Published in final edited form as:

J Mol Biol. 2009 November 13; 393(5): 1106. doi:10.1016/j.jmb.2009.08.074.

Selection and structure of hyperactive inteins: peripheral changes relayed to the catalytic center

Kaori Hiraga¹, Ikko Soga^{1,2}, John T. Dansereau¹, Brian Pereira^{1,3}, Victoria Derbyshire¹, Zhenming Du⁴, Chunyu Wang⁴, Patrick Van Roey¹, Georges Belfort³, and Marlene Belfort^{1,2,✉}

¹Wadsworth Center, New York State Department of Health, 150 New Scotland Avenue, Albany, NY 12208

²School of Public Health, State University of New York at Albany, Albany, NY 12201-2002

³Department of Chemical and Biological Engineering, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180

⁴Biology Department, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180

Abstract

Inteins are phylogenetically diverse self-splicing proteins that are of great functional, evolutionary, biotechnological and medical interest. To address the relationship between intein structure and function, particularly with respect to regulating the splicing reaction, and to groom inteins for application, we developed a phage display system to extend current *in vivo* selection for enhanced intein function to selection *in vitro*. We thereby isolated inteins that can function under excursions in temperature, pH and denaturing environment. Remarkably, most mutations mapped to the surface of the intein, remote from the active site. We chose two mutants with enhanced splicing activity for crystallography, one of which was also subjected to NMR analysis. These studies define a "ripple effect", whereby mutations in peripheral non-catalytic residues can cause subtle allosteric changes in the active site environment in a way that facilitates intein activity. Altered salt-bridge formation and chemical shift changes of the mutant inteins provide a molecular rationale for their phenotypes. These fundamental insights will advance the utility of inteins in chemical biology, biotechnology and medicine.

Keywords

Intramolecular intein communication; robust intein mutants; intein application

INTRODUCTION

Inteins interrupt genes in hundreds of organisms, in all three domains of life, archaea, bacteria and eukaryotes. They are removed by an autocatalytic splicing reaction, which joins the flanking polypeptides, called exteins, to yield a functional protein ¹. Inteins and exteins also play an increasingly important role in chemical biology and in areas such as biotechnology, drug delivery and sensor development ²⁻⁸. Inteins can be used in different extein contexts to ligate two foreign peptides ^{3; 5}. Inteins also have the potential to be used as post-translational molecular switches for the control and delivery of target proteins, because they inactivate host

✉Correspondence to: Marlene Belfort, Phone: 518-473-3345, FAX: 518-474-3181, belfort@wadsworth.org.

proteins and splicing restores protein function^{9; 10}. Furthermore, splicing can be arrested at different stages, facilitating preferential cleavage of either the N- or C-extein^{1; 11}. Cleavage too has found extensive application, in protein release and bioseparations^{2; 12–14}. Moreover, inteins are novel potential targets for antimycobacterial drugs¹⁵. There is therefore an enormous need to understand how activity of self-splicing proteins is programmed into their structure; such knowledge is required for both a mechanistic appreciation of how inteins function in a diverse array of organisms in different environments, and how intein activity can be modulated for biotechnological and biomedical application.

Directed evolution is widely used to generate proteins with modified functionality. Using a mutagenesis and selection approach, several novel inteins have been developed. These include temperature-sensitive intein variants^{16–18}, pH-sensitive intein mutants², and inteins controlled by small molecules^{4; 19}. In our laboratory, an *Escherichia coli*-based selection system was developed to isolate a number of small inteins that lack a central endonuclease domain, which is dispensable to splicing^{20; 21}. This selection was also used to isolate an intein with improved splicing activity (splicing mutant, SM), and an intein with pH controllable cleavage activity (cleavage mutant, CM)². However, such selection systems cannot be used to derive inteins that function in non-natural or extreme environments.

To engineer novel inteins, we developed a phage display system for the selection of enhanced splicing activity under a range of conditions. Because phage particles can be viable in non-natural environments, and since an intein-containing protein can be displayed on the surface of a phage particle, one can screen large libraries of displayed inteins for specific phenotypes under desirable conditions.

We engineered a T7 phage to express intein derivatives inserted into a chitin binding domain (CBD), rendering the CBD incapable of binding chitin. Selection for inteins that splice under specific conditions is based on regenerating an intact CBD. In this work, we isolated intein mutants with enhanced splicing activity, of which some have subtle changes in pH sensitivity, thereby extending the range of intein utility. Strikingly, the mutations were remote from the catalytic center and mapped broadly over the intein surface. Structural analysis by X-ray crystallography and NMR spectroscopy provides a molecular rationale for the mutant phenotypes, indicating that residues at the periphery of the intein can cause allosteric changes, to influence dynamics at the active site, a phenomenon we refer to as a "ripple effect".

RESULTS

Construction of an *in vitro* selection system for intein splicing activity

In our phage display system, designed to select for splicing activity, the intein was inserted into the CBD and the fusion precursor proteins were displayed on the surface of T7 phage (Fig. 1A). We selected the CBD of chitinase A1 from *Bacillus circulans* WL-12. This is a 52-amino acid protein that retains binding activity over a wide range of temperatures and pHs^{22; 23}. The intein used in this study was the 139-amino acid $\Delta\Delta I_{hh}$ mini-intein derived from the splicing domain of the *Mycobacterium tuberculosis* RecA intein (Mtu RecA intein), with a 7-amino acid β -turn sequence from the auto-processing domain of the *Drosophila* Hedgehog protein at the deletion junction²¹. We showed that the activity of $\Delta\Delta I_{hh}$ is much lower than that of the full-length intein, but that a single mutation V67L (SM)² can restore the splicing activity of the $\Delta\Delta I_{hh}$ inteins²¹. We also determined the X-ray structures of $\Delta\Delta I_{hh}$ and $\Delta\Delta I_{hh}$ -SM²⁴, and observed that the mutation does not affect the molecular conformation, but that it results in improved packing of side chains in a hydrophobic pocket and thus increases the stability of the molecule.

The $\Delta\Delta I_{hh}$ gene was inserted before the Cys30 codon of the CBD gene and the CBD::intein fusion was cloned into phage T7 downstream of the capsid gene. In principle, the phage that displays CBD::intein fusion proteins on the capsid is not able to bind to chitin because the CBD structure is disrupted by the intein insertion (Fig. 1A). Only if the intein is active and the full-length CBD is reconstituted by splicing will the phage particle be able to bind to chitin resin. The construct was designed with a thrombin protease cleavage site, such that the CBD can be cleaved off, and the phage particle that carries the DNA sequence of an active intein can then be eluted from the chitin beads (Fig. 1B).

To assess the selection scheme, we constructed recombinant T7 phage in which a series of inteins was inserted into the CBD, and the binding behavior of each phage was tested (Table 1). The positive control phage displaying an uninterrupted CBD bound to the chitin resin effectively, and about 25% of the phage were recovered from the resin by thrombin treatment (Table 1, line 1). The negative control phage displaying CBD containing an inactive intein (CBD:: $\Delta\Delta I_{hh}$ -AA) had low affinity for chitin resin, with only 0.03% recovery of phage, indicating that the intein insertion disrupted the CBD structure (Table 1, line 2). The binding of negative and positive control phage, cultured in the absence and the presence of zinc ion, which reversibly inhibits splicing²⁵, were consistent over a range of temperatures, from 30° to 50°C, and from pH 6.0 to 8.6, indicating that the phage display system is robust.

The binding of phage displaying the wild-type mini-intein (CBD:: $\Delta\Delta I_{hh}$) was as low as the inactive intein (0.03% recovery) (Table 1, line 3), whereas the phage displaying the highly active splicing mutant, SM (CBD:: $\Delta\Delta I_{hh}$ -SM), had higher affinity than the CBD:: $\Delta\Delta I_{hh}$ intein, with >100-fold more phage (~4%) recovered from the chitin resin (Table 1, line 4). First, it is noteworthy that although the recovery of the SM phage was >100-fold that of the wild-type $\Delta\Delta I_{hh}$, its splicing activity was ~3-fold greater (unpublished results; see also below), indicating that recovery does not correlate directly with splicing. This discrepancy is likely due to 5–10 CBD::inteins being displayed on each phage particle, resulting in high avidity for the chitin resin. Additionally, there may be enhanced stability or other selective advantage of CBD phage that have lost the intein. Second, it is worth noting that recovery of CBD:: $\Delta\Delta I_{hh}$ -SM was only ~15% of that of CBD alone, (*cf* lines 1 and 4), suggesting that splicing was incomplete, possibly due to a level of unproductive N-terminal cleavage.

In this selection system, it was necessary to prevent spontaneous splicing under physiological conditions, during phage amplification. We therefore cultured the phage in the presence of zinc ion. Indeed, most splicing of CBD:: $\Delta\Delta I_{hh}$ -SM was prevented by the addition of 0.5 mM $ZnCl_2$ to the growth medium (0.06% phage recovery) (Table 1, line 5), and splicing could be restored *in vitro* by the addition of 1 mM EDTA at pH 6.0 (1.8% recovery) (Table 1, line 6). Given the greater activity of the Mtu RecA intein at low pH (pH 6.0) than at high pH (pH >7.0)²⁵, we also wished to test our phage display system at several pHs. Phage recovery dropped at higher pH values, consistent with pH-dependent splicing (Table 1, lines 7–9).

Enrichment for improved splicing activity

To test the feasibility of using T7 phage display to select active inteins, the phage displaying low activity (CBD:: $\Delta\Delta I_{hh}$) and high activity (CBD:: $\Delta\Delta I_{hh}$ -SM) were mixed at a ratio of 100:1 and multiple rounds of affinity selection for intact CBD were carried out at various pHs (Fig. 1C). The sequences of eluted phage after each round of selection were analyzed by a modified method of the single-nucleotide-discrimination PCR using *Taq* DNA polymerase (Supplementary Methods). At all pHs, the population of phage displaying CBD:: $\Delta\Delta I_{hh}$ -SM increased at each round of selection, indicating that the active SM intein was preferentially selected.

Next, randomly mutated mini-intein libraries, consisting of $\sim 10^6$ independent clones, were constructed by error-prone PCR and displayed on the surface of T7 phage in fusion with CBD (Fig. 1B). We used a low mutation library, M_L , and a high mutation library, M_H , in which we predicted 1–2 and 3–4 mutations per gene, respectively. The phage ($\sim 10^8$ pfu) were incubated in splicing buffer under various conditions of temperature, pH and denaturant concentration, and subjected to selection by chitin binding. The number of phage increased in the eluted fraction for successive rounds of selection, indicating that selection for higher intein activity occurred for all splicing conditions and with both libraries, with up to a ~ 700 -fold increase in eluted phage at the lower pHs (Fig. 1D, and Fig. S1).

Individual phage from the second, third and fourth rounds of selection were isolated and characterized by DNA sequencing, and by percent recovery from the chitin resin, revealing mutations over the length of the intein sequence that stimulate splicing (Fig. 2A and Supplementary Table 1). Interestingly, mutations were also found in the N-terminal extein (positions -1 and -2; black bars in Fig. 2A). Intein mutations were in the N-terminal segment of the intein (V22, D24, A34, A42, R54, R60, G63, I66, K74 and G85, blue bars), the Hedgehog loop that replaced the endonuclease domain (R96, V98, E99 and T100, red bars), and the C-terminal segment of the intein (F421, yellow bar). The V67L splicing mutation (SM), which had been isolated previously², was observed repeatedly, in both single and multiple mutants (Fig. 2A).

The mutations were mapped onto the three-dimensional structure of the $\Delta\Delta I_{hh}$ Mtu RecA intein²⁴ (Fig. 2B). Strikingly, all intein mutations were outside the active site (residues C1, H439 and N440, colored green). In addition to V67, the Hedgehog loop is a hot-spot for mutations, particularly residue R96 (Fig. 2A and B, red residues), suggesting that these mutations are among the most effective at improving splicing activity of the intein. The effect of V67L and the loop mutations was additive to the other beneficial mutations, and improved the phage recovery substantially compared with the single mutants (Fig. 2A, Table S1).

We also characterized the phage recovery of the most effective double-mutant phage cultured in the presence of zinc ion followed by *in vitro* splicing at pH 6.0 and 8.6 (Fig. 2C). Almost all mutants showed higher recovery than $\Delta\Delta I_{hh}$ (WT) and $\Delta\Delta I_{hh}$ -SM (SM) at both pHs. Among them, D24Y/V67L (clone 6), V67L/V98E (clone 29), and V67L/F421Y (clone 31), which were isolated at high pH, showed the highest recovery at pH 8.6. Of these, V67L/F421Y, a dominant clone in the M_H library after four rounds of selection at pH 8.6, had the highest recovery at both pHs. In contrast, G63S/R96C (clone 17) (Fig. 2A and 2C), which was isolated from the M_L library after four rounds of selection at pH 6.0, was recovered well at pH 6.0 but not at pH 8.6. These results confirm that the system is effective at isolating different mutations under specific selection pressures.

Activity of selected mutants

To characterize protein splicing activity of mutants *in vitro*, we wished to use a fluorescence-based system based on GFP::intein precursors²⁷. However, splicing of the mutants was so rapid, that the process took place *in vivo*, despite precursor being driven into inclusion bodies. Also, folding of the green-fluorescent protein (GFP) reporter became rate-limiting. We therefore used the same GFP::intein constructs in a gel-based system to identify the precursor and GFP splice product that had accumulated *in vivo* after 2 h of induction (Fig. 3A, schematic and gel). The gene encoding one of various mini-intein mutants was inserted after codon 129 of the *GFPuv*-encoding gene²⁸ and the solubilized GFP::intein precursors were separated on polyacrylamide gels. The extent of *in vivo* splicing of the single mutants tested at 2 h post-induction ranged from 1.1- to 1.8-fold that of the wild-type GFP:: $\Delta\Delta I_{hh}$, as compared to a 3-fold increase for the SM mutant GFP:: $\Delta\Delta I_{hh}$ -SM (Fig. 3A, bar chart). In contrast, the double-

mutants selected by phage display, which also contained the V67L (SM) mutation, were 3.8- to 5.2-fold elevated in splicing with respect to the wild-type intein.

To acquire rate information on activity of the constructs, we focused on C-terminal cleavage. We elected to measure C-terminal cleavage because this is an on-pathway, rate-limiting reaction ^{1, 8}. This rationale was reinforced by the observations that the V67L enhanced-splicing mutant has elevated C-terminal cleavage, and that a D24G mutation was isolated in a mutant selected for enhanced cleavage ². To measure cleavage activity, we devised an assay in which one can visualize intact fluorescent precursors that remain soluble. The cleavage construct, HT:intein:GFP, contains a short N-extein (HT) fused to the intein, with GFP as the Cextein (Fig. 3B, schematic; Fig. S2). The intein contains a C1A mutation, to inhibit N-terminal cleavage and splicing. After cleavage, the tripartite HT:intein:GFP precursor and GFP cleavage product are clearly distinguished as separate fluorescent bands on an SDS-polyacrylamide gel, when crude extracts are not boiled prior to separation (Fig. 3B, gel). In this case GFP is uninterrupted, bypassing the folding problem. After induction of the cleavage construct at 37°C for 1 h, crude cell extracts were incubated at 25°C at pH 6.0 or 7.0. The assay was restricted to single mutants, because of extensive *in vivo* cleavage of the double mutants. The rate constants for cleavage were determined at different pHs (Fig. 3B, bar graph; Fig. S3). At pH 6.0, the cleavage rates of the splicing mutants ranged from 4-fold elevated relative to the wild type for D24Y down to ~2.5-fold elevated, compared with the SM mutant that was 3-fold increased. At pH 7.0, cleavage rates ranged from 3.4-fold elevated for V98E and D24Y to 1.5-fold elevated, which was similar to the SM mutant.

We also determined the performance of selected mutants in the cleavage construct over a broad temperature range, from 20°C to 55°C (Fig. S4). Whereas the wild-type intein, the V67L (SM) and F421 mutants had temperature optima for cleavage around 35°C, the optimum was shifted upward, to 45°C, for the D24Y and V98E mutants, and their V67L counterparts. Furthermore, whereas the single mutants D24Y, V98E and F421Y were generally more active at 35°C than at 55°C, like the wild type and V67L mutant, the double mutants D24Y/V67L, V98E/V67L and F421Y/V67L were all more active at 55°C than at 35°C. This elevation of activity at 55°C represents a 3.5- to 4-fold increase over the wild type, and demonstrates the additive effects of V67L with the other mutations.

Crystallography and NMR spectroscopy provide a molecular rationale for splicing phenotypes: the ripple effect

We focused on D24 and F421 for further study for the following reasons. First, D24 mutants emerged from two completely different selection schemes, once from an *in vivo* bacterial selection ², and multiple times in the *in vitro* phage display system (Fig. 2A). Second, F421Y was the only mutation found in the C-terminal segment of the intein (Fig. 2B, yellow). Third, both mutants were substantially more active than the wild-type intein in both splicing and C-terminal cleavage, with D24 mutants having been selected independently under different conditions, and F421Y having been selected at high pH (Fig. 3; Table S1). Finally, D24 and F421 are in close proximity in three-dimensional space, but in different environments. D24 is on the surface of the intein and fully solvent exposed, whereas F421 is buried, but interacting with side chains of solvent-exposed residues. (Fig. 2B).

To identify possible structure-activity correlations for the inteins, the crystal structures of the F421Y and D24Y mutants were determined and compared with the wild-type ΔI_{hh} intein (PDB entry 2IN9) (Fig. 4A and B). Consistent with observations for other RecA intein mutant structures, differences in the crystallization conditions and crystal packing of the three proteins do not affect the overall molecular conformation ²⁴. Least-squares fitting of all α -carbon atoms of the D24Y and F421Y molecules on the wild-type intein structure resulted in r.m.s.d. values of 0.48 Å and 0.51 Å, with largest deviations of 1.51 Å and 1.07 Å, respectively, occurring in

surface loops most likely affected by environmental factors. The similarity between the molecular conformations allows interpretation of the effects of the mutations. Interestingly, the most pronounced effect of either mutation is the disruption of a bifurcated salt bridge between both residues D24 and D50 with R419, on the surface of the molecule (Fig. 4B–D). In both mutations, the increased sizes of the residues at position 24 (Fig. 4C) or 421 (Fig. 4D) force R419 into conformations that are incompatible with the bifurcated salt bridge with residues 24 and 50. The salt bridge between amino acids 419 and 24 links the N- and C-terminal segments (Fig. 4A, turquoise and yellow) and is therefore likely involved in the structural integrity and dynamics of the inteins.

We wished to determine whether disruption of the 419-to-24 salt bridge alters stability of the mutant proteins. We therefore subjected the wild-type and mutant inteins to thermolysin digestion at a range of temperatures (34°C–70°C). Whereas the V67L mutation in SM produced increased resistance of the intein to thermolysin treatment, as previously reported²¹, the D24Y and F421Y mutations did not cause a reproducible difference in stability compared to the wild-type intein (Fig. S5).

NMR spectroscopy was therefore employed to probe the effects of one mutation, F421Y, on intein structure and dynamics in solution, using chemical shift perturbation to detect subtle changes that would be undetectable crystallographically or by enzymatic digestion. Indeed, in contrast to the single localized change observed in the crystal structure, the NMR results show that the F421Y mutation caused a ripple effect, reflected in wide-spread chemical shift changes relative to the wild-type F421 intein, even at sites far from the mutation (Fig. 5). Residues L2, K27, Q51, G52, R54, L59, I61, A62, L67, W68, T420, F421, and E434 experienced large chemical-shift changes ($\Delta\delta > 1$ ppm) (Fig. 5A–C). When these chemical shift perturbations were mapped to the three-dimensional structure of the intein (Fig. 5C), some of these affected residues were as far as 20 Å from F421Y (eg. I61). With one exception (K27), the residues with the largest $\Delta\delta$ values formed a contiguous stretch across the β -strands at the core of the molecule.

Interestingly, prominent chemical shift changes were observed at residue 67, which is mutated (V67L) in SM, the mutant that has enhanced splicing activity^{2; 21}. In addition, a large chemical shift perturbation at L2 suggests that the F421Y mutation has an effect on the N-terminal splice junction. Although C1 chemical shift could not be compared due to fast solvent exchange, significant chemical shift perturbation was observed for N440 at the C-terminal splice junction. Relay of these chemical-shift changes from the mutation to the intein termini, which are in the intein active site, likely account for the enhancement in splicing and cleavage activities of the mutant.

DISCUSSION

Through a combination of genetic, biochemical and structural studies, we have gained insight into intramolecular communication within an intein molecule. We developed a directed evolution system using phage display, to generate novel inteins with enhanced splicing and cleavage activity over a broad temperature and pH range. The mutations that fine-tune function were found scattered throughout the intein structure, without occupying the active site area (Fig. 2B). Likewise, mutations at sites distant from the active center have recently been isolated in a cell-based selection system using a split intein⁸. Interestingly, a role for exteins in splicing is apparent from mutations selected in residues Y-2 and K-1 in the N-extein (Fig. 2A), as has also been recently demonstrated for the DnaE intein from *Synechocystis* sp PCC6803⁷. Although these mutations are close to the active site, all the mutations within the intein that enhance activity are located on or near exposed surfaces of the molecule (Fig. 2B), indicating that they mediate long-range effects.

Structural perturbations affecting intein function

One of the surface residues chosen for further analysis, D24, is located at the C-terminal end of the conserved sequence block N2²⁹. Although the residue is not highly conserved, three different mutations were isolated at this position (D24Y, D24N and D24E). Remarkably, a D24G mutation was isolated previously from an *in vivo* selection scheme in a hyperactive cleavage mutant that had multiple mutations², underscoring the importance of this residue for intein activity.

The second residue chosen for in-depth analysis was F421, the only activity-enhancing mutation found in the C-terminal segment of the intein. The residue is located in block C2^{30; 31}, is highly conserved, and flanks a β -strand. The α -carbon of F421 is proximal to the N-terminal splice junction, and the residue is adjacent to D422, which is important for splicing and cleavage²⁴. Surprisingly, this position is almost always occupied by Phe or Tyr. However, mutation at the equivalent residue in the *Cne* PRP8 intein (W151A) abolishes splicing but allows C-terminal cleavage, suggesting that the mutation affects the conformation at the N-terminal splice junction³². The F421 and D24 side-chains are within 5 Å of each other in the ΔI_{hh} wild-type structure (Fig. 2B and Fig. 4B)²⁴. The relationship of these two residues to the overall protein structure is considered in the crystallography and NMR studies discussed below.

The intein is composed of an N-terminal and C-terminal segment that together form a two-subdomain structure surrounding the N- and C-termini. The two subdomains are connected by two covalent links at one end of the molecule (Fig. 2B and Fig 4A)²⁴. The crystal structures show no significant differences in the molecular conformation of the mutants, except for disruption of a salt bridge between residues of the two segments (D24 and R419), in both the D24Y and F421Y mutants (Fig. 4B–D). The effect of the salt bridge does not have a major impact on overall stability of the D24Y and F421Y mutant inteins (Fig. S4), but rather influences intein dynamics, as has been suggested for electrostatic interactions in other proteins³³. The NMR data indicate relatively large chemical shift changes introduced by the F421Y mutation for residues in the β -sheet at the interface between the two subdomains, as well as for L2. The L2 residue is not only proximal to the N-terminal splice junction (C1) but also is adjacent to the β -strand that carries R419 and F421 (Fig. 5C).

Apparent differences between the results of crystallography and NMR are related to the inherent differences between the two techniques. First, chemical shifts of amide protons are sensitive to small structural changes, such as minute alterations in hydrogen-bond length, that are not readily observed in crystal structures. Second, the crystal structure shows a single conformation that is averaged over many molecules selected by the crystal environment, whereas the chemical shift data from NMR represent a weighted-ensemble average for all conformations sampled in solution. NMR is therefore sensitive not only to conformational but also dynamic changes. Consequently, the combined results suggest that, while the overall conformation of the mutants is very similar to the wild-type intein, there is a change in the molecular dynamics or conformational flexibility that is transmitted from the F421Y mutation near the surface to residues at the interior, near the active site. This ripple effect likely causes dynamic changes in the environment of the catalytic center. Perturbed dynamics in the catalytic core could result in differences in the ability of the molecule to make the subtle conformational changes necessary for catalysis, and thus lead to altered splicing and cleavage efficiencies for these mutants. In a similar analysis, a comparative NMR study on a bacterial lipase revealed that mutations that activate the enzyme have a virtually identical structure to the wild-type lipase, but displayed differences in local dynamics³⁴.

The mutants isolated by phage display appear mechanistically distinct from the previously-isolated V67L hyperactive splicing mutant. First, the distinction is reflected in activity, where

the phage display mutants on average splice less well than the SM mutant, but cleave more rapidly (Fig. 3). Second, the SM mutant has enhanced stability, whereas D24Y and F421Y do not (Fig. S5). These differences are reflected in the additivity in catalytic properties and stimulatory effects of the two mutant types (Fig. 3A and Fig. S5).

Implications for biotechnological application of the intein mutants

Some of the multiple mutants exhibited higher splicing activity than the hyperactive SM mini-intein we had isolated using an alternate *in vivo* selection system¹⁴ (Fig. 3A), and most of the single mutants were about twice as active as SM in C-terminal cleavage (Fig. 3B). Moreover, several of the mutants are more active at high temperature (50°C) and high pH (8.6) than the wild-type intein (Fig. 2C, Fig. 3B, and S5), suggesting a robustness that will be useful for applications in rugged environments. Further, mutants that are more active than the wild-type intein at pH ≥ 7 might be used to isolate inteins that are less active at low pH, for a reverse pH profile. The natural pH responsiveness of the Mtu RecA intein was previously exploited, whereby the precursor accumulated at pH >7.5 , and cleavage or splicing of an intein fusion was triggered at pH 6 to 7.5^{10; 14}. Although this pH profile is useful for protein purification from *E. coli*, where intracellular pH is ~ 7.6 ³⁵, it is limited for proteins produced from eukaryotic cells or transgenic milk where pH is normally <7 ^{36; 37}. Further selection for reverse-pH inteins, which are active at high but not low pH, would be important for expression and purification in these eukaryotic systems.

The endonuclease loop was found to be the most focused target for improvement of mini-intein activity, aside from the V67 locus, with one-third of the hyperactive mutants mapping to four residues in the loop (Fig. 2A and B). The sensitivity of the loop to mutation suggests that this location is a good environment for manipulation for biosensor development. Indeed, foreign ligand-binding proteins inserted into the endonuclease loop change conformation upon ligand binding, and the binding signal can be transduced via the intein to different extein reporter molecules^{4; 19}. The molecular evolution strategy described here will be ideal for developing and fine-tuning such novel intein-based molecular switches, which can be designed to detect a vast number of signals and deliver desirable extein-dependent outputs in response to splicing. Thus, our phage display strategy for isolating mutant inteins is applicable to the creation of new intein-based molecular tools, while providing structural insights to intein function.

MATERIALS AND METHODS

Materials and bacterial strains

Oligonucleotides were synthesized by Integrated DNA Technologies, Inc. Restriction enzymes and ligase were obtained from New England Biolabs and Invitrogen. The Accu-Prime *pfx* DNA polymerase (Invitrogen) was used for conventional PCR and plaque PCR, the Gene-Morph II system (Stratagene) was used for error-prone PCR, and *AmpliTaq* DNA polymerase (Applied Biosystems) was used for single nucleotide discrimination PCR. The vector pET30b was purchased from Novagen. The vector pHGSap was a generous gift from Prof. Henry Paulus. All T7 phage display materials and procedures were supplied in the T7Select 10-3b phage display system (Novagen). The host for T7 phage culture was *E. coli* BLT 5403. *E. coli* XL-10 Gold (Stratagene) was used for cloning of pET vectors, and *E. coli* BL21 Star (DE3) (Invitrogen) was used for the GFP::intein expression. The chitin resin was purchased from New England Biolabs. Thrombin protease was obtained from Novagen. All plasmids and phage constructs were verified by DNA sequencing. Protein concentrations were estimated by Bio-Rad Protein Assay Reagent using BSA as a control.

Construction of T7 phage variant displaying CBD::intein fusions

The control phage displaying the 52-amino acid CBD domain of *B. circulans* chitinase A1 and Mtu RecA mini-intein variants ($\Delta\Delta I_{hh}$, $\Delta\Delta I_{hh}$ -SM, and $\Delta\Delta I_{hh}$ -AA) inserted at the Cys30 of the CBD domain were constructed as described in the Supplementary Methods. Briefly, the genes encoding the intein and the CBD were amplified by PCR, fused by overlap extension PCR³⁸, and cloned downstream of the thrombin recognition sequence of pET30b vector (Novagen). The genes encoding thrombin-CBD and thrombin-CBD::intein were amplified by PCR and cloned into the T7 Select 10-3 vector (Novagen) downstream of the capsid protein gene.

For the random mutation library, the parental gene CBD:: $\Delta\Delta I_{hh}$ was amplified by errorprone PCR under two different conditions. The average mutation frequencies for the libraries (M_L and M_H) were calculated from the relative amplification and the theoretical mutation frequency of Gene Morph II mutazyme (1 or 2 mutant residues per gene for M_L , 3 or 4 residues for M_H). The resulting libraries were ligated into the T7 phage gene and packaged. The total numbers of clones in the M_L and M_H libraries after packaging were 0.3×10^6 and 1.2×10^6 plaque forming units (pfu), respectively.

Phage affinity selection

The phage amplified in LB medium (10 g Bacto tryptone, 5 g Bacto yeast extract, 10 g NaCl per liter) without $ZnCl_2$ were diluted 100-fold with the binding buffer (50 mM Tris-HCl, pH 8.6, 1 M NaCl, 0.1% Tween20, 0.1 mM $ZnCl_2$), and 0.8 ml of diluted phage ($\sim 10^8$ pfu) was mixed with chitin resin (0.2 ml bed volume). The phage were bound to the resin for 2 h at 4°C. The mixture was then transferred to a mini-spin column and the flow-through fraction was removed by a short centrifugation at 2000 rpm. The bound phage were washed 10 times with 0.6 ml of the binding buffer and twice with 0.6 ml of the wash buffer (200 mM Tris-HCl, pH 8.0, 1.5 M NaCl). The resin was incubated in 0.2 ml of the thrombin buffer (200 mM Tris-HCl, pH 8.0, 1.5 M NaCl, 1.44 units of thrombin) overnight at 30°C, and the released phage were collected by a brief centrifugation. The recovered phage were titered, amplified by infection of BLT5403 cells to a titer of $\sim 10^{10}$ pfu/ml (see Supplementary Methods), and used as input for the subsequent round of selection.

For *in vitro* splicing selection, the phage were grown in LB medium supplemented with 0.5 mM $ZnCl_2$ to suppress *in vivo* splicing during phage amplification. The phage were diluted 10-fold with the splicing buffer at various pHs (20 mM MES, pH 6.0, or HEPES, pH 7.0, or Tris-HCl, pH 8.0 and 8.6, containing 1 M NaCl and 1 mM EDTA), and incubated at 20°C–25°C for 3–4 days. For high-temperature or denaturant incubation, the phage were diluted with the splicing buffer (20 mM HEPES, pH 7.0, containing 1 M NaCl and 1 mM EDTA) and incubated at 52°C, or diluted with the same splicing buffer containing 8 M urea and incubated at 25°C. After incubation, the phage were diluted 10-fold with binding buffer, and $\sim 10^8$ pfu of phage were tested for chitin binding. The selected phage were grown as above with 0.5 mM $ZnCl_2$. After the second, third and fourth rounds of selection, several individual clones were picked from the titer plate, amplified separately, their binding phenotype verified, and their DNA sequenced.

GFP-intein fusions to monitor splicing and cleavage

GFP::intein internal fusions to monitor splicing were amplified by PCR from T7 phage and cloned into the SmaI site of pHGSpap vector²⁷. Details of cloning, expression of the GFP::intein fusions, and monitoring splicing on SDS-polyacrylamide gels is described in Supplementary Methods.

The HT::intein:GFP fusions to monitor cleavage have a his-tag as the N-extein and GFP as the C-extein. The protein expression of the cleavage construct and visualization of the

HT:inein:GFP precursor and GFP product by fluorescence detection of unboiled samples on SDS-polyacrylamide gels is described in Supplementary Methods. C-terminal cleavage was modeled as a first order reaction with respect to the amount of precursor, yielding the following equation for the amount of free GFP:

$$G_t = G_0 + NIG_0 \times (1 - e^{-kt}) \quad (1)$$

G_t and G_0 are the amounts of GFP at times t and 0, respectively; NIG_0 is the amount of precursor at time 0; and k is the reaction rate constant. Equation 1 was fit to the cleavage data using the Solver software tool within Microsoft Excel. For each data set, the root mean square deviation was minimized through variation of G_0 , NIG_0 , and k , yielding the best-fit values for these parameters.

Crystallography and NMR

Proteins for crystallization and NMR were prepared as previously described²⁴. Crystallography and NMR spectroscopy were performed using standard protocols described in Supplementary Methods.

Protein data bank accession codes

The atomic coordinates and structure factors for the D24Y and F421Y mutants have been deposited, with codes 3IGD and 3IFJ, respectively.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Wadsworth Center's Molecular Genetics and Macromolecular Crystallography Cores are gratefully acknowledged. We thank Zhong Li for assistance with the crystallization of the proteins, and Maryellen Carl for help with the manuscript. This work was supported by grants from the NIH (GM44844 to MB and GM81408 to CW) and NSF (C1404079 to GB) and BP was supported in part by the Biomolecular Science and Engineering training grant GM067545.

References

1. Paulus H. Protein splicing and related forms of protein autoprocessing. *Annu. Rev. Biochem* 2000;69:447–496. [PubMed: 10966466]
2. Wood D, Wu W, Belfort G, Derbyshire V, Belfort M. A genetic system yields self-cleaving inteins for bioseparations. *Nat. Biotech* 1999;17:889–892.
3. Perler FB, Adam E. Protein splicing and its applications. *Curr. Opin. Biotechnol* 2000;11:377–383. [PubMed: 10975457]
4. Buskirk AR, Ong YC, Gartner ZJ, Liu DR. Directed evolution of ligand dependence: small-molecule-activated protein splicing. *Proc. Natl. Acad. Sci. USA* 2004;101:10505–10510. [PubMed: 15247421]
5. Muralidharan V, Muir TW. Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nat. Methods* 2006;3:429–438. [PubMed: 16721376]
6. Flavell RR, Kothari P, Bar-Dagan M, Synan M, Vallabhajosula S, Friedman JM, Muir TW, Ceccarini G. Site-specific 18F-labeling of the protein hormone leptin using a general two-step ligation procedure. *J Am Chem Soc* 2008;130:9106–9112. [PubMed: 18570424]
7. Amitai G, Callahan BP, Stanger MJ, Belfort G, Belfort M. Modulation of intein activity by its neighboring extein substrates. *Proc. Natl. Acad. Sci. USA* 2009;106:11005–11010. [PubMed: 19541659]

8. Lockless SW, Muir TW. Traceless protein ligation utilizing evolved split inteins. *Proc. Natl. Acad. Sci. USA* 2009;106:10999–11004. [PubMed: 19541616]
9. Noren CJ, Wang J, Perler FB. Dissecting the chemistry of protein splicing and its applications. *Angew. Chem. Int. Ed. Engl* 2000;39:450–466. [PubMed: 10671234]
10. Wu W, Wood DW, Belfort G, Derbyshire V, Belfort M. Intein-mediated purification of cytotoxic endonuclease I-TevI by insertional inactivation and pH-controllable splicing. *Nucleic Acids Res* 2002;30:4864–4871. [PubMed: 12433989]
11. Evans TC, Xu MQ. Mechanistic and kinetic considerations of protein splicing. *Chemical Reviews* 2002;102:4869–4884. [PubMed: 12475209]
12. Chong S, Mersha FB, Comb DG, Scott ME, Landry D, Vence LM, Perler FB, Benner J, Kucera RB, Hirvonen CA, Pelletier JJ, Paulus H, Xu M. Single-column purification of free recombinant proteins using a self-cleavage affinity tag derived from a protein splicing element. *Gene* 1997;192:271–281. [PubMed: 9224900]
13. Chong S, Montello GE, Zhang A, Cantor EJ, Liao W, Xu MQ, Benner J. Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. *Nucleic Acids Res* 1998;26:5109–5115. [PubMed: 9801307]
14. Wood DW, Derbyshire V, Wu W, Chartrain M, Belfort M, Belfort G. Optimized single-step affinity purification with a self-cleaving intein applied to human acidic fibroblast growth factor. *Biotechnol. Prog* 2000;16:1055–1063. [PubMed: 11101334]
15. Paulus H. Protein splicing inhibitors as a new class of antimycobacterial agents. *Drugs of the Future* 2007;32:973–984.
16. Adam E, Perler FB. Development of a positive genetic selection system for inhibition of protein splicing using mycobacterial inteins in *Escherichia coli* DNA gyrase subunit A. *J. Mol. Microbiol. Biotechnol* 2002;4:479–487. [PubMed: 12432958]
17. Cann IK, Amaya KR, Southworth MW, Perler FB. Bacteriophage-based genetic system for selection of nonsplicing inteins. *Appl. Environ. Microbiol* 2004;70:3158–3162. [PubMed: 15128583]
18. Zeidler MP, Tan C, Bellaiche Y, Cherry S, Hader S, Gayko U, Perrimon N. Temperature-sensitive control of protein activity by conditionally splicing inteins. *Nat. Biotechnol* 2004;22:871–876. [PubMed: 15184905]
19. Skretas G, Wood DW. Regulation of protein activity with small-molecule-controlled inteins. *Protein Sci* 2005;14:523–532. [PubMed: 15632292]
20. Derbyshire V, Wood DW, Wu W, Dansereau JT, Dalgaard JZ, Belfort M. Genetic definition of a protein-splicing domain: functional mini-inteins support structure predictions and a model for intein evolution. *Proc. Natl. Acad. Sci. USA* 1997;94:11466–11471. [PubMed: 9326633]
21. Hiraga K, Derbyshire V, Dansereau JT, Van Roey P, Belfort M. Minimization and stabilization of the *Mycobacterium tuberculosis recA* intein. *J. Mol. Biol* 2005;354:916–926. [PubMed: 16288917]
22. Hashimoto M, Ikegami T, Seino S, Ohuchi N, Fukada H, Sugiyama J, Shirakawa M, Watanabe T. Expression and characterization of the chitin-binding domain of chitinase A1 from *Bacillus circulans* WL-12. *J. Bacteriol* 2000;182:3045–3054. [PubMed: 10809681]
23. Wang JY, Chao YP. Immobilization of cells with surface-displayed chitin-binding domain. *Appl. Envir. Microbiol* 2006;72:927–931.
24. Van Roey P, Pereira B, Li Z, Hiraga K, Belfort M, Derbyshire V. Crystallographic and mutation studies of *Mycobacterium tuberculosis recA* mini-inteins suggest a pivotal role for a highly conserved aspartate residue. *J. Mol. Biol* 2007;367:162–173. [PubMed: 17254599]
25. Mills KV, Paulus H. Reversible inhibition of protein splicing by zinc ion. *J. Biol. Chem* 2001;276:10832–10838. [PubMed: 11152694]
26. Gangopadhyay JP, Jiang S-Q, van Berkel P, Paulus H. In vitro splicing of erythropoietin by the *Mycobacterium tuberculosis RecA* intein without substituting amino acids at the splice junctions. *Biochim. Biophys. Acta* 2003;1619:193–200. [PubMed: 12527116]
27. Gangopadhyay JP, Jiang SQ, Paulus H. An in vitro screening system for protein splicing inhibitors based on green fluorescent protein as an indicator. *Anal. Chem* 2003;75:2456–2462. [PubMed: 12918990]
28. Cramer A, Whitehorn EA, Tate E, Stemmer WP. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat Biotechnol* 1996;14:315–319. [PubMed: 9630892]

29. Pietrokovski S. Modular organization of inteins and C-terminal autocatalytic domains. *Protein Sci* 1998;7:64–71. [PubMed: 9514260]
30. Pietrokovski S. Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. *Protein Sci* 1994;3:2340–2350. [PubMed: 7756989]
31. Perler FB, Olsen GJ, Adam E. Compilation and analysis of intein sequences. *Nucleic Acids Research* 1997;25:1087–1093. [PubMed: 9092614]
32. Pearl EJ, Tyndall JD, Poulter RT, Wilbanks SM. Sequence requirements for splicing by the Cne PRP8 intein. *FEBS Lett* 2007;581:3000–3004. [PubMed: 17544410]
33. Kumar S, Nussinov R. Close-range electrostatic interactions in proteins. *Chembiochem* 2002;3:604–617. [PubMed: 12324994]
34. Sibille N, Favier A, Azuaga AI, Ganshaw G, Bott R, Bonvin AM, Boelens R, van Nuland NA. Comparative NMR study on the impact of point mutations on protein stability of *Pseudomonas mendocina* lipase. *Protein Sci* 2006;15:1915–1927. [PubMed: 16823035]
35. Slonczewski JL, Rosen BP, Alger JR, Macnab RM. pH homeostasis in *Escherichia coli*: measurement by ³¹P nuclear magnetic resonance of methylphosphonate and phosphate. *Proc. Natl. Acad. Sci. USA* 1981;78:6271–6275. [PubMed: 7031646]
36. Baruah GL, Couto D, Belfort G. A predictive aggregate transport model for microfiltration of combined macromolecular solutions and poly-disperse suspensions: testing model with transgenic goat milk. *Biotechnol. Prog* 2003;19:1533–1540. [PubMed: 14524716]
37. Hunter CV, Tiley LS, Sang HM. Developments in transgenic technology: applications for medicine. *Trends Mol. Med* 2005;11:293–298. [PubMed: 15949771]
38. Horton RM, Hunt HD, Ho SN, Pullen JK, Pease LR. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* 1989;77:61–68. [PubMed: 2744488]

Figure 1.

Phage display selection system for intein splicing. (a) Phage displaying a precursor (CBD::intein) and splice products (CBD on phage and intein). (b) Selection of highly active inteins from a randomly mutated library. TCS = thrombin cleavage site. Circles indicate mutations. (c) Enrichment of phage displaying active inteins. Phage displaying high-activity (CBD:: $\Delta\Delta I_{hh}$ -SM) and low-activity (CBD:: $\Delta\Delta I_{hh}$) inteins were mixed 1:100, and the population of CBD:: $\Delta\Delta I_{hh}$ -SM phage was tested after each round of selection under conditions specified in Table 1. (d) Increase in titer of the eluted fraction after selection as follows: -Zn, the phage were cultured in the absence of zinc without further incubation; pH 6.0–8.6, the phage were cultured in the presence of zinc ion and incubated in splicing buffer at the indicated pHs. These data are for the M_L library as described in the text. See Figure S1 for the M_H data.

Figure 2.

Intein mutations that enhance phage recovery. (a) Selected mutations (left) and recovery from chitin resin (right). The control phage displaying only CBD, CBD:: $\Delta\Delta I_{hh}$ (WT), and CBD:: $\Delta\Delta I_{hh}$ -SM (SM) are shown in gray bars. The mutants in the N-terminal extein (-2 and -1 positions), the N-terminal intein segment, the loop region, and the C-terminal intein segment are shown in black, turquoise, red and yellow bars, respectively. (b) Location of mutations on the crystal structure of the Mtu RecA mini-intein $\Delta\Delta I_{hh}$. Residues 1–94 (N-terminal segment) and 403–440 (C-terminal segment) are separated by the 7-amino acid Hedgehog sequence VRDVG (the loop residues, 95–101). The N-terminal segment and the mutated residues

within are turquoise, the loop region is red, the V67L mutations is magenta, the C-terminal segment and F421 residue are yellow, and the active site residues C1, H439 and N440 are green. (c) Phage recovery of double mutants after *in vitro* splicing at pH 6.0 (grey bars) and pH 8.6 (black bars). Numbering of mutants corresponds to panel A.

Figure 3.

Activity of selected mutants. (a) *In vivo* splicing activity. The GFP:: $\Delta\Delta I_{hh}$ precursors (pre) and GFP splice products are shown in the schematic, and labeled in the polyacrylamide gel. Lane numbering in the gel corresponds to the variants labeled in the bar graph below, which shows percent splicing of single mutants (2–8) and double mutants (9–14) compared to the wild type (1). The lane labeled GFP contains extract from cells expressing uninterrupted protein. (b) *In vitro* C-terminal cleavage activity. The precursor and products of cleavage constructs are shown in the schematic, and fluorescent species are labeled in the representative polyacrylamide gel. Wild-type and F421Y mutant cell extracts are compared after the indicated incubation times at pH 7. Time-courses of percent cleavage for all the mutants at pH 6 and 7

are shown in Figure S3. Cleavage was modeled as a first-order reaction with respect to precursor and the rates plotted in the bar graph for both pHs ($n = 3$).

Figure 4.

Structure of selected intein mutants. (a) Structure of parent intein ($\Delta\Delta I_{hh}$). Area of focus in panels (b) - (d) is boxed. (b) Wild-type intein showing a bifurcated salt bridge between D24, R419 and D50. (c) The D24Y mutant. The bifurcated salt bridge to R419 and the bridge between R419 and D50 are disrupted. (d) The F421Y mutant. The replacement of a Phe residue with Tyr also disrupts the salt bridge between D24 and R419; instead a salt bridge is formed between D50 and R419. Turquoise versus yellow backbone and residues are colored as in Fig. 2B.

Figure 5.

Chemical shift perturbation caused by the F421Y mutation. (a) Overlay of the NMR ^{15}N - ^1H HSQC spectra of $\Delta\Delta\text{I}_{\text{hh}}\text{-SM}$ (V67L) and the $\Delta\Delta\text{I}_{\text{hh}}\text{-SM-F421Y}$ (V67L/F421Y) mutant. Peaks from $\Delta\Delta\text{I}_{\text{hh}}\text{-SM}$ are in green while peaks from $\Delta\Delta\text{I}_{\text{hh}}\text{-SM-F421Y}$ mutant are in red. Examples of shifted residues are indicated. (b) Chemical shift differences ($\Delta\delta$) between $\Delta\Delta\text{I}_{\text{hh}}\text{-SM-F421Y}$ and $\Delta\Delta\text{I}_{\text{hh}}\text{-SM}$. $\Delta\delta$ is plotted against residue number. The site of the loop between residues 94 and 403 is marked by a Δ . (c) Projections of chemical shift perturbation in $\Delta\Delta\text{I}_{\text{hh}}\text{-SM-F421Y}$ onto the crystal structure of $\Delta\Delta\text{I}_{\text{hh}}\text{-SM}$ (PDB entry 2IN0). The ribbon color is linearly interpolated between red and purple for $\Delta\delta$ between 4.0 ppm and 1.0 ppm; and between

purple and cyan for $\Delta\delta$ between 1.0 ppm and 0 ppm. The ribbon is colored grey where $\Delta\delta$ is not available (for prolines, N-terminal and unassigned residues).

Table 1

The recovery of control phage after chitin resin binding.

Line	Displayed protein	Description	Culture conditions	Splicing conditions	Phage recovery (%)
1	CBD	Positive control	± Zn	pH 6.0–8.6	25.7 (±9.2) ^b
2	CBD::ΔΔI _{hh} -AA	Inactive intein	± Zn	pH 6.0–8.6	0.03 (±0.03) ^b
3	CBD::ΔΔI _{hh}	Wild-type mini-intein	± Zn	pH 6.0–8.6	0.03 (±0.03) ^b
4	CBD::ΔΔI _{hh} -SM ^a	Improved mini-intein	– Zn	– ^c	3.8 (±2.1)
5	CBD::ΔΔI _{hh} -SM	Improved mini-intein	+ Zn	– ^d	0.06 (±0.00)
6	CBD::ΔΔI _{hh} -SM	Improved mini-intein	+ Zn	pH 6.0	1.8 (±1.7) ^e
7	CBD::ΔΔI _{hh} -SM	Improved mini-intein	+ Zn	pH 7.0	0.20 (±0.01) ^e
8	CBD::ΔΔI _{hh} -SM	Improved mini-intein	+ Zn	pH 8.0	0.06 (±0.02) ^e
9	CBD::ΔΔI _{hh} -SM	Improved mini-intein	+ Zn	pH 8.6	0.08 (±0.04) ^e

^aV67L enhanced splicing mutant (SM)

^bNo change observed under the following conditions of temperature (30–50°C), pH (6.0–8.6), and ± Zn ion in the medium.

^cThe phage was cultured in LB medium without ZnCl₂, no *in vitro* incubation.

^dThe phage was cultured in LB medium with 0.5 mM ZnCl₂, no *in vitro* incubation.

^e*In vitro* splicing at indicated pH after cultured in LB medium with ZnCl₂.