Published in final edited form as:
J Mol Biol. 2008 March 21; 377(2): 323–336.

A study in molecular contingency: glutamine phosphoribosylpyrophosphate amidotransferase is a promiscuous and evolvable phosphoribosylanthranilate isomerase

Wayne M. Patrick[1,2] and **Ichiro Matsumura**[1*]

[1]*Department of Biochemistry, Center for Fundamental and Applied Molecular Evolution, Emory University, Atlanta, Georgia 30322, USA.* [2]*Institute of Molecular Biosciences, Massey University, Auckland 0745, New Zealand.*

Summary

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 collection of *Escherichia coli* open reading frames (ORFs). The over-expression of $(His)_6$-tagged glutamine phosphoribosylpyrophosphate amidotransferase (PurF) unexpectedly rescued a Δ*trpF E. coli* strain from starvation on minimal media. The wild-type PurF and TrpF enzymes are unrelated in sequence, tertiary structure or catalytic mechanism. The promiscuous phosphoribosylanthranilate isomerase (PRAI) activity of the ASKA PurF variant apparently stems from a pre-existing affinity for phosphoribosylated substrates. The relative fitness of the $(His)_6$-PurF/Δ*trpF* 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 (k_{cat}/K_M = 0.3 $s^{-1}.M^{-1}$) was ~25-fold more efficient than its ancestor, but >10^7–fold less efficient than the wild-type PRAI. These observations demonstrate in quantitative terms that the weak secondary activities of promiscuous enzymes can dramatically improve the fitness of contemporary organisms.

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 multi-step pathway for the catabolism of the *s*-triazine herbicides,[1] 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 pre-existing protein scaffolds. Nonetheless, these

*Corresponding author; email address: imatsum@emory.edu.

ground-breaking examples offer little predictive capability to those who study adaptive evolution in other systems.[3]

The substrate ambiguity and catalytic promiscuity of pre-existing enzymes have garnered attention as plausible starting points for the evolution of novel functions,[4;5;6;7] especially when mutations that cause the over-expression of promiscuous enzymes (including gene duplication events[8]) amplify weak secondary activities to physiologically relevant levels (Figure 1).[9;10;11;12;13;14;15] In a previous study, we investigated this model with the ASKA open reading frame (ORF) library, which contains each *Escherichia coli* ORF cloned into the expression vector pCA24N.[16] Aliquots of each plasmid were pooled and used to transform 104 conditional auxotrophs from the Keio collection of single-gene deletion strains.[17] 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 over-expression of 41 mostly non-homologous *E. coli* genes.[14] These results showed that many proteins are multi-functional, 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 (Figure 2a), as our model reaction. *In vitro*, the reaction occurs spontaneously under mild but non-physiological conditions (50% v/v ethanol, room temperature).[18] *In vivo*, it is ordinarily catalyzed by PRA isomerase (PRAI, EC 5.3.1.24), a well-characterized $(\beta\alpha)_8$-barrel protein (Figure 3a) that is the product of the *trpF* gene.[19;20] In turn, the product of the reaction, CdRP, is the substrate for the *trpC* gene product, indoleglycerol-phosphate synthase (IGPS). We did not consider PRAI activity in our previous study[14] because the Keio knock-out collection does not include a Δ*trpF* strain. 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.[21]

PRA isomerization is a good model reaction for several reasons. First, the frequent occurrence of $(\beta\alpha)_8$-barrel proteins within the protein database suggests that this scaffold may be particularly evolvable.[22;23;24] 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.[19;25] 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 $(\beta\alpha)_8$-barrels of histidine biosynthesis. They showed that single point mutations can impart PRAI activity on the HisA protein, which catalyzes an analogous Amadori rearrangement,[26] and also on HisF, which normally catalyzes the subsequent synthesis of the imidazole ring.[27] 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 pre-existing *E. coli* proteins with promiscuous PRAI activity. We expected that IGPS:PRAI, HisA and/or HisF would rescue Δ*trpF E. coli* cells, and also thought it possible that other $(\beta\alpha)_8$-barrel proteins (*i.e.* distantly related paralogues) would effect rescue. We transformed the auxotrophic *E. coli* strain JMB9Δ*trpF*[28] 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 over-expression of each ASKA-encoded gene. After < 24 h 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, six colonies were identified after six days of growth, and four 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 N-terminal $(His)_6$ tag and a C-terminal green fluorescent protein (GFP) tag.[16] This homotetrameric, two-domain enzyme (Figure 3b) normally catalyzes the first committed step of *de novo* purine biosynthesis, the conversion of PRPP to phosphoribosylamine (Figure 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 public.[16] 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 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). In order to avoid interference in downstream absorbance assays, the pCA24N-encoded, C-terminal GFP tag was also removed from this clone; this modification had no affect on its ability to rescue *E. coli* JMB9Δ*trpF*.

To verify the phenotype, fresh *E. coli* JMBΔ*trpF* cells were retransformed with pCA24N-*purF (I198V)* and struck on minimal selection medium. Evenly-sized colonies appeared at a uniform rate, in an identical fashion to 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[29;30]) 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 (Figure 3), but both enzymes have evolved to recognize substrates containing a ribose-5-phosphate (R5P) moiety (Figure 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 over-expressed $(His)_6$-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.[31;32] We therefore sought to demonstrate PRAI activity for the purified $(His)_6$-PurF(I198V) protein *in vitro*.

The slow growth of *E. coli* JMB9Δ*trpF* harboring pCA24N-*purF(I198V)*, even when the rescuing protein was over-expressed, 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).[25] Consequently, we turned to a coupled endpoint assay (Supplementary Data Figure S1) and 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 enzyme from *Acinetobacter baylyi* strain ADP1. This enzyme (AcTrpD), which will be described in greater detail elsewhere, proved amenable to over-expression 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, indoleglycerolphosphate (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 (Figure 4). Further, 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 (Figure 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 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,[6;33] 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 6.4×10^5 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 spectrum of mutations (Supplementary Data 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[34;35] and the 'PCR distribution' described by Drummond *et al.*[36] 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 JMB9Δ*trpF* cells were transformed with the library. The resulting transformants were propagated on selective minimal media; after only 48 h 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 Data Figure 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* JMB9Δ*trpF* cells (Table 1). This protein was solubly expressed and possessed a similar number of mutations to the selected variants, so it 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 (Figure 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 to anchor the ends of a long, flexible loop that closes over the phosphoribosyltransferase (PRTase) active site when PRPP is bound (Figure 6).[37;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 (Figure 2), was critical for improving the promiscuous PRAI activity of $(His)_6$-PurF.

A second round of directed evolution

The fittest variant from the epPCR library, $(His)_6$-PurF(1-04), had to be over-expressed to rescue *E. coli* JMB9Δ*trpF*. 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* JMB9Δ*trpF*. 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 two-fold 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 $lacI^q$ alleles and P_{T5}/*lacO* regions of pCA24N-*purF(I198V)*, -*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 (Figure 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 product, and corresponds to 22 turnovers per PurF(1-04) active site. This result suggested that the PRAI activity of $(His)_6$-PurF(1-04) was 25- to 30-fold higher than that of $(His)_6$-PurF(I198V).

We next attempted to measure the kinetic parameters of the $(His)_6$-PurF(1-04)-catalyzed reaction. A sensitive assay for PRAI activity has been described,[19] but it is limited to substrate concentrations ≤ 0.8 mM[27] because of the extremely high fluorescence quantum yields of anthranilate (the upsteam substrate) and PRA. Instead, we monitored the formation of InGP *via* its absorbance at 278 nm. The contribution of anthranilate to A_{278} was accounted for by running controls without added $(His)_6$-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 (Figure 7b), this rate of spontaneous isomerization was negligible. It was not possible to saturate the enzyme in the assays; therefore, k_{cat}/K_M was estimated from the slope

of the linear fit when the initial velocity was plotted against substrate concentration (Supplementary Data Figure S3). Assays of three separate batches of $(His)_6$-PurF(1-04) allowed us to estimate $k_{cat}/K_M = 0.3 \pm 0.2\ s^{-1}.M^{-1}$. This value is 2.3×10^7-fold below that of the wild-type *E. coli* PRAI ($k_{cat}/K_M = 6.8 \times 10^6\ s^{-1}.M^{-1}$).[19]

Our average yield of purified $(His)_6$-PurF(1-04) was 19 $mg.L^{-1}$ 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 over-expressing cell contained $> 1.8 \times 10^5$ molecules of $(His)_6$-PurF(1-04). Assuming a cytoplasmic volume of 1.25 fL,[39] this corresponds to an intracellular concentration of > 0.24 mM. The total tryptophan concentration in a healthy *E. coli* cell is estimated to be approximately 10 mM.[40] It appears that the very poor PRAI activities of $(His)_6$-PurF(I198V) and $(His)_6$-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 $(His)_6$-PurF(I198V) and $(His)_6$-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 μM.[19] Our observation that $(His)_6$-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 days to 10 days. $(His)_6$-PurF(1-04) contained eight substitutions and led to colony formation in < 30 h, yet our negative control, 1-Neg (Table 1), carried a similar mutational load in *purF*, but did not complement *E. coli* JMB9Δ*trpF* 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 it is $(His)_6$-PurF(I198V) and $(His)_6$-PurF(1-04) that are producing the observed PRAI activities, *in vivo* and *in vitro*.

Relative fitness corresponds to *in vitro* activity

The measured catalytic activity of $(His)_6$-PurF(1-04) was extremely low, yet over-expression of this protein led to colony formation at a rate that was comparable to 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 $(His)_6$-PurF(I198V), $(His)_6$-PurF(1-04) and the bifunctional IGPS:PRAI enzyme from *E. coli* (Table 2). The GFP tags on the $(His)_6$-PurF(I198V) and $(His)_6$-IGPS:PRAI clones from the ASKA library allowed us to conduct pairwise competition experiments with one tagged and one untagged clone, in turn enabling relative growth rates to be measured in a standard 24 h assay.[41]

In the first experiment, we showed that removing the GFP tag from $(His)_6$-PurF(I198V) had no affect on the fitness of the strain; that is, the fitness of the GFP-tagged $(His)_6$-PurF(I198V) clone relative to the untagged clone is not significantly different from 1 (Table 2). As expected, a second experiment showed that cells harboring $(His)_6$-PurF(I198V) were rapidly out-competed by those expressing $(His)_6$-IGPS:PRAI (W = 4.8). Interestingly, the growth rate of the $(His)_6$-PurF(I198V) strain was increased substantially in the presence of $(His)_6$-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 $(His)_6$-PurF(1-04)-expressing cells were grown in 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 Δ*trpF* suppressors because we expected a particular outcome. It seemed reasonable that HisA (EC 5.3.1.16), an homologous $(\beta\alpha)_8$-barrel that catalyzes the same isomerization reaction (albeit on a slightly different substrate), would possess the requisite promiscuous activity when over-expressed. Instead, our results showed that over-expression of $(His)_6$-PurF(I198V), an amidotransferase (EC 2.4.2.14) from the purine anabolic pathway, could suppress the tryptophan auxotrophy of the Δ*trpF* strain. The PurF and PRAI proteins are unrelated in structure (Figure 3), but the two enzymes normally recognize broadly similar substrates (Figure 2). Purified $(His)_6$-PurF(I198V) exhibits a very modest but detectable level of PRAI activity *in vitro* (Figure 4). A variant with 25–30 fold greater activity emerged after a single round of random mutagenesis and selection.

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] Further, Downs and her co-workers have shown 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 non-enzymatic synthesis of the PurF product, phosphoribosylamine.[45] With these results in mind, we considered a scenario in which over-expressed PurF was rescuing *E. coli* JMB9Δ*trpF* by catalyzing the TrpD reaction, thereby leading to hyper-accumulation of PRA for its spontaneous isomerization to CdRP. However, neither the *E. coli* nor the *A. baylyi* TrpD could rescue *E. coli* JMB9Δ*trpF* when over-expressed (data not shown).

Instead, our *in vitro* assays (Figure 4 and Figure 7) showed that $(His)_6$-PurF(I198V) and $(His)_6$-PurF(1-04) catalyze PRA isomerization directly. PRAI and PurF both recognize phosphoribosylated substrates (Figure 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[19]) enol/keto tautomerization yields CdRP, 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 (Figure 3b). There are no catalytic residues 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.[37] Moreover, binding of PRPP triggers a series of conformational changes that activate the glutaminase domain and form an inter-domain channel, excluding solvent and ensuring that the only possible nucleophile is NH_3, derived from glutamine hydrolysis.[37;47]

The recurrence of mutations at two sites in the flexible loop (Asn328 and Asn352; Figure 6) and the role of this loop in binding the phosphoribosylated substrate, PRPP, provide 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 oxycarbonium ion intermediate, which is stabilized by the negatively-charged pyrophosphate tail and probably also the carboxylate side chains of Asp367 and Asp368.[37] PRA possesses an anthranilate moiety in place of a pyrophosphate tail (Figure 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 cPRPP 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.[37] 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 (Figure 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".[48] 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 over-expression

Most extant enzymes are extremely efficient catalysts, at least with respect to their primary activities, with k_{cat}/K_M values of 10^4–10^9 $s^{-1}.M^{-1}$.[49] The catalytic efficiency of the wild-type PRAI falls squarely in this range ($k_{cat}/K_M = 6.8 \times 10^6$ $s^{-1}.M^{-1}$).[19] In contrast, we estimated the catalytic efficiency of our most efficient variant, $(His)_6$-PurF(1-04), to be over seven orders of magnitude lower ($k_{cat}/K_M = 0.3$ $s^{-1}.M^{-1}$). The *E. coli* strain over-expressing 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 over-expression mutations (including gene duplication and/or mutations in promoter regions, rare codons, repressors, activators, *etc*.) in adaptive molecular evolution (Figure 1).

Does catalytic promiscuity imply evolvability?

Several theorists have predicted that catalytic promiscuity and evolvability are correlated[4;5;6;7] and certain directed evolution experiments have produced variants with increased promiscuity.[33;50] 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 25–30 fold more efficient and that could rapidly out-compete its ancestor in relative fitness experiments. The over-expression 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.[51;52] 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), unless otherwise noted. All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA) and their sequences are listed in Supplementary Data Table S2. IPTG was from Gold Biotechnology (St. Louis, MO). Chemicals and assay components, including anthranilate and PRPP, were from Sigma Chemical Co. (St. Louis, MO), unless otherwise noted.

Bacterial strains and media

E. coli strain JMB9*hsdR*$^{-}$*M*$^{+}$ Δ*trpF* was obtained from the Coli Genetic Stock Center at Yale University. The *E. coli* mutator strain XL1-Red was from Stratagene (La Jolla, CA). *E. coli* strain KK8(pDM), which lacks the entire *trp* operon and harbors plasmid pDM for constitutive expression of the *lac* repressor,[53] was a kind gift from Thomas Schwab and Prof. Reinhard Sterner at the University of Regensburg, Germany. *Acinetobacter baylyi* strain ADP1 was from the ATCC Bacteriology Collection (#33305). LB was used as a rich growth medium, with agar added to 1.5% (w/v) for growth on solid medium. Vogel and Bonner's Medium E[54] 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 μg.mL^{-1}), $FeCl_3$ (10 μM) and agar (1.5%, w/v). Where necessary for the maintenance of plasmids, the following antibiotics were added to the growth medium: ampicillin (Amp), 100 μg.mL^{-1}; chloramphenicol (Cam), 34 μg.mL^{-1}; kanamycin (Kan), 25 μg.mL^{-1}; and spectinomycin (Spec), 100 μg.mL^{-1}.

Selection for promiscuous PRAI activities

We had previously pooled the plasmids of the ASKA ORF library.[14] *E. coli* JMB9Δ*trpF* cells were made electrocompetent and transformed with this library. The transformed cells were washed and resuspended in Vogel-Bonner salts (1 × concentration), then spread on minimal selection medium containing IPTG (50 μM) 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.[34] Plates were incubated at 30°C for 10 days, and colonies that appeared were grown in LB-Cam medium for further analysis and archiving.

Analysis of selected clones

Saturated cultures of the 10 ASKA clones that could suppress the Δ*trpF* 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 Data Table S2) identified the suppressors as *purF(I198V)* and *purF(R196H)*. Their phenotypes were confirmed by transforming fresh *E. coli* JMB9Δ*trpF* cells with purified plasmid DNA and replating on minimal selection medium.

The I198V mutation was corrected by overlap assembly PCR. Primary amplifications were with pCA24N.for and purF_WT.rev, and 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 primers pCA24N.for and tS1term.rev, and the PCR product was cloned back into pCA24N after digestion with SfiI. The result was plasmid pCA24N-*purF(I198V)*.

Construction of a $(His)_6$-PurF(I198V) epPCR library

A two-polymerase epPCR strategy[35] was employed to create a maximally diverse and minimally biased library of *purF(I198V)* variants. To begin, *purF(I198V)* was excised from pCA24N by digestion with PspOMI and NotI, and ligated with vector pCDF-1b (Novagen, Madison, WI) that had been digested with the same enzymes. The resulting plasmid, pCDF-*purF(I198V)* was used as the template for epPCR, as it encoded resistance to Spec, 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 over-estimated.

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 primers pCA24N.for and 3'pCDF, and with Mutazyme polymerase (Stratagene, La Jolla, CA), according to the manufacturer's protocol. The cycling conditions for amplification were: 95°C for 1 min; 10 cycles of 94°C for 20 s, 56° C for 20 s, 72°C for 2 min; one cycle of 72 deg for 1 min. The amplified product was purified with the PCR Purification Kit (Qiagen, Valencia, CA) 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 $MnCl_2$ (0.2 mM, final concentration). The cycling conditions were: 95°C for 1 min; 20 cycles of 94°C for 20 s, 56°C for 20 s, 72°C for 2 min; 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) that had been similarly digested, and the products of the ligation were introduced into *E. coli* JMB9Δ*trpF* 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. Primers pCA24N.for and tS1term.rev were also used to sequence the *purF* 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 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 = 6.4×10^5 unique DNA variants) was sampled.[34] After 48 h incubation at 30°C, the 40 fastest-growing colonies were used to inoculate LB-Cam medium (120 uL) and grown to saturation. The cultures were used directly in PCRs with 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 (1-04)* plasmid DNA and used to inoculate an LB-Cam culture (40 mL) directly. Dilutions plated on LB-Cam indicated that the initial inoculum was ~60 transformants. After 39 h at 37° C, the culture contained ~3×10^{10} cells, corresponding to 29 generations of growth. Plasmid DNA was purified from the cell culture using EconoSpin columns (Epoch Biolabs, Houston, TX) and used to transform *E. coli* JMB9Δ*trpF* by electroporation. Approximately 1.5×10^6 transformants were spread on selection medium containing 50 μM IPTG and Cam. After 24 h growth at 30°C, the four largest colonies were picked, grown in LB-Cam and, in the case of

variant *purF(2-02)*, sequenced using primers pCA24N.for, pCA24N.rev2, pUC rev+120, purF.seq, rev+270, purF_WT.rev and univ-70.

Restreaking experiments

The growth phenotype of each variant was confirmed by retransformation of fresh *E. coli* JMB9Δ*trpF* cells. In each case, a single colony was used to inoculate LB-Cam (400 uL) 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 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 (cells harboring the pCA24N-*purF(1-04)* variant). Cells containing either pCA24N-*purF (1-04)* or pCA24N-*purF(2-02)* were also tested on 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.[55] The *E. coli trpC* gene, encoding residues 1-259 of the bifunctional IGPS:PRAI monomer,[21] was amplified from the ASKA clone pCA24N-*trpCF* using 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 *trpF*). 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 multi-protein 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(1-04), PRAI, and IGPS. To avoid any possibility of contaminating the preparations with chromosomally-encoded PRAI, proteins were expressed in the *E. coli* deletion strains JMB9Δ*trpF* (AcTrpD) and KK8(pDM) (the remainder). 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, Amp/Kan) to A_{600} ~ 0.6. IPTG (0.4 mM, final concentration) was added as an inducer, and protein over-expression 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 buffer A (40 mM Tris•HCl, 300 mM NaCl, 10 mM imidazole, 10% v/v glycerol, 1 mM β-mercaptoethanol, pH 8.0) with lysozyme (0.2 $mg.mL^{-1}$; Sigma) and DNase I (1 $mU.mL^{-1}$; Sigma) added. Cells were lysed by sonication on ice, and the lysates were clarified by centrifugation (20400*g*, 4°C, 40 min). Binding of each $(His)_6$-tagged protein to Ni-NTA affinity resin (Qiagen, Valencia, CA) 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 $MgCl_2$, 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). Prior to filtration, extra β-mercaptoethanol (2 mM, final total concentration) and EDTA (2 mM, final concentration) were added to preparations of PRAI and IGPS, as these proteins are known to be susceptible to inactivation by metal ion-catalyzed oxidation.[25]

Protein concentrations were quantified by measuring A_{280}, using molar extinction coefficients that were calculated according to Pace *et al*.[56] Aliquots of PRAI and AcTrpD were stored at −80°C, while the other proteins were stored at 4°C and used within three days of being purified.

HPLC detection of reaction products

Endpoint assays and HPLC were used to demonstrate the PRAI activity of $(His)_6$-PurF(I198V) and $(His)_6$-PurF(1-04). Reaction mixtures (600 μL, total volume) contained: Tris•HCl, pH 8.6 (40 mM); NaCl (200 mM); $MgCl_2$ (0.5 mM); β-mercaptoethanol (0.5 mM); anthranilate (2 mM); PRPP (10 mM); and the coupled enzymes AcTrpD (for *in situ* synthesis of PRA from anthranilate and PRPP; 1 μM) and IGPS (for conversion of CdRP to the stable product InGP; 2 μM). The $(His)_6$-PurF(I198V) and $(His)_6$-PurF(1-04) enzymes were assayed at concentrations of 5 μM. A positive control contained PRAI (0.2 μM) and a negative control contained an equal volume of buffer B. IGPS is known to be rather unstable;[21] however, control reactions with higher concentrations of the coupled enzyme showed no changes in InGP formation. This demonstrated that denaturation of IGPS was not limiting the amount of InGP produced. The other enzymes (AcTrpD, $(His)_6$-PurF(I198V), $(His)_6$-PurF(1-04) and PRAI) showed no tendency to aggregate or denature under the conditions of the assay.

In order to prolong the life of the labile substrate, PRA, the tubes containing each reaction were wrapped in foil and incubated at 23°C for 16 h. Next, each reaction was aliquotted into the wells of a Nunc Immobilizer Ni-chelate microplate (Nunc A/S, Roskilde, Denmark) and incubated at 23°C for a further 30 min. This removed the proteins (all of which were $(His)_6$-tagged) from solution. The recovered mixture of substrates and products was passed through a 0.22 μm Spin-X filter (Corning). An aliquot (100 μL) was diluted with 700 μL buffer B and a Shimadzu UV-1601 spectrophotometer was used to measure its absorbance spectrum at 25°C, against a blank containing buffer B.

The remainder of each reaction was analyzed by HPLC, using a protocol modified from the one described by Kirschner *et al*.[25] A Shimadzu LC-20AT controller was used to equilibrate a 250 mm C_{18} reversed phase column (Grace Vydac, Columbia, MD) with Tris•HCl, pH 7.5 (50 mM), at room temperature. The components of each reaction were chromatographed isocratically in this buffer, at a flow rate of 1.0 $mL.min^{-1}$. A 20 μL loading loop was used, and elution profiles were monitored at 278 nm using a Shimadzu SPD-20A UV-Vis detector. Peaks were assigned by injecting an anthranilate standard, and by comparing the positive and negative controls. Peak integration was carried out using EZStart version 7.4 software (Shimadzu).

Steady-state kinetic assays

The PRAI activity of $(His)_6$-PurF(1-04) at 25°C was quantified using a spectrophotometric assay,[25] as the substrate concentrations used (up to 2 mM) were too high to allow the use of the more sensitive fluorimetric assay.[19] Assays (800 μL, total volume) 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 ($\varepsilon_{278,\ InGP} = 5590\ M^{-1}.cm^{-1}$, based on the measurement for N-Ac-Trp-NH_2[56]). 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 0–2.0 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.[41] Exponential-phase cells of two strains, one GFP-tagged and one untagged, were used to inoculate an aliquot of 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 (50 μM) 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

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 at Emory University. This work was supported by grants from the NIH (R01 GM074264), the NSF (CHE-0404677) and the Institute of Molecular Biosciences at Massey University.

References

1. Shapir N, Mongodin EF, Sadowsky MJ, Daugherty SC, Nelson KE, Wackett LP. Evolution of catabolic pathways: Genomic insights into microbial *s*-triazine metabolism. J. Bacteriol 2007;189:674–682. [PubMed: 17114259]
2. Scanlan TS, Reid RC. Evolution in action. Chem. Biol 1995;2:71–75. [PubMed: 9383406]
3. Dean AM, Thornton JW. Mechanistic approaches to the study of evolution: the functional synthesis. Nat. Rev. Genet 2007;8:675–688. [PubMed: 17703238]
4. D'Ari R, Casadesús J. Underground metabolism. Bioessays 1998;20:181–186. [PubMed: 9631663]
5. Jensen RA. Enzyme recruitment in evolution of new function. Annu. Rev. Microbiol 1976;30:409–425. [PubMed: 791073]
6. Khersonsky O, Roodveldt C, Tawfik DS. Enzyme promiscuity: evolutionary and mechanistic aspects. Curr. Opin. Chem. Biol 2006;10:498–508. [PubMed: 16939713]
7. O'Brien PJ, Herschlag D. Catalytic promiscuity and the evolution of new enzymatic activities. Chem. Biol 1999;6:R91–R105. [PubMed: 10099128]
8. Bergthorsson U, Andersson DI, Roth JR. Ohno's dilemma: evolution of new genes under continuous selection. Proc. Natl Acad. Sci. USA 2007;104:17004–17009. [PubMed: 17942681]
9. Berg CM, Wang MD, Vartak NB, Liu L. Acquisition of new metabolic capabilities: multicopy suppression by cloned transaminase genes in *Escherichia coli* K-12. Gene 1988;65:195–202. [PubMed: 3044925]
10. Cocks GT, Aguilar T, Lin EC. Evolution of l-1,2-propanediol catabolism in *Escherichia coli* by recruitment of enzymes for l-fucose and l-lactate metabolism. J. Bacteriol 1974;118:83–88. [PubMed: 4595205]
11. Hall BG. The EBG system of *E. coli*: origin and evolution of a novel β-galactosidase for the metabolism of lactose. Genetica 2003;118:143–156. [PubMed: 12868605]
12. Miller BG, Raines RT. Identifying latent enzyme activities: substrate ambiguity within modern bacterial sugar kinases. Biochemistry 2004;43:6387–6392. [PubMed: 15157072]
13. Morett E, Saab-Rincón G, Olvera L, Olvera M, Flores H, Grande R. Sensitive genome-wide screen for low secondary enzymatic activities: the YjbQ family shows thiamin phosphate synthase activity. J. Mol. Biol. 2008In press.

14. Patrick WM, Quandt EM, Swartzlander DB, Matsumura I. Multicopy suppression underpins metabolic evolvability. Mol. Biol. Evol 2007;24:2716–2722. [PubMed: 17884825]
15. St. Martin EJ, Mortlock RP. Natural and altered induction of the l-fucose catabolic enzymes in *Klebsiella aerogenes*. J. Bacteriol 1976;127:91–97. [PubMed: 179982]
16. Kitagawa M, Ara T, Arifuzzaman M, Ioka-Nakamichi T, Inamoto E, Toyonaga H, Mori H. Complete set of ORF clones of *Escherichia coli* ASKA library (A Complete Set of *E. coli* K-12 ORF Archive): unique resources for biological research. DNA Res 2005;12:291–299. [PubMed: 16769691]
17. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. Construction of *Escherichia coli* K-12 inframe, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol 2006;2:2006.0008.
18. Creighton TE. The nonenzymatic preparation in solution of N-(5′-phosphoribosyl) anthranilic acid, an intermediate in tryptophan biosynthesis. J. Biol. Chem 1968;243:5605–5609. [PubMed: 5699054]
19. Hommel U, Eberhard M, Kirschner K. Phosphoribosyl anthranilate isomerase catalyzes a reversible Amadori reaction. Biochemistry 1995;34:5429–5439. [PubMed: 7727401]
20. Wilmanns M, Priestle JP, Niermann T, Jansonius JN. Three-dimensional structure of the bifunctional enzyme phosphoribosylanthranilate isomerase: indoleglycerolphosphate synthase from *Escherichia coli* refined at 2.0 Å resolution. J. Mol. Biol 1992;223:477–507. [PubMed: 1738159]
21. Eberhard M, Tsai-Pflugfelder M, Bolewska K, Hommel U, Kirschner K. Indoleglycerol phosphate synthase-phosphoribosyl anthranilate isomerase: comparison of the bifunctional enzyme from *Escherichia coli* with engineered monofunctional domains. Biochemistry 1995;34:5419–5428. [PubMed: 7727400]
22. Gerlt JA, Raushel FM. Evolution of function in $(\beta/\alpha)_8$-barrel enzymes. Curr. Opin. Chem. Biol 2003;7:252–264. [PubMed: 12714059]
23. Sterner R, Höcker B. Catalytic versatility, stability, and evolution of the $(\beta\alpha)_8$-barrel enzyme fold. Chem. Rev 2005;105:4038–4055. [PubMed: 16277370]
24. Brändén C-I. The TIM barrel – the most frequently occurring folding motif in proteins. Curr. Opin. Struct. Biol 1991;1:978–983.
25. Kirschner K, Szadkowski H, Jardetzky TS, Hager V. Phosphoribosylanthranilate isomerase-indoleglycerol-phosphate synthase from *Escherichia coli*. Methods Enzymol 1987;142:386–397. [PubMed: 3298981]
26. Jürgens C, Strom A, Wegener D, Hettwer S, Wilmanns M, Sterner R. Directed evolution of a $(\beta\alpha)_8$-barrel enzyme to catalyze related reactions in two different metabolic pathways. Proc. Natl Acad. Sci. USA 2000;97:9925–9930. [PubMed: 10944186]
27. Leopoldseder S, Claren J, Jürgens C, Sterner R. Interconverting the catalytic activities of $(\beta\alpha)_8$-barrel enzymes from different metabolic pathways: sequence requirements and molecular analysis. J. Mol. Biol 2004;337:871–879. [PubMed: 15033357]
28. Sterner R, Dahm A, Darimont B, Ivens A, Liebl W, Kirschner K. $(\beta\alpha)_8$-barrel proteins of tryptophan biosynthesis in the hyperthermophile *Thermotoga maritima*. EMBO J 1995;14:4395–4402. [PubMed: 7556082]
29. Hastings PJ, Slack A, Petrosino JF, Rosenberg SM. Adaptive amplification and point mutation are independent mechanisms: evidence for various stress-inducible mutation mechanisms. PLoS Biol 2004;2:e399. [PubMed: 15550983]
30. Hendrickson H, Slechta ES, Bergthorsson U, Andersson DI, Roth JR. Amplification-mutagenesis: evidence that "directed" adaptive mutation and general hypermutability result from growth with a selected gene amplification. Proc. Natl Acad. Sci. USA 2002;99:2164–2169. [PubMed: 11830643]
31. Kim PJ, Lee DY, Kim TY, Lee KH, Jeong H, Lee SY, Park S. Metabolite essentiality elucidates robustness of *Escherichia coli* metabolism. Proc. Natl Acad. Sci. USA 2007;104:13638–13642. [PubMed: 17698812]
32. Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa T, Naba M, Hirai K, Hoque A, Ho PY, Kakazu Y, Sugawara K, Igarashi S, Harada S, Masuda T, Sugiyama N, Togashi T, Hasegawa M, Takai Y, Yugi K, Arakawa K, Iwata N, Toya Y, Nakayama Y, Nishioka T, Shimizu K, Mori H, Tomita M. Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. Science 2007;316:593–597. [PubMed: 17379776]

33. Aharoni A, Gaidukov L, Khersonsky O, Gould SM, Roodveldt C, Tawfik DS. The 'evolvability' of promiscuous protein functions. Nat. Genet 2005;37:73–76. [PubMed: 15568024]
34. Firth AE, Patrick WM. Statistics of protein library construction. Bioinformatics 2005;21:3314–3315. [PubMed: 15932904]
35. Patrick WM, Firth AE, Blackburn JM. User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. Protein Eng 2003;16:451–457. [PubMed: 12874379]
36. Drummond DA, Iverson BL, Georgiou G, Arnold FH. Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins. J. Mol. Biol 2005;350:806–816. [PubMed: 15939434]
37. Krahn JM, Kim JH, Burns MR, Parry RJ, Zalkin H, Smith JL. Coupled formation of an amidotransferase interdomain ammonia channel and a phosphoribosyltransferase active site. Biochemistry 1997;36:11061–11068. [PubMed: 9333323]
38. Muchmore CR, Krahn JM, Kim JH, Zalkin H, Smith JL. Crystal structure of glutamine phosphoribosylpyrophosphate amidotransferase from *Escherichia coli*. Protein Sci 1998;7:39–51. [PubMed: 9514258]
39. Shepherd N, Dennis P, Bremer H. Cytoplasmic RNA Polymerase in *Escherichia coli*. J. Bacteriol 2001;183:2527–2534. [PubMed: 11274112]
40. Neidhardt, FC.; Umbarger, HE. Chemical composition of *Escherichia coli*. In: Neidhardt, FC., editor. Escherichia coli and Salmonella. 2nd edit.. 1996. Washington, D.C.: ASM Press; p. 13-16.
41. Lenski RE, Rose MR, Simpson SC, Tadler SC. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2000 generations. Am. Nat 1991;138:1315–1341.
42. Mayans O, Ivens A, Nissen LJ, Kirschner K, Wilmanns M. Structural analysis of two enzymes catalysing reverse metabolic reactions implies common ancestry. EMBO J 2002;21:3245–3254. [PubMed: 12093726]
43. Browne BA, Ramos AI, Downs DM. PurF-independent phosphoribosyl amine formation in *yjgF* mutants of *Salmonella enterica* utilizes the tryptophan biosynthetic enzyme complex anthranilate synthase-phosphoribosyltransferase. J. Bacteriol 2006;188:6786–6792. [PubMed: 16980480]
44. Ramos I, Downs DM. Anthranilate synthase can generate sufficient phosphoribosyl amine for thiamine synthesis in *Salmonella enterica*. J. Bacteriol 2003;185:5125–5132. [PubMed: 12923085]
45. Ramos I, Vivas EI, Downs DM. Mutations in the tryptophan operon allow PurF-independent thiamine synthesis by altering flux *in vivo*. J. Bacteriol. 2008In press.
46. Henn-Sax M, Thoma R, Schmidt S, Hennig M, Kirschner K, Sterner R. Two $(\beta\alpha)_8$-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates. Biochemistry 2002;41:12032–12042. [PubMed: 12356303]
47. Messenger LJ, Zalkin H. Glutamine phosphoribosylpyrophosphate amidotransferase from *Escherichia coli*. Purification and properties. J. Biol. Chem 1979;254:3382–3392. [PubMed: 372191]
48. Gould, SJ. Wonderful Life: The Burgess Shale and the Nature of History. New York: W. W. Norton and Company; 1989.
49. Wolfenden R, Snider MJ. The depth of chemical time and the power of enzymes as catalysts. Acc. Chem. Res 2001;34:938–945. [PubMed: 11747411]
50. Roodveldt C, Tawfik DS. Shared promiscuous activities and evolutionary features in various members of the amidohydrolase superfamily. Biochemistry 2005;44:12728–12736. [PubMed: 16171387]
51. Couñago R, Chen S, Shamoo Y. *In vivo* molecular evolution reveals biophysical origins of organismal fitness. Mol. Cell 2006;22:441–449. [PubMed: 16713575]
52. Miller SP, Lunzer M, Dean AM. Direct demonstration of an adaptive constraint. Science 2006;314:458–461. [PubMed: 17053145]
53. Ivens A, Mayans O, Szadkowski H, Wilmanns M, Kirschner K. Purification, characterization and crystallization of thermostable anthranilate phosphoribosyltransferase from *Sulfolobus solfataricus*. Eur. J. Biochem 2001;268:2246–2252. [PubMed: 11298741]
54. Vogel HJ, Bonner DM. Acetylornithinase of *Escherichia coli*: partial purification and some properties. J. Biol. Chem 1956;218:97–106. [PubMed: 13278318]

55. Patrick WM, Blackburn JM. *In vitro* selection and characterization of a stable subdomain of phosphoribosylanthranilate isomerase. FEBS J 2005;272:3684–3697. [PubMed: 16008567]

56. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. How to measure and predict the molar absorption coefficient of a protein. Protein Sci 1995;4:2411–2423. [PubMed: 8563639]

Abbreviations used

AcTrpD, anthranilate phosphoribosyltransferase from *Acinetobacter baylyi* strain ADP1
Cam, chloramphenicol
CdRP, 1'-(2'-carboxyphenylamino)-1'-deoxyribulose 5'-phosphate
cPRPP, 1α-pyrophosphoryl-2α,3α-dihydroxy-4β-cyclopentanemethanol 5-phosphate
epPCR, error-prone PCR
GFP, green fluorescent protein
HPLC, high performance liquid chromatography
InGP, indoleglycerol-phosphate
IPTG, isopropyl-β-D-thiogalactoside
ORF, open reading frame
PRA, phosphoribosylanthranilate
PRAI, phosphoribosylanthranilate isomerase
PRPP, 5-phosphoribosyl-1-pyrophosphate
PurF, glutamine PRPP amidotransferase
R5P, ribose-5-phosphate
PRTase, phosphoribosyltransferase

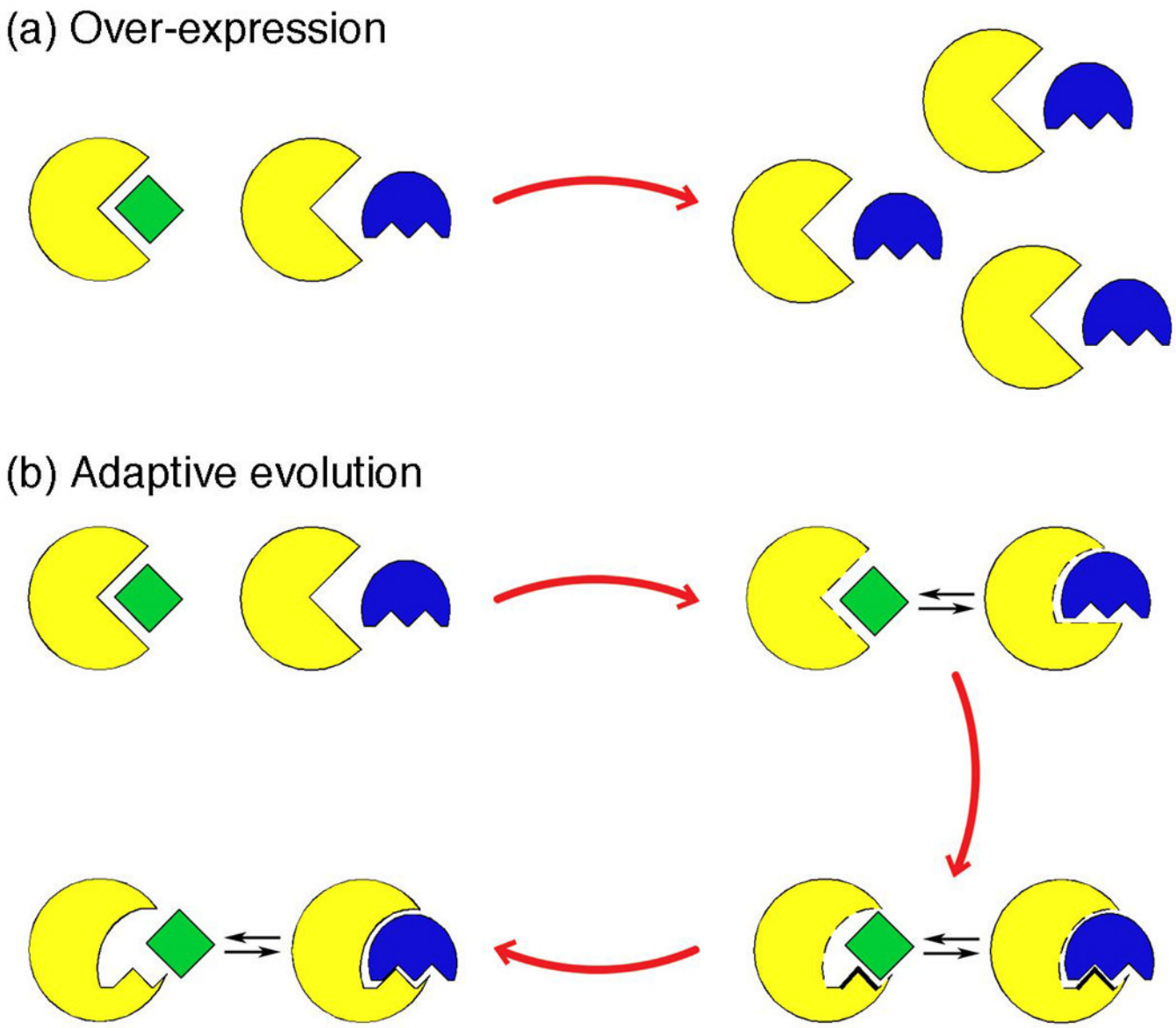

Figure 1.
A two-stage model for the evolution of novel catalytic functions. (a) A promiscuous catalyst has evolved to recognize the diamond-shaped substrate, but also has some low affinity for the alternate substrate. In a changing environment, it is activity against this alternate substrate that confers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

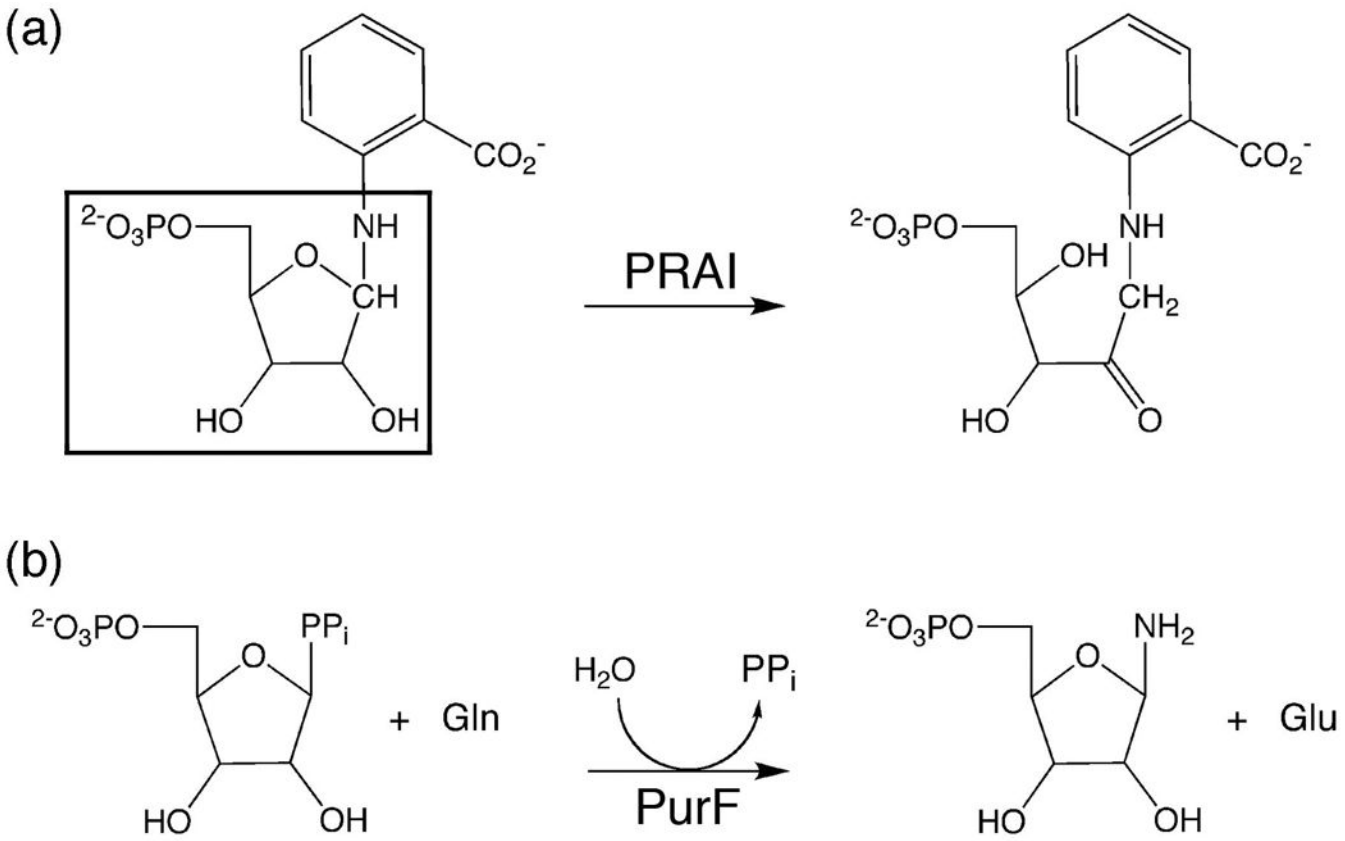

Figure 2.

The reactions catalyzed by the isomerase, PRAI, and the transferase, PurF. (a) PRAI catalyzes the Amadori rearrangement of the aminoaldose PRA to the aminoketose CdRP. The common R5P 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.

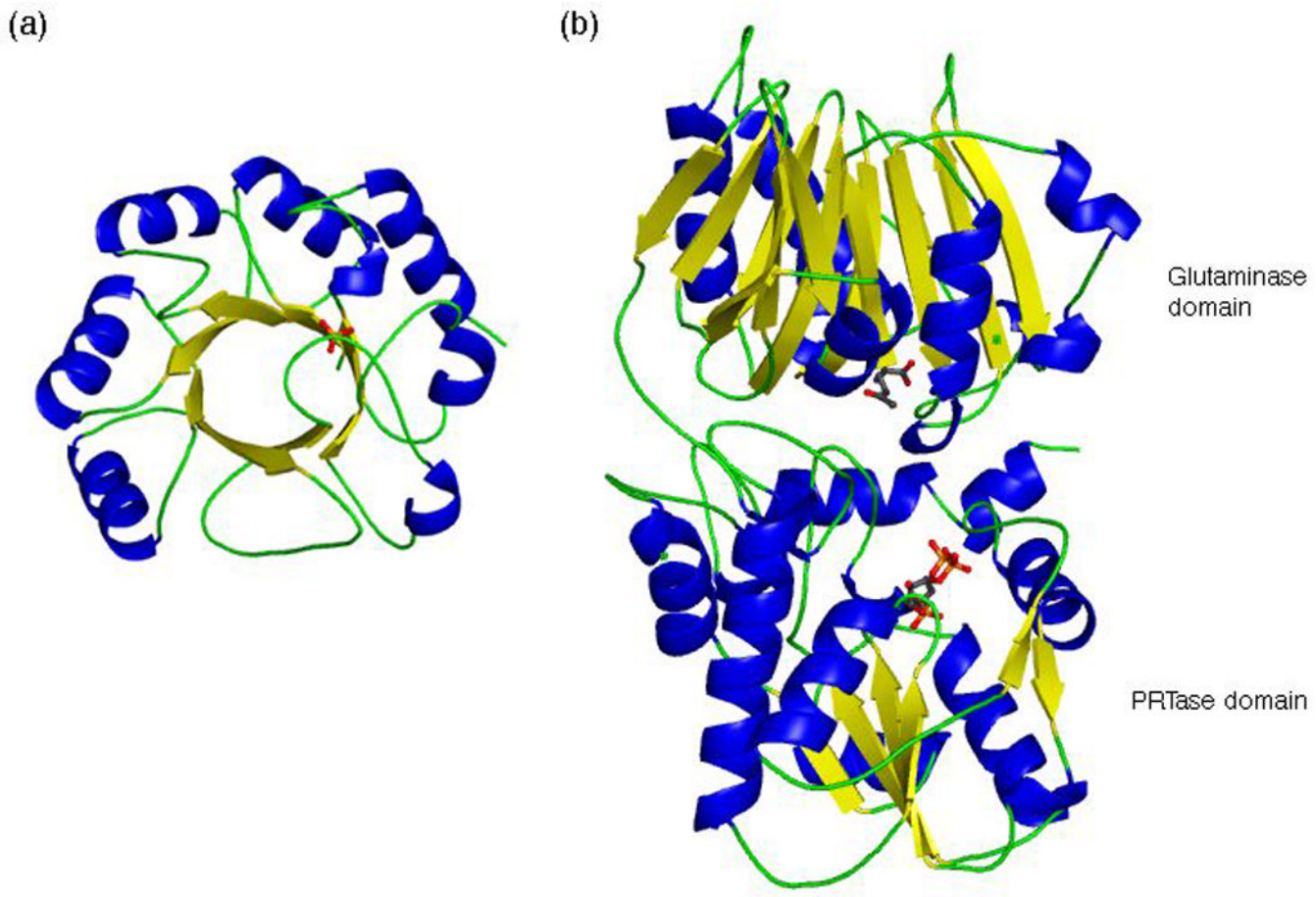

Figure 3.

PRAI and PurF are not homologues. (a) The 197-residue $(\beta\alpha)_8$-barrel protein PRAI (PDB entry 1pii[20]), with a phosphate ion shown in the position it would also occupy as part of the substrate, PRA. (b) One subunit of the homotetramer, PurF (505 amino acids; PDB entry 1ecc[37]). Each PurF monomer contains an N-terminal glutaminase domain with the Ntn hydrolase fold, and a C-terminal type I PRTase domain. The substrate analogues 6-diazo-5-oxo-L-*nor*-leucine and 1α-pyrophosphoryl-2α,3α-dihydroxy-4β-cyclopentanemethanol 5-phosphate (cPRPP) are shown in the glutaminase and PRTase active sites, respectively. Loops have been smoothed for clarity.

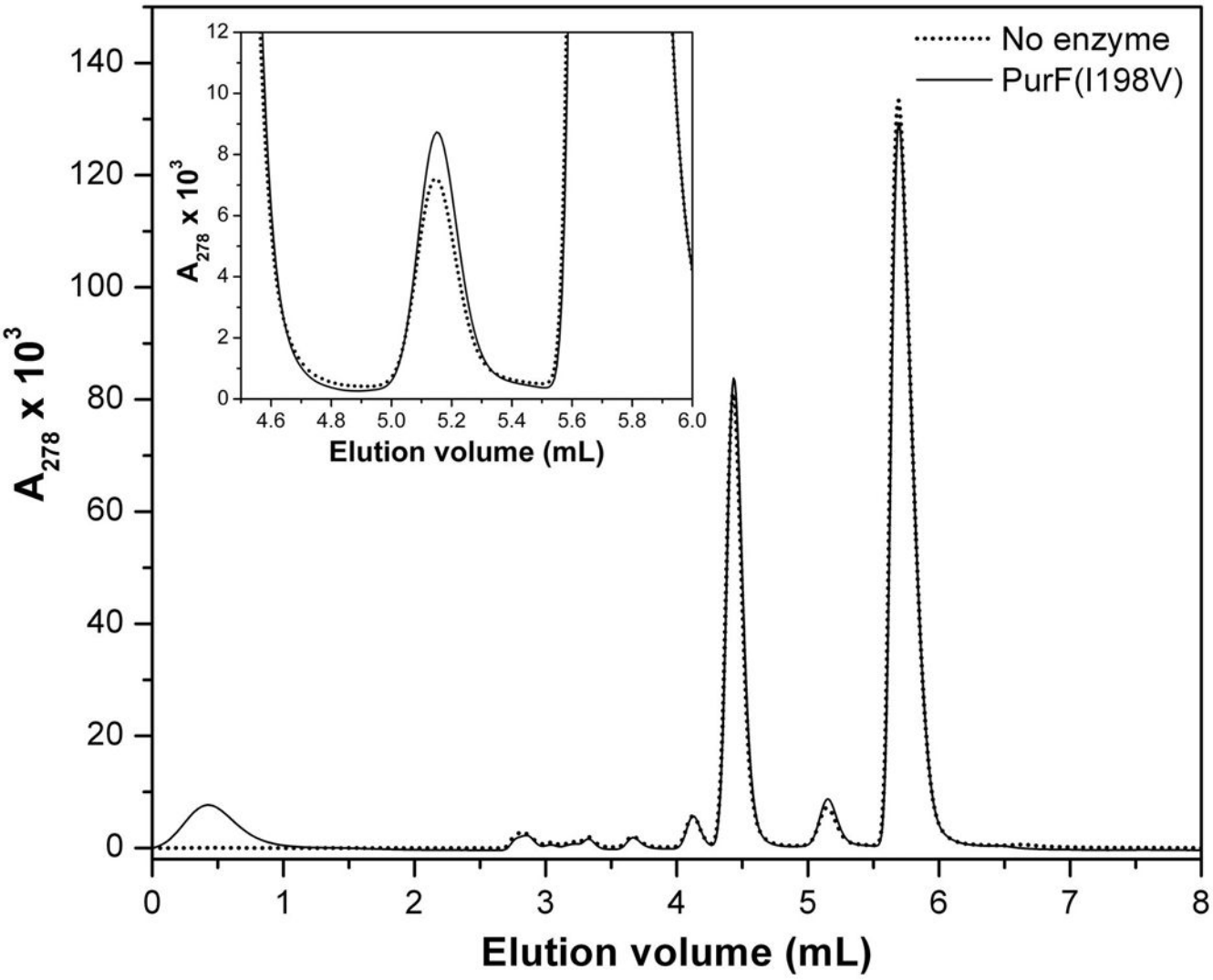

Figure 4.

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.

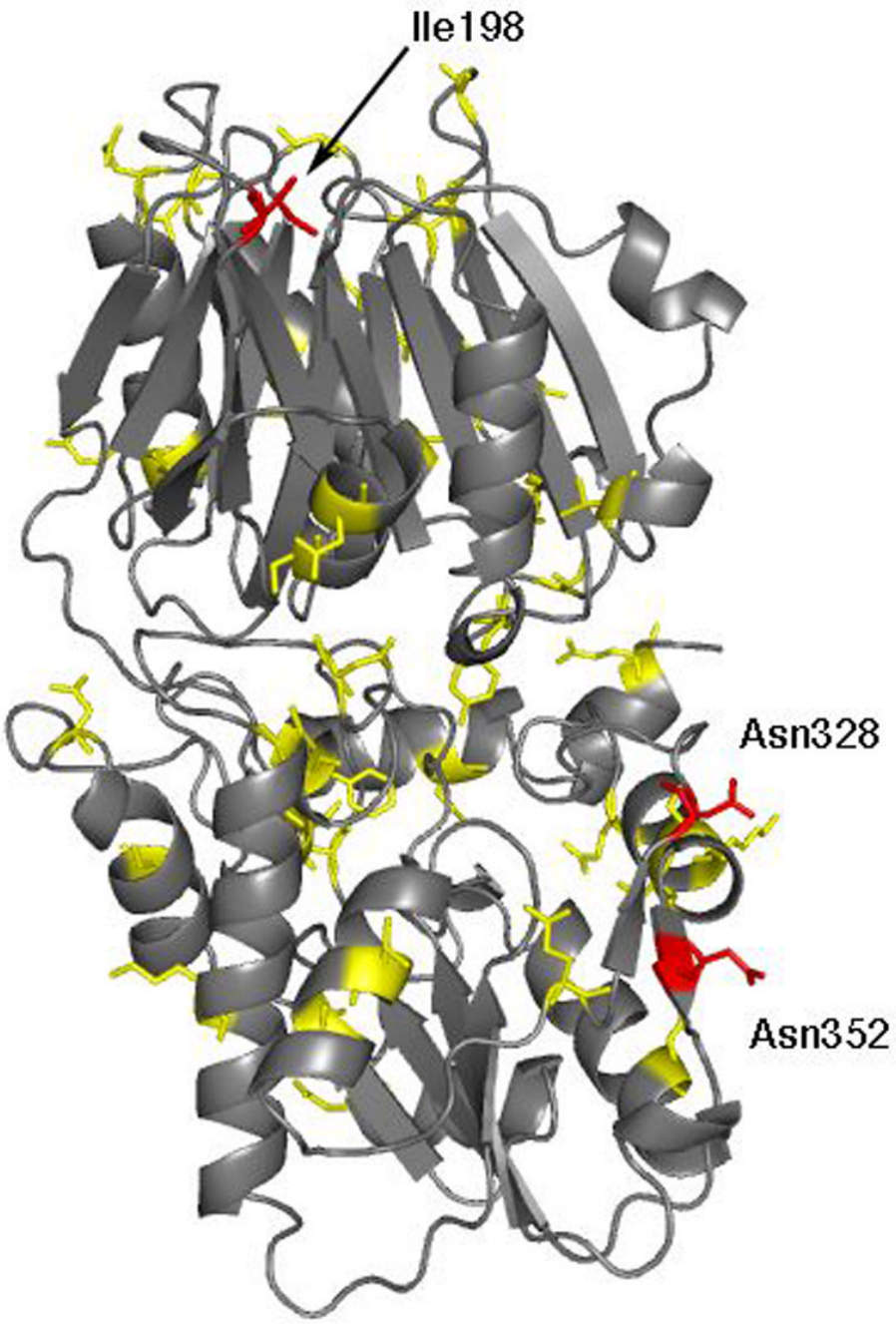

Figure 5.
Mutations in variants selected from the epPCR library, mapped onto the tertiary structure of PurF (PDB entry 1ecc[37]). Residues in yellow are mutated in at least one variant. The position mutated in the parental clone, Ile198, and the sites of recurrent mutations in the PRTase flexible loop (Asn328 and Asn352) are highlighted in red.

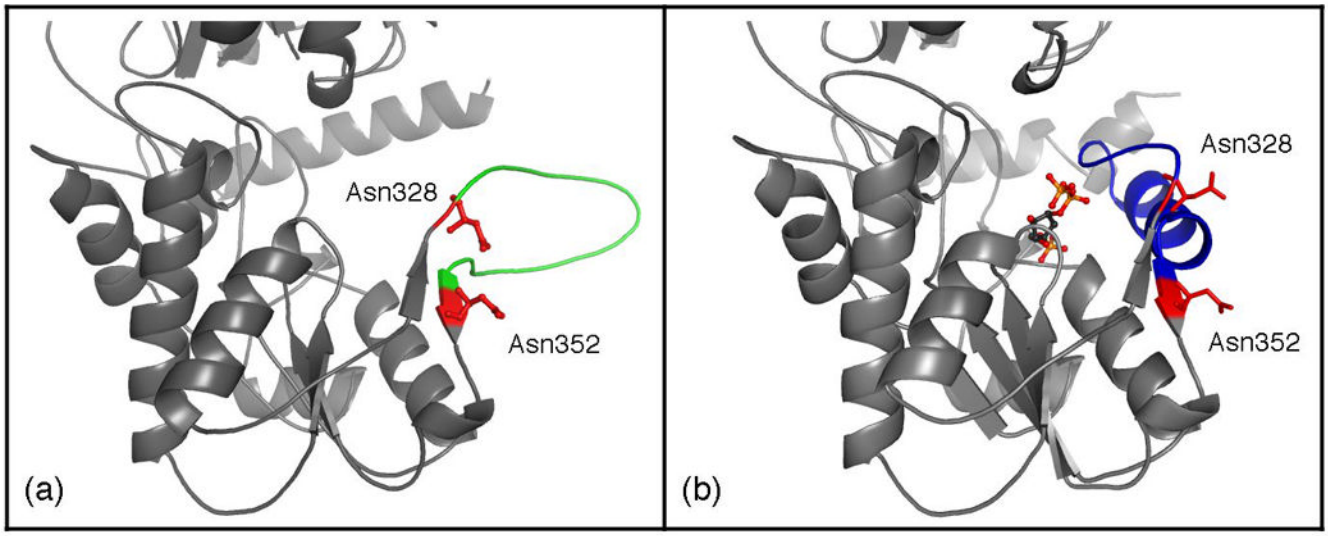

Figure 6.

Asn328 and Asn352 anchor the PRTase flexible loop in (a) the open structure of PurF (PDB entry 1ecj[38]); and (b) the closed, substrate-bound structure (PDB entry 1ecc[37]). The two asparagine residues are highlighted in red, with the loop between them highlighted in green or blue. The PRPP analogue, cPRPP, is shown in ball-and-stick representation in (b).

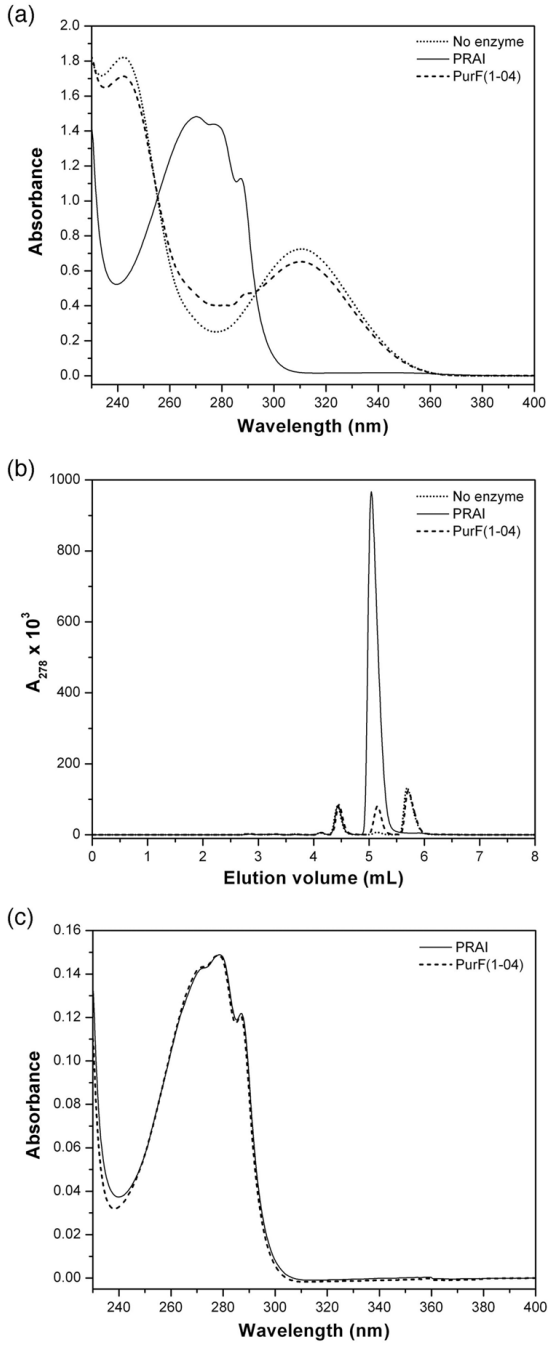

Figure 7.

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_{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,[25] confirming that both PRAI and $(His)_6$-PurF(1-04) are catalyzing the expected isomerization reaction.

Table 1

Improved $(His)_6$-PurF variants isolated from the epPCR library.

Variant	Frequency	Growth[a]	Mutations[b]
PurF(I198V)	[Template]	1	I198V
1-01	4	4	E130V, I198V, F255L, **N328S**, S346P, M382V; 2 silent
1-02	9	3	Y75C, E130V, R162H, I198V, S211G, D215N, F255L, D264E, I313T, V347A, F402Y, E422G; 1 silent
1-04	22	5	I37M, A39T, E85G, P88S, N124S, I198V, K282R, **N328S**; 1 silent
1-16	1	5	V52L, R121H, F136L, E149G, N161Y, I198V, **N328I**, N494S; 2 silent
1-22	1	3	N92Y, G176D, I198V, Q323R, **N352K**, I444V, D480G
1-26	1	2	S80N, F143I, I198V, L311P, Q323R, Q340R, **N352K**, N451Y, Q491R, E495G, N498D, L499P; 3 silent
1-27	1	2	H122L, I135T, I198V, I270V, K345R, **N352K**; 4 silent
1-30	1	2	L115M, G175D, I198V, **N352K**, V461A, H502Y; 2 silent
1-Neg[c]	—	—	A86V, I198V, R329H, D367Y; 1 silent

[a] Qualitative estimate based on rate of colony formation, scored from 1 (PurF(I198V)) to 5 (PurF(1-04)).

[b] Recurring mutations in the flexible substrate binding loop are highlighted in bold typeface.

[c] A negative control from the library that is solubly expressed, but does not complement *E. coli* JMBΔ*trpF*.

Table 2

Relative fitness of strains over-expressing $(His)_6$-PurF(I198V), $(His)_6$-PurF(1-04) and $(His)_6$-IGPS:PRAI.

Competition experiment[a]	# Cell divisions[b]		Relative fitness[c]
	Strain 1	Strain 2	
PurF(I198V) vs. PurF(I198V)	0.49	0.45	1.1 ± 0.1
IGPS:PRAI vs. PurF(I198V)	8.8	1.9	4.8 ± 0.2
PurF(1-04) vs. **PurF(I198V)**	1.8	0.37	4.8 ± 0.7

[a] The GFP-tagged protein is indicated in bold typeface.

[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 mean ± standard error (n = 3).