Supporting Information

King et al. 10.1073/pnas.1400814111

SI Text

SI Materials and Methods

Bacteria were routinely propagated in lysogeny broth (LB) or on LB agar containing 1.5% Bacto agar (Difco). All strains were grown at 37 °C with ampicillin (100 $\mu g \cdot mL^{-1}$), or kanamycin (50 $\mu g \cdot mL^{-1}$) as appropriate. *manA* strains require the addition of mannose (0.1% wt/vol) to produce O polysaccharide (OPS). Where OPS production by these strains was not specifically desired, glucose [0.4% (wt/vol)] was added to inhibit the uptake of mannose from rich medium.

For OPS production experiments, bacteria were cultured in low-phosphate medium [120 mM Tris·HCl, pH 7.4, 40 mM NaCl, 20 mM KCl, 40 mM NH_4Cl, 20 mM Na_2SO_4, 1 mM $MgCl_2$, 0.2 mM $CaCl_2$, 0.004 mM $ZnCl_2$, 0.002 mM $FeCl_3$, 0.1 mM KH_2PO_4, 0.4% (wt/vol) glycerol. This solution was autoclaved and the following filter-sterilized supplements were added: 0.5 $\mu g \cdot mL^{-1}$ thiamine and 20 $\mu g \cdot mL^{-1}$ each of adenine, uracil, histidine, and tryptophan]. Low-phosphate medium was inoculated by addition of 1/50th volume from an overnight culture in LB.

Construction of the Tetracycline-Inducible Protein Expression Vector pWQ552. The tetracycline-inducible vector was assembled in a three-component ligation reaction between the following fragments: (*i*) NdeI-*bla*-P15A*ori*-NheI, (*ii*) NheI-*tetR*-*tetO*-EcoRI, and (*iii*) EcoRI-MCS::*aph*-AseI; where MCS::*aph* refers to a multiple cloning site with the aminoglycoside 3′-phosphotransferase gene, *aph*, inserted into the XbaI site. The fragments were prepared as follows: (*i*) The P15A origin of replication and β-lactamase gene were obtained by NheI/NdeI digestion of pMBL19. (*ii*) Sequences for the *tet*-repressor gene (*tetR*), together with the upstream *tet* operator (*tetO*) sequences and downstream stem-loop transcriptional terminator, were derived from the Tn10 sequence found in the genomes of some *E. coli* BL21 strains (e.g., Accession: CP001665). This DNA was synthesized by GeneArt (Invitrogen). Translation-silent base pair substitutions were introduced to remove internal XbaI and NdeI sites and flanking EcoRI and NheI sites were included in the synthetic DNA. (*iii*) MCS sequence based on the MCS of pBAD24 (1) was included on the 5′ tails of PCR primers used to amplify the *aph* gene from pUC4K: 5′-aaaaaa**gaattc**accatggtacccggggatcctctagacgttgctgtcgctagcttcac and 5′-aaaaaa**attaat**aagcttgcatgcctgcaggtcgactctagagttgctgtcgcgaaccccag (EcoRI and AseI sites in bold).

Cells transformed with the desired ligation product were selected by growth on ampicillin and kanamycin. Subsequently, the kanamycin resistance cassette was removed by XbaI digestion and self-ligation.

Construction of WbdA- and WbdD Expression Vectors. Genes encoding His-tagged WbdA and WbdD proteins were excised from pWQ492 and pWQ470, respectively, using EcoRI and HindIII, and ligated into similarly digested pWQ552 to produce pWQ685 and pWQ686. Plasmids were verified by DNA sequencing (Guelph Advanced Analysis Centre) or restriction endonuclease digestion.

Construction of His_6-$WbdD_{1–536}$ Expression Vector. To construct plasmid pWQ502 for expression of His_6-$WbdD_{1–536}$, primers, 5′-ggag**gaattc**accatgcatcatcatcatcatcatgagaacctatattttcaaggcactaaagacttaaacacgctgg and 5′-gatc**aagctt**ttacgccagttgctccatattctcacc containing EcoRI and HindIII restriction sites, respectively, were used to amplify the *wbdD* gene by PCR using pWQ470 as a template. The PCR fragment was ligated into EcoRI/HindIII-digested pBAD24.

Determination of Number of O Units in OPS Bands. Some bacteria produce OPS in which very short chains can be detected by SDS/PAGE analysis of lipopolysaccharides (LPSs). In these cases, the number of repeat units in any OPS band can be judged by counting up the LPS "ladder" from the lipid A-core band. *Escherichia coli* O9a does not produce such short chains and so this method is not applicable. Therefore, to calculate the number of sugars in SDS/PAGE bands of *E. coli* O9a LPS we compared the electrophoretic mobility of short O9a OPS chains produced through strong overexpression of a truncated WbdD protein from pWQ502, with the positions of *E. coli* O9 LPS bands (Fig. S7). The O9 LPS represents a different serotype but its repeat unit only differs from serotype O9a in the possession of an extra mannose residue in the repeat unit, due to a mutation in the *wbdA* gene (2, 3). Unlike O9a, the O9 LPS does contain very short OPS chains so that in the low molecular weight range it is possible to count the number of O repeats. This method indicates that the lowest intense band in the "short O9a" OPS contains 5 repeat units which is equivalent to 20 mannose residues added to lipid A-core-adapter (Fig. S6).

Radiolabeling of LPS. Whole bacterial cells were radiolabeled by growth in 5 mL of low-phosphate medium containing 3.7 $MBq \cdot mL^{-1}$ of $H_3\,{}^{32}PO_4$ (Perkin-Elmer). Cultures were incubated with shaking for 6–8 h. At harvest, the cultures had optical densities between 0.2 and 0.7 at 600 nm.

Confirmation That Protein Is Induced Linearly from P_{tet}. CWG635 [pWQ686] (His_6-WbdD) or CWQ901[pWQ685] (WbdA-His_{10}) were grown at 37 °C in low-phosphate minimal medium. Except for the lack of radioisotope, the culture conditions were identical to those used to prepare radiolabeled LPS. One A_{600nm} unit of bacterial cells was collected from each culture by centrifugation during midexponential growth ($A_{600\ nm}$ = 0.5–0.8) and lysed in SDS/PAGE buffer. The whole-cell lysates were separated on 10% SDS/PAGE gels and transferred to nitrocellulose for Western blot analysis. Blots were probed with anti-His_5 (Qiagen) and goat anti-mouse-horseradish peroxidase-conjugated antibodies (Jackson Immunoresearch) and developed with Luminata Crescendo chemiluminescent substrate (Millipore). Densitometry of bands corresponding to the His-fusion proteins was performed on digitized images using ImageJ software (4). Densitometry of the WbdD samples was performed on two separate Western blots to enable all protein concentrations to be measured within the dynamic range of the chemiluminescent detection system. These data were normalized by inclusion of the 8 $ng \cdot mL^{-1}$ sample on both blots.

Isolation and Analysis of Nonradiolabeled LPS. For routine analysis, LPS was prepared by proteinase K digestion of whole cell lysates (5). The LPS was then resolved using tricine-SDS-polyacrylamide gel electrophoresis (tricine-SDS/PAGE) (6) in 65- or 80-mm-long 12.5% resolving gels. Electrophoresis was conducted at 25 mA (constant current) and then LPS was visualized by silver staining (7).

Isolation and Analysis of ^{32}P-Radiolabeled LPSs. We used a modification of the rapid LPS purification method of Yi and Hackett (8). Cells were harvested from 5 mL of culture, and washed with 500 μL of water. The cells were suspended in 200 μL of 70% EtOH, vigorously mixed by vortex, incubated for 20 min at room

temperature, and then harvested by centrifugation (12,000 × *g* for 1 min). This process was repeated with the same volume of 100% EtOH, and then acetone. The resulting pellet was air-dried for 20 min. Cells were then vortexed in 200 μL of TRIzol reagent (Invitrogen) with 20 μL of chloroform, and incubated at room temperature for 10 min. Phases were separated by centrifugation at 12,000 × *g* for 10 min and the aqueous (upper) layer was transferred to a fresh tube. Next, 100 μL of H_2O was added, the mixture was vortexed vigorously, and incubated at room temperature for 10 min. Phases were separated as before and these steps were repeated until four aqueous layers had been collected. These were combined and dried by incubation at 55 °C for 16–20 h. The residue was suspended in 500 μL of 0.375 M $MgCl_2$ in 95% EtOH and chilled for 1 h at −20 °C. The tubes were centrifuged in a chilled rotor (12,000 × *g* for 10 min), the supernatant was discarded, and the pellet was dissolved in 50 μL of 10 mM $MgCl_2$. DNase and RNase were added to 20 $\mu g \cdot mL^{-1}$ and the tubes were incubated at room temperature for 1 h. The LPS solution was then transferred to a tube containing the residue from freeze-drying 375 μL of 1 M $MgCl_2$, and 950 μL of EtOH was added. The suspension was chilled at −20 °C for 1 h, and then the LPS was collected by centrifugation in a chilled rotor (12,000 × *g* for 10 min). The pellet was dissolved in 100 μL of 4× Laemmli sample buffer (9) and proteinase K was added to 200 $\mu g \cdot mL^{-1}$. The sample was incubated at 55 °C overnight, and 3 μL was loaded per lane for tricine SDS/PAGE analysis. LPS was separated on 80-mm-long gels. The gels were dried and analyzed with a phosphorimager (Bio-Rad Personal Molecular Imager FX).

Quantitation of OPS Bands. The method for quantifying OPS bands is outlined in Fig. S7. For each SDS/PAGE lane, pixel densities in the phosphorimager output were analyzed using gel analysis tools in ImageJ (4). Peak areas were measured using the Multipeak Fitting 2 package in Igor Pro-6.21 (WaveMetrics). Gauss peaks were fitted into the data envelope and, as far as possible, the peak dimensions and positions were allowed to freely optimize. Where necessary, constraints were applied so that the peak fitting conformed to our knowledge of OPS bands in SDS/PAGE gels, in particular, that bandwidth (i.e., the Gauss peak full width at half maximum) and interband spacing both increase closer to the bottom of the gel. Baseline was fitted simultaneously with the peaks. The area under each Gauss peak was used as the measurement of the quantity of OPS with that particular length. Each radiolabeled LPS sample was loaded on three or more gel lanes, and the electrophoresis–autoradiography–densitometry–peak-fitting process was performed for each lane independently. Estimated errors based on these technical replicates ranged from 0.004 to 0.773% (SEM).

SI Mathematical Modeling

It has previously been shown that to generate modal OPS length distributions, the rates of nascent polymer elongation or termination must depend on polymer length (10). This dependence on chain length could be either implicit (e.g., with temporal regulation of biosynthesis) or explicit (e.g., with length dependence in the elongation or termination rates). We examined both of these possibilities.

A Molecular Clock Generates Poissonian OPS Distributions. We first constructed a mathematical model of polymer synthesis in which chain-length control is achieved by allowing synthesis to occur only for a specified duration. We refer to this regulation mechanism as a molecular clock. A molecular clock can regulate polymer length with either a distributive or processive polymerase. In distributive polymerization a polymer can be elongated by more than one enzyme or catalytic site, as the enzyme and substrate may dissociate and rebind during the polymerization process. In this case, all polymerases in the cell must be synchronized so that if the nascent polymer is extended by multiple enzymes, synthesis is still terminated a precise duration after it began. This implies that "start" and "stop" signals act globally on an entire cell or cell compartment. Alternatively, each polymer could be synthesized by a single polymerase (processive polymerization), in which case biosynthesis need not be synchronized.

Assumptions.

1) Synthesis of each nascent polymer occurs for a precise duration, τ.
2) The rates of elongation, termination export, and degradation do not depend on nascent polymer length.

We allowed the elongation rate $k(\mathrm{A},\mathrm{D},t)$ to depend on the concentrations of both WbdA and WbdD, and possibly time. For a single polymer, the chain extension process can be written as

$$n_i + m \xrightarrow{k(\mathrm{A},\mathrm{D},t)} n_{i+1}, \qquad \textbf{[S1]}$$

where m is a monomer and n_i is a nascent polymer of length i. For O9a OPS, this "monomer" is actually the four-sugar repeat added by the mannosyltransferase WbdA (Fig. 1).

The polymer length distribution in this model is either generated by a molecular clock that elongates a single polymer for a duration τ (processive) or an ensemble of synchronized molecular clocks that are all active for a duration τ (distributive). In both cases, polymerization proceeds with rate $k(\mathrm{A},\mathrm{D},t)$ for a duration τ, so the polymer length distribution must be the same in both cases. For calculation purposes, therefore, we considered the situation in which all "clocks" are synchronized. In this scenario, we have a set of polymers each starting with length zero and each being extended by the same rate k for a specified duration τ. Mathematically, this situation is indistinguishable from systems of in vitro chemical polymerization considered by Flory (11) and we therefore expect the polymer lengths to follow a Poisson distribution. Generally, a Poisson distribution expresses the probability that a certain number of events occur within a specified time when these events are independent and occur at a constant average rate (e.g., radioactive decay events). For the molecular clock, the length of the nascent polymer "records" the number of interactions between monomers and the growing end of a nascent polymer during one round of synthesis. Therefore, it is natural that their length distribution is Poissonian. For completeness, we include a derivation adapted from ref. 11.

Writing N_i as the number of molecules of n_i, the polymerization can be summarized by the following equation:

$$\frac{dN_i}{dt} = \begin{cases} -k(\mathrm{A},\mathrm{D})N_i & \text{for } i = 0, \\ -k(\mathrm{A},\mathrm{D})N_i + k(\mathrm{A},\mathrm{D})N_{i-1} & \text{for } i \geq 1. \end{cases} \qquad \textbf{[S2]}$$

With the definition $\nu(\tau) \equiv \int_0^\tau k(\mathrm{A},\mathrm{D},t)dt$, we can rewrite Eq. **S2** as

$$dN_0 = -N_0 d\nu, \qquad \textbf{[S3]}$$

$$dN_i = (N_{i-1} - N_i)d\nu. \qquad \textbf{[S4]}$$

Integrating Eq. **S3** gives

$$N_0(\tau) = N_0(0)e^{-\nu(\tau)}. \qquad \textbf{[S5]}$$

With this solution, Eq. **S4** with $i = 1$ then becomes

$$dN_1 = \left(N_0(0)e^{-\nu(\tau)} - N_1\right)d\nu, \qquad \textbf{[S6]}$$

which is solved by $N_1 = \nu(\tau)N_0(0)e^{-\nu(\tau)}$. We can continue this process of substituting N_{i-1} back into Eq. **S4** to iteratively integrate dN_i for increasing $i > 1$ to obtain

$$N_i = \frac{N_0(0)e^{-\nu(\tau)}\nu(\tau)^i}{i!}. \quad \textbf{[S7]}$$

This is a Poisson distribution, characterized by the single parameter $\nu(\tau)$, equal to both the mean and the variance. For the case in which the polymerization rate $k(A, D, t)$ does not change during a single round of synthesis, $\nu(\tau) = k(\mathrm{A}, \mathrm{D})\tau$ so that the mean length of polymers and the variance of the distribution both increase linearly with the time for which polymerization is allowed to proceed. For generality, we allowed $\nu(\tau)$ to depend on the concentrations of WbdA and WbdD; however, our analysis shows that a molecular clock is constrained to generate distributions with an equal mean and variance, regardless of the form of the function $k(\mathrm{A}, \mathrm{D})$. The experimentally observed OPS distributions for *E. coli* O9a all have variance much smaller than the mean (Fig. 2*D*). Therefore, the O9a OPS length is more tightly regulated that the broad spread of Poisson distributions (Fig. S2*A*) and therefore cannot be generated by a molecular clock.

Previous work on length control in template-independent polymerization stated that distributive polymerization generates product lengths according to a Poisson distribution whereas in processive polymerization, product lengths are determined by when the enzyme releases the nascent polymer (12). In the above consideration, we found that Poisson distributions are indeed generated from distributive polymerization, however only when there is no chain length dependence in the elongation rate. What is often not made clear in discussions of distributive polymerization is that these Poisson distributions are time-dependent, which means that to generate terminated polymers with lengths distributed according to a Poisson distribution, synthesis must proceed in successive rounds of elongation and termination. In addition, we have shown that processive polymerization can also generate Poisson distributions. This is because an ensemble of synchronized distributive polymerases elongating for a fixed duration is equivalent to an ensemble of nonsynchronized processive polymerases each elongating for the same duration.

In the above consideration, we assumed that elongation of a polymer by each protein complex occurred for a fixed duration τ, and found that the chain-length distributions generated would be more dispersed than those actually observed. In a real biological system, however, the elongation time τ would not be so precisely defined, which would cause polymer length distributions to be even more dispersed than Poissonian. Therefore, in terms of reproducing the observed *E. coli* O9a OPS distributions, the model above represents a best case for the molecular clock model.

To further exclude the possibility that a molecular clock underlies the concentration dependence that we see in our data, we considered the possibility that a small polymer of fixed length was the initial substrate for a molecular clock polymerase complex. This would allow the Poisson distributions to be shifted to longer mean chain lengths without increasing the variance. We used the following equation for a "shifted" Poisson distribution arising from a molecular clock that acts on an initial substrate of length j:

$$P_i = \begin{cases} 0, & i < j \\ \dfrac{e^{-\nu}\nu^{i-j}}{(i-j)!}, & i \geq j \end{cases}. \quad \textbf{[S8]}$$

We minimized the sum of squared errors between shifted Poisson distributions and the entire set of experimental data for values of initial substrate length, j, ranging from 0 to 15 mannose repeats. We found that the best fit was obtained when the substrate had a length of 6 mannose repeats ($j = 6$) (Fig. S2 *B–D*). However, regardless of how the concentrations of WbdA and WbdD affect the elongation rate or duration, this model is still constrained to produce distributions that increase in variance as the mean chain length increases. The resulting distributions are therefore still too disperse to explain the tightly regulated OPS lengths seen for high concentrations of WbdA or the wild-type *E. coli* O9a OPS distribution (Fig. S2*D*).

In summary, we have shown that a molecular clock with no explicit chain-length dependence in elongation or termination rates generates product lengths according to a Poisson distribution and therefore cannot account for the tightly regulated and concentration-dependent OPS lengths generated by *E. coli* O9a. In addition, we have clarified the assumptions under which distributive polymerization generates Poisson distributions and shown that these distributions are in fact dependent on the time for which synthesis is allowed to proceed.

Having excluded the possibility that temporal regulation of biosynthesis underlies *E. coli* O9a OPS length control, we went on to formulate a model with explicit polymer length dependence in elongation–termination rates.

Steady-State Model of OPS Biosynthesis. *E. coli* O9a OPS biosynthesis is thought to occur on the inner cytoplasmic membrane by stepwise addition of monomers to the nonreducing end of a nascent polymer. Polymers of length zero are referred to as acceptors, and are tethered to the membrane. The first step in polymerization occurs when a monomer is transferred to an acceptor to generate a polymer of length 1. In the model, this nascent polymer can either be terminated by WbdD and subsequently exported to the outer membrane, or further extended by WbdA to generate a longer nascent polymer.

Assumptions.

1) The distribution of polymer lengths on the outer membrane is independent of time.
2) Nascent polymers are terminated in the cytoplasm and are not subsequently elongated.
3) The rates of elongation and termination do not depend on time but may depend on nascent polymer length.
4) Export and degradation of polymers is independent of polymer length and is not regulated in a time-dependent manner.

Under Assumptions 2 and 4, termination is the final step that influences the chain-length distribution observed on the outer cytoplasmic membrane. As discussed in the main text, although there is currently no evidence for degradation of terminated polymers, the degradation terms also incorporate dilution of polymers on the outer membrane due to cell growth.

Definitions.

- $N_i(t)$ = Nascent polymer concentration: The concentration of a polymer of length i on the inner cytoplasmic membrane at time t.
- $T_i(t)$ = Terminated polymer concentration: The concentration of a polymer of length i on the outer cytoplasmic membrane at time t.
- $\Gamma_i(t)$ = Terminated polymer distribution: The probability that a polymer on the outer membrane has length i at time t.
- $p_i(t)$ = Elongation rate: The rate at which nascent polymers of length i are elongated to length $i+1$.
- $q_i(t)$ = Termination rate: The rate at which nascent polymers of length i are terminated.
- $s(t)$ = Acceptor addition rate: The rate at which new acceptors (nascent polymers of length zero) are added to the inner membrane.
- $r_i(t)$ = Degradation rate: The rate of degradation of terminated polymers.

The following dynamical system explicitly models the rates of change of nascent and terminated polymers of different lengths. An illustration of the process is shown in Fig. S3. Briefly, nascent polymers of length zero are created with a rate $s(t)$ and are consumed by elongation to polymers of length 1 (Eq. **S9**). Nascent polymers of length $i>0$ are created by elongation from length $i-1$, with rate $p_{i-1}(t)$, and consumed by elongation or termination with rates $p_i(t)$ and $q_i(t)$, respectively (Eq. **S10**). Finally, terminated polymers of length $i>0$ are created from nascent polymers of length i with rate $q_i(t)$ and degraded (or diluted through cell growth) with rate $r_i(t)$ (Eq. **S11**):

$$\frac{dN_0(t)}{dt}=s(t)-p_0(t)N_0(t), \quad \textbf{[S9]}$$

$$\frac{dN_i(t)}{dt}=p_{i-1}(t)N_{i-1}(t)-(p_i(t)+q_i(t))N_i(t), \quad \text{for } i>0, \quad \textbf{[S10]}$$

$$\frac{dT_i(t)}{dt}=q_i(t)N_i(t)-r_i(t)T_i(t), \quad \text{for } i>0. \quad \textbf{[S11]}$$

Under Assumption 3, that the termination and elongation rates are independent of time, and Assumption 4, that the degradation of terminated polymers is independent of time and polymer length ($r_i(t)=r$), we obtain

$$\frac{dN_0(t)}{dt}=s-p_0N_0(t), \quad \textbf{[S12]}$$

$$\frac{dN_i(t)}{dt}=p_{i-1}N_{i-1}(t)-(p_i+q_i)N_i(t), \quad \text{for } i>0, \quad \textbf{[S13]}$$

$$\frac{dT_i(t)}{dt}=q_iN_i(t)-rT_i(t), \quad \text{for } i>0. \quad \textbf{[S14]}$$

Now if the terminated polymer length distribution is at equilibrium [Assumption 1, $(dT_i(t)/dt)=0$], it follows from Eq. **S14** that the nascent polymer distribution must also be independent of time and therefore $dN_i(t)/dt=0$, which gives

$$0=s-p_0N_0, \quad \textbf{[S15]}$$

$$0=p_{i-1}N_{i-1}-(p_i+q_i)N_i, \quad \text{for } i>0, \quad \textbf{[S16]}$$

$$0=q_iN_i-rT_i, \quad \text{for } i>0. \quad \textbf{[S17]}$$

Eqs. **S15** and **S16** can be solved for N_i in terms of p_i, q_i, and s,

$$N_0=\frac{s}{p_0}, \quad \textbf{[S18]}$$

$$N_i=\frac{s}{p_i+q_i}\prod_{j=1}^{i-1}\frac{p_j}{p_j+q_j}, \quad \text{for } i>0. \quad \textbf{[S19]}$$

Using Eq. **S17**, the concentration of terminated polymers is given by

$$T_0=0, \quad \textbf{[S20]}$$

$$T_i=\frac{s}{r}\frac{q_i}{p_i+q_i}\prod_{j=1}^{i-1}\frac{p_j}{p_j+q_j}, \quad \text{for } i>0. \quad \textbf{[S21]}$$

We are interested in the terminated chain-length distribution Γ_i, rather than the actual concentrations of the chains T_i. In calculating Γ_i from T_i, we can therefore neglect s and r, because they are independent of i, which gives the unnormalized terminated chain-length distribution,

$$\Gamma_0=0, \quad \textbf{[S22]}$$

$$\Gamma_i=\frac{q_i}{p_i+q_i}\prod_{j=1}^{i-1}\frac{p_j}{p_j+q_j}, \quad \text{for } i>0. \quad \textbf{[S23]}$$

With the change of variables $\tilde{q}_i=q_i/(p_i+q_i)$, $\tilde{p}_i=p_i/(p_i+q_i)$, this result can be written

$$\Gamma_i=\tilde{q}_i\prod_{j=1}^{i-1}\tilde{p}_j, \quad \textbf{[S24]}$$

where $\tilde{q}_0=0$. Noting that $\tilde{q}_i+\tilde{p}_i=1$ for any $i>0$, we can calculate the normalization factor,

$$\sum_{i=0}^{\infty}\Gamma_i=\tilde{q}_1+\tilde{q}_2\tilde{p}_1+\tilde{q}_3\tilde{p}_2\tilde{p}_1+\cdots, \quad \textbf{[S25]}$$

$$=(1-\tilde{p}_1)+(1-\tilde{p}_2)\tilde{p}_1+\cdots, \quad \textbf{[S26]}$$

$$=1-\tilde{p}_1+\tilde{p}_1-\tilde{p}_2\tilde{p}_1+\tilde{p}_2\tilde{p}_1+\cdots, \quad \textbf{[S27]}$$

$$=1. \quad \textbf{[S28]}$$

Thus, we obtain a simple formula for the terminated chain-length distribution,

$$\Gamma_i=\tilde{q}_i\prod_{j=1}^{i-1}\tilde{p}_j, \quad \textbf{[S29]}$$

$$=\tilde{q}_i\prod_{j=1}^{i-1}(1-\tilde{q}_j), \quad \textbf{[S30]}$$

because $\tilde{p}_i+\tilde{q}_i=1$. By definition, $\tilde{p}_i,\tilde{q}_i\in[0,1]$ and quantify the relative rates of elongation and termination, respectively. It is apparent therefore that these are actually elongation and termination probabilities as a function of chain length. Eq. **S30** is similar to the result obtained by Hunt (10), although the mathematical formulation considered here differs somewhat; consequently, the resulting polymer length distribution is completely determined by specifying only the termination probabilities $\tilde{q}_i$.

Termination Probability for a Molecular Ruler. In template-independent polymerization, a molecular ruler is a complex in which the active site of polymerization and termination are spatially separated by a well-defined distance that determines the minimum polymer length generated by the complex. In terms of elongation and termination probabilities, a molecular ruler has zero probability of terminating short polymers and high probability of terminating polymers that can reach the termination site. Because there is expected to be some variability in the exact termination length, we do not expect a completely switch-like response in termination

probability and therefore model this transition using a sigmoidal function (Fig. S4),

$$\tilde{q}_i = \frac{\lambda}{1+e^{-\theta(x-l)}}. \quad \text{[S31]}$$

λ gives the maximum value of the termination probability, l gives the average length of the ruler, and θ controls the rate of transition between low and high termination probabilities.

Concentration Dependence Suggests the Existence of Complexes with Variable Stoichiometry. For a molecular ruler to explain a polymer length distribution that depends on the relative concentration of the enzymes involved, the length of the ruler must change as a function of the concentrations of the enzymes. One possibility that could generate such concentration-dependent polymer length distributions is to have several molecular ruler complexes that exist in equilibrium and each produce polymers of different lengths.

To investigate this possibility, we assume that there are molecular complexes on the inner cytoplasmic membrane that contain WbdA and WbdD at various stoichiometries, each of which acts as an independent molecular ruler. Because WbdD has been shown to form stable trimers, and WbdA is known to bind to WbdD, partial occupation of putative three WbdA binding sites on a preformed WbdD trimer could allow three distinct complexes to be present on the inner membrane (AD_3, A_2D_3, and A_3D_3). Assuming also that the length of polymers generated by these species increases with the number of WbdA bound, this may be sufficient to generate the observed concentration-dependent polymer length distribution.

WbdA–WbdD Complex Formation. To find an expression for the relative concentrations of D_3, AD_3, A_2D_3, and A_3D_3, given the total concentrations of WbdD and WbdA, we formulated the following model of WbdA–WbdD complex formation using mass action kinetics. The reaction scheme is illustrated in Fig. 3*A* and formulated mathematically in Eqs. **S32–S37**. D_{trimer} is the total concentration of WbdD trimers, A_{total} is the total concentration of WbdA, A_{free} is the concentration of unbound WbdA, and $D_3, AD_3, A_2D_3, A_3D_3$ are the concentrations of WbdD trimers with 0, 1, 2, or 3 molecules of WbdA bound, respectively. k_{ij} are rate constants for the reactions illustrated in Fig. 3*A*.

$$D_{\text{trimer}} = D_3 + AD_3 + A_2D_3 + A_3D_3, \quad \text{[S32]}$$

$$A_{\text{total}} = A_{\text{free}} + AD_3 + 2A_2D_3 + 3A_3D_3, \quad \text{[S33]}$$

$$\frac{dD_3}{dt} = -k_{01}A_{\text{free}}D_3 + k_{10}AD_3, \quad \text{[S34]}$$

$$\frac{dAD_3}{dt} = k_{01}A_{\text{free}}D_3 - k_{10}AD_3 - k_{12}AD_3 + k_{21}A_2D_3, \quad \text{[S35]}$$

$$\frac{dA_2D_3}{dt} = k_{12}A_{\text{free}}AD_3 - k_{21}A_2D_3 - k_{23}A_2D_3 + k_{32}A_3D_3, \quad \text{[S36]}$$

$$\frac{dA_3D_3}{dt} = k_{23}A_{\text{free}}A_2D_3 - k_{32}A_3D_3. \quad \text{[S37]}$$

The steady-state solution of Eqs. **S32–S37** gives a system of six algebraic equations for five unknowns ($A_{\text{free}}, D_3, AD_3, A_2D_3, A_3D_3$) with parameters $D_{\text{trimer}}, A_{\text{total}}$, and k_{ij}. With the definitions $R_1 = k_{01}/k_{10}$, $R_2 = k_{01}k_{12}/k_{10}k_{21}$, and $R_3 = k_{01}k_{12}k_{23}/k_{10}k_{21}k_{32}$, this can be simplified to yield the following quartic polynomial in A_{free}:

$$\begin{aligned} &-R_3A_{\text{free}}^4 \\ &+ (R_3A_{\text{total}} - R_2 - 3R_3D_{\text{trimer}})A_{\text{free}}^3 \\ &+ (R_2A_{\text{total}} - R_1 - 2R_2D_{\text{trimer}})A_{\text{free}}^2 \\ &+ (R_1A_{\text{total}} - 1 - R_1D_{\text{trimer}})A_{\text{free}} + A_{\text{total}} = 0. \end{aligned} \quad \text{[S38]}$$

For a given set of parameters, A_{free} is determined numerically as the unique positive real zero of Eq. **S38**. The remaining variables can then be calculated from A_{free} and the parameters D_{trimer} and R_i,

$$D_3 = \frac{D_{\text{trimer}}}{1 + R_1A_{\text{free}} + R_2A_{\text{free}}^2 + R_3A_{\text{free}}^3}, \quad \text{[S39]}$$

$$AD_3 = \frac{R_1A_{\text{free}}D_{\text{trimer}}}{1 + R_1A_{\text{free}} + R_2A_{\text{free}}^2 + R_3A_{\text{free}}^3}, \quad \text{[S40]}$$

$$A_2D_3 = \frac{R_2A_{\text{free}}^2D_{\text{trimer}}}{1 + R_1A_{\text{free}} + R_2A_{\text{free}}^2 + R_3A_{\text{free}}^3}, \quad \text{[S41]}$$

$$A_3D_3 = \frac{R_3A_{\text{free}}^3D_{\text{trimer}}}{1 + R_1A_{\text{free}} + R_2A_{\text{free}}^2 + R_3A_{\text{free}}^3}. \quad \text{[S42]}$$

The parameters R_1, R_2, and R_3 can be replaced with dissociation constants for the binding of molecules of WbdA to trimers of WbdD, $K_{d(A:D_3)} = 1/R_1$, $K_{d(A:AD_3)} = R_1/R_2$, and $K_{d(A:A_2D_3)} = R_2/R_3$. Eqs. **S39–S42** can then finally be used to find the concentrations of the various WbdA–WbdD complexes at equilibrium given the total concentration of WbdA, WbdD, and a set of three dissociation constants.

Model Fitting and Parameter Estimation. Eqs. **S40–S42** give the equilibrium concentrations of AD_3, A_2D_3, and A_3D_3. If each of these complexes represents a molecular ruler defined with termination probability $\tilde{q}_i$ given by Eq. **S31**, we can calculate the total ensemble termination probability as a function of polymer length $\tilde{q}_i$ by weighting each of the "rulers" according to their relative concentration,

$$\tilde{q}_i = \frac{AD_3(\tilde{q}_{AD_3})_i + A_2D_3(\tilde{q}_{A_2D_3})_i + A_3D_3(\tilde{q}_{A_3D_3})_i}{AD_3 + A_2D_3 + A_3D_3}. \quad \text{[S43]}$$

We have made the additional assumption, for simplicity, that WbdD trimers without WbdA bound do not terminate nascent polymers. An additional molecular ruler for D_3 could be included in this modeling framework but would require the introduction of an extra parameter l_{D_3} and would not substantially alter the predicted OPS distributions.

Eq. **S30** allows the final polymer length distribution to be determined from this ensemble termination probability and OPS length distributions predicted by the model can be compared with the experimental data.

We can then fit the model to the experimental data by optimization of the dissociation constants $K_{d(A:D_3)}$, $K_{d(A:AD_3)}$, $K_{d(A:A_2D_3)}$, the concentration A_{total}, and the concentration D_{trimer}. In the experiments where the concentrations of WbdA and WbdD were varied, the induced protein concentration was shown to be proportional to the concentration of the inducer anhydrotetracycline (AhTc) over the ranges used (Fig. S1). Therefore, the concentration of inducer was used for the concentration of the proteins in the two datasets. Because there is no quantification of absolute protein concentration or K_d measurements for

WbdA–WbdD complex formation, the scale for concentrations and dissociation constants is arbitrary.

In addition to these concentrations and dissociation constants, the parameters that control the shape of the molecular rulers are λ, θ, and the three average termination lengths l_{AD_3}, $l_{A_2D_3}$, $l_{A_3D_3}$. Although there is no reason to expect that each of the three molecular complexes has the same shaped ruler (common θ and λ), we choose this constraint to reduce the number of free parameters. The average termination lengths are specified in terms of numbers of 4-mannose repeat units.

All numerical calculations and model fitting were performed in R (13). A differential evolution algorithm (14) was used to minimize the sum of squared errors between the model and the data for all titration experiments.

As discussed in the *Results* section of the main text, the shape of the distributions differs between the experiments in which we varied the concentration of WbdA or WbdD. We found that a much better fit to the WbdD titration experiment data can be obtained by allowing a single extra parameter θ' which controls the shape of the A_2D_3 and A_3D_3 rulers in the WbdD series. Possible reasons for this are discussed in the main text. In total there are 11 free parameters optimized to fit both data series, namely A_{total}, D_{trimer}, $K_{d(\text{A:D}_3)}$, $K_{d(\text{A:AD}_3)}$, $K_{d(\text{A:A}_2\text{D}_3)}$, λ, θ, θ', l_{AD_3}, $l_{A_2D_3}$, $l_{A_3D_3}$. Parameter values are shown in Table S1.

1. Guzman LM, Belin D, Carson MJ, Beckwith J (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* 177(14):4121–4130.
2. Kido N, Kobayashi H (2000) A single amino acid substitution in a mannosyltransferase, WbdA, converts the *Escherichia coli* O9 polysaccharide into O9a: Generation of a new O-serotype group. *J Bacteriol* 182(9):2567–2573.
3. Stenutz R, Weintraub A, Widmalm G (2006) The structures of *Escherichia coli* O-polysaccharide antigens. *FEMS Microbiol Rev* 30(3):382–403.
4. Abramoff MD, Magelhaes PJ, Ram SJ (2004) Image processing with ImageJ. *Biophotonics Int* 11(7):36–42.
5. Hitchcock PJ, Brown TM (1983) Morphological heterogeneity among *Salmonella* lipopolysaccharide chemotypes in silver-stained polyacrylamide gels. *J Bacteriol* 154(1):269–277.
6. Lesse AJ, Campagnari AA, Bittner WE, Apicella MA (1990) Increased resolution of lipopolysaccharides and lipooligosaccharides utilizing tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *J Immunol Methods* 126(1):109–117.
7. Tsai CM, Frasch CE (1982) A sensitive silver stain for detecting lipopolysaccharides in polyacrylamide gels. *Anal Biochem* 119(1):115–119.
8. Yi EC, Hackett M (2000) Rapid isolation method for lipopolysaccharide and lipid A from gram-negative bacteria. *Analyst (Lond)* 125(4):651–656.
9. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227(5259):680–685.
10. Hunt F (1985) Patterns of LPS synthesis in gram negative bacteria. *J Theor Biol* 115(2): 213–219.
11. Flory PJ (1940) Molecular size distribution in ethylene oxide polymers. *J Am Chem Soc* 62(6):1561–1565.
12. May JF, Splain RA, Brotschi C, Kiessling LL (2009) A tethering mechanism for length control in a processive carbohydrate polymerization. *Proc Natl Acad Sci USA* 106(29):11851–11856.
13. R Core Team (2013) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).
14. Mullen K, Ardia D, Gil D, Windover D, Cline J (2011) DEoptim: An R package for global optimization by differential evolution. *J Stat Softw* 40(6):1–26.

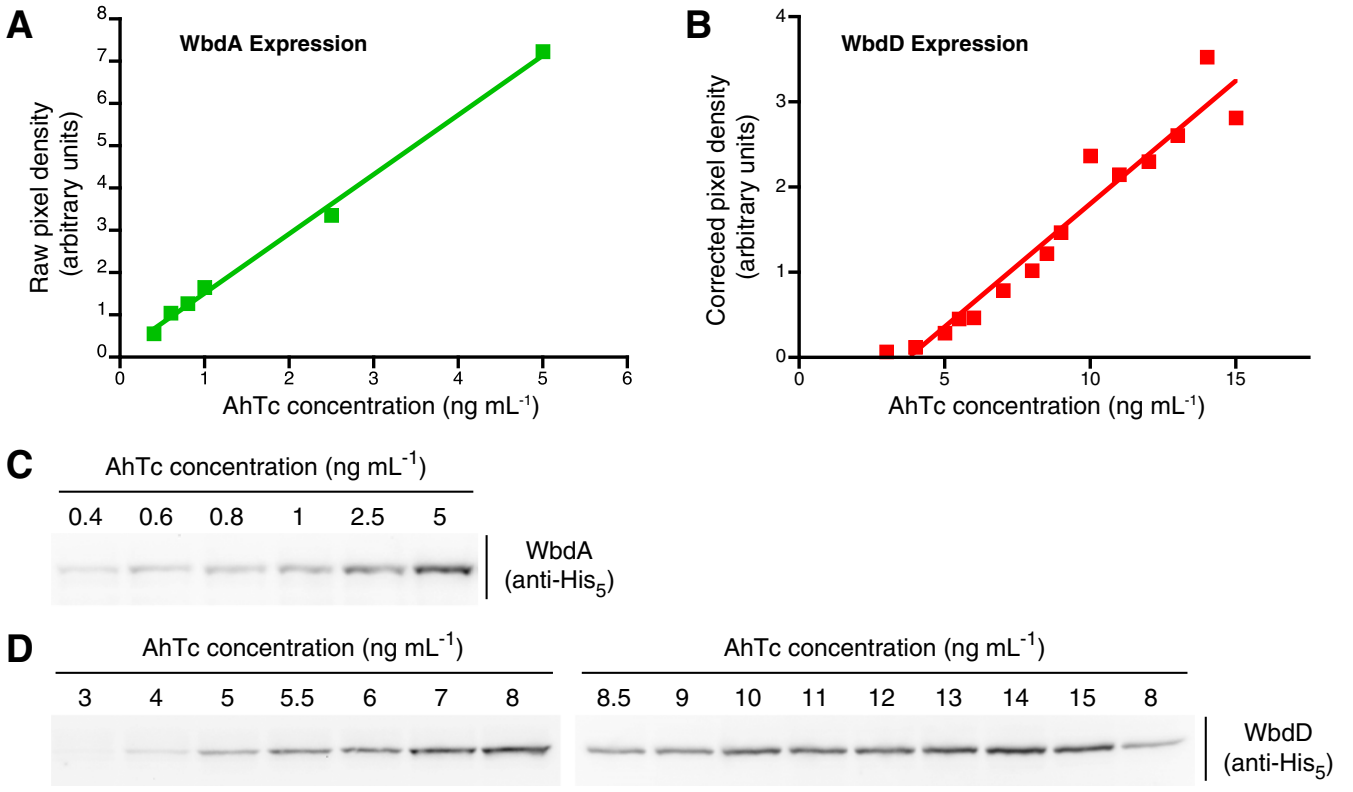

Fig. S1. Response of WbdA and WbdD protein concentration levels to the inducing agent AhTc when expressed from the tetracycline promoter of pWQ552. Protein concentrations were measured by densitometry on Western blots using an anti-His_5 primary antibody. Linear regression fit the WbdA and WbdD data with R^2 values of 0.9962 and 0.9388, respectively. *A* and *B* show representative results of densitometry performed on the Western blots shown in *C* and *D*, respectively.

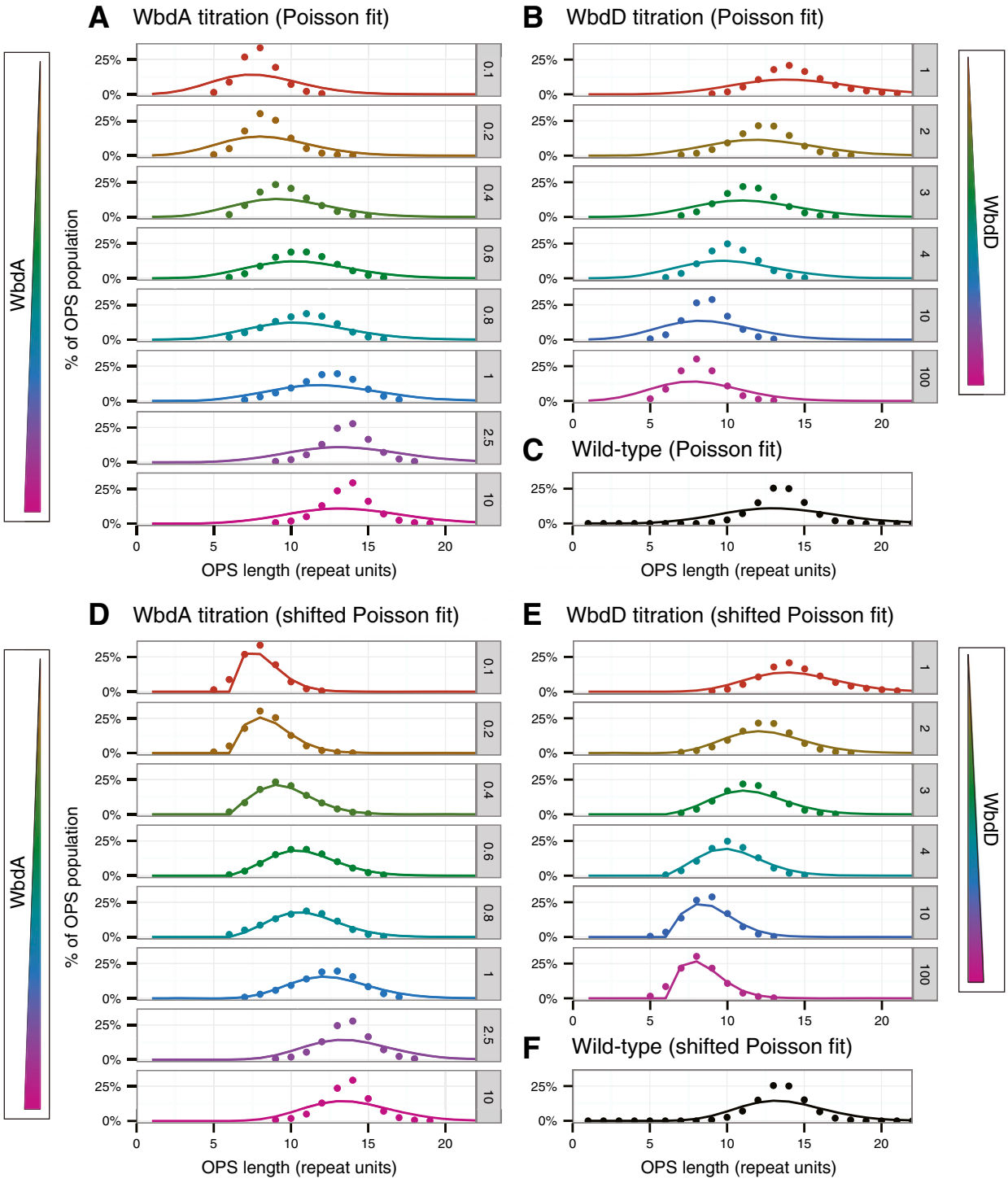

Fig. S2. *E. coli* O9a OPS distributions are more tightly regulated than those produced by a molecular clock. (*A–C*) Poisson distributions (solid lines) with mean equal to the experimentally observed OPS distribution mean, plotted with measured OPS distributions (circles). (*D–F*) Poisson distributions that are shifted away from small chain lengths by six repeat units plotted with measured OPS distributions (circles).

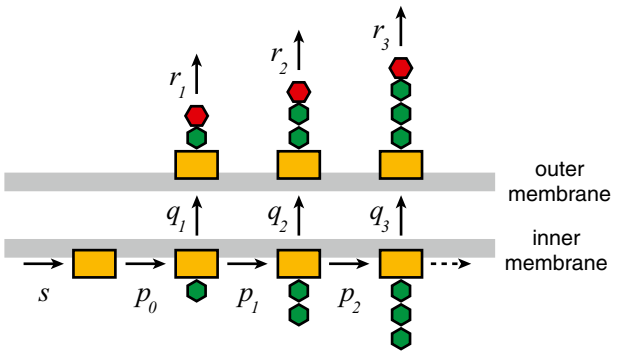

Fig. S3. Rates in the steady-state model of OPS biosynthesis. New acceptors are added with rate s. Nascent polymers of length i are extended with rate p_i and terminated–exported with rate q_i. Terminated polymers of length i are degraded with rate r_i on the outer membrane.

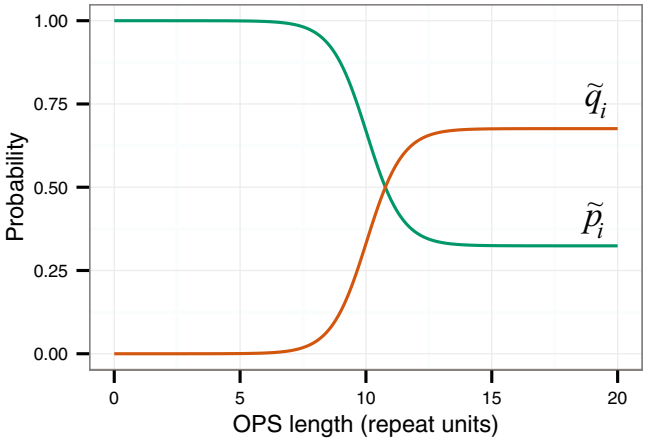

Fig. S4. Termination probability $\tilde{q}_i$ and elongation probability $\tilde{p}_i$ as a function of chain length for a molecular-ruler-type of enzyme complex.

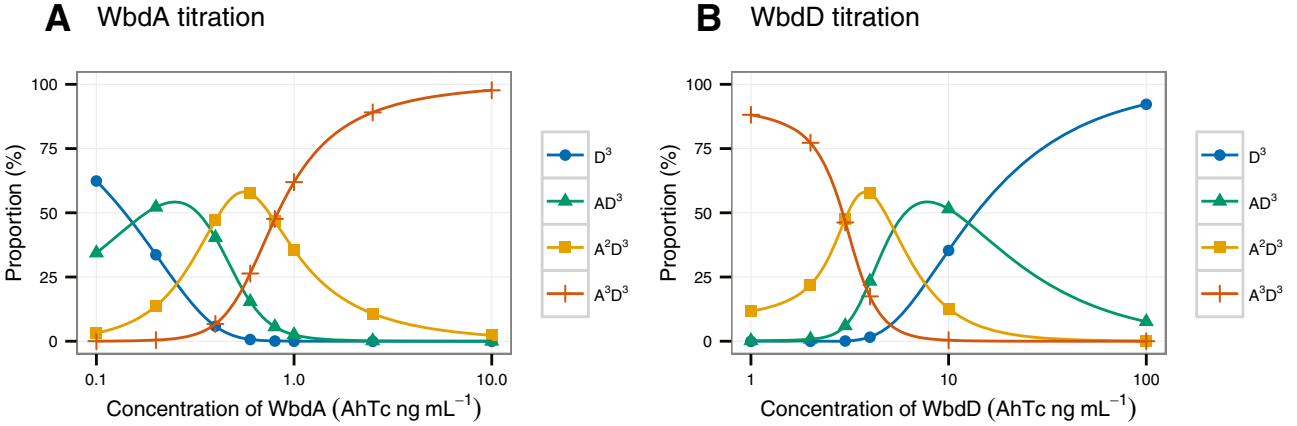

Fig. S5. Predicted relative concentrations of WbdA–WbdD complexes. Fitting the model to the OPS distributions requires estimation of the equilibrium dissociation constants K_d(A:D_3), K_d(A:AD_3), K_d(A:A_2D_3) and thus the relative concentrations of WbdA–WbdD complexes as the concentration of (*A*) WbdA or (*B*) WbdD is varied.

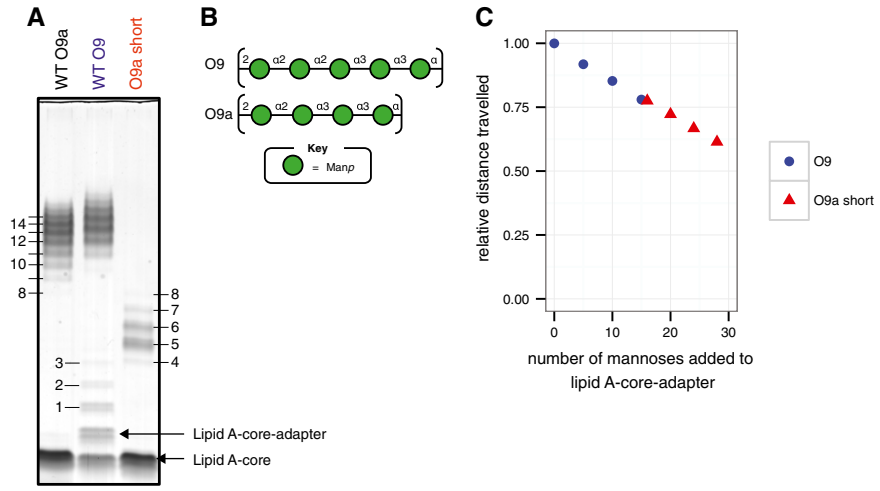

Fig. S6. Determination of the number of OPS repeat units in each band after SDS/PAGE separation of *E. coli* O9a LPS. (*A*) Silver-stained tricine-SDS/PAGE analysis to compare wild-type *E. coli* O9 LPS with that of an *E. coli* O9a strain, "O9a short," in which strong overexpression of truncated WbdD from pWQ502 produces short OPS chains. (*B*) These samples are comparable because the structure of *E. coli* O9 LPS is similar to that of O9a, except that there is an additional mannose in the O repeat. (*C*) A plot of electrophoretic mobility versus the calculated total number of mannoses in all of the repeat units, based on the assumption that each of the LPS bands has the number of O9 or O9a repeat units as annotated on the gel in *A*.

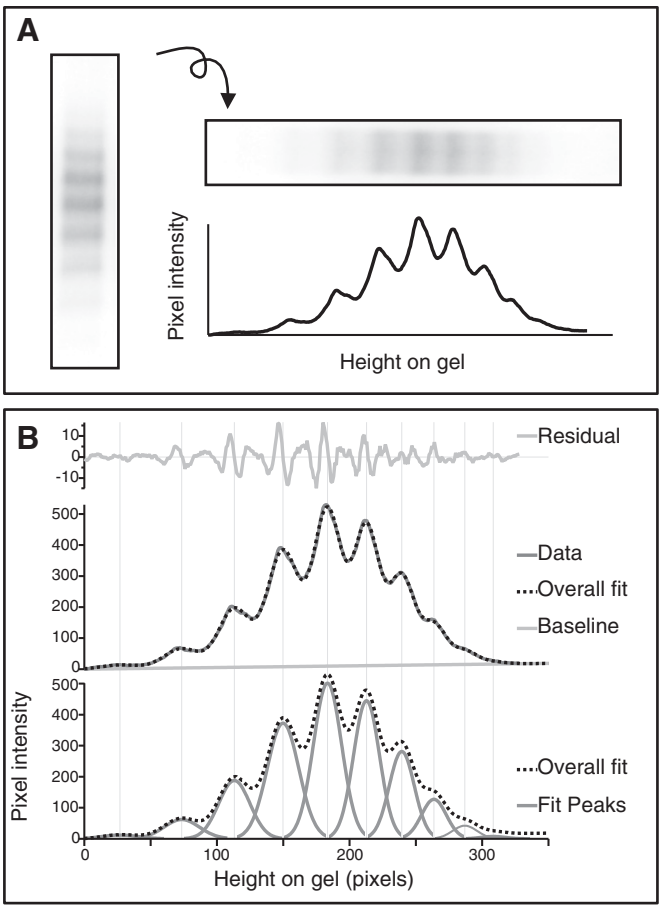

Fig. S7. Process for measuring OPS chain length distributions. (*A*) Radiolabeled LPS sample was separated by tricine-SDS/PAGE and radioisotope located by autoradiography. A pixel density plot was made for each autorad lane. (*B*) The area of each peak was calculated using multipeak curve fitting.

Table S1. Parameter values for the variable geometry model

Parameter	Fitted value	Units
A_{total}	2.6	ng·μL^{-1} AhTc
D_{trimer}	0.24	ng·μL^{-1} AhTc
l_{AD_3}	7.4	OPS repeats
$l_{A_2D_3}$	10.0	OPS repeats
$l_{A_3D_3}$	13.3	OPS repeats
λ	0.68	Probability
θ	1.41	—
θ'	1.29	—
$K_{d(A:D_3)}$	4.6×10^{-3}	ng·μL^{-1} AhTc
$K_{d(A:AD_3)}$	2.7×10^{-2}	ng·μL^{-1} AhTc
$K_{d(A:A_2D_3)}$	2.2×10^{-1}	ng·μL^{-1} AhTc

Table S2. Bacterial strains and plasmids used in this study

Strain/plasmid	Description	Reference
Strains		
E. coli Top10	Cloning strain	Invitrogen
E. coli CWG28	Derivative of E69; O9a:K $^{-}$; *trp his lac rpsL cps*$_{K30}$; Sm^r	1
E. coli CWG901	CWG28 *manA* Δ*wbdA*; Sm^r, Tc^r, Gm^r	2
E. coli CWG635	CWG28 *manA wbdD*::*aacC1*; Sm^r, Tc^r, Gm^r	3
Plasmids		
pMBL19	pACYC177 derivative; *bla* P15A-*ori*	4
pUC4K	Source of *aph*	GE Healthcare
pBAD24	L-arabinose-inducible plasmid; Ap^r	5
pWQ492	pBAD24 derivative encoding WbdA-His$_{10}$	2
pWQ470	pBAD24 derivative encoding His$_6$-WbdD	2
pWQ502	pBAD24::his$_6$-*wbdD*$_{1–536}$	This work
pWQ552	Derivative of pMBL19 with Tc-inducible promoter; *bla tetR* P15A-*ori*	This work
pWQ685	pWQ552::*wbdA*-his$_{10}$	This work
pWQ686	pWQ552::his$_6$-*wbdD*	This work

1. Whitfield C, Schoenhals G, Graham L (1989) Mutants of Escherichia coli O9:K30 with altered synthesis and expression of the capsular K30 antigen. *J Gen Microbiol* 135(10):2589–2599.
2. Clarke BR, Greenfield LK, Bouwman C, Whitfield C (2009) Coordination of polymerization, chain termination, and export in assembly of the *Escherichia coli* lipopolysaccharide O9a antigen in an ATP-binding cassette transporter-dependent pathway. *J Biol Chem* 284(44):30662–30672.
3. Clarke BR, Cuthbertson L, Whitfield C (2004) Nonreducing terminal modifications determine the chain length of polymannose O antigens of *Escherichia coli* and couple chain termination to polymer export via an ATP-binding cassette transporter. *J Biol Chem* 279(34):35709–35718.
4. Nakano Y, Yoshida Y, Yamashita Y, Koga T (1995) Construction of a series of pACYC-derived plasmid vectors. *Gene* 162(1):157–158.
5. Guzman LM, Belin D, Carson MJ, Beckwith J (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* 177(14): 4121–4130.