Domain-swap polymerization drives the self-assembly of the bacterial flagellar motor

Matthew A B Baker[1,2], Robert M G Hynson[1], Lorraine A Ganuelas[1], Nasim Shah Mohammadi[1], Chu Wai Liew[1], Anthony A Rey[1], Anthony P Duff[3], Andrew E Whitten[3], Cy M Jeffries[3,7], Nicolas J Delalez[4], Yusuke V Morimoto[5], Daniela Stock[1], Judith P Armitage[4], Andrew J Turberfield[6], Keiichi Namba[5], Richard M Berry[6] & Lawrence K Lee[1,2]

Large protein complexes assemble spontaneously, yet their subunits do not prematurely form unwanted aggregates. This paradox is epitomized in the bacterial flagellar motor, a sophisticated rotary motor and sensory switch consisting of hundreds of subunits. Here we demonstrate that _Escherichia coli_ FliG, one of the earliest-assembling flagellar motor proteins, forms ordered ring structures via domain-swap polymerization, which in other proteins has been associated with uncontrolled and deleterious protein aggregation. Solution structural data, in combination with _in vivo_ biochemical cross-linking experiments and evolutionary covariance analysis, revealed that FliG exists predominantly as a monomer in solution but only as domain-swapped polymers in assembled flagellar motors. We propose a general structural and thermodynamic model for self-assembly, in which a structural template controls assembly and shapes polymer formation into rings.

During the assembly of the bacterial flagellar motor (BFM), rotor proteins assemble into large coaxial rings. This process requires tight control to coordinate ring formation and prevent aggregation. Motor construction begins with the self-assembly of dozens of copies of the protein FliF into an ~30-nm-diameter ring (MS ring) embedded in the bacterial cytoplasmic membrane[1]. The MS ring acts as a structural scaffold that seeds the assembly of hundreds of other subunits, thus facilitating formation of the rest of the complex rotary motor. This scaffold comprises a circular array of protein-binding sites that are specific to FliG, the next protein to assemble. Two other proteins, FliM and FliN, assemble upon the FliG ring[1,2], thereby forming the cytoplasmic or C-ring rotor, which interacts with transmembrane stator complexes and converts ion flux through the stator into rotation. The C ring is also part of a type 3 secretory apparatus[3] and is an ultrasensitive and tunable sensory switch that allows the motor to change directions in response to environmental stimuli[4].

At cytoplasmic concentrations, FliG is monomeric yet is capable of forming multimeric ring structures during assembly at the site of the rotor. This property makes it a good model for understanding the mechanisms that control the assembly of monomeric proteins into robust protein superstructures such as the BFM. The crystal structure of FliG has been determined for several species with different truncations and functional mutations. FliG has distinct N-terminal (FliG$_N$), middle (FliG$_M$) and C-terminal (FliG$_C$) domains, and a repeated three-helical fold known as the Armadillo-repeat motif (ARM) occurs in FliG$_M$ (ARM$_M$) and FliG$_C$ (ARM$_C$). ARMs typically display pairs of complementary conserved hydrophobic surfaces that mediate their tandem stacking into right-handed superhelices[5,6]. However, whereas all other known ARMs are expressed as tandem repeats, the FliG ARMs are separated by approximately 30 amino acids, most of which (23 residues) form an α-helix that bridges the middle and C-terminal domains (helix$_{MC}$). In some crystal structures[7], the formation of this helix prevents intramolecular stacking, and FliG adopts an extended conformation (FliG$_{ext}$) with ARM$_M$ and ARM$_C$ domains from neighboring molecules stacked upon each other. The intermolecular ARM$_M$-ARM$_C$ interface at the heart of this domain-swap structural motif links FliG molecules and facilitates formation of polymers: this interface has been proposed as the basis of an atomic model of the FliG ring[7,8]. Other crystal structures have indicated that intramolecular stacking of the FliG ARMs is possible and results in a compact monomeric conformation[9,10]. Interaction studies involving electron spin resonance[8] have demonstrated that disruption of the ARM$_M$-ARM$_C$ interface interferes with FliM oligomerization, but these assays used concentrations well above the cytosolic monomer concentration, which is likely to be similar to that of FliM (~500 nM)[11]. Because crystal conformations are likely to also be influenced by crystal packing, the biological relevance of the ARM$_M$-ARM$_C$ interaction, including its role in the self-organization of the BFM, remains controversial[8,12,13].

To determine the molecular mechanism underlying motor assembly, we demonstrated the biological importance of the ARM$_M$-ARM$_C$ interaction by using evolutionary covariance analysis. We also

determined the solution structure of monomeric FliG by using small-angle X-ray scattering (SAXS) and used biochemical cross-linking experiments to demonstrate the role of the intermolecular ARM_M-ARM_C interaction in mediating the transition between monomeric protein and the stable self-assembled protein complex. These results allowed us to construct a simple model for the synthesis of a FliG ring around its FliF template, based on a domain-swap mechanism for polymerization. Domain-swap polymerization has been reported in pathogenic processes including amyloid formation in neurodegenerative disorders[14–17] and the aggregation of γ-crystallins in human cataracts[18]. Here we describe what is, to our knowledge, the first reported case in which domain-swap polymerization is initiated, shaped and terminated by means of a structural template, thus forming a well-defined supramolecular structure. We expect that this mechanism may be generalizable to other large multimeric protein complexes.

RESULTS

The ARM_M-ARM_C interaction is evolutionarily conserved

To determine the biological relevance of the ARM_M-ARM_C interaction, we performed covariance analysis on 2,292 unique FliG protein sequences. Covariance analysis can predict protein structure[19] and sites of protein-protein interactions[20] by identifying amino acid pairs that coevolve[21]. Our analysis identified 142 out of 48,217 possible amino acid pairs with correlation scores that were more than 5 s.d. above the mean for the remaining pairs (**Fig. 1a** and Online Methods). In the $FliG_{ext}$ crystal structure from *Aquifex aeolicus* (PDB 3HJL[7]), the correlated residues were in proximity to each other (**Fig. 1b,c**): 12.2% of

the correlated residues were in direct contact (<5 Å distance) and had key roles in the stabilization of the protein fold, through formation of the hydrophobic core of globular structural domains or through hydrogen-bonding networks (**Fig. 1d**). Among the covarying residue pairs identified, nine had one member in each of the ARM_M and ARM_C domains; these appeared as an outlying distribution with ~50-Å separation (**Fig. 1b**). However, when we measured separations across the intermolecular ARM_M-ARM_C interface in the *A. aeolicus* FliG crystal lattice, rather than between residues in a single molecule, these nine covarying pairs were in proximity (**Fig. 1b**), and five were in direct contact, forming a substantial part of the hydrophobic interface (**Fig. 1e**). A two-dimensional distance plot displaying all amino acid pairs closer than 7 Å in the extended monomer structure, with overlaid correlated covariance pairs, showed that the correlated pairs mapped well to those in proximity in this structure (**Fig. 1f**). The cluster of correlated pairs corresponding to the conserved ARM_M-ARM_C interaction did not map to the monomeric extended structure but did map to proximal residues in the intermolecular ARM_M-ARM_C interaction from the extended lattice (**Fig. 1f**). This analysis demonstrated that the ARM_M-ARM_C interaction observed in all FliG crystal structures[7,9,10,12,22] is evolutionarily conserved, thus providing additional evidence that this interaction is likely to be important in protein function.

The conformation of FliG in solution

The ARM_M-ARM_C interaction causes $FliG_M$ and $FliG_C$ to form a globular $FliG_{MC}$ domain (**Fig. 2**). In crystal structures of $FliG_{ext}$ from
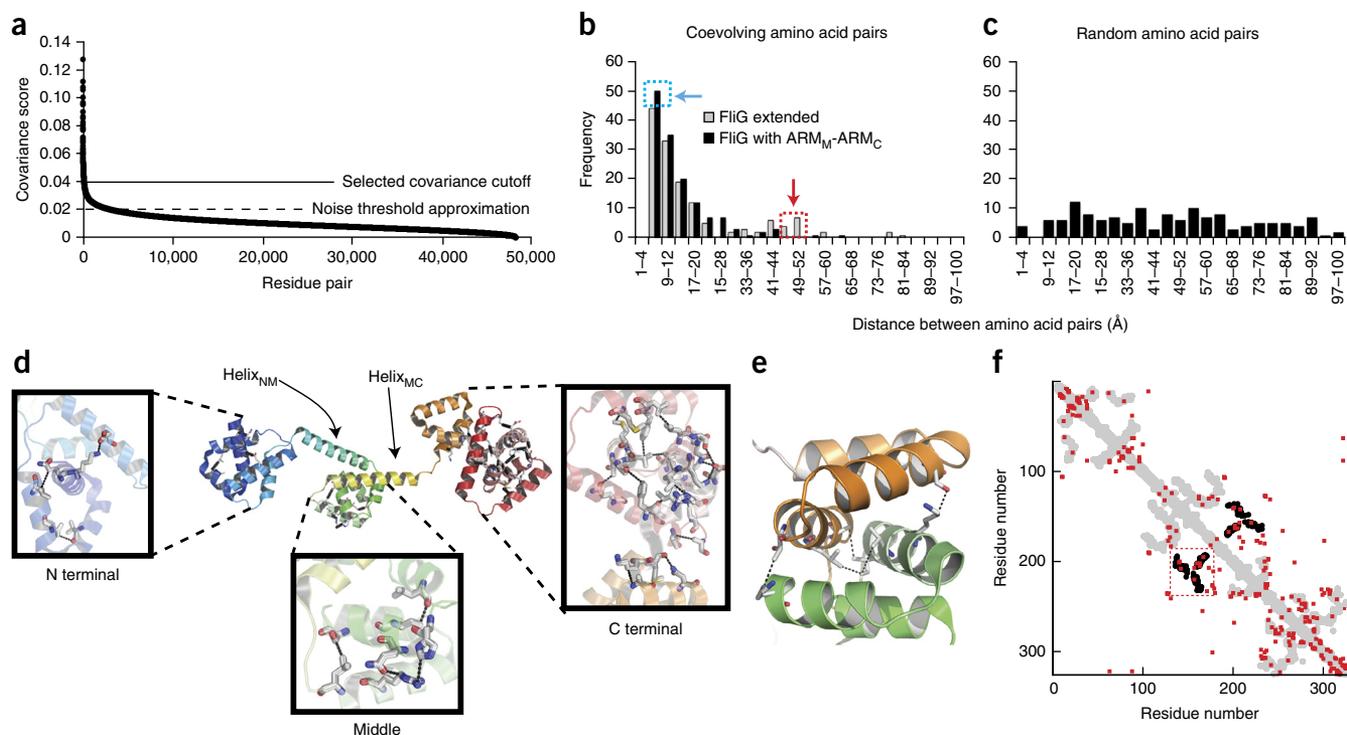
Figure 1 Evolutionary covariance of FliG. (**a**) Covariance scores of all amino acid pairs plotted in rank order. 142 residue pairs above a signal threshold (defined in Online Methods) were included in the analysis. (**b**) Histogram of intramolecular distances between these residue pairs in the crystal structure of a full-length FliG monomer from *A. aeolicus* (gray, PDB 3HJL[7]) and the corresponding histogram (black) allowing for shorter distances across the intermolecular ARM_M-ARM_C interface. (**c**) Control distribution using 150 random residue pairs. (**d**) Rendering of the *A. aeolicus* FliG crystal structure, showing the highly correlated covarying residue pairs involved in stabilizing the globular structure of each domain. (**e**) Enlarged view of the ARM_M-ARM_C superhelix, highlighting the locations of five out of the nine highly correlated pairs within the ARM_M-ARM_C superhelix that are in direct contact. (**f**) Two-dimensional distance plot showing all residue pairs closer than 7 Å in the FliG extended monomer (gray) and intermolecular ARM_M-ARM_C superhelix (black); residue pairs with high covariance scores are overlaid in red.
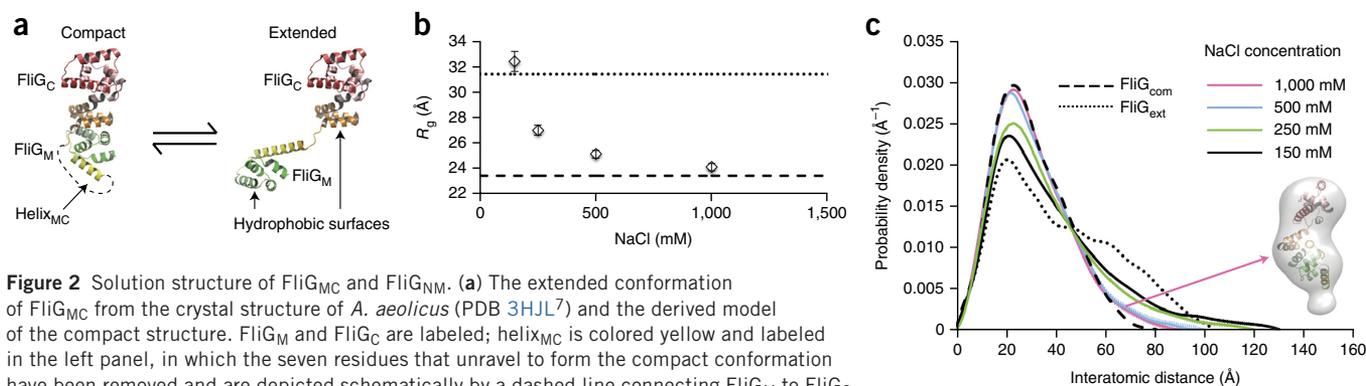
Figure 2 Solution structure of FliG$_{MC}$ and FliG$_{NM}$. (**a**) The extended conformation of FliG$_{MC}$ from the crystal structure of *A. aeolicus* (PDB 3HJL[7]) and the derived model of the compact structure. FliG$_M$ and FliG$_C$ are labeled; helix$_{MC}$ is colored yellow and labeled in the left panel, in which the seven residues that unravel to form the compact conformation have been removed and are depicted schematically by a dashed line connecting FliG$_M$ to FliG$_C$. (**b**) The R_g of FliG$_{MC}$, determined from SAXS, plotted against the NaCl concentration in the experimental buffer. The dashed and dotted lines mark the R_g values calculated for the compact and extended conformations of FliG$_{MC}$ shown in **a**, respectively. Error bars, 2× s.e.m. (**c**) Calculated $P(r)$ for FliG$_{MC}$ in compact and extended conformations (dashed and dotted lines respectively) with experimental data for different NaCl concentrations. The inset shows the average of 23 *ab initio* dummy-atom shape restorations of FliG$_{MC}$ from SAXS data collected in 1,000 mM NaCl, with the fitted crystal structure of FliG$_{com}$ superimposed.

A. aeolicus, Thermotoga maritima and *Helicobacter pylori*[7,22,23], this domain is formed between neighboring molecules in the crystal lattice (**Fig. 2a**). In contrast, in other crystal structures of FliG fragments from *T. maritima* and *H. pylori*[10,22], the ARM domains from the same molecule interact, and the monomer adopts a more compact conformation (FliG$_{com}$; modeled in **Fig. 2a**). Given that the three domains of FliG are connected by flexible linking peptides[7], their relative arrangement in crystals is almost certainly influenced by crystal packing, and it is not known how the protein behaves in solution. We therefore used SAXS to determine the conformational state of *E. coli* FliG in solution.

We collected data from the eluate of a size-exclusion chromatography column to exclude the possible influence of aggregates[24,25]: independent analysis by multiangle laser light scattering showed that the purified protein was monomeric in solution at the 10 mg/ml FliG concentrations used in all SAXS experiments (**Supplementary Fig. 1**). We further confirmed this by estimating molecular weight from Porod volumes (**Supplementary Table 1**), determining a consistent radius of gyration (R_g) and evaluating correlations between scattering profiles across elution peaks with correlation maps and reduced χ^2 analysis[26] (**Supplementary Fig. 2**). We compared measurements from SAXS experiments to the crystal structure of FliG$_{ext}$ from *A. aeolicus* (**Fig. 1d**; PDB 3HJL[7]), which is the only full-length structure available. We generated the model FliG$_{com}$ from the same structure by combining the N-terminal and middle domains (FliG$_{NM}$) of one molecule with FliG$_C$ from the neighboring protein in the crystal lattice, which forms an ARM$_M$-ARM$_C$ interaction with FliG$_{NM}$, as previously described[7]. However, for FliG to adopt a compact conformation, at least seven amino acids of an α-helix that bridges FliG$_M$ and FliG$_C$ (helix$_{MC}$; **Fig. 1d**) must unravel. These seven residues together with the five additional unstructured amino acids between helix$_{MC}$ and FliG$_C$ comprise an unstructured peptide of 12 amino acids, which provides sufficient interdomain flexibility to allow the intramolecular ARM$_M$-ARM$_C$ interaction[10,22]. The resulting loop is not resolved in crystal structures of truncated constructs containing intramolecular ARM$_M$-ARM$_C$ interfaces[9,10,27] and is not included in our model of FliG$_{com}$ (**Fig. 2a** and **Supplementary Fig. 3**). The R_g of the full-length FliG protein in solution in 150 mM NaCl was 37.2 ± 0.3 Å (mean ± 2× s.e.m.), a value closer to that of the FliG$_{ext}$ crystal structure (39.4 Å) than the FliG$_{com}$ model (27.5 Å). Extension of either or both of the flexible linkers between FliG$_N$ and FliG$_M$ and between FliG$_M$ and FliG$_C$ may account for this elongation in solution. To separate

the effects of these linkers, we characterized the solution shape of truncated FliG$_{NM}$ and FliG$_{MC}$ fragments (**Fig. 2b,c**, **Supplementary Fig. 2** and **Supplementary Table 1**). SAXS measurements indicated that FliG$_{NM}$ in solution in 150 mM NaCl ($R_g = 23.1 \pm 0.3$ Å) was more compact than the same fragment in the crystal structure ($R_g = 27.1$ Å), and the $P(r)$ plot showing the atom-pair distance distribution calculated from SAXS data revealed that in solution this construct formed a single domain (**Supplementary Fig. 4**). The R_g of FliG$_{MC}$ in solution decreased asymptotically with increasing ionic strength from 31.4 ± 0.3 Å in 150 mM NaCl to 24.1 ± 0.3 Å in 1 M NaCl (**Fig. 2b**), a value close to the 23.4 Å calculated for this fragment in the FliG$_{com}$ model. This result was consistent with higher ionic strength stabilizing the hydrophobic intramolecular ARM$_M$-ARM$_C$ interface and thus favoring the compact form[9,10]. $P(r)$ plots at 1 M NaCl were very similar to the theoretical distance distribution for the compact FliG$_{MC}$ model (**Fig. 2c**), thus indicating that in 1 M NaCl, FliG$_{MC}$ may be entirely in the compact state. Consistently with the trend in R_g (**Fig. 2b**), decreasing ionic strength shifted the distribution toward that of the FliG$_{ext}$ crystal structure (**Fig. 2c**). We aligned and averaged multiple independent *ab initio* shape restorations of FliG$_{MC}$ from SAXS data collected in 1 M NaCl, which produced a final shape that was consistent with the FliG$_{com}$ crystal structure (**Fig. 2c**), thus confirming that in 1 M NaCl, FliG$_{MC}$ is in the compact form.

To estimate the percentage of FliG$_{MC}$ in an extended conformation, we fitted experimental data collected at each NaCl concentration to a linear combination of the theoretical scattering functions $I_{com}(q)$ and $I_{ext}(q)$, calculated from the model structures FliG$_{com}$ and FliG$_{ext}$, respectively (**Supplementary Fig. 5**), thus yielding $I_{fit}(q)$:

$$I_{fit}(q) = f \times I_{ext}(q) + (1 - f) \times I_{com}(q)$$

where f is the fraction of FliG in an extended state. The motions of the two domains in FliG$_{ext}$ are restrained only by a linking peptide, so FliG$_{ext}$ is likely to adopt a continuum of states. There is therefore a large uncertainty in estimating f by fitting the SAXS data. The values of f were 48%, 35%, 7% and 1% in 150 mM, 250 mM, 500 mM and 1,000 mM NaCl, respectively. These results allowed us to draw the robust qualitative inference that f increases rapidly as the NaCl concentration increases.

The crystal structure of the full-length protein[7] has revealed that, in FliG$_{ext}$, half of helix$_{MC}$ is stabilized by contacts with FliG$_M$ and an α-helix that connects FliG$_N$ and FliG$_M$ (helix$_{NM}$; **Fig. 1d**). To determine
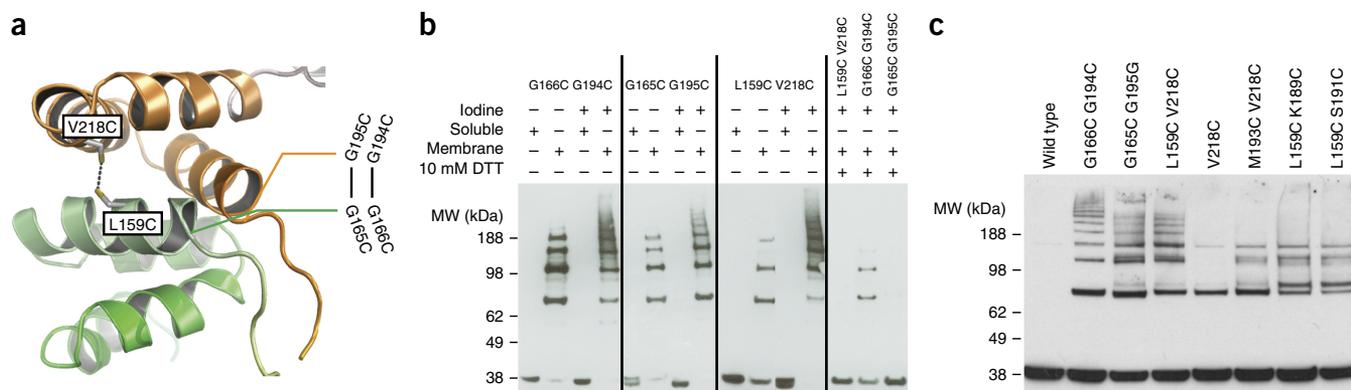
Figure 3 Cross-linking of the ARM$_M$-ARM$_C$ interface. (**a**) Rendering of the ARM$_M$-ARM$_C$ superhelix from the FliG crystal structure from *A. aeolicus* (PDB 3HJL[7]), showing the locations of three amino acid pairs (L159 V218, G166 G194 and G165 G195) which, if mutated to cysteine, are predicted to form disulfide bonds across the ARM$_M$-ARM$_C$ interface with close to ideal bond distances, angles and dihedral angles. (**b**) Anti-hemagglutinin immunoblots on denaturing SDS-PAGE with fractionated samples in the presence or absence of oxidizing agents iodine and DTT. Uncropped gels are shown in **Supplementary Figure 6**. The molecular weight (MW) of FliG is 37.9 kDa. (**c**) Immunoblots from whole-cell samples after disulfide cross-linking induced by copper–phenanthroline in live bacteria. Control disulfide mutants are V218C, M193C V218C, L159C K189C and L159C S191C.

whether helix$_{NM}$ might prevent full-length FliG from adopting a compact conformation, we also performed SAXS analysis of an expanded FliG$_{MC}$ construct including helix$_{NM}$ (FliG$_{MC-ext}$) (**Supplementary Fig. 2**). The R_g of FliG$_{MC-ext}$ also decreased with increasing NaCl concentration, thus demonstrating that, even in the presence of helix$_{NM}$, FliG adopts a more compact conformation at high ionic strengths (**Supplementary Table 1**). We concluded that formation of the intramolecular bond due to the hydrophobic intramolecular ARM interaction is likely to be reversible and that the equilibrium ratio of molecules in compact and extended states depends on ionic strength.

Targeted disulfide cross-linking of the ARM$_M$-ARM$_C$ interface

To determine the role of the ARM$_M$-ARM$_C$ interaction in functional flagellar motors, we covalently cross-linked the ARM$_M$ and ARM$_C$ domains by using targeted double-cysteine mutants. We rationally designed three double-cysteine mutations (L159C V218C, G164C G196C and G165C G195C) to form cross-links at the ARM$_M$-ARM$_C$ interface, on the basis of the crystal structure of FliG from *A. aeolicus*[7]. We replaced wild-type FliG with each mutant *in vivo* (**Fig. 3a**) and found that the mutations did not affect motor assembly, because flagellar filaments were visible under fluorescence microscopy, and cells were motile (**Supplementary Fig. 6a–f**). Because flagellar-motor basal bodies remain intact after cell fractionation and purification[28], we separated the soluble and membrane fractions of lysed cells, which contain freely diffusing and motor-associated FliG protein, respectively. We assessed cross-linking through SDS-PAGE and immunoblotting. FliG oligomers were detectable in the membrane fraction but not in the soluble fraction for all double-cysteine FliG mutations tested. The concentration of oligomers was dramatically reduced in the presence of the reducing agent DTT, thus confirming that oligomer formation was mediated by disulfide-bond formation (**Fig. 3b**). The cytoplasm is a reducing environment; therefore, we added iodine between cell lysis and fractionation to promote disulfide-bond formation and to maximize the probability of locking FliG proteins into the conformation characteristic of assembled flagellar motors. Under these oxidizing conditions, for all mutations, monomeric protein entirely disappeared from the membrane fraction, and we detected only higher-order oligomeric species. In contrast, we observed only monomeric protein in the soluble fraction (**Fig. 3b**). The presence of iodine increased the electrophoretic mobility of the monomeric

protein, and in two cases (G165C G195C and L159C V218C) the monomer appeared as two distinct bands. In contrast, in reducing conditions with DTT, all monomer protein appeared as a single band with a consistent mobility. These results demonstrated that the monomeric protein in solution readily adopted a compact state, consistently with the SAXS data described above.

We repeated cross-linking experiments *in vivo* by adding copper–phenanthroline into live cells[12]. Double-cysteine mutants also formed higher-oligomeric FliG species, beyond decamers, in these experiments (**Fig. 3c**), thus indicating that cross-linking was not an artifact of cell fractionation. Some mutants displayed faint secondary protein bands close to the expected molecular weights of FliG oligomers, possibly owing to an interaction with another unidentified cysteine-containing protein. We also performed *in vivo* characterization of three control double-cysteine mutations that would not be expected to cross-link the ARM$_M$-ARM$_C$ interface (**Fig. 3c**). We observed some cross-linking, up to tetramers, in these controls, a result possibly indicating cross-linking at other interfaces between neighboring FliG subunits or with unidentified proteins that associate with the BFM, particularly the tetramer band, which also appeared in a single-cysteine mutant control (V218C). Controls also displayed other differences such as strong doublet bands between 62 and 98 kDa.

These cross-linking data appeared to contradict the results from a previous study finding no evidence for FliG oligomerization with the L159C V218C double-cysteine mutation in the ARM$_M$-ARM$_C$ interface[29]. We reproduced the experimental conditions of this earlier study, which differed substantially from ours, in particular by having much higher expression levels of FliG. Under the conditions of the previous study, we did observe ladders of resolved oligomers, in the membrane fraction only, after cell fractionation (**Supplementary Fig. 6g**). However, we observed that under these conditions most mutant FliG formed high-molecular-weight, cross-linked aggregates that were too large to visualize by SDS PAGE, and that, in comparison with our other experimental results, there was a much higher proportion of FliG in the cytoplasmic fraction. Both of these observations may explain the apparent absence of mobile bands corresponding to cross-linked protein in the previous study, which did not separate cytoplasmic and membrane fractions. We conclude that, in assembled BFMs, FliG molecules are linked together by intermolecular ARM$_M$-ARM$_C$ interactions.

Figure 4 Schematic depiction of polymerization of the FliG molecule. (**a**) Schematic depiction of polymer extension in solution. Intra- and intermolecular ARM$_M$-ARM$_C$ interactions are shown in red. The MC linker peptide is shown in yellow when it forms an α-helix (helix$_{MC}$) in the FliG polymer and as a dashed line in the compact and extended monomers in which it is at least partly unstructured. (**b**) Schematic depiction of domain-swap polymerization on a FliF template (gray).
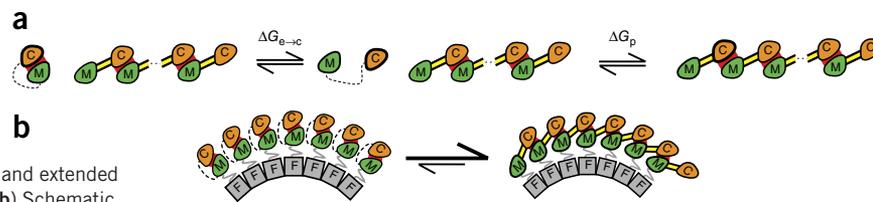
DISCUSSION

Together, these results provide structural evidence for a domain-swap polymerization mechanism for the assembly of the FliG ring. FliG assembly, to our knowledge, is the first reported system in which domain-swap polymerization has been harnessed to construct stable, finite structures. To understand how the flagellar motor self-assembles, we must consider how the spontaneous aggregation of FliG is avoided and how the MS-ring template can trigger domain-swap polymerization during the synthesis of the FliG ring.

SAXS data showed that the fraction of the monomeric protein in an extended conformation decreased with increasing salt concentration. It is difficult to estimate the fraction of monomers that are extended *in vivo*, because this population depends on the intracellular ionic strength, which varies in response to the environment of the bacterium. For example, the concentration of potassium ions alone in *E. coli* increases from 230 mM to 930 mM when the osmolality of growth medium is increased from 0.10 to 1.10 mol/kg (ref. 30). In addition, macromolecular crowding in the cytosol[31], which is thought to drive proteins toward compact configurations[32], is also likely to contribute. We postulate that FliG is likely to preferentially adopt a compact conformation *in vivo*, thus raising the possibility that polymerization is kinetically controlled, and the compact monomer is a metastable intermediate. Disruption of the intramolecular ARM stacking interaction is a candidate activation barrier to spontaneous polymerization. First, a growing oligomer has a hydrophobic ARM interface exposed at either end, thus resulting in a higher energy than that of a compact monomer, until the FliG ring is stabilized by cyclization. Second, the intramolecular ARM$_M$-ARM$_C$ interface of each FliG$_{com}$ monomer must be disrupted before it can be added to the growing chain. However, our observation of large populations of the extended monomer at intermediate salt concentrations suggested that these barriers may not be high enough to account for the very low cytoplasmic concentration of domain-swapped oligomers indicated in our cross-linking experiments. It is therefore more likely that polymerization is thermodynamically controlled, as discussed below.

A characteristic feature of the domain-swap mechanism postulated for FliG ring assembly (shown schematically in **Fig. 4**) is that the intermolecular ARM stacking interaction that binds the FliG ring together is very similar to the intramolecular ARM stacking that stabilizes a monomer in its compact conformation (**Fig. 4a**). The threshold concentration for oligomerization is therefore particularly sensitive to the properties of the flexible or weakly structured linker that joins the domains containing the two ARM motifs.

We denote a linear oligomer comprising n FliG subunits, with two exposed ARM motifs and ($n - 1$) intermolecular interfaces, as FliG$_n$ and define ΔG_P^0 as the standard-state free-energy change for the chain-extension step:

$$\text{FliG}_n + \text{FliG}_{ext} \xrightarrow{\Delta G_P^0} \text{FliG}_{n+1}$$
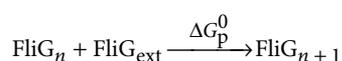
where ΔG_P^0 is assumed to be independent of n. We similarly define a standard-state free-energy change $\Delta G_{e \to c}^0$ for the intramolecular transition in which FliG$_{ext}$ is converted to FliG$_{com}$. In both of these transitions, two exposed ARM domains stack on each other and form an inter- and intramolecular ARM$_M$-ARM$_C$ interaction, respectively (**Fig. 4a**). At equilibrium:

$$\frac{[\text{FliG}_{n+1}]^*}{[\text{FliG}_n]^*[\text{FliG}_{ext}]^*} = \exp\left(-\frac{\Delta G_P^0}{k_B T}\right) \quad (1)$$

$$\frac{[\text{FliG}_{com}]^*}{[\text{FliG}_{ext}]^*} = \exp\left(-\frac{\Delta G_{e \to c}^0}{k_B T}\right) \quad (2)$$

where k_B is the Boltzmann constant, T is the absolute temperature, and $[x]^*$ is the concentration of protein x at equilibrium relative to a 1 M standard concentration. To determine a threshold concentration above which chain extension is favored over an intramolecular interaction, we compare these equations directly. To do this, we consider the intramolecular conformational transition (equation (2)) as a pseudo bimolecular reaction between the two domains, FliG$_M$ and FliG$_C$, which bind via an intramolecular ARM$_M$-ARM$_C$ interaction nearly identical to the intermolecular interaction in equation (1) (**Fig. 4**). We therefore rewrite equation (2) as for a pseudo bimolecular reaction that is analogous to the chain extension in equation (1):

$$\frac{[\text{FliG}_{com}]^*}{c_{eff}[\text{FliG}_{ext}]^*} = \exp\left(-\frac{\Delta G_P^0}{k_B T}\right) \quad (3)$$

where c_{eff} is the effective concentration of one ARM domain of FliG$_{ext}$ at the location of its linked partner domain and quantifies the difference in energy between chain extension (ΔG_P^0) and an extended monomer closing ($\Delta G_{e \to c}^0$) (full derivation in **Supplementary Note 1**). From equations (2) and (3):

$$c_{eff} = \exp((\Delta G_P^0 - \Delta G_{e \to c}^0)/k_B T) \quad (4)$$

From equations (1) and (3):

$$\frac{[\text{FliG}_{n+1}]^*}{[\text{FliG}_n]^*} = \frac{[\text{FliG}_{com}]^*}{c_{eff}} \quad (5)$$

Therefore, c_{eff} also defines a threshold monomer concentration ([FliG$_{com}$]$_{thresh}$) above which runaway polymerization occurs (**Supplementary Notes 2** and **3**):

$$[\text{FliG}_{com}]_{thresh} = c_{eff} \quad (6)$$

If polymerization in the cytosol is thermodynamically rather than kinetically disfavored, there must be a compensating free-energy drive to form rings during motor assembly. This compensation is naturally

provided by the binding of FliG subunits to a template provided by the MS ring (FliF), which increases the local concentration of FliG subunits and places them in an orientation favorable for polymerization. If we assume that an additional binding energy $\Delta G_t > 0$ is contributed for each template-bound FliG molecule, the threshold monomer concentration for polymerization on template is reduced to:

$$[FliG_{com}]_{template} = \exp((\Delta G_P^0 - \Delta G_{e \to c}^0 + \Delta G_t)/k_B T)$$
$$= c_{eff} \exp(\Delta G_t/k_B T)$$

Thus, uncontrolled aggregation can be avoided in the cytoplasm, and motor assembly on the MS-ring template is permitted if the monomer concentration lies between these two limits:

$$[FliG_{com}]_{template} < [FliG_{com}] < [FliG_{com}]_{thresh}$$

The properties of the peptide linking $FliG_M$ and $FliG_C$ determine the probability that the two ARM domains of $FliG_{ext}$ approach each other sufficiently closely for the ARM_M-ARM_C interface to form and hence dictate c_{eff}. Thus, domain swapping provides a natural way to tune the polymerization threshold by modulating the structural properties of the peptide that links $FliG_M$ and $FliG_C$. If the short peptide tether that links the two opposing ARM domains is flexible, then the threshold monomer concentration will be very high, thus protecting against premature aggregation in the cytosol, consistently with our deduction from cross-linking experiments that the cytosolic concentration of oligomers is negligible. We estimated an order-of-magnitude upper limit of c_{eff} of 80 mM by approximating the 12-residue $FliG_M$-$FliG_C$ linker as an unstructured peptide with a Kuhn length of 0.8 nm (refs. 33,34) (**Supplementary Note 3**), which is well above physiological concentrations[11]. In the *A. aeolicus* crystal structure of the FliG polymer, part of the linking peptide forms an extended $helix_{MC}$, and the remainder forms an intermolecular β-sheet[7]. Seven amino acids at the C terminus of $helix_{MC}$ and the entire five–amino acid β-sheet must unravel before FliG can adopt a compact state. The propensity of the linking peptide to form these secondary structures favors polymerization over the compact state and may be necessary to reduce c_{eff} to a level at which polymerization *in vivo* can occur on the FliF template (**Fig. 4b**).

Our results confirm that the FliG ring is made up of units connected by an intermolecular ARM_M-ARM_C interface similar to that seen in crystal structures[7,22]; this structural model perfectly matches the 34-fold, 45 nm–diameter C ring in *Salmonella enterica* observed by cryo-EM[7,35]. The nature and importance of the symmetry mismatch between this ring and its FliF template, the 26-fold MS ring, are unknown. The simplest possibility, a smooth, noncommensurate interface, is at odds with the observation that fusion of FliF and FliG does not measurably alter the observed motor structure[36] and therefore that the structure tolerates covalent linkage of 26 FliG molecules to the FliF template. If the C ring of the FliFG fusion is the same as that of wild type, 8 of its 34 FliG sites would have no covalent link to the FliF ring. It is possible that these sites are empty[29] or that a single FliG straddles two FliM subunits[8], but this would imply a fractional reduction in the electron density of 34-fold reconstructions of the C ring, which is not apparent[13]. The straddle model would require ARM_C to interact with FliM[37], which has not been observed by yeast two-hybrid interactions[38], NMR[39], pulldowns[27] or electron spin resonance[8], and which would be in competition with the strong thermodynamic drive to cap exposed ARM interfaces (**Supplementary Fig. 7a**). FliG that is not bound to FliF may occupy these gaps, which would require eight

spare FliF molecules in the FliFG fusion, which would either hang free or be cleaved from FliFG (**Supplementary Fig. 7b**).

Domain-swapped protein complexes have previously been observed to be either self-terminating small oligomers or runaway polymers[40,41]. The synthesis of the FliG ring defines a third category and is the first reported example, to our knowledge, in which a domain-swapped polymer consisting of dozens of subunits forms terminated fixed ring structures. A structural scaffold (MS ring) triggers assembly by increasing the local concentration of FliG and acts as a template ensuring that ring closure occurs at the correct polymer length. This study provides a structural and thermodynamic model that may inform the mechanisms underlying the self-assembly of other protein complexes.

METHODS
Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS
M.A.B.B. analyzed SAXS data, performed covariance and thermodynamic analysis, conceived experiments and wrote the manuscript. R.M.G.H. collected SAXS data and performed cross-linking assays. L.A.G. expressed and purified protein and collected SAXS data. N.S.M. performed cross-linking assays and *in vivo* functional assays. C.W.L. expressed and purified protein and collected SAXS data. A.A.R. expressed and purified protein. A.P.D. collected and analyzed SAXS data. A.E.W. and C.M.J. analyzed SAXS data. N.J.D. and J.P.A. constructed cell lines for this study. Y.V.M. performed fluorescence imaging experiments. D.S. conceived experiments and contributed to cross-linking experiments. A.J.T. and R.M.B. performed thermodynamic analysis, conceived experiments and wrote the manuscript. K.N. performed fluorescence imaging experiments, conceived experiments and wrote the manuscript. L.K.L. conceived experiments, performed cross-linking assays, collected and analyzed SAXS data and wrote the manuscript.

1. Sowa, Y. & Berry, R.M. Bacterial flagellar motor. *Q. Rev. Biophys.* **41**, 103–132 (2008).
2. Macnab, R.M. How bacteria assemble flagella. *Annu. Rev. Microbiol.* **57**, 77–100 (2003).
3. Erhardt, M., Namba, K. & Hughes, K.T. Bacterial nanomachines: the flagellum and type III injectisome. *Cold Spring Harb. Perspect. Biol.* **2**, a000299 (2010).
4. Lele, P.P., Hosu, B.G. & Berg, H.C. Dynamics of mechanosensing in the bacterial flagellar motor. *Proc. Natl. Acad. Sci. USA* **110**, 11839–11844 (2013).
5. Huber, A.H., Nelson, W.J. & Weis, W.I. Three-dimensional structure of the armadillo repeat region of β-catenin. *Cell* **90**, 871–882 (1997).
6. Tewari, R., Bailes, E., Bunting, K.A. & Coates, J.C. Armadillo-repeat protein functions: questions for little creatures. *Trends Cell Biol.* **20**, 470–481 (2010).
7. Lee, L.K., Ginsburg, M.A., Crovace, C., Donohoe, M. & Stock, D. Structure of the torque ring of the flagellar motor and the molecular basis for rotational switching. *Nature* **466**, 996–1000 (2010).
8. Sircar, R. *et al.* Assembly states of FliM and FliG within the flagellar switch complex. *J. Mol. Biol.* **427**, 867–886 (2015).
9. Minamino, T. *et al.* Structural insight into the rotational switching mechanism of the bacterial flagellar motor. *PLoS Biol.* **9**, e1000616 (2011).

10. Vartanian, A.S., Paz, A., Fortgang, E.A., Abramson, J. & Dahlquist, F.W. Structure of flagellar motor proteins in complex allows for insights into motor structure and switching. *J. Biol. Chem.* **287**, 35779–35783 (2012).

11. Delalez, N.J. *et al.* Signal-dependent turnover of the bacterial flagellar switch protein FliM. *Proc. Natl. Acad. Sci. USA* **107**, 11347–11351 (2010).

12. Paul, K., Brunstetter, D., Titen, S. & Blair, D.F. A molecular mechanism of direction switching in the flagellar motor of *Escherichia coli. Proc. Natl. Acad. Sci. USA* **108**, 17171–17176 (2011).

13. Stock, D., Namba, K. & Lee, L.K. Nanorotors and self-assembling macromolecular machines: the torque ring of the bacterial flagellar motor. *Curr. Opin. Biotechnol.* **23**, 545–554 (2012).

14. Sharpe, J.C. & London, E. Diphtheria toxin forms pores of different sizes depending on its concentration in membranes: probable relationship to oligomerization. *J. Membr. Biol.* **171**, 209–221 (1999).

15. Sambashivan, S., Liu, Y., Sawaya, M.R., Gingery, M. & Eisenberg, D. Amyloid-like fibrils of ribonuclease A with three-dimensional domain-swapped and native-like structure. *Nature* **437**, 266–269 (2005).

16. Yamasaki, M., Li, W., Johnson, D.J.D. & Huntington, J.A. Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization. *Nature* **455**, 1255–1258 (2008).

17. Hirota, S. *et al.* Cytochrome c polymerization by successive domain swapping at the C-terminal helix. *Proc. Natl. Acad. Sci. USA* **107**, 12854–12859 (2010).

18. Das, P., King, J.A. & Zhou, R. Aggregation of γ-crystallins associated with human cataracts via domain swapping at the C-terminal β-strands. *Proc. Natl. Acad. Sci. USA* **108**, 10514–10519 (2011).

19. Hopf, T.A. *et al.* Three-dimensional structures of membrane proteins from genomic sequencing. *Cell* **149**, 1607–1621 (2012).

20. Hopf, T.A. *et al.* Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife* **3**, e03430 (2014).

21. Neher, E. How frequent are correlated changes in families of protein sequences? *Proc. Natl. Acad. Sci. USA* **91**, 98–102 (1994).

22. Lam, K.-H. *et al.* Multiple conformations of the FliG C-terminal domain provide insight into flagellar motor switching. *Structure* **20**, 315–325 (2012).

23. Brown, P.N., Hill, C.P. & Blair, D.F. Crystal structure of the middle and C-terminal domains of the flagellar rotor protein FliG. *EMBO J.* **21**, 3225–3234 (2002).

24. David, G. & Pérez, J. Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. *J. Appl. Crystallogr.* **42**, 892–900 (2009).

25. Hynson, R.M.G., Duff, A.P., Kirby, N., Mudie, S. & Lee, L.K. Differential ultracentrifugation coupled to small-angle X-ray scattering on macromolecular complexes. *J. Appl. Crystallogr.* **48**, 769–775 (2015).

26. Franke, D., Jeffries, C.M. & Svergun, D.I. Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. *Nat. Methods* **12**, 419–422 (2015).

27. Lam, K.-H. *et al.* Structural basis of FliG-FliM interaction in *Helicobacter pylori. Mol. Microbiol.* **88**, 798–812 (2013).

28. DeRosier, D. Bacterial flagellum: visualizing the complete machine *in situ. Curr. Biol.* **16**, R928–R930 (2006).

29. Paul, K., Gonzalez-Bonet, G., Bilwes, A.M., Crane, B.R. & Blair, D. Architecture of the flagellar rotor. *EMBO J.* **30**, 2962–2971 (2011).

30. Richey, B. *et al.* Variability of the intracellular ionic environment of *Escherichia coli.* Differences between *in vitro* and *in vivo* effects of ion concentrations on protein-DNA interactions and gene expression. *J. Biol. Chem.* **262**, 7157–7164 (1987).

31. Zimmerman, S.B. & Trach, S.O. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli. J. Mol. Biol.* **222**, 599–620 (1991).

32. Kuznetsova, I.M., Turoverov, K.K. & Uversky, V.N. What macromolecular crowding can do to a protein. *Int. J. Mol. Sci.* **15**, 23090–23140 (2014).

33. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J.M. & Gaub, H.E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109–1112 (1997).

34. Dunn, K.E. *et al.* Guiding the folding pathway of DNA origami. *Nature* **525**, 82–86 (2015).

35. Thomas, D.R., Francis, N.R., Xu, C. & DeRosier, D.J. The three-dimensional structure of the flagellar rotor from a clockwise-locked mutant of *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.* **188**, 7039–7048 (2006).

36. Thomas, D., Morgan, D.G. & DeRosier, D.J. Structures of bacterial flagellar motors from two FliF-FliG gene fusion mutants. *J. Bacteriol.* **183**, 6404–6412 (2001).

37. Brown, P.N., Terrazas, M., Paul, K. & Blair, D.F. Mutational analysis of the flagellar protein FliG: sites of interaction with FliM and implications for organization of the switch complex. *J. Bacteriol.* **189**, 305–312 (2007).

38. Marykwas, D.L. & Berg, H.C. A mutational analysis of the interaction between FliG and FliM, two components of the flagellar motor of *Escherichia coli. J. Bacteriol.* **178**, 1289–1294 (1996).

39. Dyer, C.M., Vartanian, A.S., Zhou, H. & Dahlquist, F.W. A molecular mechanism of bacterial flagellar motor switching. *J. Mol. Biol.* **388**, 71–84 (2009).

40. Shameer, K., Pugalenthi, G., Kandaswamy, K.K. & Sowdhamini, R. 3dswap-pred: prediction of 3D domain swapping from protein sequence using Random Forest approach. *Protein Pept. Lett.* **18**, 1010–1020 (2011).

41. Bennett, M.J., Sawaya, M.R. & Eisenberg, D. Deposition diseases and 3D domain swapping. *Structure* **14**, 811–824 (2006).

ONLINE METHODS

Evolutionary covariance analysis. Covarying residues in Aquifex FliG were identified with a maximum-entropy approach[42]. A sequence alignment was generated from the UniProt database (http://www.uniprot.org/) by the web service EVfold[19] with HHblits[43] over two iterations with an E value of −3. Alignments are represented by rows of sequences such that aligned amino acids appear in successive 'match state' columns. To account for genetic insertions or deletions, gaps can be introduced into sequences. A large proportion of gaps is often indicative of a poor alignment; hence, sequences with more than 70% gaps were excluded, thus leaving a total of 2,292 sequences. In addition, covariance analysis was performed only on match-state columns with <50% gaps across the whole alignment (residues 5–325 in PDB 3HJL), which produced a list of ranked covariance scores for amino acid pairs. We selected 142 covarying pairs with correlation scores greater than 0.04, representing the top 0.3% of correlated pairs, for further analysis. These highly correlated pairs had correlation scores more than 5 s.d. above the mean score for those pairs below the threshold.

Protein preparation and purification. FliG constructs were obtained by PCR from genomic DNA from *E. coli* strain RP437. These were cloned into a modified pACYCDuet T7 expression vector (Novagen) with an N-terminal decahistidine tag and protease-cleavage site. Protein was expressed in T7 *E. coli* expression cells (NEB) in Luria Broth (LB) at 18 °C overnight. Cells were harvested by centrifugation and then resuspended in 20 mM Tris HCl, pH 7.5, 100 mM NaCl, 1× complete EDTA-free protease inhibitor (Roche) and DNase I before lysis in a continuous-flow cell disruptor (Constant Systems). Cell debris was removed by centrifugation, and lysate was passed over a preequilibrated Ni-NTA Sepharose column (GE Healthcare). Beads were washed thoroughly in wash buffer with 20 mM Tris-HCl, pH 7.5, 150 mM NaCl and 20 mM imidazole. The column was then reequilibrated in protease-cleavage buffer containing 50 mM Tris-HCl, pH 7.5, 150 mM NaCl and 1 mM DTT. Protein was eluted by incubation with PreScission protease (GE Healthcare), which removed the polyhistidine tag, thus producing protein with no affinity tags. Eluted protein was concentrated and further purified by size-exclusion chromatography, and its monodispersity was confirmed with multiangle laser light scattering before being used for SAXS data acquisition.

Small-angle X-ray scattering. *Data collection.* Scattering data (I(q) versus q, where q = 4πsinθ/λ, 2θ is the scattering angle, and λ = 1.127 Å) were collected immediately after elution from an inline size-exclusion chromatography column at the Australian Synchrotron. Purified monomeric protein (10 mg ml⁻¹) was injected into a 23-ml Sephacryl S-200 size-exclusion chromatography column (GE Healthcare) at a flow rate of 0.5 ml min⁻¹. The outflow was piped directly into a temperature-controlled 1.5-mm quartz capillary at 20 °C, through which monochromatic X-rays were passed at a flux of 4×10^{12} photons/s. SAXS data were collected with exposure times of 5 s on a Pilatus 1M photon counting detector (Dectris), which was placed 1.48 m from the sample capillary.

Data processing and analysis. Data reduction was performed with the beamline-specific software package 'scatterbrain' (Australian Synchrotron http://www.synchrotron.org.au/aussyncbeamlines/saxswaxs/software-saxswaxs/). Software for data processing were from the ATSAS suite of programs for SAXS data processing[44] including DATOP, PRIMUS, DATAVER, DAMMIN, DAMAVER, DATPOROD, CRYSOL and DATCMP. Custom software was written to plot the intensity of scattering data versus frame number. For each data set, plots were generated of averaged intensities over several q ranges. This was used to visualize the elution profile from SEC and select appropriate frames (20–50 frames) containing buffer only to subtract from scattering data. Averaged buffer scattering profiles were then subtracted from each acquisition frame with DATOP. The radius of gyration (R_g) was then calculated with AUTORG for each buffer-subtracted frame, and the corresponding calculated scattering, assuming a mono-disperse sample, was plotted over the average measured scattering at low q, which allowed the identification of frames in which the calculated R_g was consistent. To confirm that scattering profiles were consistent across these frames, their correlations were evaluated with DATCMP, to perform reduced χ^2 analysis and generate correlation maps. All frames that were subsequently scaled and averaged for further analysis had a *P* value for similarity of 1.000 (ref. 26) (**Supplementary Fig. 1**). Background-subtracted data from monodisperse protein were then scaled and visually inspected for consistency before being averaged, and initial Guinier plots

and probability atom distance distribution functions (P(r) versus r) were plotted with PRIMUS. Inspection of unrestrained r_{min} in P(r) plots and high-q data indicated that data were generally undersubtracted. Hence, we systematically generated frame-by-frame subtractions in which buffer was scaled by 1.0001–1.0090 with DATOP and DATAVER. The appropriate buffer scaling was chosen on the basis of criteria in which r_{min} was close to or equal to zero, and high-q data (q >0.45 Å⁻¹) fluctuated around zero. These data were used to generate P(r) plots, which were subsequently used to generate *ab initio* shape restorations for $FliG_{MC}$ in 1 M NaCl with an average normalized spatial discrepancy of 0.48 ± 0.01 (mean ± s.e.m.)[45] with DAMMIN (**Supplementary Figs. 2–4**). To allow a direct comparison, all plots of P(r) versus r were scaled to have an area under the curve of 1. At least 20 *ab initio* dummy-atom shape restorations were performed for each data set, which were aligned with DAMMIN. These were converted into pseudo electron density for averaging and scaling with pdb2vol from the Situs suite[46]. Pseudo electron density from shape restorations was then averaged in Chimera[47], and sigma levels were set such that the density volume matched the Porod volume for each data set, which was calculated with the ATSAS program DATPOROD. High-resolution X-ray crystal structures were docked into averaged pseudo electron density in Chimera by iterative maximization of the correlation between structures and density. CRYSOL was used to calculate theoretical scattering data from high-resolution atomic models to compare with SAXS data.

Cross-linking. *Plasmids and mutant design.* For cross-linking and functional assays, FliG from *E. coli* strain RP437 was cloned into an arabinose-induced expression vector[48] with an N-terminal HA peptide for immunodetection in western blots. Site-directed mutagenesis to introduce cysteine mutations was performed with long-range plasmid PCR[49]. Custom software was written to rationally identify candidate residue pairs for disulfide cross-linking experiments. The ARMm-ARMc interaction from the *A. aeolicus* crystal structure (PDB 3HJL) was used as the template structure. All possible residue pairs between ARMm and ARMc were systematically mutated to cysteine *in silico*. Then the geometries of all rotamer combinations were assessed for their similarity to ideal disulfide-bond length, angle and torsion angle.

Bacterial strains and preparation. *E. coli* cell strain JPA600 lacking FliG was created from the parent strain RP437 by allelic exchange[50]. Deletion of the FliG gene was verified with PCR and DNA sequencing, and with functional assays to confirm that motility had been eliminated. Cells were not tested for mycoplasma contamination. JPA600 was used in complementation assays to perform cross-linking and functional assays. T7 *E. coli* expression cells (NEB) were used for protein expression. Cysteine mutations were introduced into JPA600 by chemical transformation of FliG in pBAD24 expression vectors. Transformed colonies were used to inoculate a preculture in rich LB medium with 50 μg ml⁻¹ ampicillin. Precultures were incubated at 37 °C for 3–5 h (OD ≈ 1.0) and were used to noculate minimal tryptone broth (TB) medium consisting of 1% tryptone in phosphate-buffered saline and 50 μg ml⁻¹ ampicillin to a starting OD of 0.01. The minimal medium also contained 0.0025% arabinose to induce protein expression. The concentration of arabinose used was optimized by comparing the motility and FliG expression levels in complemented cells and WT cells. Inoculated TB medium was incubated at 30 °C until mid-log phase (OD = 0.6–0.8) and harvested for functional and cross-linking assays. Cell motility was assessed by transferring cells into motility buffer consisting of 10 mM potassium phosphate, pH 7.0, and 0.1 mM EDTA and observing their motility. The percentage of nonmotile, running and tumbling cells was recorded (**Supplementary Fig. 6**). Swarm assays were performed on semisolid agar plates consisting of 0.0025% arabinose, 50 μg/ml ampicillin and 0.35% agar in 1% tryptone with 85.6 mM NaCl. Swarms were inoculated by gently stabbing a transformed colony and then inoculating a swarm plate with a pipette tip. All functional assays were performed at least three times.

Cross-linking assays. For fractionated cross-linking assays, cells at mid-log phase were harvested by centrifugation at 3,000*g* for 5 min. The supernatant was removed, and cells were resuspended in lysis buffer consisting of 100 mM Tris-HCl, pH 8.0, 150 mM NaCl, lysozyme and DNase I and incubated at room temperature for 30 min. Cells were then lysed through three freeze-thaw cycles with liquid nitrogen and a water bath at room temperature. Cross-linking was performed on ice with iodine from a fresh 20 mM stock dissolved in 100% ethanol, which was added to a final concentration of 0.2 mM and was left to incubate

for 20 s. Reactions were stopped through the addition of N-ethylmaleimide from a fresh 0.4 M stock dissolved in 100% ethanol to a final concentration of 20 mM. Lysed cells were then pelleted by centrifugation at 17,000g for 15 min to harvest the soluble cytoplasmic fraction. The membrane fraction was obtained by resuspension of cell debris in 1% Triton X-100 in 0.1 M Tris-HCl, pH 8.0, and 150 mM NaCl, and subsequent incubation with gentle agitation at 4 °C for 30 min. Cell debris was pelleted again; the supernatant contained the solubilized membrane fraction. Finally, the remaining insoluble fraction was dissolved in buffer consisting of 40 mM Tris HCl, pH 8.0, 5% SDS, 8 M urea and 0.1 mM EDTA. For *in vivo* cross-linking experiments, cells were harvested and resuspended in motility buffer before addition of cross-linking reagent to a concentration of 4 mM $CuSO_4$ and 16 mM 1,10 phenanthroline, which was made fresh from a stock solution of 400 mM $CuSO_4$ in 50% ethanol and 1 M 1,10 phenanthroline in 95% ethanol, as described previously[12]. Samples were incubated with cross-linking reagent for 5 min before the reactions were stopped as described above.

Immunoblots. Samples were electrophoresed on a 4–12% Bis-Tris polyacrylamide gel (Invitrogen) for 35 min at 200 V with or without DTT as appropriate. Protein was transferred to nitrocellulose membranes and blocked with 5% skim-milk powder in PBS. Immunoblotting was performed with a 1:15,000 dilution of horseradish peroxidase–coupled anti-HA antibody (Sigma-Aldrich, H6533; validation on manufacturer's website). Protein bands were observable on nitrocellulose after incubation with HRP substrate (PerkinElmer).

Flagellar filament staining. The flagellar filaments produced by *Salmonella* cells were labeled with anti-FliC antiserum and anti-rabbit IgG conjugated with Alexa Fluor 594 (Invitrogen A-11012; validation on manufacturer's website) as previously described[51]. The cells were observed by fluorescence microscopy as previously described[51]. Fluorescence images were processed with ImageJ software, version 1.48 (http://imagej.nih.gov/ij/).

42. Marks, D.S. *et al.* Protein 3D structure computed from evolutionary sequence variation. *PLoS One* **6**, e28766 (2011).
43. Johnson, L.S., Eddy, S.R. & Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**, 431 (2010).
44. Konarev, P.V., Petoukhov, M.V., Volkov, V.V. & Svergun, D.I. ATSAS2.1, a program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **39**, 277–286 (2006).
45. Kozin, M.B. & Svergun, D.I. Automated matching of high-and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41 (2001).
46. Wriggers, W. Using Situs for the integration of multi-resolution structures. *Biophys. Rev.* **2**, 21–27 (2010).
47. Pettersen, E.F. *et al.* UCSF Chimera: a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
48. Guzman, L.M., Belin, D., Carson, M.J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
49. Papworth, C., Bauer, J.C., Braman, J. & Wright, D.A. Site-directed mutagenesis in one day with >80% efficiency. *Strategies* **9**, 3–4 (1996).
50. Parkinson, J.S. Complementation analysis and deletion mapping of *Escherichia coli* mutants defective in chemotaxis. *J. Bacteriol.* **135**, 45–53 (1978).
51. Morimoto, Y.V., Nakamura, S., Kami-ike, N., Namba, K. & Minamino, T. Charged residues in the cytoplasmic loop of MotA are required for stator assembly into the bacterial flagellar motor. *Mol. Microbiol.* **78**, 1117–1129 (2010).