A polymerizable GFP variant

Ágnes Klein¹, Balázs Tóth²,², Hajnalka Jankovics⁴, Adél Muskotáľ¹ and Ferenc Vonderviszt¹,³,⁴

¹Bio-Nanosystems Laboratory, Faculty of Information Technology, Research Institute of Chemical and Process Engineering, University of Pannonia, Egyetem u. 10, H-8200 Veszprém, Hungary, ²Agricultural Research Institute of the Hungarian Academy of Sciences, Brunsvík u. 2, H-3462 Martonvásár, Hungary and ³Research Institute for Technical Physics and Materials Science, Hungarian Academy of Sciences, Konkoly Thege u. 29-33, H-1121 Budapest, Hungary

Received October 9, 2011; revised December 6, 2011; accepted January 9, 2012

Edited by Ashutosh Chilkoti

Flagellin has the ability to polymerize into long filaments under appropriate conditions. Our work aims at the construction of flagellin-based fusion proteins which possess polymerization ability and preserve the functional properties of the fusion partner as well. The hypervariable D3 domain of Salmonella flagellin, containing residues 190–283, is a good target for genetic engineering studies since it can be extensively modified or removed without disturbing the self-assembling ability. In this work a fusion construct of flagellin and the superfolder mutant of the green fluorescent protein were created by replacing D3 with superfolder green fluorescent protein (GFP). The obtained GFP variant was capable of forming stable, highly fluorescent filamentous assemblies. Our results imply that other proteins (enzymes, binding proteins, etc.) can also be furnished by polymerization ability in a similar way. This approach paves the way for the construction of multifunctional tubular nanostructures.

Keywords: flagellin/GFP/polymerization/self-assembly/tubular nanostructure

Introduction

Bacteria swim by rotating micrometer-long helical filaments made of several tens of thousands copies of the flagellin (FliC) protein (Vonderviszt and Namba, 2008). Flagellar filaments are self-assembling systems offering a range of potential applications in nanotechnology as nanoscaffolds (Kumara et al., 2007a,b; Woods et al., 2007). We aim at developing building blocks for rationally designed assembly of flagella-based filamentous nanostructures by creating fusions of various proteins with flagellin.

In the monomeric form, flagellin possesses natively disordered terminal regions which span the first 66 and the last 44 amino acid residues (Kostyukova et al., 1988; Vonderviszt et al., 1989). The disordered terminal regions control polymerization properties of subunits and become stabilized into alpha-helical bundles upon filament formation (Vonderviszt et al., 1991; Mimori-Kiyosue et al., 1997). Flagellar filaments are resistant to physical and chemical effects as well as to proteases, and their structure is now well known at the atomic level (Samatey et al., 2001; Yonekura et al., 2003).

Flagellin from Salmonella typhimurium consists of 494 amino acid residues. Comparison of amino acid sequences revealed a high degree of homology in the terminal regions of flagellins containing ~180 N-terminal and 100 C-terminal residues, whereas the central segments are highly variable (Beatson et al., 2006). Only the conserved terminal regions of flagellin subunits are involved in filament formation whereas the hypervariable central region forms the D3 domain exposed on the filament surface (Yonekura et al., 2003). This domain is not in contact with adjacent subunits and has no significant role in the construction of the filament structure (Muskotáľ et al., 2010). The D3 domain is a good target for genetic engineering studies as it can be removed or extensively modified without disturbing self-assembling ability. The concept of our work is to replace the hypervariable D3 domain of flagellin with suitable monomeric proteins, and to use these chimeric flagellins to build filamentous objects.

The green fluorescent protein (GFP) from the jelly-fish Aequorea victoria is a widely used reporter protein in biochemical research (Tsien, 1998). GFP is composed of 238 amino acids. A key sequence of Ser–Tyr–Gly at amino acids 65–67 functions as the GFP fluorophore. These three amino acids undergo spontaneous oxidation to form a cyclized chromophore which is responsible for the fluorescence. GFP has a β-barrel architecture. The chromophore resides in the center of the barrel. Correct folding of the GFP β-barrel appears to be a prerequisite for formation of the fluorescent chromophore (Prendergast, 1999). Superfolder GFP (sfGFP) was developed for robust folding by directed evolution and contains 11-point mutations with respect to the wild-type protein (Pédelacq et al., 2006).

Recently, the prototype of flagellin-based polymerizable enzymes has been created by replacing the hypervariable central portion of the polypeptide chain of FliC forming the D3 domain with the amino acid sequence of the xylanase A (XynA) enzyme (Szabó et al., 2011). The FliC(XynA) fusion protein was capable of filament formation under appropriate conditions, and exhibited xylan degrading activity both in the monomeric and polymeric form. XynA is composed of 185 amino acids, and it has a jellyroll fold containing an α-helix and two twisted β-sheets (Murakami et al., 2005). The aim of this study was to explore whether the GFP, a massive β-barrel protein composed of 238 amino acid residues, can also be inserted into the central part of flagellin in such a way that it can fold into its native conformation and the essential functional properties of both parental proteins are preserved.

Materials and methods

Plasmid construction
A pKOT-based plasmid containing a D3 deletion mutant flagellin (ΔD3_FliC) gene was created as described previously...
(Szabó et al., 2011). In this construct the D3 coding region was deleted and replaced by a linker segment containing the recognition sites of XhoI, AgeI, XmaI and SacI enzymes. This plasmid served as a cloning vector to create the gene encoding the FlIC(sfGFP) fusion construct. The sfGFP gene was synthesized by Genscript Corporation (Piscataway, NJ, USA) based on the protein sequence data published by Pédelaq et al. (2006). The sfGFP gene was amplified using forward primer 5'-ACATCTCGAGGGTTCGGTGACAGACGCGGAGGACGTG-3' and reverse primer 5'-ACATGAGC TTCTTGTACAGCTCGTCCATGCCG-3', containing the XhoI and SacI cleavage sites, respectively. The amplified product was inserted into the multiple cloning region in the middle part of the ΔD3_FliC gene (Szabó et al., 2011) between XhoI and SacI restriction sites using the Rapid DNA Ligation Kit (Fermentas). The forward primer also coded an extra glycine–serine segment which is together with the remaining amino acids from the multiple cloning region resulted in the LEGS and EL linkers at the N- and C-terminus of the inserted sfGFP, respectively, to avoid conformational strain at the insertion site.

The obtained plasmid construct (named NT19a) containing the gene of FlIC(sfGFP) was transformed into Escherichia coli TOP10 cells (Invitrogen), checked by XhoI and SacI digestion, and sequenced to confirm the presence of the desired DNA sequence. To investigate the potential of in vivo filament formation, the NT19a construct was electroporated into the flagellin-deficient SJW2536 (Ohnishi et al., 1994) Salmonella strain. For His-tagged overexpression in E. coli hosts, the gene of the FlIC(sfGFP) fusion protein was transported into a PET23b vector (Novagen). The forward primer 5'-ACATGAATTCATGGCACAAGTCATTAATACAAACA GCC-3' and reverse primer 5'-ACATGAGC TTCTTGTACAGCTCGTCCATGCCG-3' were used to amplify the FlIC(sfGFP) gene from NT19a, and the obtained product was ligated between the EcoRI and NotI sites of the pET23b vector. Since it is known from previous studies (Chen et al., 1994) that the optimal distance between the start codon and the ribosome-binding site (RBS) is in the range of 5–13 nucleotides, our pET23b-based construct was further modified by removing a 46-nucleotide-long DNA fragment upstream of the insert bordered by the Ndel and EcoRI sites to bring closer the start codon of the FlIC(sfGFP) gene to the RBS encoded by the pET23b to improve the expression yield. Moreover, this modification resulted in the removal of redundant N-terminal amino acids from our recombinant protein. This plasmid construct was named NT19b.

**Protein expression and purification**

The pET23b-based NT19b construct was transformed into BL21-CodonPlus (DE3)-RIL cells (Stratagene) to overexpress the C-terminally His₆-tagged FlIC(sfGFP) fusion protein. A 20 ml overnight culture was inoculated into 2 l of Luria-Bertani broth, supplemented with ampicillin (100 μg/ml) and chloramphenicol (30 μg/ml) and incubated at 37°C to an OD₆₀₀ of 0.6. After the addition of isopropyl-β-D-1-thiogalactopyranoside to a final concentration of 0.5 mM, incubation was continued at 20°C for another 16 h. The bacterial pellet was collected by centrifugation at 4000 g for 30 min and stored at −20°C.

Purification of FlIC(sfGFP) was performed as follows: the bacterial pellet from 21 culture was resuspended in 80 ml 20 mM sodium phosphate buffer containing 500 mM NaCl and 50 mM imidazole, pH 7.4 and two EDTA-free Complete tablets (Roche) were added to inhibit unwanted proteolytic degradation. The sample was sonicated on ice to lyse the cells for 1 min four times at power of 10 W. The solution was centrifuged at 50 000 r.p.m. 30 min at 4°C to remove cell debris and aggregates. The supernatant was pre-purified on an SP FF cation exchange column (Amersham Pharmacia Biotech) to remove contaminating proteases (Saijo-Hamano et al., 2000). Then the flow-through fraction was applied to a 5 ml HisTrap Chelating Ni-affinity column. The absorbed proteins were eluted using a linear imidazole gradient (50–200 mM), the His₆-tagged FlIC(sfGFP) protein was eluted at ≈100 mM imidazole. After dialysis against 20 mM Tris-HCl buffer (pH 7.8) at 4°C, samples were further purified by anion-exchange chromatography on a MonoQ HR5/5 column. The elution buffer was 20 mM Tris-HCl (pH 7.8) and 0.15–0.2 M NaCl per 20 min gradient was used at a flow rate of 0.5 ml/min. Pure fractions were collected and dialyzed at 4°C against TN buffer (20 mM Tris-HCl, 150 mM NaCl, pH 7.8), and stored at −20°C.

C-terminally His-tagged sfGFP was also cloned, overexpressed and purified for use as a reference. For this construct, the gene of sfGFP was inserted into a PET23b vector using BamHI and NotI restriction enzymes. Recombinant sfGFP was overexpressed in E. coli BL21-Gold(DE3) cells and purified by Ni-affinity chromatography on a HisTrap column. sfGFP eluted at ≈150 mM imidazole concentration. Collected fractions were dialyzed against 20 mM Tris-HCl buffer and stored at −20°C. Flagellin from Salmonella SJW1103 was purified as described previously (Vonderviszt et al., 1989). Purity of protein samples was checked using SDS-PAGE followed by Coomassie blue R-250 staining.

**Fluorescence measurements**

Fluorescence emission and excitation spectra were recorded with a Fluoromax-2 (Jobin-Yvon) computer controlled fluorescence spectrophotometer. Excitation spectra were determined measuring the emission intensity at 509 nm. The emission spectra were measured applying excitation at 488 nm. Protein concentrations were usually ~50 μg/ml. Filament samples were diluted immediately before measurements to avoid significant monomer population as a result of slow depolymerization below the critical monomer concentration. All measurements were carried out in TN buffer.

**Polymerization experiments**

The polymerizing ability of the FlIC(sfGFP) fusion protein was investigated by inducing nucleation and polymerization by ammonium sulfate (AS) (Vonderviszt et al., 1991). Protein solutions of 2 mg/ml were prepared in TN buffer and 4 M AS was added to various final concentrations in the range of 0.4–0.6 M. Filament formation was observed after 24 h of incubation at 20°C. For copolymerization of flagellin and FlIC(sfGFP), monomer solutions of Salmonella SJW1103 wild-type flagellin and FlIC(sfGFP) in TN buffer were mixed at various protein ratios in the range of 1 : 1 to 1 : 5 (w/w), and polymerized by the addition of AS. The mixtures were left at 20°C for 2 days to accomplish polymerization. Filaments were observed by dark-field optical and fluorescent microscopy with an Olympus BX50 microscope equipped with a fluorescence attachment.
Results

Design and gene construction of FliC(sfGFP)

In this work we aim at inserting the superfolder GFP into the central portion of Salmonella flagellin replacing the hyper-variable D3 domain (Fig. 1). To obtain a functional flagellin-based fusion protein, it is essential to apply appropriate linkers which allow proper folding of both flagellin and the protein inserted. On removal of the D3 domain, the resulting free ends of D2 are separated only by \( \approx 6 \text{Å} \) (Muskotań et al., 2010). In the case of sfGFP the relative position of the two terminals is not known precisely, because the last six residues are not visible on the electron density map (Pe´delacq et al., 2006; PDB code: 2B3P), probably reflecting high segmental flexibility of this region. Nevertheless, the two ends of GFP are at the same end of the \( \beta \)-barrel and are supposed to be close to each other facilitating insertion into flagellin. Based on the available structural data and the high flexibility of the very C-terminal segment of sfGFP, we estimated that short linker segments LEGS and EL would be suitable at the N- and C-terminus, respectively, to avoid conformational strains on insertion. A DNA segment coding for sfGFP and the linkers at its both ends was synthesized and ligated into the middle part of the gene of the D3 deletion mutant flagellin (Szaboń et al., 2011), and this fusion construct was cloned into a pKOT-based plasmid (Muskotań et al., 2010) and a pET23b vector obtaining the NT19a and NT19b plasmids, respectively. NT19a and NT19b were used for overexpression of FliC(sfGFP) in Salmonella and E.coli, respectively.

Overexpression and purification of FliC(sfGFP)

When the NT19a plasmid vector carrying the FliC(sfGFP) gene was introduced into the flagellin-deficient non-motile SJW2536 Salmonella strain, it was expressed at a high level but neither filament formation nor recovery of motility was observed. FliC(sfGFP) was not exported by the flagellum-specific export system and remained inside the cell.

A C-terminally His\(_6\)-tagged version of FliC(sfGFP) was overexpressed in BL21-CodonPlus (DE3)-RIL E.coli cells. It was evident that the GFP portion of the overexpressed fusion protein obtained its correctly folded conformation in the cytoplasm, indicated by the green color of bacteria as observed by fluorescent microscopy. The FliC(sfGFP) fusion protein remained in soluble form in the cytosol, and no significant inclusion body formation was observed. The protein was purified from the cell lysate by Ni-chelating affinity chromatography. Further purification was performed by ion-exchange chromatography to obtain samples of high (\( >95\% \)) purity. Since the terminal regions of flagellin are disordered, extra attention was paid to use protease inhibitors preventing degradation of these regions during purification. Typically, 10–20 mg purified FliC(sfGFP) was obtained from 1 l culture.

Spectral characteristics of FliC(sfGFP)

The fluorescence excitation and emission properties of GFP strongly depend on the local environment of the fluorophore. Denatured GFP does not show fluorescence (Prendergast, 1999). To demonstrate that sfGFP can assume its native conformation even upon insertion into the middle part of flagellin, the fluorescent properties of FliC(sfGFP) were measured and compared with the recombinant sfGFP protein which has excitation and emission maxima at 488 and 509 nm, respectively (Pe´delacq et al., 2006). Spectral properties of FliC(sfGFP) both in the monomeric and polymeric forms were highly reminiscent to those of sfGFP, indicating that the sfGFP protein inserted into the central part of flagellin was properly folded. The excitation maximum was 488 nm in all cases. Monomeric FliC(sfGFP) had an emission peak at 509 nm and its fluorescent intensity increased by \( \approx 10\% \) as compared to sfGFP (Fig. 2). The emission maximum of

![Fig. 1. Construction of the FliC(sfGFP) fusion protein. (A) Arrangement of flagellin subunits within the flagellar filament. The hypervariable D3 domain is situated on the filament surface. Solid surface representation of a longitudinal- and cross section of the filament according to Mimori-Kiyosue et al. (1996). (Copyright (1996) National Academy of Sciences, USA.) The scale bar represents 10 nm. (B) The C\(_\alpha\) backbone trace of flagellin (PDB code: 1UCU) and sfGFP (PDB code: 2B3P). sfGFP was inserted into the middle part of flagellin to replace the D3 domain (residues 190–283). Short linker segments LEGS and EL were used at the N- and C-terminus of sfGFP, respectively, to facilitate proper folding.](https://academic.oup.com/peds/article-abstract/25/4/153/1514420)
polymeric FliC(sfGFP) shifted by 2 nm from 509 to 511 nm accompanied by a further 7% increase in intensity. The observed small differences between the spectra probably reflect structural constraints exposed on the GFP fold upon insertion into flagellin and formation of the filamentous assembly.

**Filament formation**

Filament formation of flagellin subunits can be induced by precipitants, such as AS. Purified C-terminally His6-tagged FliC(sfGFP) readily polymerized into a few micrometer-long filaments at 0.4 M AS (Fig. 3). Changing the applied precipitant concentration, filaments of various lengths were obtained. Filaments exhibited a straight type superhelical form indicating slight structural perturbations in the filament core (Vonderviszt et al., 1991), probably owning to the His6-tag attached to the C-terminus. FliC(sfGFP) filaments were stable in TN buffer for a long period of time as monitored by dark field and fluorescent microscopy. Their density and average length remained virtually constant at least for 3 weeks at room temperature. Copolymerization of FliC(sfGFP) with wild-type flagellin at various ratios in the range of 1:1–1:5 (w:w) also resulted in stable filaments. SDS-PAGE gels of copolymer samples indicated that FliC(sfGFP) was incorporated into the filaments in a proportion similar to the applied mixing ratio (data not shown).

**Discussion**

The ability of flagellin to form filaments on the surface of bacteria has been exploited to display foreign proteins or peptides for studying binding interactions and antigenic properties (Stocker and Newton, 1994; Westerlund-Wikstrom, 2000; Amemiya et al., 2005). Nevertheless, this approach usually worked only with oligopeptides or small proteins since the flagellum-specific export machinery was unable to efficiently translocate larger fusion proteins through the narrow central channel of the flagellar filament. In addition, a typical problem was in previous attempts that foreign proteins were introduced into the hypervariable central region of flagellin without taking into consideration its precise domain organization, and this often led to structural perturbations reflected in impaired polymerization properties and destabilization of filaments. We have recently demonstrated that removal of the hypervariable D3 domain of flagellin does not influence polymerization ability and the stability of the resulting filaments, if the other domains remain undisturbed (Muskotál et al., 2010). Thus, replacement of D3 by insertion of heterologous proteins or domains may offer a promising approach for creating flagellin-based fusions capable of forming stable filaments. It is essential to apply appropriate linkers facilitating insertion, which allow proper folding and functioning of the foreign protein without interfering with filament formation.

We realized that creating fusions with flagellin offers much wider opportunities than surface display. A large proportion of proteins are not suitable for flagellar surface display because their size and other properties prevent efficient flagellar export. However, their fusion constructs with flagellin can be readily overexpressed in bacteria, purified with ease and used for *in vitro* construction of filamentous nanostructures. This approach opens up new horizons for applications, substantially increasing the number of potential fusion partners of advantageous properties.

As a first attempt to explore the feasibility of the concept to produce flagellin-based polymerizable proteins, the fusion construct of flagellin and the xylanase A enzyme (FliC(XynA)) has been recently created (Szabó et al., 2011). The N-terminally His6-tagged FliC(XynA) chimera was enzymatically active but its polymerization ability was somewhat paralyzed. Careful inspection of the filament structure suggested that the His6-tag-containing N-terminal extra segment interfered with the proper formation of the inner core of filaments.

In this work we aimed at developing a polymerizable variant of GFP by creating fusion with flagellin. This time the purification tag was attached to the C-terminus to diminish disturbance of the filament structure. FliC(sfGFP)-His6...
was expressed at a high level in *E. coli* cells and purified by Ni-affinity and ion-exchange chromatography. The purified protein was highly fluorescent with spectral properties virtually identical to those of monomeric sfGFP. The FliC(sfGFP) fusion protein was polymerizable into long filaments which were stable at least for weeks at room temperature. Since the very C-terminal segment of flagellin subunits forms the inner wall of the central channel of filaments (Yonekura et al., 2003), it seems that the C-terminal purification tag can be accommodated into the central channel without adversely affecting filament formation. Our results demonstrate that GFP, a massive β-barrel protein, can assume its native structure on insertion into the middle part of flagellin. This is consistent with earlier studies which demonstrated that GFP is rather fusion tolerant, and can be successfully inserted and functionally accommodated within the structure of other proteins (Sheridan et al., 2002, 2004).

Engineered flagellins were used as building block modules to create flagellar filaments with functionalized surfaces tailored for nanoscaffold uses (Woods et al., 2007). Kumara et al. (2007a,b) demonstrated the versatility of employing bioengineered metal binding flagella in nanotechnology for the generation of a various nanoparticle arrays and nanotubes. In a similar way as demonstrated here, polymerization ability can be introduced into a variety of proteins (e.g. enzymes or binding proteins) as well, and building blocks for rationally designed assembly of flagellar nanостructures can be created. Directed co-polymerization of various flagellin-based fusion constructs with favorable enzymatic, signaling or molecular recognition properties allows fabrication of tubular nanostuctures that offer promising applications in nanosensorics, nanomedicine or bio-nanotechnology.

Acknowledgements

We thank K. Namba and P. Závodszky for support and encouragement. We are grateful to the Japan Science and Technology Corporation for generous donation of equipment.

Funding

This work was supported by the Hungarian Scientific Research Fund and the National Development Agency (grants CK77819, TAMOP-4.2.2/B-10/1-2010-0025 and REG_KD_09-2-2009-0022-NANOFLAG). B.T. was supported by the Bolyai Fellowship of the Hungarian Academy of Sciences.

References


