Drugs of the Future 2007, 32(11): 973-984

CCC: 0377-8282/2007
DOI: 10.1358/dof.2007.032.11.1140690

Review Article

Protein splicing inhibitors as a new class of antimycobacterial agents

Henry Paulus
Boston Biomedical Research Institute, Watertown, MA 02472, and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA. Correspondence: Paulus@bbri.org

CONTENTS

Abstract

The alarming worldwide increase in multidrug-resistant and extensively drug-resistant tuberculosis (TB) calls for new antimycobacterial drugs against new types of targets which are less likely to undergo mutations that lead to drug resistance. The discovery that three genes of *Mycobacterium tuberculosis, recA, dnaB* and *sufB*, are interrupted by inteins, which must be excised from inactive precursor proteins by the process of protein splicing to yield functional proteins, suggests protein splicing as a novel target for anti-TB drugs. Since two of these genes are essential for growth and the third predisposes to drug resistance mutations, protein splicing inhibitors, which disable two separate vital functions and at the same time suppress mutation rates, would be unlikely to elicit drug resistance. A robust, protein-based high-throughput screening system for protein splicing inhibitors has been developed and used for screening small-molecule compound libraries, leading to more than 50 confirmed positives with IC_{50} values in the low micromolar range. Most of the confirmed protein splicing inhibitors are electrophiles that inhibit protein splicing irreversibly by binding to the catalytic cysteine residue. Lead optimization, aided by docking to the known crystal structure of the RecA intein, should yield a valuable new class of antimycobacterial drugs for the treatment of drug-resistant TB.

Introduction

Tuberculosis (TB) is globally the most widespread infectious disease. Two billion people (one-third of the world's population) are infected with *Mycobacterium tuberculosis* and 5-10% of these suffer active disease, leading to nearly 3 million deaths annually. Although vaccination is the only way to eliminate TB in the long term, especially in developing countries where lengthy combination drug therapy is problematic, this does not address the needs of the 2 billion individuals who are infected right now or of the alarmingly growing number of patients who are suffering from multidrug-resistant (MDR) (1) and extensively drug-resistant (XDR) TB, which is essentially refractory to drug therapy (2). A special concern is the link between AIDS and TB, 31% of TB cases in Africa and 26% of those in the USA being associated with HIV infection; the death of 11% of all HIV patients is caused by TB (3). The mainline anti-TB drugs —isoniazid, rifampin, pyrazinamide, ethambutol and, secondarily, streptomycin— thus need be supplemented by additional drugs against new targets chosen so as to minimize the emergence of drug resistance. In addition, the ability of *M. tuberculosis* to persist in a patient for decades in a state refractory to most known antibacterial agents puts a great premium on the discovery of new drugs capable of targeting the persistent state.

In spite of these urgent needs, no new anti-TB drugs have been introduced in the past 30 years. However, the last few years have seen more activity in this field and two new drugs have recently completed the preclinical phase: PA-824, a nitroimidoazopyran (4), and R-207910, a diarylquinoline (5). The targets of both of these drugs, a step in mycolate synthesis and ATP synthase, respectively, were identified by the analysis of mutations leading to drug resistance. The fact that these mutations occur at frequencies comparable to mutations leading to resistance to the mainline anti-TB drugs is an indication that *M. tuberculosis* strains resistant to PA-824 and R-207910 will emerge in the foreseeable future and that new antibiotics will then be needed to take their place. The development of new anti-TB drugs must therefore remain a continuing effort and the search for anti-TB drugs aimed

at new targets deserves to be given high priority. Should such drugs be broad in spectrum or highly specific for *M. tuberculosis*? An important consideration in addressing this question is that effective cure of MDR TB requires long-term treatment with anti-TB drugs. In such a situation, a broad-spectrum drug would not only affect a significant fraction of the subjects' normal bacterial flora, but could stimulate the emergence of resistance among other pathogens carried by the treated population. These problems would not arise with a drug targeted specifically against *M. tuberculosis*.

The discovery that three genes of *M. tuberculosis*, *recA (Rv2737)*, *dnaB (Rv0058)* and *sufB* (*Rv1461*), are interrupted by inteins, which must be excised from inactive precursor proteins by the process of protein splicing to yield functional proteins, suggested protein splicing as a novel target for anti-TB drugs. As discussed below, protein splicing inhibitors would have unique features as anti-TB drugs that would minimize the emergence of drug-resistant mutants.

Protein splicing in *M. tuberculosis*

Protein splicing is a form of post-translational processing that consists of the excision of an intervening polypeptide sequence, the intein, from a protein, accompanied by the concomitant joining of the flanking polypeptide sequences, the exteins, by a peptide bond (Fig. 1). It proceeds by a series of four reactions that are catalyzed entirely by the intein and require no co-factors or accessory proteins (6). Comparison of the catalytic domains of inteins has revealed conserved elements that appear to play roles in intein structure and the catalysis of protein splicing. These include amino acids with a hydroxyl or thiol side-chain (Ser, Cys or Thr) adjacent to the two splice junctions, the sequence His-Asn at the *C*-terminus of the intein and 6 relatively conserved sequence motifs (Fig. 2). The polypeptide backbone conformations of the protein splicing domains, which have been solved for 9 inteins by X-ray crystallography or NMR, are almost the same. More than 400 inteins are known, distributed among unicellular representatives of eukaryotes, archaea and bacteria (8). The fact that protein splicing does not occur in higher eukaryotes and has a limited distribution among eubacteria, with none in nonpathogenic bacteria associated with humans and *Mycobacterium* as the only pathogenic representative, makes it an attractive antimycobacterial target.

M. tuberculosis harbors 3 inteins, which interrupt the DnaB, RecA and SufB proteins. As will be discussed in the next section, the DnaB and RecA proteins play important roles in DNA replication and repair, respectively, whereas SufB is a component of the Fe-S cluster assembly and repair SUF machinery (9, 10) and is essential for growth (11). We have focused our primary attention on the DnaB and RecA inteins, whose protein splicing domains, compared in Figure 3, show 29% amino acid identity and 44% similarity, not only in the conserved protein splicing motifs but also in other regions, suggesting a close evolutionary relationship. The high degree of similarity of these inteins, as well as the fact that even relatively unrelated inteins have almost identical 3-dimen-

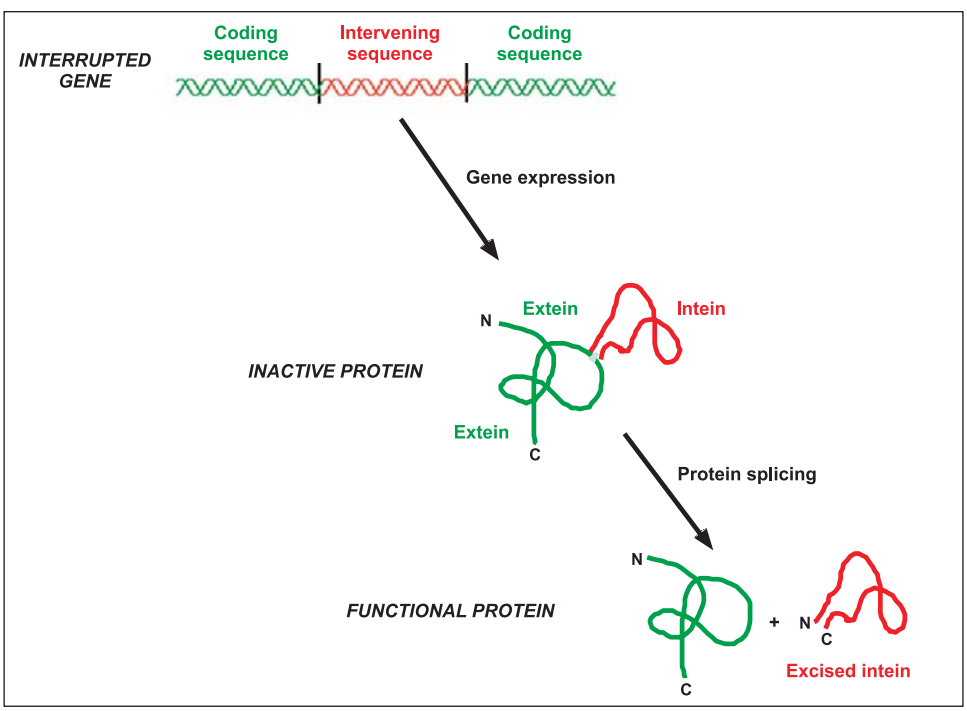

Fig. 1. Role of protein splicing in gene expression. When a gene is interrupted by an intein, its primary translation product is an inactive protein precursor that is converted to a functional protein by protein splicing.

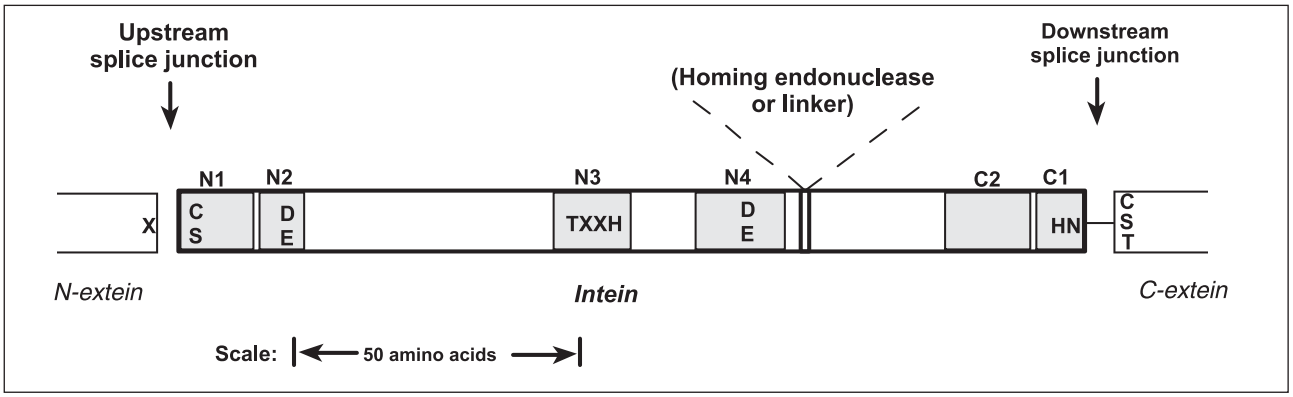

Fig. 2. Conserved elements in a typical intein. The lightly shaded regions are conserved intein motifs, labeled according to the nomenclature of Pietrokovski (7). Amino acid residues that are conserved in > 90% of known inteins are shown using the single letter notation (X = any amino acid).

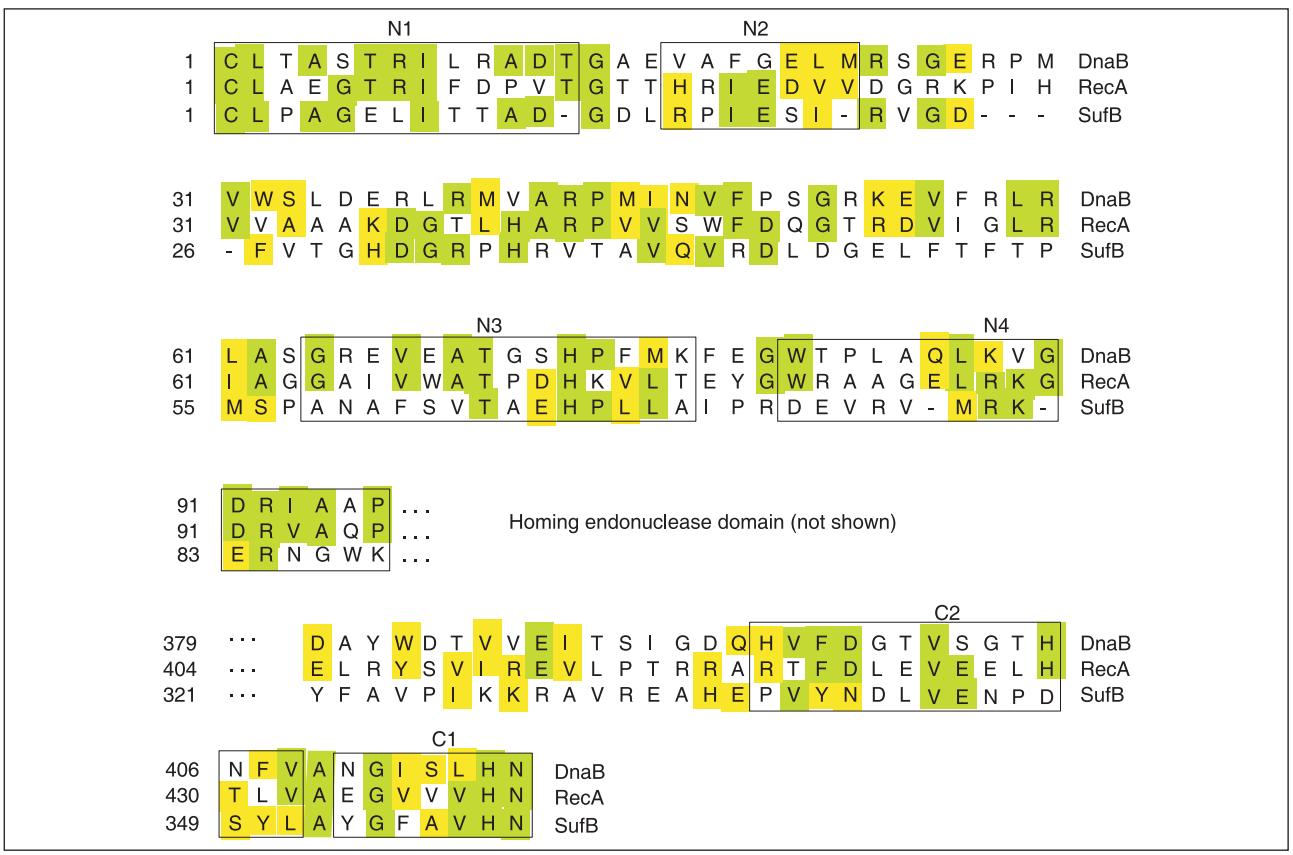

Fig. 3. Alignment of the protein splicing domains of the *M. tuberculosis* DnaB, RecA and SufB inteins. The conserved protein splicing domains (7) are shown by boxes and identical residues are highlighted in green, similar residues in yellow.

sional structures (*e.g.*, ref. 6), makes it quite likely that substances can be found which are inhibitors of protein splicing catalyzed by both of these inteins. Although the SufB (Rv1461) intein is not the primary focus of our study, its protein splicing domain has 17% identity and 31% similarity to that of the DnaB intein, suggesting that its structure is similar to the structure of the DnaB and RecA inteins and that all three inteins may be susceptible to the same inhibitors.

Role of DnaB and RecA in *M. tuberculosis*

DnaB is a replicative DNA helicase that functions in the initiation of lagging-strand DNA synthesis (12) as a component of the primosome (13). Specifically, it unwinds the DNA duplex ahead of the replication fork and recruits the DnaG primase to the primosome. The essential role of *Escherichia coli* DnaB in DNA replication is indicated by the existence of temperature-sensitive *dnaB* mutants

(14). The functional form of the DnaB protein is a hexamer which interacts directly with 2-3 molecules of the DnaG primase (15). In light of the complex protein-protein interactions in which DnaB participates, which may also include the products of the *dnaA*, *dnaC* and *dnaX* genes, it is quite unlikely that the 465-residue DnaB helicase can function without the excision of the 415-residue intein. The *M. tuberculosis dnaB* gene has been shown to be essential for growth (11).

Bacterial RecA protein (recombinase A) catalyzes various reactions related to genetic recombination, but also plays a role in DNA repair in the so-called SOS response. As in other bacteria, the *M. tuberculosis* RecA protein initiates the SOS response to DNA damage by polymerizing to filaments at the site of DNA lesions (16). However, a major difference between *M. tuberculosis* and other bacteria in the induction of the SOS response is that the *M. tuberculosis* RecA protein is synthesized in an inactive form (17), which has to undergo protein splicing before it can assume its role in the initiation of DNA repair (18). RecA is not essential for bacterial growth, except under conditions of extensive DNA damage. Since *M. tuberculosis* is an intracellular pathogen, the repair of DNA damage inflicted by the reactive oxygen and nitrogen species produced by macrophages must play an important role in its survival in the infected host. However, *recA* mutants of *Mycobacterium bovis* BCG, a nonvirulent mycobacterial strain highly susceptible to killing by DNA-damaging agents, were not compromised in terms of survival in a nude mouse model; however, when studied in an *in vitro* model for persistence, *recA* mutants were significantly more sensitive to killing by metronidazole (19).

Recent work has suggested an important role for RecA in the emergence of drug-resistant mycobacteria. One of the enzymes induced by the SOS response to DNA damage (20), and the induction of which absolutely requires functional RecA (21), is an error-prone DNA polymerase, DnaE2 (22). In *Mycobacterium smegmatis*, disruption of either *dnaE2* or *recA* reduced the number of ultraviolet (UV)-induced rifampin resistance mutations 10- and 25-fold, respectively, to nearly background levels. In *M. tuberculosis*, only the effect of *dnaE2* disruption was examined, which essentially prevented any DNA damage-induced rifampin resistance mutations and, when tested in an infected mouse model, almost doubled the median survival time and reduced lung colony-forming unit (CFU) counts 9 months post-infection 10-fold (22). DnaE2 and RecA, which is essential for DnaE2 induction, thus appear to be important players in the emergence of drug resistance mutations and in the long-term survival of *M. tuberculosis* in the infected host.

Potential advantages of protein splicing inhibitors as anti-TB drugs

Protein splicing inhibitors that inhibit the function of the DnaB, SufB and RecA inteins would have several important advantages as anti-TB drugs:

1. The inhibition of multiple targets, two of which are essential for growth (DnaB and SufB), means that multiple mutations would be required for drug resistance.
2. The inhibitors do not interact with DnaB, SufB or RecA directly and resistance can therefore not arise from mutations in the catalytic domains of these proteins, but only from mutations in the intein. However, most intein mutations lead to uncoupling of the protein splicing pathway (23) and would reduce protein splicing activity.
3. Inhibition of DnaB and SufB would directly inhibit growth, whereas inhibition of RecA would prevent induction of error-prone DNA repair, which contributes to persistence and most mutations related to drug resistance. The latter would serve to suppress the emergence of bacterial strains with inhibitor-resistant DnaB, SufB and RecA inteins even further.
4. The inhibition of DNA replication (DnaB) and DNA repair (RecA) may produce synergistic effects.
5. None of the bacteria normally associated with humans have inteins. Protein splicing inhibitors would therefore be narrow-spectrum antibiotics specific for *Mycobacterium*.
6. Protein splicing does not occur in higher eukaryotes and specific protein splicing inhibitors should therefore have no major side effects.
7. The crystal structures of several inteins has been solved and the mechanism of protein splicing is well understood in chemical terms, opening the way for lead optimization by structure-based drug design.

It should be noted that attributes 1-3 make protein splicing unique among antimycobacterial targets. A major problem with existing anti-TB drugs, and also with the new nitroimidoazopyran and diarylquinoline drugs currently under development, is the emergence of drug resistance. The fact that multiple essential protein splicing targets exist in the cell (DnaB and SufB), that intein mutations conferring drug resistance are likely to be difficult and that the inhibition of RecA splicing is antimutagenic would minimize the development of drug resistance and make protein splicing inhibitors an especially valuable complement to existing anti-TB drugs.

Screening systems for the identification of protein splicing inhibitors

Two types of assay systems for the identification of protein splicing inhibitors have been developed. One of these uses the growth of bacterial cells as an indicator for protein splicing, while the other uses purified proteins and measures a fluorescent signal.

Growth-based screening systems

Thymidylate synthase, encoded by the *thyA* locus, is essential for the synthesis of thymidylate in *E. coli* unless exogenous thymine is provided. On the other hand, in the presence of thymine and the dihydrofolate reductase

inhibitor trimethoprim, thymidylate synthase is lethal because the dihydrofolate that is produced in the course of thymidylate synthesis cannot be recycled to tetrahydrofolate, which is needed for other methylation reactions. Belfort and co-workers (24) took advantage of these selectable *thyA* phenotypes to develop an assay for protein splicing by introducing a plasmid carrying the thymidylate synthase coding region from bacteriophage T4, interrupted by the *M. tuberculosis* RecA intein, into an *E. coli thyA* host. In the absence of thymine, complementation of *thyA* by the recombinant phage gene allows growth and provides a method to select for protein splicing, but in the presence of thymine and trimethoprim, complementation leads to growth inhibition, thus allowing selection against protein splicing (25). Selection for growth in minimal medium in the absence of thymine has been used to define the minimum size of the protein splicing domain (24) and to select for inteins with enhanced splicing activity under specific conditions (26). Conversely, the ability to grow in minimal media supplemented with both thymine and trimethoprim should constitute an effective screen for protein splicing inhibitors (25).

An alternate growth-based screen, based on the *Mycobacterium xenopi gyrA* gene which is interrupted by an intein, was developed by Perler and co-workers (27). Quinolone antibiotics, which specifically bind to GyrA, a subunit of DNA gyrase, promote the covalent binding of GyrA to DNA, causing cell death by blocking the translocation of RNA and DNA polymerases along the DNA template. Although quinolone-resistant variants of GyrA can be generated by single amino acid substitutions, these cannot protect from quinolone-induced cell death if quinolone-sensitive GyrA is present in the same cell. Perler and co-workers (27) described an *in vivo* assay for the inhibition of protein splicing mediated by the *M. xenopi* GyrA intein, based on the conditional dominant lethality of quinolone-sensitive GyrA expressed in a quinolone-resistant *E. coli.* They inserted the GyrA intein of *M. xenopi* into the corresponding site of *E. coli gyrA*, cloned adjacent to an isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible promoter into a low-copy-number vector. This vector was introduced into an *E. coli* host with a quinolone-resistant *gyrA* locus, carrying the GFP coding sequence on a compatible plasmid as a fluorescent indicator of bacterial growth. The inhibition of splicing of the plasmid-encoded quinolone-sensitive GyrA allows bacterial growth in the presence of quinolone antibiotics, thereby providing a screen for protein splicing inhibitors.

A disadvantage of growth-based screens is that substances which attenuate bacterial growth would interfere with the assay. Neither of the cell-based assays for the inhibition of protein splicing has yet been applied successfully to the screening of large compound libraries.

Protein-based screening system

An *in vitro* assay system for inhibitors of protein splicing takes advantage of the observation that green fluorescent protein (GFP) fails to form its chromophore *in vivo* when interrupted by an intein inserted adjacent to residue 129, but that subsequent protein splicing allows fluorescence to develop (28). Ozawa and co-workers used this system in a trans-splicing mode as an *in vivo* assay for protein-protein interactions that bring the intein fragments together and thereby promote protein splicing (28-30). Paulus and co-workers (31) inserted the protein splicing domain of the *M. tuberculosis* RecA intein, in which the homing endonuclease domain was replaced by a His tag to yield a so-called mini-intein, adjacent to a cysteine residue replacing isoleucine-129 of an optimized GFP variant [GFPuv or Cycle 3 mutant (32)], which corresponds to the boundary of a β-sheet and loop segment. As a control, the coding sequence of a nonapeptide was inserted at the same site of GFP. When expressed in *E. coli* JM109 (DE3), both GFP fusion proteins were expressed as inclusion bodies and could be solubilized with 8 M urea as a nonfluorescent protein. Upon renaturation, only the GFP-RecA intein fusion protein developed fluorescence, but not GFP interrupted at the same site by the nonapeptide, demonstrating that the gain of fluorescence depends on protein splicing (31). Renaturation of the fusion protein could be achieved either by dialysis or dilution, but both protein splicing and the gain of fluorescence also depended on the presence of a thiol-reducing agent such as tris(2-carboxyethyl)phosphine (TCEP) (Fig. 4A) on account of the involvement of Cys residues in protein splicing mediated by the RecA intein (33). Possible premature protein splicing in the course of prolonged storage and purification of the GFP-intein fusion proteins was prevented by treating the solubilized inclusion bodies with an excess of 4,4'-dithiodipyridine to block all free Cys residues. Under these conditions, the GFP-RecA intein fusion protein could be purified by immobilized metal ion affinity chromatography (IMAC) and stored in 8 M urea for more than a month without significant loss of TCEP-dependent protein splicing activity and fluorophore formation (31). The principal steps in fusion protein preparation and the assay procedure are illustrated in Figure 5.

In a similar manner, a variant of the *M. tuberculosis* DnaB intein with a His tag near the AgeI site at the *C*-terminus of the homing endonuclease domain was inserted adjacent to a serine residue replacing isoleucine-129 of GFP_{UV}. The resulting fusion protein was also expressed as inclusion bodies, from which it could be solubilized with 8 M urea and purified by immobilized metal affinity chromatography (IMAC). The purified GFP-DnaB intein fusion protein, upon renaturation by dilution or dialysis, underwent protein splicing in the presence of TCEP to yield fluorescent GFP, with the same high signal-to-background ratio as the GFP-RecA intein fusion protein (Fig. 4B).

Library screening using the *in vitro* screening assay

Owing to the fact that the GFP-RecA intein fusion protein could be expressed and purified in a highly stable form and on account of the high reproducibility of its protein splicing characteristics, this fusion protein was cho-

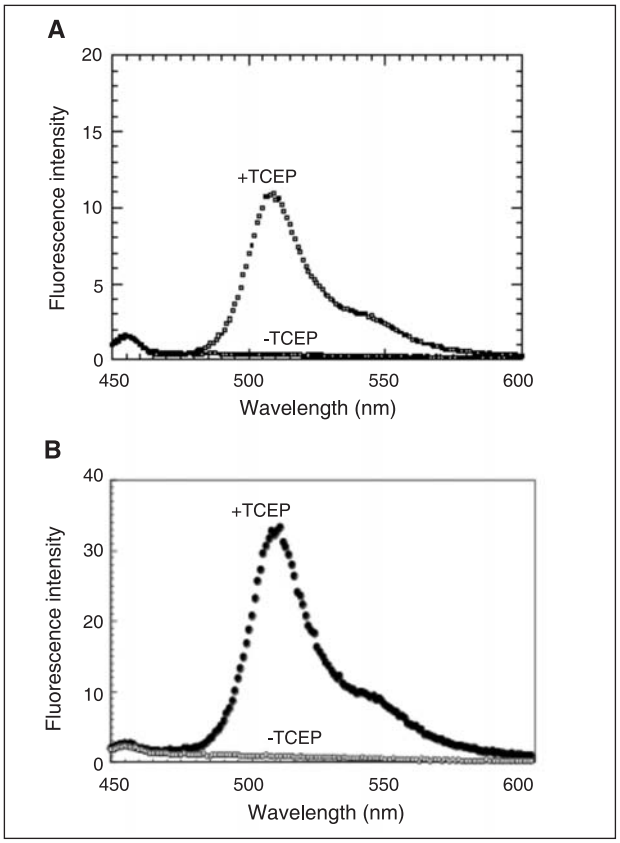

Fig. 4. **A:** Protein splicing *in vitro* with the GFP-RecA mini-intein fusion protein. Denatured inclusion bodies were treated with 4,4'-dithiopyridine and purified as described (31), renatured by dialysis and incubated for 18 h at 25 °C in the absence or presence of 1 mM TCEP. Fluorescence was measured upon excitation at 395 nm. **B:** Protein splicing *in vitro* with the GFP-DnaB intein fusion protein, labeled with a His tag inserted just downstream of the homing endonuclease domain. Denatured inclusion bodies were treated with 4,4'-dithiopyridine and purified by IMAC, renatured by dialysis and incubated for 18 h at 17 °C in the absence or presence of 1 mM TCEP. Fluorescence was measured upon excitation at 395 nm.

sen as the basis of a highly robust high-throughput screening (HTS) system for protein splicing inhibitors. The protein splicing precursor, which is purified as the 4-thiopyridine derivative in 8 M urea, can be refolded by dilution into phosphate buffer at pH 7.0 and retains full competence to undergo protein splicing upon activation by a thiol-reducing agent such as TCEP for several hours. This allowed refolding of the precursor protein before addition of the compounds to be tested, thereby precluding possible inhibition of protein refolding. After addition of the compounds to be tested for inhibition, a 30-min delay was imposed to allow for interaction of the inhibitor with the precursor protein before protein splicing was initiated by the addition of TCEP. Because the half-time of the protein splicing reaction is 6 h under the experimental conditions used, it was possible to make a "0-time" fluorescence measurement to correct for possible fluorescence of the compounds to be tested. After incubation for 2.5 half-times (15 h), the fluorescence of the GFP produced as a result of protein splicing was determined. All assays were done in duplicate and positive inhibition was recorded only if both duplicate values were more than 20% below the uninhibited control values.

Primary screening

Library screening utilized the facilities of the Institute of Chemistry and Cell Biology (ICCB) of Harvard University. The robustness of the assay and the quality of the liquid handling equipment used assured excellent reproducibility, with a Z'-factor of 0.88 (34). The libraries selected for screening were primarily commercial libraries, compiled by the staff of the ICCB and the Laboratory for Drug Discovery for Neurodegeneration at Harvard Medical School on the basis of favorable physicochemical properties and lack of functional groups that might cause toxicity and reduced stability. The producers of two libraries that account for nearly half of the compounds screened (Maybridge and ChemDiv) claim that their libraries cover a large segment of pharmacophore space and that the majority of their compounds have drug-like properties as defined by the Lipinski "Rule of Five" (35). The fact that the screening was distributed over 6 major commercial libraries provided additional diversity.

The compound libraries screened and the resulting positives are summarized in Table I. The hit rates in the various libraries tested differed significantly, but the average rate of positives of 0.4% is typical for an *in vitro* screening system that is not unduly susceptible to perturbation.

Secondary screening

With 320 positives resulting from the HTS of about 85,000 compounds at an average concentration of about 20 μM, it was important to re-analyze these hits to determine whether the inhibition seen in the original screen could also be seen at lower concentrations and whether its target was indeed protein splicing. The latter issue arises from the fact that our HTS couples protein splicing to GFP chromophore formation. The inhibition of fluorescence observed in our assay could arise either from the inhibition of protein splicing or from the inhibition of GFP chromophore formation, the self-catalyzed oxidative rearrangement of an internal Ser-Tyr-Gly tripeptide (36), *i.e.,* either of the last two steps illustrated in Figure 5. In order to distinguish between these possibilities, we devised a secondary screen in which GFP chromophore formation was assayed separately, based on the isolation of nonfluorescent GFP inclusion bodies expressed in *E. coli* at 43 °C as the 4,4'-dithiodipyridine adduct, involving their renaturation by dilution from 8 M urea, incubation with the compounds to be tested as inhibitors, followed by the addition of TCEP to initiated chromophore formation, which occurred with a half-time of about 2 h under our experimental conditions. The need to reduce the 4,4'-

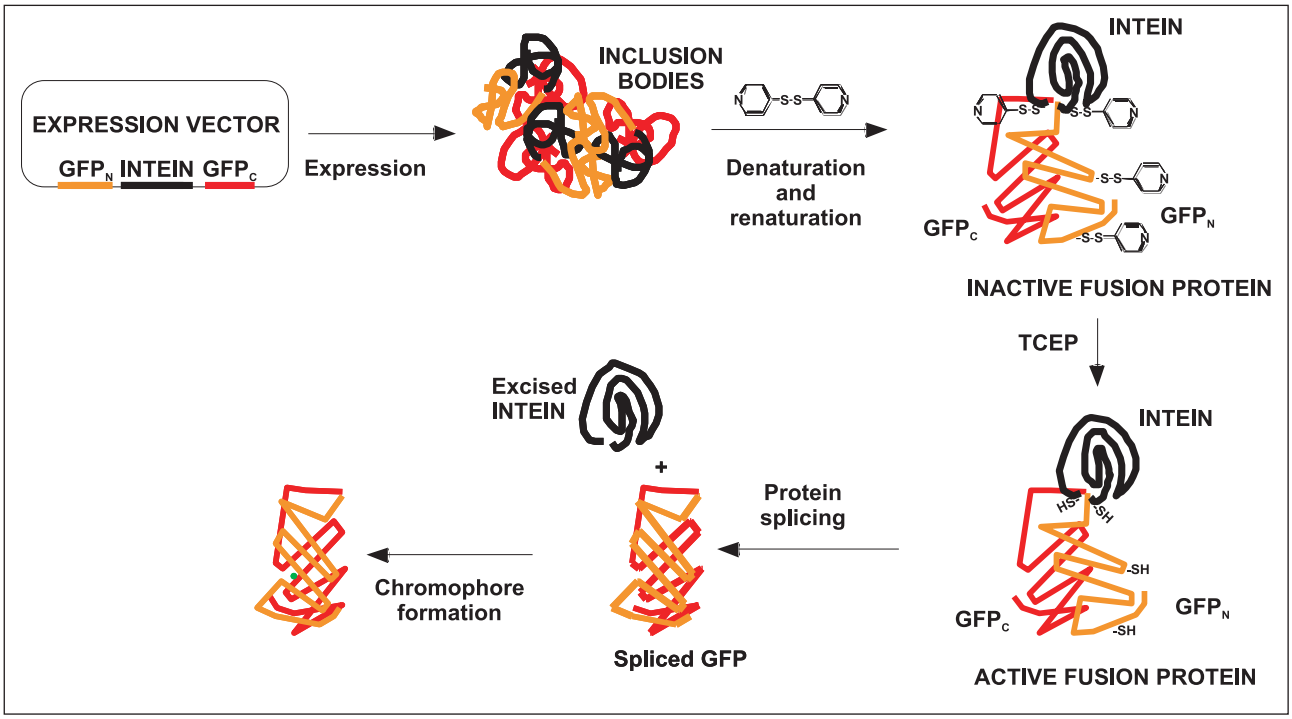

Fig. 5. Schematic representation of the steps involved in the GFP-based assay for protein splicing. The *N*- and *C*-terminal segments of GFP are indicated in beige and red, respectively, and the intein in black.

Table I: Libraries screened for protein splicing inhibitors.

Compound library	# of compounds	# of positives
Commercial Diversity Set	5,056	21 (0.42%)
Diversity Oriented Synthesis	16,030	23 (0.14%)
ICCB Bioactives	489	4 (0.82%)
ChemDiv Combilab/International	28,864	91 (0.32%)
Bionet	6,168	33 (0.54%)
Maybridge	16,807	68 (0.40%)
Peakdale	3,168	2 (0.06%)
Biomol-TimTec	8,158	72 (0.88%)
Mixed Commercial Plates	1,254	9 (0.72%)
Total	85,274	323 (0.38%)

dithiopyridine adduct of GFP prior to chromophore formation indicates that a free Cys residue is essential for this process, confirming earlier observations (37). Accordingly, inhibitors that react nonspecifically with the Cys residue would inhibit GFP chromophore formation and be eliminated by our secondary screening procedure.

The results of this secondary screening clearly revealed two classes of inhibitors: those that inhibited the intein-based assay significantly more than GFP chromophore formation and those that inhibited the two assays equally. After eliminating compounds that inhibited the intein-based assay and GFP chromophore formation to a similar extent, the remaining hits were assayed over a wide concentration range for inhibition of *in vitro* protein splicing involving both the *M. tuberculosis* RecA and DnaB inteins, which were inserted into GFP at position 129 to allow a quantitative fluorescence-based assay similar to that used in the HTS. In addition, the protein splicing reaction mixtures were analyzed by SDS-PAGE to provide a direct assay for the products of protein splicing (excised intein and spliced GFP), independent of GFP chromophore formation and fluorescence measurements. An example of such an assay is shown in Figure 6. Fifty-nine of our confirmed hits had IC_{50} values for at least one of the two inteins below 20 μM, the compound concentration used in the HTS. More than half of these compounds had IC_{50} values for both the RecA and the DnaB inteins of 5 μM or less and 25% had measurable MIC values against *M. tuberculosis* H37Rv. Antimycobacterial activity was measured against both *M. bovis* BCG and *M. tuberculosis* H37Rv by the Alamar Blue assay, which measures the reduction of Alamar Blue to a red fluorescent dye by the metabolic activity of surviving cells (38). The MIC is defined as the lowest drug concentration that prevents the color change from blue to red. The assays with *M. tuberculosis* H37Rv were done in collaboration with Dr. Chris Sassetti at the University of Massachusetts Medical School and involved testing five different compound concentrations (20, 10, 5, 2.5 and 1.25 μg/ml). With a few exceptions, antimycobacterial activity against *M. bovis* BCG and *M. tuberculosis* H37Rv agrees reasonably well, suggesting that MIC studies with *M. bovis* BCG have some predictive value for growth inhibition of virulent strains of *M. tuberculosis*. Although only about 25% of our compounds show antimycobacterial activity, it is gratifying that one in four protein splicing inhibitors can enter the mycobacterial cell and inhibit growth. Clearly, the MIC values obtained with the compounds on hand are

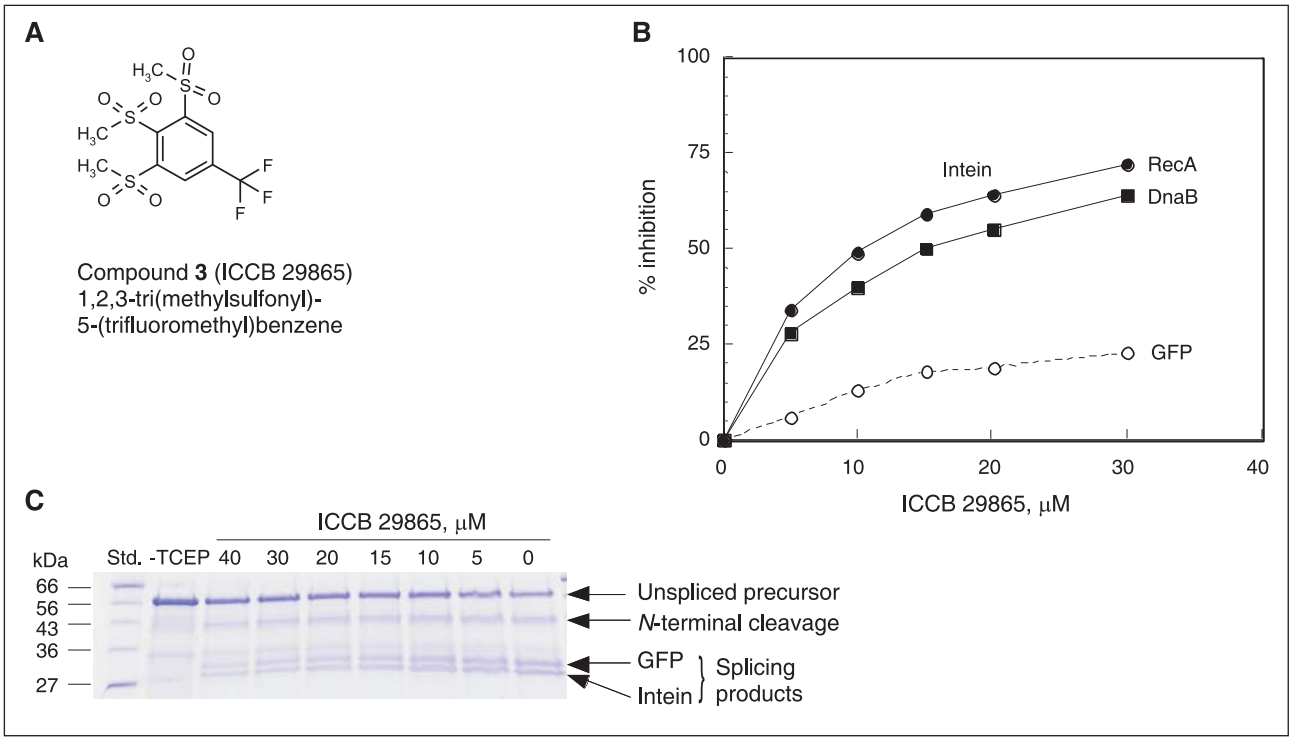

Fig. 6. Determination of the IC_{50} and specificity of compound **3** (see Table II) as a protein splicing inhibitor. **A:** Structure of compound **3.** **B:** Effect of compound **3** on the fluorometric assay of protein splicing mediated by the RecA and DnaB inteins inserted at position 129 of GFP and on chromophore formation of GFP after refolding from inclusion bodies. **C:** Effect of compound **3** on protein splicing mediated by the RecA intein inserted into GFP, measured by the conversion of the precursor protein to spliced products by denaturing polyacrylamide gel electrophoresis.

higher than would be expected of an effective drug and significant optimization will be required. Comparison of the MIC values for some of our compounds with those for the three mainline TB drugs, pyrazinamide (6.2 μg/ml), rifampin (0.2 μg/ml) and isoniazid (0.02 μg/ml), provides an indication of the lead optimization necessary to bring the MICs of the protein splicing inhibitors into a comparable range.

Mechanism of action of protein splicing inhibitors

The 59 compounds that specifically inhibit RecA- and DnaB-mediated protein splicing in the low micromolar range are listed in Table II. They include diverse pharmacophores, but as indicated, more than half of these compounds can be grouped into 9 structural classes, which are shown in Figure 7. A general characteristic of most protein splicing inhibitors is that they are electrophiles, such as Michael acceptors or α-haloketones, suggesting that they may react with a nucleophilic residue on the inteins, such as the *N*-terminal catalytic cysteine residue. This notion was supported by the observation that the inhibition of protein splicing by most compounds is attenuated by the presence of a thiol such as dithiothreitol in the 1 mM concentration range (Table III) and that the inhibition of protein splicing is irreversible (data not shown). Indeed, experiments with a model peptide corresponding to the *N*-terminus of the RecA intein showed that all but 4 of the compounds listed in Table II formed a covalent adduct with the Cys thiol group. However, the estimated second-order rate constant for the reaction of a typical inhibitor with the model peptide is approximately 10^{-5} that observed with the intact intein, suggesting that inhibitor binding near the intein active center greatly contributes to reactivity. The contribution of inhibitor binding to the intein active site is probably greater than suggested by this factor, since it is known that the *N*-terminal cysteine residue of the RecA intein has a pK_a of 8.2, indicating that it is a relatively weak nucleophile (39).

The recent elucidation of the crystal structure of the *M. tuberculosis* RecA intein (40) allows some insight into the mechanism of protein splicing inhibition. Although inteins do not have a substrate binding pocket, their "substrate" being the polypeptide bonds that link the intein to the exteins (Fig. 2), the RecA intein, as well as all other inteins whose structure is known, have a shallow cavity near their active center, to which the side-chain of the *N*-terminal cysteine residue is exposed (Fig. 8). The binding of an electrophilic compound to this cavity in an orientation that favors interaction with the reactive thiol of the intein would promote its alkylation and the irreversible inhibition of protein splicing.

In retrospect, considering that inteins are covalently linked to their "substrates" and that protein splicing does not involve substrate binding to a classical substrate-binding pocket that can be blocked by a competitive

Table II: Properties of confirmed protein splicing inhibitors.

Code #	Library	M_r	Struct. class	IC_{50} for RecA intein (μM)	IC_{50} for DnaB intein (μM)	MIC for *M. bovis* BCG (μg/ml)	MIC for *M. tuberculosis* H37Rv (μg/ml)
1	MixCom	366	M	5	30	-	-
2	Maybridge	388		8	13	-	-
3	Maybridge	380		10	17	20	20
4	Maybridge	314		5	5	-	-
6	Bionet	381	C	4	13	-	-
7	Bionet	371	C	4	6	-	-
8	Bionet	386		5	5	-	-
9	Bionet	385	D	10	7	5	5
10	Bionet	380	C	8	7	-	-
11	Bionet	356	C	4	8	5	5
12	Bionet	398	C	3	5	20	10
13	Bionet	361		4	8	20	10
17	MixCom	466		20	15	-	-
21	Maybridge	253		8	15	-	-
22	Maybridge	302		3	5	-	20
23	Maybridge	317		3	<5	-	-
24	Maybridge	174		11	12	-	5
27	Maybridge	425		2	<5	-	-
29	Maybridge	341		30	<5	-	-
30	Maybridge	392	K	5	<5	20	-
31	Maybridge	251	K	5	5	6	10
33	Maybridge	381	K	4	<5	-	-
35	Bionet	370	C	1	<<5	20	-
36	Bionet	384	C	5	5	-	-
37	ChemDiv	361		5	<5	-	-
38	MixCom	367		0,8	<5	-	-
39	Bionet	351	E	0,8	<5	-	-
40	Bionet	399	B	5	<5	-	-
41	Bionet	395	E	2	<5	-	-
43	Bionet	399	D	3	<5	1	2,5
44	Bionet	334		5	<<5	-	-
45	Bionet	347		5	<5	5	2,5
51	ChemDiv	561		2	<5	-	-
53	ChemDiv	472	A	1	<<5	~20	-
54	ChemDiv	375		6	11	5	10
55	ChemDiv	384	A	15	7	5	20
56	ChemDiv	373	N	<35	10	-	-
57	Bionet	354		<5	5	-	-
58	Bionet	385	B	<5	7	-	-
59	Bionet	385	B	<5	<5	-	-
62	Maybridge	385	B	<5	5	-	-
64	Maybridge	377	A	6	<5	-	-
65	Maybridge	427	A	<<5	<<5	-	-
66	Maybridge	414	A	<<5	<<5	>20	-
67	Maybridge	365		<5	5	-	-
68	Maybridge	271		8	9	-	-
69	Maybridge	381	M	<5	37	-	-
70	Maybridge	346		12	5	-	-
71	Maybridge	446	N	<<5	<<5	10	10
72	Maybridge	406	N	<<5	<<5	10	10
74	TimTec	198		<5	nd	-	-
75	TimTec	408	P	<5	nd	-	-
76	TimTec	267	P	8	nd	-	-
77	TimTec	309	P	30	nd	-	-
78	TimTec	325	P	13	nd	-	-
80	TimTec	369		8	nd	-	-
81	TimTec	380		12	nd	-	-
82	TimTec	295		8	nd	-	-
83	TimTec	273		7	nd	-	10

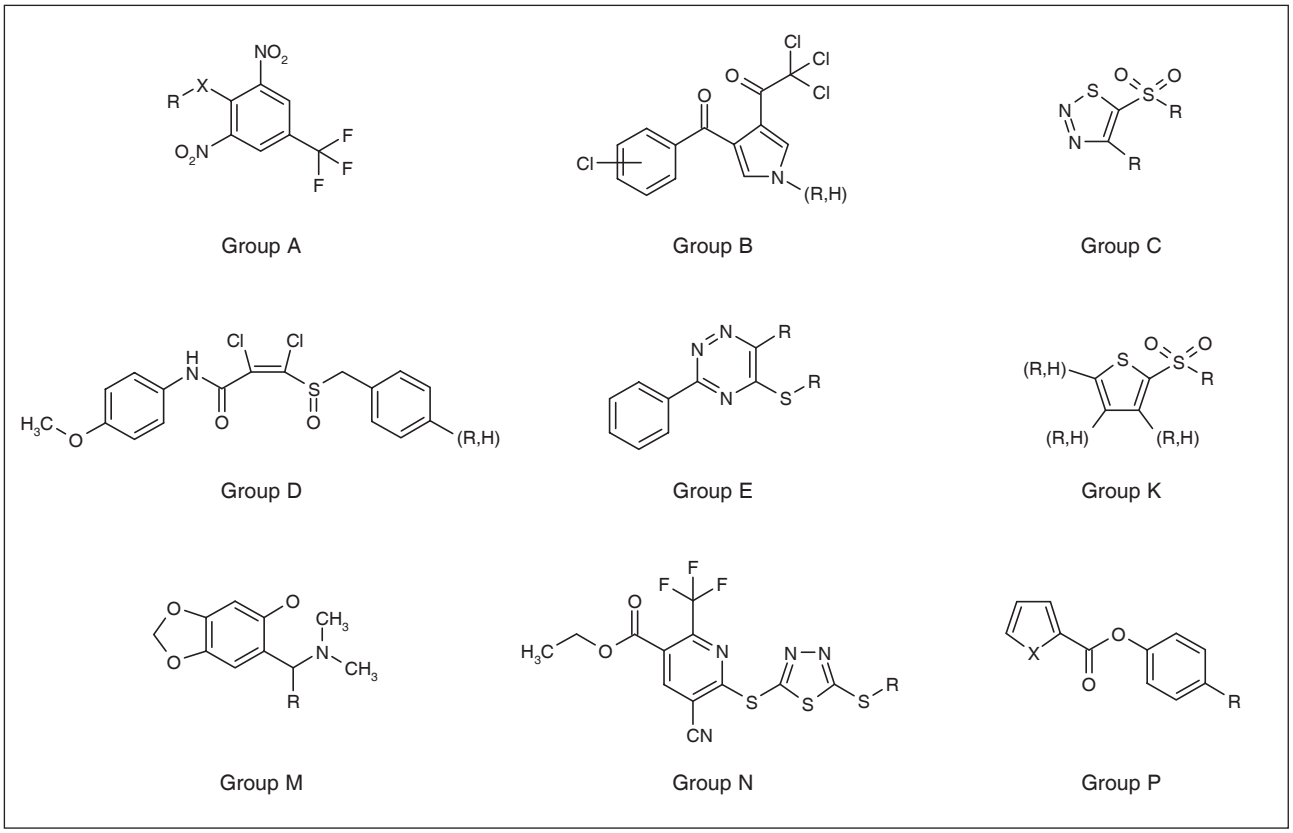

Fig. 7. Recurring structural motifs among confirmed protein splicing inhibitors. The structure groups illustrated encompass nearly 60% of the confirmed inhibitors identified in the HTS. R = alkyl or aryl substituent, X = S or O.

Table III. Effect of 1 mM DTT on the inhibition of protein splicing mediated by the RecA and DnaB inteins.

Inhibitor	RecA			DnaB		
Cpd. #	Concentration (μM)	% Inhibition with TCEP	% Inhibition with DTT	Concentration (μM)	% Inhibition with TCEP	% Inhibition with DTT
1	30	79	80	30	52	88
2	15	78	2	15	62	10
3	20	73	0	30	64	0
4	10	61	38	10	52	44
6	10	68	13	15	56	30
7	10	77	14	10	67	32
8	10	68	25	5	51	49
9	20	63	6	15	73	0
10	15	71	6	15	74	6
11	10	60	8	15	71	6
12	10	74	10	10	21	10
17	30	71	13	30	39	12

inhibitor, and in the absence of any evidence for allosteric transitions in the course of protein splicing, it is unavoidable that the mechanism for the inhibition of protein splicing is noncompetitive inhibition, *i.e.*, the covalent modification of an essential amino acid side-chain. This explains why almost every protein splicing inhibitor identified in the course of an extensive HTS is an electrophile that binds covalently to the catalytic thiol of the intein.

The use of irreversible inhibitors as drugs

Traditional thinking in drug discovery has been to disregard drug candidates that bind covalently to their target. Indeed, most drug screening protocols are carried out in the presence of dithiothreitol in order to avoid compounds such as Michael acceptors that modify cysteine residues. However, A.J. Bridges at Pfizer has been a leader in pro-

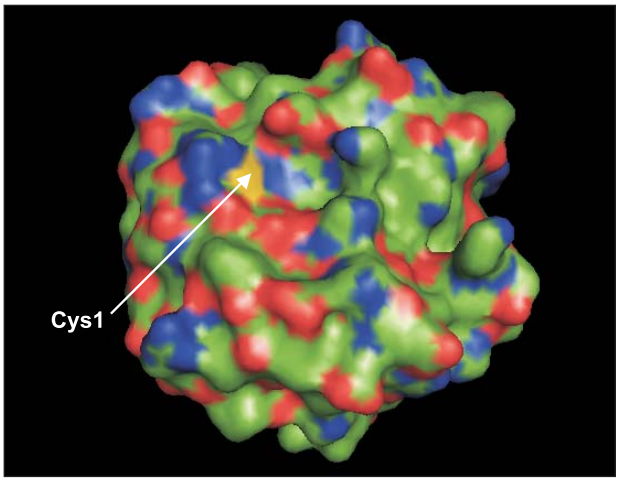

Fig. 8. Structure of the *M. tuberculosis* RecA intein (40) (Protein Data Bank structure 2INO), depicted as a surface representation using PyMol, with *N*-containing side-chains indicated in blue, *O*-containing side-chains in red and sulfur in yellow. The *N*-terminal active-site Cys residue is indicated by the arrow.

moting a change of thinking on this question in the protein kinase field. To quote from one of his papers: "The irreversible nature of these compounds may offer other potential advantages in terms of target suppression and *in vivo* pharmacokinetics. Prolonged suppression of the kinase target will likely be necessary for maximum antitumor activity, and an irreversible inhibitor will provide an advantage in this respect by permanently inhibiting kinase activity, which will return only when new receptor is synthesized" (41). This principle is now being increasingly applied to the design of protein kinase inhibitors as anticancer drugs (42).

The challenge in using reactive compounds such as Michael acceptors as drugs is to find substances with sufficiently low reactivity that they do not react indiscriminately with all nucleophiles, but depend on strong specific binding near a nucleophilic catalytic group so as to alkylate this group with a high degree of selectivity. The lead optimization efforts on protein splicing inhibitors that are now in progress are focusing on this objective. One approach is taking advantage of the known crystal structure of the *M. tuberculosis* RecA intein (Fig. 8) to optimize the binding and selectivity of the protein splicing inhibitors identified in the HTS described here.

Acknowledgements

The author thanks the Institute of Chemistry and Cell Biology (ICCB) and the Laboratory for Drug Discovery for Neurodegeneration at Harvard Medical School for making their high-throughput screening facilities available for this work; Drs. Caroline Shamu and Su Chiang for their advice and the ICCB staff for their assistance; Dr. Chris Sassetti for the MIC determinations on *M. tuberculosis* H37Rv; Dr. Jaya Gangopadhyay for her contribution to the development of the protein splicing assay; Dr. Shuqin Jiang for his help in protein purification and the screening procedures; and Drs. Marlene Belfort and Patrick van Roey for insightful discussions of the implications of this work, especially in light of the structure of the RecA intein.

References

1. Espinal, M.E., Laszlo, A., Simonsen, L. et al. *Global trends in resistance to antituberculosis drugs*. N Engl J Med 2001, 344(17): 1294-303.

2. Raviglione, M. *XDR-TB: Entering the post-antibiotic era?* Int J Tuberc Lung Dis 2006, 10(11): 1185-7.

3. Corbett, E., Watt, C., Walker, N., Maher, D., Williams, B., Raviglione, M., Dye, C. *The growing burden of tuberculosis: Global trends and interactions with the HIV epidemic*. Arch Intern Med 2003, 163(9): 1009-21.

4. Stover, C., Warrener, P., VanDevanter, D. et al. *A small-molecule nitroimidazopyran drug candidate for the treatment of tuberculosis*. Nature 2000, 405(6789): 962-6.

5. Andries, K., Verhasselt, P., Guillemont, J. et al. *A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis*. Science 2005, 307(5707): 223-7.

6. Paulus, H. *Protein splicing and related forms of protein autoprocessing*. Annu Rev Biochem 2000, 69: 447-96.

7. Pietrokovski, S. *Modular organization of inteins and C-terminal autocatalytic domains*. Protein Sci 1998, 7(1): 64-71.

8. Perler, F.B. *InBase, the intein database*. Nucl Acids Res 2002, 30(1): 383-4.

9. Huet, G., Daffe, M., Saves, I. *Identification of the Mycobacterium tuberculosis SUF machinery as the exclusive mycobacterial system of [Fe-S] cluster assembly; evidence for its implication in the pathogen's survival*. J Bacteriol 2005, 187(17): 6137-46.

10. Huet, G., Castaing, J.-P., Fournier, D., Daffe, M., Saves, I. *Protein splicing of sufB is crucial for the functionality of the Mycobacterium tuberculosis SUF machinery*. J Bacteriol 2006, 188(9): 3412-4.

11. Sassetti, C.M., Boyd, D.H., Rubin, E.J. *Genes required for mycobacterial growth defined by high density mutagenesis*. Mol Microbiol 2003, 48(1): 77-84.

12. Arai, K.I., Kornberg, A. *A general priming system employing only DnaB protein and primase for DNA replication*. Proc Natl Acad Sci USA 1979, 76(9): 4308-12.

13. Saluja, D., Godson, G.N. *Biochemical characterization of Escherichia coli temperature-sensitive dnaB mutants dnaB8, dnaB252,dnaB70, dnaB43, and dnaB454*. J Bacteriol 1995, 177(4): 1104-11.

14. Kogoma, T. *Two types of temperature sensitivity in DNA replication of an Escherichia coli dnaB mutant*. J Mol Biol 1976, 103(1): 191-7.

15. Bird, L.E., Pan, H., Soultanas, P., Wigley, D.B. *Mapping protein-protein interactions within a stable complex of DNA primase and dnaB helicase from Bacillus stearothermophilus*. Biochemistry 2000, 39(1): 171-82.

16. Movahedzadeh, F., Colston, M.J., Davis, E.O. *Determination of DNA sequences required for regulated Mycobacterium tuber-*

culosis recA expression in response to DNA-damaging agents suggests that two modes of regulation exist. J Bacteriol 1997, 179(11): 3509-18.

17. Davis, E.O., Jenner, P.J., Brooks, P.C., Colston, M.J., Sedgwick, S.G. *Protein splicing in the maturation of M. tuberculosis RecA protein: A mechanism for tolerating a novel class of intervening sequence*. Cell 1992, 71(2): 201-10.

18. Kumar, R.A., Vaze, M.B., Chandra, N.R., Vijayan, M., Muniyappa, K. *Functional chracterization of the precursor and spliced forms of RecA protein of Mycobacterium tuberculosis*. Biochemistry 1996, 35(6): 1793-802.

19. Sander, P., Papavinasundaram, K.G., Dick, T. et al. *Mycobacterium bovis BCG recA deletion mutant shows increased susceptibility to DNA-damaging agents but wild-type survival in a mouse infection model*. Infect Immun 2001, 69(6): 3562-8.

20. Davis, E., Dullaghan, E., Rand, L. *Definition of the mycobacterial SOS box and use to identify lexA-regulated genes in Mycobacterium tuberculosis*. J Bacteriol 2002, 184(12): 3287-95.

21. Rand, L., Hinds, J., Springer, B., Sander, P., Buxton, R., Davis, E. *The majority of inducible DNA repair genes in Mycobacterium tuberculosis are induced independently of recA*. Mol Microbiol 2003, 50(3): 1031-41.

22. Boshoff, H., Reed, M., Barry, C., Mizrah, V. *DnaE2 polymerase contributes to in vivo survival and the emergence of drug resistance in Mycobacterium tuberculosis*. Cell 2003, 113(2): 183-93.

23. Chong, S., Williams, K.S., Wotkovicz, C., Xu, M.Q. *Modulation of protein splicing of the Saccharomyces cerevisiae vacuolar membrane ATPase intein*. J Biol Chem 1998, 273(17): 10567-77.

24. Derbyshire, V., Wood, D.W., Wu, W., Dansereau, J.T., Dalgaard, L.Z., 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(21): 11466-71.

25. Belfort, M. (Health Research, Inc.). *Inteins as antimicrobial targets: Genetic screen for intein function*. US 5795731.

26. Wood, D.W., Wu, W., Belfort, G., Derbyshire, V., Belfort, M. *A genetic system yields self-cleaving inteins for bioseparations*. Nat Biotechnol 1999, 17(9): 889-92.

27. Adam, E., Perler, F.B. *Development of a positive genetic selection system for inhibition of protein splicing using mycobacterial inteins in Escherichia coli gyrase subunit A*. J Mol Microbiol Biotechnol 2002, 4(5): 479-87.

28. Ozawa, T., Nogami, S., Sato, M., Ohya, Y., Umezawa, Y. *A fluorescent indicator for detecting protein-protein interactions in vivo based on protein splicing*. Anal Chem 2000, 72(21): 5151-7.

29. Ozawa, T., Takeuchi, M., Kaihara, A., Sato, M., Umezawa, Y. *Protein splicing-based reconstitutition of split green fluorescent protein for monitoring protein-protein interactions in bacteria: Improved sensitivity and reduced screening time*. Anal Chem 2001, 73(24): 5866-74.

30. Ozawa, T., Umezawa, Y. *Detection of protein-protein interactions in vivo based on protein splicing*. Curr Opin Chem Biol 2001, 5(5): 578-83.

31. Gangopadhyay, J.P., Jiang, S.-Q., Paulus, H. *An in vitro screening system for protein splicing inhibitors based on green fluorescent protein as an indicator*. Anal Chem 2003, 75(10): 2456-62.

32. Crameri, A., Whitehorn, E.A., Tate, E., Stemmer, W.P.C. *Improved green fluorescent protein by molecular evolution using DNA shuffling*. Nat Biotechnol 1996, 14(3): 315-9.

33. Mills, K.V., Lew, B.M., Jiang, S., Paulus, H. *Protein splicing in trans by purified N- and C-terminal fragments of the Mycobacterium tuberculosis RecA intein*. Proc Natl Acad Sci USA 1998, 95(7): 3453-8.

34. Zhang, J., Chung, T., Oldenburg, K. *A simple statistical parameter for use in evaluation and validation of high throughput screening assays*. J Biomol Screen 1999, 4(2): 67-73.

35. Lipinski, C.A., Lombardo, F., Dominy, B.W., Febey, P.J. *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings*. Adv Drug Deliv Rev 2001, 46(1-3): 3-26.

36. Niwa, H., Inouye, S., Hirano, T. et al. *Chemical nature of the light emitter of the Aequorea green fluorescent protein*. Proc Natl Acad Sci USA 1996, 93(24): 13617-22.

37. Reid, B.G., Flynn, G.C. *Chromophore formation in green fluorescent protein*. Biochemistry 1997, 36(22): 6786-91.

38. Franzblau, S.G., Witzig, R.S., McLaughlin, J.C. et al. *Rapid, low-technology MIC determination with clinical Mycobacterium tuberculosis isolates by using the microplate alamar blue assay*. J Clin Microbiol 1998, 36(2): 362-6.

39. Shingledecker, K., Jiang, S.-Q., Paulus, H. *Reactivity of the cysteine residues in the protein splicing active center of the Mycobacterium tuberculosis RecA intein*. Arch Biochem Biophys 2000, 375(1): 138-44.

40. Van Roey, P., Pereira, B., Li, Z., Hiraga, K., Belfort, M., Derbyshire, V. *Crystallographic and mutational studies of Mycobacterium tuberculosis recA mini-inteins suggest a pivotal role for a highly conserved aspartate residue*. J Mol Biol 2007, 367(1): 162-73.

41. Fry, D., Bridges, A., Denny, W. et al. *Specific, irreversible inactivation of the epidermal growth factor receptor and erbB2, by a new class of tyrosine kinase inhibitors*. Proc Natl Acad Sci USA 1998, 95(20): 12022-7.

42. Cohen, M.S., Zhang, C., Shokat, K.M., Taunton, J. *Structural bioinformatics-based design of selective, irreversible kinase inhibitors*. Science 2005, 308(5726): 1318-21.