Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

Segmental isotopic labelling of a central domain in a multi-domain protein by protein trans-splicing using only one robust DnaE intein

Alena E.L. Busche, A. Sesilja Aranko, Mehdi Talebzadeh-Farooji, Frank Bernhard, Volker Dötsch, and Hideo Iwaï

Supplementary Methods and Materials

Construction of vectors

Three-fragment ligation with the model proteins

The N-terminal precursor protein containing H_6-Smt3-*Npu*$Int_{N\Delta C15}$ is encoded in the plasmid pHYRSF53-36, which was created by transferring Smt3 gene into the previously described plasmid pHYRSF49-36.[1,2] The gene of Smt3 was amplified by the two oligo nucleotides #HK166 (5'-ACATATGGGATCCACCAATCTGTTCTCTG) and #HK156 (5'-GACTAGTGACTCAGAAGTCAATCAAGAAGCTAAG) and cloned into pHYRSF49-36 using the *Spe*I and *BamH*I sites, resulting in the plasmid pHYRSF53-36. The plasmid pSARSF120 for the central precursor protein (H_6-*Npu*Int_{C15}-GB1-*Npu*$Int_{N\Delta C15}$) was constructed by inserting the gene of the C-intein *Npu*Int_{C15}, which was amplified from pHYBAD44 using the two oligonucleotides #SZ015 (5'-TGCCAAGCTT ATTCCGTTAC GGTG) and #HK146 (5'-TACATATGGA CCATAATTTT GCACTC), at the front of GB1-*Npu*$Int_{N\Delta C36}$ in the plasmid pHYRSF-1 using *Nde*I and *Sac*II sites. For the C-terminal precursor protein containing *Npu*Int_{C36}-YFP, the gene of yellow fluorescent protein (YFP) was amplified from pHVW (FlyBase) using oligonucleotides #SA002 (5'-GTGGTACCGGCAAGGGCGAGGAGC) and #HK031 (5'-CGC AAG CTT AAG TGA TCC CGG CGG CGG) and thereafter ligated into pSKBAD2A using *Kpn*I and *Hind*III restriction sites, resulting in pSABAD128 (*Npu*Int_{C36}-YFP).

Plasmids for the expression of T1-T2-T3 from CurA

The three consecutive ACP domains (T1, T2, and T3) located at the C-terminus of the polypeptide CurA were amplified from genomic DNA of *Lyngbya majuscula* strain 19L (a gift from Dr. C. Walsh, Harvard University, USA) using the oligonucleotides: #CurAT1T2T3-NdeI-F (5'-GGGCATATGG GCAGCAGCCA TCATCATCA), #CurAT1T2T3-XhoI-R (5'-GGCTCGAGTC ACAGCTTAGA ACCACCAGTA). The PCR product was digested with *Nde*I and *Xho*I and cloned into pET-28b(+) vector (Novagen), resulting in the new expression vector pet28-01.

Plasmid for three-fragment ligation of CurA

The first domain (T1) of the three consecutive ACP domains of the CurA protein was amplified using the primers: #I1CurAT1-NcoI-F (5'-GGCCATGGCC ACTCCTCAGG TAAATCAAGT) and #I1CurAT1-BamHI-R (5'-CCGGATCCCG GTTTGGTACC CTGAGAGCTC A). For the construction of the N-terminal precursor protein (H_6-Smt3-T1-*Npu*$Int_{N\Delta C15}$), the PCR product was digested with *Nco*I and *Bam*HI and cloned into pHYRSF66-36, resulting in the new expression vector pABRSF-1. pHYRSF66-36 was derived from pHYRSF49-36 by eliminating *Nco*I site using the two oligonucleotides #HK181 (5'-CTTTAATAAG GAGATATAACA TGGGCAGCAG) and #HK182 (5'-CTGCTGCCCA TGTTATATCT CCTTATTAAAG) and by replacing the *Nde*I site with *Nco*I site using the two oligonucleotides #HK183 (5'-CAGATTGGTG GTTCCATGGA GTACAAACTT ATCC) and #HK184 (5'-GGATAAGTTT GTACTCCATG GAACCACCAA TCTG).[1] This plasmid expresses H_6-Smt3-T1-*Npu*$Int_{N\Delta C15}$ fusion protein upon induction with isopropyl β-D-thiogalactoside (IPTG). The second domain (T2) of the CurA protein was amplified using the primers: #I2CurAT2-KpnI-F (5'-CCGGTACCCA GCAATCTCTG AAAA) and #I2CurAT2-HindIII-R (5'-CCAAGCTTTT ATTTAGTGCC CTGAGAGGCC A). The PCR product was digested with *Kpn*I and *Hind*III and cloned into the vector pSABAD120. The pSABAD120 was constructed by transferring the inserted gene in pSARSF120 into pSKBAD2 by using the restriction sites of *Nde*I and *Hind*III. The new expression vector pABBAD02 expressed the fusion protein of the C-terminal 15 residues of *Npu*DnaE-Int_C with T2 (H_6-*Npu*Int_{C15}-T2) for the ligation with T1. In order to construct the central precursor protein for the three-fragment ligation, the plasmid

pABRSF03 was constructed by transferring the gene of *Npu*Int$_{C15}$ and T2 into pHYRSF49-36 by using *Nde*I and *BamH*I sites.[1] The resulting plasmid pABRSF-3 expresses H$_6$-Smt3-*Npu*Int$_{C15}$-T2-*Npu*Int$_{N\Delta C36}$ fusion upon induction with IPTG. For the C-terminal precursor, the gene of the third ACP domain (T3) of the CurA protein was amplified using the two primers #IBCurAT3-KpnI-F (5'-CCGGTACCAA AACCCTGCAG CCGCTGCC) and #IBCurAT3-HindIII-R (5'-GGAAGCTTTC ACAGCTTAGA ACCACCAGTA GCAGC). The PCR product was digested with *Kpn*I and *Hind*III and cloned into the vector pHYRSF1-02.[2] The new expression vector pABRSF-2 expressed the fusion protein of H$_6$-*Npu*Int$_{C36}$-T3 under the control of T7 promoter. The plasmid pABBAD-15 was constructed by transferring the gene encoding *Npu*Int$_{C36}$-T3 into pBAD vector using *Nde*I and *Hind*III sites. The resulting plasmid pABBAD-15 expresses the fusion protein of *Npu*Int$_{C36}$-T3 upon induction with L-arabinose

Three-fragment ligation of the model proteins

The first step in vivo *ligation with the model proteins GB1 and YFP*

The two plasmids encoding H$_6$-*Npu*Int$_{C15}$-GB1-*Npu*Int$_{N\Delta C36}$ (pSARSF120) and *Npu*Int$_{C36}$-YFP (pSABAD128) were transformed into *E. coli* ER2566 cells (New England Biolab). 2 × 10 ml of a 20 ml overnight culture of the transformed cells was inoculated into 2 × 0.5 L LB-medium supplemented with 100 μg/ml ampicillin and 25 μg/ml kanamycin and grown at 37 ˚C. *Npu*Int$_{C36}$-YFP in pSABAD128 containing araBAD promoter was expressed by the induction with 0.08% (w/v) L-arabinose when OD_{600} = 0.6 was reached. After half an hour of the induction with arabinose, a final concentration of 1.0 mM IPTG was added in order to induce the expression of the 2nd precursor protein of H$_6$-*Npu*Int$_{C15}$-GB1-*Npu*Int$_{N\Delta C36}$ in pSARSF120 bearing T7/lac promoter and the subsequent ligation of the two precursor proteins. The induction of the cells with L-arabinose and IPTG continued for another four hours before the cells were harvested by centrifugation at 6,900 ×*g* for 10 min. The cell pellet was re-suspended with lysis buffer (50 mM Na phosphate, 300 mM NaCl, pH 8.0) and stored at –80 ˚C for further purification.

Purification of the ligated product from the first step in vivo ligation with the model proteins GB1 and YFP

The harvested cells were thawed and lysed by ultrasonification. The cell lysate was cleared by centrifugation at 40,000 ×*g* for 55 min and purified by BioRad Profinia purification station with Ni-NTA-column according to manufacturers' protocol. Purified proteins were dialyzed against 0.5 M NaCl, 10 mM Tris, 1 mM EDTA, pH 7.0. Total amount of protein was about 10 mg. About 64% of the precursor proteins were ligated.

The second step in vitro *protein ligation of the model proteins*

The ligated product (H_6-*Npu*Int_{C15}-GB1-YFP) from the first *in vivo* protein ligation was mixed at 1:1 molar ratio with H_6-Smt3-*Npu*$Int_{N\Delta C15}$, which was expressed and purified as described previously.[2] The final concentration of the proteins was 12 μM in the presence of 0.5 mM tris(2-carboxyethyl)phosphine (TCEP). The reaction mixture was incubated at 25 ºC with agitation at 350 rpm for 15 hours. For SDS-PAGE analysis a small portion (8 μl) was taken from the reaction mixture after 0 min, 3 min, 10 min, 30 min, 1 hour, 3 hours, 4 hours and 15 hours. The reaction was stopped by mixing each sample with 1 × SDS sample buffer. The samples were stored on ice and loaded on 18% SDS-gel after heating at 95 ºC for 5 min. The ligation mixture was sent for analysis by MALDI-TOF mass spectrometry.

Segmental isotopic labeling of a central domain of CurA (T1-T2-T3)

*Protein expression and purification of CurA (T1-T2-T3) and H_6-Smt3-T1-*Npu*$Int_{N\Delta C15}$*

Apo-*Cur*A (T1-T2-T3) was expressed in *E.coli* ER2566 using the plasmid pET28-01. 40 ml of overnight culture were spun down at 1000 ×*g* and inoculated into 2 L ^{15}N-labelled M9 medium supplemented with 25 μg/ml kanamycin and grown at 37 °C. The M9 media contains 0.05 mM $FeCl_2$ which inhibits the apo to holo conversion of the protein[5]. The cells bearing the expression vector were grown to an OD of 0.6 at 37 °C and then transferred to 20 °C and induced with 0.4 mM IPTG at an OD of 0.7–0.9. After 6 hours of the induction with IPTG, the cells were harvested. The protein was purified

by Immobilized Metal ion Affinity Chromatography (IMAC) using Ni-NTA superflow (QIAGEN). Total amount of purified protein was estimated to be 32 mg. Unlabeled H_6-Smt3-T1-*Npu*$Int_{N\Delta C15}$, which is a precursor protein for the three-fragment ligation, was expressed using the plasmid pABRSF-1 in 1.4 L of LB medium supplemented with 25 μg/ml kanamycin and 0.01 mM $FeSO_4$. The *E.coli* ER2566 cells bearing pABRSF-1 were grown to an OD of 0.6 at 37°C and then transferred to 30°C, followed by the induction with a final concentration of 0.4 mM IPTG at an OD of 0.7–0.8. The cells were grown for another 6 hours and harvested by centrifugation at 9,000 ×*g* for 15 min. The pellet was frozen and stored at –80 °C for further purification.

The first step in vivo *protein ligation for central fragment labeling of CurA (T1-T2-T3)*

E.coli ER2566 bearing the two plasmids pABRSF03 (H_6-Smt3-*Npu*Int_{C15}-T2-*Npu*$Int_{N\Delta C36}$) and pABBAD-15 (*Npu*Int_{C36}-T3) were grown at 37°C in 2 L of LB medium supplemented with two antibiotics 100 μg/ml ampicillin and 25 μg/ml kanamycin and 0.01 mM $FeSO_4$. The C-terminal precursor protein (*Npu*Int_{C36}-T3) was first induced with 0.1% (w/v) L-arabinose at an OD of 0.4–0.5 for 4 hours and transferred to another shaker adjusted to 30°C. The culture medium was replaced with 2 L of ^{15}N-labelled M9 medium supplemented with 25 μg/ml kanamycin, 0.05mM $FeCl_2$, and a final concentration of 0.4 mM IPTG. The replacement of the medium was done by spinning down the cells at 900 ×*g* with a rotor SLA-3000. The cell pellets were gently resuspended with 100 ml of the labelled M9 medium and then span down once more at 900 ×*g* for 10 min. The cell pellets were then resuspended with 2.0 L of the ^{15}N-labelled M9 medium. After further 5 hours of the second induction, the cells were harvested at 9,000 ×*g* for 10 min and frozen at –80 °C. 20 mg of the segmentally isotope-labelled ligated product of H_6-Smt3-*Npu*Int_{C15}-T2-T3 was obtained after IMAC purification.

The second step in vitro protein ligation for central fragment labeling of CurA (T1-T2-T3)

The purified two precursor proteins, H_6-Smt3-*Npu*Int_{C15}-T2-T3 and H_6-Smt3-T1-*Npu*$Int_{N\Delta C15}$ were dialyzed against PBS, pH 7.4. About 20 mg of

H_6-Smt3-*Npu*Int$_{C15}$-T2-T3 were mixed with a total amount of 5 mg H_6-Smt3-T1-*Npu*Int$_{N\Delta C15}$. The final concentration of H_6-Smt3-*Npu*Int$_{C15}$-T2-T3 was 50 μM in the reaction mixture. H_6-Smt3-T1-*Npu*Int$_{N\Delta C15}$ had a final concentration of 11 μM in the reaction mixture. The reaction mixture was incubated at room temperature in the presence of 0.5mM TCEP for 10 hours. In the reaction mixture yeast Ubiquitin-Like Protein-Specific Protease 1(Ulp1) was added for removal of the H_6-Smt3 fusion tag.[3] After 10 hours the buffer was exchanged with 10 mM Tris, 500 mM NaCl, 0.5 mM EDTA, 0.5 mM TCEP using a centrifugal filter unit (MWCO 5000) (Amicon). In this ligation buffer (10 mM Tris, 500 mM NaCl, 0.5 mM EDTA, 0.5 mM TCEP), the reaction was continued for further 7 hours at 25 °C with agitation at 200 rpm. After the reaction, the H_6-Smt3 fusion tag was removed by IMAC. The flow through from IMAC was collected and concentrated by a centrifugal filter unit (MWCO 5000) (Amicon). The buffer was exchanged to 10 mM Na phosphate, pH 7.2 for further purification with ion exchange chromatography. The ligated product was purified by anion exchange chromatography using MonoQ 5/50GL. Fractions were collected and dialysed for over night against 5 L of 50 mM Na phosphate, 100 mM NaCl, pH 7.2. The protein concentrated to 500 μl with a centrifugal filter unit for NMR measurements. The total amount of segmentally ^{15}N-labeled T1-[^{15}N]-T2-T3 was 1.7 mg.

Segmental isotopic labeling of T1 domain in CurA (T1-T2-T3)

Protein expression and purification of labeled H_6-Smt3-T1-Npu*Int$_{N\Delta C15}$*

The N- terminal precursor protein was expressed in *E.coli* ER2566 using the plasmid pABRSF01. 40 ml of the 100 ml pre-culture were spun down at 1,000 ×*g* and inoculated into 2 L ^{15}N-labelled M9 medium supplemented with 25 μg/ml kanamycin and grown at 37 ˚C. The M9 media contains 0.05 mM $FeCl_2$ which inhibits the apo to holo conversion of the protein.[4] The cells bearing the expression vector were grown to an OD of 0.5 at 37 °C and then transferred to 20 °C and induced with 0.4 mM IPTG at an OD of 0.6–0.7. After 6 hours of the induction with IPTG, the cells were harvested. The protein was purified by Immobilized Metal ion Affinity Chromatography (IMAC) using Ni-NTA superflow (QIAGEN). Total amount of purified protein was estimated to be 40 mg.

Coexpression and Purification of unlabeled C-terminal precursor H_6-Smt3-NpuInt_{C15}-T2-T3

E.coli ER2566 bearing the two plasmids pABRSF-3 (H_6-Smt3-*Npu*Int_{C15}-T2-*Npu*$Int_{N\Delta C36}$) and pABBAD-15 (*Npu*Int_{C36}-T3) were grown at 37 °C in 2 L of LB medium supplemented with two antibiotics 100 µg/ml ampicillin and 25 µg/ml kanamycin and 0.01 mM $FeSO_4$. At an OD of 0.4 the cells were transferred to another shaker adjusted to 30 °C. The C-terminal precursor protein (*Npu*Int_{C36}-T3) was first induced with 0.1% (w/v) L-arabinose at an OD of 0.5–0.6 for 1 hour. Then the N-terminal precursor protein was induced with 0.4 mM IPTG. The two proteins were co-expressed for further 7 hours. The cells were harvested at 9,000 ×*g* for 10 min and frozen at –80 °C. 160 mg of the unlabeled, ligated product of H_6-Smt3-*Npu*Int_{C15}-T2-T3 were obtained after IMAC purification.

The in vitro protein ligation for N-terminal T1 domain labeling of CurA (T1-T2-T3)

The purified two precursor proteins, H_6-Smt3-*Npu*Int_{C15}-T2-T3 and ^{15}N-labelled H_6-Smt3-T1-*Npu*$Int_{N\Delta C15}$ were dialyzed against ligation buffer (10 mM Tris, 500 mM NaCl, 0.5 mM EDTA) for the second step *in vitro* ligation. About 42 mg of H_6-Smt3-*Npu*Int_{C15}-T2-T3 were mixed with a total amount of 40 mg H_6-Smt3-T1-*Npu*$Int_{N\Delta C15}$. The final concentration of H_6-Smt3-*Npu*Int_{C15}-T2-T3 and H_6-Smt3-T1-*Npu*$Int_{N\Delta C15}$ was 50 µM in the presence of 0.5 mM tris(2-carboxyethyl)phosphine (TCEP). The reaction mixture was incubated at room temperature for 13 hours with agitation. Then the yeast Ubiquitin-Like Protein-Specific Protease 1 (Ulp1) was added for removal of the H_6-Smt3 fusion tag and incubated for further 12 hours. After the reaction, the H_6-Smt3 fusion tag was removed by IMAC. The flow through fraction from IMAC was collected and concentrated by a centrifugal filter unit (MWCO 10000) (Amicon). 30 mg of the ligated protein were obtained and the buffer was exchanged to 10mM Na phosphate, pH 7.2 for further purification with ion exchange chromatography. The ligated product was purified by anion exchange chromatography using MonoQ 5/50GL. The fractions containing the ligated protein were collected and dialysed for over night against 5 L of 50 mM Na phosphate, 100 mM

NaCl, pH 7.2. For the NMR measurements, the protein was concentrated to 0.4 mM using a centrifugal filter unit for NMR measurements. The total amount of segmentally [^{15}N, ^{15}C]-labelled [^{15}N/^{13}C]-T1-T2-T3 was 13 mg.

NMR spectroscopy

All the NMR measurements were performed on either a Bruker Avance Spectrometer or Varian INOVA spectrometer at ^{1}H frequency of 600 MHz, which are equipped with cryogenic probe heads. The spectra were recorded with 80 μM sample for the segmentally isotope-labelled T1-[^{15}N]-T2-T3 and with a 0.4 mM sample for uniformly ^{15}N-labelled *Cur*A (T1-T2-T3) at 17 °C. The sample conditions were 50 mM Na phosphate, pH 7.2, 100 mM NaCl with 1 mM $^{2}H_{12}$-DTT.

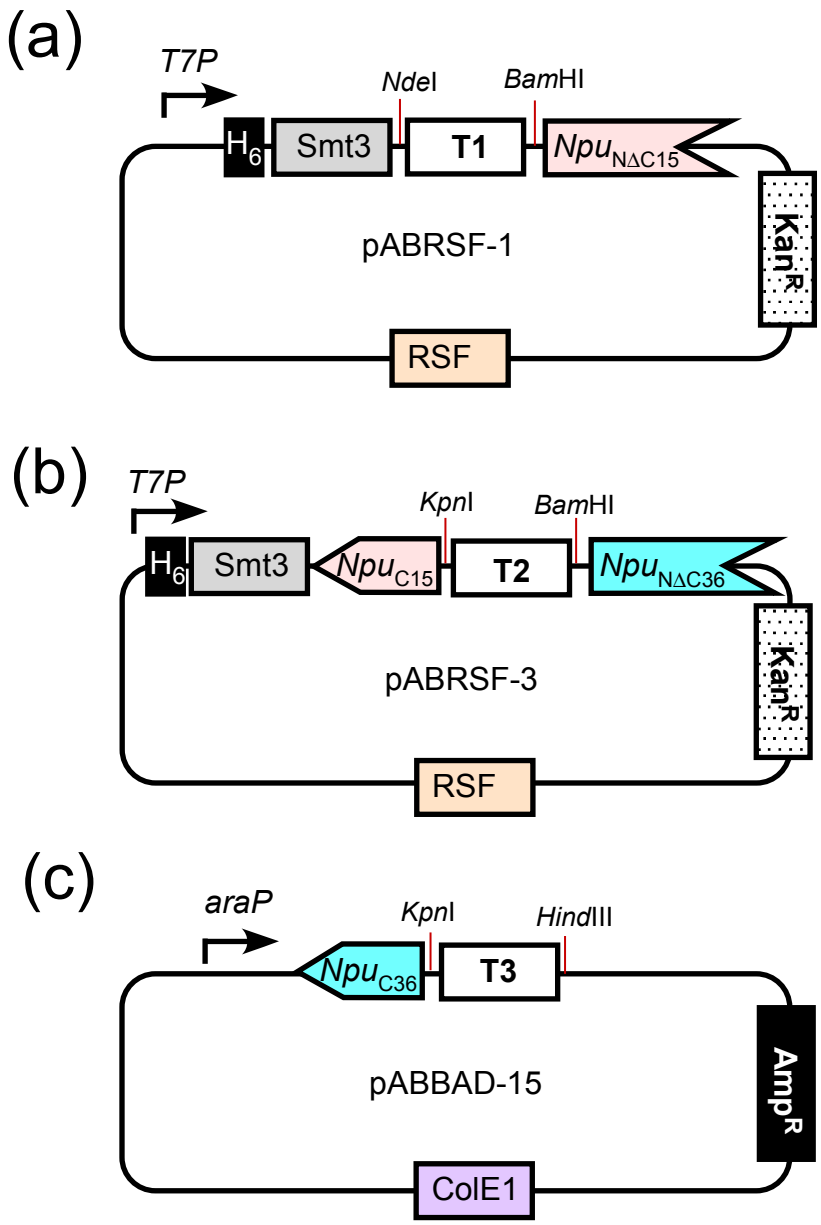
References

1. Muona, M. Aranko, A. S., and Iwai, H. (2008) Segmental isotopic labelling of a multi-domain protein by protein ligation using protein *trans*-splicing. *ChemBioChem*. **9**, 2958−2961

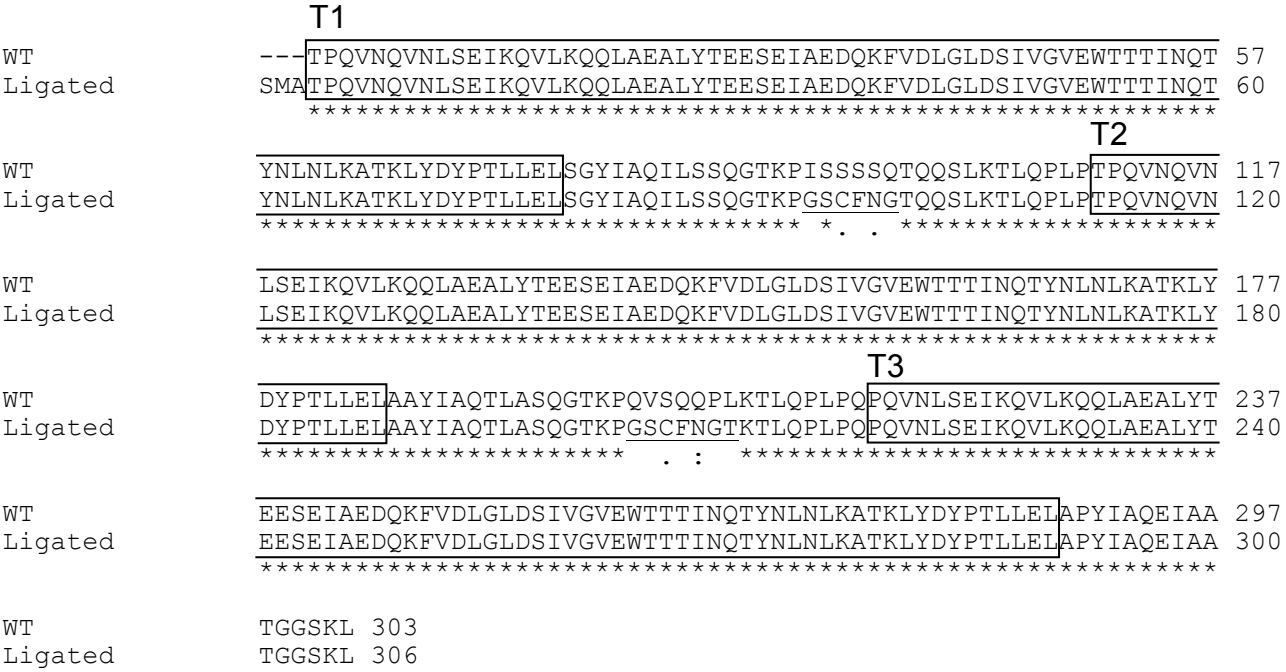
2. Aranko, A.S., Züger,S., Buchinger, E., and Iwaï, H. (2009) *In vivo* and *in vitro* protein ligation by naturally occurring and engineered split DnaE. *PLoS ONE* **4**, e5185.

3. Mossessova, E. & Lima, C.D. (2000) *Mol. Cell* **5**, 865−876.

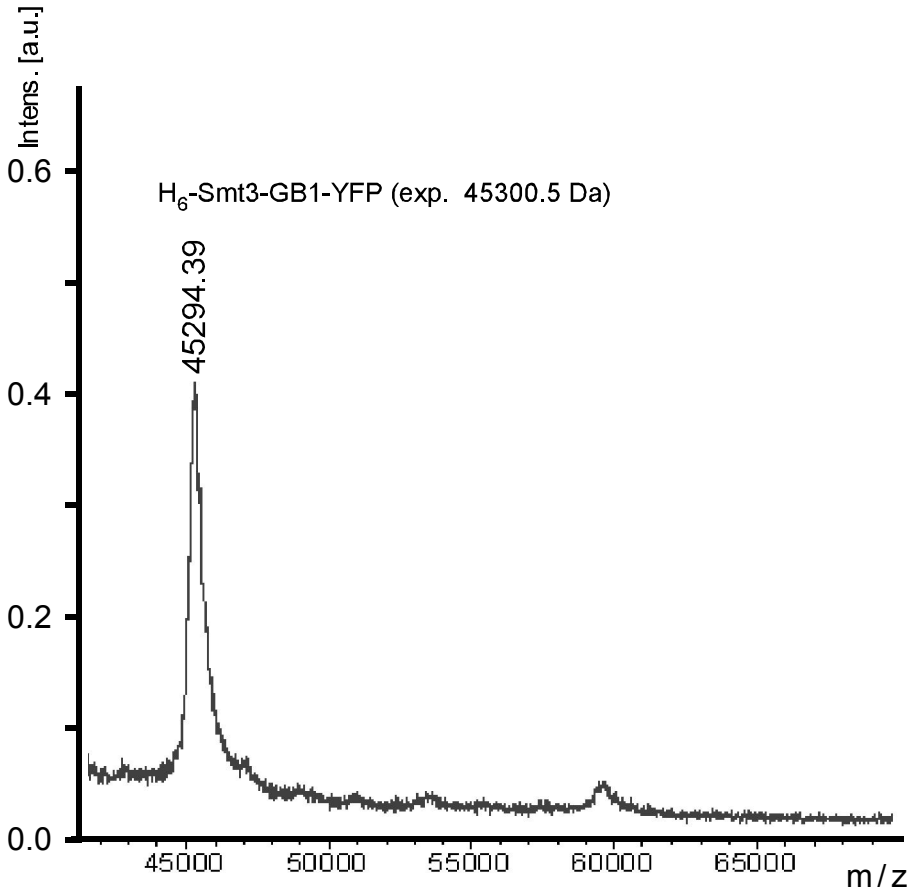
4. Weber, T., R. Baumgartner, et al. (2000) Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure* **8**, 407−18.



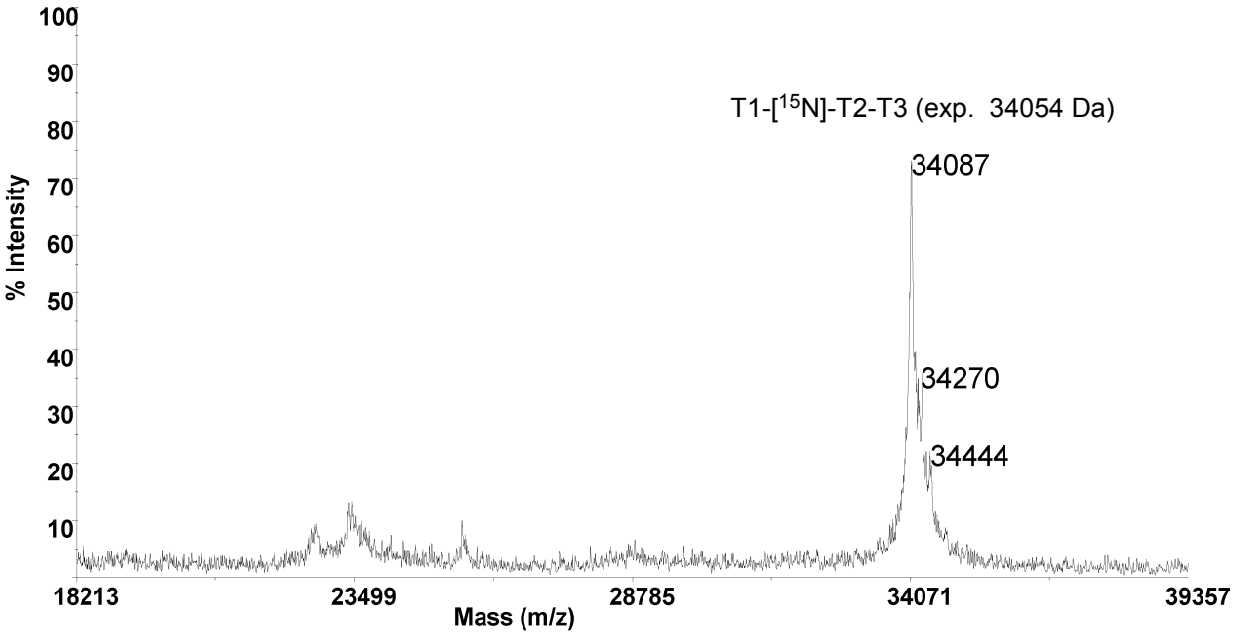
Suppl. Fig. 1: Maps of the plasmids used for segmental isotopic labelling of CurA by three-fragment ligation (a) N-terminal precursor. (b) Central precursor. (c) C-terminal precursor.



Suppl. Fig. 2 : Sequence alignment of wild-type CurA (T1-T2-T3) and the ligated CurA (T1-T2-T3) by three-fragment ligation. The sequences for the linkers used for ligation are underlined. T1,T2,T3 domains are indicated by boxes.



Suppl. Fig. 3 : MALDI-TOF mass spectrum of the *in vitro* ligated product of the three model proteins.



Suppl. Fig. 4 : Mass spectrum of the segmentally isotope-labelled T1-[^{15}N]-T2-T3.