Supplementary Information

Intermolecular domain swapping induces intein-mediated protein alternative splicing

A. Sesilja Aranko, Jesper S. Oeemig, Tommi Kajander, & Hideo Iwaï

Nature Chemical Biology: doi:10.1038/nchembio.1320

Supplementary Results

Supplementary Table 1 Data collection and refinement statistics

	Δvariant	*Npu*DnaE intein
Data collection		
Space group	P1	$P2_12_12_1$
Cell dimensions		
a, *b*, *c* (Å)	39.72, 64.01, 64.22	57.55, 66.70, 67.48
α, β, γ (°)	109.0, 99.53, 95.52	90, 90, 90
Resolution (Å)	50-2.20 (2.33-2.20)	50-1.72 (1.82-1.72)
R_{merge}	6.8 (34.2)	6.5 (59.1)
$I/\sigma I$	9.02 (2.12)	21.13 (3.67)
Completeness (%)	92.2 (90.6)	99.2 (96.3)
Redundancy	1.85 (1.85)	7.04 (7.06)
Refinement		
Resolution (Å)	2.20	1.72
No. reflections	50474	198900
R_{work}/ R_{free}	0.197/0.266	0.169/0.209
No. atoms	4580	2558
Protein	4298	2310
Ligand/ion	-	32
Water	282	217
B-factors		
Protein	25.9	21.50
Ligand/ion	-	52.03
Water	29.9	34.70
R.M.S. deviations		
Bond lengths (Å)	1.278	1.252
Bond angles (º)	0.010	0.010

One crystal was used for each structure.
*Highest resolution shell is shown in parenthesis.

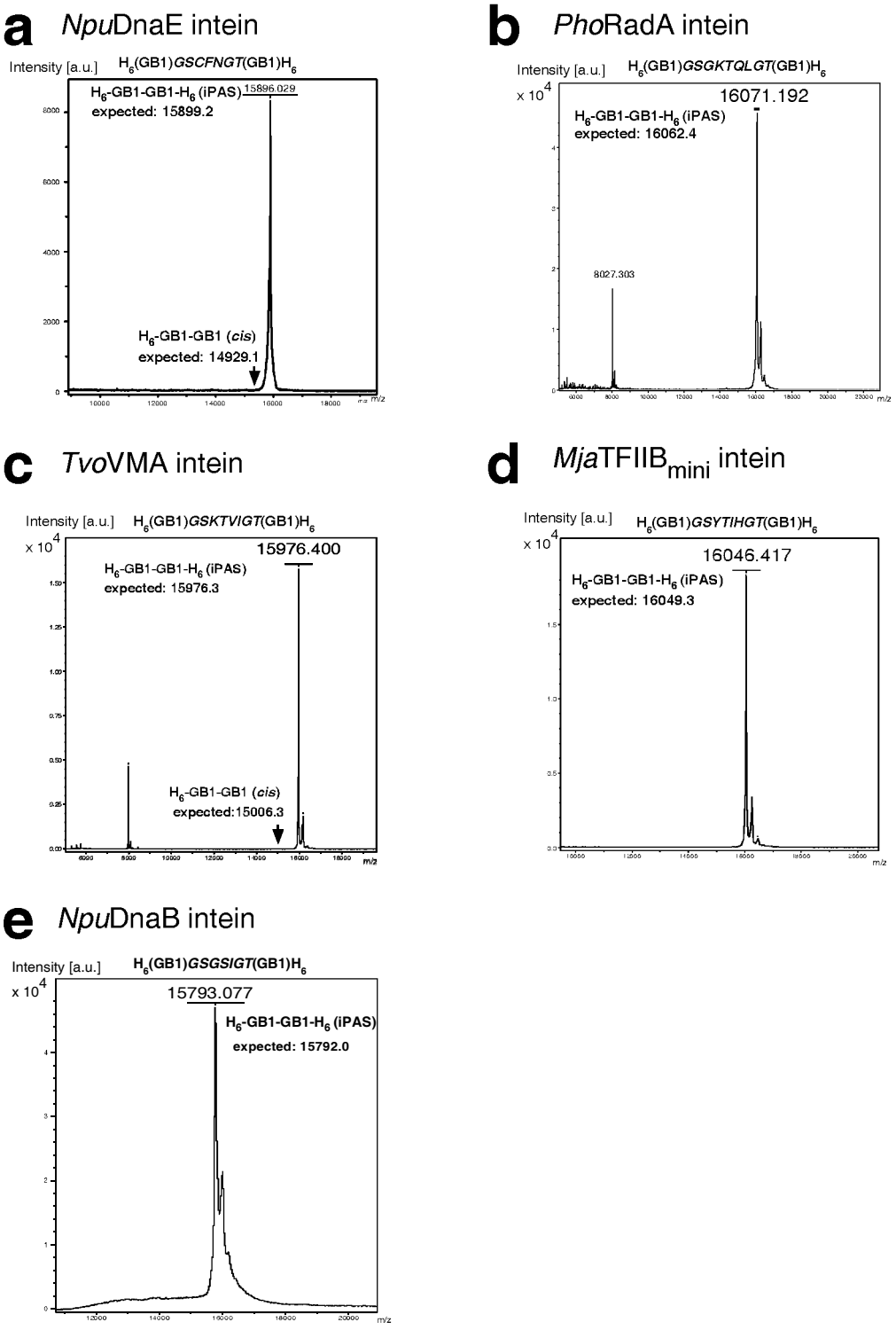
Supplementary Table 2 List of organisms harboring more than two inteins

Name of Organisms	# of inteins	# of split inteins	# of proteins with multiple intein insertions
Methanococcus jannaschii (Mja)	19	0	3
Haloquadratum walsbyi DSM 16790 (Hwa)	15	0	4
Thermococcus kodakaraensis KOD1 (Tko)	15	0	3
Trichodesmium erythraeum IMS101 (Ter)	14	1	4
Pyrococcus abyssi (Pab)	14	0	3
Pyrococcus horikoshii OT3 (Pho)	14	0	1
Thermococcus litoralis (Tli)	12	0	3
Pyrococcus furiosus (Pfu)	10	0	1
Methanococcus aeolicus Nankai-3 (Maeo)	6	0	0
Microcoleus chthonoplastes PCC7420 (Mch)	5	1	2
Methanocaldococcus vulcanius M7 (Mvu-M7)	5	0	1
Methanopyrus kandleri AV19 (Mka)	5	0	0
Crocosphaera watsonii WH 8501 (Cwa)	4	1	0
Thermus thermophilus HB8 (Tth-HB8)	4	0	2
Thermus thermophilus HB27 (Tth-HB27)	4	0	2
Natronomonas pharaonis DSM 2160 (Nph)	4	0	1
Batrachochytrium dendrobatidis JEL423 (Bde)	4	0	1
Methanocaldococcus sp. FS406-22 (Msp)	4	0	1
Thermococcus sibiricus MM 739 (Tsi)	4	0	1
Synechocystis species, strain PCC6803 (Ssp)	4	1	0
Thermococcus sp. AM4 (Tsp-AM4)	4	0	0
Cafeteria roenbergensis virus BV-PW1 (CroV)	4	0	0
Mycobacterium leprae, strain TN (Mle-TN)	4	0	0
Haloarcula marismortui ATCC 43049 (Hma)	4	0	0
Pseudomonas aeruginosa phage phiEL (PP-PhiEL)	4	0	0
Natrialba magadii ATCC 43099 (Nma)	3	0	1
Thermobifida fusca YX (Tfus)	3	0	1
Alkalilimnicola ehrlichei MLHE-1 (Aeh)	3	0	1
Nostoc punctiforme (Npu)	3	1	0
Nostoc species PCC7120 (Nsp)	3	1	0
Cyanothece sp. CCY0110 (Csp)	3	1	1
Thermococcus aggregans (Tag)	3	0	1
Nocardia farcinica IFM 10152 (Nfa)	3	0	0
Gloeobacter violaceus, PCC 7421 (Gvi)	3	0	1
Mycobacterium bovis subsp. bovis AF2122/97 (Mbo)	3	0	0
uncultured archaeon GZfos10C7	3	0	0
Mycobacterium gastri (Mga)	3	0	0
Salinibacter ruber DSM 13855 (Sru)	3	0	0

Supplementary Table 3 List of plasmids used for protein alternative splicing

No.	Plasmid	Short description of the insert	Promoter	Addgene ID	Ref.
1	pSKDuet16	H_6-GB1-*Npu*DnaE-GB1	T7	41684	1
2	pMHBAD14C	$Int^C_{NpuDnaE\_C35}$-GB1-H_6	BAD	42304	2,#
3	pMKBAD28	$Int^C_{NpuDnaE\_C14}$-GB1-H_6	BAD	45595	#
4	pSABAD219	$Int^C_{NpuDnaE\_C102}$-GB1-H_6	BAD	45596	#
5	pSKDuet26	H_6-GB1-*Tvo*VMA-GB1	T7	41689	1
6	pSABAD331	$Int^C_{TvoVMA\_C48}$-GB1-H_6	BAD	45609	#
7	pHYDuet183	H_6-GB1-*Pho*RadA-GB1	T7	41685	1
8	pSABAD332	$Int^C_{PhoRadA\_C46}$-GB1-H_6	BAD	45610	#
9	pHYDuet216A	H_6-GB1-*Δ*variant-GB1	T7	45611	#
10	pSABAD25-SK16	H_6-GB1-*Npu*DnaE-GB1	BAD	-	#
11	pSADuet194-14	$Int^C_{NpuDnaE\_C35}$-GB1-H_6	T7	-	#
12	pMMDuet19	H_6-GB1-*Npu*$DnaB_{mini}$-GB1	T7	45594	1
13	pSADuet502	H_6-GB1-*Npu*DnaB-GB1 (full-length *Npu*DnaB intein)	T7	46373	#
14	pSABAD250	$Int^C_{NpuDnaB\_C39}$-GB1-H_6	BAD	45612	#
15	pSKDuet20	H_6-GB1-*Mja*TFIIB-GB1	T7	41694	1
16	pSADuet760	H_6-GB1-*Mja*$TFIIB_{mini}$-GB1	T7	45613	#
17	pSABAD14-750	$Int^C_{MjaTFIIB\_C53}$-GB1-H_6	BAD	45616	#
18	pSEDuet40	H_6-MBP-*Tvo*VMA-GB1	T7	46378	1
19	pSABAD25-74	H_6-GB1-*Tvo*VMA-SH3	BAD	46375	#
20	pSABAD518	H_6-GB1-*Tvo*VMA_A-SH3 (N186A, T+1A) (inactive mutation of AA)	BAD	46374	#
21	pJORSF19	Full-length *Pho*RadA precursor	T7	45614	#
22	pALBRSF12	*Npu*DnaE intein for crystal structure	T7	46376	#
23	pHYRSF236	*Δ*variant for crystal structure	T7	46377	#
24	pJJBAD7	GFP (derived from pBAD/AC2)	BAD	-	3,4
25	pSABAD505	GFP^N-*Npu*DnaE-GFP^C (S30R, Y39N, F64L, S65C)	BAD	45615	#
26	pSARSF505	H_6-GFP^N-*Npu*DnaE-GFP^C (S30R, Y39N, F64L, S65C)	T7	46832	#
27	pSABAD544	GFP^N-*Npu*DnaE_AG-GFP^C (S30R, Y39N, F64L, S65C) inactive mutation of AG in *Npu*DnaE intein.	BAD	45617	#
28	pSARSF534-3	H_6-$Int^C_{NpuDnaE\_C35}$-GFP^C	T7	45618	#

described in this article.



Supplementary Figure 1: Mass spectrometry analysis. MALDI-TOF mass spectra of the iPAS product (H_6-GB1-GB1-H_6) produced between *cis*-splicing precursor (H_6-GB1-Intein-GB1) and a split intein precursor (Int_C-GB1-H_6) for (**a**) *Npu*DnaE intein, (**b**) *Pho*RadA intein, (**c**) *Tvo*VMA intein, (**d**) *Mja*TFIIB$_{mini}$ intein, and (**e**) *Npu*DnaB intein. Expected masses for iPAS products are indicated. Variations in the expected mass among the inteins are due to differences in the linker sequences between the intein and GB1s, which are shown at the top of each spectrum in italic. Met of the start codon is not completely removed for some cases which can be observed as small peaks with 131.2 dalton larger molecular weight.

a ***Npu*DnaE intein**

N35(C102)

/CLSYETEILT VEYGLLPIGK IVEKRIECTV YSVDNNGNIY TQPVAQWHDR GEQEVFEYCL

C35(*)

EDGSLIRATK DHKFMTVDGQ MLPIDEIFER ELDLMRVDNL PNIKIATRKY LGKQNVYDIG

C14

VERDHNFALK NGFIASN/C

b ***Tvo*VMA intein**

/CVSGETPVYL ADGKTIKIKD LYSSERKKED NIVEAGSGEE IIHLKDPIQI YSYVDGTIVR

SRSRLLYKGK SSYLVRIETI GGRSVSVTPV HKLFVLTEKG IEEVMASNLK VGDMIAAVAE

C48

SESEARDCGM SEECVMEAEV YTSLEATFDR VKSIAYEKGD FDVYDLSVPE YGRNFIGGEG

LLVLHN/T

c ***Pho*RadA intein**

/CFARDTEVYY ENDTVPHMES IEEMYSKYAS MNGELPFDNG YAVPLDNVFV YTLDIASGEI

KKTRASYIYR EKVEKLIEIK LSSGYSLKVT PSHPVLLFRD GLQWVPAAEV KPGDVVVGVR

C46

EEVLRRRIIS KGELEFHEVS SVRIIDYNNW VYDLVIPETH NFIAPNGLVL HN/T

d ***Npu*DnaB$_{mini}$ intein**

/CLAGDSLVTL VDSGLQVPIK ELVGKSGFAV WALNEATMQL EKAIVSNAFS TGIKPLFTLT

C39

TRLGRKIRAT GNHKFLTING WKRLDELTPK EHLALPRNLP SSGTLAKSDI YWDEIVSIEY

Δ

SGEEEVFDLT VPGLHNFVAN NIIVHN/S

e ***Mja*TFIIB$_{mini}$ intein**

/SVDYNEPIII KENGEIKVVK IGELIDKIIE NSENIRREGI LEIAKCKGIE VIAFNSNYKF

KFMPVSEVSR HPVSEMFEIV VEGNKKVRVT RSHSVFTIRD NEVVPIRVDE LKVGDILVLA

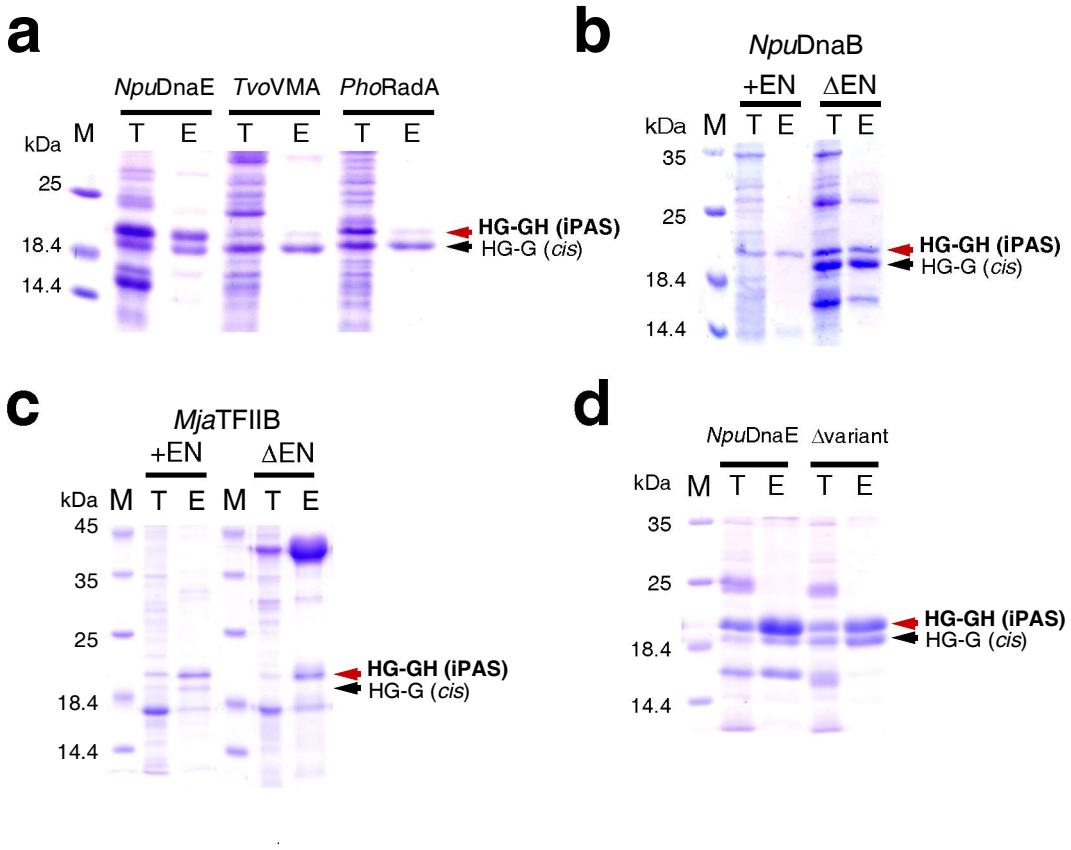
KNSDFIFLKI KEINKVEPTS GYAYDLTVPN AENFVAGFGG FVLHN/T

Δ

f ***Mja*TFIIB$_{C53}$ intein**

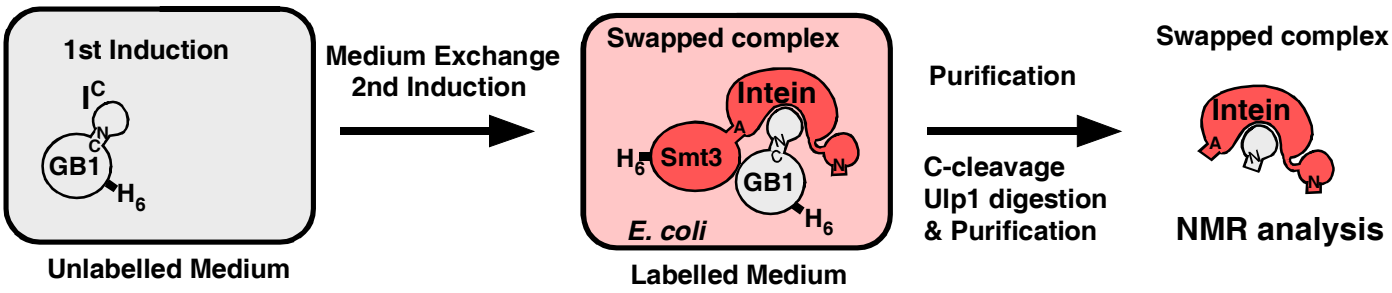
TNRKLEKLIN SDFIFLKIKE INKVEPTSGY AYDLTVPNAE NFVAGFGGFV LHN/T

Supplementary Figure 2: Protein sequences of inteins and split sites tested for iPAS. Split sites tested for iPAS in (**a**) *Npu*DnaE intein, (**b**) *Tvo*VMA intein, (**c**) *Pho*RadA intein, (**d**) *Npu*DnaB mini-intein, (**e**) *Mja*TFIIB mini-intein, and (**f**) C-terminal split intein (C53) of *Mja*TFIIB intein. The split sites are indicated by reverse triangles on the top with the abbreviated name. Δ indicates the insertion site of endonuclease domains in *Npu*DnaB and *Mja*TFIIB mini-inteins. * indicates the naturally split site. The boundaries with exteins are indicated by "/".



Supplementary Figure 3: SDS-PAGE analysis of iPAS with different inteins

(**a**) Comparison of iPAS between *Npu*DnaE, *Tvo*VMA, and *Pho*RadA inteins. (**b**) Comparison of iPAS of *Npu*DnaB intein with (+EN) and without (ΔEN) the endonuclease domain. (**c**) Comparison of iPAS of *Mja*TFIIB intein with (+EN) and without (ΔEN) the endonuclease domain. (**d**) Comparison of iPAS between *Npu*DnaE intein and the deletion variant (Δvariant). "T" indicates total cell lysate. "E" indicates elution from Ni-NTA spin columns. M indicates molecular weight marker. iPAS and *cis*-splicing products are indicated by red and black arrows, respectively.



Supplementary Figure 4: Schematic presentation of the procedure for the preparation of the fragmentally ^{13}C,^{15}N-labelled domain swapped complex.

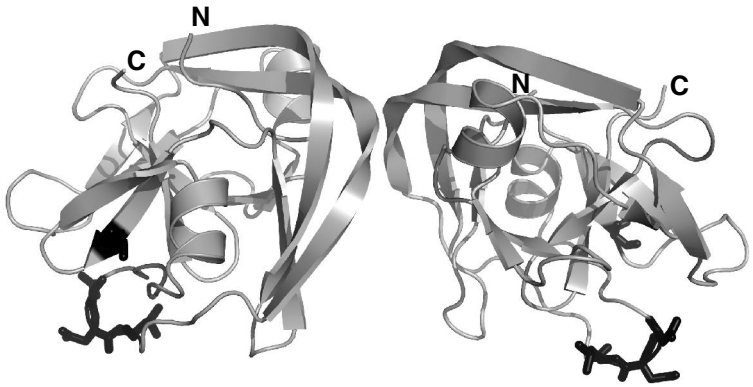
Nature Chemical Biology: doi:10.1038/nchembio.1320

a

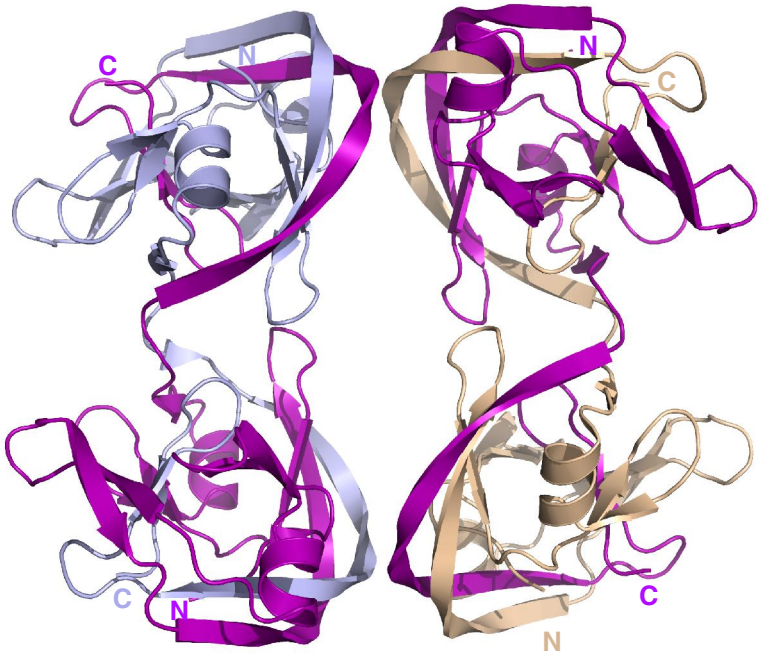
```
                N-extein  Intein
                        C
NpuDnaE     -3 SGGALSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFE  57
Δvariant    -3 SGGALSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFE  57

NpuDnaE     58 YCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPNIKIATRKYLGKQNVY 117
Δvariant    58 YCLEDGSLIRATKDHKFM-VDGQMLPIDEIFERELDLMRN---PGIKIATRKYLGKQNVY 113
                                          C-extein
                                        C
NpuDnaE    118 DIGVERDHNFALKNGFIASNADNG 141
Δvariant   114 DIGVERDHNFALKNGFIASNA--- 134
```

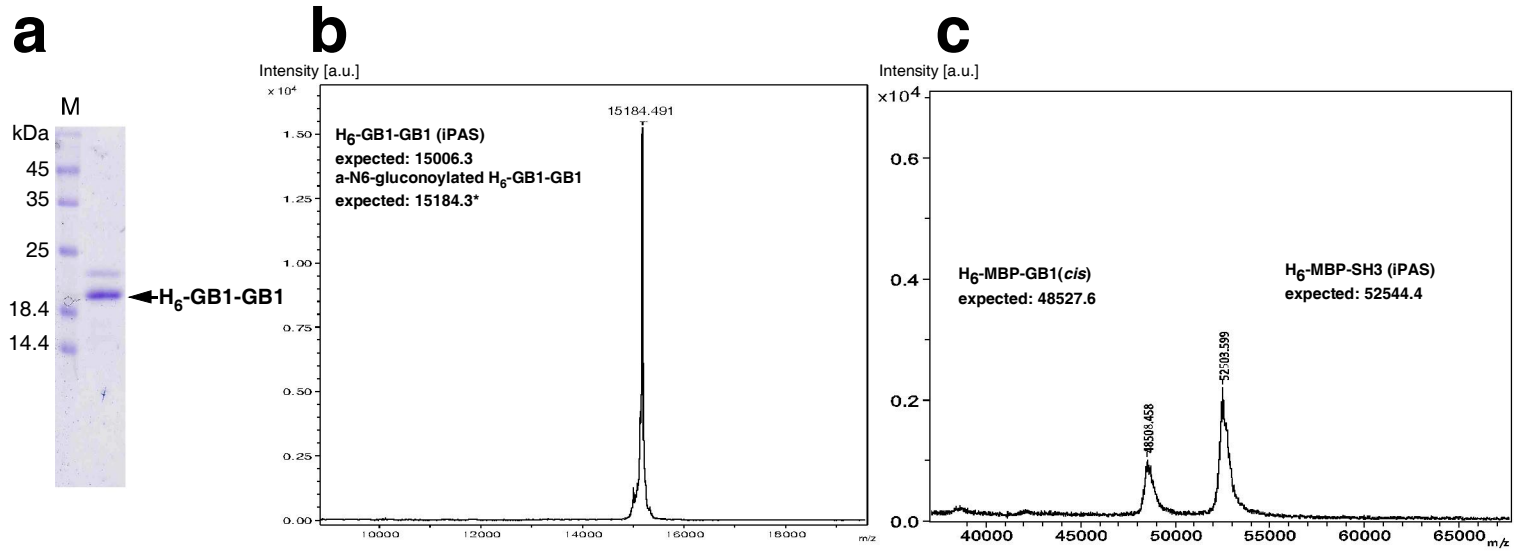
b



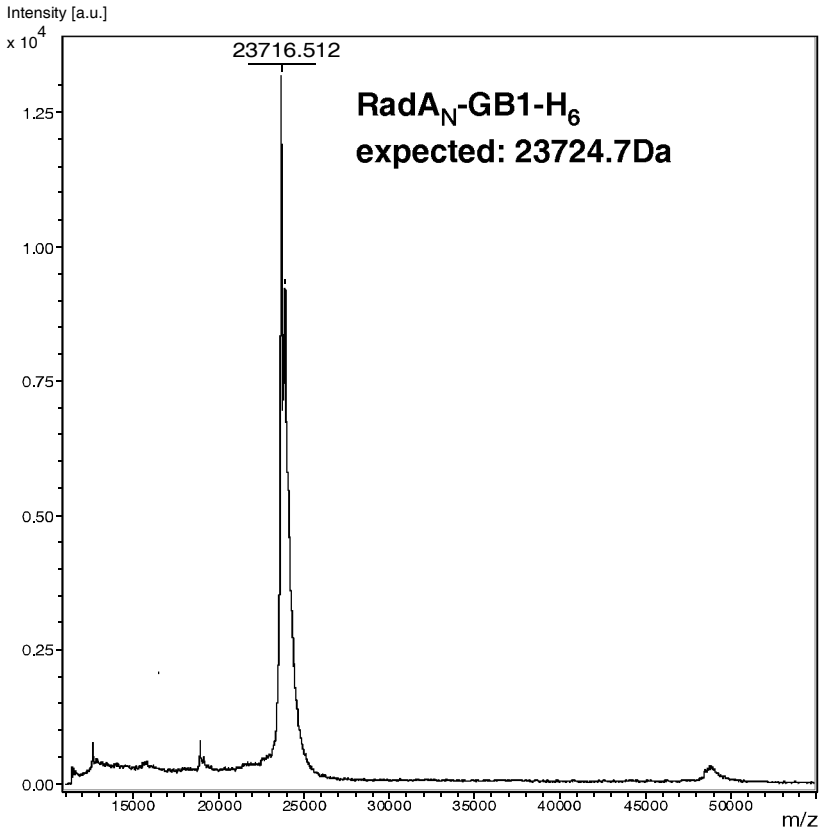
c



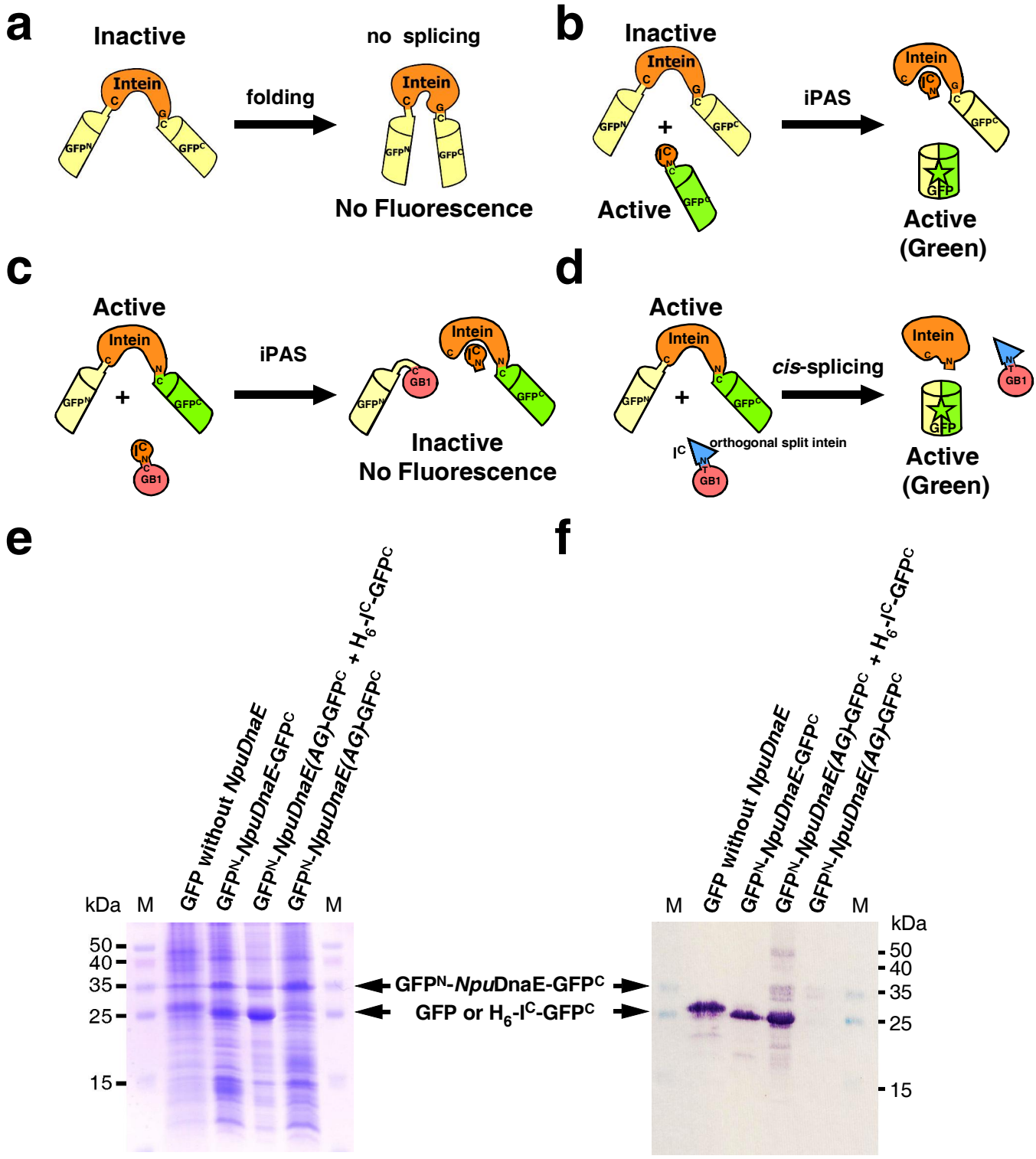
Supplementary Figure 5: X-ray structural analysis of *Npu*DnaE intein (**a**) Sequence alignment of *Npu*DnaE intein and the deletion variant (Δvariant) used for the crystal structures. Two mutated cysteine residues to inactivate the splicing are shown on the top. The deleted residues in the Δvariant (HYRSF236) are in bold and italic. (**b**) The crystal structure of monomeric *Npu*DnaE intein. Two molecules in the asymmetric unit are shown. Four residues deleted in Δvariant are shown in stick model and in dark grey. (**c**) The crystal structure of swapped dimers of Δvariant. Four molecules in the asymmetric unit are shown. One molecule in each domain-swapped dimer is highlighted in light yellow or light blue. N and C denote the N- and C-termini, respectively.



Supplementary Figure 6: MALDI-TOF mass spectra of the alternatively spliced products between the two *cis*-splicing precursors containing *Tvo*VMA intein. (**a**) Alternatively spliced product (H_6-GB1-GB1) purified by ion-exchange chromatography for mass-spectrometry analysis. (**b**) Identified H_6-GB1-GB1 in the mass spectrum. (**c**) Identified alternatively spliced product of H_6-MBP-SH3 in the mass spectrum. *the mass difference is due to alpha-N-6-phosphogluconoylation of H_6-tag.



Supplementary Figure 7: Mass spectrometry of iPAS product. MALDI-TOF mass spectrum of the alternatively ligated product (iPAS) produced between the wild-type full-length *Pho*RadA protein bearing *Pho*RadA intein (native exteins) and a split intein precursor with C-terminal 46 residues of *Pho*RadA intein with GB1-H_6 as C-extein.



Supplementary Figure 8: Western blotting analysis of iPAS with GFP (**a**) Schematic presentation of inactive GFP precursor with non-splicing intein. (**b**) Scheme for activation of splicing-incompetent GFP precursor by iPAS. (**c**) Scheme for suppressing green fluorescence by iPAS. (**d**) control of suppressing green fluorescence with orthogonal split intein. (**e**) SDS-PAGE analysis of expressions of GFP and GFP with the insertion of *cis*-splicing and non-splicing *Npu*DnaE inteins. (**f**) Western blotting with Anti-GFP antibody. GFP without *Npu*DnaE intein is 8 residues longer at the C-terminus than GFP with *Npu*DnaE intein (see Supplementary Note).

Fig. 1c

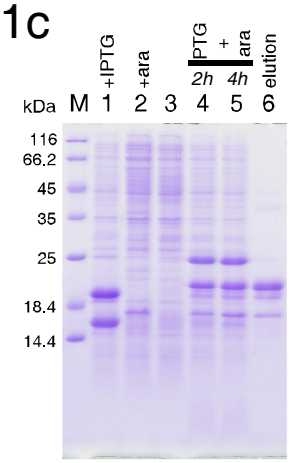


Fig. 1e

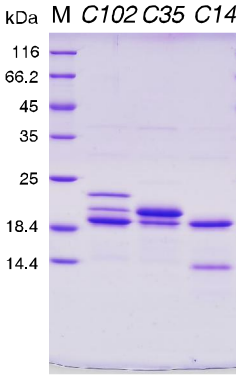


Fig. 3c,d

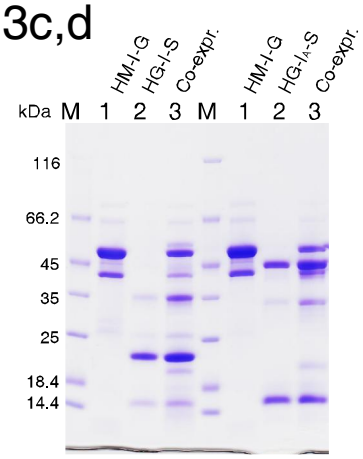


Fig. 3e

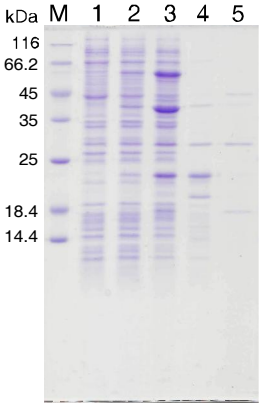


Fig. 4a

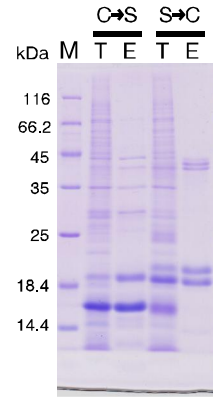
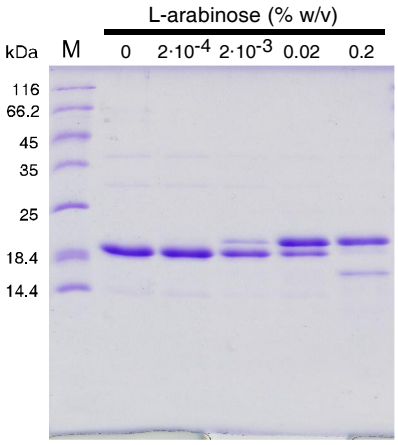


Fig. 4b



Supplementary Figure 9: Entire images of the gels in Figs. 1, 3, and 4.

Nature Chemical Biology: doi:10.1038/nchembio.1320

Supplementary Note

(Descriptions of plasmids used for protein alternative splicing)

Cis-*splicing H_6-GB1-*Npu*DnaE-GB1 (HG-I-G)*

Cis-splicing precursor bearing *Npu*DnaE intein with two GB1s (B1 domain of IgG binding protein G) as exteins was described previously (pSKDuet16)[1].

Split C-precursor $Int^C_{Npu\text{DnaE\_C35}}$-GB1-$H_6$ (I^C-GH) (C35)

The gene of C-intein of *Npu*DnaE intein with GB1 was amplified from pSKBAD2[5]with two oligonucleotides SK094: 5'-TAA CAT ATG ATC AAA ATA GCC ACA CG and HK158: 5'-AGA ATT CCG TTA CGG TGT AGG TTT TG. The amplified gene was cloned between *Nde*I and *EcoR*I sites of pSABAD92A[2], resulting in plasmid pMHBAD14C encoding $Int^C_{Npu\text{DnaE\_C35}}$-GB1-$H_6$.

Split C-precursor $Int^C_{Npu\text{DnaE\_C14}}$-GB1-$H_6$ (C14)

A split *Npu*DnaE intein precursor containing the C-terminal 14 residues of the intein was constructed as follows. This split site was previously called C15 by counting the start codon of Met[6]. The gene of the C-terminal 14 residues of *Npu*DnaE intein was amplified together with GB1-H_6 from pMHBAD14C by PCR with two oligonucleotides, HK146: 5'-TAC ATA TGG ACC ATA ATT TTG CAC TC and HK122: 5'-CTA AAG CTT AAT GAT GAT GAT GAT GAT G. The PCR product was cloned into pSKBAD2 between *Nde*I and *Hind*III sites, resulting in pMKBAD28.

Split C-precursor $Int^C_{Npu\text{DnaE\_C102}}$-GB1-$H_6$ (C102)

The C-terminal 102 residues of the full-length *Npu*DnaE intein was amplified by PCR with two oligonucleotides, SK095: 5'-TAG GTA CCA TTG AAA CAA TTA GAA GCT ATG and HK477: 5'-CCA TAT GAA TGG AAA TAT TTA TAC AC from pSKDuet16 as a template. The PCR product was digested and cloned into pMHBAD14C using *Nde*I and *Kpn*I sites, resulting in pSABAD219 encoding $Int^C_{Npu\text{DnaE\_C102}}$-GB1-$H_6$.

Nature Chemical Biology: doi:10.1038/nchembio.1320

Cis-*splicing* Tvo*VMA intein precursor (HG-I(TvoVMA)-G)*

Cis-splicing precursor containing *Tvo*VMA intein and two GB1s as exteins was described previously (pSKDuet26)[1].

Split C-precursor split with* Tvo*VMA intein (Int$^{C}_{Tvo}$VMA_C48-GB1-H$_6$)

The gene for the split intein precursor containing 48-residue C-intein of *Tvo*VMA intein was amplified from pSKDuet26 as a template by PCR with the two oligonucleotides, HK585: 5'-ACA TAT GGA AGT TTA TAC GTC ACT TGA AG and SK182: 5'-ATG GTA CCT ATA ACC GTG TTG TGA AGT ACG AG. The PCR product was cloned into pMHBAD14C using *Nde*I and *Kpn*I sites, resulting in plasmid pSABAD331.

Cis-*splicing and split* Pho*RadA intein precursor (HG-I(*Pho*RadA)-G)*

Cis-splicing precursors containing *Pho*RadA intein was previously constructed and described as pHYDuet183[1].

Split C-precursor from* Pho*RadA intein (Int$^{C}_{Pho}$RadA_C46 -GB1-H$_6$)

Split precursor containing the C-terminal 46 residues of *Pho*RadA intein was derived from pHYDuet183 by PCR with the two oligonucleotides, HK586: 5'-ACA TAT GAG AAT AAT ATC CAA AGG AG and HK376: 5'-TTG GTA CCT AGC TGA GTA TTA TGG AGA ACA AGT C. The PCR product was cloned into pMHBAD14C using *Nde*I and *Kpn*I sites, resulting in plasmid pSABAD332 encoding the split precursor of Int$^{C}_{Pho}$RadA_C46-GB1-H$_6$.

Cis-*splicing precursor containing a deletion variant of* Npu*DnaE intein*

The deletion of three residues and a mutation in *cis*-splicing *Npu*DnaE intein was introduced in pSKDuet16, using two oligonucleotides, HK438: 5'-GCG GAA CCC CGG GAT CAA AAT AGC CAC ACG and HK439: 5'-CCC GGG GTT CCG CAT CAA ATC CAA TTC, resulting in pSADuet205. An additional deletion of T76 in *Npu*DnaE intein was introduced in pSADuet205 with the two oligonucleotides, HK485: 5'-GAC CAT AAG TTT ATG GTC GAC GGT CAA ATG TTG CC and HK486: 5'- GGC AAC ATT TGA CCG TCG ACC ATA AAC TTA TGG TC, resulting in pHYDuet216A.

Cis-*splicing precursor and split C-intein precursor with* Npu*DnaE intein for swapping the order of expression*

To test the effect of the order of protein expressions, *cis*-splicing precursor was swapped to pBAD vector with arabinose promoter. The coding gene of H_6-GB1-*Npu*DnaE-GB1 was transferred into pHYBAD25 using *Nco*I and *Hind*III restriction enzymes, resulting in pSABAD25-SK16. pHYBAD25 was constructed by mutating *Nde*I site of pSFBAD19[2] to *Nco*I site with two oligonucleotides, HK48: 5'-GAA GGA GAT ATA CCC ATG GTT AAA GTT ATC G and HK49: 5'-CGA TAA CTT TAA CCA TGG GTA TAT CTC CTT C.

The split C-intein precursor was constructed in pRSF vector (Novagen) by transferring the gene of $Int^{C}_{Npu\text{DnaE\_C35}}$-GB1-$H_6$ from pMHBAD14C into pHYDuet194 using *Nde*I and *Hind*III sites, resulting in pSADuet194-14. pHYDuet194 was constructed by mutating *Nco*I to *Nde*I site in pHYRSF-1 using two oligonucleotides HK446: 5'-CTT TAA TAA GGA GAT ATC ATA TGG GCA GCA GCC and HK447: 5'-GGC TGC TGC CCA TAT GAT ATC TCC TTA TTA AAG.

Cis-*splicing precursor* Npu*DnaB intein with and without endonuclease domain (+EN and ΔEN)*

Cis-splicing precursor bearing *Npu*DnaB mini-intein lacking endonuclease domain with two GB1s as exteins was constructed previously (pMMDuet19)[1]. *Cis*-splicing precursor containing the full-length *Npu*DnaB intein including endonuclease domain was constructed by amplifying the gene of *Npu*DnaB from the genomic DNA of *Nostoc punctiforme* (ATCC 29133) by PCR with the two oligonucleotides, HK151: 5'- TAG GAT CCG GTT GTT TAG CAG GCG ATA GTC and HK212: 5'-GAG GTA CCA ATG GAA TTG TGA ACA AT. The PCR product was cloned between *Bam*HI and *Kpn*I sites of pSKDuet27[1], resulting in pSADuet502 encoding the precursor of H_6-GB1-*Npu*DnaB-GB1.

Split C-intein precursor with NpuDnaB intein ($Int^{C}_{Npu\text{DnaB\_C39}}$-GH)

The C-terminal split precursor derived from *Npu*DnaB intein was constructed by amplifying the C-terminal 39 residues of *Npu*DnaB intein from pMMDuet19 by the two oligonucleotides, HK514: 5'-GAC ATA TGA GTG ATA TTT ATT GGG and HK212 and by cloning it between *Nde*I and *Kpn*I sites in pMHBAD14C, resulting in pSABAD250. This plasmid encodes the split C-terminal precursor of $Int^{C}_{Npu\text{DnaB\_C39}}$-GB1-$H_6$.

Cis-*splicing precursor* Mja*TFIIB intein with and without endonuclease domain (+EN and ΔEN)*

Cis-splicing precursor with full-length *Mja*TFIIB (pSKDuet20) was previously described[1]. *Mja*TFIIB mini-intein without the endonuclease domain was constructed from pSKDuet20 by assembly PCR. The N-terminal half was amplified with the two oligonucleotides SK169: 5'-TTG GAT CCT ATA GTG TTG ATT ACA ACG AAC C and I292: 5'-TTT AAG AAT ATG AAA TCA GAA TTC TTT GCT AAA AC. The C-terminal part was amplified with two oligonucleotides, I291: 5'-GAT ATA TTA GTT TTA GCA AAG AAT TCT GAT TTC AT and SK170: 5'-CTG GTA CCG TGG ATG GTG TTG TGT AAT ACA AAT C. The two PCR products were assembled and re-amplified with SK169 and SK170 before cloned into pSKDuet16 between *Bam*HI and *Kpn*I, resulting in pSADuet760 that encodes H_6-GB1-*Mja*TFIIB$_{mini}$-GB1.

Split C-intein precursor with* Mja*TFIIB intein (Int$^{C}_{Mja\text{TFIIB_C53}}$-GH)

The C-terminal split intein part was amplified from pSKDuet20 with two oligonucleotides, SK170: 5'-CTG GTA CCG TGG ATG GTG TTG TGT AAT ACA AAT C and I287: 5'-AAC ATA TGA CCA ATA GAA AAC TCG A. The PCR product was cloned between *Nde*I and *Kpn*I sites of pMHBAD14C, resulting in pSABAD14-750.

Cis-*splicing precursor with* Tvo*VMA intein with MBP and GB1 as N- and C-exteins (HM-I-G)*

Cis-splicing precursor of H_6-MBP-Intein(*Tvo*VMA)-GB1 was constructed in RSF vector as previously described (pSEDuet40) [1].

Cis-*splicing precursor with* Tvo*VMA intein with GB1 and SH3 as exteins (HG-I-S)*

Cis-splicing precursor bearing H_6-GB1-Intein(*Tvo*VMA)-SH3 was constructed in pBAD vector from pSEDuet74[1] by digesting the plasmid with *Nco*I and *Hind*III and by transferring the insert into pHYBAD25, resulting in pSABAD25-74.

Inactivated* cis-*splicing precursor with* Tvo*VMA intein with GB1 and SH3 as exteins (HG-I$_A$-S)

Inactive *cis*-splicing precursor (HG-I$_A$-S) was constructed in pSABAD25-74 by introducing two mutations of N186A in *Tvo*VMA intein and T+1A in C-extein, using two oligonucleotides, HK821: 5'-GGA CTT CTC GTA CTT CAC GCT GCA GTT ATA GGT ACC C and HK822: 5'-GGG TAC CTA TAA CTG CAG CGT GAA GTA CGA GAA GTC C, which resulted in pSABAD518.

The wild-type full-length* Pho*RadA precursor bearing the* Pho*RadA intein

The entire wild-type *Pho*RadA precursor including both native intein and exteins was amplified from genomic DNA genomic DNA from *Pyrococcus horikoshii* (ATCC 700860). To delete an internal *Hind*III site, the N-terminal half was amplified with oligonucleotides HK575: 5'-A TCC ATG GCA ATG ATC ATG GTT AAG AAA GGG AGC and HK810: 5'-CCC GCC TCC CTA AGT TTT TCA GCA GTA GC. The other C-terminal half was amplified with HK809: 5'-GCT ACT GCT GAA AAA CTT AGG GAG GCG GG and HK434: 5'-ACT TAA GCT TAA TCT TCA ATT CCC TTC TCC G. These two PCR products were assembled and re-amplified with HK575 and HK434. The amplified PCR produce was digested with *Nco*I and *Hind*III and ligated into pRSF1-b (Novagen), resulting in pJORSF19.

The single-chain variant and Δvariant of* Npu*DnaE intein for structure determination

A plasmid for the single-chain variant of *Npu*DnaE intein with inactive mutations (C1A and C+1A) was constructed in pHYRSF53[6]. The gene was first amplified by PCR from pSKDuet16 as a template using two oligonucleotides HK203: 5'- GTG GAT CCG GAG GAG CTC TAA GCT ATG AAA CG and HK433: 5'-TTG AAG CTT ATC CAT TGT CGG CAT TAG AAG CTA TG. The PCR product was cloned in pHYRSF53 using *Bam*HI and *Hind*III, resulting in pALBRSF12. The gene of Δvariant of *Npu*DnaE intein was amplified from pHYDuet216A by PCR using two oligonucleotides HK203 and SK186: 5'-ATC AAG CTT AGG CAT TAG AAG CTA TGA AGC C and cloned between *Bam*HI and *Hind*III of pHYRSF53, resulting in pHYRSF236.

Green Fluorescent Proteins with *Npu*DnaE insertion

Cis-*splicing precursor GFP^N*-Npu*DnaE-GFP^C*

Cis-splicing precursor encoding GFP^N-*Npu*DnaE-GFP^C was originally derived from the cycle3 mutant of GFP (Stemmer GFP) (pJJBAD7 derived from deleting a *Nde*I site in pBAD/AC2)[3,4]. We modified the gene with site-directed mutagenesis to create and delete some restriction sites for inserting a synthetic *Npu*DnaE intein gene. We also introduced superfolder GFP mutations (S30R and Y39N) in the N-terminal half of GFP (GFP^N)[7]. The final construct of pSABAD505 has the following protein sequence under arabinose promoter in pBAD vector:

MASKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTL
VTTLCLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFE
YCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPNIKIATRKYLGKQNVYD
IGVERDHNFALKNGFIASNCYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGNY
KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIKANFKI
RHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAG
LT.

The sequence for *Npu*DnaE intein is highlighted in underline. This plasmid also contains "CYG" as the chromophore sequence instead of "SYG" of the wild-type GFP. In addition, it is missing the C-terminal sequence of HGMDELYN. The promoter of this vector was switched to T7 promoter by digesting the gene with *Nhe*I and *Hind*III and transferring into pHYRSF53 digested with *Spe*I and *Hind*III. The resulted plasmid is pSARSF505.

Inactivated cis-*splicing precursor, GFP^N*-Npu*DnaE(AG)-GFP^C*

Cis-splicing of the precursor of GFP^N-*Npu*DnaE-GFP^C (pSABAD505) was inactivated by replacing the last two residues of the intein by "AG" using the two oligonucleotides, HK883: 5'-AAA AAT GGC TTC ATA GCC GCG GGT TGT TAT GGT GTT CAA and HK884: 5'-TTG AAC ACC ATA ACA ACC CGC GGC TAT GAA GCC ATT TTT, resulting in pSABAD544.

Split C-precursor H$_6$-Int^C$_{Npu DnaE\_C35}$*-GFP^C*

A split C-intein precursor bearing the 35-residue C-intein of *Npu*DnaE intein was constructed by amplifying the gene including C-terminal split GFP from pSABAD505 using two oligonucleotides, SK094: 5'-TAA CAT ATG ATC AAA ATA GCC ACA CG and #153: 5'-CGC AAG CTT AAG

TCA ACC CAG CAG CAG and inserting the amplified gene between *Nde*I and *Hind*III sites of pHYRSF-1 (Addgene, 34549), resulting in pSARSF534-3.

References

1. Ellilä, S., Jurvansuu, J. M. & Iwaï, H. Evaluation and comparison of protein splicing by exogenous inteins with foreign exteins in Escherichia coli. *FEBS lett*. **585**, 3471–3477 (2011).
2. Buchinger, E., *et al*. Use of Protein trans-splicing to produce active and segmentally ^{2}H, ^{15}N labelled mannuronan C5-epimerase AlgE4. *Prot. Sci.* **19**, 1534–1543 (2010).
3. Crameri, A., Whitehorn, E. A., Tate, E. & Stemmer, W. P. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* **14,** 315–9 (1996).
4. Iwai, H., Lingel, A. & Plückthun A. Cyclic green fluorescent protein produced in vivo using an artificially split PI-*Pfu*I intein from *Pyrococcus furiosus*. *J. Biol. Chem*. **276**, 16548–54 (2001).
5. Iwai, H., Züger, S., Jin, J. & Tam, P. H. Highly efficient protein trans-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. *FEBS Lett*. **580,** 1853–8 (2006).
6. Muona, M., Aranko, A. S. & Iwai, H. Segmental isotopic labelling of a multi-domain protein by protein ligation using protein *trans*-splicing. *ChemBioChem.* **9**, 2958−2961 (2008).
7. Pédelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotech.* **24**, 79–88 (2005).

Nature Chemical Biology: doi:10.1038/nchembio.1320