Protein Switches Identified From Diverse Insertion Libraries Created Using S1 Nuclease Digestion of Supercoiled-Form Plasmid DNA

Jennifer Tullman, Gurkan Guntas, Matthew Dumont, Marc Ostermeier

Department of Chemical and Biomolecular Engineering, Johns Hopkins University, 3400N. Charles St., Baltimore, Maryland 21218, Telephone: +1-410-516-7144; Fax: +1-410-516-5510; e-mail: oster@jhu.edu

ABSTRACT: We demonstrate that S1 nuclease converts supercoiled plasmid DNA to unit-length, linear dsDNA through the creation of a single, double-stranded break in a plasmid molecule. These double-stranded breaks occur not only in the origin of replication near inverted repeats but also at a wide variety of locations throughout the plasmid. S1 nuclease exhibits this activity under conditions typically employed for the nuclease's single-stranded nuclease activity. Thus, S1 nuclease digestion of plasmid DNA, unlike analogous digestion with DNaseI, effectively halts after the first double-stranded break. This property makes easier the construction of large domain insertion libraries in which the goal is to insert linear DNA at a variety of locations throughout a plasmid. We used this property to create a library in which a circularly permuted TEM1 β-lactamase gene was inserted throughout a plasmid containing the gene encoding *Escherichia coli* ribose binding protein. Gene fusions that encode allosteric switch proteins in which ribose modulates β-lactamase catalytic activity were isolated from this library using a combination of a genetic selection and a screen.

Biotechnol. Bioeng. 2011;108: 2535–2543.

© 2011 Wiley Periodicals, Inc.

KEYWORDS: S1 nuclease; protein engineering; directed evolution; protein switch

Introduction

The fusion of two genes into a single gene has been widely used as a tool in protein engineering and protein

purification. Most gene fusions are end-to-end. An alternate topology is to place an "insert" gene into an "acceptor" gene in a process called domain insertion. Domain insertion has been used to couple two proteins' functions to create protein switches (Ostermeier, 2005). For example insertion of β-lactamase (BLA) into maltose binding protein (MBP) can create switch proteins in which β-lactamase activity is modulated by maltose (Guntas and Ostermeier, 2004; Guntas et al., 2004, 2005). However, the exact fusion that will possess switching activity is difficult to predict. One approach for the identification of useful fusions is to create combinatorial libraries in which the insert gene is randomly inserted into the acceptor gene (Edwards et al., 2008; Guntas and Ostermeier, 2004). The diversity-generating step in this process is the creation of a random double-stranded break in the plasmid-borne acceptor gene. Such a break can be created by DNaseI digestion (Guntas and Ostermeier, 2004) or by use of transposons (Edwards et al., 2008), but each method has its drawbacks.

Dilute concentrations of DNaseI can be used to create random, double-stranded breaks in plasmid DNA in order to create random insertion libraries (Biondi et al., 1998; Guntas and Ostermeier, 2004) and random circular permutation libraries (Graf and Schachman, 1996). DNaseI digestion can create a distribution of deletions and tandem duplications in the acceptor sequence (Guntas and Ostermeier, 2004)—a diversity that is likely to be relevant for switching activity. However, the DNaseI digestion step in this process is very sensitive to the temperature, the DNaseI/DNA ratio, and the digestion time. The delicate balance between enough DNaseI activity to produce the desired double-stranded breaks but not too much to cause large deletions at the insertion site is difficult to achieve. Typically, appropriate digestion conditions have to be determined empirically for each preparation of plasmid DNA. In addition, large quantities of starting plasmid DNA are necessary in order to recover enough DNaseI-digested DNA to create the library. Even under the

best conditions a significant fraction of large deletions are unavoidable and the creation of high-quality, large, random insertion libraries ($>10^6$ members) is difficult. The MuDel transposon method (Edwards et al., 2008) for creating random insertion libraries is limited to the creation of libraries in which exactly three nucleotides are in the acceptor sequence at the insertion site. In addition, the library size that has been demonstrated using the MuDel transposon method is limited. A more robust method for preparing random insertion libraries would facilitate their construction.

S1 nuclease is a single-strand specific endonuclease typically used for blunting the ends of dsDNA (Ghangas and Wu, 1975) or for mapping RNA transcripts (Berk, 1989; Desai and Shankar, 2003; Gite and Shankar, 1995; Weaver and Weissmann, 1979). Factors such as ionic strength, pH, and temperature affect the specificity and activity of this nuclease (Vogt, 1973). Certain reaction conditions, including those typically used for the digestion of ssDNA, result in S1 nuclease's linearization of supercoiled dsDNA (Beard et al., 1973; Germond et al., 1974; Gonikberg, 1979; Mechali et al., 1973; Panayotatos and Wells, 1981; Shishido et al., 1983; Wiegand et al., 1975). Although it has been hypothesized that S1 nuclease digests supercoiled plasmid DNA in AT rich regions (Germond et al., 1974; Wiegand et al., 1975), as has been shown with Mung Bean nuclease (Kowalski, 1984), subsequent studies demonstrated that S1 nuclease digests supercoiled plasmids primarily in regions containing inverted repeats (Lilley, 1980; Panayotatos and Wells, 1981). In these regions, which generally occur in origins of replication, the DNA adopts a cruciform structure that S1 nuclease recognizes as single-stranded and cleaves (Kato et al., 2003; Kurahashi et al., 2004). However, evidence suggests that the identity of the divalent cations cofactor for the enzyme affects the location of digestion. S1 nuclease digestion of supercoiled Simian Virus 40 DNA created a much more diverse set of reaction products when Zn^{+2} was used instead of Mn^{+2} (Shishido, 1979), but this activity has not been characterized in detail. S1 nuclease's ability to cut supercoiled DNA has been used to facilitate the cloning of the plasmid-like cucumber mitochondrial DNA (Bai et al., 2003).

We wondered if S1 nuclease could be used as an alternative to DNaseI in the creation of random insertion libraries. Ideally, S1 nuclease digestion of supercoiled plasmid would proceed under conditions typically used for removing single-stranded overhangs from linear, double-stranded DNA (e.g., high ionic strength and low temperature). Thus, the digestion would advantageously stop after the first double stranded break due to linear dsDNA's resistance to digestion by S1 nuclease under these conditions. The extent to which S1 nuclease could effectively digest supercoiled plasmid at regions other than inverted repeats was unknown, but experiments in which the use of Zn^{+2} instead of Mn^{+2} increased the number of digestion products (Shishido, 1979) encouraged us to examine this question. We reasoned that even if S1 digestion was heavily biased to occur at inverted repeats, these regions are typically in origins of replication. Insertion of long sequences of DNA within the origin of replication, which would occur in random insertion library construction, is unlikely to create viable plasmids. Thus, if S1 nuclease sufficiently created double-stranded breaks at a low level throughout the plasmid, these would be greatly enriched for by the transformation step. In previous studies that examined the sites of digestion of supercoiled plasmid DNA by S1 nuclease the DNA was analyzed right after digestion (e.g., by agarose gel electrophoresis). A low-level of digestion at sites spread out elsewhere on the plasmid, if they occurred, would not have been readily detected due to the predominance of digestion at inverted repeats.

We report here that S1 nuclease creates double-stranded breaks in supercoiled plasmid DNA at a wide variety of locations. Owing to this ability, digestion using S1 nuclease is a convenient alternative to DNaseI for the creation of random insertion libraries. These libraries are sufficiently diverse to be used in the creation of protein switches. A library in which a circularly permuted TEM1 β-lactamase gene was randomly inserted into a plasmid containing the gene encoding *Escherichia coli* ribose binding protein was created using S1 nuclease. From this library we have selected several distinct gene fusions that encode allosteric protein switches in which ribose modulates the β-lactamase catalytic activity of the fusion.

Materials and Methods

S1 Nuclease and DNaseI Cleavage of Supercoiled dsDNA

The plasmid pRH04.152–rbsb is derived from the plasmid pDIMC8–rbsb with the f1 origin removed. Plasmid pRH04.152–pfMBP is the same plasmid as pRH04.152–rbsb with the gene rbsb (encoding for the protein RBP) replaced with the gene encoding for MBP from *Pyrococcus furiosus*. Two micrograms of the plasmid pRH04.152–pfMBP, pDIMC8–rbsb, or pRH04.152–rsbsb were incubated in 50 mM sodium acetate, 280 mM NaCl, 4.5 mM $ZnSO_4$ at pH 4.5 (at 25°C) with S1 nuclease (Promega, Madison, WI) in a final volume of 25 μL at the indicated temperature (see Fig. 1A). The reaction was stopped by passing it over a DNA clean and concentrator column (Zymo Research Corp., Irvine, CA) and eluting in 20 μL of DI water. For DNaseI digestion, 25 μg of plasmid was digested with DNaseI (0.8 mU) for 8 min at 22°C in the presence of 50 mM Tris–HCl (pH 7.4), 10 mM $MnCl_2$, and 50 μg/mL BSA in a total volume of 0.5 mL. The reaction was quenched by the addition of EDTA to a concentration of 5 mM and the solution was desalted using the clean and concentrator column. Each sample was then separated by agarose gel electrophoresis on a 0.8% agarose, TAE gel. To analyze the site of S1 nuclease digestion, the linear DNA resulting from either S1 nuclease digestion or DNaseI digestion was
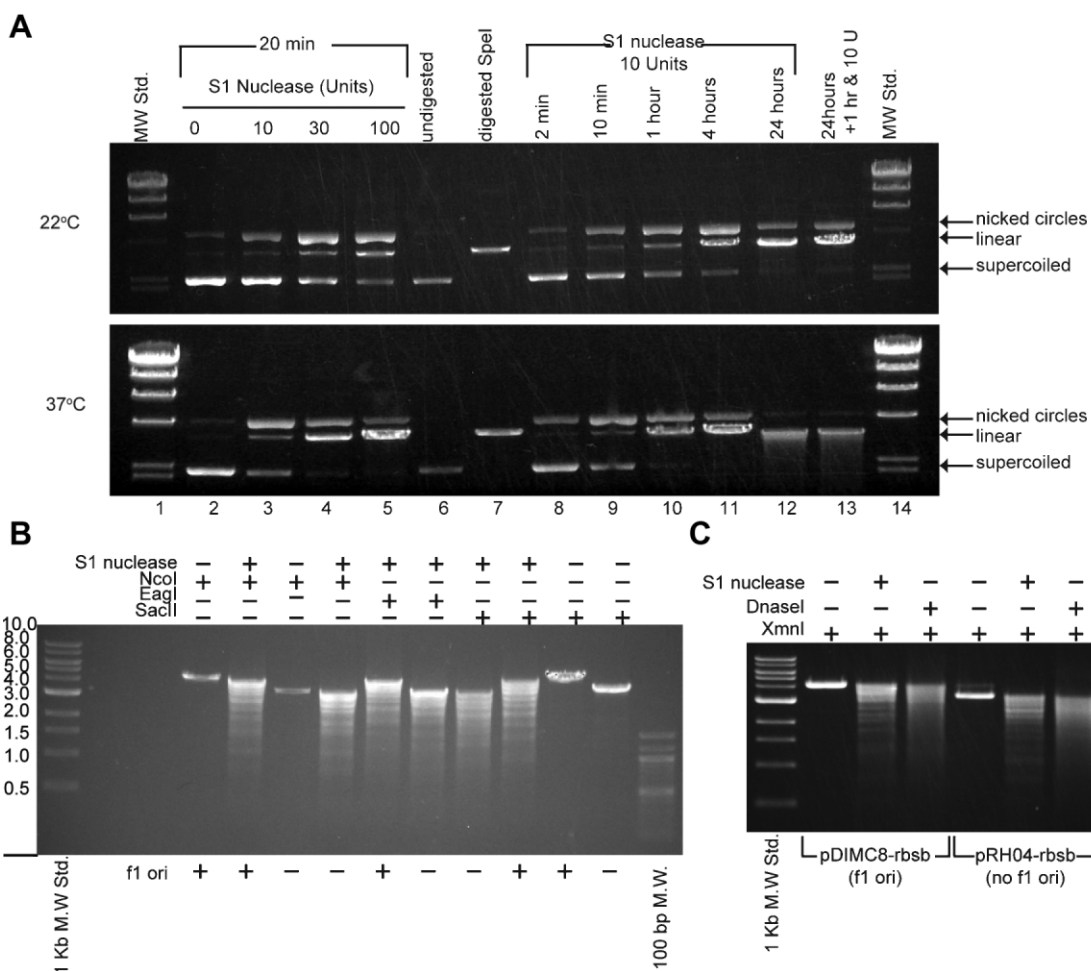
Figure 1. S1 nuclease converts supercoiled plasmid DNA to nicked circular and linear forms through digestion at a variety of locations in the plasmid. **A**: Two micrograms of supercoiled pRH04.152–rbsB was incubated at 22°C (top) or 37°C (bottom) with different amounts of S1 nuclease for different lengths of time and then analyzed by agarose gel electrophoresis. The sixth and seventh lanes contain undigested supercoiled control and SpeI-digested linear control, respectively. The band that runs slower than the linear band was presumed to correspond to nicked circular plasmid. In the 13th lane, after the DNA was digested with 10 U for 24 h an additional 10 U of S1 nuclease was added and the DNA was incubated for an additional hour. The first and last lanes contain the λDNA/HindIII molecular weight standards. **B**: Plasmids with or without the f1 origin (pDIMC8–pfMBP and pRH04–pfMBP, respectively) were incubated with or without S1 nuclease, followed by incubation with or without NcoI, EagI, and SacII and then analyzed by agarose gel electrophoresis. **C**: Plasmids with or without the f1 origin (pDIMC8–rbsb and pRH04–rbsb, respectively) were incubated with S1 nuclease or DNaseI, followed by incubation with XmnI and analyzed by agarose gel electrophoresis.

extracted from the gel using the Qiagen QIAquick gel extraction kit. This linear DNA was digested with restriction enzymes NcoI, EagI, SacI, or XmnI under the recommended conditions (New England Biolabs, Ipswich, MA) for 1 h at 37°C and analyzed by agarose gel electrophoresis on a 0.8% agarose, TAE gel.

Construction of Insertion Library

Plasmid pDIMC8–rbsB DNA was incubated at 37°C for 20 min as above in 25 separate reactions, each containing 2 μg of pDIMC8–rbsB and 10 U S1 Nuclease. The reaction was stopped through use of a Zymo Research Corp. clean and concentrator spin column and the 25 samples

subsequently pooled together. The amount of linear DNA created was quantified using the band intensity after gel electrophoresis. To blunt the ends and seal any nicks in the DNA, T4 DNA ligase, and T4 DNA polymerase were added in a ratio of 160 U (400,000 U/mL, New England Biolabs) and 1 U (300 U/mL, New England Biolabs) per 1 μg of linear DNA, in a reaction containing 1× T4 ligase buffer and 200 μM dNTPs. The mixture was incubated at 12°C for 20 min. The addition of 10 mM EDTA and heating to 75°C for 15 min stopped the reaction. The repaired, linear DNA was then isolated using agarose gel electrophoresis on a 0.8% agarose, lithium borate (Faster Better Media, Hunt Valley, MD) gel. We used the Qiagen gel extraction kit and obtained 6 μg of DNA. Five microliters of Antarctic phosphatase (New England Biolabs) was used to dephosphorylate the

DNA in Antarctic phosphatase buffer for 30 min at 37°C. Heating to 65°C for 10 min inactivated the phosphatase. Next, the product was ligated with the *bla[170–170]* insertion gene, which had been amplified by polymerase chain reaction from the vector pDIMC8–MBP317–347 with phosphorylated primers, 5'P-GCCATACCAAACGACGAG-3' and 5'P-GGCTTCATTCAGTTCCGG-3'. Each ligation reaction contained 1 µg of plasmid DNA with a 1.5:1 molar ratio of insert DNA, and 1 µL of T4 DNA ligase (1,000,000 U/mL), 1× ligase buffer, and PEG8000 to 5% final concentration. The mixture was incubated overnight at room temperature. The ligations were purified over a clean and concentrator column and eluted in 20 µL of water. The 20 µL ligation was centrifuged under vacuum to 2 µL final volume and transformed into 40 µL of DH10B cells (Invitrogen, Carlsbad, CA) by electroporation. After incubation in recovery media for 1 h at 37°C, the cells were plated on LB-agar plates containing 50 µg/mL chloramphenicol. Individual library members were picked and sequenced using the bla170seqfor and bla170seqrev primers (5'-CTTCTGACAACGATCGG-3' and 5'-CAGCAATAAACCAGCCAGC-3', respectively).

Identification and Preliminary Analysis of Switches

Library members were plated on LB-agar plates containing 128, 256, and 512 µg/mL ampicillin and incubated at 37°C overnight. The surviving colonies were screened in 96-well format for ribose-dependent β-lactamase activity using a colorimetric nitrocefin assay as described previously (Guntas et al., 2004). Clones with greater than a twofold ribose-dependent difference in initial velocity of nitrocefin hydrolysis (i.e., greater than twofold switching activity) were further examined. The minimum inhibitory concentration (MIC) of ampicillin in the presence and absence of ribose was determined for each putative switch in a similar fashion as previously described (Guntas et al., 2004) with the exception that 200 colony forming units were plated on plates containing LB-agar (lacking yeast extract), 300 µM IPTG, and 5 mM ribose.

Protein Purification and Kinetic Characterization

The protein switch p2.C5 was initially purified by exploiting the 10× C-terminal His-tag, and then further purified via FPLC. Cells harboring the plasmid were grown in 1.5 L of LB media supplemented with M9 salts at 37°C until the OD_{600} reached 0.6. IPTG was then added to a final concentration of 1 mM and the flasks transferred to 25°C for 16 h. The cells were harvested by at 6,400g at 4°C for 10 min. The cell pellet was then resuspended in 15 mL chilled 50 mM phosphate buffer with 100 mM NaCl, pH 7.3 with protease inhibitor cocktail (Sigma, P8849, St. Louis, MD). The cells were lysed using a French pressure cell press (ThermoSpectronic, Madison, WI) at 20,000 psi. The lysate was centrifuged twice for 30 min at 48,000g at 4°C before pouring over 1 mL of

Talon resin (Clontech, Mountain View, CA) in a gravity flow column. The flow-through was reapplied to the column once. The column was washed with 5 mL of phosphate buffer and then increasing amounts of imidazole in phosphate buffer (2 mL × 10 mM imidazole, 2 mL × 20 mM imidazole, 2 mL × 40 mM imidazole, 6 mL × 50 mM imidazole). The addition of 3 mL of 250 mM imidazole eluted the protein from the column. The elution fraction was then dialyzed into 50 mM Tris buffer, pH 7.5 using a 10 kDa molecular weight cut-off centrifugal filter device (Millipore, Billerica, MA). Dialysis was accomplished by pulling the imidazole buffer through the device using a swinging bucket centrifuge (Eppendorf 5804 R) at 4,500g for 10 min and then resuspending the protein in 10 mL of the Tris buffer. This resuspension/centrifugation process was repeated five times to fully exchange the buffer. Next, the sample was purified via FPLC (GE Healthcare AKTApurifier, Piscataway, NJ) using an anion-exchange column (Hiprep Q Ff 16/100) and eluted over a gradient with final concentration 50 mM Tris, 150 mM NaCl buffer, pH 7.5. The protein-containing fractions were visualized via SDS–PAGE gel electrophoresis and Coomassie blue staining. The protein was estimated to be >90% pure and its identity confirmed by analysis by western blot with an anti-BLA antibody (Millipore) and by assays for nitrocefin hydrolysis. The purified protein was dialyzed as above into 50 mM phosphate buffer (pH 7.0). The initial rate of hydrolysis of 100 µM nitrocefin by 20 nM protein was measured as previously described (Guntas et al., 2004) in 50 mM sodium phosphate buffer (pH 7.0) with and without 5 mM ribose.

Results

S1 Nuclease Digestion of Supercoiled Plasmid DNA

For the creation of random insertion libraries we desired S1 nuclease digestion to halt after a single double-stranded break in supercoiled plasmid DNA. Thus, we sought to characterize S1 nuclease digestion of supercoiled DNA under conditions that greatly favor cleavage of ssDNA over linear dsDNA. We used standard conditions for removal of ssDNA overhangs from linear dsDNA by S1 nuclease (i.e., high salt, low pH, and low temperature). These conditions should disfavor local unwinding of plasmid DNA at A + T-rich regions at neutral pH in low salt (Kowalski et al., 1988).

We tested S1 nuclease digestion of pRH04.152–pfMBP, a plasmid derived from pDIMC8 (Paschon and Ostermeier, 2004) designed for creating random insertion libraries. We found that 10 U of S1 nuclease could convert 2 µg of supercoiled DNA to a roughly equimolar mixture of linear and open circular DNA in 1 h at 37°C in a volume of 25 µL (Fig. 1A). This is approximately 10% of the amount of S1 nuclease previously reported to be necessary for this reaction (Bai et al., 2003). Agarose gel electrophoresis analysis of the reaction as a function of time was consistent with the expectation that S1 nuclease first converts the supercoiled

DNA to open circular form and then to linear form (Fig. 1A). Complete conversion to linear could be achieved in 24 h at the elevated temperature; however, smearing on the gel indicated that long incubations resulted in some undesirable further digestion of the dsDNA. Analogous S1 nuclease digestion of pDIMC8 and pUC19 with the same number of units of S1 nuclease, temperature, and time, produced equivalent amounts of linear form DNA of the expected size for each plasmid (data not shown).

We next examined the dsDNA produced by S1 nuclease digestion for the location of digestion. We first looked for any gross bias in cleavage locations in pRH04.152–pfMBP and pDIMC8–pfMBP. Plasmid pDIMC8–pfMBP is the parental plasmid for pRH04.152–pfMBP in which the *f1* origin is present. We had removed the *f1* origin in the construction of pRH04.152–pfMBP in anticipation that this origin might be a highly sensitive region for S1 nuclease digestion. The linear form DNA produced by S1 nuclease digestion of these two plasmids was isolated by agarose gel electrophoresis and digested with *Nco*I, *Eag*I, or *Sac*II— restriction endonucleases that each have one unique site in pRH04.152–pfMBP and pDIMC8–pfMBP. The size of the resulting products was analyzed by agarose gel electrophoresis (Fig. 1B). The presence of a ladder indicates that S1 nuclease cleavage is not without bias. However, the smear over a wide size range and the presence of numerous bands indicated that there was no gross bias of S1 nuclease cleavage toward only 1 or 2 specific regions, such as the *f1* and *p15a* origins.

We compared the S1 nuclease cleavage to DNaseI cleavage by digesting the plasmids pDIMC8–rbsb and pRH04.152–rbsb with either S1 nuclease or DNaseI and then subsequently digesting the linear product with *Xmn*I (Fig. 1C). DNaseI, like S1 nuclease, produced digestions throughout the plasmid. DNaseI produced a more even distribution of digestion products. However, the bias observed with S1 nuclease may be partially removed during transformation, since insertions in the *p15A* origin are unlikely to produce viable plasmids. Furthermore, whether or not a low level of digestion occurs at all locations in the plasmid cannot be determined from this coarse evaluation of the distribution of S1 nuclease digestion sites. Sequencing of individual members of a library would provide more specific data on the sites of S1 nuclease cleavage. Finally, the true test of S1 nuclease for making insertion libraries for switch creation is whether it can be used to make a library from which switches can be identified.

Creation of an Insertion Library

In order to determine the exact locations of double-stranded breaks produced by S1 nuclease, we created a library of insertions of blunt-ended, linear dsDNA into pDIMC8–rbsB. For the insert DNA we used a circularly permuted variant of the TEM1 β-lactamase gene (*bla*) that is permuted at residue 170 (*bla[170–170]*). This particular circular

permutation was identified in MBP317–347 (Guntas et al., 2005), a protein switch in which β-lactamase activity is strongly activated by maltose. The library construction protocol (Fig. 2) was based on previous protocols for
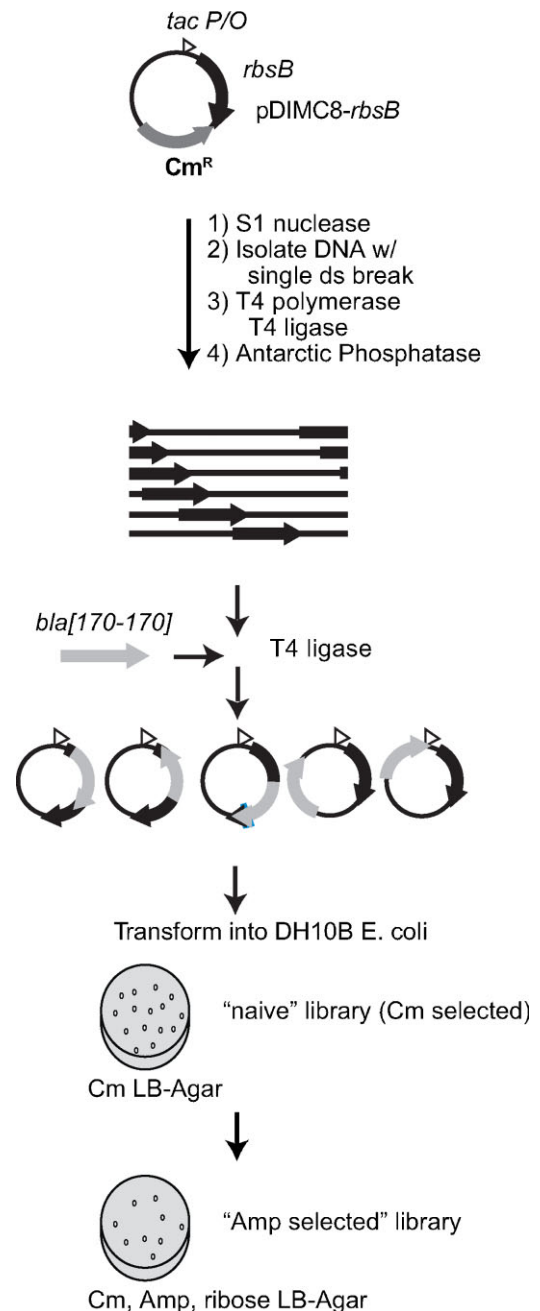

Figure 2. Schematic description of the library creation method. **Step 1**: S1 nuclease digests supercoiled plasmid DNA. **Step 2**: DNA with a single, double-stranded break is isolated by agarose gel electrophoresis. **Step 3**: Incubation with T4 polymerase and T4 ligase repairs any single strand overhangs and seals any nicks. **Step 4**: Dephosphorylation prevents recyclization during the subsequent ligation step. **Step 5**: An insert DNA (created by PCR using phosphorylated primers) is incorporated using T4 ligase. **Step 6**: The library DNA is transformed into *E. coli* cells. **Step 7**: The library is selected for ampicillin resistance and the selected clones screened for switching ability.

random insertion library creation that used DNaseI as the agent for random cleavage (Guntas et al., 2005). We anticipated that digestion with S1 nuclease, like DNaseI, would primarily create deletions at the site of double strand breaks, owing to S1 nuclease's well-known proclivity for removing ssDNA overhangs in dsDNA. However we hoped that unlike digestion with DNaseI, which can readily progress to complete digestion of dsDNA, digestion with S1 nuclease would result in smaller deletions owing to linear dsDNA's resistance to S1 nuclease.

The S1 nuclease step was followed by a repair step with T4 DNA ligase and T4 DNA polymerase that was designed to blunt any ssDNA overhangs (should they exist) and seal any nicks in the dsDNA. Subsequent treatment with alkaline phosphatase minimized recircularization of the vector without the desired insertion of the circularly permuted *bla* into the vector. The resulting naïve library consisted of 1.12×10^7 transformants, of which approximately 50% contained the inserted *bla* gene (based on plasmid size determined by agarose gel electrophoresis). This library is 10–100 times larger than those obtained using an analogous protocol with DNaseI (Guntas and Ostermeier, 2004; Guntas et al., 2004, 2005). Although we describe this library as "naïve" we note that the act of transformation and plating on agar containing chloramphenicol necessarily selects against insertions in the p15A origin that interfere with plasmid replication and insertions in the chloramphenicol resistance gene that cause loss of chloramphenicol resistance.

Characterization of the Naïve Insertion Library

The insertion of the *bla* DNA at the site of S1 nuclease digestion provided a facile way to locate the S1 nuclease digestion site in individual clones of the library: sequencing from the *bla[170–170]* DNA outward in both directions. Plasmid DNA from 30, randomly chosen, insert-containing, naïve library members were sequenced and the location of S1 nuclease cleavage identified (Fig. 3A). All sequenced library members were unique, however two members had the same four nucleotides deleted but had *bla[170–170]* inserted in opposite orientations (see Supplementary Information). No insertions were identified in the Cm^R gene (660 bp) or the *p15a* origin (483 bp)—consistent with the expectation that the vast majority of insertions in those locations would not produce a viable plasmid capable of conferring chloramphenicol resistance. Thus, of the 4,310 bp in the plasmid, we estimated that only 3,167 bp are sites where insertion of the cpBLA gene would produce a viable plasmid in the naïve library. The *f1* origin is 455 bp, so we would expect to see 455/3,167 = 14.3% of the sequenced colonies to contain insertions within the f1 origin. We find 9/30 = 30% contained an insert in the *f1* origin, which is only twofold more frequent than expected if the site of S1 cleavage were completely random. The *rbsB* gene (890 bp) represents 28% of the 3,167 bp. Twenty percent of the insertions are in the *rbsB* gene. Thus, insertions in the *rbsB*
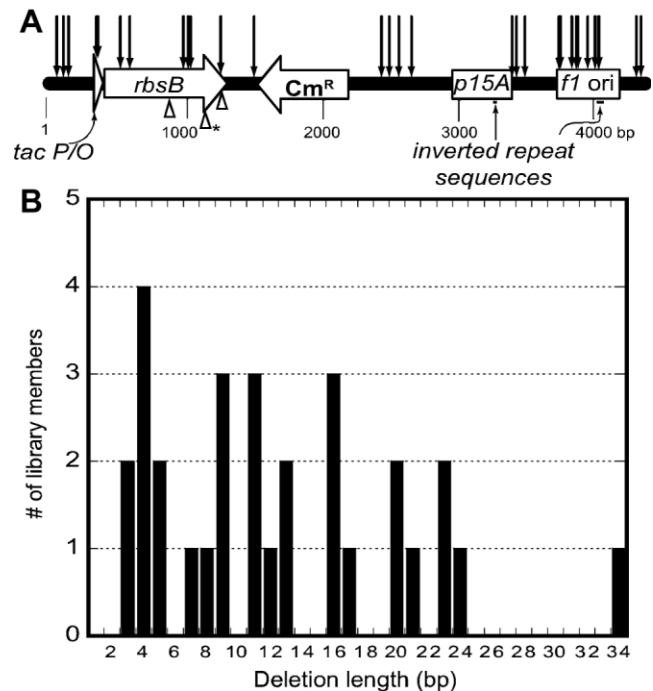

Figure 3. S1 nuclease digestion of supercoiled DNA results in a diverse set of products. **A**: To scale linear diagram of plasmid pDIMC8–*rbsB* with *rbsB* gene (including the *rbsB* signal sequence), *tac* promoter (*tacP/O*), *p15A* origin, *f1* origin, and chloramphenicol resistance gene (Cm^R) labeled. Arrows indicate the location of S1 nuclease digestion sites identified from sequencing of 30 naïve library members. Triangles indicate the location of S1 nuclease digestion sites in the switches discovered in the library, with four distinct switches discovered at the location marked by the asterisk. **B**: Distribution of deletion lengths at the site of S1 nuclease-created double stranded breaks in the same 30 naïve library members.

gene occurred about 50% less frequently than one would expect if the insertions were completely random. It is useful to compare our results here with the results of sequencing naïve library members of an analogous library created using DNaseI which had the same plasmid backbone (Guntas and Ostermeier, 2004). One difference is that 10% of the library members created with DNaseI digestion contained tandem duplications and 90% contained deletions. Here, all 30 insertions were accompanied by deletions of the vector DNA in a seemingly random distribution of sizes ranging from 3 and 34 bp (Fig. 3B). In the DNaseI-created library a wide range of deletion lengths was observed, including several clones with >100 bp deletions (Guntas and Ostermeier, 2004). Such large deletions are common in libraries created with DNaseI, especially if the amount of DNaseI and incubation time is not carefully optimized, and decrease the likelihood of creating functional fusion proteins. The sequencing results indicated that S1 nuclease has the ability to create breaks between any combination of the four bases in supercoiled DNA.

We analyzed the sequence of pDIMC8–rbsB using EMBOSS (Rice et al., 2000; Sarachu and Colet, 2005), a free, open-source DNA sequence analysis software package

that can identify inverted repeats using the "einverted" package. We identified two pairs of inverted repeats greater than eight bases in pDIMC8–rbsB—one within the p15A origin and one within the f1 origin. Insertions near the inverted repeat in the p15A origin were not found in our sampling of the naïve library presumably because insertions in that region disrupt plasmid replication. Of the nine insertions found within the f1 origin, three were within the 54 bp inverted repeat identified by EMBOSS and two of these (clones 27 and 28) were within the loop region of this repeat that would be expected to be single-stranded and had identical nucleotide deletions. Assuming a random insertion pattern, one would expect $54/3,167 = 1.7\%$ of the sequenced colonies to have been digested by S1 nuclease within the f1 origin inverted repeat site. Instead we find $3/30 = 10\%$ in the 54 bp of inverted repeat site (sixfold more prevalent than expected from a random insertion). Thus, we confirm previous results (Lilley, 1980; Panayotatos and Wells, 1981) indicating that inverted repeats are an attractive target for S1 digestion in supercoiled DNA. However, in contrast to those studies we find a predominance of digestions widespread at other locations. We believe the difference between previous results and ours lies in the different origins, and thus different inverted repeats, present in the plasmids. Panayotatos and Wells found a strong bias for digestion in the colE1 origin (Panayotatos and Wells, 1981). Indeed, we find that S1 nuclease digestion of pUC19—a plasmid containing the colE1 origin—results in a more limited set of digestion products than digestions of pDIMC8–rbsB and pRH04.152–rbsB (data not shown).

The Insertion Library Contains Gene Fusions Encoding Protein Switches

We next wondered if the library was sufficiently diverse to contain genes encoding protein switches. This is the most relevant test for whether S1 nuclease is a useful tool for making protein switch libraries. We have previously isolated allosteric switch proteins from libraries in which the gene encoding TEM1 β-lactamase was randomly inserted into a plasmid containing the gene encoding MBP (Guntas et al., 2005). The resulting switch proteins were fusions of the two proteins. The β-lactamase activity of these switches is modulated by maltose in a manner analogous to natural heterotrophic allosteric enzymes. For heterotrophic allosteric enzymes, the allosteric effector molecule that modulates the enzyme's catalytic activity is not also the substrate for the enzyme.

In the library described in this paper, switch proteins would have β-lactamase enzyme activity that is modulated by ribose. We searched the library for switches in which the presence of ribose activated β-lactamase enzyme activity. The library was first subjected to a selection for ampicillin (Amp) resistant members in the presence of ribose. Library members surviving this selection were screened for ribose-dependent β-lactamase enzyme activity using a screen for ribose-dependent nitrocefin hydrolysis activity as described (Guntas et al., 2004). We identified six switch genes that had similar, but unique sequences (Table I). The positions of insertion in the rbsB gene are indicated by triangles in Figure 3A. The lysates of cells carrying these genes contained nitrocefin hydrolysis activity that was two- to sevenfold higher when ribose was added to the reaction (Table I, second column). The nitrocefin hydrolysis activity of lysates from cells expressing wild-type β-lactamase did not depend on ribose. The genes encoding two of these switch proteins conferred to E. coli cells the ability to grow at between two- and eightfold higher concentrations of Amp when 5 mM ribose was included in the growth medium (Table I, last two columns). Ribose did not affect the Amp resistance of cells expressing wild-type β-lactamase. A one-to-one correspondence between the effect of ribose on the rate of hydrolysis of a particular concentration of nitrocefin (i.e., as in lysates) and the phenotypic effect of the switch gene on providing ribose-dependent resistance to a different antibiotic (ampicillin) is not necessarily expected or observed. The switching effect can be substrate dependent (Guntas et al., 2005) and ampicillin resistance requires effective hydrolysis at a range of concentrations of ampicillin.

Table I. Switch sequences and switching activity.

Protein	Protein sequence[a]	Switching activity[b]	MIC[c] (µg/mL)		MIC ratio
			− Ribose	+ Ribose	
p2.G8	RBP[1–131]–BLA[170–264]–DKS–BLA[24–170]–RBP[133–271]	2.0	128	512	4
p3.F4	RBP[1–215]–BLA[170–264]–DKS–BLA[24–170]–RBP[226–271]	2.7	16	128	8
p2.F4	RBP[1–220]–BLA[170–264]–DKS–BLA[24–170]–RBP[223–271]	5.3	32	128	4
p1.F10	RBP[1–220]–BLA[170–264]–DKS–BLA[24–170]–RBP[225–271]	7.2	32	128	4
p2.C5	RBP[1–220]–BLA[170–264]–DKS–BLA[24–170]–RBP[224–271]	5.1/6	16	128	8
p3.A2	RBP[1–260]–BLA[170–264]–DKS–BLA[24–170]–RBP[262–271]	2.0	512	1,024	2
BLA	Wild-type BLA[1–264]	1.0	8,192	8,192	1

[a]The number in brackets refers to the amino acid number in the corresponding parental protein. DKS is the tripeptide linker between the original N- and C-terminus of the BLA domain.
[b]The initial rate of protein-catalyzed hydrolysis of 100 µM nitrocefin in the presence of ribose divided by that in the absence of ribose. All switching activities were measure in lysates, except p2.C5 for which the value for lysate is followed by the value measure with purified protein.
[c]Minimum inhibitory concentration of ampicillin on agar plates without or with 5 mM ribose.

We chose P2.C5 (which exhibited a high ribose effect on the rate of hydrolysis of nitrocefin and on the MIC for ampicillin) to purify and confirm switching activity by testing the effect of ribose and other sugars on catalytic activity in vitro. This was facilitated by use of a C-terminal His-tag and a Ni^{+2}-affinity column. Purified p2.C5 catalyzed the hydrolysis of $100\,\mu M$ nitrocefin at a sixfold higher rate when 5 mM ribose was added to the assay buffer. Wild-type β-lactamase enzyme activity was not affected by ribose. The addition of maltose, sucrose, and glucose, sugars for which RBP lacks affinity, did not have an effect on the catalytic activity of p2.C5. The purified p2.C5 switch (at 20 nM) exhibited a rate of hydrolysis of $100\,\mu M$ nitrocefin of 9.5×10^{-5} mM/s in the presence of ribose, in comparison with 1.45×10^{-5} mM/s in the absence of ribose. At the same protein and substrate concentration, wild-type β-lactamase hydrolyzes nitrocefin at a rate of 8.6×10^{-3} mM/s (Sigal et al., 1984). Thus, p2.C5 is approximately 100-fold less active than wild-type β-lactamase but only about threefold less active than the maltose-activated switch MBP317–347 created with the same *bla[170–170]* circular permutation (Guntas et al., 2005).

RBP and MBP are both periplasmic binding proteins. Both RBP and MBP consist of two domains that rotate about a hinge region upon binding their respective ligands. The switches described here, together with previous switches created with MBP, illustrate that periplasmic binding proteins are a useful class of proteins for switch construction using domain insertion. We readily identified switches here constructed from a particular circular permutation of BLA found in the MBP-derived switches. This suggests that specific circular permutations found in switches made with one protein offer an excellent starting point for creating new protein switches with other proteins.

Conclusion

S1 nuclease creates double stranded breaks in a variety of locations in supercoiled plasmid DNA under the appropriate reaction conditions. We find that these locations are not limited to, nor overly biased toward, inverted repeats. This activity can be used to create a library in which DNA can be inserted at a diverse set of locations in the plasmid. Such libraries are sufficiently diverse such that the method can be used to create allosteric protein switches through the insertion of an enzyme domain into a ligand-binding domain. We used the method to identify fusion proteins in which β-lactamase enzyme activity is regulated by ribose. These switches are a further illustration of the utility of random domain insertion and circular permutation for the creation of protein switches. We next plan to examine whether random mutagenesis and cassette mutagenesis can be used to improve our switches both in terms of the magnitude of the ribose effect and the level of activity in the on state of the switch.

References

Bai YL, Yang ZL, Qiao MQ, Zhang XM, Zhou J, Gao CC. 2003. The action of S1 nuclease and a cloning strategy for microcircular DNAs. Sheng Wu Gong Cheng Xue Bao 19(2):240–243.

Beard P, Morrow JF, Berg P. 1973. Cleavage of circular, superhelical simian virus 40 DNA to a linear duplex by S1 nuclease. J Virol 12(6):1303–1313.

Berk AJ. 1989. Characterization of RNA molecules by S1 nuclease analysis. Methods Enzymol 180:334–347.

Biondi RM, Baehler PJ, Reymond CD, Véron M. 1998. Random insertion of GFP into the cAMP-dependent protein kinase regulatory subunit from *Dictyostelium discoideum*. Nucleic Acids Res 26(21):4946–4952.

Desai NA, Shankar V. 2003. Single-strand-specific nucleases. FEMS Microbiol Rev 26(5):457–491.

Edwards WR, Busse K, Allemann RK, Jones DD. 2008. Linking the functions of unrelated proteins using a novel directed evolution domain insertion method. Nucleic Acids Res 36(13):e78.

Germond JE, Vogt VM, Hirt B. 1974. Characterization of the single-strand-specific nuclease S1 activity on double-stranded supercoiled polyoma DNA. Eur J Biochem 43(3):591–600.

Ghangas GS, Wu R. 1975. Specific hydrolysis of the cohesive ends of bacteriophage lambda DNA by three single strand-specific nucleases. J Biol Chem 250(12):4601–4606.

Gite SU, Shankar V. 1995. Single-strand-specific nucleases. Crit Rev Microbiol 21(2):101–122.

Gonikberg EM. 1979. Kinetic parameters of superhelical DNA cleavage by endonuclease S1. Mol Biol (Mosk) 13(5):1064–1069.

Graf R, Schachman HK. 1996. Random circular permutation of genes and expressed polypeptide chains: Application of the method to the catalytic chains of aspartate transcarbamoylase. Proc Natl Acad Sci USA 93:11591–11596.

Guntas G, Ostermeier M. 2004. Creation of an allosteric enzyme by domain insertion. J Mol Biol 336(1):263–273.

Guntas G, Mitchell SF, Ostermeier M. 2004. A molecular switch created by in vitro recombination of nonhomologous genes. Chem Biol 11(11):1483–1487.

Guntas G, Mansell TJ, Kim JR, Ostermeier M. 2005. Directed evolution of protein switches and their application to the creation of ligand-binding proteins. Proc Natl Acad Sci USA 102(32):11224–11229.

Kato M, Hokabe S, Itakura S, Minoshima S, Lyubchenko YL, Gurkov TD, Okawara H, Nagayama K, Shimizu N. 2003. Interarm interaction of DNA cruciform forming at a short inverted repeat sequence. Biophys J 85(1):402–408.

Kowalski D. 1984. Changes in site specificity of single-strand-specific endonucleases on supercoiled PM2 DNA with temperature and ionic environment. Nucleic Acids Res 12(18):7071–7086.

Kowalski D, Natale DA, Eddy MJ. 1988. Stable DNA unwinding, not "breathing," accounts for single-strand-specific nuclease hypersensitivity of specific A + T-rich sequences. Proc Natl Acad Sci USA 85(24):9464–9468.

Kurahashi H, Inagaki H, Yamada K, Ohye T, Taniguchi M, Emanuel BS, Toda T. 2004. Cruciform DNA structure underlies the etiology for palindrome-mediated human chromosomal translocations. J Biol Chem 279(34):35377–35383.

Lilley DM. 1980. The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. Proc Natl Acad Sci USA 77(11):6468–6472.

Mechali M, de Recondo AM, Girard M. 1973. Action of the S1 endonuclease from *Aspergillus oryzae* on simian virus 40 supercoiled component I DNA. Biochem Biophys Res Commun 54(4):1306–1320.

Ostermeier M. 2005. Engineering allosteric protein switches by domain insertion. Protein Eng Des Sel 18(8):359–364.

Panayotatos N, Wells RD. 1981. Cruciform structures in supercoiled DNA. Nature 289(5797):466–470.

Paschon DE, Ostermeier M. 2004. Construction of protein fragment complementation libraries using incremental truncation. Methods Enzymol 388:103–116.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet 16(6):276–277.

Sarachu M, Colet M. 2005. WEMBOSS: A web interface for EMBOSS. Bioinformatics 21(4):540–541.

Shishido K. 1979. Location of S1 nuclease-cleavage sites on circular, superhelical DNAs between polyoma-virus and simian virus-40. Agric Biol Chem 43(5):1093–1102.

Shishido K, Noguchi N, Ando T. 1983. Correlation of enzyme-induced cleavage sites on negatively superhelical DNA between prokaryotic topoisomerase I and S1 nuclease. Biochim Biophys Acta 740(1):108–117.

Sigal IS, DeGrado WF, Thomas BJ, Petteway SR Jr. 1984. Purification and properties of thiol beta-lactamase. A mutant of pBR322 beta-lactamase in which the active site serine has been replaced with cysteine. J Biol Chem 259(8):5327–5332.

Vogt VM. 1973. Purification and further properties of single-strand-specific nuclease from Aspergillus oryzae. Eur J Biochem 33(1):192–200.

Weaver RF, Weissmann C. 1979. Mapping of RNA by a modification of the Berk–Sharp procedure: The 5′ termini of 15 S beta-globin mRNA precursor and mature 10 s beta-globin mRNA have identical map coordinates. Nucleic Acids Res 7(5):1175–1193.

Wiegand RC, Godson GN, Radding CM. 1975. Specificity of the S1 nuclease from Aspergillus oryzae. J Biol Chem 250(22):8848–8855.