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

Crystallization and preliminary crystallographic analysis of an Ig-domain-encompassing fragment of the giant adhesion protein SiiE from *Salmonella enterica*

Karina U. Sturm,[a]‡ Martin H. Griessl,[a]‡ Carolin Wagner,[b] Jörg Deiwick,[b] Michael Hensel[b] and Yves A. Muller[a]

[a]Lehrstuhl für Biotechnik, Department of Biology, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany, and [b]Division of Microbiology, School of Biology/Chemistry, University of Osnabrück, Osnabrück, Germany

‡ These authors contributed equally to this work.

Received 22 June 2011
Accepted 8 August 2011

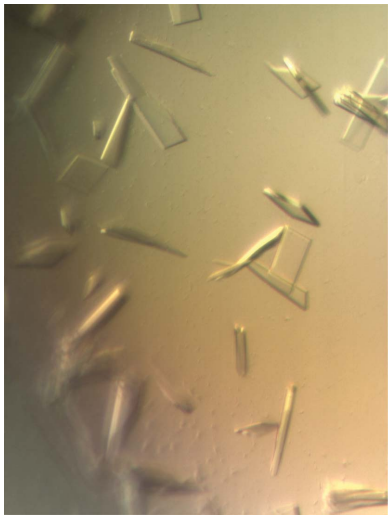
Salmonella infections can be life-threatening. SiiE is a giant adhesion molecule of 5559 amino acids that is encoded in *Salmonella* pathogenicity island 4 (SPI4) and that promotes the initial contact between the pathogen and polarized epithelial cells in the intestine of the host. Starting from an engineered deletion version of SiiE (mini-SiiE; 97 kDa), limited proteolysis was used to reproducibly generate a 30 kDa fragment that readily crystallized. Mass spectrometry hints that this fragment spans the predicted Ig domains 50–52 of SiiE. Crystals of both native and selenomethionine-labelled protein could be obtained in space group *C*2 and diffraction data were recorded to a resolution of 1.85 Å.

1. Introduction

Whereas some *Salmonella* strains such as *S. enterica* serovar Typhi drastically impact human health, other *S. enterica* serovars such as Typhimurium or Enteritidis only cause medium to mild impairments such as intestinal inflammation and diarrhoea in most individuals (World Health Organization, 2005). In *Salmonella* the early steps of invasion, *i.e.* the breaching of the intestinal epithelial barrier, are mediated by the interplay of gene products encoded in two separate gene clusters within the *Salmonella* genome. Whereas one gene cluster termed *Salmonella* pathogenicity island 1 (SPI1) encodes a type III secretion system (T3SS) that injects *Salmonella* effector proteins into eukaryotic host cells, the role of the SPI4 gene cluster is far less well understood (Gerlach *et al.*, 2008; Haraga *et al.*, 2008).

SPI4 contains six open reading frames and the corresponding gene products are termed SiiA to SiiF. SiiC, SiiD and SiiF form a type I secretion system (T1SS), while SiiE constitutes the secreted substrate protein (Gerlach *et al.*, 2007; Morgan *et al.*, 2007; Wagner *et al.*, 2011). The proteins SiiA and SiiB appear not to be required for effective secretion of SiiE (Gerlach *et al.*, 2007). SiiE is remarkable for several reasons. With 5559 amino acids, it is the largest protein encoded by *Salmonella*. Bioinformatics investigations predicted that SiiE contains as many as 53 Ig-domain repeats. Towards the N-terminus, the Ig repeats are preceded by a predicted coiled-coil region with eight heptad repeats (Gerlach *et al.*, 2007). However, homo-oligomer formation of SiiE has not been observed to date. The C-terminal region contains the secretion signal for the T1SS.

SiiE functions as a giant nonfimbrial adhesin and its proposed role is to promote initial contact between *Salmonella* and the apical side of the intestinal epithelium (Wagner *et al.*, 2011). This initial contact is required before invasion can proceed *via* the T3SS that is encoded on SPI1 (Gerlach *et al.*, 2008). The extended size of SiiE seems to be necessary in order for the C-terminus of SiiE to reach out beyond the unusually long lipopolysaccharide molecules that cover the surface of *Salmonella*. The rate of invasion becomes significantly reduced upon the deletion of ten or more Ig domains in SiiE (Gerlach *et al.*, 2008; Wagner, 2011). Currently, the identity of the binding partner of SiiE on the surface of epithelial host cells is not known. At present, no experimental data are available for the atomic structure of any portion of SiiE. In order to shed light on the molecular mechanisms that provide function and specificity to the initial adhesion of *Salmonella* to host cells, structural information would be highly welcome.



2. Materials and methods

2.1. Protein production and purification

An engineered deletion variant of SiiE (UniProtKB entry Q8ZKG6) in which the N-terminal residues 5–403 were fused to the C-terminal residues 5053–5559 of SiiE (mini-SiiE) was recombinantly produced in *Escherichia coli* strain BL21 (DE3) with plasmid p3355-1 (Wagner, 2011). This plasmid was engineered from a pGEX-6P-1 expression plasmid (GE Healthcare, Uppsala). Mini-SiiE could be produced in high amounts in *Escherichia coli* and readily purified using the following procedure (Wagner, 2011). After an initial glutathione-Sepharose affinity-chromatography step (GE Healthcare), the eluted protein was cleaved with PreScission protease. The protein mixture was then purified by Q-Sepharose ion-exchange chromatography followed by a second glutathione Sepharose affinity-chromatography step, yielding highly pure mini-SiiE protein with a theoretical molecular weight of 97 kDa.

Since crystallization trials with mini-SiiE remained unsuccessful, we performed limited proteolysis experiments with proteinase K to identify a more stable and possibly more easily crystallizable fragment of mini-SiiE. For this purpose, we incubated proteinase K and mini-SiiE at a ratio of 0.003 g:1 g for several hours at 293 K in 20 m*M* Tris–HCl buffer pH 7.4 and analysed the time course of the proteolysis reaction by SDS–PAGE. After several hours a protease-resistant fragment of about 30 kDa could be identified. Limited proteolysis proved to be highly reproducible and milligram amounts of this fragment could therefore be produced when stopping the proteolytic digestion with the serine protease inhibitor 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF) after 4 h. The fragment could be isolated with high purity by size-exclusion chromatography (Superdex 200, GE Healthcare). Mass spectrometry was used to help to identify the sequence of the peptide.

A highly similar purification protocol was also used for production of the selenomethionine-labelled SiiE fragment. The incorporation of selenomethionine was accomplished by taking advantage of the bacterial feedback-inhibition mechanism (Van Duyne *et al.*, 1993). For this purpose, *E. coli* BL21 (DE3) cells carrying the same expression plasmid as before were cultivated in M9 minimal medium and subsequently incubated with an amino-acid mixture supplemented with selenomethionine. Protein expression was induced with 0.1 m*M* IPTG and the bacterial cells were incubated overnight. In contrast to the unlabelled native protein, all purification steps and the proteolytic cleavage of the selenomethionine-labelled protein sample were performed in the presence of 5 m*M* DTT.

2.2. Mass spectrometry

The 30 kDa proteinase K fragment was separated by SDS–PAGE and visualized by Coomassie staining, and 0.5 µg was processed for trypsin digestion (310 K, 12 h). Peptides were separated at a flow rate of 50 µl min^{-1} using C18 RP-HPLC (Alltima AQ column, length 150 mm, inner diameter 1 mm) using a 30 min gradient (A = 95:5:0.03, B = 20:80:0.03 water:acetonitrile:TFA; Agilent 1100). The eluate was used directly for ESI-MS (Esquire HCT, Bruker Daltonics). MS/MS data sets were processed using Bruker Daltonics *DataAnalysis* v.4.0 and *Biotools* v.3.1. The mass deviation was 1 Da.

2.3. Crystallization

For crystallization setups, the native unlabelled and selenomethionine-labelled 30 kDa SiiE fragments were concentrated to 5 mg ml^{-1} in 20 m*M* Tris–HCl buffer pH 7.4. 2 m*M* DTT was added to the selenomethionine-containing sample only. Crystallization trials for the native fragment were set up using Crystal Screen, Crystal Screen 2, Index and PEGRx screens (Hampton Reseach, Aliso Viejo, California, USA) at 292 and 277 K. Initial crystals were optimized with the Additive Screen (Hampton Reseach). Crystals of the unlabelled protein were grown using the sitting-drop method, mixing 0.2 µl protein solution with 0.2 µl of a reservoir solution consisting of 0.1 *M* sodium cacodylate pH 6.7, 0.2 *M* calcium acetate, 16%(*w*/*v*) PEG 8000 and 3%(*w*/*v*) 6-aminohexanoic acid. The droplet was equilibrated against 70 µl reservoir solution at 292 K for several weeks. Crystals of the selenomethionine-labelled protein were obtained directly from the PEG/Ion screen (Hampton Reseach, Aliso Viejo, California, US). As for the unlabelled protein, 0.2 µl of both protein solution and reservoir solution [0.2 *M* ammonium iodide, 20%(*w*/*v*) PEG 3350] were mixed and equilibrated against 70 µl reservoir solution.

Table 1
Data-collection statistics.

Values in parentheses are for the highest resolution shell.

		Crystal form II	
	Crystal form I	Peak	Remote
Beamline	BESSY MX-14-1		
Detector	Rayonics MX225 3 × 3 CCD detector		
Temperature (K)	100		
Data set	Native	SeMet-labelled	
Space group	*C*2	*C*2	
Unit-cell parameters (Å, °)	a = 129.40, b = 69.04, c = 67.26, β = 90.92	a = 93.50, b = 153.40, c = 96.30, β = 103.9	
Wavelength (Å)	0.9180	0.9797	0.9180
Resolution range (Å)	50–1.85 (1.90–1.85)	35–2.12 (2.18–2.12)	60–2.12 (2.18–2.12)
Unique reflections	50496 (3704)	141274 (8916)	143232 (10703)
Average multiplicity	4.1 (4.1)	1.9 (1.6)	1.9 (1.9)
Completeness (%)	99.6 (99.7)	95.9 (82.2)	97.1 (98.5)
Mean $I/\sigma(I)$	13.9 (3.2)	9.3 (2.3)	9.05 (2.3)
$R_{\rm meas}$† (%)	8.3 (53.6)	7.9 (45.8)	8.2 (54.2)
$R_{\rm mrgd\text{-}F}$‡ (%)	10.8 (54.7)	14.4 (45.8)	16.2 (77.3)
Wilson B value (Å^2)	25.4	38.4	39.9

† $R_{\rm meas}$ is defined as $\sum_h [n_h/(n_h-1)]^{1/2} \sum_i^{n_h} |\langle I_h\rangle - I_{h,i}| / \sum_h \sum_i^{n_h} I_{h,i}$ (Diederichs & Karplus, 1997). ‡ $R_{\rm mrgd\text{-}F}$ is defined as $\sum |A_{I_{h,P}} - A_{I_{h,Q}}| / 0.5 \sum (A_{I_{h,P}} + A_{I_{h,Q}})$ (Diederichs & Karplus, 1997).

2.4. Data collection and preliminary phasing

A single data set from a native nonlabelled crystal and two data sets from the selenomethionine-derivatized protein fragment were collected at the BESSY synchrotron, Berlin at 100 K (Table 1). All data sets were reduced with the program *XDS* and scaled using *XSCALE* (Kabsch, 2010*a*,*b*). The two data sets for the selenomethionine-labelled protein were collected at wavelengths of 0.9797 and 0.9180 Å and the structure was solved with the SAD method. The selenium substructure could be determined with the program *SHELXD* (Sheldrick, 2008) and an initial round of density modification and autotracing was performed with *SHELXE* (Sheldrick, 2008).

3. Results and discussion

We had not previously attempted to produce and purify the complete SiiE protein since its size of about 595 kDa and the fact that it is predicted to contain in excess of 50 individual domains render its successful crystallization unlikely (Gerlach *et al.*, 2007; Wagner *et al.*, 2011). As an alternative, we generated a protein termed mini-SiiE (Wagner, 2011). This deletion mutant lacks the central amino acids 404–5052 of SiiE and has a size of about 97 kDa. However, even our attempts to crystallize mini-SiiE have remained unsuccessful to date.

In contrast, limited proteolysis experiments yielded a fragment of about 30 kDa that could be isolated reproducibly and crystallized.

Limited proteolysis of mini-SiiE with proteinase K proceeds through cleavage of the polypeptide chain at multiple positions in parallel (Fig. 1). Three prominent protein bands become apparent over time. Within the first 10 min, significant amounts of a defined fragment of about 20 kDa are temporarily formed that subsequently become rapidly degraded. In parallel, fragments of about 10 and 30 kDa are obtained that appear to remain stable under the chosen conditions for up to 7 h (Fig. 1 and data not shown). Because of this long-term stability, we were able to isolate milligram amounts of the 30 kDa fragment, whereas similar attempts to isolate the 20 kDa fragment have so far remained unsuccessful. The observed time course hints that the sequences of the 20 and 30 kDa fragments, as well as those of the 10 and 30 kDa fragments, do not overlap.

Mass-spectrometric analysis of the 30 kDa fragment has allowed us to identify peptides that span residues 5078–5287 of SiiE. These peptides extend from the beginning of predicted Ig-like domain 50 to the middle of Ig-like domain 52 of SiiE (Gerlach *et al.*, 2007). We reason that all three Ig-like domains 50–52 might be intact in the 30 kDa fragment. In the overproduced mini-SiiE construct, Ig domain 50 is preceded by an Ig domain that is constituted of a chimera between Ig domain 2 (two thirds of the sequence) and Ig domain 49 of SiiE (Wagner, 2011). To us, it appears quite likely that such a chimeric Ig domain is less stable than a native Ig domain and we therefore propose that the proteolytically obtained 30 kDa fragment starts at Ig domain 50 and extends until the end of Ig domain 52 (predicted molecular mass 28.9 kDa). This would also be consistent with the prediction that Ig domain 52 is separated from the subsequent Ig domain 53 by a 51-residue-long insertion that could be devoid of any secondary structure (Gerlach *et al.*, 2007). On assuming that the crystallized fragment contains Ig-like domains 50–52 of SiiE, the combined peptides identified by mass spectrometry provide 73% sequence coverage.

The purified fragment crystallized in two different crystal forms, both with *C*2 space-group symmetry (Fig. 2*a*, Table 1). The crystals of the native protein in crystal form I are characterized by an unusual β angle of 90.9° and diffracted to 1.85 Å resolution (Fig. 2*b*). Any attempts to explain the diffraction pattern of this crystal form in a higher symmetry Laue group always yielded unreasonably high *R* factors. In both crystal forms more than one monomer of the 30 kDa fragment can be expected. Whereas in crystal form I up to three molecules could be present (solvent content between 76 and 29%), in crystal form II between three and six molecules are presumably contained in the crystallographic asymmetric unit (solvent content between 68 and 36%). The most likely values appear to be two molecules in crystal form I (solvent content 52%, Matthews coefficient 2.6 Å^3 Da^{-1}) and four molecules in crystal form II (solvent content 57%, Matthews coefficient 2.9 Å^3 Da^{-1}; see also below).

The collection of two data sets at different wavelengths from the selenomethionine-labelled protein in crystal form II at the BESSY

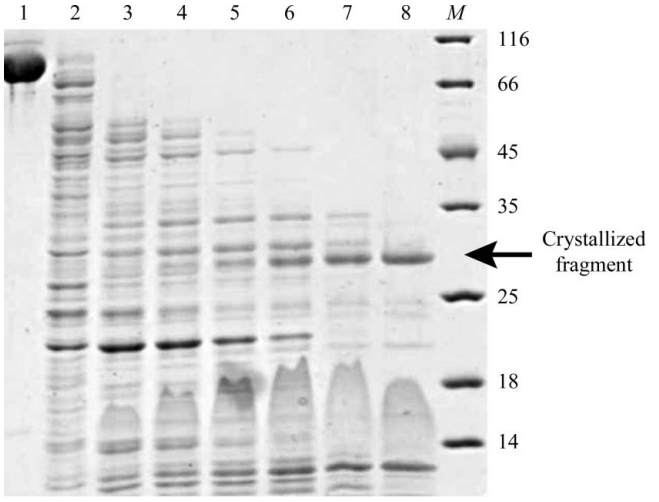
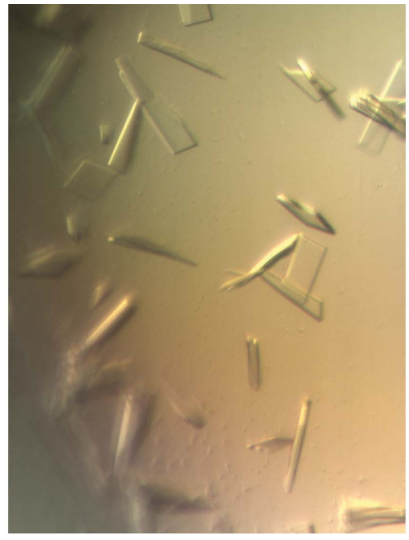
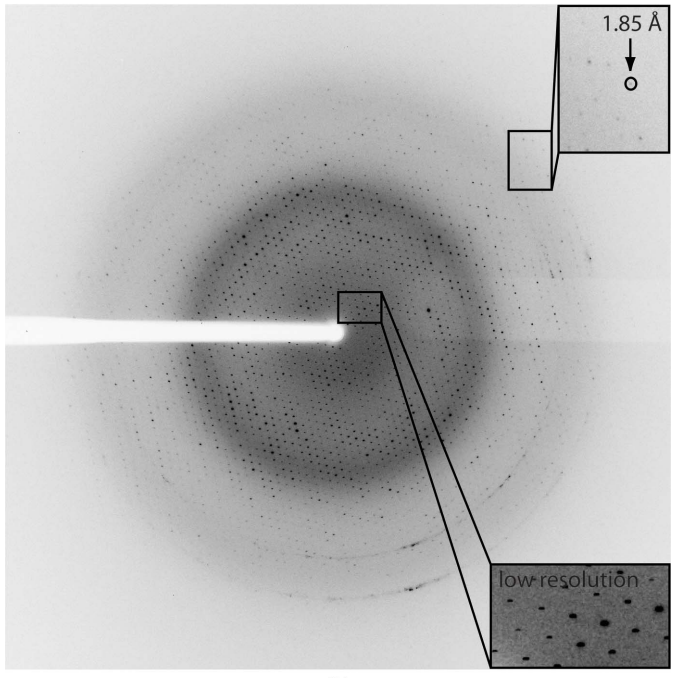


Figure 1
Time course of the limited proteolytic digestion of mini-SiiE monitored by SDS–PAGE. Upon incubation of mini-SiiE with proteinase K at a ratio of 0.003 mg proteinase per milligram of target protein, a prominent stable fragment of about 30 kDa can be observed for up to 7 h. Lane 1, t = 0; lane 2, t = 60 s; lane 3, t = 300 s; lane 4, t = 10 min; lane 5, t = 15 min; lane 6, t = 30 min; lane 7, t = 1 h; lane 8, t = 2 h. Lane *M* contains molecular-weight marker (labelled in kDa).



(*a*)



(*b*)

Figure 2
(*a*) Crystals of the selenomethionine-labelled fragment of about 30 kDa of SiiE. (*b*) Diffraction pattern recorded from crystals of crystal form I and extending to 1.85 Å resolution.

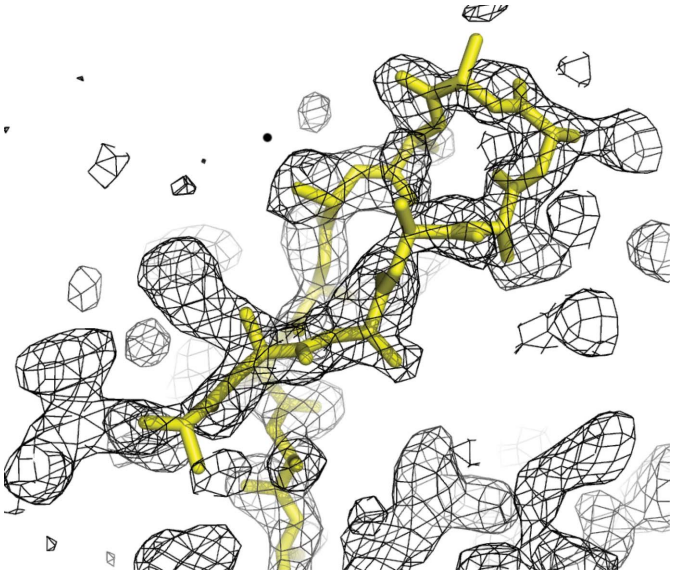


Figure 3
Preliminary electron-density map obtained after density modification with *SHELXE* (Sheldrick, 2008). Electron density is displayed at a 1.3σ cutoff and is illustrated with the program *PyMOL* (DeLano, 2002).

synchrotron, Berlin allowed us to derive initial phases for these crystals (Table 1). Although MAD phasing was attempted, better statistics were obtained with the SAD method. The anomalous signal of the peak data set proved to be useful up to a resolution of 3.0 Å [average $f''/\sigma(f'') > 1.0$]. *SHELXD* located ten Se atoms with occupancies greater than 0.3. The best solution had correlation coefficients of 47.6/28.4 (for all/weak data; Schneider & Sheldrick, 2002). Ig domains 50–52 of SiiE contain three methionines in total, so that the ten observed Se atoms identified with *SHELXD* would roughly agree with the presence of four molecules of this 30 kDa fragment in the asymmetric unit. An initial electron-density map obtained with *SHELXE* was iteratively further improved using the autotracing option. The electron-density map shows features that are clearly recognisable as β-sheets (Fig. 3).

Overall, we anticipate that the quality of the initial electron-density map in combination with the 1.85 Å resolution of the native data set will allow us to determine the crystal structure of the 30 kDa SiiE fragment, thereby enabling us to obtain the first experimental insights into the molecular-repeat structure of this rather unusual giant protein of 5559 residues in total.

We would like to thank Uwe Müller and Manfred Weiss at the BESSY synchrotron (Berlin, Germany) for help with data collection and Madhumati Sevvana for help with phasing. We thank Mattes Schulze and Rainer Buchholz from the Lehrstuhl für Bioverfahrenstechnik, University Erlangen-Nuremberg for help with the initial mass-spectrometric analysis and Stefan Walter from the Division of Microbiology, University of Osnabrück for help with LC-ESI-MS. This work was supported by the Deutsche Forschungsgemeinschaft within the framework of SFB 796 (Teilprojekt Z).

References

DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
Gerlach, R. G., Cláudio, N., Rohde, M., Jäckel, D., Wagner, C. & Hensel, M. (2008). *Cell. Microbiol.* **10**, 2364–2376.
Gerlach, R. G., Jäckel, D., Stecher, B., Wagner, C., Lupas, A., Hardt, W. D. & Hensel, M. (2007). *Cell. Microbiol.* **9**, 1834–1850.
Haraga, A., Ohlson, M. B. & Miller, S. I. (2008). *Nature Rev. Microbiol.* **6**, 53–66.
Kabsch, W. (2010*a*). *Acta Cryst.* D**66**, 125–132.
Kabsch, W. (2010*b*). *Acta Cryst.* D**66**, 133–144.
Morgan, E., Bowen, A. J., Carnell, S. C., Wallis, T. S. & Stevens, M. P. (2007). *Infect. Immun.* **75**, 1524–1533.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.
Wagner, C. (2011). PhD thesis. Friedrich-Alexander-University Erlangen-Nuremberg, Germany.
Wagner, C., Polke, M., Gerlach, R. G., Linke, D., Stierhof, Y. D., Schwarz, H. & Hensel, M. (2011). *Cell. Microbiol.* **13**, 1286–1301.
World Health Organization (2005). *Drug-resistant Salmonella*. Geneva: World Health Organization. http://www.who.int/mediacentre/factsheets/fs139/en/.