Acta Crystallographica Section F

Volume 70 (2014)

Supporting information for article:

Protein expression, characterization, crystallization and preliminary X-ray crystallographic analysis of a Fic protein from *Clostridium difficile*

Ditte Welner, Emil Dedic, Hans C. van Leeuwen, Ed Kuijper, Morten Jannik Bjerrum, Ole Østergaard and René Jørgensen

S1. Mass spectroscopy

Due to the contaminants always present in Fic samples (Figure 1a), we analyzed the sample with different mass spectrometry methods.

S1.1. Materials and Methods for MS

Sequencing grade trypsin was purchased from Promega (Madison, USA), 6.2 mg/mL alpha-cyano-4-hydroxycinnamic acid in 36:56:8 MeOH:ACN:Water was purchased from Agilent (Palo Alto, USA). All other chemicals were from Sigma-Aldrich (St. Louis, USA).

S1.2. In-gel digestion and protein identification by MALDI-TOF mass spectrometry

Protein bands equivalent to the bands shown in figure 1a were excised from the gel (band 1 and 2 as one piece), washed, in-gel reduced, S-alkylated with iodoacetic acid, and digested over night with trypsin as described (Shevchenko *et al.*, 2006). The formed peptides were extracted from the gel piece, dried in a vacuum centrifuge, redissolved in 5 % formic acid and subjected to a desalting/concentration step (Gobom *et al.*, 1999) before elution on a stainless steel target with 6.2 mg/mL alpha-cyano-4-hydroxycinnamic acid in methanol:acetonitrile:water (36:56:8) prior to analysis by MALDI-TOF MS. Delayed extraction MALDI mass spectra was recorded on a Bruker Ultraflex I Time-of-flight mass spectrometer using external calibration (Peptide Calibration Standard, Bruker-Daltonics, Bremen, Germany).

Protein identification was performed by peptide map fingerprinting using the MASCOT v 2.2 (Perkins *et al.*, 1999) searching against a custom designed sequence database including the entire *Eschericia coli* BL21 DE3 proteome from UniProt (release 2012_08) and the CdFic$^{SE/AA}$ sequence. Fixed modifications were set to carbamidomethylation of Cys and variable modifications were set to oxidation of Met and deamidation of Asn/Gln. Mass tolerances on the peptide masses were set to 200 ppm as external calibration was required due to absent trypsin autolysis products.

S1.3. Mr determination by mass spectrometry

The purified CdFic$^{SE/AA}$ was desalted using a homemade micro column made with poros 50R1 (Perseptive Biosystes, Framingham, USA) essentially as described in (Jensen & Wilm, 2002). In

brief, 2.5 μL CdFic$^{SE/AA}$ (8 mg/mL in 20 mM Tris pH 7.4, 5% glycerol, 5 mM β-mercaptoethanol) was diluted ten-fold with 5 % formic acid before loading on a pre-equilibrated acidified column. Salts were washed out by applying 2 x 20 μL 5 % formic acid to the column before elution of the desalted protein with 30 μL 50 % methanol, 5 % formic acid. The protein solution were then loaded into a 50 μL Hamilton syringe for direct injection (0.2 μL/min) into an Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a Proxeon nano-electrospray ion source (Proxeon, Odense, Denmark). Spectra were recorded for 5 min in the m/z range 700-1700 using highest mass resolution (100,000 at m/z 400). Recorded spectra were analyzed using Qual Browser (part of Xcalibur v 2.0.7) with the plug-in Xtract installed. All spectra were combined in a single summed spectrum showing charge states from 17+ to 37+ followed by deconvolution to the 1+ charge state using Xtrac with the following settings: Mass range: 800 – 1695 Da, S/N threshold: 10, max charge state 35, all other settings as default.

S1.4. Results

Protein identification by MALDI TOF analysis of tryptic peptides extracted from the two SDS-PAGE bands shown in Figure 1a confirmed both the upper band and the lower bands to be CdFic$^{SE/AA}$ (Figure S1a and b). It is therefore possible that the lower Mw bands are a result of CdFic$^{SE/AA}$ degradation yielding two peptides of similar weight as we observe peptides from both the *N*-terminal and the *C*-terminal of CdFic$^{SE/AA}$.

However, analysis by direct injection of CdFic$^{SE/AA}$ into a high resolution mass spectrometer showed that the sample primarily contained a protein with Mr 28.09 kDa as well as a compound with Mr 13.2 kDa (Fig S2a and b). In addition, trace amounts of molecules with Mr ranging from 5 kDa to approximately 35 kDa were observed. This means that the low Mw 13,2 kDa species is present in the solution and is not a consequence of SDS-PAGE. The reason for not seeing two similarly sized fragments of CdFic$^{SE/AA}$ as seen in Fig 1a could be that 1) the other fragment does not bind to the poros 50R1 under the used conditions or 2) the other fragment (if the C-terminal fragment) attracts a high number of positive charges (because of the His-tag) so that it ends up outside the m/z range where data was collected.

(*a*)

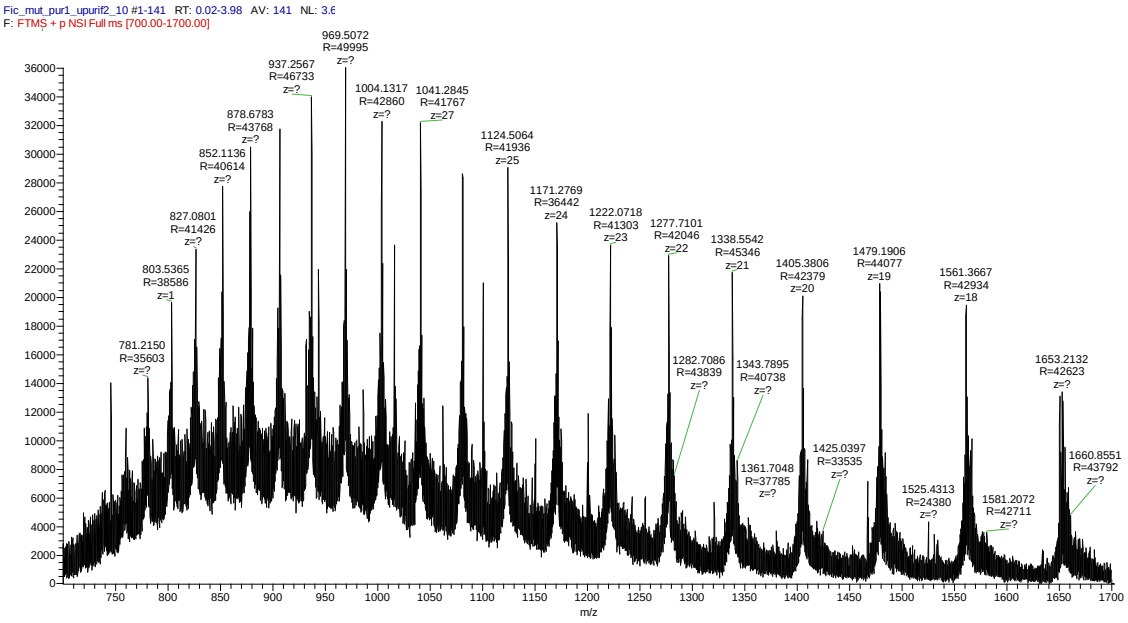
```
  1 MNNSFLDKLI ETKELKNSLY NVLKHNFLYH ANKIAGSTFT TEALALLLDK NVVTGRHTLD
 61 DVQETVNSSY VFDTVIDSLK EKITHNFLRN LHSSLIFNTT LHSRGMAGIY KTIPNMILGT
121 DVSIAQPFEV EPKLDELIEW YYSQSEVSIK VIAEFHYRFE LIHPFQDGNG RIGRFVMLKQ
181 MLENNLPIKI VSWDSEDLYR NSLNSCSLGN YVPLIEYLSS LEDFREVYKM LWKLEHHHHH
241 H
```

(*b*)

```
  1 MNNSFLDKLI ETKELKNSLY NVLKHNFLYH ANKIAGSTFT TEALALLLDK NVVTGRHTLD
 61 DVQETVNSSY VFDTVIDSLK EKITHNFLRN LHSSLIFNTT LHSRGMAGIY KTIPDMILGT
121 DVSIAQPFEV EPKLDELIEW YYSQSEVSIK VIAEFHYRFE LIHPFQDGNG RIGRFVMLKQ
181 MLENNLPIKI VSWDSEDLYR NSLNSCSLGN YVPLIEYLSS LEDFREVYKM LWKLEHHHHH
241 H
```

Figure S1 Identification of CdFic$^{SE/AA}$ by MALDI TOF analysis and database searching. Amino acids highlighted in the identification of CdFic$^{SE/AA}$ from the high Mr band (*a*) and the lower Mr bands (*b*) shown in figure 1a.

(*a*)



(*b*)

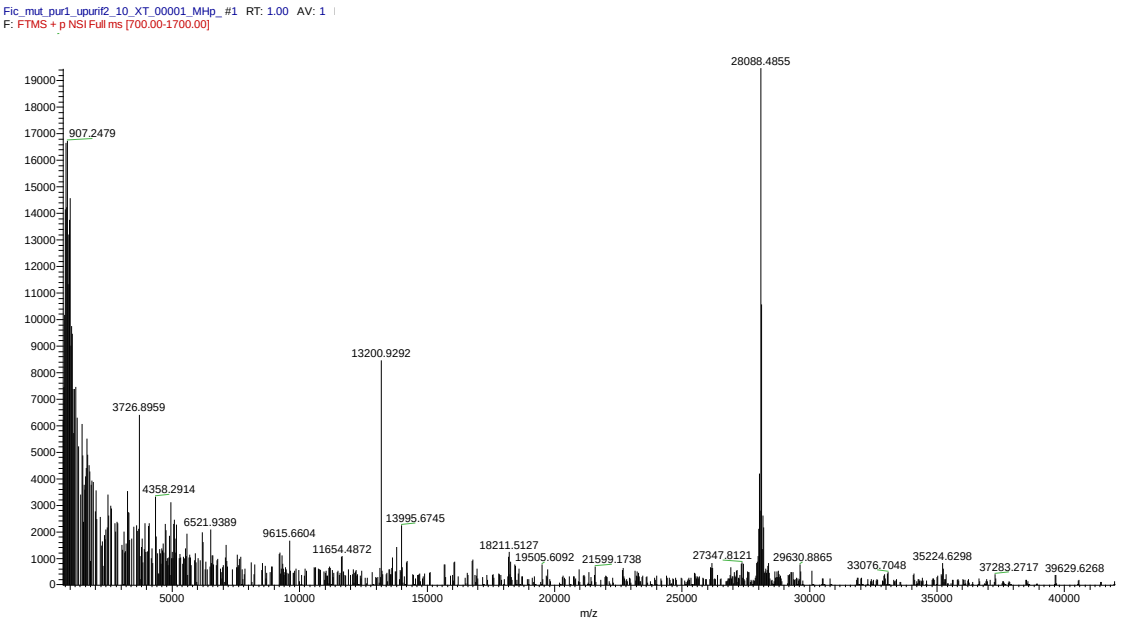


Figure S2 Mr determination of CdFic$^{SE/AA}$ by high resolution mass spectrometry (R=100,000 @ 400 m/z). (*a*) Charge envelope obtained after summation of recorded data showing charge states of CdFic$^{SE/AA}$ from 17+ to 37+. (*b*) Spectrum after deconvolution of all peaks to charge state 1+ showing CdFic$^{SE/AA}$ (Mr = 28.09 kDa) as well a smaller fragment (Mr = 13.20 kDa).

S2. Circular Dichroism

To corroborate the hypothesis that CdFic$^{SE/AA}$ as well as wild type CdFic adopt a predominantly α-helical Fic fold we analysed the samples with circular dichroism (CD) spectroscopy.

S2.1. Materials and Methods for CD spectroscopy

A J-815 CD spectrometer (Jasco) was used to record CD spectra. Spectra were recorded in diluted buffer (20 mM Tris/HCl, pH 7.4, 5 % glycerol and 5 mM β-mercaptoethanol unless otherwise stated) with protein concentrations around 1 mg/ml. Quartz glass far and near UV CD cuvettes with light paths of 0.005 and 0.020 cm were used. CD spectra were collected from 260 nm to 178 nm with 1 nm intervals and 16 s integration time at each wavelength. Minimum 2 scans were taken for each sample. The spectra were averaged and blanks were subtracted. Finally the spectra were converted to Δε per residue molar absorption units of CD (M^{-1} cm^{-1}) using $\Delta\varepsilon = \theta \ (0.1 \cdot MRW)/(3298 \cdot l \cdot c)$, where θ is machine unit in milli degree, *MRW* is protein mean weight per aa in g/mol. *l* is path length in cm and *c* is concentration of the protein in $mg \cdot mL^{-1}$.

Modeling of the CD-spectra was done at the website Dichroweb (Whitmore & Wallace, 2008) with CDSSTR using reference set 1 optimized for 178-260 nm. (Compton & Johnson, 1986; Manavalan & Johnson, 1987; Sreerama & Woody, 2000). The Dichroweb interface builds on the SELCON, CONTIN, CDSSTR, K2D and VARSLC algorithms (Sreerama & Woody, 2000), where CDSSTR is an improved version of the original Varslc program by W.C. Johnson.

S2.2. Results for CD spectroscopy

The results are summarized in Figure S3 and Table S1.

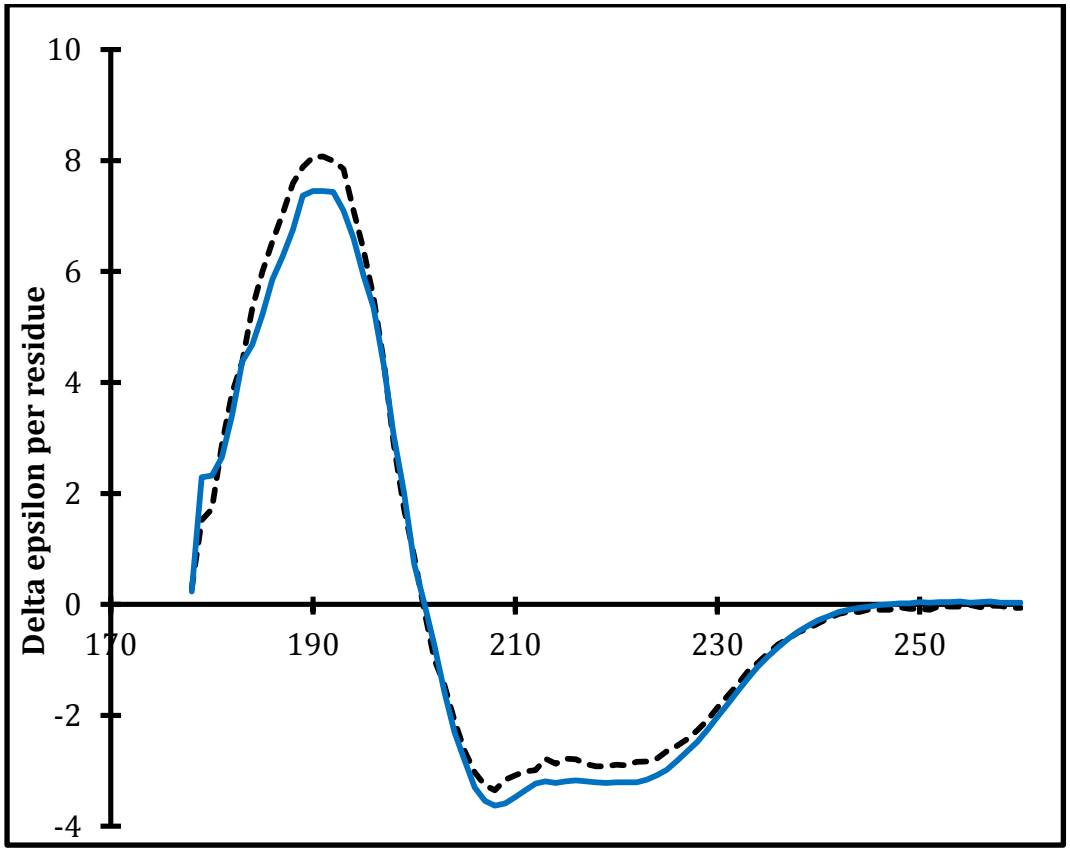


Figure S3 CD spectra of wild type CdFic and CdFic$^{SE/AA}$ confirming the α-helical nature of both proteins. Black: CD spectrum of 1 mg/ml wild type CdFic in 20 mM Tris/HCl, pH 7.5, 5% glycerol. Blue: CD spectrum of 1 mg/ml CdFic$^{SE/AA}$ in 20 mM Tris/HCl, pH 7.4, 5% glycerol, 5 mM, β-mercaptoethanol.

Table S1 Results from Modeling of secondary structure fractions obtained from the CD-spectra of wild type CdFic and CdFic$^{SE/AA}$ with CDSSTR at Dichroweb using reference set 1 optimized for 178-260 nm.

	Wt CdFic	**CdFic$^{SE/AA}$**
α-helix (regular)	24 %	24 %
α-helix (distorted)	15 %	15 %
β-strand (regular)	12 %	13 %
β-strand (distorted)	7 %	7 %
Turns	17 %	16 %
Unordered	25 %	25 %
NRMSD	0.016	0.023

References

Compton, L. A. & Johnson, W. C., Jr. (1986). *Anal Biochem* **155**, 155-167.
Gobom, J., Nordhoff, E., Mirgorodskaya, E., Ekman, R. & Roepstorff, P. (1999). *J Mass Spectrom* **34**, 105-116.
Jensen, O. r. & Wilm, M. (2002). *The Protein Protocols Handbook*, edited by J. Walker, pp. 693-710. Humana Press.
Manavalan, P. & Johnson, W. C., Jr. (1987). *Anal Biochem* **167**, 76-85.
Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. (1999). *Electrophoresis* **20**, 3551-3567.
Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. (2006). *Nat Protoc* **1**, 2856-2860.
Sreerama, N. & Woody, R. W. (2000). *Anal Biochem* **287**, 252-260.
Whitmore, L. & Wallace, B. A. (2008). *Biopolymers* **89**, 392-400.