
On the structural and functional modularity of glycinamide ribonucleotide formyltransferases

SEUNG-GOO LEE,^{1,2} STEFAN LUTZ,^{1,3} AND STEPHEN J. BENKOVIC

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

(RECEIVED May 9, 2003; FINAL REVISION July 8, 2003; ACCEPTED July 9, 2003)

Abstract

Glycinamide ribonucleotide formyltransferases (GARTs) are part of the de novo purine biosynthetic pathway, catalyzing the direct transfer of a formyl group from the tetrahydrofolate cofactor to the glycinamide ribonucleotide substrate. Despite the low amino acid-sequence identity between the GARTs from *Escherichia coli* and human, their tertiary structures are superimposable. As part of our functional studies of these enzymes, we have investigated the interchangeability of individual protein fragments or modules between the two enzymes and the functional properties of the resulting hybrids. The modular nature of GART facilitated the creation of combinatorial libraries of chimeras between the *Escherichia coli* and human enzymes, which were functionally selected through complementation of an auxotrophic *Escherichia coli* strain. From a pool of several dozen sequence distinct hybrids, six in vivo-functional fusion genes were selected, overexpressed, and purified to homogeneity. The kinetic analysis of these constructs and the comparison of their k_{cat} and K_{M} values to the parental enzymes suggest that the characteristic kinetic properties from the two parents are “modular encoded” and can be exchanged by domain swapping. The chimeras in general, however, are subject to temperature instability and misfolding; thus, they serve primarily as useful candidates for further rounds of optimization.

Keywords: Protein engineering; ITCHY; SCRATCHY; domain swapping; modularity

Modularity is an attractive concept with implications for how proteins may have evolved in nature. This hypothesis describes the build-up of complex structures from small, independent peptidic subunits—a “Lego-like” assembly of biocatalysts from a naturally existing repertoire of functional protein domains. The identification and characterization of these functional building blocks is relevant to the elucidation of questions regarding protein evolution, folding, and, last but not least, the successful engineering of proteins for tailored applications.

Examples of modular design can be identified at all levels of protein structure. Polyketide synthases and nonribosomal

peptide synthetases represent a class of covalently linked clusters consisting of multiple enzymes with individual functions. This class catalyzes consecutive steps in the synthesis of complex secondary metabolites (Cane et al. 1998; Mootz et al. 2000). The human GART, one of the parental proteins in this study, is actually part of a trifunctional enzyme complex with GAR synthetase (PurD) and aminoimidazole-ribonucleotide synthetase (PurM; Daubner et al. 1985), although for this example the catalyzed steps are not consecutive. The formation of such assemblies is believed to facilitate substrate channeling, thereby expediting substrate turnover and protecting reactive intermediates. Beyond biosynthetic chemistry, protein modularity can also contribute to the extraordinary properties of biomaterials. Macromolecules such as elastin, collagen, or titin rely on repetitive peptidic subunits to gain unique physical properties (Hinman et al. 2000; Scott et al. 2002).

Modular design is not limited to large, multifunctional protein complexes but can also be found within individual proteins. A key example is the class of β/α -barrel proteins that comprise eight repetitive strand-turn-helix motifs. The functional tolerance of this structural framework with re-

Reprint requests to: Stephen J. Benkovic, The Pennsylvania State University, Department of Chemistry, University Park, PA 16802, USA; e-mail: sjb1@psu.edu fax: (814) 865-2973.

¹These authors contributed equally to this work.

Current addresses: ²Microbial Function Lab., Korea Research Institute of Bioscience and Biotechnology (KRIBB), 52, Oun-Dong, Yusong-Gu, Taejon 305-333, South Korea; ³Department of Chemistry, Emory University, Atlanta, GA 30322, USA.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.03139603>.

spect to engineering is well documented. The fragmentation into half-barrels, as well as the substitution of individual motifs within the barrel framework does not disrupt the protein's ability to self-assemble into a functionally competent enzyme (Luger et al. 1989; Eder and Kirschner 1992; Mainfroid et al. 1993; Hocker et al. 2001). While a similar breakdown into small peptide fragments may not be as obvious for other proteins, such patterns are believed to exist for most protein scaffolds.

A variety of approaches can be followed to examine existing proteins for a functionally viable fragmentation pattern. A possible lead in the search for subunits can be the analysis of eukaryotic genomes and mapping the exon structures onto the existing structural information (Go 1981; Dorit et al. 1990; Seidel et al. 1992). In the context of our study, the natural exon boundaries of the human GART are located between amino acids 10:11, 54:55, 101:102, and 140:141. Alternatively, structural data and homology modeling in combination with databases such as CATH (Orengo et al. 1997) can guide the assignment of individual folds.

Assistance for our function-based study of engineered formyltransferases derives from the available structural data for both parental enzymes. The early crystallographic studies of the *Escherichia coli* GART (Almassy et al. 1992; Chen et al. 1992; Klein et al. 1995) have revealed a clearly defined substrate-binding cleft at the interface between the N-terminal (amino acid 1–101) and the C-terminal fragment (amino acid 102–212). The former consists of a sheet of four β -strands (β 1– β 4 in Fig. 1), sandwiched between four helices (α 1– α 4) comprising a Rossmann fold (Rossmann et al. 1974) that is involved in binding of the phosphate portion of the GAR substrate. The C-terminal portion includes a continuation of the β -sheet by adding three antiparallel strands (β 5– β 7) and is shielded on one side by the extended helix α 6. Two of the active site residues, N106 and H108, are part of β 5 that is centered within the active site cleft,

thus placing the two residues in close proximity to the substrates (Almassy et al. 1992). The third catalytically important residue, D144, is located in the loop region connecting β 6 and β 7 that caps the folate binding site. Additional residues involved in folate binding include a hydrophobic pocket composed of residues L85, F88, L92, V97, L104, and V139, as well as hydrogen-bonding network involving the amino acids R90, L92, T140, and R141 (Klein et al. 1995). The GAR binding site consists of multiple amino acid side chains derived from both the N- and C-terminal domains. The orientation of GAR in the active site is achieved through an extensive hydrogen-bonding network involving residues S12, N13, E173, and H174 (Fig. 2). Additional interactions are obtained through hydrogen bonding of the ribose to E170 as well as multiple hydrophobic interactions involving the side chains of amino acids A86, F88, I107, H121, T166, and T171 (Klein et al. 1995).

Most recently, the crystallographic analysis of human GART confirmed the predicted high degree of structural similarity with the *E. coli* enzyme (Zhang et al. 2002). Despite a low amino acid sequence identity (31%), the two proteins show a striking structural overlap with an average rms deviation of 1.5 Å (Zhang et al. 2002). These results also confirm previous predictions based on homology modeling data and the experimental correlation between the observed changes of functional properties of the human protein and PurN upon mutagenesis (Inglese et al. 1990; Kan et al. 1992).

In addition to computational and gene mapping analysis, several experimental techniques for the exploration of fragmentation and recombination sites have been introduced in recent years. Combinatorial, structure-based protein engineering methods include incremental truncation for the creation of hybrid enzymes (ITCHY; Ostermeier et al. 1999), SCRATCHY (Lutz et al. 2001b), sequence homology-independent protein recombination (SHIPREC; Sieber et al.

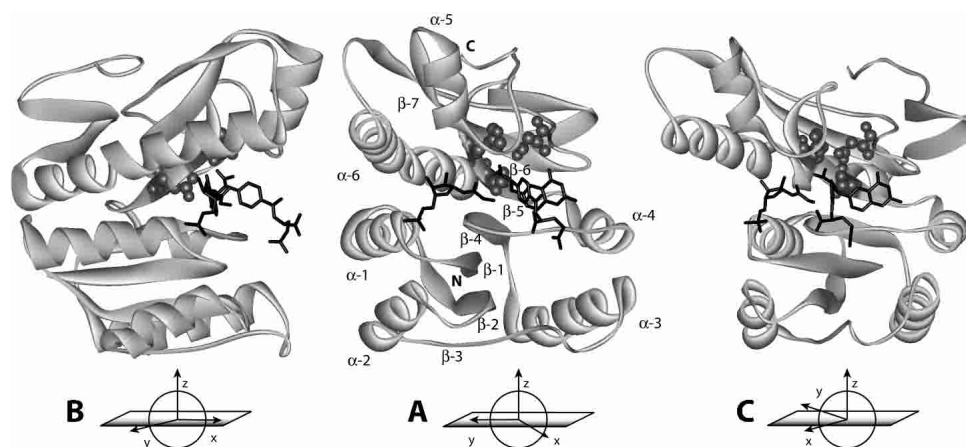


Figure 1. Structural overview and secondary structure assignment in the PurN framework. Shown in black sticks are the enzyme-bound substrate and cofactor while the active site residues N106, H108, and D144 are highlighted as gray spheres.

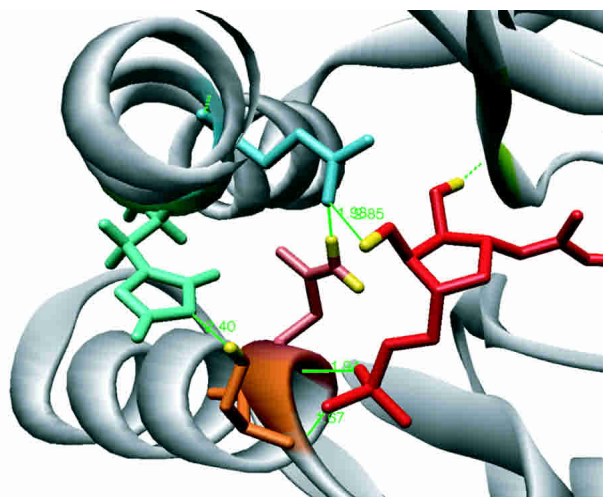


Figure 2. Hydrogen-bonding network in GAR binding site. The elaborate network of hydrogen bonds in the GAR binding site contributes substantially to catalysis through substrate preorientation. The system involves the side chains and peptide backbone of amino acid residue S12 and N13 in helix-1, E173 and H174 of helix-6, as well as the phosphate and ribose portion of GAR. The distance in the primary sequence between the involved protein residues makes the noncovalent complex a sensitive determinant for the structural integrity of the protein.

2001), and structure-based combinatorial protein engineering (SCOPE; O'Maille et al. 2002). These techniques provide the means to explore the modular nature of proteins by allowing the swapping of secondary structure elements and domains between protein homologs independent of their sequence identity. As reported previously, our laboratory has successfully implemented the ITCHY and SCRATCHY techniques on the glycinamide ribonucleotide formyltransferases (GARTs) from *E. coli* (PurN) and human (hGART; Ostermeier et al. 1999; Lutz et al. 2001b). From the large selection of functional hybrid constructs, we have selected six chimeras that were overexpressed and the protein products isolated for detailed kinetic analysis. Their characterization provides insight into the functional impartiality of individual protein fragments within the context of an existing protein framework.

Results and Discussion

In vivo activity of the hybrid GAR formyltransferases

The initial selection of functional hybrid enzymes with GART activity from the incremental truncation and SCRATCHY libraries was performed by *in vivo* complementation of the auxotrophic *E. coli* host strain TX680 (Ostermeier et al. 1999; Lutz et al. 2001b). As mentioned previously, the *in vivo* complementation of these libraries is strongly affected by the incubation temperature. Selection plates grown at ambient temperature yielded several-fold higher colony numbers as compared with incubations at

37°C. Sequence analysis also indicated an extended diversity of functional hybrid enzymes at lower temperatures where constructs with compromised structural integrity and slower folding pattern are better accommodated.

From the pool of confirmed (functional) and sequenced hybrid enzymes, we selected six single-crossover constructs and one double-crossover construct (Fig. 3) for overexpression and detailed kinetic analysis *in vitro*. The choice of these constructs was based on the nature of their crossovers. The GPX-B12 (N-terminal *purN* and C-terminal *hGART*) and GPX-M1 (N-terminal *hGART* and C-terminal *purN*) hybrids are “mirror-images,” having the crossover at position 100–102. Constructs with fusion points in that region were among the most dominant library members found in the *in vivo* assays at elevated temperatures.

The third hybrid, GPX-M36, has its fusion point at position 89/90, closely resembling GPX-M1 but carrying helix-4 of *PurN* (Fig. 1). Helix 4 is part of the cofactor-binding pocket and the comparison of the kinetic data of M1 and M36 could provide insight into the binding geometry of the fDDF and the contribution of the network of noncovalent interactions to the overall stability of the protein.

The two GPX fusions M24 and M55 both possess crossovers in the connecting loop between helices 2 and 3 (Fig. 1). Whereas GPX-M24 is an “exact” fusion at position 56/57 (the crossover between the *purN* and *hGART* sequence falls precisely at the aligned position), GPX-M55 has amino acid position 57 deleted, resulting in a shortened loop. Comparison of the kinetic properties of the two proteins could help assess the importance of this loop's flexibility in catalysis and the protein's overall structural integrity. In addition, the experimental data could provide insight on the structural independence of the N-terminal Rossmann fold in the context of a larger protein structure.

Our study also includes a hybrid structure with two crossovers, generated by SCRATCHY. *In vivo*, the GPG-N11 construct (N-terminal *hGART*, central portion from *purN*, C-terminal *hGART*) shows promising catalytic activity in the functional selection experiments by complementing the auxotrophic *E. coli* under uninduced conditions at room temperature. The location of the crossovers at position 56/57 and 131/132 creates a hybrid enzyme with the fDDF-binding site and the active site residues (N106/H108) furnished by the *E. coli* protein and with the glycinamide-ribonucleotide binding site, as well as the majority of the surrounding protein framework derived from the human homolog.

Finally, the *in vivo* complementation assays identified the rather unusual yet functional candidate GPX-M12 that has a 75-amino-acid insertion, almost completely duplicating the truncation region. The hybrid consists of the N-terminal 129 amino acids from hGART as well as the C-terminal 159 amino acids from *E. coli*, replicating the central region from amino acid 54 to 129. As shown in Figure 3, two possible

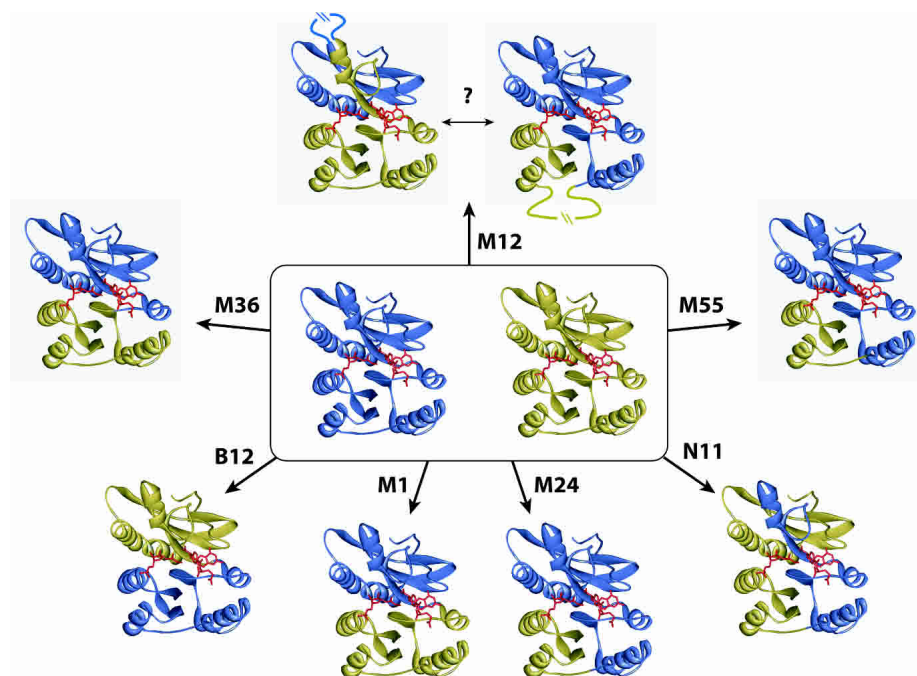


Figure 3. Schematic overview of the hybrid enzyme composition. The parental enzymes are shown in blue (PurN) and gold (hGART). The structure coordinates of PurN, cocrystallized with glycinamide ribonucleotide and 5'-deaza-tetrahydrofolate, by Almasy et al. (1992), served as template for the drawings. For orientation purposes, the bound substrate and cofactor are shown in red.

structural arrangements are conceivable where either the hGART or the PurN overlapping portion contribute to the active part of the enzyme while the remaining sequence loops away from the catalytically active framework. Considering the proposed mechanisms for protein evolution by gene duplication, GPX-M12 could resemble an initial step in the process of developing larger, more complex protein structures with extended function or regulatory domains. Note that the duplication region corresponds closely to the crossover points in GPG-N11.

Protein overexpression

The seven hybrids, as well as the parental wild-type genes were subcloned into the pET expression system and overexpressed as described in the Materials and Methods section. In preliminary experiments, optimal expression of active protein was achieved upon induction at 18°C. Tests at elevated temperatures generally resulted in the generation of inclusion bodies with no detectable enzymatic activities. Despite the lower temperature during the optimized induction, the SDS-PAGE analysis of the cell culture lysates revealed significant differences in expression level depending on the hybrid construct. Whereas wild-type PurN and PGX-B12 showed expression levels of >30% of total *E. coli* proteins, the GPX hybrids were expressed at levels below 5% of total protein and showed a strong tendency to aggregate and precipitate out of solution.

Despite their successful *in vivo* complementation, two of the hybrids, GPX-M36 and GPX-M12, failed to produce any soluble proteins upon overexpression. Both constructs are presumably compromised in their folding capabilities, generating exclusively protein inclusion bodies under the present conditions. Denaturation and refolding experiments on the inclusion bodies failed to produce any detectable enzyme activity for these two hybrid proteins. Upon closer investigation of GPX-M12, we hypothesize that the strong tendency towards aggregation is caused by the poorly defined structure of the large duplicated region that exposes large hydrophobic patches on the protein's surface.

In the case of GPX-M36, we presume that the observed misfolding is related to the helical region 4. More specifically, the performance of the various hybrid crossover combinations in this study point towards the disruption of favorable interactions between helices 3 and 4 as the cause for the observed instability of GPX-M36. Contacts between the β -strands 1, 4, and 5 seem interchangeable between the two parental structures as demonstrated in PGX-B12, GPX-M1, and GPX-M24. A critical effect of this region to markedly weaken folate-cofactor binding is unlikely, given the functional competence of GPX-M36 *in vivo*, as well as the tolerance of the binding site to fragment substitution as observed in PGX-B12 and GPX-M1. Furthermore, the highly conserved $\beta 4/\alpha 4$ loop is unaffected by the presence of the fusion point, owing to its sequence preservation in the two parents. The only remaining contact of helix- $\alpha 3$ is to

helix- α 4. In the absence of any noticeable sequence conservation in either structural element (as determined by multiple sequence alignment of 19 GART-family members—data not shown), we can only speculate on possible stabilizing interactions. From the crystal structure of hGART, a cluster of hydrogen bonds (D72/E76/K99/W100) were identified that would be disrupted in GPX-M36 and might result in loss of stability.

Purification and oligomeric states of wild-type and hybrid enzymes

The soluble, overexpressed wild-type and hybrid proteins were isolated and purified to homogeneity. The previously reported purification protocol (Shim and Benkovic 1998) was modified to include a C-terminal His-tag on the protein to facilitate a simple purification protocol via Ni-NTA affinity chromatography. Protein of >95% purity as indicated by SDS-PAGE was obtained in one step. The comparison with earlier kinetic data (Ostermeier et al. 1999) suggests that the catalytic performance of the enzymes (based on the wild-type proteins) is largely unaffected by the presence of the His-tag (Table 1).

In the subsequent purification step by gel filtration, the wild-type PurN eluted as a single peak corresponding to the monomeric protein species as predicted from previous work, which reports the prevalence of PurN's monomeric form above pH 6.8 (Mullen and Jennings 1998). In contrast, the wild-type human GART elutes earlier at a retention time corresponding to the dimeric form of the protein. These results differ from data recently published by Zhang et al. (2002) who exclusively observed the monomeric form of hGART over a wide range of pHs. Differences in the buffer systems used by the two laboratories might explain the observed experimental differences.

In the same gel filtration experiments, the hybrid proteins showed various levels of higher molecular mass contaminants. Given the material's purity in SDS-PAGE experiments, the observed side-products must originate from aggregation of partially folded hybrid proteins. Fractionation

followed by activity assays of the individual samples established that only the peaks corresponding to the monomeric and (if present) the dimeric form of the hybrids show measurable fDDF conversion to product. Furthermore, the accumulation of higher-mass products is affected by temperature. Whereas the protein remains active for several weeks under the described storage conditions at -70°C (see Materials and Methods), catalytic turnover rapidly diminishes upon prolonged storage above freezing.

Temperature sensitivity

The temperature optimum for maximum enzymatic activity of the hybrid enzymes directly correlates with the corresponding selection conditions. Hybrid enzymes isolated from selection plates that were incubated at room temperature (GPX-M24, M55, and GPG N11) showed maximal activity around 20°C (Fig. 4). The activity profile of the wild-type enzymes and GPX-M1, isolated at 37°C , reached their maximum at $30\text{--}35^{\circ}\text{C}$.

A noticeable exception to this correlation is PGX-B12. Originally selected from *in vivo* complementation experiments at 37°C , the isolated and purified protein proved extremely thermo-sensitive as evident by the rapid decline of fDDF conversion to product at temperatures above 4°C . The protein was stabilized to some extent in buffer containing either 0.1–1 g/L BSA or 20% glycerol. These results led to the hypothesis that the hybrid enzyme, although fundamentally functional based on its primary sequence, has suffered significant fold destabilization owing to the substitution of the protein fragments. Although steric crowding can compensate for fold instability *in vivo*, the isolated protein requires artificial crowding agents such as BSA or glycerol *in vitro* to maintain its functional status.

In vivo/in vitro differences of hybrid GART activity

The differences in enzymatic activity *in vivo* and *in vitro* presumably originate from the overall compromised structural integrity of the hybrids under *in vitro* conditions. As

Table 1. Summary of the kinetic data of the two wild-type enzymes (PurN/hGART) and the five soluble hybrid enzymes

	K_m (GAR) (μM)	K_m (fDDF) (μM)	k_{cat} /s	k_{cat}/K_m (GAR) ($\mu\text{M}^{-1}/\text{s}$) $\times 10^3$	k_{cat}/K_m (fDDF) ($\mu\text{M}^{-1}/\text{s}$) $\times 10^3$
PurN (wt)	57.7 ± 3.6	6.6 ± 0.3	66 ± 9	1140	10,000
PGX-B12	26.2 ± 1.4	102 ± 16	33 ± 4	1260	320
GPX-M24	262 ± 12	18 ± 0.6	21 ± 2.9	80	1166
GPX-M55	630 ± 136	31 ± 2	12 ± 3.3	19	390
GPG-N11	293 ± 48	43 ± 14	0.99 ± 0.18	3.4	23
GPX-M1	66.2 ± 9.7	1.9 ± 0.2	4.0 ± 0.6	60	2100
hGART (wt)	3.4 ± 0.7	0.59 ± 0.04	4.0 ± 0.6	1180	6780

PGX B12: PurN 1–101/hGART 102–202; GPX M1: hGART 1–100/PurN 101–212; GPX M24: hGART 1–54/PurN 55–212; GPX M55: hGART 1–56/PurN 58–212; GPG N11:hGART 1–56/PurN 57–131/hGART 132–202. Experiments were performed at 23°C with the exception of PGX-B12 (4°C).

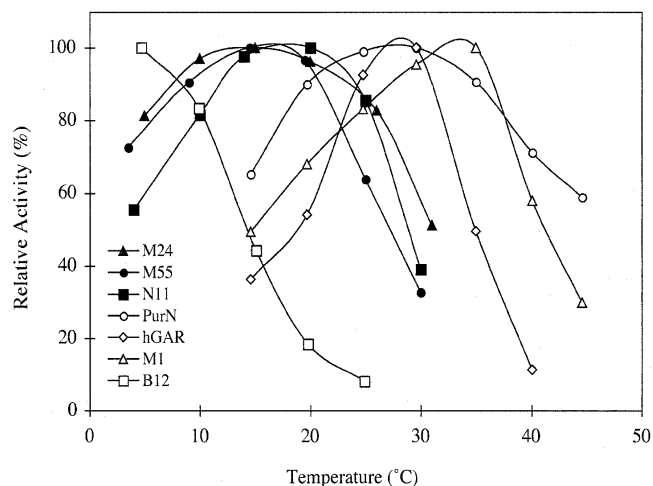


Figure 4. Temperature-activity plots for wild-type and hybrid enzymes. The selection conditions for the various enzymes are clearly reflected in the protein's functional performance.

discussed above, minor structural clashes (a consequence of the domain swapping process) are likely to disrupt portions of the extensive network of noncovalent interactions thereby weakening the protein structure as a whole. *In vivo*, the fold destabilization can be compensated at least partially by macromolecular crowding to provide sufficient enzymatic activity to complement the host strain. Taking advantage of potential stabilizing effects, we explored the functional complementation of our hybrid enzyme libraries under hyperosmotic conditions (Richey et al. 1987; Cayley et al. 1991). The results, however, neither show an increase in the percentage of functional hybrids, nor identify novel fusion constructs (S. Lutz, unpubl.). *In vitro* experiments are performed in a substantially more dilute environment compared to cellular conditions, which render the enzymes prone to denaturation and aggregation. We conclude that the discrepancy between *in vivo* and *in vitro* experiments is the result of the compromised overall stability of the hybrid enzymes.

Effects of domain swapping on catalysis

The concept of modularity as a factor in the creation of functional protein frameworks would favor the generation of more efficient, tailor-made biocatalysts through recombination of selected structure elements and domains from the best-performing parental sequences. We have, as part of this study on GART, initiated a detailed investigation into the catalytic properties of selected hybrid enzymes by steady-state kinetics. The K_M and k_{cat} values for all soluble hybrid and wild-type enzymes were determined and their corresponding specific rate constants (k_{cat}/K_M) calculated (Table 1).

Given the kinetic parameters for the wild-type enzymes, two distinct strategies can be pursued to achieve similar specific catalytic rates. Although the K_M 's for PurN are 10- to 20-fold higher than for hGART, the bacterial protein compensates with a 16-fold higher catalytic rate for the transformylation (Table 1). Substitution of the functionally more efficient protein fragments from either parent might lead to hybrid constructs with the tighter substrate binding of the human enzyme and/or the higher catalytic rates of the bacterial protein.

In the kinetic studies of the five soluble hybrid enzymes, the two mirror-image hybrids PGX-B12 and GPX-M1 show the highest catalytic activity (k_{cat}/K_M) of all the hybrids (Table 1). For GPX-M1, the K_M for GAR is within the experimental error of the wild-type PurN whereas its affinity for fDDF lies between the two parents. PGX-B12 on the other hand shows twofold higher affinity for GAR than the wild-type PurN but a 15- to 150-fold reduced affinity for fDDF relative to the parent enzymes. Note, however, that the major binding determinants for GAR lie in both N- and C-terminal domains whereas those for the folate cofactor are more focused in the C-terminal portion of the proteins. Consequently, if the modules acted highly independently, PGX-B12 should have exhibited a K_M for fDDF closer to that of hGART, and GPX-M1 a K_M for fDDF near to that of PurN. As this is clearly not the case, the kinetic properties of the hybrids must reflect interactions across the entire protein.

The comparison of the k_{cat}/K_M values for the two hybrid and wild-type enzymes shows a wild-type like behavior for PGX-B12 with respect to the GAR substrate whereas GPX-M1 possess parents-like fDDF specificity. Perhaps the k_{cat}/K_M specificity constant is a better index for the properties of the hybrid. We expect and find that PGX-B12 exhibits a k_{cat}/K_M for GAR closer to that for PurN, and GPX-M1 a k_{cat}/K_M for fDDF approaching that for hGART. In terms of k_{cat} , moreover GPX-M1 represents a less-desirable hybrid enzyme solution that "acquired" the high K_M of PurN and the low k_{cat} of hGART, resulting in a functional yet suboptimal enzyme.

A more thorough data analysis for the PGX-B12 hybrid enzyme reveals further interesting details. The twofold improved binding affinity for GAR with respect to PurN is counterbalanced by a drop in the catalytic rate by a factor of two. The 15-fold increase in K_M for fDDF further suggests a poorly performing hybrid protein. However, the data for PGX-B12 were recorded at 4°C in contrast to all other kinetic data, including that for wild-type, which were measured at 23°C. Furthermore, the lack of structural integrity of PGX-B12 as indicated by its very fast thermal denaturation and tendency to agglomerate certainly must have an adverse effect on its overall functional performance. If the fold instability is the result of the structural incompatibility of a few key residues, one can speculate that the functional

performance of a structurally “fine-tuned” version of PGX–B12 might indeed surpass the catalytic function of either parent.

The comparison of the kinetic data for PGX–B12 with previously reported results (Ostermeier et al. 1999) underlines the importance of an adequate protein purification and characterization. The removal of misfolded, soluble proteins and aggregates, as well as the careful choice of assay conditions (temperature, crowding agents) permitted a more accurate reassessment of the enzyme’s specific activities, which were underestimated by three to four orders of magnitude. Hybrid enzymes generated by domain substitution are, therefore, not per se poor catalysts but are more likely are catalytically competent enzymes that lack the structural integrity on a timescale necessary for catalysis.

Similar folding problems affected the kinetic studies of the remaining three hybrid proteins, GPX–M24, M55, and GPG–N11. GPX–M24 and M55 both carry the crossover point between $\beta 3$ and $\alpha 3$ but are distinguished by a single amino acid deletion in the latter. Despite the crossover’s distance to the active site and the general structural flexibility of loop regions, even the precise fusion of the formyltransferase fragments in GPX–M24 creates a hybrid enzyme with reduced substrate and cofactor binding affinities (30- to 80-fold for fDDF and GAR relative to PurN and hGART). Yet, the construct shows surprisingly high catalytic turnover ($k_{\text{cat}} = 21/\text{sec}$). We speculate that the binding of GAR is compromised by the disturbance of the extensive hydrogen-bonding interactions between $\alpha 1$, $\alpha 6$, and GAR (Fig. 2). The impact on function by the additional deletion mutation in GPX–M55 is comparably small. The shortening of the loop translates into a moderate reduction in binding affinity for both substrate and cofactor by two- to threefold. This effect can likely be attributed to the extended misalignment of the N-terminal Rossman fold against the remaining protein framework.

The final candidate in our study was the double-crossover hybrid enzyme GPG–N11. Its first crossover is located in the $\beta 3/\alpha 3$ loop region as in GPX–M24 and M55. Following a 74-amino-acid fragment from PurN that contains most of the folate binding site, the sequence reverts to the human protein sequence at the $\beta 6$ strand (Fig. 1). The binding affinities for both substrate and cofactor largely reflect the higher K_M values previously discussed for the GPX–M24 and M55 hybrids. The absence of an improved specificity for GAR despite the reconstitution of its $\alpha 1/\alpha 6$ binding module suggests additional effects introduced by the multiple crossovers. A significant difference between all the single crossover hybrid proteins and GPG–N11 is that the three active site residues (N106, H108, D144) in the latter stem from both parents and not one as in the cases of single crossover hybrids. The fact that a composite active site remains functional is truly remarkable and a tribute to the robust nature of biocatalysts.

Significance of cofactor hydrolysis by hybrid enzymes

As part of our enzymological studies, we also investigated the contribution of fDDF hydrolysis to the measured kinetic data. In previous work, hybrid enzymes between PurN and PurU were shown to possess sufficient formyltransferase activity to accommodate growth of *purN*-deficient *E. coli* yet showed significant levels of cofactor hydrolysis activity in the subsequent in vitro studies (Nixon and Benkovic 2000). Although native GARTs do not show any significant hydrolysis activity, hybrid enzymes between human and *E. coli* GAR formyltransferases may catalyze this side reaction at significant levels because of the changes in folding integrity of the hybrid enzymes.

The hydrolysis of fDDF by wild-type and hybrid GARTs was investigated in the presence and absence of GAR. PGX–B12’s formyltransferase activity rapidly declined at temperatures above 4°C and the observed fDDF turnover at ambient temperature was attributed entirely to cofactor hydrolysis. These data support our hypothesis that PGX–B12’s structural integrity is severely compromised at elevated temperatures and in the absence of crowding agents. Although the hybrid enzyme retains its binding affinity for the substrate and cofactor, it loses its ability to effectively transfer the formyl-group from fDHF to glycnamide ribonucleotide. With a less-compact protein structure, increased accessibility to the active site by water would consequently favor the hydrolysis of the cofactor.

Except for PGX–B12, none of the chimeric enzymes showed notable activity in the absence of GAR. The conversion of fDDF to DDF in the presence of GAR could originate from either formylation of the substrate or hydrolysis of the cofactor, whereby GAR binding only serves to maintain the structural integrity of the protein. These two pathways can be distinguished by monitoring formate formation in a coupled assay with formate dehydrogenase (FDH) and excess NAD. The reduction of NAD by FDH is directly proportional to the amounts of formate formed in the sample solution and can be measured spectrophotometrically. None of the wild-type and hybrid enzymes indicated formic acid production above background. Finally, formation of the formyl–GAR product was further verified by a coupled assay, employing fGAM synthetase and AIR synthetase (Schendel and Stubbe 1986). In summary, the hybrid enzymes described in this work were found to have no significant fDDF hydrolysis activity. The observed kinetic data for the hybrid enzymes, derived from absorption measurements that monitor the conversion of fDDF to DDF are entirely attributed to GAR transformylation.

Modularity and protein engineering

In this study, we have investigated several GART hybrid enzymes in a search for evidence of modular-encoded func-

tional properties. A comparison of the crossover positions found experimentally with the four natural exon boundaries of the human GART (between amino acid 10:11, 54:55, 101:102, and 140:141) suggests that the naturally fusion points indeed make good candidates for gene fragment recombination. The exons show considerable correlation with the crossover locations of functional hybrid enzymes identified by our combinatorial protocol. However, the formation of functional hybrids is not limited to these sites. Several additional functional hybrids with crossover positions beyond the exon boundaries have also been identified.

The creation of novel proteins and enzymes by a simple “Lego-like” principle, however, is an oversimplification in most instances. Analysis of the kinetic data for the various hybrids do indicate that particular functional properties are approximately contained within individual structural units and, as such, can be transferred through domain recombination. Most hybrid constructs, however, lack structural stability, probably because of unfavorable interdomain contacts. Among the seven hybrid enzymes discussed here, the PGX-B12 construct appears to possess the most promising framework for further optimization. Computational studies, as well as random mutagenesis, have the potential to lead to a refined hybrid structure with improved physical properties. Furthermore, the current data collection can be used to direct the design of an idealized hybrid protein. By combining the improved GAR binding affinity of PGX-B12 with the fDDF binding site from GPX-M1, a functionally superior second generation of PurN/hGART multiple crossover hybrids could be constructed.

Materials and methods

All 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 protocol. 10-Formyl-5,8-dideazafolate (fDDF) was purchased from Dr. John Hynes (Medical University of South Carolina).

The hybrid enzymes described in the study were isolated from ITCHY and SCRATCHY libraries and their sequences have been reported previously (Ostermeier et al. 1999; Lutz et al. 2001a,b).

Overexpression of hybrid enzymes

In addition to the two parental genes, seven hybrid genes (Fig. 3; **PGX-B12**, PurN 1–101/hGART 102–202; **GPX-M1**, hGART 1–100/PurN 101–212; **GPX-M12**, hGART 1–129/PurN 54–212; **GPX-M24**, hGART 1–54/PurN 55–212; **GPX-M36**, hGART 1–89/PurN 90–212; **GPX-M55**, hGART 1–56/PurN 58–212, **GPX-N11**, hGART 1–56/PurN 57–131/hGART 132–202) from *in vivo* complementation experiments (see above) were selected for protein overexpression and subsequent kinetic analysis. The individual genes were amplified off the plasmid using gene-specific primers and subcloned into pET22b (Novagen, Madison WI) via the *NdeI/HindIII* restriction sites and their sequences were con-

firmed by DNA sequencing. Subsequent to transformation into *E. coli* BL21 (DE3; Novagen, Madison WI), individual colonies were grown in LB media to OD₆₀₀ ~0.5 at 37°C. Cultures were then cooled on ice, induced with 0.1 mM IPTG, and incubated for 21 h at 18°C. After centrifugation, cell pellets were stored at –70 °C.

Protein purification

The primary purification of the overexpressed hybrid proteins was performed by affinity chromatography, using the C-terminal His-tag. Harvested cells were suspended in 8 mL of buffer L (50 mM Na-phosphate buffer at pH 8.0, 300 mM NaCl, and 10 mM imidazole) containing 0.5 mM PMSF and 1 mg/mL egg white lysozyme and subjected to two rounds of freeze and thaw, followed by sonication. After centrifugation, the supernatant was incubated with 1 mL Ni-NTA resin (Qiagen) for 1 h at 4°C. The suspension was poured into a syringe-column and washed with 10 column volumes of wash solution (20 mM imidazole in buffer L). His-tagged protein was recovered in 2 mL of elution solution (200 mM imidazole in buffer L). The protein solution (0.5 mL) was then loaded onto a Superose 12 gel filtration column (Amersham Biosciences, Piscataway, NJ; 24 mL bed volume) that had been pre-equilibrated in buffer X (50 mM Tris-HCl at pH 8.0, 0.5 mM EDTA, 150 mM NaCl). Fractions were tested for activity and pooled accordingly. After concentration using Centricon spin filters (MWCO 10K, Amicon, Bedford, MA), glycerol was added to a final concentration of 20%, and the enzyme was stored at –70°C. Protein concentrations were determined by the Bradford analysis against BSA (BioRad, Hercules, CA).

Kinetic analysis

The catalytic performance of each GART construct was measured via the deformylation of fDDF at 295 nm followed on a Cary 1 spectrophotometer (Varian, Palo Alto, CA; Smith et al. 1981). In preparation for the assay, the individual enzyme concentrations were adjusted to normalize the specific activity used per assay (0.002 Units/mL rxn) by diluting the stock solution with ice-chilled buffer (0.1 M Hepes at pH 7.5; 0.5 mM EDTA; 10 mM dithiothreitol; 1 g/L bovine serum albumin; 20% glycerol). Furthermore, the precise concentration of β -GAR was determined enzymatically by measuring DDF-formation at 295 nm ($\Delta\epsilon = 18.9/\text{mM}/\text{cm}$) under a limiting concentration of GAR (Shim and Benkovic 1998). fDDF was quantified by the absorbance at 254 nm ($\epsilon = 23.5/\text{mM}/\text{cm}$; Smith et al. 1981).

Kinetic parameters were determined in triplet over a substrate concentration range of 0.5–400 μM β -GAR and 0.5–400 μM fDDF. With the exception of PGX-B12, the enzymatic activity of all proteins was assayed at 23°C in reaction buffer (0.1 M Hepes at pH 7.5; 0.5 mM EDTA). Because of its decreased *in vitro* stability, the activity of PGX-B12 was measured in a 20% glycerol-substrate solution at 4°C. Data analysis was performed by standard Lineweaver-Burk plots.

Protein activity vs. temperature

The temperature-dependency of GART activity was investigated in a reaction buffer (0.1 M Hepes at pH 7.5; 0.5 mM EDTA),

containing 40 μM fDDF and 40 μM β -GAR (total volume: 100 μL). Reaction buffer and substrates, equilibrated at the selected temperatures for 5 min, were mixed with enzyme and incubated for 10 min (max. conversion: 20%). The reaction was quenched by heat-denaturation (95°C for 5 min). The hybrid's activity was determined by quantifying the level of DDF from its absorbance at 295 nm.

fDDF hydrolysis versus catalysis

The significance of enzymatic hydrolysis of fDDF in comparison to the formylation reaction was investigated by two coupled-enzyme assays. Cofactor hydrolysis was measured spectrophotometrically via formate oxidation by formate dehydrogenase and NAD^+ . Briefly, the assay was started by fDDF addition (40 μM final concentration) to a preequilibrated mixture of GART (0.002 Units/mL rxn), FDH (1 mg/mL; Roche Biochemicals, Indianapolis, IN) and 1 mM NAD^+ in reaction buffer (0.1 M Hepes at pH 7.5; 0.5 mM EDTA, 40 μM GAR) at 23°C. The increase in absorbance at 340 nm is a combination of NADH and DDF production and accurate NADH quantities were calculated by subtraction of DDF rates of formation, determined in previous experiments. In a second assay, the competition between fDDF hydrolysis and formylation of GAR was monitored via the enzymatic conversion of the produced fGAR by GAM synthetase and AIR synthetase. The experiments were performed essentially as described by Schendel and Stubbe (1986).

Acknowledgments

We thank Dr. Ravi Rajagopalan for providing samples of fGAR synthetase and AIR synthetase. This work was supported in part by the NIH GM 24129.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Almasy, R.J., Janson, C.A., Kan, C.C., and Hostomska, Z. 1992. Structures of apo and complexed *Escherichia coli* glycinamide ribonucleotide transformylase. *Proc. Natl. Acad. Sci.* **89**: 6114–6118.
- Cane, D.E., Walsh, C.T., and Khosla, C. 1998. Biochemistry—Harnessing the biosynthetic code: Combinations, permutations, and mutations. *Science* **282**: 63–68.
- Cayley, S., Lewis, B.A., Guttman, H.J., and Record Jr., M.T. 1991. Characterization of the cytoplasm of *Escherichia coli* K-12 as a function of external osmolarity. Implications for protein–DNA interactions in vivo. *J. Mol. Biol.* **222**: 281–300.
- Chen, P., Schulze-Gahmen, U., Stura, E.A., Inglesse, J., Johnson, D.L., Marolewski, A., Benkovic, S.J., and Wilson, I.A. 1992. Crystal structure of glycinamide ribonucleotide transformylase from *Escherichia coli* at 3.0 Å resolution. A target enzyme for chemotherapy. *J. Mol. Biol.* **227**: 283–292.
- Daubner, S.C., Schrimsher, J.L., Schendel, F.J., Young, M., Henikoff, S., Patterson, D., Stubbe, J., and Benkovic, S.J. 1985. A multifunctional protein possessing glycinamide ribonucleotide synthetase, glycinamide ribonucleotide transformylase, and aminoimidazole ribonucleotide synthetase activities in de novo purine biosynthesis. *Biochemistry* **24**: 7059–7062.
- Dorit, R.L., Schoenbach, L., and Gilbert, W. 1990. How big is the universe of exons? *Science* **250**: 1377–1382.
- Eder, J. and Kirschner, K. 1992. Stable substructures of eightfold β - α -barrel proteins—fragment complementation of phosphoribosylanthranilate isomerase. *Biochemistry* **31**: 3617–3625.
- Go, M. 1981. Correlation of DNA exonic regions with protein structural units in haemoglobin. *Nature* **291**: 90–92.
- Hinman, M.B., Jones, J.A., and Lewis, R.V. 2000. Synthetic spider silk: A modular fiber. *Trends Biotechnol.* **18**: 374–379.
- Hocker, B., Jurgens, C., Wilmanns, M., and Sterner, R. 2001. Stability, catalytic versatility and evolution of the (β α)(8)-barrel fold. *Curr. Opin. Biotechnol.* **12**: 376–381.
- Inglesse, J., Johnson, D.L., Shiau, A., Smith, J.M., and Benkovic, S.J. 1990. Subcloning, characterization, and affinity labeling of *Escherichia coli* glycinamide ribonucleotide transformylase. *Biochemistry* **29**: 1436–1443.
- Kan, C.C., Gehring, M.R., Nodes, B.R., Janson, C.A., Almasy, R.J., and Hostomska, Z. 1992. Heterologous expression and purification of active human phosphoribosylglycinamide formyltransferase as a single domain. *J. Protein Chem.* **11**: 467–473.
- Klein, C., Chen, P., Arevalo, J.H., Stura, E.A., Marolewski, A., Warren, M.S., Benkovic, S.J., and Wilson, I.A. 1995. Towards structure-based drug design: Crystal structure of a multisubstrate adduct complex of glycinamide ribonucleotide transformylase at 1.96 Å resolution. *J. Mol. Biol.* **249**: 153–175.
- Luger, K., Hommel, U., Herold, M., Hofsteenge, J., and Kirschner, K. 1989. Correct folding of circularly permuted variants of a β - α -barrel enzyme in vivo. *Science* **243**: 206–210.
- Lutz, S., Ostermeier, M., and Benkovic, S.J. 2001a. Rapid generation of incremental truncation libraries for protein engineering using α -phosphothioate nucleotides. *Nucleic Acids Res.* **29**: E16.
- Lutz, S., Ostermeier, M., Moore, G.L., Maranas, C.D., and Benkovic, S.J. 2001b. Creating multiple-crossover DNA libraries independent of sequence identity. *Proc. Natl. Acad. Sci.* **98**: 11248–11253.
- Mainfroid, V., Goraj, K., Rentier-Delrue, F., Houbrechts, A., Loiseau, A., Gohimont, A.C., Noble, M.E., Borcher, T.V., Wierenga, R.K., and Martial, J.A. 1993. Replacing the (β α)-unit 8 of *E. coli* TIM with its chicken homologue leads to a stable and active hybrid enzyme. *Protein Eng.* **6**: 893–900.
- Mootz, H.D., Schwarzer, D., and Marahiel, M.A. 2000. Construction of hybrid peptide synthetases by module and domain fusions. *Proc. Natl. Acad. Sci.* **97**: 5848–5853.
- Mullen, C.A. and Jennings, P.A. 1998. A single mutation disrupts the pH-dependent dimerization of glycinamide ribonucleotide transformylase. *J. Mol. Biol.* **276**: 819–827.
- Nixon, A.E. and Benkovic, S.J. 2000. Improvement in the efficiency of formyl transfer of a GAR transformylase hybrid enzyme. *Protein Eng.* **13**: 323–327.
- O'Maille, P.E., Bakhtina, M., and Tsai, M.D. 2002. Structure-based combinatorial protein engineering (SCOPE). *J. Mol. Biol.* **321**: 677–691.
- Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., and Thornton, J.M. 1997. CATH—a hierarchic classification of protein domain structures. *Structure* **5**: 1093–1108.
- Ostermeier, M., Shim, J.H., and Benkovic, S.J. 1999. A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.* **17**: 1205–1209.
- Richey, B., Cayley, D.S., Mossing, M.C., Kolka, C., Anderson, C.F., Farrar, T.C., and Record Jr., M.T. 1987. Variability of the intracellular ionic environment of *Escherichia coli*. Differences between in vitro and in vivo effects of ion concentrations on protein–DNA interactions and gene expression. *J. Biol. Chem.* **262**: 7157–7164.
- Rossmann, M.G., Moras, D., and Olsen, K.W. 1974. Chemical and biological evolution of nucleotide-binding protein. *Nature* **250**: 194–199.
- Schendel, F.J. and Stubbe, J. 1986. Substrate specificity of formylglycinamide synthetase. *Biochemistry* **25**: 2256–2264.
- Scott, K.A., Steward, A., Fowler, S.B., and Clarke, J. 2002. Titin; a multidomain protein that behaves as the sum of its parts. *J. Mol. Biol.* **315**: 819–829.
- Seidel, H.M., Pompliano, D.L., and Knowles, J.R. 1992. Exons as microgenes? *Science* **257**: 1489–1490.
- Shim, J.H. and Benkovic, S.J. 1998. Evaluation of the kinetic mechanism of *Escherichia coli* glycinamide ribonucleotide transformylase. *Biochemistry* **37**: 8776–8782.
- Sieber, V., Martinez, C.A., and Arnold, F.H. 2001. Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.* **19**: 456–460.
- Smith, G.K., Mueller, W.T., Benkovic, P.A., and Benkovic, S.J. 1981. On the cofactor specificity of glycinamide ribonucleotide and 5-aminoimidazole-4-carboxamide ribonucleotide transformylase from chicken liver. *Biochemistry* **20**: 1241–1245.
- Zhang, Y., Desharnais, J., Greasley, S.E., Beardsley, G.P., Boger, D.L., and Wilson, I.A. 2002. Crystal structures of human GAR Tase at low and high pH and with substrate β -GAR. *Biochemistry* **41**: 14206–14215.