Published in final edited form as:

J Mol Biol. 2012 August 3; 421(1): 85–99. doi:10.1016/j.jmb.2012.04.029.

NMR and crystal structures of the *Pyrococcus horikoshii* RadA intein guide a strategy for engineering a highly efficient and promiscuous intein

Jesper S. Oeemig¹, Dongwen Zhou², Tommi Kajander¹, Alexander Wlodawer², and Hideo Iwai^{1,*}

¹Research Program in Structural Biology and Biophysics, Institute of Biotechnology, University of Helsinki. P.O. Box 65, Helsinki, FIN-00014, Finland ²Macromolecular Crystallography Laboratory, National Cancer Institute at Frederick, Frederick, MD, 21702, USA

Abstract

In protein splicing, an intervening protein sequence (intein) in the host protein excises itself out and ligates two split host protein sequences (exteins) to produce a mature host protein. Inteins require the involvement for the splicing of the first residue of the extein that follows the intein (which is Cys, Ser, or Thr). Other extein residues near the splicing junctions could modulate splicing efficiency even when they are not directly involved in catalysis. Mutual interdependence between this molecular parasite (intein) and its host protein (exteins) is not beneficial for intein spread but could be advantageous for intein survival during evolution. Elucidating extein-intein dependency has increasingly become important since inteins are recognized as useful biotechnological tools for protein ligation. We determined the structures of one of inteins with high splicing efficiency, the RadA intein from *Pyrococcus horikoshii*. The solution NMR structure and the crystal structures elucidated the structural basis for its high efficiency and directed our efforts of engineering that led to rational design of a functional minimized RadA intein. The crystal structure of the minimized RadA intein also revealed the precise interactions between N-extein and the intein. We systematically analyzed the effects at the -1 position of N-extein and were able to significantly improve the splicing efficiency of the less robust splicing variants by eliminating the unfavorable extein-intein interactions observed in the structure. This work provides an example of how unveiling structure-function relationships of inteins offers a promising way of improving their properties as better tools for protein engineering.

Keywords

protein engineering; protein splicing; extein; *trans*-splicing; segmental labeling

1. INTRODUCTION

Protein splicing is a unique self-catalytic reaction in which an internal protein sequence (intein) removes itself from the host protein, concomitantly ligating the two flanking sequences (exteins) with a peptide bond. Inteins have been identified in bacteria, archaea, and unicellular eukaryotes, with more than 550 putative inteins currently registered in the intein database (InBase). Protein splicing can also occur in *trans* via split inteins, which could be naturally occurring or artificially engineered. Protein *trans*-splicing (PTS) has become an important tool for protein ligation *in vivo* as well as *in vitro*, enabling useful biotechnological applications that include segmental isotopic labeling, site-specific modifications, and protein cyclization.” However, extensive applications of protein ligation by PTS have been often hindered by poor tolerance of splicing junction sequences and their ligation efficiency.’

A well understood, typical mechanism of protein splicing involves four concerted steps: 1) N-S(O) acyl shift, 2) *trans*-(thio)esterification, 3) Asn cyclization, and 4) S(O)-N acyl shift. Additionally, non-canonical inteins that do not follow this standard mechanism have also been reported.’ The first residue of inteins is usually a Ser or Cys that catalyzes N-S(O) migration at the N-terminal junction to initiate protein splicing. Class II and III inteins contain as the first residue Ala that cannot initiate the first step of N-S(O) shift. These inteins utilize different pathways to achieve protein splicing. Structural studies have suggested that specific roles for other residues in the proximity of the catalytic residues differ significantly between different inteins.” Variations in the catalytic mechanism seem to be better tolerated in protein splicing than for other enzymes, possibly because intein is a single turnover enzyme with fused substrates. On the other hand, neighboring extein residues near the splicing junctions that are not directly involved in the catalytic reaction could significantly influence the splicing efficiency of inteins.” This extein dependency could restrict practical applications of inteins because it makes it hard to predict whether protein-splicing reaction would result in productive splicing or off-pathway cleavages in foreign contexts.” Therefore, it is of considerable significance for the advancement of biotechnological applications of inteins to understand the atomic details of extein dependency. Previously, we identified RadA intein from *Pyrococcus horikoshii* (*Pho*) as an intein with highly efficient splicing activity, similar in its efficiency to the *cis*-splicing variant of naturally split DnaE intein from *Nostoc punctiforme* (*Npu*). The high splicing efficiency and its relatively small size make *Pho*RadA intein attractive for developing novel biotechnological tools. In order to elucidate the structural basis of the activity of *Pho*RadA intein we have determined its structures by NMR as well as X-ray crystallography. These structural results have led us to further redesign of this intein.

Results and Discussion

Inteins with robust splicing activity are indispensable for applications of PTS such as protein ligation. *Npu*DnaE intein was found to be one of the highly efficient split inteins widely used for PTS.’ For the applications of multi-fragment ligation with PTS, it is critical to have at least one more robust split intein that is orthogonal to naturally split *Npu*DnaE intein, although a few approaches have been reported to circumvent the cross-activity between split inteins.’ We determined the three-dimensional structure of *Pho*RadA intein in order to gain better understanding of its mode of activity and to provide a basis for creating new biotechnological tools.

The NMR solution structure of *Pho*RadA intein

As a first step to elucidate the NMR structure of the full-length *Pho*RadA intein, ¹³C, ¹⁵N-labeled *Pho*RadA intein was recombinantly produced in *E.coli*. This construct contains the

intein with a mutation of the first residue (C1A), a two-residue C-extein with a mutation of the first residue of C-extein (T+1A), and no N-extein (Fig. 2). The resonance assignments of *PhoRadA* intein with two C-extein residues were nearly completed by conventional triple-resonance experiments, with 94.8% of all the atoms except for the C-extein and a few residues (Fig. 1). Based on the assignments, the NOEs peaks were iteratively analyzed by the automated algorithm implemented in Cyana 3.0 in order to determine the structure. The final solution structure of *PhoRadA* intein contains predominantly β -sheets with a short helix between residues 21 and 31 and a short 3_{10} helix (Fig. 2). The solution NMR structures were well defined and highly consistent, with a backbone root-mean-square deviation (RMSD) of 0.61 Å for residues 1–172, excluding the two C-extein residues (Table 1). The overall fold of *PhoRadA* intein belongs to the HINT (hedgehog/intein) domain superfamily (Fig. 3). As expected from its primary structure, *PhoRadA* intein is a mini-intein lacking an endonuclease domain, the biological function of which, other than catalyzing protein splicing, is still largely unknown. The most highly disordered regions of the structure are at residues 121–133 and 173–174, where their resonance assignments are missing and only a few NOEs could be identified (Fig. 3). The heteronuclear NOE and ^{15}N relaxation data also indicate that the loop around residues 121–133 is disordered (Supplemental Fig. 1).

The crystal structures of *PhoRadA* intein

The construct of *PhoRadA* intein used for the NMR structure determination contains the C1A mutation in the intein and a two-residue C-extein with a mutation of Thr to Ala (T+1A), aimed at prevention of splicing and cleavage (Fig. 2). However, we could not define the conformation around the C-extein due to the lack of sufficient assignments for these residues. In order to find out if more information about this region could be gleaned from a crystal structure, we crystallized the construct of *PhoRadA* intein identical to the one used in the NMR study. Its crystal structure was determined at 1.75 Å resolution by the structure and density optimized molecular replacement implemented in Rosetta, using the NMR structures as starting models (Fig. 4a). The final model of *PhoRadA* intein contains two polypeptide chains, one HEPES buffer molecule, four glycerol molecules, and 124 water molecules in the asymmetric unit. In the polypeptide chain A, 172 amino acid residues have been modeled, although loop regions 13–14, 55–57 and 125–129 show only very weak electron density. The last two residues of the C-extein region could not be modeled due to lack of electron density. In the polypeptide chain B, 169 residues have been modeled, including the loop region 12–15, which shows weak electron density. However, the loop region 127–129 was not included in the model of this chain due to complete lack of electron density. In addition, density was missing or incomplete for a number of side-chains. The quality of the Ramachandran plot of the final model of *PhoRadA* intein was evaluated by PROCHECK which shows 93.3%, 5.4% and 1.3% of all residues falling into the most favored, additionally allowed, and generously allowed regions, respectively. Of the three residues in the latter category, two are located in the loops with poor electron density, but Arg99 of molecule A is located in very well defined density. Both the main chain and the side-chain of this residue make a number of hydrogen bonds, which are most likely responsible for its non-ideal torsion angles. The conformations of the two molecules in the asymmetric unit of the crystals of *PhoRadA* intein are very similar, with the RMSD of 0.254 Å for the 138 superimposed C α atoms (Fig. 4a). The most significant deviation is present in the loop region containing residues 11–17 (Fig. 4a). In chain B, the loop is longer than its counterpart in chain A, since it includes residues 11, 16, and 17, which are part of β -strands in chain A. In the loop region 125–129, chain B is more disordered and three residues (Arg127, Ile128, and Ile129) are not included in the final model due to complete lack of electron density.

Comparison between the NMR and crystal structures of *PhoRadA* intein—The NMR structure of *PhoRadA* intein contains all the 174 amino acid residues of the construct,

including the last two residues at the C-terminus that are part of the extein, which, however, were not well defined due to the lack of NOE constraints. This C-extein region was also not seen in the crystal structures of *PhoRadA* intein, confirming conformational flexibility of this region in both structures. The RMSD between the NMR structure and chain A of the crystal structure is 0.905 Å for 142 superimposed residues, whereas the RMSD between the NMR structure and chain B is 0.808 Å for 135 superimposed residues. The secondary structure elements are generally the same in both the crystal and NMR structures. Relatively large deviations mainly occur at the loop regions, and the most significant one is located in the loop 120–133 (Fig. 4b). The displacement between main chains in this region is around 6.8 Å. Not surprisingly, this loop is the most flexible region of the whole structure. Nevertheless, the overall deviation between the crystal and NMR-derived coordinates is on par with other protein structures investigated by these two techniques.

Minimizing *PhoRadA* intein

Both the NMR and crystal structures revealed a disordered loop that included residues 120–133. The loop is located between blocks B and F, where many inteins contain an endonuclease domain that is presumably important for horizontal gene transfer. The presence of disorder prompted us to test if this loop could be shortened without loss of function. We removed 10 residues between β 11 and β 12 and replaced Lys131 by Asn to better accommodate a turn (Fig. 2). The newly minimized *PhoRadA* intein (RadA_{min} intein) was tested with a model system using two B1 domains of immunoglobulin binding protein G (GB1) as exteins for *cis*-splicing (Fig. 5c). Expression of the *cis*-splicing precursor protein bearing the engineered RadA_{min} intein predominantly produced the *cis*-spliced product (H₆-GB1-GB1), confirming that RadA_{min} intein is highly functional (Fig. 5c). Both precursor bearing *PhoRadA* and RadA_{min} inteins produced only the spliced product of H₆-GB1-GB1 and the excised intein after 4 hours of the induction (Fig. 5c, lanes 1 and 3). The elution fractions from the Ni-NTA column contained only the *cis*-spliced product (H₆-GB1-GB1), but neither the precursor nor cleaved products were present in the elution fraction (Fig. 5c lanes 2 and 4). Thus, *cis*-splicing of the engineered RadA_{min} intein is very efficient and indistinguishable from that of *PhoRadA* intein, strongly suggesting that the removed disordered loop is a mere remnant from the endonuclease domain that was lost during evolution.

The crystal structure of RadA_{min} intein

The structure of the newly designed RadA_{min} intein was verified by X-ray crystallography. The protein containing both four N- and two C-extein residues, with the alanine-mutations at the first residues of the intein and C-extein (Fig. 2), was crystallized and diffraction data extending to 1.58 Å resolution were obtained. The better diffraction of RadA_{min} intein compared to the parent enzyme could be due to the removal of the disordered loop. The final model of RadA_{min} intein contains a single polypeptide chain and 251 water molecules in the asymmetric unit. All 168 amino acid residues present in the expression construct have been modeled and only residues 13, 82, and 83, which are all located in the loop regions, show relatively weak electron density. The electron density for the four extein residues at the N-terminus and for two additional residues at the C-terminus that are not part of the intein is of excellent quality and the conformation of the termini is not in doubt (Fig. 7). The Ramachandran plot of the final RadA_{min} model shows 92.4%, 6.2%, 0.7%, and 0.7% of all residues falling into the most favored, additionally allowed, generously allowed and disallowed regions, respectively. Not surprisingly, the only residue that falls into the disallowed region, Asp13, is located in a flexible loop with very weak electron density.

Comparison between *PhoRadA* and *RadA_{min}* inteins

Whereas the crystal structure of *PhoRadA* intein contains two molecules in the asymmetric unit, there is only one molecule in the asymmetric unit of *RadA_{min}* structure. The RMSD between *RadA_{min}* and *PhoRadA* intein chain A is 0.493 Å for the 130 superimposed residues, and the RMSD between *RadA_{min}* and *PhoRadA* intein chain B is 0.541 Å for the 133 superimposed residues. The loop region 120–133 of the native *PhoRadA* intein structure, which includes a short α -helix, is quite flexible as demonstrated by the weak electron density and comparatively high temperature factors. In the *RadA_{min}* intein structure, because of the ten-residue deletion, the remaining four residues became quite stable and all the residues have complete electron density, suggesting that the designed deletion did not disturb the rest of the structure. As this loop region is poorly ordered and dispensable for the splicing, it is an excellent candidate for introducing a split site for developing new split inteins.

Structure of the catalytic site

The crystal structure of *RadA_{min}* intein allows us to investigate at atomic resolution the catalytic site together with the N- and C-extein residues. Unlike the *cis* conformation of the N-scissile peptide bond reported in *MxeGyrA* intein, the scissile peptide bond between Met-1 and Ala1 of *RadA_{min}* intein is in the usual *trans* conformation. In the structure of the mutant of *SceVMA* intein precursor (X10SSS), the carbonyl oxygen and the first residue O γ are 3.1 Å apart. This distance is short enough for a nucleophilic attack between them, suggesting the formation of a five-membered ring thiozolidine intermediate during the N-S acyl shift. The existence of the suggested thiozolidine intermediate was recently confirmed with *SspDnaB* intein. In the *RadA* intein structures, the first residue (Cys1) is replaced by Ala, therefore it is not possible to measure the same distance as in the mutant of *SceVMA* intein. However, modeling of Cys1 based on the experimental structure of *RadA_{min}* intein indicates that the corresponding distance could be as short as 3.0 Å, similar to the mutant of *SceVMA* intein, indicating that *PhoRadA* intein may utilize the same tetrahedral intermediate during the N-S acyl shift (Fig. 6). The structure of *RadA_{min}* intein also indicates a very short distance between the carbonyl carbon of the N-scissile peptide and the side chain of the +1 residue that is responsible for the second step of *trans*-esterification (Table 3, Fig. 6). The close proximity of these key atoms suggests that *trans*-esterification could take place without large conformational changes after the first step N-S acyl rearrangement. This structural feature might account for the highly efficient *cis*-splicing of *PhoRadA* intein.

An aspartate residue in Block F of *PhoRadA* intein, located in the vicinity of the catalytic site, is highly conserved among inteins. That residue was found to play a pivotal role during the splicing in several inteins (Fig. 6). "The sidechain of Asp153 in Block F forms hydrogen bonds with the backbone of the +1 residue, as well as with the side chain of the last residue of the intein, i.e. Asn172, which catalyzes the cleavage of the branched intermediate after *trans*-esterification step. The interaction between Asn172 and Asp153 keeps the side chain of Asn172 away from the C-scissile bond, thereby preventing the third step of Asn cyclization. This structure implies that conformational rearrangement to induce Asn cyclization after the *trans*-esterification step is necessary. This particular interaction has not been observed with other intein structures that contained extein sequences, since in these structures the last Asn residues were replaced by Ala or Ser (Table 3). The orientation of Asn172 observed in the structure could be important for the understanding of the required structural rearrangements necessary at the individual steps during the protein splicing.

Extein-Intein interactions

Since our aim is to create more promiscuous inteins with respect to the splicing junctions, we are particularly interested in understanding the roles of extein sequences in protein splicing. In the structures of full-length *PhoRadA* intein determined by X-ray crystallography and NMR spectroscopy, it was not possible to define the precise conformation of the C-extein region. The results obtained by both techniques indicate the existence of conformational flexibility in this region of this protein. We have introduced a mutation from Thr to Ala at the +1 position of C-extein and also removed the N-extein sequence, which might have increased the flexibility of the extein region. A flexible C-extein structure was also reported in another intein structure. Although local mobility is suggested to be important for several steps of chemical reactions to take place, it is still not clear whether such flexibility is functionally essential or simply a result of shortening of the C-extein or removal of the N-extein. On the other hand, we could precisely define the conformation of the extein sequences in the case of RadA_{min} intein. Despite the increasing number of intein structures deposited in the Protein Data Bank (PDB), only a few of them have been determined with both the N- and C-extein residues modeled (Table 3). Among these intein structures, the distances between the carbonyl carbon atom of the N-terminal scissile bond and S γ (or O γ) atom of the first residue of C-extein are often quite long (8–9 Å), except for the mutant (X10SSS) of *SceVMA* intein (PDB code: 1JVA), where it is ~4 Å. The RadA_{min} intein structure is similar to the latter intein as the –1 and +1 residues are closer in the three-dimensional structure (“closed” conformation). The ψ angle of the first residue of RadA_{min} intein is –35°, which is significantly different from ψ angles of 159° for *SspDnaE* (1ZDE), 151° for *SspDnaB* (1MI8), and 127±7° for *SceVMA* (1JVA).” Two of the previously reported inteins with the open conformation contain a zinc ion near the active site. It has been demonstrated that zinc can inhibit protein splicing of inteins, suggesting that its coordination observed in the crystal structures could be responsible for the longer distance between the –1 and +1 residues, locking them in the open conformation. The crystal structures using several variants of inteins probably represent snapshots of the various conformations that may occur during protein splicing, which certainly requires structural rearrangements to accommodate a few different chemical reaction step at the same site.

Implication for N-extein dependency—Based on the N-extein confirmation seen in the structure of RadA_{min} intein, it is now possible to scrutinize the structural details of the extein-intein interactions of the *PhoRadA* intein. To elucidate the structural basis of N-extein dependency of *PhoRadA* intein, we compared *cis*-splicing of *PhoRadA* intein with 20 different amino acid types at the –1 position of N-extein using a model system (Fig. 8a). *PhoRadA* intein has not only highly efficient splicing activity but also accepts many sidechains at the –1 position without significant loss of function (Fig. 8b and Supplemental Fig. 2), making this intein very attractive for biotechnological applications. This comparison revealed that proline, amino acids with negative charges (Glu and Asp), β -branched amino acids (Val, Ile, and Thr), and amino acids with smaller sidechains (Gly and Ala) have lower *cis*-splicing efficiency. It is not surprising that proline at the –1 position has the lowest splicing activity, because the scissile peptide is chemically different from other peptide bonds as it lacks H^N hydrogen atom. In the crystal structure of RadA_{min} intein, the side chain of the –1 residue makes contacts with Glu71, Val73, and Val151 (Fig. 7). The native N-extein of *PhoRadA* intein has Lys at the –1 position and its positively charged side chain is making a charge-charge interaction with the negatively charged Glu71. Whereas all other residues with longer sidechains (van der Waals volume of >70 Å³) retained good splicing activity, residues with smaller van der Waals volumes, such as Ala (67 Å³) and Gly (48 Å³), slightly lowered the splicing efficiency. This suggests that van der Waals contacts are important for the proper alignment of the active site atoms to direct the reaction towards the productive splicing pathway. It can be postulated that β -branched amino acids (Val, Ile, and

Thr) disturb the precise arrangement of the active site groups due to their bulky side-chains, lowering the productive *cis*-splicing. It is noteworthy that β -branched aminoacids are also similarly unfavorable for *trans*-esterification in native chemical ligation. It might be of interest to investigate whether the effect of β -branched residues for both reactions is due to a similar mechanism.

Engineering a promiscuous intein at the N-junction—We hypothesized that the negative charge of Glu71 could be the cause for lowering splicing efficiency of the two variants bearing Glu or Asp at the -1 position (the E-1 and D-1 variants), presumably due to unfavorable negative charge interactions. To test our hypothesis, we introduced a mutation of E71T in *PhoRadA* intein to remove the negative charge. The mutant E71T indeed significantly improved *cis*-splicing of the E-1 variant from about 40% to >90% by lowering the amount of unspliced precursor, as well as by suppressing the cleavage at the C-terminal junction (Fig. 8c, lanes 2 and 4). Moreover, the E71T mutant of *PhoRadA* intein also retains high splicing activity with the native -1 residue of Lys (Fig. 8c, lane 5). This observation confirms that the negative charge of Glu71 is responsible for the lower efficiency of the E-1 variant. In the case of the D-1 variant, the *cis*-splicing was not improved at all, despite the same negative charge of the side chain (Fig. 8c, lanes 1 and 3). Unlike the E-1 variant, the D-1 variant resulted in mostly N-cleavage, as can be seen in the elution fraction containing significant amounts of N-cleaved product of H₆-GB1 (Fig. 8c, lanes 1 and 3). This result suggests that the lower efficiency of the D-1 variant is not due to the same mechanism as the E-1 variant. A similar N-cleavage with Asp at the -1 position was reported for *MxeGyrA*, *SspDnaE*, and *SceVMA* inteins.” It is known that an aspartic acid in a polypeptide chain could hydrolyze at its C-terminal side under acidic conditions due to a nucleophilic attack of the side-chain carboxyl group, depending on peptide sequences. The N-cleavage of the D-1 variant is possibly due to a similar involvement of the aspartate side chain at the -1 position that cleaves the N-scissile peptide bond or the thioester after N-S acyl shift. The InBase lists dozens of inteins with Asp at the -1 position and it is likely that they utilize another structural mechanism to prevent the N-cleavage.

In summary, we have determined the solution and crystal structures of *PhoRadA* intein. One of the structures allowed us to investigate the intein-extein interactions for understanding the structural basis of the N-extein dependency. Inteins are often considered to be mere parasitic proteins inserted into host proteins, providing no evident benefit to the host organisms. From the perspective of parasites, strong mutual extein-intein interactions could be less beneficial for intein spreads as it might not function at different insertion points. The structures of *PhoRadA* inteins revealed the defined interactions between the -1 position of N-extein and the intein as well as conformational flexibilities in the extein residues near the splicing junction. This interaction between N-extein and *PhoRadA* intein can play an important role for the highly efficient protein splicing, as observed with the 20 variants at the -1 position. Starting with the structural results, we demonstrated that a mutation in the intein could significantly improve protein splicing efficiency of the less robust E-1 variant, suggesting that *PhoRadA* intein has acquired mutations after the insertion into the host protein, but retained the efficient splicing activity in the context of the insertion point of the host protein. Ancestral inteins that spread to different host proteins might have been less dependent on the splicing junction sequences except for the first residue of C-extein that needs to be Cys, Ser, or Thr. In the intein database, all 20 common residue types except for Trp have been reported at the -1 position of the putative inteins. Although not all inteins have been tested for protein splicing, it is likely that inteins could tolerate all naturally occurring amino acids at least at the -1 position, since Trp at the -1 position of *PhoRadA* intein could be introduced without significant loss of function. Our results suggest that it might be possible to create “super” inteins that are less extein-dependent and highly efficient in protein splicing. Such inteins may resemble an ancient intein that had spread in different host

proteins and in various insertion sites. Further atomic-detail structural information for various inteins would be essential for developing promiscuous and robust inteins for advancing protein ligation technology with protein splicing.

Materials and methods

Protein production of ^{13}C , ^{15}N labeled *PhoRadA*

Cloning, expression, and purification of *PhoRadA* intein for protein crystallization and X-ray structure determination have been described previously. The same construct of *PhoRadA* intein was also used for preparation of the stable isotope labeled samples for NMR studies. *PhoRadA* intein was expressed in *Escherichia coli* ER2566 strain where the plasmid (pKRRS15) bearing *PhoRadA* intein was co-transformed with pLysSRARE plasmid to suppress the leaky protein expression and to compensate for rare codons (Novagen). The cells were initially grown in LB-medium containing 0.05% (w/v) D-glucose, 100 $\mu\text{g/mL}$ ampicillin, and 5 $\mu\text{g/mL}$ chloramphenicol until the optical density reached $\text{OD}_{600} \approx 0.5$. The cells were collected by spinning down at 900g at 25 °C for 10 minutes and resuspended with M9 medium supplied with 1.3 g/L $^{15}\text{NH}_4\text{Cl}$ and 2 g/L 20% or 100% ^{13}C -D-glucose as sole nitrogen and carbon sources and induced with a final concentration of 1 mM isopropyl- β -D-1-thiogalactopyranoside (IPTG). $^{13}\text{C}/^{15}\text{N}$ labeled *PhoRadA* intein was purified as previously. The buffer was exchanged to 20 mM sodium phosphate buffer pH 6.0 and protein was concentrated to 0.4 mM in 250 μl . The sample was transferred to a SHIGEMI micro cell tube for NMR measurements.

Cloning, expression, and purification of *RadA_{min}* intein

The *cis*-splicing precursor bearing *RadA_{min}* intein was constructed using the full-length *cis*-splicing *PhoRadA* intein vector (pHYDuet183) as a template. Two oligonucleotides, HK640: 5'-GAT GTT GTT GGT GTT AGG AAT GGA GAA CTT GAG TTC CAT GAG GTT TC and HK641: 5'-GAA ACC TCA TGG AAC TCA AGT TCT CCA TTC CTA ACA CCA ACA ACA TC were used to remove residues 121–130 of *PhoRadA* intein by PCR, resulting in pDPDuet04.

For structural analysis, inactive *RadA_{min}* intein was produced as yeast Smt3 fusion protein with a mutation of Cys1 to Ala. In addition, the first residue of C-terminal extein (Thr+1) was replaced by Ala to inactivate the protein splicing completely (Fig. 2). The gene was amplified from pDPDuet04 with two oligonucleotides, HK409: 5'-GCA TAT GGC CTT TGC TAG GGA TAC C and HK390: 5'-GCA TAT GGC CTT TGC TAG GGA TAC C. The PCR product was digested with *NdeI* and *HindIII* and cloned into pHYRSF53LA to product a fusion protein, resulting in plasmid pCARSF15. The DNA sequence of pCARSF15 had unexpected insertions of “CAG” at the front of *NdeI* site, presumably because the 5' end was not digested by *NdeI* but directly ligated into the digested vector, possibly due to exonuclease contamination during the digestion of the vector. We decided to retain the insertion because it introduced only an insertion of Gln. The final construct contains a four-residue sequence SQHM as N-extein at the front of *PhoRadA* intein.

E. coli ER2566 strain was transformed with pCARSF15 and grown at 37 °C in LB medium supplied with 25 $\mu\text{g/mL}$ kanamycin. The cells were grown until the optical density reached $\text{OD}_{600} \approx 0.6$, followed by induction with a final concentration of 1 mM IPTG. After 3 hours of the induction, the cells were harvested by centrifugation at 6700g at 4 °C for 10 minutes. The cell pellet was resuspended in buffer A (50 mM sodium-phosphate buffer, 300 mM NaCl, pH 8.0) and subsequently flash-frozen in liquid nitrogen for further purification.

The cell suspension was thawed and lysed by ultrasonification. The cell lysate was cleared by centrifugation at 42000g at 4 °C for 45 minutes. The supernatant was loaded on a 5 mL

HisTrap FF column (GE Healthcare). The column was further washed with buffer A and eluted by applying a linear gradient of buffer B (50 mM sodium-phosphate, 300 mM NaCl, pH 8.0, 250 mM imidazole). Fractions containing RadA_{min} intein fusion protein were dialyzed against PBS buffer. The dialyzed protein was digested with His-tagged yeast Ubiquitin-like-specific protease 1 (Ulp1) at 25 °C for 5 hours and reloaded on a 5 mL HisTrap FF column (GE Healthcare) to remove digested His-tagged Smt3 tag and Ulp1. RadA_{min} intein collected from the flow through fractions was concentrated by a centrifugal concentrator (Amicon Ultra 3,000 MWCO). The buffer was replaced by Milli-Q grade water.

NMR spectroscopy and resonance assignment

All spectra used for NMR structure determination were recorded at 308 K on either Varian Innova 600 or 800 MHz spectrometers equipped with triple resonance cryogenic probe heads. The protein was concentrated to 0.15–0.4 mM in 20 mM phosphate buffer pH 6.0. For sequence specific resonance assignments, series of 2D and 3D spectra were recorded including ¹⁵N-HSQC, ¹³C-HSQC, HNCA, intraHNCA, HN(CO)CA, HNCO, HN(CA)CO, HNCACB, CBCA(CO)NH, CCC(CO)NH, HBHA(CO)NH, HBNH, ¹⁵N-edited TOCSY with a mixing time of 50 msec, H(C)CH-TOCSY, and H(C)CH-COSY. Aromatic side-chain was assigned using (HB)CB(CG)CDHD, (HB)CB(CGCD)CEHE, and aromatic CT-¹³C-HSQC spectra. For the structure calculation, an ¹⁵N-edited NOESY with a mixing time of 80 msec and a sensitive enhanced ¹³C-edited NOESY with mixing time of 100 msec were recorded at ¹H frequency of 800 MHz. ¹⁵N relaxation times were recorded at ¹H frequency of 600 MHz. T₁(¹⁵N) relaxation rates were determined using following T₁ delays: 10, 30, 50, 70, 90, 110, 130, and 150 msec. For T₂(¹⁵N) relaxation rates following T₂ delays were used: 10, 30, 50, 70, 90, and 110 ms with CPMG refocusing interval of 1.3 msec. Heteronuclear ¹⁵N{¹H}-NOEs were determined by recording ¹⁵N-HSQC spectra with and without 4.0 seconds of ¹H saturation. Peak volumes were used to determine relaxation rates by fitting and analysis in CcpNmr analysis.

All spectra were processed by program NMRPipe and specific resonance assignments were performed using the software SPARKY. The sequence specific resonance assignment of the 174-residue *PhoRadA* intein with two C-extein residues was completed by conventional triple resonance methods using ¹³C/¹⁵N doubly labeled samples with a final 94% assignment completion of H^N, N, C α , H α , and C' atoms. The aliphatic carbon and proton side-chain resonances are 97% assigned. Proline residues conformation was checked and they are all in *trans* conformation based upon C β and C γ chemical shift. Residues 121–122, 125–127, 159–160, and 173–174 could not be observed in the ¹⁵N-HSQC spectra (Fig. 1) and resonance assignments of these residues are incomplete.

NMR structure determination

NOESY peak lists were generated with automatic pick picking and integrated using CcpNmr analysis. Backbone derived angle constraints from TALOS+ were used as inputs for the structure calculation. A chemical shift list, dihedral angle constraints, and unassigned NOESY peaks list were used as inputs for the structure determination using the automatic NOESY assignment approach with CYANA 3.0. The structure determination in CYANA was based on the automatic NOE assignment and in the final cycle of structure determination 3515 NOEs were used (Table 1). The 20 final solution structures of *PhoRadA* intein with the lowest CYANA target function were further energy refined using AMBER. Molecular dynamic simulation was performed in explicit water with a 5 Å water shell surrounding the protein molecule. The structural statistic and inputs for NMR structure are summarized in Table 1. Structure validation was performed using CING server (<http://nmr.cmbi.ru.nl/cing/Home.html>) and PROCHECKNMR.,

Protein crystallization

Crystallization of the *PhoRadA* intein has been described previously. *RadA_{min}* intein was crystallized using solution of 31 mg/mL concentration. The crystallization conditions were screened using solution from Index HT screen (Hampton Research) by sitting drop vapor diffusion in Innovadyne SD2 96 well plates with drops of 100 nL protein solution and 100 nL reservoir solution at 293 K. A single crystal was picked from a drop containing 1.6 M tri-sodium citrate and cryo-protected with Paratone-N before vitrification.

Diffraction data collection and processing

Diffraction data for the crystal of *PhoRadA* intein were collected in a single pass on beamline ID23-1 at ESRF/Grenoble and were subsequently indexed, integrated, and scaled to 1.75 Å resolution using the program XDS. The crystal belongs to the space group $P2_12_12_1$ with two molecules in the asymmetric unit. Data processing statistics are listed in Table 2. The dataset was essentially complete, with high redundancy. However, the values of scaling R factors are rather high, indicating that data quality might not be optimal, but still within an acceptable range. The estimated Matthews coefficient was $1.97 \text{ Å}^3\text{Da}^{-1}$, corresponding to 37% solvent content. The structure was solved by molecular replacement using the NMR-derived coordinates as a starting model, with the weak solutions improved with the help of the program Rosetta. Further refinement was performed with Refmac5 and Phenix, using all data between 33.7 and 1.75 Å, after setting aside 3% of randomly selected reflections (~1000 total) for calculation of R_{free} . Non-crystallographic symmetry (NCS) restraints were initially applied to the two molecules, but these restraints were removed before the final refinement cycles. Isotropic individual temperature factors were refined, with the TLS parameters (one for each molecule) added in the final stages of refinement. After several further rounds of restrained and TLS refinement and manual correction using COOT, the structural model was finally refined to an R-factor of 19.0% and R_{free} of 23.4%. Data processing and structure refinement statistics are shown in Table 2.

Diffraction data for the crystal of *RadA_{min}* intein were collected on beamline ID14-1 at ESRF/Grenoble. They were indexed, integrated, and scaled to 1.58 Å resolution using the program package HKL2000. A comparatively high value of R_{sym} is due to the presence of a second crystal lattice that was partially superimposed on the main one, interfering with proper integration of a fraction of reflections. This crystal is also in space group $P2_12_12_1$ but with only one molecule in the asymmetric unit. The structure was solved by molecular replacement with the program PHASER using the coordinates of molecule A of *PhoRadA* intein as a search model. The refinement of *RadA_{min}* used a protocol similar to the one utilized for *PhoRadA* intein, and the final R-factor and R_{free} are 16.7% and 19.5%, respectively. Data processing and structure refinement statistics are shown in Table 2.

Mutagenesis of *PhoRadA* intein

Plasmids for the 20 variants of *cis*-splicing *PhoRadA* intein precursor bearing the mutations at the –1 position were constructed from pHYDuet183 by cassette mutagenesis using *Bse*RI and *Hind*III restriction sites or QuikChange Protocol (STRATAGENE) with synthetic oligonucleotides containing the mutations (the details will be reported elsewhere). The E71T mutation was introduced in the *PhoRadA* intein of the *cis*-splicing precursors containing Asp (pSCFDuet22), Glu (pSCFDuet23), and Lys (pSCFDuet4) at the –1 position. The mutation of E71T was introduced into these plasmids by QuikChange Protocol with the two oligonucleotides, HK976:5'-CAT ACA TCT ATC GCA CGA AGG TTG AGA AGC and HK977:5'-GCTTCTCAACCTTCGTGCGATAGATGTATG.

Evaluation of *cis*-splicing—*Cis*-splicing efficiency was quantified by transforming the plasmids for the 20 variants at the –1 position into *E. coli* 2566 cells. The cells were grown

in 5 mL LB-media supplied with 25 µg/mL kanamycin at 37 °C until OD₆₀₀ reached 0.5–0.6. The expression of the *cis*-splicing precursors was induced by addition of IPTG at a final concentration of 1 mM. After four-hour induction with IPTG, the cells were harvested by centrifugation at 4 °C for further purification. The cells were lysed by re-suspending the pellets in 100 µL B-PER® Bacterial Protein Extraction Reagent (Thermo Scientific) and incubation at 25 °C for 10 minutes. The cell suspension was cleared by centrifugation at 14000g for 5 minutes and the supernatant was loaded on a Ni-NTA spin column (Qiagen). Bound proteins were eluted from the column by applying a 50 mM sodium phosphate (pH 8.0), 300 mM NaCl, and 250 mM imidazole buffer. The elution fractions were analysed on 18% SDS-PAGE (Supplemental Fig. 2). The gels were stained with PhastGel™ Blue R (GE Healthcare) and protein band intensity was determined using the software ImageJ 1.45. The intensities of protein bands were used to quantify the protein splicing efficiency, given that staining dye equally binds to all proteins. The errors for the splicing efficiency were estimated from three independent experiments.

Accession numbers—The resonance assignment is deposited in the BioMagResBank (accession number:18320). Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 4E2T for *PhoRadA* intein, 4E2U for RadA_{min} intein, and 2LQM for the NMR structures of *PhoRadA* intein.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank C. Albert and S. Ferkau for technical help in the protein and plasmid preparations. J.S.O. acknowledges the National Graduate School in Informational and Structural Biology for financial support. The WeNMR project (European FP7 e-Infrastructure grant, contract no. 261572, www.wenmr.eu), supported by the national GRID Initiatives of Belgium, Italy, Germany, the Netherlands (via the Dutch BiG Grid project), Portugal, UK, South Africa, Taiwan and the Latin America GRID infrastructure via the Gisela project, is acknowledged for the use of web portals, computing, and storage facilities. This work was supported in part by the Academy of Finland (1131413, 137995) and Biocenter Finland (for H.I. the crystallization, and NMR facilities at the Institute of Biotechnology), and in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

Abbreviations

HSQC	heteronuclear single-quantum correlation
IPTG	isopropyl-β-D-1-thiogalactopyranoside
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
Npu	<i>Nostoc punctiforme</i>
Pho	<i>Pyrococcus horikoshii</i>
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
RMSD	root mean square deviation
PTS	protein <i>trans</i> -splicing
Ulp1	Ubiquitin-like-specific protease 1

References

1. Paulus H. Protein splicing and related forms of protein autoprocessing. *Annu. Rev. Biochem.* 2000; 69:447–496. [PubMed: 10966466]
2. Perler FB. InBase: the Intein Database. *Nucleic Acids Res.* 2002; 30:383–384. [PubMed: 11752343]
3. Pietrovski S. Intein spread and extinction in evolution. *Trends Genet.* 2001; 17:465–472. [PubMed: 11485819]
4. Skrisovska L, Schubert M, Allain FHT. Recent advances in segmental isotope labeling of proteins: NMR applications to large proteins and glycoproteins. *J. Biomol. NMR.* 2010; 46:51–65. [PubMed: 19690964]
5. Saleh L, Perler FB. Protein splicing in cis and in trans. *Chem. Rec.* 2006; 6:183–193. [PubMed: 16900466]
6. Evans TC, Martin D, Kolly R, Panne D, Sun L, Ghosh I, Chen L, Benner J, Liu XQ, Xu MQ. Protein trans-splicing and cyclization by a naturally split intein from the *dnaE* gene of *Synechocystis* species PCC6803. *J. Biol. Chem.* 2000; 275:9091–9094. [PubMed: 10734038]
7. Iwai H, Lingel A, Plückthun A. Cyclic green fluorescent protein produced in vivo using an artificially split PI-PfuI intein from *Pyrococcus furiosus*. *J. Biol. Chem.* 2001; 276:16548–16554. [PubMed: 11278952]
8. Iwai H, Züger S, Jin J, Tam P. Highly efficient protein trans-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. *FEBS Lett.* 2006; 580:1853–1858. [PubMed: 16516207]
9. Lockless SW, Muir TW. Traceless protein splicing utilizing evolved split inteins. *Proc. Natl. Acad. Sci. U S A.* 2009; 106:10999–11004. [PubMed: 19541616]
10. Noren CJ, Wang J, Perler FB. Dissecting the Chemistry of Protein Splicing and Its Applications. *Angew. Chem. Int. Ed. Engl.* 2000; 39:450–466. [PubMed: 10671234]
11. Southworth MW, Benner J, Perler FB. An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile. *EMBO J.* 2000; 19:5019–5026. [PubMed: 10990465]
12. Tori K, Dassa B, Johnson MA, Southworth MW, Brace LE, Ishino Y, Pietrovski S, Perler FB. Splicing of the mycobacteriophage Bethlehem DnaB intein: identification of a new mechanistic class of inteins that contain an obligate block F nucleophile. *J. Biol. Chem.* 2010; 285:2515–2526. [PubMed: 19940146]
13. Klabunde T, Sharma S, Telenti A, Jacobs WR, Sacchettini JC. Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing. *Nat. Struct. Biol.* 1998; 5:31–36. [PubMed: 9437427]
14. Ding Y, Xu M, Ghosh I, Chen X, Ferrandon S, Lesage G, Rao Z. Crystal Structure of a Mini-intein Reveals a Conserved Catalytic Module Involved in Side Chain Cyclization of Asparagine during Protein Splicing. *J. Biol. Chem.* 2003; 278:39133–39142. [PubMed: 12878593]
15. Sun P, Ye S, Ferrandon S, Evans TC, Xu MQ, Rao Z. Crystal structures of an intein from the split *dnaE* gene of *Synechocystis* sp. PCC6803 reveal the catalytic model without the penultimate histidine and the mechanism of zinc ion inhibition of protein splicing. *J. Mol. Biol.* 2005; 353:1093–1105. [PubMed: 16219320]
16. Johnson MA, Southworth MW, Herrmann T, Brace L, Perler FB, Wüthrich K. NMR structure of a KlbA intein precursor from *Methanococcus jannaschii*. *Protein Science.* 2007; 16:1316–1328. [PubMed: 17586768]
17. Amitai G, Callahan BP, Stanger MJ, Belfort G, Belfort M. Modulation of intein activity by its neighboring extein substrates. *Proc. Natl. Acad. Sci. U S A.* 2009; 106:11005–11010. [PubMed: 19541659]
18. Nogami S, Satow Y, Ohya Y, Anraku Y. Probing novel elements for protein splicing in the yeast Vma1 protozyme: a study of replacement mutagenesis and intragenic suppression. *Genetics.* 1997; 147:73–85. [PubMed: 9286669]
19. Aranko AS, Züger S, Buchinger E, Iwai H. In vivo and in vitro protein ligation by naturally occurring and engineered split DnaE inteins. *PLoS ONE.* 2009; 4:e5185.
20. Chong S, Shao Y, Paulus H, Benner J, Perler FB, Xu MQ. Protein splicing involving the *Saccharomyces cerevisiae* VMA intein. The steps in the splicing pathway, side reactions leading to

- protein cleavage, and establishment of an in vitro splicing system. *J. Biol. Chem.* 1996; 271:22159–22168. [PubMed: 8703028]
21. Ellilä S, Jurvansuu JM, Iwai H. Evaluation and comparison of protein splicing by exogenous inteins with foreign exteins in *Escherichia coli*. *FEBS Lett.* 2011; 585:3471–3477. [PubMed: 22001202]
 22. Busche AEL, Aranko AS, Talebzadeh-Farooji M, Bernhard F, Dötsch V, Iwai H. Segmental isotopic labeling of a central domain in a multidomain protein by protein trans-splicing using only one robust DnaE intein. *Angew. Chem. Int. Ed. Engl.* 2009; 48:6128–6131. [PubMed: 19591176]
 23. Shah NH, Vila-Perelló M, Muir TW. Kinetic control of one-pot trans-splicing reactions by using a wild-type and designed split intein. *Angew. Chem. Int. Ed. Engl.* 2011; 50:6511–6515. [PubMed: 21656885]
 24. Güntert P. Automated structure determination from NMR spectra. *Eur. Biophys. J.* 2009; 38:129–143. [PubMed: 18807026]
 25. Perler FB. Protein splicing of inteins and hedgehog autoproteolysis: structure, function, and evolution. *Cell.* 1998; 92:1–4. [PubMed: 9489693]
 26. Lyskowski A, Oeemig JS, Jaakkonen A, Rommi K, DiMaio F, Zhou D, Kajander T, Baker D, Wlodawer A, Goldman A, Iwai H. Cloning, expression, purification, crystallization and preliminary X-ray diffraction data of the *Pyrococcus horikoshii* RadA intein. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 2011; 67:623–626.
 27. DiMaio F, Terwilliger TC, Read RJ, Wlodawer A, Oberdorfer G, Wagner U, Valkov E, Alon A, Fass D, Axelrod HL, Debanu D, Vorobiev SM, Iwai H, Pokkuluri PR, Baker D. Increasing the Radius of Convergence of Molecular Replacement by Density and Energy Guided Protein Structure Optimization. *Nature.* 2011; 473:540–543. [PubMed: 21532589]
 28. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 1993; 26:283–291.
 29. Jaudzems K, Geralt M, Serrano P, Mohanty B, Horst R, Pedrini B, Elsliger MA, Wilson IA, Wüthrich K. NMR structure of the protein NP_247299.1: comparison with the crystal structure. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 2010; 66:1367–1380.
 30. Oeemig JS, Aranko AS, Djupsjöbacka J, Heinämäki K, Iwai 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.* 2009; 583:1451–1456. [PubMed: 19344715]
 31. Mizutani R, Nogami S, Kawasaki M, Ohya Y, Anraku Y, Satow Y. Protein-splicing Reaction via a Thiazolidine Intermediate: Crystal Structure of the VMA1-derived Endonuclease Bearing the N and C-terminal Propeptides. *J. Mol. Biol.* 2002; 316:919–929. [PubMed: 11884132]
 32. Ludwig C, Schwarzer D, Mootz HD. Interaction studies and alanine scanning analysis of a semi-synthetic split intein reveal thiazoline ring formation from an intermediate of the protein splicing reaction. *J. Biol. Chem.* 2008; 283:25264–25272. [PubMed: 18625708]
 33. Van Roey P, Pereira B, Li Z, Hiraga K, Belfort M, Derbyshire V. Crystallographic and mutational studies of *Mycobacterium tuberculosis* recA mini-inteins suggest a pivotal role for a highly conserved aspartate residue. *J. Mol. Biol.* 2007; 367:162–173. [PubMed: 17254599]
 34. Creighton, TE. *Proteins: Structures and Molecular properties*. 2nd edition. W. H. Freeman and Company; New York: 1993.
 35. Hackeng TM, Griffin JH, Dawson PE. Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. *Proc. Natl. Acad. Sci. USA.* 1999; 96:10068–10073. [PubMed: 10468563]
 36. Southworth MW, Amaya K, Evans TC, Xu MQ, Perler FB. Purification of proteins fused to either the amino or carboxy terminus of the *Mycobacterium xenopi* gyrase A intein. *BioTechniques.* 1999; 27:110–114. 116, 118–120. [PubMed: 10407673]
 37. Chong S, Montello GE, Zhang A, Cantor EJ, Liao W, Xu MQ, Benner J. Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. *Nucleic Acids Res.* 1998; 26:5109–5115. [PubMed: 9801307]
 38. Inglis AS. Cleavage at aspartic acid. *Methods Enzymol.* 1983; 91:324–332. [PubMed: 6304451]
 39. Muona M, Aranko AS, Iwai H. Segmental isotopic labelling of a multidomain protein by protein ligation by protein trans-splicing. *Chembiochem.* 2008; 9:2958–2961. [PubMed: 19031436]

40. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation. *Biochemistry*. 1994; 33:5984–6003. [PubMed: 7514039]
41. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*. 2005; 59:687–696. [PubMed: 15815974]
42. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*. 1995; 6:277–293. [PubMed: 8520220]
43. Schubert M, Labudde D, Oschkinat H, Schmieder P. A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ¹³C chemical shift statistics. *J. Biomol. NMR*. 2002; 24:149–154. [PubMed: 12495031]
44. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR*. 2009; 44:213–223. [PubMed: 19548092]
45. Herrmann T, Güntert P, Wüthrich K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* 2002; 319:209–227. [PubMed: 12051947]
46. Bertini I, Case DA, Ferella L, Giachetti A, Rosato A. A Grid-enabled web portal for NMR structure refinement with AMBER. *Bioinformatics*. 2011; 27:2384–2390. [PubMed: 21757462]
47. Doreleijers JF, Vranken WF, Schulte CS, Markley JL, Ulrich EL, Vriend G, Vuister GW. NRG-CING: integrated validation reports of remediated experimental biomolecular NMR data and coordinates in wwPDB. *Nucleic Acid Res.* 2012; 40:D519–524.
48. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*. 1996; 8:477–486. [PubMed: 9008363]
49. Kabsch W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* 1993; 26:795–800.
50. Matthews BW. Solvent content of protein crystals. *J. Mol. Biol.* 1968; 33:491–497. [PubMed: 5700707]
51. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* 1997; D53:240–255.
52. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr.* 2002; D58:1948–1954.
53. Brünger AT. The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*. 1992; 355:472–475. [PubMed: 18481394]
54. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr.* 2004; D60:2126–2132.
55. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 1997; 276:307–326.
56. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J. Appl. Cryst.* 2007; 40:658–674. [PubMed: 19461840]
57. Poland BW, Xu MQ, Quioco FA. Structural insights into the protein splicing mechanism of PI-SceI. *J. Biol. Chem.* 2000; 275:16408–16413. [PubMed: 10828056]

Highlights

- Structure-function relationship of a highly efficient intein for protein design.
- The structures of *PhoRadA* intein unveiled the extein-intein interactions.
- The extein dependency of *PhoRadA* intein was investigated at the −1 position.
- We improved the splicing efficiency of the less robust variant by rational design.

Figure 2.

Primary structures of *PhoRadA* intein and minimized *PhoRadA* intein (*RadA_{min}*) used for the structure determination. Secondary structures based on the determined NMR model are indicated above the protein sequence. Conserved intein sequence regions of blocks A, B, G, and H are marked. The mutated active site residues are in bold. Highly conserved His and Thr in block B and Asp in block F in the active site are in bold and italic (see the text).

Figure 3. The NMR solution structure of *PhoRadA* intein. a) The 20 superimposed NMR structures. b) Ribbon presentation of the NMR structure labeled with the secondary structures. Residues that were not assigned are shown in stick model and colored in orange. N and C-termini are indicated.

Figure 4.

The crystal structure of *PhoRadA* intein. a) Superposition of the two *PhoRadA* intein models in the asymmetric unit. The loop regions containing residues 11–17 and 120–133 are shown in stick representation with numbering for the beginnings and the ends. b) The mean NMR model in ribbon representation (grey) is superimposed with the two models of the *PhoRadA* intein crystal structures (blue). The N- and C-termini are indicated. Backbone atoms of residues 1–126 and 130–172 are used for the superposition of the structures.

Figure 5. Minimized *PhoRadA* intein (RadA_{min}) structure. a) Structure of RadA_{min} intein (red). b) A stereoview of the superposition of RadA_{min} intein structure (red), crystal structures of *PhoRadA* intein (blue), and the mean NMR structure of *PhoRadA* intein (grey). The N- and C-termini are indicated. c) SDS-PAGE analysis of *cis*-splicing of *PhoRadA* intein and RadA_{min} intein. M indicates protein marker. Lane 1, the total cell lysate of the cells expressing *cis*-splicing precursor with *PhoRadA* intein; lane 2, elution fraction from the Ni-NTA column loaded with the cell lysate expressing the *PhoRadA* intein precursor; lane 3, The total cell lysate of the cells expressing the precursor bearing RadA_{min} ; lane 4, elution fraction from the Ni-NTA column loaded with the cell lysate expressing the precursor containing RadA_{min} intein.

Figure 6.

The structure of the splicing site of RadA_{min} intein. A stereoview of residues located in the vicinity of the splicing site. Distances between atoms between the conserved residues are indicated with dashed lines. N, O, and S atoms are indicated with blue, red, and yellow spheres, respectively. The residues from the N- and C-exteins are colored dark grey. Residue numbers are labeled with the locations in the conserved intein motif (blocks A, B, F, and G). Only backbone atoms of Phe2 are shown.

Figure 7.

Electron density map of RadA_{min} intein and the extein residues. a) A stereoview of the map of the extein residues (blue mesh) and the first and last residues of RadA_{min} intein (red mesh). b) A stereoview of the map showing the residues within 4 Å of the last residue of N-extein (M-1). The electron density map is contoured at 1σ .

Figure 8.

N-extein dependency of *PhoRadA* intein. a) A model system used for evaluating *cis*-splicing efficiency. X stands for 20 amino acids at the -1 position tested for the comparison. The precursor protein could result in either on-pathway *cis*-splicing, off-pathway N-cleavage, or off-pathway C-cleavage. b) A plot of *cis*-splicing efficiency [%] against 20 amino acid types at the -1 position. The error bars were obtained from the analysis of three independent experiments. c) Comparison of protein splicing efficiency of *PhoRadA* intein and the E71T variant. The purified fractions from the cells expressing *cis*-splicing precursors by Ni-NTA column were analyzed by SDS-PAGE. M indicates protein markers. Lane 1, D-1 variant of *PhoRadA* intein; lane 2, E-1 variant; lane 3, D-1 variant with E71T mutation in the *PhoRadA* intein; lane 4, E-1 variant with the E71T mutation; lane 5, K-1 variant with the E71T mutation.

Table 1Statistics of the NMR structure calculation of *PhoRadA* intern.

Quantity	Value
NOE upper distance limits	3515
Short range NOE ($ i-j \leq 1$)	1654
Medium range NOE ($1 < i-j < 5$)	425
Long range NOE ($ i-j \geq 5$)	1436
Dihedral angles ^a	278
Residual CYANA target function ($\text{\AA}^2$)	1.14 ± 0.08
Residual NOE violation	
Number $0.3 \text{ \AA}$	3 ± 2
Maximum ($\text{\AA}$)	0.67 ± 0.24
Residual dihedral angle violation	
Number 5°	6 ± 2
Maximum ($\text{\AA}$)	12.28 ± 2.96
Amber energies (kcal/mol)	
Total	-6518 ± 24.18
van der Waals	-1402 ± 15.8
Electrostatic	-11520 ± 330
RSMD from ideal geometry	
Bond length ($\text{\AA}$)	0.0229 ± 0.0001
Bond angles ($^\circ$)	2.120 ± 0.015
RMSD to mean coordinate	
Backbone 1-172 ($\text{\AA}$)	0.61 ± 0.10
Heavy atoms 1-172 ($\text{\AA}$)	1.09 ± 0.16
Ramachandran plot statistics ^b	
Most favored regions (%)	91.3
Additional allowed regions (%)	8.2
Generously allowed regions (%)	0.5
Disallowed regions (%)	0.0

^aTALOS+ derived angle predictions [44].^bDerived by PROCHECK-NMR [48].

Table 2

Crystallographic data collection and structure refinement

Data collection	RadA	RadA _{min}
Beamline	ESRF/ID23-1	ESRF/ID14-1
Space group	$P2_12_12_1$	$P2_12_12_1$
Molecules/a.u.	2	1
Unit cell a , b , c (Å);	58.1, 67.4, 82.9	46.9, 63.6, 66.6
$\alpha=\beta=\gamma$ (°)	90	90
Resolution (Å) *	33.7–1.75 (1.80–1.75)	46.0–1.58 (1.64–1.58)
$R_{\text{merge}}^{\dagger}$ (%)	7.0 (95.3)	8.6 (45.4)
No. of reflections (measured/unique)	234,686/33,297	238,001/27,745
$\langle I / \sigma I \rangle$	17.3 (2.5)	12.2 (2.4)
Completeness (%)	99.3 (99.8)	98.9 (89.6)
Redundancy	7.1 (7.2)	8.6 (7.9)
Refinement		
Resolution (Å)	33.7–1.75	46.0–1.58
No. of reflections (refinement/ R_{free})	33,296/999	27,680/887
$R / R_{\text{free}}^{\ddagger}$	0.190/0.234	0.167/0.195
No. atoms		
Protein	2788	1366
Ligands	39	0
Water	121	251
R.m.s. deviations from ideal		
Bond lengths (Å)	0.012	0.011
Bond angles (°)	1.65	1.4
PDB code	4E2T	4E2U

* The highest resolution shell is shown in parentheses.

$R_{\text{merge}}^{\dagger} = \sum_h \sum_i |I_i - \langle I \rangle| / \sum_h \sum_i I_i$, where I_i is the observed intensity of the i -th measurement of reflection h , and $\langle I \rangle$ is the average intensity of that reflection obtained from multiple observations.

$R^{\ddagger} = \sum \|F_o - |F_c|\| / \sum F_o$, where F_o and F_c are the observed and calculated structure factors, respectively, calculated for all data. R_{free} was defined in [53].

Table 3

Intein structures with N- and C-extein residues

Intein	PDB	Res. [Å]	WT. N-extein	WT. C-extein	Model N-extein	Model C-extein	Distance C'(-I)-S(O) γ (+I) (C'-C β) (Å)	Mutations and additional remarks
<i>SspDnaE</i> ¹⁴	1ZDE	1.95	FAEY	CFNK	FEKY/A	A/CFN	8.0 (7.8)	C1A, N158A, C+1 is coordinated by Zn-ion
<i>SspDnaB</i> ¹³	1MI8	2.0	RESG	SIWQ	SG/A	A/SI	8.0 (7.8)	C1A, N429A, endonuclease domain is removed
<i>SceVMA</i> ³¹	1JVA	2.1	IYVG	CGER	VG/S	S/SGER	3.8 (4.1)	C1S, H79N, N454S, C+1S
<i>SceVMA</i> ⁵⁷	1EF0	2.1	IYVG	CGER	KLEG/A	A/CGER	9.1 (7.8)	C1A, N454A, C+1 is coordinated by Zn-ion
<i>MjaK1bA</i> ¹⁶	2JMZ	NMR	GHDG	CSGT	GHDG/A	A/SSGT	8.1 (9.3)	N168A, C+1S
RadA _{min}	4E2U	1.58	GSGK	TQLA	SQHM/A	N/AQ	-(3.7)	C1A, T+1A, residues 121-130 are removed