JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

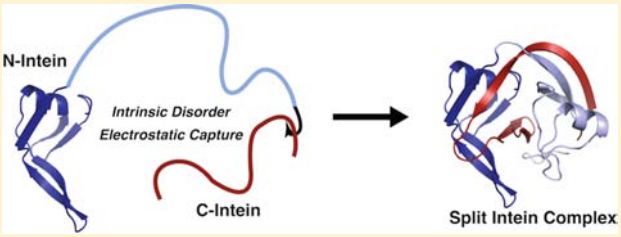
Naturally Split Inteins Assemble through a "Capture and Collapse" Mechanism

Neel H. Shah,[†,§] Ertan Eryilmaz,[‡,§] David Cowburn,[‡] and 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

Supporting Information

ABSTRACT: Split inteins are a class of naturally occurring proteins that carry out protein splicing *in trans*. The chemical mechanism of protein *trans*-splicing is well-understood and has been exploited to develop several powerful protein engineering technologies. Split intein chemistry is preceded by efficient molecular recognition between two protomers that become intertwined in their bound state. It is currently unclear how this unique topology is achieved upon fragment association. Using biophysical techniques in conjunction with protein engineering methods, including segmental isotopic labeling, we show that one split intein fragment is partly folded, while the other is completely disordered. These polypeptides capture each other through their disordered regions and form an ordered intermediate with native-like structure at their interface. This intermediate then collapses into the canonical intein fold. This mechanism provides insight into the evolutionary constraints on split intein assembly and should enhance the development of split intein-based technologies.



INTRODUCTION

Inteins are autoprocessing domains that facilitate their own excision from larger precursor proteins while ligating their flanking polypeptide sequences (exteins) through a native peptide bond.[1] This process, known as protein splicing, occurs in unicellular organisms from all domains of life.[2,3] The chemical mechanism of protein splicing is well-understood, and this has led to the development of numerous protein engineering tools that harness inteins to chemically manipulate virtually any protein of interest.[4] There are two types of naturally occurring inteins: highly abundant contiguous inteins and rarer split inteins. The latter carry out protein splicing *in trans*, where the biochemistry is preceded by the spontaneous association of two separately transcribed and translated fragments (N- and C-inteins, Figure 1a). Although naturally split inteins evolved from their contiguous counterparts and share the same fold,[5,6] it is notable that many artificially split versions of contiguous inteins cannot spontaneously assemble and splice.[7−9] The intein fold has a complex interwoven topology (Figure 1b),[10] and split inteins must bear unique features that allow them to access this topology *in trans*. Given that inteins splice essential proteins, there would have been strong selective pressure to optimize fragment complementation in early split inteins. Indeed, this evolution is evident in modern split inteins, which associate rapidly and tightly.[11−13]

The largest known class of split inteins is the cyanobacterial DnaE intein family.[5] In their natural context, these inteins assemble the catalytic subunit of DNA polymerase III (DnaE). Several members of this family carry out protein splicing with remarkable efficiency, in tens of seconds rather than hours as is the case for many other inteins, making them ideal tools for protein engineering.[14,15] Of these ultrafast *trans*-splicing domains, the DnaE intein from *Nostoc punctiforme* (Npu) is the best-characterized member, and it is rapidly becoming the intein of choice for new protein chemistry technologies.[16−19] Currently, little is known about the assembly of Npu or other split DnaE inteins. We previously demonstrated that the Npu fragments associate with low nanomolar affinity,[12] similar to orthologous split intein fragments from *Synechocystis* sp. PCC6803 (Ssp).[11,13] Atomic force microscopy measurements on three other split DnaE inteins suggest that this tight binding is a conserved property in this family.[20] Furthermore, Ssp binding is known to be extraordinarily fast and ionic strength dependent.[11] Recently, it was proposed that both Ssp fragments undergo a disorder-to-order transition upon binding; however, the molecular determinants and mechanism of this transition were not established.[13] On the basis of sequence analyses, it is clear that most split inteins differ from contiguous ones by the presence of significant charge segregation between the N- and C-inteins (Figure S1), and indeed, mutating critical ion pairs and triads at the fragment interface weakens fragment affinity.[12]

Here we elucidate the mechanism of split intein assembly, using Npu as a relevant model system. We show that split intein fragments interact through a multiphase process initiated by electrostatic interactions between extended regions of both fragments (capture) followed by compaction and stabilization of the initially disordered segments onto a prefolded region of

Received: October 10, 2013
Published: November 15, 2013

   dx.doi.org/10.1021/ja4104364 | J. Am. Chem. Soc. 2013, 135, 18673−18681

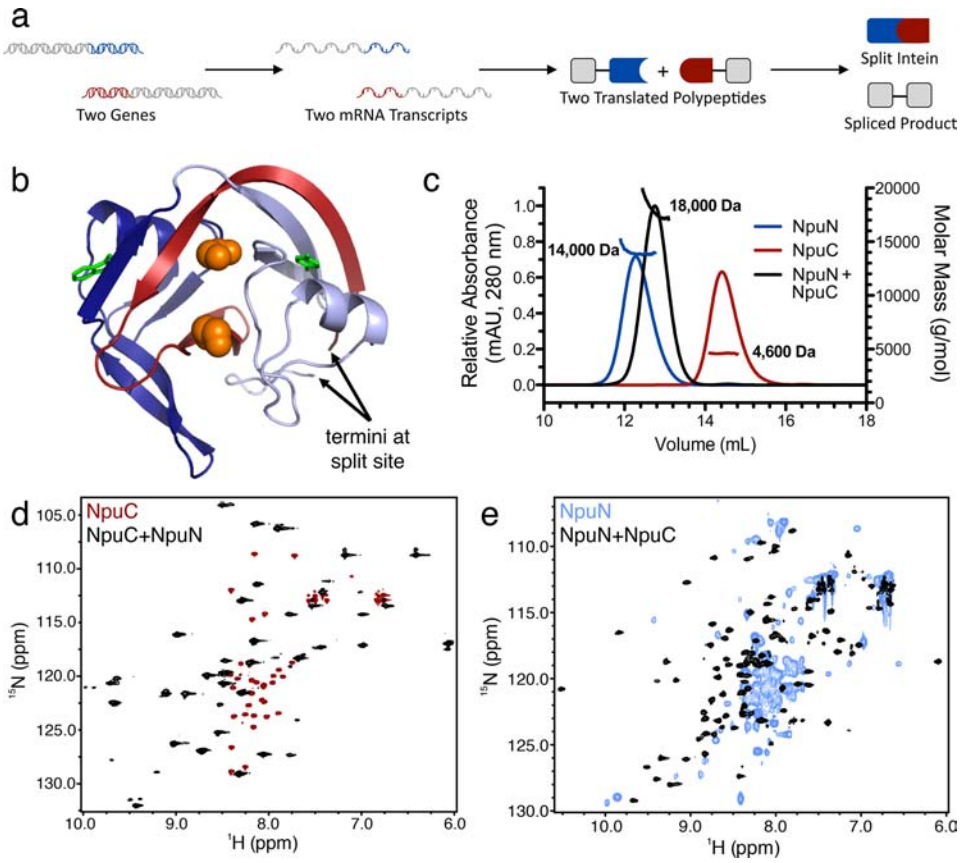


Figure 1. Characterization of Npu fragments and complex. (a) Scheme depicting the genetic origin of split intein fragments and their role in protein *trans*-splicing. (b) Rendering of the Npu intein structure, highlighting relevant NpuN lobes ($NpuN_1$ in dark blue and $NpuN_2$ in light blue) and NpuC in red. Termini to which exteins would be attached are shown as orange spheres, and sites of intrinsic fluorophores used in this study (W47 and Y58W) are shown as green sticks; rendering based on PDB 2KEQ.[10] (c) SEC-MALS of NpuN (blue, 25 μM, expected MW = 12188 Da), NpuC (red, 400 μM, expected MW = 4443 Da), and their complex (black, 25 μM, expected MW = 16631 Da). (d) ^{1}H−^{15}N HSQC spectra of NpuC alone (red, 250 μM, 600 MHz) and in complex with unlabeled NpuN (black, 250 μM, 800 MHz). (e) ^{1}H−^{15}N HSQC spectra of NpuN alone (blue, 250 μM, 500 MHz) and in complex with unlabeled NpuC (black, 250 μM, 800 MHz). NMR spectra were acquired at 25 °C in a pH 6.5 buffer containing 25 mM phosphates, 100 mM NaCl, and 1 mM DTT.

the N-intein (collapse). Given the emerging role of split inteins as powerful tools for in vitro and in vivo protein engineering and semisynthesis,[21] this biophysical mechanism should aid in the development and further improvement of split intein-based technologies.

RESULTS

Npu Fragments Undergo Dramatic Conformational Changes upon Association. To elucidate the binding mechanism of Npu and related inteins, we initially sought to characterize the structures of the 102 residue N-intein (NpuN) and 35 residue C-intein (NpuC) in isolation. Both proteins were generated and purified as previously described, bearing short tripeptide exteins that should not contribute significantly to the spectroscopic or chromatographic signals being observed (Figures S2 and S3).[22] These constructs also contained a C1A mutation in the N-intein and N137A mutation in the C-intein to prevent protein splicing during analyses of the split intein complex.[22] Circular dichroism spectra of the fragments and an equimolar mixture suggested low α and β secondary structure content in the isolated fragments and a dramatic structural change upon binding (Figure S4). Consistent with these data, NpuC was extremely susceptible to degradation by thermolysin, whereas the complex was effectively resistant to proteolysis (Figure S5a,b). NpuN showed an intermediate rate of degradation, suggesting a partially compact structure (Figure S5c). Remarkably, the N-intein eluted earlier than the complex on a size exclusion chromatography (SEC) column, indicating that it must collapse from an extended state upon binding NpuC (Figure 1c). To rule out the possibility of dimerization, we performed multiangle light scattering (MALS) online with SEC and confirmed that NpuN is a monomer in solution (Figure 1c).

Additional insight into fragment structure was gained from nuclear magnetic resonance (NMR) spectroscopy. The ^{1}H−^{15}N HSQC spectrum of uniformly ^{15}N-labeled NpuC showed sharp amide N−H resonances with poor dispersion along the ^{1}H axis, indicative of an unfolded protein (Figure 1d, red). Notably, a similar spectrum was previously observed for the homologous Ssp C-intein, suggesting conserved structural properties among split inteins, even in the unbound state.[13] Upon addition of unlabeled NpuN, the signal dispersion increased dramatically, consistent with NpuC being part of a well-folded protein domain (Figure 1d, black). In the bound and unbound states, backbone resonances were readily assigned to aid in further analyses. NMR relaxation experiments (R_1, R_2, and heteronuclear NOE measurements) on isolated NpuC confirmed the high degree of flexibility throughout the chain with no detectable residual secondary structure (Figure S6). In contrast to NpuC, the HSQC spectrum of isolated NpuN

dx.doi.org/10.1021/ja4104364 | J. Am. Chem. Soc. 2013, 135, 18673−18681

showed moderate signal dispersion, however only a fraction of the expected resonances were visible, and the observed signals varied greatly in relative intensity (Figure 1e, blue). These data depict an N-intein that is highly dynamic but at least partially folded, consistent with both the expanded dimensions observed by SEC and the intermediate protection from thermolysin proteolysis. While these experiments provide a general picture of NpuN, the nature of the NMR data precluded backbone resonance assignments and thus sequence-specific information.

NpuN Is Comprised of Two Structurally Distinct Lobes. We previously demonstrated that Npu fragment binding is dependent on electrostatic interactions between the highly acidic NpuN (calculated pI = 4.4) and highly basic NpuC (calculated pI = 9.7).[12] The majority of relevant anionic residues on NpuN are concentrated to the second half of the sequence ($NpuN_2$, residues 51−102), which is spatially distinct from the first half of the sequence ($NpuN_1$, residues 1−50) in the final intein complex (Figures 1b and S1, light and dark blue regions, respectively). This uneven distribution of charges is clearly seen in the final complex, where the interface between $NpuN_1$ and NpuC is primarily hydrophobic (Figure 2a,b) while

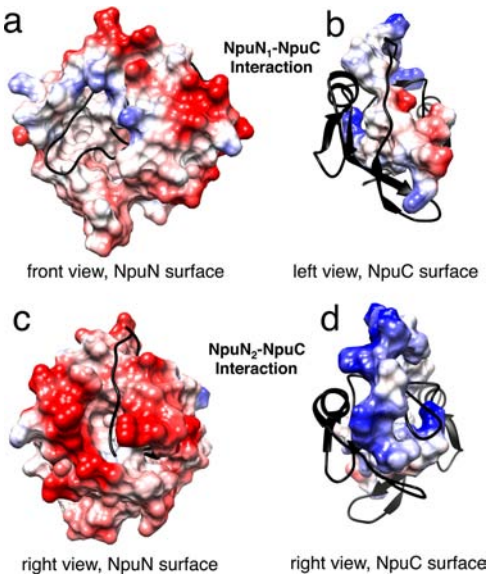


Figure 2. Electrostatic surface representations of Npu. The color scheme for these representations is red for negative charge, white for neutral, and blue for positive charge. All perspectives are given relative to panel (a), which is the same perspective as Figure 1b. The renderings are based on PDB 2KEQ.[10] Panels a and b highlight complementary hydrophobic surfaces on either fragment at the $NpuN_1$−NpuC interface. Panels c and d highlight complementary electrostatic surfaces on either fragment at the $NpuN_2$−NpuC interface.

$NpuN_2$ and NpuC have significant electrostatic complementarity (Figure 2c,d). Given these observations, we hypothesized that the different electrostatic properties of these two NpuN lobes may shed light on the overall structure of NpuN in isolation. Indeed, a closer analysis of the sequence composition of these lobes indicated that the first lobe, $NpuN_1$, has a high density of hydrophobic residues and low net charge per residue, consistent with a well-folded protein (Figure 3a). By contrast, the second lobe, $NpuN_2$, has a high net charge per residue and lower hydrophobicity, a hallmark of intrinsically disordered proteins (IDPs, Figure 3a).[23] Furthermore, despite their virtually identical molecular weights (ΔMW = 2 Da) and their symmetry-related structures in the final complex (Figure 1b), isolated $NpuN_1$ eluted later than $NpuN_2$ on a SEC column, suggesting a more compact structure for this lobe (Figures 3b and S7a−h). The more expanded lobe, $NpuN_2$, could be compacted in the presence of high salt, a known property of IDPs (Figure S7i).[24] Additionally, $NpuN_1$, but not $NpuN_2$, bound the hydrophobic dye 1-anilino-8-naphthalene sulfonate (ANS), suggesting that this compact lobe has molten globule-like properties with solvent-exposed hydrophobic surfaces (Figure S7j).[25] As discussed above, the interface between $NpuN_1$ and NpuC in the final split intein complex is dominated by hydrophobic interactions (Figure 2), which presumably satisfy these exposed patches.

We next asked if the different structural properties observed in the isolated NpuN lobes are retained in the context of full-length NpuN. To address this, we identified products from the limited proteolysis of NpuN by mass spectrometry. Early in the proteolysis reaction, we identified several pairs of products consistent with single proteolytic events (Figure S5d−f). As expected, in full-length NpuN, the second lobe was more susceptible to proteolysis than the first, supporting the notion that it is less compact. To more definitively assess the relative lobe structures, we employed Expressed Protein Ligation (EPL)[26] to generate full-length NpuN with segmental isotope labeling either on $NpuN_1$ or $NpuN_2$. Residue 59 in NpuN, eight residues downstream of the $NpuN_1$−$NpuN_2$ junction, is a native cysteine and provides a reactive handle for ligation of the lobes while retaining the wild-type NpuN sequence (Figure 3c and S8). This ligation approach allows for the selective observation of either $NpuN_1$ or $NpuN_2$ resonances in the context of full-length NpuN. The $^1H-^{15}N$ HSQC spectrum of the $NpuN_1$-labeled sample showed markedly greater proton chemical shift dispersion than that of the $NpuN_2$-labeled sample (Figure 3d). These spectra confirm the bipartite structure of the N-intein, with an at least partially folded $NpuN_1$ lobe and a disordered $NpuN_2$ lobe. Interestingly, many of the $NpuN_1$ resonances that were visible on a 500 MHz spectrometer were not visible at 800 MHz due to chemical exchange line-broadening, whereas $NpuN_2$ resonances were less affected by the field strength of the spectrometer (Figure S9). Thus, the two connected lobes have different chemical exchange rates indicating internal dynamics on different time scales,[27] consistent with their distinct states of compaction. Importantly, both segmentally labeled N-inteins could bind NpuC to yield correctly folded, functional complexes (Figure S10).

The $NpuN_2$−NpuC Interaction Represents a Binding Intermediate. Given their distinct structures, we hypothesized that $NpuN_1$ and $NpuN_2$ play unique roles in engaging NpuC. To test this, we analyzed mixtures of each isolated lobe with NpuC by SEC-MALS at concentrations ranging from 7 to 50 μM. No interaction was observed between the two N-intein lobes or between $NpuN_1$ and NpuC at any concentration tested (Figure S11a−c). Remarkably, a 1:1 complex between $NpuN_2$ and NpuC was unambiguously observed by SEC-MALS (Figures 4a and S11d). Notably, this complex eluted later than $NpuN_2$ alone, indicating that this lobe is compacted upon binding the C-intein (Figures 4a and S7f). Furthermore, binding was weakened by increasing the ionic strength, demonstrating that there is an electrostatic component to this interaction (Figure S11e,f).

dx.doi.org/10.1021/ja4104364 | J. Am. Chem. Soc. 2013, 135, 18673−18681

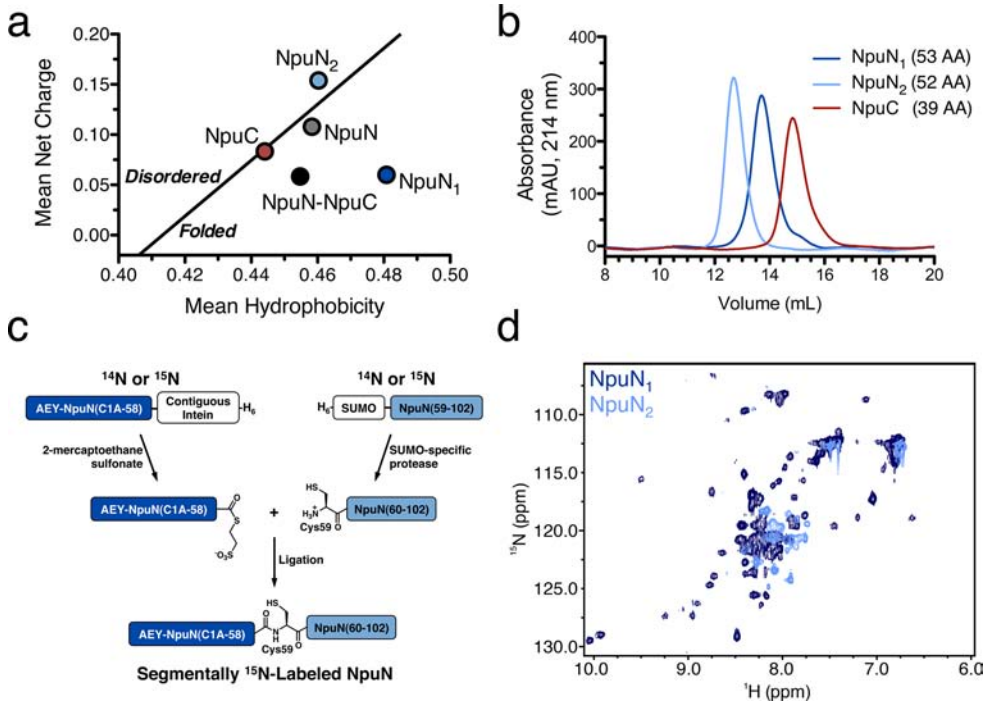


Figure 3. Characterization of the NpuN lobes. (a) Charge-hydrophobicity plot comparing Npu fragments, complex, and N-intein lobes. Mean hydrophobicity (H) is calculated on a normalized Kyte-Doolittle scale, and mean net charge (R) is the absolute value. The solid line delineating disordered and folded proteins is empirically defined as $R = 2.785 \cdot H - 1.151$.[23] (b) SEC of $NpuN_1$ (dark blue, 7 μM), $NpuN_2$ (light blue, 7 μM) and NpuC (red, 7 μM). (c) Scheme depicting the segmental isotopic labeling of NpuN using Expressed Protein Ligation. (d) $^1H-^{15}N$ HSQC spectra of NpuN segmentally ^{15}N-labeled on $NpuN_1$ (dark blue, 100 μM) and $NpuN_2$ (light blue, 100 μM) collected at 500 MHz. NMR data was acquired at 25 °C in a pH 6.5 buffer containing 25 mM phosphates, 100 mM NaCl, and 1 mM DTT.

Since NpuC preferentially binds $NpuN_2$ over $NpuN_1$, we postulate that the $NpuN_2$–NpuC complex represents an intermediate in the coupled binding and folding pathway for DnaE inteins. Thus, we sought to characterize the structure of this complex. The $^1H-^{15}N$ HSQC spectrum of isolated $NpuN_2$ in the absence and presence of NpuC suggested a global structural transition from a highly flexible state to a collapsed state (Figure 4b). On the basis of their deviation from average random coil values, the backbone ^1H and ^{13}C chemical shifts for $NpuN_2$ both alone and in complex with NpuC were used to predict secondary structure content in this lobe (Figures 4c and S12a–d). In the absence of NpuC, $NpuN_2$ had no predicted secondary structural elements, consistent with it being an IDP. In the presence of NpuC, however, $NpuN_2$ was predicted to have α-helical and β-sheet content in the same regions as in the native Npu complex, particularly for residues at the interface with NpuC (Figure 4c).

Unlike the global conformational change seen for $NpuN_2$ upon binding NpuC, only a fraction of the NpuC backbone resonances showed chemical shift perturbations upon binding $NpuN_2$, while the remainder retained the chemical shifts seen for isolated NpuC (Figure S12e,f). Notably, for those resonances that changed, the magnitude and direction of the chemical shift perturbation was similar to that seen for the same residues in full-length complex, suggesting native-like structure at these positions (Figures 4d and S12f). The perturbed residues primarily correspond to a contiguous stretch of nine amino acids at the N-terminus of NpuC (104–112), which make both ionic and β-strand pairing interactions with $NpuN_2$ in the full-length complex (Figure 4e). To further confirm that these C-intein residues were important for $NpuN_2$ binding and collapse, we generated a fusion between $NpuN_2$ and the first 13 residues of NpuC, NpuN(51–115). As expected, the $^1H-^{15}N$ HSQC spectrum of Npu(51–115) described a well-folded protein, indicating that this short stretch of NpuC was sufficient to collapse $NpuN_2$ (Figure 4f).

Split Intein Assembly Is Multiphasic and Electrostatically Driven. The presence of a well-defined binding intermediate such as the $NpuN_2$–NpuC complex indicates that split intein assembly is multiphasic and that these phases may be discretely observable. The fluorescence of the sole tryptophan in the Npu complex (W47, Figure 1b) is quenched upon Npu assembly and can be used to monitor binding (Figure 5a).[12] Using this probe, we measured an equilibrium dissociation constant (K_D) for Npu of 1.2 ± 0.8 nM (Figure S13). In kinetic measurements using stopped-flow fluorescence, only a single binding phase was observable with a k_{on} of (8.9 ± 0.3) × 10^5 M^{-1} s^{-1} (Figures 5b,c and S14a) and an off-rate too slow to extract directly from our kinetic data (when calculated from K_D and k_{on}, the k_{off} is between 10^{-4} and 10^{-3} s^{-1}). W47 lies in $NpuN_1$, which our data indicate is compact throughout split intein assembly and may not be sensitive to multiple steps in the binding process. Thus, we engineered a new tryptophan in the more dynamic $NpuN_2$ (Y58W, Figure 1b) and mutated W47 to tyrosine to silence $NpuN_1$ fluorescence. These mutations modestly affected the binding equilibrium (K_D = 3.4 ± 1.0 nM, Figure S13) and had no effect on splicing activity (Figure S15). Unlike W47 fluorescence, W58 fluorescence increased dramatically upon fragment association, consistent with $NpuN_2$ undergoing a folding transition upon binding NpuC (Figure 5d). Stopped-flow fluorescence of this mutant divulged two binding phases (Figures 5e,f and S14b,c), an initial concentration-dependent fast phase with a k_{on} of (2.7 ± 0.4) × 10^5 M^{-1} s^{-1} followed by a slower concentration-

dx.doi.org/10.1021/ja4104364 | J. Am. Chem. Soc. 2013, 135, 18673–18681

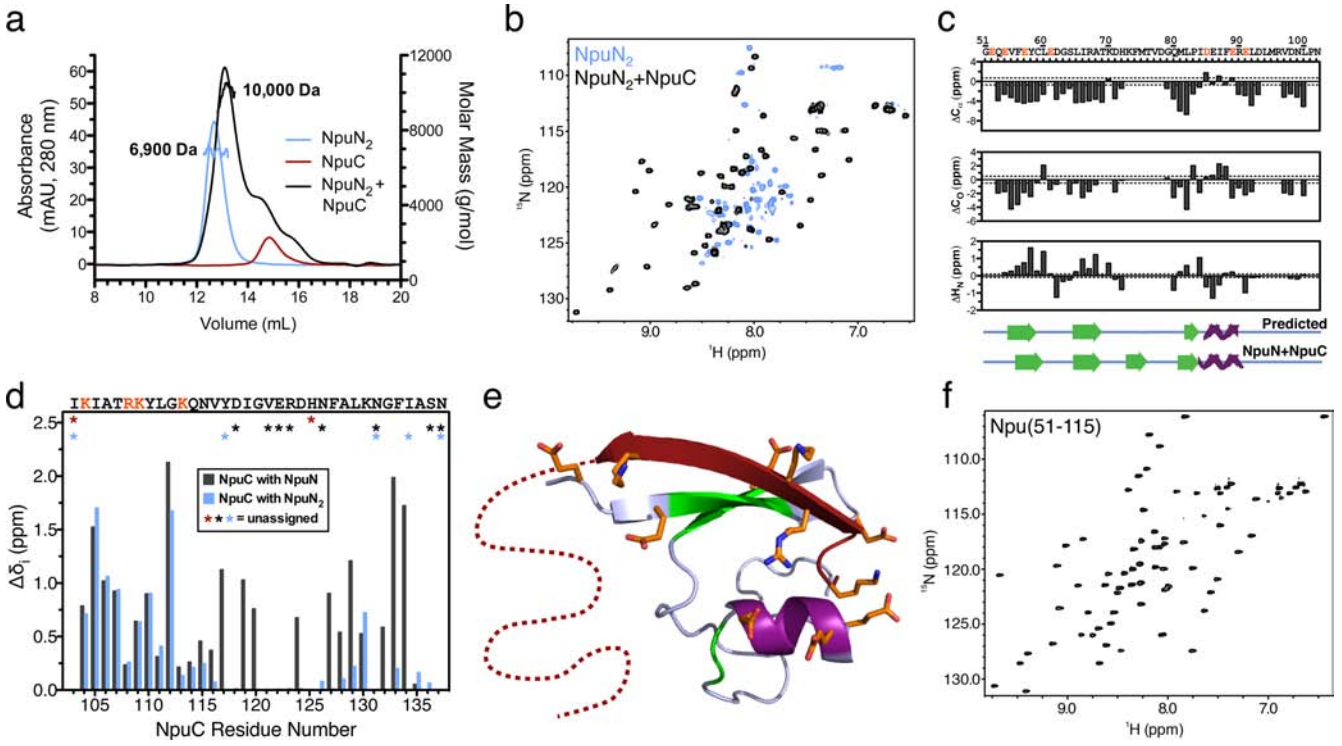


Figure 4. Characterization of the $NpuN_2$–NpuC interaction. (a) SEC-MALS of isolated $NpuN_2$ (light blue, 62.5 μM, expected MW = 6098 Da), NpuC (red, 7 μM), and an equimolar mixture of $NpuN_2$ and NpuC (black, 50 μM, expected MW = 10541 Da). (b) $^1H-^{15}N$ HSQC spectra of $NpuN_2$ alone (light blue, 250 μM) and with equimolar unlabeled NpuC (black, 250 μM) collected at 800 MHz. (c) Chemical shift index values ($\Delta C/H = \delta C/H_{Observed} - \delta C/H_{RandomCoil}$) for C$\alpha$, CO, and HN atoms for $NpuN_2$ in complex with NpuC. The consensus secondary structure prediction is shown below compared with $NpuN_2$ in the context of the full Npu complex (as seen in PDB 2KEQ).[10] The $NpuN_2$ sequence is given above, with important anionic residues in orange. (d) Composite ^1H and ^{15}N backbone chemical shift perturbation values for NpuC in complex with NpuN (black) or $NpuN_2$ (light blue) calculated relative to isolated NpuC. $\Delta\delta_i = [(\Delta\delta_{H,complex} - \Delta\delta_{H,alone})_i^2 + 0.11(\Delta\delta_{N,complex} - \Delta\delta_{N,alone})_i^2]^{1/2}$ The NpuC sequence is given above, with important cationic residues in orange. Below the sequence, an asterisk marks unassigned residues for isolated NpuC (red), the complex with NpuN (black), or the complex with $NpuN_2$ (light blue). (e) Rendering of the $NpuN_2$–NpuC interaction, as seen in the native Npu complex (PDB 2KEQ).[10] $NpuN_2$ is light blue, NpuC is red, and charged residues involved in intermolecular electrostatic interactions are highlighted as orange sticks. Predicted β-sheet and α-helical regions of $NpuN_2$ are colored in green and purple, respectively. Residues 103–115 of NpuC are rendered as a ribbon, and the remainder of the sequence is rendered as a dotted line. (f) $^1H-^{15}N$ HSQC spectrum of Npu(51–115) collected at 600 MHz. NMR data in panels b, c, d, and f were acquired at 25 °C in a pH 6.5 buffer containing 25 mM phosphates, 100 mM NaCl, and 1 mM DTT.

independent phase (k_{obs} = 0.04 ± 0.02 s^{-1}) that was only marginally faster than the overall protein splicing rate (k_{splice} = ~0.01 s^{-1}, Figure S15). Notably, for both the wild-type and mutant inteins, the K_D increased and k_{on} decreased in higher ionic strength buffers (Figures S13 and S14).

Our fluorescence binding measurements are consistent with a rapid, electrostatically driven encounter between two fragments followed by a collapse of the structure. As discussed in the preceding sections, the electrostatic component to binding most likely originates from the interaction between $NpuN_2$ and NpuC. In a previous study, we mutated residues involved in several intermolecular ionic interactions throughout the fragment interface in Npu, swapping anionic residues in NpuN for cationic ones and cationic residues in NpuC for anionic ones.[12] We found that introducing repulsive electrostatic interactions at the fragment interface often diminished splicing activity, and compensatory electrostatic mutations typically recovered activity. Importantly, this effect on activity correlated with binding affinity. Furthermore, the region of the intein most sensitive to electrostatic mutations comprised an ion triad on the face of the α-helix at the core of the $NpuN_2$–NpuC interface (D85, E89, and R108), consistent with the biophysical data presented herein.

Split Intein Fragments Associate via a "Capture and Collapse" Mechanism. On the basis of the foregoing biophysical experiments, we propose that Npu fragment assembly is a multistep process in which NpuC is first engaged by $NpuN_2$, followed by $NpuN_1$ (Figure 6). The initial encounter between the fragments is dictated by electrostatic interactions between the disordered, highly cationic NpuC and the extended, highly anionic second lobe of NpuN. This encounter complex resolves to an intermediate that has secondary structure and intermolecular interactions resembling half of the native Npu complex. After this initial "capture", the flexible regions "collapse" into an ordered state that is further stabilized by hydrophobic interactions between NpuC and the well-ordered first lobe of NpuN. The resulting complex has intertwined fragments, and this topology is presumably only accessible due to the intrinsic disorder of the starting materials. This "capture and collapse" model is meant to provide a framework for understanding intein fragment assembly, and we recognize that there could be additional folding steps within each phase that are not detectable using the biophysical approaches employed herein.

Native Topology Is Important for Split Intein Stability and Function. NpuC intercalates between the NpuN lobes,

dx.doi.org/10.1021/ja4104364 | J. Am. Chem. Soc. 2013, 135, 18673–18681

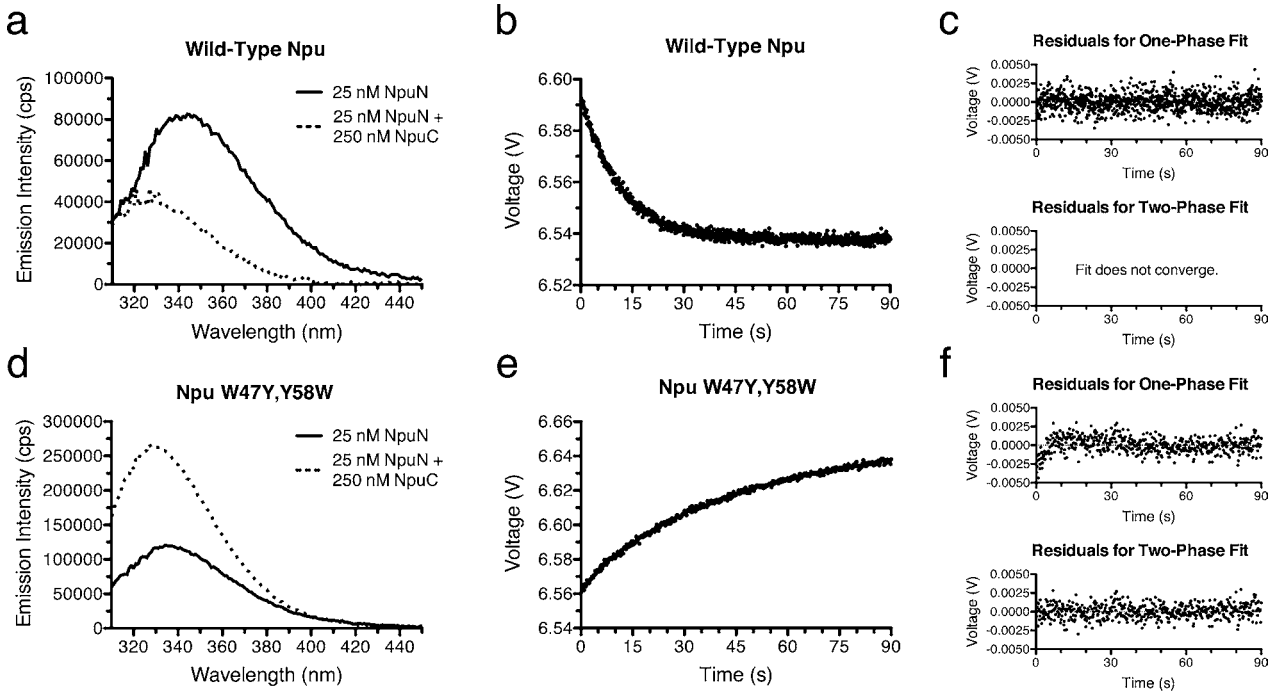


Figure 5. Intrinsic fluorescence binding measurements: (a) Tryptophan fluorescence of wild-type NpuN in the absence and presence of NpuC (λ_{ex} = 290 nm); (b) Tryptophan fluorescence of wild-type Npu upon stopped-flow mixing of fragments (λ_{ex} = 290 nm, λ_{em} = >320 nm); (c) Residuals from fits to a one-phase (top) and two-phase (bottom) binding model for wild-type Npu; (d) Tryptophan fluorescence of NpuN with the W47Y and Y58W mutations in the absence and presence of NpuC (λ_{ex} = 290 nm); (e) Tryptophan fluorescence of Npu W47Y,Y58W upon stopped-flow mixing of fragments (λ_{ex} = 290 nm, λ_{em} = >320 nm); (f) Residuals from fits to a one-phase (top) and two-phase (bottom) binding model for Npu W47Y,Y58W.

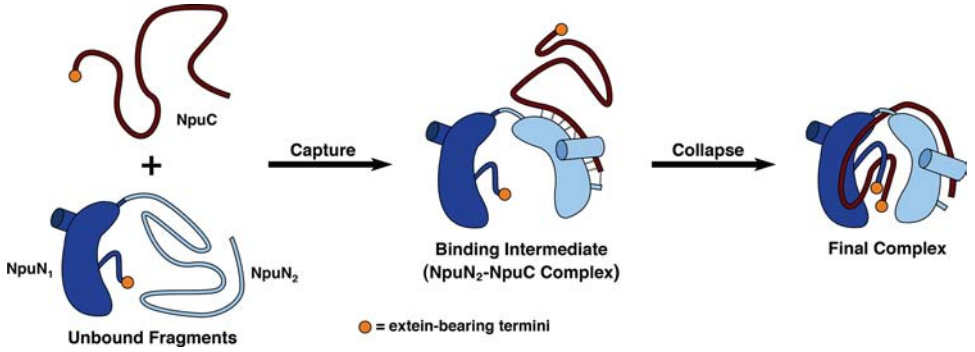


Figure 6. The "capture and collapse" mechanism of split intein assembly. An N-terminal segment of the disordered C-intein is captured by the extended second lobe of the N-intein with compaction of that lobe into a native-like structure. This intermediate then collapses as the remainder of the C-intein docks into the preorganized first lobe of the N-intein.

allowing it to engage $NpuN_2$ through electrostatic interactions and $NpuN_1$ through hydrophobic interactions (Figure 2). These segregated intermolecular interactions and the resulting interweaving of the chains should not only drive assembly, as described above, but also prevent the bound fragments from rapidly dissociating, thereby precluding splicing. To test this notion, we engineered two Npu variants with altered primary sequence topology. In the first system, we permuted the split intein sequence to yield two new proteins related by a pseudosymmetry axis (Figures 7a and S2). One protein, Npu(51−115) described above represents the well-folded binding intermediate composed of $NpuN_2$ and a segment of NpuC, where all crucial ionic interactions are presatisfied (Figure 4e,f). The second protein, a fusion of $NpuN_1$ and the remainder of NpuC, presatisfies the hydrophobic interactions between these two protomers (Figure 2a,b). A complex of these two permuted fragments should associate slowly and dissociate rapidly, given the lack of key intermolecular interactions and entwined fragments. As expected, the permuted intein fragments could assemble into the correct intein fold, as evidenced by their capacity to carry out protein splicing when fused to model extein domains, ubiquitin (Ub) and SUMO (Figures 7b,c and S16). However, the rate of splicing was extremely concentration dependent, indicating that association is rate-limiting. It is noteworthy that, even at high concentrations, splicing was dramatically slower for this permuted construct than wild-type Npu (days instead of minutes). As the complex was too unstable to measure binding parameters, we attribute this low efficiency at high concentrations to a short lifetime of the active complex caused by a fast off-rate.

dx.doi.org/10.1021/ja4104364 | J. Am. Chem. Soc. 2013, 135, 18673−18681

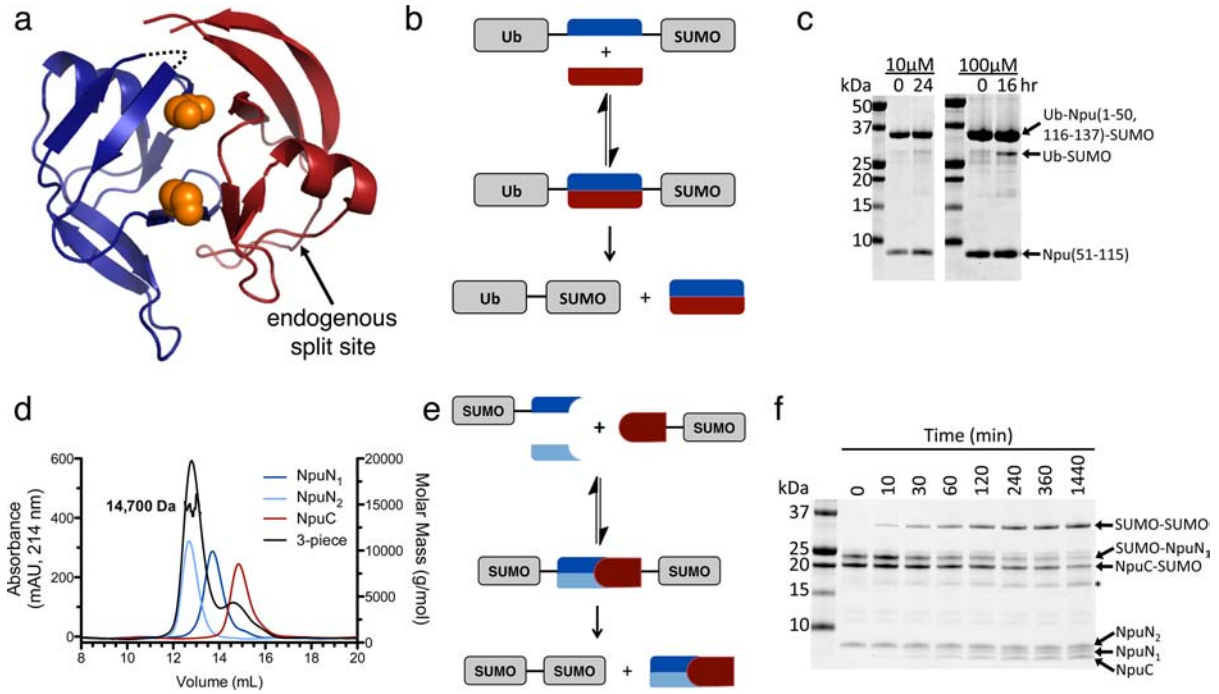


Figure 7. Split intein assembly and function with engineered sequence topologies. (a) Structural representation of the two intein regions in the permuted Npu constructs (based on PDB 2KEQ).[10] The dashed line represents a diglycine linker introduced between residues 1−50 and 116−137. (b) Scheme showing protein splicing with the permuted Npu. Ub (ubiquitin) and SUMO are model N- and C-exteins. (c) Proteins splicing of permuted Npu at 10 and 100 μM fragment concentrations. Reaction mixtures when analyzed at indicated time-points by SDS-PAGE with coomassie staining. (d) SEC-MALS of the three-piece Npu complex (black, 7 μM of each fragment, expected MW = 16649 Da) compared with the individual components ($NpuN_1$, dark blue; $NpuN_2$, light blue; NpuC, red; each at 7 μM). (e) Scheme showing protein splicing with the three-piece Npu. (f) Splicing activity of the three-piece Npu at 5 μM fragment concentrations. Reaction mixtures where analyzed at indicated time-points by SDS-PAGE with coomassie staining. The asterisk indicates a side product, most likely SUMO due to premature N-extein cleavage.

In the second system, we utilized the isolated NpuN lobes in conjunction with NpuC to generate a three-piece intein. While these constructs must contend with the entropic cost of forming a ternary complex, the key intermolecular interactions between each lobe and NpuC are still present. The three-piece complex was stable at low-micromolar concentrations (Figure 7d) where the binary interaction between $NpuN_2$ and NpuC was barely visible (Figure S11d). Since NpuC cannot independently interact with $NpuN_1$ (Figure S11b,c), this result shows that the lobes bind NpuC cooperatively. Notably, the resulting ternary complex could carry out protein splicing in the context of model SUMO extein domains with substantially higher efficiency than the permuted complex (Figures 7e,f and S17), demonstrating the kinetic stability of the active complex.

■ DISCUSSION AND CONCLUSIONS

Inteins have a unique fold in which the polypeptide chain traverses back and forth between two symmetry-related halves (Figure 1b). Given this, it is remarkable that some inteins are naturally split, as this topology precludes a binding mechanism involving natively structured fragments. In this report, we describe the structure of split intein fragments and determine their mechanism of association. Using a variety of biophysical techniques coupled with protein engineering and protein chemistry, we show that the N-intein fragment has a unique bipartite structure comprised of one at least partially folded lobe tethered to an intrinsically disordered region of equal length. This extended region of the N-intein captures the completely disordered C-intein through electrostatic interactions, leading to a native-like intermediate. This intermediate then further collapses onto the more structured N-intein lobe, resulting in a functional intein domain with entangled fragments.

To function, split intein fragments must not only find one another efficiently, but also bind to form a stable complex that can persist throughout the duration of the complex multistep splicing reaction, which can last from seconds to hours.[15] Our studies indicate that split DnaE inteins have evolved several unique structural features that fulfill these requirements: (1) The initial encounter complex is dictated by two disordered protein regions, which should provide larger capture radii for an enhanced collision probability.[28] (2) This enhanced on-rate is further reinforced by electrostatic steering of these fragments.[29] (3) The slow off-rate is presumably governed by the retention of these ionic interactions in the final complex (Figure 2c,d),[12] as well as significant hydrophobic interactions between NpuC and $NpuN_1$ (Figure 2a,b). (4) The disordered regions of the fragments provide access to an interwoven topology, which further reduces the off-rate (Figure 1b).

The significance of this interweaving for intein function was demonstrated using two engineered systems: one with a permuted primary sequence and another involving a three-piece intein. It is noteworthy that both of these systems could, in principle, be used to design novel intein-based technologies. The permuted system, with poor binding, could be enhanced by fusion to an inducible dimerization system, thus making a conditionally splicing Npu. To our knowledge, such a system has never been developed with a highly efficient intein such as Npu, as the native fragments of this and other naturally split inteins spontaneously associate. Instead, conditional protein

dx.doi.org/10.1021/ja4104364 | J. Am. Chem. Soc. 2013, 135, 18673−18681

splicing (CPS) modules have traditionally been developed using less efficient artificially split inteins, which require fusion to dimerization domains as they typically lack the key properties required to function well *in trans*: charge complementarity and well-behaved disordered fragments. Construction of a controllable fast-splicing Npu intein would be extremely desirable, as CPS systems show promise as tools to control protein function in vivo in response to external stimuli.[9] The three-piece system is particularly intriguing, as it is implicitly conditional: the central piece, $NpuN_2$, is effectively a trigger for protein oligomerization and splicing. Further design and optimization of both the permuted and three-piece Npu inteins should lead to the development of useful tools for biochemistry and cell biology.

Split inteins likely arose from their contiguous predecessors due to DNA inversions or aberrant gene insertions and excisions.[5,6] These genomic rearrangements independently occurred several times, as evidenced by the diversity of split inteins, their host proteins, and their host organisms.[5,6,30] Interestingly, even outside of the DnaE family described in this study, many split inteins are cut at the same point on the intein fold and exhibit dramatic charge segregation between fragments (Figure S1c). Furthermore, homology modeling of non-DnaE split inteins predicts that many have intermolecular electrostatic interactions at the ridge of their horseshoe-like structure, as seen in the DnaE inteins.[6] Thus, the efficient folding mechanism for Npu is most likely a general solution at the protein level to the rare but repeated fracturing of intein genes at the DNA level. Given that split inteins assemble essential proteins (e.g., DNA polymerase alpha subunit, DNA helicase, and ribonucleotide reductase), it is not surprising that this solution entails a number of traits that ensure highly efficient binding. Ultimately, the molecular details of this mechanism, laid out in this study, not only provide insight into the evolution of efficient biomolecular recognition, but they will also lay the groundwork for future engineering efforts on this unusual and useful class of proteins.

EXPERIMENTAL SECTION

Sample Preparation. NpuN and NpuC were purified as previously described.[22] Isolated $NpuN_1$ and $NpuN_2$ proteins and the permuted intein constructs were generated similarly to NpuN. For functional analyses, wild-type sequences were used, and for biophysical measurements, a C1A mutation in NpuN and N137A mutation in NpuC were introduced to inactivate the intein. For segmental labeling of NpuN, the protein was split between Y58 and C59 and the two reactive fragments were produced and ligated using standard protocols. For isotope labeling, *Escherichia coli* were grown in minimal M9 medium enriched with $^{15}NH_4Cl$ and [U−^{13}C]-D-glucose as the sole sources of nitrogen and carbon.

Biophysical Measurements. SEC-MALS experiments were carried out on a Superdex 75 10/300 column (volume = 24 mL) at 4 °C. For MALS analysis, light scattering and refractive index measurements were averaged across the entire width above half of the peak height to extract the molar mass. All NMR experiments were carried out at 25 °C in a pH 6.5 phosphate buffer using standard techniques. For ANS fluorescence, samples were excited at 370 nm and emission spectra were recorded between 400 and 720 nm. For steady-state tryptophan fluorescence, samples were excited at 290 or 295 nm and emission was recorded between 300 and 500 nm. For stopped-flow tryptophan fluorescence, samples were excited at 290 nm and emission was recorded above a 320 nm cutoff filter.

ASSOCIATED CONTENT

Supporting Information

Full methods and experimental data, including protein semisynthesis and purification protocols, characterization of proteins, and additional biophysical data. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

muir@princeton.edu

Author Contributions

§N.H.S. and E.E. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank the members of the Muir laboratory for valuable discussions and Istvan Pelczer of the Princeton University NMR facility for his generosity. This work was supported by the U.S. National Institutes of Health (NIH grant GM086868). NMR resources at NYSBC were supported by NIGMS P41-GM066354.

REFERENCES

(1) Saleh, L.; Perler, F. B. *Chem. Rec.* **2006**, *6*, 183−93.
(2) Pietrokovski, S. *Trends Genet.* **2001**, *17*, 465−72.
(3) Perler, F. B. *Nucleic Acids Res.* **2002**, *30*, 383−4.
(4) Vila-Perelló, M.; Muir, T. W. *Cell* **2010**, *143*, 191−200.
(5) Caspi, J.; Amitai, G.; Belenkiy, O.; Pietrokovski, S. *Mol. Microbiol.* **2003**, *50*, 1569−77.
(6) Dassa, B.; London, N.; Stoddard, B. L.; Schueler-Furman, O.; Pietrokovski, S. *Nucleic Acids Res.* **2009**, *37*, 2560−73.
(7) Southworth, M. W.; Adam, E.; Panne, D.; Byer, R.; Kautz, R.; Perler, F. B. *EMBO J.* **1998**, *17*, 918−26.
(8) Yamazaki, T.; Otomo, T.; Oda, N.; Kyogoku, Y.; Uegaki, K.; Ito, N.; Ishino, Y.; Nakamura, H. *J. Am. Chem. Soc.* **1998**, *120*, 5591−5592.
(9) Mootz, H. D.; Blum, E. S.; Tyszkiewicz, A. B.; Muir, T. W. *J. Am. Chem. Soc.* **2003**, *125*, 10561−9.
(10) Oeemig, J. S.; Aranko, A. S.; Djupsjöbacka, J.; Heinämäki, K.; Iwaï, H. *FEBS Lett.* **2009**, *583*, 1451−1456.
(11) Shi, J.; Muir, T. W. *J. Am. Chem. Soc.* **2005**, *127*, 6198−206.
(12) Shah, N. H.; Vila-Perelló, M.; Muir, T. W. *Angew. Chem., Int. Ed.* **2011**, *50*, 6511−5.
(13) Zheng, Y.; Wu, Q.; Wang, C.; Xu, M.-Q.; Liu, Y. *Biosci. Rep.* **2012**, *32*, 433−42.
(14) Zettler, J.; Schütz, V.; Mootz, H. D. *FEBS Lett.* **2009**, *583*, 909−14.
(15) Shah, N. H.; Dann, G. P.; Vila-Perelló, M.; Liu, Z.; Muir, T. W. *J. Am. Chem. Soc.* **2012**, *134*, 11338−41.
(16) Muona, M.; Aranko, A. S.; Raulinaitis, V.; Iwaï, H. *Nat. Protoc.* **2010**, *5*, 574−87.
(17) Borra, R.; Dong, D.; Elnagar, A. Y.; Woldemariam, G. A.; Camarero, J. A. *J. Am. Chem. Soc.* **2012**, *134*, 6344−53.
(18) Vila-Perelló, M.; Liu, Z.; Shah, N. H.; Willis, J. A.; Idoyaga, J.; Muir, T. W. *J. Am. Chem. Soc.* **2013**, *135*, 286−92.
(19) Jagadish, K.; Borra, R.; Lacey, V.; Majumder, S.; Shekhtman, A.; Wang, L.; Camarero, J. A. *Angew. Chem., Int. Ed.* **2013**, *52*, 3126−31.
(20) Sorci, M.; Dassa, B.; Liu, H.; Anand, G.; Dutta, A. K.; Pietrokovski, S.; Belfort, M.; Belfort, G. *Anal. Chem.* **2013**, *85*, 6080−8.
(21) Shah, N. H.; Muir, T. W. *Isr. J. Chem.* **2011**, *51*, 854−861.
(22) Shah, N. H.; Eryilmaz, E.; Cowburn, D.; Muir, T. W. *J. Am. Chem. Soc.* **2013**, *135*, 5839−47.
(23) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. *Proteins: Struct., Funct., Genet.* **2000**, *41*, 415−427.

(24) Müller-Späth, S.; Soranno, A.; Hirschfeld, V.; Hofmann, H.; Rüegger, S.; Reymond, L.; Nettels, D.; Schuler, B. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 14609−14.
(25) Semisotnov, G. V.; Rodionova, N. A.; Razgulyaev, O. I.; Uversky, V. N.; Gripas, A. F.; Gilmanshin, R. I. *Biopolymers* **1991**, *31*, 119−128.
(26) Xu, R.; Ayers, B.; Cowburn, D.; Muir, T. W. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 388−93.
(27) Millet, O.; Loria, J. P.; Kroenke, C. D.; Pons, M.; Palmer, A. G. *J. Am. Chem. Soc.* **2000**, *122*, 2867−2887.
(28) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8868−73.
(29) Schreiber, G.; Fersht, A. R. *Nat. Struct. Biol.* **1996**, *3*, 427−31.
(30) Choi, J. J.; Nam, K. H.; Min, B.; Kim, S.-J.; Söll, D.; Kwon, S.-T. *J. Mol. Biol.* **2006**, *356*, 1093−106.