ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research

Cite this article: dos Santos RN, Khan S, Morcos F. 2018 Characterization of C-ring component assembly in flagellar motors from amino acid coevolution. *R. Soc. open sci.* **5**: 171854.
http://dx.doi.org/10.1098/rsos.171854

Received: 11 November 2017
Accepted: 5 April 2018

Subject Category:
Biochemistry and biophysics

Subject Areas:
computational biology/structural biology/evolution

Keywords:
flagellar motor, bacterial motility, coevolution, structure-based model

Author for correspondence:
Faruck Morcos
e-mail: faruckm@utdallas.edu

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4080011.

THE ROYAL SOCIETY PUBLISHING

Characterization of C-ring component assembly in flagellar motors from amino acid coevolution

Ricardo Nascimento dos Santos[1], Shahid Khan[2] and Faruck Morcos[3,4,5]

[1]Institute of Chemistry and Center for Computational Engineering and Science, University of Campinas, Campinas, SP, Brazil
[2]Molecular Biology Consortium, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[3]Department of Biological Sciences, [4]Department of Bioengineering, and [5]Center for Systems Biology, University of Texas at Dallas, Richardson, TX, USA

FM, 0000-0001-6208-1561

Bacterial flagellar motility, an important virulence factor, is energized by a rotary motor localized within the flagellar basal body. The rotor module consists of a large framework (the C-ring), composed of the FliG, FliM and FliN proteins. FliN and FliM contacts the FliG torque ring to control the direction of flagellar rotation. We report that structure-based models constrained only by residue coevolution can recover the binding interface of atomic X-ray dimer complexes with remarkable accuracy (approx. 1 Å RMSD). We propose a model for FliM–FliN heterodimerization, which agrees accurately with homologous interfaces as well as *in situ* cross-linking experiments, and hence supports a proposed architecture for the lower portion of the C-ring. Furthermore, this approach allowed the identification of two discrete and interchangeable homodimerization interfaces between FliM middle domains that agree with experimental measurements and might be associated with C-ring directional switching dynamics triggered upon binding of CheY signal protein. Our findings provide structural details of complex formation at the C-ring that have been difficult to obtain with previous methodologies and clarify the architectural principle that underpins the ultra-sensitive allostery exhibited by this ring assembly that controls the clockwise or counterclockwise rotation of flagella.

1. Introduction

Flagellar motors (FM) are intricate molecular machines fundamental for bacterial motility. These motors, embedded in the basal body, are powered by electrochemical ion gradients. Transmembrane stator complexes couple ion flux to bidirectional rotation. Modulation of the clockwise (CW)/counter-clockwise (CCW) rotation bias by chemotactic stimuli drives migration in chemical gradients through a biased random walk [1]. CCW and CW intervals as measured in cells tethered by a single flagellum have mean durations of a second or so, but the cells switch between rotation states within milliseconds without detectable changes in rotation speed. Torque is generated when the complexes, acting independently, step along a FliG protein rotor ring attached to the transmembrane MS-ring scaffold [2,3]. A large multi-subunit cytoplasmic ring assembly, the C-ring (figure 1), also forms part of the rotor [3,5,6]. In the eubacteria, *Escherichia coli* and *Salmonella typhimurium*, the C-ring is composed of the FliM and FliN proteins (figure 1). The C-rings of other bacteria contain FliY instead of, or in addition to, FliN [7]. The chemotaxis signal protein CheY binds and tethers to the FliM N-terminal domain to trigger long-range allosteric changes that result in a large reorientation of the FliG helix contacted by the stator complexes and, thereby, rotation reversal [8]. The change in rotation bias is an ultra-sensitive function (Hill coefficient of 10.5) of activated CheY concentration [9,10]. There are about 35 copies of FliM and about three times that number for FliN [11]. The C-ring is a dynamic structure. It loses or gains FliM and FliN subunits to adaptively reset rotation bias upon chemotactic stimulation [12,13]. Accordingly, the subunit symmetry of the C-ring is also variable, but the mean number is 34 [4,14]. Conformational spread across subunits must occur to switch C-ring state within milliseconds [15]. These characteristics emphasize that switching of flagellar rotation provides a remarkable example of long-range allostery whose elucidation will be an important advance for protein science.

A library of X-ray atomic structures for the FliG, FliM and FliN proteins is available, and ring models of their organization have been constructed based on fits to electron density maps (figure 1) [3,4]. The models have been guided by mutagenesis, *in situ* cross-linking and mass spectrometry data [16–18]. More recently, computational methods, particularly residue coevolution, have guided model-building of flagellar protein complexes generally [19] and in FliG specifically [20]. The coevolution signal for FliM is distinct from the FliG middle domain. Inter-subunit contacts make a strong contribution to the signal for the latter, consistent with its proposed central location in the C-ring and its role in conformational spread [21]. There are competing models for the distal C-ring. One model proposes a periodic repeat of a 4:1 complex of FliN C-terminal homo-tetramer with a FliM C-terminal domain ($FliN_C$ and $FliM_C$) [22]. Another model proposes a repeat of a 3:1 complex with a $FliN_C$ homodimer juxtaposed with a $FliM_C$–$FliN_C$ heterodimer [16]. Both models are consistent with the sequence homology between $FliM_C$ and $FliN_C$ as well as *in situ* cross-link data. The coevolutionary signal for the distal C-ring has not been determined thus far.

Here, we present models for the architecture of FliM middle domain ($FliM_M$) and $FliM_C$–$FliN_C$ complexes that address the above issues. Our study is based on the analysis of amino acid coevolution [23–30]. Recent applications of coevolution have focused on its use as a probe of protein structure, conformational states and dynamics [25,31–48]. To study coevolutionary signals, we use direct coupling analysis (DCA), an approach advanced by us and others, for a number of years [24,25,49–52]. Most recently, we used residue coevolution, including DCA together with dynamics extracted from conformational ensembles based on bonding constraints, to detect alternative conformational states of $FliM_M$ [53]. The ensemble approach retains full-atom detail, but lacks temporal information [54]. In this study, we apply structure-based models (SBM) to rigorously evaluate the $FliM_M$ conformational heterogeneity detected by coevolutionary signals. Structure-based models provide a particularly elegant approach based minimally on heavy atom volume exclusion to evaluate coevolutionary constraints [55–57]. We demonstrated in a previous study that SBMs were effective in predicting complex formation from monomer X-ray structures when coevolved couplings are employed as distance constraints to drive MD simulations [45]. Most of these studies have been focused on validating the approach rather than providing new biological insights into the systems studied. This work aims to be an example of how coevolutionary signals and physical modelling can bring a more detailed understanding of the molecular interactions shaped by evolution and at the same time be consistent with low-resolution experimental evidence on the C-ring complexes. We report that the coevolutionary signal for distal C-ring components is enough to reproduce $FliN_C$ homo-dimer association within 1 Å root mean square deviation (RMSD) and to predict $FliM_C$–$FliN_C$ heterodimers. Moreover, two distinct dimer conformations for $FliM_M$ were obtained in the absence of chemical bonding constraints from just coevolutionary constraints and the

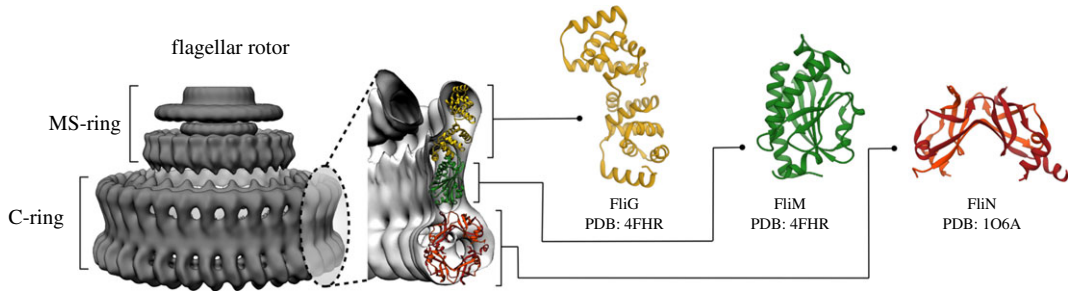

Figure 1. Architecture of flagellar rotor [4]. Proteins FliG, FliM and FliN are the primary building blocks of the C-ring in the basal body of flagella.

monomer X-ray structure. These dimer states precisely agree with *in situ* cross-linking data and are inter-convertible in a way that is consistent with switching kinetics for flagellar rotation.

2. Material and methods

2.1. Protein sequence datasets

For $FliN_C$ and $FliM_C$ proteins, 11 119 sequences were obtained from the Uniprot database and aligned using hidden Markov models (HMM) provided in the Pfam server for FliMN_C domain (family PF01052) [58,59]. The extended version of FliN domain multiple sequence alignment (MSA) was generated using the full sequence of *Thermotoga maritima* FliN (Uniprot G4FDE4) with the HMMER server, selecting only those sequences that have a protein architecture with both FliM and FliMN_C domains [60]. This procedure generated a MSA with 5151 sequences. With respect to FliM, we built a custom MSA using the HMMER server [61]. A HMM profile was created using the FliM sequence of *T. maritima* (Uniprot Q9WZE6) and selecting only sequences that are components of the bacterial flagellum. Sequences with more than 30 consecutive mispairing positions (gaps) were removed, accounting for approximately 15% of the dataset. The resulting MSA used for DCA is composed of 3939 sequences.

2.2. Coevolutionary analysis, sequence mapping and contact selection

Each MSA was processed using the mean field version of direct coupling analysis (mfDCA) [24,25,34]. We model the joint probability of a given protein sequence to be part of a family with a statistical model resulting from entropy maximization.

In order to infer the parameters of these distributions, we use an efficient mean field approach [25]. This approach estimates the parameters of a global probability distribution of amino acid occupancies. The parameters in the distribution can be used to measure how two sites are directly coupled in the family and to infer potential residue–residue interactions. A pairwise 'direct' probability is computed and used in a metric called direct information (DI):

$$\mathrm{DI}_{ij} = \sum_{A,B} P_{ij}^{(\mathrm{dir})}(A,B) \ln \frac{P_{ij}^{(\mathrm{dir})}(A,B)}{f_i(A) f_j(B)}, \tag{2.1}$$

where A and B denote two amino acid types appearing at positions i and j of the MSA (i.e. the i and j columns in an alignment matrix) [25]. In addition, $f_i(A)$ denotes the frequency at which A is observed at position i of the alignment and $P_{ij}^{(\mathrm{dir})}$ computes the pairwise probability of direct correlation [25]. Highly coupled residue pairs ranked by DI values represent pairwise amino acids that plausibly coevolved to satisfy physical requirements to structure and function (i.e. folding, oligomerization, conformational dynamics). The final residue–residue interaction map was generated by domain matching in protein sequences with HMMER, employing an in-house developed mapping script. In order to identify signals that will exclusively drive our predictions towards oligomerization, we discard interdomain folding contacts by combining solvent accessibility and crystallographic data [45]. This step comprises the use of all-atom monomeric X-ray structures from *Thermotoga maritima* (PDB IDs: 4FHR chain A for $FliM_M$ and 1O6A chain A for $FliM_C$ and $FliN_C$) to compute the solvent accessible surface area (SASA) of individual residues employing GetArea [62]. In each case, from that list of top coevolving DCA pairs,

only those pairs for which the summation of SASA values was higher that 50% were considered for modelling studies. For the $FliM_C-N_C$ dimer prediction, due to the lack of a crystallographic data for $FliM_C$ and the fact this system is a heterodimer, we filtered contacts by finding equivalent DCA contacts for $FliN_C$ generated by monomeric filtering and remapping to $FliM_C$ model. The selection of DCA contacts for prediction of oligomeric complexes involving $FliN_C$ and $FliM_C$ (tetramers in the complete ring) was performed by employing a scheme that removes all contacts between DCA pairs that were already in contacts at the predicted dimers. Contact maps were computed considering Cα atom pairs within a distance of 10 Å. In order to compute the interfacial RMSD for the predicted *T. maritima* $FliM_C$–$FliN_C$ heterodimer, the *T. maritima* sequences were aligned to the reference chimeric *Salmonella enterica* sequence (PDB ID: 4YXC), and matching inter-protein pairs within 10 Å were considered in the calculation.

2.3. Structure-based model simulations

The coevolutionary constraints obtained after mapping and filtering DCA pairs were incorporated as Gaussian-like potentials to drive complex formation along SBM MD simulations [63]. An iterative process was used that progressively reduces the equilibrium distance of the constraints to allow proper exploration of the dimeric interface space [45]. Initially, PDB coordinates with separate copies of the monomeric structures for each system were used as input to the SMOG server [56]. The parameter and topology files generated by SMOG were then modified by inserting the SBM potentials generated from the predicted coevolutionary constraints. Finally, MD simulations were developed using a modified version of GROMACS software including support to SBM Gaussian-like potentials [63,64]. The final interaction potential for each residue pair is given below.

$$V_{ij}(r_{ij}) = A_{ij}\left[\left(1 + \left(\frac{1}{A_{ij}}\right) R_{ij}(r_{ij})\right)(1 + G_{ij}(r_{ij})) - 1\right], \tag{2.2}$$

where $R_{ij}(r_{ij})$ accounts for volume exclusion repulsion factor and is given by

$$R_{ij}(r_{ij}) = \epsilon\left(\frac{d}{r_{ij}}\right)^{12}, \tag{2.3}$$

where d and ϵ are the exclusion volume and a normalization constant, respectively. $G_{ij}(r_{ij})$ is the Gaussian attractive interaction:

$$G_{ij}(r_{ij}) = -\epsilon \exp\left[-\frac{(r_{ij} - r_{ij}^{o})^2}{(2w_{ij}^2)}\right]. \tag{2.4}$$

These Gaussian-like potentials allow us to define parameters such as the interaction force amplitude (A_{ij}) and width (w_{ij}) independently of the optimal distance (r_{ij}^{o}) defined for each pair i and j [63].

The final Cα coordinates of each model were used to reconstruct all-atom models by including side-chain optimized rotamers with REMO [65]. For $FliM_C$–$FliN_C$ heterodimer, the monomeric model for $FliM_C$ was generated through the SwissModel server [66] using the $FliM_C$ sequence for *T. maritima* and a crystallographic $FliN_C$ model as a template (PDB ID: 1O6A, chain A).

Because the complexes in this work were obtained by using one domain family, the predicted pairs present a unique configuration in which residues i and j are represented by the first and second indexed monomeric chain of the system, respectively. In order to consider the possibility of residue j in the first chain interacting with i in the second, we also included a second specular constraint map (j and i) to generate the SBM potentials. To perform the step-by-step approximation between monomers during simulations, we started the system with two distant and randomly oriented separated monomers and progressively changed the parameters of equilibrium distance (r) and potential width (w) from SBM potentials. This procedure was performed in MD simulation steps of 10 ns using the following combination of parameters (r/w): 50 Å/2, 30 Å/2, 15 Å/4, 11 Å/4, 6.4 Å/4, 6.4 Å/2 and 6.4 Å/0.5. All steps were performed at a relative temperature of 120 (normalized in SMOG implementation, see [56]). The reduction in the Gaussian width constant (w) in the last step helps to refine the interface and to optimize the accuracy of the complex modelling. Furthermore, in order to isolate the real contribution of DCA pairs in complex predictions, we performed MD simulations for negative controls. These control simulations are further detailed in the electronic supplementary material.

3. Results

3.1. Coevolutionary signals recapitulate FliN dimerization and help model lower C-ring formation

As a way to explore the feasibility of using coevolutionary signals to investigate complex formation in families of Fli proteins, we analysed sequences of proteins from FliM and FliN families using DCA [25,50] to predict C-ring molecular complexes [45]. In an initial step, employing exclusively unbiased coevolutionary signals and physically separated monomeric structures, we predicted the association of a $FliN_C$ homodimer from *T. maritima* and compared it with reported structures for this complex (PDB ID: 1O6A). We analysed sequences in the family for the C-terminal domain shared by both FliM and FliN proteins (FliMN_C, Pfam PF01052) using DCA (see Material and methods section). This analysis allowed us to observe high coevolutionary signals for FliN by the identification of populated contact regions superposing to monomeric contacts and captured in the top-ranked DCA correlations (electronic supplementary material, figure S3). Figure 2*a* shows the top DI couplings for a MSA of this dataset after the application of a filtering process to remove contacts related to folding and to isolate contacts mainly originating from dimerization (see Material and methods section). When compared to the native contacts from a crystal structure of a $FliN_C$ homodimer, it is possible to observe representative clusters of coevolving residue pairs that coincide with dimeric regions (dashed circles in figure 2*a* upper-left corner). These residue pairs (200 pairs with highest coevolving signal and filtered for solvent accessibility) were used to build a physical potential in coarse-grained molecular dynamics (MD) simulations for the prediction of the homodimeric FliN complex (figure 2*b*). In this process, two copies of the X-ray crystallographic structure of chain A from FliN (PDB ID: 1O6A) were used to predict a homodimer. This methodology takes advantage of the cooperative nature of native interactions and is robust to noisy predicted interactions (see controls in electronic supplementary material, figures S1 and S2). Using this procedure, we were able to recapitulate with high accuracy the correct FliN homodimer conformation, with a lowest RMSD of only 0.81 Å (figure 2*c*; electronic supplementary material, Movie S1) and an average RMSD of 1.97 Å over the last simulation stage (see Material and methods section). Representative DCA pairs used as potentials to drive dimer association are shown in the three-dimensional depiction in electronic supplementary material, figure S4.

3.2. Characterization of the FliM–FliN interaction interface

Encouraged by our success in recovering the biologically meaningful conformation of the FliN homodimer using only coevolutionary signals, we decided to test the viability of a biologically relevant $FliM_C$–$FliN_C$ heterodimer. The $FliM_C$ monomer for *T. maritima* was modelled using the sequence corresponding to the $FliMN_C$ domain and the previously cited crystallographic data to the $FliN_C$ chain A as a template (PDB ID: 1O6A). DCA couplings from the $FliMN_C$ family were also mapped onto the $FliM_C$ sequence, and the resulting predicted coevolving interfacial residues between $FliN_C$ and $FliM_C$ were used to drive the complex formation of this heterodimer. The interface of the resulting complex (figure 3*a*; electronic supplementary material, figure S5 and Movie S2) is consistent with previous studies [67]. The geometry is preserved, and the interface we recover agrees accurately with the experimental evidence [67], with an interfacial structural alignment with RMSD of 1.31 Å when compared with a fusion protein comprising both portions of $FliM_C$ and $FliN_C$ from *S. enterica* (PDB ID: 4YXC).

The reported model describing the lower part of the C-ring as an array of closed-ring-like $FliN_C$ homotetramers intercalated by $FliM_C$ monomers [22] is supported by the interpretation of cross-linking data and observed mutational effects. In order to evaluate the agreement of our predicted heterodimer with these data, we compared the reported cross-linking experiments between $FliM_C$ and $FliN_C$ with the interfacial contacts of this complex. Figure 3*b* shows that the complex predicted for $FliM_C$ and $FliN_C$ presents a different interface from the one originally proposed based on experimental data [22]. In the predicted heterodimer, instead of a side-to-side interaction between $FliN_C$ and $FliM_C$ that would be compatible with the idea of $FliN_C$ homotetramers intercalated by single $FliM_C$ monomers, we observe a heterodimer conformation that shares the same interface of that observed in $FliN_C$ homodimers. This possibility of a common dimerization interface for $FliN_C$ and $FliM_C$–$FliN_C$ was recently evidenced by other studies [16]. Remarkably, even though the predicted binding mode differs from the one originally proposed based on experimental data, it satisfies the related observed cross-links between $FliM_C$ and $FliN_C$, with dimeric interacting pairs coinciding with the regions of most cross-linking occurring in both FM directional rotation modes (red points in figure 3*b*). Furthermore, this correspondence is more evident when we consider the agreement between cross-linking data observed only in the FM CW rotation mode,

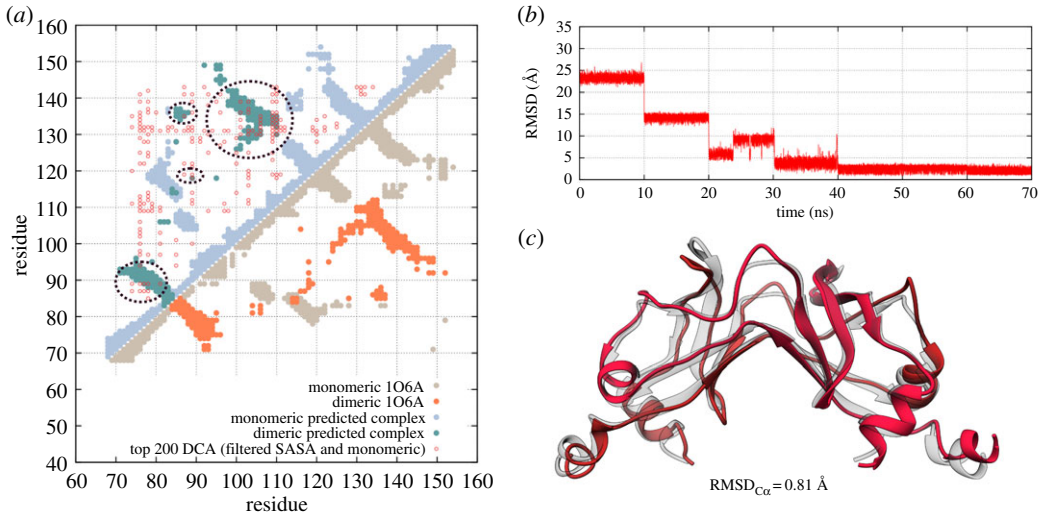

Figure 2. Prediction of the FliN$_C$ homodimer. (*a*) Comparison between residue contacts in the crystal structure (above the diagonal in blue circles) and the DCA-predicted structure (lower diagonal in orange circles) for FliN dimerization. Interactions predicted from DCA and used to generate the FliN complex from FliN monomers are also shown (red circles, upper left corner). (*b*) RMSD between the predicted FliN$_C$ homodimer and the crystal structure used as reference (PDB ID: 1O6A) for the simulation. (*c*) The best predicted model (in red) presented a RMSD of 0.81 Å when compared to the reference model (in white).

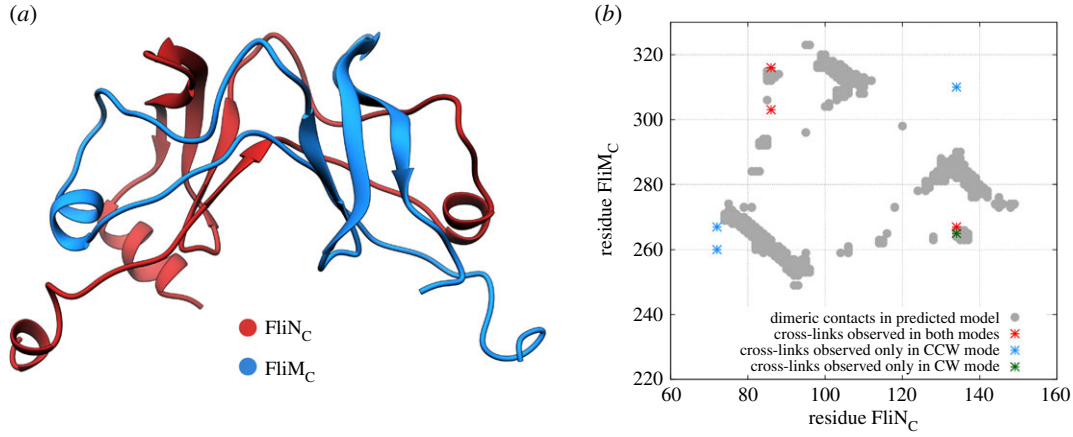

Figure 3. Prediction of the FliM$_C$–FliN$_C$ heterodimer. (*a*) Three-dimensional structure of the predicted heterodimer. (*b*) Comparison of dimeric contacts for the predicted FliM$_C$–FliN$_C$ heterodimer that were observed as cross-links [22]. Good agreement between the model contacts and cross-links is observed, particularly for links common to both rotation modes and those unique to CW rotation.

as denoted by the green point in figure 3*b*. Moreover, it has been reported that specific mutations in FliN$_C$ are able to prevent its binding to FliM$_C$ [22]. Electronic supplementary material, figure S6 shows that most of these residues (five out of six, depicted in green) are accurately located at the interface of the proposed complex, where they contribute to dimer stabilization. Altogether, these results support the idea of a coexistence of functional FliN$_C$ homodimers and FliM$_C$–FliN$_C$ heterodimers, as proposed by McDowell *et al.* [16].

3.3. FliM oligomerization and interface switching

FliM is one of the fundamental components of the flagellar basal body. Its homo-oligomerization in a 34-fold ring is reported to constitute the central core of the C-ring (figure 1), whereas the outer rings are formed mainly by FliN and FliG [68–71]. Several studies suggest FliM is directly involved in C-ring conformational changes related to the direction of flagellar rotation [53,68,69,71]. Furthermore, cryo-electron microscopy and cross-linking experiments point out a possible side-to-side symmetrical

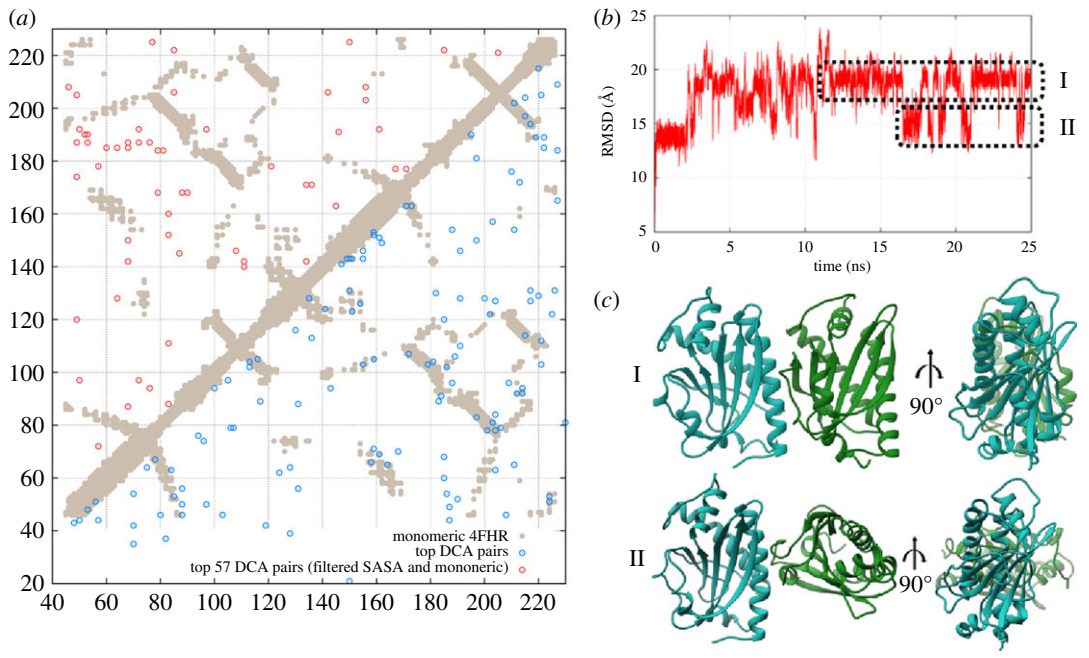

Figure 4. (*a*) Coevolutionary analysis of the $FliM_M$ domain family. DCA contacts (blue circles) reproduced the monomeric contacts occurring in the $FliM_M$ crystal structure (beige dots, PDB ID: 4FHR) [70]. Dimeric contacts (red circles) were selected from the combined monomeric/dimeric signals by excluding solvent inaccessible and monomeric contacts. (*b*) MD simulation using DCA contacts to predict $FliM_M$ homo-dimeric complex formation. At the final stage of the simulation, an oscillation of dimeric complexes was observed in two distinct modes (I and II). (*c*) In dimeric model I for $FliM_M$, monomers interact through side-to-side contacts to promote a parallel configuration. In model II, monomers interact in a twisted configuration.

association among FliM subunits that could explain ring oligomerization, but molecular details about the interfaces for this multi-complex still remain unsolved [4,21,69,70]. In order to uncover the most plausible oligomerization interface for FliM involved in the C-ring formation, we analysed the family of the middle domain of FliM ($FliM_M$) to extract evolutionary hints about dimerization to be able to infer complex formation using MD simulations with SBMs. As a first step, we identified a relatively abundant set of FliM homologues (see Material and methods section for details) with 3939 sequences corresponding to $FliM_M$ (Pfam PF02154). The MSA of this dataset served as the basis for coevolutionary analysis with DCA. The identified coevolved residues at the interface were then mapped onto the *T. maritima* FliM sequence, and the predicted contacts were compared with those reported on a crystal structure from PDB 4FHR (figure 4*a*). We recovered clear signals from monomeric contacts, but we also observed clusters of reside–residue correlations outside of the monomeric conformation. Following our methodology for dimerization prediction, signals from non-solvent accessible residue pairs and interaction pairs related to folding were excluded. The remaining directly correlated residue pairs (figure 4*a*) were then used in a simulation to drive the dimerization of two monomeric $FliM_M$ structures using chain A of the structure 4FHR (see Material and methods section).

We computed MD trajectories for two $FliM_M$ monomers using highly coupled evolutionary pairs as interaction constraints in our computational model. Figure 4*b* shows the RMSD for the trajectory of the complex formation simulation. In the last stage of the simulation, two interaction modes appear to dominate the conformational space by switching of the dimeric interface in a dual interaction mode (figure 4*b*; electronic supplementary material, Movie S3). Clustering trajectories in both modes allowed us to identify two distinct binding interfaces, characterized by a parallel and a twisted orientation between each $FliM_M$ monomer (figure 4*c*). A three-dimensional depiction of representative DCA pairs that promoted both twisted and parallel conformations is shown in electronic supplementary material, figure S7. Comparison of the dimeric contacts formed in each configuration showed that residues involved in complex formation have a common cluster region in the quadrant formed by residues 50–80 and 175–190 (figure 5). Furthermore, interfaces I and II both consist of about the same number of contacts, with further distinct secondary clusters of dimeric contacts. We have observed similar features of conformational plasticity in ligand-binding proteins using coevolutionary information, where

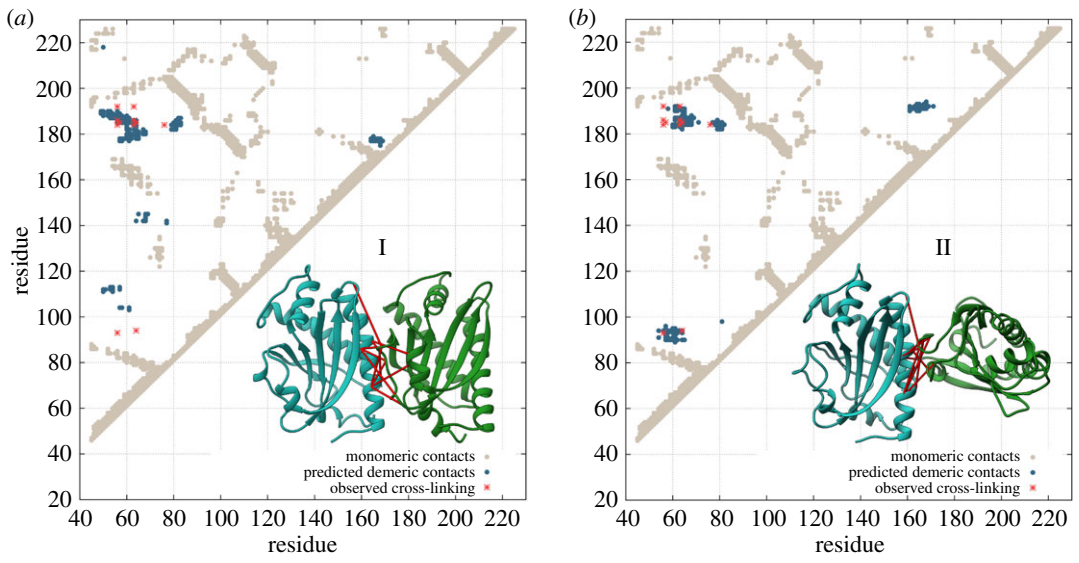

Figure 5. Dimeric contacts at the interface of the predicted models for $FliM_M$ oligomerization. (*a*) Model I, in a parallel side-to-side configuration and (*b*) model II, in a twisted perpendicular configuration. Both models present an equivalent number of contacts and a common interface interaction region that highly correlates with experimental data from cross-linking experiments (red connections in three-dimensional models) [69].

functional modes are encoded by evolution [31]. This is the first time that we have observed similar features for protein complexes.

In order to provide further evidence for the validity of our predictions, we compared the resulting dimeric contacts with reported data for cross-linking experiments with $FliM_M$ oligomers [69]. Figure 5 shows that the most populated region of predicted dimeric contacts in both complexes is located exactly at the interface of previously reported cross-linking pairs (see red markers) related to oligomerization [8,21,69]. Remarkably, although model I agrees very well with the parallel side-to-side $FliM_M$ oligomerization suggested by EM density map fitting [70], model II with a twisted dimerization interface is the one that accurately satisfies all reported cross-linking evidence, with a secondary interface region exactly superposing two disulfide cross-linking Cys pairs (56/93 and 64/94) from a mutant $FliM_M$ [69]. This higher correspondence between cross-linking and the predicted interface can also be observed as shorter pairwise distances in the three-dimensional representation of model II when compared to model I (figure 5).

It is notable that although the parallel binding mode is the one that best fits the EM map [4] (representing CW rotation state), the perpendicular mode predicted in this study is the one that accurately satisfies all cross-linking data. When we measure the distance between Cα atoms in our proposed models for all the cross-linking pairs, we observe that the perpendicular model has on average smaller distances compared with the parallel model (electronic supplementary material, table S1) [69]. In addition, both predicted complex configurations were independently obtained based solely on coevolutionary signals. Together, these facts suggest that these parallel and twisted predicted interfaces can represent functional conformations of a FliM switching mechanism related to CW and CCW directional rotation of FM, respectively.

4. Discussion

$FliN_C$ and $FliM_C$ interact to form the membrane-distal portion of the C-ring in flagellar motors. However, the precise framework and interfaces involved in the association of these proteins is still a source of questions. While some studies suggest this region is composed of $FliN_C$ homotetramers intercalated by $FliM_C$ monomers [22], other results indicate a common functional dimerization interface between $FliN_C$ and $FliM_C$ [16]. Despite these findings, only chimeric constructions with fused $FliM_C$ and $FliN_C$ domains structure have been reported [67]. Furthermore, the molecular details of FliM oligomerization to build the central portion of C-ring and its conformation involved in FM switching are still elusive. Amino acid coevolution has been in the spotlight as a breakthrough methodology to identify molecular

interactions when experimental measurements are inadequate. Methods using coevolution have finally reached a mature state that allows us to make predictions rather than focusing on validations. Moreover, we have developed an efficient methodology to identify interfaces involved in complex formation [45]. Here, we have shown a biologically relevant application of this approach to predict precisely the interface involved in the formation of FliN homodimers (figure 2). Furthermore, we show that, using our methodology, coevolutionary signals between $FliM_C$ and $FliN_C$ proteins are able to generate a stable heterodimer (figure 3) that shares features found in $FliN_C$ homodimers. The possibility of reproducing this conformation entirely from coevolution suggests that this heterodimer is a functional complex and supports the idea that $FliN_C$ and $FliM_C$ share the same interaction interface. This interaction pattern could lead to the spiral-like framework formed by the repetitive association of $FliM_C$–$FliN_C$ heterotetramers in a proportion of 1 : 3, as proposed by McDowell *et al.* [16]. The predicted model for $FliM_C$–$FliN_C$ heterodimer is also in conformity with $FliN_C$ mutations and $FliM_C$–$FliN_C$ disulfide cross-linking data that support the $FliN_C$ doughnut-shaped homotetramer model (figure 3*a*; electronic supplementary material figure S6) [22].

FliM is located in the central part of the C-ring and plays a key role in mediating signal propagation for switching of FM directional rotation through conformational changes. Currently, there are only models for $FliM_M$ oligomerization, generated from EM density fitting [70,71] and sparse cross-linking data [69]. This model describes $FliM_M$ monomers arranged in a parallel side-to-side orientation, forming a symmetrical ring-like framework, although the details of their interface for oligomerization remains unclear. Predictions of $FliM_M$ complexes using coevolutionary information allowed us to identify two stable binding interfaces that are able to switch between themselves (figure 4*b*,*c*). While one of these homodimer models (figure 5*a*) agrees very well with the arrangement of parallel side-to-side $FliM_M$ monomers and satisfies a fraction of the reported FliM disulfide cross-links [69], a second, twisted model, in which both monomers interact in a perpendicular orientation presents interface contacts that reproduce precisely those observed in disulfide cross-links experiments (figure 5*b*). Moreover, this predicted twisted interface involves contacts over a unique region encompassing residues 160–170 and 190–195 of FliM (figure 5*b*). Therefore, alternative cross-linking experiments involving this region in wild-type FliM (e.g. zero-length pairing of Ser167 and Glu193) could provide additional evidence to corroborate this alternative conformational state, as well as the correspondence between the predicted interface and the cross-linking data.

The observation of dimeric contacts that match exactly to experimental cross-linking that are not seen in the suggested parallel $FliM_M$ dimerization model [69,70] (cross-links at residues 56/93 and 64/94) suggests that the twisted $FliM_M$ complex identified (figure 5*b*) could depict a novel conformational state of the basal body when the FM is in CCW flagellar rotation. We believe that the conformation shown in model II might be connected with a transient conformational state and could represent a strained ring conformation. FliM has an $\alpha/\beta/\alpha$ domain with considerable flexibility, as we showed in an early study [53]. Contacts formed by adjacent FliG, FliM (C-terminal) and FliN proteins could lock-in into a strained ring state consistent with the dynamic loss/gain of FliM subunits. We also envision rings where not all the dimeric interfaces are identical; it is possible to find a mixture of interfaces *in vivo* where model I and model II coexist. Possible more-intricate interactions should also be considered in this system, such as the intercalation of identical dimers by switching of activation factors and the simultaneous presence of both models I and II. This information could provide further insights into the overall molecular mechanism involved in the FM switching system. Together, these studies shed light on unresolved questions about the formation of complex structures in flagellar systems. They suggest novel dynamic properties of FliM dimerization that can be tested experimentally. We hope that our results can serve as a basis for further research aimed at unraveling more molecular details about the switching mechanisms of bacterial flagellar motors.

Data accessibility. All research materials supporting the results of studies reported in this manuscript are available in the electronic supplementary material, consisting of figures, tables and simulation trajectory movies. All raw metadata used to derive the studies are deposited in Dryad Digital Repository: (http://dx.doi.org/10.5061/dryad.0mv6t) [72].

Authors' contributions. R.N.d.S., S.K. and F.M. designed the study. R.N.d.S. developed the study. R.N.d.S. and F.M. conducted analyses of data. R.N.d.S., S.K. and F.M wrote the manuscript. All the authors discussed the results and gave their final approval for publication.

Competing interests. We declare we have no competing interests.

Funding. This work received funding from the University of Texas at Dallas (F.M.); São Paulo Research Foundation (FAPESP, grant 2015/13667-9) (R.N.d.S.) and seed funds from the Molecular Biology Consortium (S.K.).

Acknowledgements. We thank Xianli Jiang from the Department of Biological Sciences at UT Dallas for her help during data collection and revision of this manuscript.

References

1. Berg HC. 2003 The rotary motor of bacterial flagella. *Ann. Rev. Biochem.* **72**, 19–54. (doi:10.1146/annurev.biochem.72.121801.161737)
2. Brown PN, Terrazas M, Paul K, Blair DF. 2007 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. (doi:10.1146/annurev.biochem.72.121801.161737)
3. Lynch MJ, Levenson R, Kim EA, Sircar R, Blair DF, Dahlquist FW, Crane BR. 2017 Co-folding of a FliF-FliG split domain forms the basis of the MS:C ring interface within the bacterial flagellar motor. *Structure* **25**, 317–328. (doi:10.1016/j.str.2016.12.006)
4. Thomas DR, Francis NR, Xu C, DeRosier DJ. 2006 The three-dimensional structure of the flagellar rotor from a clockwise-locked mutant of *Salmonella enterica* Serovar Typhimurium. *J. Bacteriol.* **188**, 7039–7048. (doi:10.1128/JB.00552-06)
5. Francis NR, Sosinsky GE, Thomas D, DeRosier DJ. 1994 Isolation, characterization and structure of bacterial flagellar motors containing the switch complex. *J. Mol. Biol.* **235**, 1261–1270. (doi:10.1006/jmbi.1994.1079)
6. Khan IH, Reese TS, Khan S. 1992 The cytoplasmic component of the bacterial flagellar motor. *Proc. Natl Acad. Sci. USA* **89**, 5956–5960. (doi:10.1073/pnas.89.13.5956)
7. Sircar R, Greenswag AR, Bilwes AM, Gonzalez-Bonet G, Crane BR. 2013 Structure and activity of the flagellar rotor protein fliy: a member of the CheC phosphatase family. *J. Biol. Chem.* **288**, 13 493–13 502. (doi:10.1074/jbc.M112.445171)
8. Paul K, Brunstetter D, Titen S, Blair DF. 2011 A molecular mechanism of direction switching in the flagellar motor of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **108**, 17 171–17 176. (doi:10.1073/pnas.1110111108)
9. Cluzel P, Surette M, Leibler S. 2000 An ultrasensitive bacterial motor revealed by monitoring signaling proteins in single cells. *Science* **287**, 1652–1655. (doi:10.1126/science.287.5458.1652)
10. Yuan J, Berg HC. 2013 Ultrasensitivity of an adaptive bacterial motor. *J. Mol. Biol.* **425**, 1760–1764. (doi:10.1016/j.jmb.2013.02.016)
11. Zhao R, Pathak N, Jaffe H, Reese TS, Khan S. 1996 FliN is a major structural protein of the C-ring in the *Salmonella typhimurium* flagellar basal body. *J. Mol. Biol.* **261**, 195–208. (doi:10.1006/jmbi.1996.0452)
12. Branch RW, Sayegh MN, Shen C, Nathan VSJ, Berg HC. 2014 Adaptive remodelling by FliN in the bacterial rotary motor. *J. Mol. Biol.* **426**, 3314–3324. (doi:10.1016/j.jmb.2014.07.009)
13. Yuan J, Branch RW, Hosu BG, Berg HC. 2012 Adaptation at the output of the chemotaxis signalling pathway. *Nature* **484**, 233–236. (doi:10.1038/nature10964)
14. Young HS, Dang H, Lai Y, DeRosier DJ, Khan S. 2003 Variable symmetry in *Salmonella typhimurium* flagellar motors. *Biophys. J.* **84**, 571–577. (doi:10.1016/S0006-3495(03)74877-2)
15. Bai F *et al.* 2010 Conformational spread as a mechanism for cooperativity in the bacterial flagellar switch. *Science* **327**, 685–689. (doi:10.1126/science.1182105)
16. McDowell MA *et al.* 2016 Characterisation of *Shigella* Spa33 and *Thermotoga* FliM/N reveals a new model for C-ring assembly in T3SS. *Mol. Microbiol.* **99**, 749–766. (doi:10.1111/mmi.13267)
17. Paul K, Gonzalez-Bonet G, Bilwes AM, Crane BR, Blair DF. 2011 Architecture of the flagellar rotor. *EMBO J.* **30**, 2962–2971. (doi:10.1038/emboj.2011.188)
18. Sircar R, Borbat PP, Lynch MJ, Bhatnagar J, Beyersdorf MS, Halkides CJ, Freed JH, Crane BR. 2015 Assembly states of FliM and FliG within the flagellar switch complex. *J. Mol. Biol.* **427**, 867–886. (doi:10.1016/j.jmb.2014.12.009)
19. Taylor WR, Matthews-Palmer TRS, Beeby M. 2016 Molecular models for the core components of the flagellar type-III secretion complex. *PLoS ONE* **11**, 1–33. (doi:10.1371/journal.pone.0164047)
20. Baker MAB *et al.* 2016 Domain-swap polymerization drives the self-assembly of the bacterial flagellar motor. *Nat. Struct. Mol. Biol.* **23**, 197–203. (doi:10.1038/nsmb.3172)
21. Pandini A, Kleinjung J, Rasool S, Khan S. 2015 Coevolved mutations reveal distinct architectures for two core proteins in the bacterial flagellar motor. *PLoS ONE* **10**, e0142407. (doi:10.1371/journal.pone.0142407)
22. Sarkar MK, Paul K, Blair DF. 2010 Subunit organization and reversal-associated movements in the flagellar switch of *Escherichia coli*. *J. Biol. Chem.* **285**, 675–684. (doi:10.1074/jbc.M109.068676)
23. Gbel UJ, Sander C, Schneider R, Valencia A. 1994 Correlated mutations and residue contacts in proteins. *Proteins: Struct. Funct. Bioinf.* **18**, 309–317. (doi:10.1002/prot.340180402)
24. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. 2009 Identification of direct residue contacts in protein-protein interaction by message passing. *Proc. Natl Acad. Sci. USA* **106**, 67–72. (doi:10.1073/pnas.0805923106)
25. Morcos F *et al.* 2011 Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl Acad. Sci. USA* **108**, E1293–E1301. (doi:10.1073/pnas.1111471108)
26. Cocco S, Monasson R, Weigt M. 2013 From principal component to direct coupling analysis of coevolution in proteins: low-eigenvalue modes are needed for structure prediction. *PLoS Comput. Biol.* **9**, 1–17. (doi:10.1371/journal.pcbi.1003176)
27. de Juan D, Pazos F, Valencia A. 2013 Emerging methods in protein co-evolution. *Nat. Rev. Genet.* **14**, 249–261. (doi:10.1038/nrg3414)
28. Skwark MJ, Abdel-Rehim A, Elofsson A. 2013 PconsC: combination of direct information methods and alignments improves contact prediction. *Bioinformatics* **29**, 1815–1816. (doi:10.1093/bioinformatics/btt259)
29. Skwark MJ, Raimondi D, Michel M, Elofsson A. 2014 Improved contact predictions using the recognition of protein like contact patterns. *PLoS Comput. Biol.* **10**, e1003889. (doi:10.1371/journal.pcbi.1003889)
30. Haldane A, Flynn WF, He P, Vijayan RSK, Levy RM. 2016 Structural propensities of kinase family proteins from a Potts model of residue co-variation. *Protein Sci.* **25**, 1378–1384. (doi:10.1002/pro.2954)
31. Morcos F, Jana B, Hwa T, Onuchic JN. 2013 Coevolutionary signals across protein lineages help capture multiple protein conformations. *Proc. Natl Acad. Sci. USA* **110**, 20 533–20 538. (doi:10.1073/pnas.1315625110)
32. Sfriso P, Duran-Frigola M, Mosca R, Emperador A, Aloy P, Orozco M. 2016 Residues coevolution guides the systematic identification of alternative functional conformations in proteins. *Structure* **24**, 116–126. (doi:10.1016/j.str.2015.10.025)
33. Pazos F, Valencia A. 2008 Protein co-evolution, co-adaptation and interactions. *EMBO J.* **27**, 2648–2655. (doi:10.1038/emboj.2008.189)
34. Sukowska JI, Morcos F, Weigt M, Hwa T, Onuchic JN. 2012 Genomics-aided structure prediction. *Proc. Natl Acad. Sci. USA* **109**, 10 340–10 345. (doi:10.1073/pnas.1207864109)
35. Marks DS, Hopf TA, Sander C. 2012 Protein structure prediction from sequence variation. *Nat. Biotechnol.* **30**, 1072–1080. (doi:10.1038/nbt.2419)
36. Hopf TA, Colwell LJ, Sheridan R, Rost B, Sander C, Marks DS. 2012 Three-dimensional structures of membrane proteins from genomic sequencing. *Cell* **149**, 1607–1621. (doi:10.1016/j.cell.2012.04.012)
37. Ferguson A, Mann J, Omarjee S, Ndung'u T, Walker B, Chakraborty A. 2013 Translating HIV sequences into quantitative fitness landscapes predicts viral vulnerabilities for rational immunogen design. *Immunity* **38**, 606–617. (doi:10.1016/j.immuni.2012.11.022)
38. Jana B, Morcos F, Onuchic JN. 2014 From structure to function: the convergence of structure based models and co-evolutionary information. *Phys. Chem. Chem. Phys.: PCCP* **16**, 6496–6507. (doi:10.1039/c3cp55275f)
39. Cheng RR, Morcos F, Levine H, Onuchic JN. 2014 Toward rationally redesigning bacterial two-component signaling systems using coevolutionary information. *Proc. Natl Acad. Sci. USA* **111**, E563–E571. (doi:10.1073/pnas.1323734111)
40. Ovchinnikov S, Kamisetty H, Baker D. 2014 Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife* **2014**, e02030. (doi:10.7554/eLife.02030)
41. Michel M, Hayat S, Skwark MJ, Sander C, Marks DS, Elofsson A. 2014 PconsFold: improved contact predictions improve protein models. *Bioinformatics* **30**, i482–i488. (doi:10.1093/bioinformatics/btu458)
42. Hayat S, Sander C, Marks DS, Elofsson A. 2015 All-atom 3D structure prediction of transmembrane β-barrel proteins from sequences. *Proc. Natl Acad. Sci. USA* **112**, 5413–5418. (doi:10.1073/pnas.1419956112)
43. Espada R, Parra RG, Mora T, Walczak AM, Ferreiro DU. 2015 Capturing coevolutionary signals inrepeat proteins. *BMC Bioinformatics* **16**, 207. (doi:10.1186/s12859-015-0648-3)
44. Ovchinnikov S *et al.* 2017 Protein structure determination using metagenome sequence data. *Science* **355**, 294–298. (doi:10.1126/science.aah4043)
45. dos Santos RN, Morcos F, Jana B, Andricopulo AD, Onuchic JN. 2015 Dimeric interactions and complex formation using direct coevolutionary

couplings. *Sci. Rep.* **5**, 13652. (doi:10.1038/srep13652)

46. Sutto L, Marsili S, Valencia A, Gervasio FL. 2015 From residue coevolution to protein conformational ensembles and functional dynamics. *Proc. Natl Acad. Sci. USA* **112**, 13 567–13 572. (doi:10.1073/pnas.1508584112)
47. Cheng RR, Nordesjö O, Hayes RL, Levine H, Flores SC, Onuchic JN, Morcos F. 2016 Connecting the sequence-space of bacterial signaling proteins to phenotypes using coevolutionary landscapes. *Mol. Biol. Evol.* **33**, 3054–3064. (doi:10.1093/molbev/msw188)
48. Malinverni D, Jost Lopez A, De Los Rios P, Hummer G, Barducci A. 2017 Modeling Hsp70/Hsp40 interaction by multi-scale molecular simulations and co-evolutionary sequence analysis. *eLife* **6**, e23471. (doi:10.7554/eLife.23471)
49. Ekeberg M, Lövkvist C, Lan Y, Weigt M, Aurell E. 2013 Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models. *Phys. Rev. E - Stat. Nonlinear Soft Matter Phys.* **87**, 012707. (doi:10.1103/PhysRevE.87.012707)
50. Morcos F, Hwa T, Onuchic JN, Weigt M. 2014 Direct coupling analysis for protein contact prediction. In *Protein structure prediction* (ed. D Kihara), pp. 55–70. New York, NY: Human Press. (doi:10.1007/978-1-4939-0366-5)
51. Morcos F, Schafer NP, Cheng RR, Onuchic JN, Wolynes PG. 2014 Coevolutionary information, protein folding landscapes, and the thermodynamics of natural selection. *Proc. Natl Acad. Sci. USA* **111**, 12 408–12 413. (doi:10.1073/pnas.1413575111)
52. Coucke A, Uguzzoni G, Oteri F, Cocco S, Monasson R, Weigt M. 2016 Direct coevolutionary couplings reflect biophysical residue interactions in proteins. *J. Chem. Phys.* **145**, 174102. (doi:10.1063/1.4966156)
53. Pandini A, Morcos F, Khan S. 2016 The gearbox of the bacterial flagellar motor switch. *Structure* **24**, 1209–1220. (doi:10.1016/j.str.2016.05.012)
54. Seeliger D, Haas J, de Groot BL. 2007 Geometry-based sampling of conformational transitions in proteins. *Structure* **15**, 1482–1492. (doi:10.1016/j.str.2007.09.017)
55. Noel JK, Levi M, Raghunathan M, Lammert H, Hayes RL, Onuchic JN, Whitford PC. 2016 SMOG 2: a versatile software package for generating structure-based models. *PLoS Comput. Biol.* **12**, e1004794. (doi:10.1371/journal.pcbi.1004794)
56. Noel JK, Whitford PC, Sanbonmatsu KY, Onuchic JN. 2010 SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res.* **38**, W657–W661. (doi:10.1093/nar/gkq498)
57. Whitford PC, Sanbonmatsu KY, Onuchic JN. 2012 Biomolecular dynamics: order-disorder transitions and energy landscapes. *Rep. Prog. Phys.* **75**, 076601. (doi:10.1088/0034-4885/75/7/076601)
58. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, Poux S, Bougueleret L, Xenarios I. 2016 UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to use the entry view. In *Plant bioinformatics: methods and protocols* (ed. D Edwards), pp. 23–54. New York, NY: Humana Press. (doi:10.1007/978-1-4939-3167-5_2)
59. Finn RD *et al.* 2014 Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–D230. (doi:10.1093/nar/gkt1223)
60. Finn RD, Clements J, Eddy SR. 2011 HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37. (doi:10.1093/nar/gkr367)
61. Finn RD, Clements J, Eddy SR. 2011 HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37. (doi:10.1093/nar/gkr367)
62. Fraczkiewicz R, Braun W. 1998 Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem.* **19**, 319–333. (doi:10.1002/(SICI)1096-987X(199802)19:3<319::AID-JCC6>3.0.CO;2-W)
63. Lammert H, Schug A, Onuchic JN. 2009 Robustness and generalization of structure-based models for protein folding and function. *Proteins: Struct. Funct. Bioinf.* **77**, 881–891. (doi:10.1002/prot.22511)
64. Pronk S *et al.* 2013 GROMACS 45: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854. (doi:10.1093/bioinformatics/btt055)
65. Li Y, Zhang Y. 2009 REMO: a new protocol to refine full atomic protein models from C-alpha traces by optimizing hydrogen-bonding networks. *Proteins* **76**, 665–676. (doi:10.1002/prot.22380)
66. Biasini M *et al.* 2014 SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, 252–258. (doi:10.1093/nar/gku340)
67. Notti RQ, Bhattacharya S, Lilic M, Stebbins CE. 2015 A common assembly module in injectisome and flagellar type III secretion sorting platforms. *Nat. Commun.* **6**, 7125. (doi:10.1038/ncomms8125)
68. Berry R. 2004 The bacterial flagellar motor. In *Molecular motors* (ed. M), pp. 111–140. Weinheim, Germany: Wiley-VCH Verlag GmbH and Co. KGaA.
69. Park SY, Lowder B, Bilwes AM, Blair DF, Crane BR. 2006 Structure of FliM provides insight into assembly of the switch complex in the bacterial flagella motor. *Proc. Natl Acad. Sci. USA* **103**, 11 886–11 891. (doi:10.1073/pnas.0602811103)
70. Vartanian AS, Paz A, Fortgang EA, Abramson J, Dahlquist FW. 2012 Structure of flagellar motor proteins in complex allows for insights into motor structure and switching. *J. Biol. Chem.* **287**, 35 779–35 783. (doi:10.1074/jbc.C112.378380)
71. Zhao X, Norris SJ, Liu J. 2014 Molecular architecture of the bacterial flagellar motor in cells. *Biochemistry* **53**, 4323–4333. (doi:10.1021/bi500059y)
72. dos Santos RN, Khan S, Morcos F. 2018 Characterization of C-ring component assembly in flagellar motors from amino acid coevolution. Dryad Digital Repository. (doi:10.5061/dryad.0mv6t)