Improvement in the efficiency of formyl transfer of a GAR transformylase hybrid enzyme

Andrew E.Nixon¹ and Stephen J.Benkovic²

1Present address: Dyax Corp., One Kendall Square, Building 600, Suite 623, 5th Floor, Cambridge, MA 02139, USA
2To whom correspondence should be addressed. Email: sjb1@psu.edu

A hybrid glycinamide ribonucleotide transformylase was assembled from two protein domains that were treated as discrete modules. One module contained the ribonucleotide binding domain from the purN glycinamide ribonucleotide transformylase; the second module contained the catalytic machinery and the formyl tetrahydrofolate binding domain from the enzyme encoded by purU, formyl tetrahydrofolate hydrolase. The resultant enzyme showed 0.1% catalytic activity of the wild-type glycinamide ribonucleotide transformylase enzyme but had a formyl transfer efficiency of 10%. A combinatorial mutagenesis approach was used to improve the solubility and formyl transfer properties of the hybrid enzyme. The mutagenized hybrid glycinamide ribonucleotide transformylase was initially expressed as a fusion to the ß-peptide of ß-galactosidase. Clones were selected for improvement in solubility by determining which clones were capable of ß-complementation using a blue/white screen. One clone was further characterized and found to have an improved efficiency of transfer of the ribonucleotide increasing this property to >95%.

Keywords: DNA shuffling/GAR transformylase/hybrid enzyme/protein engineering

Introduction

An understanding of how nature has created a vast and diverse array of enzymes offers insight into how protein engineering may be used to create enzymes with specific functions. The evolution of proteins with novel function is generally thought to be a result of gene duplication followed by diversification through the accumulation of point mutations (Petsko et al., 1993). Gene recombination, through processes such as exon shuffling (Gilbert, 1978, 1987), allows the creation of unique proteins through the recruitment of functional domains from proteins with diverse functions. Such domain recruitment has been proposed to be important in the evolution of metabolic pathways (Jensen, 1976) and is one way in which novel activities may be produced in nature.

Glycinamide ribonucleotide (GAR) transformylase is one such enzyme which may have arisen as a result of recombination of protein domains. In Escherichia coli the purN encoded GAR transformylase, a 23 kDa enzyme, catalyzes the transfer of the formyl group from 10-formyltetrahydrofolate to the free amino of GAR to give formyl-GAR and tetrahydrofolate as products (Figure 1). Because of the importance of GAR transformylase in purine biosynthesis and the potential therapeutic benefit from its inhibition, the enzyme has been thoroughly characterized. A high-resolution crystal structure of the enzyme bound to a multisubstrate adduct inhibitor (Klein et al., 1995) and protein labeling studies (Inglese et al., 1990) have allowed the identification of the regions involved in substrate and cofactor binding. From mutagenesis (Warren et al., 1996) and kinetic studies (Shim and Benkovic, 1998) a catalytic mechanism for the enzyme has been proposed.

At the DNA level there is a marked homology between the 10-formyltetrahydrofolate binding domain of the purN encoded GAR transformylase, the cofactor binding domain of formylmethionine tRNA transferase and a C-terminal region of 10-formyltetrahydrofolate hydrolase, the product of the purU gene (Nagy et al., 1995). It is tempting to suggest that all three cofactor domains originated from a common ancestor that subsequently acquired novel functions through recruitment of other protein domains. At 60% sequence homology the 10-formyltetrahydrofolate domain from purU is most closely related to that from purN. The product of the purU gene, 10-formyltetrahydrofolate hydrolase, catalyzes the hydrolysis of the 10-formyltetrahydrofolate cofactor to tetrahydrofolate and free formate (Figure 1). The free formate is then thought to be utilized by a second GAR transformylase, encoded by the purT gene, to formylate GAR in a reaction driven forward by the hydrolysis of ATP (Nygaard and Smith, 1993). The physiological requirement for a second GAR transformylase is not clearly understood, although it is possible that this system is an evolutionary precursor that is still advantageous under conditions of stress.

To investigate the possibility of utilizing evolutionary strategies such as domain recruitment in the design of biocatalysts, we have previously described the creation of a hybrid enzyme with GAR transformylase activity (Nixon et al., 1997). Using a modular approach in which the protein domains were treated as discrete elements, a hybrid enzyme was created from the GAR binding domain of the purN derived GAR transformylase of E.coli and the formyltetrahydrofolate binding domain from purU. Although this enzyme was capable of formyl transfer, there was a strong preference for hydrolysis of the formyltetrahydrofolate cofactor to give formate and tetrahydrofolate. The most likely explanation of this observation is that this hybrid enzyme exists in two states, a closed conformation that is capable of formyl transfer and an open form that is capable of cofactor hydrolysis but not formyl transfer. We reasoned that in the open form hydrophobic surfaces of the protein would be exposed to solvent, which would lead to the solubility problems we encountered. We anticipated that a shift in equilibrium towards the closed conformation should improve the ratio of formyl transfer to hydrolysis. Here we describe efforts to increase the efficiency of formyl transfer of the hybrid GAR transformylase.

Materials and methods

Materials

Restriction enzymes were obtained from Promega and New England Biolabs. T4 DNA ligase and Taq polymerase were
Fig. 1. Reactions catalyzed by the purN and purU enzymes.

obtained from Promega. Ultrapure dNTPs were obtained from Boehringer Mannheim. Agarose for analytical gel electrophoresis was obtained from Kodak. For preparative gel electrophoresis, Nusieve GTG from FMC bioproducts was used.

Bacterial strains

DH-5α (GIBCO/BRL-Life Technologies), BL21(DE3) pLys S (Novagen) and MW12 [ara Δ(gpt-pro-lac) thi rbs-221 ilvB2102 ilvHI2202 purN-lacZ−Y+::KanR purT (DE3)] were used.

Plasmid construction

The stop codon was removed from the purN/PurU fusion using the megaprimer mutagenesis method (Sarkar and Sommer, 1990; Pont-Kingdom, 1994). A first PCR reaction was performed using purU reverse (ATC GAT AAG CTT TAC GTT GAG AAA AAT GAA C) and a mutagenic oligo in which the concentration between 2 and 195 µM was varied between 4 and 160 µM. The primers purN forward (5′ GAT ATA CAT ATG AAT ATT GTG GTG CTG ATT TCC) and XhoI mutant reverse (5′ CCA GTT ATT TTT GCT GCC GAC TCG AGA ATA ATC G) were then used to amplify the assembled hybrid enzyme gene in a further 30 cycles using the same PCR program as described above.

β-Galactosidase assays

Blue/white screening was performed essentially as described (Sambrook et al., 1989). Briefly, DH5-α transformed with the fusion construct were plated on LB-agar plates containing chloramphenicol (35 µg/ml), IPTG (1 mM) and X-Gal (50 µg/ml). The plates were then incubated at 37°C for 16 h, generally sufficient time for the blue color to develop. The chemiluminescent β-galactosidase substrate Galacton-Star (Tropix, MA, USA) was used in accordance with the manufacturer’s protocol.

Protein preparation

Expression of the shuffled hybrid enzyme–lacZ fusion was in the GAR transformylase auxotroph MW12. Cells containing the hybrid enzyme–lacZ fusion expression system were grown in minimal media containing 35 mg/ml chloramphenicol at 37°C until an OD600 of 0.4–0.6 was reached. Expression was then induced by the addition of IPTG to a final concentration of 1 mM and the cells allowed to grow on for a further 4 h at 37°C. Cells were harvested by centrifugation and the pellet frozen. Expression and purification of the His-tagged hybrid enzyme was as previously described (Nixon et al., 1997).

Enzyme assays

Enzyme activity was determined by monitoring the formylation of 10-formyl-5,8-dideazafolate (fDDF) (Δε = 18.9 mM⁻¹ cm⁻¹ at 295 nm). Assays were performed in 50 mM Tris, 1 mM EDTA at pH 7.5 and 20°C using a Gifford 252 spectrophotometer. Steady-state parameters were determined using a hybrid enzyme concentration of 2 µM. The GAR concentration was varied between 4 and 160 µM and the fDDF concentration between 2 and 195 µM. For specific activity determination, an excess of substrates was used.

Formyl transfer

Hybrid enzyme (1 µM) was incubated with 0.7 mM [14C]fDDF (388 µCi/µmol) alone or in addition to 0.8 mM GAR for 12 h at 37°C. The reaction products were separated by TLC using PEI-Cellulose plates (EM Separations, Gibbstown, NJ) eluting with 50 mM K2HPO4 (pH 7.0) and quantitated using a Molecular Dynamics PhosphorImager.

Results

Construction of the hybrid enzyme used in this study has been described previously (Nixon et al., 1997). The particular construct used was the ‘long’ hybrid enzyme in which residues 1–115 from the purN gene were fused to residues 80–280 from the purU gene. Sequence comparisons of the two enzymes indicate that residues 144–237 from purU, which comprise the 10-formyltetrahydrofolate binding domain, are 60% homologous to residues 63–157 from purN. The construct in which residues 1–63 from purN were fused to residues 144–280 from purU was not catalytically active. It was therefore proposed...
that residues 63–115 from the purN domain and residues 80–144 from the purU domain serve as a flexible linker long enough to allow the two domains to interact productively with one another.

Using a substrate which contained a radiolabeled formyl group, it was possible to follow partitioning between the possible products, either by hydrolysis to free formate or transfer to GAR to yield formyl-GAR (fGAR). We determined that the hybrid enzyme favors hydrolysis of the cofactor over transfer to GAR, probably because the GAR binding domain is not aligned to facilitate capture of the formyl group from 10-formyltetrahydrofolate. This misalignment and the observed solubility problems most likely have the same root cause, faulty interactions along the domain interface. Since it is difficult to identify the key residues involved in such inter-domain contacts, we chose to use random mutagenesis, specifically DNA shuffling (Stemmer, 1994), to introduce mutations that may improve solubility and in turn improve formyl transfer efficiency.

DNA shuffling has previously been shown to have utility in improving the solubility of the green fluorescent protein (Cramer et al., 1996). In this case increased fluorescence of the bacterial colonies expressing the mutated protein was used to screen for improved solubility. We chose to use a similar strategy to improve the solubility of the GAR transformylase hybrid enzyme in this study. We developed a system to allow screening of improvements in protein solubility based on α-complementation. α-Complementation refers to the ability of the α-peptide, which consists of residues 1–146 of β-galactosidase, to restore fully the activity of an enzymatically inactive β-galactosidase created by the deletion of residues 11–41 of β-galactosidase. Our system creates a fusion between the α-peptide and the hybrid GAR transformylase. Cells which contain an insoluble fusion protein will lack β-galactosidase activity. This can be visually assessed using a blue/white screen. To ensure that the identified mutants were still capable of GAR transformylase activity, the vectors encoding the seven mutants were transformed into the auxotrophic E. coli, expressed and the specific activity determined. All mutants were found to have a similar level of GAR transformylase activity.

The DNA from the colonies was prepared and sequenced (Table I). Of the eight base substitutions identified, five resulted in amino acid replacement. Three of the mutations, two of which were neutral, fall in the region of the purU domain that has no homologous sequence in purN and that is believed to loop out into solution (Figure 2). All the mutations identified were a result of point mutations; no frameshifts were identified. The mutations were mapped on to the structure of glycinamide ribonucleotide transformylase using a combination of sequence alignments and also a model of the purU formyltetrahydrofolate binding domain. The model was prepared using the automated, knowledge-based tool Swiss Model (Peitsch, 1995). Four of the five mutations fall around the interface of the two domains, although none appear to make direct interfacial contact.

A chemiluminescent assay was used to rank the β-galactosidase activity of the hybrid enzyme–lacZ fusions. Clone 6 was found to have the highest level of β-galactosidase activity and was carried forward for more detailed kinetic studies. Production of sufficient amounts of the mutant protein required subcloning into a high level expression system. However, when the mutant protein was overexpressed using the strong T7 promoter most (>95%) of the protein formed inclusion bodies. This most likely results from using the weaker lac promoter for protein expression during selection for increased solubility but using the stronger T7 promoter for production.
of protein for kinetic analysis. Since inclusion body formation was previously observed with the GAR transformylase hybrid enzyme, we were able to purify the protein as described previously (Nixon et al., 1997). The kinetic parameters of this protein were determined using the change in absorbance of fDDF upon deformylation to dideazafolate (DDF) and found to be similar to the parent hybrid enzyme (Table II). The $K_m$ for substrates was determined to be similar to that for the parent modules 12 $\mu$M for GAR vs 19 $\mu$M for the wild-type purN and 35 $\mu$M for fDDF vs 7 $\mu$M for wild-type purU. The $k_{cat}$ for the shuffled hybrid enzyme was 0.019 s$^{-1}$, which compares with 0.016 s$^{-1}$ for the original hybrid enzyme, 16 s$^{-1}$ for the wild-type purN and 0.026 s$^{-1}$ for purU.

Since the spectroscopic assay utilising fDDF gives a measure of the rate of cofactor deformylation only, an alternative method is required to determine the amount of formyl transfer. Previously we have determined the ratio of hydrolysis to formyl transfer using fDDF which contains a $^{14}$C radiolabel in the formyl group (Nixon et al., 1997). The hybrid enzyme was incubated with an excess of $^{14}$CfDDF and GAR for 12 h at 37°C. The products were separated by TLC and imaged using a PhosphorImager (Molecular Dynamics). The major product produced was iGAR with <1% production of free formate. The ratio of hydrolysis to formyl transfer for the wild-type enzyme is ~1:10 000, for the original hybrid GAR transformylase ~40:1 and for the mutant hybrid GAR transformylase >1:100. This represents an improvement in efficiency of formyl transfer for the shuffled hybrid enzyme of >4000. The absolute rate of formyl transfer can be determined from the $k_{cat}$ value and the efficiency of formyl transfer. For the original hybrid enzyme this value is 0.0004 s$^{-1}$ and for the shuffled hybrid enzyme >0.0188 s$^{-1}$, a 47-fold improvement in the rate of formyl transfer.

**Discussion**

Rational mutagenesis has previously been utilized to modify substrate and cofactor specificity and to engineer control into enzymes. Although this approach is labor intensive and requires knowledge of protein structure and ideally catalytic mechanism, it is attractive since there is little requirement for a sophisticated screening or selection system. Stochastic methodology such as error-prone PCR and DNA shuffling is attractive since it requires no prior knowledge of structure or catalytic mechanism. Also, this approach very quickly produces a large library of variants that can be investigated. However this approach is at its most powerful when used in combination with either a genetic selection or a suitable screen. In nature it is most likely that the main functional building blocks are assembled and then these are modified to yield maximum activity through the accumulation of mutations. This situation is akin to the rational design of an enzyme followed with random mutagenesis and selection for improvement of a particular property and is the approach we have adopted.

We previously described the creation of a GAR transformylase hybrid enzyme that was capable of formyl transfer although the partitioning of the formyl group between free formate and fGAR was ~40:1 in favor of free formate. The most likely explanation of this observation is that the two domains of the hybrid enzyme are not orientated to facilitate efficient transfer of the formyl group to the ribonucleotide. Difficulties in ensuring correct interactions across protein domain interfaces has previously been noted for oligomeric protein assembly (Murata and Schachman, 1996). It seems reasonable to expect that the problems which exist in the formation of heterologous oligomeric proteins would also exist in the formation of hybrid enzymes from domains of two different proteins. Based on the observations for oligomeric protein assembly, it was expected that modification of purN–purU domain–domain interface should produce a more soluble enzyme. In addition to improved solubility, improved domain assembly should result in a hybrid GAR transformylase that is capable of efficient formyl transfer.

Given the general complexity of interactions across protein domain interfaces, rather than target discrete residues to be mutated, point mutations were introduced in a random fashion over the whole of the hybrid enzyme gene using DNA shuffling. Mutants were then screened using a β-galactosidase-based screening system that would identify those mutants that were more soluble. It was proposed that the more soluble mutants would also be more efficient in transfer of formyl to GAR. Clone 6 (Table I) was found to have the highest level of β-galactosidase activity. With the exception of one addition neutral point mutation found in clone 5, clone 6 has the same pattern of mutations as clone 5. It is likely, therefore, that the difference in β-galactosidase activity stems from differences in expression levels of the two mutant proteins. Four of the mutations in clone 6 fall around the interface of the two domains, although from an analysis of the purN crystal structure none of these residues make any direct inter-domain contacts. In the absence of a crystal structure of the hybrid enzyme or of purU it is difficult to fully explain how these mutants could act to stabilize the interaction of the two domains.

The rate of fDDF deformylation of both the parent hybrid enzyme and the shuffled hybrid enzyme is three orders of magnitude lower than that for wild-type purN. It is worth noting that although this value is close to that measured for purU, the source of the catalytic module, it is not expected that this would be an upper limit on the rate of formyl transfer. Rather, we would expect that the hybrid enzyme should be capable of activity similar to that of wild-type purN since the amino group of the ribonucleotide should be a more potent nucleophile than water.

At the simplest level domain swapping to create novel biocatalysts offers the potential of vast numbers of novel biocatalysts, many of which would have industrial and perhaps
therapeutic value. The modular approach that has been described here and elsewhere has demonstrated the power of the domain swapping technique. A more refined domain swapping technology has recently been described (Hopfner et al., 1998). This work describes the creation of a hybrid enzyme generated by combining the N-terminal sub-domain of the coagulation factor X with the C-terminal subdomain from trypsin. Although successful, this approach relies on detailed information of the proteins involved. A combinatorial approach to the fusion of protein domains has recently been described (Ostermeier et al., 1999a,b). This approach has the potential to move beyond the limitations inherent in a rational domain swapping venture.

References

Received March 24, 1999; revised December 10, 1999; accepted February 20, 2000