Doctoral Thesis

High resolution structural analysis of the flagellar hook of *Salmonella enterica serovar Typhimurium*

Horvath Peter
March 2018

Graduate School of Frontier Biosciences
Osaka University

Table of Contents

Acknowledgements

Special thanks go out to Namba-Sensei, because he invited me to Japan and gave me the rare privilege to work in an inspiring environment and motivated me to become a better scientist. Special thanks go out to Kato-Sensei, who patiently taught me everything from Japanese culture to the abstract notion of Fourier transforms and whatever may fall in between the two. Thank you very much to Miyata-Sensei, Minamino-Sensei, Namba laboratory and everyone who has contributed in some ways to this thesis.

Köszönöm a családomnak és a barátaimnak a támogatást.

Abstract

Equipments and techniques for cryo-EM methods for structural studies of macromolecules have rapidly advanced in recent years. These developments allow near-atomic resolution structure determination, which is inevitable for understanding underlying biological mechanisms. The bacterial flagellar hook functions as a universal joint. The highest resolution structure available so far is 7.1 Å, and this was achieved with a CCD camera. To gain more detailed information about the flexible mechanism of the hook, we tried to analyse the hook structure by electron cryomicroscopy using a direct electron detector camera. We succeeded in solving the structure at 4.1 Å resolution and constructed the whole atomic model of the hook. We will discuss the mechanism for flexible bending and mechanical stability of the hook.

General Introduction

Bacterial flagella

Bacteria motion is active. They can vary their paths with respect to chemical (Mesibov, 1972) and temperature gradients (Imae,1985). During "taxis" they can move from unfavourable environment to more favourable. *Salmonella enterica serovar Typhimurium* (hereafter abbreviated *Salmonella Typhimurium*) have several peripheral flagella.

The rotation of the flagellum is driven by the motor. The "fuel" of the motor is proton influx across inner cell membrane (Matsuura et al. 1977). The usual rate is 300 Hz, in *Vibrio Alginolyticus* where the driving force is Na+ the rotation rate is 1500 Hz (Magariyama et al., 1994). The average cell speed of 20-30 micrometer/sec, when the flagella is rotating in CCW direction (R. M. Macnab & Koshland, 1972). Bacteria can vary its rotational direction (from CCW to CW) and as a result the bundle falls apart and the cell will tumble. Bacterial locomotion is a kind of random walk. Addition of attractant or removal of repellent (positive stimuli) can influence the trajectory by increasing the straight trajectories. Addition of repellent and removal of attractant (negative stimuli) will exhibit the opposite effect. Motor rotation switching is governed by the receptors in the inner membrane due to signal transduction mechanism.

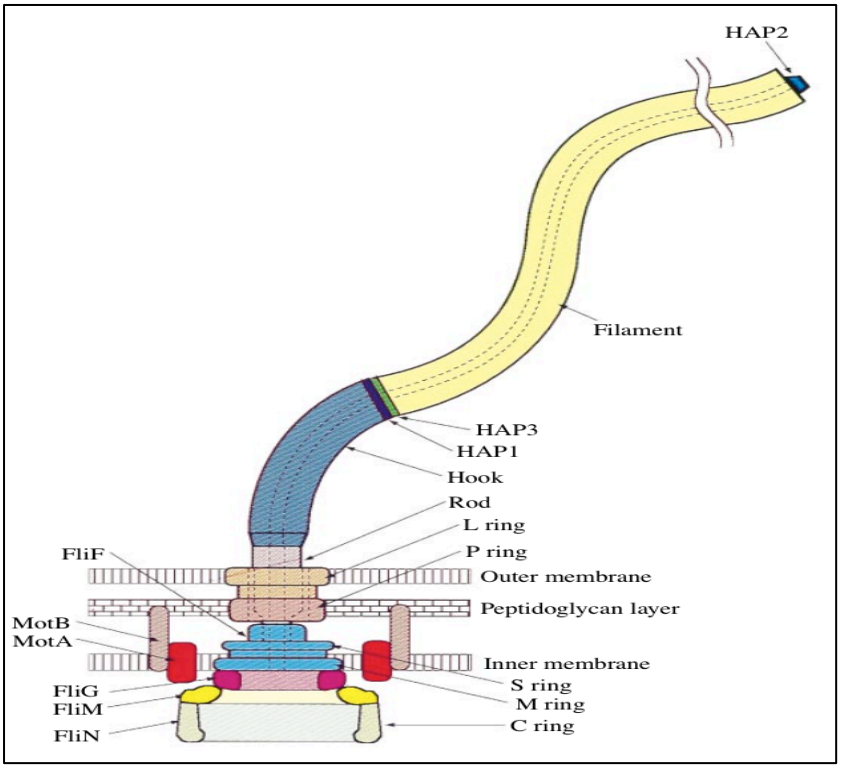


Figure 1 Schematic figure of Salmonella Typhimurium flagellum

Assemble takes place first by first forming FliF ring . Structure evolve in every direction of space from that point. The axial proteins (rod, hook, flagellar filament) assembles helically. The length of the hook and the rod are well regulated. Filament length: 15 micrometer. Axial proteins are exported by the export system. The export system consists of integral membrane proteins and cytoplasmic proteins (FlhA, FlhB, FliO, FliP, FliQ, FliR (export gate proteins) FliH, (ATPase regulator ATPase), FliI (ATPase), FLiJ (putative chaperone). The hook and the filament diameter is approximately 25-30A.

Hook protein FlgE

The bacterial flagellum is a supramolecular assembly (figure 1) a composition of 30 different kinds of proteins and can be divided into three parts from the distal cytoplasmic end: the filament, hook and the basal body respectively. The basal body is located in the cell membranes and functions as a rotary motor. Rotation can occur both in counterclockwise and clockwise direction (viewed from filament side). Rotation of CW direction results in a disrupted tumbling cell motion, whereas CCW rotation results in a bundle of filaments which can propel the cell smoothly. This is how Salmonella can move towards more favorable, change their current location. The rotation of the motor requires electrochemical energy which is supplied by the incoming flux of hydrogen and sodium ions across the cell membrane embedded stators. So called ion motive force.
The hook and the filament extend outwards. The filament functions as a propeller to produce thrust. The hook connects the filament with the basal body as a universal joint to smoothly transmit torque produced by the motor to the helical filament.
The self-assembly of such a supramolecular system begins in the cytoplasm extending through the cell all the way to the cell exterior. The external elements are needed to be exported. This type of export called Type III pathway is also responsible for the exportation of virulence factors.
Self–assembly starts with basal body, followed by the hook and finishes in the construction of the filament. A remarkable feature is that the flagellar type 3 protein export apparatus, that it coordinates flagellar gene

expression with assembly (Minamino, 2017). The flagellar type 3 export apparatus is also capable of monitoring the state of rod-hook assembly and switches its substrate export from rod substrate to hook substrate. These are the first steps of axial structure assembly. The elongation mechanism differs in the case of the two proteins. The rod component polymerizes from the inner membrane across the periplasmic space and stops at a length of 25 mm at the outer membrane. There is a significant difference between the rod elongation and the polymerization of the subsequent assembly events, namely the rod subunits are exported into the finite volume periplasmic space. So far no proof exist for any rod-length controlling mechanism (Robert M Macnab, 2003) The hook is built from approx. 120 subunits of FlgE, with ~55+/-6 nm length (Samatey et al., 2004). Opposed to the rod, in the case of the hook there exist a regulator protein called FliK which is responsible for both hook length measurement and signal transmission to the export apparatus. FliK or FlhB mutant strains the length of the of the hook is not regulated and in these so called "polyhook"(Silverman & Simon, 1972); (Patterson et al.,1973); (Komeda, Silverman, & Simon, 1978); (Suzuki & Tim), 1981) strains the length can be over 1000 micrometers. The details of this regulation mechanism is still not completely understood. Different mechanisms were hypothesized such as the moleculer ruler model (Kawagishu, Homma, Williams, & Macnab, 1996), cup model (Makishima et al., 2001).

Interestingly a FliK null strain showed same polyhook length as the wild type strain's hook (Makishima et al., 2001).

The helical packing of FLgE shows same arrangement in the mutant polyhook as in the wild type (T Wagenknecht & DeRosier, 1981). The mutant polyhook also exhibits the properties of the wild type hook e.g.: polymorphism, a phenomenon where under different environmental conditions several different helical forms can be observed (table 1). The observed forms can be classified into 4 different classes:

Type	Conditions	Pitch (nm)	Diameter (nm)	Handedness
Normal	pH 4-9, ≳12°C	95±9	35±3	Right
Coiled	pH 3-6, ≳10°C	Not Determined	57±4	Right, Left
Left-handed	pH 2-5, ≲15°C	60-100	5-35	Left
Straight	≲pH 2 or ~pH 7, ≲12°C	Not Determined	Not Determined	Not Determined

Table 1 Physical properties of the different polymorphic forms

(Satoru Kato, Okamoto, & Asakura, 1984)

Polymorphic transition of the flagellar polyhook of *Salmonella Typhimurium* can be explained by the two-state model.(Asakura, 1970); (Calladine, 1976); (Calladine, 1978); (Kamiya, Asakura, Wakabayashi, & Namba, 1979); (S. Kato, Aizawa, & Asakura, 1982) suggested that the protofilament lies along the 16-star helical line. Another physically more founded hypothesis suggests that D2-D2 interactions producing a well defined distribution of points on the twist-curvature diagram. Which can be used to define a protofilament direction, that can produce supercoils. Unfortunately there are not enough data to conclude a final, valid model. Recently

Sakai et al., 2018 (in press) showed that the axial interaction between the triangular loop of D1 domain and the D2 domain is responsible for the supercoiling mechanism in *Salmonella Typhimurium*'s hook protein.

Image analysis and electron microscopy studies have shown similarity between hook and flagellar filament, both posses tubular structure made of 11 protofilaments.

Full length FlgE, the monomeric subunit of the hook doesn't form crystals, a truncated monomer corresponding to 71-369 (out of 402) numbered residues was first crystallized and determined by X-ray crystallography (Samatey et al., 2004). Total mass is 31 kDa hence the name FlgE31. The X-ray structure of FlgE31 revealed two domains D1 (71-144) and D2 (145-284) (figure 2), which are connected by anti-parallel beta-sheet. Domain D2 is an eight strand beta barrel with additional extra loops. D1 domain consists of 3 beta-hairpins, beta-meander which consists of three antiparallel beta-sheets and a triangular shaped part which is called triangular loop (116-135).

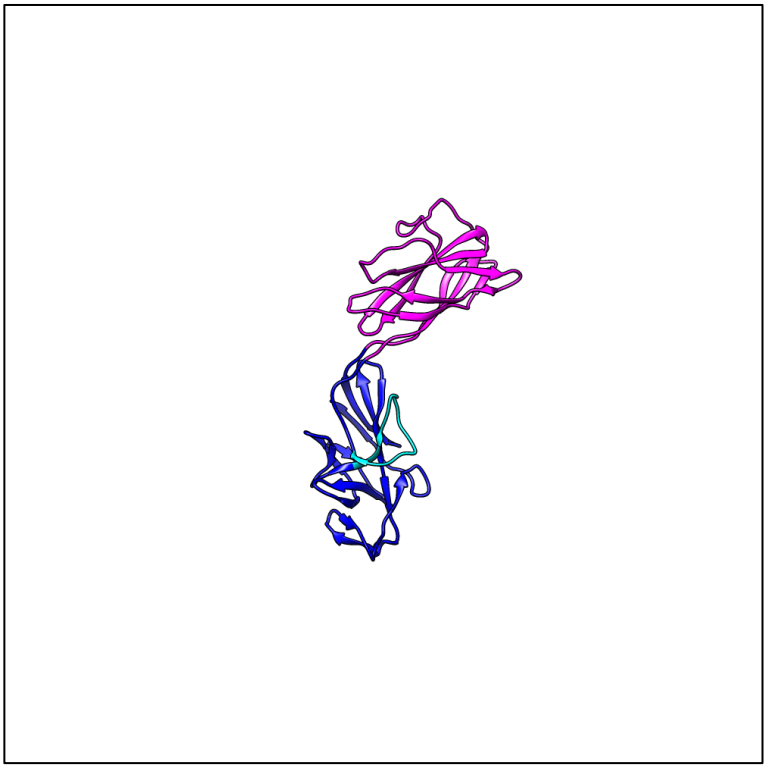


Figure 2 FlgE31 partial structure

D1 domain (71-144,285-363) colored in blue, triangular loop (116-135) part of the D1 domain colored in cyan, D2 domain (145-284) colored in magenta (Samatey et al., 2004)

In the helical lattice it is possible to define helical lattice lines along which neighboring subunits domain interaction can be analysed (figure 3),(figure 4). It has been well known from early structural studies (Terence Wagenknecht, DeRosier, Aizawa, & Macnab, 1982) of the flagellar hook that D2 domains exhibit strong interactions along the 6-start helix.

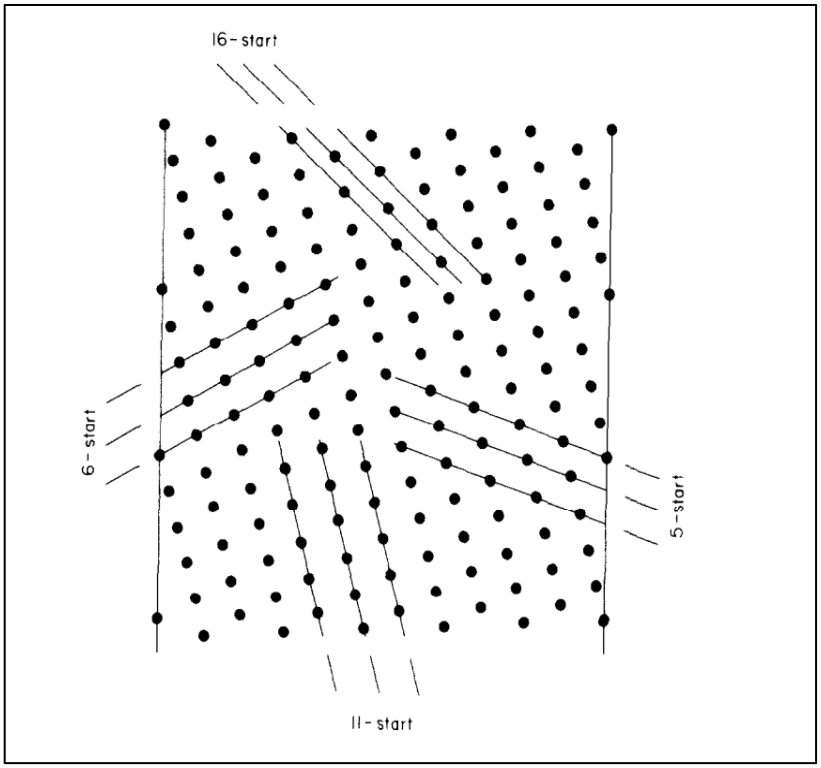


Figure 3 Radial projection of straight hook

Families of helical lattice lines are shown in the figure. (Satoru Kato et al., 1984)

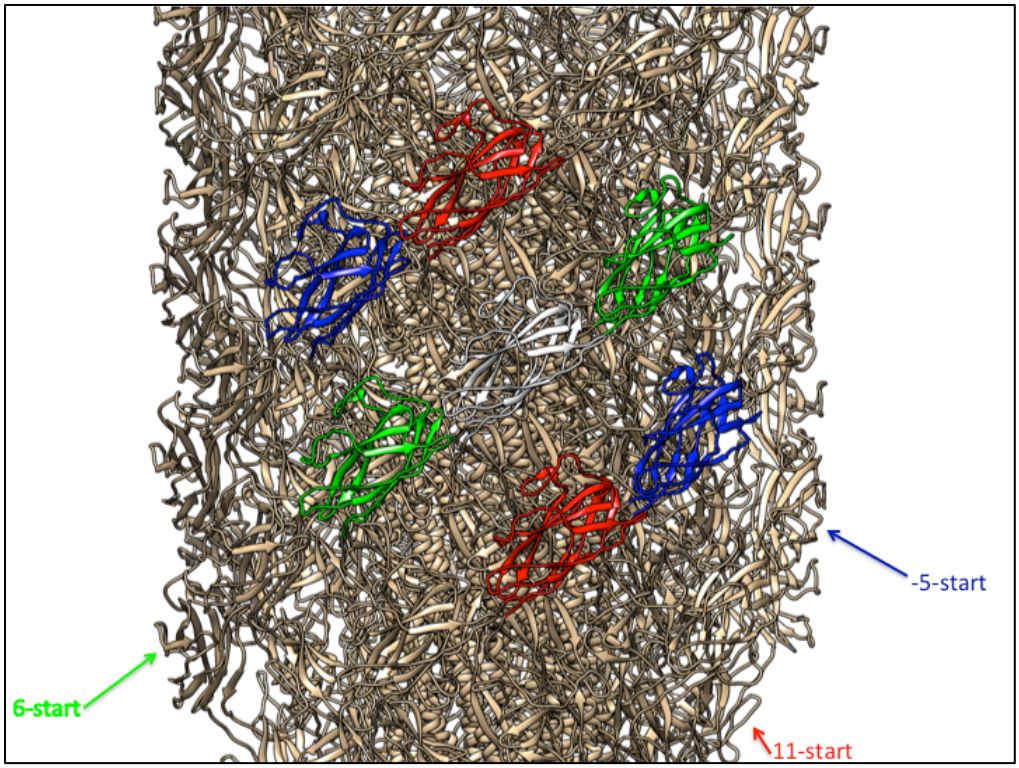


Figure 4 Intermolecular interactions along helical directions

Interactions of subunits along the helical lattice lines can be analysed. Arrows delineate different numbered helical families. 11-start interactions define the protofilaments in the hook. The hook is composed of 11 protofilaments, similarly to the filament.

All interactions formed by the D1 and D2 domains were proved to be either polar-polar or polar-charge interactions, which can easily explain why the truncated FlgE31 couldn't polymerize into a stable hook. This phenomenon also indicates that the close interactions of innermost D0 domain is responsible for the assemble and mechanical stability in a similar way to the D0 domain of the filament. (Samatey et al 2004).

Cryo EM resolution revolution

Electron microscope investigation of biological specimens started with Ruska's observations on bacteriophage around 1941 (Ruska et al., 1941).

Around that time biological samples were dehydrated or fixed which resulted in an introduction of undesired artifacts to the sample. One of these techniques called negative staining where during the sample preparation the sample is introduced to heavy metal salt solution (such as 1% uranium acetate)(figure 8). Developments towards the observation of samples that show more resemblance to their native structure were done by pioneers like Henderson and Unwin (R. Henderson, P.N. Unwin, 1975). The author obtained a 7 Å resolution 3D model of purple membrane. In 1981 Dubochet (Dubochet, Booy, Freeman, Jones, & Walter, 1981) presented a method where they embedded in amorphous ice, restoring the native structure.

Because this state was maintained at liquid nitrogen temperature (-196^{0}C), a new term was introduced, called cryo-EM.

By the year 1997 with the aid of single particle reconstruction alpha helices of hepatitis B virus could be visualized (Böttcher, Wynne, & Crowther, 1997).

In 2008 the backbone of the viruses could be traced. (Yu et al., 2008)

By 2010 de novo atomic model building was achieved in the case of viruses (X. Zhang, et al, 2010). This result was facilitated by the inherent symmetry of the target protein. Development of high resolution structural analysis of protein molecules without internal symmetry remains challenging.

Since 2013 the deposited number of high resolution structures increased in PDB as well as in other cryo-EM databases. The reason is twofold :

1, recent developments in the field of cryo-EM hardware. The direct electron detector camera (figure 5).

2, improvement of image analysis methodologies.

Let us consider the reasons in more detail.

1, The DQE (detective quantum efficiency) is a measure of the image quality defined as the number of quanta is used to produce an image with a certain SNR ratio divided by the total number of quanta that was used for the particular image. The ratio shows a Poisson-distribution. (Dainty, J.C. and Shaw, R., 1974). An ideal case is when this ratio equals to one. In reality the applied detector devices show the following approximated quotients: DQE[photographic films]~0.3
DQE[CCD (charged coupled device)] camera~0.1
DQE[DD(direct electron detector)] camera~0.5
(Patterson-delafleld, Martinez, Angeles, & Yamaguchi, 1973)
Recently state of art cameras apply MAPS (monolithic active pixel sensors)(Milazzo et al., 2005)which allow individual, direct electron detection on a thin semiconductor device.

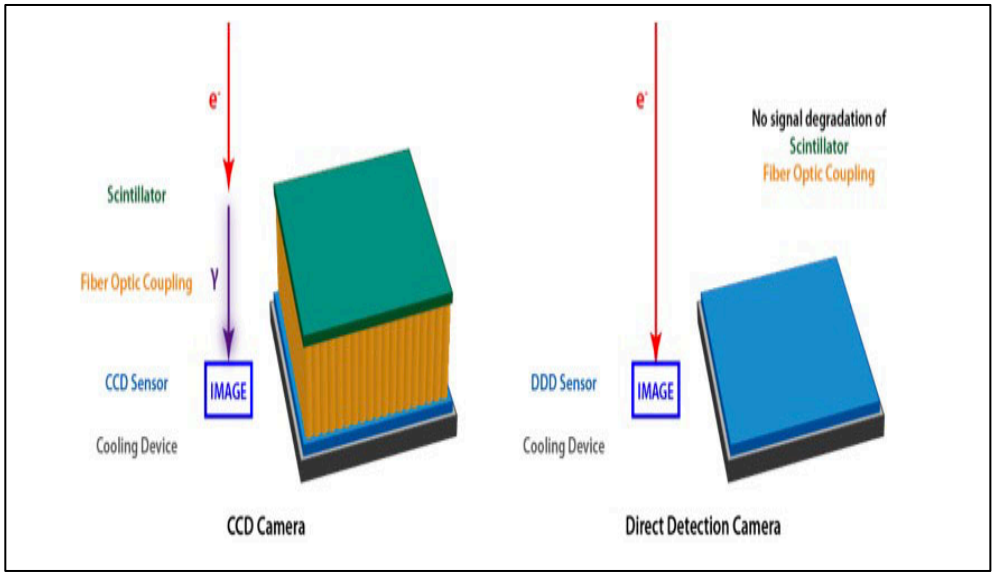
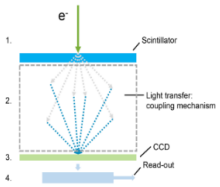


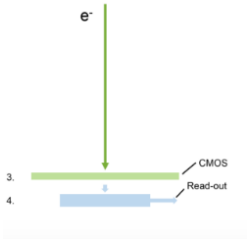
Figure 5 Different electron detection approaches

Traditional CCD cameras utilizes scintillator to convert primary electrons (red arrow, e^-) to photons (purple arrow, $\boldsymbol{\gamma}$) before being detected CCD sensor. Direct electron detectors (DED) doesn't transform electrons to photons, hence it's signal to noise ratio and sensitivity is significantly increased. Further advantages of DED: minimized image distortion, high frame rate (dose fractionation, motion correction)(directelectron.com)

CCD imaging steps
1, Convert electrons to signal
2, Transfer signal
3, Detect signal with sensor
4, Electronically transfer signal and read out to form image



DED imaging steps
1~~, Convert electrons to signal~~
2, ~~Transfer signal~~
3, Detect signal with sensor
4, Electronically transfer signal and read out to form image



The detector efficiency (DQE) can be increased by decreasing the ratio of the backscattered electrons which in turn can be decreased by removing as much substrate as possible in a process called "back-thinning" (McMullan, Chen, Henderson, & Faruqi, 2009). The state of art cameras are back thinned to 50 micrometer, this allows the electron beam to interact with the detector avoiding backscattering

events. In Namba laboratory we use a device produced by Gatan company. Direct electron detectors have the highest DQEs among detecting devices in electron microscopes. The images are formed by integrating individual movie frames.

2, improvement of image analysis methodologies:
Most macromolecular complexes exhibit conformational or compositional heterogeneity. If the heterogeneous subset can not be divided into smaller homogeneous subsets than the superposition of the inhomogeneous classes will result in a blurry, low resolution image. To solve this problem supervised classification was introduced (Gao, Valle, Ehrenberg, & Frank, 2004). The method assumes prior knowledge of structural variability in the sample.
A more general approach was suggested by (Sigworth, 1998) introducing the maximum-likelihood approach to single particle image refinement. This method allows the underlying structure to be estimated from large sets of noisy images. The algorithm forms a weighted sum over all possible in-plane rotations and translations of the image. The weighting factors calculated are the probabilities of the transformations. The result shows a reduced sensitivity to the starting reference. Other methods:
Multivariate statistical analysis (Elad, Clare, Saibil, & Orlova, 2008):
This method allows the user to find the source of variations introduced to the images which are plane orientation and local structural changes. The automated classification allows discrimination of 2D images.
Bayesian statistics:

Handling noisy data evaluation, with prevention of overfitting close to the resolution limit, where the SNR (signal-to-noise ratio) is the lowest. The statistical framework reduce the amount of subjective and heuristic decisions in the image analysis procedure, seeks the maximized single probability function, that is to find the model that has the highest possibility of being the correct one as a function of the observed data. In regularization step model parameters are imposed with prior distribution function. Optimization of the resulting posterior distribution is called regularized likelihood optimize or maximum a posteriori (MAP) function. (Scheres, 2012b) Another important contribution to the resolution revolution is an image processing method called beam induced motion correction (Li et al., 2013). This method is closely related to the developments of the new generation of cameras (figure 6). When electron encounters the specimen during exposure forces act upon sample resulting in motion, which results blurring in the reconstructed image, which leads to a resolution increase. New detectors allow the user to record movies which are composed of individual frames, that can be aligned, thereby improving the signal-to-noise ratio, what resulted in a set of high resolution structures (table 2).

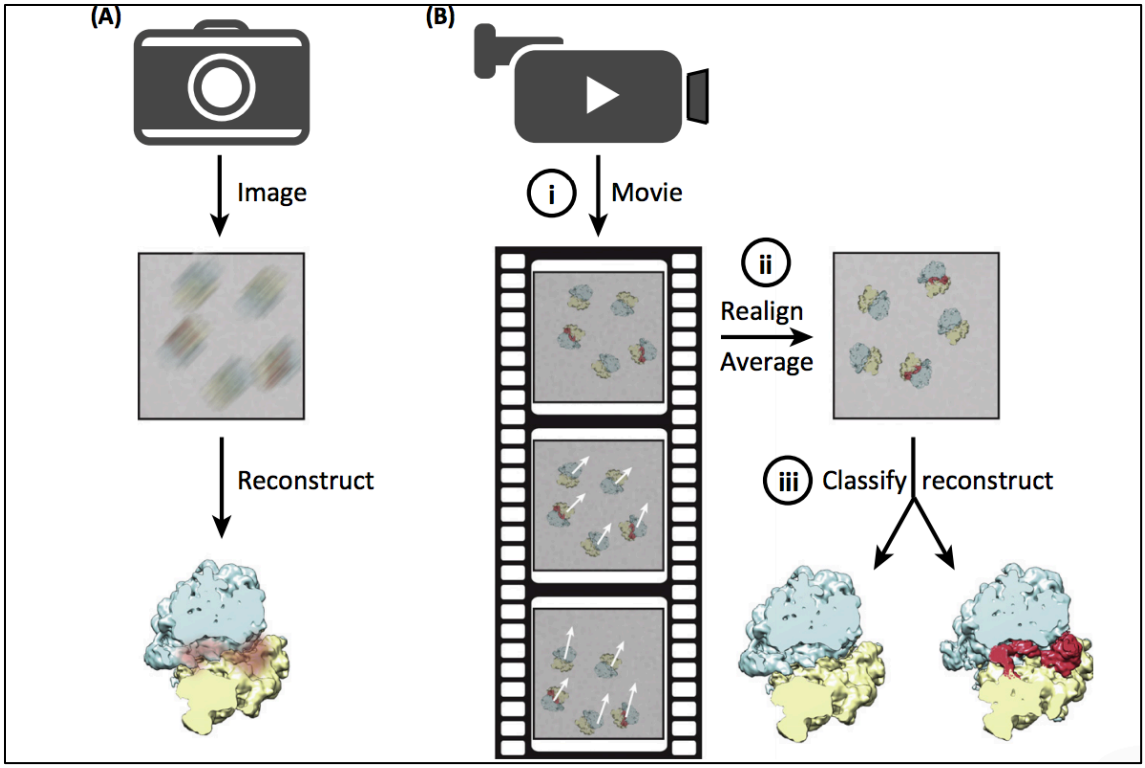


Figure 6 Past, recent method comparison

A, Noisy images recorded on photographic films. Structurally heterogeneous images were averaged together, led to low resolution 3D images. B, Recent improvements such as 1, movie recording 2, movie frames correction 3, classification methods can produce higher resolution images. (Bai, McMullan, & Scheres, 2015)

Sample	MW (Mda)	Point group	3D classification	Beam induced motion correction	Software	Resolution (Å)	Release date
80S ribosome + emetine	4.2	C1	Yes	Yes	RELION	3.2	2014 Jun-14
Yeast mitoribosome	1.9	C1	Yes	Yes	RELION	3.2	2014 Apr-14
20S proteasome	0.7	D7	No	Yes	RELION	3.3	2013 May-13
Dengue virus	50	I	No	No	EMAN,MPSA	4.1	2013 Nov-13
Rotavirus	35	I	No	Yes	FREALIGN	4.4	2012 Sep-12
80S ribosome	4.2	C1	Yes	Yes	RELION	4.5	2013 Jan-13
γ-secretase	0.17	C1	Yes	Yes	RELION	4.5	2014 Jun-14

Table 2 Cryo-EM single particle reconstructions

Image data collection with Falcon direct electron detector (Bai et al., 2015).

Previously low-resolution electron density maps relied on other high resolution maps as a complementary method to interpret electron density maps on atomic level. Such complementary methods are X-ray crystallography or NMR. Recent developments in the field makes it possible to achieve resolutions beyond 4 Ångstrom and de novo atomic model building has become possible.

Cryo-EM electron density map is inherently related to X-ray map the former preserves extra information about the phases, that are lost in the latter. This resemblance allow evaluation of the electron density maps with softwares that are used in the field of X-ray crystallography. Such as COOT (Emsley, Lohkamp, Scott, & Cowtan, 2010). Another example can be Rosetta (Frank DiMaio, Zhang, Chiu, & Baker, 2013)

Focus of this study

In this study I focused on the Cryo-EM structural analysis of the hook protein (FlgE) of the length ruler protein FliK deficient mutant of *Salmonella Typhimurium*. The mutant's hook is structurally identical to the native *Salmonella Typhimurium*'s hook. It consists of four domains the inner Do, Dc, D1 and the outer D2 domain.
Fadel Samatey (Samatey et al., 2004) determined the partial atomic structure of FlgE called FlgE31 (31kDa referring to the mass of the proteolytic fragment).
This structure was missing terminal regions (D0 domain) and the Dc domain (25-70, 364-366) and a short sequence of amino acids out of the total 402 amino acids that constitute the whole FlgE.
In 2009 (Fujii et.al, 2009) determined the partial polyhook Cryo-EM electron density map at 7.1 Ångstrom resolution. At this resolution the DO domain can be seen as a featureless cylinders. Because of the low resolution of the electron density map no atomic model could be built in Dc domain (25-77,364-366).
The goal of the study was to further improve the resolution of the hook, utilizing the new improved hardware and software tools that allow the near atomic resolution study of the biological macromolecules proteins that are difficult or impossible to analyze with other methods. I aimed to solve the full atomic structure at nearly atomic resolution. The hook protein of *Salmonella Typhimurium* is an interesting and important protein target. Unique characteristics of the hook protein such as its capability to change forms (polymorphism), length controlled elongation mechanism or the flexibility make it an interesting target. In this work I would like

to elucidate the flexibility and the mechanical stability of the hook based on the newly revealed Dc domain and the refined full atomic model of FlgE. These are key notions to understand the universal joint function of *Salmonella Typhimurium*. This function is critical for bacterial motion and torque transmission from the motor (directly from the rod) to the filament. I aim to deduce the molecular requisites of flexibility and mechanical stability.

Materials and Methods

Sample preparation

The Salmonella hook in its native form is not suitable for high resolution cryo-EM structural analysis. Instead a FliK mutant (Patterson-delafleld et al., 1973) was used for isolation of polyhook-basal body complexes (figure 7). This mutant can grow up to 1 micrometer long and structurally identical to the native hook's structure. Rendering the hook as a suitable target for cryo-EM structural analysis.

Frozen cells were grown overnight in 30mL LB broth (Bertani, 1951) (1% tryptone, 0.5% yeast extract, 1% NaCl) shaking at 37°C in an

incubator. Cells were transferred into two 2X1.3L LB broth. 2 drops of antifoam 204 non-silicone organic defoamer is added to the 2X1.3L LB broth. Cells were grown at 37°C with shaking in a Bioshaker until OD_{600} reached between 1~1.3. Cells were centrifugated at 7000 rpm, 4°C for ten minutes at 10°C. The pellet was resuspended in 80 mL sucrose solution (0.5M sucrose, 0.1M TRIS/HCL pH=8). While stirring the suspension on ice 8mL EDTA (pH=8) and 8mL 10 mg/mL lysozyme were added slowly to the resuspended sucrose solution. The solution was stirred in cold room at 4°C for 30 minutes. In this process the cells were turned into spheroplasts. Eight mL $MgSO_4$ 8mL and 8mL 10% TRITON X-10 was added to the lysate and was stirred until the viscosity of the solution had decreased (about 1 hour), indicating that the cellular DNA had been degraded by endogenous DNases (deoxyribonucleases). Added 8mL 0.1M EDTA (pH=11) to the solution. Unlysed cells and cell debris were removed by low speed, 20 minutes centrifugation step (14,000 rpm) at 4^0C. The pH of the supernatant is raised to pH=11 with 5N NaOH. At this pH outer membrane structures were dissolved. Another step of low speed centrifugation step followed to make sure that the solution doesn't contain cell debris. The lysate was subject to a following 1 hour ultracentrifugation (24.000 rpm) step. The pellet was resuspended in about 2 mL of Buffer C (10mM Tris/HCL, 5 mM EDTA, 1% Triton X-100). Insoluble fraction was removed by a 10 minute low-speed centrifugation (9000 rpm) step. The supernatant was then loaded onto a 20-50% (w/w) sucrose density gradient in buffer C and centrifuged at 4°C, for 13 hour. Fractionation was done by gradient station by Biocomp. Five times diluted fractions were analyzed by conventional SDS-PAGE method. Collected sample was diluted by S-

buffer (1M Tris, 1M NaCl, 10% TX100, H_2O) and enriched by ultrancentrifugation (24000rpm, 4°C, 60 minutes). The pellet was resuspended in approximately 40 microliter S-buffer. Insoluble fragments were removed with low-speed centrifugation (9000 rpm, 10 minutes, 4°C). Sample was kept on ice in cold room at least for one day before electron microscope observation.

Negative staining

During this process the sample is embedded in heavy metal solution (typically uranyl-acetate or phosphotungstic acid). The coated sample gains increased contrast. Heavy metal ions interact with the electron beam, thereby causing phase shift. The phase shifted beam interferes with the rest of the beam, creating phase contrast. Negative stained images are amenable to image processing. 3D reconstructions of negative stained specimens usually show artifacts (figure 8) such as flattening.
3.5 microliter 10X times diluted purified polyhook-basal body sample was applied onto a carbon coated copper grid. Sample preparation was made in the cold room at 4˚C in order to keep the polyhook-basal body strain straight. Sample was stained with 2% phosphotungstic acid. Electron micrographs were recorded at 7500X magnification. Sample parameters such as concentration and morphology were checked if they were met the necessary criteria of cryo-EM analysis or not.
Sample was observed with a JEOL transmission electron microscope operated at 100kV.

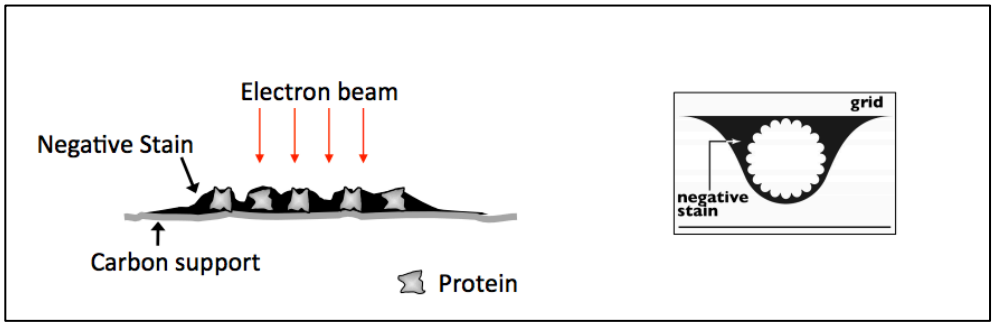


Figure 7 Schematic figure of negative staining

The process allows rapid screening for sample concentration and sample quality. Protein samples that do not meet the requirements can not enter to next level of screening. Advantages: quick screening method (size, structural homogeneity, quick screening method (size, structural homogeneity), radiation damage is irrelevant, sample preparation is easy, 3D reconstruction is feasible, high contrast Disadvantages: limited resolutions (~ **25**Å), protein distortion (flattening), damages and artifacts, uneven staining, artefacts, molecular instability, Non native conditions

Cryo-EM data collection

Optimized sample preparation conditions were applied to cryo-EM sample preparation. An aliquot of 2.4 microliter sample was applied onto a quantifoil holey carbon molybdenum grid and was plunge frozen into liquid ethane using a software driven device called Vitrobot from FEI. After numerous trial and error experiments, the following optimized blotting parameters were used: 100% humidity, 4˚C, 2 times blotting, each blot lasted 4 seconds. With these parameter settings hooks were not completely straight. In order to straighten them out I put 50 mL of liquid nitrogen in a small beaker to further decrease the temperature in the chamber of the Vitrobot. Images were taken with JEOL JEM 3200FSC field emission cryo-electron microscope equipped with omega-type energy filter, accelerating voltage 300 kV, ZrO layer/W tip Schottky type of

thermionic field emission gun and liquid helium cooled cryogenic stage. Zero-loss energy with slit width set to 10 eV to cut off higher energy dispersed electrons. Images were recorded with state of art direct electron detector K2 summit by Gatan in super resolution mode in which the pixel size 7676X7420. Nominal magnification was 50.000X. Dose fractionation mode was used in order to split the total dose into a number of frames. Total exposure time was 6 s. Frame exposure was 0.2 s. Total 30 sub-frames were recorded.

Image processing

MotionCor2

The GPU accelerated program was used to correct the beam induced movement of the 30 sub-frames. The program is an implementation of an algorithm that measures and corrects the frame motions at single pixel level across the whole size frame. The algorithm is applicable in a wide range of defocus. The software results in improved Thon-rings and improved 3D reconstruction resolution. Another option of the software is the weighting of individual sub-frames in Fourier space that enables the user to collect high dose images resulting in a high SNR (signal to noise ratio) (Zheng, Palovcak, Armache, Cheng, & Agard, 2016). Recorded gain reference was and dose weighting were applied. The first two frames were discarded.

Gctf

Precise CTF(contrast transfer function) correction is inevitable forhigh resolution structural analysis. The algorithm of the GCTF software maximizes the cross-correlation of a simulated CTF with the logarithmic amplitude spectra (LAS) of observed micrographs after background subtraction. GCTF software comes with astigmatism based rotational averaging or equiphase averaging (EPA), that results in improved visualization of the Thon rings for better analysis. Local refinement and sub-frame processing was also implemented in GCTF. The local refinement on individual particle basis makes significant improvement in the 3D reconstruction step of image processing.

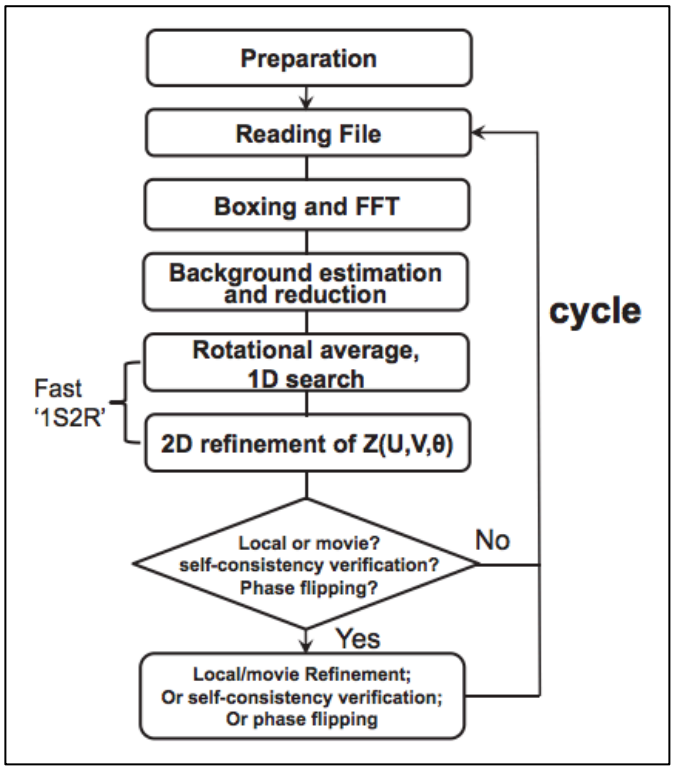


Figure 8 GCTF flowchart

(Zhang, 2016)

RELION

RELION (Regularized Likelihood Optimization) is an open source image analysis program for single particle analysis of cryo-electron micrographs. RELION deduces adequate parameters from the statistical model using Bayesian approach (figure 9). In this framework the 3D reconstruction is an optimization problem of a function. (Scheres, 2012b). The result is a model which has the highest probability of being the correct one. This method is called maximum a posteriori (MAP), or regularized likelihood optimization. Bayesian method iteratively learns most of the necessary parameters of the statistical model from the data, whereas in the case of other algorithms specific decisions are made by an expert user. A few ad hoc parameters are: initial model, spherical mask. RELION also implemented the so called gold-standard fourier shell correlations (FSC), which solves the problem of overfitting, providing reliable resolution values. This guarantees an objective evaluation and processing of the data. RELION implements a local optimization algorithm, that produces a result that is dependent on the starting model. Ab initio starting models generating algorithms are still under development.

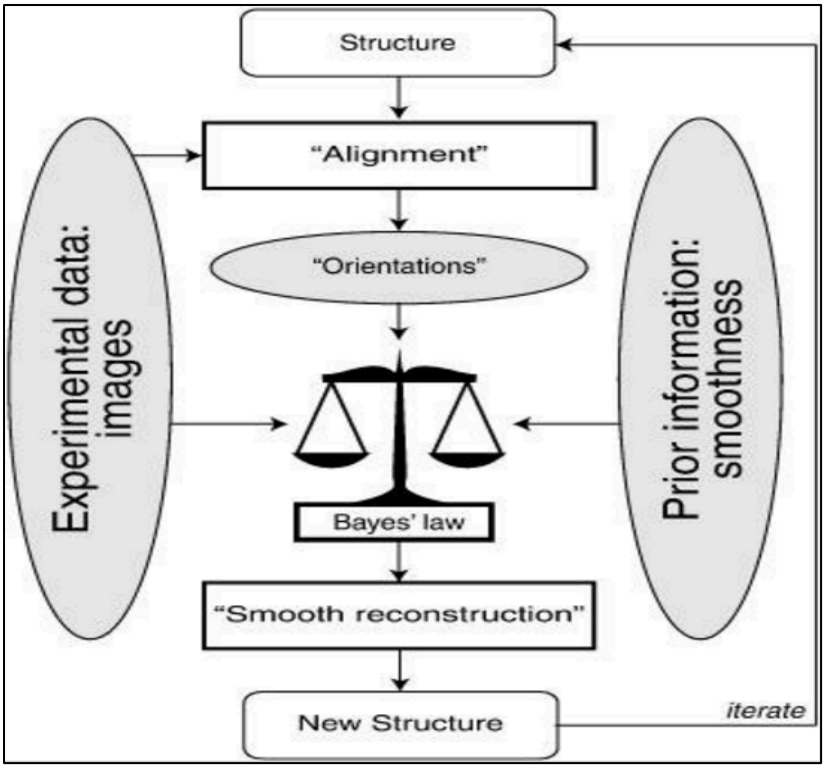


Figure 9 Schematic representation of image reconstruction methodology

First step is alignment (expectation). In this step references are aligned with experimental images, providing the relative orientations. Orientations are assigned over a probability function, taking into consideration every possible orientation. The concentration of predictive distributions depends on the power of the noise in the data. Second step is the smooth reconstruction (maximization), which is basically the 3D reconstruction step, update estimations of the noise and the signal in the dataset. The contributions of the data and the prior are determined by Bayes' Law which depends SNR ratio in the dataset. The new structure and SNR are used for the subsequent iteration step. Iteration end is decided by the user or after a number of cycles when there are no further improvements in the structure.
(Scheres, 2012a)

Atomic model

MODELLER

Atomic model was created by homology or comparative modeling with MODELLER. The goal is to obtain a three-dimensional model for the target based on single or multiple known three dimensional

structures (templates). The process consists of four main parts: fold assignment, target-template alignment, model building and model evaluation. The method assumes that small changes in the three dimensional structure. Furthermore the 3D structure within a family proteins are more conserved than the their sequences, so if two proteins are homologous on a sequence level than it is plausible to suppose 3D structural similarity. Nevertheless proteins without significant sequence similarity can have 3D structural similarity. It was estimated that at least one third of all sequences can be related to at least one known protein structure known protein structure.

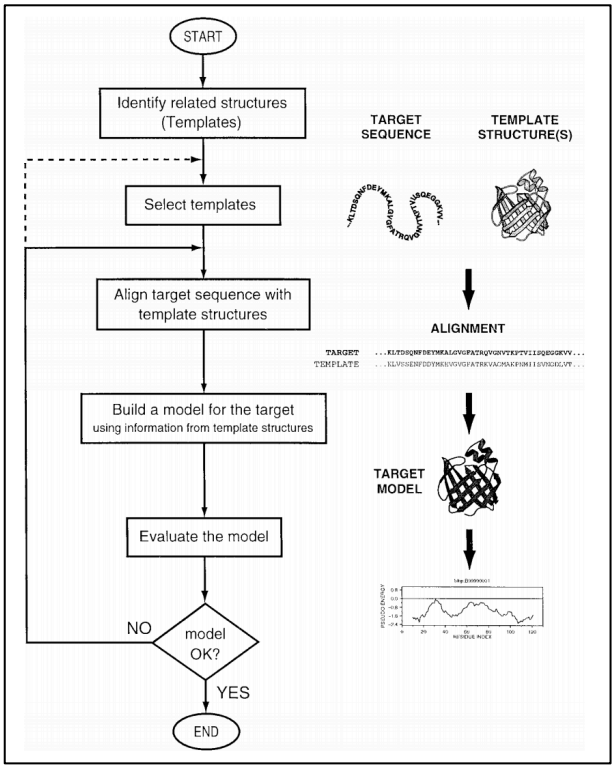


Figure 10 Homology modelling flowchart

(Webb & Sali, 2014)

Rosetta

Atomic model fitting into cryo-EM density map was carried out by ROSETTA3. The aim of this step is to refine protein model using the cryo-EM maps preferably with high resolution. The protocol is shown below (figure 27).

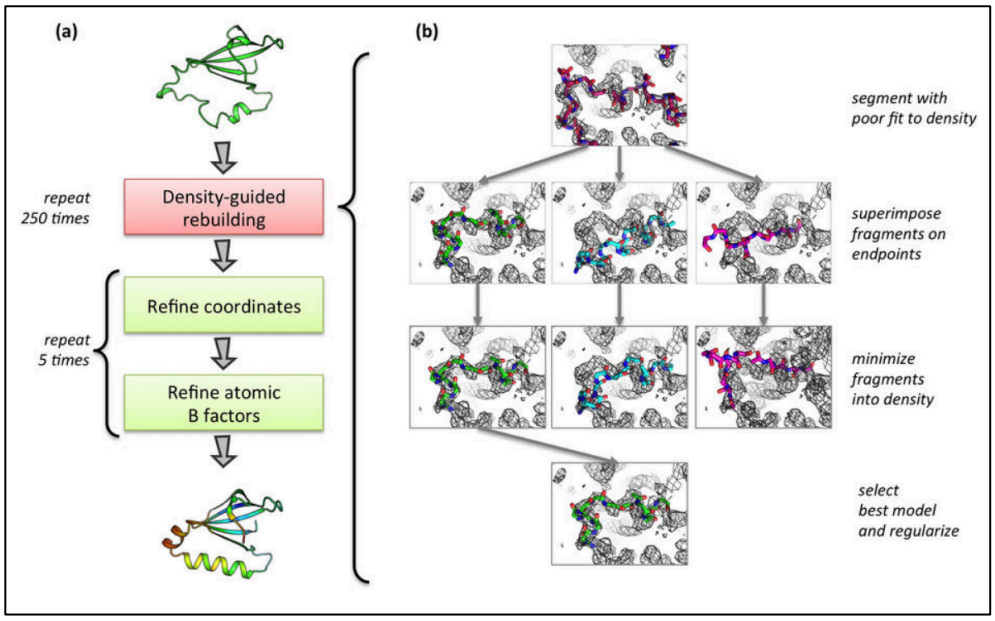


Figure 11 Flowchart of rebuilding and refinement

(F DiMaio et al., 2015)

250 cycles of local backbone rebuilding until the backbone can satisfy certain geometrical restraints, followed by a coordinate refinement step. During this step both length as well as bond angles were allowed to vary but harmonic potentials were added as constraints. Cartesian space refinement was implemented. The refinement step consists of: 1, three cycles of side-chain optimization and real-space as well as reciprocal space torsion angle minimization. 2, Five cycles of side-chain and real-space as well as reciprocal space torsion angle

minimization. Two cycles of side-chain optimization, Cartesian-reciprocal-space minimization. Last step is B factor refinement . Finally the best fitting backbone can be selected.
The above protocol is fully automated opposed to other refinement softwares. Backbone misplacements are automatically recognized and corrected by the program. The achievable resolution with cryo-EM is within the range of 3-4.5 Ångstrom. In this resolution range individual secondary structures and bulky side chains features are discernable. At this so called near atomic resolution range the refinement (as well as model building) can be error prone due to the difficulty of identifying side chain rotamers, turns or loops, regions where electron density is poor. To overcome such difficulties are desirable in order to be able to consider atomic details. ROSETTA generates a realistic geometry, does voxel size refinement for errors occurring in the magnification calibration coefficient of the electron microscope as well assignment of the backbone. Local strains can serve as an indicator where refinement should take place. Furthermore the missing density has a great impact on the convergence of conformational sampling during the refinement for the sake of the correct representation the side chains are weighted by factors. Typically the one's that can not be assigned with high confidence will be down-weighted. The voxel size is given by the physical pixel size multiplied by the scale factor corresponding to the magnification. In ROSETTA the voxel size is optimized via an iterative procedure until the voxel size is finally converged to a stable value. The advantages off Rosetta compared to other programs are automatic backbone refinement, physically realistic force fields.

Results

3D structure from negative stained images

During the sample preparation short polyhook-basal bodies that are in the size range of not being suitable for image analysis had to be separated from the longer polyhook-basal bodies that were suitable for image analysis. The separation was carried out by 20-50 w/w% linear sucrose gradient centrifugation step. Two fractions were obtained after fractionation. Short polyhook-basal bodies were unsuitable for image analysis (figure 12).

Image were taken with: Jeol 1011, TEM mode: brightfield, Image type: 8 bit, Magnification 7500X, Accelerating voltage: 100 kV, 2% PTA negative stain.

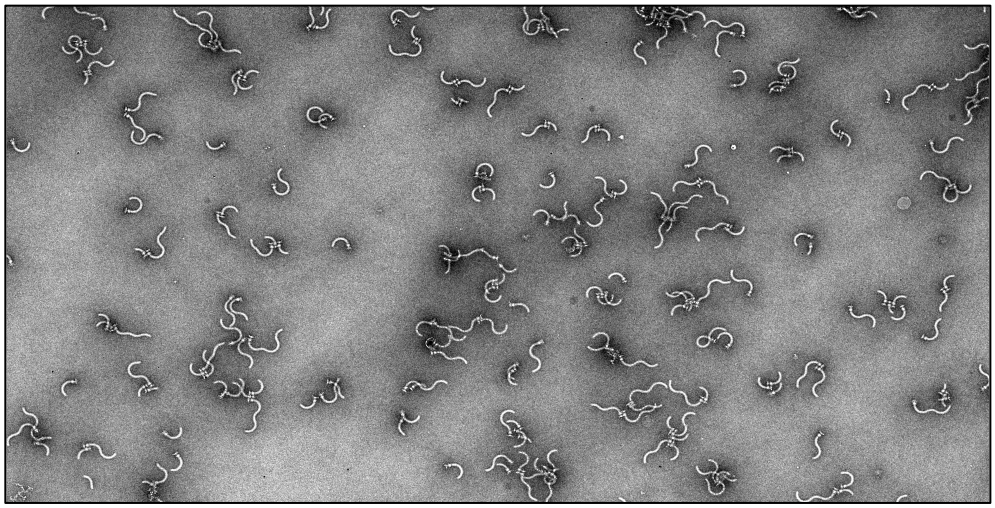


Figure 11 Electron micrograph of short polyhook basal-bodies

Long polyhook-basal bodies were suitable for image analysis (figure 13), although at room temperature they maintain a highly curved

structure. Such a highly curved structure is not suitable for image processing, therefore additional step is needed to straight them out. Image taking conditions were the following: Jeol 1011, TEM mode: brightfield, Image type: 8 bit, Magnification 7500X, Accelerating voltage: 100 kV, 2% PTA negative stain.

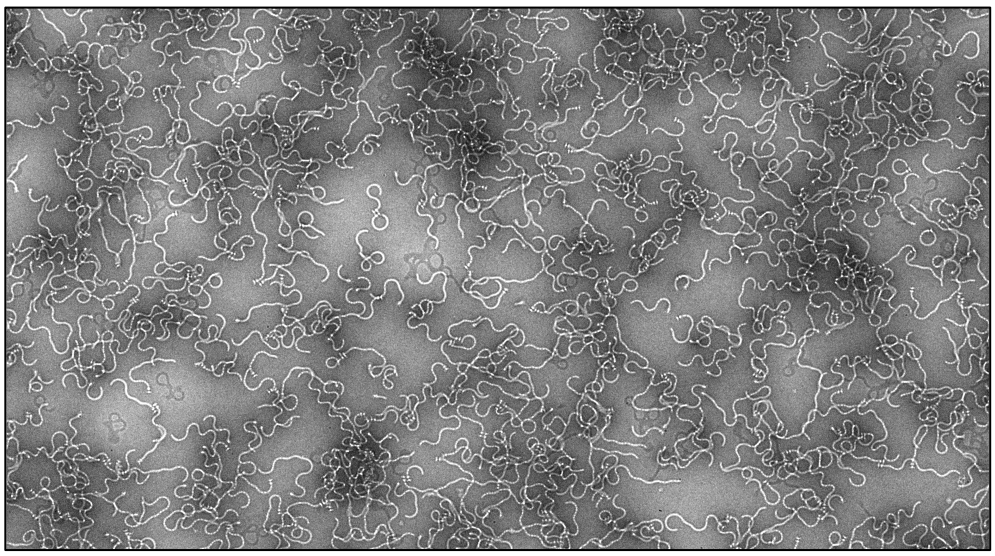


Figure 12 Electron micrograph of long polyhook-basal bodies

In order to make polyhook-basal bodies suitable for image processing, they were need to be straightened out. This was achieved by keeping the sample on ice overnight in the cold room after sample preparation. Furthermore grid preparation was done in the cold room (t≈ **4**°C) with 2% uranyl acetate (pH=2) staining solution, which made polyhook-basal bodies straight, suitable for image analysis (figure 14). Image taking conditions were the following: Jeol 1011, TEM mode: brightfield, Image type: 8 bit, Magnification: 50.000X. Accelerating voltage: 100 kV, detector 4KX4K CCD camera.

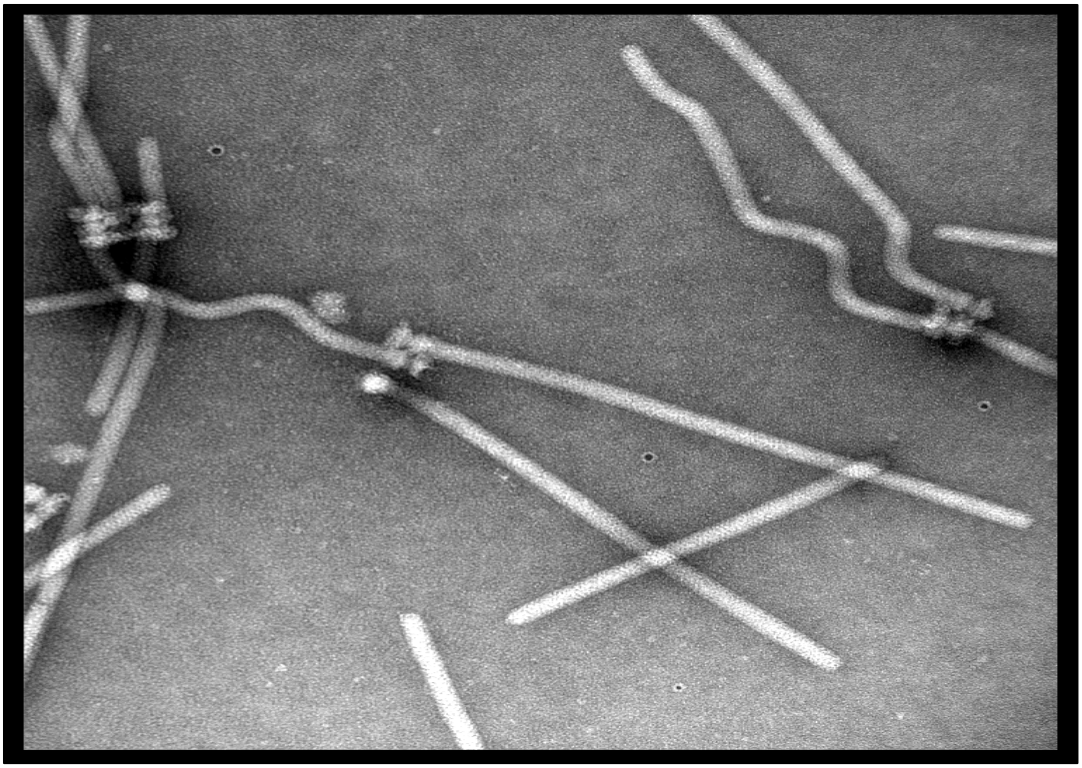


Figure 13 Electron micrograph of straightened, long polyhook-basal bodies

After successfully straightening out the polyhook-basal bodies. From each experimental micrograph, target polyhook particles were boxed out with EMAN2 helixboxer software (Tang et al., 2007), then cropped. The extracted polyhook segments were 2D classified into 10 classes by SPIDER software (Shaikh et al., 2008). Flattening artefact can be observed in the case of class averages (figure 15). Optimal shifts and rotations were determined to align the experimental images with projections of reference images. For each particle in-plane Euler rotation angle, and 2D translations were determined, thereby alignments with the reference images are possible, hence classification can be realized.

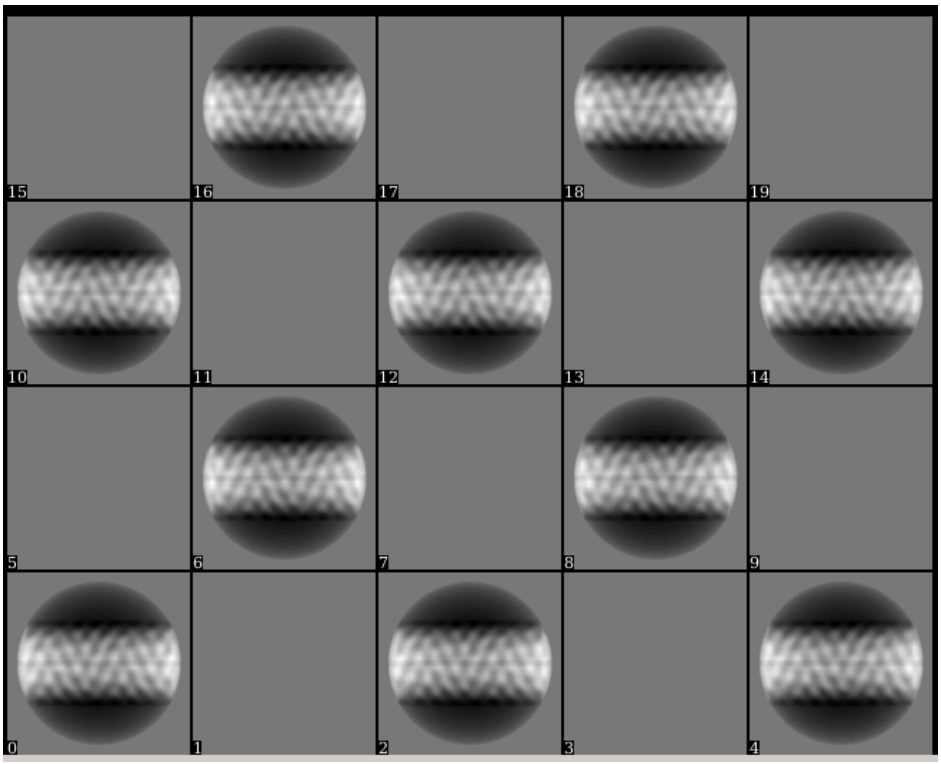


Figure 14 2D Classification of negative stained particles

In the following step volume reconstruction was carried from its projections. This step is called the back projection or 3D reconstruction by back-projection using direct Fourier method. The final 3D object was used as reference in different image analysis softwares (figure 16). The algorithm in SPIDER software constitutes of five major steps: 1, 2D FFT (fast fourier transform) of the input 2D projection images. Reverse gridding was used to resample 2D Fourier input images into 2D polar Fourier coordinates. 3, Gridding weights are calculated to make up for non–uniform distribution of the grid points. 4, Gridding using a convolution kernel with subsequent 3D inverse FFT yields samples on a 3D Cartesian grid. 5, Removal of weights in real space yields the reconstructed 3D object. (Penczek, 2010)

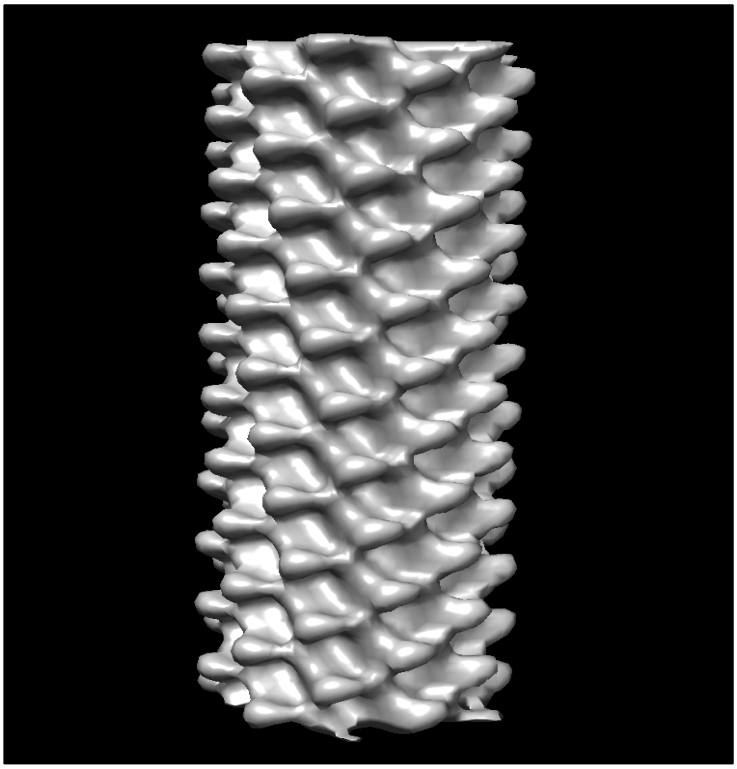


Figure 15 3D reconstruction of negative stained particles

After sample preparation and image collection conditions were optimized. The next level was structural analysis with cryo-electron microscope. 30 movie frames were recorded K2 direct electron detector. Movie frames were aligned with motioncor2 (Zheng et al., 2016) software (figure 17) and CTF corrected with Gctf software. Images were collected at 300kV. Pixel size: 0.41 A/pix. Totally 486 micrographs were taken.

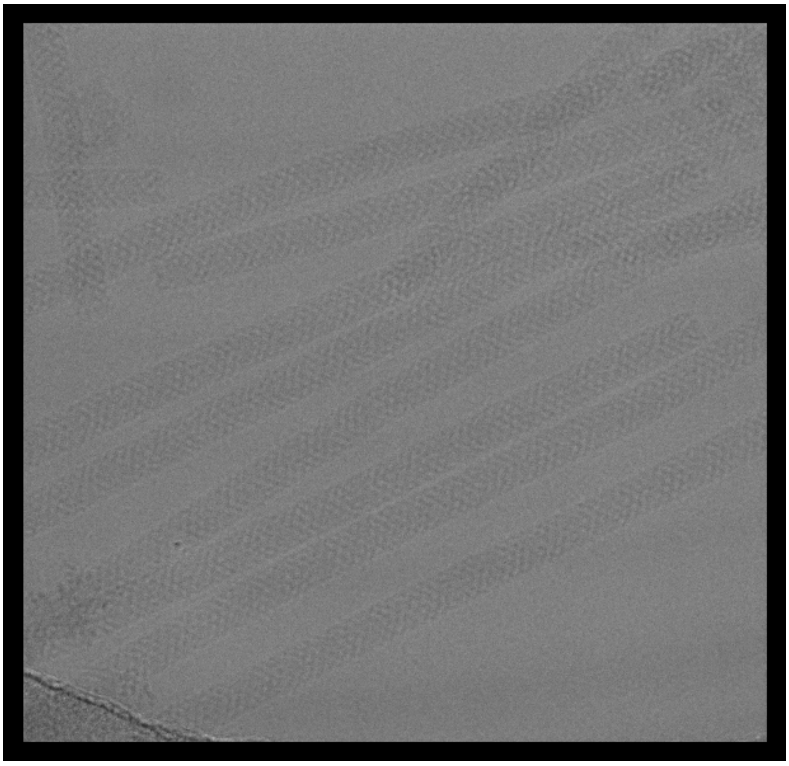


Figure 16 Motion and and CTF corrected cryo-electron micrograph

Thon rings (or power spectrum) of cryo-EM micrographs show the spatial frequency and the contrast transfer function relationship (figure 18). The left side of the image shows the theoretical (calculated) relationship and the right side shows the experimental relationship. The power spectrum is calculated by taking the absolute value of the Fourier transform of every discrete pixel in the image and then square it and finally normalize it with the total number of pixels in the image.

The oscillation of the CTF (contrast transfer function) produce a set of concentric rings. The inner rings corresponding to low resolution

information. The outermost ring corresponds to the resolution limit that is 2.521 Ångstrom on the above shown image. Image discrimination criteria was based on Thon ring evaluation.

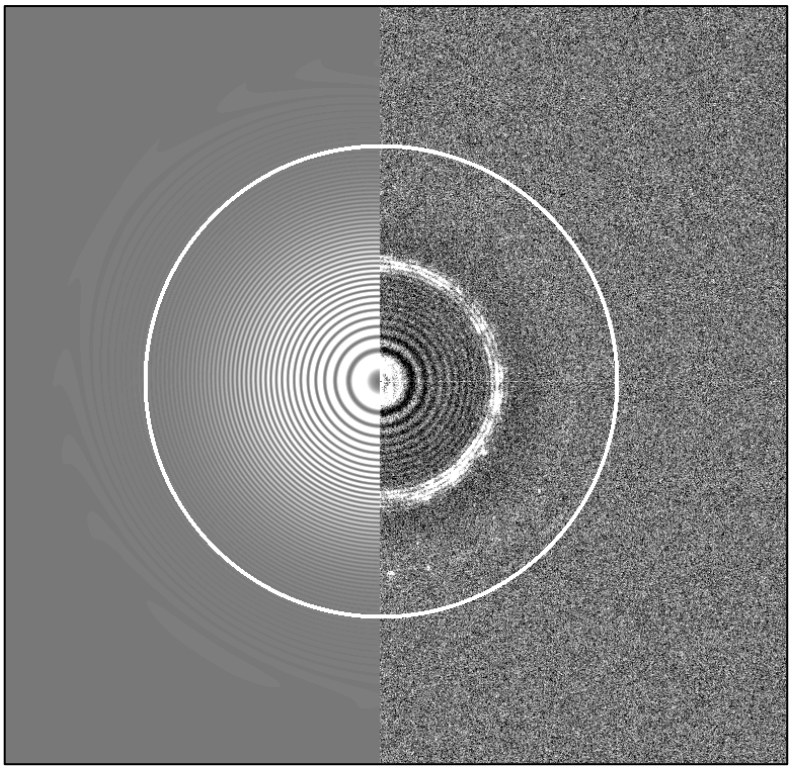


Figure 17 Thon rings (power spectrum)

2D classification

In the maximum likelihood (ML) approach the parameters describing the relative orientations of the particles are treated as hidden variables, which are integrated out in the likelihood calculation. Opposed to cross correlation, where only the highest cross correlation coefficient assignment is calculated, probability-weighted averages over all possible assignments. Using the ML method yields

better models with less bias, compared to the cross correlation method, especially in the case of low SNR images. Opposed to the most widely used cross correlation criterion (which is an extreme case of the maximum likelihood method), RELION tries to estimate the most likely model given the observed data. It was shown that maximizing the likelihood function can result in less biased model approximation than other methods (Scheres et al., 2005). 42293 particles were classified into eighty-two 2D classes (figure 19). Out of the total number of segments 41568 particles were selected out. These particles were used in the following image processing steps.

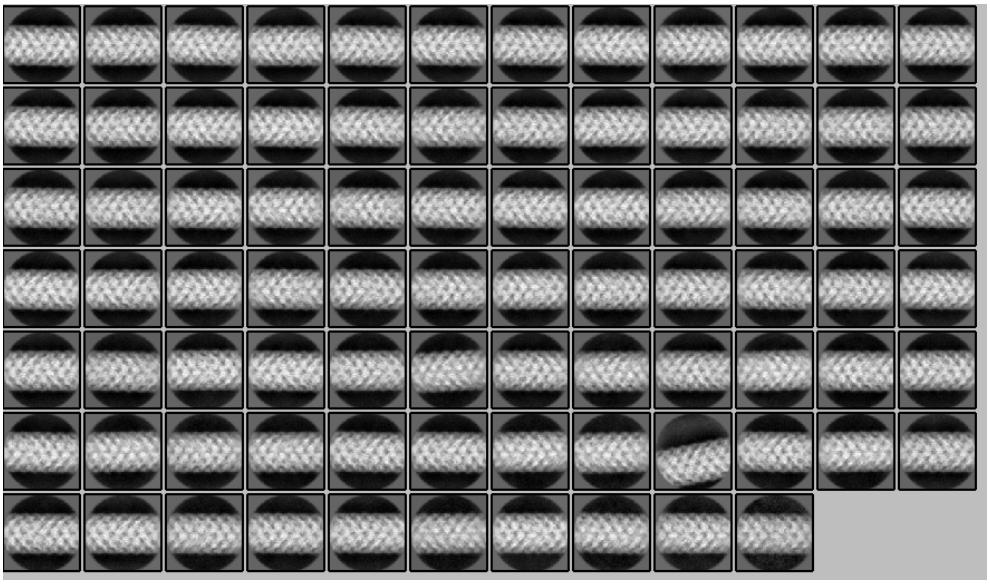


Figure 18 2D classification of polyhook cryo-EM particles

3D classification

The 3D classification technique is based on also maximum likelihood statistics. This is a suitable strategy to reveal hidden information from data sets that are large and noisy (figure 20).
The classification optimizes the following log likelihood function:

$$L(\Theta) = \sum_{i=1}^{I} \ln \sum_{k=1}^{K} \int_{\varphi} P(X_i|k, \varphi, \Theta) P(k, \varphi|\Theta) d\varphi,$$

where Θ is set of parameters which are likely to describe a heterogeneous data, X_i experimental 2D projections, $P(X_i|k, \varphi, \Theta)$ is the probability to observe X_i given model Θ, φ is rotation and translation, k (k=1,…K) are the classes, $P(k\ \varphi|\ \Theta)$ is the prior probability of φ, k. The main difference between this approach and the rest is that the discrete assignments by cross-correlation are replaced by probability weighted integrations over all assignments. Expectation-maximization algorithm is used to optimize the log-likelihood function (Scheres et al., 2007).

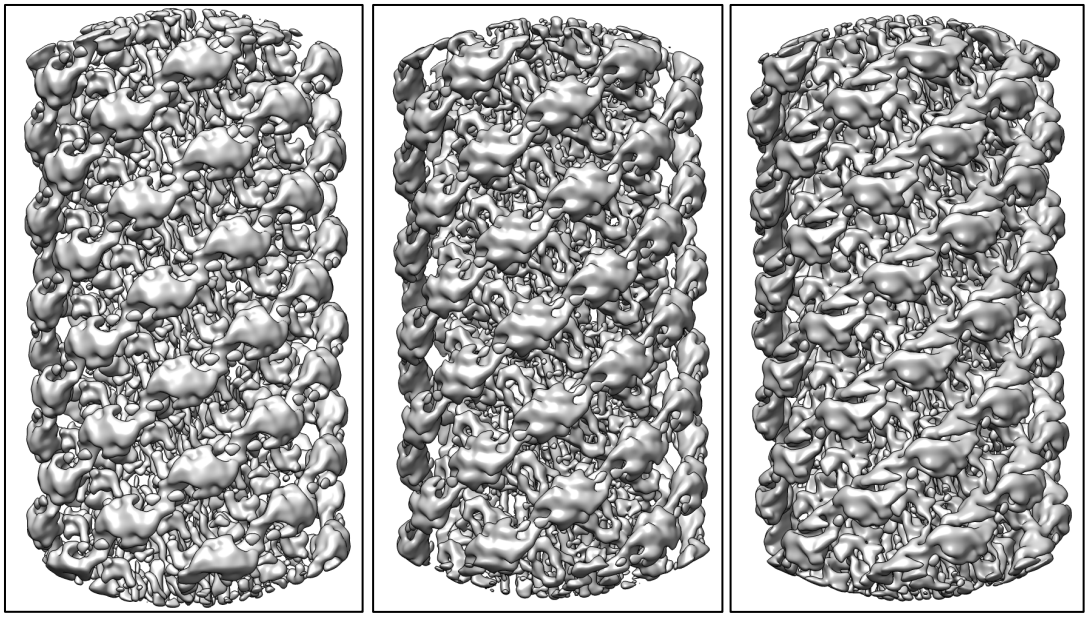


Figure 19 3D classification of cryo-EM particles

Particles were divided into three 3D classes. Number of iterations:25. Regularization parameter (T): 4. Mask diameter: 310 Å. Angular sampling interval:0.1°. Local angular search was performed. Tube diameter-inner, outer:20 Å, 200 Å

3D autorefinement

Underlying process of this is step is the so called golden standard Fourier shell correlation.

In this step the dataset is divided into two halves. The two halves are refined independently. Thereby the resulting a resolution value that is free from overfitting, representing true resolution value. Notable resolution loss as a result of dividing the dataset wasn't detectable (Scheres & Chen, 2012)

Post-processing

Semi-automated map postprocessing. Automated masking, MTF (modulation transfer function) and B-factor sharpening was done. In this the Fourier shell correlation is performed in a similar way as in 3D auto-refine but this time between the two half datasets that are masked.

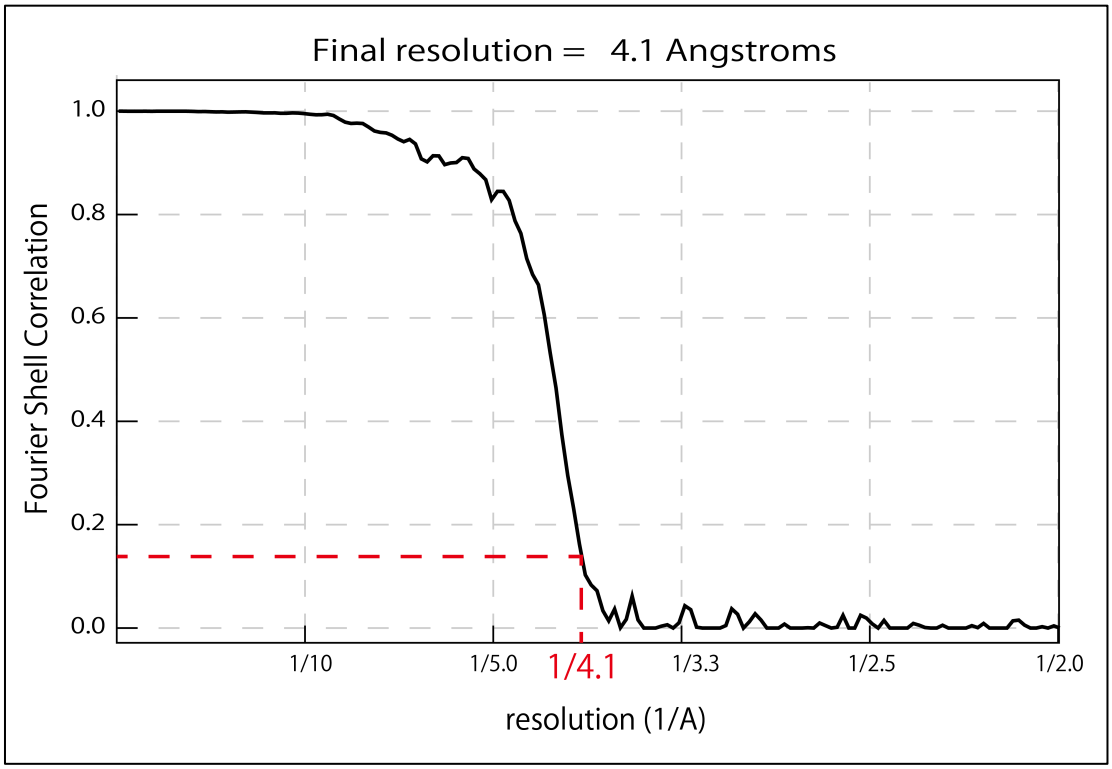


Figure 20 Fourier Shell Correlation curve

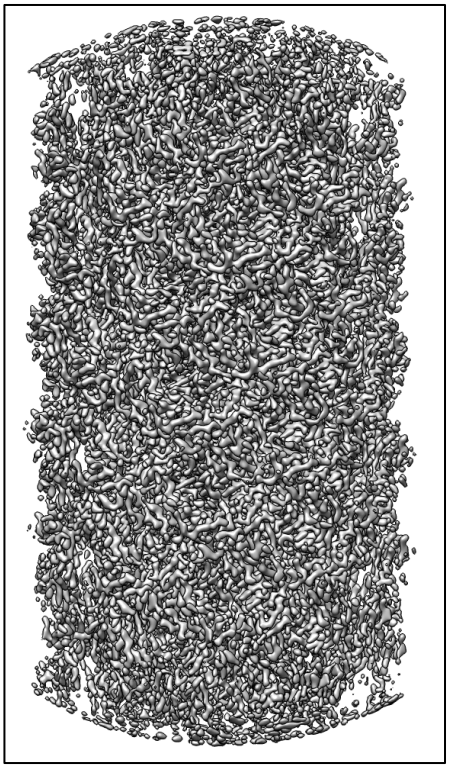


Figure 21 RELION postprocessing step

Final resolution 4.1 Ångstrom

Homology model

I used the model that was created by (Fujii, Kato, & Namba, 2009) as a target model at 7.1 Å. That model was constructed using the D2 and D1 domain structure obtained by (Samatey et al., 2004) X-ray crystallography structure of D1 and D2 domains at 1.8 Å as a target and the D0 domain of the flagellin at 4 Å (Yonekura, Maki-Yonekura, & Namba, 2003) as a template to model the atoms into the cylindrical shaped D0 domain map that was hypothesized to be a coiled coil domain.

That structure was used as a template model for building the homology model of the missing Dc domain. For this purpose *Campylobacter Jejuni*'s L-stretch was selected (Matsunami, Barker, Yoon, Wolf, & Samatey, 2016) as a target model. The *Salmonella Typhimurium*'s FlgE contains a 49 amino acid long (25-70, 364-366) unknown 3D structure portion. The aim is to rebuild this missing portion of a model by the guidance of the electron density map.
The method includes four steps:
1, identifying homologous sequences (template(s) search)
2, aligning the sequence to the template(s)
3, building model for the sequence using the template(s)
4, evaluating the model

Full atomic structure of FlgE with Dc domain obtained from a map of 4.1Å resolution cryo-EM density map (figure 23). The full sequence consists of 402 amino acids. D0 domain is: 1-25, 367-402, Dc domain: 25-70, 364-366, D1 domain: 71-144, 285-363, D2 domain: 145-284. Fujii et. al structure was used as a target model. The resolution was high enough to make an atomic model either by homology modeling or ab initio method for practical reasons the former one was chosen. *Campylobacter Jejuni* L-stretch was used as a template model for the Dc domain. (Matsunami et al., 2016).

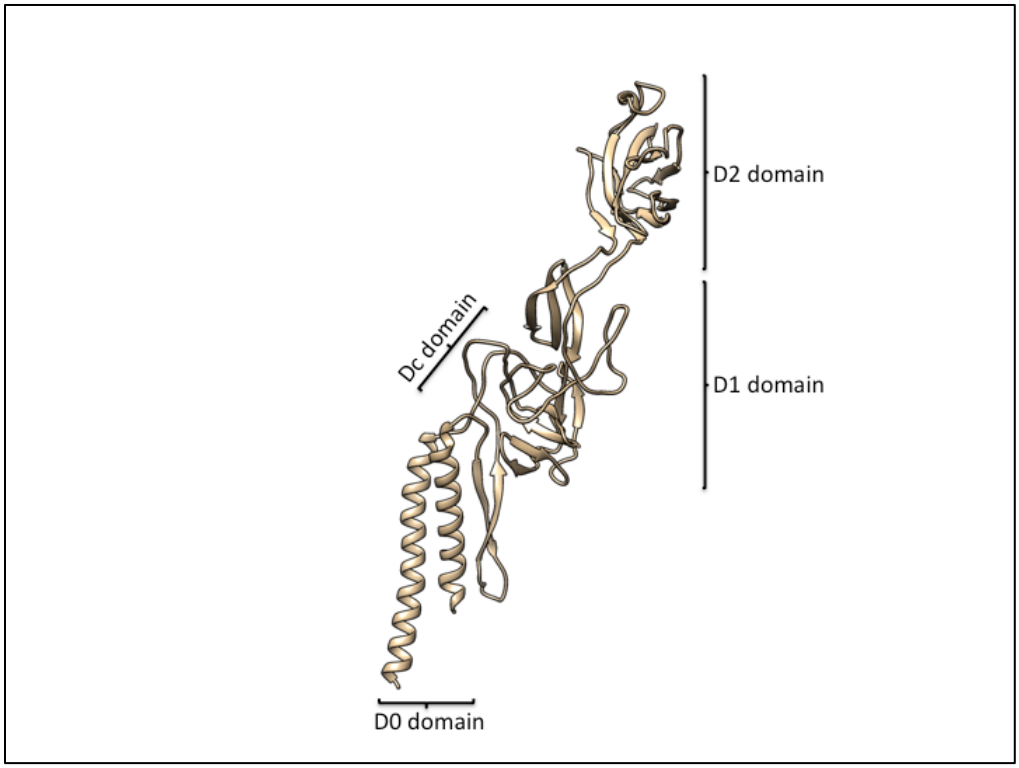


Figure 22 Full atomic structure of FlgE

Atomic model fit into electron density map

Rosetta tools for structure determination in the density was used to fit the atomic model into the electron density map. The algorithm consists of a Cartesian sampler mover. It checks backbone segments given local strain and local density agreement. Z-score is an indicator of other refined near atomic structure similarity to each other. If it is below certain threshold, it will get rebuild. 200 rebuilding cycles were set. 5 backbone segment were replaced in one rebuilding cycle. Weighting of amino acid side chain were the default values. Finally the structure was minimized by a low resolution energy function in order to improve the protein backbone geometry (figure 24).

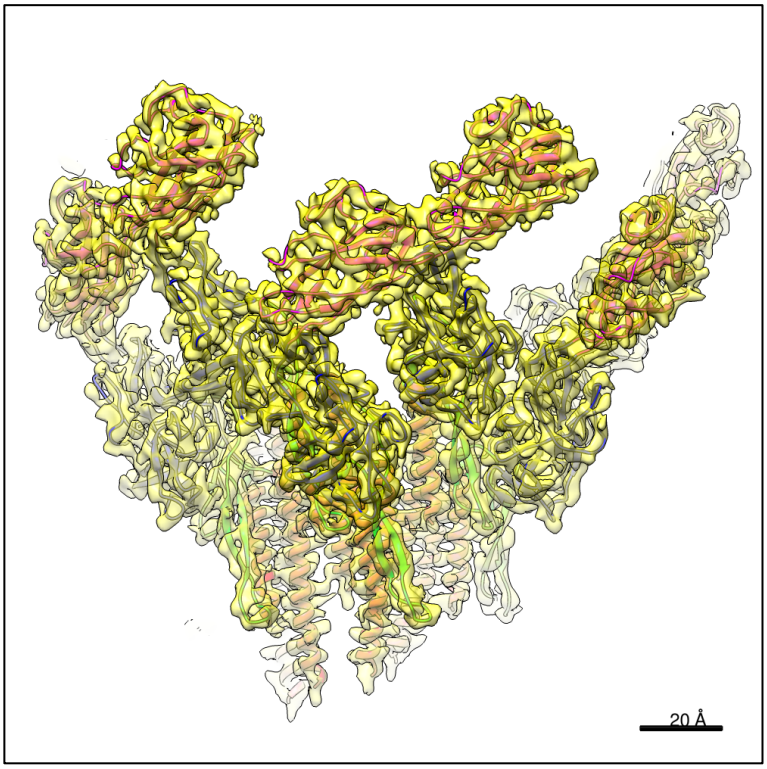


Figure 23 FlgE D0, Dc, D1, D2 atomic model fit in the electron density map

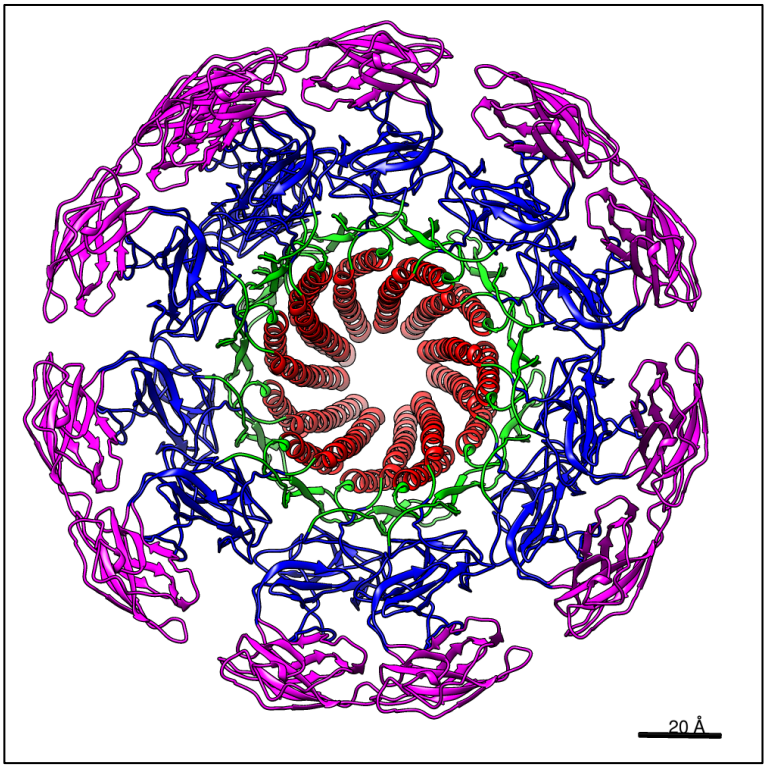


Figure 24 FlgE D0, Dc, D1, D2 domain atomic model

Side view of 11 FlgE (figure 24) atomic model domains fit in to the electron density map segmented out from the full electron density map. According to the 0.143 Fourier-shell criteria the resolution proved to be 4.1Å. Secondary structures are clearly visible. The different domains are colored with different colors. Proceeding radially from the innermost domain: D0 domain: red, Dc domain: green, D1 domain: blue, D2 domain: magenta.

The atomic structure end on view (figure 24). Domains are arranged radially. The coloring of the different domains are the same as on the figure above.

The span diameter of the full atomic structure is approximately 180Å.

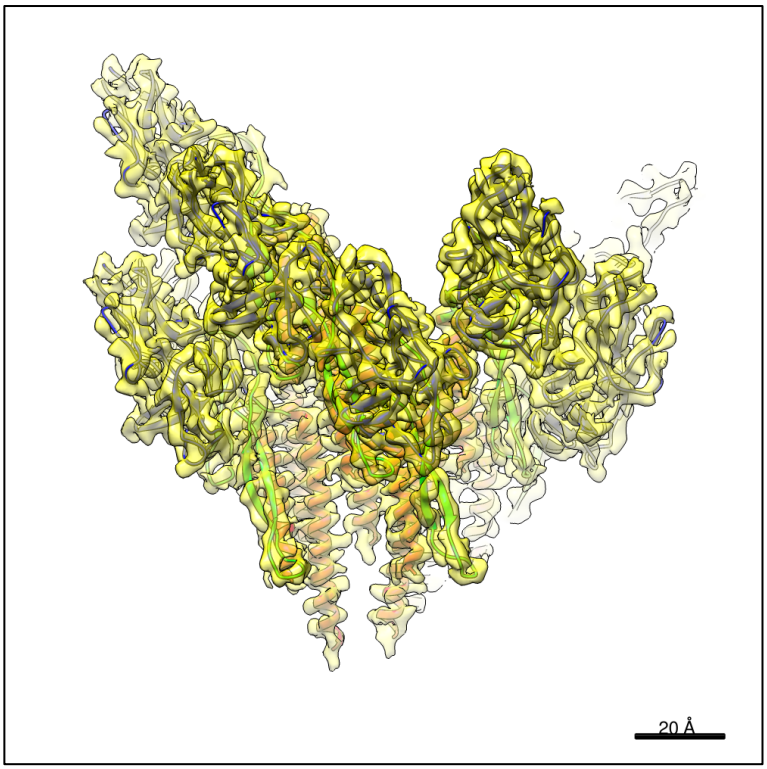


Figure 25 FlgE D0, Dc, D1 atomic model fit in the electron density map

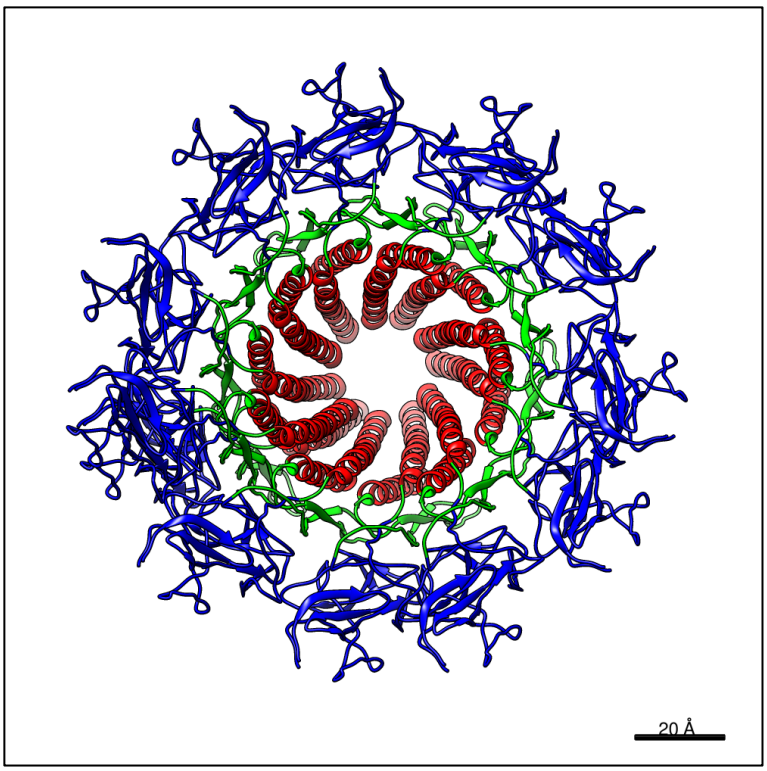


Figure 26FlgE D0, Dc, D1 domain atomic model

Side view of 11 FlgE (figure 25) atomic model fit in to the electron density map segmented out from the full electron density map. End on view of D0, Dc, D1 atomic model (figure 26). The approximate span diameter is 145Å. D0 domains are colored in red, Dc domains are green, D1 domains are blue.

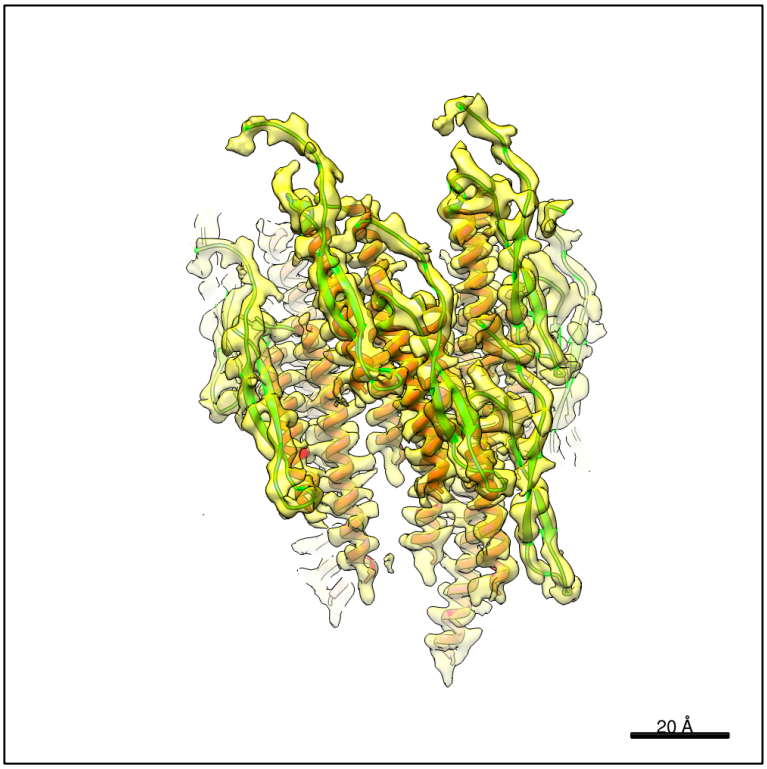


Figure 27 FlgE D0, Dc atomic model fit in electron density map

Side view of D0, Dc atomic model fit in electron density map (figure 27). Dc domains are discernable for the first time. Color code is the same as in figure (figure 31)

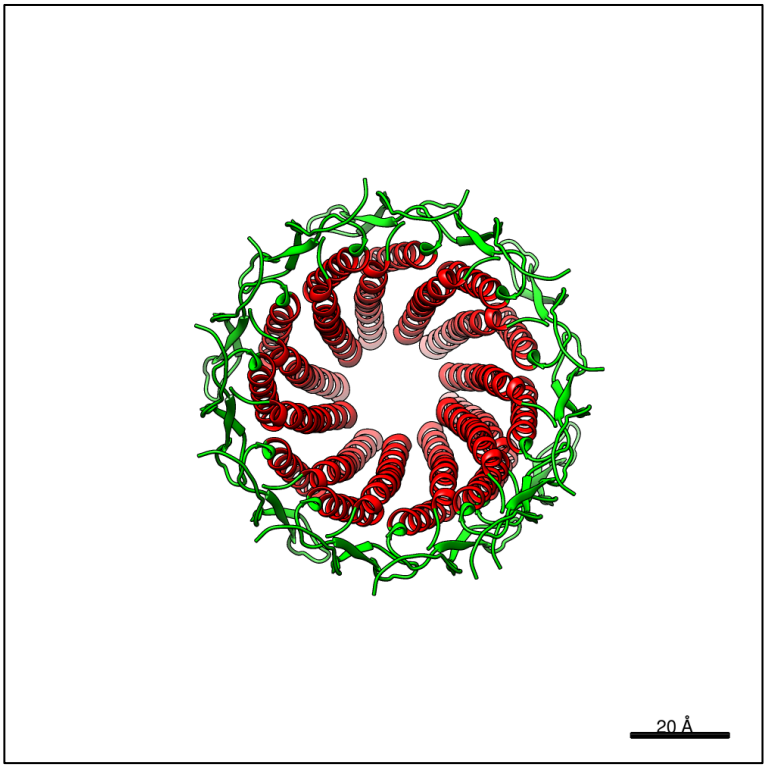


Figure 28 D0, Dc domain atomic model

End on view of Dc, Do atomic model (figure 28). The approximate span diameter is 90Å. Coloring is the same as above. The Dc domain electron density map and atomic structure were obtained for the first time. The Dc domain wraps around the cylindrical surface spanned by the D0 domains like a mesh. The role of the Dc domain is to stabilize the FlgE protein. The Dc domain shows similarity with *Campylobacter Jejuni*'s hook and with the *Salmonella Typhimurium* distal rod, FlgG Dc domain.

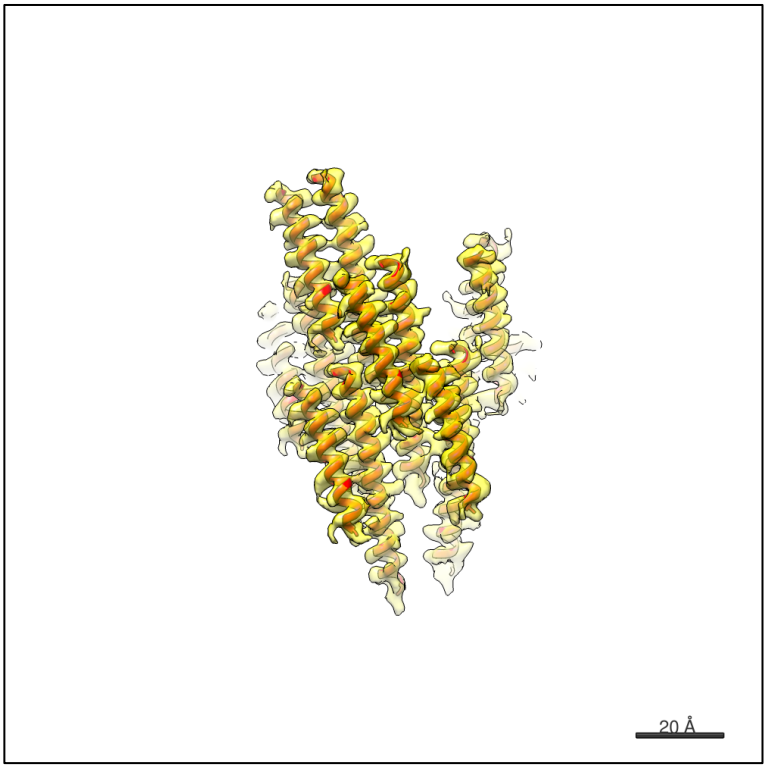


Figure 29 FlgE D0 atomic model fit in electron density map

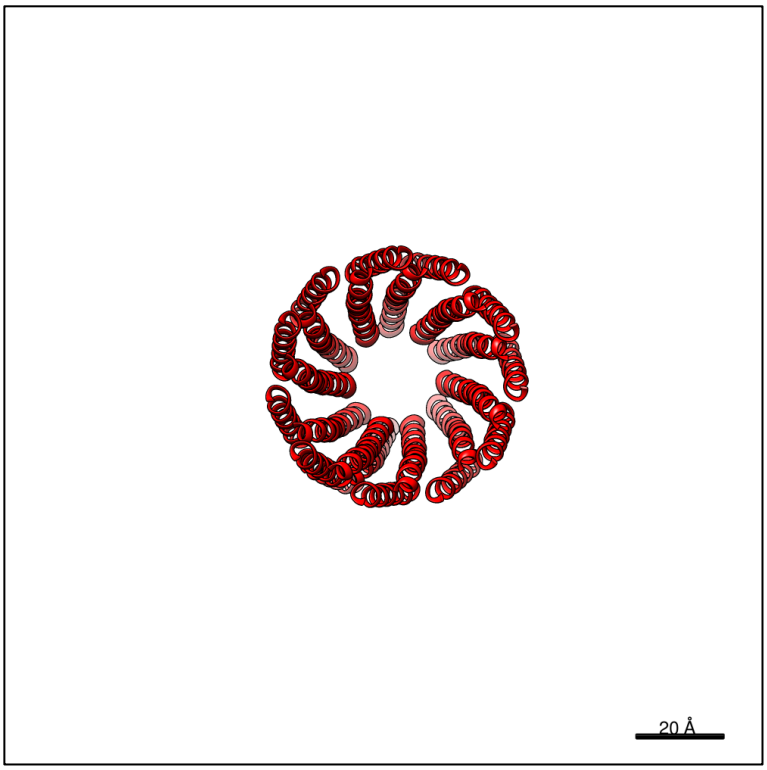


Figure 30 D0 domain atomic model

Side view of the Do domain fit in the electron density map (figure 29). Below the end on view of the atomic structure of the alpha helices arranged in coiled coil (figure 30). The inner channel diameter is approximately 18 Å. The inner side of the helix is polar in order to facilitate the subunit export. The outer D0 domain span is roughly 60 Å. This arrangement is capable to interact along the all major helical start directions so it is responsible for the mechanical stability of the hook but it is capable to accommodate extension and compression along the axis of the hook. The N-terminal consists of 26 amino acids and the C-terminal constitutes of 36 amino acids. This domain is well conserved among other bacteria's (for example: *Campylobacter Jejuni*) as well as other axial proteins of *Salmonella enterica serovar Typhimurium* (rod, flagellin show relatively high percentage identity).

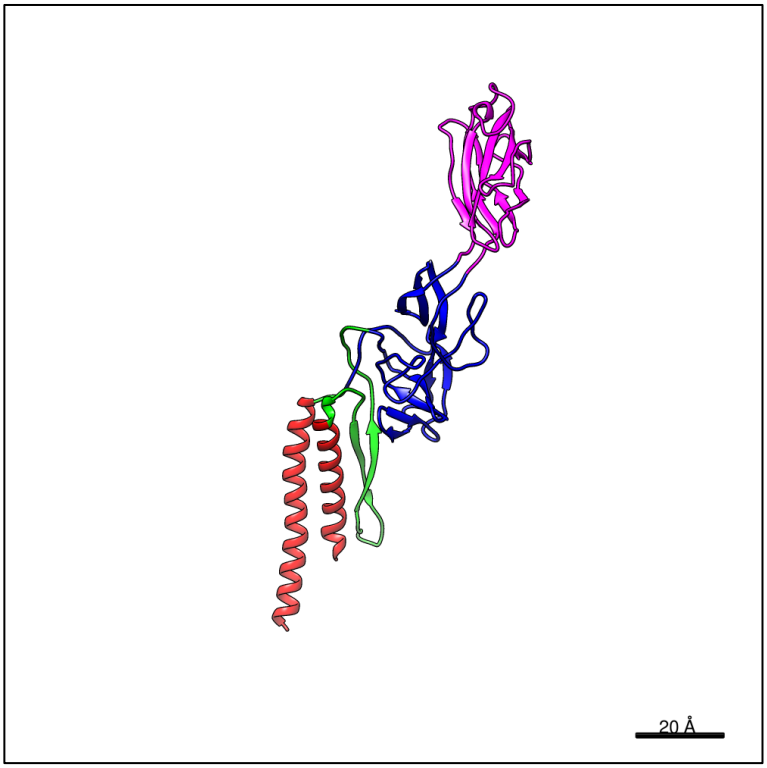


Figure 31 FlgE D0, Dc, D1, D2 domain atomic model

FlgE consist of 402 amino acids, arranged in 4 radially arranged domains. From the innermost proceeding radially outward: D0, Dc, D1, D2 respectively (figure 31).

The innermost D0 domain is composed of two coiled-coil helices. The N-terminal helix (1-25) is approximately 40Å in length and the C-terminal helix (367-402) is roughly 57Å long. In the case of the previous cryo-EM structure helices of the D0 domain were represented by two featureless cylinders, pitches were indiscernible because the relatively low resolution of the structure. The recent structure reveals traceable helical pitch and the missing Dc domain and the refined full atomic structure. The C-terminal helix is directly connected to D1 domain, whereas between the N-terminal helix and the D1 domain the Dc domain (25-70, 364-366) is inserted, named Dc domain. The Dc domain was unrevealed in previous structure

analysis works. The Dc domain is roughly 60Å in length. It has almost parallel orientation with respect to the D0 domain.
The next domain proceeding radially outward from the axis of the hook, between Dc and D2 domain is named D1 domain. D1 domain (71-144,285-363) is a sophisticated fold containing antiparallel beta sheets separated by loops and a special triangle shaped loop (116-135). The outermost domain, named D2 domain (145-284) forms a nine stranded beta barrel with strands connected to each other by loops.

Discussion

Domain organization of *Salmonella Typhimurium's* hook

The hook is a tubular fibre. It consists of 11 protofilaments. Protofilaments are cooperative subunits between the and L and R states to produce the superhelical form of the filament. It posseses helical symmetry, with helical rise and azimuthal rotation 4.05Å and

64,78° respectively. The domains of the FlgE monomer are arranged in a ring structure when viewed along the central z-axis.
The hook diameter is approximately 180Å. The helices of the DO domain form an inner tube that has a span diameter of about 18Å. The outer diameter of the DO domain is nearly 60Å. The DO domain is surrounded by the mesh like Dc domain which has a span diameter measured from the axis of the polyhook of about 90Å. D1 domain span diameter roughly 145Å. The outermost D2 domain has a span diameter of about 180Å.

D0 domain the inner channel.

The N and the C-terminal helices in the DO domain form coiled coil structures. Coiled coils are relatively rigid (although we will see later some freedom of motion along the axis of the hook is possible, axial compression, extension). The strong interaction between the coiled coil helices along every helical-start directions provides the mechanical stabilization of the hook. 11 coiled coil helices form the inner channel with an approximate span diameter of approximately 18Å. In this channel axial protein subunit transport takes place. The inner wall of the channel consists mainly of polar residues with side chains protruding towards the lumen. This arrangement facilitates the transport of the unfolded subunits through the channel and could possibly prevent the association between the walls of the tube. The DO domain shows relatively high homology with one of the axial proteins called distal rod (FlgG) which serves as the driving shaft and other bacterial species like *Campylobacter Jejuni* (figure 32). The DO

domain terminals also share some similarities with other axial proteins of *Salmonella Typhimurium* flagellum. This similarity pattern is called the heptad repeat, which manifests itself in a hydrophobic amino acid in every 7th residue position counted from the terminal.(Homma et.al)

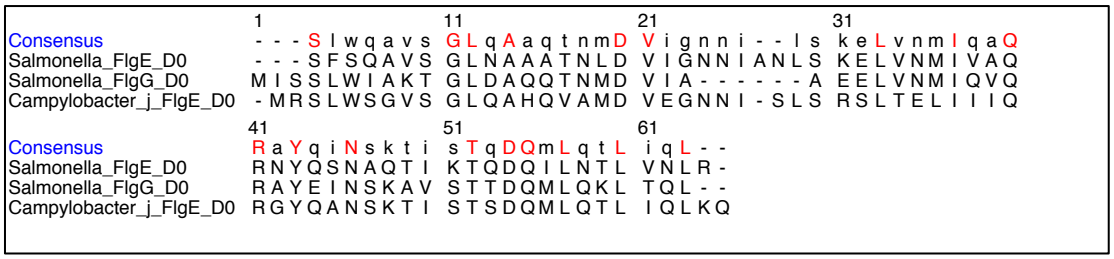


Figure 32 D0 domain multiple protein sequence alignment

Multiple sequence alignment shows the similarity between the *Salmonella Typhimurium*'s FlgE, FlgG D0 domain and *Campylobacter Jejuni*'s FlgE D0 domain. Salmonella FlgE consists of 61 amino acids, FlgG is 57 amino acid long and *Campylobacter Jejuni*'s FlgE D0 domain sequence contains 62 amino acids. The percentage identity between the Salmonella FlgE and FlgG, *Campylobacter Jejuni*'s D0 domain is 45,61%, 48.39% respectively. After a pairwise sequence alignment, a fitting is performed based on the aligned residue pairs. A good measure of the 3D structural similarity is the root mean squared deviation of the backbone atoms of the superimposed proteins. The RMSD value between the N-terminal helices of Salmonella FlgE and FlgG was 0.424Å, for the C-terminal helices it was 0.783Å.

Dc domain and its putative role in stabilization

Dc domain (25-70, 364-366) is composed of 49 amino acids. It connects the D0 domain with the D1 domain. The N-terminal end of DO domain is directly connected via a short six amino acid sequence to the Dc domain. The Dc domain contains a 6 residue long N-terminal stretch, two antiparallel beta strands (composed of seven residues) connected by a 11 amino acid loop (MFAGSKVGLGV), a twelve amino acid long flexible sequence (DFTDGTTTNTGR) and a C-terminal composed 3 residues. According to the NIH protein blast search (Altschul, Gish, Miller, Myers, & Lipman, 1990) in the PDB databank there exist a 34% identity protein, namely the hook of the *Campylobacter Jejuni*. The *Campylobacter Jejuni*'s hook composed of five distinct domains. Similarly to the Dc domain of the Salmonella's hook, a loop called L-stretch in Campylobacter is responsible for connecting the DO domain with the D1 domain. The pairwise sequence alignment shows 36.73% sequence identity between the Salmonella's hook Dc domain and the L-stretch of the *Campylobacter Jejuni*'s hook. This L-stretch consists of 75 amino acids. Amino acid composition is arranged in flexible loops and two antiparallel beta sheets and similarly to the Salmonella Dc loop the two beta sheets are connected by a 14 amino acid long (IATAPTDGRGGSNP) sequence. The Campylobacter's hook plays a significant role in the structure stabilization (Matsunami et al., 2016). The relatively high percentage residue identity and the similarity in the secondary structures implicate the similar biological function, namely the stabilization of the tertiary and quaternary structure.

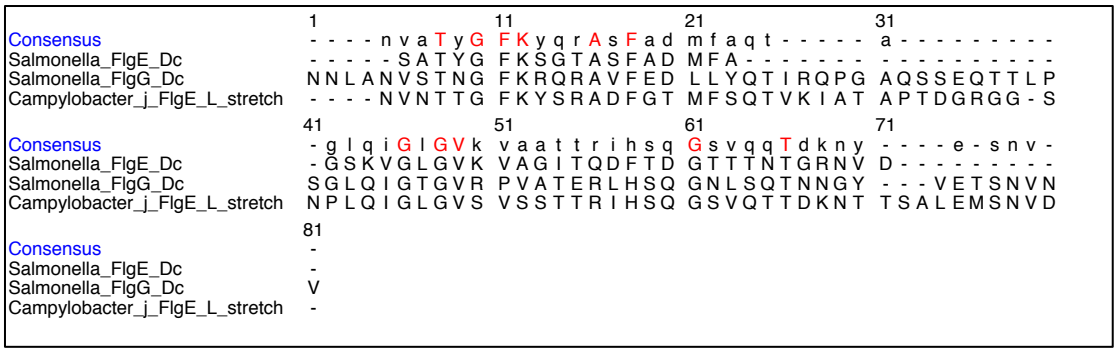


Figure 33 Dc domain multiple protein sequence alignment

Pairwise sequence alignment shows 36.73% percentage identity between the two domains Dc domain and L-stretch from *Salmonella Typhimurium* and *Campylobacter J*. respectively (figure 33). Homology entails the functional similarity and the possibility of structural relationship between the two proteins. In the case of the Dc domain both conclusions can be derived from the relatively high percentage identity. The axial proteins show also structural similarity, they also exhibit sequential correspondence, which resulted in 28.57% percent identity. The RMSD between the N-terminal of the Salmonella FlgE Dc domain and the N-terminal of the *Campylobacter J*. FlgE hook is 0.982Å. The RMSD value between the C-terminal of the same domain was 0.059Å. In the Salmonella FlgE Dc domain there is a 17 amino acid residue empty region after 19th residue. This region will play an important role in flexibility and rigidity as I will describe later on.

The 49 amino acid structure of Dc domain were completely missing from previous works on *Salmonella Typhimurium*'s hook structural analysis because of the flexible structural elements of the loop. I will show later that the Dc domain has a putative role in flexibility and stabilization of the hook.

D1 and D2 outer domains

D1 domain (71-144, 285-363) appears to be a complex fold. It contains beta hairpins and beta strands connected with longer and shorter segments of (11aa, 50aa, 7aa, 8aa, 18aa) flexible stretches. The D1 domain consists of 9 beta strands arranged in three double stranded antiparallel beta sheets which form an approximate angle of 120° with the adjacent beta sheet and a beta meander consists of three stranded antiparallel beta sheet. The D1 domain forms a triangular loop(116-135) which has a key role of interacting with the above D2 domain (Samatey et al., 2004).

The D1 is connected to the D2 domain on both the C-terminal and N-terminal via a short flexible stretch. The fold of the D2 domain is an 10 strand beta barrel structure, the strands being connected, similarly to the D1 domain via short, flexible stretches. According to (Sakai et al., 2017) recently published paper D1 domain has it's role in flexibility, while D2 domain is contributing to the polymorphic transformation of the hook. Flexibility is realized due to the loose interactions between domains along certain helical-start lattice lines. These loosely connected domains will result in a flexible quaternary structure. This will allow the protein to function as a universal joint. I will discuss these interactions in detailed below.

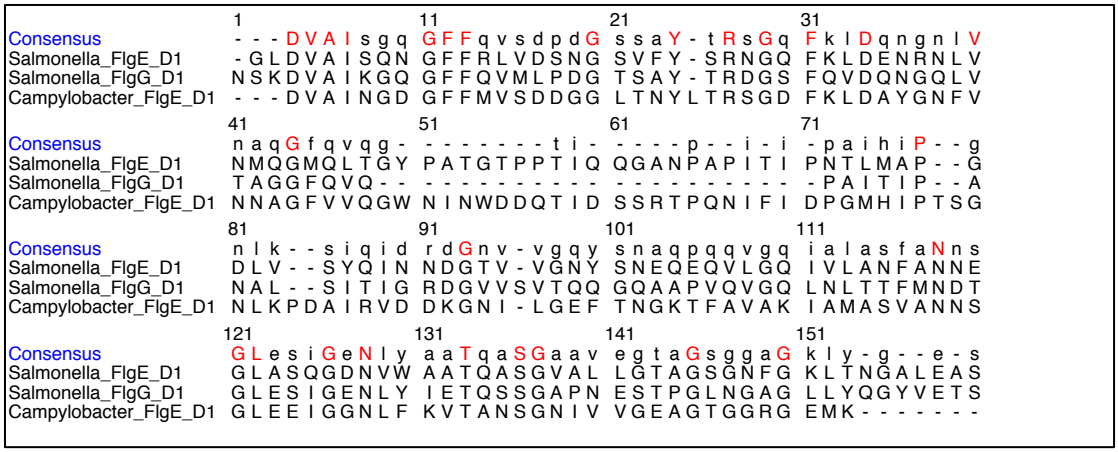


Figure 34 D1 domain multiple protein sequence alignment

Multiple sequence alignment shows sequential similarity between the Salmonella FlgE D1 domain, Salmonella FlgG D1 domain and Campylobacter FlgE D1 domain respectively (figure 34). In the former case the percent identity between the two sequences was 34.85% and in the latter case it was 32.89%.

The RMSD value between the main chain residues of Salmonella FlgE D1 domain and the main chain atoms the Salmonella FlG D1 domain is 0.770Å. This value for the same domain C-terminal was. The average distance between the C-terminal backbone is 0.819Å.

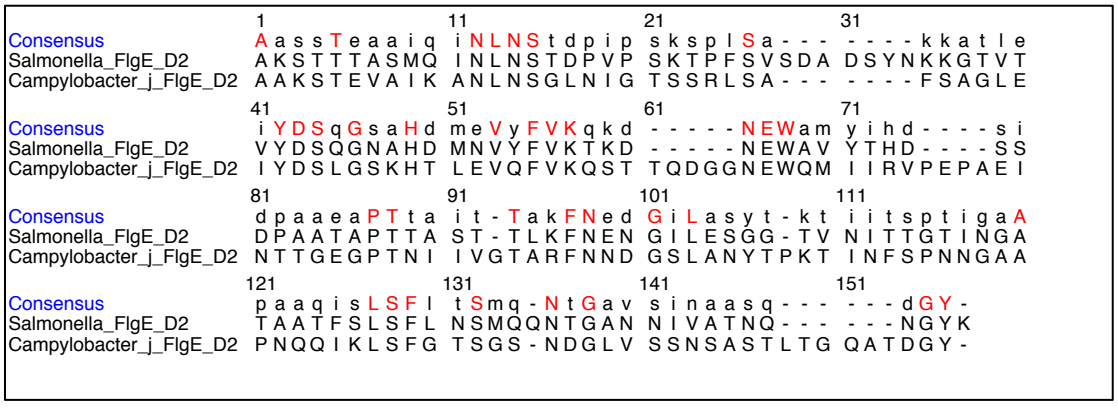


Figure 35 D2 domain pairwise protein sequence alignment

Pairwise sequence alignment between Salmonella FlgE D2 domain and *Campylobacter J*. FlgE D2 domain (figure 35). The percent identity between the two sequences is 25%. The RMSD between Salmonella FlgE and *Campylobacter J*. FlgE D2 N-terminal is 0.907Å. The RMSD between Salmonella FlgE and *Campylobacter J*. FlgE D2 C-terminal is 0.896Å.

Interactions along helical lines

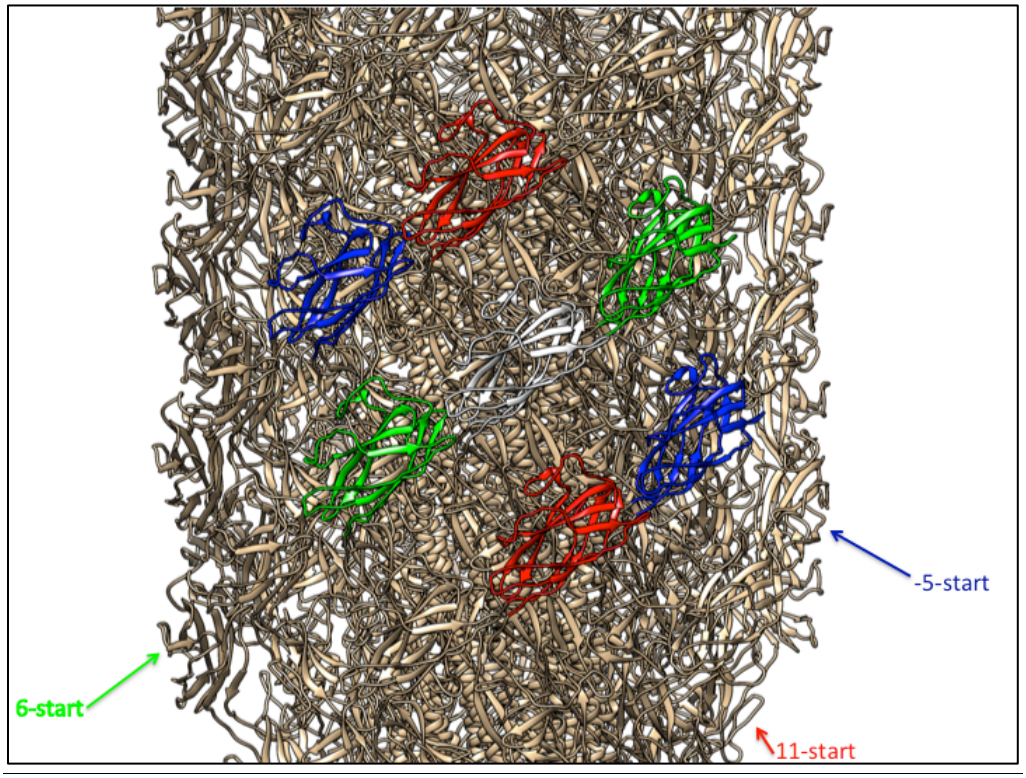


Figure 36 Helical directions

Figure 36 indicating the main interaction directions in the helix by coloring the D2 domains of the subunits that are localized along different helical lines. (For simplicity and transparency I omitted the

other domains). There are three main directions: 6-start (green arrow and green D2 domains), 11-start (red arrow, red D2 domains), -5-start (blue arrow, blue D2 domains). The white subunit in the intersection of the three main helical lines indicates that the subunit participates in each helical direction. (Note: The union of green, red and blue is white).

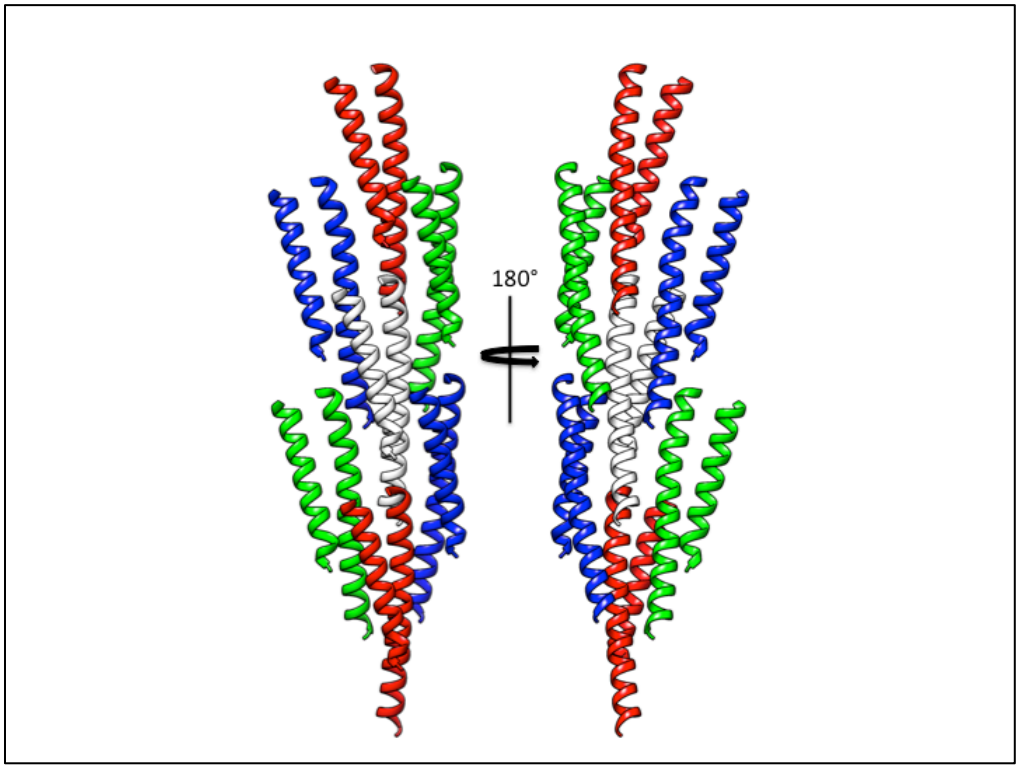


Figure 37 D0 domain interaction summary

Figure showing the summary of interactions among D0 domains along the three main helical directions (figure 37). Three domains of a FlgE subunit were selected to represent the interactions along the three main helical directions in the helix. The red helices showing interactions along the 11-start helical direction. The green helices show interactions along the 6-start direction. The blue helices

represent helices along the –5 start direction. The white helix in the middle represent the helix that participates in every helical direction

D0 interactions

Along the -5 start:

The n-th subunit's N-terminal helix of the coiled coil pairs show polar interactions along approximately 2 pitches long interaction interface with adjacent (n+5)-th subunit's D0 domain's N-terminal.
The n-th subunit's C-terminal helix has a more extensive interaction site along 6 pitches on the neighboring (n+5)-th subunit's C-terminal helix (figure 38).

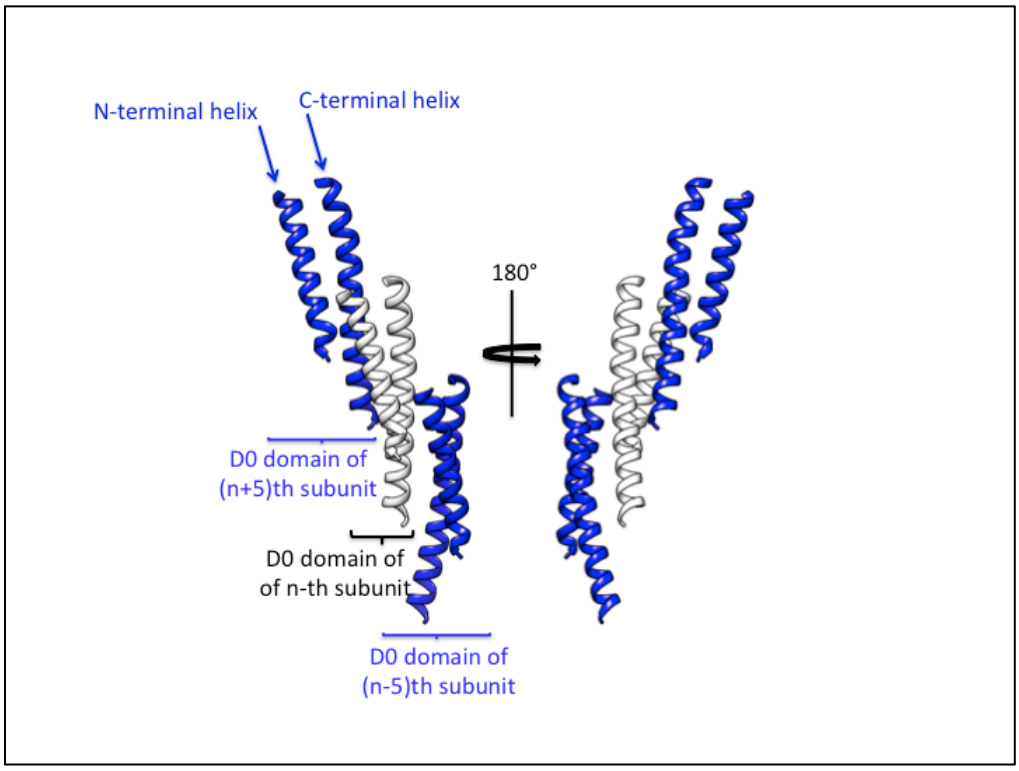


Figure 38 D0 interactions along -5 start direction

Along the 11-start:

Along this direction n-th subunit's D0 domain's C-terminal helix of the coiled coil motif interacts on either end with the neighboring subunits, namely the (n+11)th and the (n-11)th subunit's C-terminal along a short interaction site of two pitches on both ends.
Along this direction the N-terminal helices are not connected to each other. The n-th subunit's D0 domain's N-terminal helix is parallel with the (n+11)th and the (n-11)th subunit's D0 Domain's N-terminal helices with an offset of about 10Å, which is approximately the diameter of a helix.

The (n)-th subunit's D0 domain's N-terminal helix is positioned above the (n-11)-th subunit's D0 domain's C-terminal helix with a gap insertion of about one helical pitch between them, approximately 6Å. This gap plays an important role in the flexibility of the protein (figure 39).

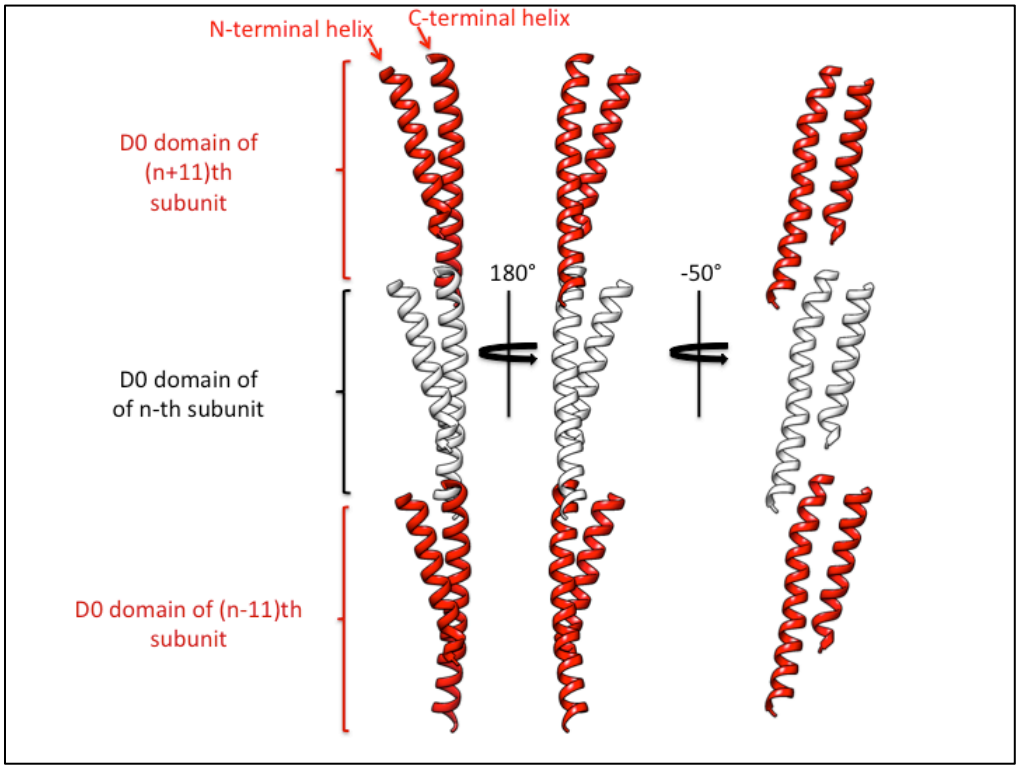


Figure 39 Do domain interactions along 11-start direction

Along the 6-start:

This interaction direction shares some similarity with the -5-start direction interactions.

The n-th subunit's D0 domain's C-terminal extensively interacts with (n+6)th subunit's Dc domain's C-terminal helix along 6 pitches (figure 40). Similarly to the -5-start helical directions. But in this direction there are no interactions observable between the n-th subunit's N-terminal helix and the (n+6)th subunit's D0 domain's N-terminal helix. Instead of that interaction the n-th subunit's D0 domain's N-terminal shows affinity to interact with the (n-6)th subunit's DO domain's C-terminal along 3 pitches through Van der Waals forces.

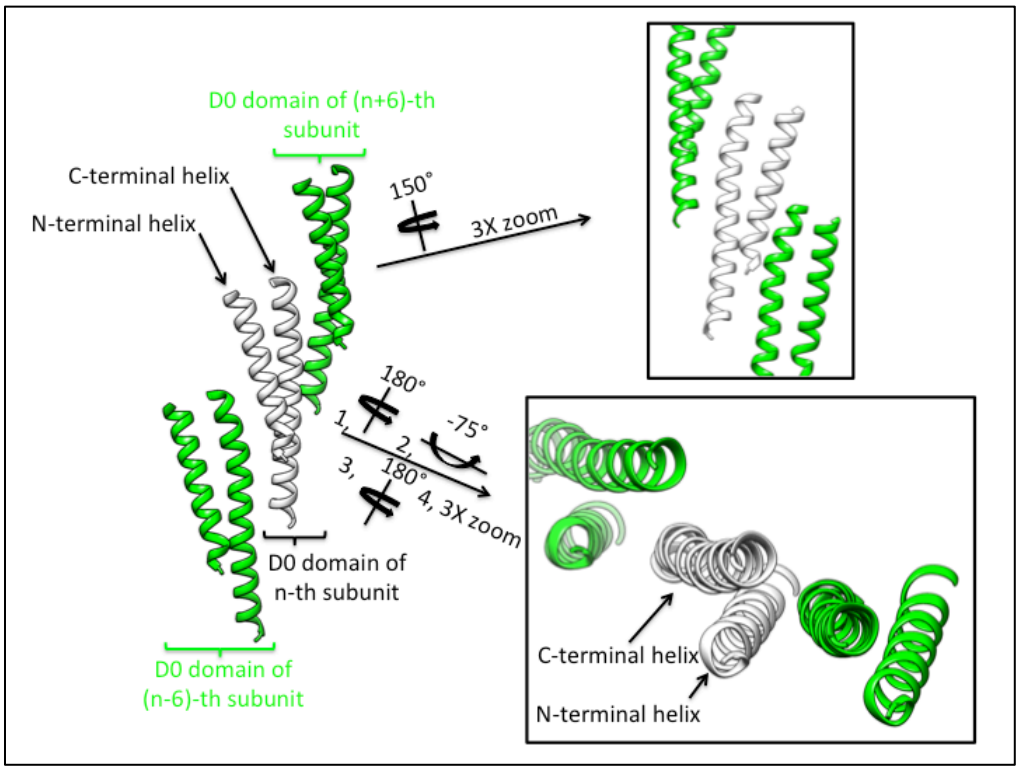


Figure 40 D0 interactions along 6-start direction

DC interations

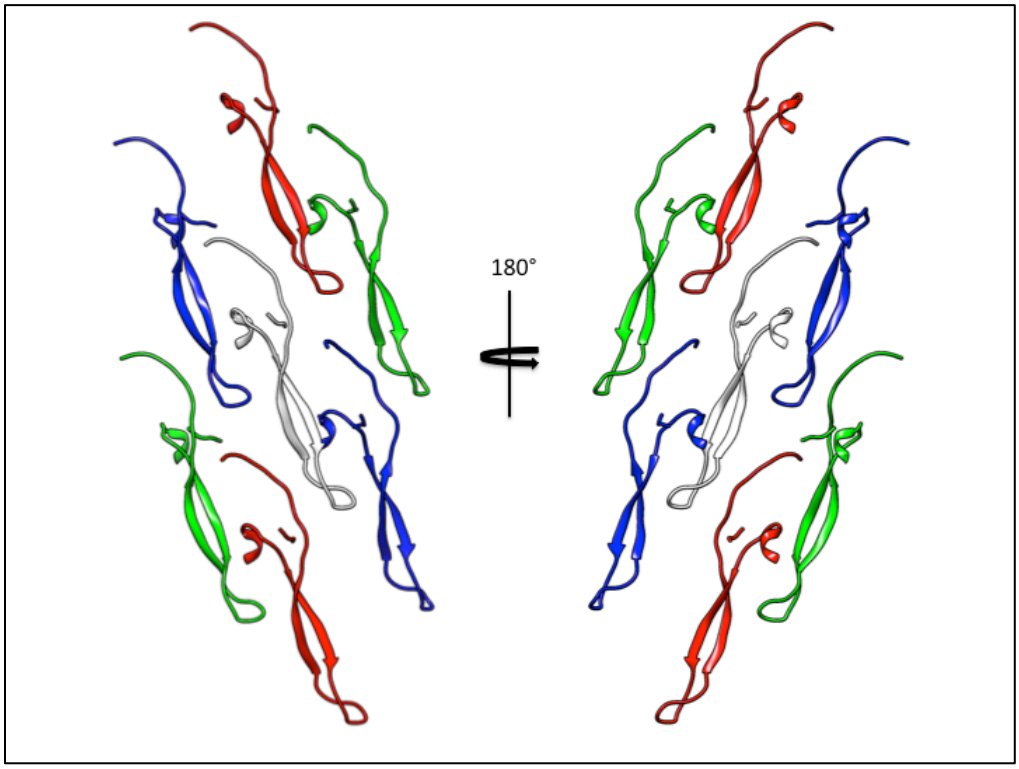


Figure 41 Dc domain interaction summary

Figure 41 showing the summary of interactions among Dc domains along the three main helical directions. Three domains of a subunit were selected to represent the interactions along the three main helical directions in the helix. The red Dc domains showing interactions along the 11-start helical direction. The green Dc domains indicating interactions along the 6-start direction. The blue Dc domains represent interactions along the –5 start direction. The white Dc domain lying in the intersection of three main helical directions represent that this Dc domain participates in every helical directions.

Along the -5 start:

In the -5 start helical direction the Dc domains show intimate packing (figure 42). The n-th subunit's Dc domain's beta sheet interacts with the (n+5)th subunit Dc domain's hairpin's turn. The n-th subunit's Dc domain's C-terminal and N-terminal stretches both show affinity to interact with the (n+5)th subunit Dc domain's beta sheet. Interactions are prevalently made by hydrophobic amino acids.

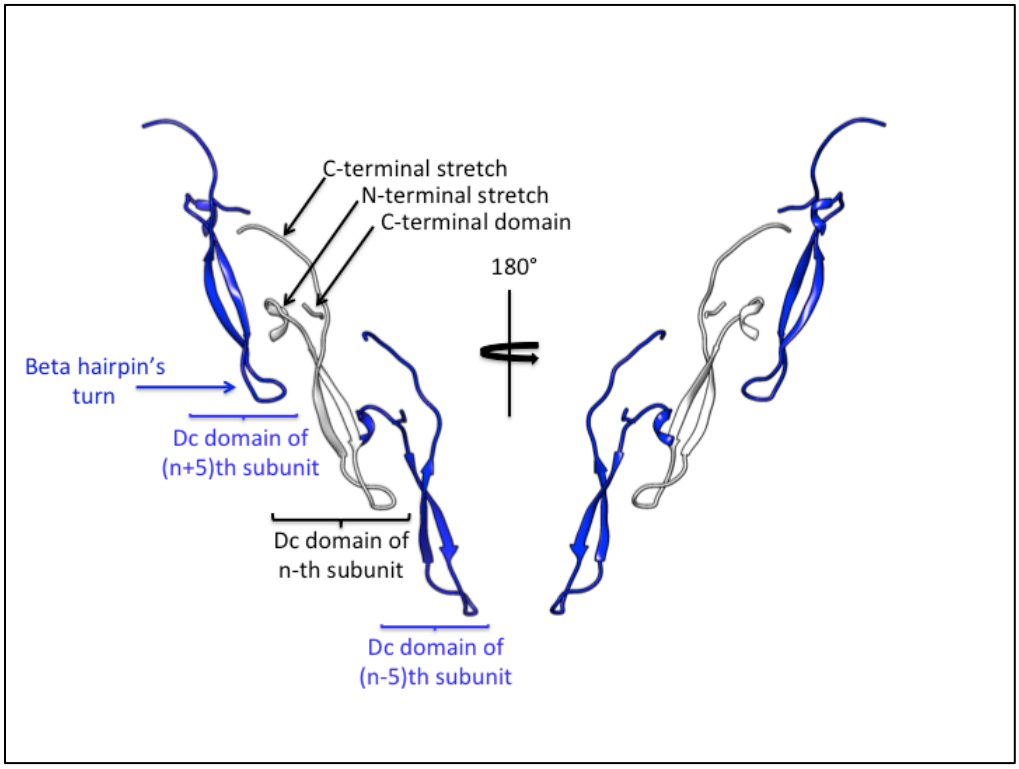


Figure 42 Dc domain interaction along -5-start direction

Along the 11-start:

The n-th subunit's Dc domain's C-terminal stretch interacts with the (n+11)th subunit's Dc domain's beta hairpin's loop via non covalent interactions (VDW forces) (figure 43). Both the stretch and the loop

are decorated with hydrophobic and polar amino acids. Since the number of hydrophobic amino acids that are capable of forming interactions are rather limited, hence the 11-start interactions seem to allow axial compression and extension with the rearrangement of the bonding interactions.

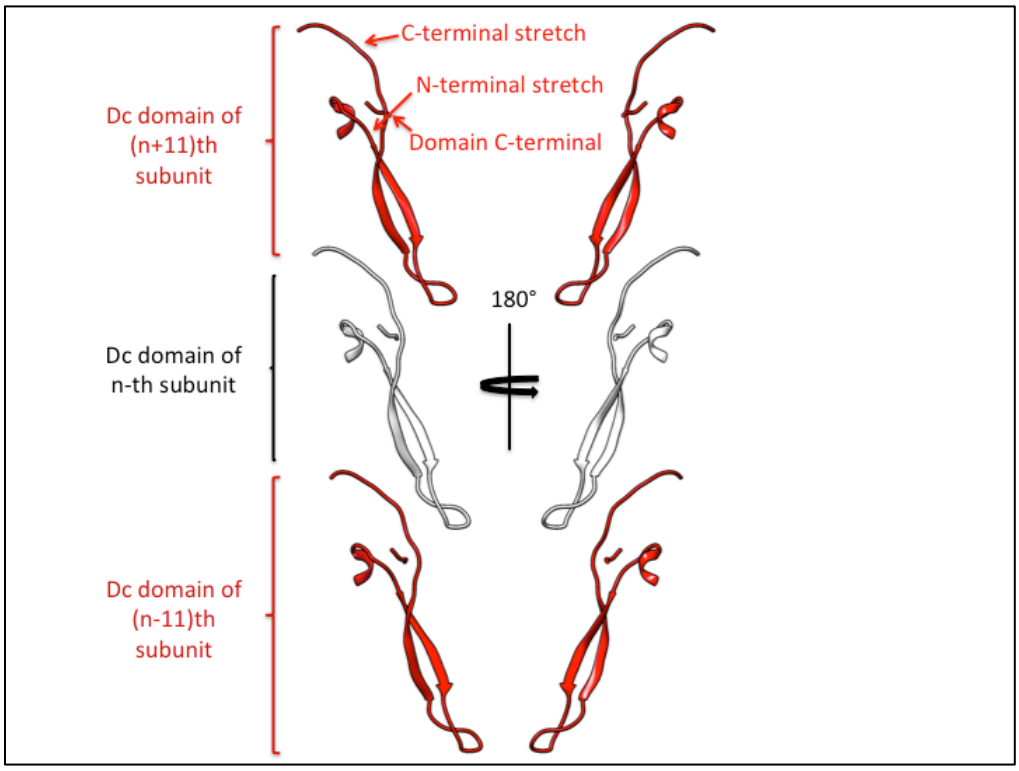


Figure 43 Dc domain interactions along 11-start direction

Along the 6-start:

There are no observable contacts along the 6-start helical direction. The neighboring subunits are unable to interact with each other because of the approximately 10Å distance between two n-th and the (n+6)th subunit's Dc domain (figure 44). This space between the

subunits along the 6-start direction allows the hook to be flexible and to bend, thereby contributing to the hook universal joint function.

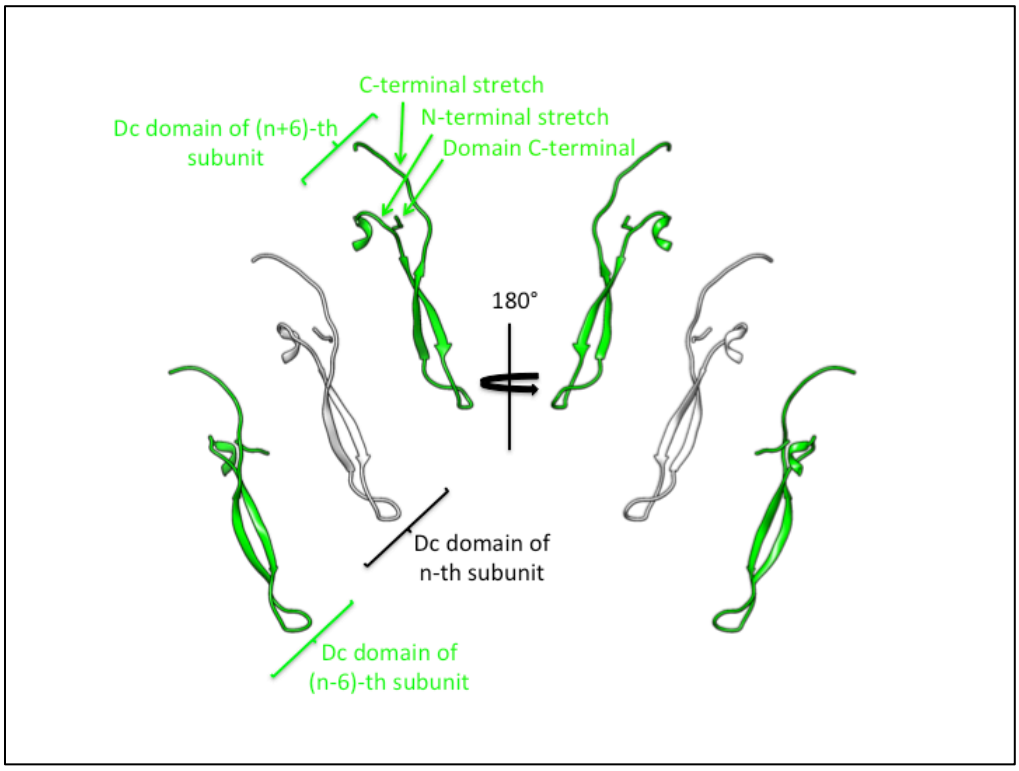


Figure 44 Dc doman interactions along 6-start directions

D1 interactions

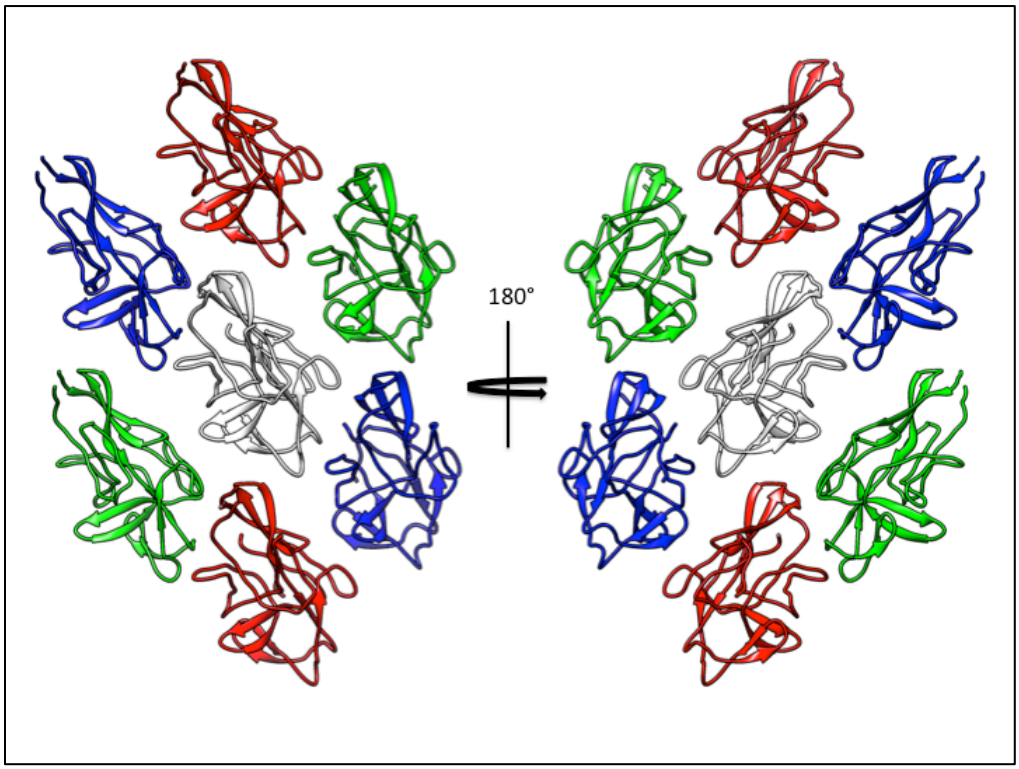


Figure 45 D1 interaction summary

Figure 45 showing the summary of interactions among D1 domains along the three main helical directions. Three domains of a subunit were selected to represent the interactions along the three main helical directions in the helix. The red D1 domains showing interactions along the 11-start helical direction. The green D1 domains indicating interactions along the 6-start direction. The blue D1 domains represent interactions along the –5 start direction. The white D1 domain lying in the intersection of three main helical directions represent that this Dc domain participates in every helical directions.

Along -5start:

-5 start helical direction exhibits the similar interaction profile as the 6-start or the 11-start direction. The interactions are limited in a few side sidechain ineractions between the n-th and (n+5)th subunit's D1 domains (figure 46). The interactions are presumably formed by the hairpin's loop region which is composed of mainly polar amino acids. For details please read (Samatey et.al 2004) paper.

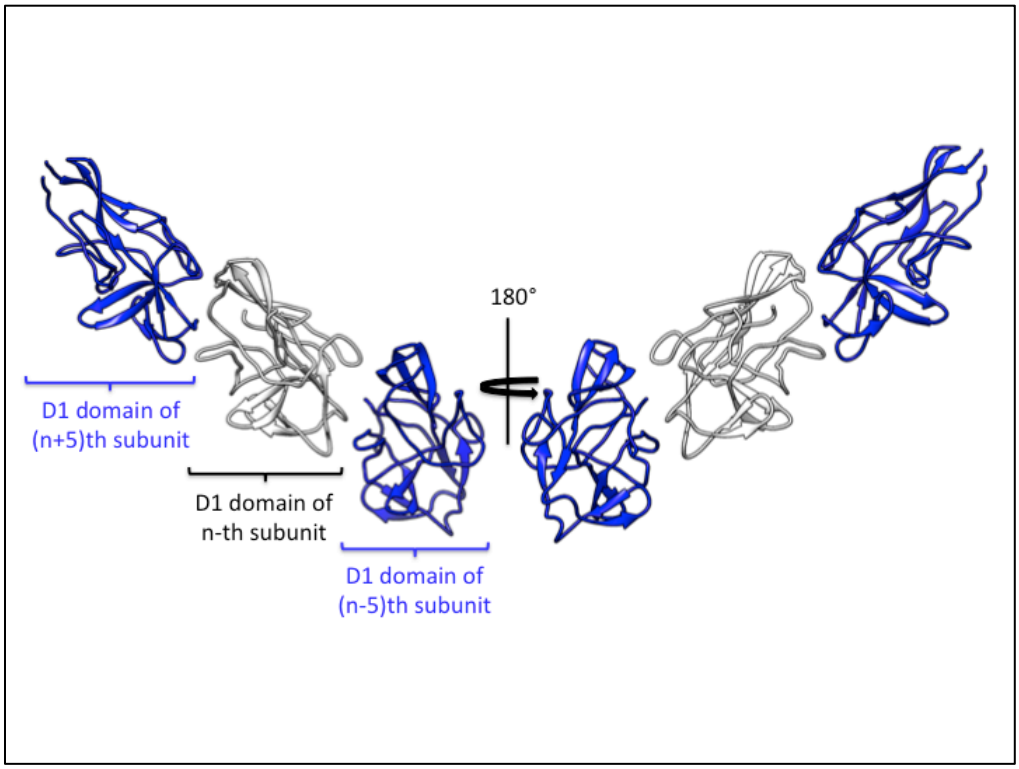


Figure 46 D1 domain interactions along -5-start direction

Along 11-start:

Average distance between the neighboring interfaces between the (n)th and the (n+11)th subunit is approximately 5,5Å (figure 47).

Nevertheless the significant space presumably there are some type of weak Van der Waals interction between the n-th and the (n+11)th subunit's D1 domain presumably by hydrogen bonding via water monolayer. For details see (Samatey et al., 2004).

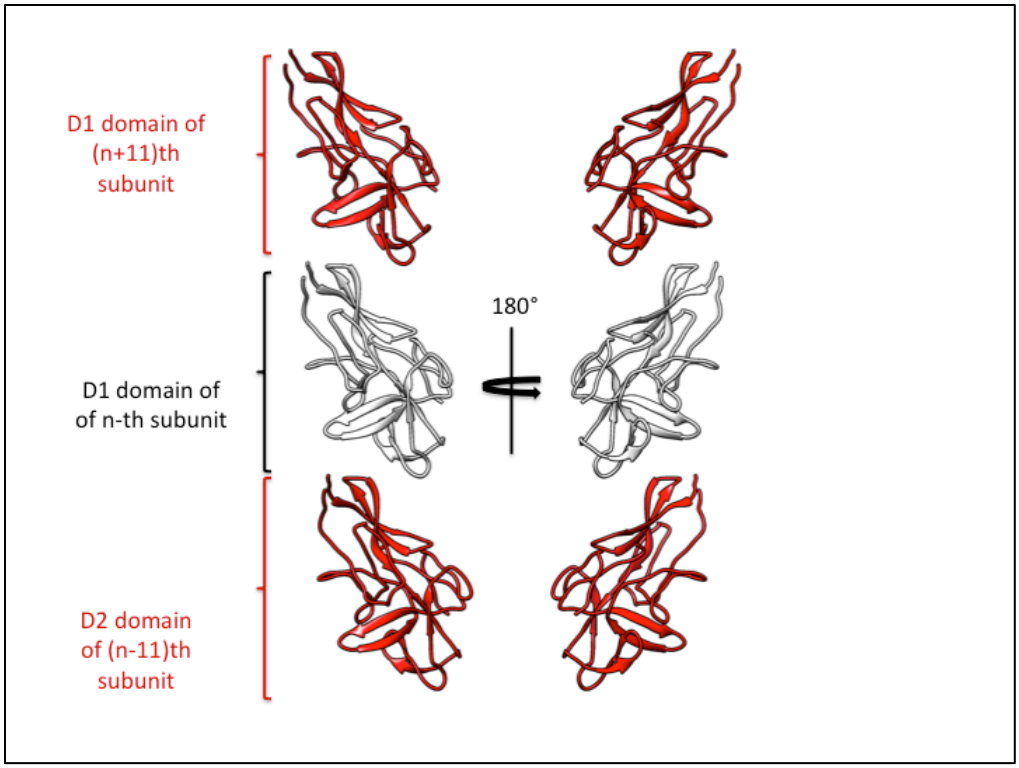


Figure 47 D1 domain interactions along 11-start direction

Along 6-start:

The n-th subunit interaction surface constitutes mainly of polar amino acids, there were no discernable contacts between n-th and the (n+6)th subunit's D1 domain (figure 48). The polar-polar interactions that Samatey and colleagues reported (Samatey et al., 2004) were not discernable by any means. Very weak VDW contacts

can be seen between the n-th subunit D1 domain's C-terminal stretch and the (n+6)th subunit D1 domain's beta hairpin's turn. The weak domain interactions caused by the distance between two adjacent domains contribute to the flexibility of FlgE.

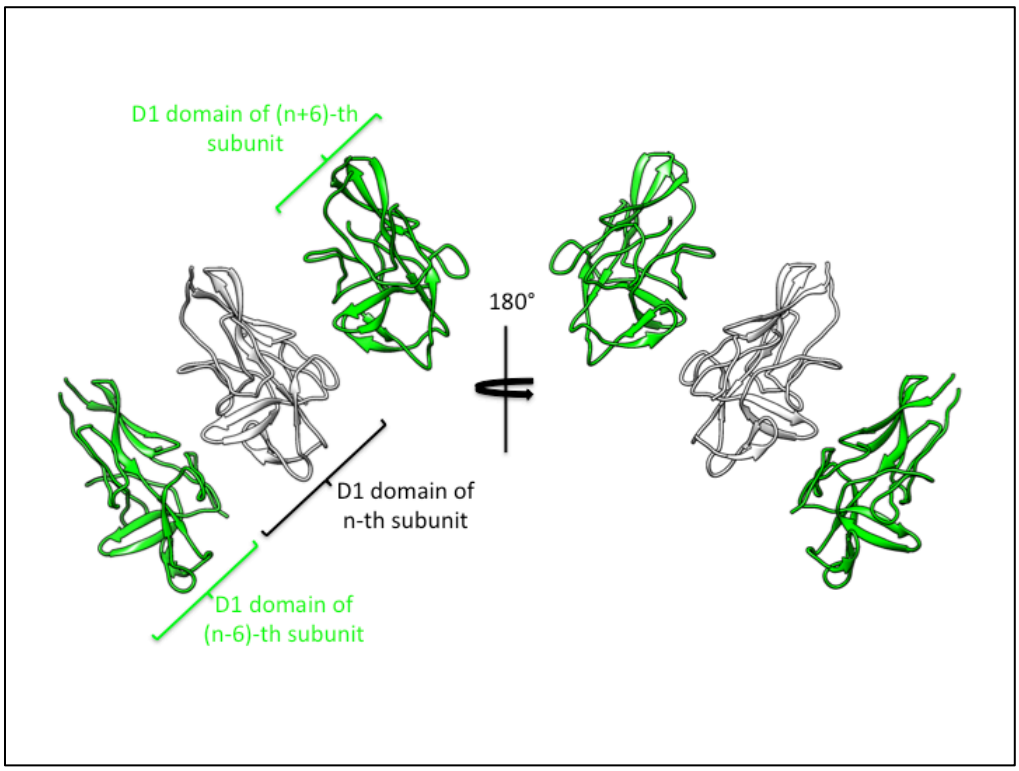


Figure 48 D1 interactions along 6-start direction

D2 interactions

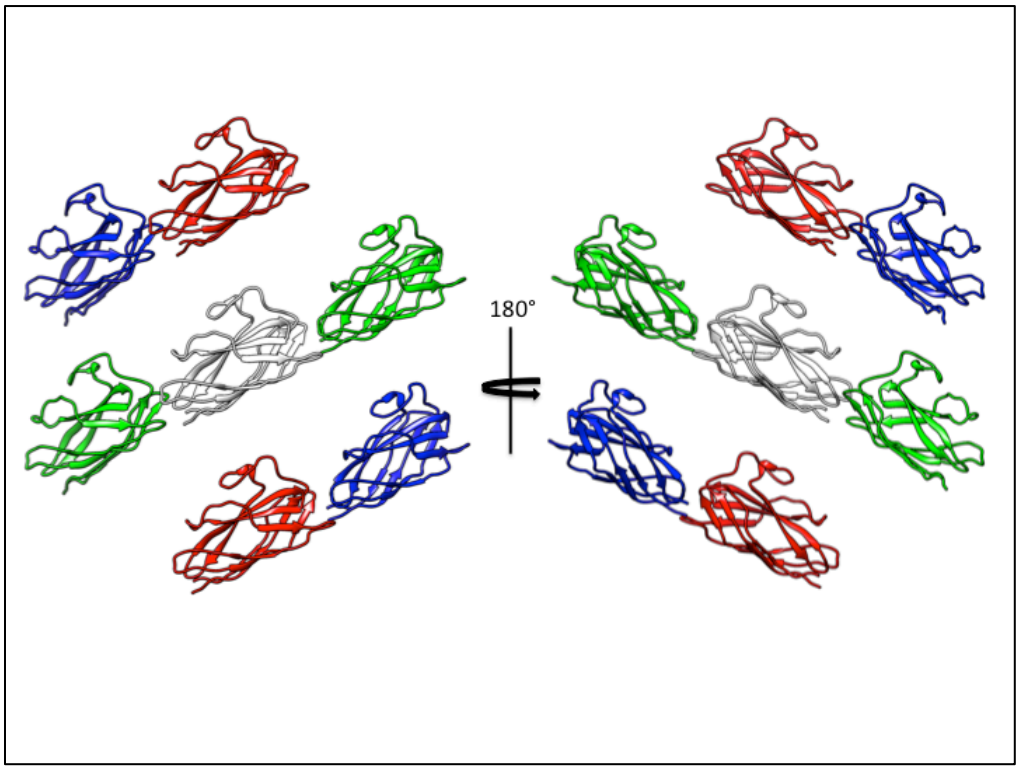


Figure 49 D2 interaction summary

Figure 49 showing the summary of interactions among D2 domains in the three main helical directions. Three subunits were selected to represent the interactions along the three main helical directions in the helix. The red D2 domains showing interactions along the 11-start helical direction. The green D2 domains indicating interactions along the 6-start direction. The blue D2 domains represent interactions along the –5 start direction. The white D1 domain lying in the intersection of three main helical directions represent that this Dc domain participates in every helical directions.

Along -5 start:

There are no discernable contacts along the -5-start helical direction since the average distance between the n-th and the (n+11)th

subunit is approximately 22 Å (figure 50). Sakai et, al.,2017 showed that D2 domain contributes to *Salmonella* FlgE to adopt the curved form of the hook. Pseudorevertant of the D2 domain mutant allowed bundle formation behind the cell indicating that flexibility was reestablished.

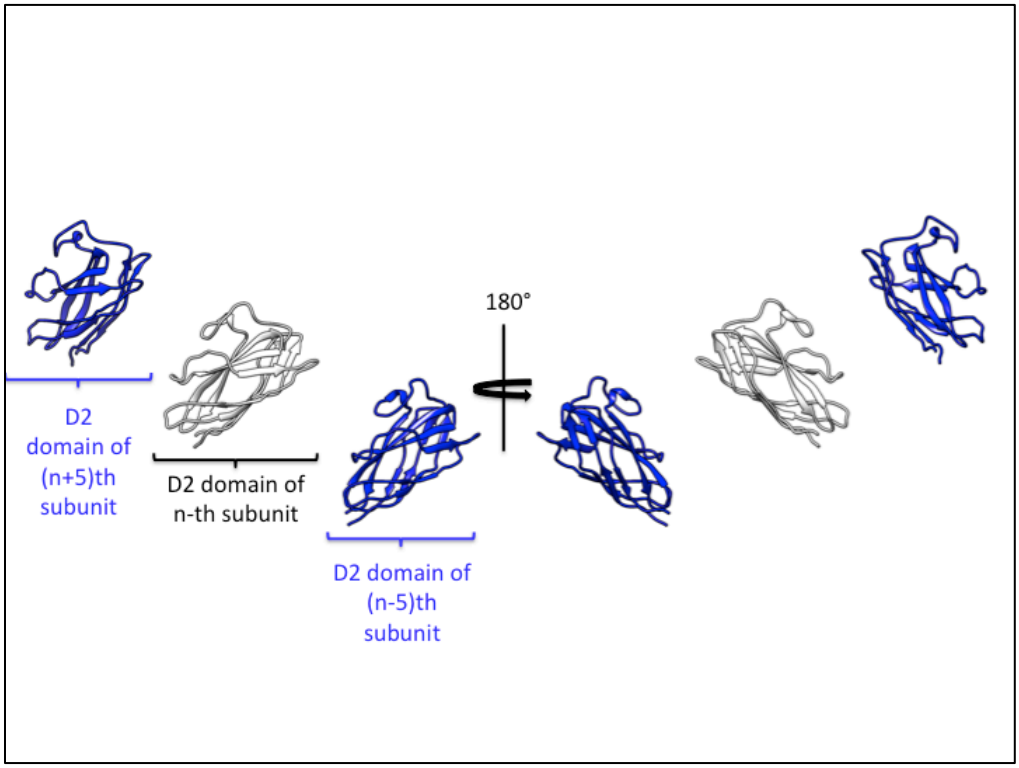


Figure 50 D2 domain interactions along -5-start direction

Along 11-start:

There are no contacts observable along the 11-start helical direction since the average distance between the n-th and the (n+11)th subunit is approximately 13 Å (figure 51).

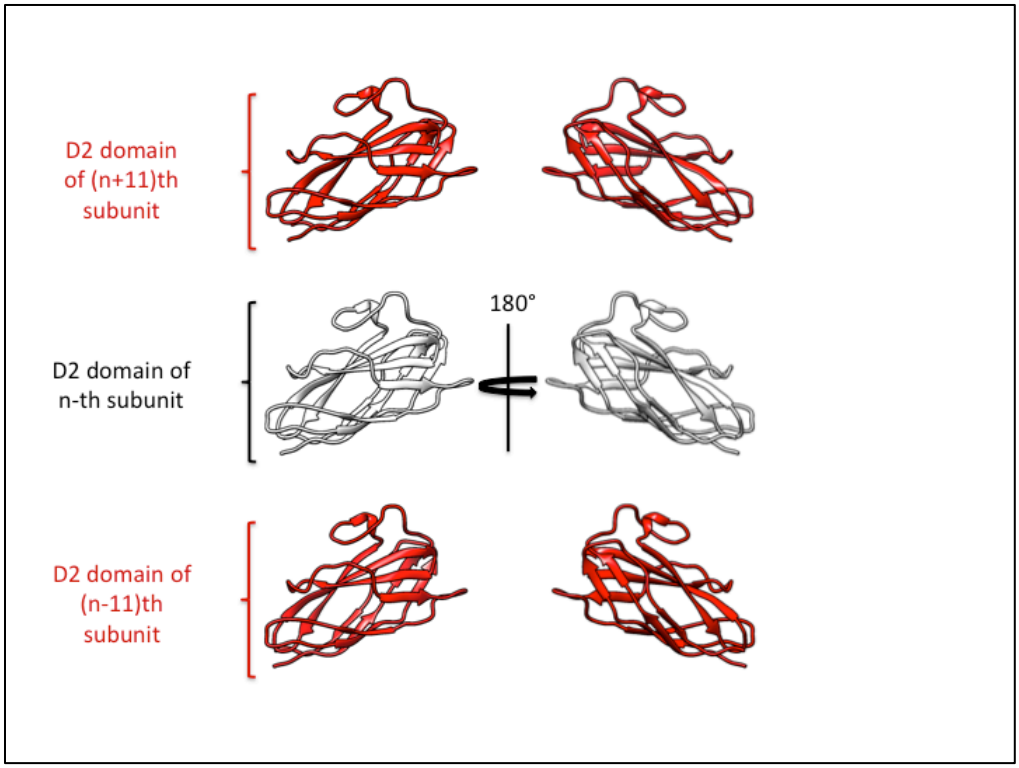


Figure 51 D2 interaction along 11-start direction

Along 6-start:

The n-th subunit shows affinity to be tightly connected to the the (n+6)th subunit's D2 domain's D2 domain along the 6-start helical direction (figure 52).

The n-th subunit's D2 domain's interacting interface is decorated with hydrophylic amino acids. The interacting residues in every case are located on the beta hairpin's turn region. The n-th subunit's D2 domain's beta hairpin's turn interacts with the (n+6)th subunit's beta hairpin's turn. The interacting turns accommodate hydrophobic amino acids as well. The turn regions are not ordered (but well characterized by dihedral angles phi,psi)

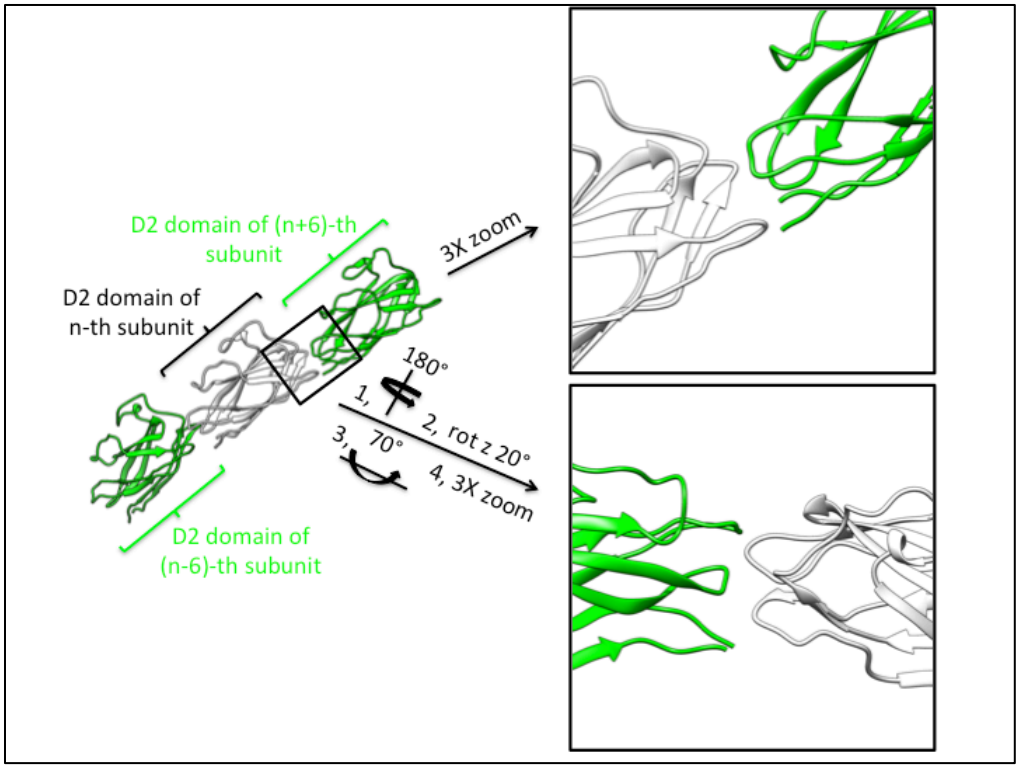


Figure 52 D2 interaction along 6-start direction

DO-DC interaction analysis

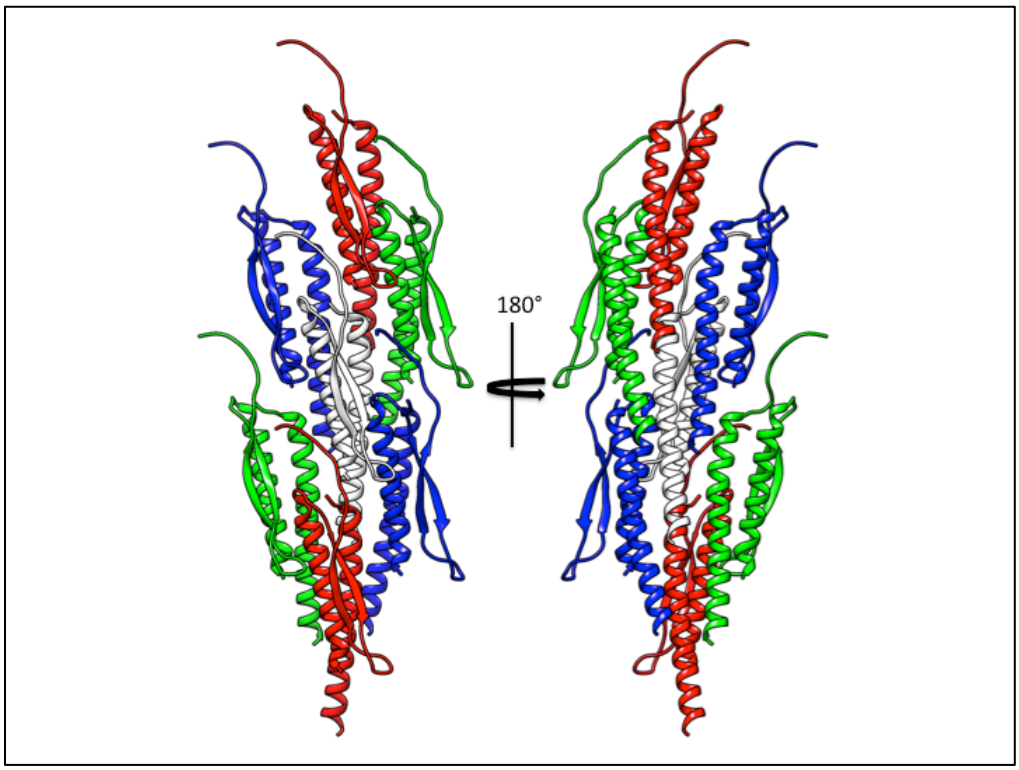


Figure 53 D0-Dc interaction summary

Figure 53 showing the summary of interactions between D0 and Dc domains in the three main helical directions. D0 and Dc domains of three subunits were selected to represent the interactions along the three main helical directions in the helix. The red D0, Dc domains showing interactions along the 11-start helical direction. The green D0, Dc domains indicating interactions along the 6-start direction. The blue D0, Dc domains represent interactions along the –5-start direction. The white D0, Dc domains lying in the intersection of three main helical directions represent that these D0, Dc domains participate in every helical directions.

Along the 11-start:

The n-th subunit's Dc domain is in close contact with the same subunit's D0 domain's N-terminal helix (figure 54). The interaction is created by hydrophobic amino acids. The interaction site spans the the length of the Dc domain, showing intimate connection between the N-terminal helix of the D0 domain and the Dc domain. Since as above discribed the helices are capable of some extension, compression along the axis of the hook it is plausible to suppose that the interaction sites between N-terminal helix of the D0 domain and Dc domain can rearrange during compression or extension.

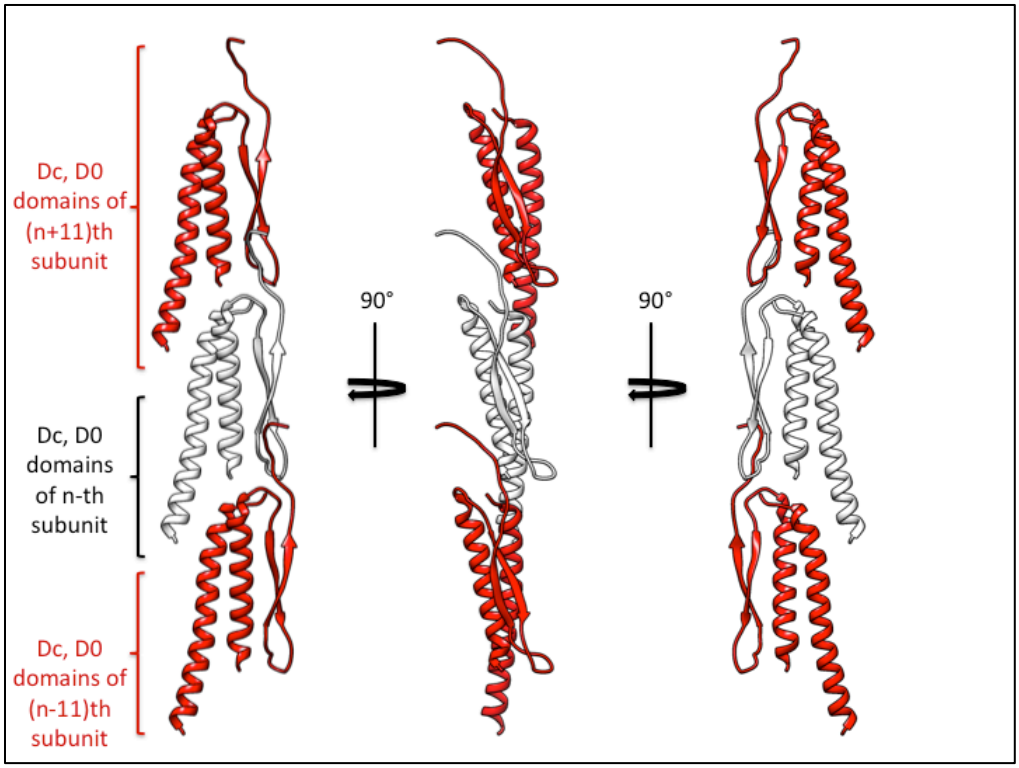


Figure 54 D0-Dc interactions along 11-start

Along the 6-start:

The n-th subunit Dc domain interacts with the same subunit's D0 domain's N-terminal (figure 55). There are no observable interactions between the adjacent (n+6)th and the (n-6)th subunit's DO domains and n-th subunit's Dc domain.

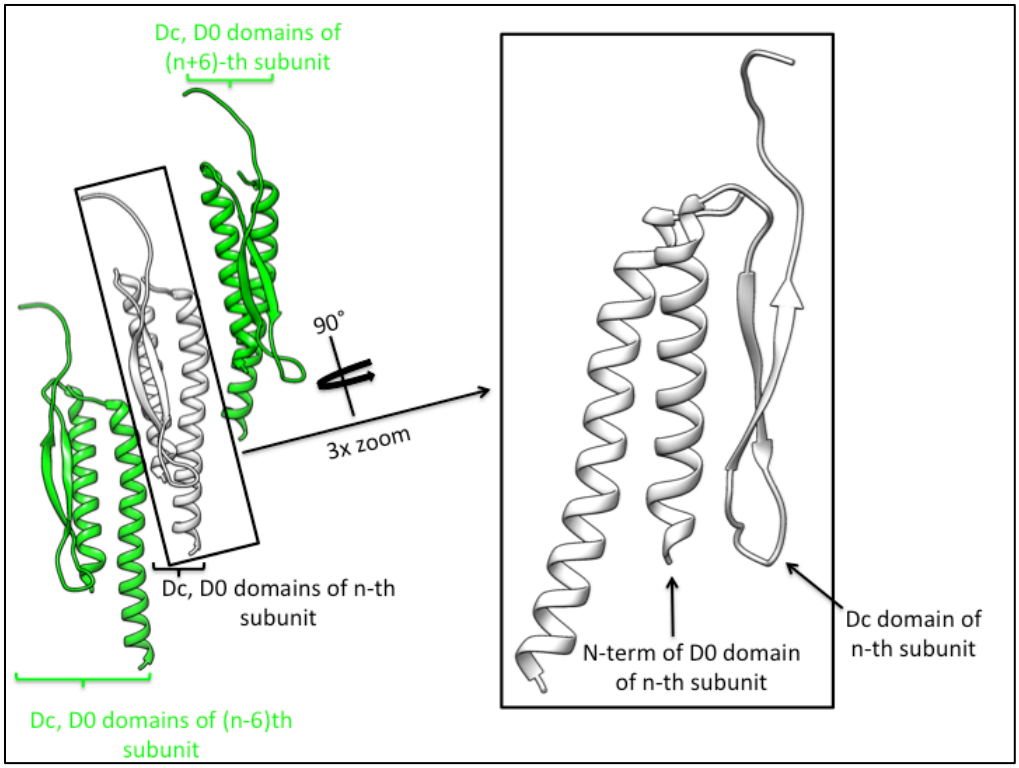


Figure 55 D0-Dc interactions along 6-start direction

Along -5 start:

The n-th subunit's Dc domain is interacting with the same subunit's D0 domain's N-terminal alpha helix (figure 56). The interface covered with both polar and hydrophobic amino acids, thereby allowing stronger and weaker interactions between the two subunits. The n-th subunit's Dc domain's beta hairpin's loop interacts with the (n-5)th subunit's D0 domain's N-terminal helix.

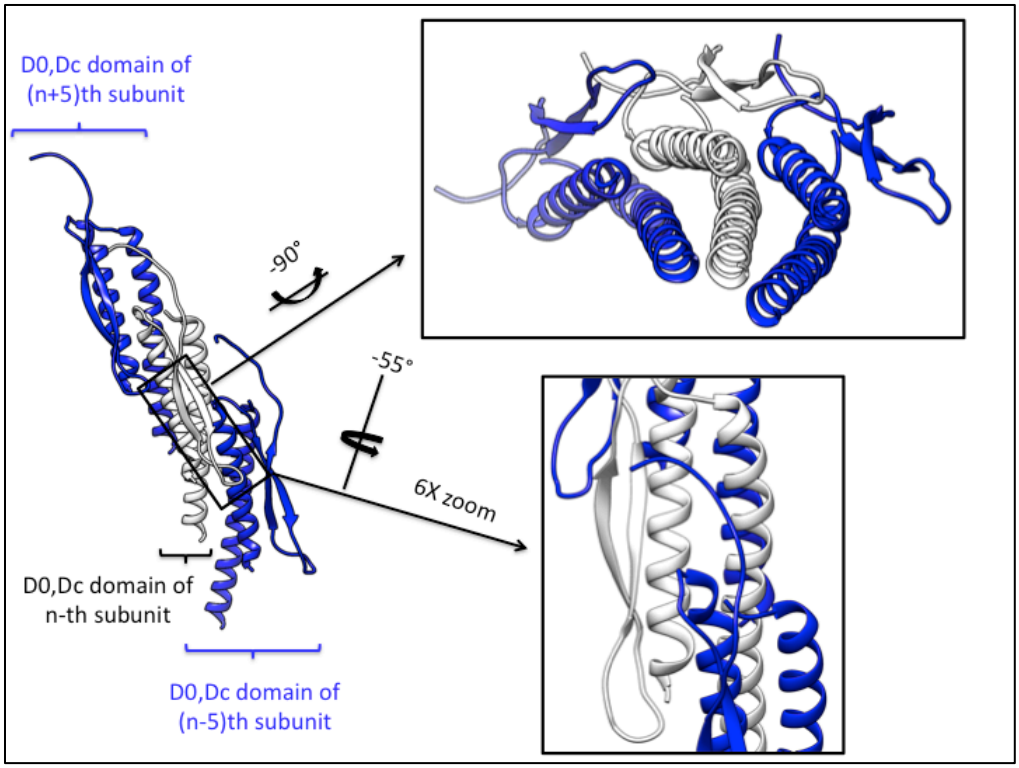


Figure 56 D0-Dc interactions along -5-start direction

DC-D1 interaction

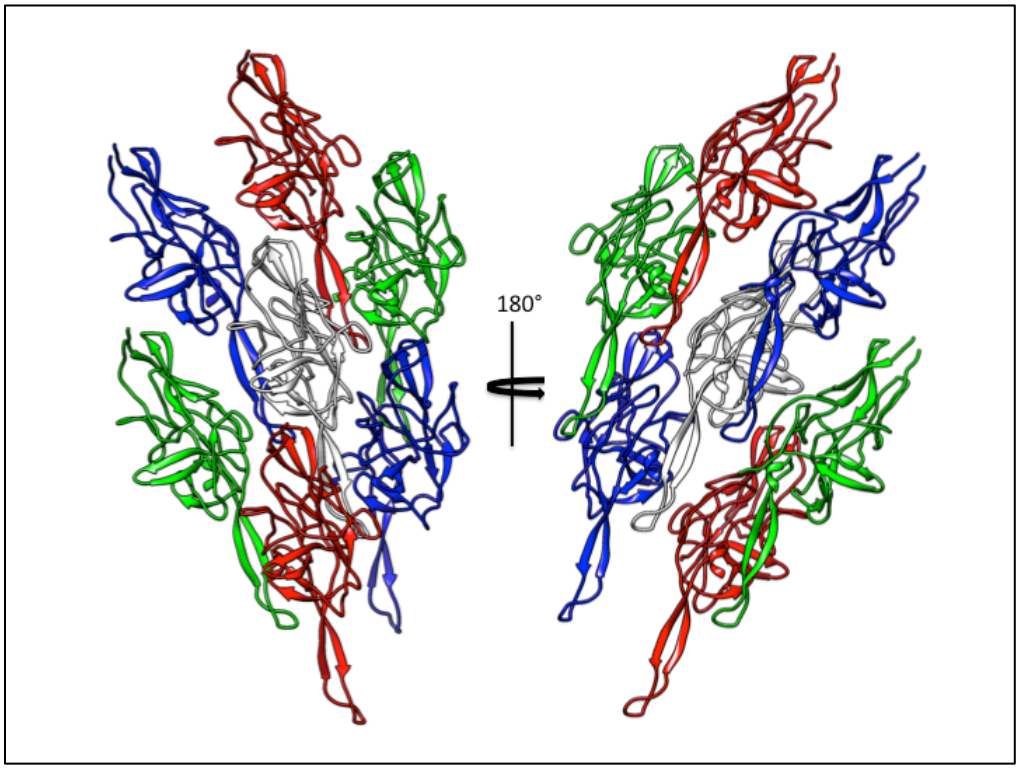


Figure 57 Dc-D1 interaction summary

Figure 57 showing the summary of interactions between Dc and D1 domains in the three main helical directions. Dc and D1 domains of three subunits were selected to represent the interactions along the three main helical directions in the helix. The red Dc, D1 domains showing interactions along the 11-start helical direction. The green Dc, D1 domains indicating interactions along the 6-start direction. The blue Dc, D1 domains represent interactions along the –5-start direction. The white Dc, D1 domains lying in the intersection of three main helical directions represent that these Dc, D1 domains participate in every helical directions.

11-start direction:

The n-th subunit's Dc domain's C-terminal stretch is interacting with the n-th subunit's D1 domain (firgure 58). Since the stretch is composed of polar and hydrophobic amino acids, both polar-polar, VDW interactions can form.
Since D1 domains have rather week interactions between them in every helical directions it is plausible to suppose that D1 and Dc interaction interface is for stabilization.

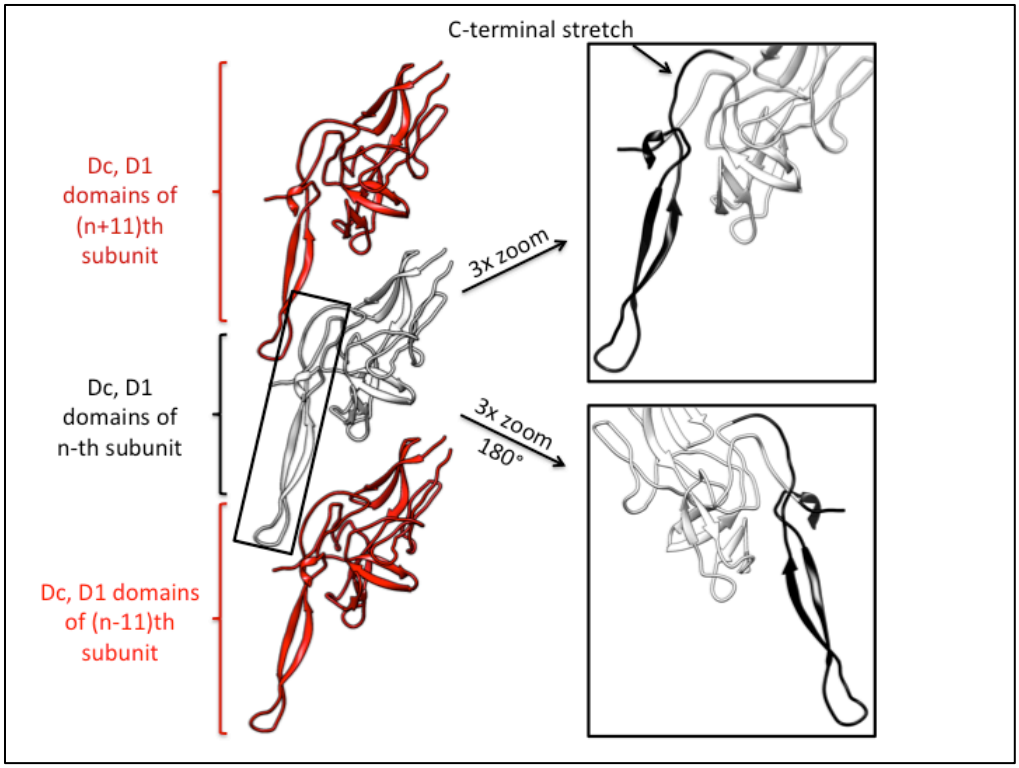


Figure 58 Dc-D1 interactions along 11-start direction

The Dc domain of the n-th subunit is colored in black.

6-start direction:

Same interaction between the Dc-D1 domains as in the case of 11-start (figure 58, figure 59).

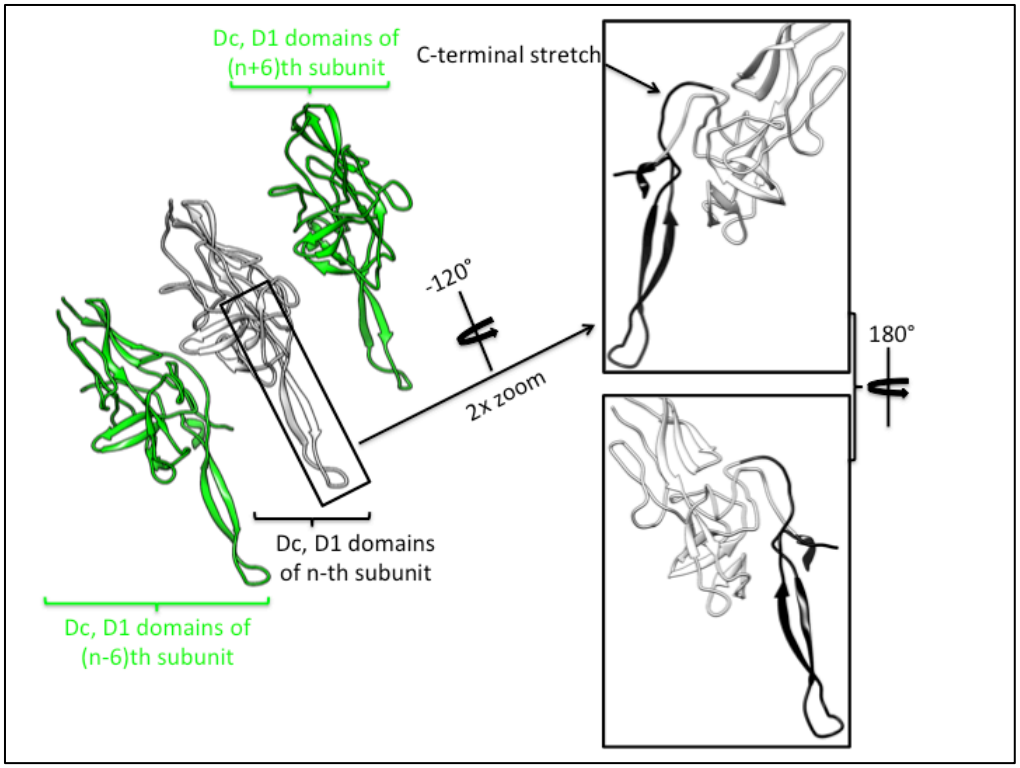


Figure 59 Dc-D1 interactions along 6-start direction

The Dc domain of the n-th subunit is colored in black.

-5 start direction:

The n-th subunit's Dc domain's C-terminal stretch is interacting with the n-th subunit's D1 domain. This set of interaction is mainly characterized by polar amino acids. Furthermore the n-th subunit's Dc domain's beta strands and the loop are impacted between the n-th and the (n-5)th subunit's D1 domains connecting the two subunits together. The spatial arrangement of the the n-th subunit's Dc domain's loop suggest that it is also capable of interacting the (n-5)th subunit's Dc domain presumably via weak non covalent interactions. Latter observations can also underpin the stabilizing effect of the Dc domain.

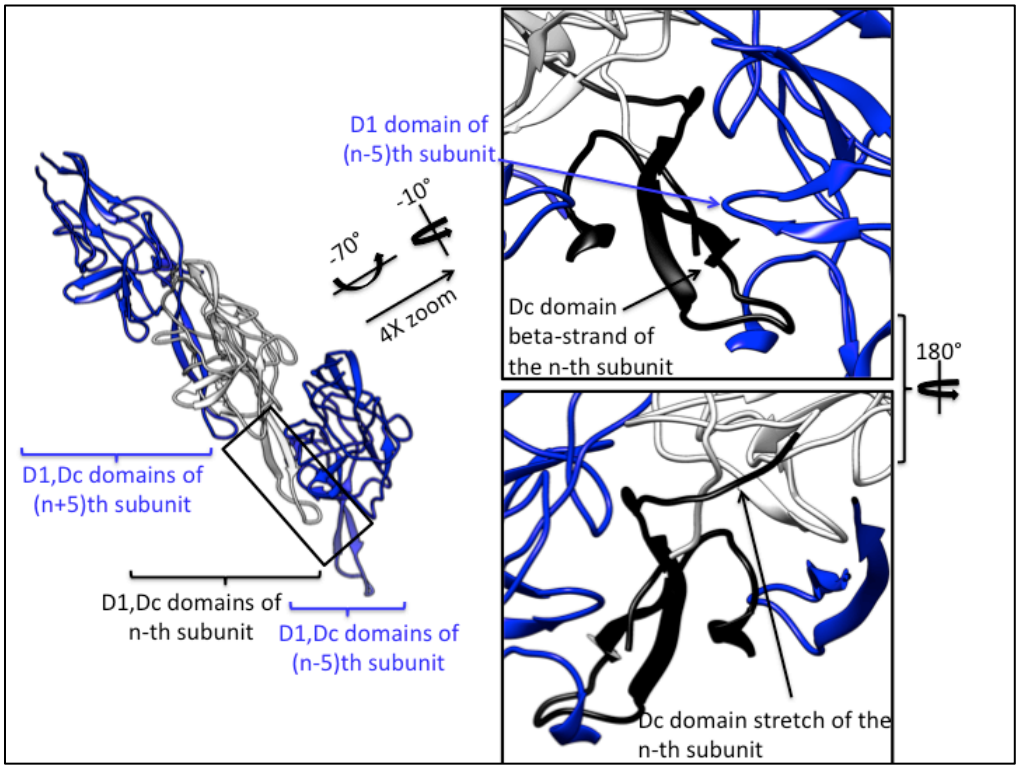


Figure 60 Dc-D1 interactions along -5-start direction

The Dc domain of the n-th subunit is colored in black.

D1-D2 interctions

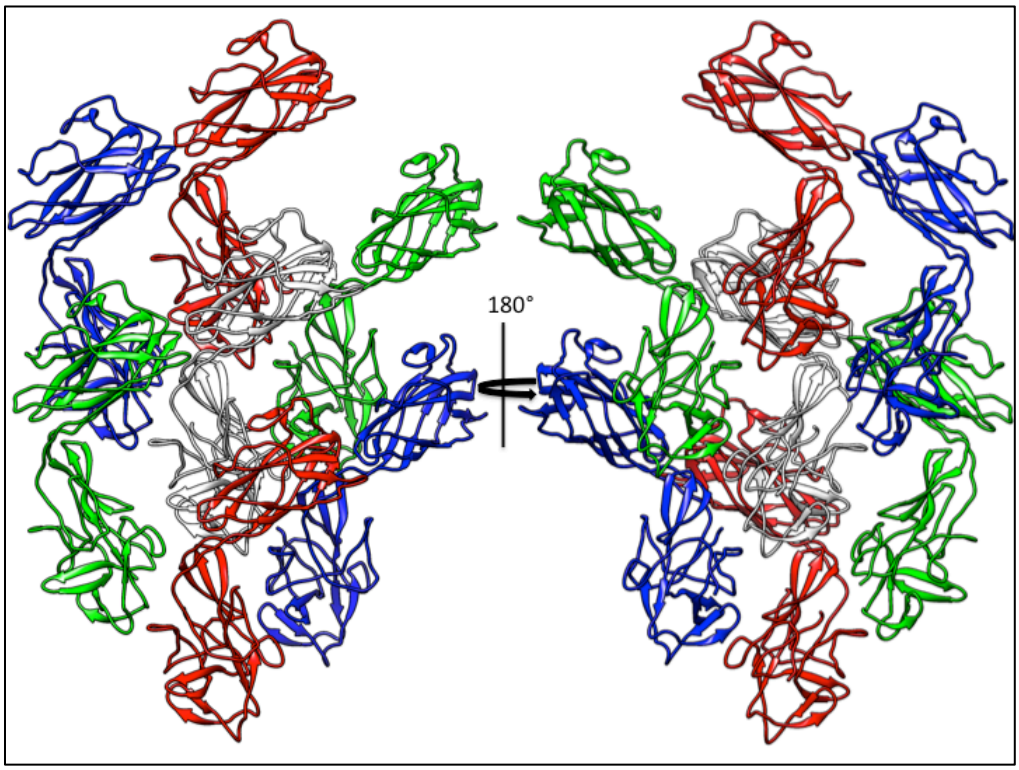


Figure 61 D1-D2 interactions summary

Figure 61 showing the summary of interactions between D1 and D2 domains in the three main helical directions. D1 and D2 domains of three subunits were selected to represent the interactions along the three main helical directions in the helix. The red D1, D2 domains showing interactions along the 11-start helical direction. The green D1, D2 domains indicating interactions along the 6-start direction. The blue D1, D2 domains represent interactions along the –5-start direction. The white D1, D2 domains lying in the intersection of three main helical directions represent that these D1, D2 domains participate in every helical directions (figure 61).

11-start direction:

The n-th subunit's D1 domain's interacting with the (n-11)th subunit's D2 domain via a special sequence (116-135) that resembles to a triangle hence the name triangular loop. The triangular loop and D2 domain interaction is responsible for the supercoiling of the hook (Samatey et al., 2004) (figure 62).

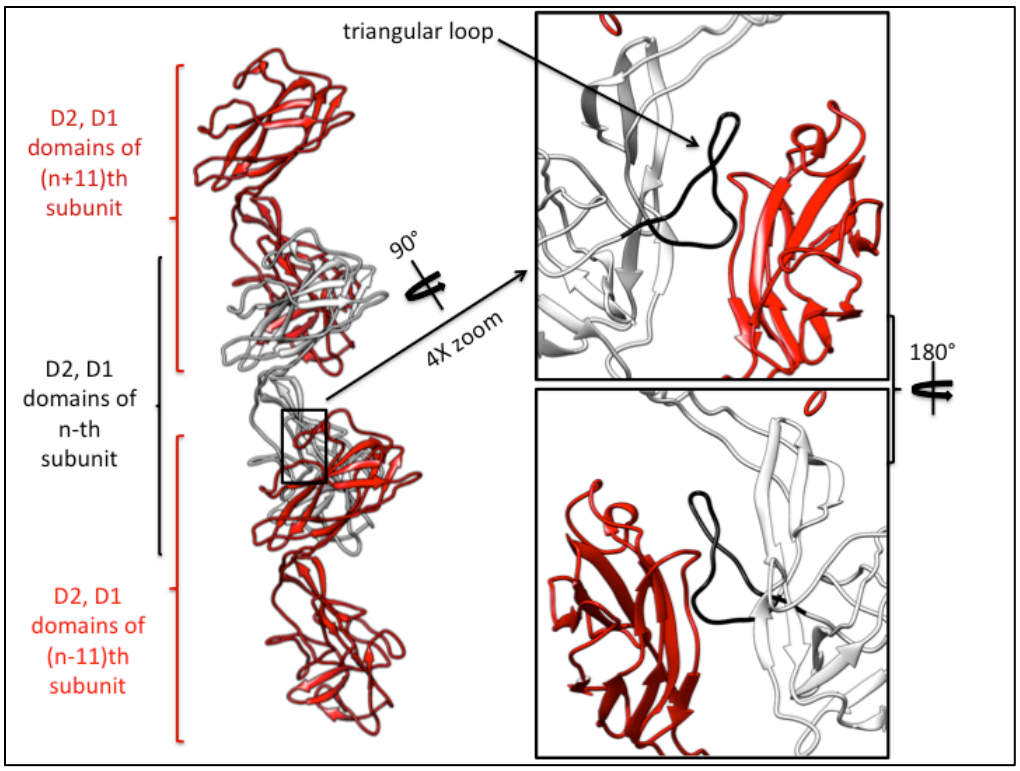


Figure 62 D1-D2 domain interactions along 11-start direction

The triangular loop (116-135) of the n-th subunit D1 domain is colored in black.

6-start direction:

The n-th subunit's D1 domain is "wedged" in the connection interface between the n-th subunit's D2 domain and (n-6)th subunit's D2 domain, showing interaction affinity with both the n-th and the (n-6)th D2 domain (figure 63).

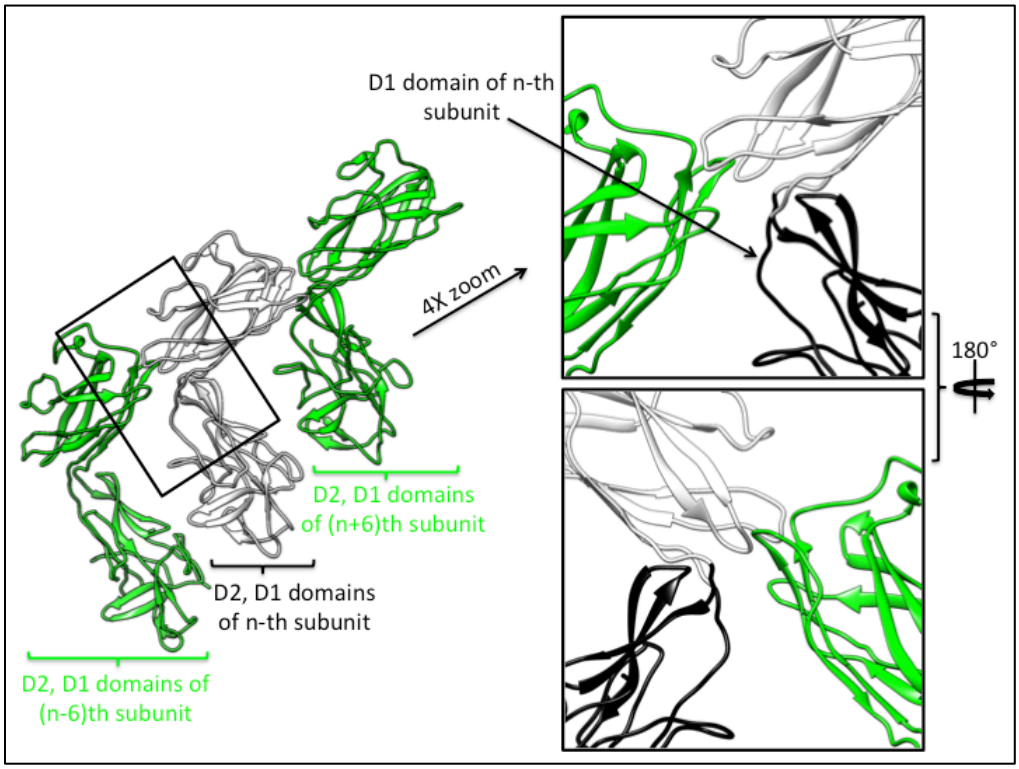


Figure 63 D1-D2 domain interactions along 6-start

The D1 domain of the n-th subunit is colored in black.

-5-start direction:

Along this helical direction the n-th subunit's D1 domain's triangular loop (116-135) interacts with the n-th D1 domain's C-terminal stretch and the N-terminal stretch that connects the D1 domain with the D2 domain (figure 64).

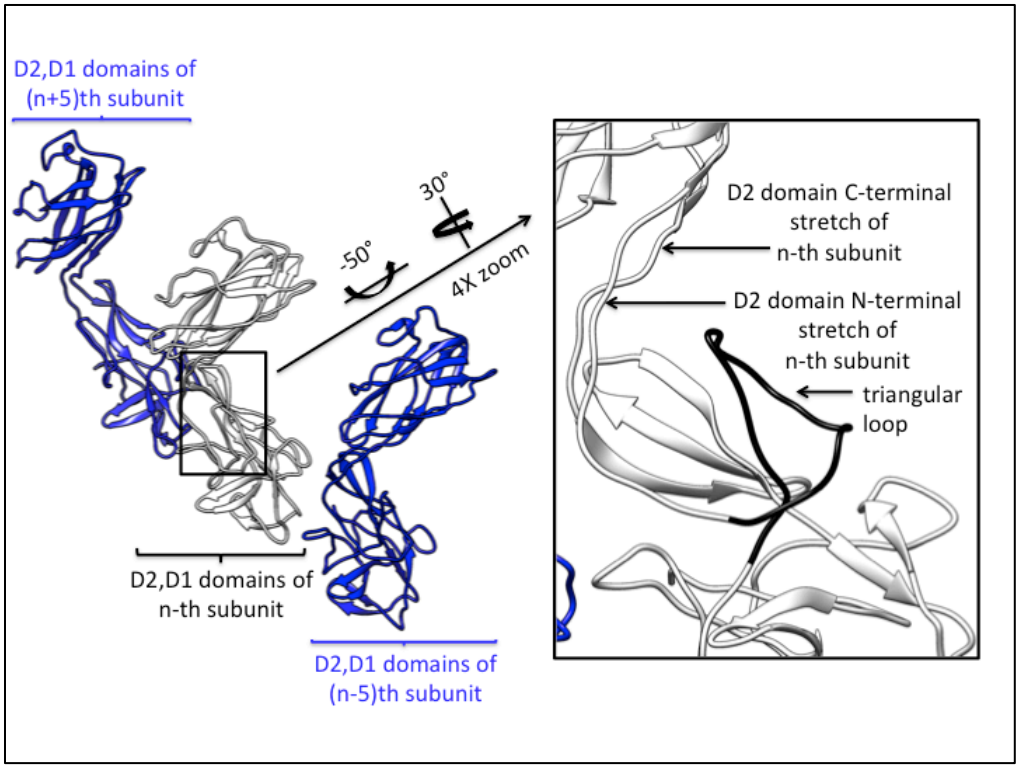


Figure 64 D1-D2 domain interactions along -5-start direction

The triangular loop (116-135) of the n-th subunit D1 domain is colored in black.

D0, Dc and D1 interactions

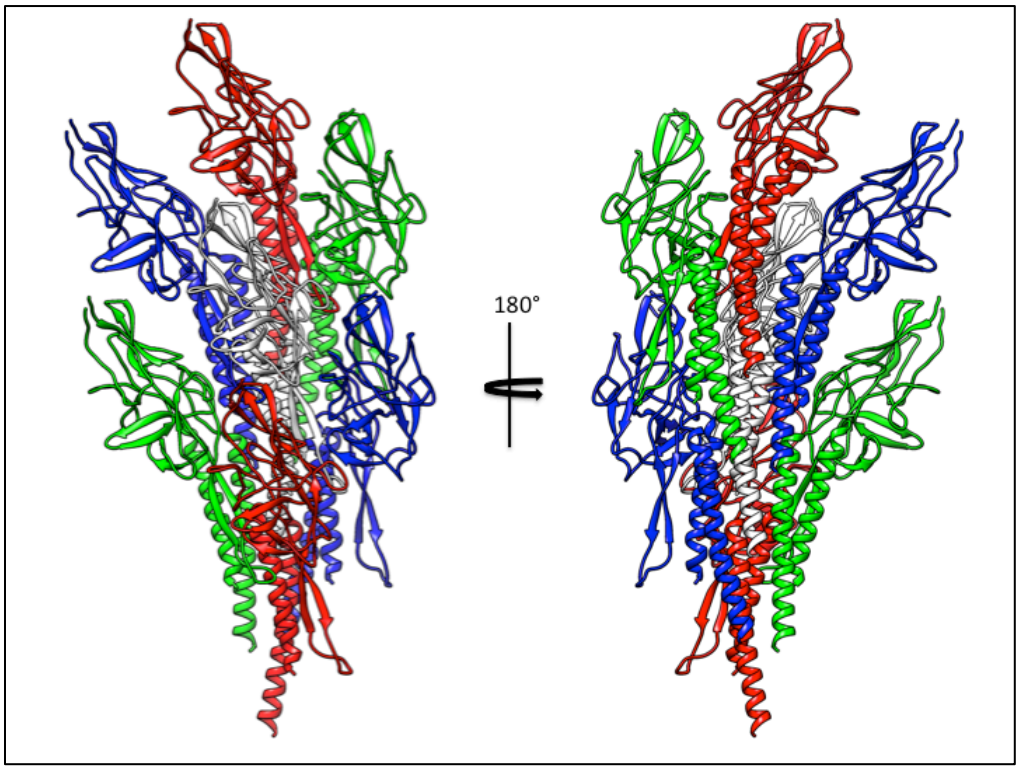


Figure 65 D0-Dc-D1 interactions summary

Figure 65 showing the summary of interactions between D0, Dc and D1 domains in the three main helical directions. D0, Dc and D1 domains of three subunits were selected to represent the interactions along the three main helical directions in the helix. The red D0, Dc and D1 domains showing interactions along the 11-start helical direction. The green D0, Dc, D1 domains indicating interactions along the 6-start direction. The blue D0, Dc and D1 domains represent interactions along the –5-start direction. The white D0, Dc and D1 domains lying in the intersection of three main helical directions represent that these D1, D2 domains participate in every helical directions

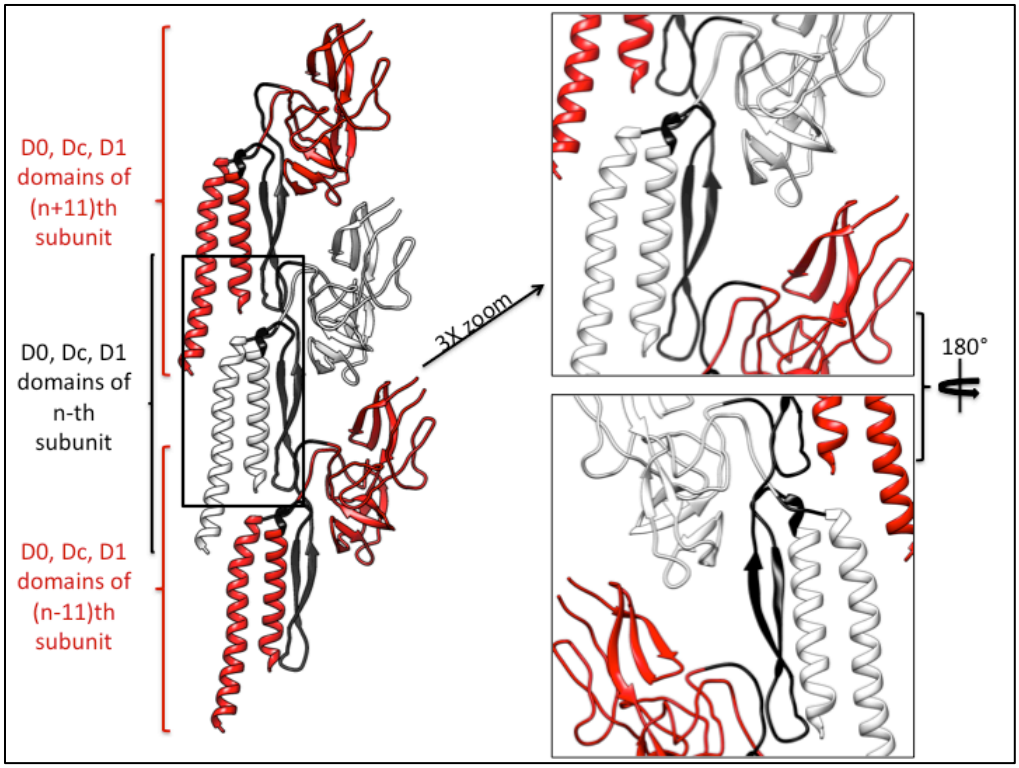


Figure 66 D0-Dc and D1 domain interactions along 11-start direction

The Dc domains are colored in black.

The N-terminal helix of the n-th subunit is interacting with the beta sheet of the Dc domain of the n-th subunit. The C-terminal stretch of the Dc domain of the n-th subunit is interacting with D1 domain of the n-th subunit and also the turn of the beta hairpin of the (n+11)th subunit. As previously discussed interaction between the C-terminal helix of the n-th subunit and the C-terminal terminal helix of the (n+11)th subunit can also be observed along the 11-start direction. Because of the flexible C and N-terminal hinges of the Dc domain it is possible that the Dc domain of the n-th subunit is capable to rearrange the interaction interface between the N-terminal helix of the n-th subunit and the D1 domain of the (n-11)th subunit (figure 66).

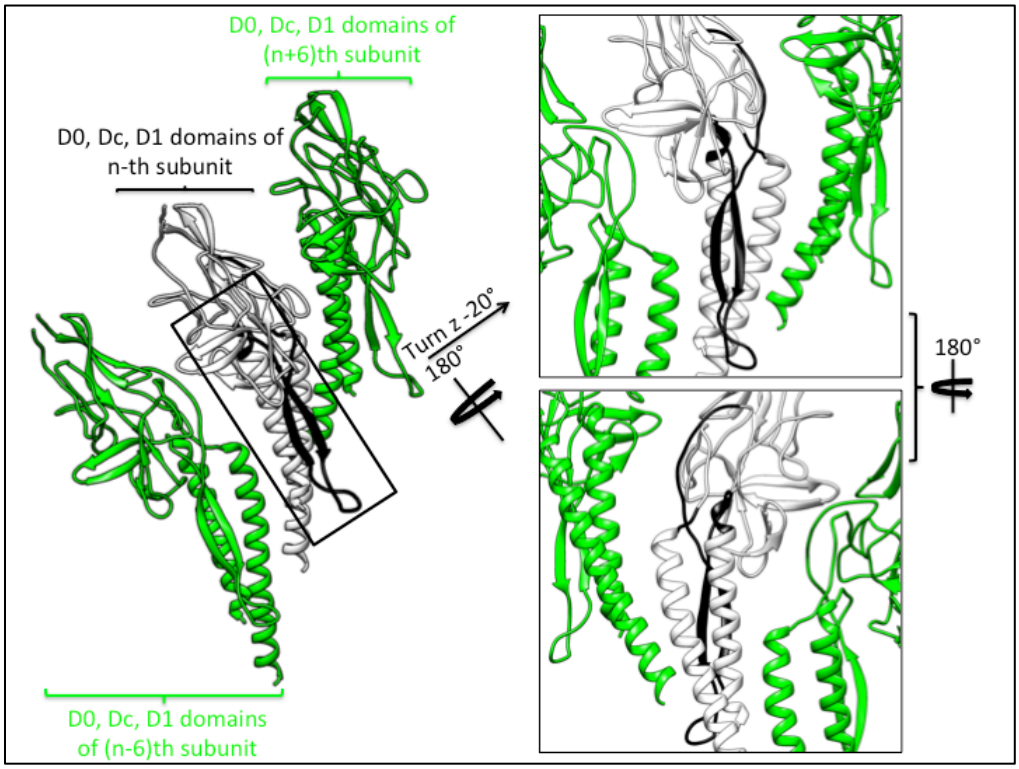


Figure 67 D0-Dc and D1 domain interactions along 6-start direction

The Dc domain of the n-th subunit is colored in black.

Same interactions can be observed as previously discussed (figure 59, figure 55).

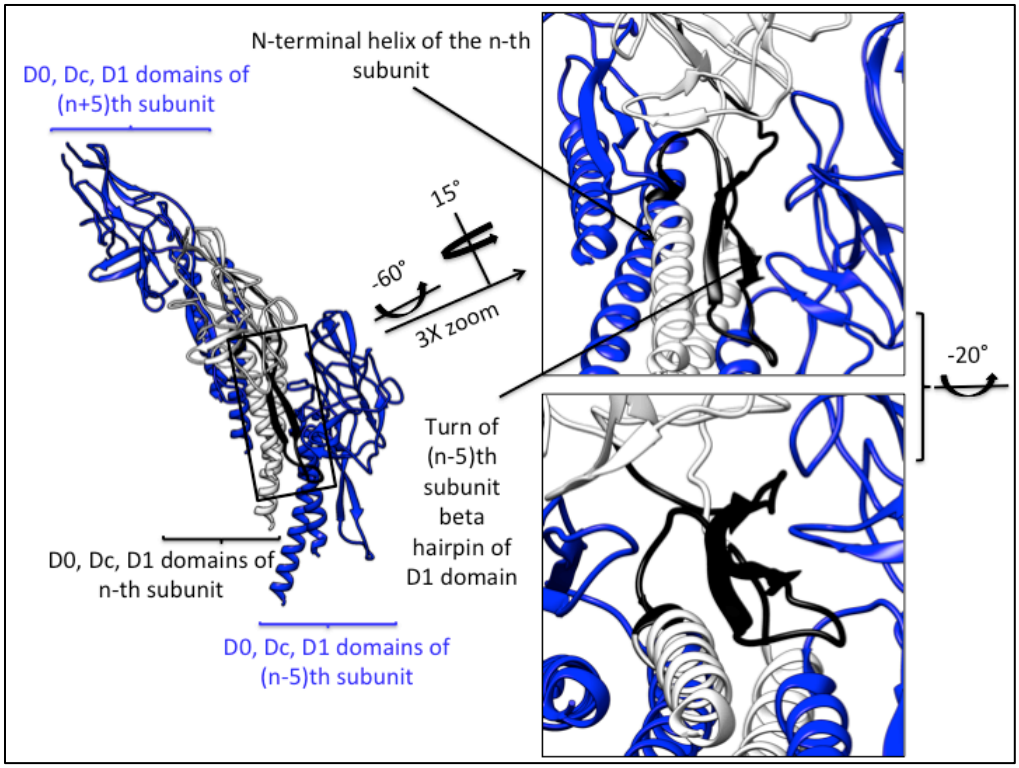


Figure 68 D0-Dc and D1 domain interactions along -5-start direction

The Dc domain of the n-th subunit is colored in black.

The N-terminal terminal helix of the n-th subunit is interacting with beta sheet of the Dc domain of the n-th subunit. The antiparallel beta sheet of the Dc domain of the n-th subunit has an interaction site with the turn of the beta hairpin of the D1 domain of the (n-5)th subunit. The Dc domain of n-th subunit maintains a position between the D1 domain of the (n-5)th subunit and the D0 domain of the (n-5)th subunit. Because the flexible stretches of the C and N-terminal of Dc domain it is possible that the Dc domain of the n-th subunit is capable to rearrange the bonding interface between the D0 domain of the (n-5)th subunit and the D1 domain of the (n-5)th subunit acting as a switch. The flexible turn of beta hairpin of the Dc domain is another structural feature the supports the hypothesis of the Dc domain acting as a switch (figure 68).

Conclusions

The bacterial flagellum axial portion can be functionally classified into 3 groups. The rod protein functions a drive shaft, the hook protein functions as a universal joint and the flagellin functions as propeller. The rod protein embedded in the peptidoglycan and the outer membrane of the cell. It functions as a drive shaft between the motor protein and the hook. It is responsible for stable 300Hz rotation speed in *Salmonella Typhimurium* (which rotation speed is comparable with a formula 1 engine (Toyota RVX-06 V8 engine is nearly 317 Hz)), without any lubricant. The rod is composed of two sets of proteins. FlgF, FlgB, FlgC constitute the proximal rod and FlgG is the single building block of the distal rod.

The axial proteins of the bacterial flagella share structural similarity. One similarity aspect is due to the identified heptad repeats of hydrophobic amino acid residues in the terminal segments. These repeats are sequence characteristics that can indicate the susceptibility of the sequence to acquire coiled coil structure (Homma, DeRosier, & Macnab, 1990). Furthermore the axial proteins arranged as concentric tubes with radially increasing diameters. *Salmonella Typhimurium* FlgE and FlgG shares relatively high 37% percent sequence identity, the overall RMSD is 1.807Å. Moreover FlgG is directly connected to FLgE without any adaptor protein, FlgG and FlgE helical parameters are identical. Rod and hook segments averaged Fourier transforms indicate almost identical layer line positions, which also implies high structural resemblance (Fujii et al., 2017). Nevertheless their function and mechanical properties show a

diverged profile. The hook is a highly flexible protein, whereas the rod is rigid.

For the first time the inner core of the *Salmonella Typhimurium* was revealed at nearly atomic resolution. The main chain was clearly traceable, amino acids with bulky side chains were also discernable. The helices are arranged in coiled coil structure. The longer C-terminal helix is approximately 55Å and the shorter N-terminal is nearly 39Å. The longer N-terminal helix is composed of 25 amino acids and the longer C-terminal is 36 amino acids. Between the n-th subunit's N-terminal helix and the n-11 th subunit's C-terminal helix a gap of approximately 6Å can be found. This gap allows flexibility along the 11-start, called the gap compression/extension mechanism that was discussed in detail in (Furuta et al., 2007). Pairwise sequence alignment showed 45,61%, 48.39% sequence identity between *Salmonella Typhimurium* FlgE D0 domain and *Salmonella Typhimurium* FlgG (distal rod protein), *Campylobacter Jejuni* D0 domain respectively. A measure of 3D similarity, the RMSD also indicated nearly identical 3D structural features in the case of D0 domains.

Salmonella D0 domain's C-terminal helix side that forms the lumen's inner surface is decorated with hydrophilic amino acids in order to facilitate the protein transport through the approximately 19Å diameter channel. D0 domain of the FlgE showed non covalent interactions along all helical directions as previously was shown. These interactions can be classified into the sliding interactions (Furuta et al., 2007). The main role of the inner core from mechanistic perspective is to ensure the stability of the structure by connecting the subunits together, making extensive interactions

between the protofilaments (the set of subunits along the 11-start helical directions). In frame deletions of the N and C-terminals of D0 domain prohibited hook assembly (Moriya, Minamino, Hughes, Macnab, & Namba, 2006). Due to sliding motions the rearrangements of the bonding between the interacting helices is a plausible speculation.

The structure of the Dc domain was also revealed for the first time in *Salmonella Typhimurium.* Dc domain forms a mesh like structure which is loosely sewed along the 6-start helical direction. This domain connects the innermost D0 with D1 domain. The Dc domain exhibits extensive interaction along the -5-start helical direction and the 11-start helical direction but there are no discernable, direct interactions between Dc domains along the 6-start helical direction. Therefore supposedly Dc domain has the role of stabilizing the structure along the -5-start and 11-start helical directions and providing flexibility along the 6-start helical direction along which no interactions were observable. Along the 11-start and the -5-start helical directions the Dc domain of the n-th subunit resides in close vicinity of the N-terminal helix of the D0 domain and the D1 domain of the (n-5)th and the (n-11)-th subunit, exhibiting interactions, mainly between hydrophobic amino acids (figure 68,66). Along the -5-start helical direction the n-th subunit's Dc domain can interact with the (n-5)th subunit's D0's N-terminal helix and the it's D1 domain as well (figure 56, 60, 68,66). The helices of the D0 domain can interact along all helical lines, as previously discussed. However D1 domains have weak interactions along the three helical directions. In fact before this study it was postulated that the only stabilizing interaction along the 11-start is the interaction between the

triangular loop of D1 domain and the D2 domain (figure 62). The near atomic resolution structure showed interactions between the Dc domains along the 11-start directions (figure 43), (figure 58). Thereby, the Dc domain exerting stabilization effect along the 11-start and the -5-start helical directions (figure 60).

The interactions also underpin the hypothesis that Dc domain serves as stabilizer of the structure. Furthermore Dc connects the inner, mechanically stable D0 core with outer relatively loosely connected domains via its C-terminal and N-terminal stretches that can serve as hinges. It plays the role of a transition domain between the inner and the outer domains. It was recently shown by (Sakai et.,al, 2017) mutational assays that D1 domain is responsible for flexibility of *Salmonella Typhimurium*. Dc domain and D1 domain interaction along -5-start, 11-start (figure 60, 62) can result in a reduced flexibility by limiting the conformational space along of the D1 domain along -5-start and 11-start helical directions, furthermore because the Dc and D1 domain interaction can induce change in the conformation of the D1 domain which will further induce changing in the conformation of the D2 domain, which is responsible for polymorphic changing resulting in a straight morphology .

It was shown in sequential in-frame deletion mutation assays of flgE ΔfliK that 30-49 of Dc domain is responsible for FlgE export, 40-59 residues for filament assembly and 50-59 for hook morphology ((Moriya et al., 2006). Furthermore the deletion of residues of 50-59 in the Dc domain sequence produced straight polyhooks. Residues 30-49 are responsible are the export signal that can be recognized by the type 3 secretion system.

An interesting survey by (Hiraoka et al., 2017) showed that when a short specific FlgG sequence (GSS) was inserted between the position Phe-42 and Ala-43 in the Dc domain of *Salmonella Typhimurium* it was found that the inserted 18 residue (YQTIRQPGAQSSEQTTLP) sequence reduced the flexibility of the hook, moreover it made the hook straight and rigid.
The aforementioned two works by (Hiraoka et al., 2017) and (Moriya et al., 2006) also pointed out the stabilization role of the Dc domain and the impact on the morphology of FlgE. In both cases mutations were capable to transform curved, bent hooks to straight and rigid (in the case of the FlgG specific sequence insertion). Rigidity is a stronger criterium than straightness. Rigid is also straight but straight is not always rigid.

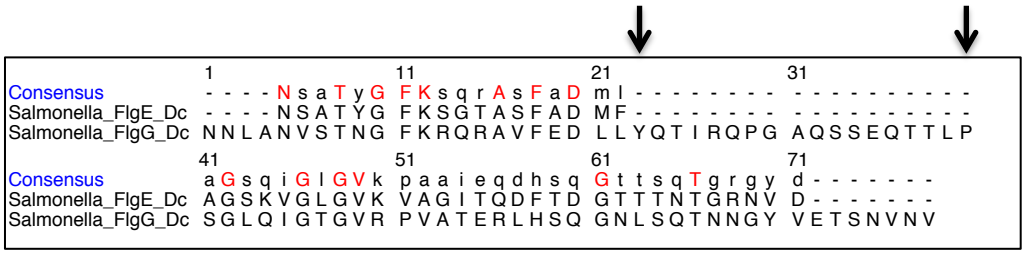


Figure 69 FlgE Dc domain and FlgG Dc domain pairwise protein sequence alignment

Pairwise sequence alignment between FlgE Dc and FlgG Dc domain. Arrows indicate the region in the Dc domain where FlgG (distal rod) specific sequence (YQTIRQPGAQSSEQTTLP) was inserted.

Figure 69 shows interactions between the C beta sheet (50-59) of the Dc domain of the n-th subunit facing the N-terminal helix of the D0 domain of the n-th subunit. The antiparallel beta sheet (34-41) is in the vicinity of the (n-5)th D1 domain. If in an in-frame deletion experiment the (50-59) sequence is removed the antiparallel beta sheet (34-41) in the vicinity of the (n-5)th D1 can attach to the

(n-5)th D1 domain reducing conformational space of it. The similar scenario can occur along 11-start direction. In that case the beta sheet and the turn of the beta hairpin of the n-th subunit can possibly interact with the flexible N and C-terminal hinges of the (n-5)th D1 domain.

Insertion of FlgG specific sequence into the turn of the beta hairpin of the Dc domain (figure 70) between the 42nd and 43rd amino acid will elongate the sequence. Presumably this causes a more intimate interaction along -5-start and 11-start helical directions between the n-th Dc domain and the D1 domain of the (n-5)th and the (n-11)th subunit.

Previous study by (Fujii et al., 2017) analysed the similarities and differences between the rod and the hook. The study claimed that the the difference between rigidity of the FlgG and flexibility of FlgE arouse from the longer N-terminal helix of FlgG, the extra density of the Dc domain of the FlgG that can connect adjacent domains along 11-start. And 7° D1 domain orientation difference in FlgG compared to FlgE D1 domain in *Salmonella Typhimurium* that allows the FlgG D1 domain to interact along 11-start more extensively than the FlgE resulting in a rigid and a flexible structure respectively. By inserting the GSS sequence orientation seems to induce changes that favours excessive D1 domain interactions along 11-start helical direction in the FlgE, hence diminishing the flexibility. (Hiraoka et al., 2017) experiments showed that elongation of the Dc domain by the insertion of 18 amino residue (GSS) into the sequence of Dc domain of *Salmonella Typhimurium* between 42nd and 43rd amino acid is enough to produce straight and rigid *Salmonella Typhimurium* mutant.

Therefore the claim was that Dc domain of FlgE has an important role in stabilizing the FlgE structure and polymorphic transition, namely to change its form from bent to straight.

It was known from previous work of (Fujii et al., 2009) that D0 domains can exert interactions along the three main helical directions therefore the D0 domain is responsible for mechanical stability. The near atomic resolution structure of *Salmonella Typhimurium* FlgE revealed new stabilizing interactions the -5 start and the 11-start helical directions. These interactions as well as previously shown interactions are essential to maintain the universal joint function of the hook. It is also important to stress that the contributions of the domains that show no discernable interactions along a helical directions directions based on the inspection of the currently analyzed three dimensional structure are equally important contributors to the universal joint mechanism allowing flexibility.

As it was shown earlier the Dc domain is capable of intimate interaction along the -5-start and along 11-atart as well. These interactions exert a stabilizing effect and manifests itself in the Recent studies by Fujii et al., showed that the otherwise supercoiled hook can be turned to straight phenotype by the deletion of the D2 domain. D2, triangular loop, D2+triangular loop deficient mutant analysis was carried out by Sakai et al. by motility assay they showed that the mutant hook without D2 domain was still capable of swimming in soft agar but with a reduced speed compared to the wild type. Triangular loop and D2+triangular loop deficient mutants were non motile. In every cases pseudorevertant mutations restored partially the bundle formation so, therefore in all mutants case

reduced motility was observed. Hence neither the D2 domain deletion nor the D1 triangular loop deletion reduced the bending flexibility. The conclusion was that even though D2 and the triangular loop deficient mutants are straight but they are not rigid. Flexibility was restored, so they were capable to form flagellar bundles behind the cell body. Bundle formation is a prerequisite criterium for the flexibility of the hook. If the hook can not function as a universal joint, impaired bundle formation can occur hence reduced swimming motility can be observed. In order to form the bundle behind the cell the hook of flagella have to be bent. Other cases where the hook is rigid the single flagellum will protrude radially outwards from the cell, causing the interference with the concerted association of flagella to form the bundle, resulting in reduced motility speed. This latter case was observed in the aboved mention study by (Hiraoka et al., 2017). This mutation made the hook rigid and straight thus preventing it from bundle formation behind the cell. The molecular background of flexibility was discussed from a structural point of view based on the near atomic resolution structure of the FlgE.

Summary

A FliK deficient, non flagellated mutant strain of *Salmonella Typhimurium, SJW880* was isolated as previously described by (Aizawa, Dean, Jones, Macnab, & Yamaguchi, 1985). Modifications and optimizations in the original purification protocol were made to make the sample suitable for cryo-EM analysis. Three dimensional reconstruction using negative stained images was carried out by homemade program using SPIDER software (Shaikh et al., 2009).

Optimized samples were investigated by JEM-3200 FSC JEOL microscope mounted with K2 direct electron detector. Grid preparation conditions were further optimized to obtain high quality images. High quality images were analyzed with different software packages: SPRING (Desfosses, Ciuffa, Gutsche, & Sachse, 2014), RELION2 (Scheres, 2012b), EMAN2 (Tang et al., 2007) and in-silico conditions to extract high resolution information from the cryo-electron micrographs. Final 3D reconstruction was obtained by RELION2. Homology model was generated by MODELLER (Webb & Sali, 2014). Model fitting into the density map was carried out by ROSETTA3 (Wang et al., 2016).

In this work I utilized the state of art hardware and software tools to obtain nearly atomic resolution information about key protein in flagellum. The 4.1 Å cryo-EM density map allowed me to obtain molecular level insights into the *Salmonella Typhimurium* hook protein. This study revealed the full atomic structure of the FlgE protein and allowed to study of the flexible property answering questions that couldn't be answered before the cryo-EM structural analysis entered to the resolution revolution era. It was shown by the detailed interaction analysis along the helical directions of the full atomic structure that the relatively flexible Dc domain is responsible for stabilization of the D1 domain which resulted in a reduced conformational space hence reduced flexibility. It was also shown that Dc domain can act as a polymorphic switch between curved and straight forms of the hook. The interaction of the Dc domain with the D1 domain can cause orientation change in D1 domain which further induce change in the orientation of the D2 domain via the triangular loop, resulting in a straight morphology.

References

Aizawa, S. I., Dean, G. E., Jones, C. J., Macnab, R. M., & Yamaguchi, S. (1985). Purification and characterization of the flagellar hook-basal body complex of Salmonella typhimurium. *Journal of Bacteriology*, *161*(3), 836–849. https://doi.org/10.1046/j.1365-2958.1997.3341714.x

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–10. https://doi.org/10.1016/S0022-2836(05)80360-2

Bai, X. chen, McMullan, G., & Scheres, S. H. W. (2015). How cryo-EM is revolutionizing structural biology. *Trends in Biochemical Sciences*, *40*(1), 49–57. https://doi.org/10.1016/j.tibs.2014.10.005

Bertani, G. (1951). Studies on Lysogenesis I. The mode of phageliberation by lysogenic eschericia coli. *Journal of Bacteriology*, *62*(3), 293–300. https://doi.org/citeulike-article-id:149214

Böttcher, B., Wynne, S. a, & Crowther, R. a. (1997). Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature*. https://doi.org/10.1038/386088a0

Calladine, C. R. (1976). Design requirements for the construction of bacterial flagella. *Journal of Theoretical Biology*, *57*(2), 469–489. https://doi.org/10.1016/0022-5193(76)90016-3

Calldine, C. R. (1978). Change of waveform in bacterial flagella: The role of mechanics at the molecular level. *Journal of Molecular Biology*, *118*(4), 457–479. https://doi.org/10.1016/0022-2836(78)90285-1

Desfosses, A., Ciuffa, R., Gutsche, I., & Sachse, C. (2014). SPRING - An image processing package for single-particle based helical reconstruction from electron cryomicrographs. *Journal of Structural Biology*, *185*(1), 15–26. https://doi.org/10.1016/j.jsb.2013.11.003

DiMaio, F., Song, Y., Li, X., Brunner, M. J., Xu, C., & Conticello, V. (2015). Atomic accuracy models from 4.5 {Å} cryo-electron microscopy data with density-guided local rebuilding. *Pdfs.Semanticscholar.Org*, *12*(4), 361–365. Retrieved from https://pdfs.semanticscholar.org/b84d/4e910b383ace29e04b639c8868827ca02160.pdf

DiMaio, F., Zhang, J., Chiu, W., & Baker, D. (2013). Cryo-EM model validation using independent map reconstructions. *Protein Science*, *22*(6), 865–868. https://doi.org/10.1002/pro.2267

Dubochet, J., Booy, F. P., Freeman, R., Jones, a V, & Walter, C. a. (1981). Low temperature electron microscopy. *Annual Review of Biophysics and Bioengineering*, *10*, 133–149. https://doi.org/10.1146/annurev.bb.10.060181.001025

Elad, N., Clare, D. K., Saibil, H. R., & Orlova, E. V. (2008). Detection and separation of heterogeneity in molecular complexes by statistical analysis of their two-dimensional projections. *Journal of Structural Biology*, *162*(1), 108–120. https://doi.org/10.1016/j.jsb.2007.11.007

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, *66*(4), 486–501. https://doi.org/10.1107/S0907444910007493

Fujii, T., Kato, T., Hiraoka, K. D., Miyata, T., Minamino, T., Chevance, F.

F. V., … Namba, K. (2017). Identical folds used for distinct mechanical functions of the bacterial flagellar rod and hook. *Nature Communications*, *8*(May 2016), 1–10. https://doi.org/10.1038/ncomms14276

Fujii, T., Kato, T., & Namba, K. (2009). Specific Arrangement of ??-Helical Coiled Coils in the Core Domain of the Bacterial Flagellar Hook for the Universal Joint Function. *Structure*, *17*(11), 1485–1493. https://doi.org/10.1016/j.str.2009.08.017

Furuta, T., Samatey, F. A., Matsunami, H., Imada, K., Namba, K., & Kitao, A. (2007). Gap compression/extension mechanism of bacterial flagellar hook as the molecular universal joint. *Journal of Structural Biology*, *157*(3), 481–490. https://doi.org/10.1016/j.jsb.2006.10.006

Gao, H., Valle, M., Ehrenberg, M., & Frank, J. (2004). Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset. *Journal of Structural Biology*, *147*(3), 283–290. https://doi.org/10.1016/j.jsb.2004.02.008

Hiraoka, K. D., Morimoto, Y. V., Inoue, Y., Fujii, T., Miyata, T., Makino, F., … Namba, K. (2017). Straight and rigid flagellar hook made by insertion of the FlgG specific sequence into FlgE. *Scientific Reports*, *7*(April), 46723. https://doi.org/10.1038/srep46723

Holley, W. (1975). © 1975 Nature Publishing Group. *Group*.

Homma, M., DeRosier, D. J., & Macnab, R. M. (1990). Flagellar hook and hook-associated proteins of Salmonella typhimurium and their relationship to other axial components of the flagellum. *Journal of Molecular Biology*, *213*(4), 819–832. https://doi.org/10.1016/S0022-2836(05)80266-9

Kamiya, R., Asakura, S., Wakabayashi, K., & Namba, K. (1979).

Transition of bacterial flagella from helical to straight forms with different subunit arrangements. *Journal of Molecular Biology*, *131*(4), 725–742. https://doi.org/10.1016/0022-2836(79)90199-2

Kato, S., Aizawa, S., & Asakura, S. (1982). Reconstruction in vitro of the flagellar polyhook from Salmonella. *Journal of Molecular Biology*, *161*(4), 551–560. https://doi.org/10.1016/0022-2836(82)90407-7

Kato, S., Okamoto, M., & Asakura, S. (1984). Polymorphic transition of the flagellar polyhook from Escherichia coli and Salmonella typhimurium. *Journal of Molecular Biology*, *173*(4), 463–476. https://doi.org/10.1016/0022-2836(84)90391-7

Kawagishu, I., Homma, M., Williams, A. W., & Macnab, R. M. (1996). Characterization of the flagellar hook length control protein FliK of Salmonella typhimurium and Escherichia coli. *Journal of Bacteriology*, *178*(10), 2954–2959.

Komeda, Y., Silverman, M., & Simon, M. (1978). Identification of the Structural Gene for the Hook Subunit Protein of Escherichia coli Flagella, *133*(1), 364–371.

Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., … Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods*, *10*(6), 584–590. https://doi.org/10.1038/nmeth.2472

Macnab, R. M. (2003). How bacteria assemble flagella. *Annual Review of Microbiology*, *57*(1), 77–100. https://doi.org/10.1146/annurev.micro.57.030502.090832

Macnab, R. M., & Koshland, D. E. (1972). The Gradient-Sensing

Mechanism in Bacterial Chemotaxis. *Proceedings of the National Academy of Sciences of the United States of America*, *69*(9), 2509–2512. https://doi.org/10.1073/pnas.69.9.2509

Magariyama, Y., Sugiyama, S., Muramoto, K., Maekawa, Y., Kawagishi, I., Imae, Y., & Kudo, S. (1994). Very fast flagellar rotation. *Nature*. https://doi.org/10.1038/371752b0

Matsunami, H., Barker, C. S., Yoon, Y.-H., Wolf, M., & Samatey, F. A. (2016). Complete structure of the bacterial flagellar hook reveals extensive set of stabilizing interactions. *Nature Communications*, *7*, 13425. https://doi.org/10.1038/ncomms13425

McMullan, G., Chen, S., Henderson, R., & Faruqi, A. R. (2009). Detective quantum efficiency of electron area detectors in electron microscopy. *Ultramicroscopy*, *109*(9), 1126–1143. https://doi.org/10.1016/j.ultramic.2009.04.002

Milazzo, A.-C., Leblanc, P., Duttweiler, F., Jin, L., Bouwer, J. C., Peltier, S., … Xuong, N.-H. (2005). Active pixel sensor array as a detector for electron microscopy. *Ultramicroscopy*, *104*(2), 152–159. https://doi.org/http://dx.doi.org/10.1016/j.ultramic.2005.03.006

Moriya, N., Minamino, T., Hughes, K. T., Macnab, R. M., & Namba, K. (2006). The Type III Flagellar Export Specificity Switch is Dependent on FliK Ruler and a Molecular Clock. *Journal of Molecular Biology*, *359*(2), 466–477. https://doi.org/10.1016/j.jmb.2006.03.025

Patterson-delafleld, J., Martinez, R. J., Angeles, L., & Yamaguchi, S. (1973). A New f l a Gene in S a l m o n e l l a t y p h i m u r i u m - - f l a R - - and Its Mutant Phenotype-Superhooks, *120*.

Penczek, P. a. (2010). Fundamentals of three-dimensional

recostruction from projections. *Methods in Enzymology*, *482*(10), 1–28. https://doi.org/10.1016/S0076-6879(10)82001-4.Fundamentals

Samatey, F. a, Matsunami, H., Imada, K., Nagashima, S., Shaikh, T. R., Thomas, D. R., … Namba, K. (2004). Structure of the bacterial flagellar hook and implication for the molecular universal joint mechanism. *Nature*, *431*(7012), 1062–1068. https://doi.org/10.1038/nature02997

Scheres, S. H. W. (2012a). A bayesian view on cryo-EM structure determination. *Journal of Molecular Biology*, *415*(2), 406–418. https://doi.org/10.1016/j.jmb.2011.11.010

Scheres, S. H. W. (2012b). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, *180*(3), 519–530. https://doi.org/10.1016/j.jsb.2012.09.006

Scheres, S. H. W., & Chen, S. (2012). Prevention of overfitting in cryo-EM structure determination. *Nature Methods*, *9*(9), 853–854. https://doi.org/10.1038/nmeth.2115

Scheres, S. H. W., Núñez-Ramírez, R., Gómez-Llorente, Y., San Martín, C., Eggermont, P. P. B., & Carazo, J. M. (2007). Modeling Experimental Image Formation for Likelihood-Based Classification of Electron Microscopy Data. *Structure*, *15*(10), 1167–1177. https://doi.org/10.1016/j.str.2007.09.003

Scheres, S. H. W., Valle, M., Nuñez, R., Sorzano, C. O. S., Marabini, R., Herman, G. T., & Carazo, J. M. (2005). Maximum-likelihood multi-reference refinement for electron microscopy images. *Journal of Molecular Biology*, *348*(1), 139–149. https://doi.org/10.1016/j.jmb.2005.02.031

Shaikh, T. R., Gao, H., Baxter, W. T., Asturias, F. J., Boisset, N., Leith, A., & Frank, J. (2008). Spider image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nature Protocols*, *3*(12), 1941–1974. https://doi.org/10.1038/nprot.2008.156

Shaikh, T. R., Gao, H., Baxter, W. T., Asturias, F. J., Leith, A., Frank, J., & Lourmel, R. De. (2009). Spider. *Nat Protoc*, *3*(12), 1941–1974. https://doi.org/10.1038/nprot.2008.156.SPIDER

Sigworth, F. J. (1998). A maximum-likelihood approach to single-particle image refinement. *Journal of Structural Biology*, *122*(3), 328–39. https://doi.org/10.1006/jsbi.1998.4014

Silverman, M. R., & Simon, M. I. (1972). Flagellar assembly mutants in Escherichia coli. (Counter Rotation Hooks). *Journal of Bacteriology*, *112*(2), 986–993.

Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., & Ludtke, S. J. (2007). EMAN2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology*, *157*(1), 38–46. https://doi.org/10.1016/j.jsb.2006.05.009

Wagenknecht, T., & DeRosier, D. (1981). Reconstruction of the Flagellar from Caulobacter crescentus. *J. Mol. Biol.*, *151*, 439–465.

Wagenknecht, T., DeRosier, D. J., Aizawa, S. I., & Macnab, R. M. (1982). Flagellar hook structures of Caulobacter and Salmonella and their relationship to filament structure. *Journal of Molecular Biology*, *162*(1), 69–87. https://doi.org/10.1016/0022-2836(82)90162-0

Wang, R. Y. R., Song, Y., Barad, B. A., Cheng, Y., Fraser, J. S., & DiMaio, F. (2016). Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *eLife*,

5(September2016), 1–22. https://doi.org/10.7554/eLife.17219

Webb, B., & Sali, A. (2014). *Comparative protein structure modeling using MODELLER*. *Current Protocols in Bioinformatics* (Vol. 2014). https://doi.org/10.1002/0471250953.bi0506s47

Yonekura, K., Maki-Yonekura, S., & Namba, K. (2003). Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature*, *424*(6949), 643–650. https://doi.org/10.1038/nature01830

Yu, I., Zhang, W., Holdaway, H. A., Li, L., Kostyuchenko, V. A., Chipman, P. R., … Chen, J. (2008). Proteolytic Maturation. *Science*, *13*(March), 1834–1838. https://doi.org/10.1126/science.1153264

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *Journal of Structural Biology*, *193*(1), 1–12. https://doi.org/10.1016/j.jsb.2015.11.003

Zheng, S., Palovcak, E., Armache, J.-P., Cheng, Y., & Agard, D. (2016). Anisotropic Correction of Beam-induced Motion for Improved Single-particle Electron Cryo-microscopy. *bioRxiv*, 1–30. https://doi.org/10.1101/061960