Article

Extein Residues Play an Intimate Role in the Rate Limiting Step of Protein Trans-Splicing

Neel H. Shah, Ertan Eryilmaz, David Cowburn, and Tom W. Muir

J. Am. Chem. Soc., **Just Accepted Manuscript** • DOI: 10.1021/ja401015p • Publication Date (Web): 18 Mar 2013

Extein Residues Play an Intimate Role in the Rate Limiting Step of Protein *Trans*-Splicing

Neel H. Shah,† Ertan Eryilmaz,‡ David Cowburn,‡ Tom W. Muir*,†

† Department of Chemistry, Princeton University, Frick Laboratory, Princeton, New Jersey 08544, United States

‡ Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York 10461, United States

ABSTRACT: Split inteins play an important role in modern protein semisynthesis techniques. These naturally occurring protein splicing domains can be used for *in vitro* and *in vivo* protein modification, peptide and protein cyclization, segmental isotopic labeling, and the construction of biosensors. The most well-characterized family of split inteins, the cyanobacterial DnaE inteins, show particular promise as many of these can splice proteins in under one minute. Despite this fact, the activity of these inteins is context-dependent: certain peptide sequences surrounding their ligation junction (called local N- and C-exteins) are strongly preferred, while other sequences cause a dramatic reduction in splicing kinetics and yields. These sequence constraints limit the utility of inteins, and thus a more detailed understanding of their participation in protein splicing is needed. Here, we present a thorough kinetic analysis of the relationship between C-extein composition and split intein activity. The results of these experiments were used to guide structural and molecular dynamics studies, which revealed that the motions of catalytic residues are constrained by the second C-extein residue, likely forcing them into an active conformation that promotes rapid protein splicing. Together, our structural and functional studies also highlight a key region of the intein structure that can be re-engineered to increase intein promiscuity.

INTRODUCTION

Protein splicing is a post-translational auto-processing event carried out by a class of proteins known as inteins.[1] During this process, an intein domain excises itself from a larger precursor protein and ligates its N- and C-terminal flanking sequences (termed exteins) through a native peptide bond. Inteins naturally exist in two forms. Most are *cis*-splicing inteins that are expressed as single polypeptide chains embedded within their host proteins. By contrast, the far less abundant *trans*-splicing inteins are transcribed and translated as two separate protomers that associate and fold into the canonical intein domain structure.[2] The association rate of naturally split inteins is rapid compared to the rate of the subsequent protein splicing reaction.[3] Regardless of whether splicing occurs in *cis* or in *trans*, the mechanism of protein splicing is the same (Figure 1). First, the N-extein/intein peptide bond is activated through an N-to-S acyl shift to form a linear thioester intermediate. Next, this activated acyl group undergoes *trans*-thioesterification to form a branched thioester intermediate (BI) on the first residue of the C-extein, Cys_{+1}. In the last chemo-enzymatic step, the C-terminal Asn residue of the intein cyclizes, thereby resolving this branched intermediate into an excised intein and an N-extein/C-extein thioester adduct. Finally, this transient thioester spontaneously rearranges to a native peptide bond to yield the spliced product, and the excised intein succinimide hydrolyzes to yield a free carboxylate.

It is noteworthy that while different families of inteins utilize subtle variations on this general biochemical mechanism (such as Ser or Thr nucleophiles, rather than Cys), the catalytic residues for protein splicing are always confined to the intein domain and the first C-extein residue.[1] Despite this fact, a growing body of experimental evidence indicates that intein splicing efficiency is highly dependent on the identity of two or three local extein residues on either side of the splice junction.[4-10] For example, introduction of non-native residues at the -3, -2, and -1 positions, located on the N-extein (Figure 1), can alter the linear thioester formation efficiency or promote hydrolysis of this intermediate. Mutation of the +1, +2, and +3 residues, located on the C-extein (Figure 1), can abolish or greatly diminish splicing activity, and even lead to premature asparagine cyclization before branched intermediate formation. For each intein family, this context dependent activity is dictated by evolutionary pressures, as inteins are naturally embedded between highly conserved residues in a number of different endogenous host proteins.[11] As a result, different inteins are biased towards different sequences at their splice junction.

The chemical synthesis of larger and more complex peptides and proteins is an ongoing challenge, and inteins are being widely used to facilitate such syntheses.[12] Thus, the sensitivity of protein splicing to local extein sequence (i.e. residues immediately flanking the intein) has significant practical implications. All intein-based technologies are premised on a single notion: the chemical perturbations that an intein carries out on its endogenous host protein can be applied in a virtually traceless manner to any exogenous protein of interest. In reality, however, efficient and traceless synthesis of complex products is not always achieved. Rather, current technologies often require either the incorporation of non-native residues surrounding the splice junction in the target molecule or the sacrifice of reaction kinetics and product yields to obtain

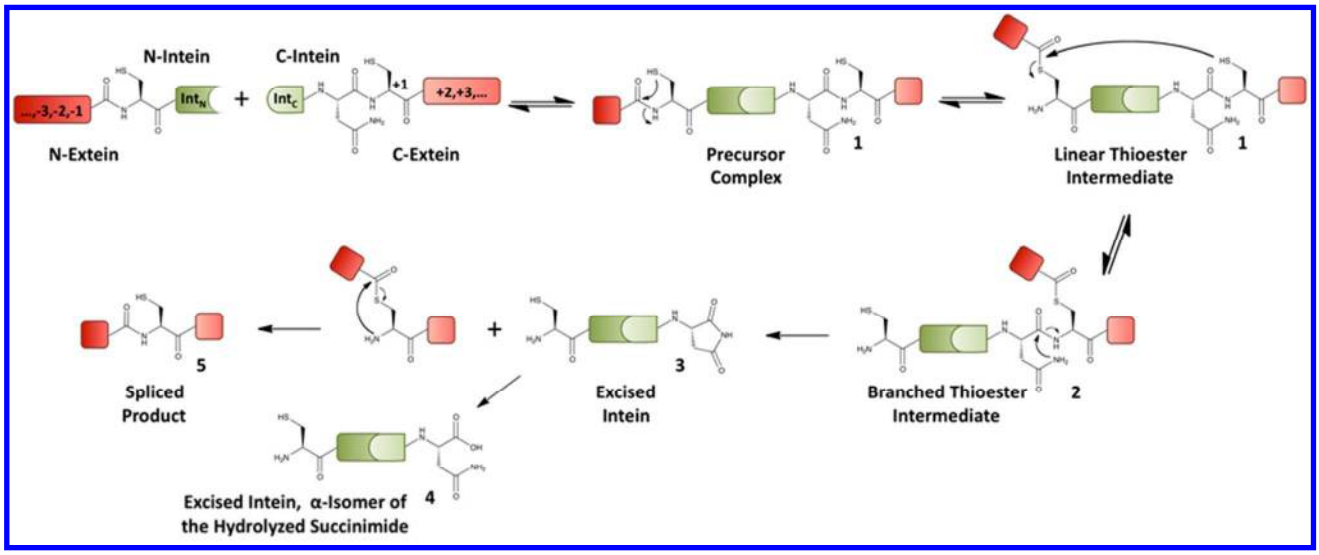


Figure 1. The mechanism of protein *trans*-splicing (PTS). Relevant species along the reaction coordinate are labeled. Numbers **1**-**4** refer to the chemically distinct C-intein adducts that can be observed in the splicing assays described in this report (see Figures 3 and S8). Note that for simplicity, only the α-amino acid isomer of **4** is shown, however species **3** can ring-open into an α- or β-amino acid form.

the desired native sequence. An improved understanding of the general splicing mechanism and of its sensitivity to local extein sequences thus remains of central concern.

Of particular interest as protein engineering tools are the split DnaE inteins, all of which endogenously generate the catalytic subunit of DNA polymerase III after protein *trans*-splicing.[13] Until recently, many split intein-based technologies relied on the founding member of this family termed Ssp, which derives its name from the model cyanobacterium that encodes it, *Synechocystis species PCC6803*.[2] However, Ssp catalyzes protein *trans*-splicing in hours, which is too slow for many practical applications.[14] With the discovery and characterization of new split DnaE inteins, such as the now prevalent *Nostoc punctiforme* (Npu) intein, it is clear that several members of this family catalyze protein splicing with extraordinary efficiency, in minutes or less.[5,9,15,16] Thus, many intein technologies are now being developed and improved with these new tools, including *in vitro* and *in vivo* protein semisynthesis,[17-19] segmental isotopic labeling,[20,21] peptide cyclization,[22] and the construction of novel biosensors.[23,24]

The DnaE split intein family is, however, also plagued by poor tolerance for non-native local extein sequences. All split DnaE inteins are naturally embedded within the local N-extein sequence AEY (Figure 1, residues -3, -2, -1) and C-extein sequence CFN (Figure 1, residues +1, +2, +3). Several reports indicate that DnaE inteins can tolerate significant deviation from this native N-extein sequence.[6,10,16,19,25] Conversely, the presence of non-native C-extein residues can lead to dramatic reductions of splicing efficiency. For example, mutation of the canonical CFN sequence to SGV inhibits branched intermediate resolution for Ssp, although the contributions of each C-extein mutation were not individually assessed.[7] Additionally, the identity of the +2 C-extein residue has a dramatic impact on splicing activity for all members of the DnaE family, but it is not clear what step in the splicing pathway is modulated by this residue.[5,9,10]

Despite the fact that C-extein-dependent splicing activity is well documented for DnaE inteins, little is known about the magnitude of this effect on reaction kinetics, nor the physical basis of this phenomenon. We envisioned that a detailed understanding of how C-extein residues participate in the splicing reaction could help guide the practical use of split inteins and help lay the foundation for the design of more promiscuous engineered inteins. To this end, we performed a detailed structure–activity analysis on the Npu intein, employing semisynthesis to systematically alter the C-extein moiety, thereby providing the raw materials for a series of kinetic and structural analyses. This effort led to the finding that the +2 residue in the C-extein plays a critical role in constraining the active site of the intein during resolution of the branched intermediate. The work also draws attention to a loop region in the intein structure that appears to sense C-extein composition and as such might be a productive focus of engineering efforts geared towards increasing intein promiscuity.

RESULTS

Semisynthesis of split inteins with varying C-extein composition. Our efforts began with the construction of a library of C-intein fragments (Int_C) bearing a variety of model C-exteins ranging from a single Cys residue with different capping groups to tri-peptides with unique sequences (Table 1). To rapidly generate the desired constructs, seventeen proteins in all, we employed a semisynthetic approach that utilized Expressed Protein Ligation (Figure 2).[26] Specifically, the Int_C fragments of Npu and Ssp (referred to as Npu_C and Ssp_C, respectively) were expressed in *E. coli* fused to the *cis*-splicing His_6-tagged GyrA intein and enriched over Ni columns (Figure S4). The crude fusion proteins were then reacted with either a large excess of a cysteine derivative (100 mM) to directly yield an Int_C-Cys adduct, or they were thiolyzed with 100 mM 2-mercaptoethanesulfonate (MES) in the presence of a 1-5 mM di- or tri-peptide to yield Int_C-peptide adducts (Figures 2A and B). The desired product from each reaction was

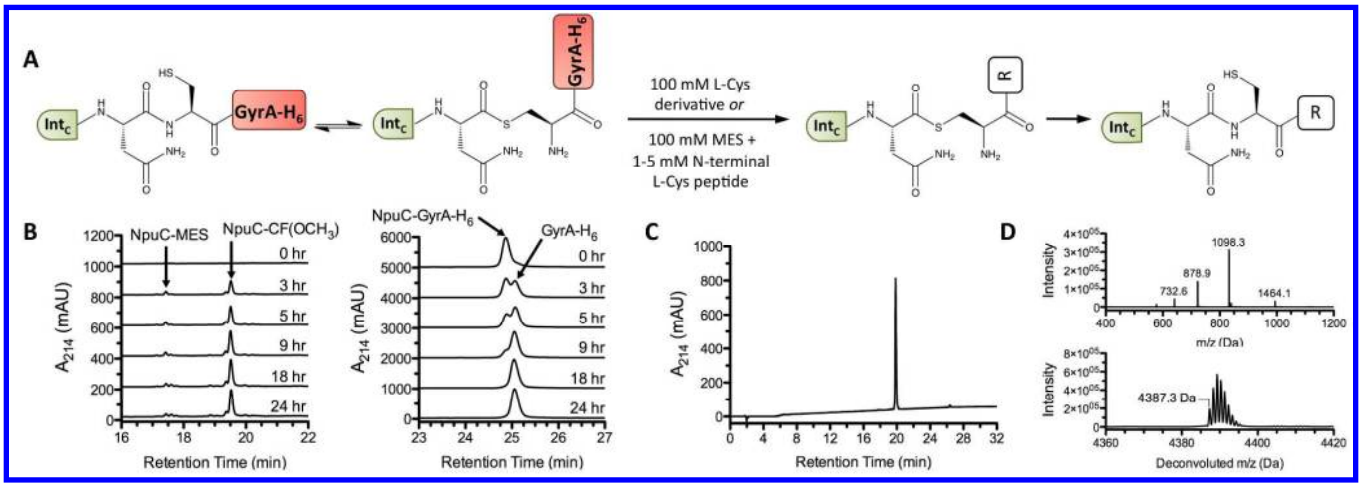


Figure 2. Semisynthesis of C-intein constructs. (A) Semisynthetic scheme (R = -OH, -OCH_3, -NH_2, -$NHCH_3$, or an additional one or two amino acids, as indicated in Table 1). (B) RP-HPLC analysis of a one-pot MES-thiolysis/ligation to synthesize Npu$_C$-CF(OCH_3). Npu$_C$-MES and ligation product accumulation are shown in the left panel and cleavage of the Npu$_C$-GyrA-H_6 fusion protein is shown in the right panel. (C) RP-HPLC and (D) ESI-MS analysis of Npu$_C$-CF(OCH_3) after purification. The raw mass spectrum is shown in the top panel and the deconvoluted spectrum is shown in the bottom panel (expected monoisotopic mass = 4387.28 Da).

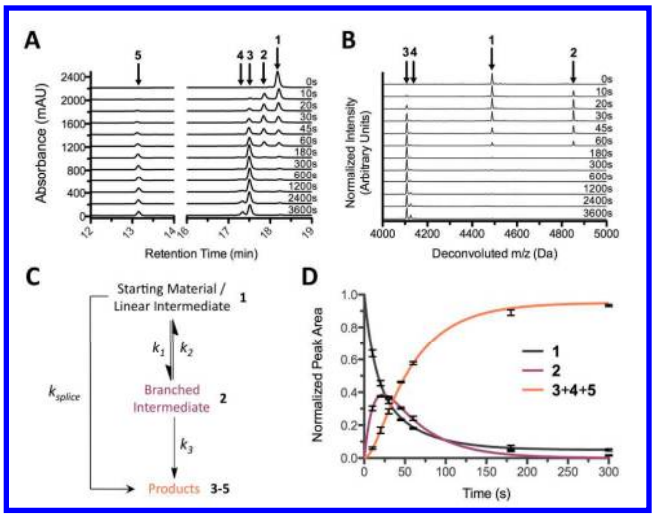


Figure 3. Splicing assays to analyze branched intermediate formation and resolution. Time-dependent RP-HPLC (A) and ESI-MS (B) analyses of the reaction between AEY-Npu$_N$ and Npu$_C$-CFN(NH_2) (Table 1, reaction 1). The deconvoluted mass spectra in panel B are normalized to the intensity of the largest peak at each time point. (C) Simplified three-state kinetic model of protein splicing compatible with the analytical techniques presented herein. (D) Quantified reaction progress data for reaction 1 fit to the kinetic model in panel C. Note that the product curve is the combined peak areas of all three products of the *trans*-splicing process, the two excised intein species (**3** and **4**) and the spliced product (**5**). Error bars represent the standard deviation from three independent reactions. Numbering corresponds to the numbered species defined in Figures 1 and S8.

readily purified by reverse-phase high performance liquid chromatography (RP-HPLC, Figures 2C and S7) and its identity was confirmed by electrospray ionization mass spectrometry (ESI-MS, Figure 2D and Table S2). Importantly, this semisynthesis approach allowed for the modular assembly of constructs with natural amino acid mutations within the C-intein and effectively any functional groups in the C-extein side-chains and backbone.

Kinetic assays to monitor branched intermediate formation and resolution. To rigorously assess C-extein effects on protein *trans*-splicing, we developed two complementary analytical approaches that allowed us to distinguish various chemical species along the reaction coordinate in a time-resolved fashion. First, N-intein (Int$_N$) proteins bearing a minimized N-extein tripeptide (AEY-Int$_N$) were generated recombinantly and purified (Figures S5-S7 and Table S2). These constructs were mixed with their Int$_C$ counterparts at 30 °C, and aliquots were removed from the reaction solution at various time points and quenched by acidification to pH 1-2. Importantly, all reactions were carried out at pH 7.2 in the absence of thiol-based reducing agents to prevent any undesired hydrolysis or thiolysis reactions that would convolute kinetic analyses. Time points were analyzed by RP-HPLC, and for most reactions the various Int$_C$-related species (Figure 1, **1**-**4**) and the spliced product (Figure 1, **5**) could be readily separated (Figures 3A and S9). For reactions where sufficient separation between species **1**-**5** was not achieved by RP-HPLC, the quenched time points were desalted and analyzed as complex mixtures by ESI-MS (Figures 3B and S10). Given the similarity in sequence composition, size, and net charge between species **1**-**4**, the molecules showed similar levels of ionization, and thus the RP-HPLC analyses and ESI-MS analyses gave virtually identical results (compare Figures 3A and B and see Figure S13 for quantitative analysis of the error between the two assays). Importantly, in both assay formats, the starting material and linear intermediate were indistinguishable, so the data were fit to a simplified kinetic model that collapsed the first two catalytic steps into a single equilibrium reaction (Figures 3C and D). The results of our kinetic analyses are summarized in Table 1 and Figure 4.

We initially carried out a series of control reactions to validate our assays. The splicing kinetics of the wild-type Npu and Ssp inteins were assessed in their native N- and C-extein contexts (Table 1, reactions 1 and 2). The overall rates of spliced product formation (k_{splice}) were 1.36 x 10^{-2} s^{-1} and 1.46 x 10^{-4} s^{-1}, respectively, consistent with

Table 1. Rate constants for individual steps and the overall splicing reaction.[a]

Rxn	Intein	C-Extein	k_1 (s^{-1})	k_2 (s^{-1})	k_3 (s^{-1})	k_{splice} (s^{-1})
1	Npu_{WT}	$CFN(NH_2)$	$(5.21 \pm 0.28) \times 10^{-2}$	$(1.77 \pm 0.38) \times 10^{-2}$	$(3.15 \pm 0.04) \times 10^{-2}$	$(1.36 \pm 0.02) \times 10^{-2}$
2	Ssp_{WT}	$CFN(NH_2)$	$(4.70 \pm 0.26) \times 10^{-3}$	$(7.03 \pm 0.44) \times 10^{-3}$	$(3.86 \pm 0.17) \times 10^{-4}$	$(1.46 \pm 0.03) \times 10^{-4}$
3 [b]	Npu_{C1A}	$CFN(NH_2)$	-	-	$(1.43 \pm 0.03) \times 10^{-4}$	-
4 [c]	Npu_{N137A}	$CFN(NH_2)$	$(1.70 \pm 0.13) \times 10^{-2}$	$(1.86 \pm 0.11) \times 10^{-3}$	-	-
5 [d]	Npu_{WT}	C(OH)	$(2.41 \pm 0.07) \times 10^{-2}$	$(4.40 \pm 0.14) \times 10^{-3}$	-	-
6	Npu_{WT}	$C(OCH_3)$	$(6.59 \pm 0.20) \times 10^{-2}$	$(1.56 \pm 0.06) \times 10^{-2}$	$(4.76 \pm 0.13) \times 10^{-4}$	$(4.32 \pm 0.16) \times 10^{-4}$
7	Npu_{WT}	$C(NH_2)$	$(3.16 \pm 0.09) \times 10^{-2}$	$(6.59 \pm 2.41) \times 10^{-3}$	$(7.31 \pm 0.26) \times 10^{-5}$	$(6.30 \pm 0.87) \times 10^{-5}$
8	Npu_{WT}	$C(NHCH_3)$	$(4.40 \pm 0.35) \times 10^{-2}$	$(1.13 \pm 0.10) \times 10^{-2}$	$(1.33 \pm 0.01) \times 10^{-4}$	$(1.08 \pm 0.03) \times 10^{-4}$
9	Npu_{WT}	$CF(OCH_3)$	$(5.90 \pm 0.85) \times 10^{-2}$	$(1.20 \pm 0.39) \times 10^{-2}$	$(1.56 \pm 0.16) \times 10^{-3}$	$(1.28 \pm 0.01) \times 10^{-3}$
10	Npu_{WT}	$CF(NH_2)$	$(6.10 \pm 1.10) \times 10^{-2}$	$(1.13 \pm 0.40) \times 10^{-2}$	$(9.30 \pm 0.42) \times 10^{-3}$	$(6.32 \pm 0.11) \times 10^{-3}$
11	Npu_{WT}	$CFA(NH_2)$	$(6.05 \pm 0.36) \times 10^{-2}$	$(1.31 \pm 0.27) \times 10^{-2}$	$(2.57 \pm 0.04) \times 10^{-2}$	$(1.31 \pm 0.02) \times 10^{-2}$
12	Npu_{WT}	$CAN(NH_2)$	$(7.11 \pm 2.11) \times 10^{-2}$	$(2.74 \pm 0.58) \times 10^{-2}$	$(3.12 \pm 0.28) \times 10^{-4}$	$(2.39 \pm 0.12) \times 10^{-4}$
13 [b]	Npu_{C1A}	$CAN(NH_2)$	-	-	$(2.41 \pm 0.02) \times 10^{-6}$	-
14	Npu_{H125N}	$CFN(NH_2)$	$(4.21 \pm 0.46) \times 10^{-2}$	$(8.96 \pm 3.22) \times 10^{-3}$	$(5.53 \pm 0.50) \times 10^{-4}$	$(4.92 \pm 0.13) \times 10^{-4}$
15 [e]	Npu_{H125N}	$CAN(NH_2)$	$(7.81 \pm 0.34) \times 10^{-2}$	$(2.94 \pm 0.01) \times 10^{-2}$	$(3.23 \pm 0.27) \times 10^{-5}$	$(3.23 \pm 0.27) \times 10^{-5}$
16	Npu_{D124Y}	$CFN(NH_2)$	$(7.75 \pm 0.55) \times 10^{-2}$	$(2.06 \pm 0.23) \times 10^{-2}$	$(3.27 \pm 0.11) \times 10^{-2}$	$(1.74 \pm 0.07) \times 10^{-2}$
17	Npu_{D124Y}	$CAN(NH_2)$	$(1.06 \pm 0.76) \times 10^{-1}$	$(3.87 \pm 0.47) \times 10^{-2}$	$(4.43 \pm 0.05) \times 10^{-4}$	$(3.61 \pm 0.21) \times 10^{-4}$

[a] k_1, k_2, and k_3 were extracted from a global fit of all three normalized curves for one reaction to the analytical solutions for the differential rate equations that describe our kinetic model. k_{splice} was extracted by fitting the product formation curve to a standard first-order rate equation. The values represent the average and standard deviation from three individually fit unique reactions.

[b] In reactions 3 and 13, the mutation of Cys_1 precludes the first steps of the splicing pathway. k_3 represents the rate of succinimide formation and thus C-extein cleavage in the absence of branch formation.

[c] In reaction 4, mutation of the catalytic asparagine abolishes succinimide formation, thus the reaction does not progress past the branched intermediate.

[d] In reaction 5, while all catalytic residues are present, no branched intermediate resolution was observed during the course of the assay.

[e] The extremely slow BI resolution in reaction 15 led to roughly 10-20% N-extein hydrolysis as a side reaction, preventing global fitting to our kinetic model. For this reaction, k_1 and k_2 were extracted from a two-state equilibrium kinetic model using only the pre-equilibrium phase of the reaction (first 10 minutes). k_3 was assumed to be identical to k_{splice}, which was determined by fitting the product formation curve to a first order rate equation.

previous measurements from gel-based assays.[9,14,16] These experiments also demonstrated that BI resolution (described by k_3) is the slow step for Ssp but the initial and latter steps of PTS are kinetically coupled for the faster Npu reaction. As additional controls, we independently mutated the first catalytic cysteine, Cys_1, and the C-terminal asparagine, Asn_{137}, in Npu to alanine and analyzed the effect of these mutations on splicing activity. As expected, the C1A mutation completely inhibited splicing, however a basal level of succinimide formation, and thus C-extein cleavage, was observed on a time scale of hours (Table 1, reaction 3). This result is consistent with the notion that C-terminal asparagine cyclization is stimulated by branched intermediate formation, as was previously shown for the GyrA intein.[27] Additionally, the N137A mutation abolished splicing and C-extein cleavage, but only modestly reduced the kinetics of the initial steps (Table 1, reaction 4).

C-extein effects on branched intermediate formation and resolution. Next, we employed our kinetic assays to determine the effect of C-extein composition on individual steps in the *trans*-splicing reaction (Table 1, reactions 5-12). These experiments revealed that C-extein variation had only a small effect on the kinetics of BI formation (k_1 and k_2), while it profoundly affected the BI resolution step (k_3) and thus the overall splicing rate (k_{splice}) (Figures 4A and B). A detailed comparison of these kinetic analyses revealed several important trends (Figure 4C). First, C-extein chain-length had a substantial effect on activity. Cys_{+1} alone could not sustain BI resolution with an uncapped carboxylate, suggesting that a negative charge near the active site is undesirable (Table 1, reaction 5). Capping the +1 residue as an amide or an ester restored a basal level of splicing activity (Table 1, reactions 6-8). Interestingly, Cys_{+1} capped with a methyl ester afforded a 4-fold rate increase over the methyl amide analog, possibly indicating an inhibitory role for this amide N-H moiety or an anomalous non-native effect of this subtle perturbation (Table 1, reactions 6 and 8). Ultimately, the effect of chain-length on BI resolution was more pronounced once the entire Phe_{+2} residue was added (Table 1, reaction 10),

however three C-extein residues were required to recapitulate the fastest reported rates for Npu (Table 1, reaction 1).

Through our kinetic analyses we also identified two specific functional groups that made major contributions to BI resolution. First, we found that the amide bond after Phe$_{+2}$ provided a 6-fold rate enhancement relative to a methyl ester (compare reactions 9 and 10). This result suggests that the amide N-H group is involved in a hydrogen bond that facilitates BI resolution, for example, by stabilizing a catalytically competent conformation. The second, more significant functional group is the Phe$_{+2}$ phenyl ring. While this residue is known to be important, as discussed above, the extent of its contribution to BI resolution was not previously known. Our measurements indicate that the addition of the bulky Phe side-chain enhances BI resolution kinetics 100-fold relative to Ala (compare reactions 1 and 12). Interestingly, the presence of the Phe side-chain also stimulated the basal rate of succiminide formation (i.e. C-extein cleavage) in the context of a C1A mutant of Npu$_N$ (Table 1, compare reactions 3 and 13), implying that the Phe side-chain is making favorable interactions even in the absence of the BI. By contrast, the side-chain of the Asn$_{+3}$ does not contribute to the *trans*-splicing reaction (compare reactions 1 and 11).

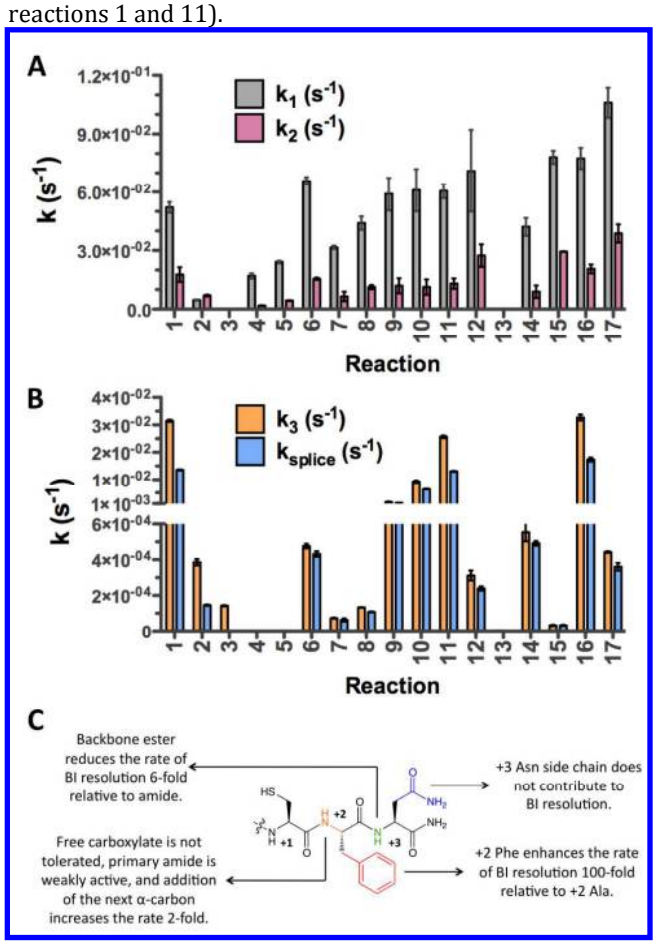


Figure 4. C-Extein contributions to splicing activity. (A) Forward (k_1) and reverse (k_2) rates of branched intermediate formation from starting materials. (B) Rate of branched intermediate resolution (k_3) and overall rate of *trans*-splicing (k_{splice}). (C) Scheme highlighting the key conclusions from the kinetic data.

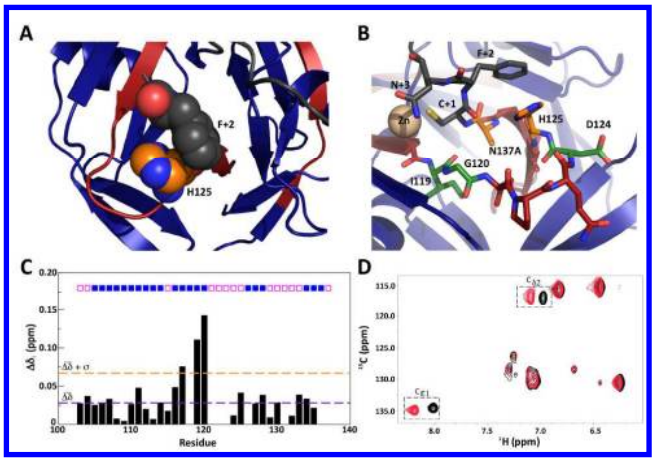


Figure 5. Structural effects of mutating the C-extein +2 residue. (A) Crystal structure of the SspDnaE intein (pdb: 1ZDE) highlighting close packing of His$_{125}$ and Phe$_{+2}$ in spheres. The N-intein and C-intein are shown as blue and red ribbons, respectively. (B) The active site of Ssp bearing catalytic Cys and Asn mutations, native extein residues, and a coordinated zinc ion. Important residues surrounding the C-intein/C-extein junction (orange/black junction) are shown as sticks. The C-extein is shown in gray, key catalytic residues are shown in orange, and other non-catalytic residues highlighted in this study are shown in green. (C) Composite ^{1}H and ^{15}N backbone chemical shift perturbations ($\Delta\delta_i$) in Npu$_C$ (^{13}C,^{15}N labeled) in complex with unlabeled Npu$_N$ as a function changing the +2 C-extein residue from Phe to Ala (see SI for calculations). The mean value is marked by a dashed purple line, and one standard deviation above the mean is marked by a dashed orange line. Residues in secondary structure elements are marked with boxes above the bars, solid blue boxes are strands and empty pink boxes are loops. (D) Overlay of the aromatic region of the ^{1}H-^{13}C-HSQC spectra of segmental labeled Npu$_N$: Npu$_C$ complexes containing either Phe (black) or Ala (red) as the +2 C-extein residue. Chemical shift perturbations of His$_{125}$ imidazole ring ^{1}H-^{13}C correlations, Cε$_1$ and Cδ$_2$, are marked in dashed boxes.

A structural role for the +2 C-extein residue. Given the significant contribution of the Phe$_{+2}$ side-chain to splicing kinetics, we next sought to understand the structural origin of its involvement in split intein chemistry. Most high resolution structures of inteins, including the only published structure of Npu,[28] do not contain C-extein residues. One important exception to this is a crystal structure of Ssp bearing five native N-extein residues (KFAEY), three native C-extein residues (CFN), and mutations of the terminal intein residues, Cys and Asn, to Ala.[29] In this structure, the Phe$_{+2}$ side-chain packs against a catalytic histidine that lies on a flexible loop (Figure 5A). This histidine (His$_{125}$ in Npu) is completely conserved in the DnaE family and has been implicated as a general acid or base in the BI resolution step of many inteins.[27,29] Mutation of His$_{125}$ in Npu to an Asn reduced the rate of BI resolution roughly 60-fold, similar to the F+2A mutation (Table 1, reactions 14 and 12, respectively). The Ssp structure suggests that Phe$_{+2}$ participates in protein *trans*-splicing by stabilizing His$_{125}$ through a direct interaction. Indeed, the effect of mutating both residues in Npu on BI resolution kinetics was non-additive ($\Delta\Delta G_{coupling}$ = 1.07 kcal mol^{-1}) indicating some co-operativity between Phe$_{+2}$ and His$_{125}$ with respect

to this step (Table 1, reaction 15, and Figure S14 for thermodynamic cycle analysis).

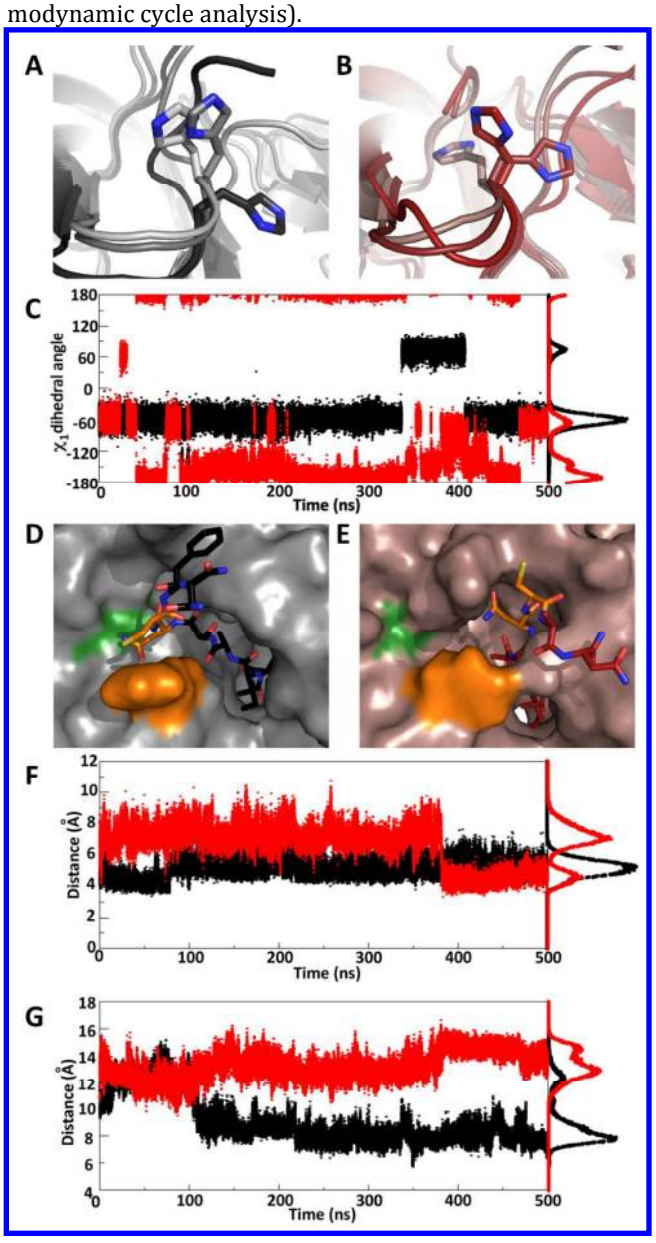


Figure 6. Molecular dynamics simulations to probe +2 amino acid-dependent active site dynamics. Three representative frames from the MD trajectories of (A) AEY-Npu$_N$ + Npu$_C$-CFN(NH_2) and (B) AEY-Npu$_N$ + Npu$_C$-CAN(NH_2) highlighting His$_{125}$ rotameric states. (C) Trajectory of the His$_{125}$ side-chain dihedral angle (χ_1) during the simulations. Representative frames highlighting the positioning of Asn$_{137}$ relative to the His$_{125}$ loop in (D) the CFN simulation and (E) the CAN simulation. Asn$_{137}$ and Cys$_{+1}$ are shown as orange sticks, His$_{125}$ is shown as an orange surface, and Ile$_{119}$ and Gly$_{120}$ are shown as green surfaces. (F) Distance between His$_{125}$ and Asn$_{137}$ Cβ atoms during the simulations. (G) Distance between Ile$_{119}$ and Cys$_{+1}$ amide nitrogens during the simulations. Data from the simulation with the CFN(NH_2) C-extein are shown in black, and analogous data from the simulation with the CAN(NH_2) C-extein are shown in red. Traces to the right of trajectory graphs are histograms indicating the distribution of angles or distances sampled throughout the simulation.

To better understand the structural impact of the +2 residue, we carried out solution NMR analyses of Npu in both a CFN(NH_2) or CAN(NH_2) C-extein context. NMR constructs were prepared analogously to those used for kinetic assays with some additional provisions. Specifically, the Npu$_N$ protein contained the native N-extein sequence (AEY) and an inactivating C1A mutation, but was not ^{13}C or ^{15}N isotopically labeled. The Npu$_C$ constructs, bearing the N137A mutation, were ^{13}C and ^{15}N enriched in the recombinant Int$_C$ portion, but not in the synthetic C-extein region. N- and C-inteins were mixed, and the complexes were purified to homogeneity by size exclusion chromatography (Figures S15 and S16). Use of this segmental labeling scheme meant that only the Npu$_C$ residues (Ile$_{103}$-Ala$_{137}$), which have identical chemical composition in both complexes, would be visible in heteronuclear correlation experiments. This was expected to simplify assignment whilst still allowing the putative interaction between the +2 residue and the catalytic His$_{125}$ to be interrogated. The inactivating mutations (C1A and N137A) ensured that chemistry would not occur during data acquisition.

With the exception of several residues in the loop containing the catalytic His$_{125}$ residue, we were able to assign the majority of the Npu$_C$ backbone resonances in the complexes using standard triple-resonance experiments (Figure S17). Most of the backbone resonances were unperturbed upon changing the +2 C-extein residue from Phe to Ala (Figure 5C). The only exceptions to this were the amide resonances from Ile$_{119}$ and Gly$_{120}$, which showed a modest perturbation. These residues are located at the beginning of the loop containing the catalytic His residue and, in the Ssp crystal structure, lie close to the C-intein/C-extein peptide bond that is ultimately attacked during branched intermediate resolution (Figure 5B). The His$_{125}$ backbone amide resonance was not itself sensitive to the nature of the +2 C-extein residue. However, the aromatic side-chain protons of this residue did exhibit significant chemical shift perturbations on mutating the +2 residue, suggesting an altered chemical environment for this side-chain in the absence of the +2 phenyl ring (Figure 5D). Together with our mutagenesis and kinetic data, these NMR studies lend support to the idea that the active site conformation of Npu is coupled to the identity of the C-extein +2 residue.

The Phe$_{+2}$ C-extein residue constrains active site motions. In order to gain additional insight into the interplay between C-extein residues and the Npu active site, we carried out molecular dynamics (MD) simulations of two wild-type intein complexes bearing either CFN(NH_2) or CAN(NH_2) as C-exteins (identical to the constructs in Table 1, reactions 1 and 12). Simulations were carried out in explicit solvent in 1 fs steps for 0.5 μs. Comparison of the two simulation trajectories afforded a more detailed picture of the coupling between the +2 residue and the intein active site. One of the more striking results from the simulation was the effect of changing +2 C-extein on the dynamics of the His$_{125}$ side-chain. In the presence of Phe$_{+2}$ the His side-chain primarily adopts a single rotameric state with only a briefly excursion to an alternate rotamer (Figures 6A and C, black trajectory). By contrast, with an Ala$_{+2}$ residue, His$_{125}$ frequently switches between three side-chain rotamers and favors a different conformation than the one

found with Phe$_{+2}$ (Figures 6B and C, red trajectory). Interestingly, the backbone ϕ and φ dihedral angles for His$_{125}$ show virtually no change as a function of C-extein composition (Figure S19). These data are consistent with the fact that there were chemical shift perturbations for the His$_{125}$ side-chain but not the backbone.

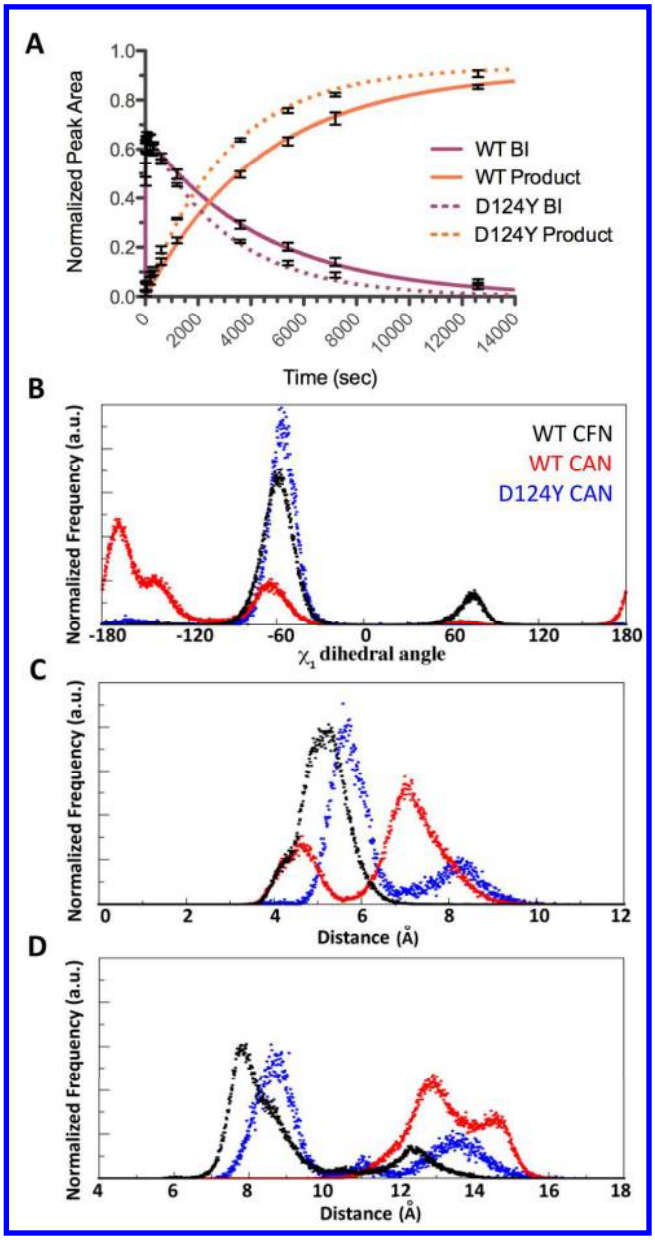


Figure 7. Structural and functional effect of the D124Y mutation in Npu. (A) Kinetic data showing rate enhancement for the BI resolution step with the D124Y mutation in the CAN(NH_2) C-extein context. Solid lines correspond to best fit kinetic curves for the wild-type intein (Table 1, reaction 12), and dotted lines correspond to the D124Y mutant (Table 1, reaction 17). Only the BI and product reaction curves are shown, and the starting material curve is omitted for clarity. Error bars represent the standard deviation from three independent reactions. Histograms showing the distribution of (B) His$_{125}$ χ_1 dihedral angles, (C) distances between His$_{125}$ and Asn$_{137}$ Cβ atoms, and (D) distances between Ile$_{119}$ and Cys$_{+1}$ amide nitrogens during the wild-type CFN (black), wild-type CAN (red) and D124Y CAN (blue) simulations.

The second major consequence of the +2 residue mutation was the overall positioning of the C-intein/C-extein junction (i.e Asn$_{137}$-Cys$_{+1}$) relative to the His$_{125}$ loop. In the simulation with CFN as the C-extein, Asn$_{137}$ remained buried in the groove above this loop, similar to the Ssp structure (Figure 6D). By contrast, in the CAN simulation, the entire strand bearing Asn$_{137}$ and the C-extein occupied space outside of this groove region (Figures 6E). An important consequence of this is that the distance between Asn$_{137}$ and His$_{125}$ (Figure 6F) and between Ile$_{119}$ and the scissile peptide bond (Figure 6G) were significantly shorter for the majority of the CFN simulation than for the CAN simulation. Overall, these MD simulations indicate that the presence of a sterically bulky amino acid at the +2 position in the C-extein acts to constrain the motions of key catalytic residues leading to a more compacted arrangement around the scissile peptide bond.

In considering the mechanistic implications of these observations it is important to emphasize that the simulations employed, by necessity, a linear precursor protein as the starting point. Use of a BI structure in the simulations would have been more desirable given that our kinetic data reveal that formation of this intermediate stimulates cleavage of the peptide bond at the C-intein/C-extein junction (Table 1, compare reactions 1 and 3). Unfortunately, there is currently no high-resolution structural information on any intein in the branched intermediate state. Thus, we were forced to extrapolate from the structures available. Despite this caveat, the major conclusion from the simulation work is broadly consistent with our mutagenesis and kinetic data. In particular, we observe coupling between the +2 residue and the catalytic His$_{125}$ both in the simulations and in the kinetics of BI resolution. We further note that the Phe side-chain stimulates C-extein cleavage even in the absence of the BI (Table 1, compare reactions 3 and 13), arguing that this bulky side-chain augments catalysis even in the linear precursor.

An activating point mutation on the His$_{125}$ loop. Local C-extein residues appear to affect the structure and dynamics of residues surrounding the flexible His$_{125}$ loop, thereby modulating BI resolution kinetics. Thus, it is conceivable that point mutations within the intein that alter loop conformation or flexibility could also modulate splicing activity and even tolerance to non-native extein residues. In a previous directed evolution study on an Npu$_N$-Ssp$_C$ chimera, we identified several mutations that make this intein more tolerant of the C-extein sequence SGV, rather than CFN.[7] Intriguingly, one of these mutations was an Asp-to-Tyr mutation adjacent to His$_{125}$ (Asp$_{124}$). We found that this mutation enhanced the rate of Npu splicing by 50% in the presence of Ala$_{+2}$ (Figure 7A, compare reactions 12 and 17). Importantly, this mutation was still tolerated when Phe$_{+2}$ was present, suggesting that it increases overall promiscuity towards C-exteins (compare reactions 1 and 16). The Npu NMR structure[28] and the Ssp crystal structures[29,30] indicate that Asp$_{124}$ packs against a β-turn from the N-intein. Given this close packing, the bulky D124Y mutation would require conformational rearrangement and possibly also rigidification of the catalytic His$_{125}$ loop, which can modulate activity. As predicted, in a 100 ns MD simulation of Npu$_{D124Y}$ with a CAN(NH_2) C-

extein, the His_{125} loop conformation was altered, His_{125} rotamer dynamics were constrained, and Asn_{137} persistently remained above the His_{125} loop, similar to the Npu_{WT}-CFN(NH_2) simulation (Figures 7B-D, S21, and S22). This simulation suggests that the D124Y mutation reduces C-extein dependence by recapitulating the constraints on active site dynamics typically applied by Phe_{+2}, specifically the stabilization of His_{125} and the appropriate positioning of the C-intein/C-extein junction close to His_{125}.

DISCUSSION AND CONCLUSIONS

In this report, we examined the molecular determinants for C-extein-dependent protein *trans*-splicing. This investigation was facilitated by the utilization of protein semisynthesis to generate inteins linked to a variety of C-exteins and by the development of novel kinetic assays that provide information about individual steps along the *trans*-splicing reaction coordinate. Through these studies, we not only extracted information on C-extein requirements, but also gained additional mechanistic insights into split DnaE intein splicing. Specifically, our experiments confirmed that branched intermediate resolution is the slowest step for PTS (k_3). They also provided evidence supporting the notion that some DnaE inteins have a highly activated N-terminal splice junction ($k_1/k_2 > 2$ for all Npu constructs), consistent with our previous report.[9] Interestingly, this N-terminal activation appears roughly ten-fold slower and is significantly less efficient ($k_1/k_2 = 0.67$) for the Ssp intein. Additionally, we found that for Npu, the rate of Asn cyclization upon BI formation is 200-fold faster than its rate in the absence of the branched structure. Stimulation of Asn cyclization upon BI formation is also found in the *cis*-splicing GyrA intein.[27] We propose that this kinetic stimulation is a common feature of inteins, in effect creating a trigger that helps ensure the proper fidelity of the reaction by minimizing premature cleavage of the C-extein. Lastly, it is particularly surprising that the H125N mutation does not completely abolish BI resolution, but rather reduces its rate 60-fold. Indeed, the splicing rate of this mutant is still faster than for wild-type Ssp. For many non-DnaE inteins, this step requires two histidine residues, one analogous to His_{125} and another immediately preceding the C-terminal Asn residue.[27,31] Given the lack of this penultimate histidine in the DnaE inteins, His_{125} has been implicated as the sole general acid/base for BI resolution.[29] Our data suggest that while His_{125} is clearly important for BI resolution, other unidentified residues must also contribute to catalysis of this step.

The current study improves our understanding of the relationship between C-extein composition and *trans*-splicing efficiency. The kinetic data indicate that the C-extein almost exclusively affects the BI resolution step. Within the C-extein, we identified specific functional groups that contribute significantly to splicing kinetics, in particular the Phe_{+2} side-chain. Our NMR experiments and MD simulations illustrate that this bulky functional group constrains active site motions, forcing catalytic histidine and asparagine residues and the scissile peptide bond in close proximity. The need for a bulky side-chain at the +2 position is further highlighted by a recent genetic selection study on the Npu intein showing that Trp is also well tolerated at this position.[10] Collectively, these data paint a picture of the Npu active site that effectively extends beyond the intein domain itself to include the +2 C-extein residue.

During protein *trans*-splicing, the N-extein is transferred from the N-terminus of the intein onto a C-extein side-chain, thereby created a unique branched protein structure. As BI resolution is the slowest and often rate limiting step for many inteins, this structure is most relevant to overall activity. To date, all published high-resolution structural data on inteins examine either a precursor or product form of the intein. While these studies, including this report, have provided substantial insights into the structural basis for protein splicing, they cannot examine interactions that are exclusively present in the branched intermediate. Indeed, our kinetic analyses revealed several important functional groups in the C-extein that affect BI resolution (Figure 4C), however only in the case of the Phe_{+2} side-chain could we postulate any kind of structural basis of this. Thus, these results reinforce the need for high-resolution structural information on the branched intermediate in the protein splicing reaction.

The fullest deployment of split inteins in protein engineering ultimately requires a truly traceless *trans*-splicing system with no sequence requirements. While bulky hydrophobic residues other than phenylalanine are tolerated at the critical +2 position for DnaE inteins, thus alleviating some sequence constraints,[5,10,25] these inteins are still only modestly promiscuous. Our results suggest that the interplay between the C-extein and the His_{125} active site loop has direct implications for the rational design of improved, more extein-tolerant split inteins. Indeed, the D124Y point mutation on this flexible loop increases the tolerance of Npu for a +2 alanine residue without affecting its activity in a native context. In a recent directed evolution endeavor on a DnaB family intein, a mutation at this position was also found to reduce C-extein sequence constraints.[32] Furthermore, we previously demonstrated that mutating other residues on this loop can generally enhance the activity of Ssp[9] and the Npu_N-Ssp_C chimera[7] in a native C-extein context. These results collectively indicate that the conformational preferences of this loop are intimately linked with inadequate BI resolution both for intrinsically slow inteins and for efficient inteins in an exogenous C-extein context. Thus, this loop is a hot-spot on the intein structure that should be explicitly targeted in future engineering efforts for the design of more high-activity, broad-specificity inteins.

EXPERIMENTAL SECTION

Semisynthesis of C-intein constructs. Semisynthetic Int_C-extein proteins were generated through Expressed Protein Ligation of a synthetic fragment, corresponding to the desired model C-extein, and a reactive recombinant fragment corresponding to the C-intein. Model C-exteins were synthesized using standard solution based or solid phase protocols (see Supporting Information for details). Reactive recombinant Int_C polypeptides were derived from the corresponding Int_C-GyrA-H_6 fusion proteins, which were expressed in *E. coli* and purified using standard methods (Figure S4). Ligation reactions involved treatment of the purified Int_C-GyrA-H_6 fusion with an excess of the model C-extein usually in the presence of an additional

thiol. Semisynthetic products were purified by preparative RP-HPLC and characterize by ESI-MS (Figure S7 and Table S2).

Expression and purification of N-Intein constructs. AEY-Npu_N and AEY-Ssp_N were expressed with an N-terminal His_6-SUMO tag in *E. coli* BL21(DE3) cells from an IPTG-inducible protein expression vector. The cells were lysed by sonication and the protein was enriched over Ni-NTA resin in a pH 8.0 phosphate buffered saline solution. The proteins were eluted from the Ni column in the presence of 250 mM imidazole (Figure S5), and the elutions were dialyzed to reduce the imidazole concentration to 5 mM. The dialyzed solutions were treated for 12 hours at room temperature with His_6-tagged Ulp1, a SUMO-specific protease, to yield the desired products. The proteolysis reactions were passed over Ni-NTA resin to deplete unreacted starting material, the cleaved His_6-SUMO tag, and Ulp1 (Figure S6). The proteins were further purified by size exclusion chromatography on a Superdex 75 column in splicing assay buffer (100 mM sodium phosphates, 150 mM NaCl, 1 mM EDTA, pH 7.2) supplemented with 1 mM DTT. Product identities were confirmed by ESI-MS, and their purities were assessed by analytical RP-HPLC (Figure S7 and Table S2).

RP-HPLC and ESI-MS analysis of splicing assays. Prior to any splicing assay, the N-intein solutions were dialyzed against splicing assay buffer (100 mM sodium phosphates, 150 mM NaCl, 1 mM EDTA, pH 7.2) overnight at 4°C. Note, thiols were omitted from this buffer since substantial N-extein cleavage was observed for reactions with a slow k_3. N-inteins and C-inteins were diluted to 15 μM and 10 μM, respectively, and TCEP added to each solution to a final concentration of 2 mM. Splicing reactions were initiated by mixing equal volumes of N- and C-inteins at 30 °C. During the reaction, aliquots of the solution were removed and mixed 3:1 (v/v) with quenching solution (8 M guanidine hydrochloride and 4% trifluoroacetic acid). For RP-HPLC analysis, 100 μL of the quenched solutions were separated over a C_{18} analytical column, recording absorbance at 214 nm, and major peaks were collected and identified by ESI-MS (Figures 3A and S9 and Table S3). For direct ESI-MS analyses, 20 μL of the quenched solutions were desalted using Millipore C_{18} Zip-Tips, diluted, and loaded on the mass spectrometer by direct infusion. The complex mixture of multiply-charged states of each species were deconvoluted into spectra depicting a well-defined mixture of singly-charged species (Figure 3B and S10 and Table S3).

Kinetic analysis. Peaks corresponding to species **1**-**5** in either the RP-HPLC chromatogram or ESI-MS spectrum were integrated and expressed as a fraction of total peak intensity for each time point. For the RP-HPLC analyses, the product was expressed as the sum of the integrated intensities for species **3**-**5** to account for changes in relative extinction coefficients. For ESI-MS analyses, the product was expressed as the sum of the integrated intensities of only species **3** and **4**, since species **5** was not visible and the ionizability of **1**-**4** was assumed to be identical. The time-dependent reaction curves for all three states of the reaction (Figures S11 and S12), starting material (**1**), branched intermediate (**2**), and products (**3**-**4** or **3**-**5**), were collectively fit to the analytical solution for the coupled differential equations describing our kinetic model (Figure 3C and Supporting Information). From this global fit, we extracted the values for k_1, k_2, and k_3 for each individual reaction. k_{splice} splice was determined by fitting the product formation curves (**3**-**4** or **3**-**5**) to a first-order rate equation. Reactions were repeated three or four times, and the average and standard deviation of all four kinetic parameters are reported in Table 1.

NMR spectroscopy. NMR experiments were carried out on uniformly ^{15}N,^{13}C-labeled Npu_C^{N137A} ligated to unlabeled C-exteins (CFN(NH_2) or CAN(NH_2)) in complex with unlabeled AEY-Npu_N^{C1A}. Experiments were run on either 600MHz (Bruker or Varian Inova), 800MHz and 900MHz Bruker spectrometers. Backbone resonance assignments of labeled Npu_C in complex with Npu_N were achieved using triple resonance experiments with standard pulse sequences.[33] The complex harbors one histidine (His_{125}). The side-chain carbons, $C\delta_2$ and $C\varepsilon_1$, of His_{125} were resolved with a standard ^{13}C,^{1}H aromatic HSQC experiment.[34-36] Standard pulse sequences were used for the measurements of R_1, R_2 and ^{15}N-^{1}H *NOE* rates.

Molecular dynamics simulations. All-atom molecular dynamics simulations were performed on Npu constructs at constant temperature and pressure (300 K and 1 atm) using the molecular dynamics suit AMBER11.[37,38] Simulations contained explicit water molecules and the net charge of the system was neutralized with sodium ions. The constructs were generated from the first representative solution NMR structure of Npu (PDB 2KEQ).[28] Prior to the simulations, this structure was modified *in silico* using UCSF Chimera[39] to generate the constructs of interest, namely, (1) a wild-type split intein complex with canonical extein sequences (AEY-Npu_N : Npu_C-CFN(NH_2)), (2) a wild-type split intein complex with a mutant C-extein (AEY-Npu_N : Npu_C-CAN(NH_2)), and (3) a D124Y mutant with the same mutant C-extein sequence (AEY-Npu_N : Npu_C^{D124Y}-CAN(NH_2)). 500 ns long simulations were run for the wild-type CFN and CAN constructs, and a 100 ns long simulation was run for the D124Y mutant. Prior to the runs, a series of minimization, heating and density equilibration steps were performed.

ASSOCIATED CONTENT

Supporting Information

Full methods and experimental data including protein semisynthesis and purification protocols, characterization of proteins, and details of kinetic analyses, NMR experiments, and MD simulations. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

* muir@princeton.edu

ACKNOWLEDGMENTS

The authors thank the members of the Muir laboratory for valuable discussions. This work was supported by the U.S. National Institutes of Health (NIH grant GM086868). The program Chimera was supported by NIGMS P41-GM103311. NMR resources at NYSBC were supported by NIGMS P41-GM066354.

REFERENCES

(1) Volkmann, G.; Mootz, H. D. Cell. Mol. Life Sci. 2012.

(2) Shah, N. H.; Muir, T. W. Israel J. Chem. 2011, 51, 854-861.

(3) Shi, J.; Muir, T. W. J. Am. Chem. Soc. 2005, 127, 6198-206.

(4) Southworth, M.; Amaya, K.; Evans, T.; Xu, M.; Perler, F. Biotechniques 1999, 27, 110-120.

(5) Iwai, H.; Züger, S.; Jin, J.; Tam, P.-H. FEBS Lett. 2006, 580, 1853-8.

(6) Amitai, G.; Callahan, B. P.; Stanger, M. J.; Belfort, G.; Belfort, M. Proc. Natl. Acad. Sci. USA 2009, 106, 11005-10.

(7) Lockless, S. W.; Muir, T. W. Proc. Natl. Acad. Sci. USA 2009, 106, 10999-11004.

(8) Shemella, P. T.; Topilina, N. I.; Soga, I.; Pereira, B.; Belfort, G.; Belfort, M.; Nayak, S. K. Biophys. J. 2011, 100, 2217-25.

(9) Shah, N. H.; Dann, G. P.; Vila-Perelló, M.; Liu, Z.; Muir, T. W. J. Am. Chem. Soc. 2012, 134, 11338-41.

(10) Cheriyan, M.; Pedamallu, C. S.; Tori, K.; Perler, F. J. Biol. Chem. 2013, 288, 6202-11.

(11) Pietrokovski, S. Trends Genet. 2001, 17, 465-72.

(12) Vila-Perelló, M.; Muir, T. W. Cell 2010, 143, 191-200.

(13) Caspi, J.; Amitai, G.; Belenkiy, O.; Pietrokovski, S. Mol. Microbiol. 2003, 50, 1569-77.

(14) Martin, D. D.; Xu, M. Q.; Evans, T. C. Biochemistry 2001, 40, 1393-402.

(15) Dassa, B.; Amitai, G.; Caspi, J.; Schueler-Furman, O.; Pietrokovski, S. Biochemistry 2007, 46, 322-330.

(16) Zettler, J.; Schütz, V.; Mootz, H. D. FEBS Lett. 2009, 583, 909-14.

(17) Dhar, T.; Mootz, H. D. Chem. Commun. 2011, 47, 3063-5.

(18) Borra, R.; Dong, D.; Elnagar, A. Y.; Woldemariam, G. A.; Camarero, J. A. J. Am. Chem. Soc. 2012, 134, 6344-53.

(19) Vila-Perelló, M.; Liu, Z.; Shah, N. H.; Willis, J. A.; Idoyaga, J.; Muir, T. W. J. Am. Chem. Soc. 2013, 135, 286-292.

(20) Busche, A. E. L.; Aranko, A. S.; Talebzadeh-Farooji, M.; Bernhard, F.; Dötsch, V.; Iwaï, H. Angew. Chem. Int. Ed. 2009, 48, 6128-31.

(21) Muona, M.; Aranko, A. S.; Raulinaitis, V.; Iwaï, H. Nat. Protoc. 2010, 5, 574-87.

(22) Jagadish, K.; Borra, R.; Lacey, V.; Majumder, S.; Shekhtman, A.; Wang, L.; Camarero, J. A. Angew. Chem. Int. Ed. 2013, 52, 3126-31.

(23) Zhang, Y.; Yang, W.; Chen, L.; Shi, Y.; Li, G.; Zhou, N. Anal. Biochem. 2011, 417, 65-72.

(24) Wong, S.; Mills, E.; Truong, K. Protein Eng. Des. Sel. 2012, 26, 207-13.

(25) Shah, N. H.; Vila-Perelló, M.; Muir, T. W. Angew. Chem. Int. Ed. 2011, 50, 6511-5.

(26) Muir, T. W.; Sondhi, D.; Cole, P. A. Proc. Natl. Acad. Sci. USA 1998, 95, 6705-10.

(27) Frutos, S.; Goger, M.; Giovani, B.; Cowburn, D.; Muir, T. W. Nat. Chem. Biol. 2010, 6, 527.

(28) Oeemig, J. S.; Aranko, A. S.; Djupsjöbacka, J.; Heinämäki, K.; Iwaï, H. FEBS Lett. 2009, 583, 1451-1456.

(29) Sun, P.; Ye, S.; Ferrandon, S.; Evans, T. C.; Xu, M.-Q.; Rao, Z. J. Mol. Biol. 2005, 353, 1093-105.

(30) Callahan, B. P.; Topilina, N. I.; Stanger, M. J.; Van Roey, P.; Belfort, M. Nat. Struct. Mol. Biol. 2011.

(31) Chen, L.; Benner, J.; Perler, F. B. J. Biol. Chem. 2000, 275, 20431-5.

(32) Appleby-Tagoe, J. H.; Thiel, I. V.; Wang, Y.; Wang, Y.; Mootz, H. D.; Liu, X.-Q. J. Biol. Chem. 2011, 286, 34440-7.

(33) Sattler, M.; Schleucher, J.; Griesinger, C. Prog. Nucl. Mag. Res. Sp. 1999, 34, 93-158.

(34) Palmer, A. G.; Cavanagh, J.; Wright, P. E.; Rance, M. J. Mag. Reson. 1991, 93, 151-170.

(35) Kay, L. E.; Keifer, P.; Saarinen, T. J. Am. Chem. Soc. 1992, 114, 10663-10665.

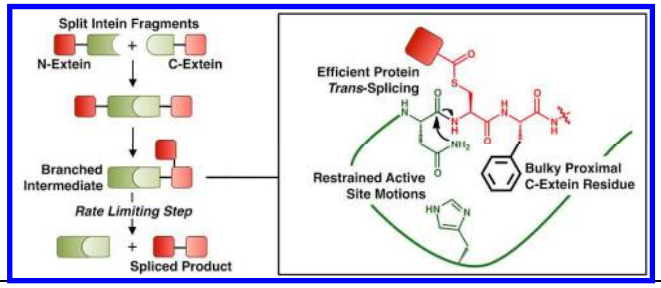
(36) Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletzky, O.; Glaser, S. J.; Sørensen, O. W.; Griesinger, C. J. Biomol. NMR 1994, 4, 301-6.

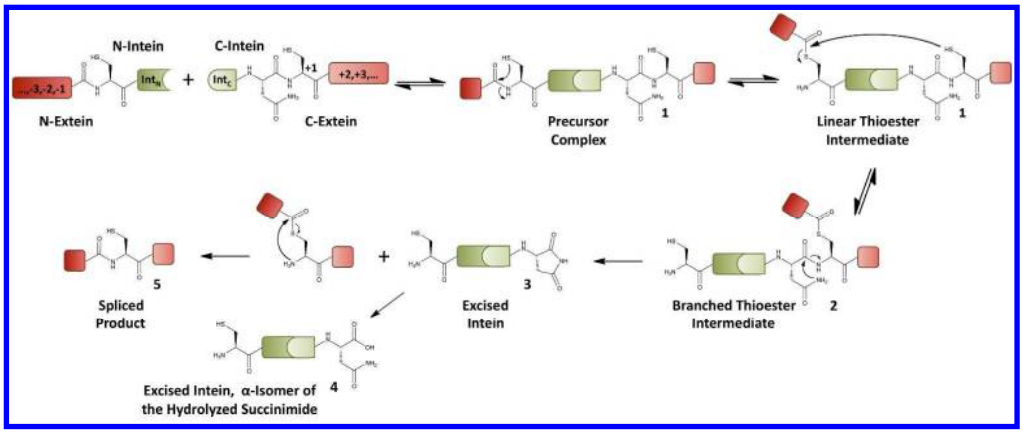
(37) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. AMBER 12. 2012, University of California, San Francisco.

(38) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. J. Chem. Theory. Comput. 2012, 8, 1542-1555.

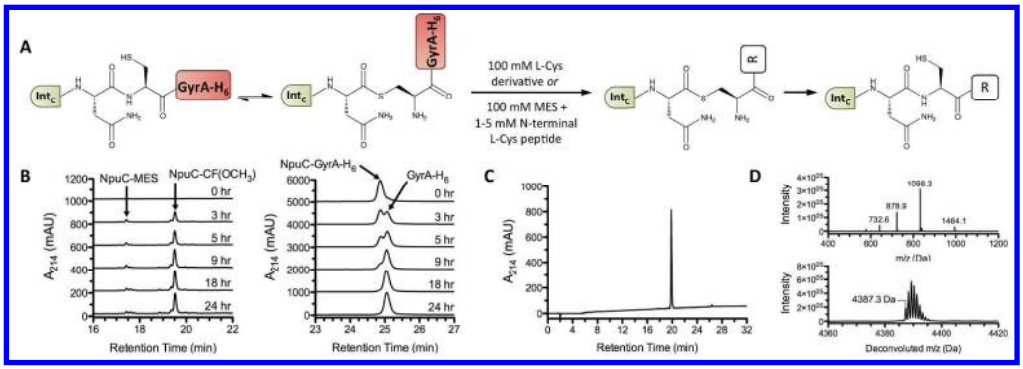
(39) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. J. Comput. Chem. 2004, 25, 1605-12.

Table of Contents Graphic

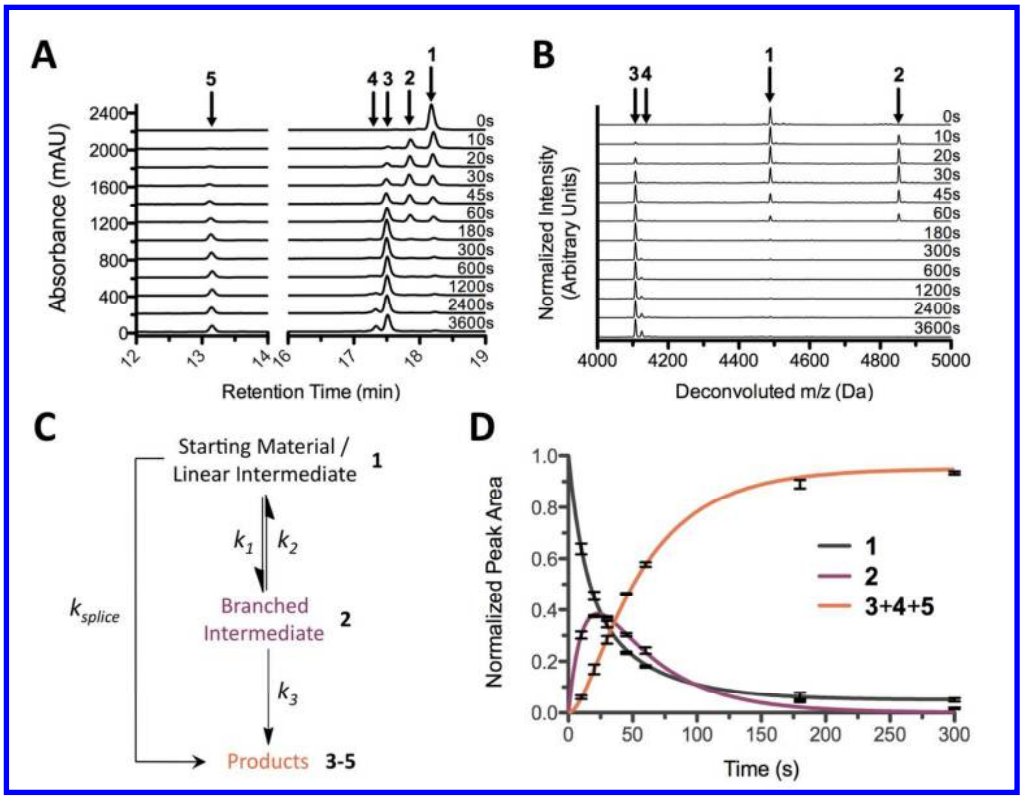




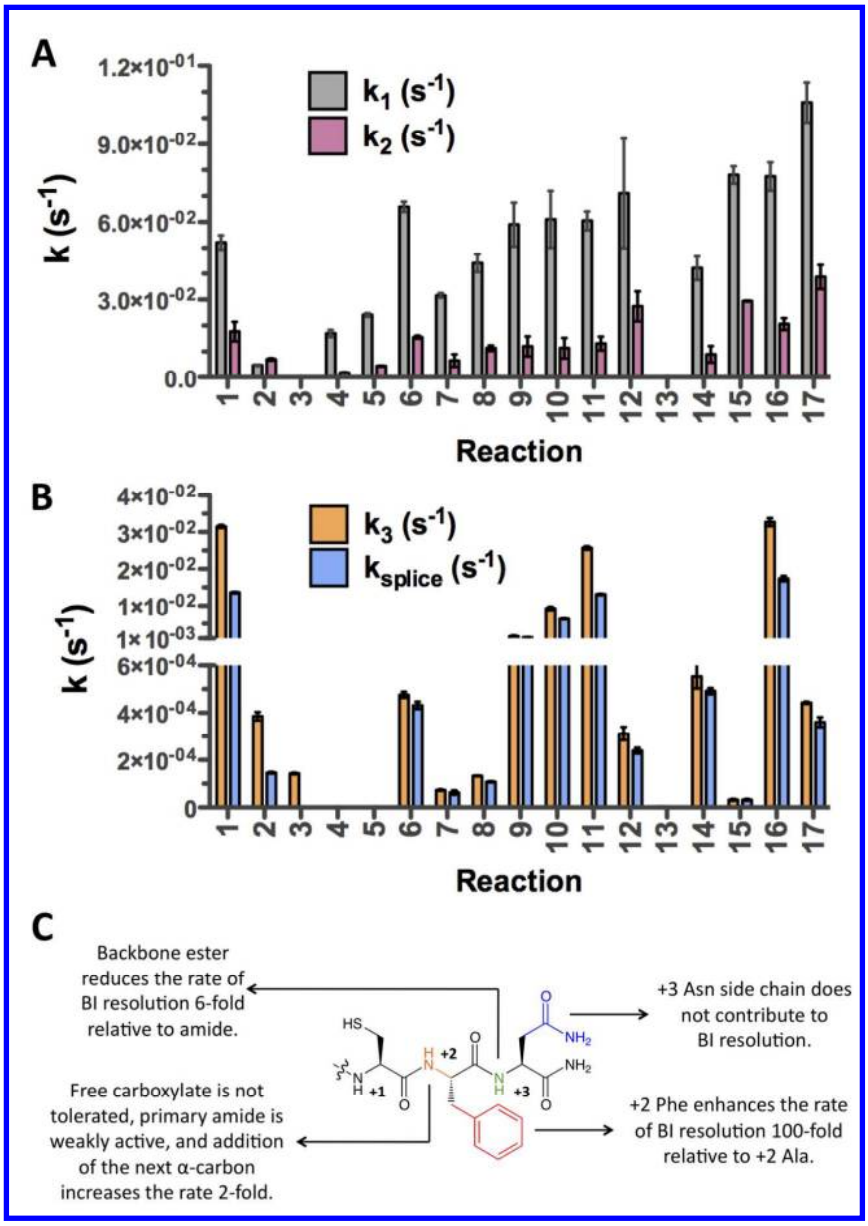
345x141mm (300 x 300 DPI)



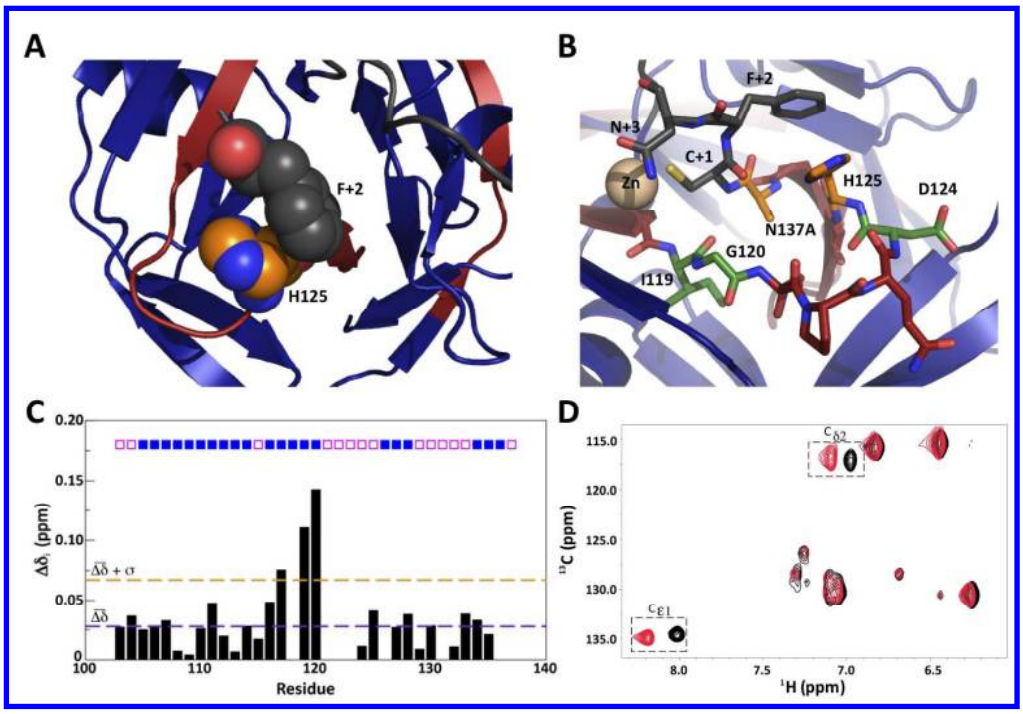
322x111mm (300 x 300 DPI)



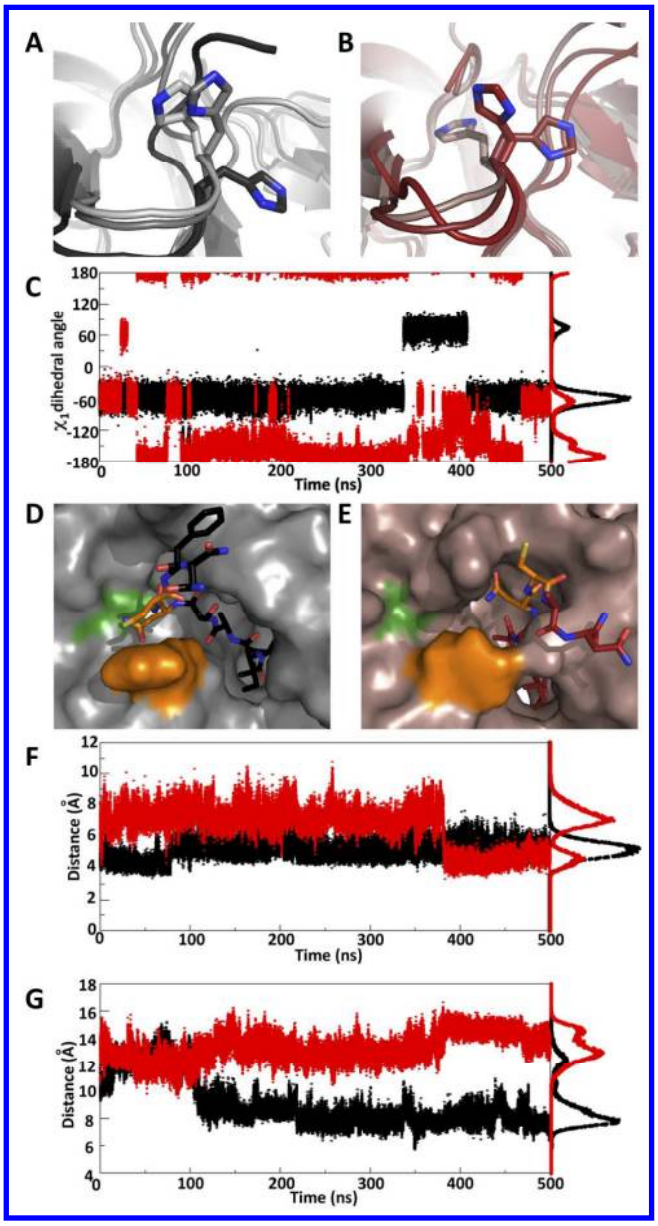
145x112mm (300 x 300 DPI)



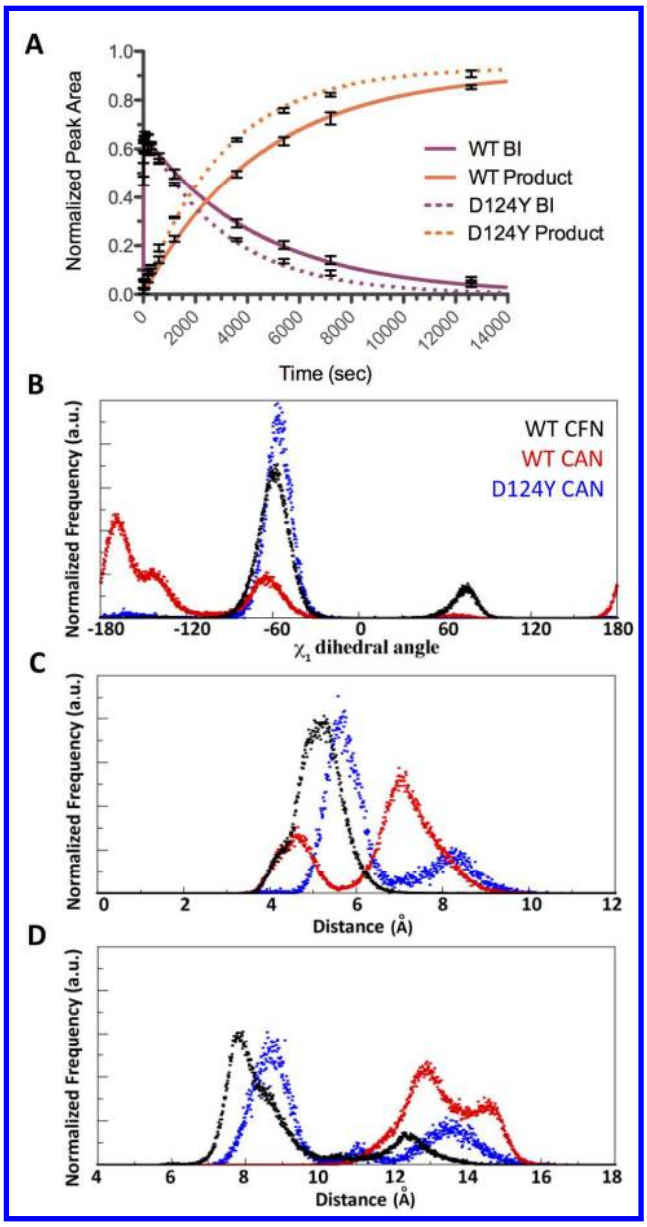
126x178mm (300 x 300 DPI)



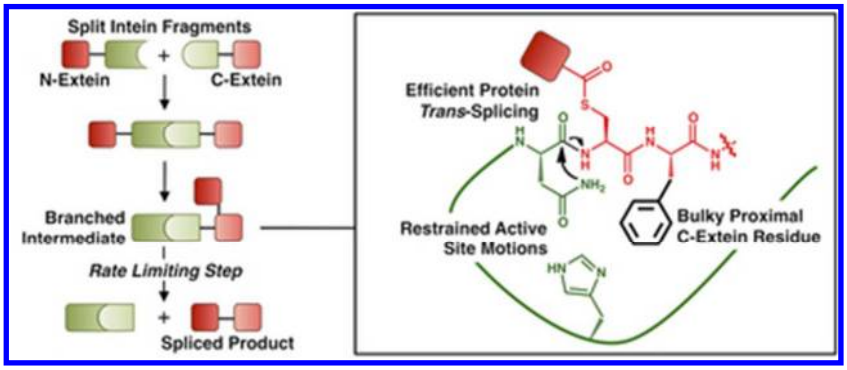
231x159mm (300 x 300 DPI)



131x248mm (300 x 300 DPI)



125x240mm (300 x 300 DPI)



34x14mm (300 x 300 DPI)