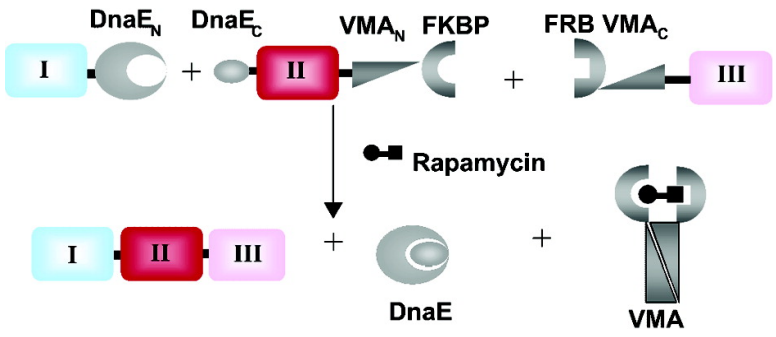
Article

Development of a Tandem Protein Trans-Splicing System Based on Native and Engineered Split Inteins

Jianxin Shi, and Tom W. Muir

J. Am. Chem. Soc., **2005**, 127 (17), 6198-6206• DOI: 10.1021/ja042287w • Publication Date (Web): 08 April 2005



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 6 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 04/08/2005

Development of a Tandem Protein Trans-Splicing System Based on Native and Engineered Split Inteins

Jianxin Shi and Tom W. Muir*

Contribution from the Laboratory of Synthetic Protein Chemistry, The Rockefeller University, 1230 York Avenue, New York, New York 10021

Received December 22, 2004; E-mail: muirt@rockefeller.edu

Abstract: Protein trans-splicing involving naturally or artificially split inteins results in two polypeptides being linked together by a peptide bond. While this phenomenon has found a variety of applications in chemical biology and biotechnology, precious little is known about the molecular recognition events governing the initial fragment association step. In this study, fluorescence approaches have been used to measure the dissociation constant for the *Ssp* DnaE split intein interaction and to determine the on and off rates of fragment association. The DnaE fragments bind with low nanomolar affinity, and our data suggest that electrostatics make an important contribution to the very rapid association of the fragments at physiological pH. This information was used to develop a tandem trans-splicing system based on native and engineered split inteins. This novel system allows the one-pot assembly of three polypeptides under native conditions and can be performed in crude cell lysates. The technology should provide a convenient approach to the segmental isotopic or fluorogenic labeling of specific domains within the context of large multidomain proteins.

Introduction

Protein splicing is an autocatalytic, post-translational process in which an intervening sequence, termed an intein, is excised from a precursor protein with subsequent ligation of the two flanking proteins, the exteins, via a native peptide bond.[1] Inteins have been identified in all three branches of life,[2] suggesting protein splicing has an ancient evolutionary origin. Moreover, a growing number of proteins have been shown to contain autoprocessing domains that are orthologous to inteins at the structural and/or mechanistic levels.[3−7] Although inteins have not been found in multicellular organisms, recent studies indicate that protein ligation can also occur by a proteasome-dependent trans-peptidation mechanism as part of antigen presentation in immunity.[8,9]

Several groups independently demonstrated that inteins can be cut into two pieces that individually have no activity, but when reconstituted will noncovalently associate to give a functional protein (Figure 1).[10−13] Protein trans-splicing, as this process is known, has found a variety of applications including in vitro protein semisynthesis,[14] segmental isotopic labeling,[13] and two-[15] and three-hybrid[16] strategies for monitoring and controlling protein activity in vivo. The identification of a naturally split intein, *Synechocytis* sp. *(Ssp)* DnaE,[17,18] has further expanded the application of trans-splicing to include protein cyclization,[19,20] in vivo protein semisynthesis,[21] plant biotechnology,[22] and cellular proteomics.[23] In addition, the split *Ssp* DnaE intein has proven useful for studying the kinetics of the chemical steps in protein splicing.[24−26]

Protein trans-splicing is an example of protein fragment complementation (PFC), referring to the well-known phenomenon in which complementary fragments of a protein can

(1) Giriat, I.; Muir, T. W.; Perler, F. B. *Genet. Eng.* **2001**, *23*, 171−199.
(2) Perler, F. B. *Nucleic Acids Res.* **1999**, *27*, 346−347.
(3) Porter, J. A.; et al. *Cell* **1996**, *86*, 21−34.
(4) Xu, Q.; Buckley, D.; Guan, C.; Guo, H.-C. *Cell* **1999**, *98*, 651−661.
(5) Rosenblum, J. S.; Blobel, G. *Proc. Natl. Acd. Sci. U.S.A.* **1999**, *96*, 11370−11375.
(6) Hodel, A. E.; Hodel, M. R.; Griffis, E. R.; Hennig, K. A.; Ratner, G. A.; Xu, S.; Powers, M. A. *Mol. Cell* **2002**, *10*, 347−358.
(7) Amitai, G.; Belenkiy, O.; Dassa, B.; Shainskaya, A.; Pietrokovski, S. *Mol Microbiol.* **2003**, *47*, 61−73.
(8) Vigneron, N.; Stroobant, V.; Chapiro, J.; Ooms, A.; Degiovanni, G.; Morel, S.; van der Bruggen, P.; Boon, T.; Van den Eynde, B. J. *Science* **2004**, *304*, 587−590.
(9) Hanada, K.-I.; Yewdell, J. W.; Yang, J. C. *Nature* **2004**, *427*, 252−256.
(10) Southworth, M. W.; Adam, E.; Panne, D.; Byer, R.; Kautz, R.; Perler, F. B. *EMBO J.* **1998**, *17*, 918−926.
(11) Mills, K. V.; Lew, B. M.; Jiang, S.-Q.; Paulus, H. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 3543−3548.
(12) Wu, H.; Xu, M.-Q.; Liu, X.-Q. *Biochim. Biophys. Acta* **1998**, *1387*, 422−432.
(13) Yamazaki, T.; Otomo, T.; Oda, N.; Kyogoku, Y.; Uegaki, K.; Ito, N.; Ishino, Y.; Nakamura, H. *J. Am. Chem. Soc.* **1998**, *120*, 5591−5592.
(14) Lew, B. M.; Mills, K. V.; Paulus, H. *J. Biol. Chem.* **1998**, *273*, 15887−15890.
(15) Ozawa, T.; Nogami, S.; Sato, M.; Ohya, Y.; Umezawa, Y. *Anal. Chem.* **2000**, *72*, 5151−5157.
(16) Mootz, H. D.; Muir, T. W. *J. Am. Chem. Soc.* **2002**, *124*, 9044−9045.
(17) Gorbalenya, A. *Nucleic Acids Res.* **1998**, *26*, 1741−1748.
(18) Wu, H.; Hu, Z.; Liu, X. Q. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9226−9231.
(19) Scott, C. P.; Abel-Santos, E.; Wall, M.; Wahnon, D. C.; Benkovic, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13638−13643.
(20) Evans, T. C., Jr.; Martin, D.; Kolly, R.; Panne, D.; Sun, L.; Ghosh, I.; Chen, L.; Benner, J.; Liu, X. Q.; Xu, M. Q. *J Biol Chem* **2000**, *275*, 9091−9094.
(21) Giriat, I.; Muir, T. W. *J. Am. Chem. Soc.* **2003**, *125*, 7180−7181.
(22) Chin, H. G.; Kim, G. D.; Marin, I.; Mersha, F.; Evans, T. C., Jr.; Chen, L.; Xu, M. Q.; Pradhan, S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4510−4515.
(23) Ozawa, T.; Sako, Y.; Sato, M.; Kitamura, T.; Umezawa, Y. *Nat. Biotechnol.* **2003**, *21*, 287−293.
(24) Martin, D. D.; Xu, M. Q.; Evans, T. C., Jr. *Biochemistry* **2001**, *40*, 1393−1402.
(25) Nichols, N. M.; Benner, J. S.; Martin, D. D.; Evans, T. C., Jr. *Biochemistry* **2003**, *42*, 5301−5311.
(26) Nichols, N. M.; Evans, T. C., Jr. *Biochemistry* **2004**, *43*, 10265−10276.

10.1021/ja042287w CCC: $30.25

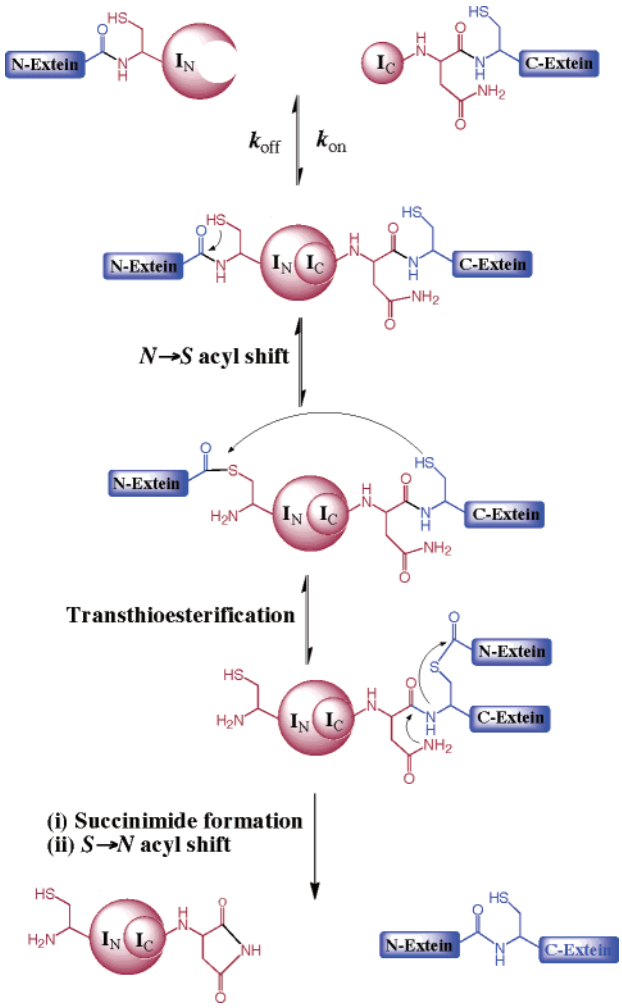


Figure 1. Protein trans-splicing. The process involves an initial binding interaction between the two split inteins, I_N and I_C, followed by a series of pseudo-intramolecular chemical steps culminating in the ligation of the N- and C-exteins and the extrusion of the split intein complex.

associate noncovalently to form a native structure.[27,28] With few exceptions, PFC is more efficient when the fragments are at high local concentration, a feature that has made these systems valuable as two-hybrid type sensors of protein−protein interactions.[29] Indeed, the split *Sce* VMA and *Ssp* DnaE inteins have both been used for this purpose.[15,30,31] The concentration-dependence is usually attributed to the fact that correct folding is favored over nonproductive processes as the effective concentration of the fragments increases. PFC has been of some interest to the protein-folding community since residual structure within the fragments might reflect local structure formed early in the folding pathway of the full-length protein.[32,33]

Despite the many applications of protein trans-splicing, essentially nothing is known about the molecular recognition events underlying the process. Artificially split inteins require a denaturation/renaturation step before splicing occurs in vitro, whereas the naturally split *Ssp* DnaE intein will support spontaneous splicing. Interestingly, the artificially split *Sce* VMA intein will support spontaneous splicing when fused to interacting protein domains.[15,16] This suggests that the difference in behavior of the artificially and naturally split inteins stems from a difference in intrinsic affinity between the complementary fragments. In this study, we have used fluorescence resonance energy transfer (FRET) measurements to measure the dissociation constant for the DnaE split intein interaction and to determine the on and off rates of fragment association. The DnaE fragments bind with low nanomolar affinity, and our data suggest that electrostatics make an important contribution to the very rapid association of the fragments at physiological pH. This information was used to develop a tandem trans-splicing system based on native and engineered split inteins. This novel system allows the one-pot assembly of three polypeptides under native conditions and should provide a convenient approach to the domain-specific incorporation of biophysical probes into large proteins.

Results and Discussion

Monitoring the Association of the *Ssp* DnaE Split Intein by FRET. It is generally believed that the two *Ssp* DnaE fragments (termed $DnaE_N$ and $DnaE_C$) interact with high affinity;[21,24,26] however, the thermodynamics and kinetics of this association have not been reported. This quantitative information would provide a starting point for understanding the molecular recognition processes underlying DnaE-mediated trans-splicing, as well as for developing additional applications of the process.

In principle, association of the $DnaE_C$ and $DnaE_N$ fragments can be monitored by changes in fluorescence resonance energy transfer (FRET) between appropriately placed fluorophores in the two polypeptides. Conceivably, the FRET probes could be conjugated to the side-chains of residues within the $DnaE_C$ and $DnaE_N$ sequences. However, the 3° structure of the complex is not known, and thus, one runs the risk of disrupting the association of the fragments through attachment of bulky fluorophores within the intein itself. A better approach would be to incorporate the probes within the N- and C-exteins, particularly since these sequences are expected to emerge from the same face of the intein complex.[34,35] Consequently, the FRET pair should be juxtaposed upon association of $DnaE_C$ and $DnaE_N$. Incorporation of the FRET pair within the exteins does lead to one major complication, namely, that they will become covalently linked following the trans-splicing reaction (Figure 1). Depending upon the on−off rates of split intein association, this could complicate determination of the dissociation constant (K_d) using steady-state fluorescence measurements. One solution to this problem might be to inactivate the split intein either by mutation of residues critical for function[26] or by adding Zn^{2+} ions, which inhibit splicing by coordinating active-site residues.[25,36] However, both of these strategies run the risk of perturbing the association of $DnaE_C$ and $DnaE_N$.

As an alternative solution to this problem, we conceived an inhibition strategy involving the incorporation of a chemical

(27) Richards, F. M. *Proc. Natl. Acad. U.S.A.* **1958**, *44*, 162−166.
(28) Taniuchi, H.; Anfinsen, C. B. *J. Biol. Chem.* **1971**, *246*, 2291−2301.
(29) Michnick, S. W.; Remy, I.; Campbell-Valois, F.-X.; Vallee-Belisle, A.; Pelletier, J. N. *Methods Enzymol.* **2000**, *328*, 208−230.
(30) Ozawa, T.; Kaihara, A.; Sato, M.; Tachihara, K.; Umezawa, Y. *Anal. Chem.* **2001**, *73*, 2516−2521.
(31) Ozawa, T.; Takeuchi, T. M.; Kaihara, A.; Sato, M.; Umezawa, Y. *Anal. Chem.* **2001**, *73*, 5866−5874.
(32) de Prat Gay, G.; Ruiz-Sanz, J.; Davis, B.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10943−10946.
(33) Kippen, A. D.; Sancho, J.; Fersht, A. R. *Biochemistry* **1994**, *33*, 3778−3786.
(34) Duan, X.; Gimble, F. S.; Quiocho, F. A. *Cell* **1997**, *89*, 555−564.
(35) Klabunde, T.; Sharma, S.; Telenti, A.; Jacobs, W. R., Jr.; Sacchettini, J. C. *Nat. Struct. Biol.* **1998**, *5*, 31−36.
(36) Mootz, H. D.; Blum, E. S.; Tyszkiewicz, A. B.; Muir, T. W. *J. Am. Chem. Soc.* **2003**, *125*, 10561−10569.

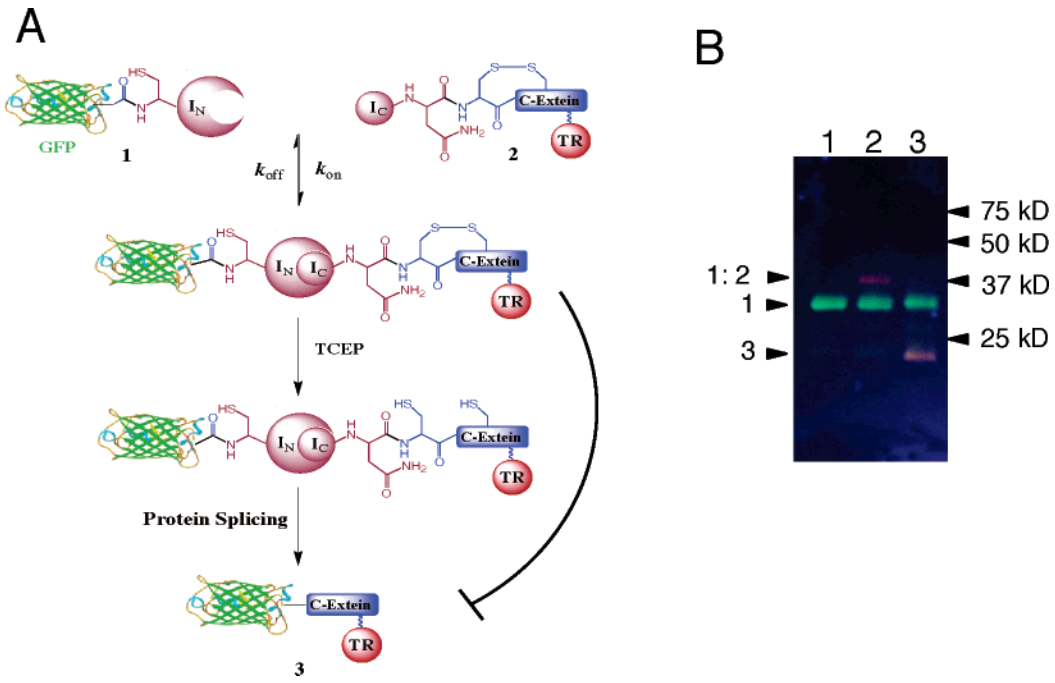


Figure 2. DnaE split intein interaction monitored by FRET. (A) Schematic showing the principle of the FRET binding assay. Protein constructs **1** and **2** contain the split intein fragments $DnaE_N$ and $DnaE_C$ and the FRET pair GFP and Texas Red (TR) within the extein sequences. Construct **2** also contains two Cys residues close in sequence. When these are oxidized, proteins **1** and **2** can interact, but splicing cannot occur (see Figure 1). Addition of a reducing agent such as TCEP triggers the splicing reaction. (B) Fluorecence image of a reducing SDS-PAGE gel of the trans-splicing reaction between **1** and **2** (illumination $\lambda = 306$ nm). The survival of GFP and TR fluorescence during SDS-PAGE allows reactants and products to be monitored. Lane 1 = purified construct **1**; lane 2 = **1** + **2** in the absence of reducing agent; lane 3 = **1** + **2** in the presence of TCEP. Note, the samples were not boiled to avoid quenching of the GFP fluorescence. This leads to a slight change in the electrophoretic mobility of the GFP fusion protein.

trigger of activity within the extein. As shown in Figure 2A, the C-extein was designed to contain two proximal cysteine residues, the first of which is the (−1) Cys involved in the splicing reaction. Since these two Cys residues are close together in sequence (sequence −CFNC−), we anticipated that they would readily form a disulfide and so prevent the splicing reaction. Note, redox switches have previously been used to control protein activity.[37,38] This inhibition strategy should be reversible and less invasive than mutagenesis of the intein or the addition of Zn^{2+}. Two fluorescent protein constructs were thus prepared. Construct **1** comprised GFP (the N-extein) fused to $DnaE_N$, whereas construct **2** comprised $DnaE_C$ fused to the aforementioned bis-cysteinyl peptide containing the organic fluorophore Texas Red (the C-extein). The former construct was fully recombinant and generated by standard bacterial expression approaches, whereas the latter was semisynthetic and prepared by expressed protein ligation[39] of a recombinant polypeptide and a synthetic peptide (see Experimental Methods).

Purified constructs **1** and **2** were combined under various conditions and the trans-splicing reactions monitored by reducing SDS-PAGE analysis (Figure 2B). Here we took advantage of the survival of GFP fluorescence during gel electrophoresis, allowing product formation to be detected by fluorescence imaging of the gel (Figure 2B). In the absence of any reducing agents in the splicing buffer, we observed a new band with a molecular weight and fluorescence emission consistent with that of a complex between **1** and **2** (Figure 2B, compare lanes 1 and 2). The formation of a $DnaE_N/DnaE_C$ complex has been previously detected by SDS-PAGE[25,26] and suggests a very tight interaction between the polypeptides and/or the presence of a cross-linked splicing intermediate. Importantly, no spliced product **3** (i.e., GFP-Texas Red peptide) could be detected under these reaction conditions. This is consistent with our design in which the two Cys residues in the C-extein segment are oxidized to a disulfide, thereby inhibiting splicing. Since there are no other cysteines in construct **2**, it suggests that the complex between **1** and **2** is indeed noncovalent and that, remarkably, it survives SDS-PAGE analysis.

Purified constructs **1** and **2** were next combined in the presence of the reducing agent TCEP (Figure 2B, lane 3). Under these conditions, a major new fluorescent protein band was observed. The molecular weight and fluorescence emission of this protein band are consistent with the expected splice product **3**. In this case we did not observe the accumulation of a complex between proteins **1** and **2**. Thus, trans-splicing can be triggered by the addition of the reducing agent TCEP, which is again consistent with our design.

Kinetics and Thermodynamics of Ssp DnaE Split Intein Association. Both equilibrium titrations and kinetic measurements revealed a very high affinity between the $DnaE_C$ and $DnaE_N$ fragments. The fluorescence-based equilibrium titration was carried out in the absence of TCEP, so that constructs **1** and **2** can associate, but splicing cannot occur due to the oxidation of the catalytic cysteine in $DnaE_C$. Upon excitation of GFP at 470 nm, we observed approximately 70% decrease in GFP fluorescence at 510 nm in the presence of saturating amounts of construct **2** (Figure 3A). This decrease was accompanied by an increase in Texas Red fluorescence at 612 nm. Titration of construct **1** with a control compound,

(37) Falcon, C. M.; Swint-Kruse, L.; Matthews, K. S. *J. Biol. Chem.* **1997**, *272*, 26818−26821.
(38) Posey, K. P.; Gimble, F. S. *Biochemistry* **2002**, *41*, 2184−2190.
(39) Muir, T. W. *Annu. Rev. Biochem.* **2003**, *72*, 249−289.

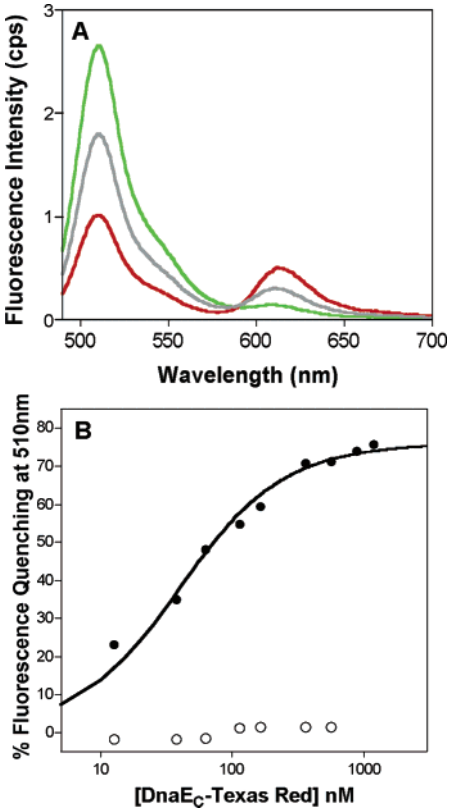


Figure 3. Fluorescence analysis of the DnaE split intein interaction. (A) Steady-state fluorescence emission spectra of **1** (15 nM) with increasing concentrations of **2** (0, 63, and 360 nM; green, gray, and red line, respectively) in the absence of TCEP (i.e., $DnaE_N$ and $DnaE_C$ can associate, but splicing cannot occur). Excitation was at 470 nm and direct (i.e., non-FRET) excitation of Texas Red was subtracted from these spectra. (B) Binding isotherm obtained from the equilibrium titration of **1** (15 nM) in the presence of increasing concentrations of **2** (solid circles) or increasing concentration of free Texas Red (open circles). Experiments were performed in triplicate.

DTT-conjugated Texas Red, did not lead to appreciable FRET between the fluorophores (Figure 3B). Thus, the FRET between constructs **1** and **2** reflects the interaction between $DnaE_C$ and $DnaE_N$. The binding isotherm was followed by the change in GFP fluorescence at 510 nm as a function of added construct **2** to construct **1** (Figure 3B). The dissociation constant (K_d) extracted from this isotherm was 36 ± 12 nM (Table 1). The same binding isotherm was obtained regardless of whether a 1 min or 10 min equilibration time was used. In other experiments, constructs **1** and **2** were combined in the presence of TCEP and the fluorescence emission spectrum of the mixture monitored over 24 h. Even though splicing was triggered under these conditions, we did not observe any discernible change in FRET over this period (data not shown). Thus, the amount of FRET associated with the initial split intein complex is similar to the final splice product **3**. This precluded further kinetic analysis of the chemical steps in of the protein splicing reaction by FRET.

Previous studies of DnaE trans-splicing suggest that association of the split intein fragments is rapid relative to the subsequent splicing reaction, which occurs in minutes to hours.[24] To monitor the kinetics of DnaE complex formation, we employed stopped-flow mixing with fluorescence detection. In these experiments, we measured the time-dependent decrease in GFP fluorescence in construct **1** upon addition of varying amounts of **2** in either the presence or absence of TCEP (Figure 4). The data were best fit by a double-exponential function. The observed rate from the initial fast phase depends linearly on the concentration of **2**, whereas the observed rate from the second slow phase does not (Figure 4). This biphasic behavior is consistent with a kinetic mechanism where more than one form of the binary complex is populated.[40] The slope and intercept extracted from the intial rate, k_{obs1}, define the association second-order rate constant k_{on} and the dissociation rate constant k_{off}, respectively, whereas the maximum rate of the second phase defines a unimolecular process within the complex.

The calculated rate constants are shown in Table 1. The kinetic data reveal that $DnaE_N$ and $DnaE_C$ associate extremely fast and dissociate moderately slowly. This accounts for the low dissociation constant of the complex; note, the K_d derived from the kinetic data is in excellent agreement with that obtained from equilibrium measurements (Table 1). Moreover, the kinetic constants are unaffected by the presence or absence of TCEP in buffer, reflecting the fact that the kinetics of association and dissociation are rapid relative to the subsequent splicing.

The second-order rate constant for association approaches that of the diffusion limit set by the Einstein−Smoluchowski equation to be 10^9-10^{10} M^{-1} s^{-1}. Association rates are dictated by macromolecular diffusion and by short-range forces that can offset orientational constraints for the interaction.[41] There is mounting evidence that nature exploits electrostatic attraction, or steering, to tune the rates of protein interactions.[42,43] Intriguingly, the $DnaE_N$ and $DnaE_C$ sequences have theoretical isoelectric points (pI) of 4.3 and 9.3, respectively (Figure 5). To examine the possibility that electrostatic steering contributes to the rapid association of $DnaE_N$ and $DnaE_C$, we repeated the kinetic experiment using a buffer of lower ionic strength. Consistent with the electrostatic steering model, we find that k_{on} increases as the ionic strength of the solution decreases, whereas k_{off} is unaffected (Table 1). The unimolecular rate constant, k_{obs2}, is also unaffected by the presence or absence of reducing agents or changes in ionic strength of the buffer (Table 1). Conceivably, the slow phase might reflect a reorganization of the initial complex into a distinctive conformational state. This type of two-phase behavior has been previously observed for the association of split proteins.[33] Further studies, perhaps involving the use of mutant $DnaE_C$ and $DnaE_N$ fragments, will be required to fully understand the significance of the second phase.

Development of a Tandem Protein Trans-Splicing System. As discussed above, the DnaE split intein fragments associate rapidly and with high affinity. These observations help explain the efficiency of DnaE-mediated trans-splicing in vitro and in vivo.[18,21−24] Another important facet of protein trans-splicing is the presumed specificity of the process. Inteins exhibit a relatively low level of sequence homology outside the conserved catalytic residues,[44] and so different split inteins are expected to be functionally orthogonal to one other (Figure 5). This is illustrated by the work of Yamazaki and co-workers, who used

(40) Wong, L.; Lieser, S.; Chie-Leon, B.; Miyashita, O.; Aubol, B.; Shaffer, J.; Onuchic, J. N.; Jennings, P. A.; Woods, V. L., Jr.; Adams, J. A. *J. Mol. Biol.* **2004**, *341*, 93−106.
(41) Berg, O. G.; von Hippel, P. H. *Annu. Rev. Biophys. Biophys. Chem.* **1985**, *14*, 131−160.
(42) Schreiber, G.; Fersht, A. R. *Nat. Struct. Biol* **1996**, *3*, 427−431.
(43) Selzer, T.; Albeck, S.; Schreiber, G. *Nat. Struct. Biol.* **2000**, *7*, 537−541.
(44) Perler, F. B. *Nucleic Acids Res.* **2000**, *28*, 344−345.

Table 1. Rate and Equilibrium Dissociation Constants for *Ssp* DnaE Split Intein Interactions in the Presence or Absence of the Reducing Agent TCEP[a]

buffer conditions	k_{on} (10^7 $M^{-1}s^{-1}$)	k_{off} (s^{-1})	K_d (nM)	k_{obs2} (s^{-1})
100 mM sodium phosphate[b]	2.8 ± 0.5	1.2 ± 0.2	43 ± 10	0.65 ± 0.03
100 mM sodium phosphate[c]			36 ± 12	
100 mM sodium phosphate, 2 mM TCEP[b]	2.5 ± 0.2	1.2 ± 0.5	47 ± 20	0.72 ± 0.05
20 mM sodium phosphate, 2 mM TCEP[b]	6.7 ± 0.2	1.7 ± 0.6	26 ± 12	0.70 ± 0.01

[a] The data are shown as the mean and standard deviation. [b] Equilibrium dissociation constants derived from the ratio of k_{off}/k_{on} using stopped-flow measurement. [c] Equilibrium dissociation constant derived from the steady-state titration.

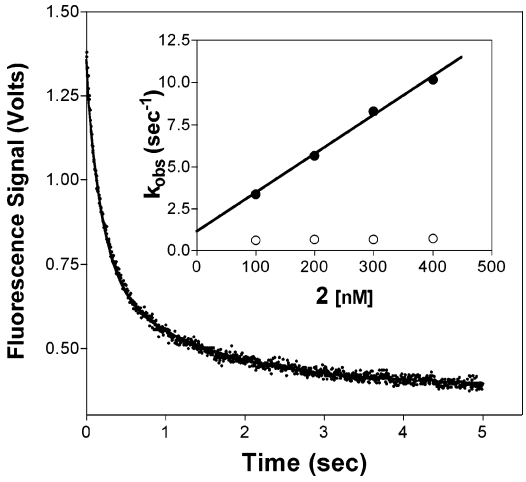


Figure 4. Stopped-flow kinetic analysis of the DnaE split intein interaction. Interaction of construct **1** (20 nM) with construct **2** (100−400 nM) was monitored using the quenching of GFP fluorescence in **1** (data shown are for [**2**] = 200 nM). The data were fit to a double-exponential decay function to give rate constants k_{obs1} and k_{obs2} ($k_{obs1} = 5.2 \pm 0.1$ s^{-1} and $k_{obs2} = 0.75 \pm 0.02$ s^{-1}; $A_1 = 0.63 \pm 0.01$ V and $A_2 = 0.34 \pm 0.01$ V in the example shown). The inset shows the observed rate constants as a function of the concentration of **2**. The observed rate from the fast phase k_{obs1} (solid circle) depends linearly on the concentration of **2**, while the observed rate from the slow phase k_{obs2} (open circles) does not change with increasing concentration of **2**. k_{on} and k_{off} were obtained using the equation $k_{obs1} = k_{on}[\mathbf{2}] + k_{off}$, where k_{on} is the slope and k_{off} the ordinate intercept.

two different split inteins to assemble three segments of a protein in a tandem trans-splicing reaction.[45] In this work, fragments from two artificially split inteins (PI-*pfuI* and PI-*pfuII*, both from the organism *Pyrococcus furiosus*) were appended to the appropriate ends of three fragments of maltose binding protein (MBP). The use of artificially split inteins meant that relatively harsh reaction conditions (chemical denaturants and elevated temperatures) were required during this procedure. Nonetheless, the two trans-splicing reactions did occur simultaneously and with high fidelity, thereby demonstrating the specificity of the process.

The ability to assemble multiple polypeptides simultaneously is highly desirable since it allows the differential incorporation of stable isotopes[13,46] or optical probes[47] into discrete regions of a protein. Ideally, procedures for doing this should work under physiological conditions and be compatible with low concentrations of reactants. With this in mind, we conceived the tandem trans-splicing strategy outlined in Figure 6A. In addition to the naturally split DnaE intein, this system incorporates an artificially split VMA intein fused to FKBP and FRB heterodimerization domains. We have previously demonstrated that this configuration allows the trans-splicing activity of the VMA intein to be triggered by addition of the small molecule rapamycin, which induces assembly of the intein complex.[16] Importantly, this "conditional protein splicing" (CPS) occurs under physiological conditions.[36] Thus, the proposed tandem trans-splicing strategy should permit the assembly of three polypeptides under physiological conditions. Furthermore, it should be possible to perform both trans-splicing reactions simultaneously (by adding rapamycin to the mix at $t = 0$) or, if desirable, in a stepwise fashion by adding rapamycin to the mix after a defined period (in this case the DnaE-mediated reaction should occur first).

As a proof of concept, we undertook the synthesis of the signaling adaptor protein c-CrkII from three recombinant protein segments. c-CrkII is a 304 amino acid protein composed of an N-terminal Src homology 2 (SH2) domain followed by two SH3 domains.[48] The target protein was divided into three pieces, each containing one of these domains (Figure 6B). The two split sites were designed to be located in the linker regions between the SH domains and are the same as those previously used in the semisynthesis of the protein by expressed protein ligation.[49]

Three recombinant fusion-proteins were generated for use in the tandem trans-splicing procedure (Figure 6C). Construct **4** contained the Crk SH2 domain (residues 1−124) fused to $DnaE_N$. This construct also contained an N-terminal Flag tag and a C-terminal MBP for analysis and purification purposes, respectively. Attachment of MBP was found to greatly improve the yield of soluble protein expressed in *E. coli*. Construct **5** contained the central SH3 domain of Crk (residues 125−267) flanked by $DnaE_C$ and VMA_N-FKBP. A C-terminal MBP was included for affinity purification and was again found to increase the amount of soluble protein expression in *E. coli*. Importantly, construct **5** also contained two native DnaE C-extein residues (Phe-Asn) between $DnaE_C$ and the SH3 domain. Previous reports have indicated that efficient DnaE-mediated trans-splicing in a foreign host requires several native extein residues surrounding the intein.[20] In preliminary studies we found insertion of just two native C-extein residues was required for efficient trans-splicing in this context (see Supporting Information). Finally, construct **6** contained the C-terminal Crk SH3 domain fused to FRB-VMA_C. In addition, MBP and a polyHis tag were included at the N- and C-termini, respectively. Note that in preliminary studies we confirmed that the individual

(45) Otomo, T.; Ito, N.; Kyogoku, Y.; Yamazaki, T. *Biochemistry* **1999**, *38*, 16040−16044.
(46) Xu, R.; Ayers, B.; Cowburn, D.; Muir, T. W. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 388−393.
(47) Muralidharan, V.; Cho, J.; Trester-Zedlitz, M.; Kowalik, L.; Chait, B. T.; Raleigh, D. P.; Muir, T. W. *J. Am. Chem. Soc.* **2004**, *126*, 14004−14012.
(48) Feller, S. M.; Ren, R.; Hanafusa, H.; Baltimore, D. *Trends Biochem. Sci.* **1994**, *19*, 453−459.
(49) Blaschke, U. K., Cotton, G. J., Muir, T. W. *Tetrahedron* **2000**, *56*, 9461−9470.

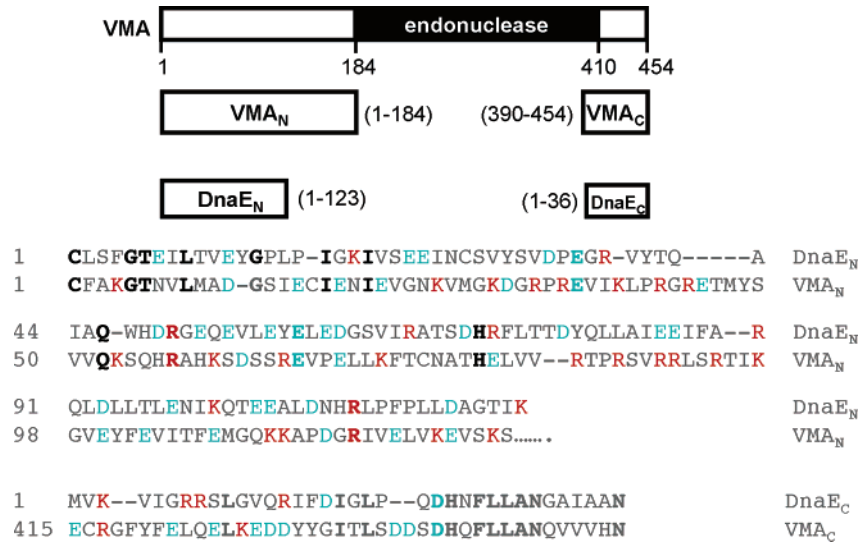


Figure 5. Comparison of the naturally split *Ssp* DnaE and artificially split *Sce* VMA inteins. Top: The VMA intein contains an internal homing endonuclease domain that is removed in the split version of the protein. Bottom: Alignment of the split DnaE and VMA inteins used in this study. Conserved residues are in bold, acidic residues in blue, and basic residues in red. The calculated isoelectric points of the VMA$_N$ and VMA$_C$ domains are 9.5 and 4.0, respectively. This is in direct contrast with DnaE$_N$ (4.3) and DnaE$_C$ (9.3).

DnaE and VMA trans-splicing reactions occurred in the context of the CrkII system (see Supporting Information).

Initially, we attempted the tandem trans-splicing reaction using purified proteins. Accordingly, proteins **4**−**6** were combined at low micromolar concentration in a buffer containing both TCEP and rapamycin, and the tandem trans-splicing reaction was monitored by SDS-PAGE. After 48 h reaction, several new protein bands were present in the reaction mixture, in addition to the three starting materials (Figure 7A, compare lanes 1 and 5). A complex reaction mixture was expected since the tandem trans-splicing reaction should generate five products, the final splice product CrkII (**7**) as well as the four residual split intein byproducts (Figure 6A). A band was observed with the expected electrophoretic mobility of **7**. Immunoblotting revealed that this component cross-reacted with both anti-Flag and anti-CrkII antibodies (Figure 7B and data not shown). Additional analysis of the mixture by RP-HPLC, followed by MALDI-TOF mass spectrometry analysis of the fractions, confirmed that the tandem splice product **7** had been generated (Figure 7C). The overall yield of the three-component reaction was estimated to be 48% based on densitometric analysis of the anti-FLAG immunoblot (Figure 7B, lane 1). The presence of a faint band around 100 kDa in the anti-FLAG immunoblot of the reaction mixture is consistent with the intermediate trans-splicing product between **4** and **5**. Importantly, no splice products were observed between constructs **4** and **6**. Thus, as expected, the VMA and DnaE split inteins are functionally orthogonal. It is worth noting that the calculated isoelectric points of the VMA$_N$ and VMA$_C$ domains are 9.5 and 4.0, respectively. This is in direct contrast with DnaE$_N$ (4.3) and DnaE$_C$ (9.3) (see Figure 5 for sequence comparison).

Encouraged by the above results, we next attempted to perform the tandem trans-splicing reactions directly in crude bacterial lysates containing the overexpressed proteins **4**−**6**. As illustrated in Figure 8A, the three bacterial lysates were incubated in the presence of TCEP and rapamycin. After 18 h of reaction, the lysates were affinity purified using Ni-NTA beads (which will bind the His-tagged proteins), and the enriched material was analyzed by SDS-PAGE and immunoblotted against an anti-FLAG antibody. Only the tandem splice product **7** would be expected to cross-react with both the Ni−NTA beads and the anti-FLAG antibody. Indeed, a band of the correct molecular weight for the tandem splice product **7** was observed in the blot (Figure 8B, lane 1). Interestingly, we also observed a higher molecular weight species consistent with the Flag-tagged starting material **4** in this sample. Since this construct does not contain a His-tag, we presume it must have been in complex with constructs **5** and **6** and hence co-affinity purified in the Ni-NTA pull-down step. Thus, a tandem trans-splicing reaction can be performed simply mixing appropriate bacterial lysates in the presence of reducing agent and rapamycin.

The ability to perform the reaction in this manner has obvious practical implications for the assembly of proteins from multiple component parts. By eliminating the need for any biochemical manipulations prior to the ligation step, the process is greatly streamlined. Moreover, by appending differential affinity tags on either end of the final splice product it should make it possible to partially purify the product in two affinity chromatography steps, although as seen in this example an additional step (e.g., size exclusion chromatography) may be needed to remove starting materials that remain in a ternary complex. Thus, we believe this strategy is an attractive route to the preparation of proteins in which discrete regions are labeled with stable isotopes or optical probes.

Conclusions

This work presents several important advances in understanding protein trans-splicing and exploiting orthogonal split inteins for use in protein engineering. Fundamental questions regarding the self-association kinetics of a split intein can now be addressed by monitoring the changes in FRET between GFP-DnaE$_N$ and DnaE$_C$-TR. *Ssp* DnaE split intein fragments display considerable affinity toward each other. The relatively low K_d results from a very fast association rate and moderately slow dissociation rate. Furthermore, the association rate depends on the ionic strength of the buffer, suggesting an electrostatic

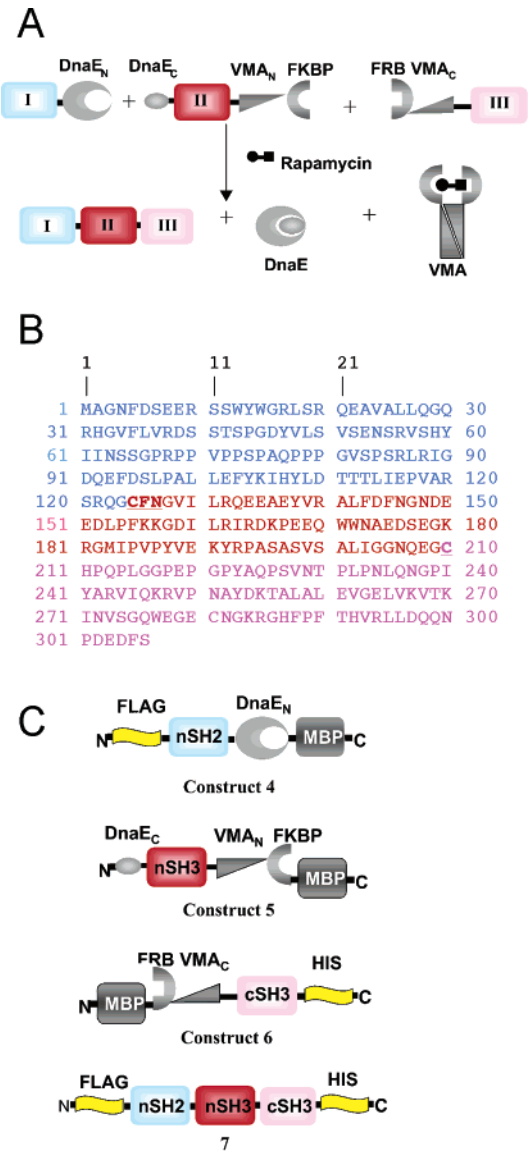


Figure 6. Tandem protein trans-splicing. (A) The system combines the naturally split DnaE and inducible split VMA inteins. Addition of rapamycin under physiological conditions triggers the ligation of proteins I−III via normal peptide bonds. (B) Sequence of murine c-CrkII. The three segments of the protein used in the trans-splicing reactions are indicated: SH2 domain (blue), nSH3 domain (red), and cSH3 domain (magenta). Mutations introduced to facilitate the two trans-splicing reactions are underlined. Note, the strategy results in the insertion of two amino acids (FN) between residues 124 and 125. (C) Domain structure of the three constructs **4**−**6** used in the tandem trans-splicing reactions, as well as the final product **7**.

influence on the kinetics of split intein interaction. The ability to monitor the rates of association and dissociation of the *Ssp* DnaE split intein may also prove useful in designing a more efficient split intein in the future.

Exploiting the orthogonality of the naturally split DnaE intein and the inducible split VMA intein, we have successfully developed a tandem trans-splicing reaction and used this to assemble the multidomain protein CrkII is a one-pot reaction under native conditions (Figures 6 and 7). Unlike earlier systems,[45] the present tandem trans-splicing technology does not require denaturing and refolding of the target protein and the use of elevated temperatures. The efficiency of protein trans-splicing is dependent on the context of host extein. In the case of *Ssp* DnaE, trans-splicing was only successful when two native

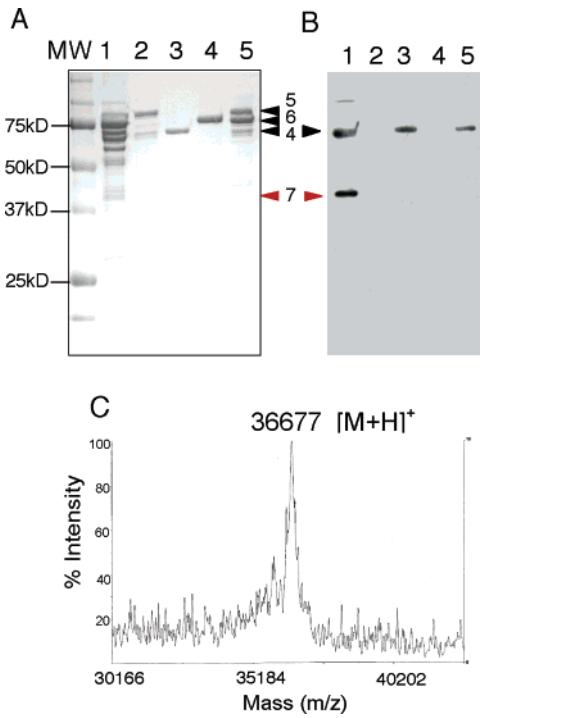


Figure 7. In vitro tandem trans-splicing of purified CrkII intein precursor proteins. Analysis of the trans-splicing reaction by Coomassie stained SDS-PAGE (A) and Western blotting using an anti-FLAG antibody (B). Purified constructs **4** (lane 3), **5** (lane 2), and **6** (lane 4) were combined (lane 5, $t = 0$), and the trans-splicing reaction was allowed to proceed for 48 h at 22 °C (lane 1). The calculated molecular masses of starting materials are as follows: **4** (69.7 kD), **5** (88.6 kD), **6** (72.7 kD). (C) MALDI-TOF mass spectrum of spliced product **7** isolated by analytical RP-HPLC of the reaction mixture (calculated mass = 36695 Da).

extein amino acids, Phe-Asn, were inserted at the splice junction. Tandem trans-splicing involving DnaE and VMA split inteins should prove generally applicable to segmental isotopic or fluorogenic labeling of specific domains within the context of a large multidomain proteins.

Experimental Procedures

General Methods. All amino acid derivatives and resins were purchased from Novabiochem (San Diego, CA). Rapamycin was purchased from Sigma-Aldrich Co. (St Louis, MO). Antibodies were purchased from Covance (Princeton, NJ; anti-His, clone MMS-156R), BD Transduction Laboratories (Lexington, KY; anti-Crk), and Sigma-Aldrich (monoclonal anti-FLAG). All other chemical reagents were purchased from Sigma-Aldrich and Fisher Scientific (Pittsburgh, PA). Analytical gradient HPLC was performed on a Hewlett-Packard 1100 series instrument with a Vydac C18 column (5 μm, 4.6 × 150 mm). Preparative HPLC was performed on a Waters DeltaPrep 4000 instrument fitted with a Waters 486 tunable absorbance detector. All runs used linear gradients of 0.1% aqueous TFA (solvent A) vs 90% acetonitrile plus 0.1% TFA (solvent B). Electrospray mass spectrometry (ESMS) was performed on a Sciex API-100 single quadrupole electrospray mass spectrometer. MALDI-TOF was performed on Applied Biosystems Voyager-DE STR MALDI-TOF using 4-hydroxy-cinnamic acid (4HCCA) as the matrix.

Peptide Synthesis. The fluorescent peptide, H-CGRGK(ϵ-Texas Red)G-NH_2, was synthesized on a 4-methylbenzhydrylamine (MBHA) resin according to the in-situ neutralization/HBTU activation protocol for Boc SPPS.[50] Following chain assembly, Texas Red succinimidyl ester (Molecular Probes, Eugene, OR) was attached to the ϵ-NH_2 group

(50) Schnölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. *Int. J. Pept. Protein Res.* **1992**, *40*, 180−193.

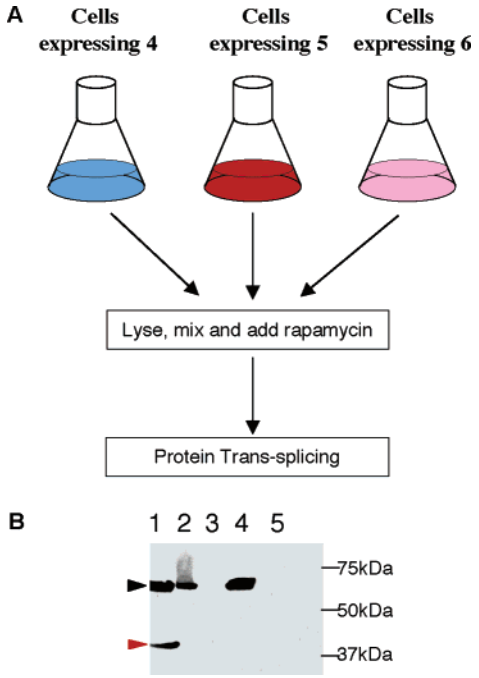


Figure 8. Tandem protein trans-splicing in whole-cell lysates. (A) Schematic showing the experimental strategy employed. Bacterial cells expressing individual constructs **4**−**6** were lysed and mixed in the presence of 100 nM rapamycin. The tandem trans-splicing reaction was allowed to proceed for 18 h at 24 °C. (B) Anti-FLAG Western blot of the reaction mixture at $t = 18$ h (lane 1) and $t = 0$ h (lane 2). Immunoblots of the input lysates are also shown: **5** (lane 3), **4** (lane 4), and **6** (lane 5).

of Lys using an orthogonal (Fmoc/Boc) protection strategy. The peptide was cleaved from the support using liquid HF at 4 °C and purified by semipreparative HPLC (Vydac C18 column) using a linear gradient of 25−45% B over 60 min at a flow rate of 5 mL/min; ESMS, observed = 1278.5 ± 1 Da, expected = 1277.5 Da.

Construction of Expression Plasmids. The plasmid encoding construct **1** (pHis6-GFP-DnaE$_C$) is described.[21] The plasmid pDnaE$_C$-GyraseA-6His was created by inserting two complementary and annealed synthetic DNA oligos encoding the polyhistidine tag, TGGAHHHHHHGA, between the *Age* and *Pst* restriction sites in the previously described template plasmid, pDnaE$_C$-GyraseA.[21] Note that this template plasmid encodes three C-terminal extein residues (CFN) between DnaE$_C$ and the GyraseA intein. Plasmids encoding CrkII split intein fusion proteins **4**−**6** were constructed as follows. pFLAG-SH2-DnaE$_N$-MBP (**4**): The template plasmid pSH2-DnaE$_N$-MBP was generated by PCR using three overlapping PCR products encoding the murine CrkII SH2 domain (residues 1−124), DnaE$_N$ and MBP (from pMal-c2X, New England Biolabs). Note, a linker sequence, Gly-Ser, was inserted between DnaE$_N$ and MBP by including the corresponding nucleotide sequence in the PCR primers. The digested PCR product was subcloned into the vector pTXB1 (New England Biolabs) using *Nde* and *Age* restriction sites. pFLAG-SH2-DnaE$_N$-MBP was created by inserting two complementary and annealed synthetic DNA oligos, encoding GHDYDKDDDKGS, into the *Nde* site of the pSH2-DnaE$_N$-MBP template. pDnaE$_c$-nSH3-VMA$_N$-FKBP-MBP (**5**): The DNA encoding nSH3-VMA$_N$-FKBP-MBP was generated by PCR using three overlapping PCR products encoding the nSH3 domain of murine CrkII (residues 125−207), VMA$_N$-FKBP (from pHM41[36]), and MBP. The digested PCR product was subcloned into the vector pDnaE$_C$-Gyrase-His using *Mfe* I and *Age* restriction sites. The cloning procedure was designed to mutate Ser125 in nSH3 to Cys and to insert the two natural extein residues (Phe-Asn) following Cys125. pMBP-FRB-VMA$_C$-cSH3-His6 (**6**): The cSH3 domain of CrkII (residues 208−304) was amplified by PCR and cloned into the *Mfe* I and *Hind* III sites of the previously described plasmid pHM45,[36] which encodes MBP-FRB-VMA$_C$. The polyHis tag was fused to the C-terminus of cSH3 through the 3′ primer, and the 5′ primer encoded the Ser208-Cys mutation in cSH3.

Protein Expression and Purification. *E. coli* BL21 (DE3) cells (Invitrogen, San Diego, CA) were transformed with individual expression plasmids. Transformed cells were grown to mid-log phase in Luria-Bertani media at 37 °C, and protein expression was induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.8 mM at 20−24 °C for 5 to 7 h. Cells were then harvested by centrifugation, resuspended in lysis buffer (25 mM sodium phosphate pH 7.5, 150 mM NaCl), and lysed by passage through a French press. Insoluble material was removed by centrifugation and, in the case of constructs **4**−**6**, the desired fusion protein purified from the supernatants by affinity chromatography using amylose affinity columns (New England Biolabs). Constructs **4** and **6** were further affinity purified using M2 anti-FLAG affinity resin (Sigma-Aldrich) and Ni-NTA resin (Novagen), respectively. Construct **1** and DnaE$_C$-Gyrase-6His were purified directly on Ni-NTA columns. Purified proteins were dialyzed against 0.1 M sodium phosphate buffer, pH 7.0 or pH 8.0, and either used directly in the trans-splicing reactions or flash frozen at −80 °C. The concentration of protein solutions was determined by Bradford assay.

Semisynthesis of DnaE$_C$-Texas Red (2). Purified DnaE$_C$-GyraseA-6His (0.5 mM) and H-CGRGK(ϵ-Texas Red)G-NH$_2$ (0.8 mM) were combined in 1 mL of ligation buffer (0.1 M sodium phosphate, pH 8.0) containing 100 mM MESNA and 10 mM TCEP at room temperature for 48 h. The ligation product, construct **2**, precipitated during the reaction and was redissolved in 30% acetic acid prior to purification by RP-HPLC using a linear gradient of 22−42% B over 30 min. The identity of ligated product was confirmed by ESMS [**2**: observed mass = 5456.0 ± 1.0 Da; expected mass = 5456.4 Da].

Characterization of Ssp DnaE Trans-Splicing by SDS-PAGE. Purified **1** and **2** were combined in splicing buffer (0.1 M sodium phosphate pH 7.0, 1 mM EDTA). The concentration of the two proteins was typically in the range 10−30 μM. The mixture was allowed to react at room temperature for 12−16 h in the presence or absence of 5 mM TCEP. The reaction was then quenched by dilution into SDS-PAGE loading buffer containing 1% v/v 2-mercaptoethanol before being analyzed by SDS-PAGE. To avoid quenching of GFP fluorescence, the sample was not boiled before loading onto the gel. The spliced product **3** was visualized using either a Typhoon 9500 fluorescence scanner (Amersham Pharmacia, Piscataway, NJ) or illumination with an ultraviolet lightbox. The concentrations of **1** and **2** were determined spectrophotometrically: **1**, $\epsilon_{488} = 58\,000\ M^{-1}\ cm^{-1}$; **2**, $\epsilon_{590} = 80\,000\ M^{-1}\ cm^{-1}$.

Spectrofluorometric Analyses of Ssp DnaE Split Inteins. Steady-state emission spectra were measured at room temperature using a Jobin Yvon/Spex Fluorolog 3 spectrofluorometer (Instrument S.A., Inc., Edison, NJ) with the excitation bandwidth set at 4 nm and emission at 5 nm. The excitation wavelength for GFP was set at 470 nm, and emission was monitored between 490 and 700 nm. The equilibrium dissociation constant (K_d) for the DnaE split intein interaction was obtained by titration of a fixed quantity of **1** (15−30 nM) with increasing concentrations of construct **2**. Dissociation constants were calculated by assuming a 1:1 complex and by fitting the fractional decrease in GFP fluorescence emission at 510 nm to eq 1:

$$\Delta F = \Delta F_{max}(N_t + C_t + K_d - \{(N_t + C_t + K_d)^2 - 4N_tC_t\}^{0.5}))(2N_t)^{-1} \quad (1)$$

where ΔF and ΔF_{max} are the change and maximum change in fluorescence, respectively, N_t is the total concentration of **1**, and C_t is the total concentration of **2**.

Association and dissociation rate constants of **1** with **2** were determined from changes in the GFP fluorescence using an Applied Photophysics SX.18MV (Leatherhead, UK) stopped-flow spectrophotometer. Time-dependent decreases in GFP fluorescence were observed

upon excitation at 480 nm by means of a 515 nm emission cutoff filter. For data analysis, the average of 3−5 individual traces was used for a given concentration of **2**. The observed fluorescence decay, $F(t)$, was fitted to a double-exponential decay function, eq 2.

$$F(t) = A_1 \exp(-k_{obs1}t) + A_2 \exp(-k_{obs2}t) + \text{Plateau} \qquad (2)$$

where A and k are the amplitude and rate of fluorescence decay, respectively. Goodness of fit was evaluated from the standard deviations of the fitted values. On average, the standard deviation represented less than 2.5% of the fitted value.

The observed rate from the first phase k_{obs1} depends linearly on the concentration of **2**, while the observed rate from the second phase k_{obs2} does not change with increasing concentration of **2**. k_{on} and k_{off} were obtained using the equation $k_{obs1} = k_{on}[\mathbf{2}] + k_{off}$, where k_{on} is the slope and k_{off} the ordinate intercept.[33] The K_d was calculated from the ratios of k_{off} to k_{on} obtained from two to four independent stopped-flow experiments.

In Vitro Protein Trans-Splicing Reactions. In a typical reaction, purified constructs **4** (∼2 μM), **5** (∼5 μM), and **6** (∼6 μM) were combined at 22 °C in 100 μL of splicing buffer in the presence of 5 mM TCEP and 8 μM rapamycin. The reaction was monitored by SDS-PAGE, immunoblotting, and MS and deemed to be complete after 48 h.

Whole Cell Lysate Protein Trans-Splicing Reactions. *E coli* BL2(DE3) cells overexpressing individual plasmids encoding **4**−**6** were harvested and resuspended in 0.1 M sodium phosphate pH 7.0, 0.25 mM EDTA, and 5 mM TCEP in the presence of protease inhibitor cocktail (Roche Applied Science, Indianapolis, IN). Each cell suspension was lysed by three passages though a French press and then clarified by centrifugation at 17000*g* for 20 min. The supernatants containing protein fragments **4**, **5**, and **6** were then mixed in a ratio of 1:2:1 to a final volume of 1 mL. Note, this ratio resulted in approximately equimolar amounts of the three constructs. The reaction was allowed to proceed in the presence of 100 nM rapamycin at 24 °C for 18 h. To ensure unambiguous identification of splicing product, we enriched the His-tagged proteins by incubation of the reaction mixture with 50 μL of Ni-NTA slurry for 15 min. Formation of the splicing product was monitored by Western blotting using an anti-FLAG antibody.

Acknowledgment. We thank Dr. Haiteng Deng for his help with MALDI-TOF of CrkII, and Dr. Zoran Radic and Prof. Palmer Taylor for instrument time and expertise in stopped-flow fluorescence. We also thank Vasant Muralidharan and Miquel Vila-Perello for their help with CrkII characterization. This work was supported by NIH grants GM55843 and EB001991 and by a Charles H. Revson fellowship to J.S.

Supporting Information Available: Details of individual split intein mediated trans-splicing of CrkII in the context of C-extein residues and ref 3 (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA042287W