[Frontiers in Bioscience 8, s1157-1165, September 1, 2003]

INTEINS AS TARGETS FOR POTENTIAL ANTIMYCOBACTERIAL DRUGS

Henry Paulus

Boston Biomedical Research Institute, Watertown, Massachusetts 02472, and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115

TABLE OF CONTENTS

1. ABSTRACT

Protein splicing is a self-catalyzed process mediated by inteins. The observation that inteins occur only in microorganisms and that they often interrupt genes that play an essential role in nucleic acid metabolism makes them attractive as potential antibacterial targets. Because mycobacteria are the only intein-containing bacteria associated with human hosts, inteins would represent highly specific antimycobacterial targets. In *Mycobacterium tuberculosis*, two important proteins of DNA repair and replication, RecA and DnaB, respectively, are interrupted by inteins that must be excised by protein splicing before these proteins can function. This review describes the screening systems for the detection of mutations or inhibitors that interfere with proteins splicing that have been developed and published to date. In three of these experimental system, inteins have been inserted into proteins that are toxic under certain conditions. Protein splicing therefore leads to conditional growth inhibition or cell death and its inhibition can be monitored in terms of bacterial growth. A fourth assay for protein splicing and its inhibition is based on purified proteins and measures the formation of Green Fluorescent Protein or its inhibition. The advantages of inteins as antimycobacterial targets are discussed.

2. INTRODUCTION

Tuberculosis is the most widespread infectious disease. One-third of the world's population (2 billion people) are infected with *Mycobacterium tuberculosis* and 5-10% of these suffer active disease, leading to nearly 3 million deaths annually. In the long-term, vaccination is the only way to reduce the incidence of tuberculosis, especially in developing countries where lengthy combination drug therapy is problematical. However, vaccination does not address the need of the 2 billion individuals who are infected right now nor of the alarmingly growing number of patients in developed countries who are suffering from multidrug-resistant (MDR) tuberculosis, which is essentially refractory to drug therapy (1). The mainline anti-tuberculosis drugs, isoniacid, rifampin, pyrazinamide, ethambutol, and, to some extent, streptomycin need thus be supplemented by additional drugs against new targets, until those, too, succumb to resistance mechanisms. It is therefore important to keep developing an armamentarium of drugs against a wide range of targets to use in combination with established drugs and replace these when they begin to fail. 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. Another context in which a new type of antimycobacterial drugs may play an important role is in response to potential bioterrorist attacks involving *M. tuberculosis*. In such a situation, it would be desirable to treat large populations prophylactically to avoid the establishment of infections, the treatment of which would require a long course of costly multidrug therapy.

Much effort is underway to identify new targets for antimycobacterial drugs. Shortly after the discovery of protein splicing in mycobacteria (2, 3), it was suggested that mycobacterial inteins may have a role in virulence (4) and may represent targets for novel antimycobacterial drugs (5). Because protein splicing has not been observed in multicellular organisms and mycobacteria are the only intein-containing bacteria associated with human hosts, protein splicing would be a highly specific antimycobacterial target. This review describes various assay systems that can be used to screen for protein splicing inhibitors.

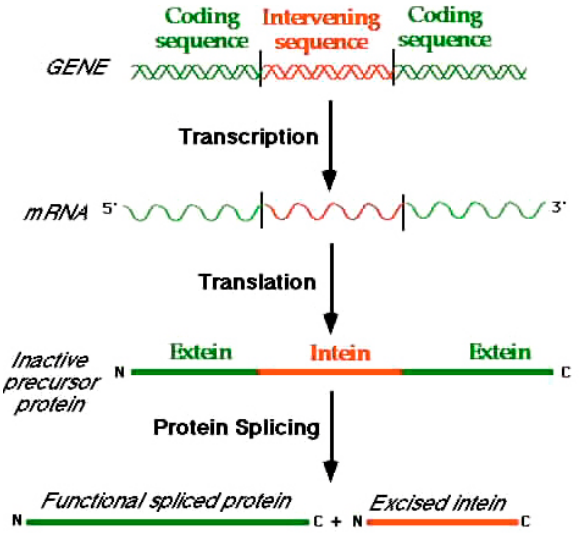

Figure 1. The role of protein splicing as a posttranslational step in gene expression.

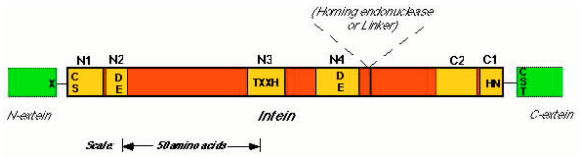

Figure 2. Conserved elements in a typical intein. The shaded areas are conserved intein motifs, identified by the nomenclature of (54). Amino acid residues that are conserved in >90% of known inteins are shown using the single-letter notation (X = any amino acid). The conserved motifs and other intein sequences are to scale, based on the *M. xenopi* GyrA intein. The site of insertion of the homing endonuclease domain or linker regions is indicated by the dark vertical. In the *S. cerevisiae* VMA intein, a DNA recognition region (DRR) is inserted just before motif N4.

3. BRIEF OVERVIEW OF PROTEIN SPLICING

Protein splicing is a form of posttranslational 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 (Figure 1). Unlike RNA splicing, another mechanism for the excision of intervening sequences, which occurs primarily in eukaryotes and multicellular organisms, protein splicing has been observed in bacteria, archaea, and unicellular eukaryotes and in eukaryotic organelles. About 150 examples of protein splicing are now known (6), and comparison of the known inteins has revealed conserved elements that appear to play roles in intein structure and the catalysis of protein splicing. As summarized in Figure 2, 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 six relatively conserved sequence motifs (Figure 2). Many inteins also harbor a homing endonuclease domain, which is functionally unrelated to protein splicing but allows the horizontal transfer of inteins between organisms. Unlike introns, inteins always interrupt highly conserved regions of their host protein and therefore disrupt its function until excised by protein splicing. Protein splicing can thus be considered a post-translational modification that leads to the conversion of an inactive precursor to a functional enzyme.

An important breakthrough in the understanding of protein splicing was the demonstration that it is a self-catalyzed process which requires neither cofactors nor accessory proteins, and that all catalytic groups that are required reside in the intein (7). Like RNA splicing, protein splicing does not occur by bond hydrolysis and re-ligation but rather by bond rearrangements, in which ester intermediates play a critical role (for recent reviews, see 8, 9, 10). It involves three intein-catalyzed reactions: (a) Formation of a linear ester intermediate by an N-O of N-S acyl rearrangement involving the nucleophilic amino acid at the N-terminal splice junction; (b) formation of a branched ester intermediate by the attack of the nucleophilic residue at the C-terminal splice junction on the linear ester intermediate; and (c) cyclization of the Asn residue adjacent to the C-terminal splice junction, coupled to cleavage of the branched intermediate to yield an excised intein and the two exteins joined by an ester bond. These reactions are followed by the rapid spontaneous rearrangement of the ester linking the exteins to the more stable amide bond.

The crystal structures of three inteins, representative of the three kingdoms of life, have been solved: the Sce VMA intein (11-14), the Mxe GyrA intein (15), and the Pfu RIR1-1 intein (16). The structures of the three inteins are similar and consist of a compact, horseshoe-shaped fold composed of antiparallel beta-strands. The fact that the disposition of the conserved residues shown in Figure 2 relative to the scissile bonds is nearly the same in these diverse inteins suggests common structural features that could serve as potential targets for protein splicing inhibitors.

4. MYCOBACTERIAL INTEINS

The first bacterial intein to be described was found to interrupt the *recA* locus *Mycobacterium tuberculosis* (2, 3). RecA inteins were subsequently also found in *Mycobacterium leprae* and closely related members of the *M. tuberculosis* complex (*M. bovis* BCG, and *M. microti*), but not in 14 saprophytic *Mycobacterium* species examined (4). The correlation between RecA inteins and pathogenesis could reflect a specific function or just be a coincidence owing to horizontal intein transfer mediated by a homing endonuclease. However, the fact that the inteins of *M. tuberculosis* and *M. leprae* are quite unrelated (27% identity) and are inserted at different sites in the closely related (92% identity) RecA proteins (Figure 3) indicates that they could not have been transmitted by the same homing endonuclease. Rather, it appears that the presence of inteins in the RecA proteins have been selected because they provide an advantage to pathogenic mycobacteria. However, a more recent survey of mycobacterial *recA* genes has shown that inteins related to

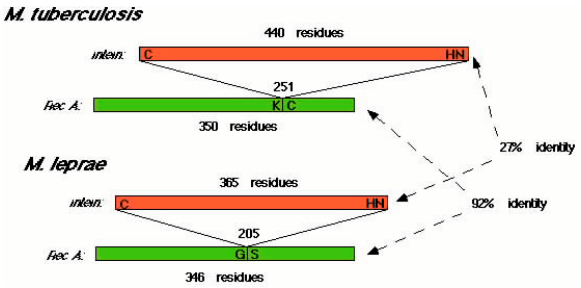

Figure 3. Comparison of the RecA proteins and their inteins of *M. tuberculosis and M. leprae*. The RecA proteins (green bars) are almost identical, whereas the inteins (red bars) are strikingly different with respect to both size and sequence and are inserted at different sites in the RecA proteins (indicated by the numbers). The amino acid residues at the splice junctions are also not conserved and are indicated in single-letter notation.

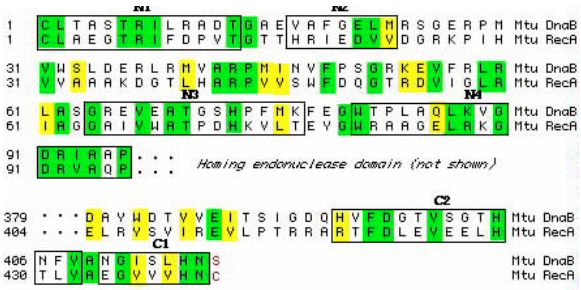

Figure 4. Alignment of the protein splicing domains of the *M. tuberculosis* DnaB and RecA inteins. The conserved protein splicing motifs (54) are shown by boxes and identical residues are highlighted in green, similar residues in yellow. The nucleophilic C-extein residue adjacent to the intein C-terminus is shown in red.

the *M. leprae* RecA intein are also present in six non-pathogenic species (17).

Analysis of the sequenced genomes of *M. tuberculosis* and *M. leprae* reveals inteins in three additional genes, *dnaB*, *ppsI*, and *gyrA*, which encode the DnaB helicase, a member of the YC24 family of ABC transporters, and subunit A of DNA gyrase, respectively (6). Interestingly, as with the RecA intein, the *M. tuberculosis* and *M. leprae* DnaB and PpsI inteins are not allelic, and the GyrA protein of *M. tuberculosis* is not interrupted by an intein. As summarized in Table 1 (Table 1), RecA, DnaB, PpsI, and GyrA inteins have also been identified in other *Mycobacterium* species (6). However, with only a few exceptions, these inteins are allelic to the those of *M. leprae* rather than to the *M. tuberculosis* inteins. The exceptions are the RecA and DnaB inteins of the closely related members of the *M. tuberculosis* complex, *M. bovis* and *M. microti*, as well as one of the two DnaB inteins found in *Mycobacterium smegmatis*. These differences suggest an interesting divergence between the *M. tuberculosis* complex and other *Mycobacterium* species.

Comparison of the protein splicing domains of the *M. tuberculosis* RecA and DnaB inteins, in contrast to their homing endonuclease domains, reveals 29% amino acid identity and 45% similarity, not only in the conserved protein splicing motifs but in other regions, suggesting a close evolutionary relationship (Figure 4). The high degree of similarity of these inteins and the observation that the 3-dimensional structures of even relatively unrelated inteins are almost the same make it quite likely that substances can be found which are inhibitors of protein splicing catalyzed by both of these inteins. On the other hand, the RecA and DnaB inteins use different C-terminal nucleophiles, a thiol and a hydroxyl, respectively, in the transesterification reaction leading to the branched ester intermediates and may therefore employ different mechanisms for nucleophile activation. Certain mechanism-based inhibitors may therefore inhibit one or the other of these inteins selectively.

5. THE ROLE OF RECA AND DNAB INTEINS IN *MYCOBACTERIUM TUBERCULOSIS*

5.1. RecA

As in enteric bacteria, the synthesis of the RecA protein is inducible by DNA damage and RecA initiates the SOS response in *M. tuberculosis* by forming filaments at the sites of DNA damage and activating the cleavage of LexA (18). However, a major difference between *M. tuberculosis* and other bacteria in the induction of the SOS response is that the RecA protein is interrupted by an intein and has to undergo protein splicing before it can assume its role in the initiation of DNA repair. This conclusion was confirmed by a detailed study of the unspliced *M. tuberculosis* RecA precursor protein, which demonstrated the absence of all functions associated with mature RecA protein, including binding to single-stranded DNA either in the absence or presence of nucleotide, DNA-dependent or independent ATPase activities, or the ability to bind 8-azido-ATP as measured by photo-crosslinking (19). Although the transcription of *recA* in *M. tuberculosis* is repressed by the binding of LexA to an SOS box and is inducible by DNA damage as in other bacteria, an unusual long lag period is seen in the appearance of RecA after the induction of DNA damage (20). Western blot analysis of the appearance of RecA protein upon DNA damage revealed only mature RecA protein and no unspliced precursor was found either before DNA damage or in the lag period after addition of DNA damaging agents (20), suggesting that protein splicing does not play a rate-limiting role in RecA expression. Splicing of the *M. tuberculosis* RecA intein has also been demonstrated in heterologous systems. Expression of the *M. tuberculosis recA* locus in *E. coli* can partially complement *E. coli recA* mutations (2) and leads to spliced RecA protein (3). The *M. tuberculosis recA* locus can also restore RecA function, measured both in terms of UV-resistance and homologous recombination, to *recA* deletion strains of *Mycobacterium smegmatis* (21, 22). The observation that splicing of *M. tuberculosis* RecA can occur in other bacterial species indicating that it does not depend on the presence of *M. tuberculosis*-specific factors. On the other hand, the possibility that RecA splicing in *M. tuberculosis* is controlled by an inhibitory factor that is inactivated by DNA damaging agents has not been ruled out.

Table 1. Distribution of inteins in *Mycobacterium* species

Organism	Extein	Intein Name	Prototype Allele
M. avium	DnaB helicase	Mav DnaB	Mle DnaB
M. bovis	DnaB helicase	Mbo DnaB	Mtu DnaB
M. bovis	RecA	Mbo RecA	Mtu RecA
M. chitae	RecA	Mch RecA	Mle RecA
M. fallax	RecA	Mfa RecA	Mle RecA
M. flavescens	DNA gyrase subunit A	Mfl GyrA	Mle GyrA
M. flavescens	RecA	Mfl RecA	Mle RecA
M. gastri	DNA gyrase subunit A	Mga GyrA	Mle GyrA
M. gastri	Pps1, YC24 family of ABC transporters	Mga Pps1	Mga Pps1
M. gastri	RecA	Mga RecA	Mle RecA
M. gordonae	DNA gyrase subunit A	Mgo GyrA	Mle GyrA
M. intracellulare	DnaB helicase	Min DnaB	Mle DnaB
M. kansasii	DNA gyrase subunit A	Mka GyrA	Mle GyrA
M. leprae	DnaB helicase	Mle DnaB	Mle DnaB
M. leprae	DNA gyrase subunit A	Mle GyrA	Mle GyrA
M. leprae	Pps1, YC24 family of ABC transporters	Mle Pps1	Mle Pps1
M. leprae	RecA	Mle RecA	Mle RecA
M. malmoense	DNA gyrase subunit A	Mma GyrA	Mle GyrA
M. microti	RecA	Mmi RecA	Mtu RecA
M. shimodei	RecA	Msh RecA	Mle RecA
M. smegmatis	DnaB helicase	Msm DnaB-1	Mle DnaB
M. smegmatis	DnaB helicase	Msm DnaB-2	Ssp DnaB
M. thermoresistibile	Rec A	Mth RecA	Mle RecA
M. tuberculosis	Pps1, YC24 family of ABC transporters	Mtu Pps1	Mtu Pps1
M. tuberculosis	DnaB helicase	Mtu DnaB	Ssp DnaB
M. tuberculosis	RecA	Mtu RecA	Mtu RecA
M. xenopi	DNA gyrase subunit A	Mxe GyrA	Mle GyrA

The intein names include only the first two letters of the name of the *Mycobacterium* species from which they are derived.

In most bacterial species, RecA is not essential for growth except under conditions of excessive DNA damage. However, upon infection of an animal host, pathogenic bacteria are exposed to reactive oxygen and nitrogen species produced during the respiratory burst of macrophages and neutrophils, and detoxification of oxidative radicals or repair of DNA damage must play an important role in bacterial virulence. In *Salmonella typhimurium*, deleterious mutations in *recA* have little effect on *in vitro* growth but significantly attenuate virulence in susceptible mice (23). Indeed, comparison of *S. typhimurium* strains with defects in catalase and RecA suggested that DNA repair is more important to virulence than detoxification of reactive oxygen species (24). *recA* was also found to play an important role in the virulence of enterohemorrhagic *E. coli* O157:H7 (25). On the other hand, *recA* mutations had no effect on the virulence of the intracellular pathogen, *Corynebacterium pseudotuberculosis*, in a mouse model (26).

Recent studies using high density mutagenesis by transposon insertion into the *M. tuberculosis* genome showed that the *recA* locus is not essential for growth on agar plates but that its disruption does significantly attenuate *in vitro* growth rates (27). Because *M. tuberculosis* is an intracellular pathogen, one would expect that the repair of DNA damage inflicted by reactive oxygen and nitrogen species produced by macrophages may play an important role in its survival in the infected host. However, although *recA* mutants of *Mycobacterium bovis* BCG, an otherwise non-virulent mycobacterial strain, are highly susceptible to killing by DNA damaging agents, they are not compromised in terms of survival in a nude mouse model (28). On the other hand, when studied in an *in vitro* model for persistence, *recA* mutants of *M. bovis* BCG had a greatly enhanced sensitivity to metronidazole (28). This interesting observation suggests that inhibitors of RecA splicing might act synergistically with metronidazole or similar drugs in eliminating dormant *M. tuberculosis* that persist in granulomas of antibiotic-treated tuberculosis patients. At this time, it is not clear whether extrapolation from non-virulent *M. bovis* BCG strains to virulent strains of *M. tuberculosis* is justified and direct studies on *recA* mutants of *M. tuberculosis* in an infection model are needed to determine whether the splicing of RecA plays an significant role in *M. tuberculosis* pathogenesis.

5.2. DnaB

DnaB is a DNA helicase that functions in the initiation of lagging-strand DNA synthesis (29) as a component of the primosome (30). Its essential role in DNA replication in *E. coli* is indicated by the existence of temperature-sensitive *dnaB* mutants (31). High density transposon mutagenesis studies in *M. tuberculosis* have shown that *dnaB* is essential for *in vitro* growth (27). The functional form of the DnaB protein is a hexamer which interacts directly with 2-3 molecules of the DnaG primase (32). In the 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 likely that DnaB can function only after excision of the 415-residue intein from the 465-residue helicase and that inhibition of DnaB splicing should therefore inhibit bacterial growth.

6. SELECTION OR SCREENING SYSTEMS FOR THE INHIBITION OF PROTEIN SPLICING

Inteins often interrupt essential proteins of microorganisms, such as DNA polymerases, DNA helicases and DNA gyrases, and, in these cases, protein splicing is essential for growth. However, some intein-containing proteins are required only conditionally, such as the vacuolar H^+-ATPase of yeast, which is essential during growth in the presence of $CaCl_2$ but not in its absence. This

makes it possible to screen for mutants that are defective in the splicing of the intein that interrupts the VMA subunit in terms of the ability to grow in the absence but not in the presence of 0.1 M CaCl2 (33) and to select for suppressors of these mutants that restore protein splicing capacity (34). It is also possible to develop artificial selection systems for protein splicing by inserting inteins into proteins that can provide a selective advantage. For example, the *M. tuberculosis* RecA intein, together with a C-terminal cysteine residue, has been inserted adjacent to residue 119 of aminoglycoside phosphotransferase, the enzyme responsible for plasmid-borne kanamycin resistance. Plasmids carrying the gene for such a fusion protein were found to convey kanamycin resistance to *E. coli* or *Mycobacterium smegmatis* under conditions when protein splicing occurs (35). This selection system has been used as an "ORFTRAP" to select for open reading frames in genomic libraries that can be inserted into a non-essential region of the RecA intein without interfering with protein splicing by disrupting the translation of the intein by frame-shifts or termination codons.

These assay systems could potentially lend themselves to the detection of protein splicing inhibitors. However, a disadvantage of such *in vivo* systems for screening antibacterial drugs against a specific target is that the screening requires comparison of growth under permissive and non-permissive conditions in order to eliminate antibacterial agents directed against other targets.

6.1. *In vivo* screening system based on the splicing of bacterial thymidylate synthase

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 DHF that is produced in the course of thymidylate synthesis cannot be recycled to THF, which is needed for other methylation reactions. Belfort and coworkers have taken 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 (36). 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 (37). Selection for growth in minimal medium in the absence of thymine has been used to define the minimum size of the protein splicing domain (36) and to select for inteins with enhanced splicing activity under specific conditions (38). Conversely, the ability to grow in minimal media supplemented with both thymine and trimethoprim should constitute an effective screen for protein splicing inhibitors (37).

6.2. *In vivo* screening system based on the splicing of bacterial DNA gyrase subunit A

DNA gyrase is a type II topoisomerase which promotes the supercoiling of DNA by causing a double-strand break, passing the DNA strand through the gap, and resealing the break. Quinolone antibiotics, such as ofloxacin, which specifically bind to DNA gyrase subunit A (GyrA), leads to 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 coworkers described an *in vivo* assay for the inhibition of protein splicing mediated by the *Mycobacterium xenopi* GyrA intein, based on the conditional dominant lethality of quinolone-sensitive GyrA expressed in a quinolone-resistant *E. coli* host (39). The GyrA intein of *M. xenopi* was introduced into the corresponding site of *E. coli gyrA*, cloned adjacent to an IPTG-inducible promoter into a low-copy-number vector, and 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 use of the fluorescent indicator allows measurement of growth rates, making it possible to screen for inhibitors that inhibit growth only partially. The occurrence of GyrA splicing was demonstrated by Western blot analysis and growth was inhibited by low concentrations of ofloxacin, confirming that spliced wild-type GyrA can induce quinolone-dependent cell death. The suitability of this experimental system for selection against protein splicing was verified by using it for the selection of temperature-sensitive protein splicing mutants (39). Its robustness was demonstrated by showing that the GyrA intein from *M. xenopi* could be replaced by the GyrA inteins from for other *Mycobacterium* species with similar results, in spite of the fact that their size varied from 198 to 420 amino acid residues. This suggest that this selection system can be adapted to assay for the inhibition of other inteins, including those that occur in *M. tuberculosis*, by inserting these into the *E. coli* GyrA protein in place of the GyrA intein.

6.3. *In vivo* screening system based on the splicing of the bacterial CcdB toxin

The toxic protein, CcdB, is the product of the *ccdAB* addiction module that serves to prevent the loss of F-factor from the *E. coli* host (for a recent review, see 40). The fact that the toxicity of CcdB is not abrogated by fusion of a 120-residue LacZalpha fragment to its N-terminus, has led to the development of positive selection vectors for recombinant DNA (41, 42). These selection vectors, which consist of multicopy plasmids carrying an IPTG-inducible *lacZalpha-ccdB* fusion, are lethal to the host in the presence of IPTG unless the *lacZalpha-ccdB* reading frame is disrupted by the insertion of a DNA segment into the *lacZalpha* polylinker or extended by more than 50 codons. Accordingly, the in-frame insertion of the 440-amino acid *M. tuberculosis* RecA intein into the *lacZalpha* polylinker will attenuate IPTG-dependent toxicity unless the intein is excised from the LacZalpha-CcdB fusion by protein splicing. This conditional phenotype of *E. coli* hosts transformed with LacZalpha-intein-CcdB vectors makes it possible to select against protein splicing in the presence of IPTG (43). Depending

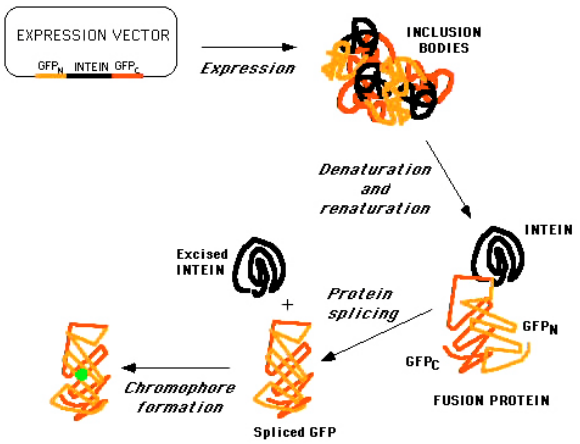

Figure 5. Schematic representation of the expression, refolding, splicing, and GFP chromophore formation of a GFP-intein fusion protein generated by the in-frame insertion of an intein coding sequence into a plasmid-encoded GFP gene.

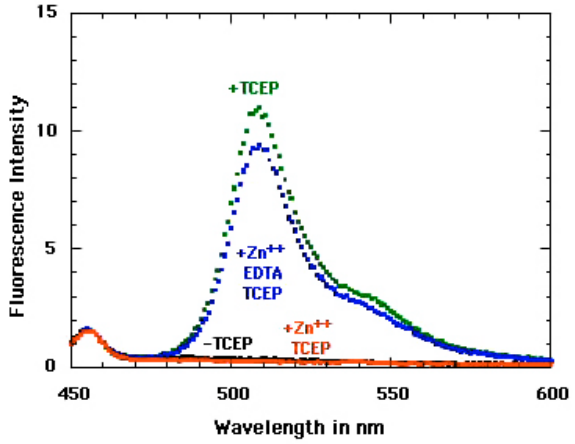

Figure 6. Fluorescent assay for protein splicing based on GFP. Renatured inclusion bodies of a GFP-intein fusion protein were incubated in the presence of the thiol reducing agent TCEP either alone (blue dots), in the presence of 2 mM $ZnCl_2$ (red dots), or in the presence of 2 mM $ZnCl_2$ and subsequent treatment with 4 mM EDTA (blue dots), illustrating the inhibition of protein splicing by zinc ion and its reversal by chelators. The black dots are the results of a control experiment without TCEP.

on the objective of the screening procedure, its stringency can be modulated by altering the level of expression of the LacZalpha-intein-CcdB fusion protein (43). In order to screen for protein splicing inhibitors that may attenuate protein splicing by less than an order of magnitude, it is advantageous to use a low-copy-number intein-CcdB plasmid so that the host cells can survive when splicing of the expressed CcdB fusion protein is only moderately suppressed.

6.4. *In vitro* screening systems based on Green Fluorescent Protein

An *in vitro* assay system for inhibitors of protein splicing was recently described. This assay 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 (44). Ozawa and coworkers 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 (44-46).

When the *M. tuberculosis* RecA intein was introduced at residue 129 of a modified GFP, the resulting fusion protein was produced as non-fluorescent inclusion bodies (47). Upon denaturation and renaturation of the GFP-intein inclusion bodies and subsequent incubation under conditions that compatible with protein splicing, green fluorescence appears in a time-dependent manner due to protein splicing followed by chromophore formation (47), as shown schematically in Figure 5 (Figure 5). GFP chromophore formation depended absolutely on protein splicing, failing to occur with a mutant intein in which a single essential asparagine residue was replaced with alanine. Refolding and subsequent protein splicing of the GFP-intein inclusion bodies was enhanced by the presence of 0.5 M L-arginine and required a thiol reducing agent such as tris(2-carboxyethylphosphine) (TCEP) or dithiothreitol, but even under optimal conditions (pH 6.5 and 20 °C), protein splicing was much slower than GFP chromophore formation, with half-times of 8 h and 1.7 h, respectively.

The high signal-to-noise ratio of this assay system is illustrated by the data in Figure 6 (Figure 6), with no detectable fluorescence in the absence of thiol reducing agent or in the presence of Zn^{2+} ion, an inhibitor of protein splicing (48, 49). It is thus possible to prevent premature splicing of the refolded fusion protein by omitting a thiol reducing agent or by inhibiting protein splicing with zinc ion. Addition of a molar excess of EDTA to the GFP-intein fusion protein renatured in the presence of $ZnCl_2$ restores 90% of its protein splicing activity (Figure 6). The renatured Zn^{2+}-complex is relatively stable, losing about 5% of its activity per hour. A robust *in vitro* screening system for protein splicing inhibitors can thus be constituted by dispensing renatured, Zn^{2+}-inhibited GFP-intein fusion protein into wells containing samples of library to be screened, followed by initiation of the protein splicing reaction by the addition of a molar excess of EDTA and measuring the rate of appearance of fluorescence (47). However, it should be noted that the appearance of fluorescence upon renaturation of the GFP-intein fusion protein involves both protein splicing and the chemical rearrangement of the spliced GFP to form the chromophore. Any substances that interferes with the formation of GFP fluorescence in this assay must therefore be re-screened in order to eliminate compounds that specifically inhibit GFP chromophore formation. Such a secondary screen involves essentially the same assay system, except that non-fluorescent inclusion bodies, produced when GFP is expressed from recombinant plasmids in *E. coli* at 37°C, are substituted for the denatured GFP-intein fusion protein (47).

Although the high-throughput screen described here is based on the RecA intein of *M. tuberculosis*, it

should be readily adaptable to other inteins that can be inserted into GFP at position 129 or other suitable insertion sites (45). Indeed, the intact *Saccharomyces cerevisiae* VMA1 intein has been inserted at position 129 and shown to be able to undergo protein splicing when expressed in *E. coli* at low temperatures (44), and similar results have recently been obtained with the full-length *M. tuberculosis* DnaB intein (S-q. Jiang & H. Paulus, unpublished observations).

7. PERSPECTIVES

The challenge of finding new antimycobacterial drugs to counter the threat posed by multi-drug resistant tuberculosis has led to the active exploration of new targets, usually proteins that have an essential role in growth or metabolism. Owing to the essential nature of such targets, similar proteins will also exist in many other organisms, which may also be affected by these drugs. Much effort in drug development will therefore have to be expended in optimizing inhibitory activity against the target organism while minimizing potential toxicity. From this point of view, inteins are ideal antimycobacterial targets. Protein splicing occurs only in microorganisms, and mycobacteria are the only intein-containing bacteria associated with human hosts. In *M. tuberculosis*, protein splicing is essential for the function of the RecA and DnaB proteins. The structure of RecA and DnaB proteins is highly conserved in nature, and if one wished to develop drugs that inhibit their function directly, one would have to be concerned with potential inhibition of other organisms. One the other hand, by targeting the inteins, one can disable mycobacterial RecA and DnaB with a drug against a target that occurs neither in the patients to be treated nor any other member of their microbial flora. Furthermore, because the RecA and DnaB inteins are structurally closely related, a major subset of inhibitors should affect both inteins. This class of inhibitors should have a more profound effect on growth or survival than inhibitors of RecA or DnaB alone. Moreover, the development of drug resistance, which would require mutations of two separate inteins, should be significantly slower.

It is not possible to make a priori predictions concerning the specificity of protein splicing inhibitors that may be isolated by the screening systems reviewed in this paper. Protein splicing inhibitors may be mechanism-based and inhibit one or several of the catalyzed steps in protein splicing, in which case they could be specific for certain groups of inteins. On the other hand, protein splicing inhibitors may interact with highly conserved structural motifs and thus have relatively broad specificity. Indeed, such structure-based inhibitors may also act on structurally related proteins such as the autoprocessing domain of the hedgehog family of proteins, which resemble inteins both in amino acid sequence (50) and three-dimensional structure (51). Hedgehog proteins play an important role in embryonic patterning of nearly all metazoans, including mammals, and interference with hedgehog signaling can have severe teratogenic effects (52, 53). It is therefore important that all protein splicing inhibitors that are considered as potential antimycobacterial drugs are first screened for possible inhibition of hedgehog autoprocessing.

It is of interest to compare the relative merits of the two types of assay systems that have been developed for screening for protein splicing inhibitors. The first involves the insertion of an intein into a protein whose expression is detrimental to survival under certain growth conditions. The ability to select for agents that prevent the conditional toxicity of the intein fusion protein is a very powerful screening procedure. It involves the measurement of cell growth which can be adapted to a high-throughput mode by incorporating a fluorescent indicator such as GFP into the bacteria whose growth is to be monitored (39). On the other hand, such *in vivo* screening systems suffer from the disadvantage that compounds that are impermeable to bacterial cells or inhibit bacterial growth nonspecifically will be eliminated from further consideration. Although in most cases such exclusion is justified, it may lead to the elimination of inhibitors whose undesirable properties could be mitigated in the course of lead optimization. The other assay system for protein splicing and its inhibition involves purified proteins. This *in vitro* system is unaffected by the ability of compounds to enter cells or by their cellular toxicity and will therefore provide a much broader spectrum of hits. Any hits, however, will have to be subjected to secondary screening in an *in vivo* assay systems to assess their usefulness as specific antibacterial agents.

No data have been published on the use of any of the assay systems described in this review for high throughput screening of large compound libraries. At this time, therefore, it is not possible to predict whether protein splicing inhibitors can indeed be found or how effective the inhibition of RecA and DnaB protein splicing will be in controlling tuberculosis.

8. REFERENCES

1. Espinal, M.E. & A.S. Laszlo, Lone. Boulahbal, Fadila. Kim, Sang Jae. Reniero, Ana. Hoffner, Sven. Rieder, Hans L.. Binkin, Nancy. Dye, Christopher. Williams, Rosamund. Raviglione, Mario C: Global trends in resistance to antituberculosis drugs. *N Engl J Med* 344, 1294-1303 (2001)

2. Davis, E.O., S.G. Sedgwick & M.J. Colston: Novel structure of the *recA* locus of *Mycobacterium tuberculosis* implies processing of the gene product. *J Bacteriol* 173, 5653-5662 (1991)

3. Davis, E.O., P.J. Jenner, P.C. Brooks, M.J. Colston & S.G. Sedgwick: Protein splicing in the maturation of *M. tuberculosis* RecA protein: A mechanism for tolerating a novel class of intervening sequence. *Cell* 71, 201-210 (1992)

4. Davis, E.O., H.S. Thangaraj, P.C. Brooks & M.J. Colston: Evidence for selection for protein introns in the RecAs of pathogenic mycobacteria. *EMBO J* 13, 699-703 (1994)

5. Colston, M.J.: The molecular basis of mycobacterial infection. *Mol Aspects Med* 17, 385-454 (1996)

6. Perler, F.B.: InBase, the intein database. *Nucl Acids Res* 30, 383-384 (2002)
7. Xu, M.-Q., M.W. Southworth, F.B. Mersha, L.J. Hornstra & F.B. Perler: In vitro protein splicing of purified precursor and the identification of a branched intermediate. *Cell* 75, 1371-1377 (1993)
8. Paulus, H.: The chemical basis of protein splicing. *Chem Soc Rev* 27, 375-386 (1998)
9. Paulus, H.: Protein splicing and related forms of protein autoprocessing. *Annu Rev Biochem* 69, 447-496 (2000)
10. Paulus, H.: Inteins as enzymes. *Bioorg Chem* 29, 119-129 (2001)
11. Duan, X., F.S. Gimble & F.A. Quiocho: Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. *Cell* 89, 555-564 (1997)
12. Poland, B.W., M.-Q. Xu & F.A. Quiocho: Structural insights into the protein splicing mechanism of PI-SceI. *J Biol Chem* 275, 16408-16413 (2000)
13. Mizutani, R., S. Nogami, M. Kawasaki, Y. Ohya, Y. Anraku & Y. Satow: Protein splicing reaction via a thiazolidine intermediate; crystal structure of the VMA1-derived endonuclease bearing the N- and C-terminal propeptides. *J Mol Biol* 316, 919-929 (2002)
14. Werner, E., W. Wende, A. Pingoud & U. Heinemann: High resolution crystal structure of domain I of the *Saccharomyces cerevisiae* homing endonuclease PI-*Sce*I. *Nucleic Acids Res* 30, 3962-3971 (2002)
15. Klabunde, T., S. Sharma, A. Telenti, W.R. Jacobs & J.C. Sacchettini: Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing. *Nature Struct Biol* 5, 31-36 (1998)
16. Ichiyanagi, K., Y. Ishino, M. Ariyoshi, K. Komori & K. Morikawa: Crystal structure of an archaeal intein-encoded homing endonuclease PI-PfuI. *J Mol Biol* 300, 889-901 (2000)
17. Saves, I., M.A. Laneelle, M. Daffe & J.M. Masson: Inteins invading mycobacterial RecA proteins. *FEBS Letters* 480, 221-225 (2000)
18. Movahedzadeh, F., M.J. Colston & E.O. Davis: Determination of DNA sequences required for regulated *Mycobacterium tuberculosis* RecA expression in response to DNA-damaging agents suggests that two modes of regulation exist. *J Bacteriol* 179, 3509-3518 (1997)
19. Kumar, R.A., M.B. Vaze, N.R. Chandra, M. Vijayan & K. Muniyappa: Functional chracterization of the precursor and spliced forms of recA protein of *Mycobacterium tuberculosis*. *Biochemistry* 35, 1793-1802 (1996)
20. Papavinasasundaram, K.G., C. Anderson, P.C. Brooks, N.A. Thomas, F. Movahedzadeh, P.J. Jenner, M.J. Colston & E.O. Davis: Slow induction of RecA by DNA damage in *Mycobacterium tuberculosis*. *Microbiology* 147, 3271-3279 (2001)
21. Papavinasasundaram, K.P., M.J. Colston & E.O. Davis: Construction and complementation of a *recA* deletion mutant of *Mycobacterium smegmatis* reveals that the intein in *Mycobacterium tuberculosis recA* does not affect RecA function. *Mol Microbiol* 30, 525-534 (1998)
22. Frischkorn, K., P. Sander, M. Scholz, K. Teschner, T. Prammananan & E.C. Bottger: Investigation of mycobacterial *recA* function: protein introns in the RecA of pathogenic mycobacteria do not affect competency for homologous recombination. *Mol Microbiol* 29, 1203-1214 (1998)
23. Buchmeier, N.A., C.L. Lipps, M.Y.H. So & F. Heffron: Recombination-deficient mutants of *Salmonella typhimurium* are avirulent and sensitive to the oxidative burst of macrophages. *Mol Microbiol* 7, 933-936 (1993)
24. Buchmeier, N.A., S.J. Libby, Y. Xu, P.C. Loewen, J. Switala, D.G. Guiney & F.C. Fang: DNA repair is more important than catalase for *Salmonella* virulence in mice. *J Clin Invest* 95 (1995)
25. Fuchs, S., I. Muhldorfer, A. Donohue-Rolfe, M. Kerenyi, L. Emody, R. Alexiev, P. Nenkov & J. Hacker: Influence of RecA on in vivo virulence and Shiga toxin 2 production in Escherichia coli pathogens. *J Microb Pathog* 27, 13-23 (1999)
26. Pogson, C.A., C.P. Simmons, R.A. Strugnell & A.L. Hodgson: Cloning and manipulation of the Corynebacterium pseudotuberculosis recA gene for live vaccine development. *FEMS Microbiol Lett* 142, 139-145 (1996)
27. Sassetti, C.M., D.H. Boyd & E.J. Rubin: Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* 48, 77-84 (2003)
28. Sander, P., K.G. Papavinasundaram, T. Dick, E. Stavropoulos, K. Ellrott, B. Springer, M.J. Colston & E.C. Bottker: *Mycobacterium bovis* BCG *recA* deletion mutant shows increased susceptibility to DNA-damaging agetns but wild-type survival in a mouse infection model. *Inf Immun* 69, 3562-3568 (2001)
29. Arai, K.I. & A. Kornberg: A general priming system employing only *dnaB* protein and primase for DNA replication. *Proc Natl Acad Sci USA* 76, 4308-4312 (1979)
30. Saluja, D. & G.N. Godson: Biochemical characterization of *Escherichia coli* temperature-sensitive *dnaB* mutants *dnaB8, dnaB252,dnaB70, dnaB43,* and *dnaB454*. *J Bacteriol* 177, 1104-1111 (1995)
31. Kogama, T.: Tw types of temperature sensitivity in DNA replication of an *Escherichia coli dnaB* mutant. *J Mol Biol* 103, 191-197 (1976)
32. Bird, L.E., H. Pan, P. Soultanas & D.B. Wigley: Mapping protein-protein interactions within a stable complex of DNA primase and DnaB helicase from *Bacillus stearothermophilus*. *Biochemistry* 39, 171-182 (2000)
33. Kawasaki, M., S. Nogami, Y. Satow, Y. Ohya & Y. Anraku: Identification of three core regions essential for protein splicing of the yeast Vma1 Protozyme. A random mutagenesis study of the entire Vma1-derived endonuclease sequence. *J Biol Chem* 272, 15668-15674 (1997)
34. Nogami, S., Y. Satow, Y. Ohya & Y. Anraku: Probing novel elements for protein splicing in the yeast VmaI protozyme: a study of replacement mutagenesis and intragenic suppression. *Genetics* 147, 73-85 (1997)
35. Daugelat, S. & W.R. Jacobs: The *Mycobacterium tuberculosis recA* intein can be used in an ORFTRAP to select for open reading frames. *Prot Sci* 8, 644-653 (1999)
36. Derbyshire, V., D.W. Wood, W. Wu, J.T. Dansereau, L.Z. Dalgaard & M. Belfort: Genetic definition of a protein-splicing domain: Functional mini-inteins support structure predictions and a model for intein evolution. *Proc Natl Acad Sci USA* 94, 11466-11471 (1997)

37. Belfort, M.: Inteins as antimicrobial targets: genetic screen for intein function. *US Patent 5,798,731* (1998)
38. Wood, D.W., W. Wu, G. Belfort, V. Derbyshire & M. Belfort: A genetic system yields self-cleaving inteins for bioseparations. *Nature Biotech* 17, 889-892 (1999)
39. Adam, E. & F.B. Perler: Development of a positive genetic selection system for inhibition of protein splicing using mycobacterial inteins in *Escherichia coli* gyrase subunit A. *J Molec Microbiol Biotech* 4, 479-487 (2002)
40. Engelberg-Kulka, H. & G. Glaser: Addiction modules and programmed cell death and antideath in bacterial cultures. *Annu Rev Microbiol* 53, 43-70 (1999)
41. Bernard, P., P. Gabarit, E.M. Bahassi & M. Couturier: Positive-selection vectors using the F plasmid *ccdB* killer gene. *Gene* 148, 71-74 (1994)
42. Bernard, P.: Positive selection of recombinant DNA by CcdB. *Biotechniques* 21, 320-323 (1996)
43. Lew, B.M. & H. Paulus: 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 (2002)
44. Ozawa, T., S. Nogami, M. Sato, Y. Ohya & Y. Umezawa: A fluorescent indicator for detecting protein-protein interactions in vivo based on protein splicing. *Anal Chem* 72, 5151-5157 (2000)
45. Ozawa, T., M. Takeuchi, A. Kaihara, M. Sato & Y. Umezawa: Protein splicing-based reconstitutition of split green fluorescent protein for monitoring protein-protein interactions in bacteria: Improved sensitivity and reduced screening time. *Anal Chem* 73, 5866-5874 (2001)
46. Ozawa, T. & Y. Umezawa: Detection of protein-protein interactions *in vivo* based on protein splicing. *Curr Opin Chem Biol* 5, 578-583 (2001)
47. Gangopadhyay, J.P., S.-q. Jiang & H. Paulus: An in itro screening system for protein splicing inhibitors based on Green Fluorescent Protein as an indicator. *Anal Chem* 75, in press (2003)
48. Mills, K.V. & H. Paulus: Reversible inhibition of protein splicing by zinc ion. *J Biol Chem* 241, 10832-10838 (2001)
49. Ghosh, I., L. Sun & M.Q. Xu: Zinc inhibition of protein *trans*-splicing and identification of regions essential for splicing and association of a split intein. *J Biol Chem* 276, 24051-24058 (2001)
50. Koonin, E.V.: A protein splice-junction motif in hedgehog family proteins. *Trends Biochem Sci* 20, 141-142 (1995)
51. Hall, T.M.T., J.A. Porter, K.E. Young, E.V. Koonin, P.E. Beachy & D.J. Leahy: Crystal structure of a hedgehog protein autoprocessing domain: homology between hedgehog and self-splicing proteins. *Cell* 91, 85-97 (1997)
52. Chiang, C., Y. Litingtung, E. Lee, K.E. Young, J.L. Corden, H. Westphal & P.A. Beachy: Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* 383, 407-413 (1996)
53. Roessler, E., E. Belloni, K. Gaudenz, P. Jay, P. Berta, S.W. Scherer, L.C. Tsui & M. Muenke: Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. *Nat Genet* 14, 357-360 (1996)
54. Pietrokovski, S.: Modular organization of inteins and C-terminal autocatalytic domains. *Protein Science* 7, 64-71 (1998)

Abbreviations: GFP, Green Fluorescent Protein; TCEP, tris (2-carboxyethylphosphine)

Key Words: Inteins, Mycobacterium Tuberculosis, Antimycobacterial Drugs, Protein Splicing, RecA, DnaB, Thymidylate Synthase, DNA gyrase, CcdB toxin, Green Fluorescent Protein, Drug Screening, Drug Targets, High Throughput Screening, Review

Send correspondence to: Dr. Henry Paulus, Boston Biomedical Research Institute, 64 Grove Street, Watertwon, MA 02472-2829, USA, Tel: 617-658-7800, Fax: 617-972-1753, E-mail: paulus@bbri.org