

Lipopolysaccharide O antigen size distribution is determined by a chain extension complex of variable stoichiometry in *Escherichia coli* O9a

Jerry D. King^{a,1}, Scott Berry^{b,1}, Bradley R. Clarke^a, Richard J. Morris^b, and Chris Whitfield^{a,2}

^aMolecular and Cellular Biology, University of Guelph, Guelph, ON, Canada N1G 2W1; and ^bComputational and Systems Biology, John Innes Centre, Norwich NR4 7UH, United Kingdom

Edited by Graham C. Walker, Massachusetts Institute of Technology, Cambridge, MA, and approved March 14, 2014 (received for review January 15, 2014)

The lengths of bacterial polysaccharides can be critical for their biological function. Unlike DNA or protein synthesis, where polymer length is implicit in the nucleic acid template, the molecular mechanisms for regulating polysaccharide length are poorly understood. Two models are commonly cited: a “molecular clock” regulates length by controlling the duration of the polymer extension process, whereas a “molecular ruler” determines length by measurement against a physical structure in the biosynthetic complex. *Escherichia coli* O9a is a prototype for the biosynthesis of O polysaccharides by ATP-binding cassette transporter-dependent processes. The length of the O9a polysaccharide is determined by two proteins: an extension enzyme, WbdA, and a termination enzyme, WbdD. WbdD is known to self-oligomerize and also to interact with WbdA. Changing either enzyme’s concentration can alter the polysaccharide length. We quantified the O9a polysaccharide length distribution and the enzyme concentration dependence *in vivo*, then made mathematical models to predict the polymer length distributions resulting from hypothetical length-regulation mechanisms. Our data show qualitative features that cannot be explained by either a molecular clock or a molecular ruler model. Therefore, we propose a “variable geometry” model, in which a postulated biosynthetic WbdA–WbdD complex assembles with variable stoichiometry dependent on relative enzyme concentration. Each stoichiometry produces polymers with a distinct, geometrically determined, modal length. This model reproduces the enzyme concentration dependence and modality of the observed polysaccharide length distributions. Our work highlights limitations of previous models and provides new insight into the mechanisms of length control in polysaccharide biosynthesis.

chain length | mathematical modeling | template-independent polymerization

Whereas nucleic acids and proteins are synthesized from a template of specified length, carbohydrates are constructed template-free, implying that the mechanisms for length control must lie in the biosynthetic machinery itself.

Lipopolysaccharide (LPS) is a glycolipid which constitutes the major and characteristic component of the outer leaflet of the outer membrane of most Gram negative bacterial cell envelopes (reviewed in ref. 1). Most bacteria produce LPS molecules that contain a conserved lipid A moiety linked to a short (core) oligosaccharide. In many species, a proportion of LPS molecules have an additional domain known as O polysaccharide (OPS), O antigen or O chain. OPS is a carbohydrate polymer in which the repeating unit (O unit) consists of a short oligosaccharide or sometimes a single sugar. OPS is synthesized template-free by glycosyltransferases and is found in lengths that vary from one repeat unit to several hundred sugars.

The biological roles of OPS vary from species to species, but for bacterial pathogens, it commonly confers resistance to the bactericidal effects of host complement. The mechanisms of resistance can vary, so that for example, the OPS of *Bordetella bronchiseptica* prevents activation of complement (2) whereas the OPS of *Salmonella*

enterica sv. Montevideo induces complement activation, but presumably at a safe distance from the bacterial membrane (3).

OPS is normally produced with a range of lengths that results in a modal distribution, i.e., a large proportion of the OPS has a length close to that of a modal value (Fig. 1A). The OPS length distribution can be critical for its biological function, for example it can affect resistance to host complement (4, 5) or the rate of invasion of, or uptake into, macrophages (5). In some species the modal OPS length can vary in response to environmental signals (6).

Most OPSs are produced by one of two mechanisms (reviewed in ref. 1). In the Wzy (polymerase)-dependent OPS biosynthesis pathway, each O repeat unit is assembled on an undecaprenyl-pyrophosphate lipid carrier at the cytoplasmic face of the inner membrane. Individual units are transported across this membrane and assembly of the OPS is catalyzed in the periplasm by the polymerase (7). The O chain is subsequently ligated to lipid A-core by the OPS ligase (8). In this pathway, chain-length modalities are conferred by the action of the polysaccharide copolymerase protein, Wzz. Authentic chain-length modalities are established *in vitro* from synthetic undecaprenyl-pyrophosphate-linked repeat units and lipid-linked substrates, using purified Wzz and the polymerase, Wzy (7), although direct biochemical evidence for the *in vivo* interaction of these proteins is lacking.

The second major OPS biosynthesis mechanism is called the ATP binding cassette (ABC) transporter-dependent pathway. In these systems, the OPS chain is completely assembled on the cytoplasmic face of the inner membrane on a single undecaprenyl-pyrophosphate lipid carrier. Polymerization involves the

Significance

Lipopolysaccharides (LPSs) are unique glycolipids that are characteristic components of outer membranes in Gram negative bacteria and play important roles in pathogenesis. Many LPSs contain a long-chain polysaccharide (known as the O antigen) whose length can be an important factor in bacterial resistance to complement-mediated killing. While components involved in chain-length determination are known in many systems, the underlying regulatory mechanism is not. Here we apply a mathematical modeling approach that integrates the existing structural and biochemical data to develop a new model, variable geometry, for chain-length regulation using the prototype for O antigens whose synthesis involves the widespread ATP-binding cassette transporter-dependent pathway.

Author contributions: J.D.K., S.B., B.R.C., R.J.M., and C.W. designed research; J.D.K., S.B., B.R.C., and R.J.M. performed research; J.D.K., S.B., B.R.C., R.J.M., and C.W. contributed new reagents/analytic tools; J.D.K., S.B., B.R.C., R.J.M., and C.W. analyzed data; and J.D.K., S.B., R.J.M., and C.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹J.D.K. and S.B. contributed equally to this work.

²To whom correspondence should be addressed. E-mail: cwhitfie@uoguelph.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1400814111/-DCSupplemental.

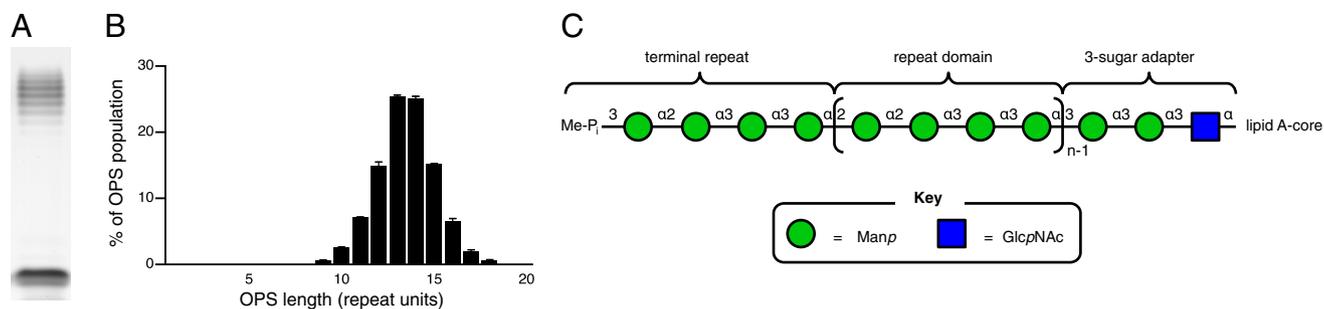


Fig. 1. Size and structure of *E. coli* O9a O polysaccharide. (A) Silver-stained tricine SDS/PAGE analysis of *E. coli* O9a LPS. The lower band is unsubstituted lipid A-core, whereas the upper cluster of bands shows LPS with different lengths of OPS. (B) The O9a OPS chain length distribution is unimodal with nearly half the chains having 13 or 14 repeat units. This distribution was measured from an autoradiogram of radiolabeled LPS (SI Text). (C) The repeat domain of O9a OPS is linked to lipid A-core via a three-sugar adapter. The terminal O repeat is modified by the addition of a methylphosphate moiety at the nonreducing terminus.

stepwise addition of single sugars, which is catalyzed by one or more glycosyltransferases (1). Then the OPS chain is transported to the periplasmic face of the inner membrane by an ABC transporter, where it is ligated to lipid A-core (9). In this pathway, there is no equivalent of the Wzz protein to control the polymer chain length, yet modality is still established.

The prototype for the ABC transporter-dependent pathway is the OPS from *Escherichia coli* O9a. This OPS has a four-sugar repeat unit containing $\alpha 1 \rightarrow 2$, and $\alpha 1 \rightarrow 3$ -linked D-mannose (Fig. 1C) (10) with a unimodal chain-length distribution (Fig. 1B). Biosynthesis of the repeating domain of this OPS is catalyzed by WbdA, a chain extension enzyme containing two glycosyltransferase domains (11). The nonreducing terminus of the OPS is modified by addition of a methyl phosphate (12), which is catalyzed by the bifunctional kinase–methyl transferase, WbdD (13, 14). The methyl transfer reaction is dependent on prior phosphorylation, and addition of the phosphate prevents further extension of the chain (15). After termination of the chain, the polysaccharide becomes a substrate for the OPS-export ABC transporter, which translocates the OPS to the periplasmic face of the inner membrane, where it is ligated to lipid A-core. The nucleotide binding component (Wzt) of this ABC transporter possesses a carbohydrate-binding module as an accessory cytoplasmic domain, which specifically binds OPS chains with the methyl-phosphate modification (16). Without these modifications the OPS is not exported (15).

A region in the C-terminal domain of WbdD localizes the protein to the membrane, and also interacts with WbdA. Without this interaction, WbdA is not correctly localized and OPS is not synthesized (17). The WbdD C-terminal domain contains coiled-coil motifs (15), and based on soluble, truncated forms of the protein, these sequences mediate the formation of WbdD trimers in protein crystals and in solution (14). Thus, WbdA and WbdD form a complex that catalyzes chain extension and chain termination, the chain length presumably being controlled by coordination of these activities. For example, overexpression of WbdD (15) or reduction in WbdA expression (18) both reduce the modal chain length. This report describes our investigations into the mechanism by which these enzymes' activities are coordinated to synthesize OPS with a highly reproducible and characteristic length distribution.

Theoretical work has shown that modal chain-length distributions cannot be generated by simple enzyme kinetic models without chain-length dependence in elongation or termination rates (19, 20). There are two main hypotheses for how such length dependence arises: by a “molecular clock,” which allows polymerization to continue for a defined amount of time before termination and export (21); or by a “molecular ruler,” which terminates chains once they reach a certain length (22). These models were established for systems involving Wzz and Wzy but the general principles apply to other polymerization mechanisms.

For the *E. coli* O9a system, the molecular clock has been formulated in terms of a conformational switching of the complex from OPS elongation to termination in response to a stochastic binding event, i.e., a timed event, whereas the molecular ruler has been hypothesized to act by spatial arrangement of the different catalytic sites such that the OPS is measured out by the physical dimensions of the biosynthetic complex (14).

In this study we show that *E. coli* O9a OPS length distribution can be quantitatively modulated by titrating the expression of either of the biosynthetic enzymes, WbdD and WbdA. This finding has strong implications for the underlying mechanism of length control and we show that previously proposed mechanisms of length control cannot generate the distributions we observe. However, by extending the molecular-ruler-type model to allow biosynthetic complexes with variable stoichiometry, we demonstrate that a variable geometry model can quantitatively explain all of our experimental data.

Results

OPS Chain Lengths Increase with WbdA Concentration. It was previously observed that the OPS modal chain length decreased when WbdA was expressed at low levels (18). To analyze this phenomenon in more detail, we sought to quantitatively control WbdA protein levels over a broad range and examine the effect on the OPS distribution. To this end, we deleted the chromosomal copy of *wbdA* and supplied a plasmid-encoded version of this gene cloned behind a tetracycline-inducible promoter. Tetracycline and presumably its analog, anhydrotetracycline (AhTc), diffuse passively across bacterial membranes (23) and thus this expression system is not subject to the all-or-nothing autocatalytic induction which may beset other systems (24, 25). We verified by Western blot that WbdA protein concentrations responded in a linear fashion to levels of AhTc in the growth medium, up to at least $5 \text{ ng}\cdot\text{mL}^{-1}$ (Fig. S1A).

In agreement with previous results (18), we found that the OPS modal chain length increased with increasing concentration of WbdA (Fig. 2A). With careful measurements of band intensity, we were also able to observe changes to the shape of the modal distributions as the concentration of WbdA was modulated. At intermediate expression levels, chain lengths exhibited the greatest variance (Fig. 2D), and at high expression levels, the distribution closely resembled that of wild-type *E. coli* O9a (Fig. 2C). Despite being able to decrease the modal chain length from the wild-type distribution by changing WbdA concentration, we found that we were unable to increase the chain length further beyond the wild-type modal length. This is not due to saturation of the inducible expression system because WbdA expression responded linearly to inducer concentration at $2.5 \text{ ng}\cdot\text{mL}^{-1}$, whereas the modal chain length did not increase when inducer concentration was further quadrupled to $10 \text{ ng}\cdot\text{mL}^{-1}$. That is, there appears to be a maximum modal chain length generated by

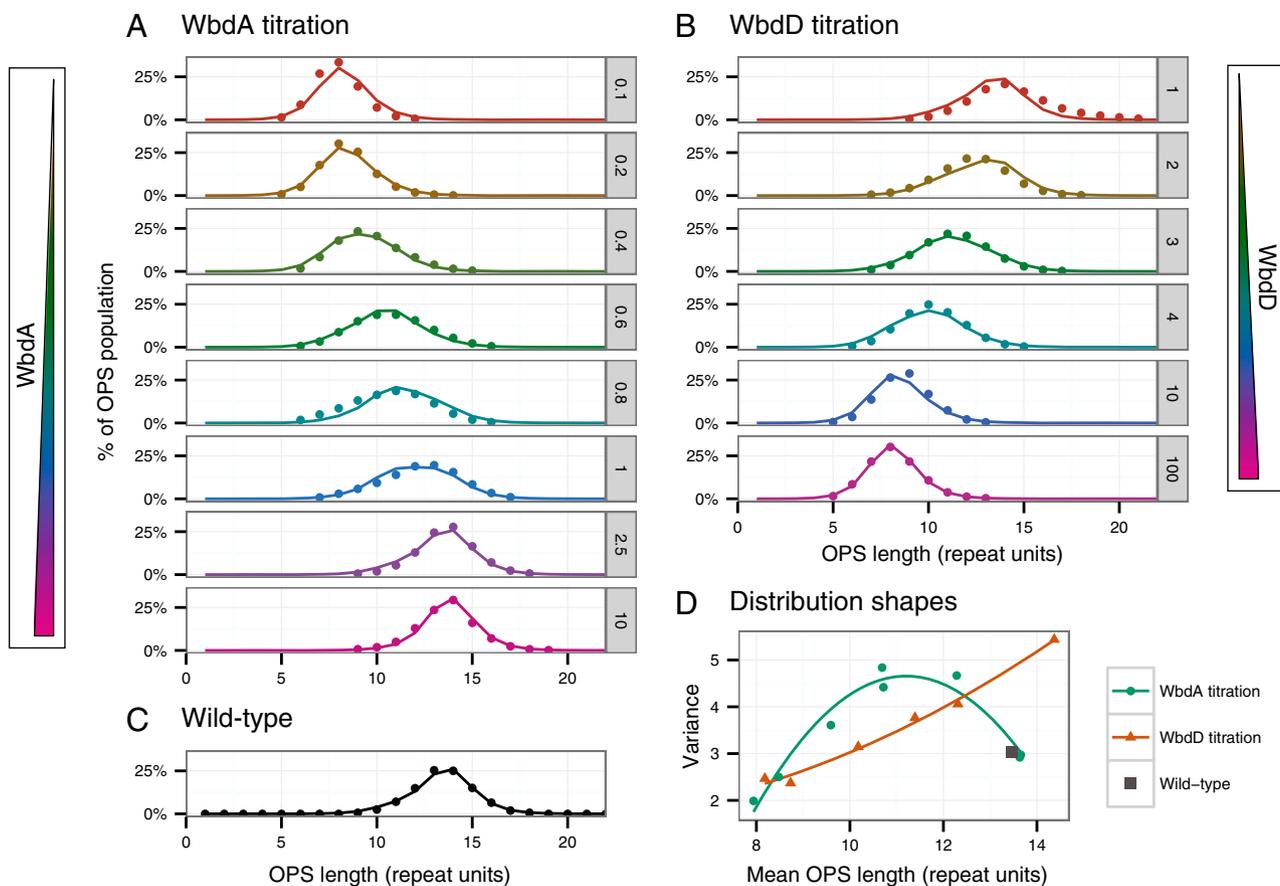


Fig. 2. Experimentally determined chain-length distributions (circles) for *E. coli* O9a expressing variable concentrations of (A) WbdA or (B) WbdD. Error bars representing SDs arising from the OPS measurement procedure were typically smaller than the data points and were therefore omitted for clarity. Results of mathematical model of OPS biosynthesis (lines) show the correct enzyme concentration dependence and quantitative agreement with the experimental data. Numbers in gray boxes indicate the concentration of AhtC used to induce protein expression. (C) Mathematical model trained on WbdA and WbdD titration data predicts the wild-type O9a OPS distribution. (D) Variance as a function of mean OPS length for each of the OPS length distributions shown in A and B. Solid lines are second-order polynomial trendlines for each titration series.

wild-type *E. coli* O9a that cannot be further increased by over-expression of WbdA.

OPS Chain Lengths Decrease with WbdD Concentration. To determine how WbdD concentration affects the final exported OPS distribution, we induced expression of WbdD in an *E. coli* O9a *wbdD::aacC1* strain. As previously described (17), this mutant expresses a C-terminal fragment of WbdD that contains domains necessary for recruitment of WbdA to the membrane but lacks kinase and methyltransferase activities. The effect of this mutation is that OPS can be expressed in this strain without expression of full-length WbdD, but is not terminated or exported. We verified that WbdD protein concentration responded to the inducer in a linear fashion up to 15 ng·mL⁻¹ (Fig. S1B).

Increasing the expression of WbdD reduced the modal chain length as reported (15) (Fig. 2B). Like the WbdA-induction experiment, we observed that the OPS distribution again responded quantitatively to the level of WbdD supplied, but in contrast with the results with WbdA, the variance of the OPS length distributions became greater with increasing modal chain length (Fig. 2D). At low WbdD concentrations, the modal length could be made to match that of the wild-type but the chain-length distribution was more disperse than wild-type.

These experiments suggest that the relative concentration of WbdA and WbdD is critical in determining both the modal chain length and the shape of the distribution. We next examined

the implications of these data on proposed models of chain-length control.

A Molecular Clock Model Predicts Chain-Length Distributions with Large Variance. A potential mechanism for chain-length determination is a molecular clock. In a molecular clock mechanism, the biosynthetic machinery can exist in two states, one that facilitates chain elongation and one that facilitates chain termination. The protein complex enters the elongation state as polymerization starts and shifts to the termination state after a predetermined time (21). The length of terminated polymers depends both on the elongation rate and the time for which elongation is allowed to proceed.

We have shown that increasing WbdA (or decreasing WbdD) generates longer chains. For a molecular clock model to fit these data, increasing the ratio of WbdA to WbdD must either increase the average elongation rate or increase the time spent in the elongation state. However, an enzymatic complex that elongates polymers at a rate independent of their length, for a certain duration, generates product lengths that follow a Poisson distribution (SI Text) (26). The OPS distributions observed in *E. coli* O9a are much less disperse than Poisson distributions (Fig. S2). Furthermore, unless all complexes within a cell are synchronized, the molecular clock model requires that each nascent polymer is elongated exclusively by a particular elongation–termination complex. That is, the polymer cannot dissociate from the biosynthetic complex during elongation. However, based on in vitro

behavior, WbdA catalyzes chain elongation with a mechanism that is distributive rather than processive (11), indicating that the nascent polymer can in fact dissociate and rebind the biosynthesis complex, at least in an *in vitro* scenario with a synthetic acceptor. Based on the non-Poissonian distribution and distributive nature of the polymerization process, we conclude that a pure molecular clock cannot produce the modal OPS distributions measured here for *E. coli* O9a.

Dimensions of Protein Complex Suggest a Geometric Model. A crystal structure of a 63-kDa fragment of WbdD, comprising residues 1–556 (of a total 708) revealed that the protein trimerizes around a three-helix bundle. This bundle contains the coiled-coil motifs from the WbdD C-terminal domain (14). This region also mediates interaction with WbdA (17). The WbdD-trimer is a triangular-based pyramid with sides of ~10 nm. When bound by one or more WbdA proteins (91 kDa), the size of the complex may be comparable to the length of the OPS chain: 12–21 nm corresponds to 8–14 O repeats. [Based on estimates of carbohydrate length made using the GLYCAM carbohydrate builder: Woods Group (2005–2013) GLYCAM Web. Complex Carbohydrate Research Center, University of Georgia, Athens, GA, www.glycam.com.]

It is therefore feasible that the length of the polysaccharide is determined by measurement against a physical dimension of the biosynthetic complex.

Formulation of a Variable Geometry Model. To summarize our experimental results, the average length of molecules increases with the concentration of WbdA and decreases with the concentration of WbdD. This enzyme-concentration dependence suggests that the mechanism of length determination involves protein complexes in dynamic equilibrium. However, earlier theoretical work has shown that enzyme kinetic models that do not have chain-length dependence in either elongation or termination rates cannot generate modal chain-length distributions (19). The measurement of OPS chains against a physical dimension of a WbdA–WbdD complex could provide a mechanism to generate modality, but in this case the modal length cannot vary with enzyme concentration unless the geometry of the complex also changes with enzyme concentration. To reconcile both enzyme concentration dependence and modality, we propose a mechanism by which alterations in protein expression levels result in incomplete formation of the WbdA–WbdD multienzyme complex. This then changes the physical dimensions against which the OPS chain is measured. We refer to this as a variable geometry model.

We hypothesize that WbdD forms trimers *in vivo*. The normal WbdD:WbdA stoichiometry is unknown, but we assume for simplicity that a single molecule of WbdA interacts with each WbdD monomer (3:3). When protein expression levels are altered, then A_3D_3 (i.e., WbdA₃–WbdD₃), A_2D_3 , AD_3 , and D_3 complexes exist in equilibrium with each other (Fig. 3A). The A_3D_3 , A_2D_3 , and AD_3 complexes each have different preferences for OPS substrate length. The fewer WbdA molecules bound in the complex, the greater the propensity to terminate shorter OPS chains.

To determine if the above model could quantitatively explain our data we turned again to mathematical modeling. We first constructed a kinetic model of the OPS polymerization process (Fig. S3). This allowed us to calculate the OPS length distribution from the chain-length preference of a biosynthetic complex, described in terms of termination probabilities for nascent polymers of different lengths. We then modeled WbdA–WbdD complex formation based on the idea that WbdA binds to preformed WbdD trimers, and allowed each WbdA–WbdD complex to terminate chains according to a geometric “length-measurement” model. Protein complexes with different stoichiometries each have an independent chain-length preference. This allowed us to calculate a predicted OPS distribution as a function of the concentrations of

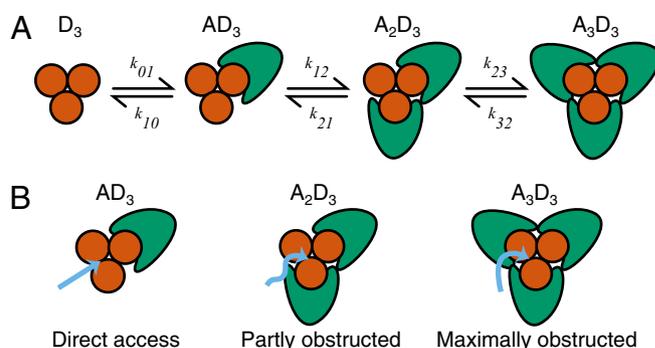


Fig. 3. (A) Model for assembly of the WbdA–WbdD complex. Up to three WbdA molecules (green) bind sequentially to a preformed WbdD trimer (orange circles). (B) Hypothetical protein complex arrangement in which binding of successive WbdA monomers progressively obstructs access to the chain termination active site on WbdD.

WbdA and WbdD. Finally, we optimized the model parameters to fit the experimental data to explore how well a variable stoichiometry model could quantitatively reproduce the experimental data. Full details are provided in *SI Text*.

The model of OPS polymerization is based on the following assumptions: (i) OPS size distribution is independent of time. (ii) OPS termination is irreversible and precedes export. (iii) Elongation and termination rates are independent of time but depend on nascent polymer length, and (iv) Export and degradation of polymers are independent of polymer length and time. There is currently no evidence for actual degradation of OPS molecules; however, the degradation terms in our model also incorporate dilution of OPS on the outer membrane due to cell growth. Under these assumptions, it can be shown (*SI Text*) that the distribution of terminated OPS lengths is given by

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

where \tilde{q}_i is the probability of terminating a nascent polymer of length i , given that such a nascent polymer exists.

We assume that the OPS is anchored in such a way that there is a negligible probability of terminating an OPS before the nascent chain has grown long enough to reach the termination site, and that once it is long enough it can be terminated. In terms of termination probability as a function of chain length, the simplest formulation of this would be as a step function that changes from zero to 1 at the cutoff length. We expect, however, that the enzyme complex will be less precise than this due to stochastic polymer bending, semiprocessive interactions with the polymerase, or possible shape changes of the protein complex itself. Consequently, we model the termination probability using a sigmoidal function (Fig. S4),

$$\tilde{q}_i = \frac{\lambda}{1 + e^{-\theta(i-l)}}. \quad [2]$$

The chain length preference is determined here by three parameters: l , the “length;” θ , the “precision;” and λ , the maximal probability of termination. We model the termination probability functions for the different complexes A_3D_3 , A_2D_3 , and AD_3 using sigmoidal functions with different characteristic length parameters l_{AD_3} , $l_{A_2D_3}$, and $l_{A_3D_3}$.

Using mass-action kinetics, we also formulate a model of WbdA–WbdD complex formation to allow determination of the concentrations of each molecular species given initial total protein concentrations of WbdA and WbdD (*SI Text*, Fig. S5).

For this implementation, we assume that the stoichiometry of WbdA:WbdD is 1:1 and that WbdD forms stable trimers *in vivo*. More complex stoichiometries could be accommodated in this model but without additional data there is no compelling reason to do so. The essential element in this type of model is that complexes with more molecules of WbdA bound tend to produce longer OPS chains.

The overall termination probability for the combined population of WbdA–WbdD complexes is modeled as an arithmetic combination of the three sigmoidal curves weighted in proportion to the concentrations of each complex type. The resulting OPS distribution is then determined from the ensemble termination probability using Eq. 1.

Parameter Optimization. We fit unconstrained model parameters to the entire WbdA- and WbdD-titration data (details in *SI Text* and *Table S1*). Unlike the molecular clock or simple geometric model, our variable stoichiometry model captures both the concentration dependence and the modality of the OPS distributions. The fitted model is shown in Fig. 2A–C.

As previously noted, the variance of the OPS length distribution was much larger for the WbdD- than the WbdA-induction series for long mean chain lengths (Fig. 2D). However, for a given set of parameters, the model predicts a single chain-length distribution for a given concentration of WbdA and WbdD. Therefore, we found that the set of parameters described above was insufficient to simultaneously fit the data from both experiments. A good fit to these data was attained by introducing an extra parameter θ' , which changes the precision of chain-length dependence in termination probabilities for A_2D_3 and A_3D_3 complexes in the WbdD expression experiments. This parameter enables the model to reproduce the variances observed in this set of data. Therefore, with the exception of θ' , all other parameters were unchanged between the two datasets.

Discussion

We have presented a model for chain-length regulation that is able to accurately reproduce our experimentally measured OPS length distributions. We now consider features of this successful model to see if they can be interpreted in biologically meaningful ways. One of the advantages of using mathematical modeling is that assumptions are made explicit and can therefore be critically evaluated. Although our model contains the assumption that WbdA- and WbdD-catalyzed reaction rates are independent of time, we also examined the case in which reaction rates do depend on time (i.e., the molecular clock) and found that such a model cannot account for the OPS length distributions we observe.

We also made the assumption that reaction probabilities depend on the OPS chain length, which has previously been shown to be essential for generating modal chain-length distributions (19). We modeled this dependence using sigmoidal curves for termination probability as a function of chain length, to mimic a simple geometric termination mechanism. For a protein complex to produce these curves, either or both of the reactions (extension or termination) may depend on chain length. The sharpness of transition from low to high termination probabilities exhibited by each type of biosynthetic complex can be plausibly explained if we hypothesize that the nonreducing terminus (i.e., the growing end) of the nascent OPS chain is constrained close to the chain extension active site but further from the site of chain termination. This may be accomplished by binding of the nonreducing terminus at a specific binding site on the complex. Alternatively, the undecaprenyl lipid anchor upon which the OPS is assembled may be constrained within two dimensions by association with the inner membrane.

To explain the enzyme-concentration dependence of the chain-length distribution, WbdA–WbdD complexes that contain fewer WbdA molecules must result in termination preferences

for shorter OPS chains. It is possible to conceive an arrangement of the multienzyme complex such that binding of the second and then third WbdA molecules progressively obstructs the shortest through-space path to the OPS chain termination active site (Fig. 3B). Complexes with more molecules of WbdA bound would have a higher probability of termination through a maximally or partly obstructed path and would therefore tend to make longer chains. Conversely, complexes with fewer molecules of WbdA bound would allow direct access of the nascent polymer to the chain termination active site on WbdD and therefore tend to make shorter chains. It may be possible to verify or disprove these hypotheses by obtaining 3D structural data on the WbdD–WbdA multienzyme complex. As previously noted, the maximum modal length in the WbdA-induction experiments did not increase beyond the wild-type modal length, despite substantial increases in WbdA concentration. In the context of our model, this suggests that in the wild-type O9a, WbdD trimers are saturated with WbdA and consequently terminate O polysaccharides via a maximally obstructed path. In this scenario, further increases in WbdA concentration in our experiments simply increase the amount of WbdA in the cytoplasm and have very little effect on the relative concentrations of WbdA–WbdD complexes (Fig. S6), and the OPS modal length is determined by the geometry of the WbdA-saturated complex.

It was necessary to introduce an additional parameter to quantitatively reproduce the chain-length distributions in the WbdD titration series. After fitting this extra parameter, the termination probabilities curves for the A_3D_3 , and A_2D_3 , complexes were flatter than the equivalent complexes in the WbdA induction series. What might this extra parameter represent, biologically? One feature of the mutant strain in which these particular measurements were made is that the chromosome encodes a catalytically inactive fragment of WbdD (WbdD*), which is sufficient for membrane localization of WbdA (17). It seems plausible that when low levels of full-length WbdD were expressed from the plasmid, a number of the biosynthetic WbdA–WbdD complexes contained one or more WbdD* molecules. Such complexes therefore contain a subunit that is not capable of chain termination, which would result in generation of longer chains on average. This idea fits well with the flatter termination probability curves for A_3D_3 , and A_2D_3 complexes that were required for model fitting.

In summary, we have formulated a model to explain the molecular mechanism of chain-length control in OPS biosynthesis in *E. coli* O9a. We have shown that a molecular clock cannot account for observed chain-length distributions and that a pure geometric molecular ruler model cannot account for OPS length distributions that depend on enzyme concentration. Instead, our variable geometry model suggests that OPS chain length is determined by the spatial separation between the undecaprenyl lipid anchor binding point and the chain termination active site. The model suggests that incomplete formation of the multienzyme complex allows alternative (shorter) pathways for the nascent OPS chains to wrap around the complex and reach the termination site. This allows the generation of modal OPS distributions that depend on enzyme concentration.

The *wbdD* and *wbdA* coding sequences are adjacent on the *E. coli* O9a chromosome, and are probably expressed from the same upstream promoter on a single transcript. Furthermore, to the extent that it has been tested, the OPS length in wild-type *E. coli* O9a is unaffected by growth phase, medium composition, temperature, or any other environmental conditions. There is therefore no reason to suppose that differential control of WbdD or WbdA protein expression levels is a mechanism that these bacteria use to modulate OPS length. Rather, it seems that the enzyme concentration dependence is an artifact of genetic manipulation. Nevertheless, these phenomena must be explained by any successful model of the system. Our model incorporates all

current data on this system and suggests several hypotheses which will be testable using structural biology.

The ABC transporter-dependent polysaccharide biosynthesis pathway is widely distributed in prokaryotes. It is not known what proportion of these pathways features a nonreducing terminal modification as in *E. coli* O9a. Whereas some definitely lack such terminal modifications, e.g., *K. pneumoniae* O2a (27), a number of different terminal modifications have been described among O polysaccharides and S-layer glycans of various bacterial species (reviewed in ref. 28). Because terminal modifications are often missed in polysaccharide structural studies, which tend to focus on the repeating units, this class of polymer is potentially more common than the current databases would suggest. Therefore, our model for OPS chain-length regulation in *E. coli* O9a has implications beyond this species and may describe the mechanisms for chain-length regulation in many biologically important polysaccharides. More generally we have introduced a new concept to the field of length control in template-independent polymerization, to explain how the length of carbohydrate polymers can be tightly controlled by a geometric measurement mechanism while remaining sensitive to the concentration of the biosynthetic enzymes involved through the formation of multisubunit complexes with variable stoichiometry.

The variable geometry model is not necessarily applicable to other types of polysaccharide biosynthesis system, for example, polymerase (Wzy)-dependent O polysaccharide biosynthesis, which lacks an analogous chain-termination reaction. However, in the process of modeling the *E. coli* O9a system we drew two conclusions about the contexts in which the molecular ruler and molecular clock mechanisms can be invoked that will be useful in considering length regulation in general. First, a molecular ruler mechanism cannot regulate length in a manner that depends on biosynthetic enzyme concentrations. Second, a molecular clock

mechanism cannot give rise to polymer length distributions that are less diffuse than a Poisson distribution. These conclusions are generally applicable, regardless of the polymer or biosynthetic machinery, and to our knowledge have not been clearly stated before.

Materials and Methods

Strains, Plasmids and Culture Conditions. C-terminally His₁₀-tagged WbdA (WbdA-His₁₀) and N-terminally His₆-tagged WbdD (His₆-WbdD) were expressed from tetracycline-inducible promoters using 0.1–100 ng·mL⁻¹ of the tetracycline analog AnTc. Further details are given in *SI Text* and *Table S2*.

Measurement of OPS Length Distributions. The measurement procedure is described fully in *SI Text* and *Fig. S7*. Briefly, whole bacterial cells were radiolabeled by growth in 5 mL of low-phosphate medium containing H₃³²PO₄ (Perkin-Elmer). After incubation at 37 °C with shaking for 6–8 h, cells were harvested. LPS was prepared using a modification of the method of Yi and Hackett (29), in which all other phosphate-containing macromolecules are removed or destroyed by extraction with TRIzol reagent (Invitrogen) and chloroform, followed by nuclease and proteinase digestions. LPS was analyzed by tricine SDS/PAGE analysis in 80-mm-long resolving gels. Gels were dried and analyzed with a phosphorimager (Bio-Rad Personal Molecular Imager FX). Because the number of phosphates in each LPS molecule is independent of OPS chain length, the intensity of LPS bands bears a linear relationship with the number of molecules of that size. Despite extensive SDS/PAGE optimization, not all LPS bands could be perfectly resolved. Therefore, densitometry was aided by use of the Multipeak fitting 2 package in Igor Pro-6.21 (WaveMetrics).

Mathematical Modeling. Mathematical Modeling is described in *SI Text*.

ACKNOWLEDGMENTS. We thank Leslie Cuthbertson for advice on the Tet promoter. C.W. acknowledges funding from the Natural Sciences and Engineering Research Council of Canada and is the recipient of a Canada Research Chair. S.B. acknowledges funding from the John Innes Foundation.

- Raetz CR, Whitfield C (2002) Lipopolysaccharide endotoxins. *Annu Rev Biochem* 71:635–700.
- Burns VC, Pishko EJ, Preston A, Maskell DJ, Harvill ET (2003) Role of *Bordetella* O antigen in respiratory tract infection. *Infect Immun* 71(1):86–94.
- Joiner KA, Grossman N, Schmetz M, Leive L (1986) C3 binds preferentially to long-chain lipopolysaccharide during alternative pathway activation by *Salmonella montevideo*. *J Immunol* 136(2):710–715.
- Burns SM, Hull SI (1998) Comparison of loss of serum resistance by defined lipopolysaccharide mutants and an acapsular mutant of uropathogenic *Escherichia coli* O75:K5. *Infect Immun* 66(9):4244–4253.
- Murray GL, Attridge SR, Morona R (2006) Altering the length of the lipopolysaccharide O antigen has an impact on the interaction of *Salmonella enterica* serovar Typhimurium with macrophages and complement. *J Bacteriol* 188(7):2735–2739.
- May JF, Groisman EA (2013) Conflicting roles for a cell surface modification in *Salmonella*. *Mol Microbiol* 88(5):970–983.
- Woodward R, et al. (2010) *In vitro* bacterial polysaccharide biosynthesis: Defining the functions of Wzy and Wzz. *Nat Chem Biol* 6(6):418–423.
- Han W, et al. (2012) Defining function of lipopolysaccharide O-antigen ligase WaaL using chemoenzymatically synthesized substrates. *J Biol Chem* 287(8):5357–5365.
- Greenfield LK, Whitfield C (2012) Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. *Carbohydr Res* 356:12–24.
- Parolis LA, Parolis H, Dutton GG (1986) Structural studies of the O-antigen polysaccharide of *Escherichia coli* O9a. *Carbohydr Res* 155:272–276.
- Greenfield LK, et al. (2012) Biosynthesis of the polymannose lipopolysaccharide O-antigens from *Escherichia coli* serotypes O8 and O9a requires a unique combination of single- and multiple-active site mannosyltransferases. *J Biol Chem* 287(42):35078–35091.
- Kubler-Kielb J, Whitfield C, Katzenellenbogen E, Vinogradov E (2012) Identification of the methyl phosphate substituent at the non-reducing terminal mannose residue of the O-specific polysaccharides of *Klebsiella pneumoniae* O3, *Hafnia alvei* PCM 1223 and *Escherichia coli* O9/O9a LPS. *Carbohydr Res* 347(1):186–188.
- Clarke BR, et al. (2011) *In vitro* reconstruction of the chain termination reaction in biosynthesis of the *Escherichia coli* O9a O-polysaccharide: The chain-length regulator, WbdD, catalyzes the addition of methyl phosphate to the non-reducing terminus of the growing glycan. *J Biol Chem* 286(48):41391–41401.
- Hagelueken G, et al. (2012) Structure of WbdD: A bifunctional kinase and methyltransferase that regulates the chain length of the O antigen in *Escherichia coli* O9a. *Mol Microbiol* 86(3):730–742.
- 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.
- Cuthbertson L, Kimber MS, Whitfield C (2007) Substrate binding by a bacterial ABC transporter involved in polysaccharide export. *Proc Natl Acad Sci USA* 104(49):19529–19534.
- 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.
- Greenfield LK, et al. (2012) Domain organization of the polymerizing mannosyltransferases involved in synthesis of the *Escherichia coli* O8 and O9a lipopolysaccharide O-antigens. *J Biol Chem* 287(45):38135–38149.
- Hunt F (1985) Patterns of LPS synthesis in gram negative bacteria. *J Theor Biol* 115(2):213–219.
- Goldman RC, Hunt F (1990) Mechanism of O-antigen distribution in lipopolysaccharide. *J Bacteriol* 172(9):5352–5359.
- Bastin DA, Stevenson G, Brown PK, Haase A, Reeves PR (1993) Repeat unit polysaccharides of bacteria: A model for polymerization resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. *Mol Microbiol* 7(5):725–734.
- Tocilj A, et al. (2008) Bacterial polysaccharide co-polymerases share a common framework for control of polymer length. *Nat Struct Mol Biol* 15(2):130–138.
- Schnappinger D, Hillen W (1996) Tetracyclines: Antibiotic action, uptake, and resistance mechanisms. *Arch Microbiol* 165(6):359–369.
- Morgan-Kiss RM, Wadler C, Cronan JE, Jr. (2002) Long-term and homogeneous regulation of the *Escherichia coli* araBAD promoter by use of a lactose transporter of relaxed specificity. *Proc Natl Acad Sci USA* 99(11):7373–7377.
- Novick A, Weiner M (1957) Enzyme induction as an all-or-none phenomenon. *Proc Natl Acad Sci USA* 43(7):553–566.
- Flory P (1940) Molecular size distribution in ethylene oxide polymers. *J Am Chem Soc* 62(6):1561–1565.
- Kos V, Cuthbertson L, Whitfield C (2009) The *Klebsiella pneumoniae* O2a antigen defines a second mechanism for O antigen ATP-binding cassette transporters. *J Biol Chem* 284(5):2947–2956.
- Cuthbertson L, Kos V, Whitfield C (2010) ABC transporters involved in export of cell surface glycoconjugates. *Microbiol Mol Biol Rev* 74(3):341–362.
- Yi EC, Hackett M (2000) Rapid isolation method for lipopolysaccharide and lipid A from gram-negative bacteria. *Analyst (Lond)* 125(4):651–656.