FuncEvoDupliGenes **- Final Publishable Summary Report (David Chou)**

Gene duplication followed by sequence divergence is recognized as a key mechanism in evolution to generate functional innovations and enrich the complexity of genetic networks. Though gene duplication is frequently observed, the ensuing process leading to functional differentiation among gene duplicates remains obscure. Lack of this critical information has limited not just our fundamental understanding of protein evolution, but the success in engineering enzymes for industrial applications and our ability to predict evolution, for instance, mutations accessible to pathogens against antibiotic treatments. To fill this knowledge gap, I used an integrative approach to bring mechanistic insights into the functional diversification among flavoprotein disulfide reductases (FDRs), namely glutathione reductase (Gor), lipoamide reductase (Lpd), soluble transhydrogenase (SthA), reactive chlorine reductase (RlcA), and mercuric reductase (MerA). Diverged before the split between prokaryotes and eukaryotes, these ancient gene duplicates now participate in distinct branches of cell metabolism but maintain remarkable sequence and structure conservation. Furthermore, they exhibit the ability to turn over secondary substrates (i.e. promiscuity). This project aims for addressing the evolutionary history, functional divergence, and mechanistic basis of promiscuity of FDRs, and deliver their utility, particularly transhydrogenase activity, to metabolic engineering and biotechnology.

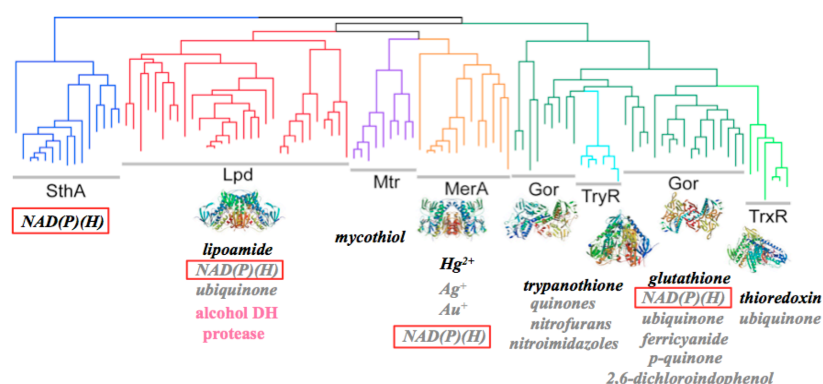

Figure 1. The gene tree of FDRs and their reacting substrates. Primary and secondary substrates of FDRs are shown below the gene tree as black and grey texts, respectively. FDRs that exhibit transhydrogenase activity are indicated by the capability to react with both NAD(H) and NADP(H) redox cofactors (red box).

The first phase of the project involved gene cloning, protein expression, and functional characterization of FDRs. Genes of four FDRs indiginous to *Escherichia coli* (i.e. soluable transhydrogenase (SthA), glutathione reductase (Gor), lipoamide reductase (Lpd), and reactive chlorine reductase (RlcA)) as well as Gor from *Saccharomyces cerevisiae* and the mercuric reductase (MerA) from *Salmonella enterica* serovar Typhi CT18 were cloned into an expression vector with a His-tag fused at the N-terminal. These FDRs were successfully expressed in wild-type (WT) *E. coli*, allowing production of FDRs in large quantity for enzyme assays. Examination of catalytic activity revealed that FDRs showed much higher catalytic activity toward their primary substrates (10^2~10^4–fold) than secondary subsrates reported in the literature.

In the second phase, I concentrated on the promiscuous transhydrogenase activity possessed by several FDRs (Figure 1). To establish a viability-based screening platform, I constructed *E. coli* (Δ*pgi* Δ*sthA*) as prior work reported this mutant unable to grow on glucose due to the lack of transhydrogenase activity. Nevertheless, when testing its capability to catabolize glucose, *E. coli* (Δ*pgi* Δ*sthA*) colonized glucose minimal agar 4 days after inoculation. The result invalidated the previous claim and indicated this mutant capable of enduring the redox cofactor disturbance, making it an unsuitable host for functional screening. To circumvent this, I acquired another *E. coli* mutant (Δ*sthA* Δ*pfkA* Δ*pfkB* Δ*qor* Δ*edd*) reported to be

leathal on glucose in a prior study. Indeed, *E. coli* (Δ*sthA* Δ*pfkA* Δ*pfkB* Δ*qor* Δ*edd*) showed no sign of growth two weeks after being inoculated in glucose minimal media. Overexpression of SthA that conferred transhydrogenase activity restored its viability by just 3.2% after two-week incubation. This result suggested that, although SthA overexpression improved the survival of *E. coli* (Δ*sthA* Δ*pfkA* Δ*pfkB* Δ*qor* Δ*edd*), this strain might exhibit physiological defects other than the imbalance of redox cofactors. On the other hand, as pointed out by prior work, long-term incubation of *E. coli* (Δ*sthA* Δ*pfkA* Δ*pfkB* Δ*qor* Δ*edd*) on glucose over three weeks led to the emergence of revertants due to spontaneous mutations in the *nuoF* gene (encodes the F chain of NADH:ubiquinone oxidoreductase). To block this mutation path, I disrupted the *nuoF* gene to generate *E. coli* (Δ*sthA* Δ*pfkA* Δ*pfkB* Δ*qor* Δ*edd* Δ*nuoF*). Surprisingly, deletion of the *nuoF* gene made the resulting strain capable of growing on glucose again, suggesting this mutation-blocking strategy unfeasible to prevent spontaneous revertants.

Given that there would still be an one-week time window to distinguish restored growth of *E. coli* (Δ*sthA* Δ*pfkA* Δ*pfkB* Δ*qor* Δ*edd*) due to extraneous transhydrogenase activity and spontaneous revertants due to *nuoF* mutations, I decided to take this chance by overexpressing Gor, Lpd, RclA, and MerA in this *E. coli* mutant and examine if these FDRs would restore its viability as SthA did. After incubation of *E. coli* (Δ*sthA* Δ*pfkA* Δ*pfkB* Δ*qor* Δ*edd*) expressing these FDRs over three weeks, none of the tested FDRs showed an effect to improve growth of this *E. coli* mutant on glucose. This led to the hypothesis that in *E. coli* (Δ*sthA* Δ*pfkA* Δ*pfkB* Δ*qor* Δ*edd*) these FDRs might not be expressed as efficiently as in *E. coli* WT. The inefficient protein production of FDRs in this mutant was confirmed by quantifying its protein expression in the nutrient-rich medium. To resolve the poor protein yield, I turned my attention to manipulate the ribosome binding site (RBS) that drives protein translation of these FDRs. Using a yellow fluorescent protein as a reporter, I developed a high-throughput method to characterize the translation potential of large RBS libraries (Figure 2) by fluorescence-activated cell sorting (FACS) coupled with next generation sequencing. This approach allowed me to investigate the translational potential of a quater of million of RBS variants and identify those that confer efficient protein expression. Future goals would be to replace the current RBS of FDRs by these efficient variants and test their capacity to boost FDRs expression in *E. coli* (Δ*sthA* Δ*pfkA* Δ*pfkB* Δ*qor* Δ*edd*).

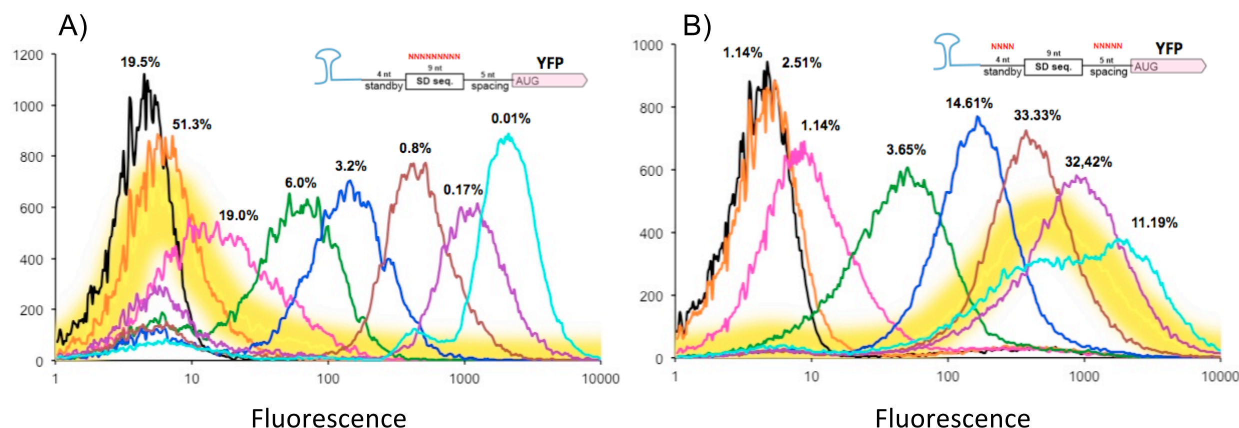

Figure 2. The distribution of translational potentials of RBS variants. Two RBS libraries were generated by randomizing either the Shine-Dalgarno sequence (A) or its surrounding 9 bp region (B). In both panels, the distribution of the whole library in terms of driving the expression of a yellow fluorescent protein is shown as the yellow halo line. Each of the two libraries was then subjected to fluorescence-activated cell sorting to collect 8 pools of RBS variants with different yellow fluorescent intensity (from weak to strong: black, orange, magenta, green, blue, brown, purple, cyan).