Intermolecular domain swapping induces intein-mediated protein alternative splicing

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

Protein sequences are diversified on the DNA level by recombination and mutation and can be further increased on the RNA level by alternative RNA splicing, involving introns that have important roles in many biological processes. The protein version of introns (inteins), which catalyze protein splicing, were first reported in the 1990s. The biological roles of protein splicing still remain elusive because inteins neither provide any clear benefits nor have an essential role in their host organisms. We now report protein alternative splicing, in which new protein sequences can be produced by protein recombination by intermolecular domain swapping of inteins, as elucidated by NMR spectroscopy and crystal structures. We demonstrate that intein-mediated protein alternative splicing could be a new strategy to increase protein diversity (that is, functions) without any modification in genetic backgrounds. We also exploited it as a post-translational protein conformation–driven switch of protein functions (for example, as highly specific protein interference).

Protein splicing is a remarkable post-translational modification in which an intervening sequence (intein) excises itself from the host precursor protein, concomitantly ligating the two flanking sequences (exteins) with a peptide bond (**Fig. 1a**)[1–3]. Currently, more than 500 inteins have been identified in all three kingdoms of life (eukaryotes, eubacteria and archaea) on the basis of conserved characteristic sequence[4]. Although some regulatory functions are suggested for specific inteins[5,6], inteins are considered to be functionless selfish genetic elements with ancient origins because they are often inserted in distinct conserved sites of various essential proteins and can be deleted without notable effects to the host organisms[7]. Protein splicing usually occurs in *cis* as an intramolecular reaction within a single precursor molecule[8]. Protein splicing can also take place in *trans* as a bimolecular reaction when a host precursor is artificially or naturally split into two fragments within the intein sequence (**Fig. 1b**)[6,9,10]. When two split fragments (I^N and I^C) derived from one intein associate together to form an active intein, two extein sequences (E^N and E^C) fused with the split intein (I^N and I^C) on two different chains will be ligated into one polypeptide chain with a covalent peptide bond by protein *trans*-splicing (PTS) upon excision of the split intein fragments (**Fig. 1b**). This PTS has been used for various biotechnological applications because protein splicing is not limited to native exteins, taking place even with foreign exteins[1,10]. As naturally occurring split inteins are discovered among cyanobacteria[6,11,12], it has been speculated that split inteins are remnants of inteins and had a primeval role in evolution in the combinatorial recombination of protein sequences for domain shuffling on the protein level[11,13]. Split inteins might be more common than presumed because split inteins have been identified in metagenomic sequences[14].

Here we demonstrate that protein splicing could also occur as a bimolecular reaction involving *cis*-splicing precursors, thereby resulting in up to four differently ligated molecules. The newly discovered intein-mediated protein alternative splicing (iPAS) could thus diversify protein repertoires after protein translation. We also exploited iPAS for modulating protein functions in a post-translational manner.

RESULTS
iPAS by split inteins
We serendipitously noticed that expression of a *cis*-splicing precursor protein together with a split intein precursor in *Escherichia coli* produced unexpected bands larger than the protein expected from the *cis*-spliced product in SDS-PAGE (**Fig. 1c**). We derived the *cis*-splicing precursor protein (HG-I-G) from naturally split DnaE intein from *Nostoc punctiforme* (*Npu*) as a *cis*-splicing intein (I) flanked with HG, the N-terminally His6-tagged B1 domains of IgG binding protein G (GB1), as the N-extein and GB1 (G) as the C-extein[15,16]. *cis*-Splicing of this precursor protein produces the spliced product, HG-G[16]. The split intein precursor (I^C-GH) contained the natural C-terminal split *Npu*DnaE intein (I^C) and the C-terminally His6-tagged GB1 (GH) as the C-extein. We interpreted these unexpected bands as HG-GH and I-G, produced by an intermolecular reaction between the two precursor proteins (HG-I-G and I^C-GH) (**Fig. 1c,d**). To prove this, we purified the spliced product using N- and C-terminal His6-tags and indeed confirmed it as the ligated product from the bimolecular reaction between the two precursor proteins by MALDI-TOF MS (**Supplementary Results, Supplementary Fig. 1a**). We termed this intermolecular reaction iPAS. We were intrigued by this observation and tested various split intein fragments of *Npu*DnaE intein, including both N- and C-terminal intein fragments for iPAS, using the identical model system of GB1 as the extein and nonsplit *Npu*DnaE as the *cis*-splicing intein[15]. In addition to naturally split I^C (C35), we also tested other I^C fragments, that is, the C-terminal 14 residues (C14) and 102 residues (C102) of *Npu*DnaE intein (**Fig. 1e** and **Supplementary Fig. 2**). These fragments are capable of *trans*-splicing together with the remaining N-terminal intein fragments[17]. Whereas C102 did produce a minute amount of the alternatively ligated product, C14 did not induce alternative splicing at a detectable level under the applied conditions (**Fig. 1e**). These data suggest that the naturally split C-terminal 35 residues of the intein (C35) induce iPAS most efficiently, which is different from *trans*-splicing[17,18]. Next we asked whether iPAS could occur with other naturally occurring *cis*-splicing inteins. We tested iPAS of highly efficient *cis*-splicing inteins, that is, RadA intein from *Pyrococcus horikoshii* (*Pho*RadA), VMA intein from *Thermoplasma volcanium* (*Tvo*VMA), TFIIB intein from *Methanocaldococcus jannaschii* (*Mja*TFIIB), *Npu*DnaB intein and engineered *Mja*TFIIB and *Npu*DnaB mini-inteins (**Supplementary Fig. 2**)[16]. We artificially created C-terminal split intein fragments from the individual *cis*-splicing inteins by splitting at their corresponding sites to the naturally split site of *Npu*DnaE intein

Research Program in Structural Biology and Biophysics, Institute of Biotechnology, University of Helsinki, Helsinki, Finland. *e-mail: hideo.iwai@helsinki.fi

Figure 1 | Protein splicing in *cis* and *trans* and iPAS. (**a**) Protein splicing in *cis*. E^N, N-extein; E^C, C-extein. (**b**) PTS. I^N and I^C represent N- and C-terminal split intein fragments, respectively. (**c**) SDS-PAGE analysis of iPAS. Lane 1, induction of *cis*-splicing precursor (HG-I-G) with IPTG; lane 2, induction of C-terminal split precursor (I^C-GH) with L-arabinose (L-Ara); lane 3, before induction; lanes 4 and 5, 2 h and 4 h after the dual induction of *cis*-splicing precursor (HG-I-G) and C-terminal split precursor (I^C-GH) with both IPTG and L-arabinose, respectively; lane 6, the elution from immobilized metal affinity chromatography (IMAC). HG-G is the *cis*-spliced product from *cis*-splicing precursor HG-I-G bearing His6-tagged GB1 (HG) and GB1 (G) as N- and C-exteins, respectively. HG-GH is the alternatively ligated product produced by iPAS. Black and red arrows indicate the bands corresponding to *cis*- and alternative splicing products, respectively. Gray arrows indicate cleaved precursor (clv) and excised intein (I). Open arrows indicate unprocessed C-terminal split precursor (I^C-GH) and the *cis*-splicing precursor (HG-I-G). MW, molecular weight. (**d**) Two pathways of protein alternative splicing (red arrow) and *cis*-splicing (black arrow). (**e**) iPAS of different split intein fragments. C102, C35 and C14 are the different C-terminal split intein fragments tested. Elution fractions from IMAC were

analyzed. Red and black arrows indicate alternatively ligated (iPAS) and *cis*-splicing products, respectively. Open arrows indicate split precursors (I^C_C102-GH and I^C_C35-GH). (**f**) Comparison of iPAS between different inteins under identical conditions. Data represent mean ± s.d., typically from three or more independent experiments. Subscript 'mini' represents engineered mini-inteins. Full gels for **c** and **e** are available in **Supplementary Figure 9**.
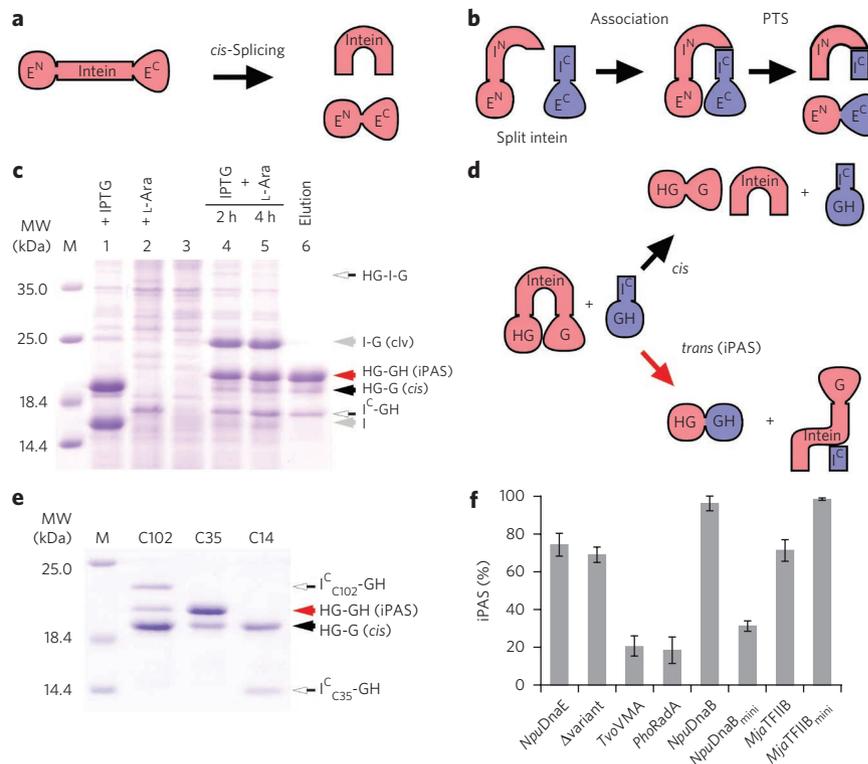
(**Supplementary Fig. 2**). The C-terminal split intein fragment fused to the C-terminally His6-tagged GB1 (I^C-GH) (C48 of *Tvo*VMA intein, C46 of *Pho*RadA intein, C53 of *Mja*TFIIB intein or C39 of *Npu*DnaB intein) was expressed together with a *cis*-splicing precursor bearing the parental nonsplit *cis*-splicing intein. We detected the alternatively ligated products (HG-GH) for all of the inteins tested, although the ratios of alternative splicing and *cis* splicing considerably varied among the inteins under the same condition (20–98%) (**Fig. 1f** and **Supplementary Figs. 1** and **3**). These data suggest that iPAS might take place with other inteins under certain conditions.

Domain swapping *in vivo* as the mechanism of iPAS
We were interested in the molecular mechanism underlying iPAS. Inducing protein splicing between two separate polypeptide chains requires a stable exchange of an identical part of the intein including a splicing junction between *cis*-splicing and split intein precursors (**Fig. 1d**), that is, three-dimensional (3D)-domain swapping. 3D-domain swapping of different proteins is confirmed in several crystal structures of other proteins exchanging a part of the protein between two molecules[19]. To elucidate the structural basis of iPAS, we developed a strategy to isolate the 3D domain–swapped complex formed *in vivo* for analysis by NMR spectroscopy, with which we differentially labeled only one of the two swapping polypeptides with stable isotopes. We analyzed the domain swapping between *cis*-splicing *Npu*DnaE intein and naturally split I^C of *Npu*DnaE intein (C35). We achieved the differential labeling by a time-delayed dual-expression system that was previously developed[20]. First, we expressed the split intein precursor (I^C-GH) in unlabeled medium, which was subsequently replaced by 15N-labeled medium, followed by overexpression of a *cis*-splicing *Npu*DnaE intein carrying the C1A mutation to prevent splicing in the labeled culture[21]. We then purified the complex with an N-terminally His6-tagged SUMO domain fused to the *Npu*DnaE intein, which was subsequently removed by ubiquitin-like specific protease 1 for NMR analysis (**Supplementary Fig. 4**)[22].

If no domain-swapped complex formed *in vivo*, the isolated molecule would be the *Npu*DnaE intein itself, thereby giving rise to an [1H,15N]-HSQC spectrum identical to that of *Npu*DnaE intein[21]. In the domain-swapped complex, the unlabeled polypeptide of I^C is expected to replace the 15N-labeled *Npu*DnaE intein, thereby causing the reduction of NMR signal intensities of the domain-swapped region, the appearance of new less-dispersed peaks in the HSQC spectrum and possible peak shifts near the swapping loop. We observed considerable changes in the HSQC spectrum of the presumable domain-swapped complex preparation compared with that of the *Npu*DnaE intein (**Fig. 2a,b**). We mapped the residues whose signal was reduced or moved on the structure of the *Npu*DnaE intein (**Fig. 2c**). These residues were located precisely in the C-terminal 35 residues and in the vicinity of the loop where an endonuclease domain is typically inserted in many inteins (**Fig. 2c**). Strong signals appeared in a narrow dispersion range of 1H chemical shift between 7.5 p.p.m. and 8.5 p.p.m., which is typical for residues in a random-coil conformation with high flexibility (**Fig. 2b**). These signals were absent in the spectrum of the *Npu*DnaE intein[21], suggesting that the region corresponding to the I^C region of the 15N-labeled *cis*-splicing *Npu*DnaE intein became unstructured and flexible (**Fig. 2d**).

The intertwined domain-swapped dimeric structure
The previously determined NMR structure of *Npu*DnaE intein is clearly monomeric in solution, as supported by 15N relaxation analysis[18]. To confirm this, we also determined the crystal structure of *Npu*DnaE intein. The crystal structure showed no sign of intertwined dimers, similarly to other reported intein structures (**Supplementary Fig. 5** and **Supplementary Table 1**)[5,23–26]. Loop deletion is a commonly used mechanism for forming 3D domain–swapped conformations, as observed in several crystal structures[19,27]. We therefore introduced a few deletions in the loops of *Npu*DnaE intein and found that a variant (Δvariant) with a three-residue deletion in the loop of the natural split site and one residue deletion in
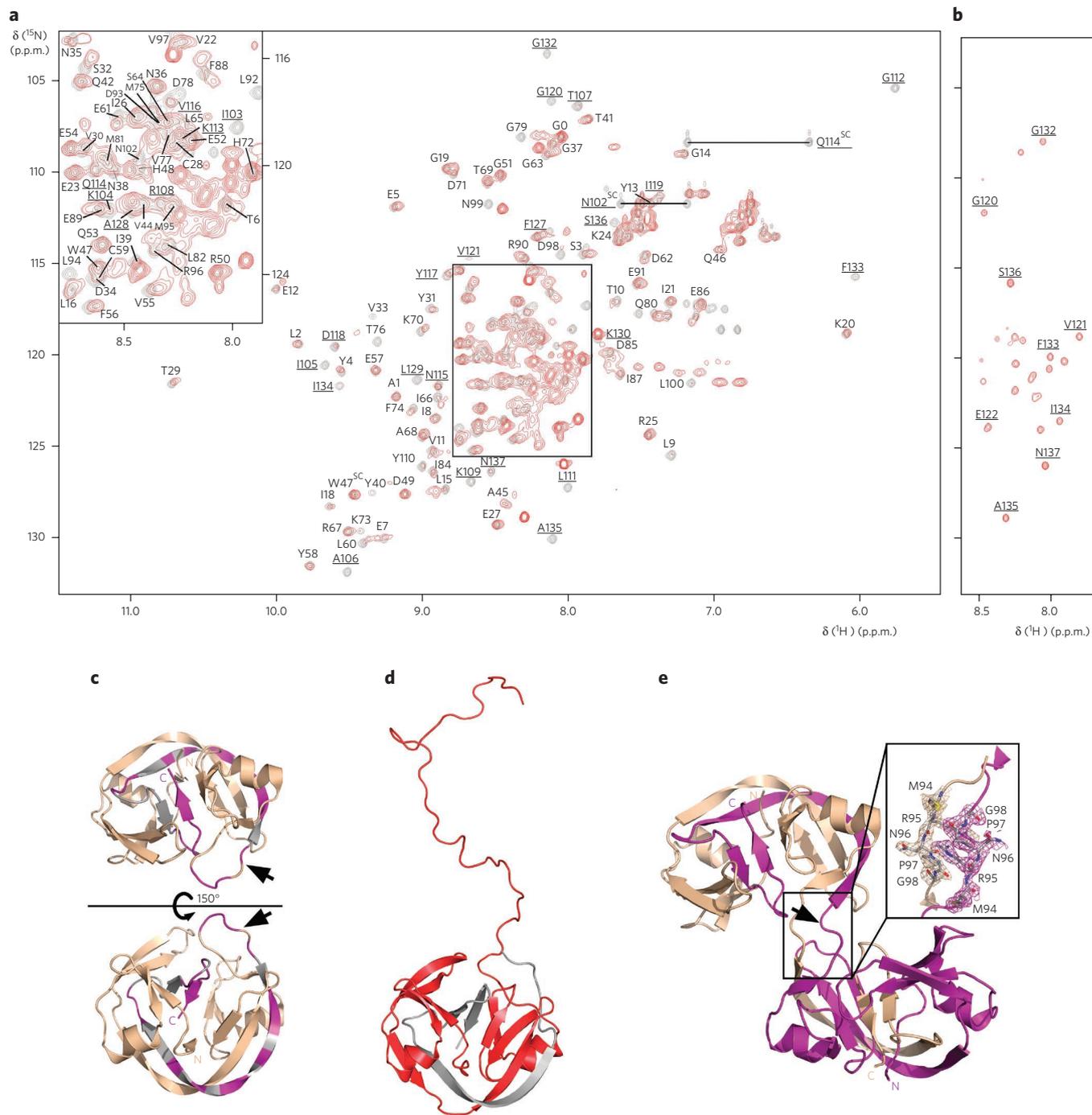
Figure 2 | Structural basis of iPAS by NMR spectroscopy and X-ray crystallography. (a) An overlay of [¹H,¹⁵N]-HSQC spectra from the single-chain full-length *Npu*DnaE intein with the C1A mutation (gray) and from the fragmentally ¹⁵N-labeled domain-swapped complex (red). The last 36 residues of the full-length *Npu*DnaE intein are underlined. **(b)** The same [¹H,¹⁵N]-HSQC spectrum of the fragmentally labeled domain-swapped complex plotted at higher counter levels, showing only strong peaks. **(c)** Mapping of *in vivo* 3D-swapped region by NMR analysis. The peaks with clearly moved (either >0.2 p.p.m. in ¹H chemical shifts or >0.3 p.p.m. in ¹⁵N chemical shifts) or reduced intensities in the [¹H,¹⁵N]-HSQC spectrum of the time-delayed preparation of *Npu*DnaE intein were mapped in the crystal structure of monomeric *Npu*DnaE intein (magenta). The residues that could not be identified unambiguously owing to lack of assignments are in gray. **(d)** A three-dimensional model of the domain-swapped complex based on the NMR data. One molecule of the full-length *Npu*DnaE intein is in red. The C-terminal 35-residue split intein fragment is in gray. **(e)** Two domain-swapped molecules of the Δvariant of *Npu*DnaE intein in the crystal. The two swapped chains are colored in purple or light yellow. The close-up shows the region (residues 94–98) where 3D-domain swapping occurs. The arrows in **c** and **e** indicate the naturally occurring split site, which also corresponds to conserved endonuclease insertion sites in many inteins.

the preceding loop formed intertwined dimers in the crystal structure (**Fig. 2e**, **Supplementary Fig. 5** and **Supplementary Table 1**). The crystal structure of the Δvariant clearly showed a 3D domain–swapped conformation, with the polypeptide chain exchange occurring after the loop where we introduced the three-residue deletion

and where the natural split site coincidently locates (**Fig. 2e** and **Supplementary Fig. 5**). This intertwined structure is in good agreement with our NMR data and the observed efficient iPAS of the I^C fragment (C35). The loop deletion seems responsible for inducing 3D-domain swapping in the crystal structure because the midpoint
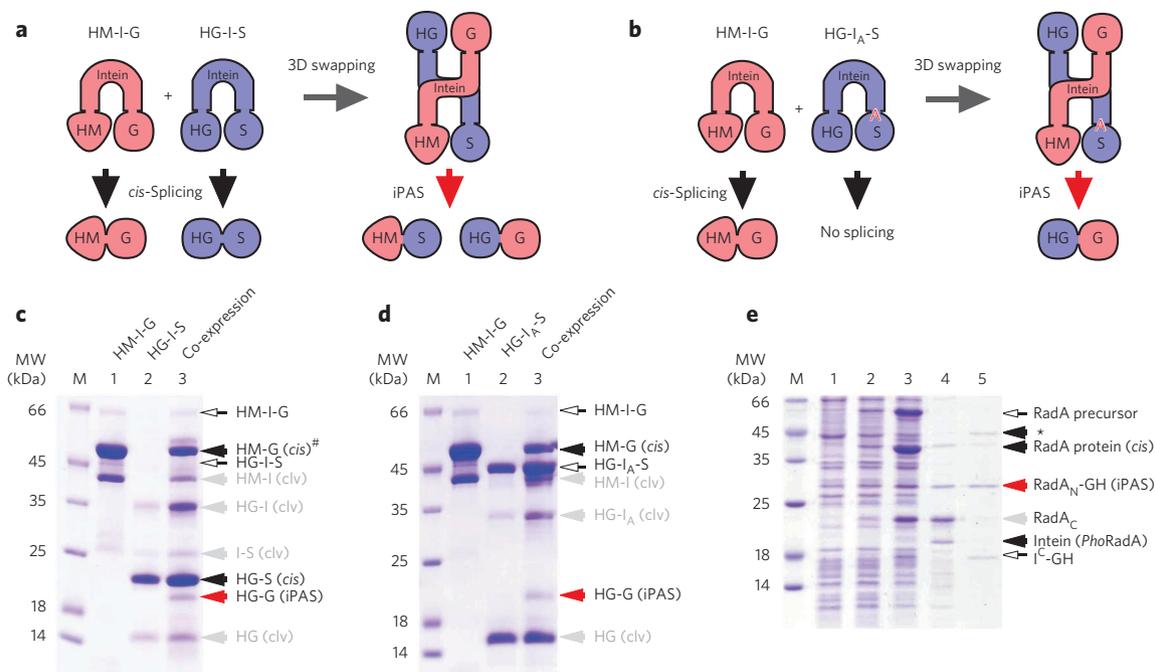
Figure 3 | iPAS without split inteins. (**a**) Schematic drawing of iPAS between two *cis*-splicing precursors (HM-I-G and HG-I-S), producing two alternatively ligated products (HM-S and HG-G). (**b**) Schematic drawing of iPAS between *cis*-splicing precursor (HM-I-G) and splicing-incompetent precursor (HG-I$_A$-S) with alanine mutations at the C-terminal splicing junction. 'A' in red indicates the mutation to alanine. (**c**) SDS-PAGE analysis of iPAS between two precursors containing a full-length *Tvo*VMA intein. Lane 1, elution fraction by IMAC from the cells expressing only HM-I-G precursor; lane 2, elution from the cells expressing only HG-I-S precursor; lane 3, elution from the cells expressing both HG-I-S and HM-I-G precursors. # indicates the position for the iPAS product of HM-S overlapping with HM-G. Side products produced by cleavages (clv) are in light gray. MW, molecular weight. (**d**) SDS-PAGE analysis of iPAS between two precursors containing active and inactive full-length *Tvo*VMA inteins. Lane 1, elution for HM-I-G precursor; lane 2, elution for inactive HG-I$_A$-S precursor; lane 3, elution from the cells expressing both HG-I$_A$-S and HM-I-G precursors. Open arrows indicate precursors. (**e**) iPAS with the wild-type *Pho*RadA protein precursor. Lane 1, before induction; lane 2, after 1-h induction with IPTG and L-arabinose; lane 3, after 6-h induction with IPTG and arabinose; lane 4, supernatant after heating at 75 °C for 25 min; lane 5, elution from IMAC. Asterisk indicates nonspecific binding to HisTrap column. Open arrows indicate unprocessed precursors. Black and red arrows indicate *cis*-splicing and iPAS products, respectively. Full gels for **c**–**e** are available in **Supplementary Figure 9**.

of thermal unfolding was decreased from 84 °C to 67 °C by the deletion, facilitating the open swapped conformation. The Δvariant, however, produced the alternatively ligated product as much as the original *cis*-splicing *Npu*DnaE intein under the same conditions, indicating that splicing kinetics is more important than stability for iPAS (**Fig. 1f** and **Supplementary Fig. 3d**)[17].

iPAS without split inteins

3D-domain swapping has been considered to be an 'entangling alliances' mechanism for creating new proteins with new properties without covalent linkage[27,28]. In the case of protein-splicing domains, 3D-domain swapping results in combinatorially ligated molecules with covalent peptide linkages, yielding diverse molecular repertoires from the identical genetic background. The 3D-swapped structure prompted us to investigate whether the intertwined dimers of two identical *cis*-splicing intein molecules could exist inside cells, which would produce two alternatively ligated products in addition to *cis*-splicing products (**Fig. 3**). 3D-domain swapping might also involve inactive precursors (**Fig. 3b**). However, we were unable to detect any iPAS product between two *cis*-splicing *Npu*DnaE intein precursors flanked with different exteins (SH3 domains (S) and maltose-binding protein (M)) as a model system under several conditions tested, presumably because *Npu*DnaE intein folds and splices too quickly for two chains to entangle to produce detectable iPAS products[29]. We detected alternatively ligated products when we replaced the intein in the two precursors with *Tvo*VMA intein (**Fig. 3c**). Although iPAS products were present in smaller amounts (~10%) with respect to *cis*-splicing products, the identified molecular masses (HG-G and HM-S) clearly

indicated the intermolecular reaction between the two separate chains, thereby increasing the diversity to four instead of two (**Supplementary Fig. 6**). We additionally confirmed the alternatively ligated product by inactivating *cis*-splicing in one of the precursors by mutating both the first residue of C-extein and the last residue of the intein to alanine (**Fig. 3b,d**). The confirmed alternatively spliced product of HG-G suggests that inactive inteins can still be involved in iPAS.

iPAS in the native context

Inteins usually do not leave any mark in the host protein upon self-excision. iPAS thus makes it even more difficult to trace the spliced product back to their origins. One important question is whether iPAS could happen with naturally occurring precursors (that is, with native exteins). We tested the wild-type 529-residue *Pho*RadA protein (UniProt accession code O58001) bearing a 172-residue intein in the Walker A motif by expressing it together with a split RadA intein precursor with C-terminally His$_6$-tagged GB1 as the C-extein (I^C-GH) for a pulldown assay[26]. We were able to efficiently isolate the alternatively ligated product of the native N-terminal fragment (N-extein) of RadA protein and C-terminally His$_6$-tagged GB1 together (RadA$_N$-GH; **Fig. 3e**) and confirmed its identity by its molecular mass (**Supplementary Fig. 7**). This observation suggests that iPAS could be a general phenomenon that could take place not only with the model system used here but also with native exteins.

Controlling iPAS

*Npu*DnaE intein is very stable, with a thermal unfolding midpoint of 84 °C once the three-dimensional fold is formed. This is consistent

with the fact that iPAS with *Npu*DnaE intein was visible only when we expressed the *cis*-splicing precursor after the expression of the other split precursor (**Fig. 4a**) and that we observed no alternative splicing of *Npu*DnaE intein *in vitro* with purified components. These observations indicate that folding and ligation kinetics are important for iPAS. The pre-existence of a split precursor, which is the driving force for the iPAS pathway, is a prerequisite for iPAS. We also found that iPAS was dependent on the relative amount of the split precursor (**Fig. 4b**). This suggests that the pre-existing intein fragment at high concentrations intercepts the intramolecular protein folding process of the nascent chain. We were able to direct >95% of the *cis*-splicing precursor to the iPAS pathway by carefully adjusting the protein expression and temperature, as previously described for *in vivo* segmental isotopic labeling by PTS[22]. The degree of iPAS occurring under identical conditions widely varies among different inteins, probably owing to the differences in folding and splicing kinetics (**Fig. 1f**). Many inteins contain an endonuclease domain that has an important role in the propagation of intein genes by horizontal gene transfer and is not required for protein splicing function[1,30]. Unlike the insertion sites of inteins, the insertion sites of endonuclease domains are well conserved among inteins and also coincide with the split sites of naturally occurring split inteins in cyanobacteria[4]. Notably, the observed 3D-swapping loop in the Δvariant structure also shares the same site (**Fig. 2e** and **Supplementary Fig. 5**), suggesting that endonuclease domains might have some influence on iPAS. We compared iPAS of *Npu*DnaB and *Mja*TFIIB inteins with and without the endonuclease domain to test whether the insertion of an endonuclease domain has any effects on iPAS. The full-length *Npu*DnaB intein, tested together with the C-terminal C39 intein fragment, showed more efficient iPAS than the *Npu*DnaB mini-intein, although the protein expression of the *Npu*DnaB intein was considerably lower owing to the cytotoxicity of the endonuclease domain (**Fig. 1f** and **Supplementary Fig. 3b**). In the case of *Mja*TFIIB intein, deletion of the endonuclease domain resulted in substantial loss of *cis*-splicing, redirecting it nearly completely to iPAS when we expressed the *cis*-splicing precursor bearing the *Mja*TFIIB mini-intein together with the split C-terminal precursor (**Fig. 1f** and **Supplementary Fig. 3c**). These data suggest that the insertion of a domain can modulate efficiency of iPAS presumably because of changes in the folding process and that an endonuclease domain might have a role in regulating iPAS. It is noteworthy that iPAS could rescue the less-active *Mja*TFIIB mini-intein. It has been shown that several inteins are not functional under specific conditions and are not necessarily optimized for the splicing function[16,17,31]. iPAS might have more important roles for poorly splicing and inactive inteins.

iPAS as a conformation-driven switch

We were interested in the capability to rescue inactive inteins by iPAS, as observed with the *Mja*TFIIB mini-intein and the inactivated *cis*-splicing precursor for controlling protein functions on the protein level (**Fig. 3b,d** and **Supplementary Fig. 3c**). To demonstrate the control of protein functions by iPAS, we developed a model system using GFP. The chromophore of GFP is autocatalytically produced with a specific sequence (Ser-Tyr-Gly in the wild-type GFP) in the central helix surrounded by β-barrels[32]. We inserted the *Npu*DnaE intein at the immediate upstream of the chromophore sequence (Cys-Tyr-Gly) with a serine-to-cysteine mutation for better protein splicing without loss of the fluorescence (**Fig. 5a**). Without excision of *Npu*DnaE intein and concomitant ligation of the flanking sequences, GFP cannot fold into a mature fluorescent protein. It is worth noting that, unlike other insertion sites in GFP, any side reactions such as cleavages caused by the intein mutants cannot form a fluorescent chromophore with this construct (that is, zero background). Mutating Ser-Asn to Ala-Gly at the C-terminal
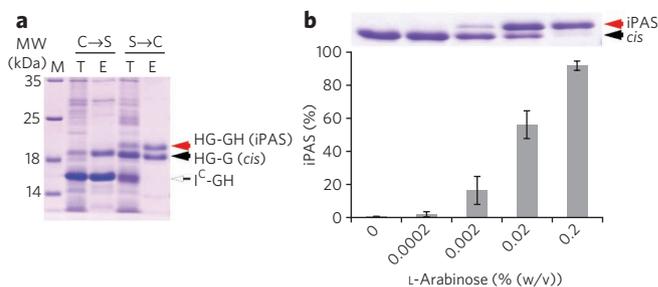
Figure 4 | Factors influencing iPAS. (**a**) Effect of the expression order on iPAS using *Npu*DnaE intein. C→S indicates that the *cis*-splicing precursor was first induced, followed by induction of the split precursor. S→C indicates that the split precursor was first induced, followed by induction of the *cis*-splicing precursor. The expression order was swapped by changing the promoters in the vectors to minimize effects from inducers. T, total cell lysate; E, elution from Ni-NTA column; MW, molecular weight. (**b**) Effects of the expression of the split precursor (C35) on iPAS of *Npu*DnaE intein tested with different concentrations of the inducer (L-arabinose). iPAS and *cis*-splicing products in SDS-PAGE are indicated by red and black arrows, respectively. Full gels are available in **Supplementary Figure 9**. Data represent mean ± s.d., typically from three or more independent experiments.

junction indeed resulted in no fluorescence owing to the lack of protein splicing (**Fig. 5b**). We were able to restore the fluorescence of the host protein (GFP) harboring the inactive intein by expressing a C-terminal split intein fragment containing the active sequence of Ser-Asn together with the C-terminal split GFP fragment (GFP^C) (**Fig. 5b** and **Supplementary Fig. 8**). Conversely, it was possible to suppress the green fluorescence of GFP bearing the active intein by adjusting the expression of an active split intein fused with an unrelated sequence (GB1) before the induction of the *cis*-splicing GFP precursor (**Fig. 5a,c**). In contrast, a split intein fragment derived from an orthogonal intein (*Mja*TFIIB) was unable to induce iPAS, showing no suppression of the fluorescence because the *Mja*TFIIB intein is unreactive (orthogonal) to the *Npu*DnaE intein (**Fig. 5a,c**). Thus, the pre-existence of appropriate amounts of the reactive intein fragment could interfere with protein folding of the *cis*-splicing precursor and direct it to the iPAS pathway in a highly specific manner.

Prevalence of inteins and split inteins

iPAS requires a split intein fragment and a *cis*-splicing intein or two *cis*-splicing inteins that are cross-reactive (nonorthogonal), as we demonstrated above. Natural split inteins have been found in DnaE polymerase in diverse cyanobacteria but not in many other species, diminishing the likelihood of iPAS as a common physiological process exploited in nature (**Supplementary Table 2**)[4]. However, recent metagenomic data suggest that split intein fragments could be more common with phage or viral origins[14,33]. Additionally, we demonstrated that split intein fragments were not always necessary for iPAS, implying some possibilities of iPAS by commonly occurring *cis*-splicing inteins as physiological processes in nature. Indeed, more than one intein (up to 19 inteins) are identified in about one-fourth of the organisms carrying inteins, including viruses and bacteriophages (**Supplementary Table 2**)[4]. Several inteins can simultaneously integrate into one protein. Forty-four proteins in InBase are interrupted by more than one intein (up to four inteins) (**Supplementary Table 2**). Thus, iPAS could potentially take place within a single organism or even within one protein and/or upon infection of viruses or bacteriophages carrying inteins. Naturally occurring iPAS with physiological relevance remains to be discovered.
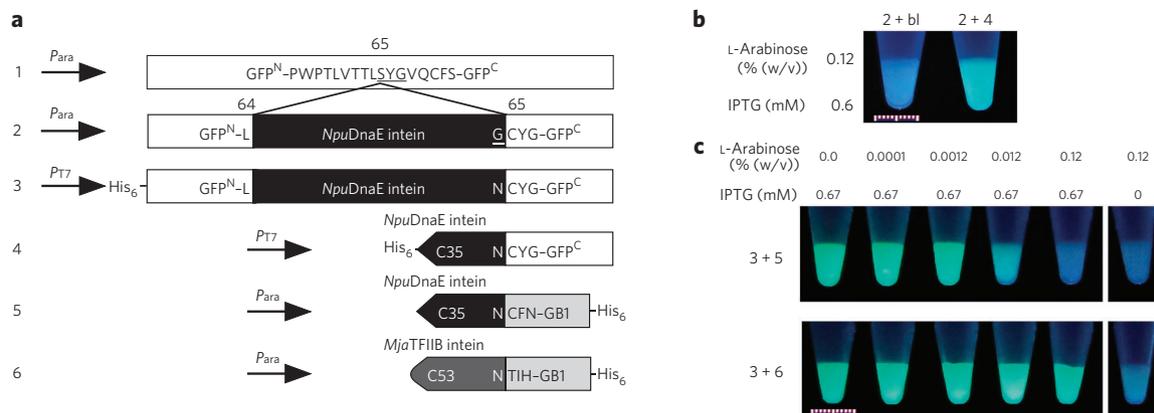
Figure 5 | Modulation of GFP fluorescence by the conformation-driven switch. (**a**) Schematic presentation of GFP constructs. (1) Original GFP without *Npu*DnaE intein and other modifications. (2) GFP with splicing-incompetent *Npu*DnaE intein. (3) GFP with *cis*-splicing *Npu*DnaE intein. (4) Active split *Npu*DnaE intein precursor (I^C) with C-terminal split GFP (GFP^C). (5) Active split *Npu*DnaE intein precursor (I^C) with C-terminally His$_6$-tagged GB1 as C-extein. (6) Orthogonal split *Mja*TFIIB intein precursor (I^C) with C-terminally His$_6$-tagged GB1 as C-extein. Construct 2 contains Ala-Gly mutations at the C-terminal splicing junction that inhibit *cis*-splicing (the glycine is shown and underlined). The sequence of the central helix is shown together with the chromophore-forming three residues underlined in construct 1. The expression of each precursor was under the control of either IPTG (P_{T7} in pRSF vector) or L-arabinose (P_{ara} in pBAD vector), as indicated. (**b**) Activation of GFP by iPAS. 2 + bl is a control showing induced *E. coli* cell suspension harboring the plasmid carrying construct 2 and a blank vector without any insert. 2 + 4 is the cell suspension expressing both constructs 2 and 4. (**c**) Suppression of GFP by iPAS. 3 + 5 and 3 + 6 show induced *E. coli* cell suspension harboring the plasmids carrying constructs 3 and 5 or 3 and 6, induced with different concentrations of L-arabinose and 0.67 mM IPTG (or 0.0 mM (far right)). A UV lamp with a wavelength of 365 nm was used to illuminate tubes for taking images. Scale bars, 1 cm.

DISCUSSION

Our discovery of iPAS as a conformation-driven switch lays new ground for increasing protein diversity, that is, functions by inteins and intein fragments without any genetic modifications, which usually require many generations owing to low mutation rates. iPAS as a post-translational protein ligation mechanism further supports the previously suggested primeval role of inteins for increasing protein diversity by combinatorial protein recombination because iPAS could take place without any split intein fragment, producing up to four combinatorially ligated molecules from two precursor genes without having multiple copies of genes[11,13]. This number can be further increased because some inteins are cross-reactive (nonorthogonal) to each other, as observed for naturally split DnaE inteins[15,34], and because inteins are generally capable of ligating foreign exteins, even though some inteins are less tolerant of changes in splicing junctions[15,26,35]. Intein distributions are sporadic because of lateral gene transfer by homing events mediated by the endonuclease domains and the loss of inteins[13,30,36]. We speculate that viruses and bacteriophages harboring inteins could function as vehicles that not only disseminate inteins by horizontal gene transfer but also might provide fitness benefits to the host by rescuing less-active inteins (or provide disadvantage by inactivating some functional inteins) by the iPAS mechanism via split or nonsplit inteins before genetic fixation. iPAS might account for the strict strain specificity of viruses carrying inteins[37].

iPAS is an irreversible process sensitive to the order of precursor expression. This requirement could be used as a biotechnological tool for studying sequences of events happening in various biological processes in living cells in developmental biology, neural circuits, translocation to specific organelles, protein folding and others because the pre-existence of one of the precursor components would result in different irreversible outputs on the protein level, as demonstrated with GFP fluorescence. The GFP reporter system developed here might also be useful for analyzing 3D-domain swapping in living cells and for identifying other factors influencing iPAS. As shown with GFP and RadA, iPAS can be potentially used for targeting inteins found in several essential proteins in various mycobacterial and fungal pathogens such as *Mycobacterium tuberculosis*

and *Cryptococcus neoformans*[38,39]. Increased protein diversity by iPAS could serve as a new tool for directed evolution of proteins because protein diversity can be increased *in vivo* by combinatorial recombination on the protein level from limited genetic libraries, of which diversity is often limited by transformation efficiency of bacterial cells[40].

Protein alternative splicing not only provides the possibility to expand molecular variability at the protein level but also creates challenges in identifying their original genes from the protein species responsible for unique functions. As protein splicing is traceless from spliced products, proteins produced by iPAS might have easily escaped from any investigation. This is further complicated by cross-reactions among nonorthogonal inteins. Additionally, inteins are flexible enzymes that can achieve protein splicing through diversified mechanisms[41]. Variations of protein-splicing mechanisms by intein and intein-related sequences are still increasing[41,42]. As more extensive genomic, metagenomic and proteomic data are available at an increasingly rapid pace, we anticipate more inteins and intein-related sequences as well as diversified natural and contrived molecules produced by iPAS, which could advance our understanding of their biochemical and physiological roles and their biotechnological applications.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Protein Data Bank (PDB): the coordinates of the structures of *Npu*DnaE intein and Δvariant are deposited under accession codes 4KL5 and 4KL6, respectively.

Addgene: all of the plasmids described in this article have been deposited under the following deposition numbers: pSK-Duet16, 41684; pMHBAD14C, 42304; pMKBAD28, 45595; pSABAD219, 45596; pSKDuet26, 41689; pSABAD331, 45609; pHYDuet183, 41685; pSABAD332, 45610; pHYDuet216A, 45611;

pMMDuet19, 45594; pSADuet502, 46373; pSABAD250, 45612; pSKDuet20, 41694; pSADuet760, 45613; pSABAD14-750, 45616; pSEDuet40, 46378; pSABAD25-74, 46375; pSABAD518, 46374; pJORSF19, 45614; pALBRSF12, 46376; pHYRSF236, 46377; pSABAD505, 45615; pSARSF505, 46832; pSABAD544, 45617; pSARSF534-3, 45618.

References

1. Paulus, H. Protein splicing and related forms of protein autoprocessing. *Annu. Rev. Biochem.* **69**, 447–496 (2000).
2. Hirata, R. *et al.* Molecular structure of a gene, *VMA1*, encoding the catalytic subunit of H⁺-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **265**, 6726–6733 (1990).
3. Kane, P.M. *et al.* Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H⁺-adenosine triphosphatase. *Science* **250**, 651–657 (1990).
4. Perler, F.B. InBase, the Intein Database. *Nucleic Acids Res.* **30**, 383–384 (2002).
5. Callahan, B.P., Topilina, N.I., Stanger, M.J., Van Roey, P. & Belfort, M. Structure of catalytically competent intein caught in a redox trap with functional and evolutionary implications. *Nat. Struct. Mol. Biol.* **18**, 630–633 (2011).
6. Wu, H., Hu, Z. & Liu, X.Q. Protein *trans*-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci. USA* **95**, 9226–9231 (1998).
7. Swithers, K.S., Senejani, A.G., Fournier, G.P. & Gogarten, J.P. Conservation of intron and intein insertion sites: implications for life histories of parasitic genetic elements. *BMC Evol. Biol.* **9**, 303 (2009).
8. Kawasaki, M., Satow, Y., Ohya, Y. & Anraku, Y. Protein splicing in the yeast Vma1 protozyme: evidence for an intramolecular reaction. *FEBS Lett.* **412**, 518–520 (1997).
9. Southworth, M.W. *et al.* Control of protein splicing by intein fragment reassembly. *EMBO J.* **17**, 918–926 (1998).
10. Mootz, H.D. Split inteins as versatile tools for protein semisynthesis. *ChemBioChem* **10**, 2579–2589 (2009).
11. Perler, F.B. A natural example of protein *trans*-splicing. *Trends Biochem. Sci.* **24**, 209–211 (1999).
12. Caspi, J., Amitai, G., Belenkiy, O. & Pietrokovski, S. Distribution of split DnaE inteins in cyanobacteria. *Mol. Microbiol.* **50**, 1569–1577 (2003).
13. Pietrokovski, S. Intein spread and extinction in evolution. *Trends Genet.* **17**, 465–472 (2001).
14. Dassa, B., London, N., Stoddard, B.L., Schueler-Furman, O. & Pietrokovski, S. Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. *Nucleic Acids Res.* **37**, 2560–2573 (2009).
15. 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**, 1853–1858 (2006).
16. 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).
17. Aranko, A.S., Züger, S., Buchinger, E. & Iwaï, H. *In vivo* and *in vitro* protein ligation by naturally occurring and engineered split DnaE inteins. *PLoS ONE* **4**, e5185 (2009).
18. Oeemig, J.S., Aranko, A.S., Djupsjöbacka, J., Heinämäki, K. & Iwaï, H. Solution structure of DnaE intein from *Nostoc punctiforme*: structural basis for the design of a new split intein suitable for site-specific chemical modification. *FEBS Lett.* **583**, 1451–1456 (2009).
19. Liu, Y. & Eisenberg, D. 3D domain swapping: as domains continue to swap. *Protein Sci.* **11**, 1285–1299 (2002).
20. Züger, S. & Iwai, H. Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. *Nat. Biotechnol.* **23**, 736–740 (2005).
21. Heinämäki, K., Oeemig, J.S., Pääkkonen, K., Djupsjöbacka, J. & Iwaï, H. NMR resonance assignment of DnaE intein from *Nostoc punctiforme*. *Biomol. NMR Assign.* **3**, 41–43 (2009); erratum **7**, 115–116 (2013).
22. Muona, M., Aranko, A.S., Raulinaitis, V. & Iwai, H. Segmental isotopic labeling of multi-domain and fusion proteins by protein *trans*-splicing *in vivo* and *in vitro*. *Nat. Protoc.* **5**, 574–587 (2010).
23. Klabunde, T., Sharma, S., Telenti, A., Jacobs, W.R. & Sacchettini, J.C. Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing. *Nat. Struct. Biol.* **5**, 31–36 (1998).
24. Mizutani, R. *et al.* Protein-splicing reaction via a thiazolidine intermediate: crystal structure of the VMA1-derived endonuclease bearing the N and C-terminal propeptides. *J. Mol. Biol.* **316**, 919–929 (2002).
25. Sun, P. *et al.* Crystal structures of an intein from the split *dnaE* gene of *Synechocystis* sp. PCC 6803 reveal the catalytic model without the penultimate histidine and the mechanism of zinc ion inhibition of protein splicing. *J. Mol. Biol.* **353**, 1093–1105 (2005).
26. Oeemig, J.S., Zhou, D., Kajander, T., Wlodawer, A. & Iwaï, H. NMR and crystal structures of the *Pyrococcus horikoshii* RadA intein guide a strategy for engineering a highly efficient and promiscuous intein. *J. Mol. Biol.* **421**, 85–99 (2012).
27. Ogihara, N.L. *et al.* Design of three-dimensional domain-swapped dimers and fibrous oligomers. *Proc. Natl. Acad. Sci. USA* **98**, 1404–1409 (2001).
28. Bennett, M.J., Choe, S. & Eisenberg, D. Domain swapping: entangling alliances between proteins. *Proc. Natl. Acad. Sci. USA* **91**, 3127–3131 (1994).
29. Zettler, J., Schütz, V. & Mootz, H.D. The naturally split *Npu* DnaE intein exhibits an extraordinarily high rate in the protein *trans*-splicing reaction. *FEBS Lett.* **583**, 909–914 (2009).
30. Nogami, S. *et al.* Homing at an extragenic locus mediated by VDE (PI-SceI) in *Saccharomyces cerevisiae*. *Yeast* **19**, 773–782 (2002).
31. Tori, K. & Perler, F.B. The *Arthrobacter* species FB24 Arth_1007 (DnaB) intein is a pseudogene. *PLoS ONE* **6**, e26361 (2011).
32. Tsien, R.Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509–544 (1998).
33. Clerissi, C., Grimsley, N. & Desdevises, Y. Genetic exchanges of inteins between prasinoviruses (phycodnaviridae). *Evolution* **67**, 18–33 (2013).
34. Dassa, B., Amitai, G., Caspi, J., Schueler-Furman, O. & Pietrokovski, S. *Trans* protein splicing of cyanobacterial split inteins in endogenous and exogenous combinations. *Biochemistry* **46**, 322–330 (2007).
35. Appleby-Tagoe, J.H. *et al.* Highly efficient and more general *cis*- and *trans*-splicing inteins through sequential directed evolution. *J. Biol. Chem.* **286**, 34440–34447 (2011).
36. Gogarten, J.P., Senejani, A.G., Zhaxybayeva, O., Olendzenski, L. & Hilario, E. Inteins: structure, function, and evolution. *Annu. Rev. Microbiol.* **56**, 263–287 (2002).
37. Clerissi, C., Desdevises, Y. & Grimsley, N. Prasinoviruses of the marine green alga *Ostreococcus tauri* are mainly species specific. *J. Virol.* **86**, 4611–4619 (2012).
38. Paulus, H. Inteins as targets for potential antimycobacterial drugs. *Front. Biosci.* **8**, s1157–s1165 (2003).
39. Liu, X.-Q. & Yang, J. Prp8 intein in fungal pathogens: target for potential antifungal drugs. *FEBS Lett.* **572**, 46–50 (2004).
40. Amstutz, P., Forrer, P., Zahnd, C. & Plückthun, A. *In vitro* display technologies: novel developments and applications. *Curr. Opin. Biotechnol.* **12**, 400–405 (2001).
41. Tori, K. & Perler, F.B. Expanding the definition of class 3 inteins and their proposed phage origin. *J. Bacteriol.* **193**, 2035–2041 (2011).
42. Aranko, A.S., Oeemig, J.S. & Iwaï, H. Structural basis for protein *trans*-splicing by a bacterial intein-like domain: protein ligation without nucleophilic residues. *FEBS J.* **280**, 3256–3269 (2013).

Author contributions

A.S.A. and H.I. designed and performed experiments, analyzed data and prepared the manuscript; T.K., J.S.O. and H.I. performed crystallographic experiments and analyzed diffraction data; H.I. and J.S.O. performed NMR measurements and analyzed NMR data.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index. html. Correspondence and requests for materials should be addressed to H.I.

ONLINE METHODS

Plasmid construction. Methods for constructing the used plasmids, which are listed in **Supplementary Table 3**, are described in the **Supplementary Note**.

Comparison of iPAS between different inteins. Each pair of the two plasmids encoding *cis*-splicing and split precursors was transformed together into *E. coli* ER2566 (New England Biolabs). The cells were grown at 37 °C in 5 ml LB medium supplemented with kanamycin (25 μg/ml) and ampicillin (100 μg/ml) and induced at $D_{600\,nm}$ = 0.4–0.6 with 0.09% (w/v) L-arabinose for 1 h, followed by addition of IPTG at a final concentration of 0.67 mM and incubation for additional 4 h. The cells were harvested by centrifugation. The expression of the proteins was analyzed as described in the following section.

Quantitative analysis of iPAS. The expressed proteins were analyzed by resuspending and lysing the cells in 100 μL of Bacterial Protein Extraction Reagent (B-PER, Thermo Scientific). After removal of cell debris by centrifugation, His$_6$-tagged proteins were purified using a Ni-NTA spin column (Qiagen). The amount of ligated products was quantified by analyzing elution fractions on 18% SDS-PAGE. The intensities of protein bands were quantified using ImageJ (US National Institutes of Health) software. The percentages of iPAS were estimated with respect to the total spliced products (iPAS and *cis*-splicing products), assuming both proteins bind staining dye equally. s.d. were estimated from at least three independent experiments.

iPAS by *Npu*DnaE intein with different split sites. Different split sites of *Npu*DnaE intein were tested for iPAS using *cis*-splicing precursor (pSK-Duet16) and split intein precursor with different split C-inteins (pMH-BAD14C (C35), pMKBAD28 (C14) or pSABAD219 (C102)). The pair of two plasmids was transformed into *E. coli* ER2566 cells, which were grown at 37 °C in 5 mL LB medium supplied with kanamycin (25 μg/ml) and ampicillin (100 μg/ml). The cells were first induced at $D_{600\,nm}$ = 0.4–0.6 with a final concentration of 0.09% (w/v) L-arabinose for 1 h. The second inducer of IPTG was added at a final concentration of 0.67 mM. The cells were incubated for additional 4 h. The cells were harvested by centrifugation and analyzed as described above.

iPAS by *Npu*DnaE intein and its Δvariant. pSKDuet16 or pHYDuet216A was transformed together with pMHBAD14C into *E. coli* ER2566 and grown at 37 °C in 5 ml LB medium supplemented with ampicillin (100 μg/ml) and kanamycin (25 μg/ml). The split C-intein precursor was first induced by addition of a final concentration of 0.09% (w/v) L-arabinose. The cells were incubated for 1 h, followed by induction by IPTG at a final concentration of 0.67 mM for additional 4 h. The cells were harvested by centrifugation and analyzed as described above.

Effect of the expression order on iPAS of *Npu*DnaE intein. A pair of two plasmids, pSKDuet16 and pMHBAD14C (for S→C) or pSABAD25-SK16 and pSADuet194-14 (for C→S), was transformed into *E. coli* ER2566 to test the effect of the expression order of the two precursors. The transformed cells were grown in 5 mL LB medium supplemented with kanamycin (25 μg/ml) and ampicillin (100 μg/ml) at 37 °C. The first protein (either split precursor or *cis*-splicing precursor) was induced with 0.09% (w/v) L-arabinose. After 1.5 h of the L-arabinose induction, the cells were spun down at 900*g* for 10 min at 25 °C, followed by a brief wash with prewarmed LB medium[22]. The cells were suspended after the wash with prewarmed 5 ml LB medium supplemented with kanamycin (25 μg/ml), ampicillin (100 μg/ml) and IPTG (at a final concentration of 0.67 mM) for the induction of the second protein. The induction of the second protein continued for additional 4 h. The cells were harvested by centrifugation and analyzed as described above.

Effect of expression levels of a split precursor on iPAS of *Npu*DnaE intein. Two plasmids of pSKDuet16 and pMHBAD14C were transformed together into *E. coli* ER2566 and grown at 37 °C in 5 ml LB medium supplemented with kanamycin (25 μg/ml) and ampicillin (100 μg/ml). The cells were induced at $D_{600\,nm}$ = 0.4–0.6 with different concentrations of L-arabinose (at a final concentration of 0%, 0.0002%, 0.002%, 0.02% or 0.2% (w/v)). The cells were incubated for 1 h before addition of IPTG at a final concentration of 0.67 mM. The cells were incubated for additional 4 h after the second induction. The cells were harvested by centrifugation and analyzed as above.

iPAS between two *cis*-splicing precursors with *Tvo*VMA intein. The plasmids pSEDuet40 and pSABAD25-74 (or pSABAD518 for the inactive precursor) were transformed together into *E. coli* ER2566. The cells were growth at 37 °C in 5 ml LB medium supplemented with kanamycin (25 μg/ml) and ampicillin (100 μg/ml). At $D_{600\,nm}$ = 0.4–0.6, the cells were induced by addition of L-arabinose at a final concentration of 0.13% (w/v) and incubated at 25 °C for 0.5 h. The cells were further induced with a final concentration of 1 mM IPTG for another ~16 h and harvested by centrifugation for the analysis.

For MS, *E. coli* cells bearing both pSEDuet40 and pSABAD25-74 were grown at 37 °C in 2 l LB medium supplemented with kanamycin (25 μg/ml) and ampicillin (100 μg/ml) and induced with a final concentration of 0.13% (w/v) L-arabinose at a $D_{600\,nm}$ of 0.6. After 0.5 h of the induction with L-arabinose, the cells were additionally induced with a final concentration of 1.1 mM IPTG for 6 h. This was followed by incubation at 25 °C for additional 12 h. The cells were harvested by centrifugation at 5,050*g* for 10 min at 4 °C and resuspended with Buffer A (50 mM Tris-HCl, 1 mM EDTA, 10 mM NaCl, pH 8). The cells were lysed by ultrasonication. The lysate was cleared by centrifugation at 38,500*g* for 1 h at 4 °C. The supernatant was loaded on a 5-ml HisTrap HP column (GE Healthcare) for IMAC purification. The elution fraction from IMAC was further purified after dialysis against 2 l of 10 mM sodium phosphate pH 8 overnight on an anion exchange column (MonoQ GL5/15, GE Healthcare) with a linear gradient of 0–500 mM NaCl. The elution fractions were analyzed on SDS-PAGE and subjected to analysis by MALDI-TOF MS (Ultraflex, Bruker Daltonics) at the Protein Chemistry Unit of Institute of Biotechnology.

MS of alternative spliced products from *cis*-splicing and split precursors. Pairs of two plasmids (pSKDuet16 and pMHBAD14C for *Npu*DnaE intein, pHYDuet183 and pSABAD332 for *Pho*RadA intein, pSKDuet26 and pSA-BAD331 for *Tvo*VMA intein, pSADuet760 and pSABAD14-750 for *Mja*TFIIB intein, pSADuet502 and pSABAD250 for *Npu*DnaB intein) were transformed together into *E. coli* strain ER2566 cells. The cells were growth in 0.5–2 l LB medium supplemented with ampicillin (100 μg/ml) and kanamycin (25 μg/ml). At $D_{600\,nm}$ = 0.4–0.6, the cells were induced for 1 h with a final concentration of 0.08–0.16% (w/v) L-arabinose to express the split intein precursor. Following the expression of the split precursor, the other *cis*-splicing precursor was induced with a final concentration of 0.7–1 mM IPTG for additional 3.5–4 h. The cells were harvested by centrifugation at 5,050*g* for 10 min at 4 °C and resuspended into Buffer A. After cell lysis by ultrasonication, the cell lysate was cleared at 4 °C by centrifugation at 38,500*g* for 1 h. The spliced products were purified by the N- and/or C-terminal His$_6$ tags by IMAC using a 5-ml HisTrap HP column. For *Mja*TFIIB intein, the elution fractions from the HisTrap HP column were dialyzed against 2 l 10 mM sodium phosphate, pH 8, overnight and further purified on a MonoQ 5/50 GL column (GE Healthcare) by applying a linear gradient of NaCl from 0 mM to 200 mM. The elution fraction containing the presumed iPAS products were analyzed by MALDI-TOF MS (Ultraflex, Bruker Daltonics) at the Protein Chemistry Unit of Institute of Biotechnology.

iPAS of native *Pho*RadA intein precursor. Plasmids of pJORSF19 and pSA-BAD332 were transformed together into *E. coli* ER2566 cells carrying an additional plasmid of pRARE for supplementing tRNA for rare codons. The transformed cells were grown in 2 l LB medium supplemented with kanamycin (25 μg/ml), ampicillin (100 μg/ml) and chloramphenicol (5 μg/ml) to $D_{600\,nm}$ = 0.4. The split intein was first induced by addition of a final concentration of 0.13% (w/v) L-arabinose for 0.5 h at 37 °C, followed by 6-h induction of the wild-type RadA precursor by addition of a final concentration of 1 mM IPTG at 37 °C. The cell culture was further incubated at 25 °C for an additional 15 h. The cells were harvested by centrifugation at 5,050*g* for 10 min and resuspended into Buffer A. The cells were lysed by incubation at 75 °C for 25 min. The cell suspension was cleared at 4 °C by centrifugation at 38,500*g* for 50 min. The supernatant was loaded on a 5-ml HisTrap HP column. The elution fraction was analyzed by MALDI-TOF MS.

NMR analysis of domain swapping with *Npu*DnaE intein. NMR sample was produced by overexpressing the two precursor proteins by using the time-delayed dual expression system[20,22]. *E. coli* ER2566 (New England Biolabs) was transformed with the two plasmids pJDJRSF05 (*Npu*DnaE intein with C1A)[21] and pMHBAD14C (35-residue C-intein with GB1 and His$_6$ tag) and grown in LB medium supplemented with ampicillin (100 μg/mL) and kanamycin

(25 μg/mL). First, the split intein precursor containing the C-intein was expressed by addition of a final concentration of 0.2% (w/v) L-arabinose for 3 h, followed by the exchange of the medium to either ^{15}N-labeled or doubly 20% ^{13}C- and 100% ^{15}N-labeled M9 medium containing ampicillin (100 μg/mL), kanamycin (25 μg/mL) and 1.0 mM IPTG[22,43]. After 3 h of the second induction, the cells were harvested by centrifugation. The presumed complex was purified by IMAC. The N-terminal His$_6$-Smt3 protein was digested by ubiquitin-like protease 1 and removed by IMAC together with other remaining His$_6$-tagged proteins as described previously[21,22]. NMR experiments were performed either on Varian INOVA 600 MHz or 800 MHz equipped with a cryogenic probehead. The resonance assignments are confirmed by [^{1}H,^{15}N]-HSQC, HNCA, HNCACB and CC(CO)NH experiments on the basis of the previously reported chemical shifts as the starting point[18,21].

Crystallographic analysis. *Crystal structure of monomeric* Npu*DnaE intein.* The protein encoded in pALBRSF12 was expressed and purified by the same procedure as that described previously[21]. The crystallization condition was screened using the Index HT screen (Hampton Research). Crystallization screens were performed using sitting-drop vapor diffusion in a 96-well plate (Innovadyne SD-2) with an 80-μL reservoir solution and protein drops of 100 nL mixed with 100 nL reservoir solution. All of the crystallization setups were performed at 293 K and at a Npu*DnaE intein concentration of 29 mg/mL. Crystallization hits from the screening were optimized by grid design using sitting-drop vapor diffusion with 200-nL protein drops mixed with 200 nL reservoir solution. The diffracting crystal was collected from a drop containing 1.4 M triammonium citrate–citric acid (pH 6.5) and cryoprotected with paratone-N. The diffraction data were collected at ESRF ID23-1 at a wavelength of 0.976 Å at 100 K, indexed, integrated and scaled to 1.72-Å resolution using the program XDS[44]. The space group was confirmed using the software Pointless[45]. Initial phases were solved using the software Phaser[46] using the Npu*DnaE intein NMR structure (PDB code 2KEQ)[18] as the search model, and two molecules were identified in the asymmetric unit. Model refinement was performed using PHENIX[47] using automatic water update in initial refinement rounds. For final refinement rounds, a TLS parameter was applied for each molecule, and the model was manually corrected using Coot[48]. The crystal structures of Npu*DnaE intein was solved at 1.72-Å resolution, with 97% of residues in the favored region of Ramachandran plot. The refinement statistics is summarized in **Supplementary Table 1**.

Crystal structure of Δ*variant of* Npu*DnaE intein.* The protein in the plasmid pHYRSF236 was expressed and purified with the same procedure as that described previously[21]. Crystallization conditions were screened as described above using a protein concentration of 30 mg/mL. A single crystal was picked from a drop containing 0.1 M Tris (pH 8.5) and 2.0 M ammonium phosphate, and X-ray diffraction data was collected at the ESRF at beamline ID29 at the wavelength 1.000 Å at 100 K. The diffraction data was indexed, integrated and scaled to 2.2-Å resolution using the program XDS[44]. Initial phases were solved by molecular replacement using Phaser[46] using the Npu*DnaE intein NMR structure (PDB code 2KEQ)[18] as a search model, and four molecules were identified in the asymmetric unit. Model refinement was performed using PHENIX[47] and Refmac5 (ref. 49). The domain-swapped dimer-loop structures and connectivity between the monomers were rebuilt on the basis of the initial F_o-F_c difference density maps and $2F_o$-F_c omit maps. For the final refinement rounds, water molecules were added with PHENIX[47] and edited with Coot[48]. The data collection and refinement statistics are summarized in **Supplementary Table 1**. Δvariant was solved at 2.20-Å resolution, with 94% of residues in the favored region of the Ramachandran plot.

Thermal stabilities of Npu**DnaE intein and** Δ**variant.** Thermal stabilities of Npu*DnaE intein and Δvariant were analyzed by Thermoflour-based assay using freshly diluted SYPRO orange dye (1:5,000)[50]. Both protein concentrations were adjusted to 31 μM in 20 mM Tris HCl (pH 7) and 1 mM EDTA. Fluorescence intensities were monitored while increasing the temperature from 20 °C to 99 °C with steps of 0.5 °C in the MiniOpticon system (Biorad). A midpoint of the melting curve (T_m) was determined by MiniOpticon Software (Biorad).

iPAS as a switch for fluorescence of GFPs. *Activation of GFP.* To demonstrate activation of the fluorescence of GFP, plasmids of pSABAD544 (encoding a splicing-incompetent precursor of GFP^{N}-Npu*DnaE(AG)-GFP^{C}) and pSARSF534-3 (encoding a splicing-competent split precursor of His$_6$-Int^{C}${}_{C35}$-GFP^{C}) or a blank plasmid of pRSFDuet-1 were transformed together into *E. coli* strain ER2566 for the protein expression. The transformed *E. coli* ER2566 cells were growth in LB medium supplemented with kanamycin (25 μg/mL) and ampicillin (100 μg/mL) at 30 °C. The split precursor in pRSF vector (or a blank vector of pRSFDuet-1) was first induced with a final concentration of 0.6 mM IPTG at $D_{600\,nm}$ = ~0.5. After 0.5 h of the induction by IPTG at 30 °C, the protein in pBAD-vectors (*cis*-splicing precursor) was induced by addition of L-arabinose at a final concentration of 0.12% (w/v). The two precursors were induced at 30 °C for another 18–20 h. The cells were harvested by centrifugation for taking images under UV light (365 nm) or for western blotting analysis. As a control for western blotting analysis, pSABAD505 (encoding a *cis*-splicing precursor of GFP^{N}-Npu*DnaE-GFP^{C}) and pRSFDuet-1 were transferred together into *E. coli* ER2566 cells and expressed, following the same protocol as described above. In addition, pJJBAD7 for expressing uninterrupted GFP was transformed into *E. coli* ER2566 cells. The cells were grown at 37 °C in LB medium supplemented with ampicillin (100 μg/mL) and induced at 30 °C for 20 h with a final concentration of 0.12% (w/v) L-arabinose when $D_{600\,nm}$ reached 0.5.

Suppression of GFP. pSARSF505 (His$_6$-GFP^{N}-Npu*DnaE-GFP^{C}) and pMHBAD14C (Int^{C}${}_{Npu\mathrm{DnaE\_C35}}$-GB1-His$_6$) or pSABAD14-750 (Int^{C}${}_{Mja\mathrm{TFIIB_C53}}$-GB1-His$_6$, a control plasmid) were transformed together into *E. coli* ER2566 cells and grown at 37 °C in LB medium supplemented with kanamycin (25 μg/mL) and ampicillin (100 μg/mL). The transformed cells were induced with different final concentrations (0%, 0.00012%, 0.0012%, 0.012% or 0.12% (v/w)) of L-arabinose when $D_{600\,nm}$ reached 0.4–0.6. Additionally, one control culture was induced with a final concentration of 0.12% (v/w) L-arabinose and without IPTG. All the cell cultures were incubated for 1 h after the arabinose induction, followed by the second induction by IPTG at a final concentration of 0.67 mM. The cells were grown for additional 4 h after the second induction and were harvested by centrifugation. The cell pellets were suspended with phosphate-buffered saline (PBS) for taking images under UV light (365 nm).

Western blotting of splicing products from GFP constructs with and without insertion of Npu*DnaE intein.* The total cell lysates were separated by 18% SDS-PAGE, which were transferred on the nitrocellulose filter membrane Hybond-C Extra membrane (GE Healthcare). Antibody to GFP (anti-GFP) recognizing residues 132–144 of the GFP (Sigma-Aldrich, G1546) was used at 1:1,500 dilution as the primary antibody. After incubation with anti-mouse IgG alkaline phosphatase conjugate (Sigma, A3484) at 1:5,000 dilution as a secondary antibody, the blotted membranes were visualized by the colorimetric method using premixed BCIP/NBT solution (Sigma, B6404).

43. 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).

44. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).

45. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).

46. McCoy, A.J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

47. Adams, P.D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1948–1954 (2002).

48. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

49. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).

50. Ericsson, U.B., Hallberg, B.M., Detitta, G.T., Dekker, N. & Nordlund, P. Thermoflour-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **357**, 289–298 (2006).