

# Structural and functional modularity of proteins in the de novo purine biosynthetic pathway

Hui Li, Walter Fast, and Stephen J. Benkovic\*

Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802

Received 17 December 2008; Revised 3 February 2009; Accepted 5 February 2009

DOI: 10.1002/pro.95

Published online 23 February 2009 proteinscience.org

**Abstract:** It is generally accepted that naturally existing functional domains can serve as building blocks for complex protein structures, and that novel functions can arise from assembly of different combinations of these functional domains. To inform our understanding of protein evolution and explore the modular nature of protein structure, two model enzymes were chosen for study, *purT*-encoded glycinamide ribonucleotide formyltransferase (PurT) and *purK*-encoded *N*<sup>5</sup>-carboxylaminoimidazole ribonucleotide synthetase (PurK). Both enzymes are found in the de novo purine biosynthetic pathway of *Escherichia coli*. In spite of their low sequence identity, PurT and PurK share significant similarity in terms of tertiary structure, active site organization, and reaction mechanism. Their characteristic three domain structures categorize both PurT and PurK as members of the ATP-grasp protein superfamily. In this study, we investigate the exchangeability of individual protein domains between these two enzymes and the *in vivo* and *in vitro* functional properties of the resulting hybrids. Six domain-swapped hybrids were unable to catalyze full wild-type reactions, but each hybrid protein could catalyze partial reactions. Notably, an additional loop replacement in one of the domain-swapped hybrid proteins was able to restore near wild-type PurK activity. Therefore, in this model system, domain-swapped proteins retained the ability to catalyze partial reactions, but further modifications were required to efficiently couple the reaction intermediates and achieve catalysis of the full reaction. Implications for understanding the role of domain swapping in protein evolution are discussed.

**Keywords:** protein evolution; de novo purine biosynthetic pathway; domain swapping

## Introduction

Protein modules are stable structural units that either have an independent function or contribute to the

function of the whole protein in cooperation with other protein modules. Modules can be defined on several different scales, ranging from linked enzymes down to small structural motifs. In eukaryotes, enzymes that catalyze consecutive reactions tend to form single, multidomain proteins, presumably to protect reactive intermediates and expedite substrate turnover. For example, GAR synthetase, GAR transformylase, and AIR synthetase catalyze the second, third, and fifth steps in the de novo purine biosynthetic pathway, respectively.<sup>1</sup> These enzymatic activities exist on separate proteins in *Escherichia coli* and *Bacillus subtilis*, but comprise a single trifunctional protein in human and chicken.<sup>2,3</sup> In this example, each linked enzyme can be considered as an individual protein module. Other macromolecules such as collagen and titin obtain unique physical properties critical for their

---

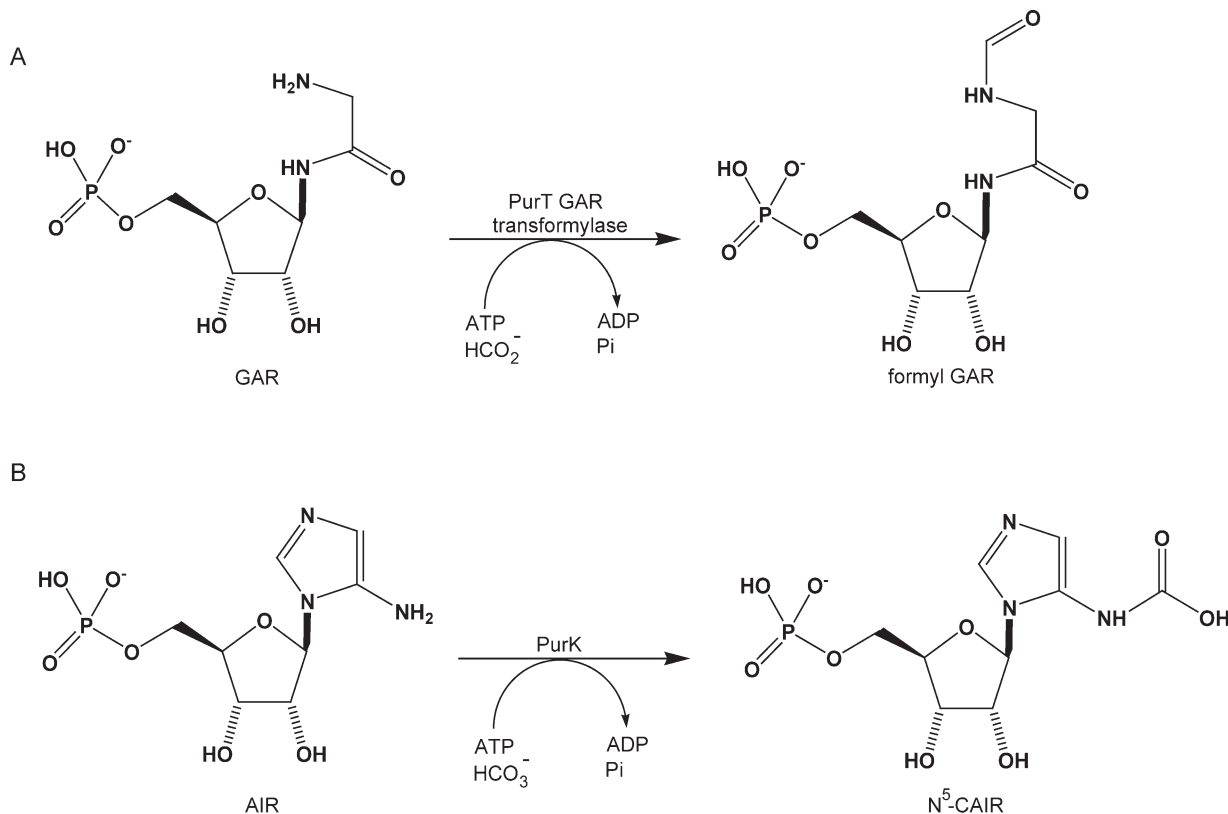
*Abbreviations:* AIR, aminoimidazole ribonucleotide; BC, biotin carboxylase; CPS, carbamoyl phosphate synthetase; GAR, glycinamide ribonucleotide; *N*<sup>5</sup>-CAIR, *N*<sup>5</sup>-carboxylaminoimidazole ribonucleotide.

Hui Li's current address is Department of Pathology, University of Washington, Seattle, Washington 98115.

Walter Fast's current address is Division of Medicinal Chemistry, College of Pharmacy, The University of Texas, Austin, Texas 78712.

Grant sponsor: NIH; Grant number: GM024129.

\*Correspondence to: Stephen J. Benkovic, Department of Chemistry, The Pennsylvania State University, 414 Wartik Laboratory, University Park, PA 16802. E-mail: sjb1@psu.edu



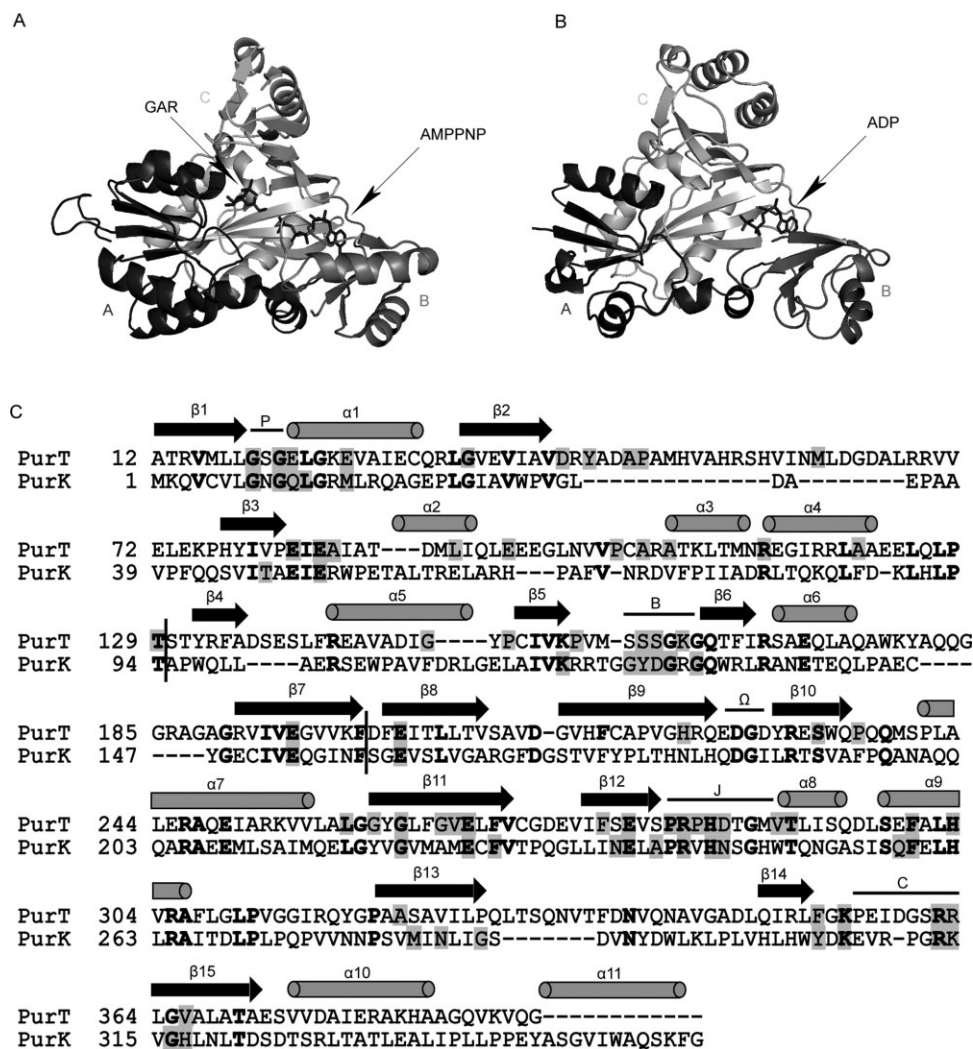
**Figure 1.** Reactions catalyzed by PurT transformylase (A) and PurK (B). The common catalytic steps are the formation of an acyl phosphate intermediate upon ATP cleavage, followed by the nucleophilic attack on the carboxyl carbon of the intermediates by the amino nitrogen of the mononucleotides. The intermediates are formyl phosphate for PurT and carboxyl phosphate (or a  $\text{CO}_2$  breakdown product) for PurK, respectively.

biological functions by forming covalently linked polymers of repetitive subunits<sup>4,5</sup> and each of the tethered monomers can be defined as a protein module. On a scale smaller than these large multidomain protein complexes, protein modules can also be defined within individual proteins. The  $(\beta/\alpha)_8$  barrel structure is the most versatile and most frequently encountered fold among enzymes. The unique barrel structure with eight successive strand-loop-helix motifs allows substitution of individual motifs within the barrel, even fragmentation into half barrels without the disruption of functional or structural integrity.<sup>6–9</sup> Here, a segment as small as the strand-loop-helix motif can be considered a protein module.

Protein modules can also be considered as evolutionary units.<sup>10</sup> It is generally believed that nature evolved novel protein functions through recombination of protein modules taken from a pre-existing repertoire, a process sometimes referred to as “domain shuffling.” Recombination followed by mutational drift can result in higher substrate specificities and enzyme activities in response to selective pressures. Identifying functional protein modules, understanding their reassembly, and determining what additional alterations are required to achieve novel functions will help us to further understand the mechanisms of protein evolu-

tion, and to formulate an efficient design methodology for the creation of new enzymes.<sup>11</sup>

To experimentally investigate how domain swapping may contribute to protein evolution, we chose two enzymes in the de novo purine biosynthetic pathway of *Escherichia coli*, glycinamide ribonucleotide (GAR) formyltransferase (PurT) and  $N^5$ -carboxyaminoimidazole ribonucleotide ( $N^5$ -CAIR) synthetase (PurK), as model enzymes. PurT catalyzes the formylation of GAR using ATP and formate, and PurK catalyzes the carboxylation of aminoimidazole ribonucleotide (AIR) to  $N^5$ -CAIR using ATP and bicarbonate (see Fig. 1). Crystallographic studies and amino acid sequence analysis indicate that both PurT and PurK belong to a newly emerged superfamily of proteins that contain an ATP grasp fold.<sup>12</sup> Other members of this protein superfamily include biotin carboxylase,<sup>13</sup> carbamoyl phosphate synthetase,<sup>14</sup> glutathione synthetase,<sup>15</sup> succinyl-CoA synthetase,<sup>16</sup> D-alanine:D-alanine ligase,<sup>17</sup> and GAR synthetase (PurD), another enzyme in the purine biosynthetic pathway.<sup>18</sup> The ATP-grasp proteins, including PurT and PurK, share a similar molecular architecture that consists of three motifs termed the A-, B-, and C-domains (see Fig. 2).<sup>20,21</sup> These enzymes also share common features in their reaction mechanisms, notably including formation of

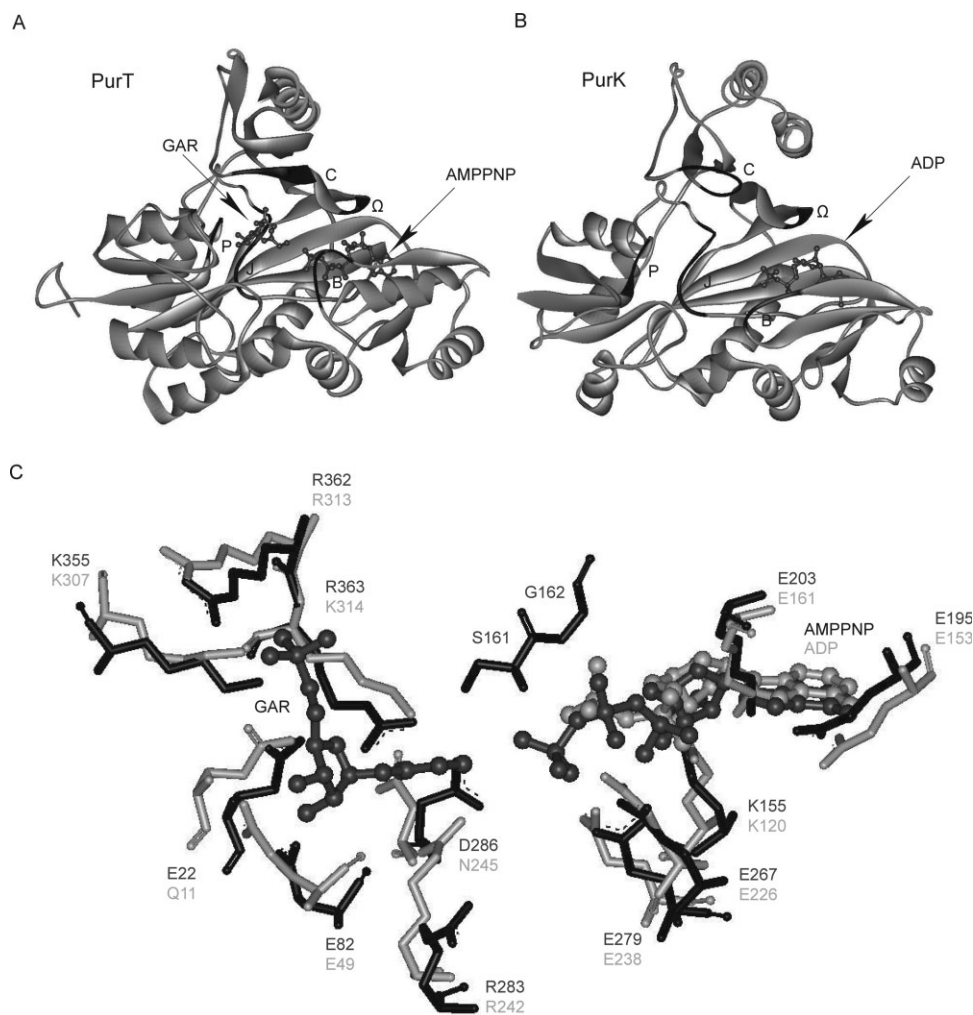


**Figure 2.** Three-dimensional structures of PurT with GAR and 5'-adenylyl imidodiphosphate (AMPPNP) (A), PurK with ADP (B), and a structure-based sequence alignment (C). In (A) and (B), each enzyme is divided into three structural motifs, the A-, B-, and C-domains which are color-coded in black, dark gray, and gray, respectively. The stick representation of the substrates in complex with enzymes is displayed in dark gray. The representations were prepared with Weblab Viewer Pro 3.7 (Molecular Simulations Inc.). In (C), a structure-based alignment was prepared using CE (combinatorial extension of the optimal path).<sup>19</sup> The arrows and cylinders represent the  $\beta$ -sheet and  $\alpha$ -helix secondary structures, respectively. Conserved residues in each enzyme are masked in gray and identical residues between PurT and PurK are shown bold. The vertical black bars label the boundaries of the A-, B-, and C-domains. Additionally, five conserved loop regions, the P-loop, B-loop,  $\Omega$ -loop, J-loop, and C-loop, are labeled.

reactive acyl phosphate intermediates.<sup>12</sup> Despite the low amino acid sequence identity between PurT and PurK (27%), their three-dimensional structures are superimposable with a root-mean-square deviation of 1.1 Å over 177 structurally equivalent  $\alpha$ -carbons.<sup>21</sup> As shown in Figure 2(C), a majority of the conserved residues involved in substrate binding and/or catalysis are positioned on five structurally conserved loop regions, namely, the P-loop, B-loop,  $\Omega$ -loop, J-loop, and C-loop. In both PurT and PurK, the residues on the B- and  $\Omega$ -loop participate in binding ATP. In PurT, GAR is bound in a pocket formed by residues of the P-, J-, and C-loops. At the time this work was initiated, the structure of PurK was only determined in the presence of bound ADP or sulfate, but it is likely that the ribo-

tide substrates for PurT and PurK (GAR and AIR, respectively) bind in similar manners [Fig. 3(C)]. The similarity between a sulfate-bound PurK structure with BC and CPS structures led to the speculation that the residues in the J-loop are involved in binding bicarbonate, formate and the acyl phosphate intermediates.<sup>20</sup> The B-loop (also called the T-loop) is also proposed to participate in formate binding and sequestration of reaction intermediates.<sup>22</sup>

In addition to structural homology, PurT and PurK also share common features in their catalytic mechanisms: ATP cleavage, formation of acyl phosphate intermediates, and nucleophilic attack by the substrates' amino groups, all features also shared by other ATP-grasp enzymes. Isotope labeling, positional



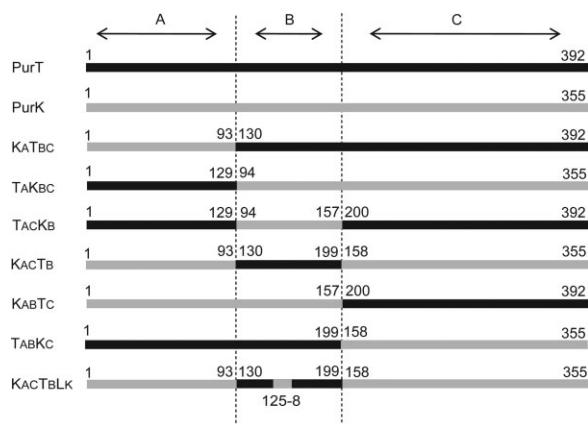
**Figure 3.** Overall architecture of the active sites in PurT (A), PurK (B), and residues within the PurT and PurK active sites (C). In (A) and (B), the B-loop, the Ω-loop, the P-loop, the J-loop, and the C-loop, are shown in black, with the backbones in gray. The GAR, AMPPNP, and ADP molecules are shown in ball-and-stick representation, and colored in dark gray. In (C), the residues involved in substrate binding and catalysis in the PurT (black) and PurK (gray) active sites are shown. Residues are numbered based on the *Escherichia coli* enzymes. Superpositioning of the PurT and PurK structures was done manually using Weblab Viewer Pro 3.7 (Molecular Simulations Inc.).

isotope exchange studies, and tests of chemical and kinetic competence are consistent with formyl phosphate serving as an intermediate in the PurT reaction.<sup>23,24</sup> The observation of quantitative <sup>18</sup>O transfer from [<sup>18</sup>O]-bicarbonate into [<sup>18</sup>O]-Pi by PurK in the presence of AIR and ATP led to the proposal that carboxyl phosphate is an intermediate in the PurK reaction.<sup>25</sup> Carboxyl phosphate has not been successfully trapped, most likely due to the low stability of this compound in water (half life <70 ms).<sup>26</sup> The structural and mechanistic similarities between PurT and PurK make them attractive model proteins for understanding how enzymatic functions might evolve through domain shuffling.

Our interest in PurT and PurK as a model system was also inspired by previous successful domain swapping between two other proteins in the purine biosynthetic pathway, PurN and PurU.<sup>27</sup> PurN, an alternative GAR transformylase that operates by a different mech-

anism than PurT, catalyzes the transfer of a formyl group from *N*<sub>10</sub>-formyl-tetrahydrofolate to the free amino group of GAR to yield formyl-GAR and tetrahydrofolate. PurU, an *N*<sub>10</sub>-formyl-tetrahydrofolate hydrolase, hydrolyses *N*<sub>10</sub>-formyl-tetrahydrofolate to formate and tetrahydrofolate. Sequence analysis of *E. coli purN* and *purU* genes shows a significant amino acid identity (~60%) in the C-terminal region, in which the *N*<sub>10</sub>-formyl-tetrahydrofolate binding pocket of PurN is situated.<sup>28,29</sup> Two functional hybrid proteins with PurN activity were created with the C-terminal region of PurN replaced by various lengths of the C-terminal region of PurU,<sup>27</sup> and the solubility and specific activity of these hybrids were further improved using a combinatorial mutagenesis approach.<sup>30</sup>

Herein, we sought to create rationally designed domain-swapped hybrid proteins using a different pair of homologous proteins selected from the purine biosynthetic pathway. Each individual A-, B-, and C-



**Figure 4.** Schematic overview of hybrid enzyme composition. The sequences of the two parental genes, PurT and PurK, are depicted in black and gray, respectively. The boundary of each domain is marked by dotted lines. The numbers represent amino acid residue positions in each parental protein.

domain in PurT and PurK was substituted by its counterpart from the other protein in order to generate six hybrid proteins. These constructs are characterized by *in vivo* functional assays and *in vitro* steady-state kinetics. Although none of the hybrids can catalyze the full reactions seen with wild-type enzymes, they can catalyze partial reactions. Additional loop swapping modifications of the hybrid proteins successfully result in one construct that displays near wild-type PurK activity. Implications for understanding how enzymes may evolve by using domain shuffling are discussed.

## Results and Discussion

### *In vivo* activity of the hybrid proteins

To swap each of the functional domains of PurK and PurT, six chimeric proteins,  $T_A K_{BC}$  (PurT 1-129 + PurK 94-355),  $K_A T_{BC}$  (PurK 1-93 + PurT 130-392),  $K_A C T_B$  (PurK 1-93 + PurT 130-199 + PurK 158-355),  $T_A C K_B$  (PurT 1-129 + PurK 94-157 + PurT 200-392),  $T_{AB} K_C$  (PurT 1-199 + PurK 158-355), and  $K_{AB} T_C$  (PurK 1-157 + PurT 200-392), were generated on the basis of structural and sequence alignments between PurK and PurT (see Fig. 4). Each hybrid has one of the A-, B-, or C-domains replaced by the corresponding domain from the other protein.

The functions of the hybrid proteins, as well as the two parental proteins, were initially tested by *in vivo* complementation in *E. coli* auxotrophic strains. TX680F' was used to test for PurT activity,<sup>31</sup> and CSH26 for PurK activity.<sup>32</sup> With inactivated chromosomal copies of *purT* and *purN*, the TX680F' strain has been shown to detect a GAR transformylase activity ten thousand times less active than wild-type enzyme.<sup>31</sup> The limit of detection for the CSH26 strain has not been determined. When bearing plasmids encoding wild-type PurT or PurK enzymes, the

TX680F' and CSH26 strains, respectively, grew reproducibly on minimal medium after a 1 day incubation. In each case, complementation of each deficiency was specific as each wild-type enzyme was only able to support the growth of its respective auxotrophic strain. However, none of the six domain-swapped hybrid proteins described earlier was able to complement the growth of either auxotrophic strain, indicating that these hybrids are either nonfunctional, or that they have activities lower than the detection limit of the *in vivo* complementation assays.

To improve the *in vivo* complementation assays, several confounding variables were minimized. Previous results indicated that with low frequency (usually 1 in  $\sim 10^5$ ) some auxotrophic cells were able to bypass selective pressures and grow on minimal media by recovering their wild-type genes through recombination, a phenomena known as gene reversion. Recombination usually occurs between the inactive chromosomal copy and a gene fragment in the plasmid to provide a fully functional gene. To eliminate background growth caused by gene reversions, two new auxotrophic *E. coli* strains, PurT(-) and PurK(-), were generated by a complete deletion of the chromosomal *purT* and *purK* genes, respectively.<sup>33</sup> Variability in protein expression was also considered. In many cases, poor protein expression and/or stability, rather than the lack of the sufficient catalytic activity, can cause a negative result during *in vivo* complementation assays. The situation is more problematic for hybrid proteins due to the unfavorable interactions likely formed at interfaces between domains derived from different parental proteins—these may lead to incorrect folding and aggregation. Here, hybrids were expressed with an N-terminal maltose binding protein fusion (MBP) and at lower incubation temperatures (25°C) to minimize inclusion body formation. The combination of lower temperatures and the use of N-terminal fusion proteins increased expression levels and solubility of hybrid proteins, as gauged by SDS-PAGE. Expression levels of all test proteins in the auxotrophic strains are in a similar range, also estimated by SDS PAGE. These improvements notwithstanding the six hybrid proteins described earlier do not complement either auxotrophic strain.

Structural and biochemical studies suggest that precise alignments of the three substrate binding pockets is critical for the PurT or PurK reactions to occur, partially due to the fact that the active site is situated at the center of the three domains' interfaces.<sup>20,21,23-25</sup> Considering the low amino acid sequence identity (27%) between PurT and PurK, it is likely that unfavorable interactions made between regions in the hybrids that were derived from different parental proteins may disrupt the relative positioning of the three protein modules. Misalignment of the substrate binding pockets could subsequently lead to inefficient coupling of reaction intermediates and a loss of overall

**Table I.** Summary of Kinetic Data for PurT, PurK, and the Hybrid Enzyme  $K_{AB}T_C$ 

Enzyme	$K_m$ (GAR) ( $\mu M$ )	$K_m$ (AIR) ( $\mu M$ )	$K_m$ (ATP) ( $\mu M$ )	$K_m$ ( $HCO_2^-$ ) ( $\mu M$ )	$K_m$ ( $HCO_3^-$ ) ( $\mu M$ )	$k_{cat}$ ( $s^{-1}$ )	$k_{cat}/K_m$ (GAR) ( $\mu M^{-1} s^{-1}$ )	$k_{cat}/K_m$ (AIR) ( $\mu M^{-1} s^{-1}$ )	$k_{cat}/K_m$ ( $HCO_2^-$ ) ( $\mu M^{-1} s^{-1}$ )	$k_{cat}/K_m$ ( $HCO_3^-$ ) ( $\mu M^{-1} s^{-1}$ )
PurT <sup>a</sup>	10.1 ± 0.5	ND <sup>b</sup>	45 ± 12	319 ± 15	ND	37.6 ± 0.8	38	ND	0.84	0.12
PurT <sup>c</sup>	36 ± 5	ND	85 ± 14	1700 ± 310	ND	2.1 ± 0.1	0.058	ND	0.025	1.2 × 10 <sup>-3</sup>
$K_{AB}T_C$ <sup>c</sup>	ND	ND	1400 ± 300	ND	ND	(9 ± 0.9) × 10 <sup>-3</sup>	ND	ND	6.4 × 10 <sup>-6</sup>	ND
PurK <sup>d</sup>	ND	26	90	ND	ND	52	ND	2.0	0.58	ND
PurK <sup>c</sup>	ND	232 ± 20	ND	ND	ND	26.4 ± 1.5	ND	0.11	ND	ND

<sup>a</sup> Adapted from Ref. 24.

<sup>b</sup> ND indicates the value was not determined.

<sup>c</sup> Fused with an N-terminal MBP domain.

<sup>d</sup> Adapted from Ref. 34.

catalytic activity. Therefore, the ability of the hybrid proteins to catalyze partial reactions was investigated.

### The ATPase activity of wild-type and hybrid proteins

The overall reactions catalyzed by wild-type PurT and PurK can be broken down into three partial reactions: an ATPase activity that cleaves ATP to ADP, a kinase activity that forms the acyl phosphate intermediates (formyl phosphate for PurT and carboxyl phosphate for PurK), and a transferase activity that catalyzes the attack of a mononucleotide on the reactive acyl phosphate intermediate (or, in the case of purK, possibly the CO<sub>2</sub> breakdown product). *In vivo* studies indicate that all of the six hybrid proteins generated by rational domain swapping between PurT and PurK lack the capability to complement auxotrophic strains, which select for the full reaction. However, *in vitro* studies can reveal whether any of these hybrids can catalyze the ATPase or acyl phosphate-forming partial reactions.

Wild-type PurT and PurK and the six hybrid proteins were overexpressed as N-terminal MBP fusions and purified to homogeneity, as described in Materials and Methods. The kinetic properties of the wild-type enzymes (PurT and PurK) were studied in detail (Table I). With an N-terminal MBP fusion, PurT and PurK still display a dependence of their ATPase activity on the presence of ribonucleotides, GAR for PurT and AIR for PurK (Table II), which has been previously demonstrated using wild-type PurT and PurK constructs that lack the MBP fusion.<sup>24,34</sup> However, the catalytic capabilities of both wild-type enzymes were significantly impaired by addition of the N-terminal MBP fusion (Table I). In the most severe case, the  $k_{cat}$  value of PurT was lowered ~19-fold when fused to MBP. Structural and biochemical studies show that wild-type PurT and PurK form homodimers in solution, with oligomerization mediated by the A- and C-domains.<sup>20,21,34</sup> It is possible that the MBP domain could disrupt oligomerization and result in a decrease of catalytic activity.

For the two wild-type and six hybrid proteins, all which bear N-terminal MBP fusions, ATPase activities were determined using different combinations of substrates (Table II). Because of the limited availability of substrates, the  $K_M$  and  $k_{cat}$  values for ATPase activity of only one hybrid protein,  $K_{AB}T_C$ , was determined (Table I). Although each of the six hybrids do not catalyze the full wild-type reaction, all hybrids do catalyze the ATPase partial reaction. The observation of ATPase activity in all six hybrid proteins strongly suggest that a functional hybrid ATP binding pocket can be constructed between B- and C-domains taken from different parental proteins. However, catalysis by the hybrid ATP binding pockets is severely impaired. When compared with wild-type enzymes, the ATPase activity of each hybrid protein is typically lower by one or two

**Table II.** *ATPase Activity of Wild-Type and Hybrid Enzymes*

Enzyme	Specific activity ( $\mu\text{mol}/\text{min}/\text{mg}$ ) for production of ADP in the presence of various substrates <sup>a</sup>						
	ATP	ATP HCO <sub>3</sub> <sup>-</sup>	ATP AIR	ATP AIR HCO <sub>3</sub> <sup>-</sup>	ATP HCO <sub>2</sub> <sup>-</sup>	ATP GAR	ATP GAR HCO <sub>2</sub> <sup>-</sup>
PurT	0.021	0.031	0.033	0.030	0.16	4.64	5.14
PurK	0.23	1.09	8.72	10.0	0.35	0.46	0.40
K <sub>A</sub> T <sub>BC</sub>	0.45	0.40	0.39	0.41	0.47	0.54	0.61
T <sub>A</sub> K <sub>BC</sub>	0.66	0.69	0.91	0.96	0.65	0.61	0.67
T <sub>AC</sub> K <sub>B</sub>	0.073	0.073	0.087	0.13	0.061	0.10	0.099
K <sub>AC</sub> T <sub>B</sub>	0.028	0.026	0.053	0.077	0.022	0.033	0.037
K <sub>AB</sub> T <sub>C</sub>	0.16	0.15	0.14	0.14	0.14	0.10	0.12
T <sub>AB</sub> K <sub>C</sub>	0.26	0.23	0.25	0.27	0.20	0.22	0.28

<sup>a</sup> Specific activities were determined by measuring ATPase activities under saturating concentrations of each substrate. The standard error for each value is less than 20%.

orders of magnitude. Additionally, the  $K_M$  value for ATP is increased more than 15-fold in the K<sub>AB</sub>T<sub>C</sub> hybrid.

The hybrid proteins with the highest ATPase activities, T<sub>A</sub>K<sub>BC</sub> and K<sub>A</sub>T<sub>BC</sub>, only have swapped A-domains. Therefore, the ATP binding pocket in these hybrids is formed by B- and C-domains taken from the same parental protein, and likely form a near-native binding pocket. Hybrid proteins with swapped C-domains, T<sub>AB</sub>K<sub>C</sub> and K<sub>AB</sub>T<sub>C</sub>, have ATPase activities about 3–5 fold less active than T<sub>A</sub>K<sub>BC</sub> and K<sub>A</sub>T<sub>BC</sub>, presumably because the ATP binding pocket is distorted due to an imperfect interface formed between the B- and C-domains taken from different parental proteins. Hybrids with the lowest ATPase activity, T<sub>AC</sub>K<sub>B</sub> and K<sub>AC</sub>T<sub>B</sub>, are ~10–20 fold slower than T<sub>AB</sub>K<sub>C</sub> and K<sub>AB</sub>T<sub>C</sub>, and have the most potentially disruptive substitutions including a hybrid ATP binding pocket and the insertion of a foreign domain. Interestingly, the comparable ATPase activities exhibited by each pair of hybrid proteins suggest that a similar impact of each domain swap in PurT and PurK is made on the ATPase activity. Unlike wild-type enzymes, the hybrid proteins have lost the dependence of their ATPase activities on the presence of ribonucleotides, suggesting that the observed synergy between ATP and ribonucleotide binding pockets in wild-type proteins is now uncoupled in the hybrids (Table I).

### Formation of acyl phosphate intermediates

Like reactions catalyzed by other ATP grasp proteins, the proposed PurT and PurK mechanisms involve formation of an acyl phosphate intermediate. Marolewski and coworkers have shown that during PurT catalysis, formyl phosphate is a chemically and kinetically competent intermediate.<sup>23</sup> Notably, the generation of formyl phosphate was detected by using a PurT mutant with a single mutation, G162I. G162 is located on the B-loop (also called the T-loop) of PurT, which contains residues that interact with the phosphate groups of ATP. In wild-type PurT, the formyl phosphate intermediate is well protected in the active site. However,

the G162I mutation is proposed to disrupt the B-loop structure and lead to dissociation of formyl phosphate into solution, enabling its detection. In addition to the biologically relevant transformation of GAR into formyl GAR, PurT can catalyze a related reaction. In the absence of GAR, PurT can catalyze the conversion of ATP and acetate to form acetyl phosphate. This observed reaction is also consistent with the proposal of an acyl phosphate reaction intermediate. In wild-type PurK, the proposed carboxyl phosphate reaction intermediate has not been trapped. This putative intermediate would be very reactive, and can break down to release Pi and CO<sub>2</sub>, which has also been suggested as the operative electrophile.

The capabilities of the wild-type and hybrid proteins to produce acetyl phosphate, formyl phosphate, and carboxyl phosphate were assayed by trapping of free acyl phosphates as hydroxamates, which can be visualized spectrophotometrically as described in the

**Table III.** *Production of Acyl Phosphates by Wild-Type and Hybrid Enzymes*

Enzyme	Specific activity ( $\mu\text{mol}/\text{min}/\text{mg}$ ) $\times 10^{-4}$ for production of acyl phosphates using various substrates <sup>a</sup>		
	Acetyl phosphate	Formyl phosphate	Carboxyl phosphate
PurT <sup>b</sup>	$4.4 \times 10^3$	NA <sup>c</sup>	NA
PurT <sup>d</sup>	715	ND <sup>e</sup>	ND
PurK <sup>d</sup>	ND	ND	ND
K <sub>A</sub> T <sub>BC</sub> <sup>d</sup>	3.8	2.8	ND
T <sub>A</sub> K <sub>BC</sub> <sup>d</sup>	8.6	4.0	ND
T <sub>AC</sub> K <sub>B</sub> <sup>d</sup>	11.1	1.7	ND
K <sub>AC</sub> T <sub>B</sub> <sup>d</sup>	0.61	0.43	ND
K <sub>AB</sub> T <sub>C</sub> <sup>d</sup>	10.7	7.7	ND
T <sub>AB</sub> K <sub>C</sub> <sup>d</sup>	0.80	0.75	ND

<sup>a</sup> Specific activities were determined under saturating concentrations of each substrate. The standard error is less than 10%.

<sup>b</sup> Adapted from Ref. 24.

<sup>c</sup> NA not applicable.

<sup>d</sup> Fusion protein with an N-terminal MBP domain.

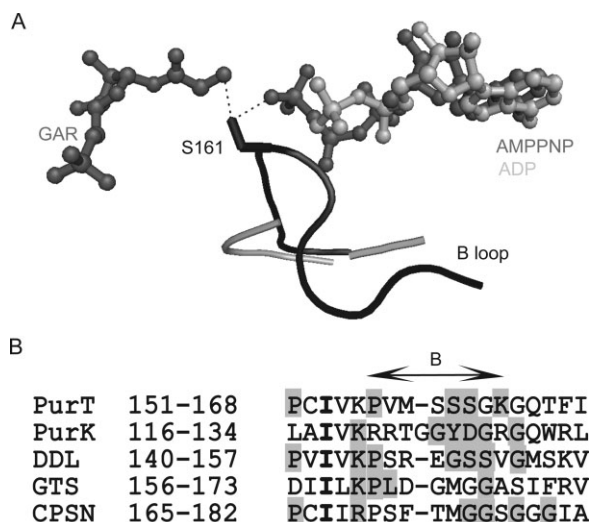
<sup>e</sup> ND not detected.

Materials and Methods section (Table III). Using this methodology, no evidence for carboxyl phosphate was observed, but this could be explained by the short lifetime of this reactive species. Wild-type PurK also does not produce detectable amounts of formyl phosphate or acetyl phosphate, but formate and acetate are not the natural substrates of this enzyme. Wild-type PurT, both with and without an N-terminal MBP fusion, is able to produce detectable levels of acetyl phosphate but not formyl phosphate. Presumably no formyl phosphate is trapped because this intermediate does not dissociate from the wild-type enzyme at appreciable levels.

In contrast to the wild-type proteins, all hybrid proteins showed the ability to generate acetyl phosphate and formyl phosphate. Additionally, wild-type PurT and PurK proteins do not produce ADP in the absence of GAR or AIR, respectively, but the rates of acyl phosphate production by the hybrid proteins were not affected by the presence of either GAR or AIR. For each hybrid protein except T<sub>ACK</sub><sub>B</sub>, the catalytic activity for production of acetyl phosphate is in reasonable agreement with that for the production of formyl phosphate, suggesting no preference for formate or acetate as recipient. Compared to their ATPase activities, the hybrid proteins catalyze the production of acyl phosphate less efficiently, by two orders of magnitude. For wild-type PurT, the ratio of trapped acetyl phosphate to ADP was  $0.70 \pm 0.15$ .<sup>24</sup> Because of a highly optimized architecture in the active site of the wild-type enzyme, the unaccounted for acetyl phosphate is likely hydrolyzed in solution or at the active site before being trapped. In the hybrid proteins, the ratio of trapped acetyl phosphate to ADP decreased to less than 0.01. Therefore, more than 99% of ATP was hydrolyzed and released into solution as ADP and Pi without being transferred to produce detectable acetyl or formyl phosphate. The significant drop in the efficiency of Pi transfer in the hybrid proteins is likely a result of incompatibilities in the swapped protein domains or other missing factors.

### Generation of an active chimeric protein with PurK activity

Although the six hybrid proteins described earlier can all catalyze partial reactions, none can efficiently couple the reaction intermediates to achieve catalysis of the full reaction. Presumably, either misalignment of the swapped domains or other missing protein features leads to the inefficient coupling and prerelease of intermediates into solution. As reported previously, a single mutation in PurT, G162I, leads to reaction uncoupling similar to what is observed in these hybrid proteins. Mutation of this glycine to a bulky isoleucine residue greatly reduces the overall catalytic activity of PurT, increases the  $K_M$  of formate, and allows “leaking” of the formyl phosphate intermediate into solution where it can be trapped.<sup>23</sup> This amino acid position is located on the B-loop (also called the T-loop),



**Figure 5.** Comparison of the B loop (also called the T-loop) in PurT and PurK structures (A), and the consensus sequences from five ATP grasp proteins (B). In (A), the B loop regions in PurT and PurK are colored in black and gray, respectively. The GAR, AMPPNP, and ADP molecules are shown in ball-and-stick representation, with GAR, AMPPNP colored in dark gray and ADP colored in gray. The residue S161 in PurT is depicted in stick representations, and the hydrogen bonds between its hydroxyl group and GAR and AMPPNP are shown as dotted lines. The B loop of PurK (gray) is partially disordered. Panel (B) shows structure-based alignment of the B-loop sequences from PurT, PurK, D-alanine:D-alanine ligase (DDL), glutathione synthetase (GTS), and carbamoyl phosphate synthetase (CPS N-domain). Conserved residues in each enzyme are masked in gray, and identical residues among all five sequences are shown bold.

and is where further efforts were directed to improve the domain-swapped hybrids.

In ATP grasp proteins, the B-loop connects the second and third strands of the B-domain's  $\beta$ -sheet, and is disordered when a bound ligand is not present. The conformation of the B-loop is even dependent on the moiety occupying the  $\gamma$ -phosphate position of ATP analogs.<sup>22</sup> The B-loop is located near the central interfaces of all three domains and likely helps to sequester reaction intermediates. Of the five loop regions constituting the active site in PurT, only the residues in the B-loop are involved in the binding of two substrates, ATP and GAR, as the hydroxyl group of Ser161 is in hydrogen bond distance to both the  $\gamma$ -phosphate of ATP and the amine group of GAR (see Fig. 5). Although the B-loop is not visible in the ADP-bound PurK structure (presumably due to the lack of a  $\gamma$ -phosphate group), this loop is believed to play a similar role in binding of ATP and AIR in PurK. In general, the sequence of B-loops are rich in glycine and serine, but the PurK sequence is atypical and bears conserved tyrosine and aspartate residues. In summary, because of the proximity of the B-loop to both substrate-binding sites, its proposed role in



**Table IV.** Kinetic Parameters PurK and  $K_{AC}T_B L_K^a$ 

Enzyme	$k_{cat}$ ( $s^{-1}$ )	$K_M$ (ATP) ( $\mu M$ )	$K_M$ (AIR) ( $\mu M$ )	$k_{cat}/K_M$ (AIR) ( $\mu M^{-1} s^{-1}$ )	$k_{cat}/K_M$ (ATP) ( $\mu M^{-1} s^{-1}$ )
PurK	$5.9 \pm 0.5$	$48 \pm 8.5$	$66 \pm 16$	0.089	0.12
$K_{AC}T_B L_K$	$0.87 \pm 0.02$	$6.8 \pm 0.7$	$10.3 \pm 1.2$	0.084	0.13

<sup>a</sup> Experimental procedures described in Materials and Methods.

sequestering reactive intermediates and the sequence divergence in this loop between PurT and PurK enzymes, we focused on making further modifications to the B-loops in the partially functional domain-swapped hybrid proteins.

To evaluate the potential of the B-loop to improve catalysis, six additional protein constructs were created and selected *in vivo* for PurT and PurK activities. In each protein, a small portion of the B-loop derived from PurT (S159, Ser160, and Ser161, abbreviated as  $L_T$ ) is replaced by the corresponding sequence in PurK (Gly125, Tyr126, Asp127, and Gly128, abbreviated as  $L_K$ ), and vice versa. This procedure generated six proteins,  $T_{AC}K_B L_T$ ,  $K_{AC}T_B L_K$ ,  $K_{AB}T_C L_T$ ,  $T_{AB}K_C L_K$ ,  $K_{ABC}L_T$ , and  $T_{ABC}L_K$  that are the counterparts of  $T_{AC}K_B$ ,  $K_{AC}T_B$ ,  $K_{AB}T_C$ ,  $T_{AB}K_C$ , wild-type PurK, and wild-type PurT, respectively, but with swapped B-loops. This short amino acid stretch is smaller than can be defined as a protein module, but is within the scope of changes accessible by mutational drift.

Of the six loop-swapped proteins tested with both PurK(−) and PurT(−) auxotrophic strains, one construct does enable cell survival. The domain- and loop-swapped hybrid protein  $K_{AC}T_B L_K$  can complement the PurK(−) auxotrophic strain, indicating that this hybrid can now catalyze the full reaction. Under selective conditions, cells expressing the  $K_{AC}T_B L_K$  hybrid grow at a rate comparable to cells expressing wild-type PurK. This contrasts with the  $T_{AC}K_B L_T$  hybrid, which does not confer survival to the PurT(−) auxotrophic strain. To investigate the properties of the active hybrid enzyme in more detail, wild-type PurK and  $K_{AC}T_B L_K$  were purified to homogeneity as C-terminal His-tagged constructs. Steady-state kinetic parameters of these two purified proteins were determined by following the production of ADP (Table IV). Compared to wild-type PurK,  $K_{AC}T_B L_K$  has a sevenfold decrease in the  $k_{cat}$ , a sevenfold decrease in the  $K_M$  for ATP, and a 6.4-fold decrease in  $K_M$  for AIR. These changes are offsetting, so that the specificity constants ( $k_{cat}/K_M$ ) for both ATP and AIR match those of wild-type PurK, within error. This striking result suggests that the only element in the B-domain that is unique for PurK catalysis is the swapped four amino acid sequence found in the B-loop. It is also notable that this small change allows the hybrid protein to regain the dependence of its ATPase activity on the presence of AIR, as seen with wild-type PurK.

During preparation of this manuscript, a publication by Thoden et al. reported the structures of *E. coli* PurK in complex with MgATP and with MgATP/Pi.<sup>35</sup> In these two structures, the B-loop is observed to inter-

act with the  $\gamma$ -phosphoryl group of ATP, or with Pi, respectively. Bound Pi is proposed to mimic the carboxy-phosphate intermediate, and is within hydrogen bonding distance of the backbone amide of Asp127, a residue located within the B-loop. Additionally, the side chain of Asp127, in conjunction with Lys314 (found in the C-domain), is proposed to activate the exocyclic amine of AIR for its attack. This recent work provides a possible mechanistic basis for understanding why swapping the B-loop is essential to recover PurK activity in the hybrid protein reported herein.

It is not entirely surprising that from all of the possible successful outcomes, swapping the B-domain and recovery of PurK activity was observed. In contrast to formylation of GAR, carboxylation of AIR can occur readily through a nonenzymic process, and growth of PurK auxotrophic strains can be rescued by growth in conditions that enrich CO<sub>2</sub> concentrations.<sup>25,36</sup> The B-domains of ATP grasp proteins are flexible domains that close down over the ATP substrate, and this flexibility could minimize unfavorable effects due to nonoptimized interactions between domains that have not coevolved. Also, carboxylation of AIR by either carboxyl phosphate or CO<sub>2</sub> likely represents an easier task than formylation of GAR. Even if carboxyl phosphate is not stabilized in the hybrid, the resulting CO<sub>2</sub> breakdown product could serve as an electrophile. Therefore, swapping the B-domains and recovery of PurK activity may represent the “lowest hurdle” to overcome in this model system.

There have been many other successful examples of engineering novel protein catalysts by modular domain swapping, most notably those including various enzymes fused to zinc-finger domains, or polyketide synthases and nonribosomal peptide synthases with deleted, swapped, or added modules (reviewed in Refs. 36 and 37). Unlike these examples where the binding and catalytic domains are somewhat distinct, domain swapping between PurT and PurK requires compatibility between all three modules to reconstitute functional binding and active sites. Therefore, achieving one active domain- and loop-swapped hybrid protein from only 14 rationally designed constructs is notable and is consistent with the proposal of domain swapping as a viable mechanism for protein evolution.

## Materials and Methods

All restriction enzymes used were purchased from New England Biolabs (Beverly, MA) unless otherwise indicated. DNA samples were purified using the QIAprep, as well as the QIAquick Gel and PCR purification

kit (all from Qiagen, Valencia, CA), following the manufacturer's protocols.

### **Bacterial strains**

The *E. coli* strains, DH-5 $\alpha$  and BL21(DE3), were obtained from Invitrogen (Carlsbad, CA) and Novagen (Madison, WI), respectively. TX680F' [ara  $\Delta$  (gpt-rpo-lac) thi rbs-221 ilvB2102 ilvH1202 *purN*'-lacZ<sup>+</sup>Y<sup>+</sup>::Kan<sup>R</sup> *purT*] *E. coli*, a PurT auxotrophic strain, is a gift from Dr. J.M. Smith.<sup>39</sup> CSH26, *purK* [*purK*, *zba*::*tn10*] *E. coli*, a PurK auxotrophic strain, was a gift from Dr. Gert Dandanell.<sup>32</sup> The PurT(-) strain is *E. coli* K-12 MG1655 with chromosomal *purN* and *purT* deleted, and the PurK(-) strain is *E. coli* K-12 MG1655 with chromosomal *purK* deleted. Both PurT(-) and PurK(-) strains were constructed by a one-step knockout method like that described elsewhere.<sup>33</sup>

### **Construction of hybrid enzymes**

All hybrid enzymes were constructed using the PCR overlapping extension method<sup>40</sup> (see Fig. 4). For *in vivo* complementation assays, the wild-type genes encoding PurT and PurK genes and the hybrid gene fragments were cloned into pDIM-PGX vector using *NdeI*/*SpeI* sites.<sup>41</sup> For protein expression and purification, the gene fragments encoding six chimeras (T<sub>AK</sub>BC, K<sub>AT</sub>BC, T<sub>AC</sub>K<sub>B</sub>, K<sub>AC</sub>T<sub>B</sub>, T<sub>AB</sub>K<sub>C</sub>, and K<sub>AB</sub>T<sub>C</sub>), as well as the sequences encoding PurT and PurK, were amplified from pDIM vectors using the PCR, and subcloned into pMAL-c2x vectors (New England Biolabs, Beverly MA) between *XmnI* and *SpeI* sites. The gene fragments encoding wild-type PurK and K<sub>AC</sub>T<sub>B</sub>L<sub>K</sub> were also subcloned into a pET22b expression vector (Navogen, Madison, WI) between *NdeI* and *HindIII* sites. All of the final products were confirmed by DNA sequencing of the inserts.

### **Screening of PurT and PurK activities by *in vivo* complementation**

pDIM vectors encoding each of the hybrid and wild-type enzymes were transferred into auxotrophic *E. coli* strains. After each transformation and an initial ampicillin (100 mg/mL) selection on LB agar, three single colonies were randomly picked and streaked in an "X" pattern on LB ampicillin (100 mg/mL) agar to make a master printing plate. After incubation at 37°C overnight, the cells on master plates were replica printed onto selective plates (M9 salt, 0.2% glucose, 0.06% caseine, 2  $\mu$ g/mL thiamine, 1.5% agar, 100  $\mu$ g/mL ampicillin) with 0.3 mM isopropyl  $\beta$ -D-thiogalactoside (IPTG). Functional complementation tests were conducted for up to 48 h at 30°C in TX680F' and CSH26 strains and at 25°C in PurT(-) and PurK(-) strains. Colonies that grew on selective plates were restreaked on new selective plates to obtain single colonies, from which the plasmids were extracted, and retransformed into the respective auxotrophic strain, followed by the

same selection procedure to affirm complementation. Plasmids were extracted from the first and second selections and the gene inserts were sequenced to confirm that there were no additional mutations.

### **Overexpression of hybrid enzymes and protein purification**

N-terminal maltose-binding-protein (MBP) fusion proteins of PurT, PurK, and the six hybrid proteins were expressed using the pMAL-c2x derivative described earlier. To eliminate background contamination by native enzymes, the hybrid proteins K<sub>AT</sub>BC, T<sub>AC</sub>K<sub>B</sub>, K<sub>AB</sub>T<sub>C</sub>, and wild-type PurT were expressed in the PurT(-) auxotrophic strain, while T<sub>AK</sub>BC, K<sub>AC</sub>T<sub>B</sub>, T<sub>AB</sub>K<sub>C</sub>, and wild-type PurK were expressed in the PurK(-) auxotrophic strain. The pET22b expression vectors containing sequences encoding wild-type PurK and K<sub>AC</sub>T<sub>B</sub>L<sub>K</sub> were transformed into BL21 (DE3) *E. coli* for overexpression. In each case, a single colony was used to start an overnight culture, 1% of which was inoculated into fresh LB media (100  $\mu$ g/mL ampicillin) and grown at 37°C to an OD<sub>600</sub> of ~0.5 before addition of 0.3 mM IPTG and cooling of the incubations to 18°C. After 8 h of induction at this lower temperature, cells were harvested by centrifugation and stored at -70°C.

Utilizing the specific binding of the N-terminal MBP domain, primary purification of the proteins overexpressed from the pMAL-c2x vector was performed on amylose resin, following the manufacturer's instructions. Subsequent to elution from the amylose resin using Column Buffer (20 mM Tris-HCl, pH 7.4, 200 mM NaCl, and 1 mM EDTA) with 10 mM maltose, the protein solution was loaded onto a Superose 12 gel filtration column (24 mL bed volume) pre-equilibrated in Buffer X (20 mM Tris-HCl, pH 7.4, 100 mM NaCl), and eluted using five bed volumes of Buffer X. Fractions containing hybrid proteins were identified by apparent size using SDS-PAGE, and pooled accordingly.

Proteins expressed from derivatives of the pET22b expression vector were purified using Ni-NTA agarose (Qiagen, Valencia CA), following the manufacturer's instructions. After removal of excessive imidazole by dialysis, the protein solutions were loaded onto a 5 mL HiTrap<sup>TM</sup> DEAE column (Amersham-Pharmacia Biotech, Piscataway, NJ) pre-equilibrated in Buffer X. During application of a linear salt gradient (0-1M NaCl), the desired proteins eluted at 0.3-0.4M NaCl. Fractions containing desired proteins were identified by apparent size using SDS-PAGE, and pooled accordingly.

Purified proteins (>95% homogeneous by SDS-PAGE) were concentrated using Centricon spin filters (MWCO 10 kDa, Amicon, Bedford, MA) and stored frozen at -70°C. Protein concentrations were determined by the Bradford analysis using bovine serum albumin as a standard.

### Measurement of ATP hydrolysis

Because the effect of  $\text{HCO}_3^-$  on ATP cleavage rates of hybrid proteins is investigated, excess  $\text{HCO}_3^-$  was removed from buffers as described elsewhere.<sup>34</sup> The ATPase activity of each hybrid protein was measured by monitoring the production of ADP via the coupling reactions of pyruvate kinase and lactate dehydrogenase (PK/LDH), as previously described with some modifications.<sup>24</sup> In a total volume of 100  $\mu\text{L}$ , 100 mM HEPES, pH 7.4, 20 mM KCl, 8 mM  $\text{MgCl}_2$ , 2 mM PEP, 0.2 mM NADH, and 10 units of PK/LDH were combined and allowed to incubate at 25°C. Reactions were initiated by addition of purified enzymes, and the absorbance decrease at 340 nm due to the consumption of NADH ( $\epsilon = 6.22 \text{ mM}^{-1} \text{ cm}^{-1}$ ) was monitored on a Cary I UV-Vis spectrophotometer (Varian, Palo Alto CA). For wild-type PurT, PurK and the hybrid  $\text{K}_{\text{AB}}\text{T}_{\text{C}}$  proteins, the kinetic parameters of ATP cleavage were determined in triplicate over a concentration range of 10–200  $\mu\text{M}$  AIR, 5–100  $\mu\text{M}$  GAR, 0.1–10 mM ATP, and 0.1–2 mM formate. By varying the concentration of one substrate at saturating concentrations of the other substrates, the Michaelis constants for each substrate were determined by fitting to a double reciprocal plot using the program KaleidaGraph (Synergy Software). For all other hybrid fusion proteins, the ATPase specific activity ( $\mu\text{mol min}^{-1} \text{ mg}^{-1}$ ) was determined using the different substrates at their respective saturating concentrations (ATP, 5 mM; GAR, 100  $\mu\text{M}$ ; AIR, 260  $\mu\text{M}$ ; formate, 3 mM;  $\text{HCO}_3^-$ , 3 mM).

### Detection of acyl phosphate

Free acyl phosphate was assayed by trapping as the hydroxamate derivative, using a previously described method.<sup>42</sup> Briefly, in a total volume of 300  $\mu\text{L}$ , 100 mM HEPES, pH 7.4, 20 mM KCl, 8 mM  $\text{MgCl}_2$ , and saturating concentrations of substrates were combined and incubated at 25°C for 5 min. Reactions were initiated upon addition of purified enzymes. At various time points, aliquots of the incubating reaction were taken and mixed with an equal volume of a freshly made 1:1 (v/v) mixture of 3M hydroxylamine and 3M NaOH. Following a 10-min incubation with the alkaline hydroxylamine solution, 1 volume of 0.74M trichloroacetate (TCA) and 2 volumes of 0.22 mM  $\text{FeCl}_3$  in 0.5M HCl were added. The resulting derivatives are yellow in color and their absorbance at 490 nm is used for quantification in comparison to a succinohydroxamate standard curve.

### Kinetic assay for the PurK activity

The PurK activity of wild-type PurK and the hybrid  $\text{K}_{\text{AC}}\text{T}_{\text{B}}\text{L}_{\text{K}}$  protein, each bearing a C-terminal His tag, were measured by monitoring the production of ADP via the coupling reactions of pyruvate kinase and lactate dehydrogenase (PK/LDH), as described elsewhere with some modifications.<sup>25</sup> Briefly, in a total volume

of 100  $\mu\text{L}$ , 100 mM HEPES, pH 8.0, 20 mM KCl, 8 mM  $\text{MgCl}_2$ , 2 mM PEP, 0.2 mM NADH, and 10 units of PK/LDH were combined, and incubated at 25°C for 5 min. Reactions were initiated by the addition of purified enzymes, and the absorbance decrease at 340 nm due to the consumption of NADH ( $\epsilon = 6.22 \text{ mM}^{-1} \text{ cm}^{-1}$ ) was monitored on a Cary I UV-Vis spectrophotometer (Varian). Kinetic parameters were determined in triplicate using a concentration range of 5–40  $\mu\text{M}$  AIR, 5–100  $\mu\text{M}$  ATP. By varying the concentration of one substrate at the saturating concentrations of the other substrates, the Michaelis constants for each substrate were determined by fitting to a double reciprocal plot using the program KaleidaGraph (Synergy Software).

### Conclusions

In this study, we have investigated the modular nature of protein structure and function using PurT and PurK as a model system. Features of these results help to illuminate aspects of protein domain swapping that may occur during enzyme evolution. In this system, swapping one domain between two homologous three-domain proteins consistently resulted in less stable hybrid proteins that were unable to catalyze a full reaction. Presumably, the interface between the swapped domain and the rest of the hybrid is not well optimized. Despite these imperfections, each of the domain-swapped hybrids is able to catalyze partial reactions, suggesting that functional binding and catalytic sites are partially maintained. Additional minor modifications of the domain-swapped hybrids, here the substitution of a short loop region suspected to sequester reaction intermediates, are required to enhance coupling of the partial reactions to reconstitute a full reaction sufficient to impact cell survival. This study illustrates how a mechanistic understanding of intermediary hybrid proteins can improve the design and construction of a fully functional enzyme.

### References

1. Aimi J, Qiu H, Williams J, Zalkin H, Dixon JE (1990) De novo purine nucleotide biosynthesis: cloning of human and avian cDNAs encoding the trifunctional glycylamide ribonucleotide synthetase-aminoimidazole ribonucleotide synthetase-glycylamide ribonucleotide transformylase by functional complementation in *E. coli*. *Nucleic Acids Res* 18:6665–6672.
2. Kan JL, Jannatipour M, Taylor SM, Moran RG (1993) Mouse cDNAs encoding a trifunctional protein of de novo purine synthesis and a related single-domain glycylamide ribonucleotide synthetase. *Gene* 137:195–202.
3. Doolittle RF, Bork P (1993) Evolutionarily mobile modules in proteins. *Sci Am* 269:50–56.
4. Scott KA, Steward A, Fowler SB, Clarke J (2002) Titin: a multidomain protein that behaves as the sum of its parts. *J Mol Biol* 315:819–829.
5. Mainfroid V, Goraj K, Rentier-Delrue F, Houbrechts A, Loiseau A, Gohimont AC, Noble ME, Borchert TV, Wierenga RK, Martial JA (1993) Replacing the (beta alpha)-

- unit 8 of *E. coli* TIM with its chicken homologue leads to a stable and active hybrid enzyme. *Protein Eng* 6: 893–900.
6. Lang D, Thoma R, Henn-Sax M, Sterner R, Wilmanns M (2000) Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion. *Science* 289:1546–1550.
  7. Hocker B, Beismann-Driemeyer S, Hettwer S, Lustig A, Sterner R (2001) Dissection of a (betaalpha)8-barrel enzyme into two folded halves. *Nat Struct Biol* 8:32–36.
  8. Hocker B, Claren J, Sterner R (2004) Mimicking enzyme evolution by generating new (betaalpha)8-barrels from (betaalpha)4-half-barrels. *Proc Natl Acad Sci USA* 101: 16448–16453.
  9. Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA (2004) Structure, function and evolution of multidomain proteins. *Curr Opin Struct Biol* 14:208–216.
  10. Ostermeier M, Benkovic SJ (2000) Evolution of protein function by domain swapping. *Adv Protein Chem* 55: 29–77.
  11. Galperin MY, Koonin EV (1997) A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity. *Protein Sci* 6:2639–2643.
  12. Waldrop GL, Rayment I, Holden HM (1994) Three-dimensional structure of the biotin carboxylase subunit of acetyl-CoA carboxylase. *Biochemistry* 33:10249–10256.
  13. Thoden JB, Holden HM, Wesenberg G, Raushel FM, Rayment I (1997) Structure of carbamoyl phosphate synthetase: a journey of 96 Å from substrate to product. *Biochemistry* 36:6305–6316.
  14. Yamaguchi H, Kato H, Hata Y, Nishioka T, Kimura A, Oda J, Katsube Y (1993) Three-dimensional structure of the glutathione synthetase from *Escherichia coli* B at 2.0 Å resolution. *J Mol Biol* 229:1083–1100.
  15. Wolodko WT, Fraser ME, James MN, Bridger WA (1994) The crystal structure of succinyl-CoA synthetase from *Escherichia coli* at 2.5-Å resolution. *J Biol Chem* 269: 10883–10890.
  16. Fan C, Moews PC, Walsh CT, Knox JR (1994) Vancomycin resistance: structure of D-alanine:D-alanine ligase at 2.3 Å resolution. *Science* 266:439–443.
  17. Wang W, Kappock TJ, Stubbe J, Ealick SE (1998) X-ray crystal structure of glycinamide ribonucleotide synthetase from *Escherichia coli*. *Biochemistry* 37:15647–15662.
  18. Shindyalov IN, Bourne PE (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng* 11:739–747.
  19. Thoden JB, Kappock TJ, Stubbe J, Holden HM (1999) Three-dimensional structure of N5-carboxyaminoimidazole ribonucleotide synthetase: a member of the ATP grasp protein superfamily. *Biochemistry* 38:15480–15492.
  20. Thoden JB, Firestine SM, Nixon A, Benkovic SJ, Holden HM (2000) Molecular structure of *Escherichia coli* PurT-encoded glycinamide ribonucleotide transformylase. *Biochemistry* 39:8791–8802.
  21. Thoden JB, Firestine SM, Benkovic SJ, Holden HM (2002) PurT-encoded glycinamide ribonucleotide transformylase: accommodation of adenosine nucleotide analogs within the active site. *J Biol Chem* 277:23898–23908.
  22. Marolewski AE, Mattia KM, Warren MS, Benkovic SJ (1997) Formyl phosphate: a proposed intermediate in the reaction catalyzed by *Escherichia coli* PurT GAR transformylase. *Biochemistry* 36:6709–6716.
  23. Marolewski A, Smith JM, Benkovic SJ (1994) Cloning and characterization of a new purine biosynthetic enzyme: a non-folate glycinamide ribonucleotide transformylase from *E. coli*. *Biochemistry* 33:2531–2537.
  24. Mueller EJ, Meyer E, Rudolph J, Davisson VJ, Stubbe J (1994) N5-carboxyaminoimidazole ribonucleotide: evidence for a new intermediate and two new enzymatic activities in the de novo purine biosynthetic pathway of *Escherichia coli*. *Biochemistry* 33:2269–2278.
  25. Sauers CK, Jencks WP, Groh S (1975) Alcohol-bicarbonate-water system. Structure-reactivity studies on the equilibria for formation of alkyl monocarbonates and on the rates of their decomposition in aqueous alkali. *J Am Chem Soc* 97:5546–5553.
  26. Nixon AE, Warren MS, Benkovic SJ (1997) Assembly of an active enzyme by the linkage of two protein modules. *Proc Natl Acad Sci USA* 94:1069–1073.
  27. Inglese J, Johnson DL, Shiau A, Smith JM, Benkovic SJ (1990) Subcloning, characterization, and affinity labeling of *Escherichia coli* glycinamide ribonucleotide transformylase. *Biochemistry* 29:1436–1443.
  28. Klein C, Chen P, Arevalo JH, Stura EA, Marolewski A, Warren MS, Benkovic SJ, Wilson IA (1995) Towards structure-based drug design: crystal structure of a multi-substrate adduct complex of glycinamide ribonucleotide transformylase at 1.96 Å resolution. *J Mol Biol* 249: 153–175.
  29. Nixon AE, Benkovic SJ (2000) Improvement in the efficiency of formyl transfer of a GAR transformylase hybrid enzyme. *Protein Eng* 13:323–327.
  30. Ostermeier M, Shim JH, Benkovic SJ (1999) A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat Biotechnol* 17:1205–1209.
  31. Sorensen IS, Dandanell G (1997) Identification and sequence analysis of *Sulfolobus solfataricus* purE and purK genes. *FEMS Microbiol Lett* 154:173–180.
  32. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* 97: 6640–6645.
  33. Meyer E, Leonard NJ, Bhat B, Stubbe J, Smith JM (1992) Purification and characterization of the purE, purK, and purC gene products: identification of a previously unrecognized energy requirement in the purine biosynthetic pathway. *Biochemistry* 31:5022–5032.
  34. Thoden JB, Holden HM, Firestine SM (2008) Structural analysis of the active site geometry of N5-carboxyaminoimidazole ribonucleotide synthetase from *Escherichia coli*. *Biochemistry* 47:13346–13353.
  35. Gots JS, Benson CE, Jochimson B, Koduri KR. Microbial models and regulatory elements in the control of purine metabolism. In: Elliott K, Fitzsimons DW, Eds. (1976) Purine and pyrimidine metabolism. Amsterdam: Elsevier, pp 23–42.
  36. Lutz S, Benkovic SJ (2000) Homology-independent protein engineering. *Curr Opin Biotechnol* 11:319–324.
  37. Mootz HD, Marahiel MR (1999) Design and application of multimodular peptide synthetases. *Curr Opin Biotechnol* 10:341–348.
  38. Nygaard P, Smith JM (1993) Evidence for a novel glycinamide ribonucleotide transformylase in *Escherichia coli*. *J Bacteriol* 175:3591–3597.
  39. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77:51–59.
  40. Lutz S, Ostermeier M, Benkovic SJ (2001) Rapid generation of incremental truncation libraries for protein engineering using alpha-phosphothioate nucleotides. *Nucleic Acids Res* 29(4):E16.
  41. Pechere JF, Capony JP (1967) On the colorimetric determination of acyl phosphate. *Anal Biochem* 22:536–539.
  42. Kappock TJ, Ealick SE, Stubbe J (2000) Modular evolution of the purine biosynthetic pathway. *Curr Opin Chem Biol* 4:567–572.