A combinatorial approach to hybrid enzymes independent of DNA homology

Marc Ostermeier, Jae Hoon Shim, and Stephen J. Benkovic*

*The Pennsylvania State University, Department of Chemistry, 414 Wartik Laboratory, University Park, PA 16802. *Corresponding author (sjb1@psu.edu).*

Received 6 September 1999; accepted 30 September 1999

We present a methodology, termed incremental truncation for the creation of hybrid enzymes (ITCHY), that creates combinatorial fusion libraries between genes in a manner that is independent of DNA homology. We compared the ability of ITCHY and DNA shuffling to create interspecies fusion libraries between fragments of the *Escherichia coli* and human glycinamide ribonucleotide transformylase genes, which have only 50% identity on the DNA level. Sequencing of several randomly selected positives from each library illustrated that ITCHY identified a more diverse set of active fusion points including those in regions of nonhomology and those with crossover points that diverged from the sequence alignment. Furthermore, some of the hybrids found by ITCHY that were fused at nonhomologous locations had activities that were greater than or equal to the activity of the hybrids found by DNA shuffling.

Keywords: incremental truncation, DNA shuffling, hybrid enzymes, in vitro evolution, glycinade ribonucleotide transformylase

The introduction of DNA shuffling[1] in 1994 has had an enormous impact on the field of protein engineering. DNA shuffling or variations of the technique[2–4] have been used to improve enzyme activity[2,5–7], stability[8–10], and folding[11,12], and to alter substrate specificity[13]. All of these methods require relatively high levels of DNA homology to recombine genes in vitro. They are effective tools for the directed evolution of proteins because of the rational, combinatorial manner in which sequence and structural space is explored. However, DNA shuffling cannot exploit a large portion of the total combinatorial space, because crossover points between shuffled genes occurs only in regions of relatively high-level DNA homology and at loci of identity. It can be argued that crossovers between structurally homologous proteins at sites lacking DNA homology are likely to be productive for protein engineering. In fact, Monte Carlo simulations of directed evolution strategies found that exchange of nonhomologous low-energy structures was a more productive strategy than DNA shuffling[14]. However, no combinatorial strategy for creating hybrids between genes that lack DNA homology has been demonstrated.

We have described how incremental gene truncation libraries can be used to identify loci for the functional bisection of proteins[15] and have proposed a number of protein engineering strategies that utilize incremental truncation[16]. Here we present a methodology termed incremental truncation for the creation of hybrid enzymes (ITCHY) that creates combinatorial libraries between two genes in a manner that is independent of DNA sequence homology. We examined the utility of ITCHY to create functional interspecies hybrids between the *Escherichia coli* and human glycinamide ribonucleotide (GAR) formyltransferases, which have only 50% identity on the DNA level. Using this model system, we show that ITCHY libraries allow the identification of a more diverse set of functional fusions than DNA shuffling. Furthermore, a subset of the functional fusion points uniquely identified using ITCHY contained fusions at nonhomologous locations, and the corresponding hybrid enzymes were as active or more active than any of the functional hybrids found by DNA shuffling.

Results

ITCHY allows the creation of hybrid enzyme libraries between a random-length 5′ fragment of the gene encoding protein A and a random-length 3′ fragment of the gene encoding protein B (Fig. 1). A key step in this process is the digestion of the parent genes with exonuclease III (ExoIII) in the presence of NaCl such that the reaction rate is limited to ≤10 bases/min (ref. 17). During ExoIII digestion, small aliquots are removed at short intervals and quenched by addition to a low-pH, high-salt buffer. As ExoIII digests DNA at a relatively uniform rate[18], members of the library ostensibly correspond to progressive 1 bp deletions.

Description of model system. We tested our method by creating interspecies hybrids between the *E. coli* and human genes for GAR transformylase. In *E. coli*, the product of the *purN* gene encodes a monofunctional GAR transformylase of 212 amino acids that catalyzes the transfer of the formyl group from the cofactor N^{10}-formyltetrahydrofolate (fTHF) to the amino group of GAR to yield formylglycinamide ribonucleotide (fGAR)[19]. In humans, GAR transformylase activity resides on the trifunctional enzyme glycinamide ribonucleotide synthetase–aminoimidazole ribonucleotide synthetase–glycinamide ribonucleotide transformylase (GARS-AIRS-GART)[20]. The GART segment, like PurN, utilizes the cofactor fTHF and is functional as a separate domain[21,22]. There is 50% identity at the DNA level between the two genes and 41% identity (60% homology) on the amino acid level. An amino acid alignment between PurN and the GART segment reveals no gaps, although GART lacks nine amino acids at the C terminus. Based on this alignment, we refer to the GART segment of the human trifunctional enzyme as GART[1–203] for simplicity. The structures of the active sites of the two enzymes have been reported to be essentially identical[23], but the structure of GART is not available in the Protein Data Bank[24].

ITCHY libraries. Two ITCHY libraries were created between 5′ fragments of *purN*[1–144] and 3′ fragments of *GART*[54–203]. The first (IT-A; 1.3×10^5 independent clones) was created by electroporation into DH5α (2×10^8 transformants/μg DNA). In order to obtain a larger library, a second library (IT-B, 1.2×10^6 clones) was created by electroporation into DH5α-E (approximately 1×10^{10} transformants/μg DNA). The chosen gene fragments have 1–270 bp of overlap; thus, as each gene fragment would have 1 to 270 bp truncated, an ideal library containing one member of each of the desired fusions would have $270 \times 270 = 72{,}900$ members.

Provided there is no significant bias in the libraries, IT-B, with 16-fold more library members than the ideal library size, should contain all possible fusions between the two gene fragments in the region of overlap. The size diversity of IT-A and IT-B was evaluated on randomly selected library members and found to be essentially random, but with a bias against small fusions (Fig. 2).

Hybrids of PurN and GART with GAR transformylase activity were selected for on minimal media using *E. coli* auxotroph TX680F', which lacks a functional GAR transformylase[15]. Because GAR transformylase activity is essential for purine biosynthesis, TX680F' is unable to grow on minimal media in the absence of purines. Plasmid DNA was purified from the libraries and transformed into TX680F'. Several steps were taken to ensure that complementation resulted from a *purN–GART* gene fusion including PCR screening for recombinants and retransformation of the plasmid back into the auxotroph to confirm complementation. The number of active fusions per library of IT-A and IT-B was estimated to be 9 and 111, respectively, by multiplying the number of colonies on the selective plates by the library size and dividing by the number of colony-forming units (c.f.u.; rich media) plated on the selective plates.

The amino acid sequences of randomly selected active fusions of IT-A and IT-B were determined by DNA sequencing (Table 1). Active genes were found with fusions in regions of high and low homology (Fig. 3) and within loops, α-helices and β-sheets of PurN (Fig. 4). Somewhat surprisingly, almost all fusion points of active hybrids occurred at sites of exact alignment. The exception is IT-A18, in which an extra glycine results from a fusion of the first base of A101 of *purN* to the last two bases of Trp100 of *GART*. Fusions IT-A12 and IT-B9 result from a fusion of the entire length of *purN*[1–144], including the first base of a stop codon located immediately after *purN*[1–144] to the last two bases of codon Ala145 of *GART* plus *GART*[146–203] to create serine at position 145 instead of glycine (in PurN) or alanine (in GART). None of the sequenced, active members of IT-A or IT-B had any other mutations. Covalent fusion of the two fragments was found to be necessary as no interspecies heterodimers between any combination of truncated fragments of PurN[1–144] and GART[54–203] were able to complement the auxotroph (data not shown).

DNA shuffling libraries. To compare ITCHY to DNA shuffling, we created *purN–GART* gene fusions by DNA shuffling (library SH-A). The same gene fragments used in ITCHY were used to cre-

Table 1. Active PurN–GART fusions.

Library	Representative positive number	Amino acid sequence[a]			Number of times found[b]
		PurN	Link	GART	
ITCHY A	IT-A18	1–100	G	101–203	2 (1)
	IT-A8	1–113	–	114–203	9 (4)
	IT-A5	1–129	–	130–203	3 (1)
	IT-A26	1–136	–	137–203	1 (1)
	IT-A3	1–140	–	141–203	3 (1)
	IT-A1	1–144	–	145–203	3 (1)
	IT-A12	1–144	S	146–203	3 (1)
ITCHY B	IT-B12	1–102	–	103–203	1 (1)
	IT-B4	1–113	–	114–203	8 (5)
	IT-B3	1–129	–	130–203	3 (1)
	IT-B17	1–130	–	131–203	1 (1)
	IT-B6	1–134	–	135–203	1 (1)
	IT-B5	1–140	–	140–203	2 (1)
	IT-B9	1–144	S	146-203	4 (1)
Shuffled	SH-5	1–113[c]	–	114–203[c]	24 (5)

[a]When the fusion point occurs within a region of amino acid identity, the region of identity is listed as from PurN.
[b]Number in parentheses is the number of different DNA sequences found coding for a particular protein sequence (ignoring point mutations for the shuffled library).
[c]Some members had point mutations.

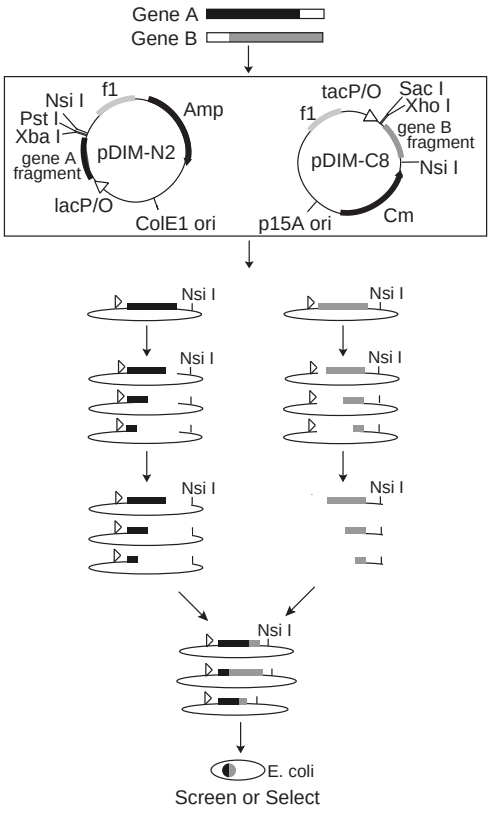

Figure 1. ITCHY combinatorial protein engineering. A large 5′ fragment of gene A is cloned into phagemid pDIM-N2, and a large 3′ fragment of gene B is cloned into phagemid pDIM-C8. pDIM-N2 and pDIM-C8 contain different antibiotic resistance genes (*Amp*, *Cm*) and different origins of replication (ColE1 ori and p15A ori), so that both can be maintained in the same cell. The phagemids also have restriction sites designed for creating incremental truncation libraries from the 3′ end of the gene fragment in pDIM-N2 and the 5′ end of the gene fragment in pDIM-C8. Incremental truncation is followed by digestion with restriction enzyme *NsiI*, isolation of the indicated DNA fragments, and ligation to form a library of gene fusions.

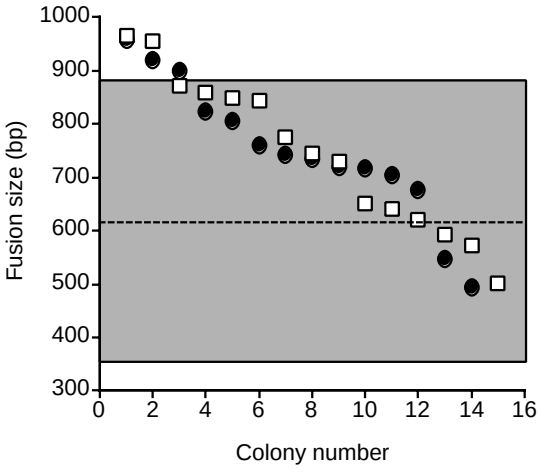

Figure 2. Size distribution of libraries. The sizes of the gene fusions of randomly selected members of IT-A (□) and IT-B (●) were estimated by gel electrophoresis and arranged in descending size order. The shaded area represents the theoretical size range based on the deletion of 1–270 bases of each fragment. Fusions larger than the desired size range result from fusion of gene fragments in which truncation has stopped in the approximately 30 bp spacer between the start of truncation and the gene to be truncated. The dashed line indicates the size of hybrid genes that are fused where their parents' sequences align.

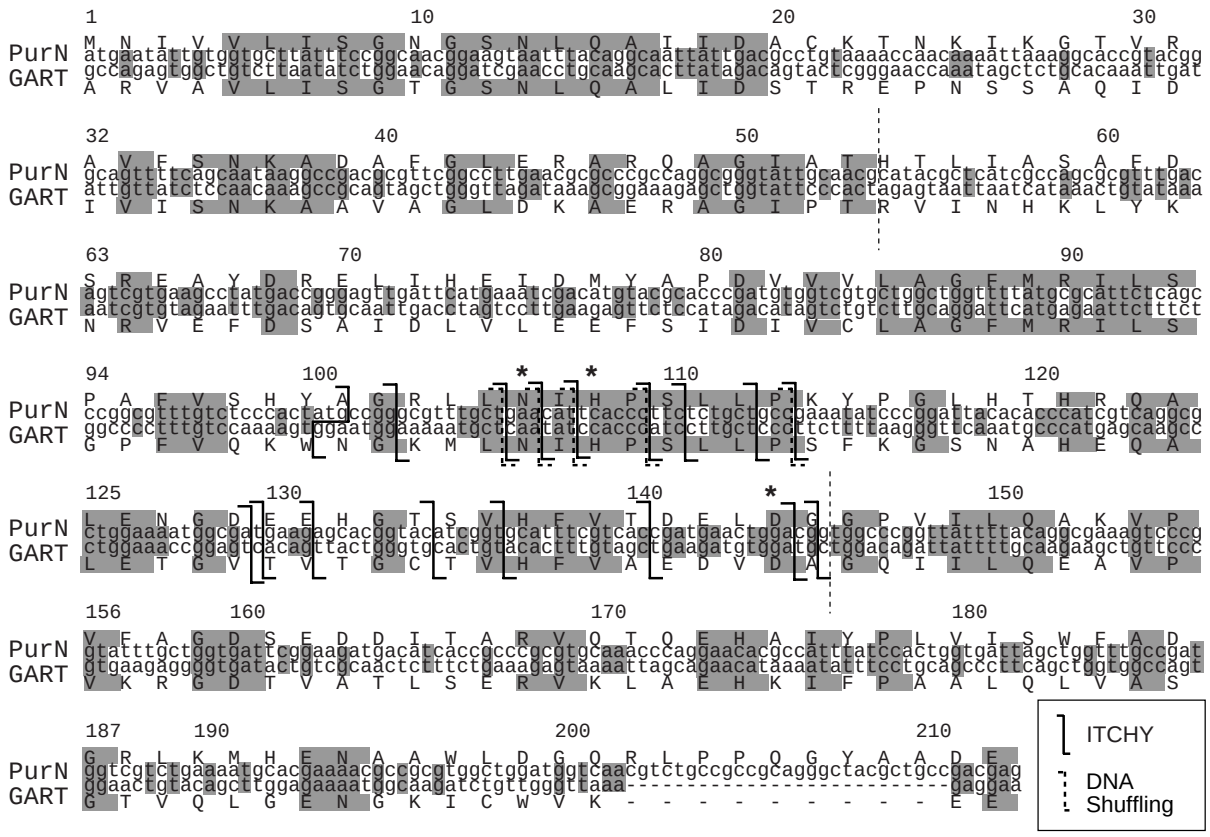

Figure 3. Fusion points of active PurN–GART hybrids relative to the alignment of PurN and GART. Crossovers of active fusions found by ITCHY are shown by a solid line and those found by DNA shuffling are shown by a dashed line. To the left of the crossover point, the fusion has sequence from PurN. To the right of the crossover point, the fusion has sequence from GART. If the crossover occurred in a region of DNA identity, the exact fusion point could not be assigned and is shown at the 3′ end of the region of identity. Regions of sequence identity are indicated in gray. *Three key active site residues. The region chosen to search for active fusions is shown between the long dashed lines.

ate SH-A, limiting crossovers to points between amino acids 54 and 144. A low annealing temperature for the reassembly PCR was necessary to create fusions between these genes. From a library of 2.4×10^5 transformants, 10 random, naive library members were selected and sequenced to evaluate the diversity of crossover points. Although this DNA shuffling procedure could, in theory, create hybrids with any odd number of crossovers, only single crossovers were found, ostensibly because of the low homology between the two genes and the relatively short (270 bp) segment of overlap. Of the 10 randomly selected, naive library members, three had crossovers at base 320 of *purN*, one at base 371, and five at base 379. In addition, one crossover was the result of mispriming between nonaligned sequences to produce a hybrid gene shorter than either of the two parent genes. This lack of diversity was not surprising considering that the two gene fragments have only 47% identity in the region of overlap and have an average length of identity of less than two bases. The fusion resulting from mispriming illustrates that an annealing temperature of 43°C is a reasonable compromise between having a low enough annealing temperature to allow for the creation of crossovers at low homology and having a high enough annealing temperature to prevent extensive mispriming.

From complementation of TX680F' it was determined that ~7.5% of the members of library SH-A contained active fusions. The amino acid sequences of randomly selected active fusions were determined by DNA sequencing (Table 1). In contrast to the variety of fusions found by ITCHY, 24 of 24 active SH-A members were fused in one region of high homology (Fig. 3). Although five different fusion points in the DNA sequence were found, all of them resulted in fusion proteins with identical amino acid sequences. Of the 24 active members of SH-A, 50% had one or more missense mutations.

Kinetic characterization of active hybrids. Initial velocity studies were initiated to determine the kinetic parameters of representative active PurN–GART fusions of IT-A, IT-B, and SH-A (Table 2). Based on k_{cat}/K_m(GAR), all characterized fusions were found to have activities of at least 500-fold less than wild-type PurN. However, in a very simplistic view, one might expect the K_m of the hybrids to be similar to that of PurN (because the PurN fragment contains most of the residues important for binding) and the k_{cat} to be similar to that of GART (because all or part of the key active site residues derive from GART). Thus, the hybrids would be expected to have a k_{cat} of 4.9 s^{-1}, a k_{cat}/K_m(GAR) of 42×10^{-3} μM^{-1} s^{-1} and a k_{cat}/K_m(fDDF) of 0.4 μM^{-1} s^{-1}. These values are only 5-fold, 28-fold, and 4-fold higher, respectively, than that of the most active fusion (IT-A5). The IT-A5 fusion was identified by ITCHY and was created by fusion at a region of nonidentity.

Table 2. Kinetic constants of selected PurN–GART fusions.

Fusion point	Enzyme	K_m (GAR) (μM)	K_m (fDDF) (μM)	k_{cat} (s^{-1})	K_{cat}/K_m (GAR) ($\mu M^{-1} \cdot s^{-1}$) X 10^3	k_{cat}/K_m (fDDF) ($\mu M^{-1} \cdot s^{-1}$)
	Wild type (human)	4.9 ± 0.4	1.0 ± 0.2	~4.9	~1,000	~4.9
102	IT-B12	3,800 ± 1,300	500 ± 280	~0.3	~0.08	~0.0006
113	IT-A8, SH-5	2,300 ± 500	116 ± 19	~2.1	~0.9	~0.02
129	IT-A5	584 ± 37	10 ± 1	~0.9	~1.5	~0.09
144	IT-A1	589 ± 56	50.5 ± 5.3	~1.0	~1.7	~0.02
	Wild type (*E. coli*)	118 ± 3	12.3 ± 1.3	90 ± 2	760	7.5

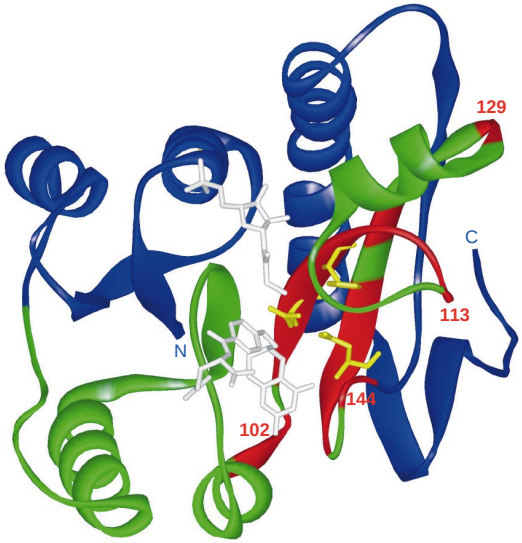

Figure 4. Fusion points of active PurN–GART hybrids mapped onto the structure of PurN[31]. The region searched for active fusions (residues 54–144) is shown in green, and the area outside the search region is shown in blue. The side chains of the three key active site residues, Asn106, His108, and Asp144, are shown in yellow. The substrate GAR (top) and a cofactor analog (5-deaza-5,6,7,8-tetrahydrofolate) are shown in white. The amino acids of PurN to which fusion of a GART fragment results in an active enzyme are shown in red. These include fusion points identified by sequencing and those that can be inferred given that an active fusion was found elsewhere in a region of amino acid identity (e.g., 104–113). The location of fusion points of hybrids characterized in Table 2 are indicated by the numbers.

We observed substrate inhibition by GAR with the wild-type human trifunctional enzyme at ≥15 μM, as has been observed for the GART domain alone[21]. Neither PurN nor any of the active PurN–GART fusions exhibited substrate inhibition by GAR, suggesting that the site for substrate inhibition does not lie solely on GART[100–203].

Discussion

We have demonstrated that ITCHY can create combinatorial libraries of genes in a manner that is independent of DNA sequence homology by identifying 10 active PurN–GART fusion proteins between N-terminal fragments of *E. coli* GAR transformylase PurN and C-terminal fragments of human GAR transformylase GART. Fusions of the N terminus of GART to the C terminus of PurN could have been explored in an analogous fashion. However, the optimum start of the GART domain may have been difficult to define, and nonnative *E. coli* codon usage at the N terminus of the fusions might have led to poor expression.

It appears that fusions within the active site may be less disruptive because of structural similarity within this region, whereas fusions distal to the active site, with lesser homology, may be more disruptive. All active fusions except one resulted from the crossover of precisely aligned genes. The only active crossover at a point of nonalignment was ITA-18, which had an extra glycine inserted between PurN 100 and GART 101. An otherwise identical fusion lacking the inserted glycine (constructed by overlap extension[25]) was found to be equally active by complementation of TX680F'. Insertions of a few amino acids[26,27] or even entire proteins[28,29] have been shown to be compatible with activity in other enzymes. Thus, we had expected that fusions with a few extra amino acids would not only be active, but that they would be the predominant active species.

There are two possible explanations for the predominance of crossovers at positions of precise alignment. First, the linear distances between conserved residues may have some importance for structure and/or function. This is supported by the fact that although *E. coli* and humans are evolutionarily distant, the linear distance between any two pairs of aligned residues is identical throughout protein. A second possibility is that the decrease in activity caused by extra amino acids is small, but this small decrease may be enough to prevent complementation of the auxotroph since k_{cat}/K_m is already reduced 500- to 10,000-fold for active fusions. Previous work with mutants of PurN suggested that the cutoff for successful complementation of the auxotroph appeared to be a difference in k_{cat}/K_m of somewhere between 2,000-fold and 12,000-fold when the mutants were expressed from the stronger T7 promoter[30]. It is conceivable that the insertion or deletion of residues in hybrids could prove advantageous for the engineering of other proteins.

Previous work suggested that residues 63 and 112 of PurN might be good choices as fusion points for creating hybrid enzymes[15]. Although fusion between PurN and GART at residue 112 proved to be active, no active fusions were found with a PurN fragment shorter than 1–100 residues. The lack of active fusions between residues 54 and 100 was confirmed with a third ITCHY library (IT-C) using fragments PurN[1–99] and GART[54–203]. Kinetic characterization of four active hybrids from IT-A and IT-B (Table 2) suggests that hybrids fused between residues 54 and 100 are not active because of weak binding of the substrate and cofactor, as the K_m values for both substrate and cofactor increase with the length of the GART in the fusion. The exception is that IT-A1 has a higher K_m(fDDF) than IT-A5, even though IT-A1 contains less of GART. Presumably, this is attributable to IT-A1 being fused in the 140–144 loop, which has been shown to have a role in binding the cofactor[31].

The ITCHY method enabled identification of a more diverse set of active chimeras than DNA shuffling, principally as a result of the relatively nonbiased and non-homology based method that creates the fusions. The active fusions identified by ITCHY demonstrate that crossovers between genes at regions of structural homology, irrespective of DNA sequence homology, are important for creating functional hybrid enzymes. The most active fusion (IT-A5) was fused at a nonhomologous location and was produced in *E. coli* at levels fourfold higher than those fusions found by DNA shuffling. Presumably, the better expression is due to efficient translation, efficient folding, or proteolytic stability, which are desired properties of an enzyme.

Although the library created by DNA shuffling had a higher frequency of positives, it was not very diverse. Approximately 30% (3/10) of the naive library and 100% (24/24) of the selected library had crossovers in the region of identity between amino acids 104 and 114, crossovers previously shown by ITCHY to create active hybrids. Outside of this region, the frequency of active hybrids must be ≥0.3% and may be zero, since none were found. Fused genes of ITCHY libraries could have been initially selected for size (e.g. the size of the original genes) resulting in an enrichment for active members of probably 10- to 100-fold to give a frequency of 0.1-1.0%.

DNA shuffling can create hybrids with multiple crossovers, whereas ITCHY libraries are limited to one crossover point per library member. One can envision an iterative method for ITCHY in order to create library members with multiple crossovers. However, as ITCHY libraries create all possible crossovers between two genes, DNA shuffling of ITCHY libraries should allow one to create a library of genes with multiple crossovers that include crossovers at regions of no homology, thus accessing a more diverse sequence space[16]. In a fashion analogous to DNA family shuffling, which improves directed evolution by accessing a more diverse yet functional sequence space[2], such a strategy should prove useful for the directed evolution of proteins. In addition, ITCHY libraries should have applications in the creation of novel enzymes by domain and subdomain swapping[32,33], as well as in the determination of structure/function relationships by characterizing hybrids of interspecies homologs[34].

Experimental protocol

Plasmids. Phagemids pDIM-N2 and pDIM-C6 have been described[15]. Phagemid pDIM-C8 is identical to pDIM-C6 except for the substitution of a *Bgl*II site for the *Bam*HI site and the substitution of a *Nsi*I site for a *Pst*I site 10 bp downstream from the *Spe*I site.

Incremental truncation for the creation of hybrid enzymes (ITCHY). Incremental truncation was performed essentially as described[15] with the following modifications. Supercoiled pDIM-N2 and pDIM-C8 were linearized by digestion with *Xba*I/*Pst*I and *Sac*I/*Xho*I, respectively. The ExoIII digestion was performed at 22°C in 60 µl of 66 mM Tris (pH 8.0), 0.66 mM $MgCl_2$, 100 mM NaCl. After inactivation of ExoIII and S1 nuclease, the ethanol-precipitated DNA was resuspended in 70 µl water. After the addition of 10 µl of 0.125 mM each dNTP, 2.5 U of Klenow fragment (in 2 mM Tris-HCl, 10 mM $MgCl_2$, pH 8.0) were added, and the mixture was incubated for 5 min at 37°C followed by heat inactivation of Klenow fragment at 72°C for 20 min. The DNA was digested with *Nsi*I (15 U) at 37°C for 2 h, and the desired fragments were isolated by gel electrophoresis using Elutrap (Schleicher & Schuell, Keene, NH), combined, and concentrated by ethanol precipitation. Ligation was carried out at 15°C overnight in a total volume of 20 µl using 6 Weiss units of T4 DNA ligase. The ligated DNA was desalted by ethanol precipitation into 30 µl of water and was electroporated into DH5α cells by six electroporations of 5 µl DNA each or into DH5α-E (Life Technologies, Rockville, MD) by two electroporations of 4 µl each. Libraries were recovered and stored as described[15].

DNA shuffling. DNA shuffling was performed as described[35] using the gene fragments PurN[1–144] and GART[54–203] with the following modifications: (1) the DNA was digested with 0.045 U of DNase for 1 min; (2) reassembly PCR used an annealing temperature of 43°C; and (3) the amplification of reassembled products (using a forward primer for the N terminus of the *purN* gene and a reverse primer for the C terminus of the *GART* gene) used the following program: 94°C for 5 min, 25 cycles of 1 min at 94°C, 1 min at 52°C, and 1 min at 72°C, followed by 7 min at 72°C. This protocol produced three visible bands. The correct-sized band was isolated by gel electrophoresis using Wizard PCR Preps (Promega, Madison, WI) and amplified further, as described above, by 10 additional cycles of PCR. All PCR was performed using *Taq* polymerase. The PCR product was ligated into vector pDIMN2 between the *Nde*I and *Spe*I sites, and the ligated plasmid was transformed into DH5α. The library was recovered and frozen in small aliquots.

Selection of active hybrids. Plasmid DNA of the ITCHY and DNA shuffling libraries were transformed into TX680F', recovered, and frozen as described above. In a 250 ml shake flask, 50 ml of 2× TY/Amp/Kan/0.2% glucose was inoculated with 10 µl of the frozen library (>10^8 c.f.u.) and grown at 37°C until OD_{600nm} = 0.2. Cells from 10 ml of culture were pelleted by centrifugation, washed once with 10 ml selective media[15], and resuspended in 2 ml selective media. After 2 h of shaking at 37°C, approximately 2.5×10^6 c.f.u. (rich media) were plated onto selective plates containing 0.3 mM isopropylthiogalactoside. Plates were incubated at 37°C for up to 48 h. Randomly chosen colonies were processed and sequenced, and complementation was verified as described[15].

Kinetic characterization. Kinetic measurements using GAR and fDDF were performed as described[36]. Wild-type *E. coli* PurN was prepared as described[31]. The PurN–GART fusions were prepared by the same method using the vector isolated from the positives (pDIM-N2) and TX680F' cells. Fusion concentrations were estimated by densitometry of SDS–PAGE separation of the most active gel filtration fraction. Purified GARS-AIRS-GART was a gift from L.T. Gooljarsingh (The Pennsylvania State University, University Park, PA).

Acknowledgments

We thank A.E. Nixon, C.P. Scott, and S.M. Firestine for many helpful discussions, and L.T. Gooljarsingh for purified GARS-AIRS-GART. This work was supported in part by National Institutes of Health Grant GM24129 (S.J.B.) and National Institutes of Health Postdoctoral Fellowship GM18560 (M.O.).

1. Stemmer, W.P.C. D.N.A shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91**, 10747–10751 (1994).
2. Crameri, A., Raillard, S.-A., Bermudez, E. & Stemmer, W.P.C. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288–291 (1998).
3. Zhao, H., Giver, L., Shao, Z. & Arnold, F.H. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol.* **16**, 258–261 (1998).
4. Shao, Z., Zhao, H., Giver, L. & Arnold, F.H. Random-priming *in vitro* recombination: an effective tool for directed evolution. *Nucleic Acids Res.* **26**, 681–683 (1998).
5. Joo, H., Lin, Z. & Arnold, F.H. Laboratory evolution of peroxide-mediated cytochrome P450 hydroxylation. *Nature* **399**, 670–673 (1999).
6. Stemmer, W.P.C. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370**, 389–391 (1994).
7. Arnold, F.H. & Moore, J.C. Optimizing industrial enzymes by directed evolution. *Adv. Biochem. Eng. Biotechnol.* **58**, 1–14 (1997).
8. Proba, K., Wörn, A., Honegger, A. & Plückthun, A. Antibody scFv fragments without disulfide bonds made by molecular evolution. *J. Mol. Biol* **275**, 245–253 (1998).
9. Giver, L., Gershenson, A., Freskgard, P.-O. & Arnold, F.H. Directed evolution of a thermostable esterase. *Proc. Natl. Acad. Sci. USA* **95**, 12809–12813 (1998).
10. Zhao, H. & Arnold, F.H. Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Eng.* **12**, 47–53 (1999).
11. Buchholz, F., Angrand, P.-O. & Stewart, A.F. Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat. Biotechnol.* **16**, 657–662 (1998).
12. Crameri, A., Whitehorn, E.A., Tate, E. & Stemmer, W.P.C. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* **14**, 315–319 (1996).
13. Zhang, J.-H., Dawes, G. & Stemmer, W.P.C. Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. *Proc. Natl. Acad. Sci. USA* **94**, 4504–4509 (1997).
14. Bogorad, L.D. & Deem, M.W. A hierarchical approach to protein molecular evolution. *Proc. Natl. Acad. Sci. USA* **96**, 2591–2595 (1999).
15. Ostermeier, M., Nixon, A.E., Shim, J.H. & Benkovic, S.J. Combinatorial protein engineering by incremental truncation. *Proc. Natl. Acad. Sci. USA* **96**, 3562–3567 (1999).
16. Ostermeier, M., Nixon, A.E. & Benkovic, S.J. Incremental truncation as a strategy in the engineering of novel biocatalysts. *Bioorg. Med. Chem.* **7**, 2139–2144 (1999).
17. Tomb, J.-F. & Barcak, G.J. Regulating the 3′-5′ activity of exonuclease III by varying the sodium chloride concentration. *Biotechniques* **7,** 932–933 (1989).
18. Wu, R. et al. Synchronous digestion of SV40 DNA by exonuclease III. *Biochemistry* **15**, 734–740 (1976).
19. Smith, J.M. & Daum, H.A. Identification and nucleotide sequence of a gene encoding 5′-phosphoribosylglycinamide transformylase in *Escherichia coli* K12. *J. Biol. Chem.* **262**, 10565–10569 (1987).
20. Aimi, J., Qiu, H., Williams, J., Zalkin, H. & Dixon, J.E. De novo purine nucleotide biosynthesis: cloning of human and avian cDNAs encoding the trifunctional glycinamide ribonucleotide synthetase-aminoimidazole ribonucleotide synthetase-glycinamide ribonucleotide transformylase by functional complementation in *E. coli*. *Nucleic Acids Res.* **18**, 6665–6672 (1990).
21. Caperelli, C.A. & Giroux, E.L. The human glycinamide ribonucleotide transformylase domain: purification, characterization, and kinetic mechanism. *Arch. Biochem. Biophys.* **341**, 98–103 (1997).
22. Kan, C.-C. et al.. Heterologous expression and purification of active human phosphoribosylglycinamide formyltrasnferase as a single domain. *J. Protein. Chem.* **11**, 467–473 (1992).
23. Varney, M.D. et al. Protein structure based design, synthesis, and biological evaluation of 5-thia-2,6-diamino-4(3*H*)-oxopyrimidines: potent inhibitors of glycinamide ribonucleotide transformylase with potent cell growth inhibition. *J. Med. Chem.* **40**, 2502–2524 (1997).
24. Bernstein, F.C. et al. The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol* **112**, 535–542 (1977).
25. Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. & Pease, L.R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51–59 (1989).
26. Hallet, B., Sherratt, D.J. & Hayes, F. Pentapeptide scanning mutagenesis: random insertion of a variable five amino acid cassette in a target protein. *Nucleic Acids Res.* **25**, 1866–1867 (1997).
27. Zebala, J. & Barany, F. Mapping catalytically important regions of an enzyme using two-codon insertion mutagenesis: a case study correlating β-lactamase mutants with the three-dimensional structure. *Gene* **100**, 51–57 (1991).
28. Betton, J.-M., Jacob, J.P., Hofnung, M. & Broome-Smith, J.K. Creating a bifunctional protein by insertion of β-lactamase into the maltodextrin-binding protein. *Nat. Biotechnol.***15**, 1276–1279 (1997).
29. Biondi, R.M., Baehler, P.J., Reymond, C.D. & Véron, M. Random insertion of GFP into the cAMP-dependent protein kinase regulatory subunit from *Dictyostelium discoideum*. *Nucleic Acids Res.* **26**, 4946–4952 (1998).
30. Warren, M.S., Marolewski, A.E. & Benkovic, S.J. A rapid screen of active site mutants in glycinamide ribonucleotide transformylase. *Biochemistry* **35**, 8855–8862 (1996).
31. Almassy, R.J., Janson, C.A., Kan, C.-C. & Hostomska, Z. Structures of apo and complexed *Escherichia coli* glycinamide ribonucleotide transformylase. *Proc. Natl. Acad. Sci. USA* **89**, 6114–6118 (1992).
32. Hopfner, K.-P. et al. New enzyme lineages by subdomain shuffling. *Proc. Natl. Acad. Sci. USA* **95**, 9813–9818 (1998).
33. Nixon, A.E., Warren, M.S. & Benkovic, S.J. Assembly of an active enzyme by the linkage of two protein modules. *Proc. Natl. Acad. Sci. USA* **94**, 1069–1073 (1997).
34. Nixon, A.E., Ostermeier, M. & Benkovic, S.J. Hybrid enzymes: manipulating enzyme design. *Trends Biotechnol.* **16**, 258–264 (1998).
35. Zhao, H. & Arnold, F.H. Optimization of DNA shuffling for high fidelity. *Nucleic Acids Res.* **25**, 1307–1308 (1997).
36. Shim, J.H. & Benkovic, S.J. Evaluation of the kinetic mechanism of *Escherichia coli* glycinamide ribonucleotide transformylase. *Biochemistry* **37**, 8776–8782 (1998).