Maia De Kerpel,^a Inge Van Molle,^a Lea Brys,^b Lode Wyns,^a Henri De Greve^a and Julie Bouckaert^{a*}

^aDepartment of Ultrastructure, Vrije Universiteit Brussel (VUB), Flanders Interuniversity Institute for Biotechnology (VIB), Pleinlaan 2, 1050 Brussels, Belgium, and ^bDepartment of Cellular and Molecular Immunology, Vrije Universiteit Brussel (VUB), Flanders Interuniversity Institute for Biotechnology (VIB), Pleinlaan 2, 1050 Brussels, Belgium

Correspondence e-mail: bouckaetj@vub.ac.be

Received 28 September 2006

Accepted 17 November 2006

N-terminal truncation enables crystallization of the receptor-binding domain of the FedF bacterial adhesin

FedF is the two-domain tip adhesin of F18 fimbriae from enterotoxigenic *Escherichia coli*. Bacterial adherence, mediated by the N-terminal receptor-binding domain of FedF to carbohydrate receptors on intestinal microvilli, causes diarrhoea and oedema disease in newly weaned piglets and induces the secretion of Shiga toxins. A truncate containing only the receptor-binding domain of FedF was found to be further cleaved at its N-terminus. Reconstruction of this N-terminal truncate rendered FedF amenable to crystallization, resulting in crystals with space group $P2_12_12_1$ and unit-cell parameters $a = 36.20$, $b = 74.64$, $c = 99.03$ Å that diffracted to beyond 2 Å resolution. The binding specificity of FedF was screened for on a glycan array, exposing 264 glycoconjugates, to identify specific receptors for cocrystallization with FedF.

1. Introduction

Post-weaning diarrhoea and oedema disease in piglets are caused by pathogenic *Escherichia coli* strains, including enterotoxigenic *E. coli* (ETEC) and Shiga toxin-producing *E. coli* (STEC). These strains produce F18 fimbriae that mediate the adherence of the bacteria to the microvilli of the epithelial surface cells in the duodenum (Bertschinger *et al.*, 1990; Gyles *et al.*, 1998) and also secrete toxins during the infection process. The Shiga toxin Stx_{2e} variant is secreted by F18ab-positive *E. coli* and reaches its target host cell *via* the bloodstream, resulting in the accumulation of fluid (oedema disease) and neurological disorders. F18ac-positive *E. coli* cause post-weaning diarrhoea by the expression of heat-stable enterotoxins (Kapitany *et al.*, 1979).

F18 fimbriae are expressed by the *fed* gene cluster, with *fedA* encoding the major subunit (Imberechts *et al.*, 1992), *fedB* the outer membrane usher, *fedC* the periplasmic chaperone and *fedE* and *fedF* minor pilus subunits. FedE and FedF are the F18 minor subunits essential for fimbrial adhesion and also regulate the length of the F18 fimbriae, which are normally 1–2 µm in length (Imberechts *et al.*, 1996). FedF was shown to be the F18 tip adhesin (Smeds *et al.*, 2001). Its pilin domain connects to the FedA polymer *via* multiple copies of FedE (Hahn *et al.*, 2000). Tip adhesins can typically be divided into two immunoglobulin-like domains: an N-terminal lectin or receptor-binding module and a C-terminal pilin. Like all major and minor pilins, the pilin domains of adhesins also lack the last β-strand G of their immunoglobulin fold (Ig-fold). Therefore, the adhesin is not a stable protein by itself, but needs to be donor-strand complemented (Choudhury *et al.*, 1999; Sauer *et al.*, 1999). The missing strand is donated by a specific chaperone in the periplasm and by the next pilin in the pilus.

In the absence of the chaperone, soluble self-complementing and swapped pilin dimers (Jędrzejczak *et al.*, 2006; Van Molle *et al.*, in preparation) and trimers (Pettigrew *et al.*, 2004; Korotkova *et al.*, 2006) are formed and their crystal structures have been reported. Here, the pilins use different methods to complement their Ig-fold other than using the N-terminal extension of the neighbouring pilin domain. Because of misfolding, the pilins in these oligomers are incompetent for assembly into fimbriae or pili. On the other hand, crystal structures of typical tip adhesins have been mostly determined

Table 1

Data-collection and processing for FedF_{15–165}.

Values in parentheses are for the highest resolution shell (2.07–2.00 Å).

Space group	$P2_12_12_1$
Unit-cell parameters (Å)	$a = 36.20, b = 74.64, c = 99.03$
Unit-cell volume (Å ³)	267578
Resolution range (Å)	42–2.0
Reflections (total/unique)	100446/18666
Completeness (%)	98.9 (98.8)
$R_{\text{merge}}^{\dagger}$ (%)	7.5 (12.1)
$\langle I/\sigma(I) \rangle$	18.5 (12.4)
Matthews coefficient (Å ³ Da ^{−1})	2.05
Solvent content (%)	40.1
FedF molecules per ASU	2

$\dagger R_{\text{merge}} = \sum_{\mathbf{h}} \sum_l |I_{\mathbf{h}l} - \langle I_{\mathbf{h}} \rangle| / \sum_{\mathbf{h}} \sum_l \langle I_{\mathbf{h}} \rangle$, where $I_{\mathbf{h}l}$ is the l th observation of reflection $\mathbf{h}$ and $\langle I_{\mathbf{h}} \rangle$ is the weighted average intensity for all observations l of reflection $\mathbf{h}$.

ProBlot PVDF membrane (Applied Biosystems) and excision of the bands corresponding to the estimated molecular weight upon staining with Coomassie R350.

2.2. Crystallization and crystal analysis

Screening for crystallization conditions for FedF_{15–165} took place at the high-throughput crystallization facility at the EMBL in Hamburg (Germany) after in-house screening with commercially available screens (Crystal Screens 1 and 2 from Hampton Research) remained unsuccessful. Using the robot, 300 nl droplets of protein solution (24 mg ml^{−1} in 20 mM HEPES pH 7.5 and 140 mM NaCl) were mixed with 300 nl precipitant solution. Five sets of 96 precipitant conditions from both commercial and academically developed screens were used: Index, Crystal Screens I and II, Natrix and MembFac (Hampton Research), the JCSG screen (Newman *et al.*, 2005) and Jena 1–4 (Jena Bioscience). Crystals of the FedF_{15–165} truncate were only obtained using the JCSG screen with 0.17 M ammonium sulfate, 25.5% (w/v) PEG 4000 and 15% (v/v) glycerol. Larger crystals were obtained at home using the same conditions and microseeding.

Prior to data collection, a microcrystal of FedF_{15–165} was directly mounted from the JCSG screen in a nylon cryoloop (Hampton Research) and placed directly into a nitrogen-gas stream at 100 K at beamline X12 of the EMBL Outstation Hamburg, Germany. The data were processed using *DENZO*, *XDISPLAYF* and *SCALEPACK* from the *HKL* package (Otwinowski & Minor, 1997) and *TRUNCATE* from the *CCP4* suite (Collaborative Computational Project, Number 4, 1994). Crystal parameters and data-processing statistics are summarized in Table 1.

Figure 2

FedF_{1–165} (lane 2) showing two bands on 12% SDS–PAGE, the N-terminal truncate FedF_{15–165} (lane 3) and PageRuler protein ladder (Fermentas SM0661; lane 1).

2.3. Screening of carbohydrate binding of FedF on a glycan microarray

The carbohydrate-specificity of the FedF_{1–165} receptor-binding domain of the FedF fimbrial adhesin from the F18⁺ *E. coli* 107/86 strain was screened against printed array version 2. Slides (Schott Nexterion H slides, catalogue No. 1070936B) were printed at the Core D carbohydrate-synthesis/protein-expression consortium located at the Scripps Research Institute. The slides exposed 264 immobilized glycans spotted at a concentration of 100 µM and in six replicates (Blixt *et al.*, 2004). Preprinted slides were soaked in deionized water for 5 min at room temperature and dried in a stream of nitrogen. FedF_{1–165} was diluted to the assay concentration of 100 µg ml^{−1} in 20 mM Tris–HCl pH 7.4, 150 mM NaCl, 2 mM CaCl₂ and 2 mM MgCl₂, the wash buffer, in the presence of 0.05% Tween-20 and 1% BSA, and 50 µl was applied to the printed surface and covered with a cover slip. The slide was incubated, protected from light, in a humidified chamber for 1 h at room temperature. A first wash step used the wash buffer with 0.05% Tween-20 and wash buffer alone for 10 s each. For detection, 50 µl rabbit anti-FedF was applied at 10 µg ml^{−1} and the slide was subsequently incubated and washed as described above. Finally, 50 µl labelled goat anti-rabbit IgG-488 was applied, incubated and washed again in the same way. After another 10 s wash in deionized water, the slide was dried in a nitrogen stream. The binding image was read by a Perkin–Elmer Microscanarray XL4000 scanner. Image analysis was performed using *Image* v.6 image-analysis software. The raw data results files were uploaded to the Consortium for Functional Glycomics database and webpages (<http://www.functionalglycomics.org>).

3. Results and discussion

Of the two FedF proteins, FedF_{1–110H6} and FedF_{1–165H6}, that were C-terminal truncates followed by a six-histidine tag, only FedF_{1–165H6} was soluble (Fig. 1). However, it did not crystallize. Therefore, FedF_{1–165} and two more truncates, FedF_{1–160} and FedF_{1–162}, which are five and three amino-acid residues shorter (Fig. 1), respectively, were produced without a His tag. The construct for the shortest C-terminal truncate, FedF_{1–160}, did not express. FedF_{1–162} and FedF_{1–165} were purified as soluble proteins and set up for crystallization. However, no crystals could be obtained. In an attempt to make the FedF receptor-binding domain more amenable to crystallization (Buts *et*

Figure 3

Crystals of FedF_{15–165} grown at the EMBL high-throughput crystallization facility.

al., 2005), we purified FedF_{v1-165} (Fig. 1) originating from a natural variant of the adhesin from the F18⁺ *E. coli* strain 189P (Imberechts, CODA, Belgium). FedF_{v1-165} also did not crystallize.

Remarkably, for all of the expressed soluble FedF truncates, the protein appeared as two close bands on SDS-PAGE, near the expected molecular weight of 18 kDa (Fig. 2). Moreover, both bands were also detected on Western blotting using the polyclonal rabbit antibody against FedF. The addition of reducing agents, such as 20 mM DTT, to the sample-loading buffer did not change this feature, indicating that this was not a consequence of incomplete disulfide reduction. N-terminal sequencing revealed two different N-termini, starting either at the expected residue or 14 residues later (Fig. 1). These results demonstrate that the FedF constructs were subject to a naturally occurring N-terminal cleavage. It was reasoned that the N-terminal heterogeneity could disturb the crystallization process. Two constructs encoding FedF₁₃₋₁₆₅ and FedF₁₅₋₁₆₅ were designed to match the shorter sequence of FedF₁₋₁₆₅ (Fig. 1). The signal sequence of the F4 fimbrial adhesin from *E. coli* was used because it is known to allow high levels of periplasmic FaeG. In FedF₁₃₋₁₆₅, the two residues preceding Asn15 were kept to resemble the cleavage site for the signal peptidase in FaeG. Nevertheless, the N-terminally truncated FedF₁₅₋₁₆₅ led to higher yields of soluble protein (Fig. 2). It was purified and set up for crystallization.

Crystals were obtained in one of the 500 conditions screened (Fig. 3). Although these crystals were smaller (300 × 40 × 40 µm) than those later reproduced in-house (1 × 0.2 × 0.1 mm), they diffracted considerably better (Table 1). It is most likely that the variable nonthermostatic conditions during setup are partly responsible for these differences, as the protein droplets tend to precipitate rapidly upon mixing at temperatures above 298 K in the laboratory. On the other hand, the temperature is well controlled in the crystallization robot and storage room.

Sugars that bind specifically to FedF may serve as a tool to solve its structure, as exemplified by the use of an Se derivative of GlcNAc (Buts, Loris *et al.*, 2003), or to obtain crystals in different space groups. The glycan array pointed out the recognition of N-acetyl-lactosamine units, Gal(β1-4)GlcNAc, that were either sulfated at position 3 and/or position 6 or fucosylated at position 2 of the galactose (Fig. 4). α1-Acid glycoprotein, which carries Gal(β1-4)-[Fuc(α1-3)]GlcNAc, was the best binding glycoconjugate. Other recognized carbohydrates were α1-2-fucosylated, namely Gal(α1-3)-[Fuc(α1-2)]Gal and the blood group H type II trisaccharide Fuc(α1-2)Gal(β1-4)GlcNAc (Fig. 4). Blood group H type II trisaccharide has previously been characterized as an F18 receptor (Snoeck *et al.*, 2004). (α1-2)-Fucosyltransferases have been implicated as being of importance for the presence of F18 receptors in piglets (Meijerink *et al.*, 1997, 2000). Trials to cocrystallize FedF in complex with several of these carbohydrates have so far remained unsuccessful.

Natural C-terminal truncates have been observed when over-producing recombinant tip adhesins, for example FimH and F17G. This possibly occurs as noncomplemented pilin domain of the adhesin becomes degraded. To date, N-terminal cleavage of tip adhesins has

not been observed. It is therefore currently not understood what the mechanism and role of the cleavage of the 14-residue N-terminal truncation of FedF₁₋₁₆₅ is (Fig. 1). The N-terminal 14-residue sequence is very susceptible to proteolysis (<http://www.expasy.org/tools/peptidecutter>), whereas the amino-acid sequence just after the N-terminal truncation site (Asn15) is predicted not to be susceptible. Its sensitivity to proteolysis or hydrolysis could indicate that the structure of the N-terminal 14 residues behaves independently of the globular receptor-binding domain. Its length is reminiscent of the typical N-terminal extension that serves as a donor strand for the complementation of the incomplete Ig-fold of pilins. Moreover, its amino-acid sequence has an alternating hydrophobic-hydrophilic pattern and is conserved among variants (Fig. 1). Both those features are not present for the new N-terminus of the N-terminal truncate. If the 14 residues were a donor strand, this would have important implications. It would allow FedF to intercalate in the middle of F18 fimbriae by donating this N-terminal extension to a previous pilin in the pilus. In the absence of the 14-residue N-terminal sequence, FedF would be destined to reside only at the tip of F18 fimbriae. Solving the FedF crystal structure could shed light on this mechanism.

We thank Liu Qiang from the Research Center of Centre Hospitalier de l'Université de Montreal, Canada for his help during the initiation of this project. We wish to thank Richard A. Alvarez, Director of Core H, for performing the glycan-array analysis, conducted by the Protein-Glycan Interaction Core H of the Consortium for Functional Glycomics funded by NIGMS grant GM62116. We are grateful to Dr Manfred Weiss, EMBL beamline scientist at X12 at the DESY synchrotron, Hamburg. JB is a post-doctoral fellow of the Fonds voor Wetenschappelijk Onderzoek-Vlaanderen, which also provided financial support (FWO-G.0513.04) and the DNA-sequencing equipment (FWO-AL215).

References

- Bertschinger, H. U., Bachmann, M., Mettler, C., Pospischil, A., Schraner, E. M., Stamm, M., Sydler, T. & Wild, P. (1990). *Vet. Microbiol.* **25**, 267–281.
- Blixt, O. *et al.* (2004). *Proc. Natl Acad. Sci. USA*, **101**, 17033–17038.
- Bouckaert, J., Berglund, J., Schembri, M., De Genst, E., Cools, L., Wührer, M., Hung, C. S., Pinkner, J., Slättegård, R., Zavialov, A., Choudhury, D., Langermann, S., Hultgren, S., Wyns, L., Klemm, P., Oscarson, S., Knight, S. & De Greve, H. (2005). *Mol. Microbiol.* **55**, 441–455.
- Buts, L., Bouckaert, J., De Genst, E., Loris, R., Oscarson, S., Lahmann, M., Messens, J., Brosens, E., Wyns, L. & De Greve, H. (2003). *Mol. Microbiol.* **49**, 705–715.
- Buts, L., Loris, R., De Genst, E., Oscarson, S., Lahmann, M., Messens, J., Brosens, E., Wyns, L., De Greve, H. & Bouckaert, J. (2003). *Acta Cryst.* **D59**, 1012–1015.
- Buts, L., Wellens, A., Van Molle, I., Wyns, L., Loris, R., Lahmann, M., Oscarson, S., De Greve, H. & Bouckaert, J. (2005). *Acta Cryst.* **D61**, 1149–1159.
- Choudhury, D., Thompson, A., Stojanoff, V., Langermann, S., Pinkner, J., Hultgren, S. J. & Knight, D. S. (1999). *Science*, **285**, 1061–1065.
- Coligan, J. E., Dunn, B. M., Ploegh, H. L., Speicher, D. W. & Wingfield, P. T. (2002). Editors. *Current Protocols in Protein Science*. New York: Wiley.
- Collaborative Computational Project, Number 4. (1994). *Acta Cryst.* **D50**, 760–763.
- Dodson, K. W., Pinkner, J. S., Rose, T., Magnusson, G., Hultgren, S. J. & Waksman, G. (2001). *Cell*, **105**, 733–743.
- Gyles, C., Johnson, R., Gao, A. L., Ziebell, K., Pierard, D., Aleksic, S. & Boerlin, P. (1998). *Appl. Environ. Microbiol.* **64**, 4134–4141.
- Hahn, E., Wild, P., Schraner, E. M., Bertschinger, H. U., Häner, M., Müller, S. A. & Aebi, U. (2000). *J. Struct. Biol.* **132**, 241–250.
- Imberechts, H., De Greve, H., Schlicker, C., Bouchet, H., Pohl, P., Charlier, G., Bertschinger, H., Wild, P., Vandekerckhove, J., Van Damme, J., Van Montagu, M. & Lintermans, P. (1992). *Infect. Immun.* **61**, 1963–1971.
- Imberechts, H., Wild, P., Charlier, G., De Greve, H., Lintermans, P. & Pohl, P. (1996). *Microb. Pathog.* **21**, 183–192.

Figure 4

Glycoconjugates recognized by FedF₁₋₁₆₅ on a glycan array by the Consortium for Functional Glycomics.

- Jędrzejczak, R., Dauter, Z., Dauter, M., Piatek, R., Zalewska, B., Mroz, M., Bury, K., Nowicki, B. & Kur, J. (2006). *Acta Cryst.* **D62**, 157–164.
- Kapitany, R. A., Forsyth, G. W., Scoot, A., McKenzie, S. F. & Worthington, R. W. (1979). *Infect. Immun.* **26**, 173–177.
- Korotkova, N., Le Trong, I., Samudrala, R., Korotkov, K., Van Loy, C. P., Bui, A. L., Moseley, S. L. & Stenkamp, R. E. (2006). *J. Biol. Chem.* **281**, 22367–22377.
- Meijerink, E., Fries, R., Vogeli, P., Masabanda, J., Wigger, G., Stricker, C., Neuenschwander, S., Bertschinger, H. U. & Stranzinger, G. (1997). *Mamm. Genome*, **8**, 736–741.
- Meijerink, E., Neuenschwander, S., Fries, R., Dinter, A., Bertschinger, H. U., Stranzinger, G. & Vogeli, P. (2000). *Immunogenetics*, **52**, 129–136.
- Miroux, B. & Walker, J. E. (1996). *J. Mol. Biol.* **260**, 289–298.
- Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I. & Perrakis, A. (2005). *Acta Cryst.* **D61**, 1426–1431.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Pettigrew, D., Anderson, K. L., Billington, J., Cota, E., Simpson, P., Urvil, P., Rabuzin, F., Roversi, P., Nowicki, B., du Merle, L., Le Bouguenec, C., Matthews, S. & Lea, S. M. (2004). *J. Biol. Chem.* **279**, 46851–46857.
- Rost, B. & Liu, J. (2003). *Nucleic Acids Res.* **31**, 3300–3304.
- Sauer, F. G., Fütterer, K., Pinkner, J. S., Dodson, K. W., Hultgren, J. S. & Waksman, G. (1999). *Science*, **285**, 1058–1061.
- Smeds, A., Hemmann, K., Jakava-Viljanen, M., Pelkonen, S., Imberechts, H. & Palva, A. (2001). *Infect. Immun.* **69**, 7941–7945.
- Snoeck, V., Verdonck, F., Cox, E. & Goddeeris, B. M. (2004). *Vet. Microbiol.* **100**, 241–246.