

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

Authors: Matthew A. B. Baker^{1,2}, Robert M. G. Hynson¹, Lorraine A. Ganuelas¹, Nasim Shah Mohammadi¹, Chu Wai Liew¹, Anthony A. Rey¹, Anthony P. Duff³, Andrew E. Whitten³, Cy M. Jeffries^{3,4}, Nicolas J. Delalez⁵, Yusuke V. Morimoto⁶, Daniela Stock¹, Judith P. Armitage⁵, Andrew J. Turberfield⁷, Keiichi Namba⁶, Richard M. Berry⁷, Lawrence K. Lee^{1,2}

Affiliations:

¹Structural and Computational Biology Division, The Victor Chang Cardiac Research Institute, Darlinghurst, New South Wales, Australia

²European Molecular Biology Laboratory Australia Node for Single Molecule Science, School of Medical Sciences, The University of New South Wales, Sydney, New South Wales, Australia

³Australian Nuclear and Science Technology Organisation, Lucas Heights, New South Wales, Australia,

⁴Present Address: European Molecular Biology Laboratory, Hamburg, Germany

⁵Department of Biochemistry, Oxford University, Oxford, UK.

⁶Graduate School of Frontier Biosciences, Osaka University, Osaka, Japan

⁷Department of Physics, Oxford University, Oxford, UK

Correspondence to Lawrence K. Lee: lawrence.lee@unsw.edu.au

Abstract:

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 FliG from *Escherichia coli*, one of the first motor proteins to assemble, forms ordered ring structures via domain-swap polymerization, which in other proteins has been associated with uncontrolled and deleterious protein aggregation. Solution and crystal structural data, in combination with *in vivo* biochemical crosslinking experiments and evolutionary covariance analysis, reveal 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, where a structural template controls assembly and shapes polymer formation into rings.

One Sentence Summary:

Self-assembly of the bacterial flagellar motor protein FliG involves the formation of a domain-swapped polymeric ring, which is initiated and shaped by a structural template.

Main Text:

During the assembly of the bacterial flagellar motor (BFM), rotor proteins assemble into large, coaxial rings. This 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 a ~30 nm diameter ring (MS-ring) embedded in the bacterial cytoplasmic membrane¹. The MS-ring acts as a structural scaffold that seeds the assembly of hundreds of other subunits to form 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} to form the cytoplasmic or C-ring rotor that interacts with transmembrane stator complexes to convert ion flux through the stator into rotation. The C-ring is also part of a type-3 secretory apparatus³ and an ultrasensitive and tunable sensory switch that allows the motor to change directions in response to environmental stimuli⁴.

At cytoplasmic concentrations, FliG is monomeric yet capable of forming multimeric ring structures during assembly at the site of the rotor. This makes it a good model for understanding the mechanisms that control the assembly of monomeric proteins to form robust protein superstructures such as the BFM. The crystal structure of FliG has been determined in 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 an 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 to form right-handed superhelices^{5,6}. However, while all other known ARMs are expressed as tandem repeats, the FliG ARMs are separated by around 30 amino acids, most of which (23 residues) can form an alpha helix that bridges the middle and C-terminal domains (helix_{MC}). In some crystal structures⁷ the formation of this helix prevents intramolecular stacking and FliG adopts an extended conformation, with ARM_M and ARM_C domains from neighboring molecules stacked on each other. The intermolecular ARM_M – ARM_C interface at the heart of this domain-swap structural motif links FliG molecules to form polymers: this

has been proposed as the basis for an atomic model of the FliG ring^{7,8}. Other crystal structures indicate that intramolecular stacking of the FliG ARMs is possible and results in a compact monomeric conformation^{9,10}. Interaction studies involving electron spin resonance⁸ have demonstrated that disruption of the ARM_M-ARM_C interface interferes with FliM oligomerization, but these assays were executed at well above the cytosolic monomer concentration, which is likely to be similar to FliM at ~500 nM¹¹. As crystal conformations are likely to be also influenced by crystal packing, the biological relevance of the ARM_M-ARM_C interaction, including its role in the self-organization of the BFM, remain controversial^{8,12,13}.

To determine the molecular mechanism behind motor assembly, we demonstrated the biological importance of the ARM_M-ARM_C interaction using evolutionary covariance analysis. We also determined the solution structure of monomeric FliG using small angle X-ray scattering (SAXS) and used biochemical crosslinking 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 disorders¹⁴⁻¹⁷ and the aggregation of γ -crystallins in human cataracts¹⁸. Here we describe the first reported case where domain-swap polymerization is initiated, shaped, and terminated by means of a structural template to form a well-defined supramolecular structure. We expect that this mechanism may generalize 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 2292 unique FliG protein sequences. Covariance analysis can predict protein structure¹⁹ and protein-protein interaction sites²⁰ by identifying amino acid pairs that co-evolve²¹. Our analysis identified 142 out of 48,217

possible amino acid pairs with correlation scores that were more than 5 standard deviations above the mean for the remaining pairs (Fig. 1A. See Materials and Methods). In the extended conformation FliG crystal structure from *A. aeolicus* (PDBID:3HJL⁷), these residues correlate to those that are in close proximity (Fig. 1B and C): 12.2% of the correlated residues are in direct contact ($< 5 \text{ \AA}$ distance) and play key roles in the stabilization of the protein fold by forming the hydrophobic core of globular structural domains or through hydrogen bonding networks (Fig. 1D). Among the co-varying residue pairs identified, nine have one member in each of the ARM_M and ARM_C domains; these appear as an outlying distribution with around 50 Å separation (Fig. 1B). However, when separations are measured 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 co-varying pairs are in proximity (Fig. 1B) and five are in direct contact, forming a substantial part of the hydrophobic interface (Fig. 1E). A 2D distance plot displaying all amino acid pairs closer than 7 Å in the extended monomer structure, with overlaid correlated covariance pairs, shows that the correlated pairs map well to those in close proximity in this structure (Fig. 1F). The cluster of correlated pairs corresponding to the conserved ARM_M-ARM_C interaction does not map to the monomeric extended structure, but does map to proximate residues in the intermolecular ARM_M-ARM_C interaction from the extended lattice (Fig. 1F). This analysis demonstrates that the ARM_M-ARM_C interaction observed in all FliG crystal structures^{7,9,10,12,22} is evolutionarily conserved, providing additional evidence that this interaction is likely to be important for 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, depicted in Figure 2. In crystal structures of FliG in the extended state (FliG_{ext}, Fig. 2A) from *A. aeolicus*, *T. maritima* and *H. pylori*^{7,22,23}, this domain is formed between neighboring molecules in the crystal lattice. Whereas 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}, modelled in Fig.

2A). Given that the three domains of FliG are connected by flexible linking peptides⁷, 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 small-angle X-ray scattering (SAXS) to determine the conformational state of *E. coli* FliG in solution.

Data were collected from the eluate of a size-exclusion chromatography column to exclude the possible influence of aggregates^{24,25}: independent analysis by multi-angle laser light scattering (MALS) showed that the purified protein was monomeric in solution at the 10 mg/ml FliG concentrations used in all SAXS experiments (Supplementary Fig. 1). This was confirmed by estimates of molecular weight from Porod volumes (Supplementary Table 1), a consistent radius of gyration (R_g) and by evaluating correlations between scattering profiles across elution peaks with correlation maps and reduced χ^2 analysis²⁶ (Supplementary Fig. 2). We compared measurements from SAXS experiments to the crystal structure of FliG_{ext} from *A. aeolicus* (Fig. 1D, PDBID: 3HJL⁷), which is the only full-length structure available⁷. A model FliG_{com} was generated 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 described⁷. 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. Together with the five additional unstructured amino acids between helix_{MC} and FliG_C, this results in an unstructured peptide of 12 amino acids, which provides sufficient inter-domain flexibility to form 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). In a solution with 150 mM NaCl, the R_g of the full-length FliG protein is 37.2 ± 0.3 Å, 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 and 2C, 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 revealed that in solution this construct formed a single domain (Supplementary Fig. 4). The radius of gyration of FliG_{MC} in solution decreased asymptotically with increasing ionic strength from 31.4 ± 0.3 Å at 150 mM NaCl to 24.1 ± 0.3 Å in 1 M NaCl (Fig. 2B), close to the 23.4 Å calculated for this fragment in the FliG_{com} model. This 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) indicating that in 1M NaCl FliG_{MC} may be entirely in the compact state. Consistent with the trend in R_g (Fig. 2B), decreasing ionic strength shifted the distribution towards that of the FliG_{ext} crystal structure (Fig. 2C). Multiple independent *ab initio* shape restorations of FliG_{MC} from SAXS data collected in 1 M NaCl were aligned and averaged to produce a final shape that was consistent with the FliG_{com} crystal structure (Fig. 2C), 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 $I_{com}(q)$, $I_{ext}(q)$ calculated from the model structures FliG_{com} and FliG_{ext} respectively (Supplementary Fig. 5):

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

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 the fraction, f , of FliG that is in an extended state by fitting the SAXS data. Fitted fractions f are 48%, 35%, 7% and 1% in 150 mM, 250 mM, 500 mM and 1000 mM NaCl respectively. These results allow us to draw the robust qualitative

inference that the fraction of FliG that is in an extended conformation increases rapidly as the NaCl concentration increases.

The crystal structure of the full-length protein reveals 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)⁷. To determine whether helix_{NM} might prevent full-length FliG from adopting a compact conformation, we also analyzed an expanded FliG_{MC} construct that includes helix_{NM} (FliG_{MC-EXT}) with SAXS (Supplementary Fig. 2). The R_g of FliG_{MC-EXT} also decreased with increasing NaCl concentration demonstrating that, even in the presence of helix_{NM}, FliG adopts a more compact conformation at high ionic strengths (Supplementary Table 1). We conclude that 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 crosslinking of the ARM_M-ARM_C interface

To determine the role of the ARM_M-ARM_C interaction in functional flagellar motors, we covalently crosslinked the ARM_M and ARM_C domains using targeted double cysteine mutants. Three double cysteine mutations (L159C V218C, G164C G196C and G165C G195C) were rationally designed to form crosslinks at the ARM_M-ARM_C interface on the basis of the crystal structure of FliG from *A. aeolicus*⁷. Wild-type FliG was replaced by each mutant *in vivo* (Fig. 3A). Mutations did not affect motor assembly since flagellar filaments were observed with fluorescence microscopy and cells were motile (Supplementary Fig. 6A-F). Since flagellar motor basal bodies remain intact after cell fractionation and purification²⁸, we separated the soluble and membrane fractions of lysed cells which contain freely diffusing and motor-associated FliG protein, respectively. Crosslinking was assessed with SDS PAGE and immunoblots. FliG oligomers were observed 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 dithiothreitol (DTT), confirming that oligomer formation was mediated by disulfide bond formation

(Fig. 3B). The cytoplasm is a reducing environment so we therefore added iodine between cell lysis and fractionation to promote disulfide bond formation 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 only higher order oligomeric species were observed. In contrast, in the soluble fraction only monomeric protein was observed (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. This demonstrated that a monomeric protein in solution could readily adopt a compact state, in keeping with SAXS data above.

We repeated crosslinking experiments *in vivo* by adding Cu-phenanthroline into live cells¹². Double cysteine mutants also formed higher oligomeric FliG species, beyond 10-mers, in these experiments (Fig. 3C), indicating that crosslinking was not an artifact of cell fractionation. Some mutations displayed faint secondary protein bands close to the expected molecular weights of FliG oligomers, possibly caused by an interaction with another unidentified cysteine-containing protein. We also characterized *in vivo* three control double cysteine mutations that would not be expected to crosslink the ARM_M-ARM_C interface (Fig. 3C). Some crosslinking up to tetramers were observed in these controls, possibly indicating crosslinking at other interfaces between neighboring FliG subunits, or with unidentified proteins that associate with the BFM, in particular the tetramer band, which also appears in a single-cysteine mutant control (V218C). Controls also displayed other differences such as strong doublet bands between 62 and 98 kDa.

These crosslinking data appear to contradict a previous study in which no evidence for FliG oligomerization was observed with the L159C V218C double-cysteine mutation in the ARM_M-ARM_C interface²⁹. 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 forms high molecular weight, crosslinked aggregates that are too large to visualize by SDS PAGE, and that, compared with our experiments, there is a much higher proportion of FliG in the cytoplasmic fraction. Both of these observations could explain the apparent absence of mobile bands corresponding to crosslinked 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.

Discussion

Together, these results provide structural evidence for a domain-swap polymerization mechanism for the assembly of the FliG ring. FliG assembly appears to be the first reported system where 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.

Mechanism for FliG assembly

SAXS data show that the fraction of the monomeric protein in an extended conformation decreases 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/kg³⁰. In addition, macromolecular crowding in the cytosol³¹, which is thought to drive proteins towards compact configurations³², is also likely to contribute. We postulate that, *in vivo*, FliG is likely to preferentially adopt a compact conformation, which raises the possibility that polymerization is kinetically controlled with the

compact monomer as 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, which has a higher energy than a compact monomer, until the FliG ring is stabilized by cyclization. Secondly, 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 suggests that these barriers may not be high enough to account for the very low cytoplasmic concentration of domain-swapped oligomers indicated by 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 Figure 4, is that the intermolecular ARM stacking interaction that binds together the FliG ring 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:



where ΔG_p^0 is assumed to be independent of n . We similarly define a standard-state free energy change $\Delta G_{e \rightarrow c}^0$ 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 to form an inter- and intra-molecular ARM_M-ARM_C interaction respectively (Fig. 4A). At equilibrium:

$$\frac{[FliG_{n+1}]^*}{[FliG_n]^*[FliG_{ext}]^*} = \exp\left(-\frac{\Delta G_p^0}{k_B T}\right) \quad (1)$$

$$\frac{[FliG_{com}]^*}{[FliG_{ext}]^*} = \exp\left(-\frac{\Delta G_{e \rightarrow c}^0}{k_B T}\right), \quad (2)$$

where k_B is the Boltzmann constant, T is the absolute temperature and $[X]^*$ is the concentration of 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 (Eq. 2) as a pseudo-bimolecular reaction between the two domains, FliG_M and FliG_C, that bind via an intramolecular ARM_M-ARM_C interaction which is nearly identical to the intermolecular interaction of Eq. 1 (Fig. 4). We therefore rewrite Eq. 2 as for a pseudo-bimolecular reaction that is analogous to chain extension, Eq. 1:

$$\frac{[FliG_{com}]^*}{c_{eff}[FliG_{ext}]^*} = \exp\left(-\frac{\Delta G_p^0}{k_B T}\right). \quad (3)$$

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 ($\Delta G_{e \rightarrow c}^0$) (see SI Note 1 for full derivation). From Eqs 2 and 3:

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

From Eqs 1 and 3:

$$\frac{[FliG_{n+1}]^*}{[FliG_n]^*} = \frac{[FliG_{com}]^*}{c_{eff}} \quad (5)$$

c_{eff} therefore also defines a threshold monomer concentration ($[FliG_{com}]_{thresh}$) above which runaway polymerization will occur (SI Notes 2 and 3):

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

If polymerization in the cytosol is thermodynamically rather than kinetically disfavoured, there must be a compensating free-energy drive to form rings during motor assembly. This 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 and places them in an orientation favourable for polymerization. If we assume that this contributes an additional binding energy $\Delta G_t < 0$ for each template-bound FliG molecule, the threshold monomer concentration for polymerization on-template is reduced to:

$$[FliG_{com}]_{template} = \exp((\Delta G_p^0 - \Delta G_{e \rightarrow c}^0 + \Delta G_t)/k_B T) = c_{eff} \exp(\Delta G_t/k_B T) \quad (7)$$

Thus uncontrolled aggregation can be avoided in the cytoplasm, and motor assembly on the MS-ring template 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 swap 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, protecting against premature aggregation in the cytosol, consistent with our deduction from crosslinking experiments that the cytosolic concentration of oligomers is negligible. We estimate an order-of-magnitude upper limit to c_{eff} of 80 mM by approximating the 12-residue FliG_M-FliG_C linker as an unstructured peptide with Kuhn length of 0.8 nm^{33,34} (SI Note 3), well above physiological concentrations¹¹. In the *A. aeolicus* crystal structure of the FliG polymer, part of the linking peptide forms an extended helix_{MC} and the remainder an

intermolecular β -sheet⁷. 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 favours polymerization over the compact state and may be necessary to reduce c_{eff} to the level where polymerization *in vivo* can occur on the FliF template (Fig. 4B).

Comments on symmetry mismatch

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 matches perfectly the 34-fold, 45 nm diameter C-ring observed by cryo electron microscopy (EM) in *Salmonella sp.*^{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, non-commensurate interface, is at odds with the observation that fusion of FliF and FliG does not measurably alter the observed motor structure³⁶, 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 wild type, 8 of its 34 FliG sites have no covalent link to the FliF ring. It is possible that these sites are empty²⁹, or that a single FliG straddles two FliM subunits⁸, but this would imply a fractional reduction in the electron density of 34-fold reconstructions of the C-ring, which is not apparent¹³. The straddle model would require ARM_C to interact with FliM³⁷, which is not been seen by yeast two-hybrid interactions³⁸, NMR³⁹, pull-downs²⁷ or electron spin resonance⁸, 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 there to be 8 spare FliF molecules in the FliFG fusion which either hang free or are cleaved from FliFG (Supplementary Fig. 7B).

Concluding remarks

Domain-swapped protein complexes were previously 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 where a domain-swap 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 that ensures that ring closure occurs at the correct polymer length. This provides a structural and thermodynamic model which we envisage will inform the mechanisms behind self-assembly of other protein complexes.

Acknowledgments:

SAXS data was collected on the SAXS/WAXS beamline at the Australian Synchrotron, Victoria, Australia. This work was supported by the Australian Research Council (Grant DP130102219, L.K.L) and the Human Frontiers Science Program (Grant RGP0030/2013 L.K.L, K.N, R.M.B, A.J.T.). L.K.L is supported by the Australian Research Council Discovery Early Career Research Award (Grant DE140100262).

Author Contributions:

M.A.B.B analyzed SAXS data, performed covariance and thermodynamic analysis, conceived experiments and wrote manuscript. R.M.G.H collected SAXS data and performed crosslinking assays. L.A.G expressed and purified protein and collected SAXS data. N.S.M performed crosslinking 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 analyzed SAXS data. C.M.J analyzed SAXS data. N.J.D and J.P.A constructed cell lines for this study. Y.M performed fluorescence imaging experiments. D.S conceived experiments and contributed to crosslinking experiments. A.J.T performed thermodynamic analysis, conceived experiments and wrote manuscript. K.N performed fluorescence imaging experiments, conceived experiments and wrote manuscript. R.M.B performed thermodynamic analysis, conceived

experiments and wrote manuscript. L.K.L conceived experiments, performed crosslinking assays, collected and analyzed SAXS data and wrote manuscript.

References and notes:

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. U. S. A.* **110**, 11839–11844 (2013).
5. Huber, A. 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. 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. U. S. A.* **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. U. S. A.* **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. U. S. A.* **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. U. S. A.* **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. U. S. A.* **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. *Current Biology* **16**, R928–R930–R928–R930 (2006).
29. Paul, K., Gonzalez-Bonet, G., Bilwes, A. M., Crane, B. R. & Blair, D. F. 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. R., 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., Pugalenth, G., Kandaswamy, K. K. & Sowdhamini, R. 3dswap-pred: prediction of 3D domain swapping from protein sequence using Random Forest approach. *Protein and Peptide Letters* **18**, 1010–1020 (2011).
 41. Bennett, M. J., Sawaya, M. R. & Eisenberg, D. Deposition diseases and 3D domain swapping. *Structure* **14**, 811–824 (2006).

Figure Legends

Fig. 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 materials and methods were included in analysis. B) Histogram of intramolecular distances between these residue pairs in the crystal structure of a full-length FliG monomer from *A. aeolicus* (grey, PDBID: 3HJL⁷) 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 of highly correlated covarying residue pairs that are involved in stabilizing the globular structure of each domain. E) Enlarged view of the ARM_M-ARM_C superhelix highlighting the locations of 5 out of the 9 highly correlated pairs within the ARM_M-ARM_C superhelix that are in direct contact. F) 2D distance plot showing all residue pairs closer than 7 Å in the FliG extended monomer (grey) and intermolecular ARM_M-ARM_C superhelix (black); residue pairs with high covariance scores are overlaid in red.

Fig. 2. Solution structure of FliG_{MC} and FliG_{NM}. (A) The extended conformation of FliG_{MC} from the crystal structure of *Aquifex aeolicus* (PDBID:3HJL⁷) 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 where the seven residues that are required to unravel to form the compact conformation are removed and depicted schematically by a dashed line that connects FliG_M to FliG_C. (B) The radius of gyration of FliG_{MC} determined from SAXS is plotted against the sodium chloride concentration in the experimental buffer. The dashed and dotted lines mark the radii of gyration calculated for the compact and extended conformations of FliG_{MC} shown in (A), respectively. Errors shown are $2 \times \text{s.e.m.}$ (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 1000 mM NaCl, with the fitted crystal structure of FliG_{com} superimposed.

Fig. 3. Crosslinking of the ARM_M-ARM_C interface. (A) Rendering of the ARM_M-ARM_C superhelix from the FliG crystal structure from *Aquifex aeolicus* (PDBID:3HJL⁷) 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-HA immunoblots on denaturing SDS PAGE with fractionated samples in the presence and absence of oxidizing agent iodine and DTT. Uncropped gels are in Supplementary Figure 6. The molecular weight of FliG is 37.9 kDa. (C) Immunoblots from whole-cell samples after Cu-phenanthroline-induced disulfide crosslinking in live bacteria. Control disulfide mutants are V218C, M193C V218C, L159C L188C and L159C S191C.

Fig. 4. Schematic depiction of polymerization for the FliG molecule. (A) Schematic depiction of polymer extension in solution. Intra- and inter-molecular ARM_M-ARM_C interactions are shown in red. The MC linker peptide is shown in yellow when it forms an alpha 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 (grey).

Online Methods

Evolutionary covariance analysis

Covarying residues in Aquifex FliG were identified using a maximum entropy approach⁴². A sequence alignment was generated from the uniprot database (www.uniprot.org) by the webservice EVfold¹⁹ using HHblits⁴³ over 2 iterations with an E value of -3 . Alignments are represented by a 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 and sequences with more than 70% gaps were then excluded, leaving a total of 2292 sequences. In addition, covariance analysis was only performed on match state columns, which have less than 50% gaps across the whole alignment (residues 5-325 in PDBID: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 have correlation scores more than 5 standard deviations 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 modified pACYCDuet T7 expression vector (Novagen) with an N-terminal 10×Histidine tag and protease cleavage site. Protein was expressed in T7 expression *E. coli* cells (NEB) in Luria Broth (LB) at 18 degrees 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 passed over a pre-equilibrated NiNTA Sepharose column (GE Healthcare). Beads were washed thoroughly in wash buffer with 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 20 mM Imidazole. The column was then re-equilibration 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 poly-histidine tag producing protein with no affinity tags. Eluted protein was concentrated and further purified by size exclusion chromatography and its monodispersity confirmed with multi-angle laser light scattering (MALS) before being used for SAXS data acquisition.

Small Angle X-ray Scattering

Data collection

Scattering data ($I(q)$ vs q where $q = 4\pi\sin\theta/\lambda$, 2θ is the scattering angle and $\lambda = 1.127 \text{ \AA}$) 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 per second. SAXS data was collected with exposure times of five seconds on a Pilatus 1M photon counting detector (Dectris, Baden Switzerland), which was placed 1.48 meters from the sample capillary.

Data processing and analysis

Data reduction was performed using beamline-specific software package ‘scatterbrain’ (Australian Synchrotron, Clayton Australia, <https://www.synchrotron.org.au/aussyncbeamlines/saxswaxs/software-saxswaxs>). Software for data processing were from the ATSAS suite of programs for SAXS data processing⁴⁴ including DATOP, PRIMUS, DATAVER, DAMMIN, DAMAVER, DATPOROD, CRY SOL and DATCMP. Custom software was written to plot the intensity of scattering data versus frame number. For each dataset, plots were generated of averaged intensities over several q -ranges. This was used to visualise 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 using the program DATOP. The radius of gyration (R_g) was then calculated using AUTORG for each buffer-subtracted frame and the corresponding calculated scattering, assuming a monodisperse sample, plotted over the average measured scattering at low q , which allowed the identification of frames where the calculated R_g was consistent. To confirm that scattering profiles were consistent across these frames, their correlations were evaluated using DATCMP, to perform reduced χ^2 analysis and correlation maps. All frames that were subsequently scaled and averaged for further analysis had a p-value for similarity of 1.000 (27) (Fig S1). Background-subtracted data from monodisperse protein was then scaled and visually inspected for consistency before being averaged and initial Guinier plots and probability atom distance distribution functions ($P(r)$ vs r) plotted using the program PRIMUS. Inspection of unrestrained r_{min} in $P(r)$ plots and high- q data indicated that data were generally under-subtracted. Hence we systematically generated frame-by-frame subtractions where buffer was scaled by 1.0001 – 1.0090 using DATOP and DATAVER. The appropriate buffer scaling was chosen based on the criteria where r_{min} was close to or equal to zero and where high q data ($q > 0.45 \text{ \AA}^{-1}$) 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 1M NaCl with average normalized spatial discrepancy of $0.48 \pm 0.01 \text{ SEM}^{45}$ using the program DAMMIN (Figs. S2-4). To allow a direct comparison, all $P(r)$ vs r plots were scaled to have an area under the curve of 1. At least 20 *ab initio* dummy-atom shape restorations were performed for each dataset, which were aligned with the software DAMMIN. These were converted into pseudo electron density for averaging and scaling using the program pdb2vol from the Situs suite of programs⁴⁶. Pseudo electron density from shape restorations were then averaged in Chimera⁴⁷ and sigma levels set such that the density volume matched the Porod volume for each dataset, which was calculated using the ATSAS program DATPOROD. High resolution X-ray crystal structures were docked into averaged pseudo electron density in Chimera by iteratively maximising the

correlation between structures and density. CRY SOL was used to calculate theoretical scattering data from high-resolution atomic models to compare with SAXS data.

Cross-linking

Plasmids

For crosslinking and functional assays, FliG from *E. coli* strain RP437 was cloned into an arabinose-induced expression vector⁴⁸ with an N-terminal HA peptide for immunodetection in Western Blots. Site-directed mutagenesis to introduce cysteine mutations was performed using long-range plasmid PCR⁴⁹.

Rational design of crosslinking mutants

Custom software was written to rationally identify candidate residue pairs for disulphide crosslinking experiments. The ARMm-ARMc interaction from the *Aquifex aeolicus* crystal structure (PDBID: 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 rotomer combinations were assessed for their similarity to ideal disulphide bond length, angle and torsion angle.

Cell strains and preparation

E. coli cell strain JPA600 lacking FliG was created from the parent strain RP437 by allelic exchange⁵⁰. Deletion of the FliG gene was verified using 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 crosslinking 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 pre-culture in rich Luria Broth (LB) media with 50 ug ml⁻¹ ampicillin. Precultures were incubated at 37°C for 3-5 hours (OD ≈ 1.0) and were used to inoculate minimal tryptone broth (TB) media consisting of 1% tryptone

in phosphate buffered saline and 50 $\mu\text{g ml}^{-1}$ ampicillin to a starting OD of 0.01. The minimal media also contained 0.0025% arabinose to induce protein expression. The concentration of arabinose used was optimised by comparing motility and FliG expression levels in complemented cells with WT cells. Inoculated TB medium was incubated at 30°C until mid log phase (OD = 0.6-0.8) and harvested for functional and crosslinking assays.

Motility 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 non-motile, running and tumbling cells was recorded (Supplementary Fig. 6). Swarm assays were performed on semi-solid agar plates consisting of 0.0025% arabinose, 50 $\mu\text{g/ml}$ ampicillin and 0.35% agar in 1% tryptone with 85.6 mM NaCl. Swarms were inoculated by gently stabbing first a transformed colony and then a swarm plate with a pipette tip. All functional assays were performed at least three times.

Crosslinking assays

For fractionated crosslinking assays, cells at mid-log phase were harvested by centrifugation at $3000 \times g$ for 5 minutes. The supernatant was removed and cells resuspended in lysis buffer consisting of 100 mM Tris-HCl pH 8.0, 150 mM NaCl, lysozyme and DNase 1 and incubated at room temperature for 30 minutes. Cells were then lysed with three freeze-thaw cycles using liquid nitrogen and a water bath at room temperature. Crosslinking 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 left to incubate for 20 seconds before the reaction was stopped with 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 $17000 \times g$ for 15 minutes to harvest the soluble cytoplasmic fraction. The membrane fraction was obtained by resuspending cell debris in 1% triton in 0.1 M Tris-HCl, pH 8.0, and 150 mM NaCl and incubating with gentle agitation at 4°C for 30

minutes. Cell debris was pelleted again: the supernatant contained the solubilised 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* crosslinking experiments, cells were harvested and resuspended in motility buffer before addition of crosslinking reagent to a concentration of 4 mM CuSO₄ and 16 mM 1,10 phenanthroline which was made fresh from stock solution of 400 mM CuSO₄ in 50% ethanol and 1 M 1,10 phenanthroline in 95% ethanol as described previously¹². Samples were incubated with crosslinking reagent for 5 minutes before the reaction was stopped as described above.

Immunoblots

Samples were electrophoresed on a 4-12% Bis-Tris polyacrylamide gel (Invitrogen) for 35 minutes at 200 volts with and without dithiothreitol as appropriate. Protein was transferred to nitrocellulose membranes and blocked with 5% skim milk powder in PBS. Immunoblotting was performed using a 1:15000 dilution of horseradish peroxidase coupled anti HA antibody (Sigma-Aldrich, catalogue number: H6533). Protein bands were observable on nitrocellulose after incubation with HRP substrate (Perkin Elmer)

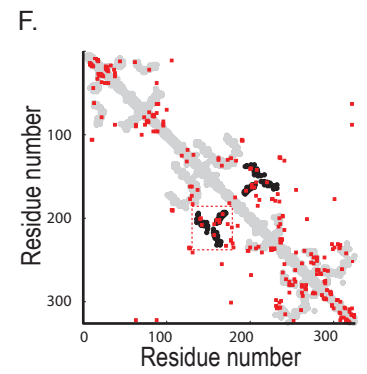
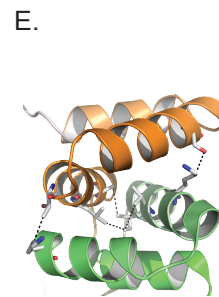
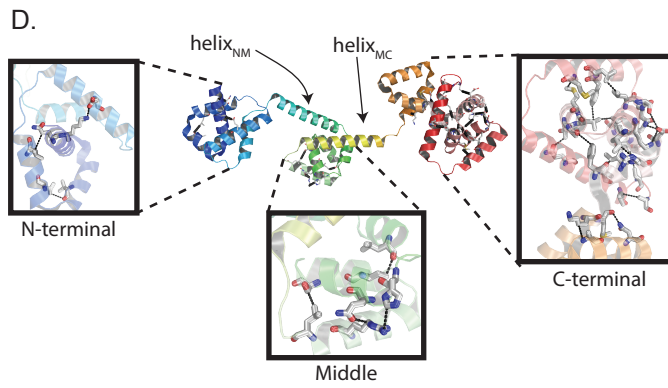
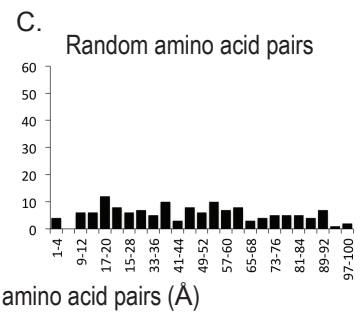
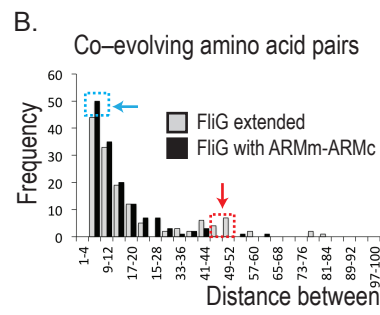
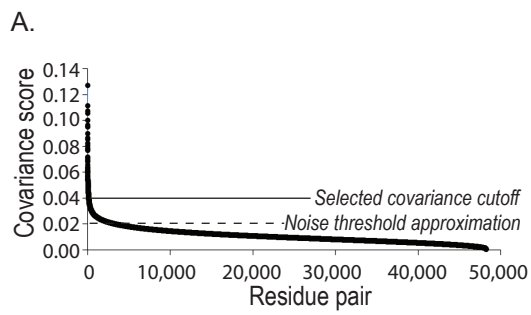
Flagellar filament staining

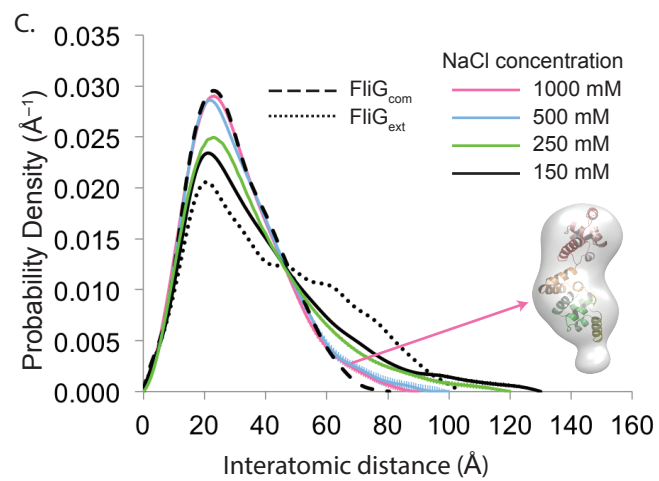
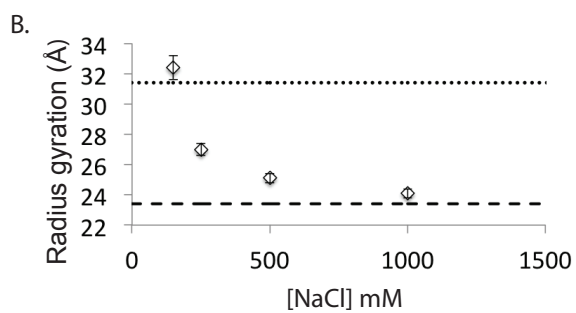
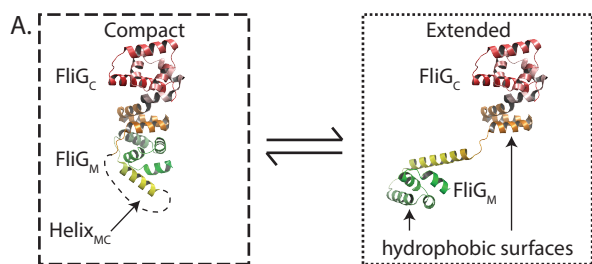
The flagellar filaments produced by *Salmonella* cells were labelled using anti-FliC antiserum and anti-rabbit IgG conjugated with Alexa Fluor 594 (Invitrogen) as described⁵¹. The cells were observed by fluorescence microscopy as described⁵¹. Fluorescence images were processed using ImageJ software version 1.48 (National Institutes of Health).

References for Online Methods

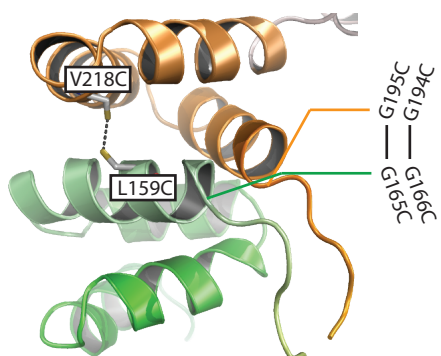
42. Marks, D. S. *et al.* Protein 3D Structure Computed from Evolutionary Sequence Variation. *PLoS One* **6**, e28766 (2011).
43. Johnson, L. S., Eddy, S. R. & Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**, 431 (2010).

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. *The Journal of Bacteriology* **177**, 4121–4130 (1995).
49. Papworth, C., Bauer JC, 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).

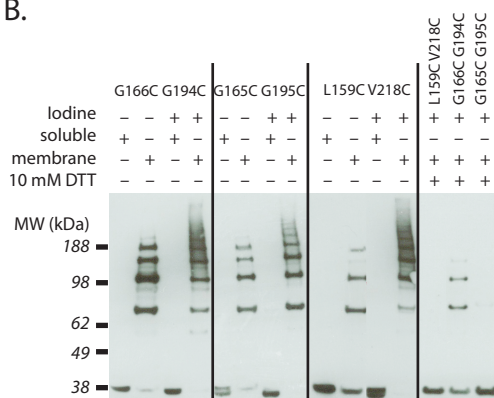




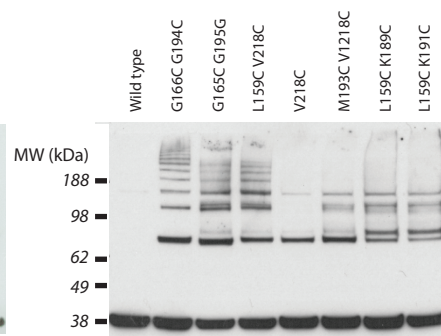
A.



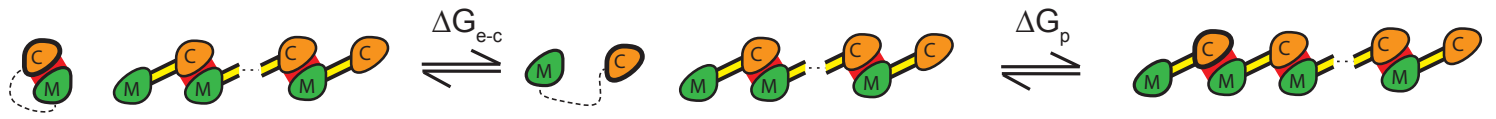
B.



C.



A.



B.

