A Study in Molecular Contingency: Glutamine Phosphoribosylpyrophosphate Amidotransferase is a Promiscuous and Evolvable Phosphoribosylanthranilate Isomerase

Wayne M. Patrick¹,² and Ichiro Matsumura¹*  

¹Department of Biochemistry, Center for Fundamental and Applied Molecular Evolution, Emory University, Atlanta, GA 30322, USA  
²Institute of Molecular Biosciences, Massey University, Auckland 0745, New Zealand  
Received 18 October 2007; received in revised form 14 January 2008; accepted 16 January 2008  
Available online 26 January 2008  

The prevalence of paralogous enzymes implies that novel catalytic functions can evolve on preexisting protein scaffolds. The weak secondary activities of proteins, which reflect catalytic promiscuity and substrate ambiguity, are plausible starting points for this evolutionary process. In this study, we observed the emergence of a new enzyme from the ASKA (A Complete Set of E. coli K-12 ORF Archive) collection of Escherichia coli open reading frames. The overexpression of (His)₆-tagged glutamine phosphoribosylpyrophosphate amidotransferase (PurF) unexpectedly rescued a ΔtrpF E. coli strain from starvation on minimal media. The wild-type PurF and TrpF enzymes are unrelated in sequence, tertiary structure and catalytic mechanism. The promiscuous phosphoribosylanthranilate isomerase activity of the ASKA PurF variant apparently stems from a preexisting affinity for phosphoribosylated substrates. The relative fitness of the (His)₆-PurF/ΔtrpF strain was improved 4.8-fold to nearly wild-type levels by random mutagenesis of purF and genetic selection. The evolved and ancestral PurF proteins were purified and reacted with phosphoribosylanthranilate in vitro. The best evolvant (kcat/KM = 0.3 s⁻¹ M⁻¹) was ~25-fold more efficient than its ancestor but >10⁷-fold less efficient than the wild-type phosphoribosylanthranilate isomerase. These observations demonstrate in quantitative terms that the weak secondary activities of promiscuous enzymes can dramatically improve the fitness of contemporary organisms.

© 2008 Elsevier Ltd. All rights reserved.

Keywords: catalytic promiscuity; contingency; directed evolution; enzyme evolution; multicopy suppressor

Introduction

New enzymatic functions arise as organisms adapt to environmental changes, but the biochemical mechanisms that underlie these adaptations remain incompletely understood. Well-studied examples include the evolution of a multistep pathway for the catabolism of the s-triazine herbicides¹ and of a phosphotriesterase to degrade organophosphate insecticides such as paraoxon.² In these cases, adaptive molecular evolution was rapid: the s-triazines and the organophosphates were first synthesized and introduced into the environment in the mid-20th century. Biochemical and genetic investigations of these systems have shown that novel catalytic activities can evolve on preexisting protein scaffolds. Nonetheless, these ground-breaking examples offer...
The substrate ambiguity and catalytic promiscuity of preexisting enzymes have garnered attention as plausible starting points for the evolution of novel functions, especially when mutations that cause the overexpression of promiscuous enzymes (including gene duplication events) amplify weak secondary activities to physiologically relevant levels (Fig. 1). In a previous study, we investigated this model with the ASKA (A Complete Set of *E. coli* K-12 ORF Archive) library, which contains each *Escherichia coli* open reading frame (ORF) cloned into the expression vector pCA24N. Aliquots of each plasmid were pooled and used to transform 104 conditional auxotrophs from the Keio collection of single-gene deletion strains. The colonies that formed under selective (minimal medium) conditions were propagated; sequence analysis of the associated plasmids showed that 21 auxotrophies were specifically suppressed by the overexpression of 41 mostly nonhomologous *E. coli* genes. These results showed that many proteins are multifunctional and demonstrated an easy way to set up directed evolution experiments.

Here, we focus upon a particularly compelling case of catalytic promiscuity. We chose the isomerization of phosphoribosylanthranilate (PRA) to 1′-(2′-carboxyphenylamino)-1′-deoxyribulose 5′-phosphate (CdRP), an Amadori rearrangement in the biosynthesis of tryptophan (Fig. 2a), as our model reaction. *In vitro*, the reaction occurs spontaneously under mild but nonphysiological conditions (50% v/v ethanol, room temperature). *In vivo*, it is ordinarily catalyzed by PRA isomerase (PRAI, EC 5.3.1.24), a well-characterized (βα)₈-barrel protein (Fig. 3a) that is the product of the *trpF* gene. In turn, the product of the reaction, CdRP, is the substrate for the *trpC* gene product indoleglycerol phosphate (InGP) synthase (IGPS). We did not consider PRAI activity in our previous study because the Keio knockout collection does not include a Δ*trpF* allele.

(a) Over-expression

(b) Adaptive evolution

---

**Fig. 1.** A two-stage model for the evolution of novel catalytic functions. (a) A promiscuous catalyst has evolved to recognize the diamond-shaped substrate but also has some low affinity for the alternate substrate. In a changing environment, it is activity against this alternate substrate that confers a selective advantage. Therefore, a regulatory mutation or gene amplification event leading to overexpression of the promiscuous catalyst confers fitness improvement. (b) Under continued selection pressure, the promiscuous catalyst evolves specificity for the new substrate (ultimately, at the expense of the ancestral activity).

**Fig. 2.** The reactions catalyzed by the isomerase, PRAI, and the transferase, PurF. (a) PRAI catalyzes the Amadori rearrangement of the aminoaalose PRA to the amino-ketose CdRP. The common RSP moiety recognized by each enzyme is boxed. (b) PurF catalyzes the transfer of a glutamine amide nitrogen to PRPP, yielding 5-phosphoribosyl-1-amine, glutamate and pyrophosphate.
In E. coli, PRAI forms the C-terminal domain of a bifunctional IGPS:PRAI enzyme; however, the IGPS and PRAI domains can be separated genetically and expressed as stable monomeric proteins, each with essentially wild-type catalytic activity.

PRA isomerization is a good model reaction for several reasons. First, the frequent occurrence of (βα)_8-barrel proteins within the protein database suggests that this scaffold may be particularly evolvable. Second, tryptophan is the least abundant amino acid, so only modest levels of enzymatic activity are sufficient to suppress tryptophan auxotrophies. Third, in vitro assays of tryptophan biosynthetic enzymes have already been developed. Most importantly, previous directed evolution studies of this activity provide standards of comparison. Sterner and his co-workers probed the functional relationships between PRAI and the homologous (βα)_8-barrel proteins of histidine biosynthesis. They showed that single point mutations can impart PRAI activity on the HisA protein, which catalyzes an analogous Amadori rearrangement, and on the HisF protein, which normally catalyzes the subsequent synthesis of the imidazole ring. Here, we searched the ASKA library for a promiscuous PRAI and then enhanced its activity by directed evolution. We describe the structure and function of an alternative scaffold for PRAI activity, which began as a completely unrelated enzyme from the de novo purine biosynthesis pathway.

**Results**

**Identification of a promiscuous isomerase**

Our first objective was to identify preexisting E. coli proteins with promiscuous PRAI activity. We expected that IGPS:PRAI, HisA and/or HisF would rescue ΔtrpF E. coli cells; we also thought it possible that other (βα)_8-barrel proteins (i.e., distantly related paralogues) would effect rescue. We transformed the auxotrophic E. coli strain JMB9ΔtrpF with the 5272 pooled plasmids of the ASKA library. Transformed cells were plated on minimal medium lacking tryptophan but supplemented with isopropyl-β-D-thiogalactoside (IPTG; 50 μM) to induce the overexpression of each ASKA-encoded gene. After 24 h of incubation at 30 °C, a number of colonies had appeared on the selection plates. As expected, these all harbored the plasmid carrying the bifunctional trpCF gene (coding for IGPS:PRAI) and therefore served as an internal positive control for our sampling of the ASKA library.

More interestingly, 6 colonies were identified after 6 days of growth, and 4 more were picked on day 10 of the selection experiment. DNA sequencing showed that each of these 10 clones contained the ASKA purF expression plasmid, encoding glutamine 5-phosphoribosyl-1-pyrophosphate (PRPP) amidotransferase (EC 2.4.2.14) fused to an

![Image](https://example.com/image.png)
N-terminal (His)₆ tag and a C-terminal green fluorescent protein (GFP) tag. This homotetrameric two-domain enzyme (Fig. 3b) normally catalyzes the first committed step of de novo purine biosynthesis, the conversion of PRPP to phosphoribosylamine (Fig. 2b). Moreover, we subsequently discovered that the two observed phenotypes (6-day growth and 10-day growth) corresponded to two genotypes. The faster-growing clones contained a missense mutation (A592G) resulting in the replacement of Ile198 with Val, while the slower-growing clones had a G587A mutation and therefore encoded (His)₆-PurF(R196H)-GFP.

During construction of the ASKA library, four independent clones containing each ORF were pooled, stored and subsequently distributed to the independent clones containing each ORF were pooled, stored and subsequently distributed to the public. In this case, further sequencing confirmed that the ancestral ASKA ‘clone’ was in fact a heterogeneous mixture of the purF(I198V) and purF(R196H) mutants; that is, these mutations did not arise in the course of our selection experiment. Nonetheless, we constructed the corrected pCA24N-purF plasmid and showed that it enabled growth on minimal selection medium at the same rate as did pCA24N-purF(I198V). We therefore chose to focus on (His)₆-PurF(I198V), with its apparently neutral substitution, rather than the impaired variant (His)₆-PurF(R196H). The pCA24N-encoded C-terminal GFP tag was also removed from this clone in order to avoid interference in downstream absorbance assays; this modification had no effect on its ability to rescue E. coli JMB9ΔtrpF.

Fresh E. coli JMBΔtrpF cells were retransformed with pCA24N-purF(I198V) and struck on minimal selection medium to verify the phenotype. Evenly sized colonies appeared at a uniform rate, in a fashion identical with the initial selection experiment from the ASKA library. The lack of stochastic variation in cell growth suggested that gene amplification and point mutation of chromosomal genes (adaptive mutability) were not contributing to the observed phenotype. Instead, the results were consistent with a genuine example of catalytic promiscuity.

(His)₆-PurF(I198V) possesses detectable PRAI activity

PRAI and PurF are not related in sequence or structure (Fig. 3), but both enzymes have evolved to recognize substrates containing a ribose-5-phosphate (R5P) moiety (Fig. 2). We hypothesized that this shared substrate affinity was the basis of (His)₆-PurF’s promiscuous PRAI activity. However, it was also possible that the overexpressed (His)₆-PurF protein was somehow enhancing the flux of an alternative pathway for tryptophan biosynthesis or otherwise bypassing the need for a functional PRAI in vivo. Indeed, the remarkable robustness of the E. coli metabolic network to perturbations such as single-gene deletions is largely ascribed to altering fluxes through alternate metabolic pathways. We therefore sought to demonstrate PRAI activity for the purified (His)₆-PurF(I198V) protein in vitro.

The slow growth of E. coli JMB9ΔtrpF harboring pCA24N-purF(I198V), even when the rescuing protein was overexpressed, suggested that its PRAI activity was low. Unfortunately, the substrate of PRAI, PRA, is extremely prone to spontaneous and irreversible hydrolysis into R5P and anthranilate under conditions conducive to in vitro enzyme assays (T > 4 °C; pH < 8.6; exposure to light in a spectrophotometer). Consequently, we turned to a coupled endpoint assay (Supplementary Fig. S1) and high-performance liquid chromatography (HPLC) resolution of the reaction products to demonstrate catalytic turnover by purified (His)₆-PurF(I198V). In order to synthesize sufficiently high concentrations of PRA from anthranilate and PRPP (the upstream substrates) in situ, we began by cloning the anthranilate phosphoribosyltransferase (PRTase) enzyme from Acinetobacter baylyi strain ADP1 (AcTrpD). This enzyme, which will be described in greater detail elsewhere, proved amenable to overexpression in E. coli and was highly active at the assay temperatures employed. The PRA isomerization reaction was also coupled to the formation of the stable chromophore InGP (the downstream product) by the addition of purified IGPS to the assay.

The appropriate substrates (PRPP and anthranilate) and coupling enzymes (AcTrpD and IGPS) were incubated in the presence or absence of (His)₆-PurF(I198V) for 16 h. As expected, separation of the assay mixture by reversed-phase HPLC showed that the detectable product of spontaneous PRA hydrolysis, anthranilate, predominated (Fig. 4). Furthermore, the high pH of the assay buffer (pH 8.6) led to spontaneous conversion of PRA to CdRP (and therefore to InGP, the downstream product, via IGPS catalysis) at a low but detectable rate. However, the presence of (His)₆-PurF(I198V) led to increased InGP formation (Fig. 4, inset), with a concomitant decrease in the size of the anthranilate peak. Integration analysis suggested that InGP had accumulated to a final concentration of 13 μM in the negative control reaction and to that of 17 μM in the reaction containing (His)₆-PurF(I198V). The difference, 4 μM, corresponded closely to the concentration of (His)₆-PurF(I198V) present in the assay (5 μM). The implication was that the (His)₆-PurF(I198V) protein only modestly accelerated product formation over the rate of spontaneous substrate hydrolysis, which occurred on the order k−1 h−1.

Selection of improved variants

Some promiscuous activities of enzymes are evolvable, so we attempted to improve the PRAI activity of (His)₆-PurF(I198V) by directed evolution. In the first round of mutagenesis, we subjected the 1515-bp purF(I198V) gene to error-prone PCR (epPCR). The effective size of the resulting library (from which a vector-derived background of <3% had been subtracted) was 6.4 × 10⁷ cloned variants. A total of 7549 bp of DNA sequence was obtained from randomly selected members of the library. This allowed us to confirm a reasonably unbiased spec-
trum of mutations (Supplementary Table S1) and to estimate the mean mutation rate at 15.5 mutations per variant, or 0.010 bp$^{-1}$. Analysis of the library’s composition using PEDEL$^{35,36}$ and the ‘PCR distribution’ described by Drummond et al.$^{37}$ showed that the library was maximally diverse. Fewer than 200 clones in the library were expected to be redundant, and most of these (∼150) were predicted to be the unmutated template, purF(I198V).

*E. coli* JM9ΔtrpF cells were transformed with the library. The resulting transformants were propagated on selective minimal media; after only 48 h of incubation at 30 °C, 40 colonies had appeared on the selection plates. A range of growth rates, scored on an arbitrary scale from 1 to 5, were observed when the clones were restreaked on selection medium (examples are shown in Supplementary Fig. S2). The fastest-growing clones formed colonies in less than 30 h. Sequencing showed that eight unique variants had been selected, one of which, purF(1–04), dominated the pool with 22 occurrences (Table 1). We also sequenced a variant, 1-Neg, that was not able to complement the tryptophan auxotrophy of *E. coli* JM9ΔtrpF cells (Table 1). This protein was solubly expressed and possessed a number of mutations similar to the selected variants; it thus suggests specificity in the activities of the selected variants.

The selected variants each contained 6–14 base substitutions, encoding 5–11 amino acid changes (in addition to I198V from the template; Table 1). The mutations were found throughout the tertiary structure of PurF (Fig. 5). Most notably, the three fittest variants (1–01, 1–04 and 1–16) all contained mutations at Asn328, while four of the other mutants (1–22, 1–26, 1–27 and 1–30) shared an N352K mutation. In the native PurF structure, Asn328 and Asn352 help anchor the ends of a long flexible loop that closes over the PRTase active site when PRPP is bound (Fig. 6).$^{20,38}$ The recurrence of mutations at Asn328 and Asn352 suggested that the reorganization of this loop, presumably to accommodate the bulkier anthranilate moiety of PRA (Fig. 2), was critical for improving the promiscuous PRAI activity of (His)$_6$-PurF.

![Fig. 4. HPLC profiles demonstrating the PRAI activity of (His)$_6$-PurF(I198V) in a coupled endpoint assay, monitored at 278 nm. The major peak eluting at 5.7 mL corresponds to anthranilate, which is one of the products of PRA hydrolysis. The peak eluting at 4.4 mL is an unidentified highly absorbing contaminant of the AcTrpD substrate PRPP. The inset shows the enzyme-catalyzed formation of InGP, the downstream product of PRAI and the coupled IGPS enzyme, over and above the spontaneous background rate.](image-url)
A second round of directed evolution

The fittest variant from the epPCR library, (His)_{6} PurF(1–04), had to be overexpressed to rescue *E. coli* JMB9ΔtrpF. We therefore attempted to improve its activity further, by random mutagenesis of pCA24N-purF(1–04) in the *E. coli* mutator strain XL1-Red. The mutagenized plasmid DNA was recovered from a saturated culture of *E. coli* XL1-Red and used to transform *E. coli* JMB9ΔtrpF. The selection yielded four clones that appeared to perform better than clone 1–04; however, only one of these, 2–02, displayed a reproducible phenotype when fresh cells were transformed and restreaked. Cells harboring pCA24N-purF(2–02) formed colonies on minimal selection medium (containing 50 μM IPTG) in 17 h at 30 °C. This represented an apparent twofold improvement in fitness over pCA24N-purF(1–04). Cells expressing purF(2–02), unlike those expressing purF (1–04), were also viable when the IPTG concentration was lowered to 5 μM, although not when IPTG was omitted from the selection medium altogether. In contrast, leaky expression from pCA24N-trpCF was sufficient for rapid growth, even in the absence of IPTG.

Despite the improved growth rate of cells expressing PurF(2–02), sequence analysis showed that the purF(1–04) and purF(2–02) ORFs were identical. Further sequencing of the lacP alleles and P_{TS}/lacO regions of pCA24N-purF(1198V), -purF(1–04) and -purF(2–02) revealed a point mutation (C → T) in the promoter region of the latter clone. While the mutation fell between the lac operator site and the -10 region of the promoter, it nonetheless appeared to destabilize the LacI^q/lacO interaction. Therefore, the improvement in fitness of this clone seems to result from an optimization of transcription.

**Kinetic characterization of (His)_{6}-PurF(1–04)**

The evolved (His)$_6$-PurF(1–04) protein was purified, and its PRAI activity was measured *in vitro* by the 16-h coupled assay described above. The absorbance spectrum of the completed reaction showed clear evidence for the production of InGP; this was confirmed by HPLC and further spectroscopic analysis of the eluted fraction containing the product (Fig. 7). Peak integration indicated that (His)$_6$-PurF (1–04) had catalyzed the formation of 110 μM InGP over the course of the incubation. This value represents the conversion of 5.5% of the available substrate to the product and corresponds to 22 turnovers per

---

**Fig. 5.** Mutations in variants selected from the epPCR library mapped onto the tertiary structure of PurF (PDB entry 1ec). Residues in yellow are mutated in at least one variant. The position mutated in the parental clone Ile198 and the sites of recurrent mutations in the PRTase flexible loop (Asn328 and Asn352) are highlighted in red.

**Fig. 6.** Asn328 and Asn352 anchor the PRTase flexible loop in (a) the open structure of PurF (PDB entry 1ec) and (b) the closed substrate-bound structure (PDB entry 1ecc). The two asparagine residues are highlighted in red, with the loop between them highlighted in green or blue. The PRPP analogue cPRPP (1α-pyrophosphoryl-2α,3α-dihydroxy-4β-cyclopentanemethanol 5-phosphate) is shown in ball-and-stick representation in (b).
Fig. 7. Characterization of the PRAI activity of (His)$_6$-PurF(1–04). (a) After a 16-h incubation, the absorbance spectrum of the products of the (His)$_6$-PurF(1–04) reaction shows evidence for the appearance of InGP (i.e., an increase in absorbance around 280 nm) and the disappearance of the AcTrpD substrate anthranilate ($A_{\text{max}} = 310$ nm). (b) HPLC separation of the reaction products. The peak corresponding to InGP elutes at ~5.2 mL. (c) Absorbance spectra of the InGP-containing fractions from (b), normalized with respect to concentration. The spectra match the known spectrum of InGP,$^{26}$ confirming that both PRAI and (His)$_6$-PurF(1–04) are catalyzing the expected isomerization reaction.
PurF(1–04) active site. This result suggested that the PRAI activity of (His)\(_e\)-PurF(1–04) was 25- to 30-fold higher than that of (His)\(_e\)-PurF(I98V).

We next attempted to measure the kinetic parameters of the (His)\(_e\)-PurF(1–04)-catalyzed reaction. A sensitive assay for PRAI activity has been described,\(^{21}\) but it is limited to substrate concentrations \(\leq 0.8\) mM\(^{26}\) because of the extremely high fluorescence quantum yields of anthranilate (the upstream substrate) and PRA. Instead, we monitored the formation of InGP via its absorbance at 278 nm. The contribution of anthranilate to the measured fluorescence quantum yields of anthranilate (the upstream substrate) was accounted for by running controls without added (His)\(_e\)-PurF(1–04), in parallel with every assay, at all substrate concentrations. These controls also allowed any observed rate of spontaneous PRA-to-CdRP conversion to be subtracted from the enzyme-catalyzed rate. Consistent with the HPLC-based endpoint assays (Fig. 7b), this rate of spontaneous isomerization was negligible. It was not possible to saturate the enzyme in the assays; therefore, \(k_{\text{cat}}/K_M\) was estimated from the slope of the linear fit when the initial velocity was plotted against substrate concentration (Supplementary Fig. S3). Assays of three separate batches of (His)\(_e\)-PurF(1–04) allowed us to estimate \(k_{\text{cat}}/K_M = 0.3\pm 0.2\) s\(^{-1}\) M\(^{-1}\). This value is 2.3 \times 10\(^7\)-fold below that for the wild-type E. coli PRAI (\(k_{\text{cat}}/K_M = 6.8\times 10^8\) s\(^{-1}\) M\(^{-1}\)).\(^{21}\)

Our average yield of purified (His)\(_e\)-PurF(1–04) was 19 mg L\(^{-1}\) of culture. Based on the cell density of the culture at the end of the induction (and the rather optimistic assumption that our purification was 100% efficient), it can be estimated that every overexpressing cell contained \(> 1.8\times 10^9\) molecules of (His)\(_e\)-PurF(1–04). Assuming a cytoplasmic volume of 1.25 fl, this corresponds to an intracellular concentration of \(> 0.24\) mM. The total tryptophan concentration in a healthy E. coli cell is estimated to be 10 mM.\(^{25}\) It appears that the very poor PRAI activities of (His)\(_e\)-PurF(I98V) and (His)\(_e\)-PurF(1–04) were overcome by synthesis of near-stoichiometric quantities of the catalysts. This energetically costly solution leaves significant room for improvement through further adaptive evolution.

**Specificity of the promiscuous PRAI activity**

The weak promiscuous activities of (His)\(_e\)-PurF(I98V) and (His)\(_e\)-PurF(1–04) raised the possibility that our assays were contaminated with trace amounts of PRAI or, indeed, some other E. coli enzyme possessing a stronger promiscuous PRAI activity. Several lines of evidence argue against this alternative explanation. First, all of the enzymes in the assays were purified from E. coli strains in which the chromosomal copy of trpF had been deleted. Second, the \(K_M\) of E. coli PRAI for PRA is 4.7 \mu M.\(^{21}\) Our observation that (His)\(_e\)-PurF(1–04) was not saturated at substrate concentrations of up to 2 mM strongly suggests that we were not measuring the activity of a PRAI contaminant. Third, we observed a wide range of phenotypes (i.e., growth rates) in strains that were isogenic, except for the presence or absence of mutations in the plasmid-encoded purF gene. In our initial selection, the conservative substitution R196H slowed colony formation from 6 to 10 days. (His)\(_e\)-PurF(1–04) contained eight substitutions and led to colony formation in \(< 30\) h, yet our negative control, I-Neg (Table 1), carried a similar mutational load in purF but did not complement *E. coli* JMB94ΔtrpF at all. These sequence variations affect enzyme activity but not protein folding or the chemical composition of the intracellular milieu and therefore provide compelling evidence that (His)\(_e\)-PurF(I98V) and (His)\(_e\)-PurF(1–04) are producing the observed PRAI activities, *in vitro* and *in vivo*.

**Relative fitness corresponds to *in vitro* activity**

The measured catalytic activity of (His)\(_e\)-PurF(1–04) was extremely low, yet overexpression of this protein led to colony formation at a rate that was comparable with the growth of wild-type cells. In order to relate enzyme activity to cellular phenotype, we determined the relative fitness values (W) of cells expressing (His)\(_e\)-PurF(I98V), (His)\(_e\)-PurF(1–04) and the bifunctional IGPS:PRAI enzyme from *E. coli* (Table 2). The GFP tags on the (His)\(_e\)-PurF(I98V) and (His)\(_e\)-IGPS:PRAI clones from the ASKA library allowed us to conduct pairwise competition experiments with one tagged clone and one untagged clone, in turn enabling relative growth rates to be measured in a standard 24-h assay.\(^{31}\)

In the first experiment, we showed that removing the GFP tag from (His)\(_e\)-PurF(I98V) had no effect on the fitness of the strain; that is, the fitness of the GFP-tagged (His)\(_e\)-PurF(I98V) clone relative to the untagged clone is not significantly different from 1 (Table 2). As expected, a second experiment showed that cells harboring (His)\(_e\)-PurF(I98V) were rapidly outcompeted by those expressing (His)\(_e\)-IGPS:PRAI (W = 4.8). Interestingly, the growth rate of the (His)\(_e\)-PurF(I98V) strain was increased substantially in the presence of (His)\(_e\)-IGPS:PRAI-expressing cells (from <1 to almost 2 divisions in 24 h; Table 2), consistent with cross-feeding of a downstream intermediate (possibly membrane-permeable indole) and/or tryptophan itself, from the latter strain to the former. This cross-feeding was not observed when (His)\(_e\)-PurF(1–04)-expressing cells were grown in

**Table 2. Relative fitness of strains overexpressing (His)\(_e\)-PurF(I98V), (His)\(_e\)-PurF(1–04) and (His)\(_e\)-IGPS:PRAI**

<table>
<thead>
<tr>
<th>Competition experiment(^a)</th>
<th>No. of cell divisions(^b)</th>
<th>Relative fitness(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PurF(I98V) versus PurF(I98V)</td>
<td>0.49 0.45</td>
<td>1.1±0.1</td>
</tr>
<tr>
<td>IGPS:PRAI versus PurF(I98V)</td>
<td>8.8 1.9</td>
<td>4.8±0.2</td>
</tr>
<tr>
<td>PurF(1–04) versus PurF(I98V)</td>
<td>1.8 0.37</td>
<td>4.8±0.7</td>
</tr>
</tbody>
</table>

\(^a\) The GFP-tagged protein is indicated in boldface.

\(^b\) Mean number of cell divisions over the course of the 24-h competition experiment (n = 3).

\(^c\) Determined from the relative growth rates of the two strains. Values are expressed as mean±SE (n = 3).
competition with the (His)$_6$-PurF(I198V) reference strain. While the (His)$_6$-PurF(1–04) strain did not grow as rapidly as (His)$_6$-IGPS:PRAI-expressing cells, (His)$_6$-PurF(1–04) was therefore estimated to be as fit as (His)$_6$-IGPS:PRAI, relative to (His)$_6$-PurF (I198V) (Table 2). Overall, these experiments show in quantitative terms that marginal enzyme activities can dramatically improve organismal fitness.

**Discussion**

**An unexpected source of PRAI activity**

We initially sought $\Delta$trpF suppressors because we expected a particular outcome. It seemed reasonable that HisA (EC 5.3.1.16), a homologous ($\beta\alpha$)$_6$-barrel that catalyzes the same isomerization reaction (albeit on a slightly different substrate), would possess the requisite promiscuous activity when overexpressed. Instead, our results showed that overexpression of (His)$_6$-PurF(I198V), an amidotransferase (EC 2.4.2.14) from the purine anabolic pathway, could suppress the tryptophan auxotrophy of the $\Delta$trpF strain. The PurF and PRAI proteins are unrelated in structure (Fig. 3), but the two enzymes normally recognize broadly similar substrates (Fig. 2). Purified (His)$_6$-PurF(I198V) exhibits a very modest but detectable level of PRAI activity in vitro (Fig. 4). A variant with 25–30-fold greater activity emerged after a single round of random mutagenesis and selection (Fig. 7).

We are not the first to observe functional relationships between the pathways of tryptophan and purine biosynthesis. Despite being structurally unrelated, the enzyme that precedes PRAI in tryptophan biosynthesis (TrpD) acts via the same mechanism as PurF to form PRA from PRPP and anthranilate.42 Furthermore, Downs et al. showed that expression of the TrpD:TrpE complex can rescue purf mutants of Salmonella enterica.43,44 The rescuing activity was improved in strains containing trpF mutations, suggesting that increased TrpD:TrpE expression, combined with decreased PRAI activity (through mutation), led to accumulation of the labile intermediate PRA. Spontaneous hydrolysis of PRA gives R5P, which is free to react with NH$_3$ in a nonenzymatic synthesis of the PurF product phosphoribosylaminel.45 With these results in mind, we considered a scenario in which overexpressed PurF was rescuing *E. coli* JMB9$\Delta$trpF by catalyzing the TrpD reaction, thereby leading to hyperaccumulation of PRA for its spontaneous isomerization to CdR5P. However, neither the *E. coli* nor the *A. baylyi* TrpD could rescue *E. coli* JMB9$\Delta$trpF when overexpressed (data not shown).

Instead, our *in vitro* assays (Figs. 4 and 7) showed that (His)$_6$-PurF(I198V) and (His)$_6$-PurF(1–04) catalyze PRA isomerization directly. PRAI and PurF both recognize phosphoribosylated substrates (Fig. 2), suggesting that shared molecular recognition properties, rather than common reaction mechanisms, can account for the promiscuous activity of PurF. In PRAI, a general acid (Asp379 in the bifunctional *E. coli* IGPS:PRAI enzyme) protonates the furanose ring oxygen of PRA, while a general base (Cys260) abstracts the proton from C2’ of the ribose.46 A spontaneous (and rate-limiting) enol/keto tautomerization yields CdR5P, the substrate of IGPS. In contrast, PurF has active sites in two separate domains: an N-terminal glutaminase domain and a C-terminal PRTase domain (Fig. 3b). There is no catalytic residue in the PRTase domain. Instead, PRPP is bound in an activated conformation, such that any suitably positioned nucleophile will attack it at C1 and displace pyrophosphate.47 Moreover, binding of PRPP triggers a series of conformational changes that activate the glutaminase domain and form an interdomain channel, excluding solvent and ensuring that the only possible nucleophile is NH$_3$, derived from glutamine hydrolysis.48,49

The recurrence of mutations at two sites in the flexible loop (Asn328 and Asn352; Fig. 6) and the role of this loop in binding the phosphoribosylated substrate PRPP provides strong circumstantial evidence that the PRTase active site is responsible for the observed PRAI activity. PurF-catalyzed synthesis of phosphoribosylamine from PRPP proceeds via an oxycarbonion intermediate, which is stabilized by the negatively charged pyrophosphate tail and probably the carboxylate side chains of Asp367 and Asp368 as well.50 PRAI possesses an antranilate moiety in place of a pyrophosphate tail (Fig. 2). This is clearly too large to be accommodated when the flexible PRTase loop is fully closed; however, it seems reasonable that the mutations selected at positions 328 and 352 play roles in restructuring this loop for more productive PRA binding. In the absence of pyrophosphate to effect substrate-assisted catalysis, a carboxylate oxygen of Asp368 is well positioned to abstract the proton from C2’ of PRA (C–O distance of 3.2 Å, using 1α-pyrophosphoryl-2α,3α-dihydroxy-4α-cyclopentanemethanol 5-phosphate as a proxy for PRA in the closed structure). A candidate for donating a proton to the furanose ring oxygen is less obvious, although the inner surface of the flexible PRTase loop is rich in charged residues.50 Further biochemical and structural studies will be required to verify the roles of PRTase active-site residues in catalyzing the isomerization of PRA.

**Catalytic promiscuity is a form of evolutionary contingency**

(His)$_6$-PurF possesses promiscuous PRAI activity because it serendipitously evolved to recognize a phosphoribosylated substrate, broadly similar to PRA (Fig. 2). It is an example of ‘contingency,’ defined as an instance when “a feature evolved long ago for a different use has fortuitously permitted survival during a sudden and unpredictable change in rules.”43 The occurrence of catalytic promiscuity, particularly at levels sufficient to alter organismal fitness, is difficult to predict *a priori*. Tryptophan is the least abundant amino acid, and the PRAI
reaction occurs spontaneously at a detectable rate. We tentatively predict that the weak PRAI activity of (His)_6-PurF(I198V), which was barely detectable in vitro, represents a lower limit for biologically relevant promiscuous activities.

**Adaptation through overexpression**

Most extant enzymes are extremely efficient catalysts, at least with respect to their primary activities, with \( k_{cat}/K_M \) Values of \( 10^{-5}−10^{-9} \) \( \text{M}^{-1} \text{s}^{-1} \). The catalytic efficiency of the wild-type PRAI falls squarely in this range (\( k_{cat}/K_M = 6.8 \times 10^8 \) \( \text{M}^{-1} \text{s}^{-1} \)). In contrast, we estimated the catalytic efficiency of our most efficient variant, (His)_6-PurF(1-04), to be over 7 orders of magnitude lower (\( k_{cat}/K_M = 0.3 \) \( \text{M}^{-1} \text{s}^{-1} \)). The *E. coli* strain overexpressing this variant is nevertheless similar in phenotype to wild-type cells (colony formation in ~30 h; c.f. 12 h for wild-type). Fitness was improved further in clone purF(2-02), via a mutation in the promoter region. These results reemphasize the potential importance of overexpression mutations (including gene duplication and/or mutations in promoter regions, rare codons, repressors, activators, etc.) in adaptive molecular evolution (Fig. 1).

**Does catalytic promiscuity imply evolvability?**

Several theorists have predicted that catalytic promiscuity and evolvability are correlated, and certain directed evolution experiments have produced variants with increased promiscuity. In this study, the catalytic promiscuity of (His)_6-PurF (I198V) was clearly a prerequisite for its subsequent adaptive evolution. Random mutagenesis and selection produced a variant that exhibited PRAI activity that was 25- to 30-fold more efficient and that could rapidly outcompete its ancestor in relative fitness experiments. The overexpression of (His)_6-PurF(1-04) must be energetically costly, but it remains unclear whether further improvement to the protein is possible. Recently, two influential studies have demonstrated that the adaptive landscapes of proteins can be highly constrained by contingency, the predominance of single nucleotide mutations, stochasticity and clonal interference. The optimization of the PurF active site to stabilize the PRAI transition state might require large-scale architectural restructuring not accessible via neutral or adaptive evolutionary pathways. It seems possible that evolvability, in addition to catalytic promiscuity, might also be a function of contingency.

**Materials and Methods**

**Materials**

All enzymes for molecular biology were from New England Biolabs (Ipswich, MA, USA), unless otherwise noted. All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA, USA), and their sequences are listed in Supplementary Table S2. IPTG was from Gold Biotechnology (St. Louis, MO, USA). Chemicals and assay components, including anthranilate and PRPP, were from Sigma Chemical Co. (St. Louis, MO, USA), unless otherwise noted.

**Bacterial strains and media**

*E. coli* strain JMB9sdR^t^- ΔtrpF was obtained from the Coli Genetic Stock Center at Yale University. The *E. coli* mutator strain XL1-Red was from Stratagene (La Jolla, CA, USA). *E. coli* strain KK8(pDM), which lacks the entire *trp* operon and harbors plasmid pDM for constitutive expression of the lac repressor, was a kind gift from Thomas Schwab and Prof. Reinhard Sterner of the University of Regensburg (Germany). *A. baylyi* strain ADP1 was from the ATCC Bacteriology Collection (ATCC). LB was used as a rich growth medium, with agar added to 1.5% (w/v) for growth on solid medium, and Vogel and Bonner’s Medium was used as the minimal medium for selection experiments. It was further supplemented with glucose (0.2%, w/v), casamino acids (0.5%, w/v), thiamine HCl (2 \( \mu\)g mL\(^{-1}\)), FeCl\(_3\) (10 \( \mu\)M) and agar (1.5%, w/v). Where necessary for the maintenance of plasmids, the following antibiotics were added to the growth medium: ampicillin, 100 \( \mu\)g mL\(^{-1}\); chloramphenicol (Cam), 34 \( \mu\)g mL\(^{-1}\); kanamycin (Kan), 25 \( \mu\)g mL\(^{-1}\); and spectinomycin, 100 \( \mu\)g mL\(^{-1}\).

**Selection for promiscuous PRAI activities**

We had previously pooled the plasmids of the ASKA ORF library. *E. coli* JMB9ΔtrpF cells were made electro-competent and transformed with this library. The transformed cells were washed and resuspended in Vogel–Bonner salts (1× concentration) and then spread on the minimal selection medium containing IPTG (50 \( \mu\)M) and Cam. Dilutions were also spread on LB-Cam agar, enabling us to estimate that approximately 147,000 transformed cells were plated on the selection medium. This ensured that the entire ASKA library was interrogated. Plates were incubated at 30 °C for 10 days, and colonies that appeared were grown in LB-Cam medium for storage and further analysis.

**Analysis of selected clones**

Saturated cultures of the 10 ASKA clones that could suppress the ΔtrpF lesion were used as the source of template DNA for PCR, using the pCA24N-specific primers pCA24N.for and pCA24N.rev. DNA sequencing of the PCR products using these two primers and purF-seq (Supplementary Table S2) identified the suppressor as purF(I198V) and purF(R196H). Their phenotypes were confirmed by transforming fresh *E. coli* JMB9ΔtrpF cells with purified plasmid DNA and replating on the minimal selection medium. The I198V mutation was corrected by overlap-assembly PCR. Primary amplifications were with pCA24N.for and purF_WT.rev as well as with purF_WT.for and pCA24N.rev. Overlap assembly of the two products in the presence of pCA24N.for and pCA24N.rev yielded an amplified fragment that could be digested with SfiI and ligated with pCA24N, to generate pCA24N-purF. The C-terminal GFP tag of the purF(I198V) ASKA clone was also removed by inserting a stop codon after the final codon of purF (encoding G505). Plasmid DNA was amplified with the
Construction of a (His)_6–PurF(I198V) epPCR library

A two-polymerase epPCR strategy was employed to create a maximally diverse and minimally biased library of purF(I198V) variants. PurF(I198V) was excised from pCA24N by digestion with PspOMI and NotI and ligated with vector pCDF-Ib (Novagen, Madison, WI, USA) that had been digested with the same enzymes to begin. The resulting plasmid, pCDF-purF(I198V), was used as the template for epPCR, as it encoded resistance to spectinomycin, rather than Cam (on pCA24N). Therefore, there was no possibility that the presence of template plasmid could lead to the diversity of the final library being overestimated.

In the first round of epPCR, 5.0 ng of pCDF-purF(I198V) DNA, corresponding to 1.7 ng of the target region (1890 bp), was amplified with the primers pCA24N.for and 3'pCDF and with Mutazyme polymerase (Stratagene), according to the manufacturer’s protocol. The cycling conditions for amplification were as follows: 95 °C for 1 min; 10 cycles of 94 °C for 20 s, 56 °C for 20 s and 72 °C for 2 min; and one cycle of 72 deg for 1 min. The amplified product was purified with a PCR Purification Kit (Qiagen, Valencia, CA, USA) before 30 ng of it was used in a second epPCR with Taq DNA polymerase and the same primers. The reaction was made more mutagenic by supplementing the Thermopol reaction buffer (New England Biolabs) with MnCl_2 (final concentration, 0.2 mM). The cycling conditions were as follows: 95 °C for 1 min; 20 cycles of 94 °C for 20 s, 56 °C for 20 s and 72 °C for 2 min; and one cycle of 72 deg for 1 min. The final epPCR product was purified with the PCR Purification Kit (Qiagen) and digested with SfiI. An aliquot (270 ng) was ligated with pCA24N (400 ng), which had been similarly digested, and the products of the ligation were introduced into E. coli JMB9_AtpF cells by electroporation. Transformed cells were spread on LB-Cam agar, and after 16 h at 30 °C, the resulting lawn of colonies was recovered. The primers pCA24N.for and tS1term.rev were also used to sequence the inserts of randomly selected clones; this enabled the library error rate and mutation spectrum to be assessed.

Selection of variants with improved PRAI activity

The recovered cells of the epPCR library were regrown in LB-Cam to mid-log phase, washed in 1× Vogel–Bonner salts and plated on the minimal selection medium containing 50 μM IPTG and Cam. Over 3.8 × 10^6 cells were plated at this step, ensuring that >99.5% of the library (total size = 6.4 × 10^10 unique DNA variants) was sampled. After 48 h of incubation at 30 °C, the 40 fastest-growing colonies were used to inoculate the LB-Cam medium (120 μL) and grown to saturation. The cultures were used directly in PCRs with the primers pCA24N.for and pCA24N.rev2, to generate templates for DNA sequencing. In turn, sequencing was carried out with these two primers and purf.seq.

Mutagenesis using E. coli XL1-Red

A second round of directed evolution was performed by passaging the purF(1–04) variant through the E. coli mutator strain XL1-Red. Cells were transformed with pCA24N-purF(I198V).

Restreaking experiments

The growth phenotype of each variant was confirmed by restreaking of fresh E. coli JMB9_AtpF cells. In each case, a single colony was used to inoculate LB-Cam (400 μL) and grown to saturation. An aliquot of the culture was pelleted, washed in 1× Vogel–Bonner salts and resuspended to A_600=0.1. A 3-μL aliquot of the resuspended cells was streaked on the minimal selection medium with IPTG (50 μM) and Cam, and the plate was incubated at 30 °C. In the case of the 40 selected variants from the epPCR library, the rate of growth was assigned a qualitative score from 1 [equivalent to cells harboring the pCA24N-purF(I198V) control] to 5 [equivalent to cells harboring the pCA24N-purF(1–04) variant]. Cells containing either pCA24N-purF(1–04) or pCA24N-purF(2–02) were also tested on the selection medium containing 5 μM IPTG.

Cloning trp genes

In addition to (His)_6–PurF(I198V) and (His)_6–PurF(1–04), it was necessary to produce a number of other enzymes for in vitro assays. The construction of pMS401, for the expression of (His)_6-tagged E. coli PRAI, was described previously. The E. coli trpC gene, encoding residues 1–259 of the bifunctional IGPS:PRAI monomer, was amplified from the ASKA clone pCA24N-trpC using the primers trpC_NcoI.for and trpC_BamHI.rev. These oligonucleotides also encoded restriction sites, enabling the digestion of the PCR product with NcoI and BamHI and its ligation with pMS401 (which had been treated with the same enzymes to excise trpC). The resulting plasmid, encoding C-terminally (His)_6-tagged IGPS, was verified by DNA sequencing and named pWP120.

In situ synthesis of PRAI’s substrate, PRA, can be accomplished using TrpD. However, this enzyme is part of a multiprotein complex in E. coli. Therefore, we cloned the homologue from A. baylyi ADP1 (GenBank accession no. CR543861). Whole cells were used as the source of chromosomal DNA, for PCR with the primers acTrpD.for and acTrpD.rev (which also encoded restriction sites for SfiI). The PCR product was restricted with SfiI and ligated into pCA24N that had been treated with the same enzyme. DNA sequencing confirmed the construction of plasmid pCA24N-AcTrpD.

Enzyme purification

Five (His)_6-tagged proteins were purified by metal affinity chromatography: AcTrpD, PurF(I198V), PurF(1–
PurF is a Promiscuous and Evolvable PRAI

04), PRAI and IGPS. Proteins were expressed in the E. coli deletion strains JMB9ΔtrpF (AcTrpD) and KK8(pDM) (the remainder) to avoid any possibility of contaminating the preparations with chromosomally encoded PRAI. Cells were grown at 37 °C in LB broth (350 mL) containing the appropriate antibiotics [AcTrpD, Cam; PurF(I198V) and PurF(1–04), Cam/Kan; PRAI and IGPS, ampicillin/Kan] to $A_{600} \approx 0.6$. IPTG (final concentration, 0.4 mM) was added as an inducer, and protein overexpression was at 30 °C for 6–8 h. Cells were harvested by centrifugation, and the pellets were stored at –80 °C. For purification, each pellet was thawed and resuspended in 10 mL of buffer A (40 mM Tris–HCl, 500 mM NaCl, 10 mM imidazole, 10% v/v glycerol and 1 mM β-mercaptoethanol, pH 8.0), with lysozyme (0.2 mg mL$^{-1}$; Sigma) and DNase I (1 μM mL$^{-1}$; Sigma) added. Cells were lysed by sonication on ice, and the lysates were clarified by centrifugation (20,400g, 4 °C, 40 min). Binding of each (His)$_6$ tagged protein to Ni-nitrilotriacetic acid affinity resin (Qiagen) was effected by incubation at 4 °C for 2 h, with rocking. Each resin was washed twice with six volumes of buffer A, before being transferred to a gravity flow column. After two further washes with six column volumes of buffer A containing 20 mM imidazole, the purified protein was eluted in buffer A containing 250 mM imidazole. Each protein was dialyzed extensively against buffer B (40 mM Tris–HCl, 200 mM NaCl, 0.5 mM MgCl$_2$ and 0.5 mM β-mercaptoethanol, pH 8.0), at 4 °C. Finally, any possible aggregates were removed by passing the recovered proteins through 0.22-μm Spin-X centrifuge tube filters (Costar, Corning, NY, USA). Prior to filtration, extra β-mercaptoethanol (final total concentration, 2 mM) and ethylenediaminetetraacetic acid (final concentration, 2 mM) were added to preparations of PRAI and IGPS, as these proteins are known to be susceptible to inactivation by metal ion-catalysed oxidation.

Steady-state kinetic assays

The PRAI activity of (His)$_6$-PurF(1–04) at 25 °C was quantified using a spectrophotometric assay, as the substrate concentrations used (up to 2 mM) were too high to allow the use of the more sensitive fluorimetric assay. Assays (total volume, 800 μL) were carried out in buffer B, containing 5 μM (His)$_6$-PurF(1–04) and 2 μM IGPS. The latter allowed reaction velocities to be estimated from the rate of InGP formation; this could be followed by an increase in absorbance at 278 nm ($\varepsilon_{278\text{m},\text{InGP}}=5590$ M$^{-1}$ cm$^{-1}$), based on the measurement for N-Ac-Trp-NH$_2$. Doubling the IGPS concentration had no effect on the rate of InGP formation. AcTrpD (1 μM) was used to synthesize the substrate PRA in the cuvette at concentrations from 0 to 2.0 mM, from corresponding amounts of anthranilate and a molar excess of PRPP. Controls showed activity to be proportional to enzyme concentration. Finally, assays were repeated with three separate batches of (His)$_6$-PurF (1–04), purified from separate cultures.

Competition experiments

The fitness values of strains expressing (His)$_6$-IGPS:PRAI and (His)$_6$-PurF(1–04), relative to (His)$_6$-PurF(I198V), were measured in competition experiments, as Exponential-phase cells of two strains, one GFP tagged and one untagged, were used to inoculate an aliquot of the minimal selection medium (8 mL). The liquid culture was incubated under constant conditions (30 °C, shaking at 230 rpm) for exactly 24 h. Dilutions were plated on LB medium containing Cam and IPTG (30 μg mL$^{-1}$) at the beginning and the end of the incubation, allowing the relative change in population of each strain to be quantified by counting the frequencies of green fluorescent colonies on the plates. All experiments were performed in triplicate.

Acknowledgements

This work was supported by grants from the National Institutes of Health (R01 GM074264), the
National Science Foundation (CHE-0404677) and the Massey University Institute of Molecular Biosciences. We are greatly indebted to Dr. Baoyun Xia for his guidance in the HPLC analysis. We also thank Dr. Monica Gerth for her assistance with the preparation of figures. DNA sequencing was performed at the Center for Fundamental and Applied Molecular Evolution (NSF DBI-0320786a) at Emory University.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2008.01.043

References

PurF is a Promiscuous and Evolvable PRAI


