Substrate recognition by a carbohydrate-binding module in the prototypical ABC transporter for lipopolysaccharide O antigen from *Escherichia coli* O9a

**Evan Mann, Steven D. Kelly, M. Sameer Al-Abdul-Wahid, Bradley R. Clarke, Olga G. Ovchinnikova, Bin Liu and Chris Whitfield**

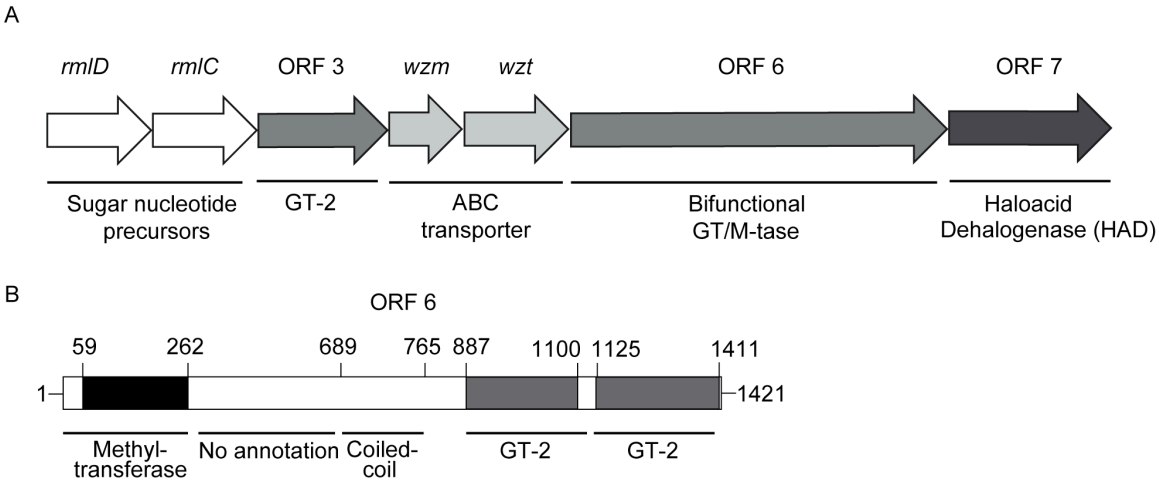


**Figure S1. Organization of the O7 *wb** locus. (***A***)** Genes within the O7 O-PS biosynthesis cluster are labeled with known gene names (where convincing characterized homologs are known), or by ORF designations characterization (above), and biosynthesis functions (below). Putative functions were assigned used the NCBI Blastp server (1). (*B*) The predicted domain organization of ORF 6 identifies two C-terminal GT-2 domains separated by a coiled coil from a predicted methyltransferase domain and an unannotated spanning ~47 kDa in the gene product. Domain boundaries are approximate and were predicted by PHYRE2 using PDB IDs: 5T39 and 2Z86 as templates for the methyltransferase and GT domains respectively, as well as with the CDD (2–4). Coiled-coil prediction was achieved using COILS (5).

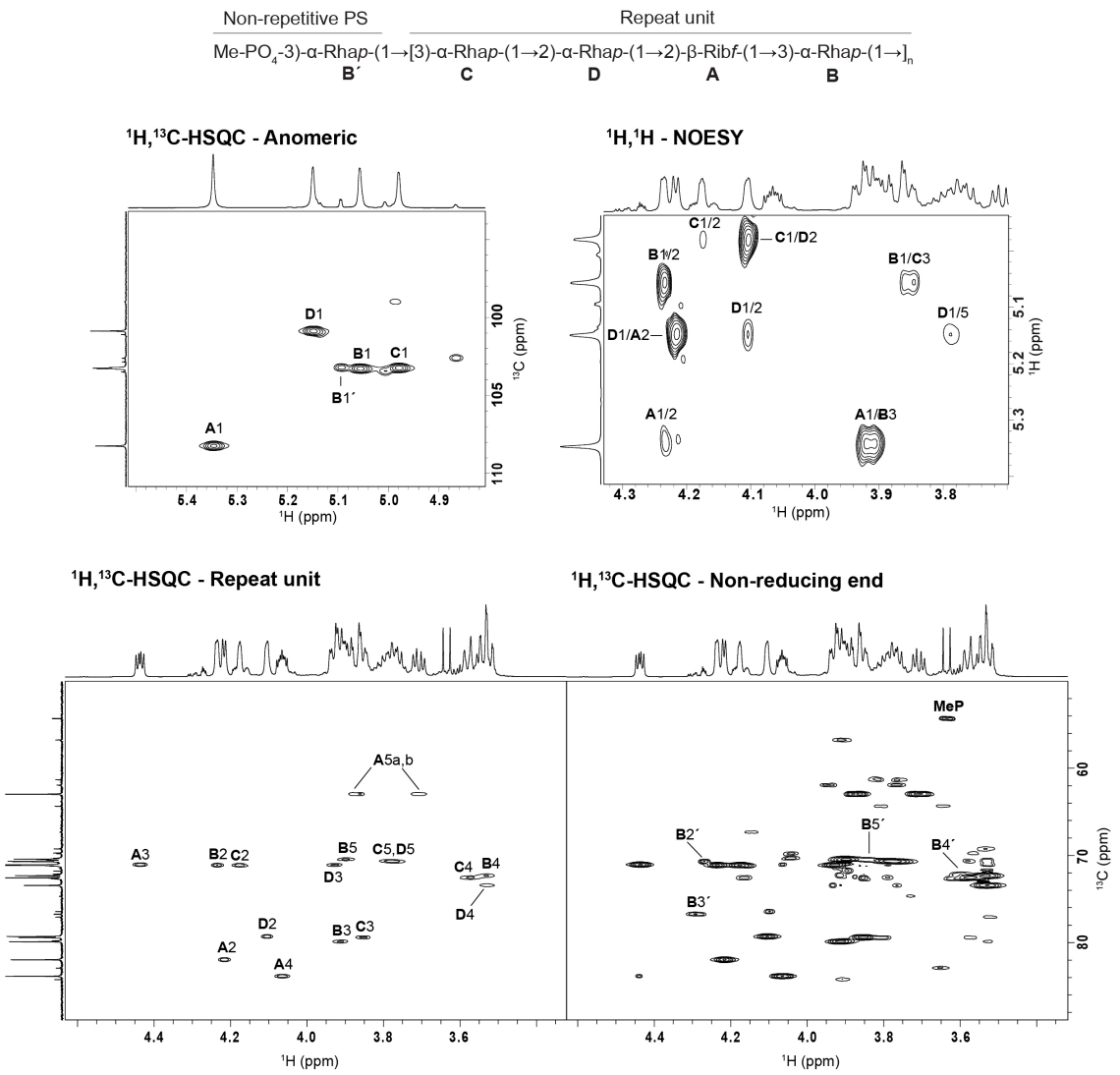


**Figure S2. Parts of $^{1}$H, $^{13}$C-HSQC and $^{1}$H, $^{1}$H-NOESY spectra of purified O7 O-PS.** The letters indicate the sugar residues denoted as shown in the structure (*above*). The anomeric region of HSQC spectrum (*upper-left*) demonstrates the tetrasaccharide repeat unit. The two *lower* panels show the same region of HSQC spectrum: the high level cut (*left*) contains the signals of the internal repeat units, whereas the low level cut (*right*) also shows the minor signals belonging to terminal residues. The signals for the terminal Rha residue at the non-reducing end (**B`**) are indicated. The NOESY spectrum (*upper-right* panel) demonstrates inter-residue correlations between anomeric protons and protons at linkage carbons, which were used to determine the sequence of the sugars in the repeat unit.

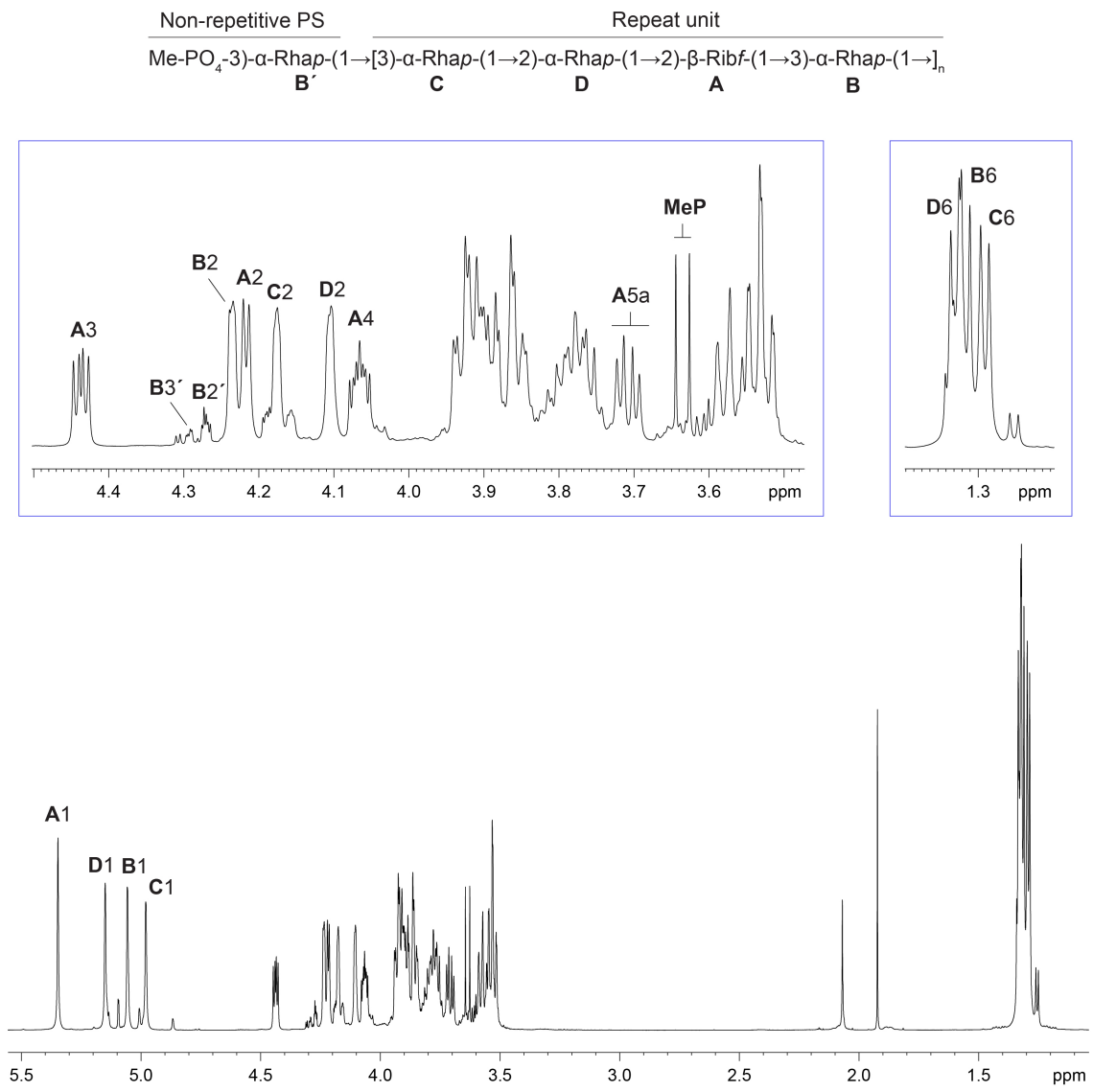


**Figure S3. $^{1}$H NMR spectrum of purified O7 O-PS.** Zoomed-in regions are shown in the inserts. Letters indicate the sugar residues denoted as shown in the structure (*above*).

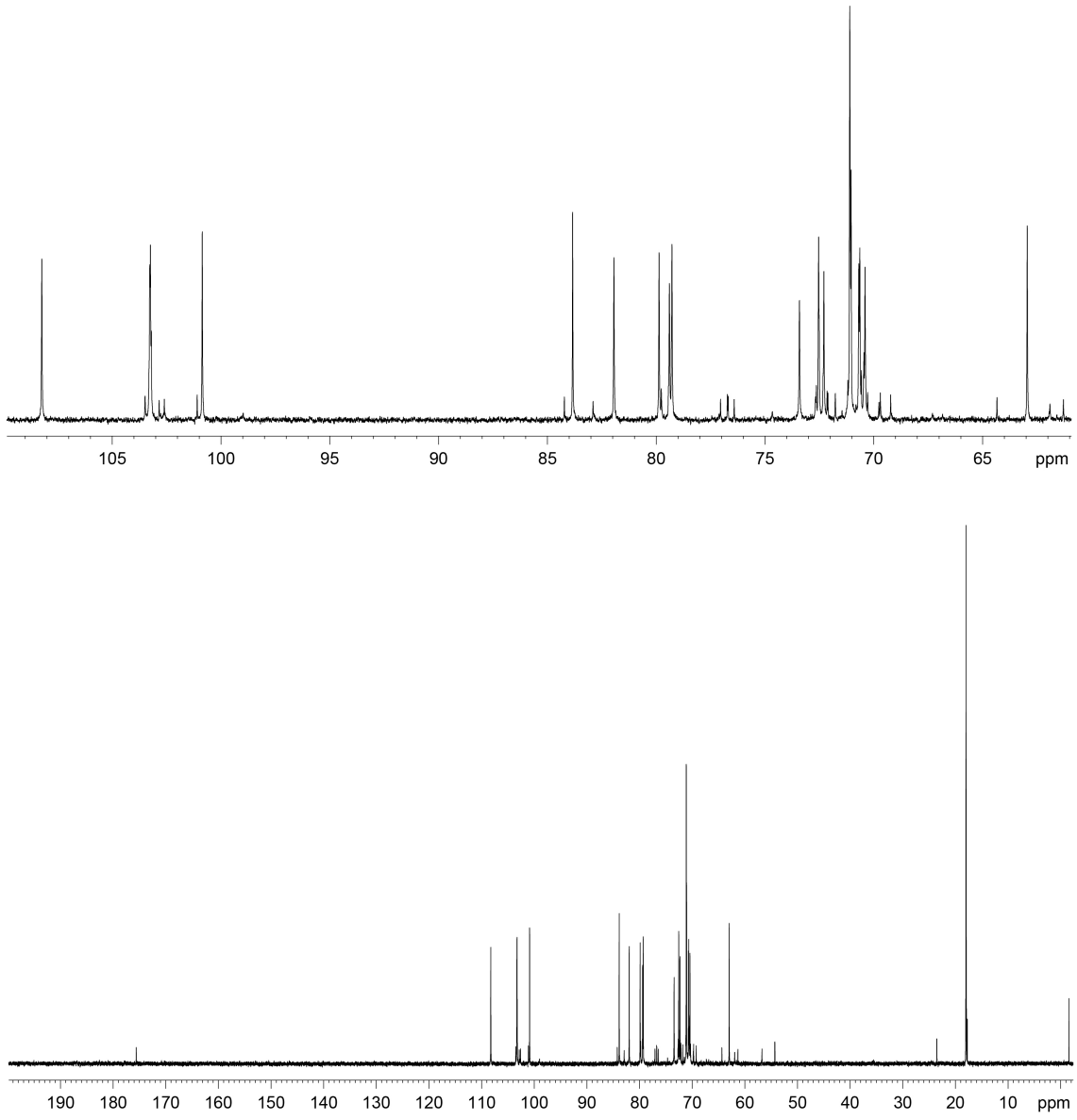


**Figure S4. Full $^{13}$C NMR spectrum (*bottom*) and zoomed-in region (*top*) of purified O7 O-PS.**

**Table S1.** Oligonucleotide primers. Upper case nucleotides indicate regions of gene homology. Underlined nucleotides represent restriction enzyme sites introduced to facilitate cloning.

| Primer | Sequence | Plasmid |
|---|---|---|
| $Wzt_{O7}F$ | 5′-gatcgctagcaggaggaattcaccATGTCATCTAACGATTATGCAATTGAAGTGG-3′ | pWQ1013, pWQ1014 |
| $CBM_{O7}R$ | 5′-gatcccatggTTAGTGATGGTGATGGTGATGCATCGGCGTACAATTCATTTTCACG-3′ | pWQ1013 |
| $Wzm_{O7}R$ | 5′-gatcgctagcaggaggaattcaccATGGCCTCAATCGAACAGGCG-3′ | pWQ1014 |

## References


1. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**, 403–410
2. Osawa, T., Sugiura, N., Shimada, H., Hirooka, R., Tsuji, A., Shirakawa, T., Fukuyama, K., Kimura, M., Kimata, K., and Kakuta, Y. (2009) Crystal structure of chondroitin polymerase from *Escherichia coli* K4. *Biochem. Biophys. Res. Commun.* **378**, 10–14
3. Starbird, C. A., Perry, N. A., Chen, Q., Berndt, S., Yamakawa, I., Loukachevitch, L. V., Limbrick, E. M., Bachmann, B. O., Iverson, T. M., and McCulloch, K. M. (2018) The structure of the bifunctional everninomicin biosynthetic enzyme EvdMO1 suggests independent activity of the fused methyltransferase-oxidase domains. *Biochemistry*. **57**, 6827–6837
4. Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M. K., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Lu, S., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Zhang, D., and Bryant, S. H. (2013) CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* **41**, D348-52
5. Lupas, A., Van Dyke, M., and Stock, J. (1991) Predicting coiled coils from protein sequences. *Science.* **252**, 1162–1164