Supplementary Information for

Time-resolved DNA release from an O-antigen specific *Salmonella* bacteriophage with a contractile tail

Nina K. Broeker, Yvette Roske, Angelo Valleriani, Mareike S. Stephan, Dorothee Andres, Joachim Koetz, Udo Heinemann and Stefanie Barbirz

Stefanie Barbirz
Email: barbirz@uni-potsdam.de

This PDF file includes:

Supplementary methods
Figs. S1 to S11
Tables S1 to S4
References for SI reference citations

Supplementary Methods

Fitting procedures for kinetic traces of DNA release from myovirus Det7.

Data set

The full data set obtained for DNA ejection kinetics of bacteriophage Det7 triggered by *S.* Typhimurium LPS at different temperatures is given in Figure S7.

Models and equations

We have implemented a kinetic analysis of each time course of the signal intensity by using three different multi-step chemical kinetics processes:

1) $A \rightarrow B \rightarrow C$ (with rates k_1 and k_2 for the first and second transition)
2) $A \rightarrow B \rightarrow C \rightarrow D$ (with rates k_1, k_2, and k_3)
3) $A \leftrightarrow B \rightarrow C$ (with rates k_+, k_-, and k_2)

We compared the models by using the Akaike Information Criterion (AIC) to weigh the quality of the fit with the number of parameters. For example, models 2 and 3 have three rate parameters whereas model 1 has only two. AIC then can assess whether a higher number of parameters and a more complex model are justified by an increased precision found for the fit. A first analysis showed that the AIC supported the use of model 2 at higher temperatures and of model 1 at lower temperatures. Model 3 resulted in poorer fittings for all curves compared to the other two models. For this reason, we will not discuss model 3 any longer in this work. The following equations were used:

Model 1

$$C(t) = A_0 \left(1 - \frac{k_2 e^{-k_1 t}}{k_2 - k_1} + \frac{k_1 e^{-k_2 t}}{k_2 - k_1} \right)$$

Model 2

$$D(t) = A_0 \left(1 - \frac{k_2 k_3 (k_3 - k_2) e^{-k_1 t} - k_1 k_3 (k_3 - k_1) e^{-k_2 t} + k_1 k_2 (k_2 - k_1) e^{-k_3 t}}{(k_2 - k_1)(k_3 - k_1)(k_3 - k_2)} \right)$$

Adjusting the initial time

As can be seen from Figure S7, at the beginning of the measurements the signal intensity becomes negative for some little amount of time. This negative signal is an artifact of the subtraction of the background noise caused by binding of small amounts of the fluorescent dye to the phage particles. To fit the data to the above models, we therefore set the first time point with a positive signal as time $t = 0$ in the fitting algorithm.

Fitting algorithms

The fitting procedure has been done entirely under MatLab by using the routing *lsqnonlin* which gives as output the best values of the parameters, the residuals (MSE), and the Jacobian of the objective function with respect to the parameters computed at the found minimum. The Jacobian, the values of the parameters and the residuals are used by the MatLab routine *nlparci* to compute the 95% confidence intervals of the parameters. These routines have been implemented in a custom designed MatLab script available under request. The residuals were also used in the AIC formula

$$AIC = 2^n + m \cdot \log(\text{MSE})$$

where n is the number of parameters, and m the size of the dataset (*i.e.* number of measurements) used for the fit. The AIC's have been then used to estimate the probabilities $\pi_i \propto \exp(-\frac{AIC_i}{AIC_{min}})$ for model i to be the model that cross-validates the data best.

Covariances and confidence intervals

Both models 1 and 2 have some general characteristics that result important for the choice of the final model. First, model 2 can reduce to model 1 by just taking any of the rates in model 2 to a very large value (say, infinity). This means that the solution provided by model 1 is most likely a local minimum also for the fitting with model 2. Second, in each model the rates can be swapped without changing the fitting quality. This can be seen also by noticing that the equations are symmetric with respect to the two and three rates for model 1 and 2, respectively. Third, given the data, the rates turn out to covariate. This results in more or less large confidence intervals.

It turned out that the covariance of the rates in model 2 is very large and so leads to meaningless and very large confidence intervals. This happens also for those experimental curves for which the AIC had indicated model 2 as the favored one. We have tried to overcome this problem implementing a Machine Learning (ML) approach to create a continuous regression line for each experiment as a polynomial function of the time and of the previous values of the curve. The aim was to create a continuous valued curve from which we could extract a large number of data points to improve the fit and reduce the covariance of the parameters. All various attempts led to a strong overfitting of the curve so that it did generalize very badly to time points between measurements. Adjusting with a regularization term did not help either.

Turning back to the fitting of the experimental results, we realized that the fitting with model 1 was in general differing by the fitting with model 2 by a few decimal points after the commas, thus not justifying taking a more complex model after all. We also found out that thanks to a reduced number of parameters, the covariances of the rates were small enough to allow for 95% confidence intervals (CI) for all measured curves except for the curves measured at 29, 30 and 31 degrees centigrade (see Table S4).

For each fitted curve there is one larger and one smaller rate. We organized the data so that the smaller rate is always k_1. Plotting k_1 and k_2 as function of the temperature in degree centigrade shows that their values for 29, 30, 31 ° may be outliers. At temperatures below 31 °C the number of ejecting particles is 20 % or lower. We assume that both the phage particle or the LPS receptor have properties altered with lower temperatures resulting in different kinetics not comparable with those found at higher temperatures. We thus excluded from the further analysis the data at temperatures below or equal to 31 degrees centigrade.

Activation energies

For both rates as measured for temperatures larger or equal to 31.5 degrees centigrade, we have fit the linear function

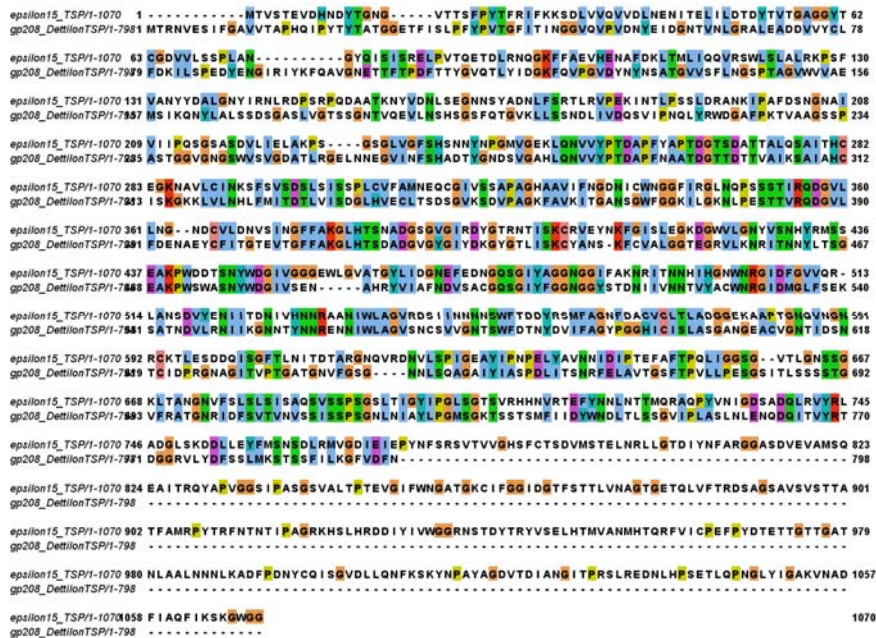$$\ln\left(\frac{k}{k_0}\right) = -\frac{E_A}{k_B T}$$

to obtain the respective activation energies (see Figure 4C). The energies are expressed in units of $k_B T = 4.116\ pN\ nm$). It results: $\ln(k_0^{(1)}) = 19.5$, $E_A^{(1)} = 27.7\ k_B T$; $\ln(k_0^{(2)}) = 31.8$, $E_A^{(2)} = 38.7\ k_B T$.

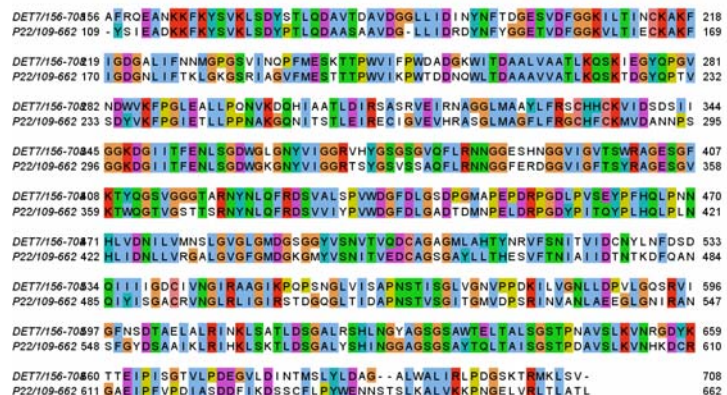**Fig. S1**: **Alignments of tailspike sequences in myovirus Det7.**
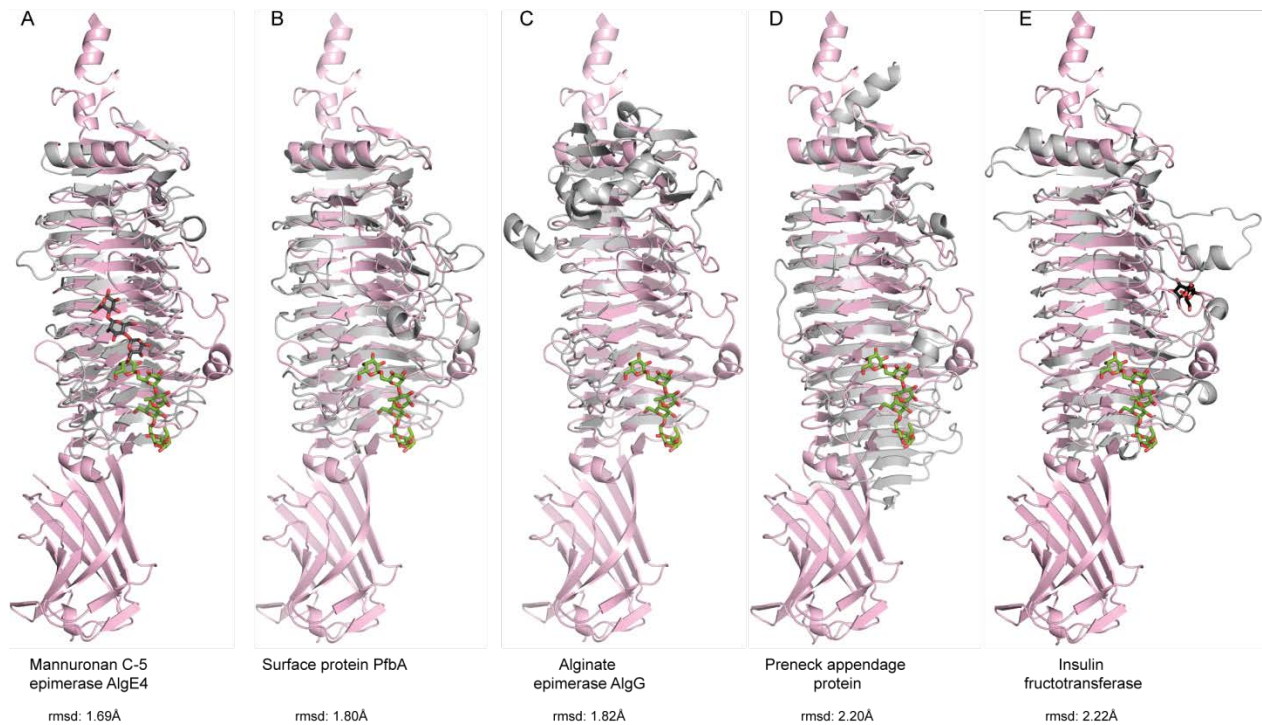
A
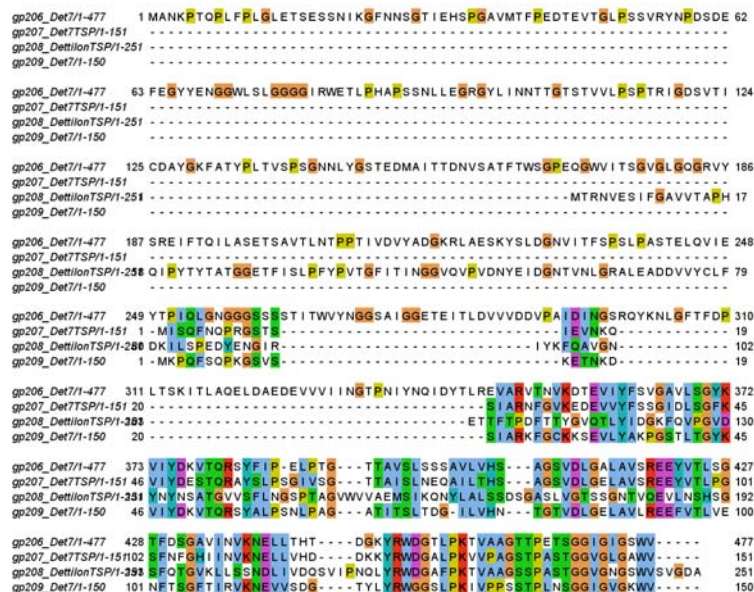
B

C

ClustalW (1) amino acid sequence alignments (default settings). **A.** N-terminal part of Det7 tailspike proteins (gp206-gp209). **B.** Det7 gp 208 (DettilonTSP) with the tailspike of *Salmonella* phage epsilon 15. **C.** Receptor binding domains of Det7 gp207 (Det7TSP) with podovirus gp9 (P22TSP). Figures were prepared with Jalview (2).

5

**Fig. S2. Topology of the DettilonTSPΔN C-terminal domain**
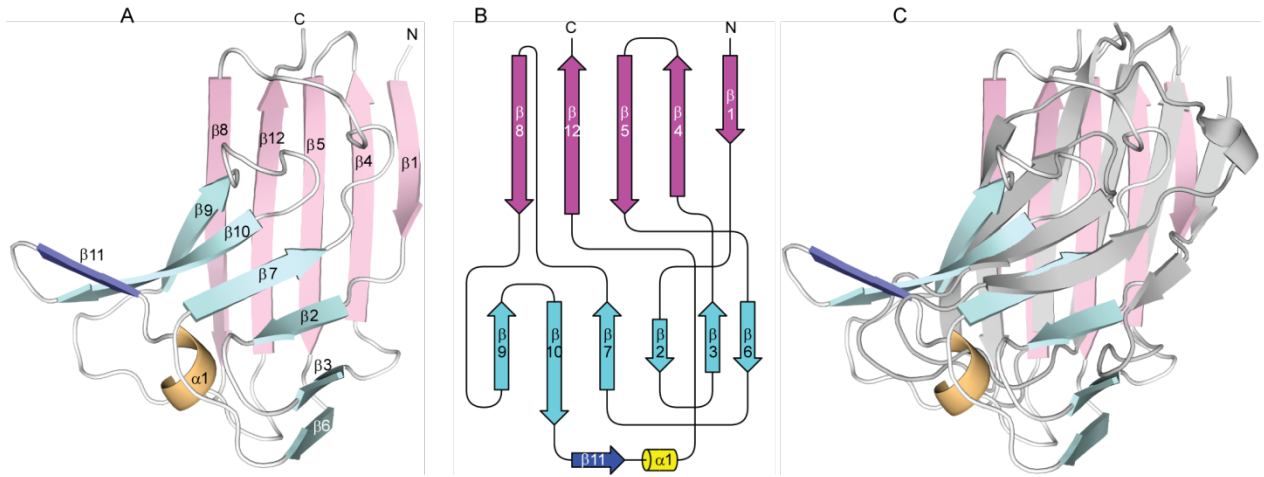
β-sandwich structure of the C-terminal DettilonTSPΔN domain. **A.** Detailed view of the β-sandwich structure with one five stranded β-sheet shown in magenta and the second 6 stranded β-sheet depicted in cyan. The additional β-strand #11 is colored blue and the short α1 helix in orange. **B.** Topology diagram of the C-terminal DettilonTSPΔN domain with colors as shown in A. **C.** Superposition of the structurally most similar protein, the receptor-binding protein head domain from *Lactococcus lactis* Phage bIL170 (PDB: 2FSD), shown in gray onto the Dettilon ΔN C-terminal domain with colors as used in A. Both structures match with an C-alpha rmsd of 3.9Å with 110 aligned residues.

Fig. S3. Structural homologs of DettilonTSPΔN

| Mannuronan C-5 epimerase AlgE4	Surface protein PfbA	Alginate epimerase AlgG	Preneck appendage protein	Insulin fructotransferase
rmsd: 1.69Å	rmsd: 1.80Å	rmsd: 1.82Å	rmsd: 2.20Å	rmsd: 2.22Å

Superimposition of structurally similar proteins onto the DettilonTSPΔN monomer structure found by similarity search with the alignment tools Dali and PDBeFold (3, 4). The protein backbone C-alpha atoms were used for superimposition with the program COOT (5). The DettilonTSPΔN structure is shown in magenta and the aligned protein structures in gray. The bound hexasaccharide in DettilonTSPΔN is presented as green stick model and the saccharides bound to the aligned proteins are shown as black sticks. **A.** Highest structural similarity to DettilonTSPΔN was found in the mannuronan C-5 epimerase AlgE4 of *Azotobacter* bound to mannuronan trisaccharide (PDB: 2PYH) (rmsd = 1.69Å over 314 residues). **B.** Plasmin- and fibronectin-binding protein A of *Streptococcus pneumoniae* (PDB: 3ZPP) (rmsd = 1.80Å over 310 residues). **C.** Alginate epimerase AlgG of *Pseudomonas aeruginosa* (PDB: 4NK6) (rmsd = 1.82Å over 254 residues). **D.** Teichoic acid binding tailspike gp12* from Bacillus subtilis phage φ29 (PDB: 3SUC) (rmsd = 2.20Å over 306 residues). **E.** The inulin fructotransferase (PDB: 2INV) of *Bacillus sp. snu-7* with bound difructose (rmsd = 2.22Å over 307 residues).

Fig. S4. Comparison of hexasaccharide-bound and ligand-free DettilonTSPΔN
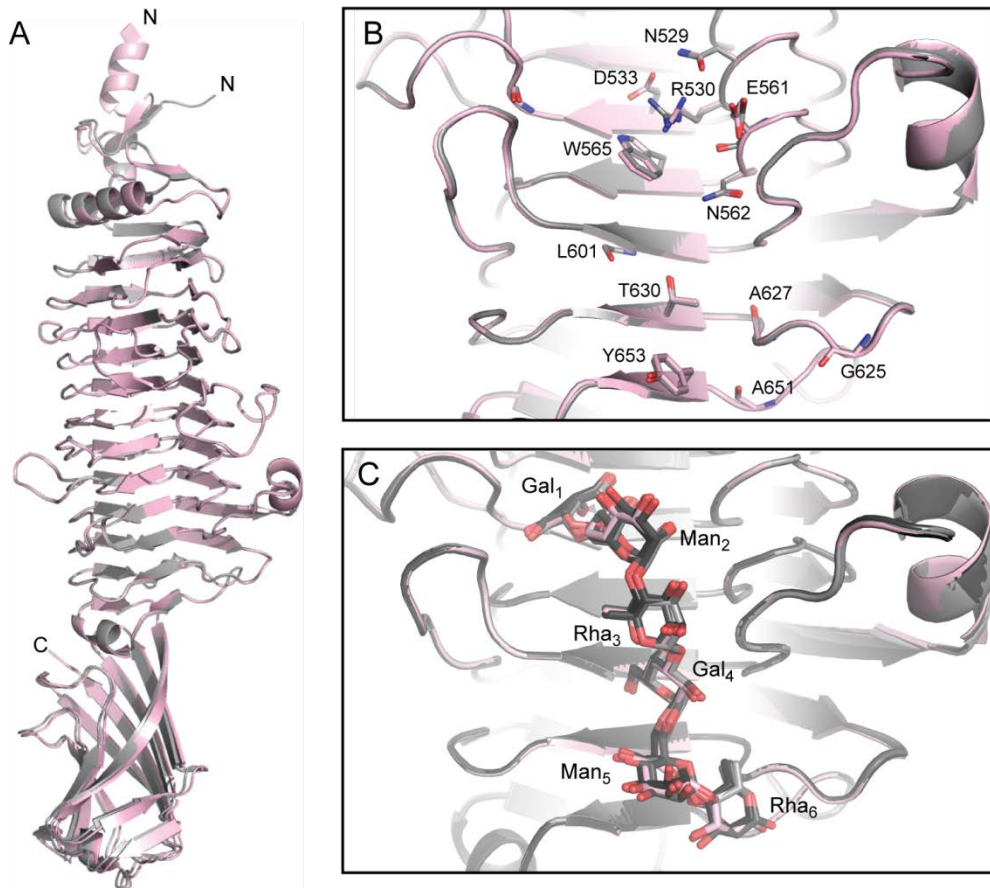

Superimposition of the DettilonTSPΔN structure in complex with the hexasaccharide (magenta) and the ligand-free structure (gray). **A.** The alignment of 536 residues results in an rmsd for Cα atoms of 0.355Å. **B.** Residues involved in binding of the hexasaccharide are depicted in stick presentation, showing the rigid binding pocket of DettilonTSPΔN. **C.** Superimposition of the carbohydrate binding pocket including the bound hexasaccharide of all six DettilonTSPΔN molecules forming two trimers in the asymmetric unit. The binding pocket of chain F with its corresponding hexasaccharide is shown in magenta and all other five DettilonTSPΔN chains with their bound carbohydrate chains are depicted in different grey levels.

Fig. S5. Fo-Fc difference density map of hexasaccharide ligand in the DettilonTSPΔN binding site
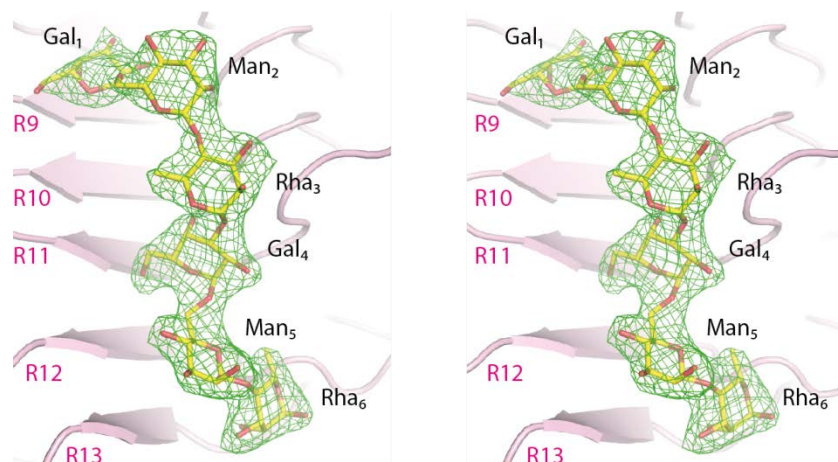

Stereo view of the α-D-Galactose-(1,6)-β-D-Mannose-(1,4)-α-L-Rhamnose-(1,3)-α-D-Galactose-(1,6)-β-D-Mannose-(1,4)-α-L-Rhamnose-(1,3) hexasaccharide binding pocket showing the Fo-Fc difference density map (contoured at 3σ in green colour) of the hexasaccharide (shown in yellow stick representation). The rungs of the β-helix involved hexasaccharide binding are numbered (R9-R13).

Fig. S6. Surface plasmon resonance signals upon binding of *S.* Anatum hexasaccharides to DettilonTSP
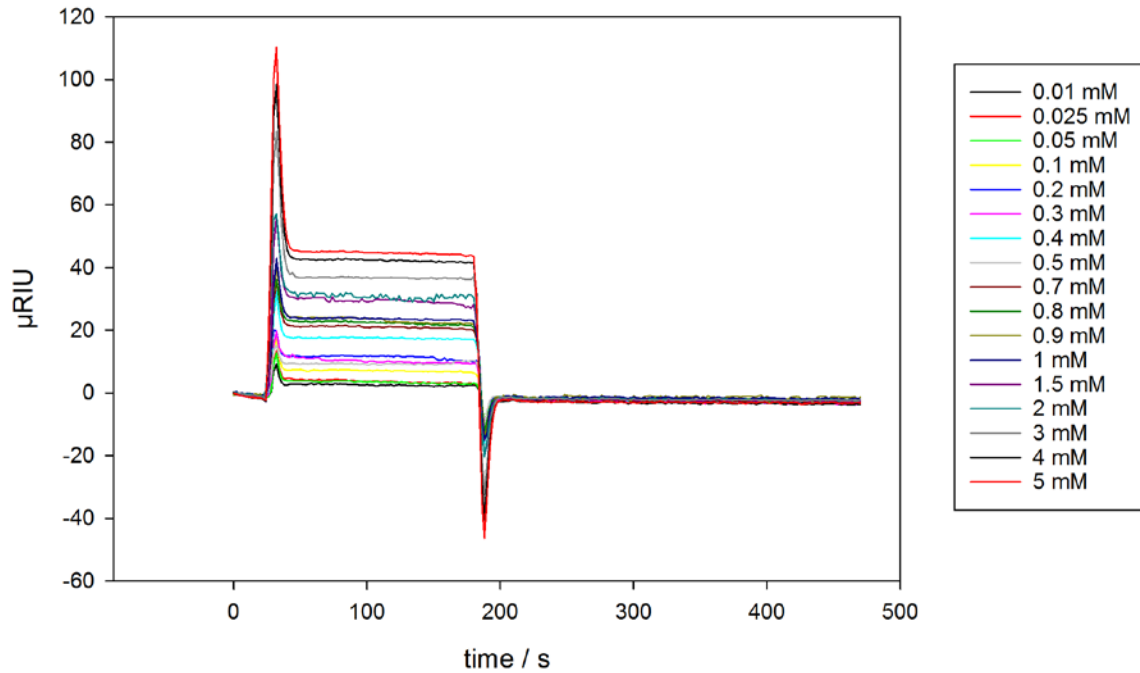
Overlay of buffer-corrected time courses of surface plasmon resonance signals upon hexasaccharide injections at different concentrations over a carboxymethyldextran surface with DettilonTSP immobilized. The hexasaccharide corresponds to two repeat units (RU) of the S. Anatum O-polysaccharide. Mean values of equilibrium signals were used to construct the binding isotherm given in Figure 3C. Signals were reference corrected with an empty carboxymethyldextran surface with no protein immobilized.

Fig. S7. *In vitro* DNA ejection from phage Det7 triggered by different lipopolysaccharide building blocks
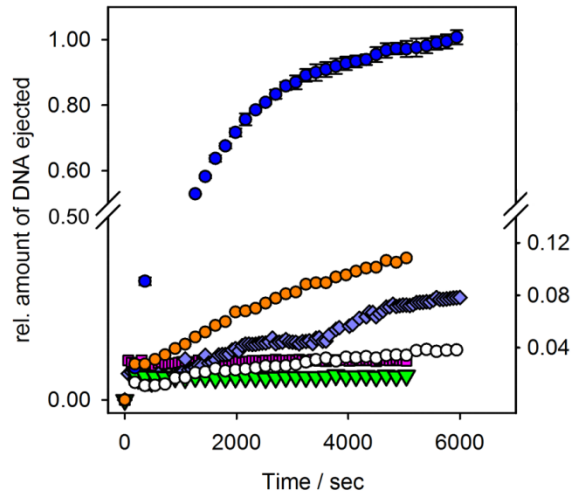
Above: $4 \cdot 10^9$ Det7 phage particles were mixed with 50 µg ml^{-1} LPS in presence of a fluorescent DNA-binding dye. Fluorescence signal obtained at 37 °C with *S.* Typhimurium LPS (blue circles), with 50 µg purified *S.* Typhimurium O-polysaccharide (light-blue diamonds), with a mixture of 50 µg purified *S.* Typhimurium O-polysaccharide and 50 µg purified *S.* Typhimurium lipid A (orange circles), with smooth LPS from *E. coli* H TD2158 (6) (green triangles), with *S.* Typhimurium LPS lacking O-antigen (white circles), 50 µg purified *S.* Typhimurium lipid A (pink squares). **Below**: Same experiment with LPS components from *S.* Anatum. Addition of DNAse after 5000 sec DNA release (white circles).
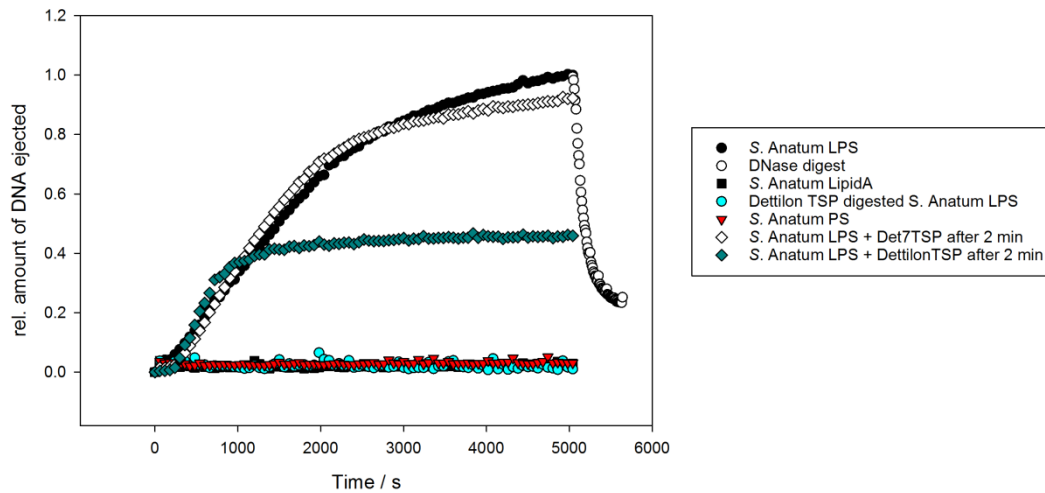
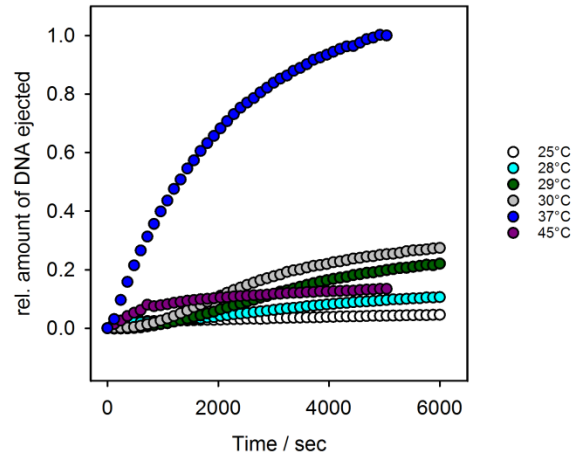**Fig. S8.** *In vitro* **DNA ejection from phage Det7 at high and low temperatures**

$8 \cdot 10^9$ pfu Det7 mixed with 67 µg ml^{-1} *S*. Typhimurium LPS in presence of a fluorescent DNA-binding dye. Fluorescence signals obtained at different temperatures in °C: 25 (white), 28 (cyan), 29 (dark green), 30 (grey), 45 (violet). The ejection curve at 37 °C is shown in blue for comparison.

Fig. S9. *In vitro* DNA ejection from phage Det7 triggered at different lipopolysaccharide concentrations
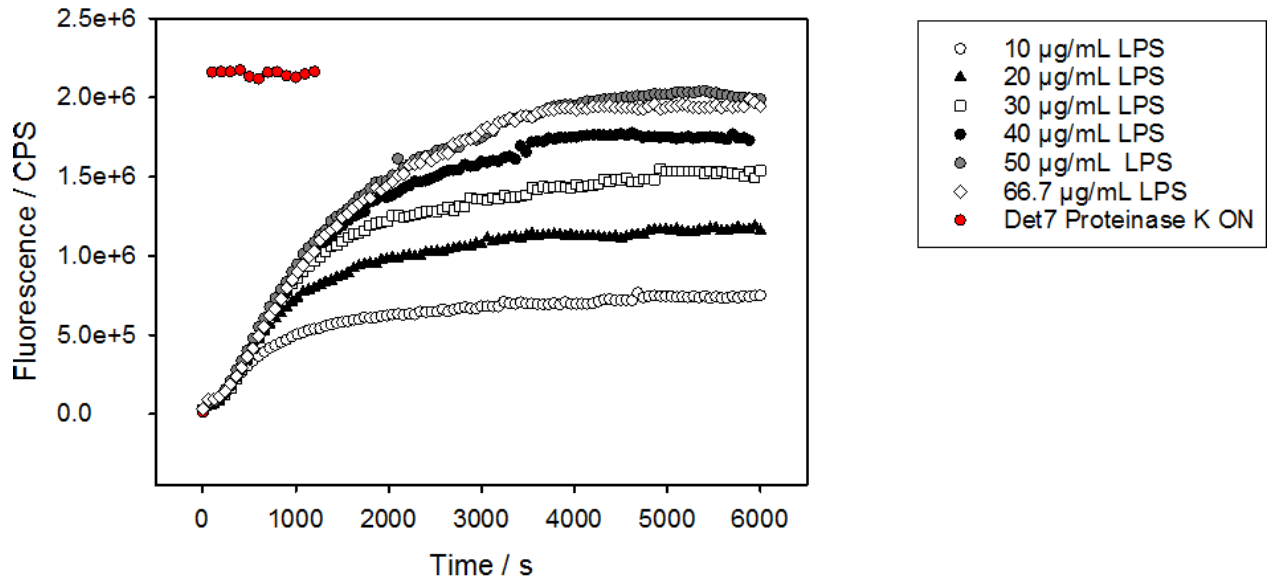
$4 \cdot 10^9$ Det7 phage particles were mixed with 50 µg ml^{-1} LPS in presence of a fluorescent DNA-binding dye. Fluorescence signal obtained at 37 °C with *S*. Typhimurium LPS at the different concentrations indicated. Red circles show the signal obtained after treating the same amount of phage particles over night with 5 mg ml^{-1} proteinase K.

Fig. S10. *S.* **Typhimurium O-antigen cleavage kinetics of Det7 TSP and P22TSP**
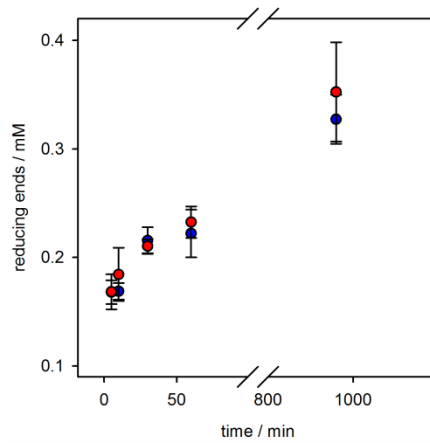
Reducing end formation was analyzed with 3-methyl-2-benzothiazolinonehydrazone (MBTH) as described (7). 1 mg ml^{-1} *S.* Typhimurium LPS were mixed with 50 µg ml Det7TSP (8) at 25 °C (blue) or 37 °C (red). The O-antigen cleavage reaction was stopped after different times by mixing aliquots with the same volume of 0.5 M NaOH.

Fig. S11. *S.* **Typhimurium and** *S.* **Anatum O-antigen chain length distribution on lipopolysaccharide grown at different temperatures**
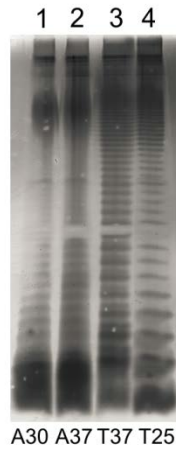
SDS-PAGE analysis of lipopolysaccharides (LPS) from *S.* Anatum grown at 30 °C (lane 1), at 37 °C (lane 2), or from *S.* Typhimurium (T) grown at 37 °C (lane 3) or at 25 °C (lane 4). Sample preparation and silver staining of the 15 % polyacrylamide gels were performed as described (9)

Table S1. MALDI-TOF mass spectrometry of oligosaccharides after size exclusion chromatography

Samples were mixed in equal volumes with matrix containing 100 mg ml^{-1} (w/v) 2,5-dihydroxybenzoic acid in 1:1:0.1 (v/v/v) water/acetonitrile/trifluoroacetic acid. Mass spectra were collected on a BrukerMicroflex (Bruker Daltonics, Bremen, Germany) and evaluated with the software mMass (10). Monoisotopic oligosaccharide masses were calculated from the S. anatum O-polysaccharide repeat unit structure [→3)-α-D-Gal$p$-(1→6)-α-D-Man$p$-(1→4)-β-L-Rha$p$-(1→] (11) containing zero, one, two or three acetylations.

Peak elution volume / ml[a]	[M+Na]$^+_{calc.}$ / Da[b]	[M+Na]$^+_{exp.}$ / Da[b]	Corresponding *S.* Anatum O-antigen repeat units (RU)
250	511.16	511.38	1RU
	553.17	553.43	1RU+Ac
235	981.32	982.00	2RU
	1023.34	1024.05	2RU+Ac
	1065.34	1066.10	2RU+2Ac
204	1451.49	1452.41	3RU
	1493.51	1494.46	3RU+Ac
	1535.52	1536.50	3RU+2Ac
	1577.17	1578.52	3RU+3Ac

[a]Elution volume on Superdex Peptide 26/60 at 1 ml min^{-1}
[b]Monoisotopic mass

Table S2. X-ray crystallographic data collection and refinement statistics

	Dettilon ΔN (Hg)	Dettilon ΔN (Pt)	Dettilon ΔN native	Dettilon ΔN + hexasaccharide
Data collection				
Beamline	BESSY 14.1	BESSY 14.1	BESSY 14.1	BESSY 14.1
Wavelength (Å)	1.009	1.072	0.9184	0.9184
Space group	$I2_13$	$I2_13$	$I2_13$	$P2_12_12_1$
Cell dimensions				
$a = b = c$ (Å)	152.1, 152.1, 152.1	151.9, 151.9, 151.9	151.6, 151.6, 151.6	153.4, 160.8, 161.7
$\alpha = \beta = \gamma$ (°)	90.0	90.0	90.0	90.0
Resolution (Å)*	48.11 - 1.65	48.05 - 1.97	47.95 – 1.63	48.76 – 2.1
	(1.75 - 1.65)	(2.09 - 1.97)	(1.73 – 1.63)	(2.23 – 2.10)
R_{meas}*	11.3 (73.3)	11.9 (84.7)	9.8 (83.1)	16.6 (229.2)
$< I / \sigma(I) >$*	12.33 (2.25)	12.68 (2.20)	9.57 (1.59)	10.60 (0.78)
CC1/2*	0.997 (0.745)	0.998 (0.732)	0.997 (0.639)	0.998 (0.452)
Completeness* (%)*	99.9 (99.3)	99.8 (99.1)	99.0 (97.5)	99.5 (98.7)
Redundancy	5.7	6.0	3.51	6.72
Refinement				
Resolution (Å)	-	-	1.63	2.1
No. reflections	-	-	138012	231486
R_{work} / R_{free} (%)	-	-	15.75 / 18.81	18.36 / 21.38
No. atoms				
Protein	-	-	4037	24548
Ligand	-	-	12	392
Water	-	-	480	1433
Mean B factor (Å^2)	-	-	19.03	39.48
R.m.s deviations				
Bond lengths (Å)	-	-	0.015	0.003
Bond angles (°)	-	-	1.551	0.777
Mol/AU	1	1	1	6

* Data in highest resolution shell are indicated in parenthesis.

Table S3. Arrhenius barriers for in vitro particle opening in tailed bacteriophages

Phage	Receptor	Method	Barrier / kcal mol^{-1}	Reference
Podovirus P22	O-antigen	Yo-Pro Fluorescence bulk	41	(12)
Siphovirus 9NA	O-antigen	Yo-Pro Fluorescence bulk	53	(12)
Siphovirus SPP1	Wall teichoic acid	Light scattering	29.5	(13)
Lambda	LamB	Light scattering	26.4	(13)
T5	FhuA	Yo-Pro Fluorescence single particle	41.6	(14)
Det7	O-antigen	Yo-Pro Fluorescence bulk	19.2 / 26.8	this work

Table S4. Fitted parameters of Det7 DNA ejection curves

For datasets at 29, 30 and 31 °C (italic, boxed in red) it was not possible to compute confidence intervals.

Temp	Parameters	Value	Lowerbound	Higherbound
28				
	k1	0.00039	0.00031	0.00046
	k2	0.00085	0.00070	0.00100
	A0	10.3	9.8	10.7
29	*k1*	*0.00057*	*-0.13212*	*0.13325*
	k2	*0.00057*	*-0.13218*	*0.13331*
	A0	*24.8*	*24.0*	*25.7*
30	*k1*	*0.00062*	*-0.39958*	*0.40082*
	k2	*0.00062*	*-0.39975*	*0.40099*
	A0	*28.8*	*26.9*	*30.8*
31	*k1*	*0.00067*	*-0.12385*	*0.12519*
	k2	*0.00067*	*-0.12395*	*0.12529*
	A0	*56.8*	*54.9*	*58.6*
31.5	k1	0.00050	0.00048	0.00052
	k2	0.00247	0.00232	0.00262
	A0	51.7	51.3	52.1
32	k1	0.00055	0.00054	0.00057
	k2	0.00280	0.00266	0.00294
	A0	55.9	55.6	56.2
33	k1	0.00057	0.00056	0.00058
	k2	0.00263	0.00252	0.00274
	A0	73.7	73.3	74.0
34	k1	0.00059	0.00058	0.00060
	k2	0.00381	0.00363	0.00398
	A0	88.2	87.8	88.5
35	k1	0.00068	0.00066	0.00069
	k2	0.00355	0.00339	0.00372
	A0	90.3	90.0	90.6
36	k1	0.00089	0.00081	0.00097
	k2	0.00177	0.00154	0.00200
	A0	88.3	87.6	89.0
37	k1	0.00075	0.00073	0.00077
	k2	0.00367	0.00339	0.00395
	A0	91.5	91.1	92.0
38	k1	0.00088	0.00087	0.00090
	k2	0.00576	0.00550	0.00602
	A0	98.7	98.5	98.9
39	k1	0.00096	0.00095	0.00098
	k2	0.00696	0.00644	0.00747
	A0	97.4	97.2	97.7
40	k1	0.00107	0.00106	0.00109
	k2	0.00740	0.00689	0.00791
	A0	97.6	97.4	97.8

References

1. Thompson JD, Higgins DG, & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22):4673-4680.
2. Clamp M, Cuff J, Searle SM, & Barton GJ (2004) The Jalview Java alignment editor. *Bioinformatics* 20(3):426-427.
3. Holm L & Sander C (1996) Mapping the protein universe. *Science* 273(5275):595-603.
4. Krissinel E & Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D* 60(Pt 12 Pt 1):2256-2268.
5. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography* 66:486-501.
6. Broeker NK*, et al.* (2018) In Vitro Studies of Lipopolysaccharide-Mediated DNA Release of Podovirus HK620. *Viruses-Basel* 10(6).
7. Anthon GE & Barrett DM (2002) Determination of reducing sugars with 3-methyl-2-benzothiazolinonehydrazone. *Analytical Biochemistry* 305(2):287-289.
8. Walter M*, et al.* (2008) Structure of the receptor-binding protein of bacteriophage det7: a podoviral tail spike in a myovirus. *J Virol* 82(5):2265-2273.
9. Kulikov EE, Golomidova AK, Prokhorov NS, Ivanov PA, & Letarov AV (2019) High-throughput LPS profiling as a tool for revealing of bacteriophage infection strategies. *Scientific Reports* 9:2958.
10. Strohalm M, Kavan D, Novak P, Volny M, & Havlicek V (2010) mMass 3: A Cross-Platform Software Environment for Precise Analysis of Mass Spectrometric Data. *Anal Chem* 82(11):4648-4651.
11. Robbins PW & Uchida T (1962) Studies on chemical basis of phage conversion of O-antigens in E-group Salmonellae. *Biochemistry* 1(2):323-335.
12. Andres D*, et al.* (2012) Tail morphology controls DNA release in two Salmonella phages with one lipopolysaccharide receptor recognition system. *Mol. Microbiol.* 83(6):1244-1253.
13. Raspaud E, Forth T, Sao-Jose C, Tavares P, & de Frutos M (2007) A kinetic analysis of DNA ejection from tailed phages revealing the prerequisite activation energy. *Biophys. J.* 93(11):3999-4005.
14. Mangenot S, Hochrein M, Radler J, & Letellier L (2005) Real-time imaging of DNA ejection from single phage particles. *Curr Biol* 15(5):430-435.