glc **C**$_1$	**D**1$_4$	**D**1$_4$, **C**$_3$, **C**$_5$
α-Glc **D**1$_1$	**D**1$_5$, reducing β-Gal-4	Reducing β-Gal-4
α-Glc **D**2$_1$	Reducing α-Gal-4	Reducing α-Gal-4
Reducing β-Gal	Nd	Reducing β-Gal-5, reducing β-Gal-3
Reducing α-Gal	Nd	Nd
Reducing β-Glc	Nd	Nd
Reducing α-Glc	Nd	Nd

Table S6. Site 1 K_D values (mM) calculated from fits (Figure 6)

	LY03-OS		**1187-OS**	
Residue	K_D	S.E	K_D	S.E
K70	2.8	1.2	0.5	0.1
I71	42.1	21.4	11.8	2.9
I72	6.4	5.9	17.0	6.4
L87			7.9	4.7

Table S7. Site 1 average K_D by global fitting of selected residues with CSPs above AVG+1STD.

Ligand	Residues	K_D (mM)		
LY03-OS	K70, I71, I72	5.4	±	1.8
1187-OS	K70, I71, I72, L87	3.5	±	1.2

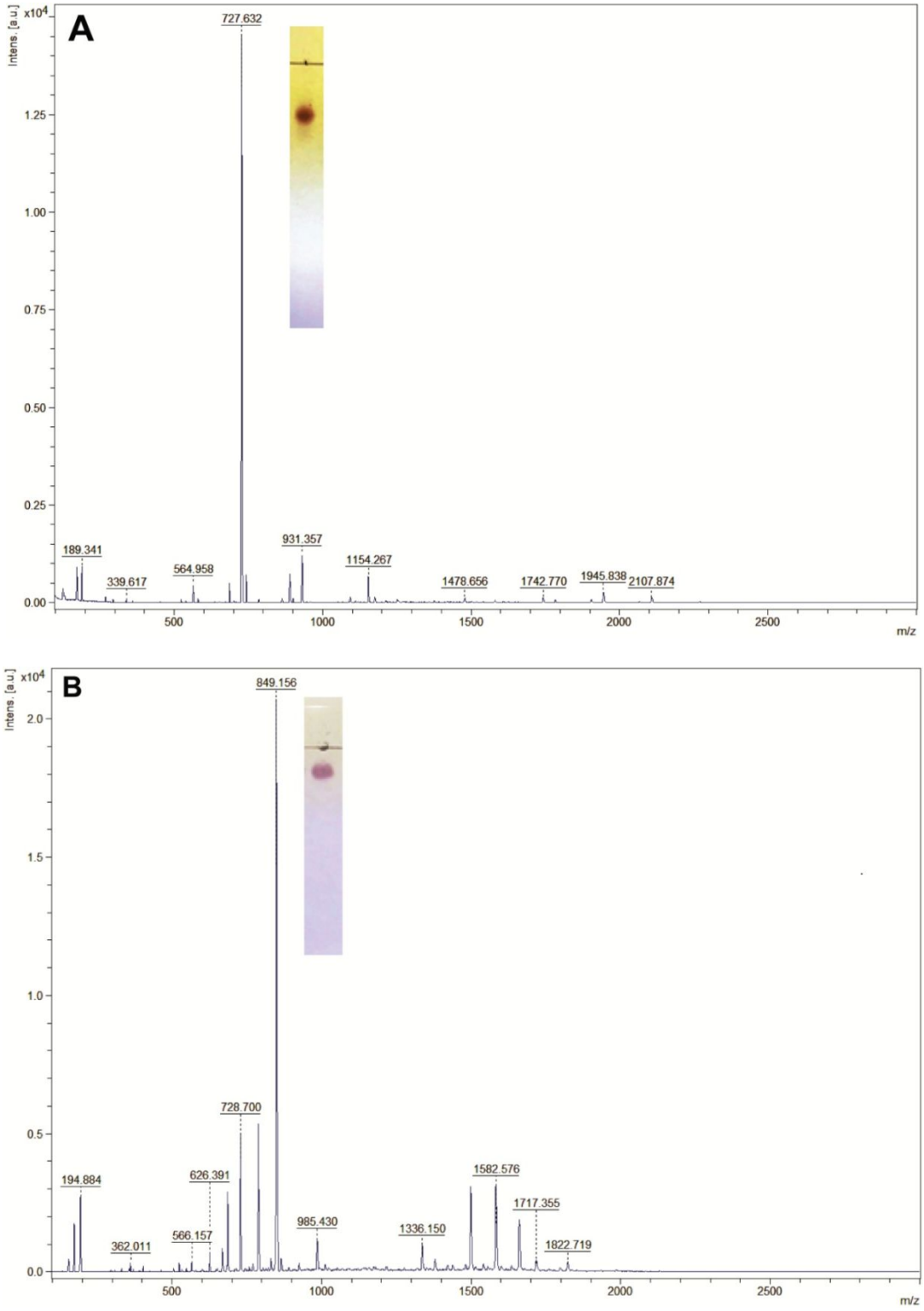


Figure S1. MALDI-TOF and TLC analysis of LY03-OS (A) and 1187-OS (B) oligosaccharides. The insets show TLC analysis of HPLC purified oligosaccharides.

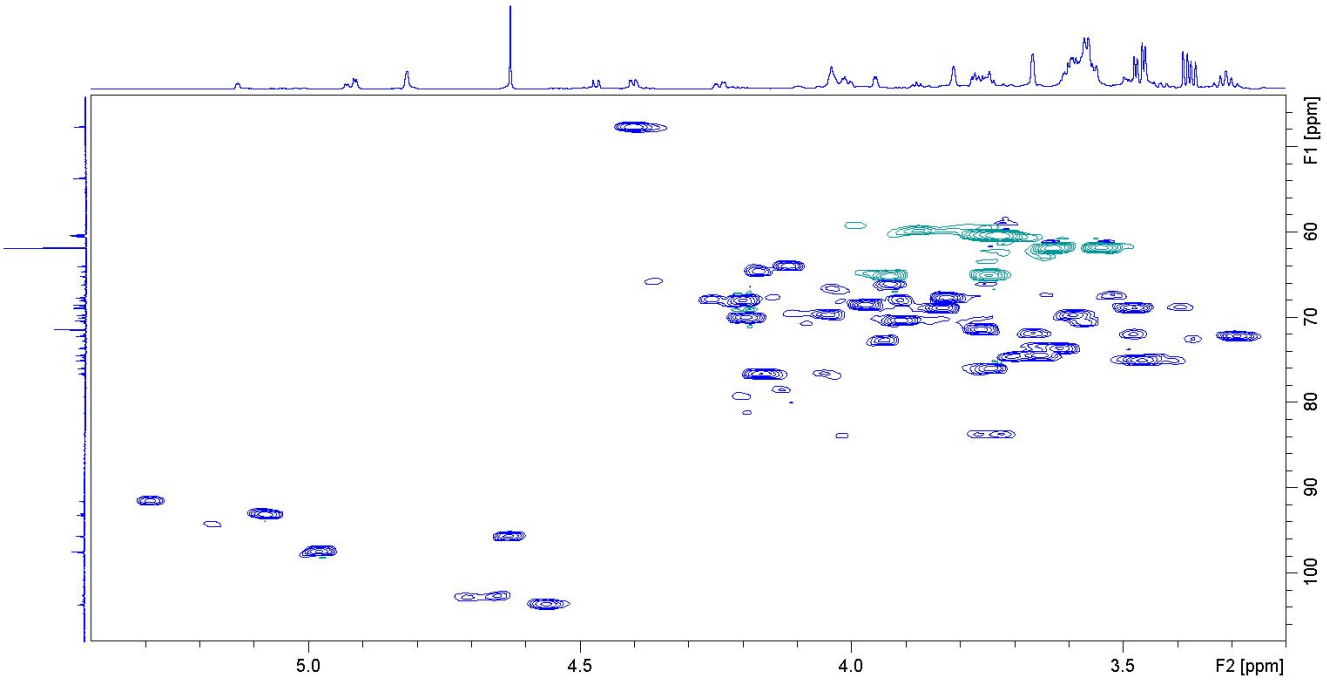


Figure S2. HSQC of the LY03-OS

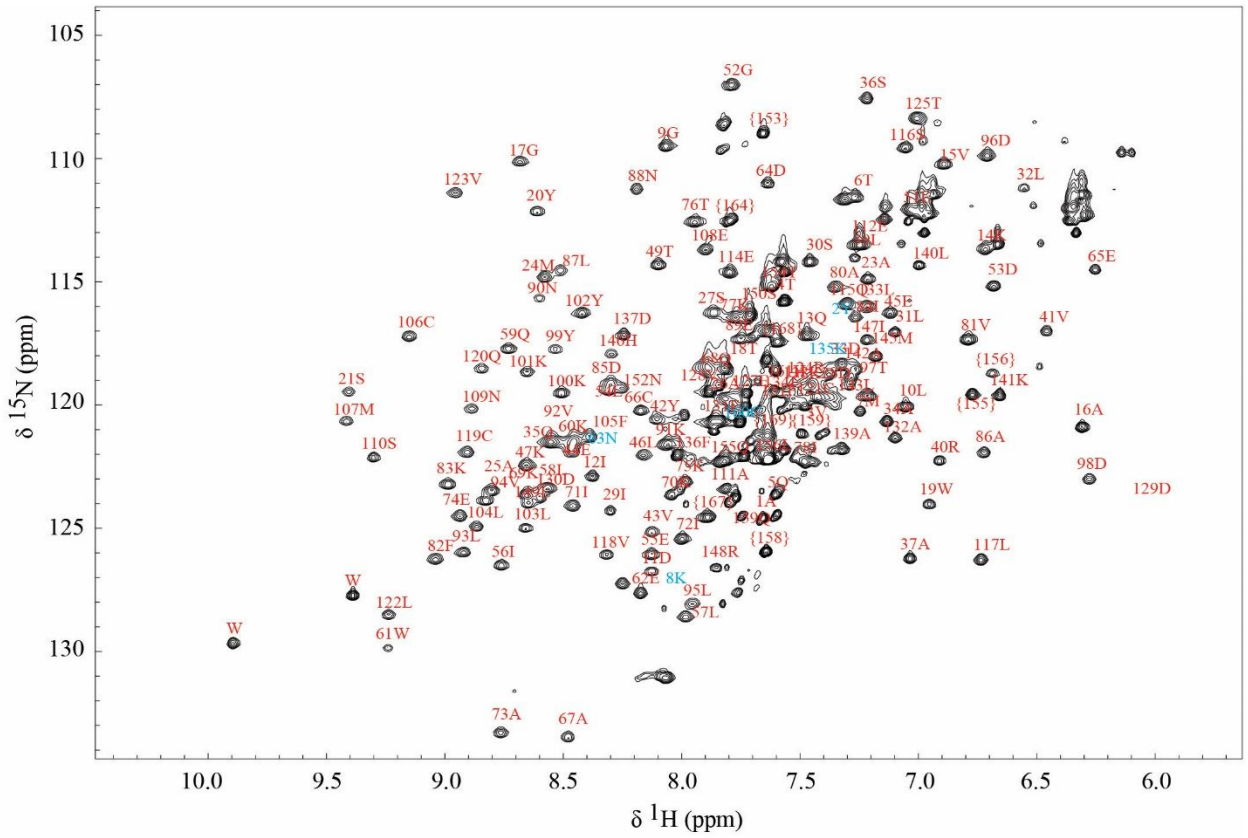


Figure S3. ^{1}N,^{15}N-HSQC spectrum 350 µM BLGA at 37°C in 50 mM potassium phosphate pH 2.65 with assignments. Residues (Y2, K8, N63, K135, and C160) assigned in the literature[s1] cannot be assigned here due to peak overlap are labelled in blue. Unassignable spin systems are numbered in curly brackets.

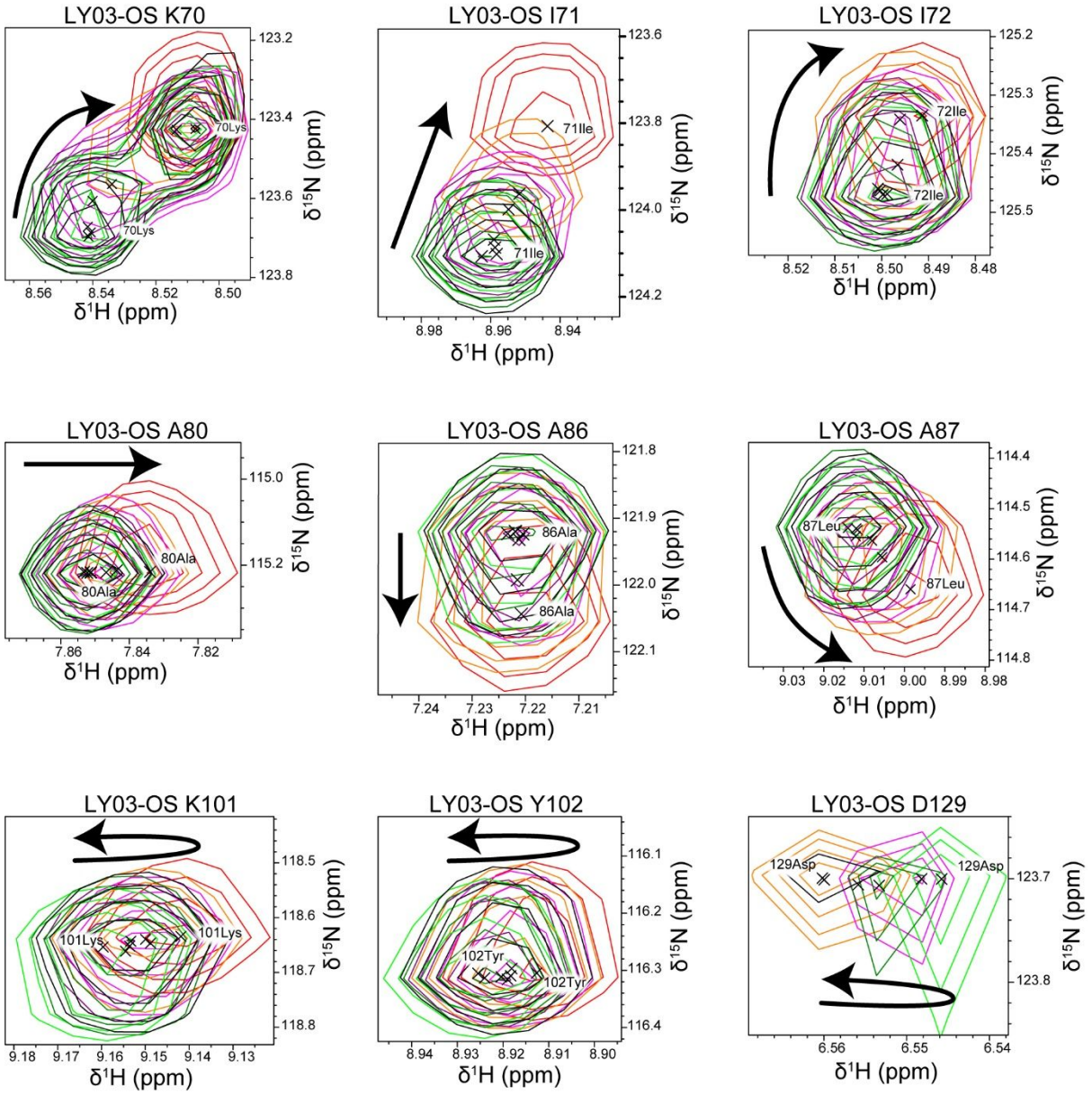


Figure S4. Changes in peak position upon titration with LY03-OS (0.0–6.3 mM), chemical shift changes of K70, I71, I72, A80, A86 and L87 corresponding to Site 1. Chemical shift changes of K101, T102 and D129 corresponding to Site 2. Arrows indicate the direction of the changes with Site 1 showing single directional movement, and Site 2 showing double directional movement.

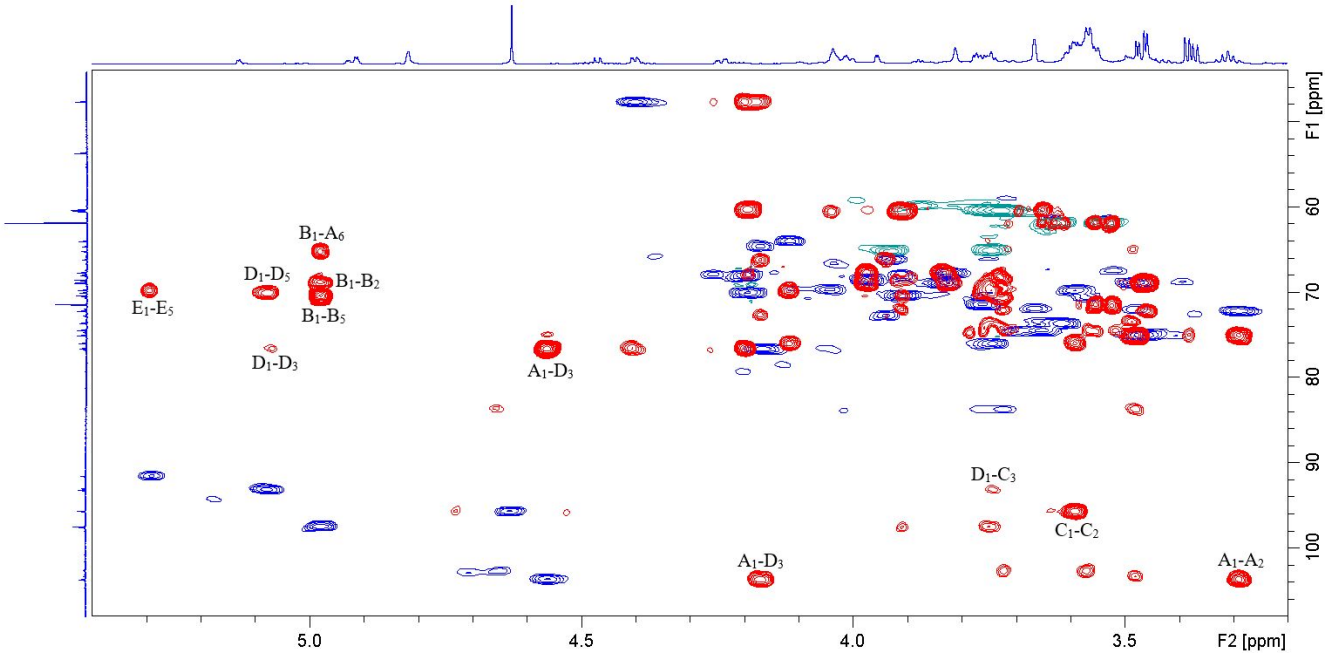


Figure S5. HSQC spectrum (blue/green) with HMBC spectrum (red) overlayed of LY03-OS with the correlations summarized in Table S2 highlighted.

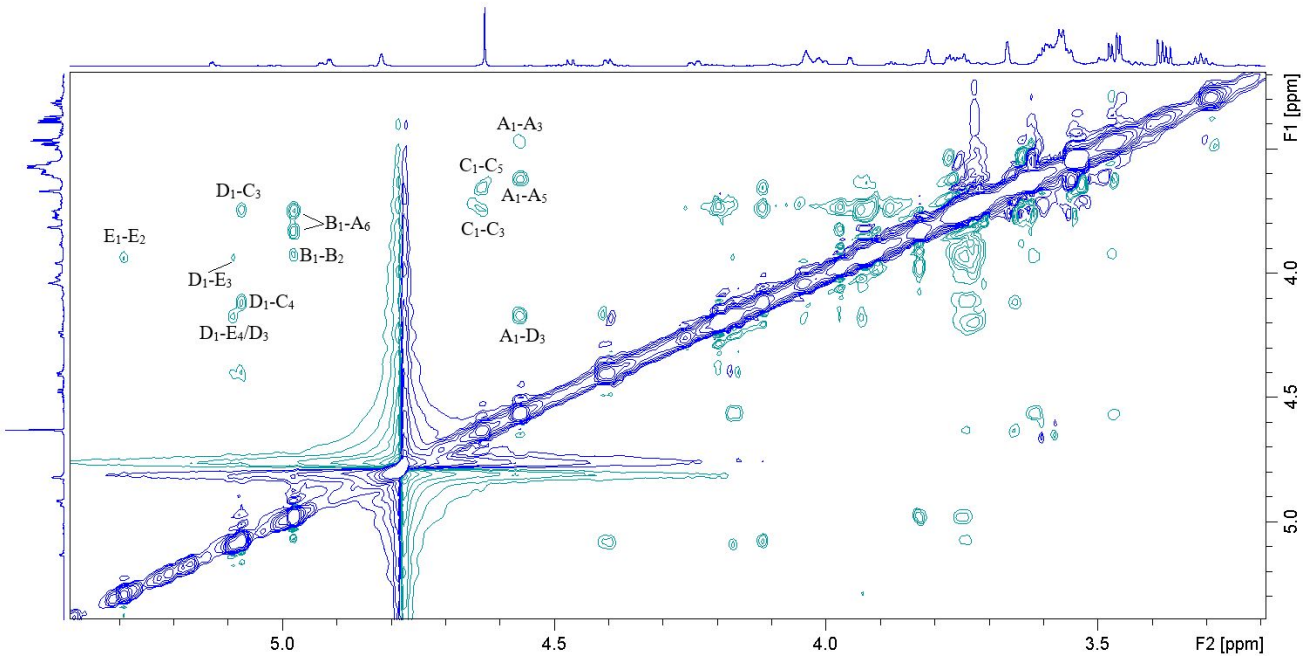


Figure S6. ROESY spectrum of LY03-OS with the correlations in Table S2 highlighted.

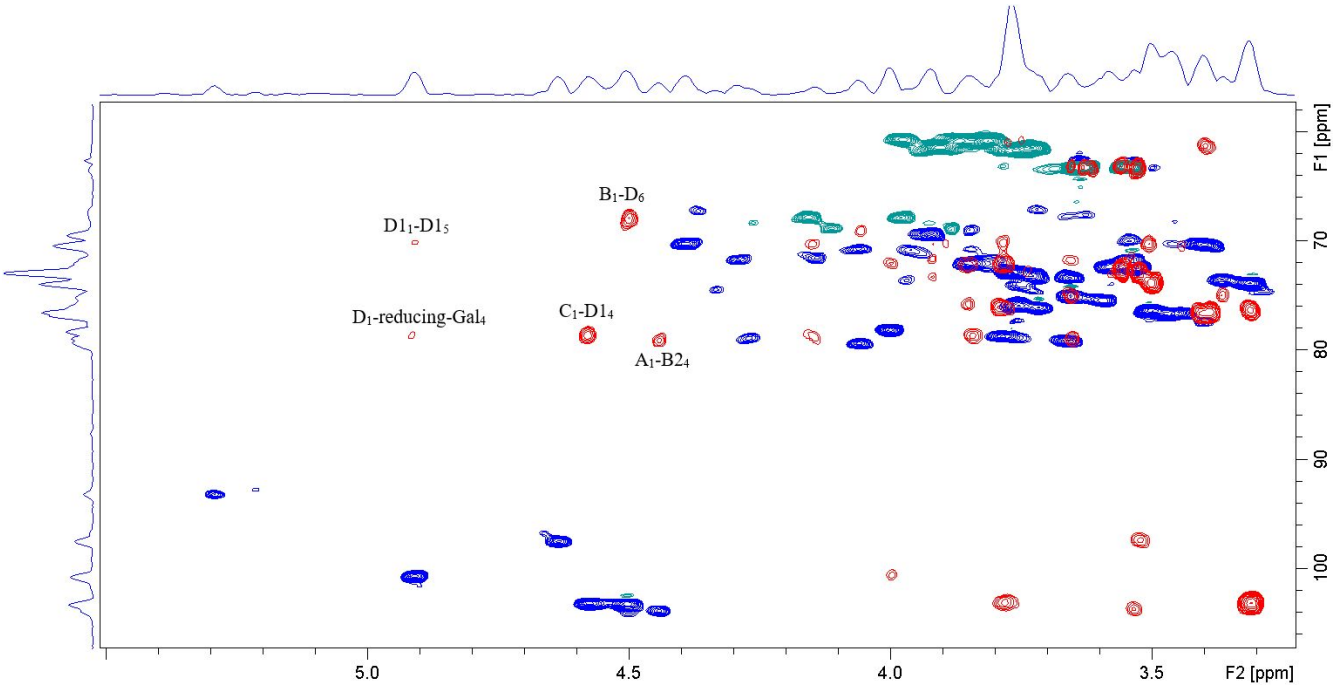


Figure S7. HSQC spectrum (blue/green) with HMBC spectrum (red) overlayed of 1187-OS with the correlations summarized in Table S5 highlighted.

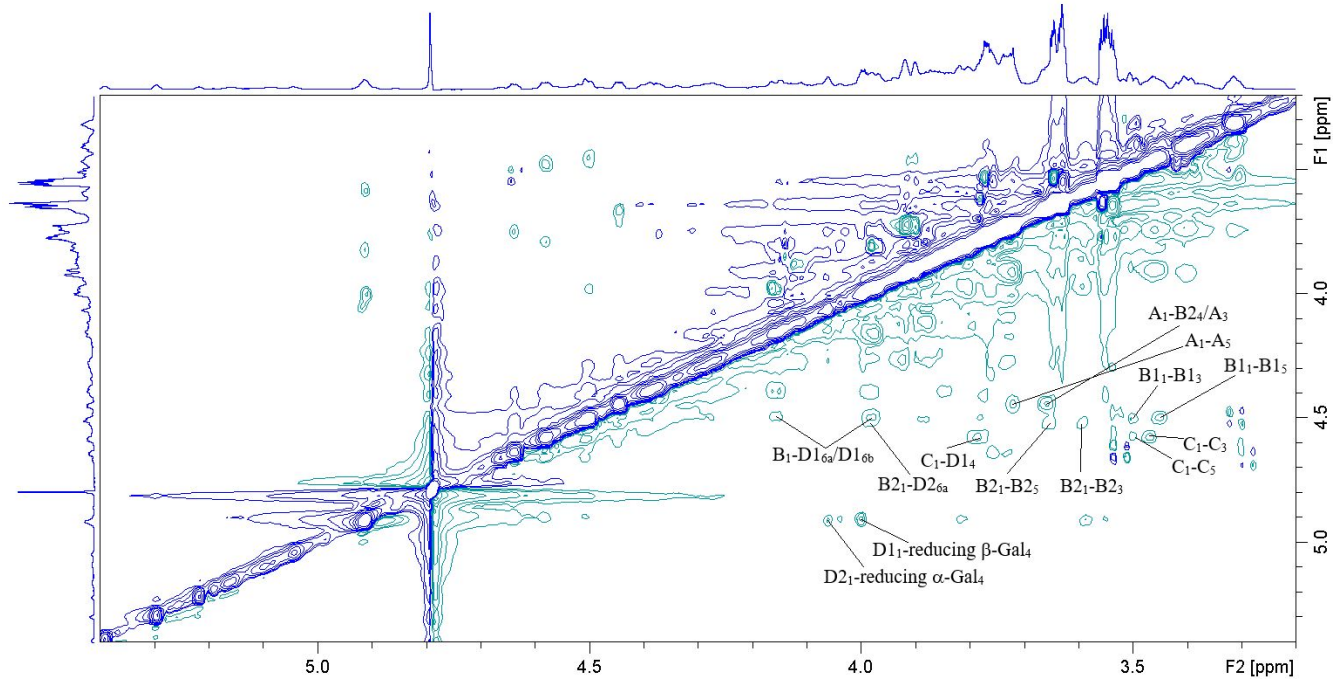


Figure S8. ROESY spectrum of 1187-OS with the correlations in Table S5 highlighted.

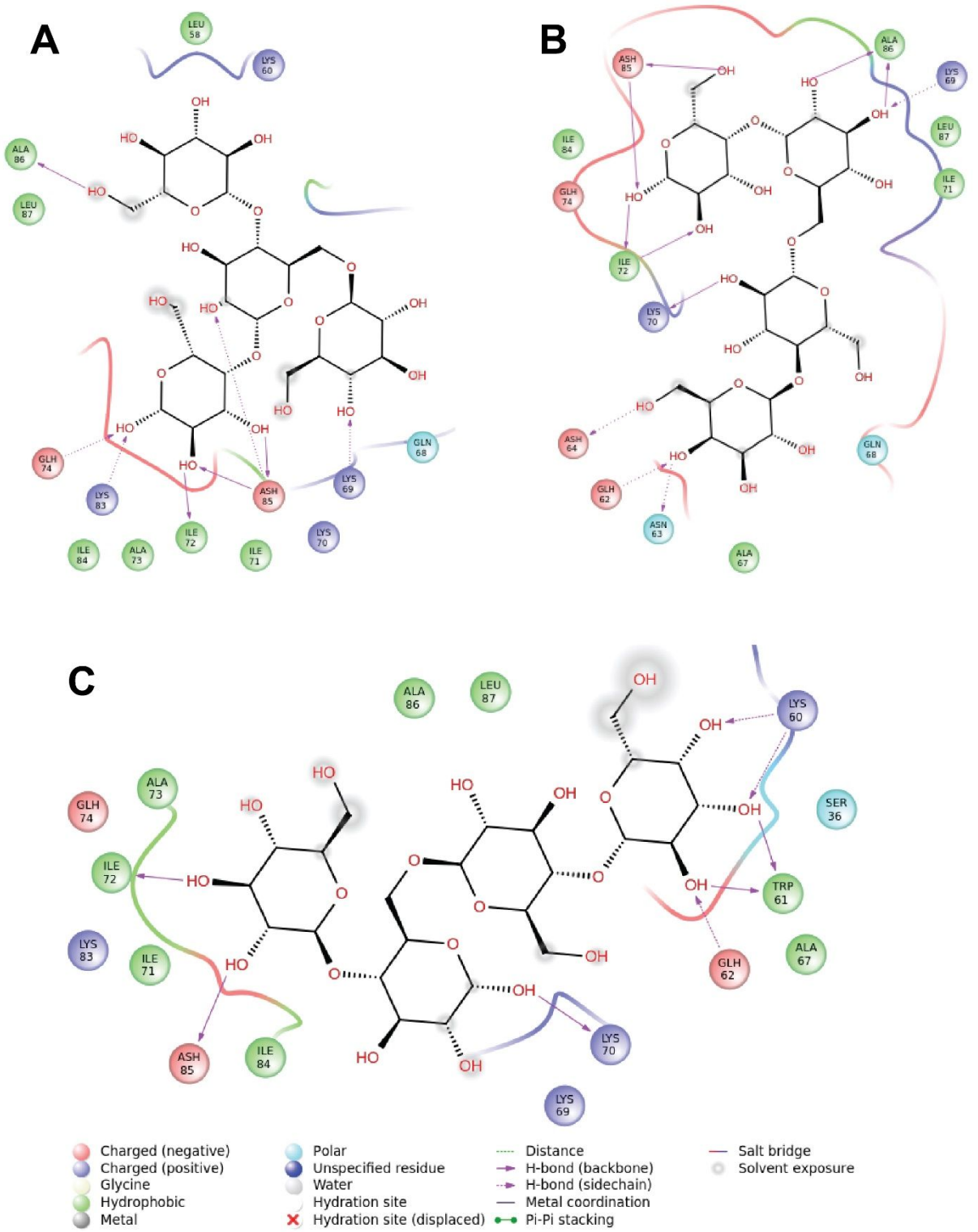


Figure S9. Ligand interaction diagram of A) 1187-OS1 B) 1187-OS2 and C) 1187-OS3 with Site 1 (K70, I71, I72)

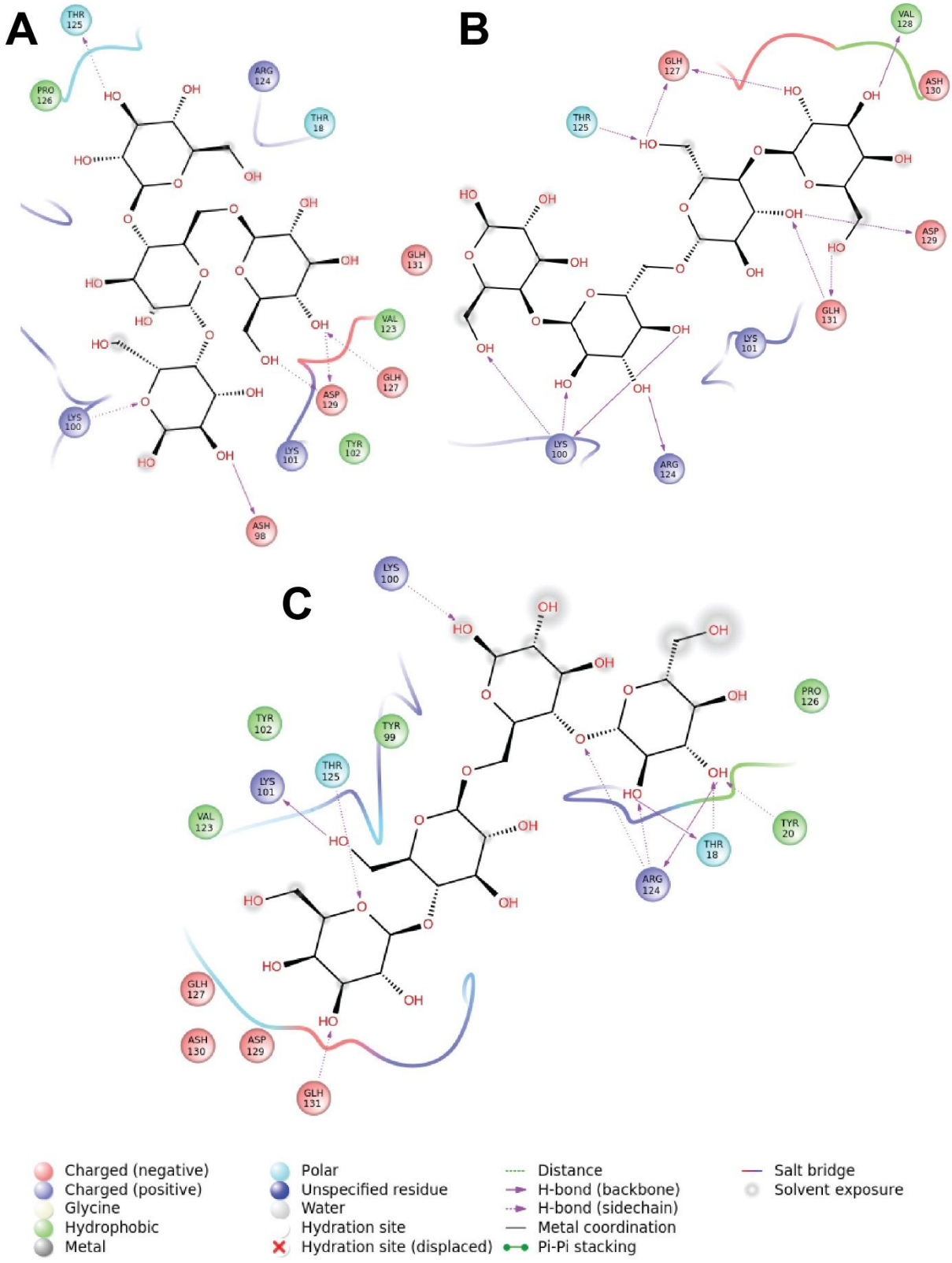


Figure S10. Ligand interaction diagram of A) 1187-OS1 B) 1187-OS2 and C) 1187-OS3 with Site 2 (K100, K101, Y102, D129)

References

S1 Stender, E. G. P.; Birch, J.; Kjeldsen, C.; Nielsen, L. D.; Duus, J. Ø.; Kragelund, B. B.; Svensson, B. Alginate Trisaccharide Binding Sites on the Surface of β-Lactoglobulin Identified by NMR Spectroscopy: Implications for Molecular Network Formation. *ACS Omega* **2019**, *4* (4), 6165–6174. https://doi.org/10.1021/acsomega.8b03532.