Proc. KMB Annual Meeting
June 24-26, 2003, Muju, Korea

On the Structural and Functional Modularity in Proteins

Seung-Goo Lee[1], Stefan Lutz[2], and Stephen J. Benkovic

Department of Chemistry, Pennsylvania State University, University Park, PA 16802; Present address:
[1]Laboratory of Microbial Function, KRIBB, 52, Yusong, Taejon 305-333;
[2]Department of Chemistry, Emory University, Atlanta, GA 30322;

Modularity is an attractive concept for protein evolution in Nature and in the laboratory. The hypothesis centers on the build-up of complex structures from small, independent peptidylic 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 for such a modular design and organization can be identified on all levels of protein structure. Polyketide syntheases and nonribosomal peptide synthetases are representatives for covalently-linked clusters of multiple enzymes with individual functions, catalyzing consecutive steps in the synthesis of complex secondary metabolites.

Glycinamide ribonucleotide formyl-transferases (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. Assisting in our function-based study of engineered GAR formyl-transferases, structural data for *Escherichia coli* and human both enzymes are available. Despite of a low amino acid sequence identity (31%), the two proteins show striking structure overlaps with an average rms deviation of 1.5 Å (Zhang et al. 2002).

Beyond computational and gene mapping analysis, several experimental techniques for the exploration of fragmentation and recombination sites have been introduced in recent years. Summarized as combinatorial, structure-based protein engineering methods, incremental truncation for the creation of hybrid enzymes (ITCHY) (Ostermeier et al. 1999), SCRATCHY (Lutz et al. 2001), sequence homology-independent protein recombination (SHIPREC) (Sieber et al. 2001), and structure-based combinatorial protein engineering (SCOPE) (O'Maille et al. 2002) provide the means to explore the modular nature of proteins by allowing the swapping of secondary structure elements and domains between structure homologues 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 *Escherichia coli* (PurN) and human (hGART) (Fig.1).

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. From the pool of confirmed (functional) and sequenced hybrid enzymes, we selected six single-crossover and one double-crossover construct for overexpression and detailed kinetic analysis *in vitro*. The choice of these constructs was based on the nature of their crossovers. The PGX-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.

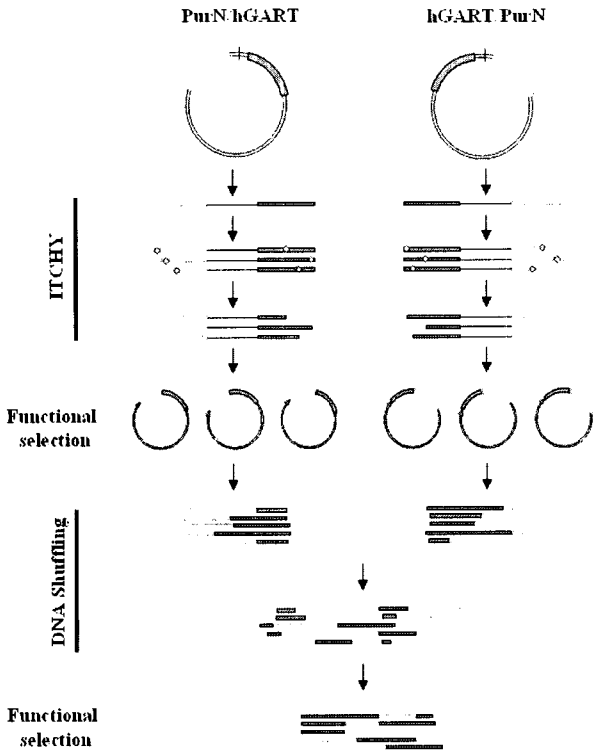

Fig. 1. Schematic overview of SCRATCHY. Initially, individual incremental truncation libraries of the two complementary constructs (pDIM-PGX and pDIM-GPX) were created. After the first functional selection to recover hybrids of parental size and in-frame, the libraries were mixed and submitted to DNA shuffling. A final selection identifies functional constructs.

The third hybrid, GPX-M36 has its fusion point at position 89/90, closely resembling GPX-M1 but carrying helix-4 of PurN (Fig. 2). 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 formyl-dihydrofolate (fDHF) and the contribution of the network of non-covalent interactions on the overall stability of the protein.

The two GPX fusions M24 and M55 both possess crossovers in the connecting loop between helix-2 and 3 (Fig. 2). While 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.

Our study also includes a hybrid structure with two crossovers, the GPG-N11 construct (N-terminal *hGART*, central portion from *purN*, C-terminal *hGART*). The location of the crossovers at position 56/57 and 131/132 creates a hybrid enzyme with the fDHF-binding site and the active site residues (N106/H108) from the *E. coli* protein but the glycinamide-ribonucleotide binding site, as well as the majority of the surrounding protein framework from its human homolog.

Finally, the *in vivo* complementation assays identified the rather unusual yet functional candidate GPX-M12 that has a 75 amino acid insertion. 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.

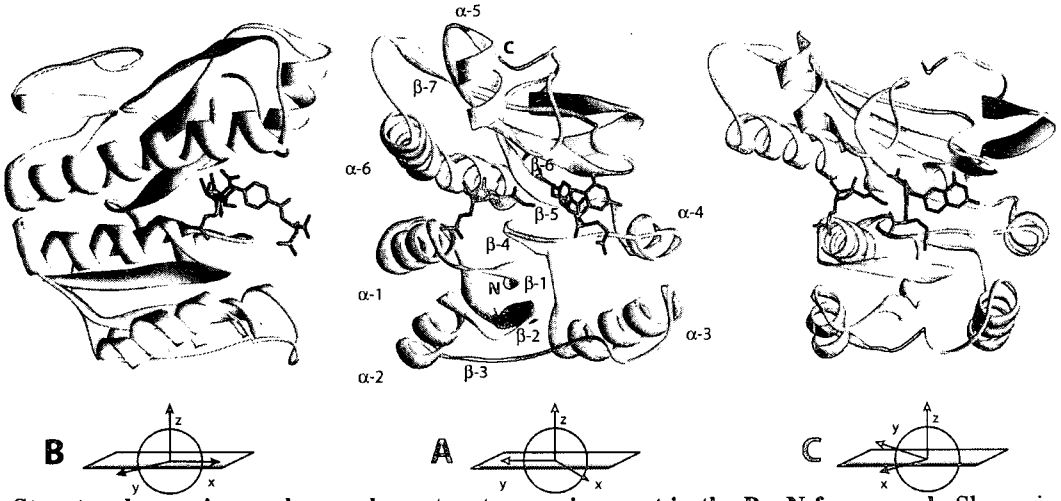

Fig. 2. Structural overview and secondary structure assignment in the PurN framework. Shown in red are the enzyme-bound substrate and cofactor while the active site residues N106, H108, and D144 are highlighted in green.

Effects of Domain Swapping on Catalysis

The concept of modularity as a determining factor in protein frameworks makes an interesting target for protein engineering and the studies of natural protein evolution. Such a notion would facilitate the generation of more efficient, tailor-made biocatalysts by recombination of selected structure elements and domains from the best-performing parental sequences, creating functionally superior hybrid enzymes. We have, as part of this study on GART, initiated a detailed investigation into the catalytic properties of the 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.

Starting with the wild-type enzymes, the kinetic data for the parents outline two distinct strategies to achieve similar specific catalytic rates. While the K_M's for PurN are 10 to 20-fold higher than for hGART, the bacterial protein compensates with a sixteen-fold higher catalytic rate of the transformylation. In reference to the above hypothesis, substitution of the functionally more efficient protein fragments from either parent could lead to hybrid constructs with the tight substrate binding of the human enzyme *and* the high 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 strongest catalytic activity of all hybrids. The comparison of the k_{cat}/ K_M values for the two hybrid and wild-type enzymes shows a wild-type like behavior of PGX-B12 in respect to the GAR substrate while GPX-M1 possess parents-like fDDF affinity. From these data, we conclude that GPX-M1 represents a less desirable hybrid enzyme solution that "acquired" the high K_M's of PurN and the low k_{cat} of hGART, resulting in a functional yet suboptimal enzyme. Details of the modular behavior of other hybrids will be addressed in presentation.

Modularity and Protein Engineering

Our study supports the notion that the creation of novel proteins and enzymes by a simple "Lego-like" principle is an oversimplification in most instances. While the analysis of the hybrid's kinetic data do indicate that particular functional properties are contained within individual structural units and can be transfer through domain recombination, most such hybrid constructs lack overall structural integrity due to unfavorable interdomain contacts. From the seven hybrid enzymes discussed here, the PGX-B12 construct is

considered 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, which could translate, into superior functional performance.

References

1. Lutz, S., Ostermeier, M., Moore, G.L., Maranas, C.D., and Benkovic, S.J. 2001. Creating multiple-crossover DNA libraries independent of sequence identity. *Proc Natl Acad Sci U S A* 98: 11248-11253.
2. O'Maille, P.E., Bakhtina, M., and Tsai, M.D. 2002. Structure-based combinatorial protein engineering (SCOPE). *J Mol Biol* 321: 677-691.
3. 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.
4. Sieber, V., Martinez, C.A., and Arnold, F.H. 2001. Libraries of hybrid proteins from distantly related sequences. *Nat Biotechnol* 19: 456-460.
5. Zhang, Y., Desharnais, J., Greasley, S.E., Beardsley, G.P., Boger, D.L., and Wilson, I.A. 2002. Crystal structures of human GAR Tfase at low and high pH and with substrate beta-GAR. *Biochemistry-Us* 41: 14206-14215.