Directed Evolution of a Designer Enzyme Featuring an Unnatural Catalytic Amino Acid

Clemens Mayer,* Christopher Dulson, Eswar Reddem, Andy-Mark W. H. Thunnissen, and Gerard Roelfes*

Abstract: The impressive rate accelerations that enzymes display in nature often result from boosting the inherent catalytic activities of side chains by their precise positioning inside a protein binding pocket. Such fine-tuning is also possible for catalytic unnatural amino acids. Specifically, the directed evolution of a recently described designer enzyme, which utilizes an aniline side chain to promote a model hydrazone formation reaction, is reported. Consecutive rounds of directed evolution identified several mutations in the promiscuous binding pocket, in which the unnatural amino acid is embedded in the starting catalyst. When combined, these mutations boost the turnover frequency (kcat) of the designer enzyme by almost 100-fold. This results from strengthening the catalytic contribution of the unnatural amino acid, as the engineered designer enzymes outperform variants, in which the aniline side chain is replaced with a catalytically inactive tyrosine residue, by more than 200-fold.

The enviable rates and selectivities with which enzymes catalyze their transformations in nature have fueled efforts to create designer enzymes that can promote new-to-nature reactions with comparable efficiencies. Toward this goal, several approaches for enzyme design have been developed over the past decades; those include de novo design, (computationally aided) protein redesign, or the recruitment of (un)natural cofactors to appropriate binding pockets. Irrespective of the approach, the catalytic activities of the resulting designer enzymes pale in comparison to those found in nature. However, one important prospect of installing abiogenic activities into proteinaceous scaffolds is the ability to boost low starting activities by mimicking the Darwinian algorithm in the laboratory. Indeed, the iterative cycle of 1) introducing diversity through mutations, 2) identifying improved catalysts, and 3) amplifying more efficient enzyme variants, collectively referred to as directed evolution, has given rise to engineered designer enzymes that display rate accelerations akin to those found in nature.

In this report, we demonstrate that directed evolution is also a means to boost the proficiency of a novel class of designer enzymes, those that feature an unnatural amino acid as a catalytic residue. Such designer catalysts mimic natural enzymes that employ posttranslational modifications of active site residues to install uniquely reactive functionalities to promote their target reactions (i.e., formylglycine in type-I-sulfatases). Designer enzymes that make use of catalytic unnatural amino acids are also distinct from protein engineering efforts, in which genetic-code expansion strategies have been used to install non-standard side chains to improve/alter hydrophobic packing or substrate recognition.

The design and characterization of a designer enzyme, which utilizes a uniquely reactive p-aminophenylalanine (pAF; Figure 1A) residue to promote abiological hydrazide and oxime formation reactions, was recently reported. More specifically, introducing pAF at position 15 in the multidrug resistance regulator from Lactococcus lactis (LmrR), resulted in LmrR_pAF (previously assigned as LmrR_V15pAF), which promotes the condensation reactions of aldehydes with hydrazines or hydroxylamines through the formation of an iminium ion intermediate with the unnatural side chain (Figure 1B). While the designer enzyme outperformed aniline in solution by a factor of approximately 560, the catalytic contribution of the unnatural side chain remained modest. In fact, an LmrR_pAF variant, which features a structurally similar, but catalytically inactive, tyrosine residue instead of pAF (LmrR_Y), was only about 10-times less efficient than the parent designer enzyme. Given the power of the evolutionary algorithm, we surmised that the role pAF for catalysis could be optimized by identifying beneficial mutations of residues surrounding the unnatural side chain.

Since our initial report of LmrR_pAF, we were able to elucidate its structure (Figure 1C; see the Supporting Information for details). As anticipated, the uniquely reactive aniline side chains are embedded in the hydrophobic pore at the homodimer interface of LmrR_pAF. Moreover, pAF residues flank the central tryptophans (W96s), which are key to recruiting planar, aromatic (substrate) molecules. Upon inspection of this rudimentary active site, 13 additional residues were identified that line the hydrophobic pore and, as a result, are in proximity to the aniline side chain...
We anticipated that targeting these residues by site-directed mutagenesis could give rise to more proficient LmrR_pAF variants.

In order to rapidly identify beneficial mutations in the LmrR_pAF binding pocket, we aimed to establish a medium-throughput screening assay that would allow for the parallel evaluation of LmrR_pAF variants. For this, production of LmrR_pAF and a catalytically-inactive control protein (RamR)\cite{29} in 96-well format was optimized and cleared lysates were prepared. The incorporation of pAF was achieved by first introducing p-azidophenylalanine in response to a stop codon at position 15 (using the helper plasmid pEVOL-pAzF)\cite{30} and subsequently reducing the azido group with tris(2-carboxyethyl)phosphine in cell lysates (see the Supporting Information). To assess the catalytic activity of the resulting LmrR_pAF in this complex mixture, we took advantage of the chromogenic hydrazone formation reaction between 4-hydrazino-7-nitro-2,1,3-benzoxadiazole (NBD-H) and 4-hydroxybenzaldehyde (4-HBA, Figure 2A)\cite{31}. These two substrates (final concentrations: 50 μM NBD-H and 5 mM 4-HBA) were added to the cleared lysates and product formation was followed at 472 nm in a 96-well plate reader for 2 hours. Indeed, lysates containing LmrR_pAF displayed 30% higher product-formation rates than those containing RamR (Supporting Information, Figure S1). Moreover, consistent with its lower activity, LmrR_Y (the variant containing tyrosine instead of pAF) did not provide significant rate acceleration with respect to RamR (Supporting Information, Figure S1). Taken together, these results suggest that it should be feasible to accurately assess the activity of LmrR_pAF variants in parallel.

Figure 1. A) Chemical structure of p-aminophenylalanine. B) Formation of an iminium ion intermediate in the presence of anilines accelerates hydrazone (X=NH) and oxime (X=O) formation reactions (for clarity, the reversibility of these reactions is not shown). C) Crystal structure of the LmrR_pAF homodimer (PDB: 6I8N). Catalytic aniline side chains (red) and Trp96 (pink) are shown as sticks. The positions of the β-carbons of 13 additional residues that line the binding pocket of LmrR_pAF are shown as spheres. A 3-(N-morpholino)propanesulfonic acid (MOPS) buffer molecule that was found to be sandwiched between the central tryptophans is omitted for clarity (see Supporting Information).

Figure 2. A) Reaction conditions for the model hydrazone formation between NBD-H and 4-HBA in cleared lysates. B) Close-up of the hydrophobic pore in LmrR_pAF. Trp96s and pAF15s shown as sticks; β-carbons of positions that gave rise to improved variants in round one and two are shown as spheres (color code as indicated). C) Evolutionary optimization of LmrR_pAF. Apparent catalytic efficiencies (kcat/Km)app of selected variants and their improvement with respect to LmrR_pAF (see Table S1 and Figure S3 in the Supporting Information for a more detailed analysis); errors are standard deviations of at least three independent experiments; color code as in Figure 2B. D) Comparison of saturation kinetics at a 4-HBA concentration of 5 mM for LmrR_pAF and the best variants obtained after two rounds of directed evolution. Errors are standard deviations of at least three independent experiments.
To identify beneficial mutations, eight residues were targeted (Supporting Information, Figure S2), which are either in close proximity to pAF15 or have previously been identified to improve catalytic parameters of unrelated, LmrR-based designer enzymes.\textsuperscript{[23–34]} Libraries targeting each position were constructed using degenerate primers (NNK codons, which allow for all 20 canonical amino acids) and approximately 400 library members were evaluated using the previously established screen. Two variants LmrR\textsubscript{pAF}\_A92R and LmrR\textsubscript{pAF}\_L18K, which gave large improvements of 2.46 ± 0.13- and 2.04 ± 0.11-times, respectively, when compared to \textit{LmrR\_pAF}, were subsequently produced and purified (Supporting Information, Figure S2). Upon determining their catalytic parameters, both variants displayed improved apparent catalytic efficiencies ((\(k_{\text{cat}}/K_{\text{m}}\))app), when compared to the parent designer enzyme (2.2-times for A92R and 1.5-times for L18K, Figure 2A and Supporting Information, Figure S3 and Table S1). Combined, these results attest that reproducible gains in the screen translate into improved catalytic parameters for purified LmrR\textsubscript{pAF} variants.

Encouraged by these results, LmrR\textsubscript{pAF}\_R (containing the A92R mutation) was chosen as the template for a new round of mutagenesis. In total, 12 additional positions were randomized and 84 library members per targeted position screened (approximately 1000 variants in total; Supporting Information, Figure S2). Another three mutations (A11L, N19M, and F93H, but not L18K) were identified, which independently led to significant rate accelerations with respect to \textit{LmrR\_pAF}\_R (Figure 2B,C). To identify potential synergistic effects between these three mutations, LmrR\textsubscript{pAF}\_R variants were constructed that combined two or all three of these mutations. Strikingly, the combination of N19M and F93H (LmrR\textsubscript{pAF}\_RMH from here onward) provided an exceptionally large improvement compared to either of the other variants tested (Figure 2CD and Supporting Information, Figure S3 and Table S1). In fact, when comparing the apparent catalytic efficiencies, LmrR\textsubscript{pAF}\_RMH was 57-times more efficient than the parent designer enzyme (Figure 2BC). To identify potential synergistic effects between these three mutations, LmrR\textsubscript{pAF}\_R variants were constructed that combined two or all three of these mutations. Strikingly, the combination of N19M and F93H (LmrR\textsubscript{pAF}\_RMH from here onward) provided an exceptionally large improvement compared to either of the other variants tested (Figure 2CD and Supporting Information, Figure S3 and Table S1). In fact, when comparing the apparent catalytic efficiencies, LmrR\textsubscript{pAF}\_RMH was 57-times more efficient than the parent designer enzyme (Figure 2BC). To identify potential synergistic effects between these three mutations, LmrR\textsubscript{pAF}\_R variants were constructed that combined two or all three of these mutations. Strikingly, the combination of N19M and F93H (LmrR\textsubscript{pAF}\_RMH from here onward) provided an exceptionally large improvement compared to either of the other variants tested (Figure 2CD and Supporting Information, Figure S3 and Table S1). In fact, when comparing the apparent catalytic efficiencies, LmrR\textsubscript{pAF}\_RMH was 57-times more efficient than the parent designer enzyme (Figure 2BC). To identify potential synergistic effects between these three mutations, LmrR\textsubscript{pAF}\_R variants were constructed that combined two or all three of these mutations. Strikingly, the combination of N19M and F93H (LmrR\textsubscript{pAF}\_RMH from here onward) provided an exceptionally large improvement compared to either of the other variants tested (Figure 2CD and Supporting Information, Figure S3 and Table S1). In fact, when comparing the apparent catalytic efficiencies, LmrR\textsubscript{pAF}\_RMH was 57-times more efficient than the parent designer enzyme (Figure 2BC). To identify potential synergistic effects between these three mutations, LmrR\textsubscript{pAF}\_R variants were constructed that combined two or all three of these mutations. Strikingly, the combination of N19M and F93H (LmrR\textsubscript{pAF}\_RMH from here onward) provided an exceptionally large improvement compared to either of the other variants tested (Figure 2CD and Supporting Information, Figure S3 and Table S1). In fact, when comparing the apparent catalytic efficiencies, LmrR\textsubscript{pAF}\_RMH was 57-times more efficient than the parent designer enzyme (Figure 2BC). To identify potential synergistic effects between these three mutations, LmrR\textsubscript{pAF}\_R variants were constructed that combined two or all three of these mutations. Strikingly, the combination of N19M and F93H (LmrR\textsubscript{pAF}\_RMH from here onward) provided an exceptionally large improvement compared to either of the other variants tested (Figure 2CD and Supporting Information, Figure S3 and Table S1). In fact, when comparing the apparent catalytic efficiencies, LmrR\textsubscript{pAF}\_RMH was 57-times more efficient than the parent designer enzyme (Figure 2BC). To identify potential synergistic effects between these three mutations, LmrR\textsubscript{pAF}\_R variants were constructed that combined two or all three of these mutations. Strikingly, the combination of N19M and F93H (LmrR\textsubscript{pAF}\_RMH from here onward) provided an exceptionally large improvement compared to either of the other variants tested (Figure 2CD and Supporting Information, Figure S3 and Table S1). In fact, when comparing the apparent catalytic efficiencies, LmrR\textsubscript{pAF}\_RMH was 57-times more efficient than the parent designer enzyme (Figure 2BC). To identify potential synergistic effects between these three mutations, LmrR\textsubscript{pAF}\_R variants were constructed that combined two or all three of these mutations. Strikingly, the combination of N19M and F93H (LmrR\textsubscript{pAF}\_RMH from here onward) provided an exceptionally large improvement compared to either of the other variants tested (Figure 2CD and Supporting Information, Figure S3 and Table S1). In fact, when comparing the apparent catalytic efficiencies, LmrR\textsubscript{pAF}\_RMH was 57-times more efficient than the parent designer enzyme (Figure 2BC). To identify potential synergistic effects between these three mutations, LmrR\textsubscript{pAF}\_R variants were constructed that combined two or all three of these mutations. Strikingly, the combination of N19M and F93H (LmrR\textsubscript{pAF}\_RMH from here onward) provided an exceptionally large improvement compared to either of the other variants tested (Figure 2CD and Supporting Information, Figure S3 and Table S1). In fact, when comparing the apparent catalytic efficiencies, LmrR\textsubscript{pAF}\_RMH was 57-times more efficient than the parent designer enzyme (Figure 2BC). To identify potential synergistic effects between these three mutations, LmrR\textsubscript{pAF}\_R variants were constructed that combined two or all three of these mutations. Strikingly, the combination of N19M and F93H (LmrR\textsubscript{pAF}\_RMH from here onward) provided an exceptionally large improvement compared to either of the other variants tested (Figure 2CD and Supporting Information, Figure S3 and Table S1). In fact, when comparing the apparent catalytic efficiencies, LmrR\textsubscript{pAF}\_RMH was 57-times more efficient than the parent designer enzyme (Figure 2BC). To identify potential synergistic effects between these three mutations, LmrR\textsubscript{pAF}\_R variants were constructed that combined two or all three of these mutations. Strikingly, the combination of N19M and F93H (LmrR\textsubscript{pAF}\_RMH from here onward) provided an exceptionally large improvement compared to either of the other variants tested (Figure 2CD and Supporting Information, Figure S3 and Table S1). In fact, when comparing the apparent catalytic efficiencies, LmrR\textsubscript{pAF}\_RMH was 57-times more efficient than the parent designer enzyme (Figure 2BC). To identify potential synergistic effects between these three mutations, LmrR\textsubscript{pAF}\_R variants were constructed that combined two or all three of these mutations. Strikingly, the combination of N19M and F93H (LmrR\textsubscript{pAF}\_RMH from here onward) provided an exceptionally large improvement compared to either of the other variants tested (Figure 2CD and Supporting Information, Figure S3 and Table S1). In fact, when comparing the apparent catalytic efficiencies, LmrR\textsubscript{pAF}\_RMH was 57-times more efficient than the parent designer enzyme (Figure 2BC). To identify potential synergistic effects between these three mutations, LmrR\textsubscript{pAF}\_R variants were constructed that combined two or all three of these mutations. Strikingly, the combination of N19M and F93H (LmrR\textsubscript{pAF}\_RMH from here onward) provided an exceptionally large improvement compared to either of the other variants tested (Figure 2CD and Supporting Information, Figure S3 and Table S1). In fact, when comparing the apparent catalytic efficiencies, LmrR\textsubscript{pAF}\_RMH was 57-times more efficient than the parent designer enzyme (Figure 2BC). To identify potential synergistic effects between these three mutations, LmrR\textsubscript{pAF}\_R variants were constructed that combined two or all three of these mutations. Strikingly, the combination of N19M and F93H (LmrR\textsubscript{pAF}\_RMH from here onward) provided an exceptionally large improvement compared to either of the other variants tested (Figure 2CD and Supporting Information, Figure S3 and Table S1). In fact, when comparing the apparent catalytic efficiencies, LmrR\textsubscript{pAF}\_RMH was 57-times more efficient than the parent designer enzyme (Figure 2BC). To identify potential synergistic effects between these three mutations, LmrR\textsubscript{pAF}\_R variants were constructed that combined two or all three of these mutations. Strikingly, the combination of N19M and F93H (LmrR\textsubscript{pAF}\_RMH from here onward) provided an exceptionally large improvement compared to either of the other variants tested (Figure 2CD and Supporting Information, Figure S3 and Table S1). In fact, when comparing the apparent catalytic efficiencies, LmrR\textsubscript{pAF}\_RMH was 57-times more efficient than the parent designer enzyme (Figure 2BC).
results indicate that the directed evolution of unnatural-amino-acid-containing designer enzymes can generate highly active, as well as versatile, catalysts.

In the absence of available structural information, it is difficult to rationalize how the identified mutations boost the catalytic potential of the aniline side chain. Both A11L and N19M are located one helical turn away from pAF15 (Figure 2B) and are likely involved in positioning the unnatural side chain in a productive conformation. A92R and F93H are located opposite of pAF15. The introduction of a permanent positive charge by A92R is somewhat puzzling, as it could negatively impact the population of the crucial iminium ion (Figure 1A), due to charge–charge repulsion.\[8\]

However, in the formation of this covalent, transient intermediate, negative species are formed, which in turn could be stabilized by the guanidinium side chain.\[3,8\] The F93H mutation, in combination with N19M, appears critical for the boosts observed in LmrR_pAF_RMHL and LmrR_pAF_RMHL, as demonstrated by the drastic loss of activity in the respective F93 or N19 reversion variants (Figure 2C and Supporting Information, Table S1). It is, therefore, tempting to suggest that histidine aides in the formation of the iminium ion and/or transamination through proton shuffling or positioning of ordered water molecules.\[3,8,31,36,37]\n
In conclusion, our work demonstrates that designer enzymes featuring an unnatural amino acid as a catalytic residue can be privileged starting points for directed evolution campaigns. Enhancing the inherent catalytic activity of an unnatural side chain is feasible by identifying beneficial mutations in a protein scaffold. Future efforts will focus on the structural and computational analysis of tailored LmrR_pAF variants in order to pinpoint the exact mechanisms by which these engineered designer enzymes can display an almost 100-times higher turnover frequency when compared to the parent variant. Lastly, we surmise that the impressive improvements observed for LmrR_pAF are not limited to an aniline side chain. Instead, we suggest that the introduction and fine-tuning of other organocatalysts,\[38–40\] which are versatile yet notoriously slow, through genetic-code expansion will result in proficient designer enzymes for a wide variety of new-to-nature reactions. Ultimately, such efforts could provide a promising route for developing efficient protein catalysts for synthetically relevant transformations.

Acknowledgements

The authors thank Dr. I. Drienovská for helpful advice throughout the project. This work was supported by the European Research Council (ERC starting grant no. 280010) and the Netherlands Organisation for Scientific Research (NWO, Vici grant 724.013.003, and Veni grant 722.017.007). G.R. acknowledges support from the Ministry of Education Culture and Science (Gravitation programme no. 024.001.035). C.M. acknowledges a Marie Sklodowska Curie Individual Fellowship (project no. 75159). We acknowledge DESY (Hamburg, Germany), a member of the Helmholtz Association HGF, for the provision of experimental facilities. Parts of this research were carried out at PETRA III and we would like to thank beamline staff for assistance in using beamline P11.

Conflict of interest

The authors declare no conflict of interest.

Keywords: directed evolution · enzyme catalysis · enzyme design · hydrazones · organocatalysis

How to cite: Angew. Chem. Int. Ed. 2019, 58, 2083–2087
Angew. Chem. 2019, 131, 2105–2109
