Biotechnology Letters **23:** 303–310, 2001.

Construction of hybrid gene libraries involving the circular permutation of DNA

Marc Ostermeier** & Stephen J. Benkovic*
The Pennsylvania State University, Department of Chemistry, 414 Wartik Laboratory, University Park, PA 16802, USA
**Author for correspondence (Fax: 814 865 2973; E-mail: sjb1@psu.edu)*
***Present address: Department of Chemical Engineering, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218-2694, USA*

Received 3 October 2000; Revisions requested 19 October 2000; Revisions received 30 November 2000; Accepted 1 December 2000

Key words: circular permutation, glycinamide ribonucleotide transformylase, hybrid enzymes, incremental truncation, temperature-sensitive mutants

Abstract

The method of incremental truncation for the creation of hybrid enzymes (ITCHY) allows the creation of comprehensive fusion libraries between 5′ and 3′ fragments of two genes in a manner that is independent of DNA sequence homology. A methodology is presented for the creation of ITCHY libraries called circularly permuted ITCHY (CP-ITCHY) that allows the creation of ITCHY libraries in a manner that does not require extensive time point sampling. In addition, CP-ITCHY requires only a single vector and productively biases the library towards those fusions that are approximately the same size as the original genes. In the model system of creating fusions between fragments of the *Escherichia coli* and human glycinamide ribonucleotide transformylase genes, the CP-ITCHY libraries are shown to contain a diverse set of active fusions including those in regions of low-homology. In addition, a high percentage of active fusions were temperature-sensitive as they complemented an auxotrophic strain of *Escherichia coli* at 22 °C but not at 37 °C.

Introduction

The creation of proteins with desired properties is an important goal of biotechnology. Owing to the difficulties in a purely rational design approach, the use of combinatorial methods to obtain proteins with desired properties has increased dramatically. Central to these combinatorial strategies are (a) methods to construct combinatorial DNA libraries, and (b) methods that facilitate identification of those rare proteins with desired properties within these large libraries. We recently developed and implemented a strategy called incremental truncation for the creation of hybrid enzymes (ITCHY) to construct gene fusion libraries (Ostermeier *et al*. 1999a).

ITCHY is a combinatorial method that can generate extensive fusion libraries between fragments of two or more genes. Unlike methods such as DNA shuffling (Stemmer 1994) and staggered extension process (Zhao *et al*. 1998) that generate combinatorial gene libraries in a manner that depends on sufficient DNA homology between the genes, ITCHY is DNA-sequence independent. As such, it can generate libraries with a more diverse set of crossovers than homology based methods, including crossovers that result in internal ‘deletions’ or ‘duplications’. A key step in the creation of ITCHY libraries is the digestion of DNA by exonuclease III (*Exo*III). The rate of *Exo*III digestion is controlled such that the frequent removal of small aliquots to a digestion-quenching buffer results in a library of ostensibly every one base deletion of a gene or gene fragment. The incremental truncation libraries created by such digestion are then fused by blunt end ligation to create the ITCHY library.

The original method used to create such ITCHY libraries, which we are now calling *t*wo *v*ector-ITCHY

or TV-ITCHY, resulted in every possible 5′ fragment of one gene fused to every 3′ fragment of a second gene with equal probability. As gene fusions between very small fragments or very large fragments are unlikely to produce active fusions, a method that eliminated or at least minimized their population in ITCHY libraries would be beneficial. One possibility is to fractionate the ITCHY libraries based on size of the gene fusions and sub-clone those members with a length approximately equal to that of the parental genes. However, we sought a method of creating the library that inherently would bias ITCHY libraries towards fusions of the same size as the original gene as well as eliminate the labor-intensive sampling during *Exo*III digestion necessary to create TV-ITCHY libraries. Here we present such a method called *c*ircularly *p*ermuted ITCHY (CP-ITCHY).

Materials and methods

All enzymes used are from New England BioLabs (Beverley, MA) unless otherwise indicated.

Plasmids

Phagemid pDIM-N5 (Figure 1A) was created by replacing the short *Bam*HI-*Nsi*I fragment of pDIM-N2 (Ostermeier *et al*. 1999b) with an oligonucleotide as described in Figure 1A. Phagemid pDIM-N5-PurN[1-202*]/GART[20-203] contains a fragment of the *E. coli purN* gene that encodes amino acids 1-202 (with the mutation D144A) between the *Nde*I and *BamH*I sites of pDIM-N5 and a fragment of the human *GART* gene that encodes amino acids 20-203 between the *Bgl*II and *Spe*I sites of pDIM-N5. The vector has a stop codon between codon 202 of *purN* and the *BamH*I site.

Creation of the circularly permuted insert

A 528 bp fragment of the *E. coli purK* gene was amplified by PCR using oligos Xba-for (5′-TTAGGCCGTCTAGAGCGTCAGGCAGGCGAAC-CG-3′) and Xba-528 (5′-GCGGAAAATCTAGACT-GGTGCGCAAAATACCG-3′) such that it was flanked by *Xba*I sites (underlined). This fragment was digested with *Xba*I and cloned into the unique *Xba*I site of pUC19 to create pUC19-Xba528. Seventy micrograms of pCU19-Xba528 were digested with 1500 units *Xba*I and the shorter fragment isolated by gel

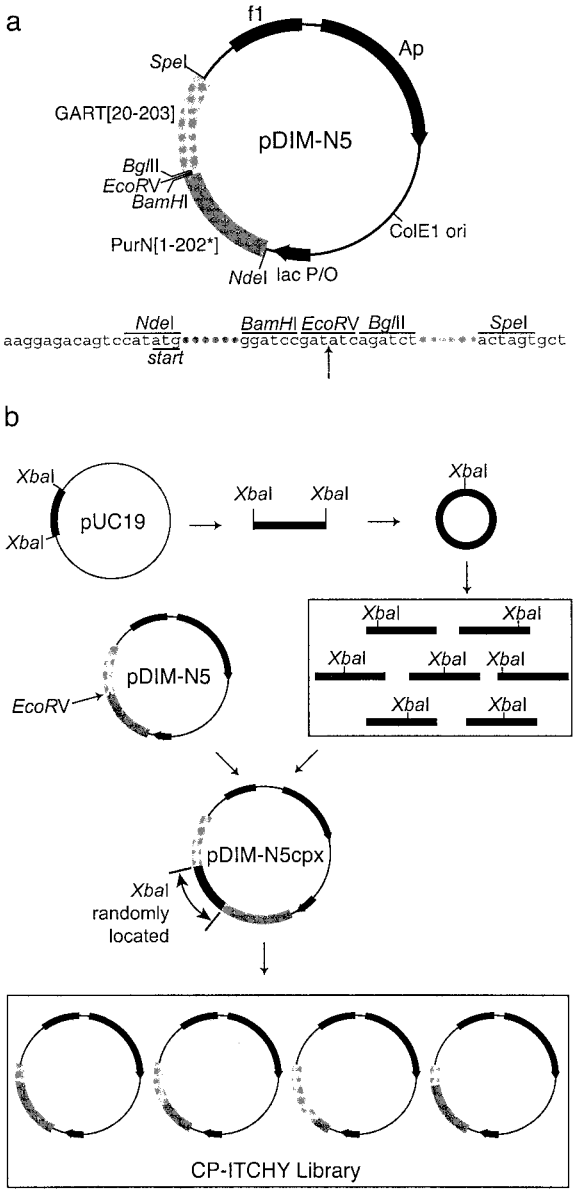

Fig. 1.

electrophoresis using QIAEX II (QIAGEN, Valencia, CA). Six μg of this fragment were treated with 200 Weiss units T4 DNA ligase in 1.7 ml ligase buffer (50 mM Tris/HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM ATP, bovine serum albumin 25 μg ml^{-1}) for 18 h at 16 °C. The ligation mixture was diluted with water up to 4 ml and concentrated approximately fifty-fold using centricon-30 spin columns (Millipore, Bedford, MA). The DNA was then digested with 600 units exonuclease III (*Exo*III, Promega, Madison, WI) in *Exo*III buffer (66 mM

Fig. 1. (facing page) (a) Description of pDIM-N5. The two gene fragments of interest (PurN[1-202*] and GART[20-203]) are cloned downstream from a IPTG inducible lac promoter (lacP/O) between the indicated restriction enzyme sites. Between the two gene fragments is located a unique restriction site that produces blunt ends (*Eco*RV). This is the site of insertion for a circularly permuted piece of DNA. The vector also has an antibiotic resistance gene (ampicillin, Ap). (b) Creation of CP-ITCHY libraries. A piece of DNA equal in length to the overlap between the two genes is amplified by PCR creating *Xba*I sites at both ends and cloned into pUC19. The DNA is excised from pUC19 using *Xba*I and treated with ligase under dilute conditions such that a significant amount of closed circular DNA is formed. The closed circular DNA is linearized at random locations by digestion with very small amounts of DNase I. The randomly linearized DNA is repaired using a DNA polymerase and a DNA ligase and cloned into the *Eco*RV site of pDIM-N5. This library of *Xba*I sites is the starting point for incremental truncation. The library is digested with *Exo*III for the desired length of time necessary to digest the same number of bp as the overlap between the two genes. The single strand overhangs are removed by mung bean nuclease, the ends are blunted with Klenow and ligation under dilute conditions results in the creation of a CP-ITCHY library.

Tris/HCl, pH 8, 0.66 mM $MgCl_2$) in a volume of 200 μl for 30 min at 37 °C to remove any unligated linear DNA. The *Exo*III was inactivated by incubation at 72 °C for 20 min. The circular DNA was desalted using QIAEX II into a final volume of 50 μl EB buffer (10 mM Tris/HCl, pH 8.5).

A series of test digestions were performed to determine the concentration of DNase I that provided the highest yield of linear product. The DNase I (RNase-free from Roche Molecular Biochemicals, Indianapolis, IN) was prepared by creating a working stock of 1 unit μl^{-1} in 50 mM Tris/HCl, pH 7.5 and 50% glycerol (v/v) that was stored at -20 °C. On the day of use, the working stock was diluted into 50 mM Tris/HCl (pH 7.5), 1 mM $MnCl_2$ and bovine serum albumin (50 μg/ml). For this experiment, 30 μl circular DNA was digested with 0.83 milliunits DNase I at 22 °C for 15 min in 50 mM Tris/HCl (pH 7.5) and 1 mM $MnCl_2$ in 400 μl. The digestion was stopped by the addition of 20 μl 50 mM EDTA, pH 8.0, and desalted using QIAquick columns (QIAGEN) into 50 μl EB buffer (see above). The linearized DNA was repaired using 3 units T4 DNA polymerase and 6 Weiss units T4 DNA ligase in ligase buffer that included 125 μM each dNTP. The repaired, linearized DNA (e.g., the circularly permuted insert) was isolated by agarose gel electrophoresis using QIAEX II into 20 μl EB buffer.

The vector was prepared by digesting 10 μg pDIM-N5-PurN[1-202*]/GART[20-203] with 50 units of *Eco*RV in 100 μl for 2.5 h. Subsequently, 90 μl water, 10 μl CIAP buffer (500 mM TrisHCl (pH 9.3), 10 mM $MgCl_2$, 1 mM $ZnCl_2$, 10 mM spermidine) and 7 units of calf intestinal alkaline phosphatase (Promega) were added and the solution incubated for an additional 1 h at 37 °C. To inactivate the alkaline phosphatase, 2 μl 500 mM EDTA, pH 8.0 was added and the DNA incubated at 72 °C for 15 min. The DNA was purified by agarose gel electrophoresis using QIAEX II into a total of 50 μl EB buffer.

One hundred ng of *Eco*RV treated, dephosphorylated vector was ligated to 10 μl circularly permuted insert with 30 Weiss units T4 DNA ligase in 15 μl at 22 °C for about 20 h. Eight electroporations of 1 μl ligation mix into 50 μl DH5α-E electrocompetent cells (rated at $\sim 10^{10}$ transformants per μg DNA) resulted in a library of 1.1×10^6 transformants on a 243 × 243 mm plate. The library was recovered and stored as previously described (Ostermeier *et al.* 1999b).

PCR characterization of circularly permuted insert

Individual colonies resulting from plating a dilution of the frozen library were analyzed by PCR to determine the location of the *Xba*I site in individual members of the library. Since for any given colony it is unknown which orientation the circularly permuted insert exists, three oligos were used in the PCR reaction: Xba-for, Xba-528 and PurN-for (5′-GATATACATATGAATATTGTGGTGCTTATTTCC-3′), an oligo that annealed to the beginning of the *purN* gene. Depending on which orientation the circularly permuted insert was ligated either (PurN-for and Xba-528) or (PurN-for and Xba-for) would produce an exponential amplification. The size of the PCR product was determined by agarose gel electrophoresis and the location of the *Xba*I site was then determined by subtracting the size of *purN*[1-202*].

Incremental truncation

A plasmid prep (QIAGEN Midiprep) on 40% of the frozen library yielded 54 μg of supercoiled plasmid. The plasmid DNA (20 μg) was digested with 40 units of *Xba*I for 1.5 h at 37 °C. The linearized vector was isolated from any uncut vector and any vector not containing the circularly permuted insert by agarose gel electrophoresis using QIAEX II. *Exo*III digestion was performed on 4 μg linearized vector at 22 °C in 120 μl 66 mM Tris/HCl (pH 8.0)/0.66 mM $MgCl_2$/50 mM NaCl using 800 units of *Exo*III (Promega). Twenty-four μl aliquots were removed at 24, 25, 26, 27 and 28 min and added to 72 μl of 40.5 mM potassium

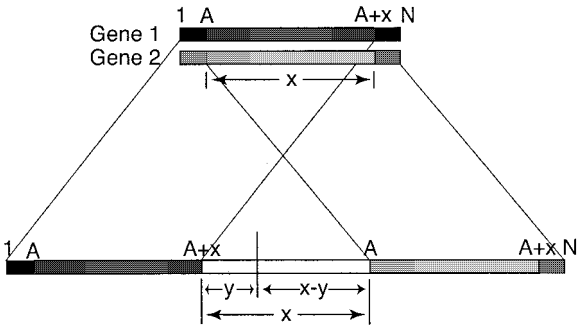

Fig. 2. Mathematical description of CP-ITCHY. This example is for creating an ITCHY library between two genes of length N. The truncation is performed on the indicated *N*-terminal fragment of gene 1 and *C*-terminal fragment of gene 2 where the region of truncation is located between A and $A + x$. The indicated fragments of the two genes, having an overlap of length x, are cloned into the same vector. A third piece of DNA of length x with a randomly located unique restriction site y bases from one end is inserted between the two gene fragments. If one then truncates x bases in either direction from this site, one will arrive at a location in either gene whose fragments sum to the length of the original genes. Thus, for gene 1, one starts at position $A + x + y$ (relative to gene 1) and truncates x bases (a decrease of the position) to arrive at $A+x+y-x = A+y$. For gene 2 one starts at position $A-(x-y)$ (relative to gene 2) and truncates x bases (an increase in the position) to arrive at $A-(x-y)+x = A+y$. The starting point for CP-ITCHY libraries is a library in which y is all possible integers between 0 and x.

acetate (pH 4.6), 338 mM NaCl, 1.35 mM $ZnSO_4$, 6.76% (v/v) glycerol at 4 °C to quench the digestion. After all the samples had been quenched, 0.5 ml of QIAquick buffer PB (QIAGEN) was added and the DNA purified using the QIAquick protocol with one modification: after the addition of the wash PE buffer the samples were incubated for 5 min at room temperature before spinning to insure removal of any salt. The DNA was eluted from the QIAquick column using 47 μl EB buffer. To this eluate, 5 μl 10× mung bean buffer (500 mM sodium acetate (pH 5), 300 mM NaCl, 10 mM $ZnCl_2$) and 4 units mung bean nuclease were added and the solution incubated at 30 °C for 30 min. Next, 0.25 ml of QIAquick buffer PB (QIAGEN) was added and the DNA purified using the QIAquick protocol with the modification listed above. The DNA was eluted from the QIAquick column using 47 μl buffer EB. To this eluate, 5 μl dNTP mix (0.125 mM each dNTP) and 5 μl 10× EcoPol buffer (100 mM Tris-HCl (pH 7.5), 50 mM $MgCl_2$, 75 mM dithiothreitol) was added and the solution equilibrated to 37 °C. Next, one unit of Klenow DNA polymerase was added. Following incubation at 37 °C for 5 min, the DNA polymerase was heat inactivated at 75 °C for 20 min. To this solution was added 98.7 μl water, 20 μl 10 × ligation buffer, 20 μl 50% (w/v) PEG and 1.33 μl T4 DNA ligase (8 Weiss units) and the solution was incubated at room temperature overnight. The DNA was concentrated by ethanol precipitation (with ammonium acetate as salt) into 10 μl of water. A single electroporation of 3 μl DNA into 50 μl TX680F′ (Ostermeier *et al.* 1999b) electrocompetent cells (rated at 1×10^8 transformants per μg pUC19) resulted in a library of approximately 1×10^6 for each of the five time points. Active PurN-GART fusions were identified by complementation of a GAR transformylase auxotrophic E. coli strain (TX680F′) as previously described (Ostermeier *et al.* 1999a).

Results and discussion

In contrast to TV-ITCHY, in which variation in truncation length is created by initiating truncation at a single point on the vector and ceasing truncation at various time points, variation in truncation length in CP-ITCHY is created by initiating truncation at different points on the vector and ceasing truncation at one time point. The different sites for initiating truncation are created by the circular permutation of a piece of DNA that is equal in length to the length of overlap between the two gene fragments and contains a unique restriction site (Figure 1B). The library of circular permutations of this piece of DNA is inserted between the two genes to be truncated (Figure 1A) and the unique restriction site serves as the start of truncation in both directions. As illustrated in Figure 2, CP-ITCHY will create all possible fusions between the two gene fragments at or near where the sequences align.

Description of model system

We have previously shown how TV-ITCHY can be used to identify active fusions between an *N*-terminal fragment of PurN (*E. coli* glycinamide ribonucleotide formyltransferase) and a *C*-terminal fragment of GART (human glycinamide ribonucleotide formyltransferase) (Ostermeier *et al.* 1999a). Although the experiment was designed to search for active hybrids fused anywhere between amino acids 54 and 144, all of the active hybrids were found to be fused between amino acids 100 and 144, almost all of them fused exactly where the sequences align.

We decided to use this same system to test CP-ITCHY and at the same time expand the range of

our search to between amino acids 20 and 144. We also sought to demonstrate an expanded range of incremental truncation from 270 bp to over 500 bp. However, we thought that fragments of PurN larger than PurN[1-144] may be active by themselves, regardless of what fragment of GART was fused to them. So to expand the range of truncation, without having fusions between PurN and GART active solely due to PurN residues, we decided to use fragment of PurN in which residue 144 had been mutated from aspartate to alanine: a mutation that inactivates PurN (Shim & Benkovic 1999). Thus the fragments used were GART [20-203] and PurN[1-202*], with the star symbolizing the D144A mutation. This gives a range of overlap between the two fragments of 182 amino acids (546 bp), almost the entire length of the two genes. However, because of the D144A mutation in PurN[1-202*], we did not anticipate finding any active hybrids fused within the range 145-202.

Circular permuted library

Fragments *purN*[1-202*] and *GART*[20-203] were cloned into phagemid pDIM-N5 as shown in Figure 1A. This phagemid was linearized by digestion between the two gene fragments with *Eco*RV, treated with alkaline phosphatase and purified by agarose gel electrophoresis in preparation for cloning in the circularly permuted insert.

The method for circular permutation was adapted from a recently published protocol (Baird *et al*. 1999). A fragment of the *purK* gene was amplified such that it was flanked with *Xba*I sites. The length of the *purK* fragment was such that once it was circularly permuted and cloned into pDIM-N5, the distance between the end of the end of *purN*[1-202*] and the beginning of *GART*[20-203] would be equal to the overlap between the *purN*[1-202*] and *GART*[20-203] in a sequence alignment. Although in principal the PCR product could be used directly in the circular permutation scheme, better results were obtained by first cloning it into the *Xba*I site of pUC19, digesting it out with *Xba*I and isolating the fragment by agarose gel electrophoresis. The fragment of *purK* with *Xba*I overhangs was cyclized by ligation under dilute conditions so that the major product was closed, circular DNA. To prevent biasing the CP-ITCHY library, the ligase-treated DNA was incubated with *Exo*III to remove any remaining linear starting material. Next, the circular DNA was digested with a small amount of DNase I to introduce, on average, one double-stranded break per molecule. The DNase I-digested DNA was treated with T4 DNA ligase and T4 DNA polymerase in the presence of dNTPs to repair gaps and nicks in the linearized product and to produce blunt ends. This repair step is crucial for the successful recovery of functional linearized DNA (Graf & Schachman 1996). This blunt-end, circularly permuted DNA was ligated at 22 °C into the previously prepared *EcoR*V/alkaline phosphatase treated pDIM-N5-PurN[1-202*]/GART[20-203].

Electroporation into DH5α-E resulted in a library of 1.1×10^6 transformants, making it all but certain that the approximately 500 possible circular permutations were present. To confer a random distribution of *Xba*I sites in the library, PCR was performed and randomly selected colonies. As expected, the location of the *Xba*I site was essentially random (Figure 3A).

General improvements to incremental truncation method

Since our initial description of TV-ITCHY, we have made two improvements to TV-ITCHY, both of which are incorporated into CP-ITCHY. First, we have switched from S1 nuclease to mung bean nuclease for removing the single stranded tail after *Exo*III digestion. We have found that S1 nuclease sometimes would fail to remove the single stranded tail from all of the DNA molecules and this primarily accounted for the bias towards shorter truncations noted previously. The second improvement was replacing the heat inactivation and ethanol precipitation with a DNA affinity column (QIAquick) to purify the DNA away from *Exo*III and the single stranded nuclease. These two modifications significantly improved the yield and quality of the truncated DNA for both TV-ITCHY and CP-ITCHY.

CP-ITCHY library

Plasmid from circularly permuted library was digested with *Xba*I and purified by agarose gel electrophoresis in preparation for truncation. Using control digestions on this DNA, we found that under the conditions used (4 μg DNA in 120 μl of 66 mM Tris/HCl (pH 8)/0.66 mM $MgCl_2$/50 mM NaCl with 800 units of *Exo*III at 22 °C) the rate of *Exo*III digestion was approximately 21 bp min^{-1} in each direction. Thus, to digest 546 bp in each direction would require a digestion time of 26 min.

The *Xba*I-digested DNA was digested with *Exo*III for 24, 25, 26, 27 or 28 min before quenching in a

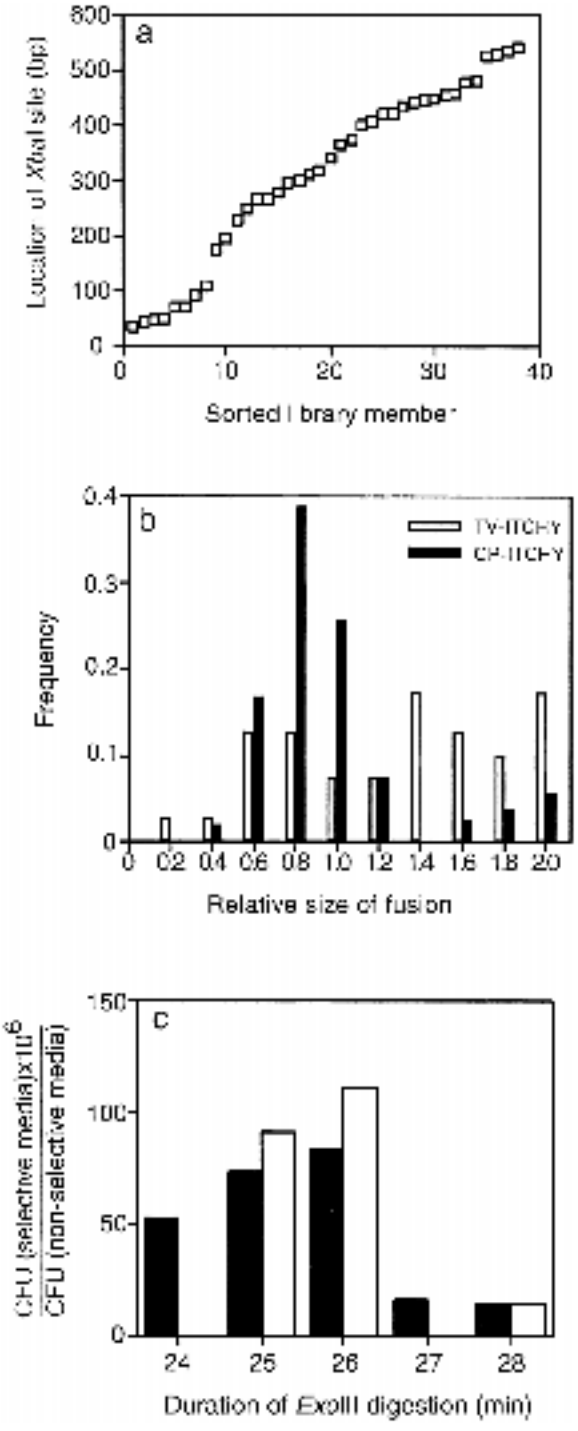

Fig. 3.

Fig. 3. (a) Distribution of *Xba*I sites in the circularly permuted insert. The location of the *Xba*I sites in 38 randomly selected members of the library were determine by PCR as described in the methods, and sorted into ascending order for presentation in this graph. The expected range of location of the *Xba*I site is between 0 and 528 bp. (b) Size distribution of fused genes in TV-ITCHY (open bars) and CP-ITCHY (solid bars) libraries. The size was determined by agarose gel electrophoresis of PCR reactions using outside primers on 40–60 randomly selected colonies. The size has been normalized such that a relative size of 1 is the same size as the starting genes, 0 is the size where both gene fragments have been truncated the maximum amount desired and 2 is the size where both gene fragments have not been truncated at all. The TV-ITCHY data combines two previously published libraries (Ostermeier *et al*. 1999a) and two unpublished libraries. (c) In each of the five CP-ITCHY libraries the frequencies of fusions capable of complementing the auxotroph at 37 °C are shown by the solid bars. A second independent, but otherwise identical, complementation experiment (open bars) was performed on three of the five libraries (CP-25, CP-26 and CP-28).

high salt, low pH buffer. The five libraries are subsequently referred to as CP-24, CP-25, etc. The DNA was desalted and purified away from the *Exo*III using a QIAquick affinity column. After treatment with mung bean nuclease to remove the single stranded tail, the DNA was treated with DNA polymerase to assure blunt ends. Ligation at 22 °C under dilute conditions circularized the truncated DNA library. The DNA was concentrated by ethanol precipitation into 10 μl and 3 μl of this was electroporated into 50 μl of electrocompetent TX680F′ cells (determined to transform with pUC19 at 1×10^8 transformants μg^{-1}). The size of the five libraries (the number of transformants) ranged from 9×10^5 to 1.1×10^6.

The size distribution in the five libraries was determined by agarose gel electrophoresis of PCR reactions on 55 randomly selected members of the five libraries using PurN forward and GART reverse primers. As can be seen in Figure 3B, CP-ITCHY creates a library biased towards those fusions that are about the same size as the original genes, whereas TV-ITCHY has a more even distribution over the size range.

Selection of active fusions

Active members of the five libraries were identified by complementation of an *E. coli* auxotroph grown at 37 °C. As expected, the highest frequency of active fusions was found in CP-26 (Figure 3C). However, owing to the size of the standard deviation in truncation length, which increases linearly with the length of truncation as 22 bp per 100 bp truncated (Hoheisel 1993), active fusions were found in the other four libraries as well. The frequency of fusions in CP-

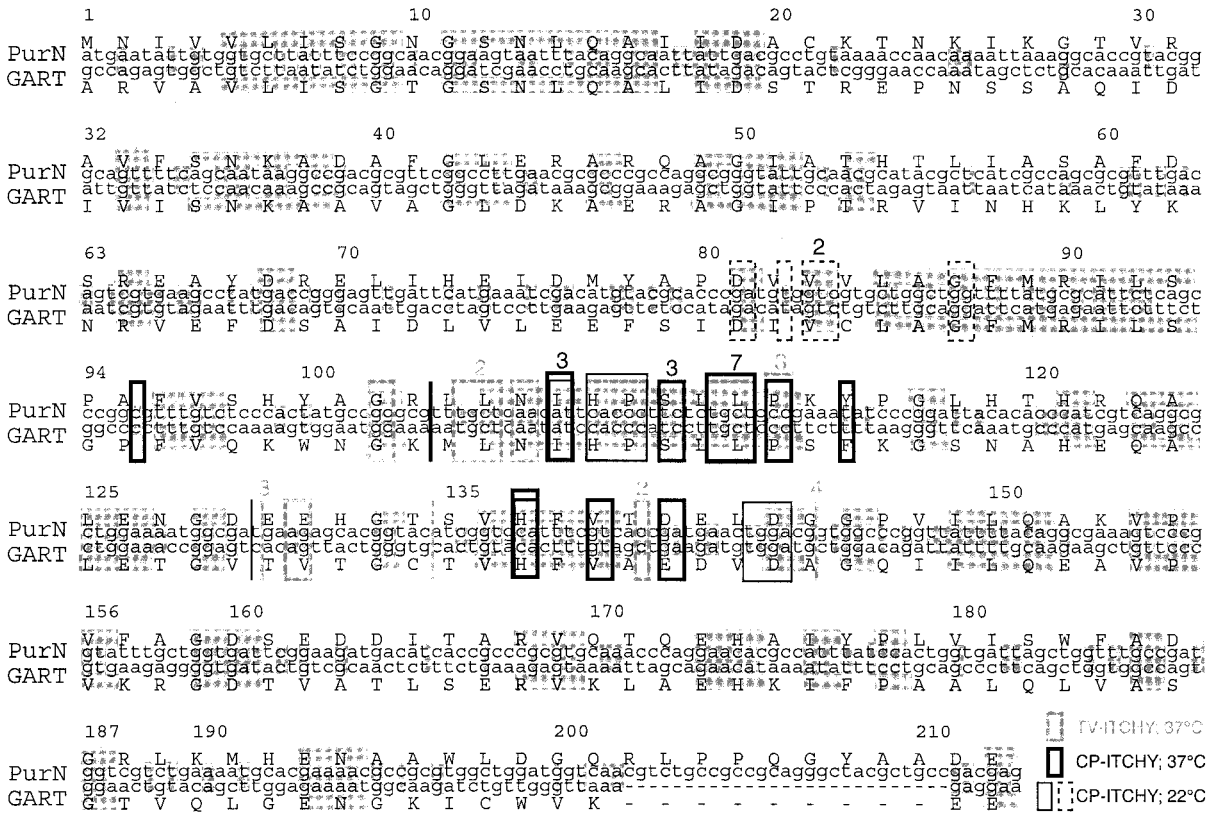

Fig. 4. Fusion points of active PurN-GART hybrids relative to the alignment of PurN and GART. The location of fusions of active hybrids found by TV-ITCHY (thick grey lines) (Ostermeier *et al.* 1999a) and CP-ITCHY at 37 °C (thick black lines) and CP-ITCHY at 22 °C (thin black lines) are shown. An unambiguous fusion point is indicated by a single line. If the fusion occurred within a region of identity, the exact fusion point cannot be known, thus the fusion point is shown as a box within which fusion must have occurred. To the left of the fusion point the sequence of the fusion is that of PurN and to the right it is that of GART. A dashed box indicates that fusions at this location complement the auxotroph when grown at 22 °C but not at 37 °C. The number above the line (or box) indicates the number of times that sequence was found in a random sampling of twenty (TV-ITCHY and CP-ITCHY at 37 °C) or ten (CP-ITCHY at 22 °C) positives. If no number is shown, the number of times the sequence was found is one. The CP-ITCHY libraries were searched between amino acids 20 and 144. The TV-ITCHY library searched between amino acids 54 and 144.

26 is approximately four fold higher than that which would be expected in a TV-ITCHY library constructed over the same size range. The frequency of positives expected in a TV-ITCHY library over the same size range was estimated by taking the frequency in a smaller library where truncations occurred over 270 bp (Ostermeier *et al.* 1999a) and, knowing that no new fusions are found outside the range of this library when the truncation range is 546 bp (see below), dividing by the ratio of the theoretical library sizes for truncations of 546 and 270 bp ($546^2/270^2$).

Twenty random active fusions were sequenced. Like the 20 randomly selected active members of TV-ITCHY library IT-B (which identified eleven different DNA sequences and seven different proteins) (Ostermeier *et al.* 1999a), CP-ITCHY identifies a variety of different fusion points (ten different DNA sequences and six different proteins) at homologous and non-homologous locations (Figure 4). Three of the six proteins identified by CP-ITCHY are newly identified active fusions.

Temperature sensitive fusions

CP-ITCHY libraries CP-24 and CP-27 were also tested for complementation of the auxotroph at 22 °C. The frequency of positives at 22 °C was found to be 8 and 5.4-fold higher, respectively, than the frequency of positives at 37 °C. Of ten randomly chosen positives of CP-24 selected at 22 °C, five were unable to grow at 37 °C. The gene fusions from the ten randomly-chosen positives were sequenced (Figure 4). The five temperature sensitive fusions were fused in a

region between amino acids 80 and 90, a region where no active fusions had previously been identified. The five non-temperature sensitive fusions were fused in regions previously identified by selection at 37 °C.

Conclusions

We have demonstrated that CP-ITCHY can be used to create extensive protein fusions libraries in manner that is independent of DNA homology. In CP-ITCHY, incremental truncation is performed starting from a circularly permuted insert located between the two genes to be fused. The major advantages of CP-ITCHY are that it (a) uses only one vector, (b) does not require extensive time-point sampling to generate the library, and (c) it biases the library towards fusions where the sequences align for similarly sized genes that do not have major gaps in their alignment. In addition, CP-ITCHY (a) requires that only one restriction enzyme site not be in the genes instead of 3-5, (b) decreases by a factor of 5-6 the amount of time between the start of truncation and the start of ligation, and (c) eliminates the need for isolation of DNA fragments by agarose gel electrophoresis. Recently we have developed a third method for creating ITCHY libraries called THIO-ITCHY involving the low level incorporation of α-phosphorothioate dexoynucleotides that shares many of the same advantages as CP-ITCHY (S. Lutz, M. Ostermeier and S.J. Benkovic, submitted). These three methods for combinatorializing genes offer many opportunities for protein engineers to explore.

Acknowledgements

We thank S. Lutz for many helpful discussions and for a critical reading of this manuscript and G. S. Baird for advice on circular permutation. This work was supported in part by National Institutes of Health Grant GM24129 (SJB).

References

Baird GS, Zacharias DA, Tsien RY (1999) Circular permutations and receptor insertions within green fluorescent proteins. *Proc. Natl. Acad. Sci. USA* **96**: 11241–11246.

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.

Hoheisel JD (1993) On the activities of *Escherichia coli* exonuclease III. *Anal. Biochem.* **209**: 238–246.

Ostermeier M, Shim JH, Benkovic SJ, (1999a) A combinatorial approach to hybrid enzymes independent of DNA homology. *Nature Biotechnol.* **17**: 1205–1209.

Ostermeier M, Nixon AE, Shim JH, Benkovic SJ (1999b) Combinatorial protein engineering by incremental truncation. *Proc. Natl. Acad. Sci. USA* **96**: 3562–3567.

Shim JH, Benkovic SJ (1999) Catalytic mechanism of *Escherichia coli* glycinamide ribonucleotide transformylase by sit-directed mutagenesis and pH-dependent studies *Biochemistry* **38**: 10024–10031.

Stemmer WPC (1994) DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91**: 10747–10751.

Zhao H, Giver L, Shao Z, Arnold FH (1998) Molecular evolution by staggered extension process (StEP) *in vitro* recombination. *Nature Biotechnol.* **16**: 258–261.