Received: 6 October 2023 | Revised: 20 May 2024 | Accepted: 21 May 2024

DOI: 10.1002/pro.5070

RESEARCH ARTICLE

Ligation of multiple protein domains using orthogonal inteins with non-native splice junctions

Alejandro Romero-Casañas[1] | **Andrea García-Lizarribar**[1] | **Jessica Castro**[1,2] | **Maria Vilanova**[1,2] | **Antoni Benito**[1,2] | **Marc Ribó**[1,2]

[1]Laboratori d'Enginyeria de Proteïnes, Departament de Biologia, Universitat de Girona, Girona, Spain

[2]Institut d'Investigació Biomèdica de Girona Josep Trueta (IdIBGi), Salt, Spain

Correspondence
Antoni Benito and Marc Ribó, Laboratori d'Enginyeria de Proteïnes, Departament de Biologia, Universitat de Girona, 17003 Girona, Spain.
Email: antoni.benito@udg.edu and marc.ribo@udg.edu

Funding information
Ministerio de Economia y Competitividad, Grant/Award Number: BIO2013-43517; Universitat de Girona, Grant/Award Number: MPCUdG2016-18; DGU, Generalitat de Catalunya, Grant/Award Number: 2017-SGR 1720

Review Editor: Aitziber L. Cortajarena

Abstract

Protein splicing is a self-catalyzed process in which an internal protein domain (the intein) is excised from its flanking sequences, linking them together with a canonical peptide bond. *Trans*-inteins are separated in two different precursor polypeptide chains that must assemble to catalytically self-excise and ligate the corresponding flanking exteins to join even when expressed separately either *in vitro* or *in vivo*. They are very interesting to construct full proteins from separate domains because their common small size favors chemical synthesis approaches. Therefore, *trans*-inteins have multiple applications such as protein modification and purification, structural characterization of protein domains or production of intein-based biosensors, among others. For many of these applications, when using more than one *trans*-intein, orthogonality between them is a critical issue to ensure the proper ligation of the exteins. Here, we confirm the orthogonality (lack of cross-reactivity) of four different *trans*- or split inteins, gp41-1, gp41-8, IMPDH-1 and NrdJ-1 both *in vivo* and *in vitro*, and built different constructs that allow for the sequential fusion of up to four protein fragments into one final spliced product. We have characterized the splicing efficiency of these constructs. All harbor non-native extein residues at the splice junction between the *trans*-intein and the neighboring exteins, except for the essential Ser + 1. Our results show that it is possible to ligate four different protein domains using inteins gp41-1, IMPDH-1 and NrdJ-1 with non-native extein residues to obtain a final four-domain spliced product with a not negligible yield that keeps its native sequence.

KEYWORDS
orthogonal inteins, protein engineering, protein splicing, sequential protein ligation

1 | INTRODUCTION

Protein splicing is a post-translational modification catalyzed by a part of the translated protein, termed intein, which catalytically self-excises and concomitantly ligates the flanking sequences, termed exteins. According to their structural organization there are three broad categories of inteins: (1) bi-functional *inteins* that contain a homing endonuclease domain inserted into the core of intein sequence (Chong et al., 1997; Chong et al., 1998), involved in intein gene horizontal transfer; (2) *mini-inteins*, smaller in size because they lack the homing endonuclease domain; and (3) *trans-splicing inteins* that have the intein domain split in two subunits. In this latter case, the N-intein (I_N) and C-intein (I_C) domains must associate to reconstitute the active intein before splicing the surrounding extein fragments (N-extein and C-extein).

Inteins share a common biochemical mechanism for the protein splicing process, in which the key residues are placed in several conserved motifs or blocks located through the primary amino acid sequence in the polypeptide precursor (Figure 1a). Most of the inteins react following a series of nucleophile-driven bond rearrangements and cleavage reactions (Figure 1b). Apart of the key residues described in Figure 1a, protein *trans*-splicing is also dependent on the extein junction sequences (Figure 1a) and in the case of the *trans*-inteins, on the association kinetic of the split molecules (Aranko et al., 2014).

From a biotechnological perspective, protein splicing provides a wide array of applications (Cheriyan & Perler, 2009; Shah & Muir, 2015). Split inteins are notably interesting due to the lack of splicing activity when expressed separately, which constitutes an optional mechanism of splicing control, and because they are much smaller than regular inteins. Both characteristics favor either the chemical synthesis of one of the subunits for protein ligation approaches and have less negative impact on the expression level or the overall stability of the construct. Split inteins have opened the door to many applications, from basic science research (Liu & Cowburn, 2017), expressed protein ligation (Berrade & Camarero, 2009; Blaschke et al., 2000; Muir et al., 1998), fusion protein cleavage or affinity tag removal (Chong et al., 1997; Chong et al., 1998; Li, 2011; Lu et al., 2011), to fields with actual applications in drug discovery (Fong et al., 2010; Tavassoli & Benkovic, 2007) or material assembly (Bowen et al., 2019). Protein splicing can help to solve biological problems that require the NMR structural characterization. The specific labeling of a particular domain or region within a protein would allow the study of interdomain interactions, conformational changes and facilitate protein structure determination of protein segments or domains in the context of the global protein of large size. Using protein *trans*-splicing methods is possible to specifically label a region of a protein by producing this region fused to a split intein fragment (N- or C-) in isotopic labeling conditions followed by the ligation to the rest of the protein produced in non-labeling conditions taking advantage of the splicing activity.

After the discovery of the first naturally split intein, efforts have been undertaken to find and characterize new native split inteins (Bachmann & Mootz, 2015; Gordo et al., 2018; Thiel et al., 2014). Although initially *Npu* DnaE from *Nostoc punctiforme* appeared to be the most advantageous native split intein (Evans et al., 2000; Zettler et al., 2009), other inteins have been described to display superior splicing kinetics and higher efficiency. Specifically, gp41-1, gp41-8, IMPDH-1 and NrdJ-1 split inteins display high apparent rate constants for *trans*-splicing reaction and are able to react with high yields at higher temperatures than 37°C and even in the presence of 4M urea (Carvajal-Vallejos et al., 2012). Furthermore, no side reactions neither cross-reactivity among them was observed (Carvajal-Vallejos et al., 2012; Pinto et al., 2020). Orthogonality (lack of cross-reactivity) between inteins is a critical issue when it is necessary to join more than two fragments, since different inteins are required to ligate each one of the fragments in the correct order (Busche et al., 2009; Iwai et al., 2006; Muona et al., 2010).

Although there is a wide range of applications where inteins are used, the I_N and I_C domains cannot be fused anywhere in the protein sequence since the presence of some residues in the junction between extein and intein can be crucial to carry out the reaction. Therefore, due to the sequence requirements of the ligation site, introduction of the intein can require the presence of non-native residues that would remain in the final splicing product, which is especially problematic when using this technology to address structure/function studies. As an approach to bypass this problem, Pinto et al. (2020) successfully built a library of orthogonal inteins with different sequence requirements on the ligation site, where to find suitable candidates whose native splice junction sequences match the desired target extein. An alternative approach has been to identify from natural sources (Cheriyan et al., 2013; Ellilä et al., 2011) or engineer (Appleby-Tagoe et al., 2011; Aranko et al., 2014; Beyer et al., 2020; Beyer & Iwaï, 2019; Stevens et al., 2017) more promiscuous inteins able to efficiently react in non-natural sequence contexts. Interestingly, the group of Iwaï carried out a systematic investigation to characterize the splicing efficiencies of a panel of inteins when the −1 and +2 positions were replaced by each of the remaining 19 amino acids (Oeemig et al., 2020). They showed that for some inteins, such as the previously mentioned

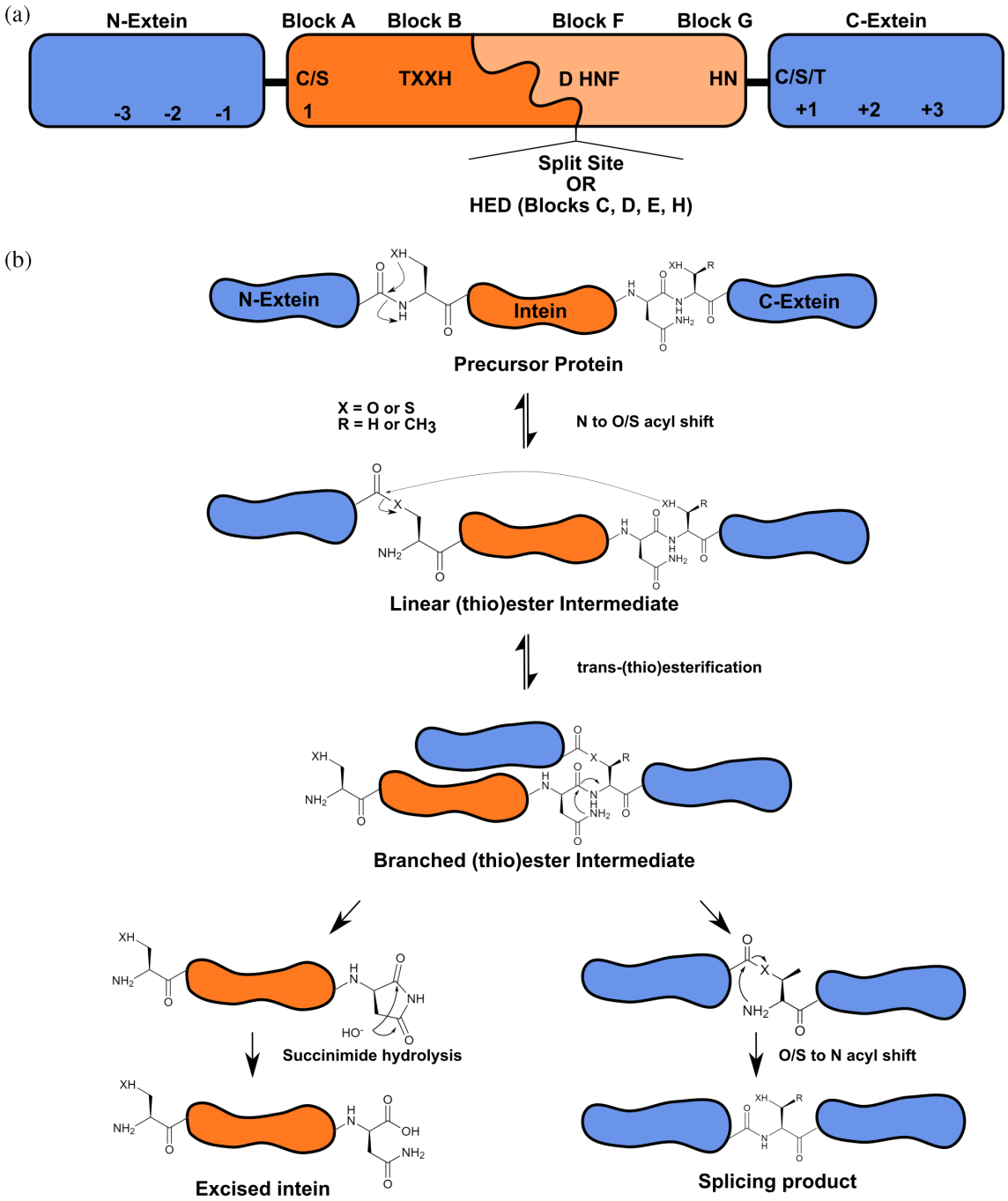


FIGURE 1 (a) Classical intein sequence, conserved motifs, and residue numbering. Conserved sequence blocks are represented, highlighting the most important residues for protein splicing. Blocks C, D, E and H are found in inteins containing the homing endonuclease domain (HED). The two colors in the intein are valid both, to show the two domains in a mini intein or the existence of the two subunits in the *trans*-inteins. (b) Canonical mechanism of protein splicing for standard or Class 1 inteins. Arrows are not intended to depict a substituting reaction but simply to show how the addition/elimination reactions are carried out.

gp41-1, these positions can be exchanged by nearly any other amino acid residues without greatly affecting the yield. The purpose of our study has been to study the possibility of ligating four different fragments using non-natural splicing contexts. We show that it is possible to produce a spliced protein from four precursors even when non-optimized splicing junctions are used.

2 | RESULTS

2.1 | Validation of inteins to produce the constructs

To produce a chimeric protein from four precursors, different orthogonal inteins had to be used. We selected four

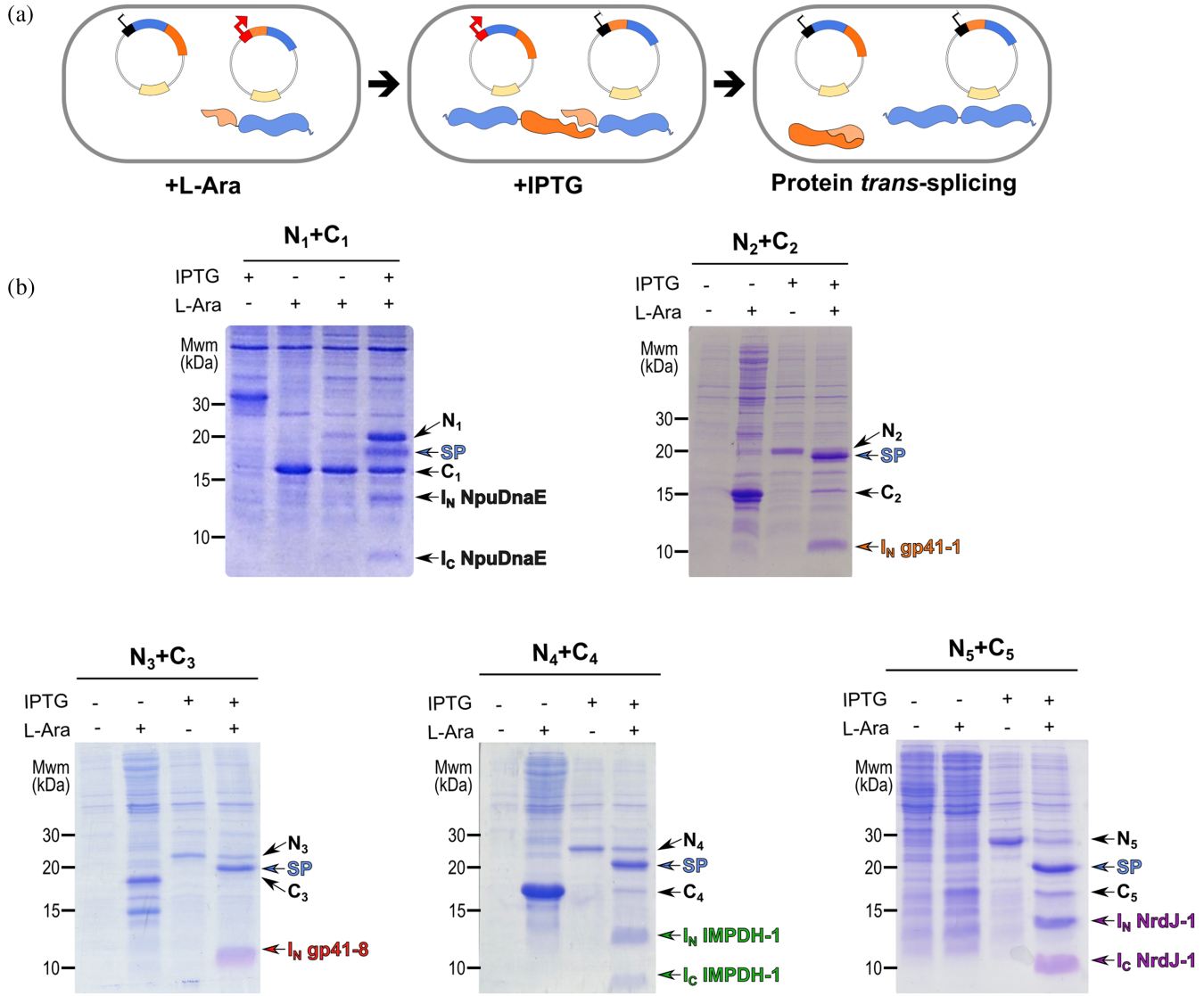


FIGURE 2 Formation of splicing products *in vivo* using the GB1-Nx and Cx-GB1 constructs with native extein residues at the splice junction. (a) Representation of the *in vivo* procedure. The C-terminal subunit of the intein is expressed upon addition of L-Arabinose fused to a GB1 domain containing a hexa-histidine tag. Later, the N-terminal subunit is expressed by adding IPTG and protein *trans*-splicing occurs from the association of both intein subunits, yielding H_6GB1-GB1H_6 as the spliced product (SP). (b) SDS-PAGE analysis of the expression of single precursors and the formation of the splicing product after adding the corresponding inductors. Single precursors are observed in cultures carrying a single vector whereas the expected splicing product (SP) is observed in co-transformed cells except in the case of N1 + C1, lane 3 (L-Ara*) that corresponds to the expression of C1 (before expressing N1) in the co-transformed culture with N1 and C1. Intein definitions can be found in Supplementary Table 1. gp41-1 and gp41-8 I_C bands are not visible in the gel due to their small molecular weight.

naturally split inteins, gp41-1, gp41-8, IMPDH-1 and NrdJ-1. For simplicity, the name of each precursor is identified with a number that indicates to which intein (N_X or C_X) the extein is fused, being Npu DnaE, gp41-1, gp41-8, IMPDH-1 and NrdJ-1, numbered as 1, 2, 3, 4 and 5, respectively (see Supplementary Table 1). Initially, we assayed the splicing activity of each endogenous I_N–I_C pair linked to GB1 as extein *in vivo* by simultaneously co-expressing H_6-GB1-I_N and I_C-GB1-H_6 pairs (Figure 2a). The split intein Npu DnaE, whose activity using this *in vivo* methodology had been previously demonstrated (Iwai et al., 2006), was included as a positive control in this early stage.

As it can be seen (Figure 2b), after the expression of both precursor proteins, *in vivo* splicing occurred for the five endogenous pairs. To confirm that in our hands these inteins did not cross-react *in vivo* each of the 20 possible combinations of I_N and I_C heterologous pairs were co-transformed in *E. coli* cells. The SDS-PAGE analysis of the co-expression of the different non-endogenous combinations showed that splicing did not occurred *in vivo* in any of them (Supplementary Figure 1 and Figure 3a for

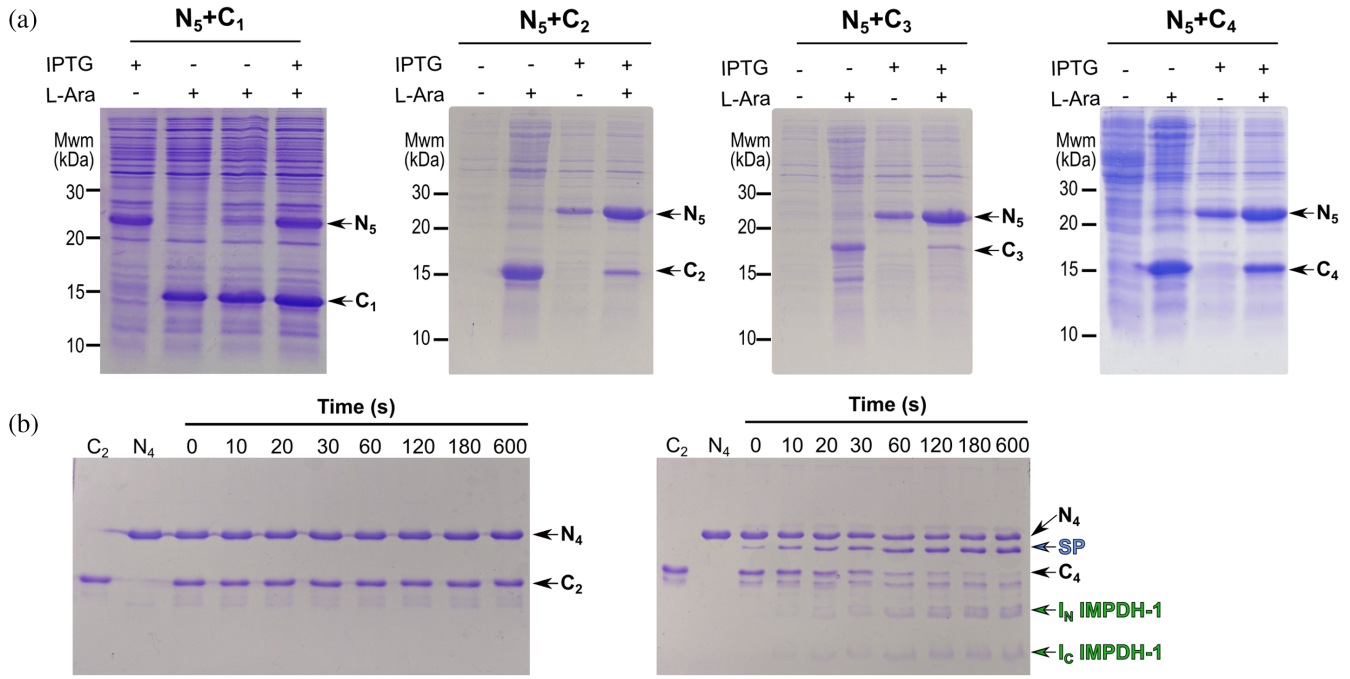


FIGURE 3 Orthogonality of split inteins *in vivo* and *in vitro*. (a) Orthogonal behavior of NrdJ-1 *in vivo*. None of the heterologous combinations yield cross-reacted spliced product (SP) after expressing both precursors. (b) Orthogonal behavior of IMPDH-1 (5 μM), as an example, *in vitro*. The heterologous combination of N4 + C2 did not produce any reaction, whereas in the endogenous N4 + C4 reaction the formation of the expected SP (as well as the excision of both intein subunits) can be observed.

N-intein of NrdJ-1 (N5) as illustrative example), confirming previous results (Carvajal-Vallejos et al., 2012; Pinto et al., 2020).

For isotopically labeling internal protein segments for NMR studies, the *in vitro* reactions can be more convenient. Thus, we also studied the splicing behavior of our constructs *in vitro* to confirm their previously described orthogonality. All H_6-GB1-Nx and Cx-GB1-H_6 precursors were expressed and purified, and splicing reactions in homologous and heterologous intein combinations were assayed and analyzed by SDS-PAGE. As an example, in Figure 3b it is shown the SDS-PAGE analysis of the endogenous combination for IMPDH-1 intein, N4 + C4, and an exogenous combination, N4 + C2, that illustrates, also *in vitro*, that splicing only occurs when Nx and Cx are the components of the same intein pair. Analysis of the rest of the reactions can be found in Supplementary Figure 2. Altogether, these results confirm that gp41-1, gp41-8, IMPDH-1, NrdJ-1, as well as Npu DnaE, behave orthogonally and can be used to ligate different protein fragments or domains in an orderly and sequential way.

2.2 | Sequential fusion of protein fragments

Precursors 1, 2, 3a and 4a (Table 1) were designed with the aim of ligating their extein domains to produce GB1-Ub-Trx-SUMO (Figure 4). Inteins gp41-1, IMPDH-1 and gp41-8 were used for the sequential ligation of the four domains orthogonally. NrdJ-1 and Npu DnaE were initially discarded because, according to previous work (Carvajal-Vallejos et al., 2012), they were described as less efficient. The first three constructs harbor a hexahistidine tag attached to the C-terminal of I_N subunit, whereas in the last one it is placed in the N-terminal of the I_C domain. These tags are removed with the inteins in a traceless manner after the splicing reaction.

We tested the promiscuity (the ability for an intein to retain its splicing activity regardless of the extein residues present at the splice junction) of the selected inteins and compared their protein splicing rate and yield when changing the native residues. We chose to maintain only the key catalytic residues Cys 1 and Ser + 1, while changing completely the surrounding sequences of the N- and C-extein (see sequence comparison in Supplementary Table 2). The non-native residues were selected because they correspond to residue sequences that joint the different domains of 43 kDa TAR DNA binding protein (TDP-43), involved in different proteinopathies (De Boer et al., 2021), that we intend to characterize by NMR in further studies.

Before assembling the full protein, each pair of I_N–I_C with the non-native extein residues (see Table 1) were tested *in vivo* using BL21 (DE3) cells to ensure that their activity was still detectable after the mentioned changes.

TABLE 1 Protein constructs used for the sequential fusion of protein fragments.

Constructs	Abbreviation	Precursor content	Theoretical Mw (Da)	Observed Mw (Da)
1	GB1N2H_6	GB1[ESGDA<u>C</u>]gp41.1N-6xHis	19,446.70	19,315.94
2	C2UbN4H_6	gp41.1C[<u>S</u>AGMQ]Ub[LRGDA<u>C</u>]IMPDH1N6xHis	28,078.79	27,941.17
3a	C4TrxN3H_6	IMPDH1C[<u>S</u>AGSD]Trx[FGCIV<u>C</u>]gp41.8N6xHis	28,213.58	ND
3b	C4SUMON5H_6	IMPDH1C[<u>S</u>AGSD]SUMO[QIGGG<u>C</u>]NrdJ1N6xHis	28,956.51	28,949.45
4a	H_{6TEV}C3SUMO	6xHisTEVgp41.8C[<u>S</u>AGSD]SUMO	18,598.60	ND
4b	H_{6TEV}C5SUMO	6xHisTEVNrdJ1C[<u>S</u>AGSD]SUMO	18,095.13	17,962.65

Note: Non-native extein residues located at the splice junction are in brackets and the catalytic Cys1 and Ser +1 are underlined.

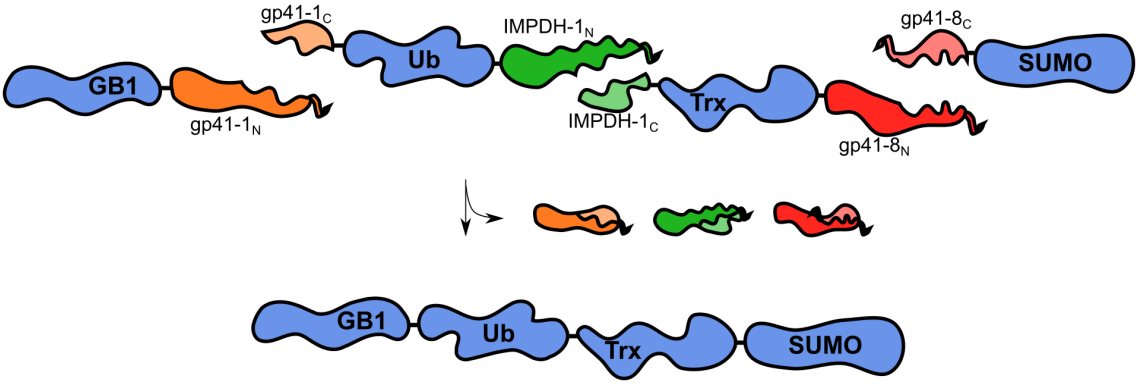


FIGURE 4 Schematic representation of the fusion of four domains using three orthogonal inteins (gp41-1, IMPDH-1, and gp41-8). Each precursor carries a hexa-histidine tag that is removed with the excised intein upon splicing. The formation of the final spliced product does not necessarily take place in a single reaction given that every individual intein has a different reaction rate. Intermediate products and unreacted precursors are omitted from the figure for simplicity.

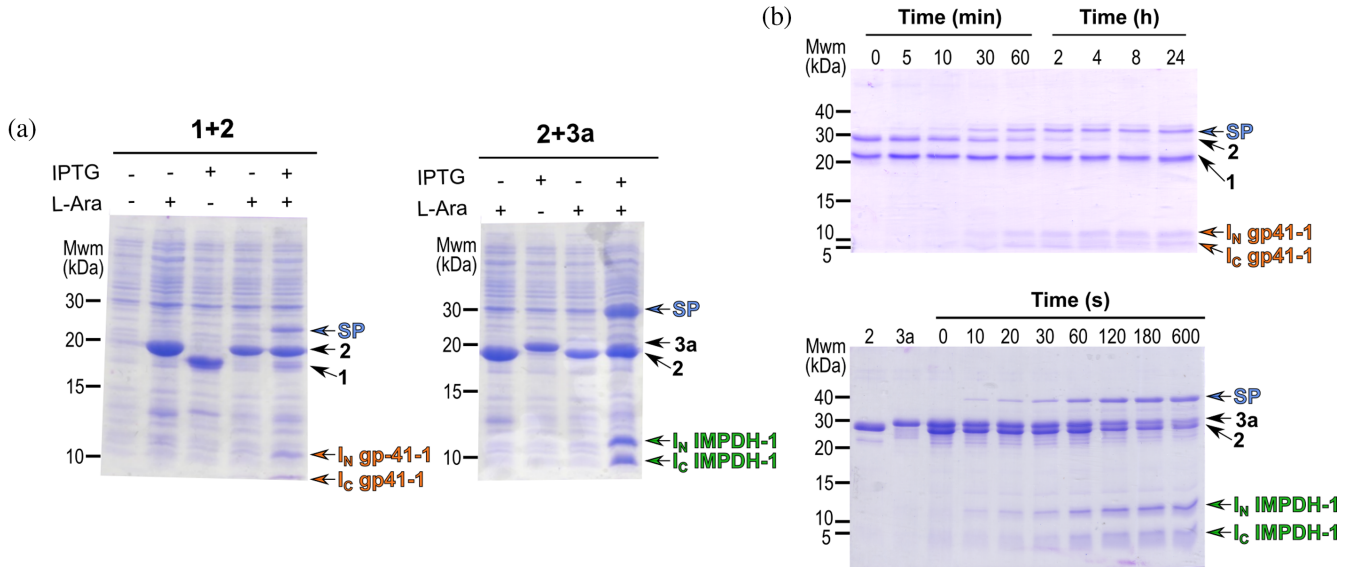


FIGURE 5 SDS-PAGE analysis of splicing reactions between precursors 1 + 2 and 2 + 3a. (a) Protein *trans*-splicing *in vivo* with constructs 1 + 2 and 2 + 3a. The expected spliced products (SP) as well as the excised intein subunits are observed after induction of both precursors. (b) Time-course of *in vitro* protein (5 μM) splicing for 1 + 2 and 2 + 3a.

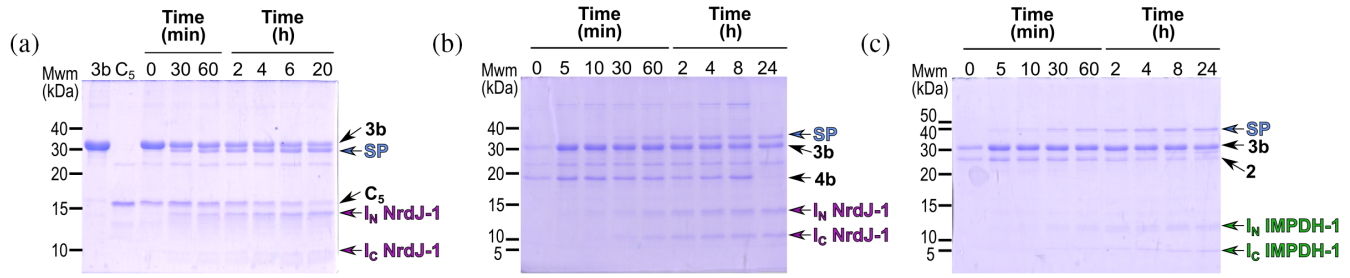
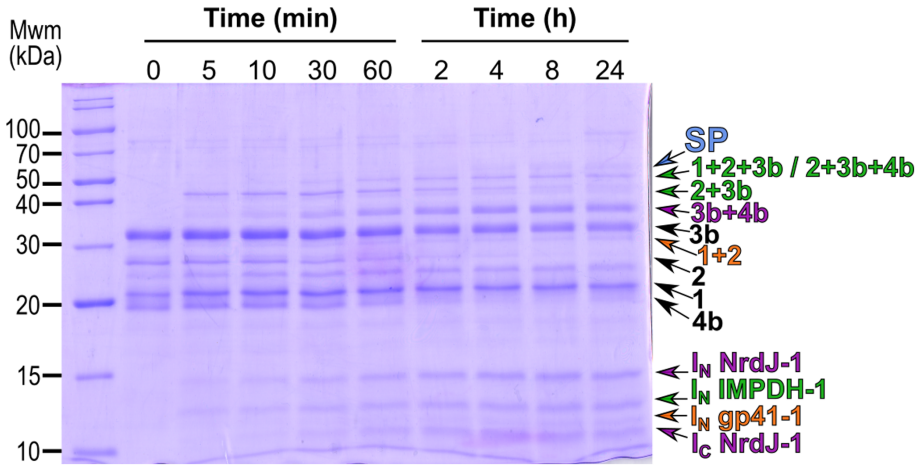


FIGURE 6 SDS-PAGE analysis of *in vitro* splicing reactions with the new constructs 3b and 4b at 37°C. (a) Time-course of *in vitro* protein splicing for 3b + C5 (5 μM) in the presence of 1M urea. (b) Time-course of *in vitro* protein splicing for 3b + 4b (5 μM) at 4M urea. (c) Time course of *in vitro* protein splicing for 3b + 2 (5 μM) at 2M urea.

FIGURE 7 Time-course SDS-PAGE analysis of the one-pot splicing reaction (2.5 μM) at 37°C and 2M urea. The full-size GB1-Ub-SUMO-SUMO spliced product (SP) is obtained, starting at 2 h. Some intermediate products can be observed (purple arrow: 3b + 4b; green arrow: 2 + 3b), as well as the excised intein subunits.



The formation of the desired splicing product was observed after the induction of both precursors for the combination of 1 + 2 and 2 + 3a (Figure 5a) but not for 3a + 4a (Supplementary Figure 3A). Then, we expressed these precursors from pSKDuet01 and purified them as described in the Material and Methods section. *In vitro* splicing assays were performed at 25 and 37°C (as well as at 2M, 4M, and 6M urea), with results resembling those observed *in vivo*. It should be noted however, that the splicing rate is remarkably slower than when using native residues as can be seen when comparing results in Figures 3b and 5b. The combination of 3a + 4a did not show splicing activity under these conditions (Supplementary Figure 3B). To overcome the possible non-functionality of gp41-8 in this context, it was decided to replace it by NrdJ-1 intein (constructs 3b and 4b, Table 1).

Due to the low solubility of the new construct 3b, splicing reactions using this precursor were carried at different urea concentrations and 37°C. Mixing 3b construct with C5 (Supplementary Table 1), which harbors the native extein residues, or 4b or 2 constructs always rendered the corresponding splicing product (Figure 6a–c, respectively). The splicing activity of 3b + 2 (Figure 6c) takes place at a slower rate than the activity of 3a + 2 (Figure 5b). Since the only difference is the extein (Trx or SUMO) contained in 3a or 3b, not only the residues at the splice junction are important, but also the whole extein can affect the intein activity.

The one-pot reaction combining 1 + 2 + 3b + 4b (see Table 1) to obtain the full GB1-Ub-SUMO-SUMO splicing product was tested *in vitro*. This one-pot reaction was performed using 2.5 μM of each precursor at 37°C and 1M urea (Figure 7). The determination of the final and intermediate products was carried out by SDS-PAGE analysis of the reactions 1 + 2; 2 + 3b; 3b + 4b; 1 + 2 + 3b; and 2 + 3b + 4b and further confirmed by Western blotting analysis using monoclonal antibodies against His-tag and SUMO (see Supplementary Figure 4). As seen in Figure 7, the splicing reaction rate when using these precursors was slower than the observed for the constructions harboring native extein residues, and therefore longer incubation times are required to observe the reaction progress. The first splicing product to appear was 2 + 3b (time = 5 min). The expected final splicing product was observed after 4 h and up to 24 h of reaction (Figure 7). Densitometric analysis of the reaction at 24 h indicated that the amount of final product corresponded

to a 1 ± 0.07% of the total reaction. Taken together, the accumulation of these intermediates indicates that the reaction of the different inteins takes place at different rates, being the reaction carried out between 2 and 3b much faster than the others.

To further study the effect of the non-native extein residues at the intein insertion sites, different time-course protein splicing reactions (Figure 8 and Supplementary Figure 2) were used to calculate the apparent first order rate constants and yield for gp41-1, IMPDH-1 and NrdJ-1 constructs indicated in Table 1 and the yield in the native and non-native extein context (Table 2). The obtained rate constants of the inteins carrying their native extein residues were very similar to those obtained in previous works (Carvajal-Vallejos et al., 2012). Overall, gp41-1 suffered a ~ 100-fold reduction in its reaction rate, although only a reduction of the yield of the splicing product to the half. NrdJ-1 underwent even a higher reduction in rate and a 3.5-fold decrease in the final splicing product yield. IMPDH-1 was affected to a lesser extent, showing higher yield and rate than the other two inteins. Even though we observed this decrease in reaction efficiency, it is important to remark that a non-negligible amount of final splicing product is still obtained with these non-native splicing junction sequences.

Since the formation of 1 + 2 takes place at a lower rate, it is likely that intermediate products could accumulate sequestering 4b or 1, which would not be able to react with 1 + 2 + 3b or 2 + 3b + 4b. If that were the case, larger amounts of final splice product could be attained if 1 + 2 and 3b + 4b reacted separately to exhaustion and were then mixed to obtain the final splicing product. To check this possibility, we reacted 5 μM of 1 + 2 and 3b + 4b for 24 h at 37°C. Then, we mixed equal volumes of both reactions and allowed the mixture to react for a further 24 h. The yield of the final product was compared on an SDS-PAGE to that obtained when the four precursors were reacted together in parallel (Figure 9a). Densitometric analysis revealed an increase of the final yield to 2 ± 0.3% (two-fold increase) using

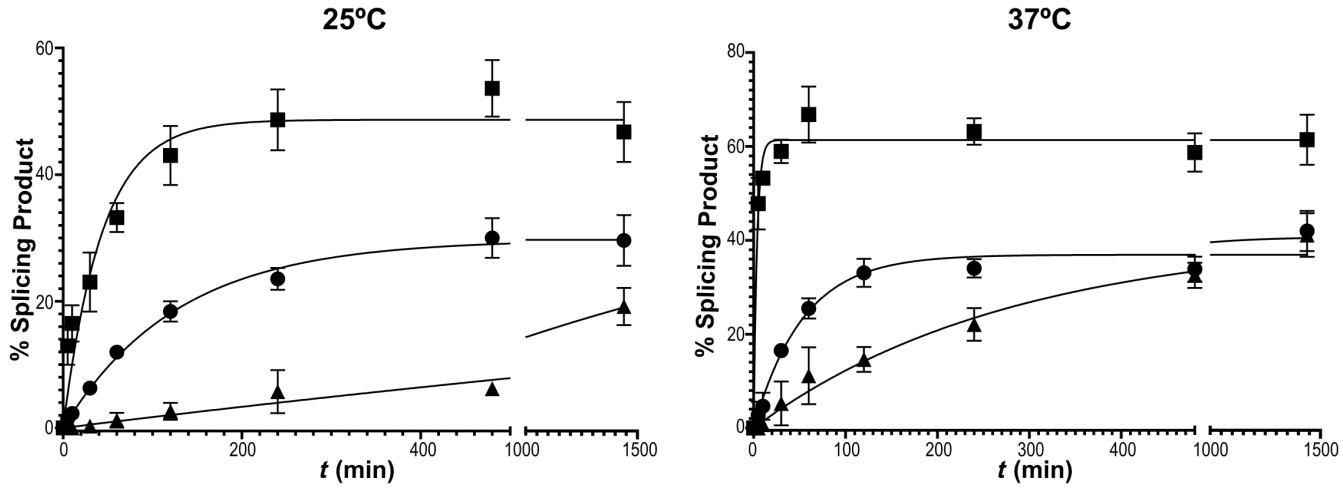


FIGURE 8 Time-course of protein *trans*-splicing reactions with non-native extein residues. The percentage of splicing product (SP) is represented in front of the time elapsed from the addition of the second precursor to the reaction mix. All reactions were assayed in triplicate. (■) IMPDH-1; (●) gp41-1; (▲) NrdJ-1.

TABLE 2 Effects of the replacements of the extein splice junctions on the apparent first order rate constants and yields ($n = 3$) for gp41-1, IMPDH-1 and NrdJ-1 *trans*-splicing reactions at 25 and 37°C.

		gp41-1		IMPDH-1		NrdJ-1	
Extein junction	**Temperature (°C)**	***k*** **(s^{-1})**	**Yield (%)**	***k*** **(s^{-1})**	**Yield (%)**	***k*** **(s^{-1})**	**Yield (%)**
Mutated[a]	25	$(1.3 \pm 0.1) \times 10^{-4}$	30 ± 1	$(3.8 \pm 0.5) \times 10^{-4}$	49 ± 2	$(2.9 \pm 0.6) \times 10^{-5}$	21 ± 2
	37	$(3.2 \pm 0.3) \times 10^{-4}$	37 ± 1	$(4.9 \pm 0.3) \times 10^{-3}$	60 ± 1	$(6.2 \pm 0.7) \times 10^{-5}$	40 ± 2
Wild type	25	$(7.0 \pm 1.2) \times 10^{-2}$	83 ± 3	$(4.7 \pm 9.9) \times 10^{-3}$	73 ± 4	$(1.87 \pm 2.9) \times 10^{-2}$	100 ± 7

Note: The rate was calculated as described in the corresponding Material and Methods section. Time courses for protein splicing were monitored for 24 h upon addition of the second precursor to the reaction mix at an equimolar concentration.

[a]Mean ± SD are indicated.

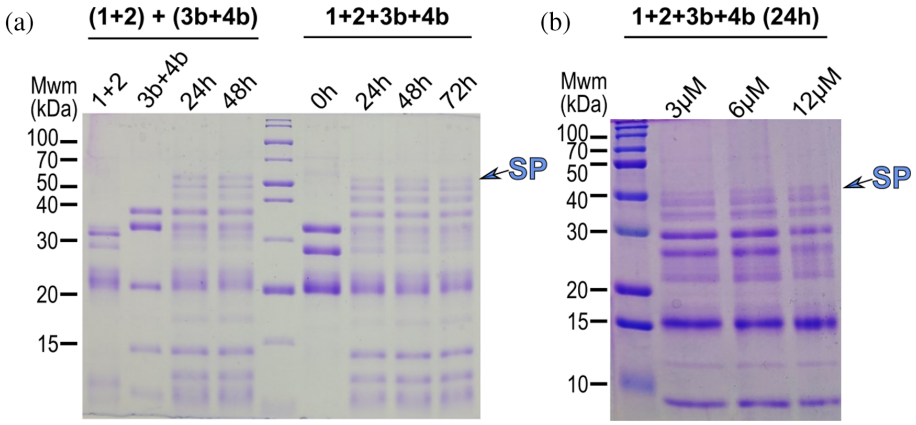


FIGURE 9 (a) SDS-PAGE analysis of the yields obtained by sequential splicing (reacting first of 1 + 2 and 3b + 4c and then mixing both reactions) and simultaneous splicing (all four precursors are initially mixed together) at 37°C. At the end, the concentration of each precursor was of 2.5 μM. (b) SDS-PAGE analysis of the splicing reactions performed for 24 h at 37°C at different precursor concentrations. The same amount of total protein was loaded in each lane.

this strategy. It is likely that a higher yield is not attained because, as is shown in Figure 9a, even after 24 h some precursors are not exhausted and therefore some intermediate products could continue to accumulate sequestering 4b or 1. Incubating 1 + 2 and 3b + 4b longer times (48 or 72 h) hardly reduced the amount of unreacted precursors (Figure 9a). The splicing reaction was tested at higher concentrations up to 12 μM. Nevertheless, the increase in precursors concentrations did not translate into additional bands that would be indicative of cross-reactivity (Figure 9b).

3 | DISCUSSION

A key to produce segmentally labeled samples is to establish protein ligation by protein *trans*-splicing with a high yield. To test protein ligation in vivo, we have used the dual-expression system used in Iwai's laboratory (Iwai et al., 2006; Muona et al., 2010) in which N- and C-terminal precursors are cloned into two separate plasmids and can be expressed by two different inducers (i.e., L-Ara and IPTG). To produce the desired mature protein from more than two precursors, at least two orthogonal split inteins are needed. In addition, these inteins cannot show cross-reactivity, that is the splicing reaction has to take place only within the endogenous I_N and I_C pair. This is necessary to ensure the sequential ligation of exteins and avoid the formation of undesired fragments or the occurrence of cyclization reactions. Obtaining a full splicing product from four independently cloned segments requires prior testing of the orthogonality of three selected inteins. Initially, we selected the four naturally split inteins gp41-1, gp41-8, IMPDH-1 and NrdJ-1 because these are the fastest and most efficient naturally split inteins that have been found and they were advocated to be orthogonal (Carvajal-Vallejos et al., 2012). In this study we confirm the previously suggested orthogonality (Carvajal-Vallejos et al., 2012; Pinto et al., 2020) of these four split intein, both, *in vitro* and *in vivo*.

We have studied the potential use of these inteins to ligate multiple protein domains expressed independently and in a sequential manner. One of the problems that arise when ligating protein segments using inteins is that they usually require the presence of non-native residues in the final splicing product due to the sequence requirements of the ligation site (i.e., native extein–intein junctions). This is especially problematic when using this technology to address structure/function studies, because the presence of these undesired residues in the final ligated protein might alter the folding, the final structure, or the interaction between domains. When using inteins for whatever purpose it must be considered that the +1 catalytic residue (Figure 1a) will appear in the final product. The inteins used in this work offer the added advantage of using Ser instead of Cys in this critical position, which allows searching within the residue sequence of the protein of interest for different insertion sites using internal Ser residues. In the worst situation, it might be assumable to substitute a certain wild-type residue by Ser. However, the selection of an intein insertion site is complicated by the fact that it has been thoroughly reported that there is a preference for extein residues found in the natural intein insertion site (Shah et al., 2013; Stevens et al., 2017). As an example, previous research on NpuDnaE (Shah et al., 2013) has shown how a single change at position −1 (Glu in that case) has a very high impact on the splicing reaction. Recently, this phenomenon has similarly been reported in other inteins (Sekar et al., 2022). Some works have shown that changes in the extein splice junction residues can drive the splicing mechanism to the formation of side-products or accumulation of the branched intermediate (Shah et al., 2012), significantly reducing splicing efficiency. On the other side, the alternative is to maintain some critical homologous native extein residues which constitutes a major inconvenience when present in the final protein.

Other studies (Pinto et al., 2020) successfully built a library of inteins where to find a suitable candidate whose native splice junction sequences match their desired target extein.

Another approach could be the substitution of native extein residues by the heterologous target protein sequence. Our approach was to study the effect of substantial changes in these sequences in order to assess the influence of the splice junction residues in the activity of the selected inteins. We have tested the promiscuity (the ability for an intein to retain its splicing activity regardless of the extein residues present at the splice junction) of the selected inteins and compared their protein-splicing rate and yield in the native and non-native context. We chose to fuse four unrelated proteins, with extein sequences in the splice junctions belonging to TDP-43 protein. Therefore, we did not expect favorable interactions between the exteins that could facilitate the association step and so the splicing reaction. Even with these sub-optimal conditions, the final product is clearly visible in the gel, roughly representing around 1%–2% of the total protein, depending on the used strategy. This fact illustrates that our approach is applicable to general situations.

Here, we demonstrate that gp41-1, IMPDH-1 and NrdJ-1 are tolerant to substantial changes in the extein residue sequences surrounding the catalytic residues and that although the reaction rate is clearly affected, the final yield after 24 h reaction remain important. We show that it is possible to ligate four different protein domains, using gp41-1, IMPDH-1 and NrdJ-1, without altering the sequence of the native final spliced product and successfully producing non-negligible amounts of the final splicing product.

To design intein constructions that provide a high yield of ligated protein becomes a critical issue when multiple reactions are considered. This fact is clearly illustrated in Figure 7. Although, individually, the three selected inteins can yield between 35% and 60% of ligated protein (Table 2), the amount of splicing product when combining 1 + 2 + 3b + 4b to obtain the full GB1-Ub-SUMO-SUMO is far lower. There, different intermediates are accumulated indicating that the reaction of the different inteins takes place at different rates, being the reaction carried out between 2 and 3b much faster than the others. This observation could indicate that some intermediates such as 2 + 3b + 4b could be sequestering precursors 4b or 1 which would not be able to react with 1 + 2 + 3b or 2 + 3b + 4b, lowering the amount of the final product. To test this situation we reacted separately to exhaustion 1 + 2 and 3b + 4b and then mixed them to obtain the final splicing product. In this case, a two-fold increase of the final splicing product was attained. Regardless, the first order rates with non-native residues are not slower than most split inteins assayed with native extein residues. As examples, we can mention Npu DnaE, with a kinetic constant around 10^{-5} s^{-1} or the Ssp family of inteins which display rates around 10^{-4} s^{-1} (Aranko et al., 2014).

We also demonstrate that inteins IMPDH-1 and NrdJ-1 work in the presence of even 3M urea. This offers the opportunity to ligate domains that may lack solubility (as is the case of the Trx extein), allowing the splicing reaction to take place under mild-denaturing or fully denaturing conditions and achieve its folding once of the final product has been formed. We also confirmed that these inteins retained the aforementioned orthogonality even with the removal of the native extein residues (Supplementary Figure 5).

Some observations can be made regarding the physicochemical properties of the non-native residues introduced in our constructs. Considering gp41-1, extein residues at the N-terminal splice junction clearly changed, switching from basic to acidic. At the C-terminal side, the sequence turned less polar, maybe reducing weak forces that stabilize intermediate intein conformations where catalytic residues position themselves to carry out the splicing reaction. Oeemig et al. (2020) have reported the effect of systematically replacing positions −1 (Tyr) and +2 (Ser) in gp41-1 by each one of the other 19 amino acids. While the replacement of Ser by Ala at the +2 position (see Supplementary Table 2) has not an effect on the yield of the spliced product, that of Tyr to Ala at −1 position drastically affects this yield (reduction to 7%). The yield of spliced product of gp41-1 carrying the non-native extein residues is reduced to around 33% (Table 2). This would indicate that other changes might compensate the substitution at position −1. Other works have shown that some of the effects of replacing splice junction residues can be reverted by mutating important residues in key catalytic blocks inside the intein (Sekar et al., 2022), restoring the interaction between these non-catalytic residues. NrdJ-1 changed in a similar way at its N-terminus, from mostly polar to mostly hydrophobic residues. IMPDH-1 retained its activity better than the two previous specimens did, which is unsurprising since the extein residues at the N-terminal splice junction remained unchanged from positions −1 to −4. At the C-terminal junction, the first residues of the extein sequence conserved the nature of the native residues at each position.

We consider our results as a valuable proof of concept for further applications of this system such as the fusion of individual domains of a protein of interest expressed independently and potentially, isotopically labeled for NMR studies. In addition, this strategy could also be used to generate native fusion proteins that show folding

problems when expressed as a single polypeptide chain or proteins that are toxic for the host cells that are used to express them. Inteins gp41-1, IMPDH-1 and NrdJ-1 retained the ability to carry out the protein splicing process, even in presence of urea, in an orthogonal manner using non-optimal sequences around the splicing junction. Although these inteins display slower reaction rates and lower yields than when using native extein residues, they display a high efficiency. Inteins of the DnaE family have been successfully modified to improve their tolerance to non-native extein residues by mutating residues around Block C (Stevens et al., 2017) or Block F (Du et al., 2009; Sekar et al., 2022). This might also be an interesting approach to consider aimed to recover the original, faster and more efficient activity of the inteins used in this study. Finally, although the residues at the splice junction between the intein and the extein are affecting the splicing reaction to some extent, it also seems important to have an extein whose structure or folding does not interfere with the appropriate association of the intein subunits and/or the positioning of the splicing catalytic residues. To overcome this handicap, it is useful to arrange different intein combinations and explore the most appropriate one. Altogether, we show a different strategy to expand the places where to insert a split intein when conducting segmental isotopic labeling.

4 | MATERIALS AND METHODS

4.1 | Constructs for intein orthogonality studies

We used pskDuet01 (AddGene Plasmid 12,172) and pskBAD2 (AddGene Plasmid 15,335) containing the N-intein (pskDuet01) and C-intein subunits (pskBAD2) of naturally split intein of the gene DnaE from the *Nostoc punctiforme* (Npu DnaE) cloned in frame with the B1 domain of IgG binding protein, protein G (GB1) (Iwai et al., 2006) as the starting backbone for our constructs.

The synthetic genes coding for gp41-1, gp41-8, IMPDH-1 and NrdJ-1N- and C-inteins including five residues of the native N-exteins and C-exteins, were purchased from Integrated DNA Technologies (Coralville, IW, USA). These intein subunits were cloned in frame with a 6xHis tag and the GB1 domain in place of the Npu DnaE extein subunits, rendering the various H_6-GB1-I_N and I_C-GB1-H_6 constructs (Supplementary Figure 1), which were named H_6GB1N1 to H_6GB1N5 and C1GB1H_6 to C5GB1H_6 for simplicity (Supplementary Table 1).

N_X and C_X refer to the I_N or I_C subunits of the five inteins used: N1-C1 refers to the subunits of Npu DnaE; N2-C2 to those of gp41-1; N3-C3 to those of gp41-8; N4-C4 to those of IMPDH-1; and N5-C5 to those of NrdJ-1.

4.2 | Constructs for the sequential fusion of protein fragments

Different precursors were designed to allow their successive ligation (Table 1). In these constructs the B1 domain of protein G (GB1), ubiquitin (Ub), thioredoxin (Trx) and Small Ubiquitin-like Modifier (SUMO) act as exteins, whereas N2/C2, N3/C3, N4/C4 and N5/C5 are the N- and C-intein subunits of gp41-1, gp41-8, IMPDH-1 and NrdJ-1 inteins, respectively. In addition, all precursors possess a 6xHis (H_6) tag to facilitate their purification using Ni-NTA affinity chromatography, which after splicing will remain bound to one of the intein subunits and thus will not be part of the final spliced product. Also, H_{6TEV}C3SUMO and H_{6TEV}C5SUMO contain the recognition sequence of the cysteine-protease from Tobacco Etch Virus (TEV) between the His-tag and the rest of the protein sequence.

Genes coding for GB1N2H_6, C2N4H_6, C4N3H_6, C4N5H_6, Ub-Trx, H_{6TEV}C5SUMO and H_{6TEV}C3SUMO were purchased from Integrated DNA Technologies (Coralville, IW, USA). All these genes contain flanking *Nde*I and *Hin*dIII restriction sites for subsequent cloning into the desired vector backbones.

4.3 | *In vivo* splicing reactions

Splicing reactions were performed *in vivo*, where pairs of intein precursors were co-transformed into BL21 (DE3) *E. coli* cells and later co-expressed to observe whether the reaction took place *in vivo*. This was made possible by using the pskDuet01 and pskBAD2 compatible vectors carrying kanamycin and ampicillin resistance genes, respectively. Expression of each construct was conducted independently by the addition of IPTG (for pskDuet01) or L-Arabinose (for pskBAD2).

For each pair, an overnight culture was established from a single co-transformed colony and transferred into a 2 mL 96 deep well plate (Fisher Scientific, New Hampshire, USA) containing 400 μL of LB media and both antibiotics. The plate was sealed with a breathable sealing film and cultures were grown overnight at 37°C and 150 rpm. These cultures were then used to inoculate 2.5 mL of fresh LB media in 10 mL 24-deep well plates containing the corresponding antibiotics and incubated at 37°C and 200 rpm. After 3 h, cells were supplemented with 0.2% L-Ara and 1 h later with 1 mM IPTG. Samples

were taken just before each inducer supplementation and 3 h after the addition of the last inducer. Cells were collected by centrifugation at 12,000 rpm, resuspended in 1x Laemmli buffer and stored at −20°C until analyzed in SDS-PAGE.

4.4 | Protein expression and purification

In order to assay intein's splicing reactions *in vitro*, all constructs were individually expressed and purified. Overnight cultures were established from single colonies and grown at 37°C and 180 rpm. Then, they were used to inoculate fresh LB media at a 1:100 ratio in the presence of the appropriate antibiotic. Cultures were then grown at 37°C and 250 rpm until an OD_{550} nm ~0.6–0.8 was reached. At this point, protein expression was induced by addition of either 1 mM IPTG or 0.2% L-Ara, depending on each vector's backbone, and further incubated for 4 h at 37°C. At the end of induction, cells were harvested by centrifugation and cell pellets were resuspended in binding buffer containing 20 mM sodium phosphate (pH 7.6), 500 mM NaCl and 20 mM imidazole. Cells were lysed by two passages though a French press (SLM-Aminco, Spectronic Instruments, IK) at 1100 psi. Soluble and insoluble cell material was fractioned via centrifugation at 11,000 rpm for 45 min at 4°C and purified as follows:

Filtered soluble proteins were loaded into the His-trap™ HP column (GE Healthcare Bio-Sciences AB, Sweden), previously equilibrated with binding buffer, and eluted with a lineal gradient from 20 to 500 mM imidazole in 30 min. Eluted fractions containing the purified protein were dialyzed against splicing buffer (20 mM sodium phosphate (pH 7), 300 mM NaCl, 1 mM EDTA and 2 mM DTT) and stored at 4°C until use (or at −80°C in the presence of 10% (v/v) glycerol).

Insoluble proteins were resuspended in binding buffer containing 7M urea to denature inclusion bodies and incubated with equilibrated Ni-NTA resin (Qiagen, Netherlands) under atmospheric conditions for 2 h at room temperature. The mixture was loaded into a column and washed with 4 column volumes of binding buffer with 7M urea. Finally, bound proteins were eluted with 20 mM sodium phosphate (pH 7.2), 500 mM NaCl, 250 mM imidazole, 7M urea. Eluted fractions were collected, and urea was removed by 2 h stepwise dialysis at room temperature against 20 mM sodium phosphate (pH 7), 300 mM NaCl containing 4M, 2M and 0M urea, progressively. Lastly, samples were dialyzed against splicing buffer and stored at −80°C in the presence of 10% (v/v) glycerol.

The identity of the precursors that were used for the sequential fusion of the full splicing product was confirmed by MALDI-TOF. In the case of $GB1N2H_6$, $C2UbN4H_6$ and $H_{6TEV}C5SUMO$, the observed molecular weight was 131 Da lower than expected, due to the spontaneous removal of the initial Methionine.

4.5 | *In vitro* splicing reactions

Pairs of purified precursors were adjusted at a concentration of 10 μM before mixing equal volumes. The reactions were carried out at 25 and 37°C, and samples were taken at different time intervals, stopping the reaction by the addition of 4x Laemmli buffer and boiling at 95°C for 5 min. Samples were analyzed by SDS-PAGE. The densitometric analysis of Coomassie stained SDS-PAGE gels was performed using ImageJ (Schneider et al., 2012). The reaction rate was calculated by normalizing each band's intensity to its molecular weight. The formed splicing product of each reaction was then normalized against the intensity of the initial band of the most consumed precursor and the data was fit to a 1st order kinetic reaction using SigmaPlot version 14.0 (Systat Software, CA, USA) as previously described (Carvajal-Vallejos et al., 2012; Zettler et al., 2009). The reaction assays were performed in triplicate.

4.6 | Western blotting

Five microliters of the splicing reaction (reaction time of 24 or 48 h) together with a mix of unreacted precursors was subjected to 15% SDS-PAGE and transferred to PDVF membranes (Millipore, Bedford, MA, USA). Proteins were detected by incubating with monoclonal antibodies against tetra-His (Qiagen, Germany; ref 34670, 1:50,000) and against SUMO-tag (GenScript, NJ, USA; antibody 4G11E9, 1:250.000) and subsequently with anti-mouse horseradish peroxidase-conjugated secondary antibody (Calbiochem, La Jolla, CA, USA). Blots were developed with Immobilon HPR substrate (Millipore, Bedford, MA, USA) and visualized with a FluorChem SP system (Alpha Innotech, San Leandro, CA, USA).

AUTHOR CONTRIBUTIONS

Alejandro Romero-Casañas: Validation; investigation; formal analysis. **Andrea García-Lizarribar:** Investigation; methodology. **Jessica Castro:** Investigation. **Maria Vilanova:** Validation; supervision. **Antoni Benito:** Writing – original draft; validation; formal analysis; supervision; writing – review and editing; project administration; funding acquisition. **Marc Ribó:** Writing – original draft; validation; conceptualization; formal analysis; supervision; writing – review and editing; project administration; funding acquisition.

ACKNOWLEDGMENTS

This research was financed by Ministerio de Economia y Competitividad, through project BIO2013-43517, by Universitat de Girona through projects MPCUdG2016-18, and by DGU, Generalitat de Catalunya through project 2017-SGR 1720. Open Access funding provided thanks to the CRUE-CSIC agreement with Wiley.

ORCID

Antoni Benito https://orcid.org/0000-0002-9776-1003

REFERENCES

Appleby-Tagoe JH, Thiel IV, Wang Y, Wang Y, Mootz HD, Liu X-Q. Highly efficient and more general *cis*- and *trans*-splicing inteins through sequential directed evolution. J Biol Chem. 2011;286(39):34440–7. https://doi.org/10.1074/jbc.M111.277350

Aranko AS, Wlodawer A, Iwaï H. Nature's recipe for splitting inteins. Protein Eng Des Sel. 2014;27(8):263–71. https://doi.org/10.1093/protein/gzu028

Bachmann A-L, Mootz HD. An unprecedented combination of serine and cysteine nucleophiles in a split intein with an atypical split site. J Biol Chem. 2015;290(48):28792–804. https://doi.org/10.1074/jbc.M115.677237

Berrade L, Camarero JA. Expressed protein ligation: a resourceful tool to study protein structure and function. Cell Mol Life Sci. 2009;66(24):3909–22. https://doi.org/10.1007/s00018-009-0122-3

Beyer HM, Iwaï H. Off-pathway-sensitive protein-splicing screening based on a toxin/antitoxin system. Chembiochem. 2019;20(15):1933–8. https://doi.org/10.1002/cbic.201900139

Beyer HM, Mikula KM, Li M, Wlodawer A, Iwaï H. The crystal structure of the naturally split Gp41-1 intein guides the engineering of orthogonal split inteins from *cis*-splicing inteins. FEBS J. 2020;287(9):1886–98. https://doi.org/10.1111/febs.15113

Blaschke UK, Cotton GJ, Muir TW. Synthesis of multi-domain proteins using expressed protein ligation: strategies for segmental isotopic labeling of internal regions. Tetrahedron. 2000;56(48):9461–70. https://doi.org/10.1016/S0040-4020(00)00830-9

Bowen CH, Reed TJ, Sargent CJ, Mpamo B, Galazka JM, Zhang F. Seeded chain-growth polymerization of proteins in living bacterial cells. ACS Synth Biol. 2019;8(12):2651–8. https://doi.org/10.1021/acssynbio.9b00362

Busche AEL, Aranko AS, Talebzadeh-Farooji M, Bernhard F, Dötsch V, Iwaï 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. 2009;48:4–6131.

Carvajal-Vallejos P, Pallissé R, Mootz HD, Schmidt SR. Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources. J Biol Chem. 2012;287(34):28686–96. https://doi.org/10.1074/jbc.M112.372680

Cheriyan M, Pedamallu CS, Tori K, Perler F. Faster protein splicing with the *Nostoc punctiforme* DnaE intein using non-native extein residues. J Biol Chem. 2013;288(9):6202–11. https://doi.org/10.1074/jbc.M112.433094

Cheriyan M, Perler FB. Protein splicing: a versatile tool for drug discovery. Adv Drug Deliv Rev. 2009;61(11):899–907. https://doi.org/10.1016/j.addr.2009.04.021

Chong S, Montello GE, Zhang A, Cantor EJ, Liao W, Xu M-Q, et al. 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(22):5109–15. https://doi.org/10.1093/nar/26.22.5109

Chong S, Vence LM, Hirvonen CA. Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. Gene. 1997;192(2):271–81. https://doi.org/10.1016/s0378-1119(97)00105-4

De Boer EMJ, Orie VK, Williams T, Baker MR, De Oliveira HM, Polvikoski T, et al. TDP-43 proteinopathies: a new wave of neurodegenerative diseases. J Neurol Neurosurg Psychiatry. 2021;92(1):86–95. https://doi.org/10.1136/jnnp-2020-322983

Du Z, Shemella PT, Liu Y, McCallum SA, Pereira B, Nayak SK, et al. Highly conserved histidine plays a dual catalytic role in protein splicing: a PKa shift mechanism. J Am Chem Soc. 2009;131(32):11581–9. https://doi.org/10.1021/ja904318w

Elliläl S, Jurvansuu JM, Iwaï H. Evaluation and comparison of protein splicing by exogenous inteins with foreign exteins in *Escherichia Coli*. FEBS Lett. 2011;585(21):3471–7. https://doi.org/10.1016/j.febslet.2011.10.005

Evans TC, Martin D, Kolly R, Panne D, Sun L, Ghosh I, et al. Protein *trans*-splicing and cyclization by a naturally split intein from the DnaE gene of *Synechocystis* species PCC6803. J Biol Chem. 2000;275(13):9091–4. https://doi.org/10.1074/jbc.275.13.9091

Fong BA, Wu W-Y, Wood DW. The potential role of self-cleaving purification tags in commercial-scale processes. Trends Biotechnol. 2010;28(5):272–9. https://doi.org/10.1016/j.tibtech.2010.02.003

Gordo V, Aparicio D, Pérez-Luque R, Benito A, Vilanova M, Usón I, et al. Structural insights into subunits assembly and the oxyester splicing mechanism of Neq pol split intein. Cell Chem Biol. 2018;25(7):871–879.e2. https://doi.org/10.1016/j.chembiol.2018.04.008

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. 2006;580(7):1853–8. https://doi.org/10.1016/j.febslet.2006.02.045

Li Y. Self-cleaving fusion tags for recombinant protein production. Biotechnol Lett. 2011;33(5):869–81. https://doi.org/10.1007/s10529-011-0533-8

Liu D, Cowburn D. Segmental isotopic labeling of proteins for NMR study using intein technology. In: Mootz HD, editor. Split inteins. Methods in molecular biology. Volume 1495. New York, NY: Springer New York; 2017. p. 131–45. https://doi.org/10.1007/978-1-4939-6451-2_9

Lu W, Sun Z, Tang Y, Chen J, Tang F, Zhang J, et al. Split intein facilitated tag affinity purification for recombinant proteins with controllable tag removal by inducible auto-cleavage. J Chromatogr A. 2011;1218(18):2553–60. https://doi.org/10.1016/j.chroma.2011.02.053

Muir TW, Sondhi D, Cole PA. Expressed protein ligation: a general method for protein engineering. Proc Natl Acad Sci. 1998;95(12):6705–10. https://doi.org/10.1073/pnas.95.12.6705

Muona M, Aranko AS, Raulinaitis V, Iwaï H. Segmental isotopic labeling of multi-domain and fusion proteins by protein *trans*-splicing in vivo and in vitro. Nat Protoc. 2010;5(3):574–87. https://doi.org/10.1038/nprot.2009.240

Oeemig JS, Beyer HM, Aranko AS, Mutanen J, Iwaï H. Substrate specificities of inteins investigated by QuickDrop-cassette mutagenesis. FEBS Lett. 2020;594(20):3338–55. https://doi.org/10.1002/1873-3468.13909

Pinto F, Thornton EL, Wang B. An expanded library of orthogonal split inteins enables modular multi-peptide assemblies. Nat Commun. 2020;11(1):1529. https://doi.org/10.1038/s41467-020-15272-2

Schneider CA, Rasband WS, Eliceiri KW. NIH image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671–5. https://doi.org/10.1038/nmeth.2089

Sekar G, Stevens AJ, Mostafavi AZ, Sashi P, Muir TW, Cowburn D. A conserved histidine residue drives extein dependence in an enhanced atypically split intein. J Am Chem Soc. 2022;144(41):19196–203. https://doi.org/10.1021/jacs.2c08985

Shah NH, Dann GP, Vila-Perelló M, Liu Z, Muir TW. Ultrafast protein splicing is common among cyanobacterial split inteins: implications for protein engineering. J Am Chem Soc. 2012;134(28):11338–41. https://doi.org/10.1021/ja303226x

Shah NH, Eryilmaz E, Cowburn D, Muir TW. Extein residues play an intimate role in the rate-limiting step of protein *trans*-splicing. J Am Chem Soc. 2013;135(15):5839–47. https://doi.org/10.1021/ja401015p

Shah NH, Muir TW. Inteins: nature's gift to protein chemists. Chem Sci. 2015;5(1):446–61. https://doi.org/10.1039/C3SC52951G

Stevens AJ, Sekar G, Shah NH, Mostafavi AZ, Cowburn D, Muir TW. A promiscuous split intein with expanded protein engineering applications. Proc Natl Acad Sci U S A. 2017;114(32):8538–43. https://doi.org/10.1073/pnas.1701083114

Tavassoli A, Benkovic SJ. Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in *E. coli*. Nat Protoc. 2007;2(5):1126–33. https://doi.org/10.1038/nprot.2007.152

Thiel IV, Volkmann G, Pietrokovski S, Mootz HD. An atypical naturally split intein engineered for highly efficient protein labeling. Angew Chem. 2014;126(5):1330–4. https://doi.org/10.1002/ange.201307969

Zettler J, Schütz V, Mootz HD. The naturally split *Npu* DnaE intein exhibits an extraordinarily high rate in the protein *trans*-splicing reaction. FEBS Lett. 2009;583(5):909–14. https://doi.org/10.1016/j.febslet.2009.02.003

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Romero-Casañas A, García-Lizarribar A, Castro J, Vilanova M, Benito A, Ribó M. Ligation of multiple protein domains using orthogonal inteins with non-native splice junctions. Protein Science. 2024;33(7):e5070. https://doi.org/10.1002/pro.5070