Protein Expression and Purification 86 (2012) 127–134

Contents lists available at SciVerse ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

Expression and purification of $SfaX_{II}$, a protein involved in regulating adhesion and motility genes in extraintestinal pathogenic *Escherichia coli*

Patricia Paracuellos [a], Anders Öhman [b,1], A. Elisabeth Sauer-Eriksson [b], Bernt Eric Uhlin [a,*]

[a] Department of Molecular Biology, Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå Centre for Microbial Research (UCMR), Umeå University, SE-901 87 Umeå, Sweden
[b] Department of Chemistry, Umeå Centre for Microbial Research (UCMR), Umeå University, SE-901 87 Umeå, Sweden

ARTICLE INFO

Article history:
Received 22 June 2012
and in revised form 14 September 2012
Available online 25 September 2012

Keywords:
$SfaX_{II}$
ExPEC
Fimbriae
Transcription regulation

ABSTRACT

Pathogenic *Escherichia coli* strains commonly harbor genes involved in formation of fimbriae, such as the sfa_{II} fimbrial gene cluster found in uropathogenic and newborn meningitis isolates. The $sfaX_{II}$ gene, located at the distal end of the sfa_{II} operon, was recently shown to play a role in controlling virulence-related gene expression in extraintestinal pathogenic *E. coli* (ExPEC). Until now, detailed characterization of the $SfaX_{II}$ protein has been hampered by difficulties in obtaining large quantities of soluble protein. By a rational modeling approach, we engineered a Cys70Ser mutation, which successfully improved solubility of the protein. Here, we present the expression, purification, and initial characterization of the recombinant $SfaX_{IIC70S}$ mutant. The protein was produced in *E. coli* BL21 (DE3) cells grown in autoinduction culture media. The plasmid vector harbored DNA encoding the $SfaX_{IIC70S}$ protein N-terminally fused with a six histidine (H6) sequence followed by a ZZ tag (a derivative of the *Staphylococcus* protein A) (H6-ZZ tag). The H6-ZZ tag was cleaved off with Tobacco Etch Virus (TEV) protease and the 166 amino acid full-length homo-dimeric protein was purified using affinity and size-exclusion chromatography. Electrophoretic mobility gel shift assays and atomic force microscopy demonstrated that the protein possesses DNA-binding properties, suggesting that the transcriptional regulatory activity of $SfaX_{II}$ can be mediated via direct binding to DNA.

Introduction

Escherichia coli (*E. coli*)[2] is a commensal bacterium in the human intestinal microbiota, nevertheless several pathogenic variants exist that can cause various intestinal or extraintestinal infections in man and animals. Extraintestinal pathogenic *E. coli* (ExPEC) strains are a common cause of urinary tract infections but can also be the causative agents of other human diseases, such as meningitis, sepsis, pneumonia, and surgical site infections [1,2]. The ExPEC strains possess several virulence traits that facilitate colonization, invasion and pathogenesis in specific bodily locations [3]. The interaction of ExPEC with host cells is generally mediated by fimbriae, hair-like structures protruding from the bacterial cell surface with an adhesin at the distal end that enable bacterial adhesion to different host receptors. The bacteria also often have the ability to express different fimbriae in an alternative manner and in response to signals from the environment [4].

Proteins needed for production of different types of fimbriae, e.g., P fimbriae, F1C fimbriae, Type 1 and S fimbriae, are encoded by separate gene clusters, such as *pap*, *foc*, *fim* and *sfa*, respectively. Usually, these gene clusters, acquired by horizontal gene transfer [5,6], consist of a set of genes for biogenesis of the actual fimbriae components-a chaperone, an usher, the fimbrial subunits, and an adhesin – as well as regulatory genes. Recent studies revealed a regulatory network among fimbrial adhesin gene systems in ExPEC, which suggest that there may be a hierarchy in their expression [7].

S-fimbriae is predominantly expressed by newborn meningitis *E. coli* (NMEC) isolates, which cause neonatal meningitis. However, S-fimbriae are also found in uropathogenic *E. coli* (UPEC) strains, enabling the bacteria to bind to human bladder and kidney epithelium [8]. The gene $sfaX_{II}$, located immediately downstream of the S-fimbrial gene cluster sfa_{II}, was first characterized in the NMEC isolate IHE3034 [9,10]. Genetic analyses have shown that the $SfaX_{II}$ protein acts on the phase variation switch of the mannose-binding Type 1 fimbriae, and that it has a regulatory role both in fimbrial expression and in flagella expression [11].

* Corresponding author. Fax: +46 90 772630.
E-mail address: bernt.eric.uhlin@mims.umu.se (B.E. Uhlin).
[1] Present address: Department of Pharmacology and Clinical Neuroscience, Umeå University, SE-901 85 Umeå, Sweden.
[2] *Abbreviations used: E. coli*, *Escherichia coli*; RMSD, root-mean-square deviation; LB, Luria–Bertani; SDS–PAGE, dodecyl sulfate–polyacrylamide gel electrophoresis; PVDF, polyvinylidene fluoride; PMSF, phenylmethanesulfonylfluoride; DTT, dithiothreitol; TEV, Tobacco Etch Virus; CD, circular-dichroism; GST, glutathione-*S*-transferase; MBP, maltose-binding protein; IPTG, isopropyl-β-D-thiogalactopyranoside.

1046-5928 © 2012 Elsevier Inc. Open access under CC BY-NC-ND license.
http://dx.doi.org/10.1016/j.pep.2012.09.006

Analysis of the $SfaX_{II}$ protein sequence suggests that it belongs to the MarR family of transcriptional regulators, a protein family involved in mechanisms such us the control of antibiotic resistance, virulence factor production, as well as the response to oxidative stress [12]. These proteins are found throughout the bacterial and archaeal domains [13] and bind palindromic or pseudopalindromic DNA as homodimers, resulting in either transcriptional repression, activation or both [14]. Crystal structures have been determined for several members of the MarR family, including the structures of apoproteins [15] and proteins in complex with cognate DNA [16,17]. The MarR family of proteins are alpha-helical and interact with DNA via a conserved, winged, helix-turn-helix motif [18].

$SfaX_{II}$ is also a member of a subfamily of 17-kDa proteins encoded by genes downstream of various fimbrial operons, such as *pap*, *foc*, *prs* and *prf*, in both UPEC and NMEC strains. In keeping with the localization of these genes, the proteins have been named PapX, FocX, PrsX and PrfX, respectively, after the corresponding upstream operon [11,19]. The sequence identity between $SfaX_{II}$ and these 17-kDa proteins varies between 87–96% [9,11]. A target DNA sequence for the PapX protein was recently determined in the *flhD* regulatory region [20], but otherwise little is known about the properties of this family of proteins. To begin characterizing the biological function of this family of proteins, we have developed an efficient protocol for expression and purification of recombinant $SfaX_{II}$ from *E. coli*. In addition, we present here a molecular model of the protein structure of $SfaX_{II}$.

Materials and methods

Protein sequence analysis and molecular modeling of $SfaX_{II}$ in silico

$SfaX_{II}$ was modeled *in silico* with the homology molecular modeling program MODELLER 9v6 [21]. The software identified the following set of four crystal structures as templates, allowing optimal modeling of the $SfaX_{II}$ protein from residues 26 to 166: PDB IDs: 2FBH (*Pseudomonas aeruginosa* transcriptional regulator MarR family), **1JGS** (*E. coli* multiple antibiotic-resistance protein MarR), **2A61** (*Thermotoga maritima* transcriptional regulator MarR family), **3BJA** (*Bacillus cereus* putative MarR-like transcription regulator). Their sequences were aligned using ClustalW [22]. Five distinct models were generated, and their geometry was assessed by a Ramachandran plot calculated with the program RAMPAGE [23].

Calculation of root-mean-square deviation (RMSD) of Cα atoms and superimposition of structures were performed with the program DaliLite [24]. Figures of the structures were prepared with the programs PyMOL [25] and ESPript [26]. Predictions of the secondary structure were made using Jpred software [27].

Bacterial strains, plasmids and growth conditions

The *E. coli* strains and plasmids used in the present work are described in Table 1. Unless otherwise stated, the strains were grown at 37 °C in either Luria–Bertani (LB) broth or autoinduction medium [28] with vigorous shaking as well as on tryptone yeast agar. When necessary, antibiotics were added at the following concentrations: carbenicillin 50 μg/mL and kanamycin 50 μg/mL.

Cloning of the gene encoding $SfaX_{II}$ and generation of the $SfaX_{IICys70}$ mutant

The *sfaX*$_{II}$ gene encoding the $SfaX_{II}$ protein was amplified by PCR from genomic DNA of the extraintestinal pathogenic *E. coli* strain IHE3034. NcoI and Acc65I restriction sites were introduced into the oligonucleotide primers for the upstream and downstream sequences. PCR was carried out with High Fidelity *Pfu* DNA polymerase under the following conditions: 3 min at 95 °C and 40 cycles of 30 s at 95 °C, 30 s at 68 °C and 1 min at 72 °C, followed by 5 min at 72 °C for final elongation of the products. The PCR fragment was digested with NcoI and Acc65I and subcloned into pET24d vectors including a H6 tag combined with GST, MBP, thioredoxin, NusA, the Gb1-domain, and a double Z-domain (ZZ) (kindly provided by Gunter Stier) as N-terminal fusions, respectively [29]. Ligation was carried out with T4 DNA ligase (Promega) overnight at 15 °C.

Cysteine 70 was replaced by serine using the QuickChange 2 Site Directed Mutagenesis kit (Stratagene). The pETZ2_1a-$SfaX_{II}$ template, a pET-based plasmid encoding the H6-ZZ-SfaXII protein, was denatured at 95 °C. The primers for mutagenesis (containing the desired mutation) were annealed at 55 °C and extended using PfuUltra DNA polymerase at 68 °C. The parental DNA was digested with the restriction endonuclease DpnI. Purified mutant DNA (pPPT2) was introduced by transformation into competent cells and subsequently the plasmid clone with the *sfaX*$_{IIC70S}$ allele was harvested. The sequences of both the native parental and mutated genes were confirmed by DNA sequencing. For the *in vivo* analysis, the plasmid pAES1 was used as template and following the same protocol as above, we generated the plasmid pPPT1 (pAES1-*sfaX*$_{IIC70S}$). Likewise, the mutation was confirmed by DNA sequence analysis.

Table 1
E. coli strains and plasmids used in this study.

Strain/ plasmid	Description/relevant characteristics	Reference/ source
E. coli strains		
DH5α	F$^-$, *recA1*, *endA1*, *hsdR17*, *supE44*, *thi-1*, *gyrA96*, *relA1*	Laboratory stock
BL21(DE3)	F$^-$ *ompT hsdSB*(rB−, mB−) *gal dcm* (DE3)	Novagen
IHE3034	NMEC clinical isolate, O18K1 :H7	[10]
AES4	IHE3034 Δ*sfaX*$_{II}$	[11]
Plasmids		
pET28a(+)	Expression vector with hexahistidine tag, Kan^r	Novagen
pBR322	Cloning vector, Cb^R, Tc^R	[45]
pAES1	pBR322, *sfaX*$_{II}$ gene from *sfa*$_{II}$ operon	[11]
pPPT1	pAES1 with the gene encoding for the mutated protein $SfaX_{IIC70S}$	This study
pETZZ_1a	pET24d with the ZZ tag. Protein A binding domain, double domain.	[29,46]
pPPT2	pETZZ_1a- $sfaX_{IIC70S}$ gene	This study
pETGST_1a	pET24d, glutathion-S-transferase	[29,47]
pETMBP_1a	pET24d, maltose binding protein	[29,48]
pETTrx_1a	pET24d, *E. coli* thioredoxin A	[29,49]
pETNus_1a	pET24d, N-Utilization substance	[29,38]
pETGB1_1a	pET24d, protein G binding domain	[29,50]

Western blot analysis

3.5 mL of bacteria with an OD_{600} of 2.0 were pelleted by centrifugation and resuspended in sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) sample buffer. Identical amounts were subsequently subjected to SDS–PAGE (15% polyacrylamide concentration) and thereafter transferred to a polyvinylidene fluoride (PVDF) micro-porus membrane. To detect expression of $SfaX_{II}$ and $SfaX_{IIc70S}$ proteins, polyclonal rabbit antiserum obtained by immunization with gel purified protein was used as primary antibody. Further visualization was carried out using the ECL+ method described by the manufacturer (Amersham Biosciences). Detection of bands was performed using the Fujifilm LAS4000 system exposing the chemiluminescent membrane.

Motility assay

Overnight cultures were diluted in LB and adjusted to an OD_{600} of 1.0. A drop (5 μL) was put on semi-solid plates containing 0.3% agar and incubated at 37 °C to compare the diameter of the area of spreading bacteria.

Overproduction and purification of the mutant $SfaX_{IIC70S}$

E. coli BL21 (DE3) bacteria transformed by electroporation with the pPPT2 plasmid (pETZZ_1a- *$sfaX_{IIC70S}$*) were grown overnight at 37 °C in Luria–Bertani (LB) medium containing 50 μg/mL kanamycin. The overnight culture was used to inoculate (20 mL/L) 4 L of autoinduction medium [28] (containing 50 μg/mL kanamycin) and incubated for 12 h at 30 °C with shaking. Cells were harvested by centrifugation at 4000*g* for 10 min. The bacterial pellet (usually ~10 g) was resuspended in sodium phosphate buffer pH 7.4 (136 mM NaCl, 2.6 mM KCl, 10.2 mM Na_2HPO_4, 1.7 mM KH_2PO_4), 23 mM ammonium sulfate, 300 mM NaCl (buffer *A*), in the presence of 1 mM protease inhibitor phenylmethanesulfonylfluoride (PMSF) and 1 mM dithiothreitol (DTT). Cells were disrupted by sonication (20 min, 30% of amplitude, 5 s ON, 5 s OFF). The lysate was cleared by centrifugation (12000*g*, 1 h at 4 °C). The supernatant containing the soluble H6-ZZ tagged protein was filtered through a 0.22-μm pore filter and loaded onto a column containing Ni–NTA resin (Qiagen). The lysate/Ni–NTA-agarose mixture was washed eight times with a total of 40 mL of buffer *A* containing 10 mM imidazole. Protein elution was carried out with ten portions (5 mL) buffer *A* containing 200 mM imidazole. The eluted fractions were collected and analyzed by SDS–PAGE. The fractions containing the protein were collected and dialyzed overnight against buffer *A*. Protein concentration was determined by measuring the absorbance at 280 nm (A_{280}) using a theoretical extinction coefficient of 18.910 M^{-1} cm^{-1}. To prepare the untagged protein, the H6-ZZ tag was cleaved off by digestion with H6-AcTEV [30], an enhanced form of Tobacco Etch Virus (TEV) protease (kindly provided by Gunter Stier and Christin Grundström), using 1 unit of enzyme per 3 μg of H6-ZZ-$SfaX_{IIC70S}$ at 4 °C overnight. Digestion was done in buffer *A* with 1 mM DTT and 0.5 mM EDTA in order to enhance the proteolytic efficiency. Digested samples were loaded onto a column containing the resin IgG Sepharose 6 fast flow (GE Healthcare), which binds protein A fusion conjugates. H6-ZZ and undigested protein were bound to the resin and flow-through, containing the native protein, was collected. H6-AcTEV protease, also present in the flow-through, was removed by passing the flow-through over a Ni-agarose resin, which retained the protease. Finally, $SfaX_{IIC70S}$ was subjected to size-exclusion chromatography with an ÄKTA purifier (GE Healthcare) and a HiLoad 16/60 superdex 75 prep grade column equilibrated with buffer *A*. Elution corresponded to a 40-kDa protein indicating a dimeric form of the protein in solution. Calculating protein concentration with a theoretical extinction coefficient of 14.440 $M^{-1}cm^{-1}$, $SfaX_{IIC70S}$ was concentrated to 4 mg/mL using a 10-kDa molecular-mass cutoff membrane (Vivascience) and stored at 4 °C.

Circular dichroism spectroscopy

Far-UV circular-dichroism (CD) spectra were collected on samples containing 10 μM $SfaX_{IIC70S}$ in sodium phosphate buffer pH 7.4, 300 mM NaCl, and 23 mM ammonium sulfate using a JASCO J-810 spectropolarimeter and 1-mm quartz cuvettes. Spectra were recorded between 196 and 260 nm, averaged over five scans, with a bandwidth of 2 nm, a response time of 2 s, a pitch of 0.5 nm and a scan rate of 20 nm/min. Thermal unfolding and refolding of the protein was investigated by monitoring the ellipticity at 220 nm in the temperature interval 4–80 °C, using a temperature gradient of 1 °C/min. The melting temperature was determined from the inflection point of the transition. Evaluation of the secondary structure content of the protein was carried out by using the CDNN program [31].

Electrophoretic mobility shift assays

Electrophoretic mobility shift assays [32] were performed with the purified mutant $SfaX_{IIC70S}$ using a gradient of protein concentration from 0.4 to 2 μM, and 125 ng DNA representing 300 bp from the regulatory region of the *sfa* operon in each test of 10 μL (DNA concentration was determined using the NanoDrop system). Binding was allowed to occur at 37 °C in a buffer containing 25 mM Hepes pH 7.5, 100 mM KCl, 0.1 mM EDTA, 5 mM DTT, and 10% glycerol. After 15 min of incubation, aliquots of the assay mixture were mixed with loading solution and run on a 6% w/v polyacrylamide native gel at 12 mA for 1 h. Staining with ethidium bromide was sufficiently sensitive to visualize the DNA fragments.

Atomic force microscopy

DNA–protein complexes were diluted with ultrapure water (Millipore) and immediately placed on a freshly cleaved mica surface as described previously [33]. The samples were incubated at 37 °C for 5 min, gently washed with ultrapure water, and dried in a desiccator for at least 2 h. Imaging was performed on a Nanoscope V Atomic Force Microscope (Bruker) using Tapping mode. The images are presented in amplitude mode.

Results and discussion

We have characterized the transcriptional regulator $SfaX_{II}$ encoded within the *sfa* gene cluster in *E. coli* strains that cause newborn meningitis and uropathogenic diseases. $SfaX_{II}$ shows weak sequence similarity to the MarR family of transcriptional regulators and is a member of a 17-kDa protein family whose genes are located at the distal end of the *sfa*, *pap*, *foc*, *prs* and *prf* fimbrial gene clusters. The proteins in this family, SfaX, PapX, FocX PrsX and PrfX, have highly conserved primary amino acid sequences (>90%), but their structures have not yet been determined experimentally. Nevertheless, predicted secondary structures suggest that, despite low sequence identity (<20%), the proteins share structural features with the MarR family. Fig. 1 shows the predicted secondary structural motifs of $SfaX_{II}$ and MarR proteins whose structures have been determined. Overall, the distributions of helices and strands in $SfaX_{II}$ are predicted to be close to identical to those of the MarR proteins, with the exception of an N-terminal 26-amino-acid insertion in the $SfaX_{II}$ protein predicted to form an extra alpha-helix (α1). Using the program MODELLER and four previously determined structures of MarR representatives as templates (Figs. 1 and 2A), we built five molecular models of $SfaX_{II}$. Stereochemically, the quality of the polypeptide backbone and side chains of the modeled structures was assessed using Ramachandran plots obtained from the RAMPAGE server. The model with the most stable structure (Fig. 2B) had 133 residues (95.7%) in the most favored regions; four (2.9%) residues in additionally allowed regions and two (1.4%) residues in disallowed regions. Pairwise superimpositions of the $SfaX_{II}$ model with those of the MarR reference proteins yield average root-mean-square deviations (RMSD) of ~2 Å for 140 equivalent Cα positions in residues 26–166 of $SfaX_{II}$.

Initial attempts to purify $SfaX_{II}$ overexpressed in bacteria were unsuccessful due to insolubility of the protein. In order to increase solubility, we therefore turned to site-directed mutagenesis. $SfaX_{II}$ contains two cysteine residues at positions 32 and 70 that could

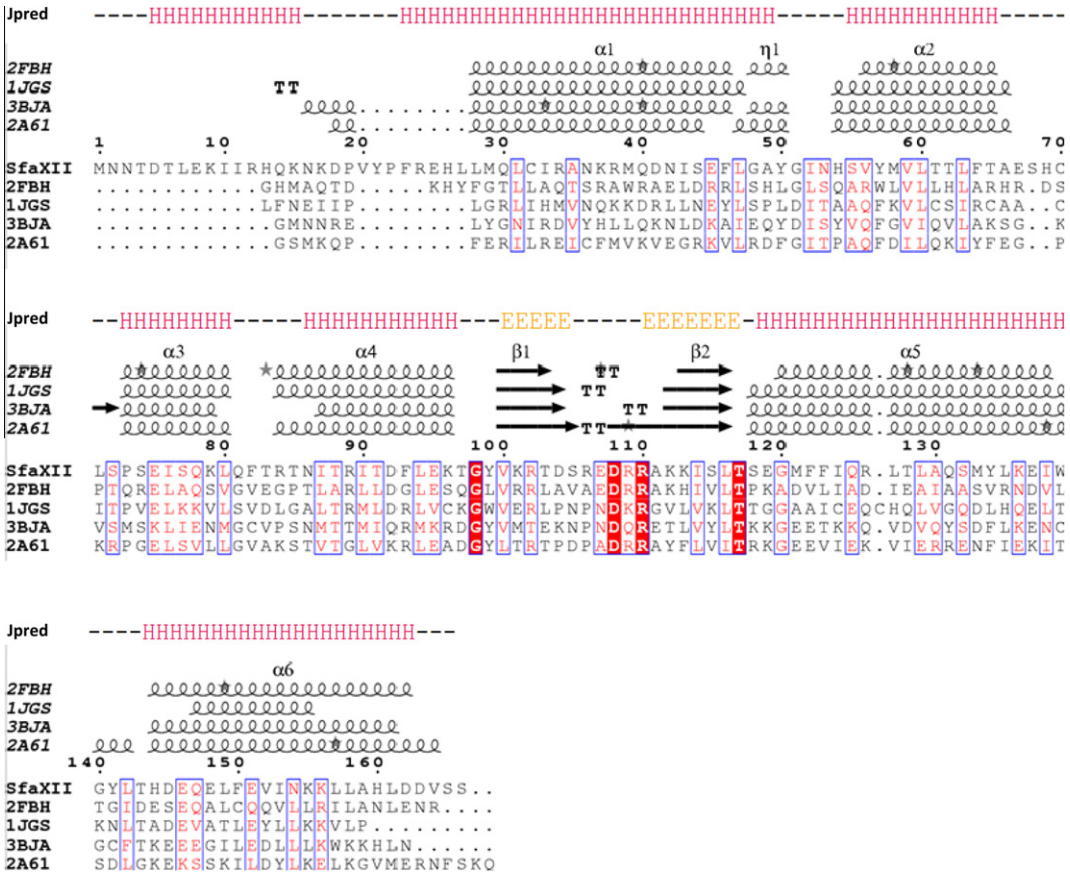

Fig. 1. Multiple sequence alignment of $SfaX_{II}$ with proteins of the MarR family identified by BLAST analysis. The amino acid sequences of $SfaX_{II}$ and transcriptional regulators of the MarR family with code PDB 2FBH , 1JGS, 2A61 and 3BJA for *Pseudomonas aeruginosa*, *Escherichia coli*, *Thermotoga maritima* and *Bacillus cereus*, respectively, were aligned using ClustalW. Identical residues are depicted in white on a red background; similar residues are shown in red type in an open box. The figure was drawn with ESPript and the secondary structure for each of the crystallized proteins is shown above the sequences, with arrows for beta-strands and ribbons for alpha-helices. Jpred software was employed for secondary structure prediction (H, α-helices and E, β-strands) of the protein $SfaX_{II}$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

reduce protein solubility by forming non-native disulphide bonds. From analyses of the model structure and the monomer–monomer interface in the biological dimer of the MarR proteins, we located the positions of the two cysteine residues within the presumed biological dimer of $SfaX_{II}$ (Fig. 2C). Cys70 is located at the surface of the structure and not in the vicinity of the helix-turn-helix DNA-binding motif. In contrast, the model suggested that Cys32 is involved in the formation of dimers. Based on this data, we therefore decided to mutate only the Cys70 residue to serine.

In order to analyze whether the C70S substitution would affect the phenotype of strains expressing $SfaX_{IIC70S}$, we mutated the *sfaX$_{II}$* gene contained in the plasmid pAES1 by site-directed mutagenesis. We transformed with this plasmid the extraintestinal pathogenic *E. coli* strain IHE3034 as well as the $SfaX_{II}$-deficient IHE3034 derivative (strain AES4) and performed tests similar to earlier published studies with the pAES1 clone [11]. Fig. 3A shows the expression levels of $SfaX_{II}$ and $SfaX_{IIC70S}$ in cells at the OD_{600} of 2.0. Strain "b" (IHE3034 transformed with the vector control pBR322) did not express a detectable level of $SfaX_{II}$. By contrast, with the strains transformed with the plasmids pAES1 or pPPT1 we could readily detect the $SfaX_{II}$ and $SfaX_{IIC70S}$ protein expression. Moreover, these two proteins were expressed at equal levels and their functionality was analyzed by bacteria motility assays. In studies of the effect of the $SfaX_{II}$ protein on flagella expression it was previously reported that $SfaX_{II}$ represses the flagella expression and therefore bacterial motility [11]. Fig. 3B shows the motility of the different (a–e) strains 5 h after inoculation. In correlation with western blot results, strains with low level or lacking $SfaX_{II}$ (b and c) were more motile. Furthermore, strains which expressed $SfaX_{II}$ (a and d) or $SfaX_{IIC70S}$ (e) at equal levels showed similarly reduced motility. Strain "e" showed the same phenotype as "d" and thus we can conclude that the substitution C70S is not affecting this *in vivo* regulatory property of the protein.

To further improve solubility, we tested co-expression of the protein with several carrier proteins. These carrier proteins are known to influence the expression, solubility and stability of target proteins because of their unique folding properties and high solubility [34]. For ease of purification, we used pET vectors with an H6 tag (six histidine) bound to the carrier proteins: glutathione-*S*-transferase (GST) [35], the maltose-binding protein (MBP) [36], thioredoxin [37], NusA [38,39], the Gb1-domain from protein G (Gb1) [40], and a double Z-domain from protein A (ZZ) [41]. All carrier proteins were fused to the N-terminal end of the $SfaX_{IIC70S}$ protein.

From co-expression tests of the $SfaX_{IIC70S}$ protein with the various carrier proteins [42], the best overexpression and solubility was obtained using the H6-ZZ-$SfaX_{IIC70S}$ construct. ZZ is a repetition of the "Z" domain based on the "B" IgG binding domain of protein A from *Staphylococcus aureus* and this domain can be easily purified using an IgG Sepharose™ 6 Fast Flow column to which the "ZZ" domain binds tightly [43]. This feature was used as an advantage to separate $SfaX_{IIC70S}$ after digestion from the H6-ZZ tag as well as from undigested protein.

To overexpress the protein, we initially used the conventional method with Luria–Bertani (LB) medium and the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) at the logarithmic phase. Many attempts were made, with changes to several different

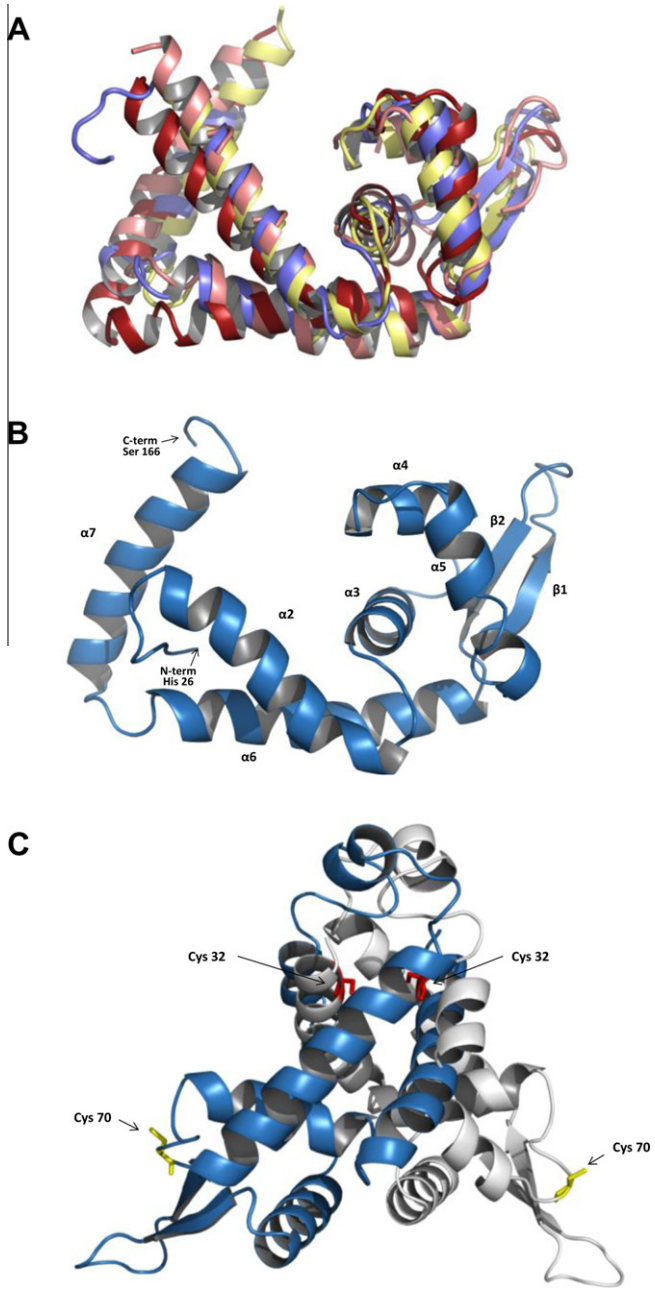

Fig. 2. Structural model of the protein $SfaX_{II}$ from *Escherichia coli* strain IHE3034. (A) Superimposition made with the program DaliLite of the four structures used as templates for molecular modeling of $SfaX_{II}$ (PDB codes: 1JGS, 2FBH, 2A61, 3BJA). (B) Cartoon representation of the structural model of $SfaX_{II}$ (residues 26–166). (C) Cartoon representation showing the model of the $SfaX_{II}$ dimer. The monomers are shown in blue and grey, respectively. The interface is modeled from the *Escherichia coli* MarR structure (PDB code 1JGS). The side chains of Cys32 and Cys70 are shown as sticks in red and yellow, respectively. Cys 32 in the two monomers are in close contact across the dimer interface and could be involved in an inter-subunit disulfide bridge. Cys70 residues, on the other hand, are situated at the surface of the complex. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

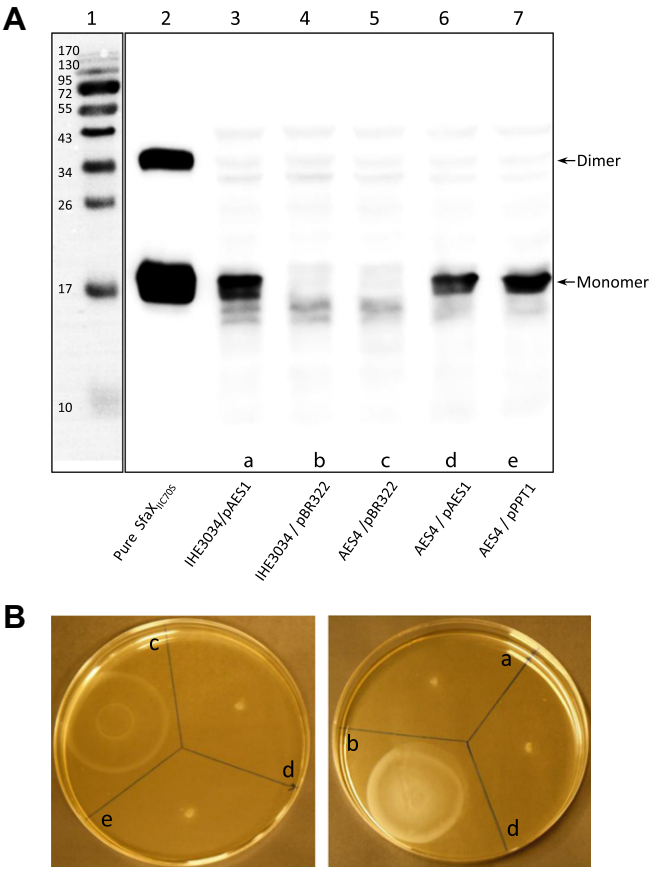

Fig. 3. Phenotypic tests of strains expressing the $SfaX_{IIC70S}$ protein. (A) Western blot analysis of $SfaX_{II}$ and $SfaX_{IIC70S}$ expression level. Lane1, molecular mass marker; Lane 2, pure SfaXIIC70S, control; Lane 3, IHE3034/pAES1 "a"; Lane 4, IHE3034/pBR322 "b"; Lane 5, AES4/pBR322 "c"; Lane 6, AES4/pAES1 "d"; Lane 7, AES4/pPPT1 "e". Bacteria were grown in LB at 37 °C to early stationary phase and samples were prepared to examine levels of proteins. (B) Analysis of motility of strains "a, b, c, d and e" corresponding to IHE3034/pAES1, IHE3034/pBR322, AES4/pBR322, AES4/pAES1 and AES4/pPPT1 respectively. Strains were growing on plates containing 0.3% agar at 37 °C.

aspects of the expression protocol; none of these attempts yielded sufficient amounts of homogeneous and pure protein. The use of an autoinduction medium did not change the expression level of the protein, although it did significantly improve the solubility [28,44]. The advantage of this medium is that it induces automatic and gradual protein expression, unlike the traditional method of adding IPTG. In the autoinduction medium, stringent control of the promoter in the early phase of growth occurs via catabolite repression by glucose. Glycerol and lactose are used as additional carbon sources and P*lac*-controlled protein production is induced when the bacterial cells have consumed the glucose. In addition, this medium contains 23 mM of ammonium sulfate, which generally helps promote solubilization of DNA-binding proteins. After cleavage of H6-ZZ with the TEV protease (Fig. 4A), gel filtration of $SfaX_{IIC70S}$ yielded a peak corresponding to the size of the dimeric protein (Fig. 4B). Protein purity was verified by SDS–PAGE (Fig. 4C). Typically, we obtain ~10 mg of pure protein from expression in four liters of autoinduction media.

In order to investigate the structural properties of the protein and its thermostability, we recorded far-UV circular-dichroism spectra. The observed spectrum (Fig. 5A) is characteristic of a protein with a significant fraction of α-helical secondary structure; by using the program CDNN, the secondary structure content was estimated to be 37% α-helix, 15% β-strand, 16% β-turn and 29% random coil. The thermal unfolding and refolding experiments (Fig. 5B) show that the protein unfolds irreversibly into an aggregation-prone random coil/β-sheet conformation, with a melting point of 51 °C.

We also investigated the DNA-binding properties of the $SfaX_{IIC70S}$ protein. Electrophoretic mobility shift assays confirmed DNA binding to the 300-bp fragment of the *sfa* operon regulatory region. We have not characterized in detail the binding specificity of $SfaX_{IIC70S}$ (Fig. 6A), however the mobility shift results are

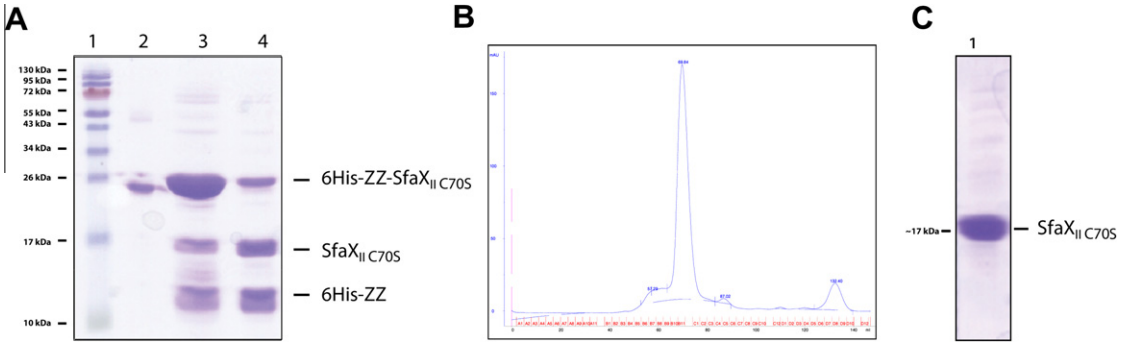

Fig. 4. Analysis of purified SfaX$_{IIC70S}$ preparations. (A) SDS–PAGE showing samples obtained after TEV digestion. Lane 1, molecular mass markers; Lane 2, TEV protease; Lane 3, partial digestion of H6-ZZ-SfaX$_{IIC70S}$ by TEV protease in buffer *A*; Lane 4, protein sample after ON cleavage with TEV protease in buffer *A* containing 1 mM DTT and 0.5 mM EDTA. (B) FPLC gel filtration on HiLoad 16/60 Superdex 75 prep grade column. Sample: SfaX$_{IIC70S}$. Buffer: Sodium phosphate buffer (pH 7.4), 300 mM NaCl, 23 mM ammonium sulfate. Fraction size: 1 mL. Flow rate: 1 mL/min. (C) SDS–PAGE analysis of the final purification stage of SfaX$_{IIC70S}$. Fractions were analyzed by 15% SDS–PAGE and stained with Coomassie Blue.

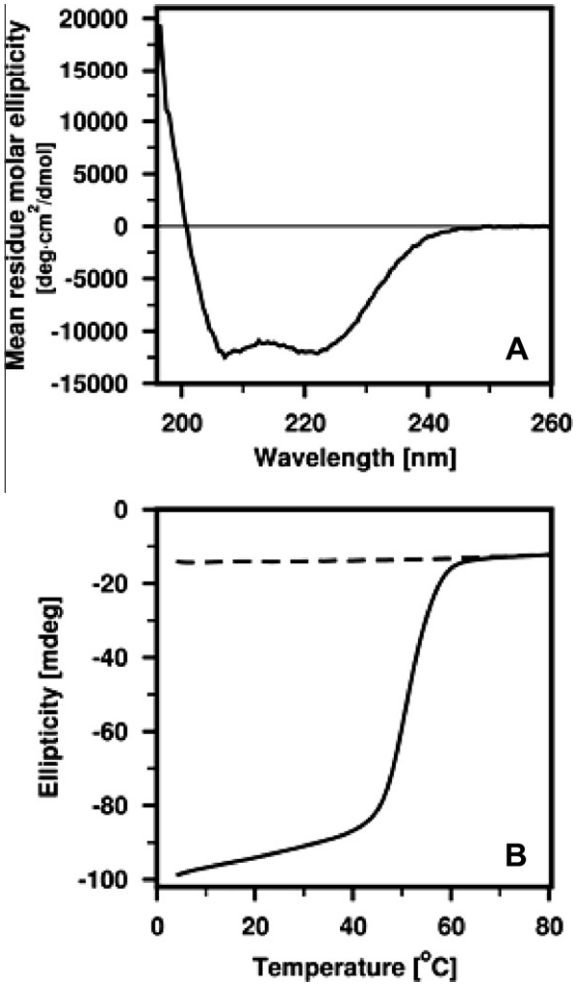

Fig. 5. Circular Dichroism. (A) Far-UV CD spectrum of a sample containing 10 μM SfaXII$_{IIC70S}$. The spectrum was analyzed using the CDNN program [31] to give estimates of 37% α-helix, 15% β-sheet, 16% β-turn and 29% random coil. (B) Thermal unfolding (solid line) and refolding (dashed line) of the protein was investigated by monitoring the ellipticity at 220 nm in a temperature range of 4–80 °C.

consistent with findings from single-molecule analysis by atomic force microscopy. Compared to images of the mica surface with only DNA (Fig. 6B) we see that, after incubation with SfaX$_{IIC70S}$, the DNA was partially bound by the protein (figure 6C).

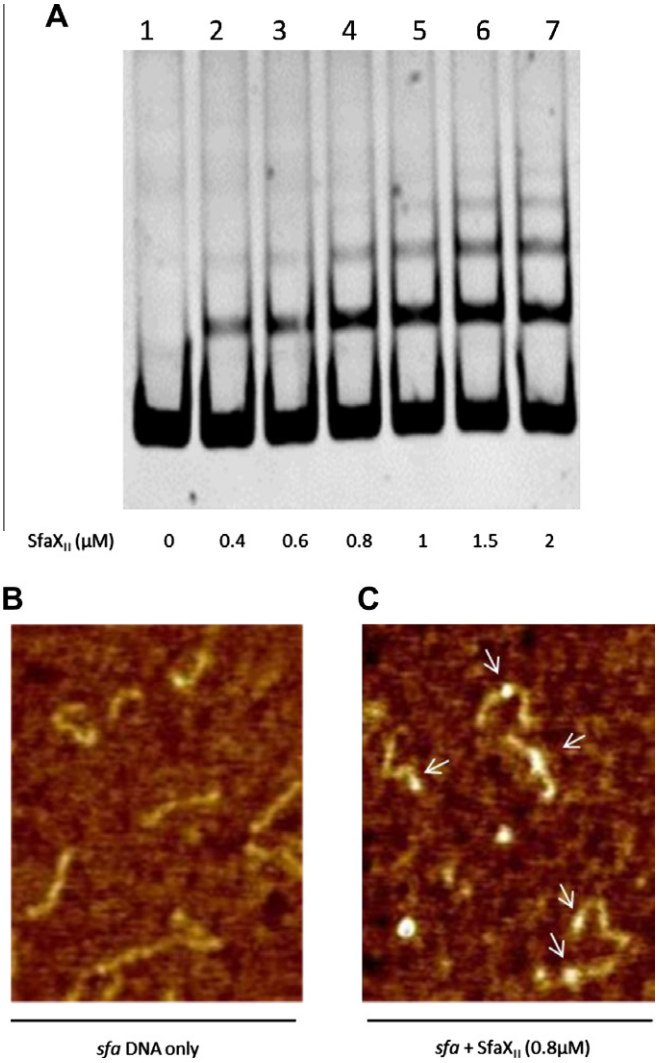

Fig. 6. *In vitro* DNA binding analyses with SfaX$_{IIC70S}$. (A) Gel mobility shift assays with a gradient of purified SfaX$_{IIC70S}$ protein from 0.4 to 2 μM with a 300-bp fragment of DNA from the *sfa* operon. Lane 1 is the negative control containing only DNA. (B) Representative atomic force micrographs of the 300-bp fragment of DNA from the *sfa* operon. (C) Atomic force micrographs showing the binding of SfaX$_{IIC70S}$ to the DNA fragment. The area of DNA that was bound by the SfaX$_{IIC70S}$ protein is indicated by arrows. SfaX$_{IIS70S}$ (0.8 μM) was incubated with 50 ng of DNA on the mica at 37 °C.

Conclusions

In this study, we created a site-directed mutant of $SfaX_{II}$, a transcriptional regulator involved in ExPEC fimbriae formation, and developed an efficient protocol for expression and purification of the recombinant protein. Insolubility of the native protein prompted us to analyze a molecular model of $SfaX_{II}$, which suggested that a Cys70Ser substitution could improve protein solubility and *in vivo* studies indicated that this substitution is not affecting the $SfaX_{II}$ regulatory function within the cell. In agreement with this, the $SfaX_{IIC70S}$ mutant showed a fivefold increase in solubility, presumably by avoiding non-native intermolecular disulfide bonds. Autoinduction medium as an alternative to standard induction by addition of IPTG, combined with an H6-ZZ carrier fused to the N-terminal end of the protein, gave us a high yield of soluble protein: four liters of autoinduction media typically yielded ~10 mg of pure protein.

The structural properties and thermostability of purified $SfaX_{IIC70S}$ were determined by far-UV CD spectroscopy and showed that the protein has a predominant α-helical secondary structure, in agreement with our molecular model, and a melting temperature of 51 °C. Gel shift electrophoretic mobility assays as well as atomic force microscopy showed that the protein binds DNA at the regulatory region of the *sfa* operon.

The approach described herein might be applicable to most members of the family of 17-kDa proteins whose genes are associated with fimbriae operons (e.g., PapX, FocX, PrsX and PrfX proteins) due to their highly conserved sequences. In summary, this robust protocol for expression and purification of $SfaX_{IIC70S}$, as well as our molecular model, open new perspectives and possibilities to better understand the functionality and specificity in terms of protein–DNA interactions of this family of proteins whose genes and localization near operons involved in fimbriae expression are highly conserved in ExPEC.

Acknowledgments

This work was financially supported by the Swedish Research Council and the Kempe Foundations. We express our gratitude to Prof. Gunter Stier, Mrs. Christin Grundström, Dr. Annika E. Sjöström, Dr. Tobias Hainzl and Mr. Aaron Dhaya Edwin for their interest and valuable discussions. We thank Mrs. Monica Persson for skilful technical assistance with atomic force microscopy.

References

[1] J.R. Johnson, T.A. Russo, Extraintestinal pathogenic *Escherichia coli*: "the other bad *E. coli*", J. Lab. Clin. Med. 139 (2002) 155–162.
[2] J.L. Smith, P.M. Fratamico, N.W. Gunther, Extraintestinal pathogenic *Escherichia coli*, Foodborne Pathog. Dis. 4 (2007) 134–163.
[3] E.M. Antao, L.H. Wieler, C. Ewers, Adhesive threads of extraintestinal pathogenic *Escherichia coli*, Gut Pathog. 1 (2009) 22.
[4] N.J. Holden, D.L. Gally, Switches, cross-talk and memory in *Escherichia coli* adherence, J. Med. Microbiol. 53 (2004) 585–593.
[5] U. Dobrindt, B. Hochhut, U. Hentschel, J. Hacker, Genomic islands in pathogenic and environmental microorganisms, Nat. Rev. Microbiol. 2 (2004) 414–424.
[6] J. Hacker, G. Blum-Oehler, B. Hochhut, U. Dobrindt, The molecular basis of infectious diseases: pathogenicity islands and other mobile genetic elements. a review, Acta Microbiol. Immunol Hung. 50 (2003) 321–330.
[7] S. Lindberg, Y. Xia, B. Sonden, M. Goransson, J. Hacker, B.E. Uhlin, Regulatory Interactions among adhesin gene systems of uropathogenic *Escherichia coli*, Infect. Immun. 76 (2008) 771–780.
[8] M. Castelain, A.E. Sjostrom, E. Fallman, B.E. Uhlin, M. Andersson, Unfolding and refolding properties of S pili on extraintestinal pathogenic *Escherichia coli*, Eur. Biophys. J 39 (2010) 1105–1115.
[9] A.E. Sjostrom, B. Sonden, C. Muller, A. Rydstrom, U. Dobrindt, S.N. Wai, B.E. Uhlin, Analysis of the sfaX(II) locus in the *Escherichia coli* meningitis isolate IHE3034 reveals two novel regulatory genes within the promoter-distal region of the main S fimbrial operon, Microb. Pathog. 46 (2009) 150–158.
[10] T.K. Korhonen, M.V. Valtonen, J. Parkkinen, V. Vaisanen-Rhen, J. Finne, F. Orskov, I. Orskov, S.B. Svenson, P.H. Makela, Serotypes, hemolysin production, and receptor recognition of *Escherichia coli* strains associated with neonatal sepsis and meningitis, Infect. Immun. 48 (1985) 486–491.
[11] A.E. Sjostrom, C. Balsalobre, L. Emody, B. Westerlund-Wikstrom, J. Hacker, B.E. Uhlin, The SfaXII protein from newborn meningitis *E. coli* is involved in regulation of motility and type 1 fimbriae expression, Microb. Pathog. 46 (2009) 243–252.
[12] N. Buchmeier, S. Bossie, C.Y. Chen, F.C. Fang, D.G. Guiney, S.J. Libby, SlyA, a transcriptional regulator of *Salmonella typhimurium*, is required for resistance to oxidative stress and is expressed in the intracellular environment of macrophages, Infect. Immun. 65 (1997) 3725–3730.
[13] E. Perez-Rueda, J. Collado-Vides, L. Segovia, Phylogenetic distribution of DNA-binding transcription factors in bacteria and archaea, Comput. Biol. Chem. 28 (2004) 341–350.
[14] S.P. Wilkinson, A. Grove, Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins, Curr. Issues Mol. Biol. 8 (2006) 51–62.
[15] M.N. Alekshun, S.B. Levy, T.R. Mealy, B.A. Seaton, J.F. Head, The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 A resolution, Nat. Struct. Biol. 8 (2001) 710–714.
[16] T. Kumarevel, T. Tanaka, T. Umehara, S. Yokoyama, ST1710-DNA complex crystal structure reveals the DNA binding mechanism of the MarR family of regulators, Nucleic Acids Res. 37 (2009) 4723–4735.
[17] M. Hong, M. Fuangthong, J.D. Helmann, R.G. Brennan, Structure of an OhrR–ohrA operator complex reveals the DNA binding mechanism of the MarR family, Mol. Cell 20 (2005) 131–141.
[18] V. Saridakis, D. Shahinas, X. Xu, D. Christendat, Structural insight on the mechanism of regulation of the MarR family of proteins: high-resolution crystal structure of a transcriptional repressor from *Methanobacterium thermoautotrophicum*, J. Mol. Biol. 377 (2008) 655–667.
[19] A.N. Simms, H.L. Mobley, PapX, a P fimbrial operon-encoded inhibitor of motility in uropathogenic *Escherichia coli*, Infect. Immun. 76 (2008) 4833–4841.
[20] D.J. Reiss, H.L. Mobley, Determination of target sequence bound by PapX motility, in flhD promoter using systematic evolution of ligands by exponential enrichment (SELEX) and high throughput sequencing, J. Biol. Chem. 286 (2011) 44726–44738.
[21] N. Eswar, B. Webb, M.A. Marti-Renom, M.S. Madhusudhan, D. Eramian, M.Y. Shen, U. Pieper, A. Sali, Comparative protein structure modeling using MODELLER, Curr. Protoc. Protein Sci. (2007) (Chapter 2, Unit 2 9).
[22] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The CLUSTAL_X windows interface. flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res. 25 (1997) 4876–4882.
[23] S.C. Lovell, I.W. Davis, W.B. Arendall 3rd, P.I. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson, D.C. Richardson, Structure validation by Calpha geometry: phi, psi and Cbeta deviation, Proteins 50 (2003) 437–450.
[24] L. Holm, J. Park, DaliLite workbench for protein structure comparison, Bioinformatics 16 (2000) 566–567.
[25] W.L. DeLano, A.T. Brunger, The direct rotation function: patterson correlation search applied to molecular replacement, Acta Crystallogr. D: Biol. Crystallogr. 51 (1995) 740–748.
[26] P. Gouet, E. Courcelle, D.I. Stuart, F. Metoz, ESPript: analysis of multiple sequence alignments in PostScript, Bioinformatics 15 (1999) 305–308.
[27] C. Cole, J.D. Barber, G.J. Barton, The Jpred 3 secondary structure prediction server, Nucleic Acids Res. 36 (2008) W197–201.
[28] Z. Li, W. Kessler, J. van den Heuvel, U. Rinas, Simple defined autoinduction medium for high-level recombinant protein production using T7-based *Escherichia coli* expression systems, Appl. Microbiol. Biotechnol. 91 (2011) 1203–1213.
[29] J. Bogomolovas, B. Simon, M. Sattler, G. Stier, Screening of fusion partners for high yield expression and purification of bioactive viscotoxins, Protein Expr. Purif. 64 (2009) 16–23.
[30] S. Nallamsetty, R.B. Kapust, J. Tozser, S. Cherry, J.E. Tropea, T.D. Copeland, D.S. Waugh, Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease *in vivo* and *in vitro*, Protein Expr. Purif. 38 (2004) 108–115.
[31] G. Bohm, R. Muhr, R. Jaenicke, Quantitative analysis of protein far UV circular dichroism spectra by neural networks, Protein Eng. 5 (1992) 191–195.
[32] L.D. Kerr, Electrophoretic mobility shift assay, Methods Enzymol. 254 (1995) 619–632.
[33] Y. Xia, K. Forsman, J. Jass, B.E. Uhlin, Oligomeric interaction of the PapB transcriptional regulator with the upstream activating region of pili adhesin gene promoters in *Escherichia coli*, Mol. Microbiol. 30 (1998) 513–523.
[34] M. Uhlen, G. Forsberg, T. Moks, M. Hartmanis, B. Nilsson, Fusion proteins in biotechnology, Curr. Opin. Biotechnol. 3 (1992) 363–369.
[35] D.B. Smith, K.S. Johnson, Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase, Gene 67 (1988) 31–40.
[36] H. Bedouelle, P. Duplay, Production in *Escherichia coli* and one-step purification of bifunctional hybrid proteins which bind maltose. Export of the Klenow polymerase into the periplasmic space, Eur. J. Biochem. 171 (1988) 541–549.
[37] E.R. LaVallie, E.A. DiBlasio, S. Kovacic, K.L. Grant, P.F. Schendel, J.M. McCoy, A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm, Biotechnology (NY) 11 (1993) 187–193.
[38] G.D. Davis, C. Elisee, D.M. Newham, R.G. Harrison, New fusion protein systems designed to give soluble expression in *Escherichia coli*, Biotechnol. Bioeng. 65 (1999) 382–388.

[39] A. de Marco, Two-step metal affinity purification of double-tagged (NusA-His6) fusion proteins, Nat. Protoc. 1 (2006) 1538–1543.
[40] J.R. Huth, C.A. Bewley, B.M. Jackson, A.G. Hinnebusch, G.M. Clore, A.M. Gronenborn, Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR, Protein Sci. 6 (1997) 2359–2364.
[41] J.A. Brockelbank, V. Peters, B.H. Rehm, Recombinant *Escherichia coli* strain produces a ZZ domain displaying biopolyester granules suitable for immunoglobulin G purification, Appl. Environ. Microbiol. 72 (2006) 7394–7397.
[42] M. Hammarstrom, N. Hellgren, S. van Den Berg, H. Berglund, T. Hard, Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*, Protein Sci. 11 (2002) 313–321.
[43] T. Moks, L. Abrahmsen, B. Nilsson, U. Hellman, J. Sjoquist, M. Uhlen, Staphylococcal protein A consists of five IgG-binding domains, Eur. J. Biochem. 156 (1986) 637–643.
[44] F.W. Studier, Protein production by auto-induction in high density shaking cultures, Protein Expr. Purif. 41 (2005) 207–234.
[45] F. Bolivar, R.L. Rodriguez, P.J. Greene, M.C. Betlach, H.L. Heyneker, H.W. Boyer, J.H. Crosa, S. Falkow, Construction and characterization of new cloning vehicles II. A multipurpose cloning system, Gene 2 (1977) 95–113.
[46] Y. Zhao, Y. Benita, M. Lok, B. Kuipers, P. van der Ley, W. Jiskoot, W.E. Hennink, D.J. Crommelin, R.S. Oosting, Multi-antigen immunization using IgG binding domain ZZ as carrier, Vaccine 23 (2005) 5082–5090.
[47] P.A. Nygren, S. Stahl, M. Uhlen, Engineering proteins to facilitate bioprocessing, Trends Biotechnol. 12 (1994) 184–188.
[48] K.D. Pryor, B. Leiting, High-level expression of soluble protein in *Escherichia coli* using a His6-tag and maltose-binding-protein double-affinity fusion system, Protein Expr. Purif. 10 (1997) 309–319.
[49] P. Jurado, V. de Lorenzo, L.A. Fernandez, Thioredoxin fusions increase folding of single chain Fv antibodies in the cytoplasm of *Escherichia coli*: evidence that chaperone activity is the prime effect of thioredoxin, J. Mol. Biol. 357 (2006) 49–61.
[50] D.A. Lindhout, A. Thiessen, D. Schieve, B.D. Sykes, High-yield expression of isotopically labeled peptides for use in NMR studies, Protein Sci. 12 (2003) 1786–1791.