A designer enzyme for hydrazone and oxime formation featuring an unnatural catalytic aniline residue

Drienovská, Ivana; Mayer, Clemens; Dulson, Christopher; Roelfes, Gerard

Published in:
Nature Chemistry

DOI:
10.1038/s41557-018-0082-z

Publication date:
2018

A designer enzyme for hydrazone and oxime formation featuring an unnatural catalytic aniline residue

Ivana Drienovská, Clemens Mayer, Christopher Dulson and Gerard Roelfes*

Stratingh Institute for Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

email: j.g.roelfes@rug.nl,

Abstract

Creating designer enzymes with the ability to catalyze abiological transformations is a formidable challenge. Efforts toward this goal typically consider only canonical amino acids in the initial design process. However, incorporating unnatural amino acids that feature uniquely reactive side chains could significantly expand the catalytic repertoire of designer enzymes. To explore the potential of such artificial building blocks for enzyme design, we selected *p*-aminophenylalanine (pAF) as a potentially novel catalytic residue. Here, we demonstrate that the catalytic activity of the aniline side chain for hydrazone and oxime formation reactions is increased by embedding pAF into the hydrophobic pore of the multidrug transcriptional regulator from *Lactococcus lactis*. Both the recruitment of reactants by the promiscuous binding pocket and a judiciously placed aniline that functions as a catalytic residue contribute to the success of the identified artificial enzyme. We anticipate that our design strategy will prove rewarding to significantly expand the catalytic repertoire of designer enzymes in the future.

Main Text

The prospects of harnessing the catalytic prowess of enzyme catalysis for chemical synthesis have fueled efforts to create made-to-order enzymes with the ability to catalyze abiological transformations.[1,2] Toward this end, diverse engineering strategies have been employed to design artificial enzymes with novel activities and specificities. These include enzyme design from first principles,[3,4] raising catalytic antibodies from the immune system,[5] exploiting computational algorithms[6] or chemical intuition[7] for protein redesign, and introducing (synthetic) transition metal cofactors into biomacromolecules.[8] While these strategies differ in how to create designer enzymes, they typically only consider canonical amino acids and cofactors in the initial stages of the design process.

Notably, enzymes found in nature routinely perform reactions that cannot be catalyzed by the limited number of functionalities present in canonical amino acid side chains. To overcome these limitations, enzymes do not only recruit reactive cofactors but also post-translationally modify existing amino acids in a given active site (**Fig. 1a**).[9] Examples for the latter include the use of an N-terminal pyruvoyl group to facilitate the decarboxylation of histidine,[10] the introduction of formylglycine as a catalytic residue in type I sulfatases,[11] or the formation of 4-methylideneimidazole-5-one as an electrophilic catalyst in ammonia lyases and aminomutases.[12]

In principle, mimicking such modified side chains by incorporating nonstandard amino acids into proteins could significantly expand the catalytic repertoire of designer enzymes.[13] To date more than 150 unnatural amino acids have been successfully incorporated by robust *in vivo* stop codon suppression strategies.[14,15] These non-canonical building blocks feature a wide range of side chains with abiological functional groups and have been utilized as handles for protein modification, imaging or spectroscopy, metal chelators, as well as redox mediators.[16] More recently, unnatural amino acids have also been installed into existing enzyme active sites to probe

or improve catalysis.[13] A topical example by the Hilvert group employs N_δ-methyl histidine as a proximal heme ligand in ascorbate peroxidase to lock the side chain into an active conformation, thus improving catalytic parameters.[17] In contrast, utilizing unnatural amino acids for designing enzymes with novel activities has remained unexplored. Toward this goal, our group has demonstrated that by incorporating the metal chelating amino acid 2,2-bipyridine, catalytically-active transition metals can be recruited to a promiscuous binding pocket, thereby installing new reactivities in protein scaffolds.[18,19] Notably though, in this and similar examples the unnatural amino acid does not actively participate in catalysis, but rather positions the active species in a well-defined environment (**Fig. 1b**).[20]

To explore the potential of unnatural amino acids to act as novel, catalytic residue in designer enzymes we selected *p*-aminophenylalanine (pAF, **Fig. 1c**).[21] While pAF has previously been incorporated into proteins as a handle for chemoselective bioconjugations,[22] the well-known ability of the aniline side chain to act as a nucleophilic catalyst has never been explored for enzyme design.[23] Indeed, over the past decade anilines have been demonstrated to catalyze the formation of hydrazones and oximes by condensation of carbonyls with hydrazines and hydroxylamines (**Fig. 1d**).[24,25] Rate accelerations in presence of millimolar concentrations of aniline can be rationalized by formation of a protonated Schiff base intermediate that facilitates transimination. Modulating the catalytic activity of anilines was achieved by perturbing the pK_a of the aromatic amine or placing an ionizable group in its close proximity.[26–28] As such, this type of catalysis is reminiscent to the mode of action by which natural enzymes tailor the reactivity of functional groups, and we hypothesized that the catalytic activity of anilines could be fine-tuned by introducing pAF into an appropriate protein environment.

Here, we report on the creation of such an artificial enzyme by embedding pAF into the promiscuous binding pocket of the multidrug transcriptional regulator from *Lactococcus lactis*,

LmrR. Strikingly, the resulting designer enzyme outperforms aniline and more reactive aniline derivatives by more than two orders of magnitude in model hydrazone and oxime formation reactions. This catalytic proficiency is achieved by the combination of a uniquely reactive pAF residue and the hydrophobic interactions provided by the proteinaceous scaffold.

Results

Design, synthesis and characterization of pAF-containing LmrR variants

LmrR is a small, homodimeric protein of about 15 kDa that contains a large hydrophobic pore at its dimer interface and displays remarkable promiscuity for binding hydrophobic

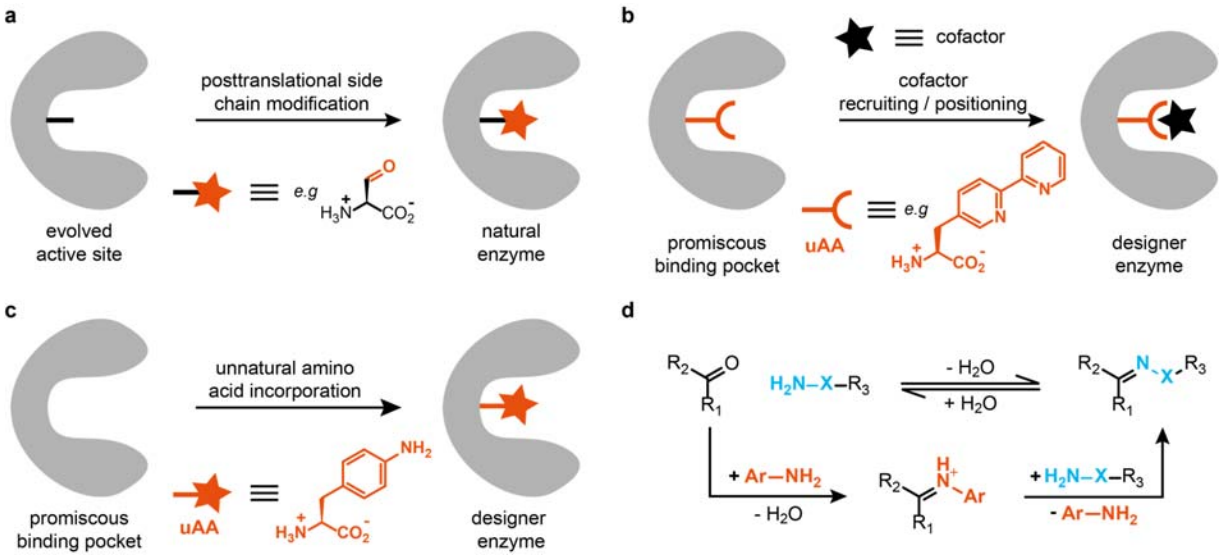

Figure 1: Unnatural amino acids as a means to expand the scope of enzyme catalysis. a: Posttranslational side chain modifications of canonical amino acids in the active sites of natural enzyme can introduce new activities into a given protein scaffold. Formylglycine, a catalytic residue in type-I sulfatases, is shown as a representative example. **b:** Catalytically active transition metal (complexes) can be recruited to promiscuous binding pockets by introduction of a metal-binding, unnatural amino acid (uAA). For example, binding of Cu^{2+} to a genetically-incorporated 2,2'-bipyridylalanine (depicted in red) resulted in the formation of enzymes with abiological activities. **c:** When placed in a promiscuous binding pocket, genetic incorporation of an unnatural amino acid featuring a side chain with unique reactivity could expand the reaction scope of designer enzymes. As a proof-of-concept that unnatural amino acids can act as catalytic residues, we chose the amino acid *p*-aminophenylalanine that features a nucleophilic aniline side chain. **d:** Anilines are well-known nucleophilic catalysts for hydrazone (X=NH) and oxime (X=O) formations. Primary aromatic amines accelerate these reactions by formation of a protonated Schiff base intermediate, which subsequently undergoes transimination to yield the desired products (for clarity, the reversibility of these reactions is not shown).

molecules such as drugs or antibiotics (**Fig. 2a**).[29,30] This feature makes LmrR an attractive scaffold for introducing novel activities into a protein scaffold, as its promiscuous binding pocket has been shown to provide a suitable environment for planar molecules to bind and undergo catalysis.[18,31,32] For the incorporation of pAF we selected four different positions (V15, N19, M89 and F93) that line the anticipated binding pocket. N19 and M89 are localized at the edges of the pore, while position V15 and F93 are located closer to the center (**Fig. 2b**). With the exception of F93 that points toward the solvent, all side chains face the inside of the pore. The incorporation of pAF into proteins has previously been achieved with the pDule_pAF plasmid.[21] However, for LmrR variants, attempts to directly incorporate pAF led to low expression yields and resulted in the misincorporation of tyrosine and phenylalanine. Instead, proteins featuring pAF at the indicated positions were prepared by first introducing *p*-azidophenylalanine (pAzF, pEVOL-pAzF)[33], and subsequently reducing the azido group with tris(2-carboxyethyl)phosphine (TCEP, see **Materials and Methods** for details). This strategy afforded pAF-containing variants after affinity chromatography and the reduction step in 9-14 mg protein per liter culture, which corresponds to 45-70% yield when compared to LmrR.[32] High resolution mass spectrometry confirmed the successful introduction of pAzF and its complete reduction to pAF upon treatment with TCEP **(Supplementary Fig. 1**). When performing size-exclusion chromatography all proteins eluted as single peak at an elution volume of 11.6 (± 0.2) ml (corresponding to a molecular weight around 30 kDa), indicating that the introduction of the unnatural amino acid does not disrupt homodimer formation (**Supplementary Fig. 2**).

Evaluating the catalytic potential of pAF-containing LmrR variants for nucleophilic catalysis

The catalytic properties of the novel artificial enzymes were initially examined in the chromogenic hydrazone formation reaction between 4-nitrobenzaldehyde (4-NBA) and 4-hydrazino-7-nitro-

2,1,3-benzoxadiazole (NBD-H, **Fig. 2c**). NBD-H undergoes a distinct red shift upon condensation with 4-NBA and product formation can conveniently be monitored at 504 nm.[26] While the uncatalyzed reaction proceeds sluggishly to give 6 ± 3% yields after 2 hours, in presence of 1 mM aniline or pAF yields are boosted to around 40% in the same time (**Fig. 2d** and **Supplementary Table 1**). Strikingly, LmrR provided similar levels of conversion (46 ± 6%) at a dimer concentration of 10 µM, corresponding to a 100 times lower catalyst loading. This previously unreported catalytic activity of the wild type protein is consistent with our design, in which we surmised that both substrates would be able to bind to the hydrophobic pocket of

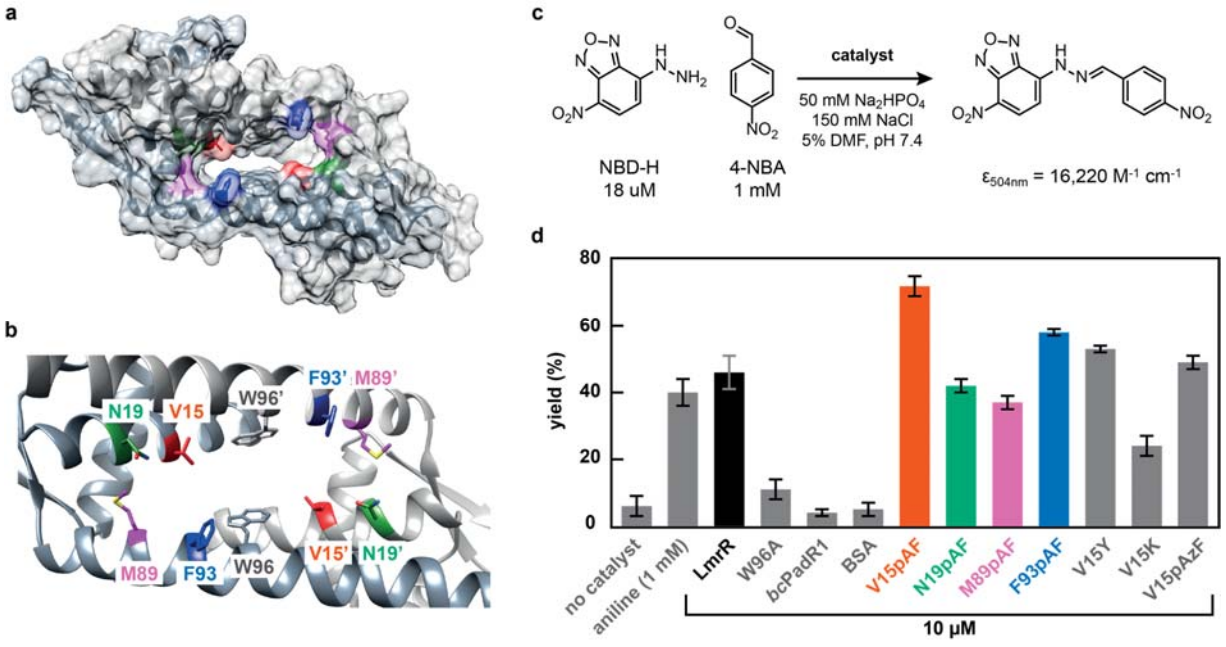

Figure 2: Both the promiscuous binding pocket of LmrR and an accurately placed aniline side chain contribute to the catalytic activity of pAF-containing designer enzymes. a–b: Surface view of the dimeric LmrR and a close-up of the hydrophobic pore. In both representations, positions (from the two monomeric chains) that were used for pAF introduction in this study are highlighted in color. W96 and W96' are the central tryptophans, which provide π-π interactions to recruit aromatic compounds. **c:** Reaction conditions for the chromogenic, model hydrazone formation between NBD-H and 4-NBA. The progress of the reaction was continuously monitored for 120 minutes by following product formation at 504 nm. **d:** Bar graph displaying reaction yields in presence of different catalysts. At a 100-fold lower catalyst concentration (dimer concentration), LmrR gives rise to comparable levels of activity as aniline. Further improvements are achieved when pAF is introduced at positions V15 or F93. Conversions represent the average of at least 3 independent measurements with standard deviations displayed. The values plotted here are listed in **Supplementary Table 1**.

LmrR. Such recruiting of reactants will increase their "effective molarity", which is a means to accelerate the bimolecular hydrazone formation. To find further support that catalysis occurs inside the hydrophobic pore, the reaction was performed with LmrR_W96A. Mutation of this tryptophan, which is critical for recruiting planar aromatic molecules to the hydrophobic pocket, drastically decreased yields (11 ± 3%, **Fig. 2d**) to levels that are similar those obtained for the uncatalyzed reaction. Similarly, *bc*PadR1, an LmrR homolog with 26% of sequence identity that does not feature a pore at its dimer interface,[34] did not show rate accelerations beyond the background reaction. Lastly, the observed rate accelerations seem specific to LmrR and its large hydrophobic cavity. BSA, a protein with a hydrophobic binding site, which contains two lysines known to act as nucleophiles in a number of reactions,[35] also did not catalyze the model hydrazone formation (**Fig. 2d**).

Next, proteins containing pAF as a potential catalytic residue were tested. Notably, at a dimer concentration of 10 μM all variants significantly accelerated the reaction (**Supplementary Table 1**), with LmrR_V15pAF performing best (72 ± 3% yield) and LmrR_M89pAF giving the lowest yields (37 ± 2%, **Fig. 2d**). These results suggest that potential gains of introducing pAF as catalytic residue depend on appropriately placing the unnatural amino acid inside the hydrophobic pore. The lower yields observed with M89pAF presumably reflect that either this position is less accessible when compared to V15pAF or that proximity to the central tryptophans is important for catalysis. To gain further insight into the role of pAF at position 15 in catalysis, we prepared variants that contain a lysine or tyrosine in place of the unnatural amino acids. Both mutants performed significantly worse; V15K gave rise to the hydrazone in 24 ± 3% yield, while V15Y displayed levels comparable to the parent LmrR (53 ± 1% yield, **Fig. 2d**). Further support for an active role of pAF in catalysis is provided by the fact that yields prior to reduction – that is pAzF

is present at position 15 – are significantly lower (49 ± 2%) when compared to the fully reduced protein (72 ± 3% yield, **Fig. 2d**). This trend, although less pronounced, is also apparent for the other pAF-containing proteins (**Supplementary Table 1**).

In the proposed catalytic cycle, anilines accelerate hydrazone formation reactions through the formation of an iminium ion upon condensation with an aldehyde (**Fig. 1d**). To trap this intermediate, we first incubated LmrR and pAF-containing variants with 4-NBA, before adding $NaCNBH_3$ to reduce the transiently forming Schiff base intermediates to amines, thereby forming an irreversible, covalent bond (see **Supplementary Information** for details). When subjecting these samples to high resolution mass spectrometry LmrR_V15pAF displayed a dominant peak corresponding to the singly modified monomer (**Fig. 3a**). Conversely, all other pAF-containing variants featured prominent peaks for both the modified and unmodified protein, indicative of the presence of a uniquely reactive aniline side chain in LmrR_V15pAF (**Supplementary Fig. 3**). The extent of double modification in LmrR_V15pAF was comparable to the amount of single modification observed for the parent LmrR and presumably reflects unspecific labeling of surface lysines (**Fig. 3a**). To pinpoint the position(s) of modification(s) we performed tryptic digests on labeled LmrR and LmrR_V15pAF followed by LC-MS/MS analysis of the resulting peptides (see **Supplementary Information** for details). While position 15 was exclusively identified as the predominant site of modification for the V15pAF variant, both LmrR and LmrR_V15pAF displayed a small extent of unspecific modifications on different lysine residues throughout the scaffold (**Supplementary Fig. 4**). To study whether these modifications affect catalysis, our model hydrazone formation was performed with LmrR and LmrR_V15pAF before and after the reduction. While for LmrR conversions were comparable (46% and 45%, respectively), LmrR_V15pAF displayed significantly higher yields when unmodified (72%) in comparison to

only 15% when modified, **Supplementary Table 1**). These results demonstrate that modifying the aniline side chain leads to inactivation, while unspecific modifications of different lysine residues throughout the LmrR scaffold have no effect on catalysis.

Kinetic characterization of LmrR_V15pAF

To further evaluate the catalytic contributions provided by the unnatural amino acid, we subjected LmrR_V15pAF to a more rigorous kinetic characterization. Unfortunately, the low

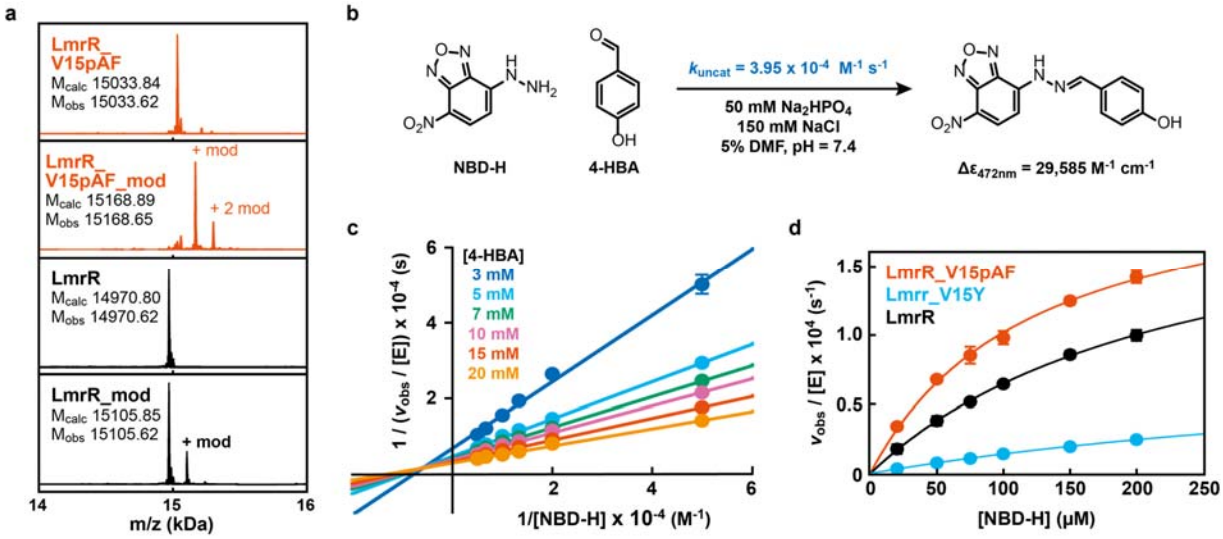

Figure 3: LmrR_V15pAF is a designer catalyst for hydrazone formation that features a uniquely reactive aniline side chain. a: Deconvoluted high resolution mass spectrometry results for LmrR and LmrR_V15pAF before and after trapping of the transiently formed iminium ion in presence of 4-NBA and $NaCNBH_3$. Consistent with the envisioned role of pAF as nucleophilic catalyst trapping is more efficient for LmrR_V15pAF than the parent protein. The unique reactivity of the aniline side chain at this position is further highlighted by the near quantitative conversion to modified species, which is not observed for any of the other pAF-containing variants (**Supplementary Fig. 3**). The extent of double modification in V15pAF is comparable to the amount of modified protein observed in LmrR; both are the result of iminium ion formation at different lysine side chains throughout the scaffold (**Supplementary Fig. 4** for MS/MS results after tryptic digest of both proteins). **b:** Reaction conditions for the hydrazone formation between NBD-H and the more water-soluble 4-HBA, which was employed for a detailed kinetic characterization of LmrR_V15pAF by following product formation at 472 nm. The uncatalyzed rate was determined by performing the reaction under pseudo first order conditions (see **Supplementary Information** for details). **c:** Double-reciprocal plot displaying the dependence of reaction velocity on NBD-H concentration at different 4-HBA concentrations. The pattern of intersecting lines left of the y-axis confirms a sequential reaction mechanism. **d:** Comparison between the initial reaction rates of LmrR_V15pAF, LmrR_V15Y and the parent protein at a 4-HBA concentration of 5 mM. Kinetic values derived from these plots can be found in **Table 1**.

solubility of 4-NBA and precipitation of the product at high substrate concentrations prevented a full characterization of the enzyme (see **Supplementary Fig. 5** for a saturation kinetic study performed at 250 μM 4-NBA). Instead, we replaced 4-NBA with the more water-soluble 4-hydroxybenzaldehyde (4-HBA), which also drastically improved solubility of the hydrazone product under our assay conditions (the reaction was followed at 472 nm, **Fig. 3b**). Under pseudo first order reaction conditions, the uncatalyzed hydrazone formation proceeds with a second order rate constant of 3.95 x 10^{-4} M^{-1} s^{-1}. The apparent third order rate constant in presence of aniline was determined to be 1.12 M^{-2} s^{-1}, corresponding to a 2.8-fold rate acceleration at 1 mM catalyst present.

LmrR_V15paF was characterized by measuring the dependence of the reaction velocity on NBD-H concentration (20–200 μM) at several fixed 4-HBA concentrations (3–20 mM). Independent binding of both substrates is indicated by the pattern of intersecting lines in a double-reciprocal plot (**Fig. 3c**), confirming a sequential mechanism for the reaction. Steady state parameters (**Supplementary Table 2**) could be derived by globally fitting the data to a random binding mechanism. We determined K_m values for NBD-H and 4-HBA to be 100 ± 13 μM and 7.92 ± 0.97 mM, respectively. The drastically lower K_m value for the former is consistent with a more efficient recruitment of this substrate by the central tryptophans (W96 and W96') in the LmrR binding pocket. Moreover, LmrR_V15pAF displays a k_{cat} of 5.0 ± 0.39 x 10^{-4} s^{-1} and an "effective molarity" (k_{cat}/k_{uncat}) of 1.3 ± 0.1 M. With an apparent third order rate constant ($k_{cat}/(K_{NBD\text{-}H}\ K_{4\text{-}HBA})$) of 6.3 ± 1.1 x 10^2 M^{-2} s^{-1} outperforms aniline by a factor >550. Similarly, LmrR_V15pAF outperforms more effective aniline derivatives, such as 5-methoxyanthrinilic acid or 3,5-diaminobenzoic acid, by more than two orders of magnitude (see **Supplementary Table 3** for details).[26] In terms of chemical proficiency ($1/K_{TS}$ = 1.6 ± 0.3 x 10^6 M^{-1}) LmrR_V15pAF compares favorably with other designer enzymes for bimolecular reactions – such as computationally

designed Diels Alderases or catalytic antibodies elicited for the same type of reaction[36–40] – and attests the effectiveness of our design approach (**Supplementary Table 2**).

Table 1: Catalytic parameters obtained from saturation kinetic studies for LmrR, LmrR_V15Y and LmrR_V15pAF.

	hydrazone formation*				NBD-ONH_2 hydrolysis				oxime formation†			
	k_{cat} ($x10^{-4}$ s^{-1})	K_{NBD-H} ($x10^{-6}$ M)	k_{cat} / K_{NBD-H} (M^{-1} s^{-1})	log(1/K_{TS})	k_{cat} ($x10^{-3}$ s^{-1})	$K_{NBD-ONH2}$ ($x10^{-6}$ M)	k_{cat} / $K_{NBD-ONH2}$ (M^{-1} s^{-1})	log(1/K_{TS})	k_{cat} ($x10^{-3}$ s^{-1})	$K_{NBD-ONH2}$ ($x10^{-6}$ M)	k_{cat} / $K_{NBD-ONH2}$ (M^{-1} s^{-1})	log(1/K_{TS})
V15pAF	2.26 (0.11)	122 (12)	1.85 (0.10)	6.20	5.70 (0.32)	181 (24)	31.4 (4.7)	6.26	51.4 (2.90)	146 (15)	352 (19)	6.73
V15Y	0.90 (0.05)	538 (49)	0.17 (0.01)	n.d	1.76 (0.21)	246 (49)	7.15 (0.73)	5.62	6.01 (0.41)	154 (19)	39.0 (2.50)	5.78
LmrR	2.20 (0.07)	239 (12)	0.92 (0.02)	n.d	1.80 (0.14)	230 (29)	7.84 (0.42)	5.66	0.78 (0.03)	81 (8)	9.61 (0.58)	5.18

Catalytic parameters were obtained by fitting the data points from plots **Fig. 3d**, **Fig. 4b** and **Fig 4c** to the Michaelis-Menten equation and are calculated per LmrR dimer. * hydrazone formation reactions were performed at a 4-HBA concentration of 5 mM. † oxime formation reactions were carried out in presence of 250 μM 4-NBA. ‡ In absence of an available $K_{4\text{-NBA}}$ apparent chemical proficiencies were obtained by substituting this term with the concentration of 4-NBA under the assay conditions (250 μM). As this concentration is well below the $K_{4\text{-HBA}}$ obtained from our studies for the hydrazone formation reaction, this estimate is likely to accurately reflect the true transition state stabilization. If the $K_{4\text{-HBA}}$ was significantly lower the apparent $1/K_{TS}$ would underestimate the true value. n.d. not determined.

A comparison of the pAF-containing protein to LmrR_V15Y and LmrR itself also revealed some noteworthy insight. At a fixed 4-HBA concentration of 5 mM, LmrR_V15Y displays only ~20% of the parent protein's activity, indicating that mutating the valine at position 15 to an aromatic amino acid is not beneficial (**Table 1**). However, introducing pAF in this position can overcome this penalty. Under the same conditions, LmrR_V15pAF, does not only outperform LmrR_V15Y by a factor of 11 but also the LmrR itself by a factor of 2 (**Fig. 3d** and **Table 1**). The favorable comparison to the tyrosine-containing mutant is particularly striking, as higher k_{cat} (2.5-fold) and lower K_M (4.4-fold) values are indicative for the unnatural amino acid acting as a catalytic residue in the hydrazone formation.

To evaluate the performance our designer enzyme as nucleophilic catalyst in other hydrazone formation reactions, we determined the relative rate accelerations provided by LmrR_V15pAF (at a concentration of 5 μM) over the uncatalyzed reaction for a total of eight aldehydes (see **Supplementary Information** for details). Benzaldehyde and *para*-substituted derivatives were well accepted by the enzyme and provided relative rate increases ranging from 5.38 for benzaldehyde to 2.48 for 4-carboxybenzaldehyde (**Supplementary Fig. 6**). The lower reactivity toward the latter is indicative of the preference of the LmrR pore for hydrophobic cations rather than anions.[29,30] This notion is further underscored by the fact that LmrR_V15pAF does not provide appreciable levels of rate acceleration for 2-carboxybenzaldehyde under otherwise identical conditions. Providing only a 10% increase at a 10 μM concentration (**Supplementary Fig. 6**), LmrR_V15pAF is able to differentiate between *para* and *ortho* substituted carboxybenzaldehyde. This feature is not shared by aniline or its derivatives, which provide comparable rate accelerations for both reactants.[27]

Performance of LmrR_V15pAF in oxime formation

To further probe whether the introduced aniline side chain in LmrR can act as a general nucleophilic catalyst we evaluated the proficiency of LmrR_V15pAF for an oxime formation. Toward this end, we synthesized O-(7-Nitrobenzo-[2,1,3-d]-oxadiazol-4-yl)hydroxylamine (NBD-ONH_2), a compound significantly more reactive than the equivalent hydrazine, NBD-H.[41] Under our standard conditions, NBD-ONH_2 spontaneously hydrolyzed, a formal nucleophilic aromatic substitution, to give rise to the *p*-nitrophenolate, NBD-O^- (k_{uncat} = 1.7 x 10^{-5} s^{-1}, **Fig. 4a**). Somewhat unexpectedly, when attempting to form an oxime by adding 4-NBA to the reaction mixture, we observed that NBD-O^- was produced much more readily (k_{uncat} = 0.26 M^{-1} s^{-1}). A 10^4-fold difference in the respective rate constants between the model oxime and hydrazone

formation reactions is attributed to a cascade reaction, in which a transiently formed oxime undergoes hydrolysis (**Fig. 4a**). In this cascade, oxime formation appears to be the rate-limiting step, as at no time we were able to detect the oxime but could instead observe the concomitant formation of NBD-O$^-$ and 4-nitrobenzaldimine (**Supplementary Fig. 7**).

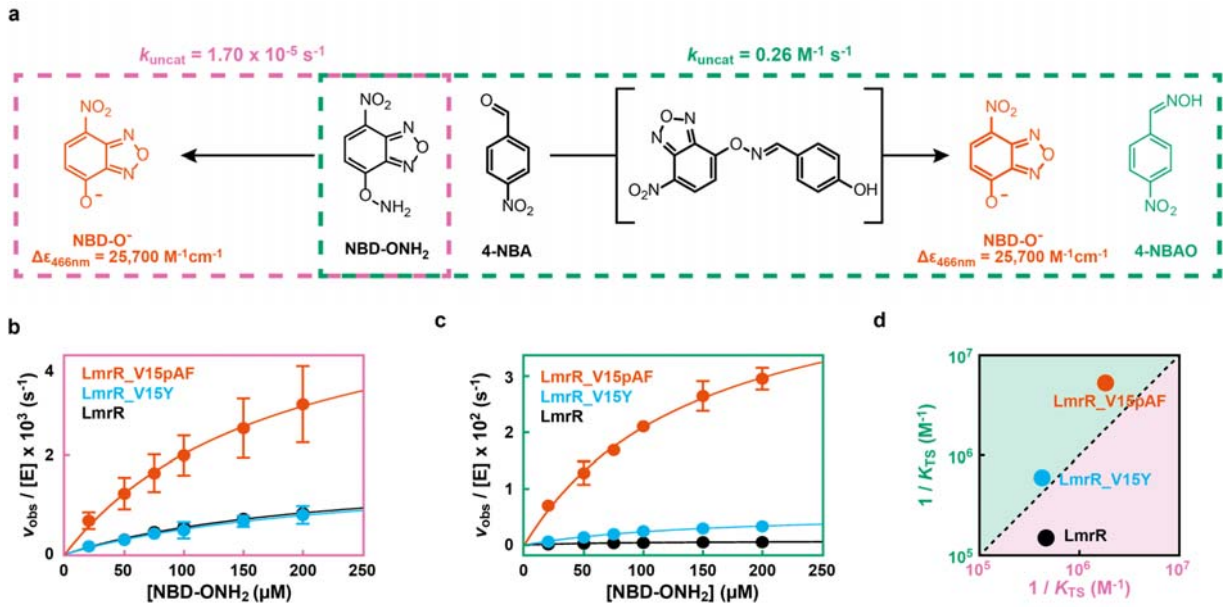

Figure 4: The aniline side chain in LmrR_V15pAF is crucial for boosting activities for a model oxime formation reaction. a: NBD-ONH$_2$ is a labile hydroxylamine that, in buffer, hydrolyzes spontaneously in a nucleophilic aromatic substitution reaction to give the *p*-nitrophenolate, NBD-O$^-$ (pink rectangle). In presence of 4-NBA a transient oxime is formed which hydrolyzes more rapidly to NBD-O$^-$, and 4-nitrobenzaldoxime (4-NBAO, green rectangle). Uncatalyzed rates were determined for both reactions following the formation of NBD-O$^-$ at 466 nm (see **Supporting Information**). **b-c:** Comparison of initial reaction velocities of LmrR, LmrR_V15Y and LmrR_V15pAF for the depicted hydrolysis (**b**) and oxime formation (**c**). LmrR_V15pAF outperforms the other proteins in both reactions; however, the contribution of the aniline side chain is significantly larger for the oxime formation. Kinetic parameters derived from these saturation kinetic studies can be found in **Supplementary Table 4**. **d:** A comparison for the observed (hydrolysis) and apparent (oxime formation) transition state stabilizations for all three proteins. Consistent with its role as a nucleophilic catalyst, LmrR_V15pAF is a better catalyst for the oxime formation, underscoring that the unnatural amino acid acts as a catalytic residue.

Upon addition of 1 mM aniline to NBD-ONH$_2$ (50 μM, phosphate buffer pH 7.4) we noticed the appearance of a color distinct from the one we observed for NBD-O$^-$ in the uncatalyzed reaction. Further analysis revealed that this difference is the result of the nucleophilic aromatic substitution of NBD-ONH$_2$ with aniline instead of water (see **Supplementary Information** for details,

Supplementary Figure 8). The same product was also observed in the oxime formation together with 4-nitrobenzaldimine, indicating that the transiently formed oxime undergoes aminolysis instead of hydrolysis. As the addition of aniline changes the pathway of both reactions and is also consumed during the reaction, it was not possible to obtain values for the rate acceleration with respect to the uncatalyzed reaction.

Conversely, we found LmrR_V15pAF to catalyze both the hydrolysis and oxime formation reaction (**Fig. 4b-c**). This difference between the pAF-containing enzyme and a small organocatalysts, such as aniline, can be attributed to the ability of the proteinaceous environment to bind the reactants. Specifically, we suggest that the central tryptophans (W96 and W96') recruit NBD-ONH_2 and subsequently position the protonated Schiff base (or water when 4-NBA is not present) in its close proximity. For the nucleophilic aromatic substitution reaction of NBD-ONH_2 with water, LmrR_V15pAF proved superior to LmrR and LmrR_V15Y, with chemical efficiencies being 4.0 and 4.4-fold higher, respectively (**Fig. 4b-c**, **Table 1**). In comparison, LmrR_V15pAF proved more efficient in catalyzing the oxime formation and outperformed the parent variant by a factor of 37 and LmrR_V15Y by a factor of 9.0 (**Table 1**). Strikingly, the improved chemical efficiency of LmrR_V15pAF for the cascade reaction with respect to the tyrosine-containing variant can exclusively be attributed to higher turnover frequencies, underscoring that the aniline side chain indeed acts as a nucleophilic catalyst. Low solubility of 4-NBA prevented us to determine its K_m value and, as a result, an apparent third order rate constant. Nevertheless, as the concentration of 4-NBA under assay conditions (250 μM) is much lower than the K_m value for the aldehyde substrate measured in the model hydrazone formation, we were able to estimate an apparent transition state stabilization (**Table 1**). When comparing the chemical proficiencies ($1/K_{TS}$) of all three LmrR variants for the two reactions (**Fig. 4d**), LmrR_V15pAF is the only

enzyme that has a clear preference for the oxime formation over the hydrolysis reaction, which is consistent with a catalytic role of pAF in the oxime formation reaction.

Conclusion

Here, we presented the design and characterization of a novel type of artificial enzyme that utilizes an unnatural amino acid as a catalytic residue. Two factors contribute to the proficiency of the identified designer enzyme, LmrR_V15pAF, in hydrazone and oxime formation reactions. On the one hand, the promiscuous binding pocket of LmrR provides a suitable environment to recruit the reactants and increase their "effective molarity". This mode of action is confirmed by the previously unidentified activity of the parent LmrR for the reactions performed in our study. On the other hand, judiciously placing of *p*-aminophenylalanine inside the hydrophobic pore increases the reactivity of the aniline side chain and results in a more efficient nucleophilic catalyst. Such merging of catalytic strategies is reminiscent of natural enzymes, which typically achieve their unmatched catalytic proficiencies by positioning uniquely functional groups in a suitable binding pocket. As a result, LmrR_V15pAF outperformed aniline by a factor >550 in a model hydrazone formation and the unnatural amino acid also proved crucial for catalyzing a related oxime formation reaction.

Thus far however, the contribution of the unnatural amino acid to catalysis is modest, as it only outperforms the corresponding tyrosine variant by about a factor of 10. Nevertheless, the combination of a promiscuous binding pocket and a uniquely active side chain provides an ideal starting point for future efforts that will focus on exploring directed evolution strategies to optimize the role of the unnatural amino acid for catalysis. Lastly, we envision that placing other unnatural amino acids that feature unique reactivities in promiscuous binding pockets will prove a rewarding strategy to significantly expand the repertoire of reactions catalyzed by designer enzymes.

Methods

Protein purification: Flasks containing 500 mL LB-medium with 100 μg/mL of ampicillin and 34 μg/ml of chloramphenicol were inoculated with 2 mL of a densely grown overnight culture of *E. coli* BL21(DE3) cells harboring the plasmids pEVOL-pAzF and pET17b_LmrR_X (see **Supplementary Information** for details). Cells were incubated at 37 °C and 135 rpm until an optical density at 600 nm of 0.8–0.9 was reached. At this point gene expression was induced with isopropyl β-D-1-thiogalactopyranoside (IPTG, final concentration 1 mM) and L-Arabinose (final concentration 0.02%) and 100 mg/L of pAzF was added as solid. Incubation was continued overnight at 30 °C, after which cells were harvested by centrifugation (6000 rpm, JA10, 20 min, 4 °C, Beckman). The cell pellet was resuspended in 20 mL washing buffer (50 mM NaH_2PO_4, 150 mM NaCl, pH 8.0) containing a protease inhibitor cocktail (cOmplete^{TM} Mini, Roche) and subsequently lysed by sonication (70% (200 W) for 7 min, 10 sec on, 15 sec off). The lysed cells were incubated with DNase I (final concentration 0.1 mg/mL with 10 mM $MgCl_2$) and phenylmethanesulfonyl fluoride solution (final concentration 0.1 mM) for 30 min at 4 °C. After removing cell debris by centrifugation (15000 rpm, JA-17, 1h, 4 °C, Beckman). The cleared lysate was loaded on a Strep-Tactin column (Strep-Tactin® Superflow® high capacity), incubated for 60 min at 4 °C and the protein was purified according to the manufacturer's guidelines. Protein-containing fractions, as judged by SDS-PAGE electrophoresis, were pooled, concentrated and variants harboring a pAzF residue reduced by addition of TCEP (final concentration 10 mM). The reduction was performed for at least 2 hours at 4 °C and the mixture was subsequently dialyzed against the reaction buffer (50 mM NaH_2PO_4, 150mM NaCl, pH 7.4, 2-times 1 L). The concentration of the protein variants was determined by using the calculated extinction coefficient for LmrR, corrected for the absorbance of the pAF (ε_{280} = 1333 M^{-1} cm^{-1}).

Initial catalytic studies with 4-NBA: The hydrazone formation between 4-NBA and NBD-H was performed under pseudo-first-order conditions in quartz cuvettes (path length 1 cm) at 25 °C. Product formation was monitored spectroscopically (Jasco V-660 spectrophotometer) for two hours following the absorption at 504 nm (for the uncatalyzed reaction), at 495 nm (when catalyzed with aniline) or at 528 nm (in the presence of proteins at a concentration of 10 μM). Unless otherwise stated, all reactions were performed in reaction buffer containing 5% (v/v) DMF, 10 μM protein (concentration of the dimer), 18 μM NBD-H and 1 mM 4-NBA. Reactions were started by addition of 8 μL of a 60 mM solution of 4-NBA in DMF and reaction progress continuously monitored (one data point per minute). The recorded absorbance data was converted to the concentration of the hydrazone by using the determined extinction factors as followed: ε_{504} = 16,545 M^{-1} cm^{-1} (background reaction), ε_{495} = 16,319 M^{-1} cm^{-1} (in presence of aniline) and ε_{528} = 15,675 M^{-1} cm^{-1} (in presence of 10 μM protein). All reported values are the average of 3-5 independent measurements.

Kinetic assays: Kinetic assays for the hydrazone formation between 4-HBA and NBD-H were performed following the formation of product at 472 nm (ε_{472} = 25,985 M^{-1} cm^{-1}). Kinetic assays for NBD-ONH_2 hydrolysis and the oxime formation in presence of 4-NBA (250 μM) were performed following the formation of the *p*-nitrophenolate, NBD-O^-, at 466 nm (ε_{466} = 25,700 M^{-1} cm^{-1}). All measurements were conducted at 25 °C in reaction buffer (pH 7.4) containing 5% (v/v) DMF. Kinetic measurements were carried out at enzyme concentrations from 1 – 5 μM, depending on the activity the variants displayed for the reaction. NBD-H and NBD-ONH_2 concentrations were varied ((20, 50, 75, 100, 150, 200 μM) while 4-HBA (3, 5, 7.5, 10, 15, 20 mM) and 4-NBA concentrations were fixed (250 μM). All reactions were started by addition of NBD-H or NBD-ONH_2 to a master mix containing all components but the enzyme. Background rates (hydrazone formation = 10 – 30 min, hydrolysis = 300 s, oxime formation = 90 s) were initially recorded for

each measurement before addition of the enzyme. Initial velocities were obtained by correcting for the background rate and the resulting values, which are the average of at least two independent measurements, fitted to the Michaelis-Menten equation. For the full kinetic characterization of LmrR_V15pAF in the hydrazone formation, initial velocities were globally fitted (*RStudio*) to the following equation:

$$\frac{v_0}{[E]} = \frac{k_{cat}\,[4HBA][NBDH]}{(K_{4HBA} + [NBDH])(K_{NBDH} + [4HBA])}$$

More information on the experimental design, protocols and data analysis are provided in the **Supplementary Information.**

Acknowledgments

This project 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). G.R. acknowledges support from the Ministry of Education Culture and Science (Gravity programme no. 024.001.035). C.M. gratefully acknowledges a Marie Skłodowska Curie Individual Fellowship (project no. 751509). The authors thank Annika Borg for preparing the plasmid harboring the *bc*PadR1 gene and M.P. de Vries from Interfaculty Mass Spectrometry Center, University of Groningen, for performing trypsin digestion and LC-MS/MS analyses.

Author Contributions

ID and CM contributed equally to this work.

GR and ID conceived the project. ID, CM and GR planned the experiments and wrote the manuscript. ID, CM, and CD performed the experimental work. All authors discussed the results and commented on the manuscript.

References

1. Hilvert, D. Design of protein catalysts. *Annu. Rev. Biochem.* **82,** 447–470 (2013).
2. Nanda, V. & Koder, R. L. Designing artificial enzymes by intuition and computation. *Nat. Chem.* **2,** 15–24 (2010).
3. Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of *de novo* protein design. *Nature* **537,** 320–327 (2016).
4. Woolfson, D. N. *et al. De novo* protein design: how do we expand into the universe of possible protein structures? *Curr. Opin. Struct. Biol.* **33,** 16–26 (2015).
5. Hilvert, D. Critical analysis of antibody catalysis. *Annu. Rev. Biochem.* **69,** 751–793 (2000).
6. Kiss, G., Çelebi-Ölçüm, N., Moretti, R., Baker, D. & Houk, K. N. Computational enzyme design. *Angew. Chem. Int. Ed.* **52,** 5700–5725 (2013).
7. Renata, H., Wang, Z. J. & Arnold, F. H. Expanding the enzyme universe: accessing non-natural reactions by mechanism-guided directed evolution. *Angew. Chem. Int. Ed.* **54,** 3351–3367 (2015).
8. Schwizer, F. *et al.* Artificial Metalloenzymes: reaction scope and optimization strategies. *Chem. Rev.* (2017). doi:10.1021/acs.chemrev.7b00014.
9. Okeley, N. M. & van der Donk, W. A. Novel cofactors via post-translational modifications of enzyme active sites. *Chem. Biol.* **7,** 159–171 (2000).
10. van Poelje, P. D. & Snell, E. E. Pyruvoyl-dependent enzymes. *Annu. Rev. Biochem.* **59,** 29–59 (1990).
11. Appel, M. J. & Bertozzi, C. R. Formylglycine, a post-translationally generated residue with unique catalytic capabilities and biotechnology applications. *ACS Chem. Biol.* **10,** 72–84 (2015).
12. Cooke, H. A., Christianson, C. V & Bruner, S. D. Structure and chemistry of 4-methylideneimidazole-5-one containing enzymes. *Curr. Opin. Chem. Biol.* **13,** 460–468 (2009).
13. Budisa, N. *et al.* Xenobiology meets enzymology: exploring the potential of unnatural building blocks in biocatalysis. *Angew. Chem. Int. Ed.* (2017). doi:10.1002/anie.201610129.
14. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79,** 413–444 (2010).
15. Dumas, A. *et al.* Designing logical codon reassignment – Expanding the chemistry in biology. *Chem. Sci.* **6,** 50–69 (2015).
16. Neumann-Staubitz, P. & Neumann, H. The use of unnatural amino acids to study and engineer protein function. *Curr. Opin. Struct. Biol.* **38,** 119–128 (2016).
17. Green, A. P., Hayashi, T., Mittl, P. R. E. & Hilvert, D. A chemically programmed proximal ligand enhances the catalytic properties of a heme enzyme. *J. Am. Chem. Soc.* **138,** 11344–11352 (2016).
18. Drienovská, I., Rioz-Martinez, A., Draksharapu, A. & Roelfes, G. Novel artificial metalloenzymes by *in vivo* incorporation of metal-binding unnatural amino acids. *Chem.*

Sci. **6,** 770–776 (2015).
19. Bersellini, M. & Roelfes, G. Multidrug resistance regulators (MDRs) as scaffolds for the design of artificial metalloenzymes. *Org. Biomol. Chem.* **15,** 3069–3073 (2017).
20. Hu, C., Chan, S. I., Sawyer, E. B., Yu, Y. & Wang, J. Metalloprotein design using genetic code expansion. *Chem. Soc. Rev.* **43,** 6498 (2014).
21. Mehl, R. A. *et al.* Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc.* **125**, 935-939 (2003).
22. Carrico, Z. M. *et al.* Oxidative coupling of peptides to a virus capsid containing unnatural amino acids. *Chem. Commun.* **369,** 1205-1207 (2008).
23. Cordes, E. H. & Jencks, W. P. Nucleophilic catalysis of semicarbazone formation by anilines. *J. Am. Chem. Soc.* **84,** 826–831 (1962).
24. Dirksen, A., Dirksen, S., Hackeng, T. M. & Dawson, P.E. Nucleophilic catalysis of hydrazone formation and transimination: implications for dynamic covalent chemistry. *J. Am. Chem. Soc.* **128**, 15602-15603 (2006).
25. Dirksen, A., Hackeng, T. M. & Dawson, P. E. Nucleophilic catalysis of oxime ligation. *Angew. Chem. Int. Ed.* **45,** 7581–7584 (2006).
26. Crisalli, P. & Kool, E. T. Water-soluble organocatalysts for hydrazone and oxime formation. *J. Org. Chem.* **78,** 1184–1189 (2013).
27. Crisalli, P. & Kool, E. T. Importance of *ortho* proton donors in catalysis of hydrazone formation. *Org. Lett.* **15,** 1646–1649 (2013).
28. Wendeler, M., Grinberg, L., Wang, X., Dawson, P. E. & Baca, M. Enhanced catalysis of oxime-based bioconjugations by substituted anilines. *Bioconjug. Chem.* **25,** 93–101 (2014).
29. Agustiandari, H., Lubelski, J., van den Berg van Saparoea, H. B., Kuipers, O. P. & Driessen, A. J. M. LmrR is a transcriptional repressor of expression of the multidrug ABC transporter LmrCD in *Lactococcus lactis*. *J. Bacteriol.* **190,** 759–763 (2008).
30. Madoori, P. K., Agustiandari, H., Driessen, A. J. M. & Thunnissen, A.-M. W. H. Structure of the transcriptional regulator LmrR and its mechanism of multidrug recognition. *EMBO J.* **28,** 156–166 (2009).
31. Bos, J., Fusetti, F., Driessen, A. J. M. & Roelfes, G. Enantioselective artificial metalloenzymes by creation of a novel active site at the protein dimer interface. *Angew. Chem. Int. Ed.* **51,** 7472–7475 (2012).
32. Bos, J., Browne, W. R., Driessen, A. J. M. & Roelfes, G. Supramolecular assembly of artificial metalloenzymes based on the dimeric protein LmrR as promiscuous scaffold. *J. Am. Chem. Soc.* **137,** 9796–9799 (2015).
33. Chin, J. W. *et al.* Addition of p-azido-L-phenylalanine to the genetic code of *Escherichia coli*. *J. Am. Chem. Soc.* **124**, 9026–9027 (2002).
34. Fibriansah, G. *et al.* Crystal structures of two transcriptional regulators from *Bacillus cereus* define the conserved structural features of a PadR subfamily. *PLoS One* **7,** e48015 (2012).
35. Albanese, D. C. M. & Gaggero, N. Albumin as a promiscuous biocatalyst in organic synthesis. *RSC Adv.* **5,** 10588–10598 (2015).
36. Braisted, A. C. & Schultz, P. G. An antibody-catalyzed bimolecular Diels-Alder reaction. *J. Am. Chem. Soc.* **112,** 7430–7431 (1990).
37. Gouverneur, V. E. *et al.* Control of the exo and endo pathways of the Diels-Alder reaction by antibody catalysis. *Science* **262,** 204–208 (1993).
38. Yli-Kauhaluoma, J. T. *et al.* Anti-metallocene antibodies: a new approach to enantioselective catalysis of the Diels-Alder reaction. *J. Am. Chem. Soc.* **117,** 7041–7047 (1995).

39. Xu, J. *et al.* Evolution of shape complementarity and catalytic efficiency from a primordial antibody template. *Science* **286,** 2345-2348 (1999).
40. Siegel, J. B. *et al.* Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **329,** 309-313 (2010).
41. Key, J. A., Li, C. & Cairo, C. W. Detection of cellular sialic acid content using nitrobenzoxadiazole carbonyl-reactive chromophores. *Bioconjug. Chem.* **23,** 363–371 (2012).