Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Purification, crystallization and preliminary data analysis of FocB, a transcription factor regulating fimbrial adhesin expression in uropathogenic *Escherichia coli*

Ulrika W. Hultdin,[a] Stina Lindberg,[b] Christin Grundström,[a] Anders Allgardsson,[a,b] Shenghua Huang,[a] Gunter Stier,[a,c] Anders Öhman,[a] Bernt Eric Uhlin[b]* and A. Elisabeth Sauer-Eriksson[a]*

[a]Department of Chemistry, Umeå University, SE-901 87 Umeå, Sweden, [b]Department of Molecular Biology, The Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, SE-901 87 Umeå, Sweden, and [c]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, DE-69117 Heidelberg, Germany

Correspondence e-mail: bernt.eric.uhlin@molbiol.umu.se, elisabeth.sauer-eriksson@chem.umu.se

Received 4 November 2009
Accepted 18 January 2010

The transcription factor FocB belongs to a family of regulators encoded by several different fimbriae gene clusters in uropathogenic *Escherichia coli*. Recent findings suggest that FocB-family proteins may form different protein–protein complexes and that they may exert both positive and negative effects on the transcription of fimbriae genes. However, little is known about the actual role and mode of action when these proteins interact with the fimbriae operons. The 109-amino-acid FocB transcription factor from the *foc* gene cluster in *E. coli* strain J96 has been cloned, expressed and purified. The His_6-tagged fusion protein was captured by Ni^{2+}-affinity chromatography, cleaved with tobacco etch virus protease and purified by gel filtration. The purified protein is oligomeric, most likely in the form of dimers. NMR analysis guided the crystallization attempts by showing that probable conformational exchange or oligomerization is reduced at temperatures above 293 K and that removal of the highly flexible His_6 tag is advantageous. The protein was crystallized using the hanging-drop vapour-diffusion method at 295 K. A native data set to 2.0 Å resolution was collected at 100 K using synchrotron radiation.

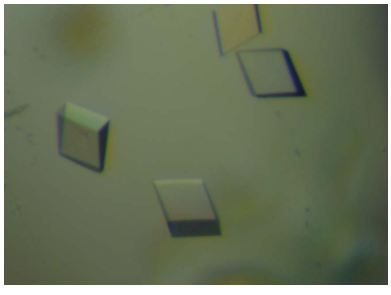
1. Introduction

Pathogenic isolates of the bacterium *Escherichia coli* commonly express surface organelles (fimbriae) that are responsible for adhesion to specific receptors in host tissue (Hacker, 1992). Typical isolates of uropathogenic *E. coli* (UPEC) carry multiple sets of fimbriae genes; it is presumably advantageous for the bacterium to express only the adhesins that interact with the specific receptors that a particular host cell expresses, thereby evoking a minimum of the host immune defence responses (Holden & Gally, 2004). The expression of different fimbrial adhesins by UPEC is regulated at the transcriptional level and the PapB/FocB family of transcription factors has been found to play a key role in the regulatory cross-talk among fimbriae-gene systems in typical UPEC isolates such as *E. coli* strain J96 (Xia *et al.*, 2000; Lindberg *et al.*, 2008).

FocB is a transcription factor in the *foc* gene cluster that is responsible for F1C fimbriae production in UPEC strains. This type of fimbriae mediate specific adherence to the collecting ducts and distal tubules of the kidney and are often expressed in connection with type 1 and P fimbriae encoded by the *fim* and *pap* genes, respectively (Klemm *et al.*, 1982; Virkola *et al.*, 1988).

The organization of the *foc* and *pap* gene clusters is very similar and the clusters include structural genes for the fimbriae subunits, the assembly machinery and the factors involved in control of expression (Ott *et al.*, 1987, 1988; Sauer *et al.*, 2000).

The transcription factors FocB and PapB (encoded by the *papB* gene in the *pap* gene clusters) are representative proteins of the growing family of fimbriae regulatory proteins. Members of this family have been identified from *E. coli* and *Salmonella* strains and are small proteins that share sequence homology and consensus residues for DNA binding and oligomerization (Xia & Uhlin, 1999; Holden *et al.*, 2001). In order to understand the mode of action and regulatory mechanism of these transcription factors, detailed information about their structure–function properties is needed. In this



paper, we describe the cloning, overexpression, purification, crystallization and preliminary X-ray diffraction studies of *E. coli* FocB.

2. Materials and methods

2.1. Cloning

focB DNA was PCR-amplified from vector pYN89 (Lindberg *et al.*, 2008) using forward primer 5′-ATCACCATGGCACAGCATGAAGTTATTACCCG-3′ and reverse primer 5′-AGATGGTACCTTATTACAGGGAAGCAGCTTCAG-3′. In order to create an FocB protein with a removable N-terminal His tag, the PCR product was digested with *Nco*I and *Acc*65I and ligated into the equivalent sites of the pETM11 expression vector. The plasmids were transformed into *E. coli* DH5α and transformants were selected on kanamycin plates. The positive clones were verified by DNA sequencing.

2.2. Protein expression and purification

Competent *E. coli* BL21(DE3)pLysS (Novagen) cells were transformed with the FocB-pETM11 plasmid and grown overnight on LB agar plates supplemented with kanamycin (30 µg ml^{-1}). A preculture of 10 ml LB medium supplemented with 100 µg ml^{-1} kanamycin was inoculated and grown at 310 K overnight with orbital shaking. For expression, a total of 4 l LB medium supplemented with 100 µg ml^{-1} kanamycin was inoculated with 2.5 ml of preculture per litre and grown with orbital shaking at 310 K until the A_{600} reached 0.5. At this point, protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.2 m*M*. After 3 h of expression, the cells were harvested by centrifugation and washed in 10 m*M* HEPES pH 7.9 and 0.5 *M* NaCl (buffer *A*) containing 5 m*M* imidazole before freezing and storage at 253 K. Uniformly ^{15}N isotope-labelled FocB protein was prepared according to the protocol above by growing *E. coli* BL21(DE3)pLysS-FocB in 1 l M9 minimal medium containing $^{15}NH_4Cl$ (Cambridge Isotopes Laboratories) as the sole source of nitrogen.

Thawing of the cells was carried out at 288 K, while all subsequent purification steps were carried out at room temperature. Thawed cells were resuspended in buffer *A* containing 5 m*M* imidazole and lysed by sonication on ice. Cellular debris was removed by centrifugation at 53 000*g* for 30 min. The supernatant was loaded onto a 5 ml Ni–NTA agarose column (Qiagen) pre-equilibrated with buffer *A* containing 5 m*M* imidazole. After washing the column in buffer *A* containing 50 m*M* imidazole, FocB was eluted in 1 ml fractions with buffer *A* containing 300 m*M* imidazole. The protein-containing fractions were pooled and EDTA and β-mercaptoethanol were added to final concentrations of 5 m*M* and 0.1%(*v*/*v*), respectively. The concentration of the protein solution was determined using the Bradford method (Bradford, 1976; Thermo Scientific) and the total yield of protein was estimated to be 90 mg. To remove the His_6 tag, 1 mg tobacco etch virus (TEV) protease was added to the FocB solution. The cleavage reaction was performed for 20 h at room temperature. The heavy precipitate, which included most of the TEV protease, was spun down at 16 000*g* for 20 min and the supernatant was filtered through a 0.22 µm syringe filter (Millipore) and loaded onto a Superdex 75 gel-filtration column (GE Healthcare) equilibrated with buffer *A* containing 5 m*M* EDTA and 0.1%(*v*/*v*) β-mercaptoethanol. The peak fractions of FocB were pooled and concentrated to a final concentration of 15 mg ml^{-1} using a Centriprep-10 centrifugal concentrator (Millipore), with a total yield of approximately 15 mg pure protein. The purified protein was stored at 277 K.

2.3. Cross-linking of purified protein

In vitro cross-linking was performed with purified protein using the cross-linking agent glutaraldehyde (Lindberg *et al.*, 2008). 250 µ*M* purified FocB protein was incubated in two concentrations (0.025% and 0.05%) of glutaraldehyde at room temperature or 310 K for 10 min. The reaction was stopped by the addition of 1 *M* Tris–HCl pH 8.0. The samples were analyzed by SDS–PAGE and the protein sizes were detected by immunoblotting using antiserum raised against the FocB protein.

2.4. CD and NMR analysis of purified FocB protein

Far-UV circular-dichroism (CD) spectra were collected at 295 K on samples containing 16.6 µ*M* FocB in 10 m*M* sodium phosphate buffer pH 6.5, 200 m*M* NaCl using a JASCO J-810 spectropolarimeter. Spectra were recorded between 195 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 s^{-1} in a 1 mm quartz cuvette.

NMR spectra were recorded at temperatures from 283 to 315 K on a Bruker DRX600 spectrometer equipped with a 5 mm triple-resonance *z*-gradient cryo-probe. Samples contained approximately 0.3 m*M* protein in a solvent consisting of either 10 m*M* sodium acetate buffer pH 5.5 or 10 m*M* HEPES buffer pH 7.5. Both buffers also contained 500 m*M* NaCl, 5 m*M* EDTA and 10 m*M* DTT. Acquisition, processing and analysis were carried out in *TOPSPIN* (Bruker Biospin).

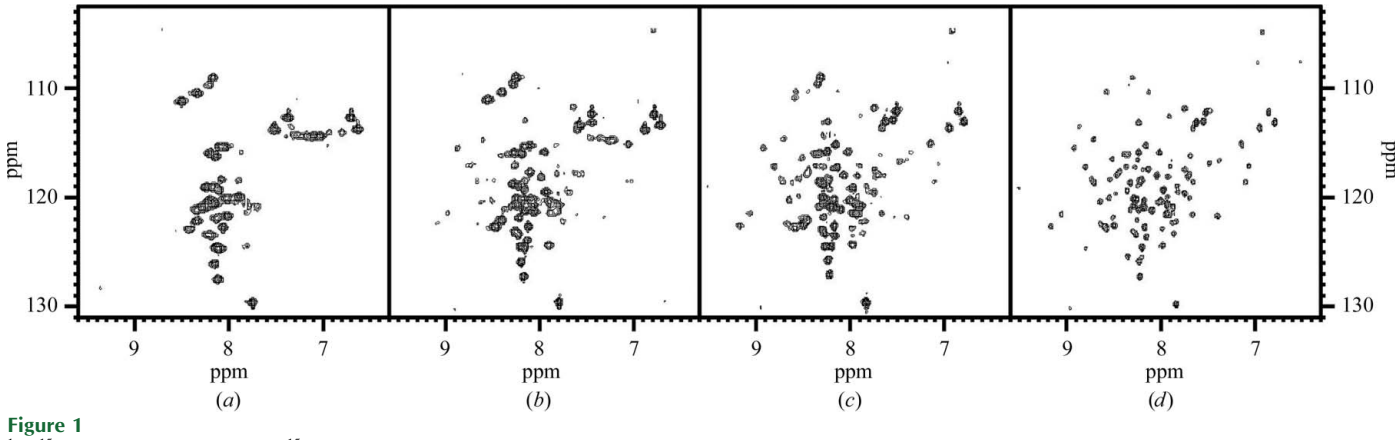


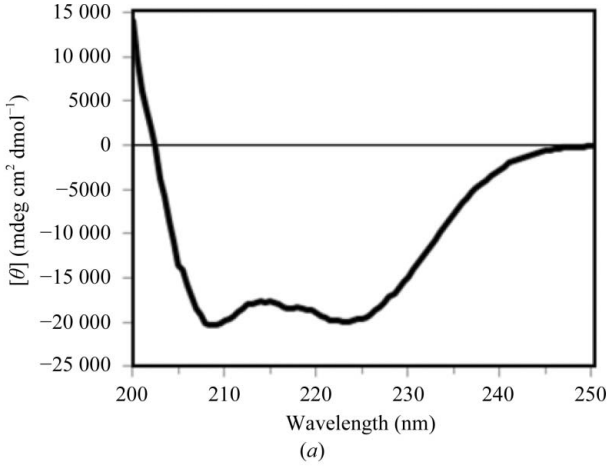
Figure 1
^{1}H–^{15}N HSQC spectra of 0.3 m*M* ^{15}N-labelled FocB (*a*)–(*c*) with and (*d*) without His_6 tag recorded at 600 MHz at (*a*) 283 K, (*b*) 298 K and (*c*, *d*) 310 K.

2.5. Crystallization

Initial crystallization trials were performed by the sitting-drop vapour-diffusion method in 96-well SD-2 (Innovadyne Technologies Inc.) crystallization plates. Using a Mosquito (TTP Labtech) nanodrop crystallization robot, 0.1 µl droplets of protein solution were mixed with an equal volume of reservoir solution. Several crystal hits were obtained in crystallization screens from Hampton Research (Crystal Screen and Crystal Screen II). The reservoir composition was optimized using the hanging-drop vapour-diffusion technique. For optimization, droplets of 1–2 µl protein solution were mixed with an equal volume of reservoir solution in XRL plates (Molecular Dimensions). The largest and best diffracting crystals were obtained in 1.5–1.75 *M* $MgSO_4$ and 0.1 *M* MES pH 6.5 at 295 K and grew to dimensions of 0.05 × 0.02 × 0.01 mm within two weeks.

2.6. Data collection and reduction

Crystals were flash-cooled in liquid nitrogen at 100 K without the addition of cryoprotectant. Diffraction data were collected on beamline ID23-1 at the ESRF to a maximum resolution of 2.0 Å. A total of 180 frames of data were collected with an oscillation angle of 1°. The exposure time was 0.5 s per frame and the crystal-to-detector distance was 175 mm. The diffraction data were processed with *XDS* (Kabsch, 1993) and scaled with *SCALA* from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). A summary of the data-collection statistics is given in Table 1. Calculation of the self-rotation function was carried out using the program *MOLREP* (Collaborative Computational Project, Number 4, 1994).



(*a*)

```
1
MAQHEVITRG GDAFLLKLRE SALSSGSMSE EQFFLLIGIS
---HHHHH-- -HHHHHHH-- ---------H HHHHHHH---

41
SIHSDRVILA MKDYLVSGHS RKDVCEKYQM NNGYFSTTLG
----HHHHHH HHHHH----- HHHHHHH--- ---HHHHHHH

91
RLTRLNVLVA RLAPYYTDSV SAIAEAASL
HHHHHHHHHH HHHHHHH--- --HHHH---
```

(*b*)

Figure 2
(*a*) Far-UV CD spectrum of a sample containing 16.6 µ*M* FocB. The spectrum was analyzed using the *CDNN* program (Bohm *et al.*, 1992) to give estimates of 65% α-helix, 6.5% β-sheet, 12% β-turn and 16.5% random coil. (*b*) Prediction of the secondary structure of FocB by the program *JPRED* (Cole *et al.*, 2008). The predicted α-helices (H) are shown below the sequence.

Table 1
Data-collection statistics.

Values in parentheses are for the highest resolution shell.

Wavelength (Å)	0.975
Space group	$P2_12_12_1$
Unit-cell parameters (Å, °)	a = 46.82, b = 58.30, c = 65.41, $\alpha = \beta = \gamma$ = 90.00
Resolution range (Å)	29.15–2.00 (2.11–2.00)
No. of observations	98159
No. of unique reflections	12636
Redundancy	7.8 (7.9)
Completeness (%)	99.9 (100.0)
R_{merge}†	0.056 (0.298)
$I/\sigma(I)$	23.6 (6.9)
No. of molecules per ASU	2
Matthews coefficient (Å^3 Da^{-1})	3.65 [1.82 with 2 molecules in the ASU]

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I_i(hkl)\rangle$ is the mean intensity of the i observations of reflection hkl.

3. Results and discussion

His-tagged full-length FocB from *E. coli* (theoretical molecular weight 12.2 kDa) could be overexpressed in *E. coli* BL21(DE3) cells and was purified by Ni–NTA agarose and gel-filtration chromatography. The His_6-tagged protein crystallized in ammonium sulfate but the crystals were of poor quality. Further manipulation of the crystallization conditions, including the use of additives, seeding and crystallization at different temperatures, did not improve the crystals. NMR analysis revealed the presence of a group of intense resonances positioned at characteristic random-coil shifts, strongly suggesting that part of the protein, presumably the His_6 tag, was unfolded (Figs. 1*a*–1*c*). Apart from these very intense resonances, the majority of residues displayed broad resonances below a temperature of approximately 293 K (Fig. 1*a*), most likely owing to conformational exchange or possibly owing to oligomerization. On increasing the temperature to above 293 K the resonances became significantly narrower and several new peaks started to appear (Figs. 1*b* and 1*c*), suggesting that FocB attained a more homogeneous well folded conformation. In view of these NMR results, a new construct of the full-length FocB with a removable His_6 tag was produced. The protein was overexpressed in *E. coli* BL21(DE3)pLysS and purified by Ni–NTA agarose chromatography. The His_6 tag was cleaved off by TEV

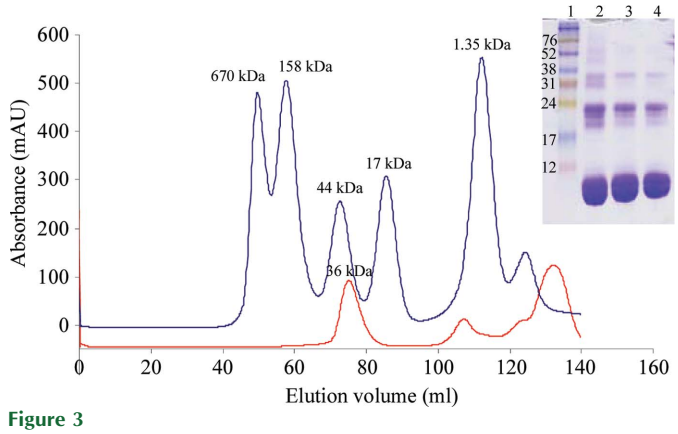


Figure 3
Gel-filtration (Superdex 75, 120 ml) chromatogram of FocB. Molecular-weight markers are shown as a blue line and include vitamin B_{12} (1.35 kDa), myoglobin (horse; 17 kDa), ovalbumin (chicken; 44 kDa), γ-globulin (bovine; 158 kDa) and thyroglobulin (bovine; 670 kDa). FocB elutes at 36 kDa (red line), which corresponds to a trimer. Inset: SDS–PAGE of purified FocB. Lane 1, molecular-weight markers (kDa); lane 2, unboiled sample without β-mercaptoethanol; lane 3, unboiled sample with β-mercaptoethanol; lane 4, boiled sample with β-mercaptoethanol.

protease, leaving two extra residues, Gly and Ala, followed by Met1 at the N-terminus. The protein was then purified to homogeneity by gel-filtration chromatography. Approximately 2.5 mg pure protein was obtained from 1 l culture. NMR analysis confirmed the removal of the His_6 tag and verified the improved structural homogeneity at temperatures above 293 K; the HSQC spectrum of FocB without the His_6 tag showed well dispersed resonances of similar intensity and was nearly identical to the spectrum of the tagged protein, apart from the expected lack of intense resonances from the flexible His_6 tag (compare Figs. 1*c* and 1*d*). The far-UV CD spectrum (Fig. 2*a*) showed that FocB has a high content of α-helical secondary structure, which was estimated as 65% using the CD deconvolution program *CDNN* (Bohm *et al.*, 1992). Secondary-structure prediction with *JPRED* (Cole *et al.*, 2008) also suggested that FocB is an all-helical structure, possibly comprising seven helices (Fig. 2*b*).

Gel-filtration chromatography showed that the protein was oligomeric and formed dimers or eventually trimers in solution (Fig. 3). However, it should be noted that the shape of protein molecules also plays an important role in gel filtration, since a gel-filtration column fractionates proteins on the basis of their Stokes radius not their molecular weight. Consequently, long extended polypeptides and proteins tend to behave as though they were larger globular protein molecules. SDS–PAGE analysis of FocB suggested that various types of dimer or trimer interfaces can form and that these forms are not dependent on disulfide-bridge formation (Fig. 3, inset). To further investigate the ability of the FocB protein to oligomerize, we performed *in vitro* cross-linking using two different concentrations of glutaraldehyde (GA). GA reacts predominantly with free amino groups, especially that of lysine, although the mechanism and chemistry involved in the GA cross-linking reaction is not yet fully understood. In the presence of GA cross-linker, the protein mainly formed complexes corresponding to the size of dimers, but higher oligomeric structures ranging from trimers to hexamers were also formed (Fig. 4*a*). Curiously, putative cross-linked dimers migrated in a ladder-like pattern. This behaviour is puzzling, but is most likely to be a consequence of the presence of different cross-linked dimeric products. Close to 40% of the oligomers were in the form of dimers, compared with 5–10% that were in the trimeric form (Fig. 4*b*). The tetrameric to hexameric forms were mainly observed when cross-linked at 310 K and comprised a minor protein fraction.

Since the structural behaviour of FocB improves at temperatures above 293 K, as revealed by NMR analysis, crystallization trials were only pursued at 295 K or higher. Crystals of FocB were obtained in 1.5–1.75 *M* $MgSO_4$ and 0.1 *M* MES pH 6.5 (Fig. 5). The crystals were very stable in the X-ray beam and diffracted to better than 2.0 Å resolution; they were thus suitable for high-resolution structure analysis (see Table 1 for the data-collection and data-processing statistics of a 2.0 Å data set). The crystals belonged to space group $P2_12_12_1$. A calculated Matthews coefficient of 3.65 $Å^3$ Da^{-1} suggested that there are most likely to be two molecules in the asymmetric unit. Calculated self-rotation functions showed no peak in the $\chi = 120^\circ$ section, which suggested that there was no threefold axis present in the crystal and that the packing of trimeric FocB molecules in the crystal was highly unlikely. The $\chi = 180^\circ$ section showed several weak peaks with heights less than 25% relative to the origin peak height. A noncrystallographic twofold axis could therefore be present in the crystal. Another possibility is that the noncrystallographic twofold axis is parallel to one of the twofold crystallographic axes.

Taken together, the biochemical data and, in particular, the results from the self-rotation function suggest that the protein is most likely to be dimeric both in solution and in the crystal, which is in agreement with most prokaryotic transcription factors. This is also in accordance with identified FocB oligomeric complexes formed in *in vivo*, where the stated oligomerization is predominantly dimeric (Lindberg *et al.*, 2008).

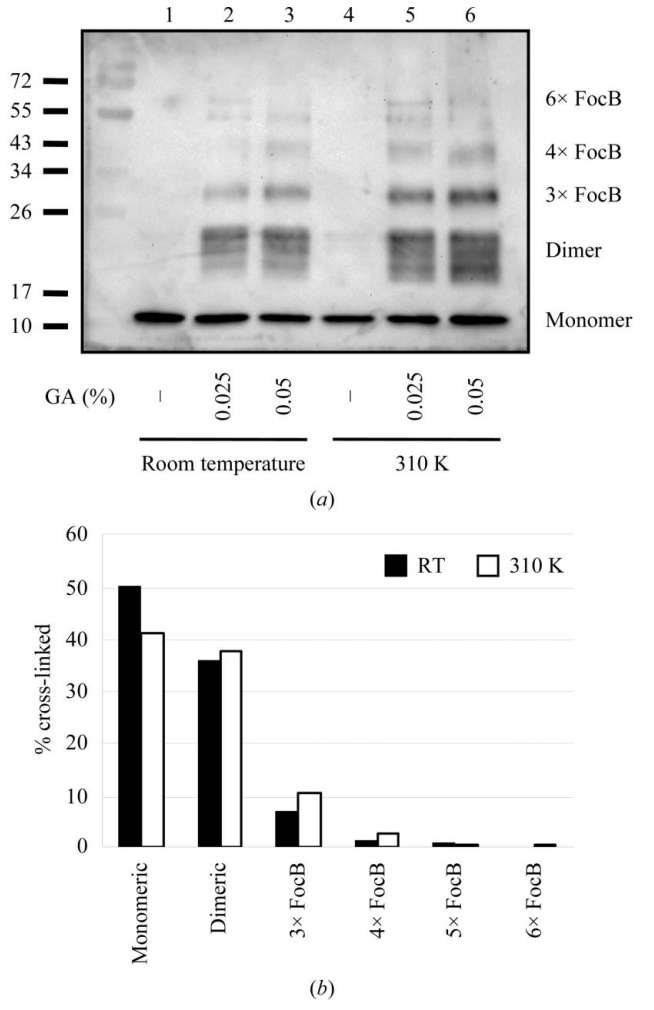


Figure 4
(*a*) *In vitro* cross-linking of FocB with glutaraldehyde. The protein bands were detected by immunoblotting using antisera raised against FocB. The positions of the molecular-weight markers are indicated (in kDa) as well as the different FocB complexes formed. (*b*) Quantification of the different FocB complexes in relation to the total protein level of FocB in the whole well. The black bars represent the cross-linking in lane 3 in (*a*) and the white bars represent the cross-linking in lane 6 in (*a*).

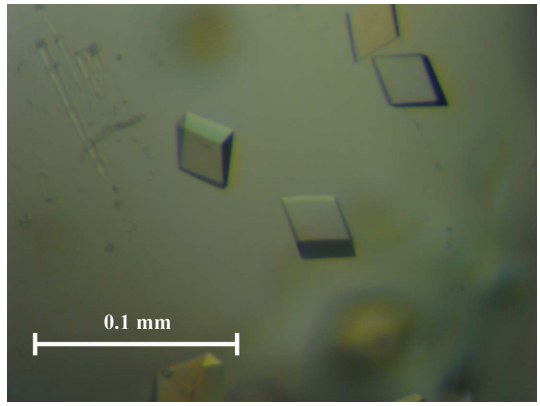


Figure 5
Crystals of *E. coli* FocB.

We expect to solve the structure by the SAD method and selenomethionine-labelled protein is currently being prepared.

We thank the European Synchrotron Radiation Facility (Grenoble, France) for provision of synchrotron-radiation facilities and Michael Murphy for critical reading of the manuscript. This work was performed within the Umeå Centre for Microbial Research (UCMR) and was supported by the Swedish Research Council, project K2007-66X-13001-09-3 and project 621-2005-3151.

References

Bohm, G., Muhr, R. & Jaenicke, R. (1992). *Protein Eng.* **5**, 191–195.
Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
Cole, C., Barber, J. D. & Barton, G. J. (2008). *Nucleic Acids Res.* **36**, W197–W201.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Hacker, J. (1992). *Can. J. Microbiol.* **38**, 720–727.
Holden, N. J. & Gally, D. L. (2004). *J. Med. Microbiol.* **53**, 585–593.
Holden, N. J., Uhlin, B. E. & Gally, D. L. (2001). *Mol. Microbiol.* **42**, 319–330.
Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
Klemm, P., Orskov, I. & Orskov, F. (1982). *Infect. Immun.* **36**, 462–468.
Lindberg, S., Xia, Y., Sonden, B., Goransson, M., Hacker, J. & Uhlin, B. E. (2008). *Infect. Immun.* **76**, 771–780.
Ott, M., Hoschutzky, H., Jann, K., Van Die, I. & Hacker, J. (1988). *J. Bacteriol.* **170**, 3983–3990.
Ott, M., Schmoll, T., Goebel, W., Van Die, I. & Hacker, J. (1987). *Infect. Immun.* **55**, 1940–1943.
Sauer, F. G., Mulvey, M. A., Schilling, J. D., Martinez, J. J. & Hultgren, S. J. (2000). *Curr. Opin. Microbiol.* **3**, 65–72.
Virkola, R., Westerlund, B., Holthofer, H., Parkkinen, J., Kekomaki, M. & Korhonen, T. K. (1988). *Infect. Immun.* **56**, 2615–2622.
Xia, Y., Gally, D., Forsman-Semb, K. & Uhlin, B. E. (2000). *EMBO J.* **19**, 1450–1457.
Xia, Y. & Uhlin, B. E. (1999). *J. Biol. Chem.* **274**, 19723–19730.