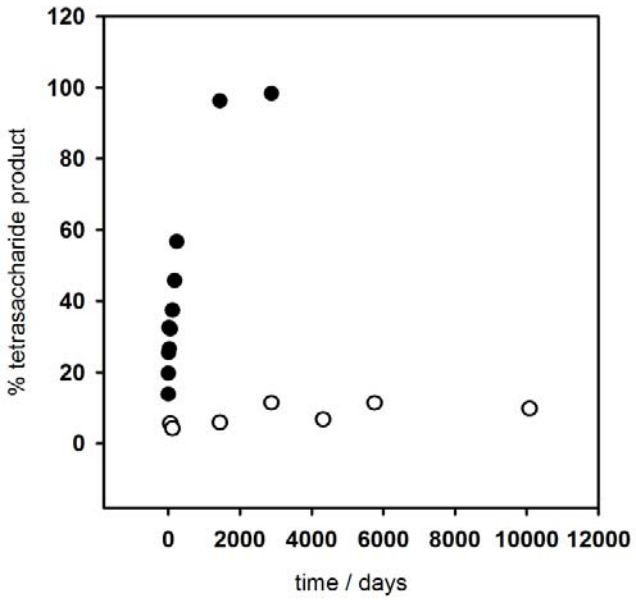
Supporting Information: Bacteriophage tailspikes and bacterial O-antigens as a model system to study weak-affinity protein-polysaccharide interactions

Yu Kang, Ulrich Gohlke, Olof Engström, Christoffer Hamark, Tom Scheidt, Sonja Kunstmann, Udo Heinemann, Göran Widmalm, Mark Santer, and Stefanie Barbirz

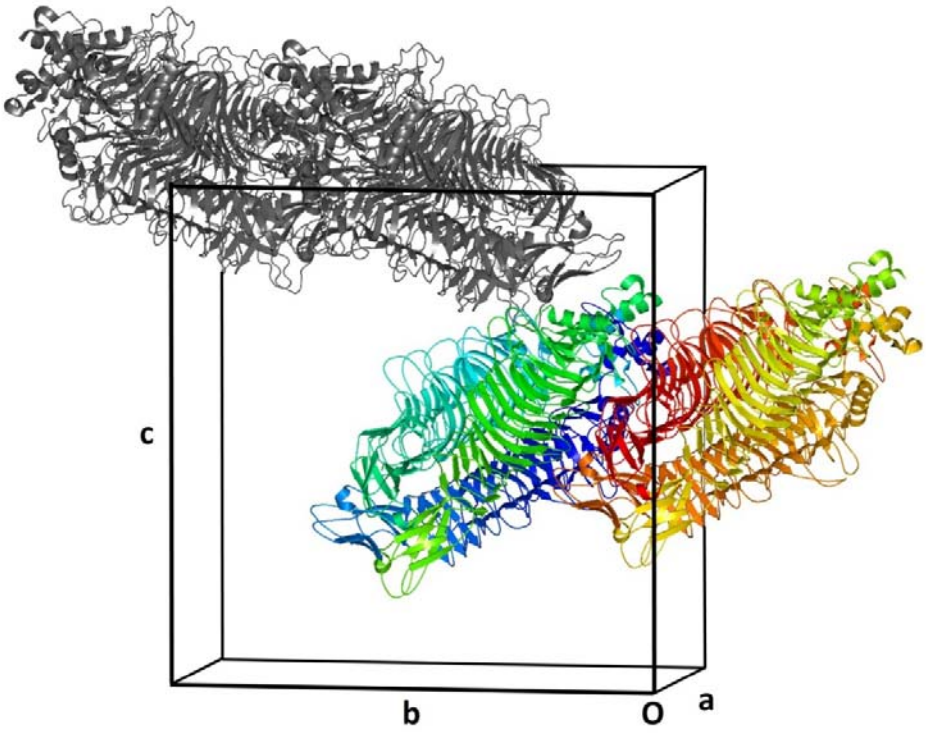
Table of contents

Figure S1: Analysis of catalytic activity of Sf6TSP mutants

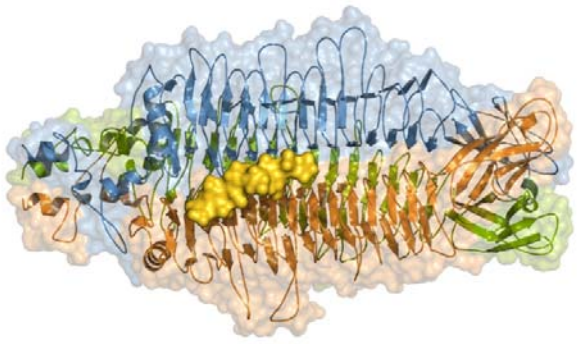


Cleavage of a 7-aminomethylcoumarin labelled dodecasaccharide by Sf6TSP WT (filled circles) or Sf6TSP E366A/D399A at 15 °C. The method has been described in Freiberg, A. *et al.* (2003) *J Biol Chem* **278**: 1542-1548.

Figure S2: Crystal structure of Sf6TSP in complex with *Shigella flexneri* O-serogroup Y octasaccharide

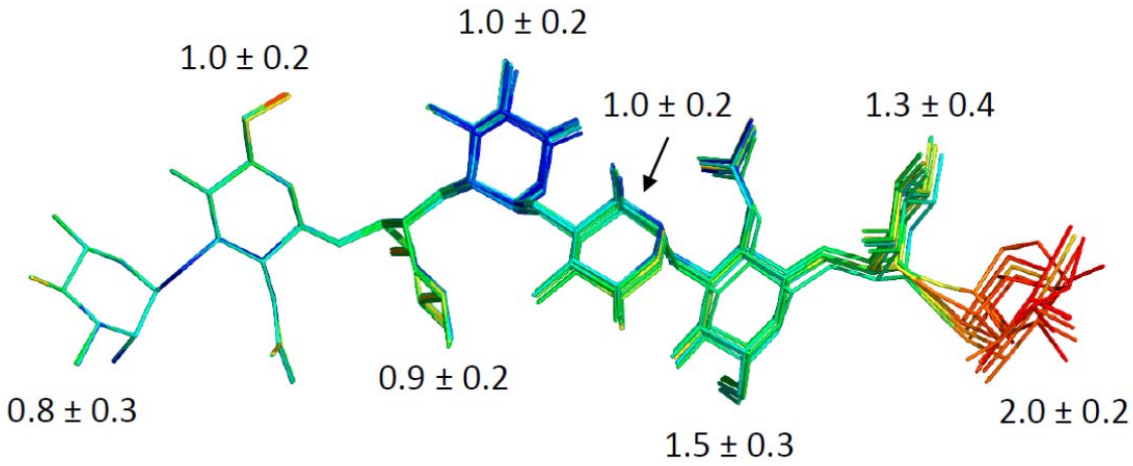


Six polypeptide chains (shown in different colors) resulting in two native trimers were found in the asymmetric unit of Sf6TSP (mutant D399A/E366A). The protein crystallized in in space group $P2_1$. The orientation of symmetry equivalent chains in the next AU is shown in grey.



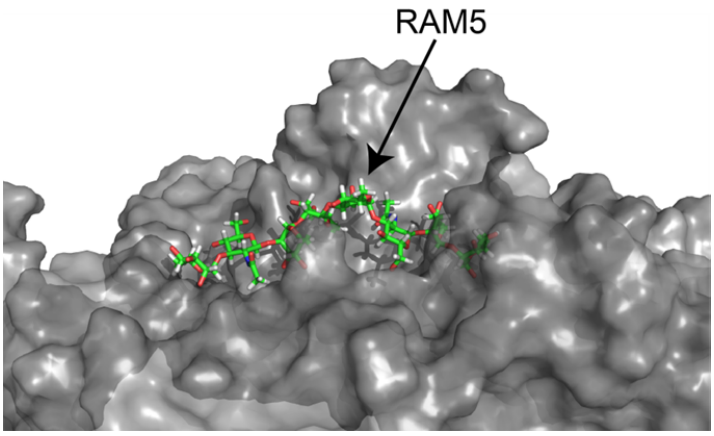
Surface view of Sf6TSP with bound octasaccharide. The N-terminus is pointing to the left with the three subunits of the homotrimer shown as ribbons. The active sites are located between adjacent β-helix domains of two subunits. One of the three bound octasaccharide ligands is shown in gold with the reducing end pointing to the right.

Figure S3: Superposition and alignment of ligands found in the binding sites of the six Sf6TSP chains in the crystallographic asymmetric unit



Numbers annotated are the root mean square deviations (in Ångström) of the single monosaccharide units of an MD simulation at 300K with Sf6TSP D399A/E366A and bound octasaccharide.

Figure S4: Binding pose of octasaccharide ligand B in the X-ray crystal structure of the complex with Sf6TSP D399A/E366A



Binding pose of ligand B of the X-ray crystal structure. Note the longer ligand-protein surface distance for RAM5 due to a pucker in the saccharide chain.

Figures S5 to S9: The CORCEMA ST-CSD approach

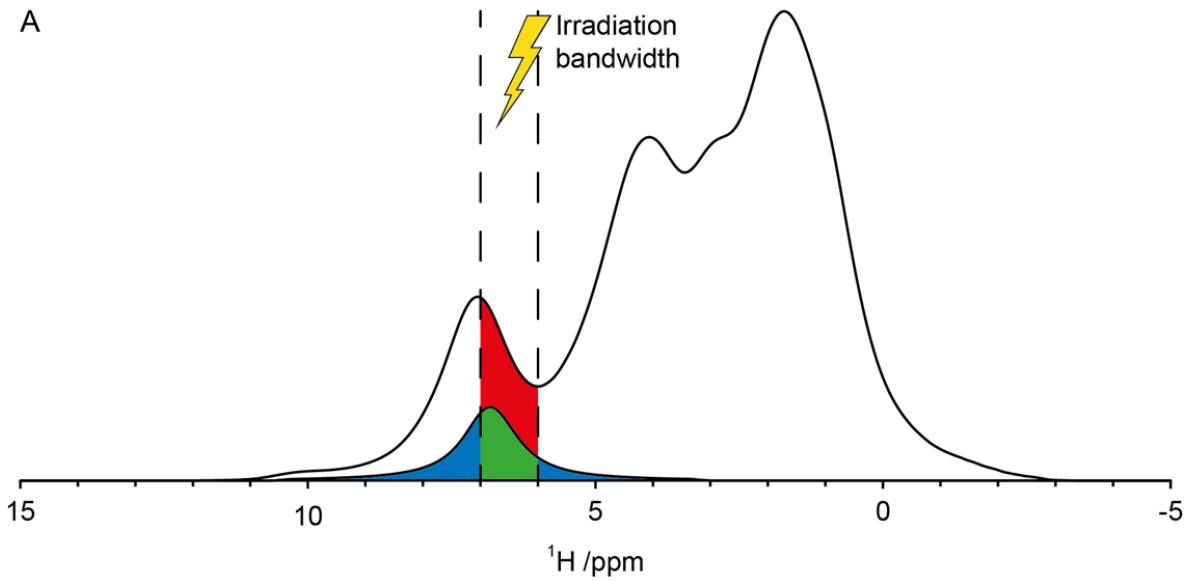


Figure S5. A simulated ^{1}H NMR spectrum of Sf6TSP$_{mut}$; protons within an irradiation bandwidth between 6 to 7 ppm are depicted separately (in front). The green area derives from protons with chemical shifts within the irradiation bandwidth that are still within the irradiation bandwidth whereas the blue areas derive from protons with chemical shifts within the irradiation bandwidth but populate parts of the spectrum beyond the irradiation bandwidth due to relaxation. The red area derives from protons with chemical shifts beyond the irradiation bandwidth but still populate the irradiation bandwidth.

A new approach to analyze NMR-STD data has been developed. The method, designated CORCEMA-ST Chemical Shift Dispersion (CSD), utilizes the full-matrix relaxation software CORCEMA-ST (Jayalakshmi & Krishna, 2002) which is able to simulate NMR-STD data from a protein-ligand structure file (pdb) and a chemical shift file (bmrb). If the chemical shifts of the protein protons are not known, then they can be predicted by various programs (*e.g.* SHIFTX2 (Han *et al.*, 2011)). In a classical CORCEMA-ST approach, STD-AF are calculated by simulating the transfer of magnetization from saturated protein protons with chemical shifts within a defined frequency interval (ideally corresponding to the NMR experiment

saturation pulse bandwidth) to ligand protons. It is generally assumed (Jayalakshmi & Krishna, 2005) that the protein protons are

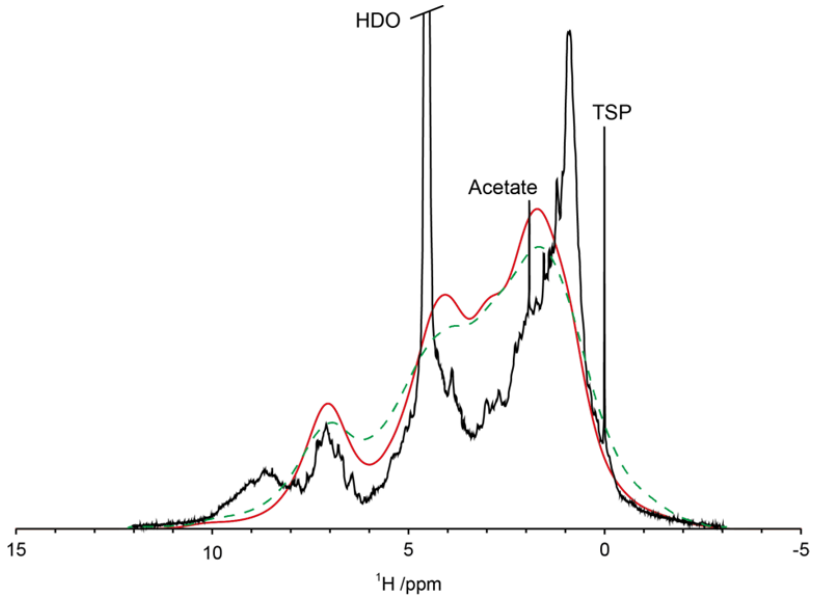


Figure S6. Experimental (black) and simulated ^{1}H NMR spectra of Sf6TSP$_{mut}$ based on the SHIFTX2 chemical shifts predictions of two protein domains (the A and B chains) using either $T_{2,protein}$ = 3.85 ms (red, solid) or $T_{2,protein}$ = 2.5 ms (green, dashed) in the simulations.

instantaneously and fully saturated. However, in the case of a large protein (e.g. 165 kDa for Sf6TSP) together with narrow saturation pulses (24 Hz) the assumption that the protons within the chemical shift interval should be fully saturated is too simplistic. An NMR peak in a spectrum is not a discrete signal at the chemical shift (as reported in the bmrb file) but rather has a Lorentzian distribution, being dependent of the transverse relaxation time (T_2), with the center at the chemical shift. The slow diffusion of a large protein causes significant broadening of the protein proton resonances which can be broader than the bandwidth of the saturation pulses. To use a classical CORCEMA-ST approach to our Sf6TSP system would results in errors of the calculated STD-AF for two reasons: (*i*) protein resonances with chemical shifts within the defined irradiation frequency interval would not be fully saturated as assumed

and (*ii*) protons with chemical shifts close to the irradiation frequency interval would in fact be partly saturated which is unaccounted for in a classical CORCEMA ST approach (Figure S5).

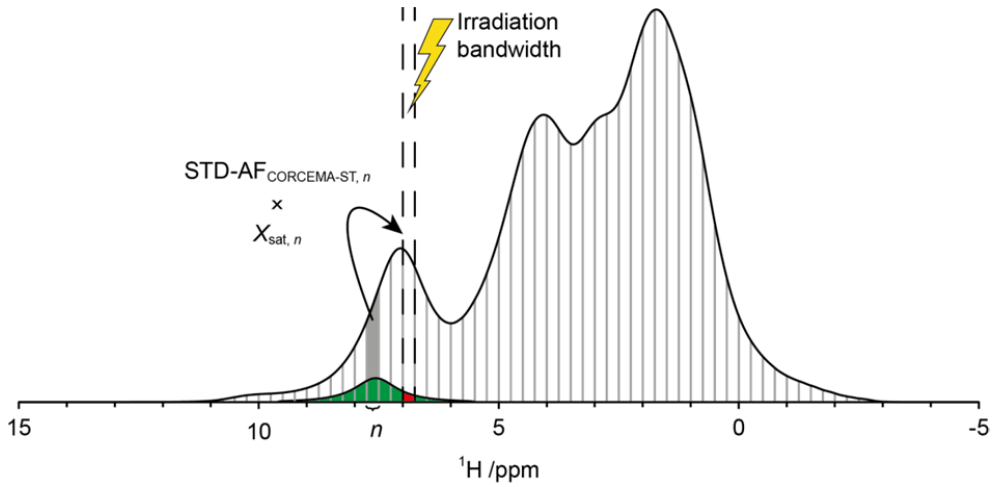


Figure S7. The concept of the CORCEMA-ST-CSD approach. The spectrum is divided into separate regions (*n*) and for each one a CORCEMA-ST simulation is performed to generate a set of STD-$AF_{CORCEMA\text{-}ST,n}$. Each region is then multiplied with $X_{sat,n}$, i.e., the part of the chemical shift distribution for the specific region that is within the irradiation bandwidth (the red area) divided with full distribution (the sum of the red and green areas). The final STD-AF_{CSD} is calculated as the sum of all STD-$AF_{CORCEMA\text{-}ST,n}$ in the spectrum.

In the CORCEMA-ST-CSD approach the NMR line shapes (influenced by T_2) are taken under consideration. The transverse relaxation rate of the protein protons ($T_{2,protein}$) can be predicted from the protein molecular weight by estimating a correlation time using Stokes' law (Cavanagh *et al.*, 1996) and applying the truncated model-free approach (Lipari & Szabo, 1982) in order to generate the spectral densities that are needed to calculate the $T_{2,protein}$. The obtained $T_{2,protein}$ is a rough estimate as it is very sensitive to the definition of the distances for the dipolar relaxation mechanism, but it still gives a good indication of the magnitude of the relaxation time. Another approach is to perform an NMR spectrum simulation of the

predicted chemical shifts and estimate $T_{2,\text{protein}}$ by comparing the simulated spectrum with an experimental one. Such comparisons between two simulated spectra of different $T_{2,\text{protein}}$ and an experimental spectrum of Sf6TSP is shown in Figure S6.

The resolution of the simulated spectrum is modest compared to the experimental NMR spectrum, but the broadening of the peaks is of equal magnitude. In the CORCEMA-ST-CSD approach, the CORCEMA-ST program is looped for narrow intervals (*n*) over the whole spectrum and the contributions to the spectral intensity within the defined irradiation bandwidth ($X_{\text{sat},n}$) are calculated for protons at the active site and with chemical shifts within an interval *n* (Figure S7). The simulated CORCEMA-ST output for each interval *n* (STD-$AF_{\text{CORCEMA-ST},n}$) are multiplied with the corresponding $X_{\text{sat},n}$ and the sum over all *n* is reported as the final simulated result (STD-AF_{sim}, equation 1)

$$\text{STD-AF}_{\text{sim}} = \sum_{n=1}^{N} X_{\text{sat},n} \cdot \text{STD-AF}_{\text{CORCEMA-ST},n} \quad (1)$$

The R_{NOE} factor is a measurement of goodness of fit (a low value indicating a good fit) and it is calculated from simulated and experimental data according to equation 2 where *i* corresponds to a data point at a given saturation time.(Krishna *et al.*, 1978)

$$R_{\text{NOE}} = \sqrt{\frac{\sum_{i=1}^{N}\left(\text{STD-AF}_{\text{sim},i} - \text{STD-AF}_{\text{expt},i}\right)^2}{\sum_{i=1}^{N}\left(\text{STD-AF}_{\text{expt},i}\right)^2}} \quad (2)$$

An R_{NOE} below 0.5 is generally considered acceptable in the literature, but values as low as 0.25 have been reported for oligosaccharide ligands.(Enríquez-Navas *et al.*, 2011) R_{NOE} from CORCEMA-ST-CSD simulations of the Sf6TSP X-ray models (ligands A – F) and the $Sf6TSP_{\text{mut}}$ MD model are reported in Table S1.

Table S1. R_{NOE} factors for the CORCEMA-ST-CSD simulations

Irradiation	$T_{2,prot}$	excluded	MD	X-rayA	X-rayB	X-rayC	X-rayD	X-rayE	X-rayF
6.95-7.00	3.85 ms	-	0.87	1.06	1.41	1.12	1.11	1.18	1.33
6.95-7.00	2.50 ms	-	0.53	0.63	0.91	0.73	0.71	0.76	0.86
6.95-7.00	2.50 ms	A	0.44	0.49	0.58	0.41	0.44	0.46	0.58
6.95-7.00	2.50 ms	RAM8+NAc	0.35	0.51	0.61	0.38	0.43	0.45	0.50
−0.35-(−0.40)	3.85 ms	-	0.58	0.47	0.51	0.49	0.50	0.51	0.49
−0.35-(−0.40)	2.50 ms	-	0.43	0.39	0.47	0.50	0.46	0.47	0.49
−0.05-(−0.10)	2.50 ms	-	0.32	0.51	0.64	0.76	0.62	0.61	0.74
−0.05-(−0.10)	2.50 ms	A	0.33	0.28	0.31	0.41	0.31	0.30	0.40
−0.05-(−0.10)	2.50 ms	RAM8+NAc	0.28	0.27	0.28	0.32	0.26	0.27	0.34

The results using $T_{2,protein}$ = 3.85 ms give moderate R_{NOE} factors (0.5 – 1). The reason for this is a general over-estimation of the STD-AF_{sim} when irradiating at 7 ppm and under-estimation for the −0.4 ppm irradiation. This can be explained by an inaccurate prediction of the protein chemical shifts and of the $T_{2,protein}$. A decrease of the $T_{2,protein}$ will reduce the STD-AF_{sim} from the 7 ppm simulation since this spectral region has a high quantity of predicted chemical shifts. On the other hand, the irradiation at –0.4 ppm is targeting not the predicted chemical shift directly, but rather the broadening of the peaks which will increase with a longer $T_{2,protein}$, thus increasing the STD-AF_{sim} at this irradiation frequency. It can also be seen in Figure S6 that the predicted chemical shifts at low frequency (around 1 ppm) seems to be over-estimated. Therefore, a new CORCEMA-ST-CSD simulations with $T_{2,protein}$ = 2.5 ms and the low frequency irradiation shifted by +0.3 ppm were carried out. The R_{NOE} factor improved (0.3 – 0.9) but was still large for certain X-ray crystal models (namely ligands B – F) which can be explained by an over-estimation of STD-AF_{sim} for the RAM8 residue (Figure S8). The reason why the results are different for the resonances of the RAM8 residue in the A chain model is that the sugar has

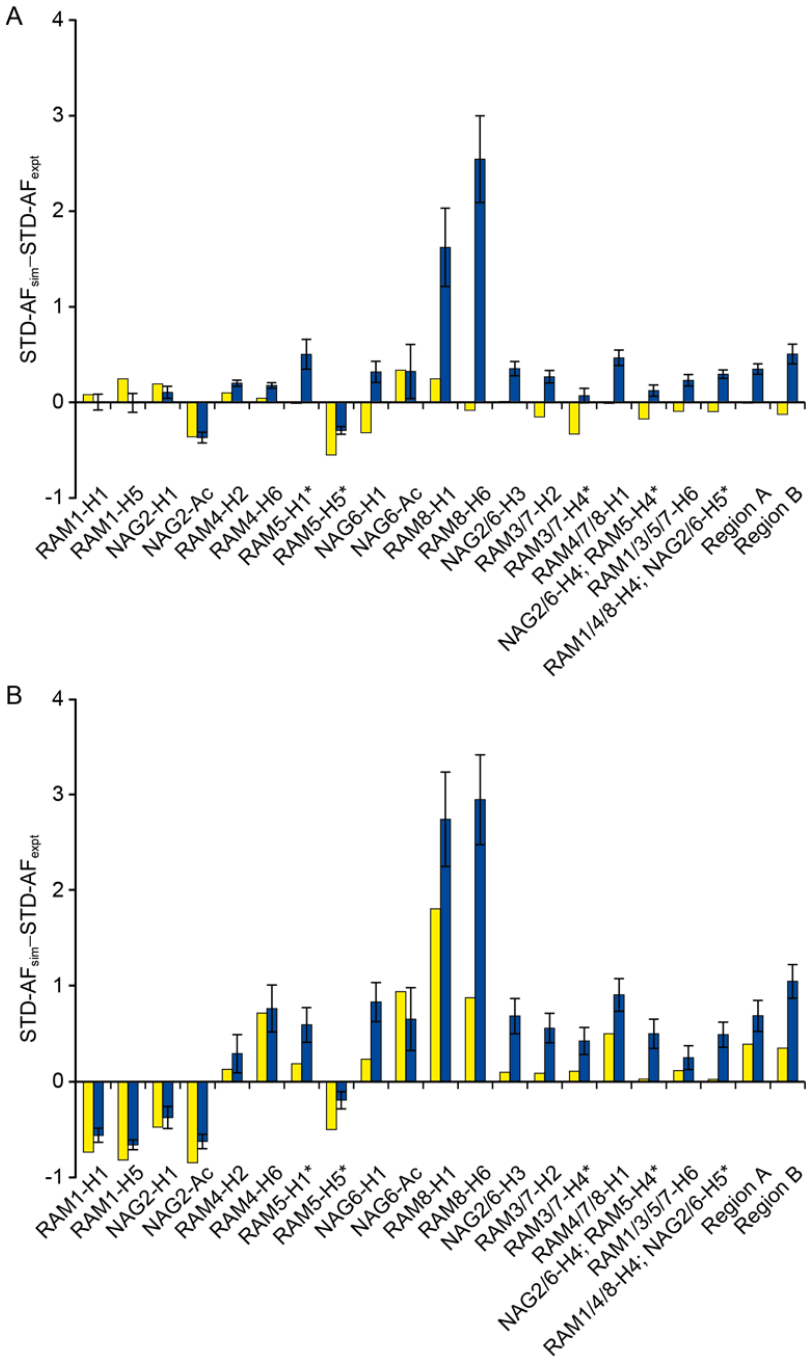


Figure S8. Average differences between simulated and experimental STD-AF of individual resonances for CORCEMA-ST-CSD simulations with $T_{2,protein}$ = 2.5 ms and the irradiation frequency set to the region between –0.05 and –0.10 ppm (A) and between 6.95 and 7.0 ppm (B) calculated from the MD (yellow) and the average for the X-ray ligands A – F (blue). Note the strong simulated intensities of RAM8-H1 and RAM8-H6 of X-ray ligands.

adopted a $^{3}S_{1}$ ring conformation instead of the $^{1}C_{4}$ seen in the other models and as a consequence protons RAM8-H1 and RAM8-H6 are closer to the protein surface in the latter models resulting in too strong STD-AF_{sim} for these protons (Figure S8). If the two protons are excluded from the simulations then the R_{NOE} decrease with between 20% and 50% for all X-ray models. It should be noted that this result does not necessarily support a $^{3}S_{1}$ ring conformation. The low intensity of the experimental STD-AF of RAM8-H1 and RAM8-H6 could also be due to flexibility in this part of the ligand, which is in good agreement with the significantly larger B-factors in the crystal structures and the large RMSD-values for the RAM8 residue than for the rest of the ligand. The R_{NOE} can be reduced further by excluding the STD-AF_{sim} of the *N*-acetyl groups which seems to be under-estimated for NAG2 and over-estimated to the same extent for the NAG6 residue. However, these deviations between experimental and simulated data could arise from less accurate integration of the NMR peaks, due to spectral overlap. It should be noted though that the simulation predicts that the STD-AF of NAG6-Ac is even stronger than measured in the experiment. Build-up curves of STD-AF_{sim} of the MD and X-ray ligand B models and experimental STD-AF are presented for selected resonances in Figure S9. The results from the CORCEMA-ST-CSD simulations are in very good agreement with the X-ray models when the RAM8 residue is excluded. The agreement is likewise good when the MD model is evaluated including all residues (R_{NOE} = 0.32 and 0.53 for irradiations at −0.4 and 7 ppm, respectively). The reason for this is that the dynamic nature of the reducing-end is accounted for in the MD simulations. It should be noted that the CORCEMA-ST-CSD results are calculated without any major iteration of physical parameters (*e.g.* $\tau_{C\text{-protein}}$, $\tau_{C\text{-ligand}}$, K_D etc.) but instead used as initially estimated.

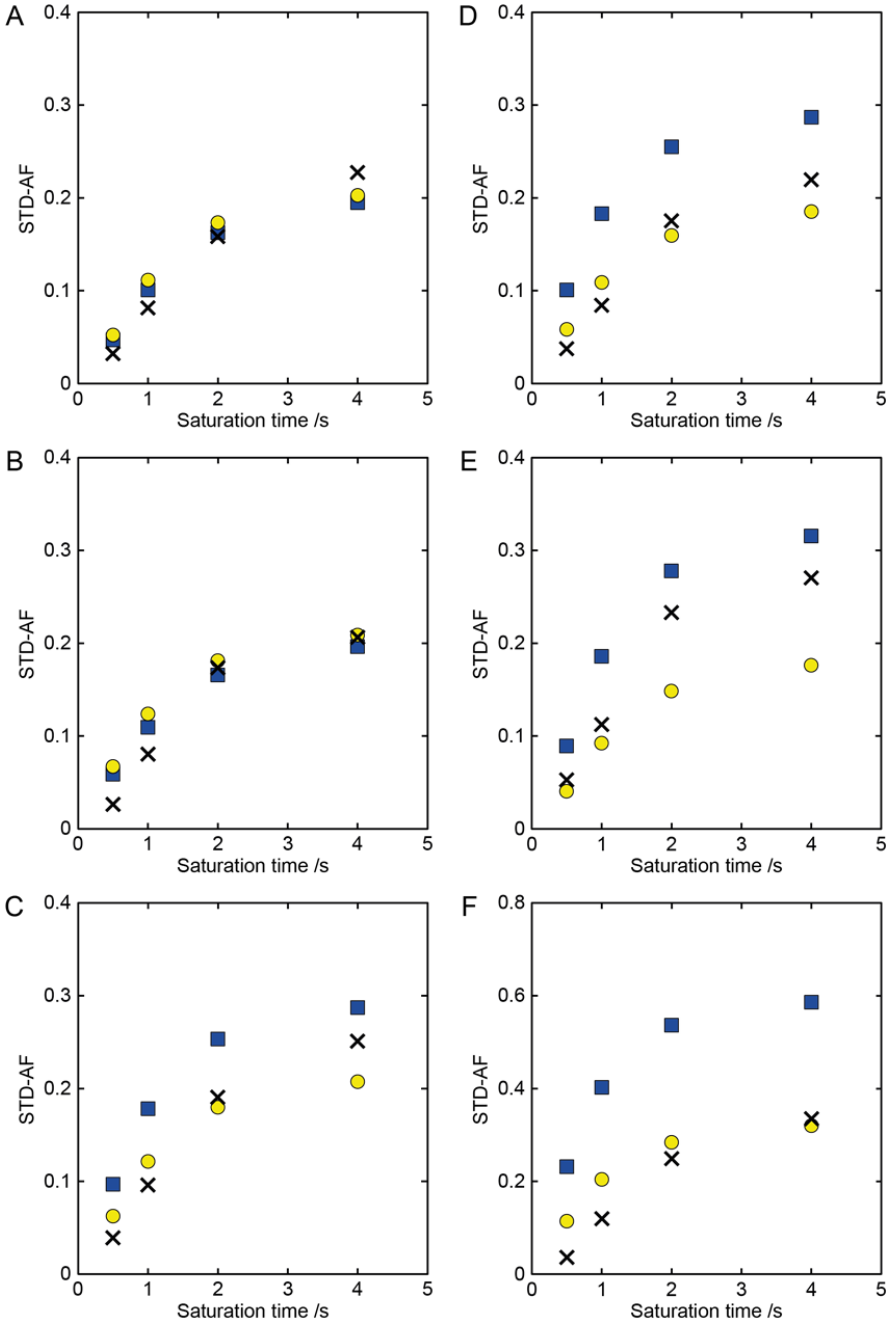


Figure S9. CORCEMA-ST-CSD simulated (MD in yellow circles and X-ray ligand B in blue squares) and experimental (crosses) STD-AF for the anomeric resonances of the octasaccharide (irradiation frequency interval set to the region between –0.05 and –0.10 ppm; $T_{2,protein}$ = 2.5 ms). Panels A – F correspond to resonance RAM1-H1, NAG2-H1, the overlapping signals of RAM3/4/7-H1, RAM5-H1, NAG6-H1, RAM7-H1 and RAM8-H1 in consecutive order.

References for the description of the CORCEMA ST-CSD approach

Cavanagh, J., W.J. Fairbrother, I. Arthur G. Palmer & N.J. Skeleton, (1996) *Protein NMR spectroscopy*. Accademic press, Inc., San Diego, California.

Enríquez-Navas, P.M., M. Marradi, D. Padro, J. Angulo & S. Penadés, (2011) A solution NMR study of the interactions of oligomannosides and the anti-HIV-1 2G12 antibody reveals distinct binding modes for branched ligands. *Chem. Eur. J.* **17**: 1547-1560.

Han, B., Y. Liu, S.W. Ginzinger & D.S. Wishart, (2011) SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* **50**: 43-57.

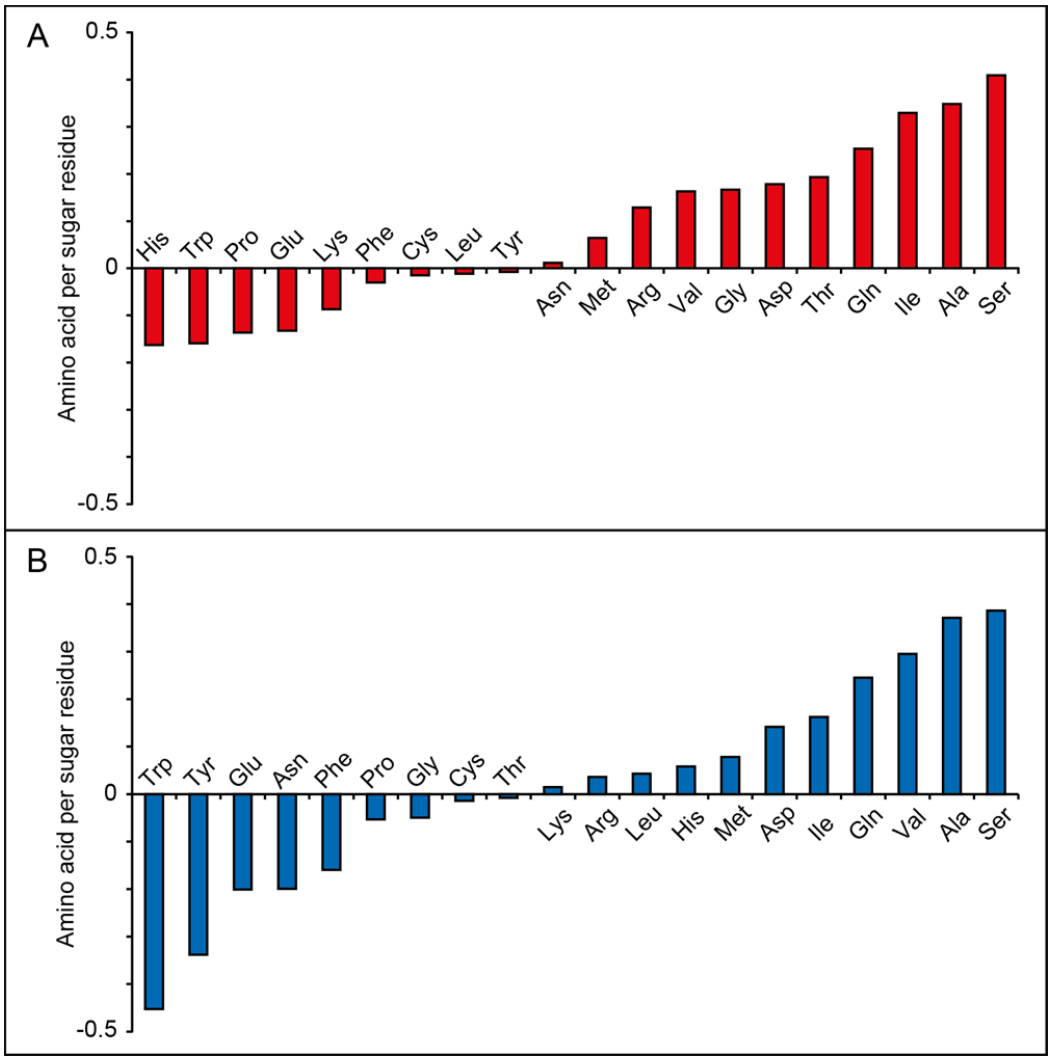
Jayalakshmi, V. & N.R. Krishna, (2002) Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes. *J. Magn. Reson.* **155**: 106-118.

Jayalakshmi, V. & N.R. Krishna, (2005) Determination of the conformation of trimethoprim in the binding pocket of bovine dihydrofolate reductase from a STD-NMR intensity-restrained CORCEMA-ST optimization. *J. Am. Chem. Soc.* **127**: 14080-14084.

Krishna, N.R., D.G. Agresti, J.D. Glickson & R. Walter, (1978) Solution conformation of peptides by intra-molecular nuclear Overhauser effect experiment - Study of valinomycin-K+. *Biophys. J.* **24**: 791-814.

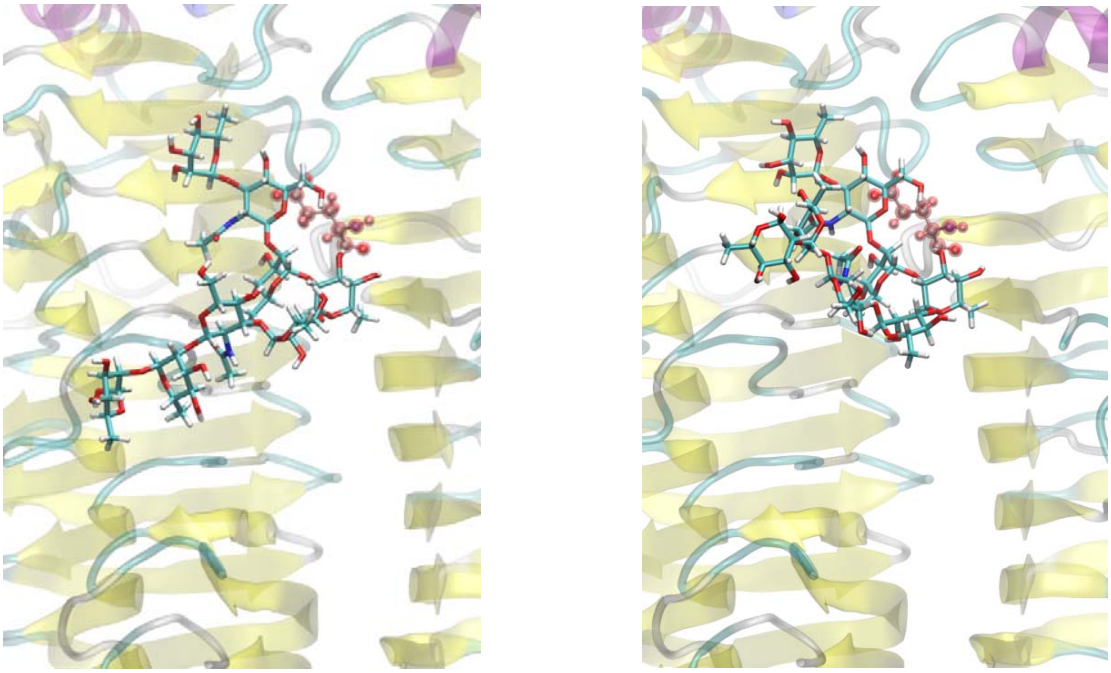
Lipari, G. & A. Szabo, (1982) Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**: 4546-4559.

Figure S10: Amino acid occurrences in the Sf6TSP octasaccharide binding site analyzed with GlyVicinityDB



Differences between the amino acid occurrences in the active site (within 6 Å to the ligand) of the Sf6TSP-octasaccharide X-ray crystal structure and the average amino acid occurrences of all active sites of either α-L-Rha*p* (A) or β-D-Glc*p*NAc (B) binding proteins found in RCSB protein data bank using *GlyVicinityDB* (Lütteke T, Frank M, von der Lieth CW. 2005. Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3D structures derived from the PDB. *Nucleic Acids Res*, 33:D242-246.).

Figure S11: Snapshots from an MD simulation of Sf6TSP D247E in complex with octasaccharide (Amber/Glycam)



Snapshot left: 50 ns, Snapshot right: 80 ns, pink spheres represent residue 247 which was mutated to glutatmate.

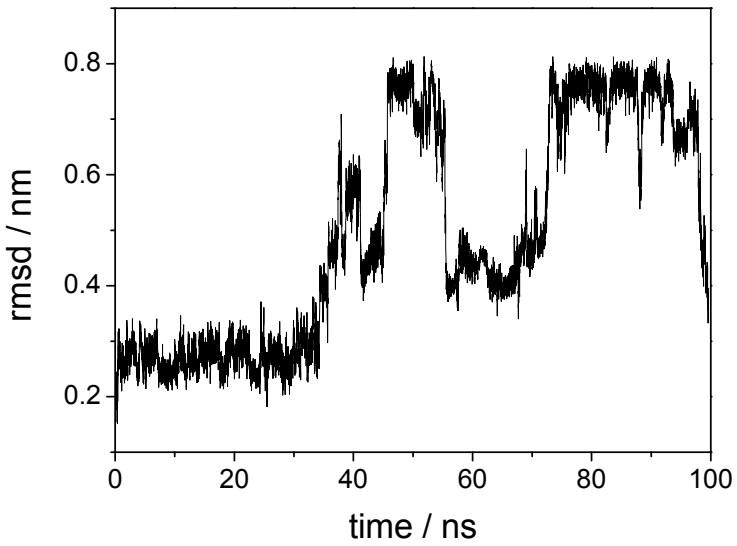
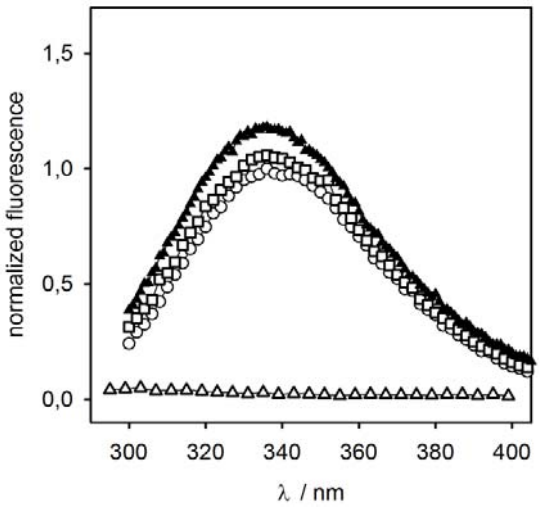


Figure S12: Fluorescence spectra of Sf6TSP mutants in the presence of *S. flexneri* O-serogroup Y polysaccharide and analysis of dissociation kinetics of Sf6TSP mutants from LPS surfaces



(above) Normalized protein fluorescence spectra of 184 nM Sf6TSP E366Q/D399N in the absence (open circles) or presence of 33.2 µg ml^{-1} (open squares) or 66.2 µg ml^{-1} (filled triangles) polysaccharide. Control: 66.2 µg ml^{-1} polysaccharide without protein (open triangles).

(below) Sf6TSPD399A dissociation kinetics from LPS surface at 25 °C. The red line represents the best fit of the data to a single exponential decay yielding a dissociation rate constant of 0.078±0.04 s^{-1}. Each point is the mean value of three independent injections with error bars representing the standard deviation. Valve technology of the SPR instrument used prevented to record reliable responses at earlier time points.

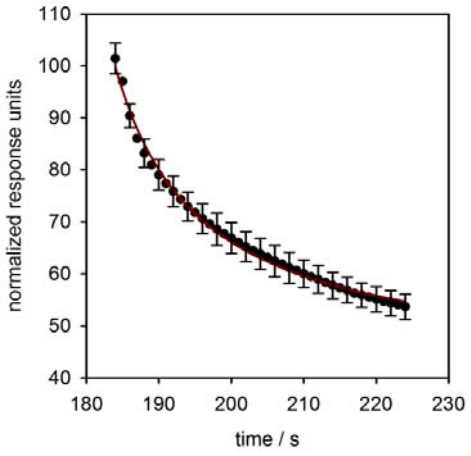
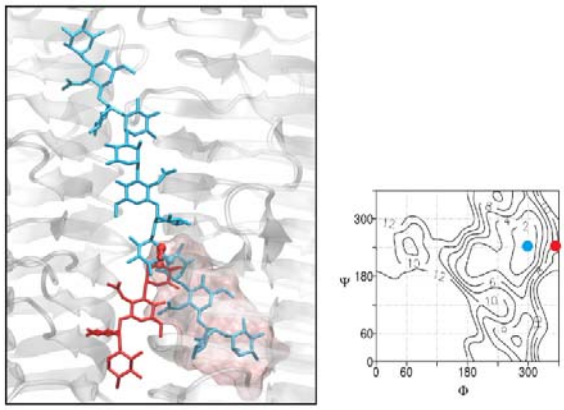


Figure S13: Choice of dodecasaccharide starting structure for molecular dynamics simulations of Sf6TSP-complexes



(Left) *Shigella flexneri* serogroup Y dodecasaccharide as low energy (cyan) or high energy (red) starting conformers for molecular dynamics simulations in complex with Sf6TSP wild type. Regions of protein steric clashes with oligosaccharide are shown in space filling representation. Atoms defining torsions of linkage II shown as spheres. (Right) ϕ-Ψ-positions of linkage II in both starting structures plotted onto free energy landscapes of the unbound disaccharide.

Figure S14: Oligosaccharide ligand dynamics in MD simulations of Sf6TSP complexes

Snapshots from simulations with Glycam/Amber. Color coding of snapshots: Blue sticks: Octasaccharide from crystal structure, Red lines: Dodecasaccharide initial frame at 0.1 ns Yellow sticks: Dodecasaccharide pose (final frame) at the end of the simulation. A: Octasaccharide, B: Dodecasaccharide (22 ns green, 60 ns blue), protein loops free, C: Dodecasaccharide (35 ns green, 60 ns blue), protein loops rigid. rmsd in Ångström calculated from carbohydrate ring atoms.

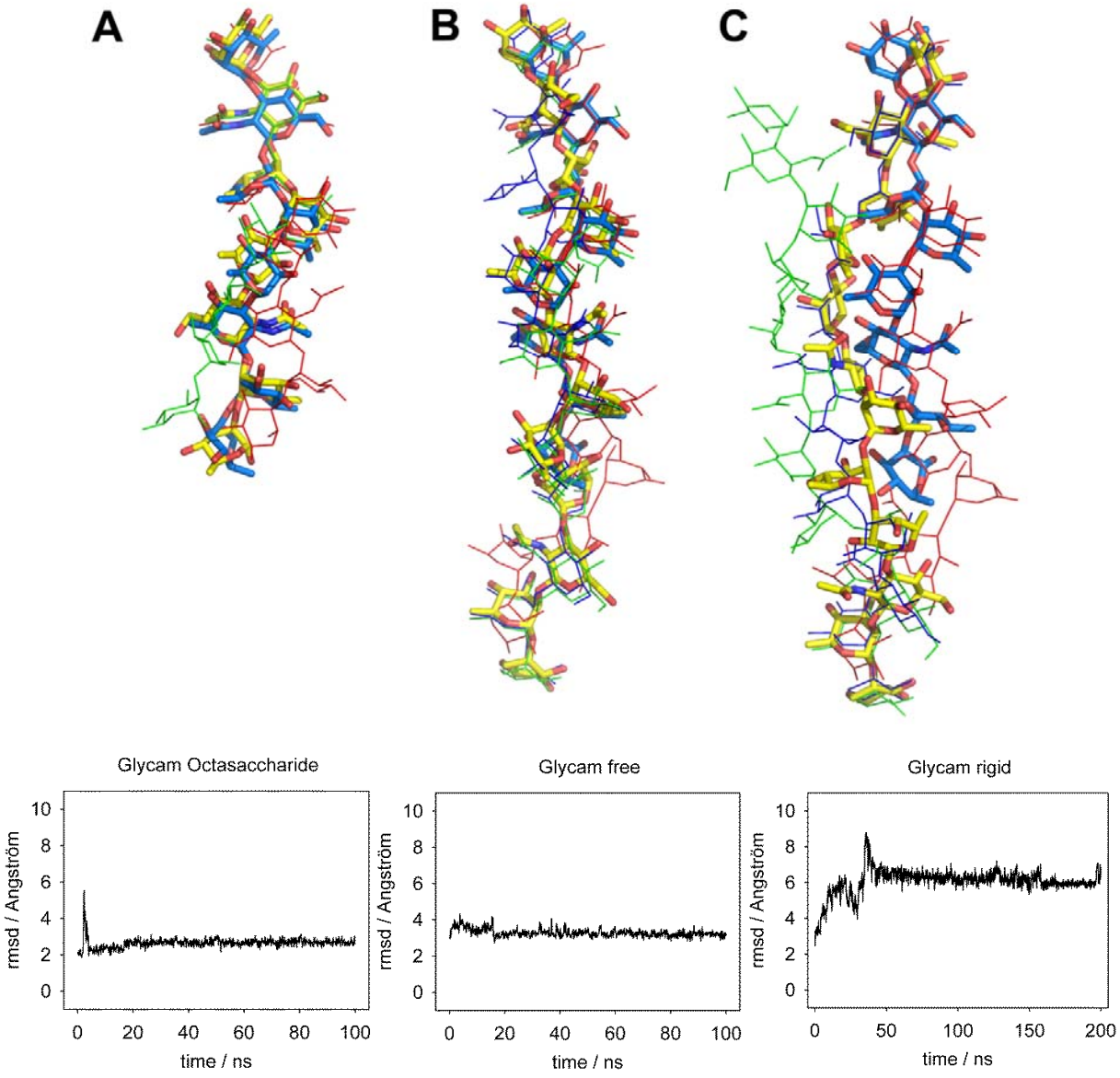


Figure S14 (continued): Oligosaccharide ligand dynamics in MD simulations of Sf6TSP complexes

Snapshots from simulations with CHARMM36. Color coding of snapshots: Blue sticks: Octasaccharide from crystal structure, Red lines: Dodecasaccharide initial frame at 0.1 ns Yellow sticks: Dodecasaccharide pose (final frame) at the end of the simulation. A: Octasaccharide (70 ns green), B: Dodecasaccharide (15 ns green, 20 ns blue, 25 ns grey), protein loops free, C: Dodecasaccharide (35 ns green, 60 ns blue, 150 ns yellow sticks), protein loops rigid. rmsd in Ångström calculated from carbohydrate ring atoms

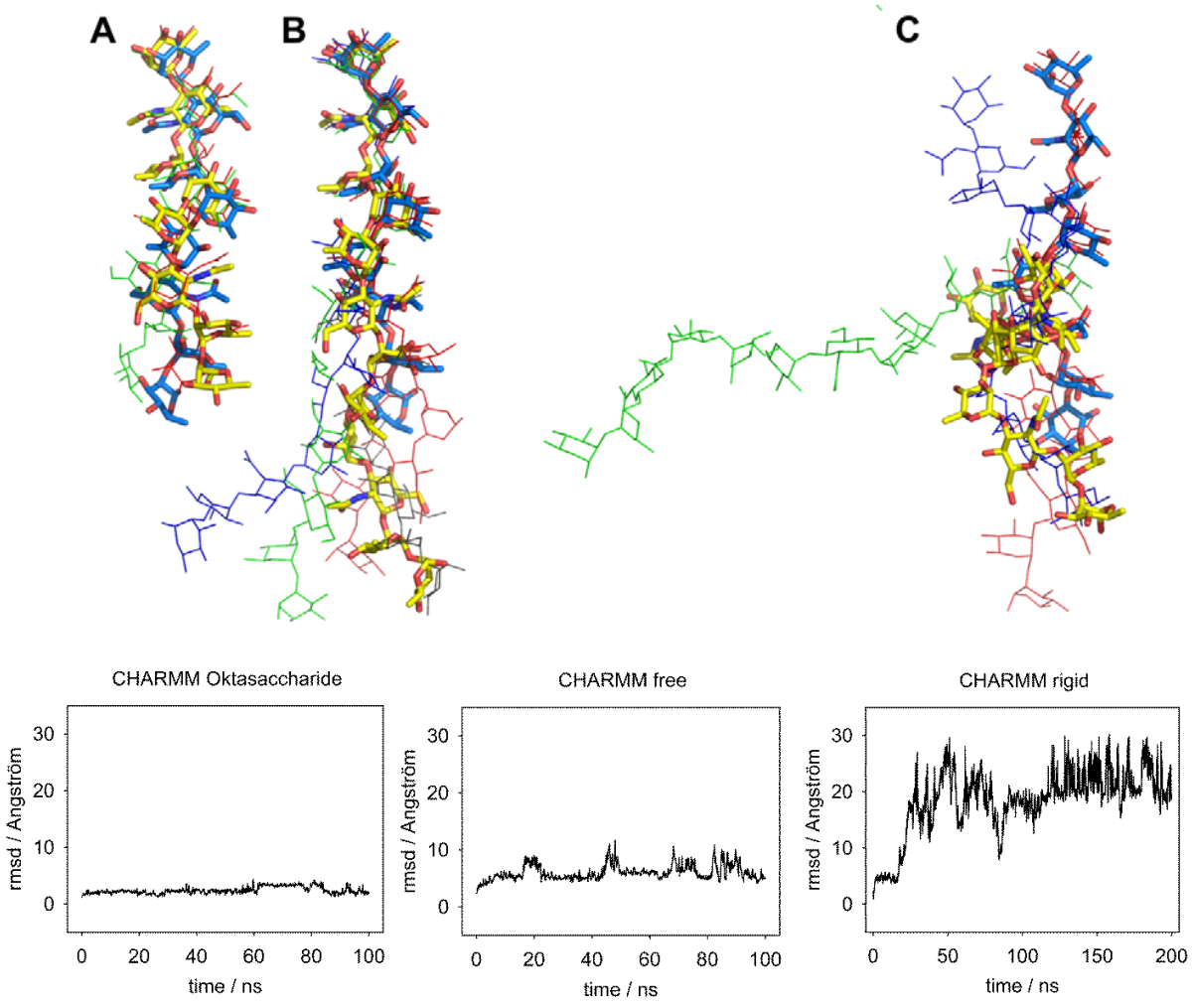


Figure S15: Dynamics of selected β-strand connecting loops in Sf6TSP during MD runs in complex with dodecasaccharide

Rmsd traces of C-α atoms of selected loops simulated with Amber/Glycam either with backbone atoms in protein loops restricted ("rigid") or with flexible backbone atoms. Color coding:

Trace 1 (red) residues 270-280	Trace 2 (blue) residues 353-364
Trace 3 (green) residues 439-444	Trace 4 (magenta) residues 458-465
Trace 5 (orange) residues 425-432	Trace 6 (cyan) residues 339-349

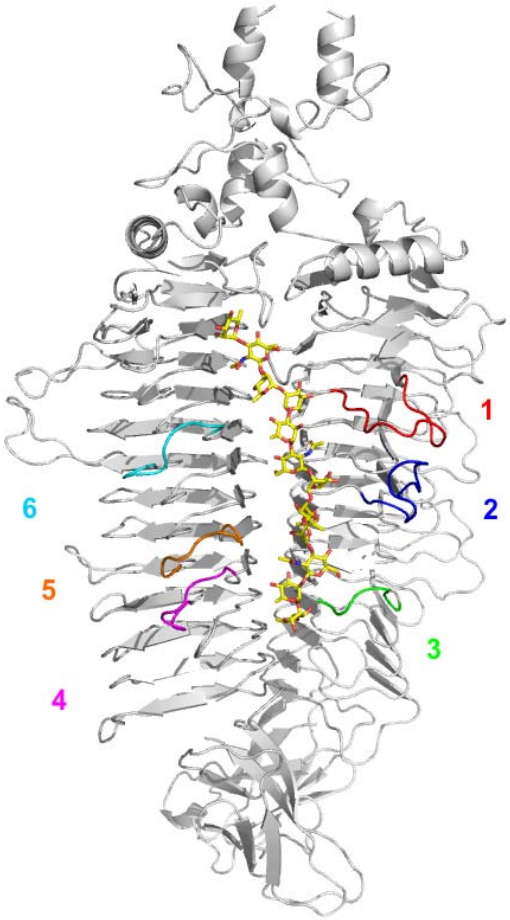


Figure S15 (continued): Dynamics of selected β-strand connecting loops in Sf6TSP during MD runs in complex with dodecasaccharide

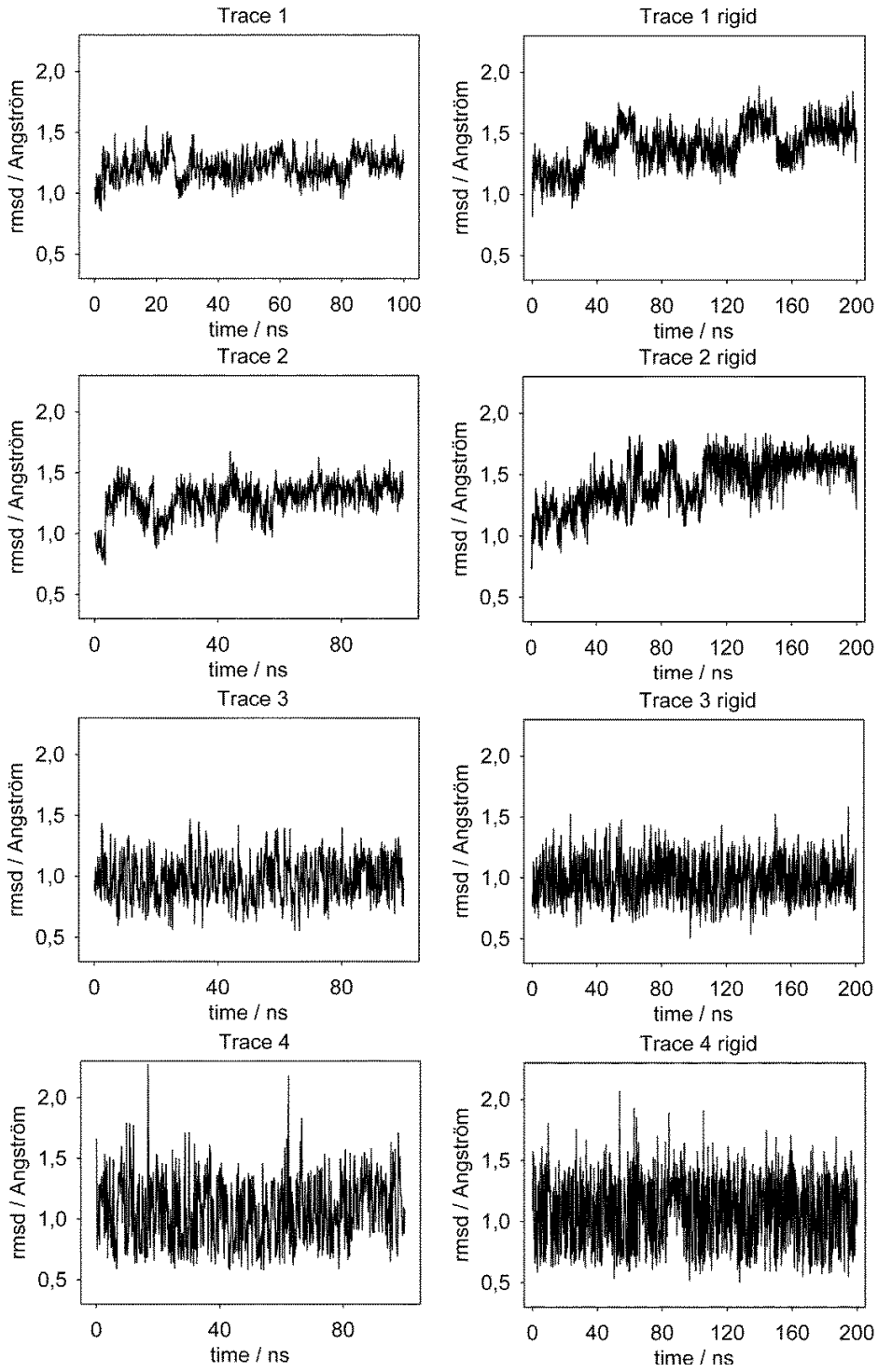


Figure S15 (continued): Dynamics of selected β-strand connecting loops in Sf6TSP during MD runs in complex with dodecasaccharide

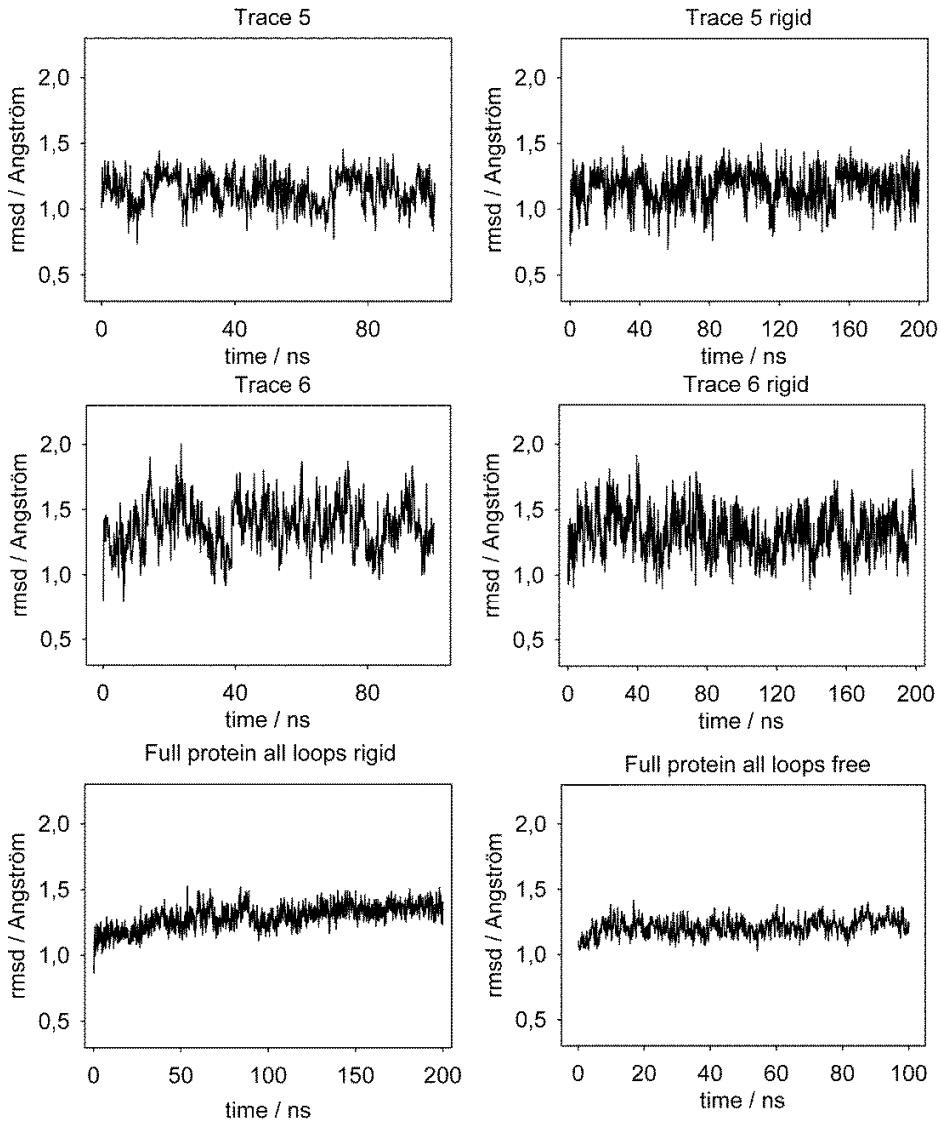
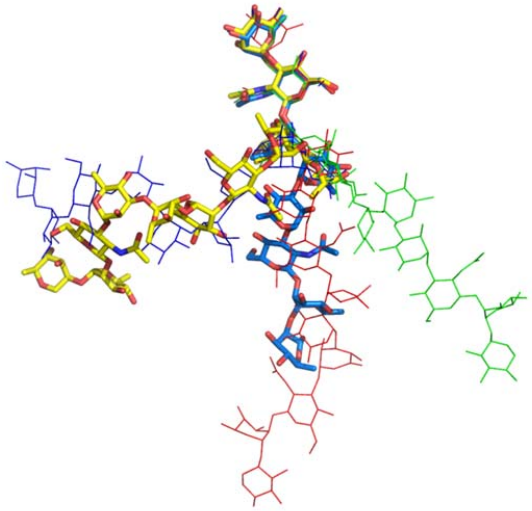


Figure S16: MD simulation of a Sf6TSP dodecasaccharide complex with Glycam/Amber, restrained protein loop backbone and restrained position of RU1

Snapshots from simulations with Glycam/Amber. Color coding of snapshots: Blue sticks: Octasaccharide from crystal structure, Red lines: Dodecasaccharide initial frame at 0.1 ns Yellow sticks: Dodecasaccharide pose at 80 ns of simulation. Green lines 20 ns, blue lines 40 ns of simulation. rmsd in Å calculated from carbohydrate ring atoms.



Glyc_1stRU_restricted

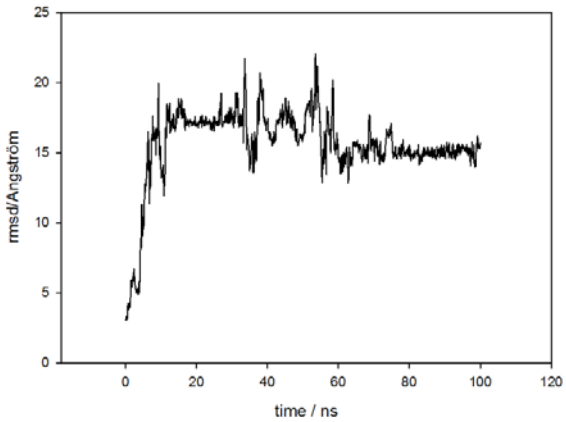
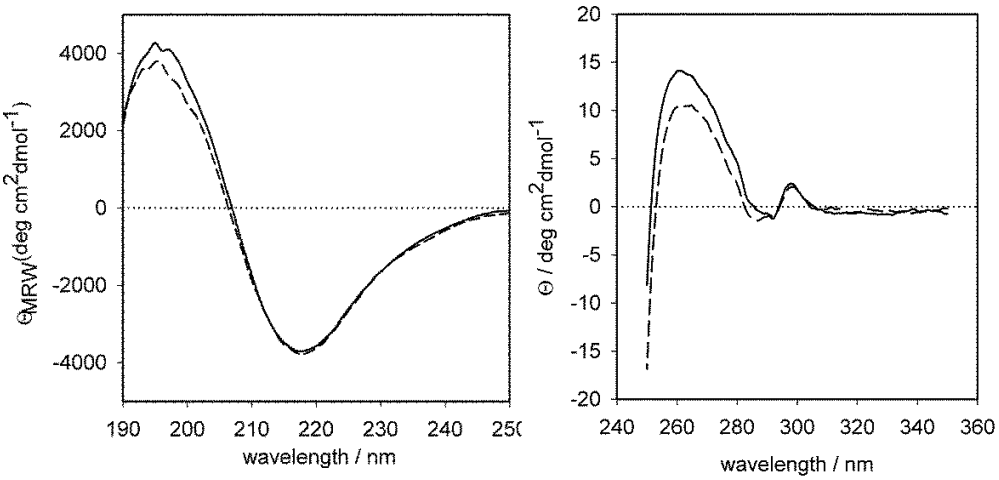


Figure S17: Stability of Sf6TSPD399N at 56 °C analyzed by near and far UV CD-spectroscopy



In near-UV CD spectroscopy (right) a 36.8 µM solution (2 mg/ml) of TSP ΔN D399N in 50 mM sodium phosphate buffer pH 7 was measured in a 5 mm cuvette, CD spectra at 20°C and 56°C were recorded and corrected with spectra of 50 mM sodium phosphate pH7 and the respective temperatures. Far-UV-spectroscopy (left) was performed in a 1 mm cuvette with a 3.68 µM solution (0.2 mg/ml) of TSP ΔN D399N in 50 mM Sodium phosphate pH 7 buffer in wavelength range from 250-190 nm.

Table S2
Crystallographic data collection and model refinement statistics for Sf6TSPΔN, E366A/D399A + *Shigella flexneri* O-serogroup Y octasaccharide

Data Collection	
Wavelength [Å]	0.91841
Temperature [K]	100
Space group	P 2_1
Unit Cell Parameters	
a, b, c [Å]	82.99, 138.90, 136.07
β [°]	92.05
Resolution range [Å][a]	46.30-1.95 (2.05-1.95)
Reflections [a]	
Unique	225,609 (32,338)
Completeness [%]	99.7 (98.5)
Multiplicity	4.5 (4.6)
Data quality [a]	
Intensity [I/σ(I)]	6.2 (2.1)
R_{meas} [%]	24.3 (83.7)
CC*	0.994 (0.786)
Wilson B value [$Å^2$]	14.3

[a]Data for the highest resolution shell in parenthesis

Continued on next page.

Refinement	
Resolution range [Å][a]	45.33-1.95 (2.00-1.95)
Reflections[a]	
Number	223,380 (16,453)
Completeness [%]	99.9 (99.9)
Test Set (5%)	11,044 (163)
R_{work}	0.195 (0.220)
R_{free}	0.235 (0.264)
Contents of the Asymmetric Unit	
Protein, Molecules, Residues, Atoms	6, 3,050, 25,191
Ligands (Molecules)	octasaccharide (6), ethylene glycol (5), Mn (10),PO_4 (4)
Water, Molecules	1,559
Mean Temperature factors [Å^2][b]	
All Atoms	18.8
Macromolecules	18.6
Ligands	28.3
Water Oxygens	17.6
RMSD from Target Geometry	
Bond Lengths [Å]	0.01
Bond Angles [°]	1.16
RMSD Temperature Factors [Å^2]	
Main-chain	1.8
Side-chain	4.0
DPI[c] [Å], based on R_{free}	0.154
Validation Statistics[d]	
Ramachandran Plot	
Residues in Allowed Regions [%, No.]	100.0, 3,038
Residues in Favoured Regions [%, No.]	96.0, 2,915
MOLPROBITY Clashscore[e]	0.64

[a] Data for the highest resolution shell in parenthesis
[b] calculated with PHENIX (Adams *et al.*, 2010)
[c] diffraction-component precision index (Cruickshank, 1999)
[d] calculated with MOLPROBITY (Chen *et al.*, 2010)
[e] Clashscore is the number of serious steric overlaps (> 0.4) per 1,000 atoms.

Adams, P.D., P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.-W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger & P.H. Zwart, (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D-Biological Crystallography* **66**: 213-221.

Chen, V.B., W.B. Arendall, 3rd, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson & D.C. Richardson, (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D* **66**: 12-21.

Cruickshank, D.W.J., (1999) Remarks about protein structure precision. *Acta Crystallographica Section D-Biological Crystallography* **55**: 583-601.

Table S3:
^{1}H and ^{13}C NMR chemical shift assignments (ppm) for a *Shigella flexneri* O-serogroup Y octasaccharide / Additional NMR experiments

^{1}H and ^{13}C NMR chemical shift assignments in ppm for the α-anomeric form of a *S. flexneri* O-serogroup Y octasaccharide cleavage product from Sf6TSP in D_2O solution at 56 °C using internal sodium 3-trimethylsilyl-(2,2,3,3-$^{2}H_4$)-propanoate (TSP, δ_H 0.00) or external 1,4-dioxane in D_2O (δ_C 67.40) as references.

Residue	Nuclei	H1	H2	H3	H4	H5	Me6/H6a	H6b	NAc
RAM1	^{1}H	4.90	3.83	3.75	3.44	3.98	1.26		
	^{13}C	101.98	71.56	71.17	72.83	69.84	17.29		
NAG2	^{1}H	4.77	3.84	3.66	3.54	3.47	3.90	3.77	2.08
	^{13}C	102.81	56.46	82.24	69.35	76.75	61.62	61.69	23.11
RAM3	^{1}H	5.16	4.15	3.88	3.35	3.73	1.26		
	^{13}C	101.71	79.53	70.67	73.23	70.02	17.46		
RAM4	^{1}H	5.16	4.07	3.92	3.48	3.74	1.32		
	^{13}C	101.71	78.94	70.79	73.17	69.98	17.48		
RAM5	^{1}H	4.88	3.87	3.80	3.56	4.02	1.26		
	^{13}C	102.01	71.48	78.26	72.47	69.9	17.22		
NAG6	^{1}H	4.74	3.84	3.66	3.54	3.47	3.90	3.77	2.07
	^{13}C	102.91	56.46	82.24	69.35	76.75	61.62	61.69	23.12
RAM7	^{1}H	5.16	4.15	3.88	3.35	3.73	1.26		
	^{13}C	101.71	79.53	70.67	73.23	70.02	17.46		
RAM8	^{1}H	5.21	3.92	3.90	3.46	3.86	1.28		
	^{13}C	93.49	80.09	70.65	73.04	69.30	17.68		

The translational diffusion coefficient of the octasaccharide was measured with four pulsed field gradient (PFG) ^{1}H NMR experiments on a sample containing the octasaccharide (1.9 mM) in D_2O (0.55 mL). The experiments were performed at 25 °C on a 600 MHz Bruker Avance III NMR spectrometer equipped with a TXI (^{1}H/^{13}C/^{31}P) probe, where the Z-gradient had been calibrated to compensate for gradient inhomogeneities by using a Gadolinium doped water sample (1 % H_2O in D_2O + 1 mg mL^{-1} $GdCl_3$) and a literature value of D_t = 1.90 × 10^{-9} m^2 s^{-1} for the HDO resonance (Mills, 1973). The diffusion time delay (Δ) was set to 150 ms and the gradient pulse length (δ) was set to 4.0 ms; each experiment was acquired with 32 data points with gradient strengths going from 5 % to 95 % of the maximum (55.79 G cm^{-1}). The decay of the resonances from the sugar bulk region (3.0 - 4.5 ppm) was used to calculate the diffusion

coefficient of the free ligand to $2.31 \cdot 10^{-6}$ $cm^2 \cdot s^{-1}$ in D_2O at 25 °C by fitting a Stejskal-Tanner type equation to the data (Stejskal & Tanner, 1965).

The ^1H and ^{13}C chemical shifts of the α-anomeric form of the octasaccharide were assigned using CASPER (Rönnols *et al.*, 2013) and standard NMR $^1H,^1H$-COSY, $^1H,^1H$-NOESY, $^1H,^{13}C$-HSQC, $^1H,^{13}C$-HMBC, $^1H,^{13}C$-HSQC-TOCSY and $^1H,^{13}C$-H2BC experiments employed for carbohydrates (Widmalm, 2007). The anomeric equilibrium at the reducing end was determined to 82 % in favor of the α-diastereomer in solution obtained by integration of the RAM8-H1α and RAM7-H1β resonances. Longitudinal relaxation times (T_1) of the octasaccharide's protons were measured with ten inversion recovery experiments with relaxation times varying between 0.01 s to 4 s. The spectra were acquired with 32k data points using 16 scans, a spectral width of 16 ppm and a recycle delay of 8 s. The spectra were zero-filled to 128k data points and a line broadening window function of 0.3 Hz was applied prior to Fourier Transformation. T_1 relaxation times were calculated using the nonlinear fitting routines provided by the instrument manufacturer.

Mills, R., (1973) Self-diffusion in normal and heavy-water in range 1-45°. *Journal of Physical Chemistry* **77**: 685-688.

Rönnols, J., R. Pendrill, C. Fontana, C. Hamark, T.A. d'Ortoli, O. Engström, J. Ståhle, M.V. Zaccheus, E. Säwén, L.E. Hahn, S. Iqbal & G. Widmalm, (2013) Complete ^1H and ^{13}C NMR chemical shift assignments of mono- to tetrasaccharides as basis for NMR chemical shift predictions of oligosaccharides using the computer program CASPER. *Carbohydr. Res.* **380**: 156-166.

Stejskal, E.O. & J.E. Tanner, (1965) Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *J. Chem. Phys.* **42**: 288-292.

Widmalm, G., (2007) General NMR Spectroscopy of Carbohydrates and Conformational Analysis in Solution. In: Comprehensive Glycoscience. J.P. Kamerling (ed). Oxford: Elsevier, pp. 101-110.

Table S4: STD-AF_0 data in the Sf6TSP octasaccharide complex

STD-AF_0 data in the Sf6TSP octasaccharide complex for two irradiation frequencies, 7.0 and −0.4 ppm, of Sf6TSP (0.12 mM) and the octasaccharide (1.87 mM) in D_2O at 56 °C and a spectrometer frequency of 500 MHz. The build-up curves that were used to calculate the STD-AF_0 have been fitted to STD-AF from ^{1}H STD NMR experiments of four different mixing times (0.5, 1, 2 and 4 s).

^{1}H resonance	T_1 /s	STD-AF_0 /s^{-1}		relative STD-AF_0 /%	
		irr. at 7.0 ppm	irr. at −0.4 ppm	irr. at 7.0 ppm	irr. at −0.4 ppm
RAM1-H1	1.40	0.19	0.11	44	31
RAM1-H5	1.62	0.23	0.17	53	49
NAG2-H1	0.80	0.25	0.13	58	38
NAG2-Ac	1.08	0.32	0.26	74	77
RAM4-H2	1.00	0.16	0.11	37	31
RAM4-H6	0.71	0.43	0.31	100	92
RAM5-H1*	1.14	0.21	0.12	48	36
RAM5-H5*	1.26	0.16	0.11	37	33
NAG6-H1	0.75	0.25	0.17	59	51
NAG6-Ac	1.04	0.37	0.34	87	100
RAM8-H1	1.88	0.23	0.17	55	51
RAM8-H6	0.83	0.23	0.19	53	55
NAG2/6-H3	0.97	0.15	0.12	34	34
RAM3/7-H2	0.99	0.32	0.22	76	64
RAM3/7-H4*	1.73	0.36	0.29	85	86
RAM4/7/8-H1	1.02	0.20	0.14	47	40
NAG2/6-H4; RAM5-H4*	1.45	0.14	0.11	33	32
RAM1/3/5/7-H6	0.70	0.23	0.21	55	62
RAM1/4/8-H4; NAG2/6-H5*	1.32	0.25	0.19	58	57
RAM1/5-H3; RAM3/7/8-H5; NAG2/6-H6b*	0.88	0.19	0.15	45	43
RAM1/5/8-H2; NAG2/6-H2; RAM3/4/7/8-H3; RAM8-H5; NAG2/6-H6a*	1.18	0.23	0.17	54	51

*Overlapping with resonances of the β-anomeric form of the octasaccharide.

Table S5: Glycosidic torsion angles obtained from molecular dynamics simulations

For tetrasaccharides and octasaccharides the torsions were compared with the torsions in the crystal structure of the complex (this work and (Muller *et al.*, 2008)). For nomenclature of linkages refer to Scheme 1 in the main document.

Tetrasaccharide			
linkage		**MD**[a]	**pdb: 2vbm**
III	phi	273.7(7.9)	277.3
	psi	128.3(8.6)	128.2
IV	phi	286.9(8.5)	304.4
	psi	266.5(7.9)	249.4
I	phi	291.4(6.5)	280.7
	psi	229.3(12.9)	242.7

[a] Calculated with Glycam. Standard deviation in parentheses

Octasaccharide									
linkage		**Glycam**[a]	**CHARMM**[a]	**pdb: 4urr**[b]					
				A	**B**	**C**	**D**	**E**	**F**
III'	phi	276.0(10.2)	270.5(8.1)	276.3	276.5	278.3	277.4	277.5	276.8
	psi	122.8(9.7)	119.0(10.5)	124.5	125.3	124.3	123.5	123.8	124.5
IV'	phi	283.7(8.8)	286.2(10.8)	286.7	286.6	287.2	286.9	287.3	287.7
	psi	270.0(8.3)	251.1(14.8)	256.6	259.1	256.9	257.8	257.1	257.6
I'	phi	291.0(6.9)	287.1(9.6)	299.4	301.4	300.3	300.0	299.3	298.3
	psi	223.6(10.4)	244.0(13.5)	229.0	228.5	230.2	228.9	229.2	231.4
II	phi	289.8(6.9)	283.7(8.4)	298.8	299.6	297.0	298.9	298.3	299.8
	psi	233.2(11.6)	250.8(11.1)	228.5	228.8	228.5	228.0	228.0	227.4
III	phi	290.7(6.7)	281.1(9.4)	287.9	290.8	287.2	287.4	287.1	286.4
	psi	140.7(8.7)	126.7(18.8)	141.8	144.9	142.3	142.6	143.3	142.8
IV	phi	268.6(8.8)	274.0(13.6)	288.8	286.9	289.7	289.4	290.7	287.4
	psi	268.9(6.7)	247.4(24.3)	254.7	244.8	253.9	254.4	253.0	256.5
I	phi	299.2(7.2)	285.1(12.0)	304.6	294.3	297.7	295.1	293.2	298.1
	psi	249.0(12.1)	236.3(30.2)	171.4	156.4	154.5	155.7	156.8	208.6

[a] Standard deviation in parentheses, [b] A-F refer to individual glycan chains in the asymmetric unit

Table S5 (continued): Glycosidic torsion angles obtained from molecular dynamics simulations

Dodecasaccharide					
Linkage		**GLYCAM(R)[a,b]**	**CHARMM(R)[a,b]**	**GLYCAM(F)[a,c]**	**CHARMM(F)[a,c]**
III'	phi	259.0(29.1)	273.9(13.0)	286.5(18.8)	267.5(9.4)
	psi	122.6(18.5)	106.4(28.3)	137.1(20.3)	119.6(10.8)
IV'	phi	273.4(10.9)	283.2(18.7)	283.9(11.6)	293.5(12.5)
	psi	240.9(23.1)	227.4(25.5)	264.0(15.8)	237.4(16.4)
I'	phi	288.2(10.8)	283.4(11.8)	281.5(10.8)	287.2(8.7)
	psi	242.1(29.9)	243.2(24.2)	218.4(20.9)	251.6(10.4)
II	phi	287.8(11.2)	284.5(11.6)	293.5(8.5)	277.7(10.8)
	psi	256.3(16.3)	251.5(15.2)	258.2(14.6)	252.7(17.5)
III	phi	291.3(12.9)	271.3(12.0)	278.2(9.1)	280.2(15.3)
	psi	139.5(20.3)	104.6(47.2)	134.8(14.1)	134.7(58.5)
IV	phi	271.4(14.1)	281.4(14.1)	275.0(13.1)	280.3(14.5)
	psi	223.3(27.3)	228.2(28.2)	210.3(20.1)	220.0(24.8)
I	phi	286.7(12.0)	284.4(12.4)	287.1(13.5)	280.3(12.9)
	psi	230.2(24.6)	237.0(28.1)	236.3(17.5)	245.2(23.2)
II''	phi	280.3(12.7)	284.2(11.6)	289.7(10.3)	288.4(11.3)
	psi	256.1(19.6)	246.4(30.7)	274.2(13.2)	251.7(16.0)
III''	phi	287.7(11.2)	275.4(13.2)	288.7(7.3)	275.5(13.9)
	psi	126.4(22.5)	104.8(28.1)	139.9(8.9)	111.1(23.9)
IV''	phi	282.8(11.6)	282.2(13.4)	261.8(8.1)	279.0(14.7)
	psi	256.9(19.2)	231.2(26.9)	265.2(9.5)	226.4(27.7)
I''	phi	283.6(7.9)	280.8(11.1)	284.5(7.5)	279.6(10.7)
	psi	209.3(11.9)	227.3(28.6)	213.1(9.3)	224.1(29.5)

[a] Standard deviation in parentheses, [b] restrained backbone (R), [c] flexible backbone (F)

Muller, J.J., S. Barbirz, K. Heinle, A. Freiberg, R. Seckler & U. Heinemann, (2008) An intersubunit active site between supercoiled parallel beta helices in the trimeric tailspike endorhamnosidase of *Shigella flexneri* Phage Sf6. *Structure* **16**: 766-775.

Table S6: Hydrogen bonds in Sf6TSP – oligosaccharide complexes

First repeat unit (RU1)								
Donor	**Acceptor**	**distances octasaccharide from MD[a] / Å**		**distances dodecasaccharide from MD (rigid)[a] / Å**		**distances dodecasaccharide from MD (flexible)[a] / Å**		**X-ray distance[b] / Å**
		AMBER03/Glycam[c]	CHARMM36	AMBER03/Glycam[c]	CHARMM36	AMBER03/Glycam[c]	CHARMM36	
ARG230-NH1	RAM1-O2	2.92	3.01	7.82	19.16	2.93	3.02	2.9
ARG230-NH2	RAM1-O2	3.46	3.22	6.89	19.00	3.65	3.08	3.0
ARG230 -NH2	RAM1-O5	3.49	3.36	6.31	18.26	3.49	3.61	3.3
ARG230 -NH1	RAM1-O3	4.25	4.44	6.87	18.65	4.36	4.27	4.0
RAM1-O2	GLU293 -OE2	3.44	3.00	8.91	15.99	3.86	3.16	2.7
	GLU293 -OE1	3.46	3.17	8.90	16.04	3.94	2.95	3.5
ARG257 -NH2	NAG2-O4	7.26	5.5	8.29	16.27	7.56	5.61	5.3
NAG2-O6	ASP247-OD1	7.52	4.4	10.62	14.54	7.64	5.55	4.2
	ASP247-OD2	7.47	4.61	10.64	14.70	7.61	5.72	4.1
RAM3 -O3	ASP247-OD1	3.46	6.92	6.46	17.27	4.53	6.94	3.9
	ASP247-OD2	4.23	7.38	6.51	17.49	4.38	7.37	5.9
RAM3-O4	THR248 -OG1	3.31	5.36	3.26	15.46	4.03	6.58	4.8
	SER246 -O	6.51	8.17	7.04	14.16	6.83	8.81	3.7
THR248 -OG1	RAM3 -O4	3.31	5.36	3.26	15.46	4.03	6.58	4.8
RAM3-O3	THR248 -OG1	4.1	3.88	3.57	14.55	4.79	4.26	3.0
	SER246 -O	5.01	6.33	5.13	13.69	5.40	7.06	2.6

[a] calculated for Sf6TSP wildtype, [b] X-ray model obtained for Sf6TSP D399A/E366A

[c] For each simulation run GLYCAM06h was used for the carbohydrates, and Amber ff03 for the protein part.

Hydrogen bonds were defined as donor-acceptor distances less than 3.5 Å and hydrogen-donor-acceptor angles less than 30 °. In the table all donor-acceptor pairs are listed that complied with this condition at least for one frame and for one simulation setup (Glycam/Amber or CHARMM), so pairs with very low occupancy are represented with relatively large average distances (e.g., > 7 Å). Distances smaller than 4 Å represent hydrogen bridges with notable occupancies (i.e., > 10 %).

Table S6: Hydrogen bonds in Sf6TSP – oligosaccharide complexes (contd)

Second repeat unit (RU2)								
Donor	**Acceptor**	**distances octasaccharide from MD[a] / Å**		**distances dodecasaccharide from MD (rigid)[a] / Å**		**distances dodecasaccharide from MD (flexible)[a] / Å**		**X-ray distance[c] / Å**
		AMBER03/Glycam[c]	CHARMM36	AMBER03/Glycam[c]	CHARMM36	AMBER03/Glycam[c]	CHARMM36	
ARG338-NH1	RAM5-O2	14.82	15.08	7.61	18.71	14.74	16.34	
GLN280-NE2	NAG6-O2N	3.27	4.07	10.05	11.99	3.77	7.79	3.1
NAG6-O6	TYR400-OH	5.31	5.55	6.36	17.02	4.99	6.62	
	ASP399-OD2	6.41	9.18	6.11	21.36	7.71	11.70	
NAG6-N2	TYR400-OH	8.98	8.34	5.70	13.74	8.69	8.40	
RAM7-O4	GLN325-NE2	4.43	5.60	10.85	14.24	3.57	7.07	
RAM8 -O3	ASP399 -OD1	4.15	7.29	7.59	22.62	8.90	11.32	
	ASP399 -OD2	4.06	7.52	7.77	22.21	8.96	11.37	
RAM8-O1	ASP399 -OD1	6.76	6.95					
RAM8-O4	ASP399 -OD1	5.27	8.61	9.49	24.07	10.51	12.69	
	ASP399 -OD2	5.39	8.77	9.69	23.62	10.46	12.78	
RAM8-O1	GLU366 -OE1	3.63	4.97					
	GLU366 -OE2	3.00	4.92					
RAM8-O3	GLU366 -OE1	6.88	7.54	9.71	20.25	9.41	11.38	
	GLU366 -OE2	5.71	7.54	9.44	20.06	9.50	11.32	
TYR400-OH	RAM8-O3	8.12	4.43	6.30	17.87	5.15	6.20	2.9
GLN325-NE2	RAM8-O5	7.32	7.43	10.87	17.28	7.63	10.01	

[a] calculated for Sf6TSP wildtype, [b] X-ray model obtained for Sf6TSP D399A/E366A

[c]For each simulation run GLYCAM06h was used for the carbohydrates, and Amber ff03 for the protein part.

Hydrogen bonds were defined as donor-acceptor distances less than 3.5 Å and hydrogen-donor-acceptor angles less than 30 °. In the table all donor-acceptor pairs are listed that complied with this condition at least for one frame and for one simulation setup (Glycam/Amber or CHARMM), so pairs with very low occupancy are represented with relatively large average distances (e.g., > 7 Å). Distances smaller than 4 Å represent hydrogen bridges with notable occupancies (i.e., > 10 %).

Table S6: Hydrogen bonds in Sf6TSP – oligosaccharide complexes (contd)

Third repeat unit (RU3)								
Donor	**Acceptor**	**distances octasaccharide from MD[a] / Å**		**distances dodecasaccharide from MD (rigid)[a] / Å**		**distances dodecasaccharide from MD (flexible)[a] / Å**		**X-ray distance[c] / Å**
		AMBER03/Glycam[c]	CHARMM36	AMBER03/Glycam[c]	CHARMM36	AMBER03/Glycam[c]	CHARMM36	
RAM9-O4	GLU366-OE1	n/a	n/a	8.43	20.02	3.07	8.57	n/a
	GLU366-OE2	n/a	n/a	8.26	19.87	3.12	8.63	n/a
ARG364-NH2	RAM9-O5	n/a	n/a	6.75	21.99	3.47	10.70	n/a
	RAM9-O4	n/a	n/a	7.88	20.95	4.18	11.13	n/a
ARG364-NH2	NAG10-O4	n/a	n/a	5.36	21.78	3.15	9.33	n/a
ARG364-NH1		n/a	n/a	4.77	23.19	3.24	11.36	n/a
ASN412-ND2	NAG10-O6	n/a	n/a	4.22	24.67	4.58	10.06	n/a
RAM11-O4	ASN455-OD1	n/a	n/a	3.90	25.63	2.81	11.93	n/a
RAM11-O3		n/a	n/a	3.77	25.96	2.87	10.63	n/a
HIS453-NE2	RAM11-O3	n/a	n/a	5.61	27.83	3.87	12.88	n/a
THR443-OG1	RAM12-O3	n/a	n/a	5.59	24.09	5.01	7.88	n/a
RAM12-O1	ASN508-OD1	n/a	n/a	5.58	27.25	4.42	12.57	n/a
RAM12-O4	SER445-OG	n/a	n/a	5.03	28.49	4.82	13.41	n/a
RAM12-O4	GLN444-O	n/a	n/a	2.76	27.89	3.22	11.48	n/a
RAM12-O3		n/a	n/a	2.78	27.63	3.49	11.46	n/a
ASN508-ND2	RAM12-O5	n/a	n/a	3.93	28.18	4.31	13.15	n/a

[a] calculated for Sf6TSP wildtype, [b] X-ray model obtained for Sf6TSP D399A/E366A

[c]For each simulation run GLYCAM06h was used for the carbohydrates, and Amber ff03 for the protein part.

Hydrogen bonds were defined as donor-acceptor distances less than 3.5 Å and hydrogen-donor-acceptor angles less than 30 °. In the table all donor-acceptor pairs are listed that complied with this condition at least for one frame and for one simulation setup (Glycam/Amber or CHARMM), so pairs with very low occupancy are represented with relatively large average distances (e.g., > 7 Å). Distances smaller than 4 Å represent hydrogen bridges with notable occupancies (i.e., > 10 %).