

## SUPPLEMENTARY INFORMATION

**Supplementary Table 1. Data collection and refinement statistics.**

	SeMet Inflection (native)	Neu5Gc	Neu5Gc $\alpha$ 2-3Gal $\beta$ 1- 3GlcNAc
<b>Data collection</b>	GMCA-CAT	In-house	In-house
Wavelength (Å)	0.9794	1.5418	1.5418
Cell parameters <i>a</i> , <i>c</i> (Å)	97.2, 165.9	96.90, 165.62	97.3, 163.5
Space group	<i>P</i> 6 <sub>5</sub>	<i>P</i> 6 <sub>5</sub>	<i>P</i> 6 <sub>5</sub>
Resolution (Å) / highest resolution bin	2.0 (2.08-2.0)	2.2 (2.20-2.32)	2.1 (2.10-2.22)
No. of reflections	518,836	95,020	99,870
No. of unique reflections	45,985	39,693	45,181
Redundancy	11.4 (10.0)	2.4 (2.2)	2.2 (2.3)
Data completeness (%)	100 (100)	89.0 (93.1)	88.8 (93.3)
Anomalous completeness (%)	87.0 (80.9)	-	-
<i>I</i> / $\sigma$ ( <i>I</i> )	5.0 (1.8)	7.7 (1.8)	6.6 (1.9)
<i>R</i> <sub>merge</sub> (%)	10.9 (41.7)	9.2 (40.6)	10.4 (38.6)
Wilson B (Å <sup>2</sup> )	27.6	28.6	23.7
<b>Refinement</b>			
<i>R</i> <sub>factor</sub> / <i>R</i> <sub>free</sub> (%)	18.7 / 23.3	18.8 / 24.0	21.2 / 26.2
Ramachandran plot			
Favoured (%)	88.6	87.8	90.0
Additionally allowed (%)	10.5	11.2	8.9
Generously allowed (%)	0.8	1.0	1.1
Protein residues	588	586	586
Sugar molecules	-	5	4
PEG 400 molecules	5	5	5
Water molecules	340	480	556
r.m.s.d. bond lengths (Å) /angles (°)	0.019 / 1.706	0.014 / 1.468	0.007 / 1.484
Average B-factors (Å <sup>2</sup> )			
Protein	26.36	20.48	20.54
Sugar molecules	-	32.75	51.19
PEG 400	59.75	63.67	66.68
Water molecules	37.92	31.95	39.86

Data are not shown for Neu5Ac data collection and refinement since there was no evidence for Neu5Ac in any of the SubB binding sites.

**Supplementary Table 2. Comparison of SubB with B subunits of other AB<sub>5</sub> toxin families.**

<b>SubB (118 aa)</b> <b>superimposed on:*</b>	<b>RMSD (Å) / alignment length</b> <b>(C<math>\alpha</math> atoms)</b>	<b>% sequence identity</b> <b>shared with SubB</b>
Ptx S2 (200 aa)	2.17 / 95	18
Ptx S3 (200 aa)	1.86 / 87	15
Ptx S4 (110 aa)	1.65 / 83	6
Ptx S5 (98 aa)	1.57 / 81	15
Ctx (103 aa)	2.26 / 84	7
Stx (69 aa)	2.96 / 63	5
LT (103 aa)	2.29 / 85	3

\*Superpositions were performed using secondary structure matching in the program Superpose, part of the CCP4 suite. The number of amino acids in each subunit are given in parentheses.

Supplementary Table 3. Table of contacts for binary complexes.

<b>Neu5Gc</b>	<b>SubB</b>	<b>Nature of interaction</b>
C1 COOH	Ser 12 OG1 Thr 107 O Ser12 N	H bond Water-mediated H-bond H-bond
C4 OH	Arg 114 O Lys 116 N Asp 8 OD1	Water-mediated H-bond Water-mediated H-bond Water-mediated H-bond
C8 OH	Gln36 NE2	H-bond
C9 OH	Gln36 OE1	H-bond
Amide N	Met 10 O Arg114 O Lys 116 N Asp 8 OD1	H-bond Water-mediated H-bond Water-mediated H-bond Water-mediated H-bond
C11 OH	Met 10 N Tyr 78 OH	H-bond H-bond
Sugar ring	Phe 11	van der Waals
<b>Neu5Gc<math>\alpha</math>2-3Gal<math>\beta</math>1-3GlcNAc (in addition to those above)</b>	<b>SubB</b>	<b>Nature of interaction</b>
C1 COOH (of Neu5Gc)	Ser 12 O	Water-mediated H-bond
C1 COOH (of Neu5Gc)	Ser 106 O	Water-mediated H-bond
Galactose ring O	Thr107 O	Water-mediated H-bond
GlcNAc C=O	Thr 107 OG1 Glu 108 OE2	H-bond Water-mediated H-bond

**Supplementary Table 4. Effect of SubB mutations on cytotoxicity of SubAB for Vero cells**

<b>SubB mutation</b>	<b>Holotoxin designation</b>	<b>Cytotoxicity<sup>a</sup> (CD<sub>50</sub>/μg)</b>	<b>% Inhibition</b>
native	SubAB	819,200	-
S12A	SubAB <sub>A12</sub>	< 200	> 99.98
Q36A	SubAB <sub>A36</sub>	102,400	87.5
Y78F	SubAB <sub>F78</sub>	25,600	96.9

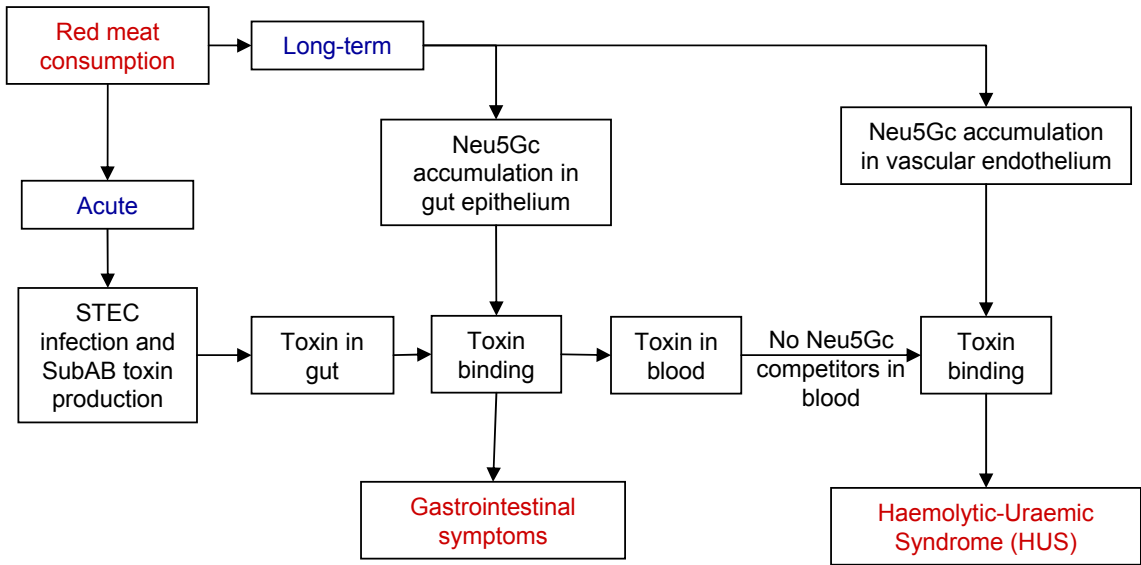
a Cytotoxicity was assayed three times on separate days with identical results.

**Supplementary Table 5. Primers for construction of SubB mutations**

<b>Primer</b>	<b>Sequence (5'-3')</b>
pETsubAF	TTGTAAGGATCCGGAGGAGCTTATGCTTAAG
pETsubBR	ATTATCTCGAGTGAGTTCTTTTTCTGTCAGG
SubBA12F	GGCATGTTTGCAGGCGTTGTTATTACCC
SubBA12R	CAACGCCTGCAAACATGCCATCCCGGGC
SubBA36F	GAGGGGAAAGCATCGGCAGGCTCCTCCAT
SubBA36R	CCTGCCGATGCTTTCCCCTCAATACAAAAATAAG
SubBF78F	GGATTTATTTTAAACCCGGAGTATGGAC
SubBF78R	CCGGGTTTAAAATAAATCCTGACCGGCTG

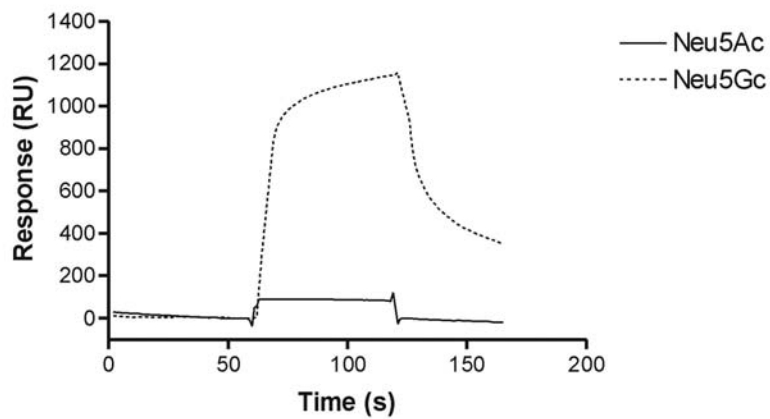
**Supplementary Figure 1. How humans who consume Neu5Gc-rich foods become susceptible to SubAB.** Consumption of Neu5Gc-rich foods such as red meats and dairy products results in metabolic incorporation of Neu5Gc into glycoconjugates on the surface of gut epithelial cells, as well as in microvascular endothelial cells in the kidney, and elsewhere. This enables expression of high affinity receptors for SubAB on the cell surface. The same foods are the commonest source of STEC bacteria, including those strains capable of producing SubAB. In the event of STEC infection due to consumption of contaminated food, toxin produced in the gastrointestinal tract can now bind to receptors on the gut epithelium and cause GI symptoms. Toxin may also be absorbed systemically, and the lack of competing Neu5Gc-glycoproteins in human serum will allow maximal binding of SubAB to high affinity receptors on the endothelium, thereby triggering severe disease such as haemolytic uraemic syndrome (HUS).

**Supplementary Figure 2. Specific binding of SubAB to Neu5Gc.** **A,** Representative sensorgram of 10  $\mu$ M SubAB binding to Neu5Ac- $\alpha$ -2-3Lac $\beta$  and Neu5Gc- $\alpha$ -2-3Lac $\beta$ . **B,** SubAB was passed over immobilised Neu5Gc- $\alpha$ -2-3Lac $\beta$  in the presence of the indicated concentrations of Neu5Gc- $\alpha$ -2-3Lac $\beta$ . The amount of SubAB bound at equilibrium in the presence of the glycan was used to generate the curve shown.  $K_i$  value was determined from the concentration of Neu5Gc- $\alpha$ -2-3Lac $\beta$  that inhibited 50% of binding. The experiments were performed in duplicate and the error bars refer to the standard error of the mean.

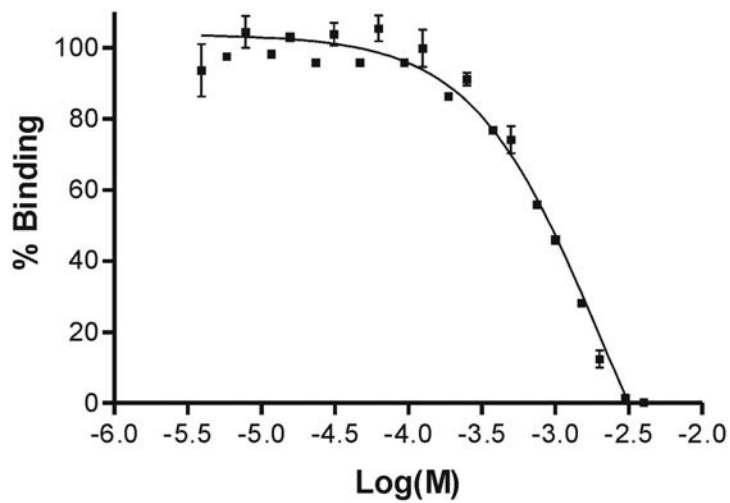


**Supplementary Fig. 1**

A



B



Supplementary Fig. 2