YesU from *Bacillus subtilis* preferentially binds fucosylated glycans

Joe Tiralongo, Oren Cooper, Tom Litfin, Yuedong Yang, Rebecca King, Jian Zhan, Huiying Zhao, Nicolai Bovin, Christopher J. Day & Yaoqi Zhou

The interaction of carbohydrate-binding proteins (CBPs) with their corresponding glycan ligands is challenging to study both experimentally and computationally. This is in part due to their low binding affinity, high flexibility, and the lack of a linear sequence in carbohydrates, as exists in nucleic acids and proteins. We recently described a function-prediction technique called SPOT-Struc that identifies CBPs by global structural alignment and binding-affinity prediction. Here we experimentally determined the carbohydrate specificity and binding affinity of YesU (RCSB PDB ID: 1oq1), an uncharacterized protein from *Bacillus subtilis* that SPOT-Struc predicted would bind high mannose-type glycans. Glycan array analyses however revealed glycan binding patterns similar to those exhibited by fucose (Fuc)-binding lectins, with SPR analysis revealing high affinity binding to Lewis^a^ and lacto-N-fucopentaose III. Structure-based alignment of YesU revealed high similarity to the legume lectins UEA-I and GS-IV, and docking of Lewis^a^ into YesU revealed a complex structure model with predicted binding affinity of $-4.3 \text{ kcal/mol}$. Moreover the adherence of *B. subtilis* to intestinal cells was significantly inhibited by Le^a^ and Le^b^ but by not non-fucosylated glycans, suggesting the interaction of YesU to fucosylated glycans may be involved in the adhesion of *B. subtilis* to the gastrointestinal tract of mammals.

To date many classes of carbohydrate binding proteins (CBPs) have been identified, including lectins such as F-type, C-type, and Galectins, and carbohydrate binding modules (CBMs) associated with glycoside hydrolases or glycosidases that occur ubiquitously in nature. Lectins regulate numerous crucial biological processes including pattern recognition of pathogens, correct folding of glycoproteins and cell-cell communication. Lectin-oligosaccharide interactions are highly specific due to the branching motifs of either homo- or hetero-polymers of monosaccharide units allowing for structural complexity that are crucial determinants for many biological interactions at a cellular level. Over 50% of all newly synthesized proteins and lipids are glycosylated, resulting in cell surface decoration of glycoproteins and glycolipids that are known to play central roles in cell development, tumour progression and metastasis.

Intensive research aimed at recognizing altered cell surface glycosylation during disease development using CBPs is becoming increasingly important for both biomarker discovery and inhibitor design. As a result it is essential that CBPs encoded within any given genome be identified and their carbohydrate specificity determined. Employing computational function prediction to guide experimental analysis has the potential to significantly aid in this endeavour.

The first generation techniques used to automate protein function annotation was based on sequence homology. Modern approaches now employ machine-learning classifiers based on the protein evolutionary data, structure geometric, or other sequence or structure information. Recently, we introduced a template-based method (SPOT-Struc) that predicts CBPs by making structural alignment between a query structure and the template structure of a known CBP. Highly aligned structures are followed by a binding affinity prediction to further remove potentially false positive CBPs. Using this method we predicted several structural genome targets as CBPs, one of which was YesU from *Bacillus subtilis* (RCSB Protein Data Bank code 1oq1).

1Institute for Glycomics, Griffith University, Gold Coast Campus, QLD 4222, Australia. 2School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, People’s Republic of China. 3Queensland Institute of Medical Research, Brisbane, Queensland, Australia. 4Shermyakin Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia. Correspondence and requests for materials should be addressed to J.T. (email: j.tiralongo@griffith.edu.au) or Y.Z. (email: yaoqi.zhou@griffith.edu.au)
YesU, even after a decade of genome engineering\textsuperscript{14}, remains an uncharacterized protein with unknown function. In 2005, Structural Classification of Proteins\textendash extended (SCOPe) classed YesU as a beta protein with a galectin-like fold, being a member of the concanavalin-A (ConA) like lectins/glucanses superfamily\textsuperscript{15} possessing a $\beta$-sandwich structure comprising 12\textendash14 strands organised as 2 sheets to form a jellyroll topology\textsuperscript{16}. Our SPOT-Struc analysis specifically matched YesU to the integral membrane mammalian protein VIP36 (2e6v)\textsuperscript{13}, a known leguminous type lectin with a $\beta$-sandwich and jellyroll fold\textsuperscript{17,18} that recognizes high Man-type glycans\textsuperscript{19}.

Here we describe the experimental verification of YesU as a CBP, and its functional annotation as a new Lewis\textsuperscript{x} (Le\textsuperscript{x}) binding lectin using a combination of glycan array profiling followed by detailed affinity analysis using surface plasmon resonance (SPR). Our data demonstrates the complementary role of computational prediction and experimental validation in function annotation.

**Results**

**Expression of Recombinant YesU and Glycan Array Analysis.** Recombinant *B. subtilis* hypothetical cytosolic protein 031524 (YesU) was expressed in *E. coli* BL21 (DE3) cells using the vector pMCSG68 obtained from DNASU (Arizona State University), and purified to homogeneity. Figure 1 illustrates the successful purification of highly pure His-tagged YesU (220 residues, 25.2 kDa) using His-select nickel affinity resin (lane 7 coomassie stain and lane 8 Western blot) that was used in subsequent analyses. Size exclusion chromatography showed the YesU in solution is predominantly monomeric with a molecular mass of approximately 25 kDa (Fig. S1), which is consistent with its structural annotation as a monomeric protein. The glycan binding potential and specificity of recombinant YesU was initially assessed using glycan array analysis, a now well-established and widely utilised tool for high-throughput screening of CBP specificity, and whose relative densities are comparable to that seen in a biological context\textsuperscript{26,21}. Interactions were verified through both visual inspection of scanned images and statistical analysis (P < 0.05). The evaluation of YesU by glycan array analysis revealed a limited glycan binding profile, with YesU exhibiting statistically significant binding only to fucosylated glycans (Fig. 2). Importantly, YesU displayed no significant binding to any Man structures present on the array, even though VIP36, the lectin matched to YesU using SPOT-Struc, exclusively binds high-Man type glycans\textsuperscript{17}. Of the fucosylated structures bound, significant binding was observed to fucosylated glycans that share a lacto-N-neotetraose (LNnT) backbone and at least one non-terminal $\alpha$-1-3 fucose (Fuc) (8C, 8H, 8I, 8L, 8N), and to Lewis (8A, 539, 542) and Blood group B trisaccharide (BGBT, 7 M, 540) and non-terminal Fuc-containing glycans, interaction with LnNT (383) and N,N\textsuperscript{″}, N\textsuperscript{″}, N\textsuperscript{‴}, N\textsuperscript{″″}, N\textsuperscript{″″}′ Hexaacetyl chitohexaose ((GlcNAC)\textsubscript{6}, 4D) was also observed (Fig. 2), and as such were also assessed by SPR.

**Surface Plasmon Resonance Analysis.** Glycan array analysis strongly suggested that YesU recognises terminal and non-terminal Fuc-containing glycans. To further quantitatively explore the interaction of YesU with these glycans SPR was performed. SPR analysis revealed preferential binding to Lewis type, Blood Group type and fucosylated LNnT glycans, in particular non-terminal Fuc variants (Table 1 and Fig. S3). A comparison of YesU\textsuperscript{′}s binding to Lewis and related structures demonstrates an increased affinity for non-terminal $\alpha$-1-2/3-Fuc, revealed through a 3-fold greater affinity for Lewis\textsuperscript{x} (Le\textsuperscript{x}, 7I, $K_D$: 0.15 ± 0.006 \mu M), lacto-$N$-fucopentaose III (LNFP III, 7C: $K_D$: 0.14 ± 0.001 \mu M) and Blood group B trisaccharide (BGBT, 7M, $K_D$: 0.19 ± 0.003 \mu M) over Lewis\textsuperscript{a} (Le\textsuperscript{a}, 7J, $K_D$: 0.42 ± 0.004 \mu M) and lacto-$N$-difucohexaose II (LNDFH II, 7E: $K_D$: 0.34 ± 0.01 \mu M) that both

---

**Figure 1.** Coomassie stained gel and subsequent western blot of the expression and purification of *B. subtilis* hypothetical cytosolic protein 031524 (YesU). Lane 1: Biorad precision plus protein standard, Lane 2: uninduced culture, Lane 3: induced culture, Lane 4: slurry mix (protein lysate and resin), Lane 5 wash 1 (binding buffer), Lane 6: wash 2 (wash buffer), Lane 7: concentrated purified YesU protein (220 residues) concentrated using Amicon 3 K 15 mL centrifugal filter (2 mg/mL), Lane 8: Subsequent western blot of purified YesU protein (220 residues) concentrated using Amicon 3 K 15 mL centrifugal filter at the expected 25.2 kDa.
possess a non-terminal α1-4 branched Fuc. Similarly, monofucosyl(1-3)-iso-lacto-N-octaose (MFiLNO, 8 N, \(K_D: 0.24 \pm 0.02 \mu M\)) also showed higher affinity over glycans with terminal Fuc, although with a slight reduction in \(K_D\) probably due to the presence of a terminal lacto-N-biose I (Galβ1-3GlcNAc). The observed decrease in affinity for Lewis^a (Le^a, 7 N, \(K_D: 0.34 \pm 0.002 \mu M\)) could be due to the presence of a terminal α1-2-Fuc. This is supported

| ID | Name | Structure | \(K_D (\mu M)\) |
|----|------|-----------|----------------|
| 7C | LNFP III | Galβ1-4(Fucα1-3)GlcNAcβ1-3Galβ1-4Glc | 0.14 ± 0.001 |
| 7I | Le^a | Galβ1-4(Fucα1-3)GlcNAc | 0.15 ± 0.006 |
| 7M | BGBT | Galα1-3(Fucα1-2)Gal | 0.19 ± 0.003 |
| 8N | MFiLNO | Galβ1-3GlcNAcβ1-3Galβ1-4(Fucα1-3)GlcNAcβ1-6(Galβ1-3GlcNAcβ1-3)Galβ1-4Glc | 0.24 ± 0.02 |
| 7N | Le^b | Fucα1-2Galβ1-4(Fucα1-3)GlcNAc | 0.34 ± 0.002 |
| 7O | BGHT II | Fucα1-2Galβ1-4GlcNAc | 0.34 ± 0.004 |
| 7E | LNDNFH II | Galβ1-3(Fucα1-4)GlcNAcβ1-3Galβ1-4(Fucα1-3)Glc | 0.34 ± 0.01 |
| 7F | HDi | Fucα1-2Gal | 0.36 ± 0.01 |
| 7I | Le^a | Galβ1-3(Fucα1-4)GlcNAc | 0.42 ± 0.004 |
| 1H | LNGT | Galβ1-4GlcNAcβ1-3Galβ1-4Glc | 0.44 ± 0.02 |
| 4D | (GlcNAc)_4 | GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ1-3GlcNAcβ1-3GlcNAcβ1-4Glc | 1.13 ± 0.02 |
| 8H | LNnDFH I | Fucα1-2Galβ1-4(Fucα1-3)GlcNAcβ1-3Galβ1-4Glc | 1.14 ± 0.03 |
| 7A | LNFP I | Fucα1-2Galβ1-3GlcNAcβ1-3Galβ1-4Glc | 1.23 ± 0.02 |
| 8A | Sulfo-Le^a | SO_3-3Galβ1-3(Fucα1-4)GlcNAc | 2.07 ± 0.02 |
| 7K | BGAT | GalNAcα1-3(Fucα1-2)Gal | 7.63 ± 0.30 |
| 1N | α1-3-Galactobiose | Galα1-3Gal | — |
| 10B | Sialyl-Le^a | Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAc | — |
| 5C | α1-2-Mannobiose | Manα1-2Man | — |
| 5F | α1-6-Mannobiose | Manα1-6Man | — |
| 1G | LNT | Galβ1-3GlcNAcβ1-3Galβ1-4Glc | — |

Table 1. Dissociation constants (\(K_D\)) of YesU (50 μg/mL) His tagged to NTA chip for selected glycans. Glycans were tested in triplicate (FC 2/3/4). FC 1 was used as a reference blank to calculate mean \(K_D\) and SD from each flow cell. Chi Square variance is less than 10% for all calculated affinities. Sensorgrams and plots of response at equilibrium against concentration for each glycan for which \(K_D\)’s were determined are given in Fig. S3. LNFP III, Lacto-N-fucopentaose III; BGBT, Blood Group B Trisaccharide; BGHT II, Blood Group H Type II trisaccharide; HDi, H-disaccharide; LNDNFH II, Lacto-N-difucohexaose II; MFiLNO, Monofucosyl(1-3)-iso-lacto-N-octaose; LNnDFH I, Lacto-N-neodifucohexaose I; (GlcNAc)_4, N,N,N′,N″,N″′,N″″-Hexaacetyl chitohexaose; LNFP I, Lacto-N-fucopentaose I; BGAT, Blood Group A trisaccharide; LNnT, Lacto-N-neotetraose; –, No concentration dependent binding.

Figure 2. Glycan array analysis of YesU. Binding of YesU to glycan structures present on the array was assessed both for statistical significance using T-tests (\(p < 0.05\), red bars), and by determining fold-change over background. Black bars indicated statistical insignificant (\(p > 0.05\)) fold-change over background. The Symbol Nomenclature for Glycans (SNFG)<sup>55</sup> was used to represent glycan structures (yellow circle, Gal; blue square, GlcNAc; yellow square, GalNAc; red triangle, Fuc). The full list of glycans present on our array and their corresponding glycan ID and structure is provided in Table S1.
by the observation that the structurally related histo-blood group antigens ABH(O); H-disaccharide (HDi, 7 F, $K_{c}$: 0.36 ± 0.01 μM) and Blood Group H type II trisaccharide (BGHT II, 7 O, $K_{c}$: 0.34 ± 0.004 μM) exhibited a 2–3-fold lower affinity. Similarly, 8 H (LNNFH 1), 7 A (LNFPI), and 7 K (BGAT) displayed a reduced affinity compared to saccharides with only a non-terminal α1-3-Fuc residue. Further the absence of Fuc as in 1 H (LNNT, Gal31-4GlcNAc31-3Gal31-4Glc), or a charged group as in 8 A (sulpho-Lewis*) and 10B (sialyl-Lewis*) to the core Lewis type structure lead to a reduction or complete loss of binding. Glycan array analysis also revealed potential binding of YesU to $N_{1}$, $N_{1}$, $N_{1}$, $N_{1}$, $N_{1}$, $N_{1}$, $N_{1}$, Hexaacetyl chitoheoxase (GlcNAc₆, 4D). Although this interaction was verified by SPR, it was of significantly weaker affinity ($K_{c}$: 1.13 ± 0.02 μM) compared to Le$^{a}$ and LNFPI (Table 1). As was also observed by glycan array analysis, no interaction of YesU with α1-2- and α1-6-mannobiose (5 C, 5 F), α1-3-galactotriose (1N), and LNT (1G) was detected by SPR. Taken together, SPR analysis clearly highlights the importance of a non-terminal Fuc and a lesser extent terminal Fuc for glyc can binding by YesU.

As previously mentioned large fucosylated glycans that share an LNNT backbone and at least one non-terminal α1-3 Fuc (eg. 8 C, 8 H, 8 I, 8L, 8 N) were bound by YesU on our glycan array but the corresponding truncated fucosylated LNNT structures (7A–7E) were not. Quantitative analysis by SPR verified the interaction of YesU with monofucosyl(1-3)-is 2-lacto-N-octoase (MFILO, 8 N; $K_{c}$: 0.24 ± 0.02 μM) and lacto-N-neofucohexaose I (LNNFH 1, 8 H: $K_{c}$: 1.14 ± 0.03 μM), further highlighting the need to consider restricted glycan presentation when assessing glycan array data. Despite this, glycan array analysis successfully predicted the glycan binding specificity of YesU that was subsequently verified and refined by SPR.

**Structure-based alignment and Le$^{a}$ docking analysis.** Glycan array and SPR analyses confirmed the prediction that YesU is a carbohydrate binding protein. However, it was expected that YesU would display a similar Man-binding profile to that described for VIP36, the template that was best matched structurally to YesU (SP-score of 0.80, Table 2). To better understand the discrepancy between the YesU predicted and observed carbohydrate binding specificity, we structurally compared YesU to five Fuc-binding lectins (Table 2 and Fig. 3A), lectin II from *Pseudomonas aeruginosa* (PA-II, pdbID: 1gtz), the *Anguilla anguilla* agglutinin (AAA, pdbID: 1k12), SP2159 from *Streptococcus pneumoniae* (pdbID: 2j1u), lectin I from *Ulex europaeus* (UEA-I, pdbID: 1f5s), and lectin IV from *Griffonia simplicifolia* (GS-IV, pdbID: 1gsl), by SPalign. This analysis revealed that YesU is more structurally similar to the legume lectins UEA-I and GS-IV than it is to the bacterial lectins PA-II and SP2159. Importantly, YesU’s similarity to UEA-1 and GS-IV is comparable to that observed with VIP36. As was also observed by glycan array analysis, no interaction of YesU with α1-2- and α1-6-mannobiose (5 C, 5 F), α1-3-galactotriose (1N), and LNT (1G) was detected by SPR. Taken together, SPR analysis clearly highlights the importance of a non-terminal Fuc and a lesser extent terminal Fuc for glyc can binding by YesU.

| Lectin | Ligand | pbdID | SP-score | Nali* | % alignment | RMSD | % sequence identity |
|--------|--------|--------|----------|-------|-------------|------|---------------------|
| VIP36  | Man01-2Man01-2Man (High Man type) | 2e6v | 0.80 | 183 | 82.0% | 2.46 | 15.5% |
| UEA-I  | Fuc01-2Gal31-4GlcNAc (BGHT II) | 1k5  | 0.75 | 185 | 82.9% | 2.86 | 15.2% |
| GS-IV  | Fuc01-2Gal31-4(Fuc01-4-3) GlcNAc (Le$^{b}$/Le$^{a}$) | 1gsl | 0.75 | 177 | 79.4% | 2.68 | 13.7% |
| SP2159 | Fuc01-2Gal31-4(Fuc01-3-3) GlcNAc (Le$^{a}$) | 2j1u | 0.52 | 113 | 50.7% | 2.97 | 12.2% |
| AAA    | Gal31-3(Fuc01-4)GlcNAc (Le$^{a}$) | 1k12 | 0.50 | 120 | 53.8% | 3.18 | 17.4% |
| PA-II  | Gal31-3(Fuc01-4)GlcNAc (Le$^{a}$) | 1gtz | 0.39 | 79  | 35.4% | 3.06 | 13.8% |

Table 2. Structure based alignment of YesU to five Fuc-binding lectins using SP-align. *Nali, Number of aligned residue pairs; RMSD, Average root-mean-squared distance.

*B. subtilis* adherence to intestinal cells is inhibited by fucosylated glycans. Our data suggests that YesU from *B. subtilis* preferentially binds fucosylated glycans such as Le$^{a}$. Fucosylated glycans, particularly Lewis...
blood group antigens are known to play a functional role in the adhesion of a number of bacteria to the gastrointestinal tract. In order to explore the possibility that YesU has a similar function in B. subtilis we investigated the ability of fucosylated and non-fucosylated glycans to inhibit the adherence of B. subtilis ATCC6633 (expression of YesU in B. subtilis ATCC6633 was confirmed by RT-PCR, data not shown) to the intestinal cell line Caco-2. Figure 5 shows that Lex and Ley at 1 μM were able to significantly inhibit B. subtilis adherence by greater than 50%, whereas the high Man-type glycan, Manα1-3(Manα1-3(Manα1-6)Manα1-6)Man (Man5) showed no significant inhibition at concentrations up to 10 μM, and Galα1-3 Gal and BGBT (Galα1-3(Fucα1-2)Gal) only exhibited significant inhibition at 10 μM (approximately 50% at 10 μM). LNnT (Galβ1-4GlcNAcβ1-3Galβ1-4Glc) was also inhibitory but not to the same extent as Lea and Leb, which mimics the SPR data that showed preferential binding of YesU to Lea and Leb over BGBT and LNnT.

Figure 3. Side-by-side representation of YesU with five Fuc-binding lectins (GS-IV, pdbID: 1gs; UEA-I, pdbID: 1fx5; SP2159 pdbID: 2j1u; AAA, pdbID: 1k12; PA-III, pdbID: 1gzt) and VIP36 (pdbid: 2e6v) that were used for structure-based alignment by SPalign. All bound ligands shown (with the exception of YesU (1oq1)) are those associated with the corresponding PDB entries (A). YesU (magenta) aligned to the structure of GS-IV (cyan) by SPalign (B).

Figure 4. The docking of Lea into YesU using Vina-Carb with default parameters revealed a potential Lea binding orientation in the final complex structure model with predicted binding affinity of −4.3 kcal/mol. Predicted interacting residues are highlighted in yellow (A). The YesU-Lea binding site was also independently predicted by the purely sequence-based technique SPRINT-CBH. Predicted binding residues within 7 Å of the docked Lea are highlighted in yellow (B).
The H, A and B blood group antigens and Le y (Fuc\(^\alpha\) binding site is a depression made up of residues Glu44, Thr86, Asp87, Gly104-Gly105, Ile130, Val134, Asn135, binding domain CBM47 31 of SP2159 34 (also referred to as Man\(^{\beta}\) binding protein VIP36 (pdbID: 2e6v) 17,18. VIP36 exhibits a typical \(\beta\)-sandwich and jellyroll fold17,18 that is a common structural characteristic of the ConA-like lectin/glucanse superfamily15. However, wide scale glycan array screening of Y esU did not display any binding to terminal mannosyl structures or even oligosaccharides containing sub-terminal Man. This lack of binding was validated by SPR with no binding observed to both \(\alpha\)-2-mannobiose and \(\alpha\)-6-mannobiose structures. Our glycan array and SPR analysis instead suggests that Y esU\'s carbohydrate specificity is more closely related to that of Fuc-binding lectins than Man-binding lectins.

**Discussion**

YesU is located between a rhamnogalacturonan acetyylesterase (RhgT) and a DUF624 domain containing protein in the \(B.\ subtilis\) genome. The location of YesU between these two genes is fairly well conserved with 20 complete \(B.\ subtilis\) genomes containing the identical alignment and an additional 32 complete genomes containing the same up and downstream genes around Y esU, with overlap between the gene start and finish codons for YesU and the downstream DUF624 protein. YesU in all genomes assessed had its own promoter (BPPROM27) with the upstream RhgT gene possessing a transcription terminator (using RNAmotif28 and Erpin29).

We had previously predicted, using SPOT-struc, YesU to be a carbohydrate binding protein13 with specificity similar to the Man-binding protein VIP36 (pdbID: 2e6v)17,18. VIP36 exhibits a typical \(\beta\)-sandwich and jellyroll fold17,18 that is a common structural characteristic of the ConA-like lectin/glucanse superfamily15. However, wide scale glycan array screening of Y esU did not display any binding to terminal mannosyl structures or even oligosaccharides containing sub-terminal Man. This lack of binding was validated by SPR with no binding observed to both \(\alpha\)-2-mannobiose and \(\alpha\)-6-mannobiose structures. Our glycan array and SPR analysis instead suggests that Y esU's carbohydrate specificity is more closely related to that of Fuc-binding lectins than Man-binding lectins.

Fuc-binding lectins have been identified in a range of organisms from bacteria to vertebrates. Interestingly, similar to VIP36 and other members of the ConA-like lectin/glucanse superfamily, F-type lectins (fucolectins) also possess a \(\beta\)-barrel with jellyroll topology that defines a typical \(\beta\)-type fold. Two of the best-characterized F-type lectins, both structurally and biochemically, are AAA (\(Anguilla\ anguilla\) agglutinin) from the European eel, and SP2159 from the Gram-positive pathogen \(S.\ pneumoniae\). AAA binds terminal Fuc residues in certain blood group antigens, including H type 1 (Fuc(1-2)Gal\(^{\alpha}\)1-4GlcNAc) and Le\(^\alpha\) (Gal\(^{\alpha}\)1-3(Fuc(1-4)GlcNAc)\(^{\beta}\)), while SP2159 binds to a more distinct set of fucosylated oligosaccharides compared to AAA, including the H, A and B blood group antigens and Le\(^\alpha\) (Fuc(1-2)Gal\(^{\alpha}\)1-4GlcNAc) -Sp8 and Fuc\(^{\alpha}\)1-3GlcNAc\(^{\beta}\) -Sp8 with any significance, and Le\(^\alpha\) (Gal\(^{\alpha}\)1-4)GlcNAc\(^{\beta}\) -Sp2), with any significance, and

Further examples of Fuc-binding lectins include UEA-1 and GS-IV from the plants \(U.\ europaeus\) and \(G.\ simplicifolia\), respectively. GS-IV is known to bind Le\(^\beta\) and Le\(^\alpha\) with high affinity36, and UEA-1 interacts strongly with \(\alpha\)-1,2 linked Fuc specifically H type 2 trisaccharides25. CFG Glycan array data is only available for UEA-1 and GS-IV are similar to other leguminous lectins, including members of the ConA-like lectin/glucanse superfamily, possessing a conventional legume lectin fold that comprises three \(\beta\)-sheets, a six-stranded back sheet, a seven-stranded front sheet, and a five-stranded S sheet which connects the front and back sheets24,25. Although leguminous lectins are structurally similar, their carbohydrate binding specificities and binding site location can vary widely. In the case of UEA-1 and GS-IV even though the binding sites are found at different locations of the protein there is significant homology between the binding site residues (Fig. 3A). Residues involved in carbohydrate binding in the GS-IV lectin are Arg48, Ser49, Asp89, Tyr105, Gly106, Gly107, Phe108, His114, Asn135, Trp138, Tyr22325, and in UEA-1 the binding site is a depression made up of residues Glu44, Thr86, Asp87, Gly104-Gly105, Ile130, Val134, Asn135, Trp137, Tyr220, and Arg223. Structure-based alignment of YesU with GS-IV (Fig. 3) followed by Le\(^\alpha\) docking...
analysis using Vina-Carb (Fig. 4A) revealed a complex structure model for YesU and Leα with binding affinity of $-4.3 \text{kcal/mol}$. Leα was located within a binding pocket comprising residues Leu74, Tyr101, Pro102, His105, Arg118, Tyr121, Glu124, Arg133, Phe138 and Met202 with 5/11 residues in common with GS-IV binding residues. An additional sequence-based carbohydrate-binding site prediction technique SPRINT-CBH further supported the binding site predicted by Vina-Carb. Based on the complex structure models of YesU and Leα shown in Fig. 4, we propose that Tyr101, Pro102, Arg133 and Phe138 stabilize the GlcNAc moiety, while Leu74, Tyr101, His105, Tyr121, Glu124 and Met202 are involved in Fuc binding. The stabilization of the GlcNAc moiety by Tyr101, Pro102, Arg133 and Phe138 more than likely accounts for the observed weak but significant binding to LnNT (1 H, Galα1-3GlcNAcβ1-3Galα1-4Glc) and (GlcNAc)4 (4D) observed by glycan array and SPR analyses. Further experimental studies are needed to unequivocally validate the proposed binding sites.

Similar to the leguminous lectins GS-IV and UEA-I, AAA and SP2159 are comprised of two main β-sheets and four-five anti-parallel β-strands, two forming a shallow positively charged pocket that forms the Fuc-binding pocket36,38. The Fuc-binding pockets in AAA and SP2159 have been attributed to hydrogen bonding with basic residues and Van der Waals contact with hydrophobic residues in the proposed binding pocket36,38. The ability of some GAGs and GAG fragments to interact with YesU may be due to the anionic sulfate interacting with YesU positively charged residues in the proposed Leα binding pocket.

Given our finding that YesU from B. subtilis preferentially binds fucosylated glycans such as Leα it is possible that YesU may be involved in the adhesion of B. subtilis to the gastrointestinal tract of mammals. Fucosylated glycans, particularly Lewis blood group antigens have been proposed to play a functional role in the colonisation and virulence of a number of gastrointestinal pathogens, including Helicobacter pylori41, Campylobacter jejuni42, Salmonella enterica sv. Typhimurium43, and Pseudomonas aeruginosa44. P. aeruginosa is particularly interesting as one of its two soluble lectins, PA-IIL binds host cell membrane44, it does not possess a signal sequence/peptide that predicts subcellular localization as determined using SignalP 4.1 (Gram-negative organism group)45. Similarly, YesU also does not possess a signal sequence/peptide as determined using SignalP 4.1 (Gram-positive organism group), suggesting that PA-IIL and YesU are secreted in a signal sequence/Sec independent process. Although it is unclear exactly how YesU is secreted in B. subtilis, evidence to support signal sequence/Sec independent secretion in Gram-positive bacteria is available. The carbohydrate binding toxin46, pneumolysin (Ply) from Streptococcus pneumoniae is secreted in a signal sequence/Sec independent process. This signal sequence/Sec independent pathway is also present in B. subtilis as Ply knock-in mutants can also export Ply in a signal sequence/autolysis independent manner47. It is therefore possible that the export of YesU utilise this signal peptide/Sec independent process in B. subtilis.

The ability of a range of fucosylated glycans including BGBT, Leα and Leβ, but not the high Man-type glycan, Manα1-3(Manα1-6)Manα1-6Man, to inhibit the adherence of B. subtilis to intestinal cells (Fig. 5) provides experimental evidence to support the potential role of YesU in the colonisation of the gastrointestinal tract by B. subtilis. Leα, in particular, is a very common glycan motif in mammals, and as such represents an important receptor for pathogenic and commensal gut microbes. Although usually considered soil organisms, members of the genus Bacillus have been found to inhabit the gastrointestinal tract of insects and animals46. In fact there is now significant evidence that species such as B. subtilis should be considered gut commensals in humans rather than purely environmental microorganisms48,49. Interestingly, a homolog of YesU is also present in B. cereus, strains of which cause foodborne illnesses in humans48, but is not present in B. thuringiensis and B. sphaericus, both of which are gut commensals in insects. Insects do not express Leα, although they do possess a Lewis-like structure, α1,3-fucosylated N-acetylgalactosaminyl-(3,4-N-acetylglucosamine (GalNAc3,4(Fucα1-3)GlcNAc)48 that resembles Leα except that a GalNAc is present rather than a Gal. The absence of a YesU homolog in Bacillus species that are insect gut commensals together with the lack of Leα epitopes in insects provides further evidence for the potential importance of YesU in the colonisation of the human gut by B. subtilis (and potentially B. cereus).

This paper validates the use of computational prediction from a method like SPOT-Struc as the first step to screen and identify potential new CBPs. We have shown that SPOT-Struc is highly predictive and ideal to uncover novel carbohydrate-binding proteins. However, prediction of the carbohydrate specificities such as reported here for YesU is more challenging for computational techniques as the sidechains rather than backbone structure play a more active role in determining lectin-binding specificity.

Methods

Expression and Purification of B. subtilis YesU Protein. The vector pMCSG68 (clone ID BSCD00606331) was obtained from DNASe plasmid repository (http://dinasu.org). The vector incorporated the B. subtilis hypothetical cysteine protein 031524 (YesU; 220 residues) with a N-terminal 6xHis tag under the control of a T7 promoter. It was transformed into competent E. coli BL21 (DE3) cells and used for recombinant protein expression. An overnight culture of BL21 (DE3)/pMCSG68(BSCD00606331) was used to inoculate LB broth containing ampicillin (100 µg/mL) and incubated at 37 °C with aeration. Once OD600nm reached 0.4–0.6, protein expression was induced using 1 mM IPTG for a further 4 hrs and the cell pellet resuspended in binding buffer (50 mM NaHPO4, 300 mM NaCl; pH 8.0), lysozyme (2 mg/mL), DNase I and protease inhibitor cocktail mix (50 µL). An additional freeze/thaw step was performed to aid in cell lysis followed by sonication and removal of the insoluble cell debris by centrifugation at 100,000 × g for 90 min. The clarified supernatant was added to 1 mL of His-select nickel affinity resin (Sigma) and rotated overnight at 4 °C using a rotational mixer. The slurry mix was then packed by gravity into 10 mL Bio-Rad chromatography column. The column was washed once with binding buffer, then washed with 50 mM NaHPO4, 1 M NaCl; pH 8.0 and the bound His-tagged protein eluted with 50 mM NaHPO4, 300 mM NaCl, 500 mM imidazole; pH 8.0 in 1 mL volume. Imidazole was removed from the sample by dialysis in PBS overnight at 4 °C. Purity was confirmed by SDS-PAGE using 12% SDS-polyacrylamide gels and Western
Blot using a mouse monoclonal anti-His$_6$ (1:10,000 dilution, Cell Signaling Technologies) and a goat anti-mouse horseradish peroxidase conjugated (1:10,000 dilution, Bio-Rad Laboratories) as the primary and secondary antibodies, respectively.

**Glycan Array Analysis.** Glycan arrays consisting of 367 diverse glycans with and without the presence of one of three spacers (sp2, sp3 or sp4) were prepared from two previously described glycan libraries$^{20,48,49}$. Amine containing glycans with spacer’s sp2, sp3 or sp4 were synthesised as previously described$^{19}$ and glycans without spacers were amine functionalised as previously published$^{20}$. All glycans were suspended in 1:1 DMF:DMSO at a concentration of 500 µM and were printed onto SuperEpoxy 2 glass slides (ArrayIt, Sunnyvale, CA) using a ArrayIt SpotBot Extreme array spotter in a six pin subarray print per glass slide format. All glycans were printed in replicates of four, including four FITC control spots, per subarray using 946µM4 pins and a contact time of 1 second at 60% relative humidity, with pins being reloaded after every 12 spots.

The arrays were printed, and subsequently neutralized in 1:1 ethanolamine:DMF, followed by blocking with 0.1% BSA in 50 mM phosphate buffered saline (PBS), pH 7.4 for 5 min at 22°C. After washing with PBS, each slide was dried by placing them in an empty 50 mL tube and centrifuging for 5 min at 200 x g. Recombinant YesU (2 µg) was incubated at a molar ratio of 1:2:4 with anti His-tag mouse monoclonal antibody (10 µg/mL, Cell Signalling Technology), anti-mouse-IgG-Alexa555 conjugated rabbit polyclonal antibody (2 mg/mL, Life Technologies) and goat conjugated anti-rabbit-IgG-Alexa555 polyclonal antibody (2 mg/mL, Life Technologies) in 50 mM Array PBS (PBS with 1.8 mM MgCl$_2$ and 1.8 mM CaCl$_2$), pH 7.4 containing 0.1% BSA for 15 min on ice protected from light. All subarrays on the slide were isolated using a Gene Frame (1.5 × 1.6 cm, 65 µL, Abgene, Epsom, UK) prior to the addition of the YesU-antibody mix to the array. A coverslip was applied to the GeneFrame and array slides incubated in a humidified incubator for 20 min at 22°C in the dark. The GeneFrame and coverslip were subsequently removed and the slide gently washed twice with 50 mM PBS, pH 7.4 containing 0.001% TWEEEN® 20, and twice with 50 mM PBS, pH 7.4. Slides were dried by centrifugation for 5 min at 200 x g and allowed to air dried for a further 5 min.

**Fluorescent Image Acquisition and Data Processing.** Fluorescent intensities of the array spots were measured using the ProScanArray microarray reader (Perkin Elmer, Waltham, MA) using the Helium-Neon 543 green excitation laser set to the Alexa Fluor 555 setting (555 nm excitation and 580 nm emission). Image analysis was carried out using the inbuilt ProScanArray imaging software, ScanArray Express (Perkin Elmer). Raw glycan signals were exported into Microsoft EXCEL. The mean background was calculated from DMF/DMSO blanks on the array plus three standard deviations. This was subtracted from each glycan to generate an adjusted signal. A one tailed T-test was performed with significance set at $p=0.05$. T-test and fold change were generated using Microsoft EXCEL.

**Surface Plasmon Resonance Detection.** Surface plasmon resonance (SPR) experiments were performed using a BLACore T100 biosensor system (GE Healthcare) at 25°C in 10 mM PBS-MgCl$_2$ (pH 7.4) at a flow rate of 30 mL/min. Purified His-YesU was diluted to 50 µg/mL in PBS (pH 7.4) and loaded on flow cell 2 (FC2) of a Ni$^{2+}$-nitrilo-triacetic acid (NTA) Series S sensor chip with 5 min of contact time. This was repeated for FC3 and FC4 to allow for triplicates and validate the reproducibility of the SPR response. FC1 had no protein loaded and was used as a blank reference. Five-fold serial dilutions of selected glycans were prepared in 10 mM PBS-MgCl$_2$ and His-YesU before the injection of the next glycan dilution series. The specificity of the glycan binding was recorded as the response signal difference between each YesU loaded FC and the reference FC1. A 10-min dissociation time was allowed after the addition of each concentration of analyte. SPR signals were analysed using the T100 BLACore Evaluation software and dissociation constants ($K_D$) determine from steady-state analysis.

**Structure-based alignment and Le$^e$ docking analysis.** We utilized the structural alignment of YesU (pdbid: 1q01) over GS-IV (pdbid: 1gs1) to build a candidate complex structure with the lectin Le$^e$/Le$^o$ of 1gs1 and identified a potential binding site. The selected binding site is supported by binding residue predictions generated by SPRINT-CBH$^{26}$. SPRINT-CBH is an orthogonal, sequence-based method, which employs a learned, SVM model to predict carbohydrate-binding residues. The final complex structure model was generated by Vina-Carb$^{51}$; a modified version of Autodock Vina (version 1.1.2)$^{52}$ designed to reproduce native glycosidic torsion angle preferences. In order to identify a potential binding mode, Le$^e$ was relaxed with default parameters inside a 30 Å box at the predicted binding site. We did not perform a de novo docking over the whole structure as docking methods were found to be less accurate than template-based methods if binding sites are not known a priori$^{53}$.

**Caco-2 adherence and glycan inhibition assays.** Adherence assays were performed using Caco-2 human intestinal cell lines essentially as previously described$^{27}$. Briefly, cells were seeded at 10$^5$ cells/well in a black wall 96-well cell culture plates in minimal essential medium (MEM) for 48–72 h prior to bacterial challenge. Caco-2 cells were monitored prior to the assay to ensure a confluent monolayer of cells and that the cells had formed tight junctions. *Bacillus subtilis* ATCC6633 was fluorescently labelled with carboxyfluorescein diacetate, succinimidyl ester (CFDA-SE), and 10$^7$ labelled bacteria were applied to the Caco-2 cells in the presence and absence of selected glycans (at final concentrations between 0.1 and 10 mM). Following incubation at 37°C for 60 min, protected from light, cells were carefully washed 3 times with pre-warmed PBS and the fluorescence measured at an excitation and emission wavelength of 485 nm and 535 nm, respectively using an Infinite® 200 PRO (Tecan) fluorescence plate reader.
53. Ghersi, D. & Sanchez, R. Improving accuracy and efficiency of blind protein-ligand docking by focusing on predicted binding sites.

42. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions.

40. Imberty, A., Wimmerova, M., Mitchell, E. P. & Gilboa-Garber, N. Structures of the lectins from Pseudomonas aeruginosa: insight into the molecular basis for host glyccan recognition. Microbes Infect 6, 221–228 (2004).

41. Consortium for Functional Glycomics. Glycan array screening of a set of soluble lectins from opportunistic bacteria. http://www.functionalglycomics.org/glycomics/HServlet?operation=view&sideMenu=no&psId=primscreen_GLYCAN_v3.221_02142006 (2006).

42. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8, 785–786, https://doi.org/10.1038/nmeth.1701 (2011).

43. Shewell, L. K. et al. The cholesterol-dependent cytolsins pneumolysin and streptolysin O require binding to red blood cell glycans for hemolytic activity. Proc Natl Acad Sci USA 111, E5312–5320, https://doi.org/10.1073/pnas.1412703111 (2014).

44. Price, K. E., Greene, N. G. & Camilli, A. Export requirements of pneumolysin in Streptococcus pneumoniae. J Bacteriol 194, 3651–3660, https://doi.org/10.1128/JB.00114-12 (2012).

45. Nicholson, W. L. Roles of Bacillus endospores in the environment. J Chem Theory Comput 12, 892–901, https://doi.org/10.1021/acs.jctc.5b00834 (2016).

46. Hong, H. A. et al. Bacillus subtilis isolated from the human gastrointestinal tract. Res Microbiol 160, 134–143, https://doi.org/10.1016/j.resmic.2008.11.002 (2009).

47. Kubelka, V., Altmann, F. & Marz, L. The asparagine-linked carbohydrate of honeybee venom hyaluronidase. Glycoconj J 12, 77–83 (1995).

51. Nivedha, A. K., Thieker, D. F., Makeneni, S., Hu, H. & Woods, R. J. Vina-Carb: Improving Glycosidic Angles during Carbohydrate optimization, and multithreading.

Acknowledgements
This work was supported in part by National Health and Medical Research Council (1059775 and 1083450) of Australia to YZ, Natural Science Foundation of China (61271378) to YY, and National Health and Medical Research Council Early Career Fellowship (1091816) to HZ. The authors thank the Australian Research Council grant LE150100161 for infrastructure support. We also gratefully acknowledge the use of the High Performance Computing Cluster “Gowonda” to complete this research. This research/project has also been undertaken with the aid of the research cloud resources provided by the Queensland Cyber Infrastructure Foundation (QCIF). NB was supported by grant No. 14-50-00131 of Russian Science Foundation.

Author Contributions
J.T. and Y.Z. conceived and coordinated the paper. O.C. performed the glycan array and SPR experiments, provided initial data analysis, and contributed to the writing of the paper and the generation of figures. J.T. analyzed, generated the final figures and tables and wrote the majority of the paper. R.K. and J.Z. expressed and purified YesU. Y.Y., H.Z. and Y.Z. provided the initial computational identification of YesU and subsequent analysis. T.L. performed molecular docking. C.J.D., N.B. and LEH provided assistance with the glycan array and C.J.D. with SPR experiments. All authors reviewed the results and approved the final manuscript for submission.

Additional Information
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-31241-8.

Competing Interests: The authors declare no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2018