# Regulation of protein activity with small-molecule-controlled inteins

GEORGIOS SKRETAS AND DAVID W. WOOD

Department of Chemical Engineering, Princeton University, Princeton, New Jersey 08544, USA

## Abstract

Inteins are the protein analogs of self-splicing RNA introns, as they post-translationally excise themselves from a variety of protein hosts. Intein insertion abolishes, in general, the activity of its host protein, which is subsequently restored upon intein excision. These protein elements therefore have the potential to be used as general molecular "switches" for the control of arbitrary target proteins. Based on rational design, an intein-based protein switch has been constructed whose splicing activity is conditionally triggered in vivo by the presence of thyroid hormone or synthetic analogs. This modified intein was used in *Escherichia coli* to demonstrate that a number of different proteins can be inactivated by intein insertion and then reactivated by the addition of thyroid hormone via ligand-induced splicing. This conditional activation was also found to occur in a dose-dependent manner. Rational protein engineering was then combined with genetic selection to evolve an additional intein whose activity is controlled by the presence of synthetic estrogen ligands. The ability to regulate protein function post-translationally through the use of ligand-controlled intein splicing will most likely find applications in metabolic engineering, drug discovery and delivery, biosensing, molecular computation, as well as many additional areas of biotechnology.

**Keywords:** protein function regulation; intein; protein splicing; allosteric enzyme; domain insertion; small molecule

The development of efficient tools for controlling protein function in living cells is one of the greatest challenges of chemical biology. Traditionally, genetic approaches such as regulation of transcription and gene knockouts have been used extensively as they enable control of gene function in a highly specific and generalizable manner (Shogren-Knaak et al. 2001). Genetics, though, are hindered by transcriptional and translational delays before a functional product is formed as well as by compensation phenomena which can occur when the genetic information of a cell is perturbed (Shogren-Knaak et al. 2001). An alternate strategy is chemical genetics, where a small molecule is used to directly activate or inactivate a particular protein target (Schreiber 2003). Although this method can provide dose-dependent control over gene products of interest at the post-translational level, a different small-molecule regulator must be discovered for each individual protein target (Shogren-Knaak et al. 2001).

Inteins are protein elements that are found as in-frame insertions within the sequences of particular host genes (Noren et al. 2000; Paulus 2000). They possess the ability to excise themselves post-translationally from their protein hosts and ligate the flanking peptide sequences (exteins) via an efficient self-catalyzed reaction called protein splicing (Noren et al. 2000; Paulus 2000). In general, inteins inactivate their host proteins but restore their activity upon splicing (Daugelat and Jacobs 1999; Wood et al. 1999; Lew and Paulus 2002; Zeidler et al. 2004). Furthermore, several inteins have been shown to retain their activity when transferred to nonnative genes or cell hosts (Derbyshire et al. 1997b; Wu et al. 2002; Gangopadhyay et al. 2003; Zeidler et al. 2004). The only host protein requirement for efficient splicing is a Cys, Ser, or Thr residue as the first amino acid of the C-terminal extein (Noren et al. 2000). This flexibility

suggests that inteins could be used as molecular "switches" for the control of arbitrary target proteins. However, construction of a practical in vivo protein switch requires that the intein splicing reaction be readily controlled by an easily manipulated signal that does not affect the general cellular metabolism.

Previous attempts to generate inteins with tunable behavior include the evolution of temperature-sensitive intein alleles (Adam and Perler 2002; Cann et al. 2004; Zeidler et al. 2004), inteins incorporating nonnative amino acids whose activity is regulated by photolysis (Cook et al. 1995) as well as ligand-controlled reconstitution and splicing of split inteins (Mills and Paulus 2001; Mootz and Muir 2002; Mootz et al. 2003). Although all of these approaches provide valuable insights on how to engineer controllable protein splicing they have not yet demonstrated a general ability to activate different proteins in vivo through controllable intein splicing. Very recently, two studies have shown that controllable intein splicing in *cis* can be used to achieve this goal. In the first case, regulation of protein activity was accomplished with the isolation and use of temperature-sensitive variants of the *Saccharomyces cerevisiae* vacuolar ATPase subunit intein (Zeidler et al. 2004) and, in the second case, with the construction of a novel chimeric intein whose splicing activity is triggered by the addition of 4-hydroxytamoxifen (Buskirk et al. 2004).

This work describes a rational protein engineering approach for generating inteins whose splicing activity is regulated in vivo by the presence of human thyroid hormone. Designed according to other engineered allosteric enzyme prototypes (Baird et al. 1999; Doi and Yanagawa 1999; Tucker and Fields 2001; Guntas and Ostermeier 2004), an artificial intein chimera was created by fusing a thyroid–hormone-binding domain within a previously engineered mini-intein. The insertion of the binding domain abolished the splicing activity of the intein, but allowed it to be later restored by addition of thyroid hormone or synthetic analogs. The resulting allosteric intein was then used to conditionally activate a variety of different proteins in *Escherichia coli* in a dose-dependent manner. Finally, a combination of directed evolution and genetic selection was employed to engineer an additional controllable intein whose splicing activity is inhibited by the presence of an orthogonal set of synthetic estrogen ligands.

## Results

### Design and construction of a ligand-activated intein

The previously engineered ΔI-SM mini-intein (Wood et al. 1999) was chosen to be the protein splicing element of the chimeric fusion. This artificial mini-intein was originally derived from the *Mycobacterium tuberculosis* RecA wild-type intein (Mtu RecA) through the deletion of its entire central endonuclease domain (ΔI mini-intein) (Derbyshire et al. 1997b). A single amino acid substitution was then used to enhance its stability and restore splicing efficiency back to the level of the wild-type protein (ΔI-Splicing Mutant) (Wood et al. 1999). The Mtu RecA intein and the ΔI-SM mutant have been shown to retain high levels of splicing efficiency when inserted within a variety of non-native gene hosts (Derbyshire et al. 1997b; Daugelat and Jacobs 1999; Wood et al. 1999; Lew and Paulus 2002; Wu et al. 2002; Gangopadhyay et al. 2003). The ligand-binding domain of the human thyroid hormone receptor β (TR) was selected as the small-molecule-binding protein. This binding protein, a member of the nuclear receptor superfamily, has been shown to undergo a conformational change as well as a stability enhancement upon thyroid hormone binding (Apriletti et al. 1995; Wagner et al. 1995; Ribeiro et al. 1998). The solved crystal structure of TR reveals that its N- and C-termini are in relative proximity (Wagner et al. 1995), a factor that may play an important role in avoiding large steric constraints in the constructed chimera. TR was inserted into the ΔI-SM intein in place of its deleted endonuclease domain. This decision was based on a priori knowledge that this is a permissive site for insertions of short polypeptides (Gangopadhyay et al. 2003) as well as entire folded protein domains (Fitzsimons Hall et al. 2002; Wu et al. 2002). Furthermore, it has been shown that mutations in the endonuclease domain of the full-length Mtu RecA intein can affect splicing efficiency (unpubl. results). This implies that this structure allows for significant functional cross-talk between the two comprising domains. The resulting chimeric intein is referred to as ΔI-SM$^{TR}$ (Fig. 1A).

In order to evaluate the splicing performance as well as the controllability of the ΔI-SM$^{TR}$ intein, the chimeric construct was inserted into the sequence of the bacteriophage T4 thymidylate synthase gene (TS) of the plasmid pKT to yield pKT:ΔI-SM$^{TR}$. The pKT vector was developed as part of a previous genetic selection system where intein insertion was shown to abolish TS activity and restore it upon splicing (Derbyshire et al. 1997b; Wood et al. 1999; Fig. 1B). Intein activity can therefore be monitored through the growth phenotype of TS-knockout cells expressing these fusions in thymineless growth media. Because the requirement for active TS increases rapidly with increasing temperature, observation of TS phenotypes at different temperatures provides semiquantitative information about intein splicing efficiency (Derbyshire et al. 1997b). This genetic selection system has been calibrated in such a way that TS$^-$ phenotypes at low temperatures imply very low splicing activity, whereas TS$^+$ phenotypes at high temperatures correspond to a highly active intein (Derbyshire et al. 1997b; Wood et al. 1999).

*Escherichia coli* D1210Δ*thyA* cells transformed with pKT:ΔI-SM$^{TR}$ and grown on thymineless agar plates exhib-
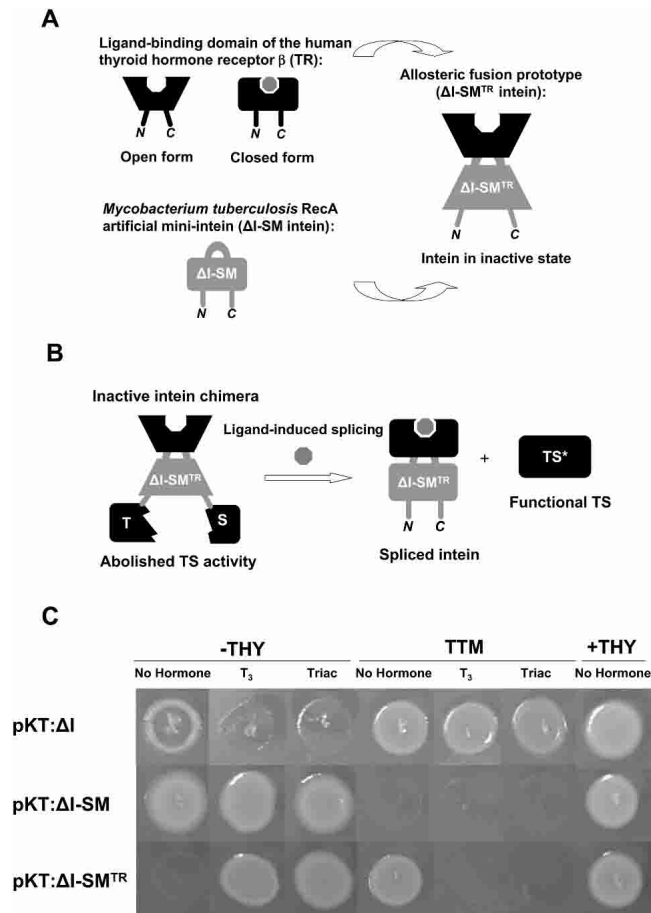
**Figure 1.** Construction of a thyroid–hormone-controlled intein. (*A*) Design of the ligand-binding domain/intein chimera. ΔI: *Mycobacterium tuberculosis* RecA artificial mini-intein; ΔI-SM: fast-splicing mutant of the ΔI intein; TR: ligand-binding domain of the human thyroid hormone β; ΔI-SM$^{TR}$: chimeric intein. (*B*) Insertion of the ΔI-SM$^{TR}$ intein into the TS reporter system for evaluation of its splicing efficiency and controllability. (*C*) Growth phenotypes of *E. coli* D1210Δ*thyA* cells transformed either with pKT:ΔI (*top* row), pKT:ΔI-SM (*middle* row), or pKT:ΔI-SM$^{TR}$ (*bottom* row) and incubated on thymineless (–THY) and trimethoprim-containing (TTM) agar plates for 24 h at 37°C with and without 50 μM thyroid hormone (T$_3$) or synthetic analogs (Triac). A thymine-rich medium (+THY) was used as a positive control.

ited TS$^-$ phenotypes in the absence of thyroid hormone at all temperatures tested (Fig. 1C; data not shown). Remarkably, the addition of the natural thyroid hormone 3,3′,5′-triiodo-L-thyronine (T$_3$) or its synthetic analog 3,3′,5-triiodothyroacetic acid (Triac) yielded strong TS$^+$ phenotypes even at temperatures as high as 37°C and 40°C (Fig. 1C; data not shown). Cells expressing a TS fusion with the original slow-splicing ΔI mini-intein (Derbyshire et al. 1997b) were unable to grow at the same conditions irrespective of the presence of thyroid hormone (Fig. 1C), thus showing that T$_3$ and Triac are not general enhancers of intein activity. In order to verify that the hormone sensitivity of the cells carrying the pKT:ΔI-SM$^{TR}$ arises from in-

creased amounts of available functional TS, the medium was supplemented with thymine to rescue TS$^-$ and poorly TS$^+$ cells. In this nonselective medium, the addition of thyroid hormone no longer had an impact on cell growth (Fig. 1C). Furthermore, in a thymine-rich medium supplemented with the dihydrofolate reductase inhibitor trimethoprim, where selection against cells that carry active TS takes place (Belfort and Pedersen-Lane 1984), the addition of T$_3$ or Triac had an inhibitory effect on cell growth (Fig. 1C). This provides strong evidence that the observed thyroid–hormone-dependent phenotypes arise from ligand-triggered splicing activity and not from a general splicing-activating or growth-enhancing property of thyroid hormone.

## Visualization of ΔI-SM$^{TR}$ intein splicing with SDS-PAGE

To directly visualize the hormone-regulated splicing activity suggested by the TS growth phenotypes, the chimeric intein was inserted within a tripartite fusion composed of a maltose-binding protein (M) as the N-terminal extein, fused to the ΔI-SM$^{TR}$ intein, fused to the C-terminal DNA-binding domain of the I-TevI homing endonuclease (C) (Derbyshire et al. 1997a) as the C-terminal extein (Fig. 2A). The high solubility of the M and C domains and the fact that the intein is inserted as an end-to-end fusion allows for the overexpression of a soluble precursor protein and, thus, for the direct analysis of the splicing reaction products via SDS-PAGE. Furthermore, the identities of the products resulting from intein splicing in this context have been verified by Western blot analysis and mass spectrometry in previous work (Derbyshire et al. 1997b). Stable and soluble MΔI-SM$^{TR}$C precursor protein was initially overexpressed in the absence of hormone for 1 h, at which point chloramphenicol was added to arrest expression of additional precursor. Cells containing the overexpressed MΔI-SM$^{TR}$C precursor were then incubated at 37°C in the presence and absence of 40 μM T$_3$, and samples were analyzed for appearance of the spliced ΔI-SM$^{TR}$ intein and the ligated MC extein product via Coomassie-stained SDS-PAGE (Fig. 2B). Although some premature splicing was observed during overexpression, no additional formation of ligated MC or spliced ΔI-SM$^{TR}$ was observed in the absence of thyroid hormone. However, the addition of T$_3$ resulted in increased formation of the spliced intein as well as of the ligated extein product. In the presence of hormone, product accumulation could clearly be detected 1 h after the addition of T$_3$ and the reaction had gone effectively to completion after ~4 h as judged by loss of precursor (data not shown). It was noted that the precursor protein was also consumed in the absence of hormone, presumably by isolated N- and C-terminal cleavage reactions or by simple degradation, although no additional spliced products following overexpression were formed in this case.
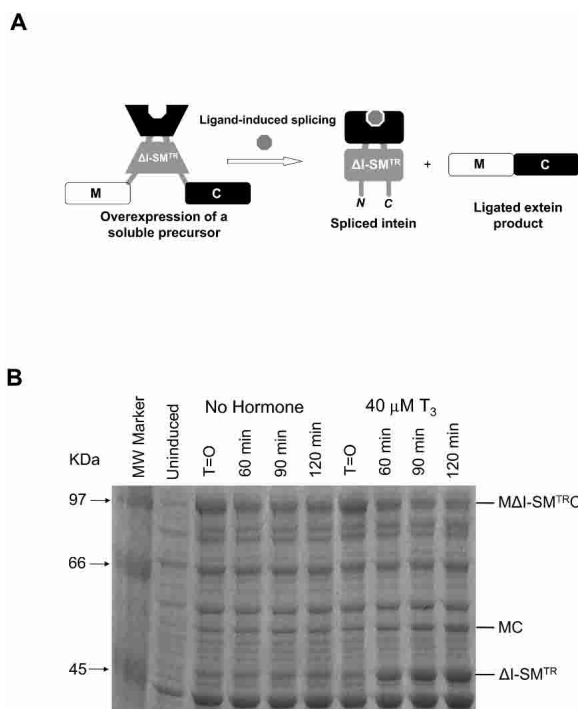
**A**



**B**



**Figure 2.** Visualization of thyroid–hormone-triggered splicing with SDS-PAGE. (*A*) Construction of a soluble precursor protein. M: Maltose-binding protein; C: C-terminal DNA-binding domain of the I-TevI homing endonuclease. (*B*) Direct observation of the accumulation of the splicing reaction product MC (expected molecular weight 56.1 kDa) and of the spliced intein $\Delta$I-SM$^{TR}$ (expected molecular weight 47.8 kDa) in the presence of thyroid hormone on Coomassie-stained 12% polyacrylamide gels.

## Thyroid hormone-induced activation of other test proteins

The general ability of the $\Delta$I-SM$^{TR}$ intein to controllably activate additional host proteins was investigated by creating insertional fusions of this intein with $\beta$-lactamase and $\beta$-galactosidase ($\beta$-gal). The chimeric intein was initially inserted into the sequence of TEM-1 $\beta$-lactamase immediately upstream of Cys121. The initial 23-residue periplasmic signal peptide was removed to ensure cytoplasmic accumulation of the precursor protein (Zlokarnik et al. 1998). The constructed intein-$\beta$-lactamase fusion was expressed in the *E. coli* strain BL21(DE3) in the presence and absence of 40 $\mu$M T$_3$ and cell lysates were evaluated for $\beta$-lactamase activity using the chromogenic substrate nitrocefin (Galarneau et al. 2002). In cells grown in the presence of thyroid hormone, lysate addition resulted in a rapid color change from yellow to red, thus revealing the presence of active $\beta$-lactamase (Fig. 3A). In contrast, no significant nitrocefin color change could be observed in the absence of hormone over the same time period.

The $\Delta$I-SM$^{TR}$ intein was additionally inserted upstream of Cys11 of the *lac*Z$\alpha$ gene in the vector pUC18. The

resulting intein fusion was expressed in the *E. coli* strain XL1-Blue with and without 40 $\mu$M T$_3$. This strain constitutively expresses the $\omega$-fragment of $\beta$-gal and successful $\alpha$-complementation can be detected with a number of colorimetric assays (Shuman and Silhavy 2003). Yellow color formation resulting from the hydrolysis of the chromogenic compound *o*-nitrophenyl-$\beta$-D-galactopyranoside (ONPG) was observed only when hormone had been added (Fig. 3B). Thyroid–hormone-triggered $\beta$-gal activity was also observed when a white-blue assay was performed with 5-bromo-4-chloro-3-indolyl-$\beta$-D-galactopyranoside (X-gal) (data not shown). This intein fusion, as well as the $\beta$-lactamase fusion, could similarly be activated by addition of Triac (data not shown).

## Protein activation is dosable

The ability to regulate the function of a target protein with a single small-molecule activator in a dosable manner would offer remarkable flexibility. The previously observed hormone-dependent TS phenotypes were used to evaluate the potential of using the $\Delta$I-SM$^{TR}$ intein to achieve controlled levels of protein activation. *E. coli* D1210$\Delta$*thyA* cells were transformed with pKT:$\Delta$I-SM$^{TR}$ and grown in liquid
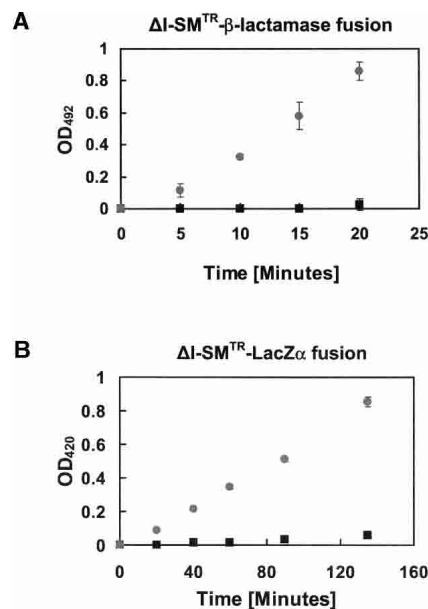


**Figure 3.** Thyroid–hormone-induced activation of $\beta$-lactamase and of $\alpha$-complementation of $\beta$-galactosidase. Colorimetric monitoring of the $\beta$-lactamase-catalyzed hydrolysis of nitrocefin (*A*) and of the $\beta$-galactosidase-catalyzed hydrolysis of ONPG (*B*) in the presence (gray circles) and absence (black squares) of 40 $\mu$M T$_3$. OD measurements were referenced with respect to lysates from cells not expressing the corresponding fusions. Experiments were carried out at least in duplicate. Error bars represent two standard deviations from the mean value.

thymineless medium at 37°C in the presence of increasing concentrations of $T_3$. Once again, cell survival relied on the presence of hormone while levels of cell growth were graded with respect to the amount of added $T_3$ (Fig. 4A). The generated dose–response curve revealed an apparent thyroid hormone detection limit of ~2.5 $\mu$M and a half-maximal effective concentration ($EC_{50}$) of ~12 $\mu$M. Cell growth appeared to saturate at ~80 $\mu$M $T_3$ but concentrations >100 $\mu$M were not tested as the solubility of the hormone in the growth medium became limiting. Thyroid hormone concentrations between 5 and 40 $\mu$M appear to pro-

vide a good linear range of protein activation. As expected, control experiments with unmodified inteins in cells harboring either the original pKT:$\Delta$I or pKT:$\Delta$I-SM plasmids indicated that these inteins are insensitive to thyroid hormone in this concentration range (Fig. 4A; data not shown). Furthermore, the $\Delta$I-SM$^{TR}$ intein was found to be less active than the slow-splicing $\Delta$I intein at low thyroid hormone concentrations, but splicing activity could be substantially increased by addition of $T_3$ to concentrations >5 $\mu$M (Fig. 4A). However, these growth phenotypes further suggest that the fully induced $\Delta$I-SM$^{TR}$ intein is not able to reach as high a level of splicing activity as the unmodified fast-splicing $\Delta$I-SM intein.

In order to verify that the observed dose-dependent TS activation occurs due to tunable intein splicing uniformly throughout the entire cell population, TS phenotypes of cells carrying the pKT:$\Delta$I-SM$^{TR}$ vector were investigated on thymineless agar plates at varying temperatures and $T_3$ concentrations. As described previously, intermediate levels of active TS lead to temperature-sensitive TS phenotypes, whereas high levels of active TS are required for growth at higher temperatures. The TS growth phenotype of cells containing pKT:$\Delta$I-SM$^{TR}$ was found to be temperature-sensitive, with this sensitivity being increased as the concentration of thyroid hormone was decreased (Fig. 4B). Furthermore, the same number of bacterial colonies was formed under all conditions that permitted cell growth (Fig. 4B). These results imply that the entire cell population is being homogeneously affected by thyroid hormone and that the observed dose-dependent growth phenotypes arise from tunable intein splicing as opposed to differential expression among subpopulations of cells.

### Generating an estrogen-regulated intein

A generally applicable strategy for creating artificial inteins whose splicing activity is modulated by a variety of small molecules would allow the individual control of multiple protein targets in the same cell. The potential to generate an estrogen-regulated intein was investigated by replacing TR with the ligand-binding domain of the human estrogen receptor $\alpha$ (ER). The resulting chimeric intein, termed $\Delta$I-SM$^{ER}$, was found to be nonfunctional under all conditions tested (data not shown). A directed evolution approach was then employed that combined random mutagenesis and genetic selection with the TS reporter system. Initially, a library of $\Delta$I-SM$^{ER}$ mutants was prepared with error-prone PCR. Since both functional splicing and estrogen-regulated splicing activity were not expected to arise from point mutations, the mutagenic reaction conditions were tuned so as to yield a relatively high frequency of mutations. The constructed $\Delta$I-SM$^{ER}$ mutant library was then cloned into the pKT selection vector and functional intein variants were then selected at room temperature on thymineless agar
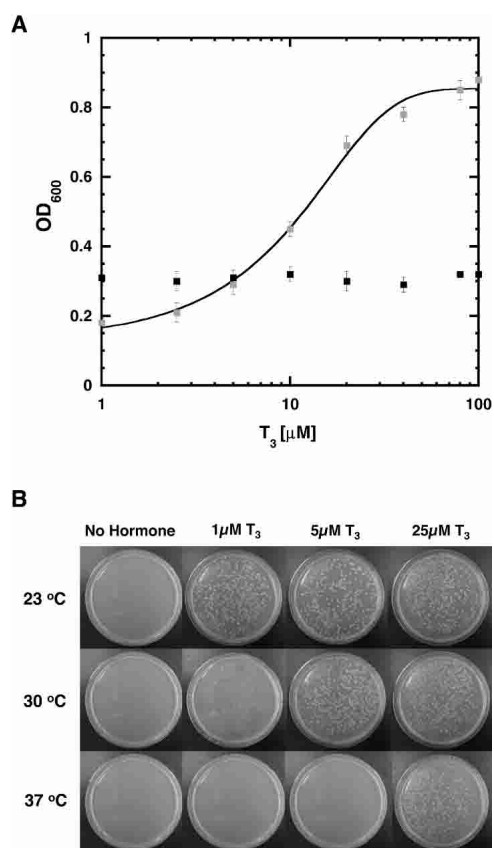


**Figure 4.** Dependence of protein activation on thyroid hormone concentration. (*A*) Dose–response curve of *E. coli* D1210$\Delta$thyA cells transformed either with pKT:$\Delta$I (black) or pKT:$\Delta$I-SM$^{TR}$ (gray) and grown in 5-mL cultures of liquid thymineless medium at 37°C for 13 h. Cells transfected with pKT:$\Delta$I-SM yielded a saturated culture after ~7 h of incubation time (data not shown). Varying concentrations of thyroid hormone were dissolved in ethanol to allow precise control of dosing. Experiments were carried out in duplicate and error bars represent one standard deviation from the mean value. A four-parameter sigmoidal fit performed with KaleidaGraph was used to generate the dose–response curve of the $\Delta$I-SM$^{TR}$ intein and to calculate the corresponding $EC_{50}$ for thyroid hormone. (*B*) Dose-dependent TS activation at different incubation temperatures. *E. coli* D1210$\Delta$thyA cells transformed with pKT:$\Delta$I-SM$^{TR}$ and incubated on thymineless agar plates containing variable concentrations of thyroid hormone at different temperatures for 40 h (30°C and 37°C ) and 60 h (23°C ).

plates in the presence and absence of a 50-μM concentration of the high-affinity synthetic estrogen analog diethylstilbestrol. After 48 h of incubation, survivors were identified on all plates, thus implying ΔI-SM$^{ER}$ variants with splicing activity. A number of mutants from both plates were isolated and evaluated for estrogen-regulated splicing activity by TS phenotype in the presence and absence of estrogen and synthetic analogs. Three of the mutants that exhibited estrogen sensitivity were selected for further characterization. When expressed in *E. coli* D1210Δ*thyA* cells and incubated in liquid thymineless medium at 34°C, cell growth was inhibited in the presence of synthetic high-affinity ER binders (Fig. 5). A control TS phenotype experiment with cells harboring either the pKT:ΔI or pKT:ΔI-SM vectors showed that the addition of the estrogenic compounds tested does not have a significant general impact on cell growth (Fig. 5; data not shown). The TS phenotypes exhibited by the isolated ΔI-SM$^{ER}$ mutants reveal that, in contrast to the ΔI-SM$^{TR}$ intein, the evolved variants are functional in the absence of synthetic estrogen but can be inactivated upon tight ligand binding. All of the mutants were found to include multiple mutations in the estrogen-binding as well as in the splicing domain of the fusion (Table 1). The specific role of these mutations is currently under investigation but the ones found in the ligand-binding domain appear to be clustered in areas either important for hormone binding (Wrenn and Katzenellenbogen 1993; Ekena et al. 1996; Brzozowski et al. 1997) or in the vicinity of the contact points between the binding and the splicing domains.
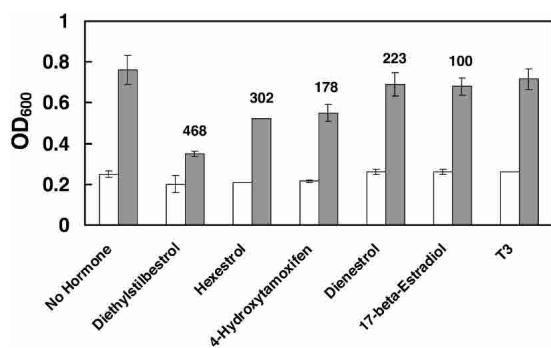


**Figure 5.** TS phenotypes of an estrogen-regulated intein. Growth of *E. coli* D1210Δ*thyA* cells transformed either with pKT:ΔI-SM$^{ER}$ mutant 2 (gray) or pKT:ΔI (white) and grown in 5-mL cultures of liquid thymineless medium at 34°C for 14 h. Cells transfected with pKT:ΔI-SM yielded a saturated culture after ~6 h of incubation time (data not shown). 17-β-Estradiol, diethylstilbestrol, hexestrol, and dienestrol were added at a concentration of 40 μM and 4-hydroxytamoxifen at 10 μM. All cultures contained an equal amount of ethanol that was kept below 1%. The numbers *above* the growth bars represent the relative binding affinity of each compound for ER (17-β-estradiol is arbitrarily set to 100) as reported by Kuiper et al. (1997). Experiments were carried out in duplicate. Error bars represent one standard deviation from the mean value.

## Discussion

Based on simple rules of protein design, an engineered self-splicing protein element (intein) has been constructed that is conditionally triggered in vivo by the presence of natural or synthetic thyroid hormone. The potential of this controllable intein to be used as a conditional activator for arbitrary target proteins was investigated by inserting the engineered intein into the sequences of a number of different proteins. It was shown in all cases that these proteins are functional only in the presence of hormone after ligand-induced splicing takes place. Protein splicing is triggered rapidly, allowing detection of the ligated extein product in <60 min by SDS-PAGE. It is also demonstrated that intein insertion can disrupt the assembly and activation of homo- and hetero-multimeric protein complexes (TS and α-complementation of β-gal, respectively), suggesting a possible use in modulating protein–protein interactions in more complex systems. The activation of intein splicing is conveniently dosable, and thus the amount of functional target protein can be easily specified in vivo by adding the required amount of the activator ligand. This offers an advantage over systems under transcriptional control where the concentration of inducer can at best regulate transcription levels and not necessarily the specific amount of functional protein. The described approach can also be combined with random mutagenesis and selection to evolve inteins which are regulated by other small molecules, thus enabling the simultaneous control of different protein targets in the same cell.

The engineered intein was constructed by fusing the ligand-binding domain of the human thyroid hormone receptor within a previously developed minimal splicing domain. This hormone-binding domain is known to undergo a conformational change and a stability enhancement upon ligand binding (Apriletti et al. 1995; Wagner et al. 1995; Ribeiro et al. 1998). The insertion of the foreign domain into the splicing domain is designed to cause a structural perturbation, and most likely affects its stability as well. These effects abolish the splicing activity of the intein, but allow it to be restored upon hormone binding. A number of similarly engineered allosteric enzymes have been previously reported, including a dihydrofolate reductase enzyme regulated by estrogen or FK506 in yeast (Tucker and Fields 2001), and the recently reported insertional fusion of β-lactamase with a maltose-binding domain (MBD) where β-lactamase activity can be activated by a variety of carbohydrates (Guntas and Ostermeier 2004). Interestingly, from a number of tested MBD binders in the latter example, only those that induce conformational changes upon binding were found to have an impact on the enzymatic activity of the chimeric fusion (Guntas and Ostermeier 2004). In other reported artificial chimeric fusions, the enzymatic activity of one domain is allosterically controlled by the binding of small ions (Baird et al. 1999), specific antibodies (Brennan et al. 1995;

**Table 1.** *Summary of the mutations found in the evolved $\Delta I$-$SM^{ER}$ inteins*

| Mutant | Splicing domain | |
|---|---|---|
| | N-terminal | C-terminal |
| 1 | I29M | L119P; E131G |
| 2 | Y79F | L119Q; M127V |
| 3 | A42V | |

| | Estrogen-binding domain | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | N-term | Helix 3 | Helix 4/5 | | Helix 6/7 | Helix 8/9 | Helix 10/11 | Helix 12 |
| 1 | L310P<br>Y331C | N359Y | W383R | | | L448P<br>L479Q | H501R<br>Q502R<br>M522V | |
| 2 | T311S<br>D332E | M343K | V392D | H398Y | G415E | | H513P<br>H516R<br>Y526N<br>K531E | |
| 3 | S305N | V368A | | | | I475N<br>K481R | K520E | L541Q<br>L544M |
| Estrogen Contact Points[a] | | M343<br>L346<br>T347<br>L349<br>A350<br>E353<br>L354 | W383<br>L384<br>L387<br>M388<br>L391<br>R394 | | M421<br>I424<br>F425<br>L428 | | G521<br>M522<br>H525<br>L525<br>M528 | L540 |

[a] Taken from Brozozowski et al. 1997.

Benito et al. 1996), or entire folded protein domains (Doi and Yanagawa 1999). Similar mechanisms of allosteric modulation may be responsible for the ligand sensitivity of the constructed chimeric intein.

Another possible explanation for the thyroid–hormone-triggered protein activation mediated by the $\Delta I$-$SM^{TR}$ intein is that the presence of hormone stabilizes the intein–extein fusion precursor. According to this mechanism, a precursor protein is formed with the ability to splice by itself in the absence of ligand, but hormone addition stabilizes this fusion, thus resulting in an increased amount of released extein product. Although this simple mechanism would still enable our desired regulation of protein activity in vivo, this mode of control is not consistent with some of the results presented in this study. First, a highly soluble M$\Delta I$-$SM^{TR}$C precursor protein could be overexpressed in the absence of hormone (Fig. 2), while the addition of $T_3$ or Triac was found to have an insignificant impact on the expression levels of the precursor protein (data not shown). These data suggest that a moderately stable and well-folded precursor protein can be expressed with this intein regardless of the presence of hormone. Furthermore, the post-translational splicing activation shown in Figure 2 indicates the presence of "splicing competent" precursor protein at least 2 h after expression. Although some background splicing was observed during overexpression, the precursor was unable to continue splicing without thyroid hormone. This leads us to hypothesize that some splicing can be attributed to a possible folding intermediate that may be able to splice at low frequency. In addition, the highly optimized M$\Delta I$-$SM^{TR}$C splicing context is likely to exacerbate any splicing leakage. In contrast, there is no evidence of background splicing in the absence of thyroid hormone with all the insertional reporter fusions tested (TS, β-gal and β-lactamase), as revealed by phenotypic and enzymatic assays. Finally, the identification of variants of the $\Delta I$-$SM^{ER}$ intein, whose splicing activity is inhibited by tight binding of estrogen analogs, indicates that ligand-induced stabilization is not sufficient to explain the behavior of the constructed inteins.

A similar chimeric intein that can be triggered by the presence of 4-hydroxytamoxifen in yeast has been very recently reported (Buskirk et al. 2004). In this case, a mutant estrogen-binding domain was inserted into a minimal Mtu RecA intein and the resulting chimera was evolved for allosteric activity using a combination of positive and negative selections. Similar to the $\Delta I$-$SM^{TR}$ chimera, the evolved ER-intein was then used to post-translationally activate a number of unrelated proteins in the yeast *S. cerevisiae* in a dose-dependent way. These results, taken together with our findings, demonstrate that the described protein engineering approaches can be used to evolve families of small-molecule-controlled inteins that respond to different com-

pounds, and that these chimeric inteins can be used to regulate protein activity in different host organisms. Ultimately such intein-based molecular protein switches could be applied to the post-translational control of protein activity in higher eukaryotes.

An interesting contrast between the evolved ER-intein variants described by Buskirk et al. (2004) and the ΔI-SM$^{ER}$ inteins reported here is that the former are activated by 4-hydroxytamoxifen, whereas the latter are inhibited upon ligand binding. Although the design of the ER-splicing domain fusions were very similar in the two studies, the specifics of how they were constructed and of how the genetic selection was carried out might be responsible for the very different outcomes. First, in this work, residues 301–553 of ER were used to replace residues 110–383 of the Mtu RecA intein, whereas Buskirk et al. (2004) used residues 304–551 of the receptor sequence along with six-aa-long linkers flanking the receptor's termini to replace residues 95–382 of the intein. It is very likely that the length of the linkers that join the two domains together is important for the structural and functional communication of the two domains in the resulting chimera. Second, our ΔI-SM$^{ER}$ fusion was created by inserting the wild-type ER sequence into a splicing mutant of a minimal Mtu RecA intein, whereas Buskirk et al. (2004) inserted a mutant ER, that binds 4-hydroxytamoxifen but not natural estrogen, into a wild-type minimal Mtu RecA intein. Starting from different points in sequence space may have guided the evolution experiments in different directions with different results.

The way the genetic selection was carried out might also explain why no estrogen-inhibited ER-intein variants were identified in the Buskirk et al. (2004) study. In this case, round one variants were selected for active splicing only in the presence of 4-hydroxytamoxifen, so mutants with the ability to splice only in the absence of ligand would not have survived. Finally, our work took place entirely in *E. coli*, while the evolved ER-intein reported by Buskirk et al. (2004) was used only in yeast. An important difference between these two organisms is the molecular chaperone Hsp90, which has been shown to form complexes with the ligand-binding domains of steroid receptors (Pratt and Toft 1997). As verified by Buskirk et al. (2004), Hsp90 increases expression and apparent stability of their ER-intein precursors. Although a bacterial homolog of Hsp90 exists in *E. coli*, it has many functional differences (Buchner 1999), and in contrast to steroid receptors expressed in mammalian cell lines (Pratt and Toft 1997), it does not copurify with ΔI-SM$^{ER}$ in the absence of estrogen (unpubl. results). The Hsp90 chaperone is only one example of the many differences between yeast and *E. coli* that might explain differences between the behaviors of these two inteins. An important point to note is simply that controllable inteins could be generated in both *E. coli* and yeast, suggesting widespread future applications.

Small-molecule-controlled inteins are likely to find utility as conditional activators for gene products of unknown function in proteomics studies as well as in numerous other applications in biotechnology. For example, ligand-induced intein splicing out of a strong reporter protein could be used as a highly specific biosensor. In principle, since intein splicing can be used to activate any host protein, the reporter protein can be chosen according to convenience. Because the activation is irreversible, the reporter would provide high sensitivity to the presence of the ligand. In drug discovery, allosteric activation of a reporter based on ligand binding to a drug target could become a rapid method for chemical library screening. This method could be combined with in vivo combinatorial biosynthesis and a genetic selection to directly identify cells that synthesize potential drug compounds against the target. Finally, these inteins might find applications in cellular and molecular computation. Groups of allosteric inteins that react to various secondary metabolites might be organized in complex networks to regulate gene expression, enzyme activity, or the production of various reporter proteins. In this case, the inteins could provide a new type of computational tool, or an advanced control strategy for cellular metabolism.

## Materials and methods

### Reagents and strains

All experiments were carried out with *E. coli* strains. XL1-Blue cells (Stratagene) were used in all cases for plasmid constructions as well as for the expression of the intein-LacZα fusion. The strain D1210Δ*thyA*::Kan$^R$ [F$^−$Δ(*gpt-proA*)62 *leuB*6 *supE*44 *ara*-14 *galK*2 *lacY*1 Δ(*mcrC-mrr*) *rpsL*20 (*Str$^r$*) *xyl*-5 *mtl*-1 *recA*13 *lacI$^q$*] was used for the determination of TS growth phenotypes (Derbyshire et al. 1997b). BL21(DE3) (Novagen) was used for the expression of the intein-β-lactamase fusion, while ER2566 (New England Biolabs) was used for the overexpression of the tripartite MΔI-SM$^{TR}$C fusion. Estrogen, thyroid hormone, and their synthetic analogs (17-β-estradiol, diethylstilbestrol, hexestrol, dienestrol, 4-hydroxytamoxifen, 3,3′,5′-triiodo-L-thyronine and 3,3′,5-triiodothyroacetic acid) were purchased from Sigma. All compounds were prepared as 10 mM stock solutions in EtOH. Nitrocefin was purchased from Becton-Dickinson and X-gal from Fisher Scientific. Both compounds were dissolved in dimethylsulfoxide to form 2 mM and 1% solutions, respectively.

### Plasmid construction in the TS context

The coding sequence corresponding to residues Glu203–Asp461 of the human thyroid hormone receptor β were amplified from the vector pCMV-hTRβ (provided by Paul Webb, Metabolic Research Unit, University of California, San Francisco), while the one coding for residues Ser301–Thr553 of the human estrogen receptor α were derived from Gal4-ER-VP16 (a gift from Randall Morse, Wadsworth Center). In each case, PCR primers were used to add a BssHII restriction site at each of the N- and C-termini of the hormone-binding domains. The ligand-binding domains of the thyroid hormone and estrogen receptors were then inserted within the

BssHII site of the mini-intein of the vector pKT:ΔI-SM to replace residues 110–383 of the Mtu RecA intein (Wood et al. 1999). The sequences of the constructed mini-intein/binding domain chimeras were verified with nucleotide sequencing.

### Determination of TS phenotypes

Cells were grown overnight in LB medium containing 200 μg/mL ampicillin and supplemented with 50 μg/mL thymine. For experiments in liquid media these saturated cultures were used with a 1:200 dilution to inoculate 5 mL of defined thymineless medium (Belfort and Pedersen-Lane 1984) with 200 μg/mL ampicillin and the specified concentrations of each of the hormone ligands. Stock solutions of various concentrations of hormone analogs in ethanol were prepared to ensure a constant final ethanol concentration in phenotype tests. In all growth experiments ethanol concentrations were kept below 1%. Levels of cell growth were measured as $OD_{600}$ on a GENESYS 2 spectrophotometer. Thymine-rich media contained at least 50 μg/mL thymine, and trimethoprim was added to a 10-μg/mL concentration. All experiments were carried out at least in duplicate.

### Splicing kinetics with SDS-PAGE

The MΔI-SM$^{TR}$C fusion was constructed by replacing the 110Δ383 mini-intein of the previously described pMIC vector (Derbyshire et al. 1997b) with the ΔI-SM$^{TR}$ intein. Cells carrying the resulting plasmid were grown in LB medium with 200 μg/mL ampicillin to mid-log phase ($OD_{600}$ ~ 0.5) and overexpression of the fusion was induced by the addition of 1 mM isopropyl-β-D-thiogalactoside (IPTG) for 1 h at 37°C. Chloramphenicol was then added to a concentration of 70 μg/mL to terminate protein production, and the specified concentrations of thyroid hormone were supplied. The progress of the splicing reaction was monitored by analyzing clarified cell lysates via SDS polyacrylamide gel electrophoresis.

### β-Lactamase and β-galactosidase colorimetric assays

Silent SacI and PstI restriction sites flanking Cys121 of the TEM-1 β-lactamase gene of the vector pGEM-11Zf(+) (Promega) were introduced with the GeneEditor in vitro site-directed mutagenesis system (Promega). These sites were used to insert the ΔI-SM$^{TR}$ intein immediately upstream of the Cys121 residue of this gene. The constructed intein-β-lactamase fusion was then transferred to the polylinker of pET-24a(+) (Novagen) using the available EcoRI and HindIII restriction sites, along with a N-terminal maltose-binding protein affinity tag. Cells carrying the resulting plasmid were grown overnight in LB medium with 20 μg/mL kanamycin, and this saturated culture was used with a 1:100 dilution to inoculate fresh 5-mL cultures of LB media with 20 μg/mL kanamycin and the specified concentrations of thyroid hormone. Cell were harvested after ~6 h of incubation, washed with PBS, and resuspended in 1 mL of 100 mM phosphate buffer at pH 7.0. The cells were then lysed by sonication and cell debris was removed by centrifugation at 4°C for 5 min. Nitrocefin was then added to a final concentration of 100 μM and its color change was monitored as the change in $OD_{492}$ on a GENESYS 2 spectrophotometer.

Cells transfected with the vector carrying the intein-LacZα fusion were grown overnight in LB medium with 200 μg/mL ampicillin and diluted 1:100 into fresh 5-mL cultures of LB media with 200 μg/mL ampicillin. When the cell density reached an $OD_{600}$ of 0.2–0.3, 1 mM IPTG was added to the cultures along with the specified concentrations of thyroid hormone. Cells were harvested after ~2 h of incubation and ONPG hydrolysis was monitored according to a β-galacosidase assay kit (Stratagene).

### ΔI-SM$^{ER}$ intein mutagenesis and selection

Mutants of the ΔI-SM$^{ER}$ intein were generated with error-prone PCR as previously described (Wood et al. 1999) except that 0.1 mM $MnCl_2$ was added in four reactions each one containing a reduced concentration of one of the four nucleotides from 200 to 80 μM. A library of ~10,000 variants was constructed and selected for functional mutants.

## References

Adam, E. and Perler, F.B. 2002. 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.* **4:** 479–487.

Apriletti, J.W., Baxter, J.D., Lau, K.H., and West, B.L. 1995. Expression of the rat α 1 thyroid hormone receptor ligand binding domain in *Escherichia coli* and the use of a ligand-induced conformation change as a method for its purification to homogeneity. *Protein Expr. Purif.* **6:** 363–370.

Baird, G.S., Zacharias, D.A., and Tsien, R.Y. 1999. Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl. Acad. Sci.* **96:** 11241–11246.

Belfort, M. and Pedersen-Lane, J. 1984. Genetic system for analyzing *Escherichia coli* thymidylate synthase. *J. Bacteriol.* **160:** 371–378.

Benito, A., Feliu, J.X., and Villaverde, A. 1996. β-galactosidase enzymatic activity as a molecular probe to detect specific antibodies. *J. Biol. Chem.* **271:** 21251–21256.

Brennan, C.A., Christianson, K., La Fleur, M.A., and Mandecki, W. 1995. A molecular sensor system based on genetically engineered alkaline phosphatase. *Proc. Natl. Acad. Sci.* **92:** 5783–5787.

Brzozowski, A.M., Pike, A.C., Dauter, Z., Hubbard, R.E., Bonn, T., Engstrom, O., Ohman, L., Greene, G.L., Gustafsson, J.A., and Carlquist, M. 1997. Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **389:** 753–758.

Buchner, J. 1999. Hsp90 & Co.—A holding for folding. *Trends Biochem. Sci.* **24:** 136–141.

Buskirk, A.R., Ong, Y.C., Gartner, Z.J., and Liu, D.R. 2004. Directed evolution of ligand dependence: Small-molecule-activated protein splicing. *Proc. Natl. Acad. Sci.* **101:** 10505–10510.

Cann, I.K., Amaya, K.R., Southworth, M.W., and Perler, F.B. 2004. Bacteriophage-based genetic system for selection of nonsplicing inteins. *Appl. Environ. Microbiol.* **70:** 3158–3162.

Cook, S.N., Jack, W.E., Xiong, X., Danley, L.E., Ellman, J.A., Schultz, P.G., and Noren, C.J. 1995. Photochemically initiated protein splicing. *Angew. Chem. Int. Ed. Engl.* **34:** 1629–1630.

Daugelat, S. and Jacobs Jr., W.R. 1999. The *Mycobacterium tuberculosis* recA intein can be used in an ORFTRAP to select for open reading frames. *Protein Sci.* **8:** 644–653.

Derbyshire, V., Kowalski, J.C., Dansereau, J.T., Hauer, C.R., and Belfort, M. 1997a. Two-domain structure of the td intron-encoded endonuclease I-TevI correlates with the two-domain configuration of the homing site. *J. Mol. Biol.* **265:** 494–506.

Derbyshire, V., Wood, D.W., Wu, W., Dansereau, J.T., Dalgaard, J.Z., and Belfort, M. 1997b. Genetic definition of a protein-splicing domain: Functional mini-inteins support structure predictions and a model for intein evolution. *Proc. Natl. Acad. Sci.* **94:** 11466–11471.

Doi, N. and Yanagawa, H. 1999. Design of generic biosensors based on green

fluorescent proteins with allosteric sites by directed evolution. *FEBS Lett.* **453:** 305–307.

Ekena, K., Weis, K.E., Katzenellenbogen, J.A., and Katzenellenbogen, B.S. 1996. Identification of amino acids in the hormone binding domain of the human estrogen receptor important in estrogen binding. *J. Biol. Chem.* **271:** 20053–20059.

Fitzsimons Hall, M., Noren, C.J., Perler, F.B., and Schildkraut, I. 2002. Creation of an artificial bifunctional intein by grafting a homing endonuclease into a mini-intein. *J. Mol. Biol.* **323:** 173–179.

Galarneau, A., Primeau, M., Trudeau, L.E., and Michnick, S.W. 2002. β-lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein protein interactions. *Nat. Biotechnol.* **20:** 619–622.

Gangopadhyay, J.P., Jiang, S.Q., van Berkel, P., and Paulus, H. 2003. In vitro splicing of erythropoietin by the *Mycobacterium tuberculosis* RecA intein without substituting amino acids at the splice junctions. *Biochim. Biophys. Acta* **1619:** 193–200.

Guntas, G. and Ostermeier, M. 2004. Creation of an allosteric enzyme by domain insertion. *J. Mol. Biol.* **336:** 263–273.

Kuiper, G.G., Carlsson, B., Grandien, K., Enmark, E., Haggblad, J., Nilsson, S., and Gustafsson, J.A. 1997. Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors α and β. *Endocrinology* **138:** 863–870.

Lew, B.M. and Paulus, H. 2002. An in vivo screening system against protein splicing useful for the isolation of non-splicing mutants or inhibitors of the RecA intein of *Mycobacterium tuberculosis*. *Gene* **282:** 169–177.

Mills, K.V. and Paulus, H. 2001. Reversible inhibition of protein splicing by zinc ion. *J. Biol. Chem.* **276:** 10832–10838.

Mootz, H.D. and Muir, T.W. 2002. Protein splicing triggered by a small molecule. *J. Am. Chem. Soc.* **124:** 9044–9045.

Mootz, H.D., Blum, E.S., Tyszkiewicz, A.B., and Muir, T.W. 2003. Conditional protein splicing: A new tool to control protein structure and function in vitro and in vivo. *J. Am. Chem. Soc.* **125:** 10561–10569.

Noren, C.J., Wang, J., and Perler, F.B. 2000. Dissecting the chemistry of protein splicing and its applications. *Angew. Chem. Int. Ed. Engl.* **39:** 450–466.

Paulus, H. 2000. Protein splicing and related forms of protein autoprocessing. *Annu. Rev. Biochem.* **69:** 447–496.

Pratt, W.B. and Toft, D.O. 1997. Steroid receptor interactions with heat shock protein and immunophilin chaperones. *Endocr. Rev.* **18:** 306–360.

Ribeiro, R.C., Apriletti, J.W., Wagner, R.L., Feng, W., Kushner, P.J., Nilsson, S., Scanlan, T.S., West, B.L., Fletterick, R.J., and Baxter, J.D. 1998. X-ray crystallographic and functional studies of thyroid hormone receptor. *J. Steroid Biochem. Mol. Biol.* **65:** 133–141.

Schreiber, S.L. 2003. The small-molecule approach to biology: Chemical genetics and diversity-oriented organic synthesis make possible the systematic exploration of biology. *Chem. Eng. News* **81:** 51–61.

Shogren-Knaak, M.A., Alaimo, P.J., and Shokat, K.M. 2001. Recent advances in chemical approaches to the study of biological systems. *Annu. Rev. Cell. Dev. Biol.* **17:** 405–433.

Shuman, H.A. and Silhavy, T.J. 2003. The art and design of genetic screens: *Escherichia coli*. *Nat. Rev. Genet.* **4:** 419–431.

Tucker, C.L. and Fields, S. 2001. A yeast sensor of ligand binding. *Nat. Biotechnol.* **19:** 1042–1046.

Wagner, R.L., Apriletti, J.W., McGrath, M.E., West, B.L., Baxter, J.D., and Fletterick, R.J. 1995. A structural role for hormone in the thyroid hormone receptor. *Nature* **378:** 690–697.

Wood, D.W., Wu, W., Belfort, G., Derbyshire, V., and Belfort, M. 1999. A genetic system yields self-cleaving inteins for bioseparations. *Nat. Biotechnol.* **17:** 889–892.

Wrenn, C.K. and Katzenellenbogen, B.S. 1993. Structure–function analysis of the hormone binding domain of the human estrogen receptor by region-specific mutagenesis and phenotypic screening in yeast. *J. Biol. Chem.* **268:** 24089–24098.

Wu, W., Wood, D.W., Belfort, G., Derbyshire, V., and Belfort, M. 2002. Intein-mediated purification of cytotoxic endonuclease I-TevI by insertional inactivation and pH-controllable splicing. *Nucleic Acids Res.* **30:** 4864–4871.

Zeidler, M.P., Tan, C., Bellaiche, Y., Cherry, S., Hader, S., Gayko, U., and Perrimon, N. 2004. Temperature-sensitive control of protein activity by conditionally splicing inteins. *Nat. Biotechnol.* **22:** 871–876.

Zlokarnik, G., Negulescu, P.A., Knapp, T.E., Mere, L., Burres, N., Feng, L., Whitney, M., Roemer, K., and Tsien, R.Y. 1998. Quantitation of transcription and clonal selection of single living cells with β-lactamase as reporter. *Science* **279:** 84–88.