STRUCTURAL BIOLOGY COMMUNICATIONS

Volume 75 (2019)

Supporting information for article:

The structure of CgnJ, a domain of unknown function protein from the crocagin gene cluster

Sebastian Adam, Andreas Klein, Frank Surup and Jesko Koehnke

Expression and purification of CgnA

For the production of the crocagin precursor peptide CgnA, its coding sequence was cloned into the pHisSUMOTEV vector (a kind gift from Dr. David Owen), and the expression plasmid was transformed into chemically competent *E. coli* Lemo21(DE3) cells (New England Biolabs) using a standard heat shock procedure. For expression, a fresh colony was added into 100 mL of LB Broth starter culture using the Luria-Miller formulation with Kanamycin (50 µg/mL) and Chloramphenicol (34 µg/mL) and was incubated at 37 °C and 200 rpm for 16 h. Expression of CgnA was carried out by inoculating LB Broth with Kanamycin (50 µg/mL) and Chloramphenicol (34 µg/mL) with the starter culture using a 1:100 dilution. Cultures were grown at 37 °C and 200 rpm until the OD_{600} reached 0.8. At this point, protein expression was induced by adding IPTG to a final concentration of 0.4 mM. Cultures were further incubated at 37 °C and 200 rpm for 16 h.

LC-MS

LC-MS was performed on a Dionex Ultimate 3000 RSLC system using BEH C18 column (100 mm x 2.1 mm, 1.7 µm) equipped with a C18 precolumn (Waters). Solvent A was H_2O containing 0.1% formic acid, and solvent B was acetonitrile containing 0.1 % formic acid. Gradient: 0-2.5 min, 5-35 % B; 2.5-5.5 min, 35-42.5 % B; 42.5-95 % B, 5.5-6.0 min or 0-0.5 min, 5% B; 0.5 – 18.5 min, 5 – 95% B; 18.5 – 20.5 min, 95% B; 20.5 – 21 min, 95 – 5% B; 21-22.5 min, 5% B. After a 2 min step at 95 % B the system was re-quilibrated to the initial conditions (5 % B). The UV spectrum was recorded by a DAD in the range from 200 to 600 nm.

MS was performed using an amaZon speed mass spectrometer (Bruker Daltonics). The LC flow was split 1:8 before entering the mass spectrometer using the Apollo ESI source. The following conditions were used: capillary voltage 4500 V, temperature 300 °C, dry-gas flow rate 10 L/min and nebulizer 30 psi. Data was recorded in the mass range from 500 to 2000 m/z.

Dynamic light scattering (DLS)

Dynamic light scattering (DLS) experiments were carried out on a Zetasizer Nano ZS (Malvern Panalytical) at room temperature. CgnJ was diluted to 50 µM in water and size measurements were carried out in triplicates with 15 measurements each. The mean sizes of 15 runs were calculated for each run with the *Zetasizer Nano* software. The mean size and standard deviation of 45 runs was calculated using Microsoft Excel and compared to the apparent sizes of monomer or dimer formation of the protein.

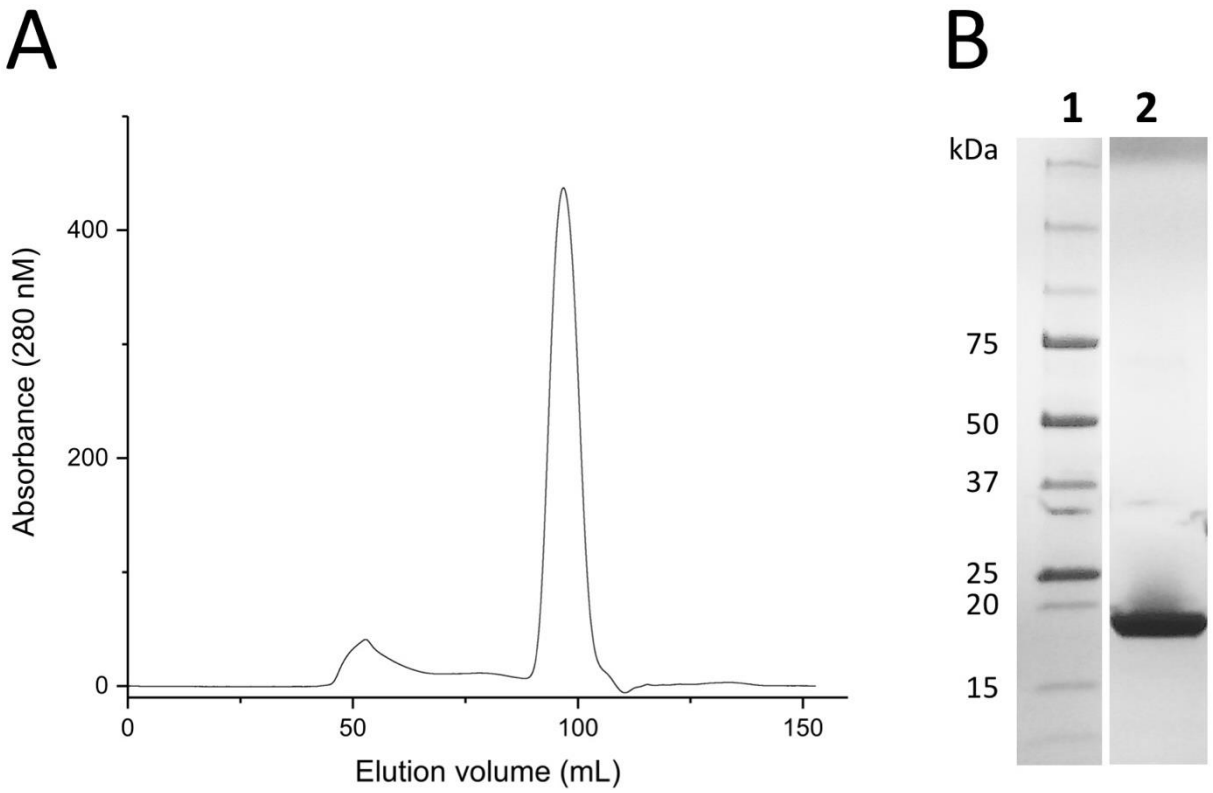


Figure S1 A Size exclusion chromatography of CgnJ using a Superdex 200pg 16/60 column. A minor portion of the protein aggregates during purification (~50 mL elution volume) while the majority of the protein elutes as a symmetrical, single peak (~100 mL elution volume) shortly before the elution volume reaches one column volume (121 mL) indicating a monomer in solution). **B** SDS-PAGE of CgnJ after size-exclusion chromatography.

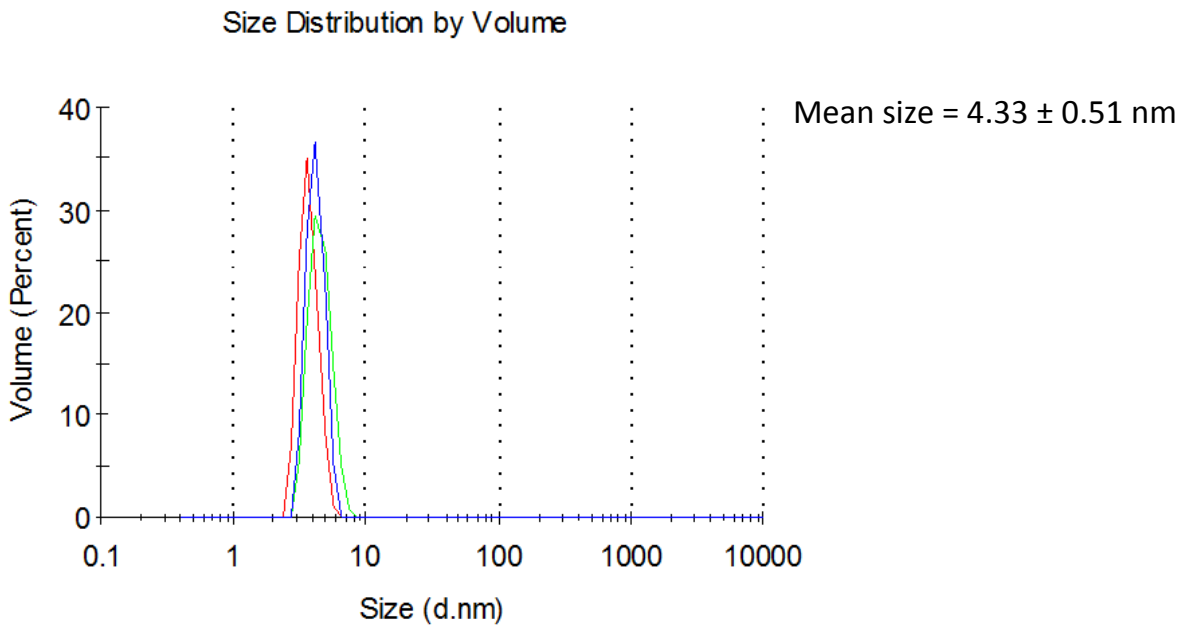


Figure S2 Dynamic light scattering measurement of CgnJ. The mean size and standard deviation of 45 runs was calculated using Microsoft Excel. The maximum size of a CgnJ monomer was measured to be ~ 45 Å using *Pymol*.

```
CgnJ      MAE---EVAEIILPASTWILFFDASCSINSPAFWSTNDAVDRIWRLKIAHELVLLQVVLE      57
ComJ      MAIKSWKPQELSISYHQFTVFQKDS--TPPVMDWTD-EAIE-----------------K      39
          **    :  *: :    : :* . *        *:  :*::                  :

CgnJ      GYFKVRCIL----RSSAPAFEMVNADVSELVS-----IVLPSGRLV-ACTTDEPTLNRHV      107
ComJ      GYAAADGAISFEAQRNTKAFILFRLNSSETVNSYEKKVTVPFHVTENGIHIESIMSKRLS      99
          **  .   :    : .: ** :.. : ** *.     :.:*      .   :.   :*

CgnJ      LTVPPGRYRVLREWSVHEESKHYDVESAEAYPADEGPDGIITLWPER  154
ComJ      FDLPKGDYQL-TCWTVPAEMSDLHAD---TYIIDAVSV---------  133
          : :* * *::   *:*  * .. ..:   :*  *
```

Figure S3 Sequence alignment of CgnJ and ComJ calculated by Clustal Omega. The sequence identity of both proteins was calculated to be 15.6 %.

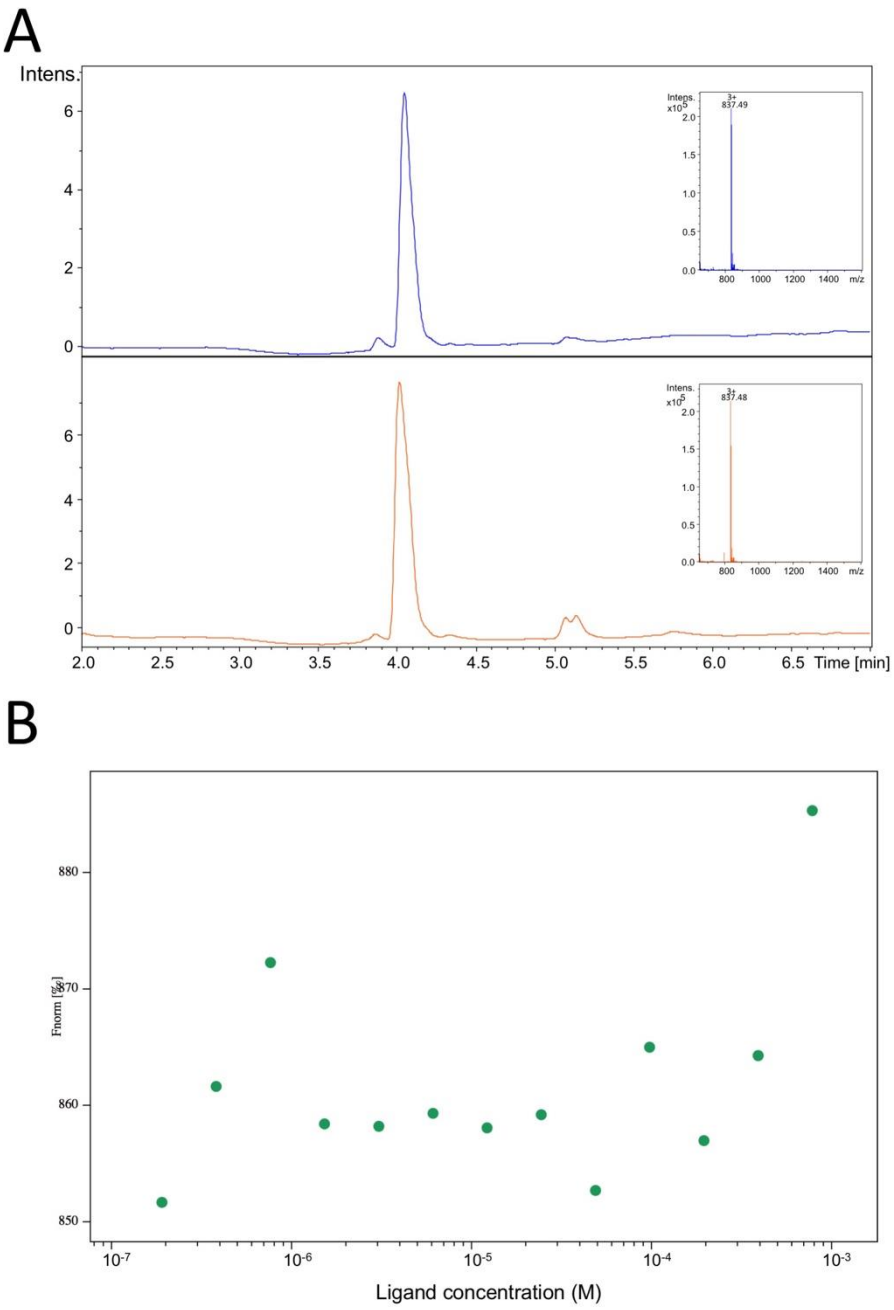
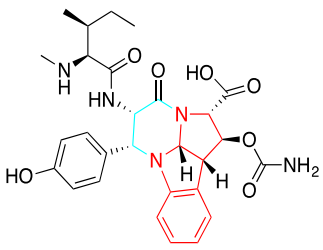
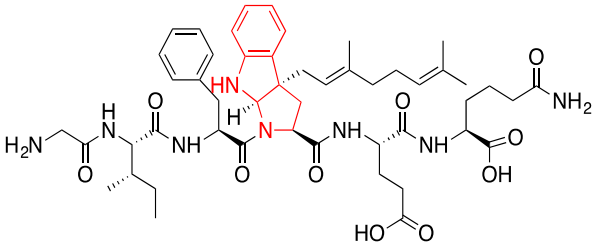


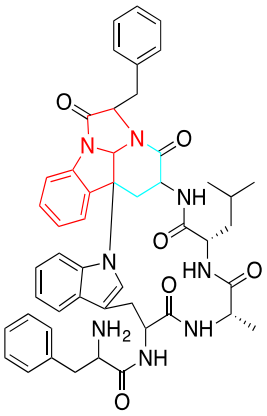
Figure S4 A LC-MS analysis of the precursor peptide CgnA (blue, top) and CgnA after the addition of CgnJ (orange, bottom). Both samples were incubated at 37 °C for 24 h prior to analysis. **B** MST measurement of CgnA with CgnJ. No binding was observed.



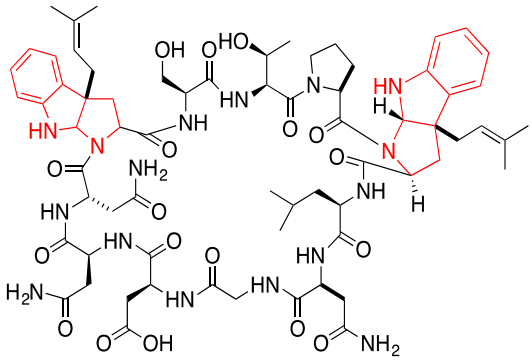
Crocagin A



ComX (RO-E-2)



Kapakahine B



Kawaguchipeptin A

Figure S5 Chemical structures of Crocagin A, ComX (RO-E-2), Kapakahine B and Kawaguchipeptin A. Similar tryptophan residues are highlighted in red whereas a similar core structure in Crocagin A and Kapakahine B is marked in cyan.