# An *In Vivo* Biocatalytic Cascade Featuring an Artificial Enzyme Catalyzed New-to-Nature Reaction

Linda Ofori Atta [a]†, Zhi Zhou [a]†‡ and Gerard Roelfes*[a]

[a] L. Ofori Atta, Dr. Z. Zhou, Prof. Dr. G. Roelfes
Stratingh Institute for Chemistry, University of Groningen
Nijenborgh 4, 9747 AG Groningen (The Netherlands)
E-mail: j.g.roelfes@rug.nl

† These authors contributed equally.
‡ current address: School of Pharmaceutical Sciences, Jiangnan University, Wuxi 214122, China.

**Abstract:** Artificial enzymes utilizing the genetically encoded non-proteinogenic amino acid p-aminophenylalanine (pAF) as catalytic residue are able to react with carbonyl compounds through an iminium ion mechanism, making reactions possible that have no equivalent in nature. Here, we report an *in vivo* biocatalytic cascade that is augmented with such an artificial enzyme catalyzed new-to-nature reaction. The artificial enzyme in this study is a pAF containing evolved variant of the Lactococcal multidrug resistance Regulator, designated LmrR_V15pAF_RMH, which efficiently converts in vivo produced benzaldehyde derivatives into the corresponding hydrazone products inside *E. coli* cells. These in vivo biocatalytic cascades comprising an artificial enzyme catalyzed reactions are an important step towards achieving a hybrid metabolism.

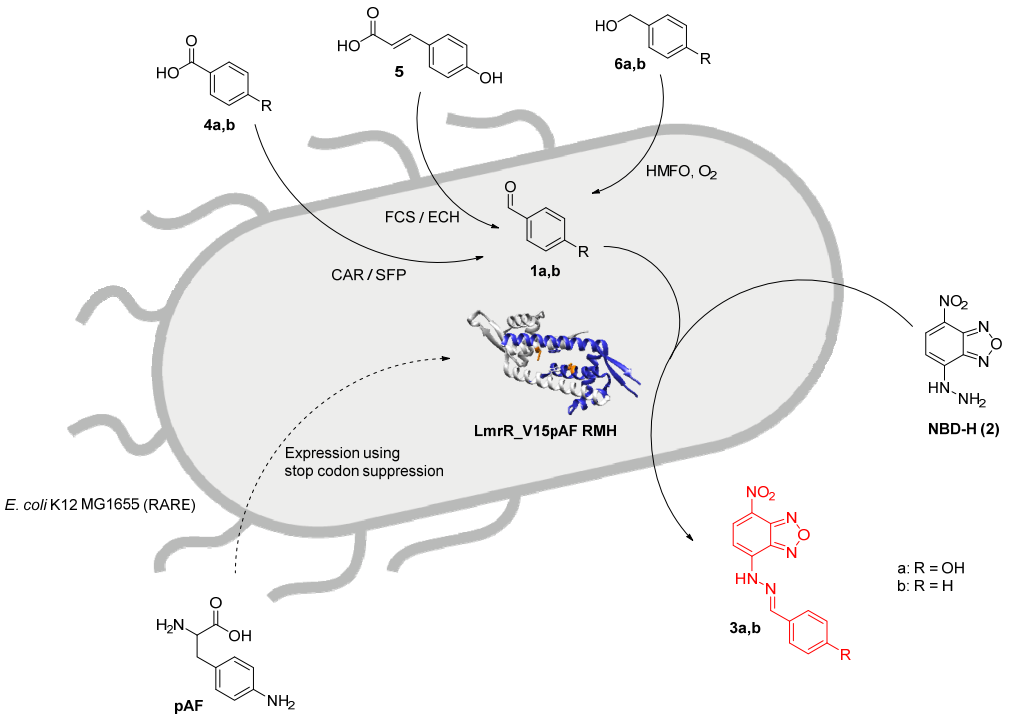
## Introduction

The cellular metabolism uniquely demonstrates the power of biocatalytic cascade reactions in living cells. Molecular structures of remarkable complexity are produced using an intricate network system of interconnected biocatalytic reactions. Such biological systems have proven amenable to modification by metabolic engineering.[1] However, the structural diversity that can be achieved is inherently limited by the synthetic repertoire available to nature. In contrast, chemical synthesis offers virtually unlimited versatility in reaction scope, but to date synthetic chemical systems cannot rival the sophistication of nature. Hence, an attractive solution would be to combine the best of both worlds to create a “hybrid metabolism”, which augments biological synthesis with abiological catalytic chemistry.[2]

Most efforts to date have focused on supplementing biological synthesis with a transition metal- or organocatalysed catalysed reaction.[3–10] While some encouraging results have been obtained, low activity of transition metal complexes and organocatalysts in biological systems, as well as potential mutual incompatibility, is often a limiting factor. An alternative approach involves using artificial enzymes, created by integrating a synthetic catalysts into a protein scaffold.[11] The protein environment helps to accelerate the reaction by providing additional interactions, but also protects the chemical catalyst from the cellular environment.[2] Recently, the first examples of application of artificial metalloenzymes in living cells were reported.[12–17] Also, biocatalytic cascades involving an artificial metalloenzyme was reported, albeit that this was carried out in vitro, using isolated proteins.[18,19] Finally, iridium-heme substituted P450 enzyme catalysed cyclopropanation of a biosynthetically produced terpene in a heavily engineered *E. coli* strain was recently reported.[20]

We have recently introduced the concept of using genetically encoded non-canonical amino acids directly as organocatalytic residue in artificial enzymes.[21,22] Here, we report the application of a catalytic p-aminophenylalanine (pAF) containing artificial enzyme in *E. coli* and its integration into an *in vivo* biocatalytic cascade. Thus biocatalytic cascades comprising natural enzymes, producing benzaldehyde derivatives from different starting materials, is augmented with an artificial enzyme catalyzed abiological reaction to create a hydrazone product.

The non-canonical amino acid p-aminophenyl alanine (pAF) into the Lactococcal multidrug resistance Regulator (LmrR)[23] using the amber stop codon suppression methodology.[21,24,25] pAF contains an aniline side chain that can react with aldehydes to form transient iminium ion species that are versatile reactive intermediates in organocatalysis.[26] LmrR with pAF at position 15, designated LmrR_V15pAF, was shown to catalyse the abiotic reaction of benzaldehyde derivatives with hydrazines to form hydrazones.[21,27,28] Subsequent directed evolution gave rise to the variant with mutations A92R_N19M_F93H, designated LmrR_V15pAF_RMH, which showed a 57-fold increase in catalytic efficiency compared to the parent LmrR_V15pAF.[29] LmrR_V15pAF and evolved variants have also been used successfully in conjugate addition reactions to enals, also exploiting the iminium ion activation strategy.[30,31]




**Scheme 1.** Schematic representation of *in vivo* biocatalytic cascade featuring an abiological hydrazone formation reaction catalyzed by the designer enzyme LmrR_V15pAF_RMH.

Since LmrR_pAF is an artificial enzyme that is biosynthesized in living cells, we surmised that the catalysed hydrazone formation reaction could also be carried out *in vivo*. Moreover, since a variety of enzymatic synthesis of aldehydes from carboxylic acid or alcohol precursors are known,[32–34] we envisioned an *in vivo* biocatalytic cascade comprising the biosynthesis of aldehydes using canonical enzymes followed by the LmrR_V15pAF_RMH catalyzed new-to-nature hydrazone formation reaction.

## Results and Discussion

### Direct Incorporation of para-amino phenylalanine

In our earlier work, it proved most efficient to introduce pAF indirectly, by first introducing p-azidophenylalanine (pAF) followed by Staudinger reduction with tris(2-carboxyethyl) phosphine (TCEP), added during the purifcation process.[21] Direct incorporation of pAF by amber stop codon suppression, using the dedicated orthogonal translation system (pDULE2-para-aminoPhe),[25] was possible, but proved to give significant misincorporation of other amino acids under standard expression conditions. However, for applications in vivo indirect incorporation via pAzF cannot be used. Hence, our initial effort focused on improving the direct incorporation of pAF by optimizing the expression conditions.

For this purpose, we tested the expression of the evolved variant of LmrR-V15pAF containing three mutations, N19M_A92R_F93H (LmrR_V15pAF_RMH).[29] Additionally, the protein includes two mutations in the DNA binding domain, D55K_Q59K, and contains a C-terminal strep tag. Three different media, that are, Lysogeny Broth (LB), M9 minimal medium and minimal medium with vitamins (MMV), were evaluated for protein expression under different conditions.

In all cases, overexpression of full length protein was observed. MS analysis of the isolated proteins showed a mixture of the desired LmrR_V15pAF_RMH and variants resulting from mis-incorporation of, in particular, phenylalanine at position 15, albeit with varying ratios depending on the conditions used. Judging from the relative intensities of the peaks, it was suggested that expression in MMV gave rise to the largest fraction of pAF incorporation. As a more quantitative measure for incorporation of pAF, the kinetics of the catalyzed hydrazone formation reaction between p-hydroxybenzaldehyde (**1a**) and 4-hydrazino-7-nitro-2,1,3-benzoxadiazole (NBD-H, **2**) were determined and compared to those of LmrR_V15pAF_RMH prepared independently via indirect incorporation.[29] The protein produced in LB and M9 media only showed low activity, suggesting that in these cases mostly misincorporation of other amino acids occurred, consistent with the MS results. In contrast, the protein expressed in MMV medium at 24 ºC for 48 hours gave good activity. Based on the comparison of the catalytic efficiency it was concluded that the incorporation efficiency of pAF was ~80% (Table S2). This was deemed sufficient for *in vivo* catalysis experiments. Additionally, the incorporation of pAF at position 15 was confirmed by trypsin digest (Figure S4).

**Hydrazone formation reactions by an artificial enzyme in whole cells**

The activity of whole *E. coli* cells expressing wild type LmrR, i.e. without pAF, LmrR_V15pAF and LmrR_V15pAF_RMH from direct incorporation in the hydrazone formation reaction was tested by using exogeneous aldehydes and NBD-H. First, a set of LmrR variants were tested for the hydrazone formation from both para-hydroxybenzaldehyde (**1a**) and benzaldehyde (**1b**) in MMV and phosphate buffer pH 7 for 3 hours in whole K-12 MG1655 (DE3) Cells. This strain allows for aromatic aldehyde accumulation because it has a reduced aromatic aldehyde reduction (RARE) activity.[35] This is due to the fact that it has been engineered to reduce the reduction of aldehydes by endogenous ketoreductases and alcohol dehydrogenases. Control experiments without cells or with wild type LmrR gave up to ~20% yield of the corresponding hydrazone products **3a** and **3b**, respectively, in MMV medium, while only a low yield was observed in buffer (Figure 1). This shows that there is some background reaction in MMV medium. Using cells expressing

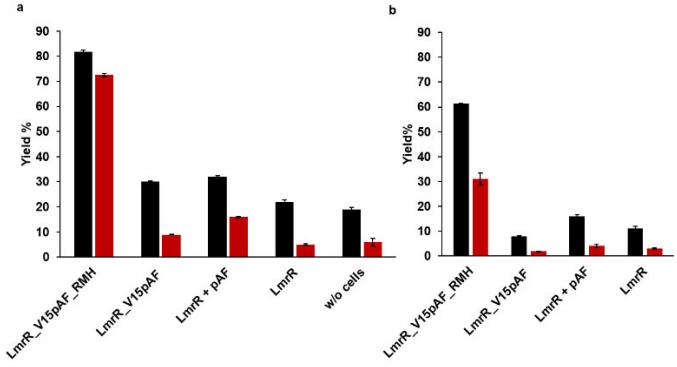



**Figure 1.** Yield of hydrazone products using MMV medium (black) and phosphate buffer (orange) in *E.coli* K12 MG1655 (RARE), (**a**) yield of **3b**; conditions: 5 mM **1b** and 50 µM **2** in both MMV medium and phosphate buffer pH 7, at 24 ℃/ 135 rpm for 3 h. (**b**) yield of **3a**; conditions: 5 mM **1a** and 50 µM **2** in both MMV medium and phosphate buffer pH 7, at 24 ℃/ 135 rpm for 3 h. Error margins represent standard deviations from experiments performed in triplicate.

LmrR_V15pAF or LmrR supplemented with pAF showed a somewhat higher yield, indicating a catalytic effect, albeit quite small. In contrast, when using cells containing the evolved variant LmrR_V15pAF_RMH, the hydrazone products **3a,b** were obtained with significantly higher yield of 62 % and 82 % yield in MMV and 31 % and 73 % in buffer, respectively. These results show that LmrR_V15pAF_RMH catalyzes the hydrazone formation in the cell. The lower yields obtained for **3a** reflect the lower reactivity of **1a** due to the strongly electron donating -OH substituent.

Next, the biocatalytic synthesis of aromatic aldehydes in *E.coli* K-12 MG1655 cells was investigated. We focused on production of benzaldehyde and para-hydroxy-benzaldehyde, using three different approaches:

i. Carboxylic acid reductase (CAR) from *Nocardia iwensis,* along with its one time activator phosphopantetheinyl transferase (SFP) from *Bacillus subtilis,* reduces benzoic acids into their corresponding benzaldehydes in a NADPH and ATP dependent fashion.[36,37] A codon optimised gene for *CAR_SFP* was harboured on a pETDuet-1 vector.

ii. The conversion of p-coumaric acid to p-hydroxybenzaldehyde by a combination of feruloyl CoA synthase (encoded by *Atfcs*) and enoyl CoA hydratase/aldolase (encoded by *Atech*) from *A. thermoflava* N1165.[38] The conversion proceeds via the initial formation of the SCoA derivative by FCS, followed by the conjugate addition of water by ECH/hydratase and retro-aldol condensation by ECH/aldolase to produce the corresponding aldehyde.[39–41] The required enzymes were expressed heterogeneously using a pETDuet-1 vector harbouring *Atfcs_Atech*.

iii. 5-hydroxymethylfurfural (HMF) oxidase (HMFO) from the white rot basidiomycete *phanerochaete chrysosporium,* an engineered flavin containing enzyme which uses molecular oxygen to oxidize alcohols to the corresponding carboxylic acids, via the intermediate formation of the corresponding aldehydes,[42,43] with concomitant formation of $H_2O_2$. The HMFO gene was cloned into a pBAD vector.

The respective enzymes were expressed at 24 ºC /135 rpm for 16 hours in MMV medium, since this was the medium that gave the expression, the required substrates were added immediately to FCS-ECH (Scheme 1. **5**) and CAR-SFP (Scheme 1. **4a,b**) containing cells to a final concentration of 5 mM. For HMFO containing cells, 5 mM substrate (Scheme 1. **6a,b**), were added after 16 hours of expression and then reacted for 2 h.

Both *AtFcs-AtEch*/*E. coli RARE*, starting from p-coumaric acid, and *CAR-SFP*/*E. coli RARE*, from p-hydroxybenzoic acid, produced a good yield of p-hydroxybenzaldehyde. In contrast, the yield of the reduction of benzoic acid to benzaldehyde catalysed by CAR-SFP/*E. coli RARE* was significantly lower. FCS-ECH was not used for production of benzaldehyde, as these enzymes do not accept cinnamic acid as substrate. The whole cell HMFO

catalysed oxidation of benzyl alcohol and p-hydroxybenzylalcohol gave **3a** and **3b** in 30 and 79 % yield, respectively (Figure S11). Having demonstrated that both LmrR_V15pAF RMH catalyzed hydrazone formation and the production of its substrates benzaldehyde and p-hydroxybenzaldehyde can be achieved in *E. coli RARE* cells, we sought to combine these processes to create the *in vivo* biocatalytic cascade. Initial attempts to integrate the genes that encode CAR-SFP or FCS-ECH, LmrR_V15TAG and the orthogonal translation system on two plasmids proved unsuccessful due to the large size of *CAR_SFP*. For this reason, we then constructed the system by integrating three plasmids: pETDuet-1 harboring either the genes for *CAR_SFP* or *FCS-ECH*, pET28b+ harbouring the gene for LmrR_V15pAF_RMH and pDULE-para-aminoPhe 2 containing the genes for the orthogonal translation system, p-aminoPhe tRNA synthetase/tRNA. The three plasmids were were co transformed into *E. Coli K-12 MG1655* (RARE). HMFO, which is much smaller, and LmrR_V15pAF RMH were both cloned on pBAD. Hence, in this case a two plasmid system comprising the pBAD vector and pDULE-para- aminoPhe 2 could be used.

First, the different cell variants were tested in minimal media in the absence of NBD-H, to evaluate how the production of aldehyde is affected by the presence of all genes (Figure S3). *Car-Sfp* gave rise to a yield around 20 % in both benzaldehyde and para hydroxy benzaldehyde production, which is lower than obtained before in the single plasmid system. In contrast, with the *Fcs-Ech* containing cells the yield of p-hydroxy benzaldehyde was around 90 %. Finally, the oxidation of benzyl alcohols with the HMFO containing cells resulted in 60 % yield of benzaldehyde and ~ 50 % of *p*-hydroxybenzaldehyde.

The in vivo cascade hydrazone formation were performed using cells from a 2.5 mL cell culture that were resuspended in freshly prepared MMV medium and the cascade reaction was started by adding NBD-H (final concentration 50 μM) and aldehyde precursor (final concentration 5 mM). Then the mixture was incubated at 24 °C in a shaker at 135 rpm for 3 to 24 hours. The whole cells expressing both CAR with LmrR_V15pAF_RMH showed good reactivity with a hydrazone yield of 57% after 3 hours in whole cells. The control experiments with only LmrR_V15pAF_RMH showed no reactivity, which was expected because no aldehyde was produced. Surprisingly, using only CAR, in the absence of LmrR_V15pAF_RMH, resulted in 28% yield of the hydrazone. This suggests that CAR also to a small extent can catalyze hydrazone formation, albeit that this needs to be investigated further. p-hydroxybenzoic acid as substrate also gave rise to formation of the corresponding hydrazone product.

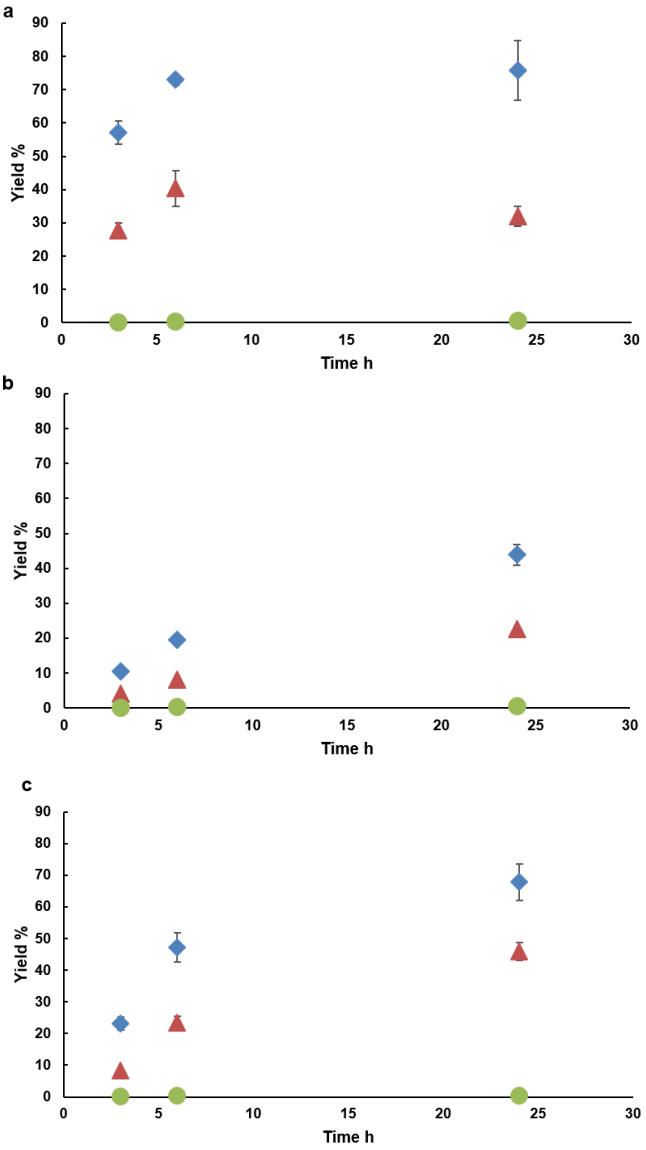



**Figure 2.** Hydrazone formation *in vivo* after the integration of CAR-SFP and FCS-ECH with LmrR_V15pAF_RMH in *E.coli* K12 MG1655 (RARE), using MMV at 24 ℃ with135 rpm for 24 h. (**a**) hydrazone formation using 5 mM **4b** and 50 μM **2.** CAR-SFP/LmrR_V15pAF_RMH (blue) gave 57 %, 73 % and 76 % after 3, 6, and 24 h, respectively. Although CAR only (red) gave some background, LmrR_V15pAF_RMH only (green) did not yield any product. (**b**) 5 mM **4a** and 50 μM **2** (**c**) 5 mM **5** and 50 μM **2**, the trend was not different from **(a).** FCS-ECH_LmrRV15pAF_RMH (blue) gave 23 %, 47 %, and 68 % of **3a** after 3, 6, and 24 h, respectively. The FCS-ECH only (red) gave a background of 46 % after 24 h. The results for LmrR_V15pAF_RMH (green) were the same for all substrates (**4a, b, 5**). Error margins represent standard deviations from experiments performed in triplicate.

Since significantly higher yields of product were observed when both CAR and LmrR_V15pAF_RMH were present, as compared to CAR alone, it can be concluded that the hydrazone product is formed predominantly from the combination of the heterologous biosynthesis of aldehyde and hydrazone formation catalyzed by the artificial enzyme.

Using whole cells expressing FCS-ECH / LmrR_V15pAF_RMH did give rise to good yields of hydrazone products based on exogenously added p-coumaric acid as starting material. The yield of hydrazone product was higher than obtained with the combination of CAR with LmrR_V15pAF_RMH. However, also in this case a non-neglible amount of product was formed in the control experiment without LmrR_V15pAF, showing the presence of significant background reactions.

***In vivo* cascade of HMFO and LmrR_V15pAF_RMH in cells**. The cascade alcohol oxidation and hydrazone formation were first tested in vitro, using purified HMFO, in absence and presence of catalase, and LmrR_V15pAF_RMH with excellent yield, to establish if the $H_2O_2$ produced by HMFO would negatively affects the reaction (Figure.SI.9).Then we performed the in vivo oxidation and hydrazone formation cascade reactions using the whole cells expressing HMFO and LmrR_V15pAF_RMH. The reactions were set up with NBD-H (final concentration 50 µM) and benzyl alcohol (final concentration 5 mM) in cells for two hours in phosphate buffer pH 6.5. The cells expressing HMFO and LmrR_V15pAF_RMH gave the corresponding hydrazone product **3b** with 81 % yield after 2 hours. In contrast, the control experiment with cells only expressing HMFO gave only 6 % yield, which is the background reaction. Using para-hydroxyl benzyl alcohol with 41 % yield of the hydrazone product was obtained after 20 hours. Those results suggest the successful and efficient in vivo cascade pathway by the combination of a natural enzyme and an artificial enzyme. Since the HFMO LmrR_V15pAF_RMH gave such a large difference compared to background reaction, it ispossible to observe the difference in reactivity by visual inspection (Figure 3b).

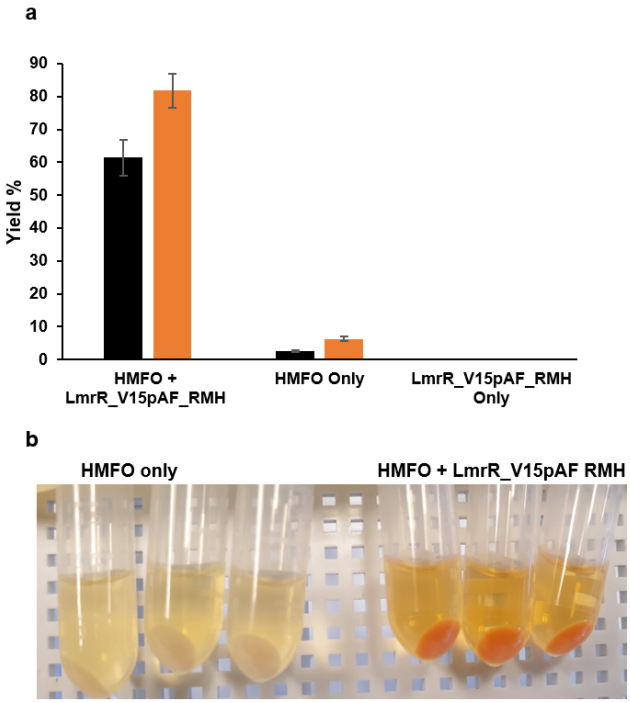



**Figure 3** (**a**) The *in vivo* cascade hydrazone formation reaction in 2.5 mL cell culture with fresh phosphate buffer (pH 6.5) by adding 50 µM **2** and 5 mM **6b** at 24 °C with 135 rpm for 1h (black) to 2 h (orange). After 1 h 61 % of **3b** was detected and 81% after 2h. However, HMFO only and LmrR_V15pAF only yielded no product. Error margins represent standard deviations from experiments performed in triplicate. (**b**) Visual representation of hydrazone formation in phosphate buffer pH 6.5, at 24 °C with 135 rpm for 2 h with HMFO_LmrR_V15pAF_RMH (right) compared to HMFO only (left).

To confirm that the catalysis occurs in the cell and that the cells remain intact during and after catalysis a recycling experiment was performed. In the first round, we performed the cascade reaction by using 5 mM of benzylalcohol and 50 uM of NBD-H. The reaction was kept at 24 ℃ /135 rpm for 3 hours, resulting in 63 % yield of the hydrazone product **3b** (Figure SI.12). The cells were then washed with 50 mM Kpi buffer at pH 6.5 to remove all extracellular components and resuspended in reaction buffer. 5 mM p-hydroxybenzyl alcohol and 50 uM NBD-H were added. After reaction for 16 hours, the corresponding product (Figure SI.12) was obtained in 22 % yield, accompanied by ~ 10 % residual product from the first round. This experiment confirms that the cells remain intact and that both the HMFO and the artificial enzyme remained inside the cells during the catalysis.

The viability of the remaining cells after catalysis was tested by spinning them down and resuspending in 1 ml LB, after which serial dilutions were plated on agar containing the required antibiotics. 336 colonies of *HMFO/LmrR_V15pAF_RMH* / *E. coli* RARE were still present on plates with cells diluted to $10^{-6}$ (Figure S17). This means that the reaction conditions, i.e. in presence of both the canonical and artificial enzymes, as well as the substrates and the hydrazone product, are well tolerated.

## Conclusion

In conclusion, we have created in vivo biocatalytic cascades in *E. coli* that comprise a combination of natural and artificial enzymes. These results show that a heterologous biosynthetic pathway can be augmented with a new-to-nature catalytic reaction for the production of novel compounds. This work is a step towards the creation of a hybrid metabolism, which combines multiple natural and artificial enzymes in living cells to produce complex molecules from simple bio-based starting materials.

## Acknowledgements

We wish to thank Prof M.W. Fraaije for the HMFO gene and the pBAD plasmid, Gecco Biotech for providing purified HMFO used in this study and R. Leveson-Gower for synthetic assistance. This work was supported by the European Research Council (ERC advanced grant 885396), the Ministry of Education, Culture and Science (Gravitation programme no. 024.001.035) and The Netherlands Organisation for Scientific Research (NWO, project 724.013.003)

**Keywords:** biocatalysis • artificial enzyme • in vivo cascade• biosynthetic pathways • hydrazone formation

[1] J. Nielsen, J. D. Keasling, *Cell* **2016**, *164*, 1185–1197.

[2] M. Jeschek, S. Panke, T. R. Ward, *Trends Biotechnol.* **2018**, *36*, 60–72.

[3] R. V. Maaskant, S. Chordia, G. Roelfes, *ChemCatChem* **2021**, 1–8.

[4] S. Wu, Y. Zhou, D. Gerngross, M. Jeschek, T. R. Ward, *Nat. Commun.* **2019**, *10*, 1–10.

[5] G. Sirasani, L. Tong, E. P. Balskus, *Angew. Chem. Int. Ed.* **2014**, *53*, 7785–7788.

[6] S. Wallace, E. P. Balskus, *Angew. Chem. Int. Ed.* **2015**, *54*, 7106–7109.

[7] S. V. Sharma, X. Tong, C. Pubill-Ulldemolins, C. Cartmell, E. J. A. Bogosyan, E. J. Rackham, E. Marelli, R. B. Hamed, R. J. M. Goss, *Nat. Commun.* **2017**, *8*, 1–9.

[8] P. Destito, C. Vidal, F. López, J. L. Mascareñas, *Chem. Eur. J.* **2021**, *27*, 4789–4816.

[9] C. Adamson, M. Kanai, *Org. Biomol. Chem.* **2021**, *19*, 37–45.

[10] J. G. Rebelein, T. R. Ward, *Curr. Opin. Biotechnol.* **2018**, *53*, 106–114.

[11] F. Schwizer, Y. Okamoto, T. Heinisch, Y. Gu, M. M. Pellizzoni, V. Lebrun, R. Reuter, V. Köhler, J. C. Lewis, T. R. Ward, *Chem. Rev.* **2018**, *118*, 142–231.

[12] S. Chordia, S. Narasimhan, A. Lucini Paioni, M. Baldus, G. Roelfes, *Angew. Chem. Int. Ed.* **2021**, 2–10.

[13] W. Ghattas, V. Dubosclard, A. Wick, A. Bendelac, R. Guillot, R. Ricoux, J. P. Mahy, *J. Am. Chem. Soc.* **2018**, *140*, 8756–8762.

[14] A. R. Grimm, D. F. Sauer, T. Polen, L. Zhu, T. Hayashi, J. Okuda, U. Schwaneberg, *ACS Catal.* **2018**, *8*, 2611–2614.

[15] M. Jeschek, R. Reuter, T. Heinisch, C. Trindler, J. Klehr, S. Panke, T. R. Ward, *Nature* **2016**, *537*, 661–665.

[16] T. Heinisch, F. Schwizer, B. Garabedian, E. Csibra, M. Jeschek, J. Vallapurackal, V. B. Pinheiro, P. Marlière, S. Panke, T. R. Ward, *Chem. Sci.* **2018**, *9*, 5383–5388.

[17] M. Wittwer, U. Markel, J. Schiffels, J. Okuda, D. F. Sauer, U. Schwaneberg, *Nat. Catal.* **2021**, *4*, 814–827.

[18] M. A. S. Mertens, D. F. Sauer, U. Markel, J. Schiffels, J. Okuda, U. Schwaneberg, *Catal. Sci. Technol.* **2019**, *9*, 5572–5576.

[19] V. Köhler, Y. M. Wilson, M. Dürrenberger, D. Ghislieri, E. Churakova, T. Quinto, L. Knörr, D. Häussinger, F. Hollmann, N. J. Turner, T. R. Ward, *Nat. Chem.* **2013**, *5*, 93–99.

[20] J. Huang, Z. Liu, B. J. Bloomer, D. S. Clark, A. Mukhopadhyay, J. D. Keasling, J. F. Hartwig, *Nat. Chem.* **2021**, DOI 10.1038/s41557-021-00801-3.

[21] I. Drienovská, C. Mayer, C. Dulson, G. Roelfes, *Nat. Chem.* **2018**, *10*, 946–952.

[22] I. Drienovská, G. Roelfes, *Nat. Catal.* **2020**, *3*, 193–202.

[23] P. K. Madoori, H. Agustiandari, A. J. M. Driessen, A. M. W. H. Thunnissen, *EMBO J.* **2009**, *28*, 156–166.

[24] L. Wang, A. Brock, P. G. Schultz, *Science* **2001**, *292,* 498–500.

[25] R. A. Mehl, J. C. Anderson, S. W. Santoro, L. Wang, A. B. Martin, D. S. King, D. M. Horn, P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 935–939.

[26] A. Erkkila¨, I. Majander, P. M. Pihko, *Chem. Rev.* **2007**, *107*, 5416–5470.

[27] D. K. Kölmel, E. T. Kool, *Chem. Rev.* **2017**, *117*, 10358–10376.

[28] E. H. Cordes, W. P. Jencks, *J. Am. Chem. Soc.* **1962**, *84*, 826–831.

[29] C. Mayer, C. Dulson, E. Reddem, A. M. W. H. Thunnissen,

G. Roelfes, *Angew. Chem. Int. Ed.* **2019**, *58*, 2083–2087.
[30] R. B. Leveson-Gower, Z. Zhou, I. Drienovská, G. Roelfes, *ACS Catal.* **2021**, *11*, 6763–6770.
[31] Z. Zhou, G. Roelfes, *Nat. Catal.* **2020**, *3*, 289–294.
[32] M. Winkler, *Curr. Opin. Chem. Biol.* **2018**, *43*, 23–29.
[33] S. Wu, R. Snajdrova, J. C. Moore, K. Baldenius, U. T. Bornscheuer, *Angew. Chem. Int. Ed.* **2021**, *60*, 88–119.
[34] J. J. Dong, E. Fernández-Fueyo, F. Hollmann, C. E. Paul, M. Pesic, S. Schmidt, Y. Wang, S. Younes, W. Zhang, *Angew. Chem. Int. Ed.* **2018**, *57*, 9238–9261.
[35] A. M. Kunjapur, Y. Tarasova, K. L. J. Prather, *J. Am. Chem. Soc.* **2014**, *136*, 11644–11654.
[36] H. Stolterfoht, D. Schwendenwein, C. W. Sensen, F. Rudroff, M. Winkler, *J. Biotechnol.* **2017**, *257*, 222–232.
[37] P. Venkitasubramanian, L. Daniels, J. P. N. Rosazza, *J. Biol. Chem.* **2007**, *282*, 478–485.
[38] J. Ni, Y. Y. Gao, F. Tao, H. Y. Liu, P. Xu, *Angew. Chem. Int. Ed.* **2018**, *57*, 1214–1217.
[39] J. Overhage, H. Priefert, A. Steinbüchel, *Appl. Environ. Microbiol.* **1999**, *65*, 4837–4847.
[40] W. Yang, H. Tang, J. Ni, Q. Wu, D. Hua, F. Tao, P. Xu, *PLoS One* **2013**, *8*, e67339.
[41] J. P. Bennett, L. Bertin, B. Moulton, I. J. S. Fairlamb, A. M. Brzozowski, N. J. Walton, G. Grogan, *Biochem. J.* **2008**, *414*, 281–289.
[42] C. Martin, A. Ovalle Maqueo, H. J. Wijma, M. W. Fraaije, *Biotechnol. Biofuels* **2018**, *11*, 56.
[43] E. Romero, G. Gadda, *Biomol. Concepts* **2014**, *5*, 299–318.

Supporting information

# An In Vivo Biocatalytic Cascade featuring an artificial enzyme catalyzed new-to nature reaction.

Linda Ofori Atta, Zhi Zhou and Gerard Roelfes*

Table of Contents

## Supplementary tables

| ***E. coli* Strains** | **Plasmids** |
|---|---|
| K-12 MG1655 (RARE) for expression | pETDuet-1 harbouring *CAR-SFP* |
| NEB 5 alpha for cloning | pETDuet-1 harbouring *FCS-ECH* |
| NEB 10 beta for cloning | pET28b+ harbouring LmrR-V15pAF_RMH |
| | pBAD harbouring HMFO and LmrR-V15pAF_RMH |
| | pDULE 2 harbouring pAF OTS |

**Table S1.** Bacteria strains and plasmid used in this study.

| **Protein Sample** | **Expression plasmid/OTS** | **$k_{cat}/K_m$ (app) $M^{-1}/s^{-1}$** |
|---|---|---|
| LmrR_V15pAF_RMH | pEVOL_pAzF | 70 |
| LmrR_V15pAF_RMH | pDULE 2 pAF | 54 |

**Table S2.** Comparison of catalytic efficiency in the hydrazone formation reaction of **1a** with **2** to form **3**, by artificial enzymes produced using direct incorporation of pAF and indirect incorporation of pAF into LmrR_V15X RMH using the hydrazone formation reaction.[2] Based on the relative activitiy it is concluded that the pAF incorporation efficiency is ~80% when using the direct incorporation method. **$k_{cat}/K_m$ (app) $M^{-1}/S^{-1}$** = catalytic efficiency **Extinction Coefficient (ε)** =29,585 $M^{-1}$ $cm^{-1}$ .

## Supplementary figure

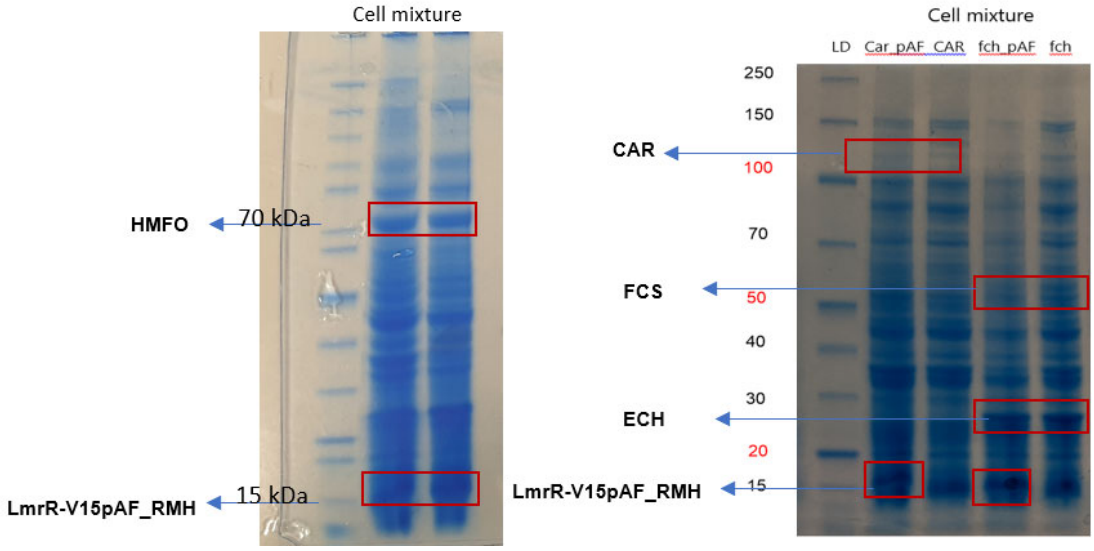



**Figure S1.** Expression of LmrR_V15pAF_RMH, CAR, FCS-ECH and HMFO after integration of the plasmids in E. coli RARE. CAR showed a ow expression, most likely caused by its large gene size of 4000 bp, making it difficult for RNA polymerase to synthesise the protein.

## Methods and Materials

**Strains and Plasmids:** *E. coli* strains and plasmids used in this study are listed in Table S1. Molecular biology techniques were performed using standard practices unless otherwise stated. Molecular cloning and vector propagation were performed in K-12 MG1655**.** The *CAR* gene from *Nocardia iowensis* and *SFP* gene from *Bacillus subtilis*, were synthesized and codon-optimized for expression in E. coli (GenScript USA). The pDULE plasmid harbouring pAF synthase\tRNA genes was purchased from addgene (pDule2-para-aminoPhe was a gift from Ryan Mehl, Addgene plasmid #85503)[1], the pBad plasmid harbouring *HMFO* was a gift from the Fraaije group (University of Groningen). Purified HMFO used in this study was a gift from Gecco Biotech (www.gecco-biotech.com).

**Cell culture conditions:** A standard recipe for Lysogenic Broth (LB) medium was used (10 g/L Bacto tryptone, 5 g/L Yeast extract, 10 g/L NaCl and 15 g/L Agar for LB-Agar plates). MMV medium was prepared with the following recipe, 5.3 g/L $Na_2HPO_4$ × 12 $H_2O$, 1.4 g/L $KH_2PO_4$, 0.2 g/L $MgSO_4$ × 7 $H_2O$, 1.0 g/L $(NH_4)_2SO_4$. The following components were sterile-filtered using 0.2 µm filter before adding the medium: 1 ml/L vitamin solution (1000 ×), 5 ml/L trace solution (200 ×), 4 ml/L glucose (50%). M9 medium was prepared using the following recipe, 56 g/L of M9 salt (mix by Sigma) and a 0.2 µm filter was used to filter the following component before addition 100 µl of 1 M $CaCl_2$, 2 mL of 1 M $MgSO_4$, 8 mL of 50% glucose solution, 5 mL of 4% vitamin B1 (thiamine), 50 mL of 20% Cas-amino acids.

## Protein expression

For experiments testing the direct incorporation of pAF into LmrRV15TAG, pDULE vector harboring pAF synthase\tRNA and pET17b+ vector harboring LmrRV15TAG were co-transformed into *E. coli* k-12 MG1655 (RARE). Culture media were supplemented with 100 µg/L ampicillin and 10 µg/L streptomycin to provide selective pressure for plasmid maintenance. Except for cultures grown in LB medium, overnight pre-cultures were pelleted and re-suspended in either M9 medium or MMV medium. This step was done to ensure that nutrient rich LB medium was removed. Experimental cultures were then initiated by adding 2 % (v/v) of re-suspended pre-cultures into 50 ml either M9 or MMV medium while 1 % (v/v) pre-culture was added to LB medium. Gene expression was then induced with 1 M isopropyl-β-d-1-thiogalactopyranoside (IPTG) (1 mM final concentration) and 20 % L- arabinose (0.02 % final concentration) between an $OD_{600}$ of 0.4 and 1. During induction, 1 M pAF (1 mM final concentration) was added in solid form, incubated at 24 °C, and agitated at 135 rpm for 48 hrs. These experiments were performed in 250 ml baffled Erlenmeyer flasks for better aeration.

| Media/Sample | Strain<br>*E.Coli* K-12 MG1655 | pAF | Temperature | Time |
|---|---|---|---|---|
| LB | RARE | 1mM | 30 °C | 16 h |
| LB control | RARE | No pAF added | 30 °C | 16 h |
| M9 | RARE | 1 mM | 30 °C | 16 h |
| M9 control | RARE | No pAF added | 30 °C | 16 h |
| MMV | RARE | 1 mM | 24 °C | 48 hours |
| MMV control | RARE | No pAF added | 24 °C | 48 hours |

**Table S3.** Expression conditions for pAF incorporation into LmrR_V15X_RMH.

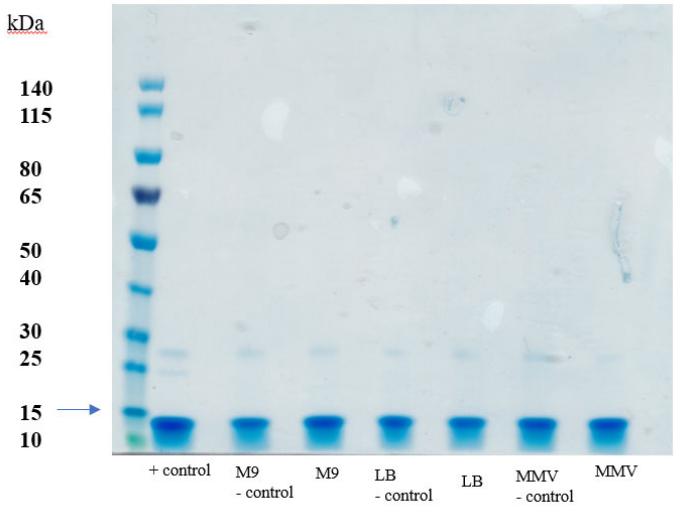



**Figure S2.** SDS-PAGE analysis of pAF incorporation into LmrR-V15X_RMH. Controls without pAF added do yield protein, which is caused by misincorporation of canonical amino acids due to the promiscuity of the orthogonal translation system.

## Metabolites analysis:

For aromatic aldehydes synthesis, experiments were performed in 250/500 ml Erlenmeyer flasks (Sigma), using 50 mL MMV, supplemented with 100 µg ampicillin. Experimental cultures were initiated by adding 1 % (v/v) inoculum volumes of overnight pre-culture to 50 ml volume of MMV, incubated at 37 °C, and agitated at 135 rpm.

Expression was induced with 1 mM isopropyl-β-d-1-thiogalactopyranoside (IPTG) or 0.02 % arabinose between optical density 600 ($OD_{600}$) of 0.8−1.0. The cultures were incubated at 24 °C and 135 rpm for 16 h. Cultures were spun down at 4000 rpm for 10 mins, pellets were resuspended with fresh MMV and 5 mM of substrate was added at 24 °C and 135 rpm for 16 h. For benzaldehyde producing cultures, samples were spun down and filtered with a 0.2 µM filter before HPLC analysis. In cultures producing para hydroxy benzaldehyde, the products were extracted with ethyl acetate.

HPLC analysis using an Agilent series instrument equipped with a diode array detector was used to detect products at wavelengths of 247 nm and 281 nm for benzaldehyde and para-hydroxy benzaldehyde, respectively. A kinetex-C18 column was used and the gradient method used the following solvents:

(A1) acetonitrile + 0.1% trifluoroacetic acid (TFA); (B1) water +0.1% TFA. The gradient started with 5% Solvent A1 and 95% Solvent B1 with a flow rate of 1.0 mL/min, and all compounds of interest eluted within 30 mins. The column temperature was maintained at 25°C.

| Gene | Strain *E.Coli* K-12 MG1655 | Substrate (5 mM ) | Temperature | Time (hours) |
|---|---|---|---|---|
| *CAR_SFP* | RARE | Benzoic Acid | 24°C | 16 h / 48 h |
| *CAR_SFP* | RARE | 4 hydroxy-benzoic Acid | 24 °C | 16 h / 48 h |
| *FCS-ECH* | RARE | 4-coumaric acid | 24 °C | 16 h / 48 h |
| *HMFO* | RARE | benzyl alcohol | 24 °C | 16 h / 48 h |
| *HMFO* | RARE | 4 hydroxy-benzyl alcohol | 24 °C | 16 h / 48 h |

**Table S4.** Expression conditions for *in vivo* aldehydes production using *E.coli* K12 MG1655 (RARE).

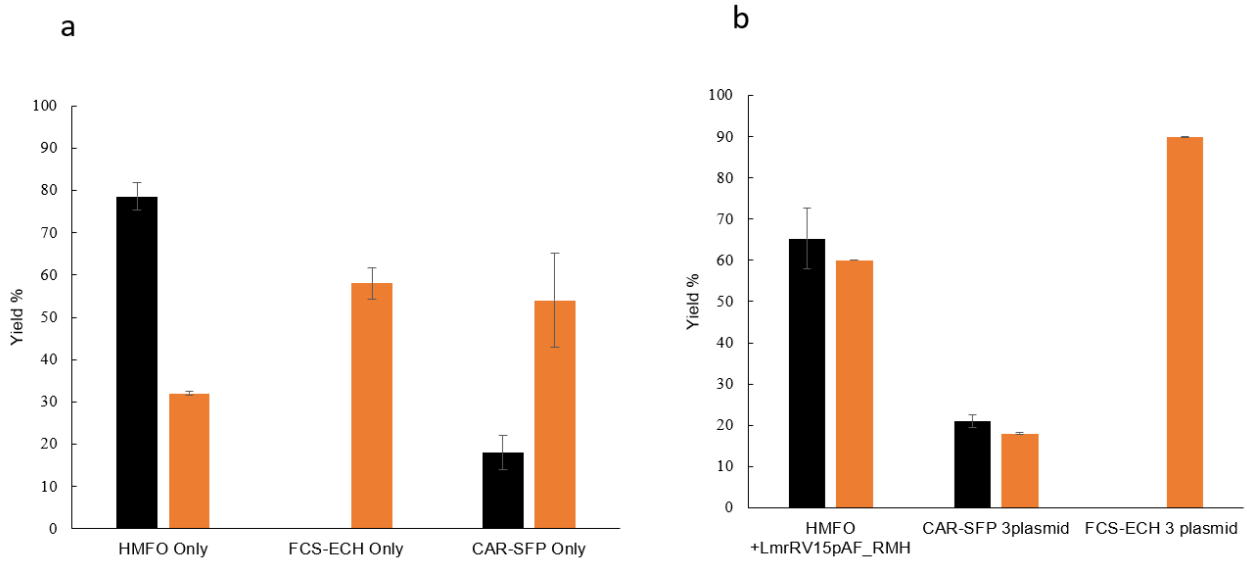



**Figure S3.** *In vivo* production of aldehydes before and after the integration of natural enzymes with LmrR-V15pAF_RMH in *E.coli* K12 MG1655 (RARE). **(a)** Benzaldehyde (black) was produced by HMFO and CAR-SFP and para hydroxy benzaldehyde (orange) was produced by HMFO, CAR-SFP and FCS- ECH. Expression was done at 24℃/135 rpm for 24 h. (**b**) Aldehydes production using 3 plasmid system (CAR-SFP, FCS=ECH) and 2 plasmid system (HMFO), after the integration of natural enzymes with artificial enzyme (LmrR-V15pAF_RMH). Expression was done at 24℃/135 rpm for 48 h.

## HPLC Chromatogram for aldehydes

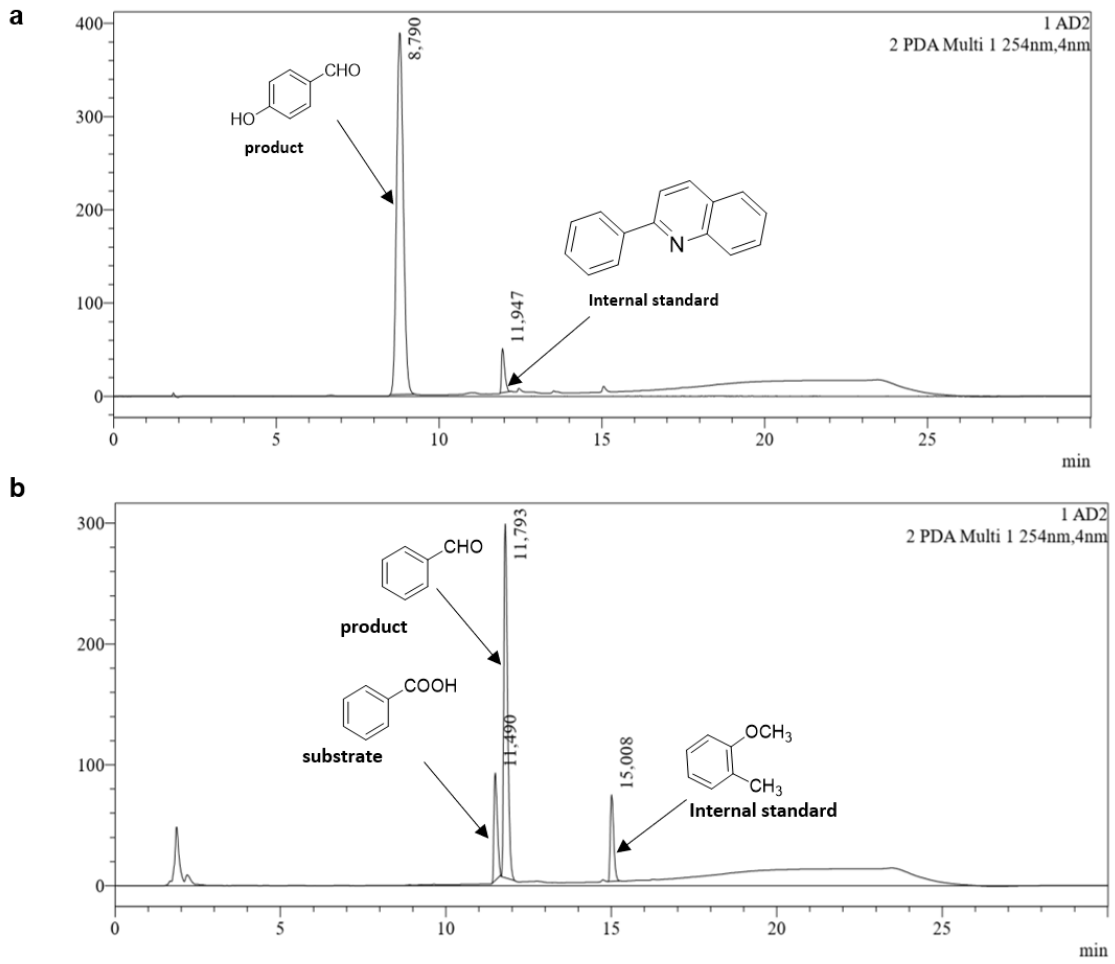



**Figure S4.** HPLC chromatogram of aldehydes biosynthesised in *E.coli* K12 MG1655 (RARE). (**a**) p-OH benzaldehyde product **3a** peak detected at 8,7 min while the internal standard 2-Phenylquinoline had a retention time around 11,9 min. (**b**) Benzaldehyde product **3b** peak detected at 11,7 min while the benzoic acid substrate peak had a retention time of 11,4 mins. Here 2-methylanisole was used as an internal standard.

## Protein Purification:

Sodium phosphate buffer (in milliQ water) was prepared with the following recipe: 50 mM of $Na_2HPO_4$; 12 $H_2O$, 150 mM of NaCl at pH 8.The buffer was filtered with 0.2 µm sterile filters. After 48 hours of expression at 24 °C, cells were harvested by centrifugation (6,000 rpm., JA10, Beckman) for 20 min, at 4 °C. The cell pellet was resuspended in 20 ml Sodium phosphate buffer and half a tablet “mini complete protease inhibitor cocktail” (Roche) was added. The solutions were sonicated on ice (8 min, pulse of 10 s ON and 15 s OFF, amplitude of 70 %).

The lysed cells were incubated on ice with 0.1 mg /ml of DNase I and 10 mM of $MgCl_2$ for 30 min. The mixture was spun down with rotor JA17(Beckman), 12,000 rpm, for 60 min, 4 °C. The supernatant was filtered using a 0.45 µm filter and 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. Eluents were collected and dialysed against the reaction buffer (50 mM $NaH_2PO_4$, 150 mM NaCl, pH 7.4, two times against 1 L). The concentration of the protein was determined by using the calculated extinction coefficient for LmrR, corrected for the absorbance of pAF (ε280 = 1,333 $M^{-1}$ $cm^{-1}$).

**Trypsin digest**

Protein for trypsin digest was initially purified by Streptag affinity chromatography and desalted with dialysis. 4 µl of the LmrRv15pAF_RMH protein solution was diluted to 20 µl using a 100 mM ammonium bicarbonate (ABC) solution. 20 µl of a 10 ng/µl trypsin solution in 100 mM ABC was added and incubated overnight at 37 °C. The digestion mixture was acidified by adding 4 µl formic acid and desalted using a C18 stage tip. The eluted peptides were dried in a speedvac and dissolved in 20 µl 0.1% formic acid and injected into the nanoLC system for LCMSMS analysis. Resulting data were analysed using PeaksX software (BioSolutions Inc., Waterloo) Ontario, Canada).

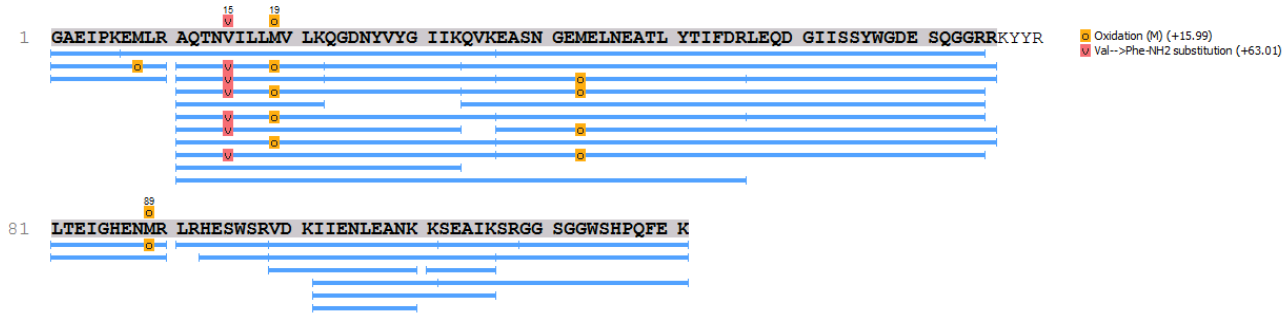



**Figure S5.** Trypsin digest of LmrR-V15pAF_RMH expressed in MMV.

## Mass Spectrometry Analysis

The mass spectrometry results showed protein expressed in M9 media had phenylalanine misincorporated and protein expressed in LB had a mixture of pAF and phenylalanine incorporated. Protein produced using the MMV media had pAF incorporated with high selectivity..

### Expected Mass = 15126 ; Observed Mass = 15126.64000

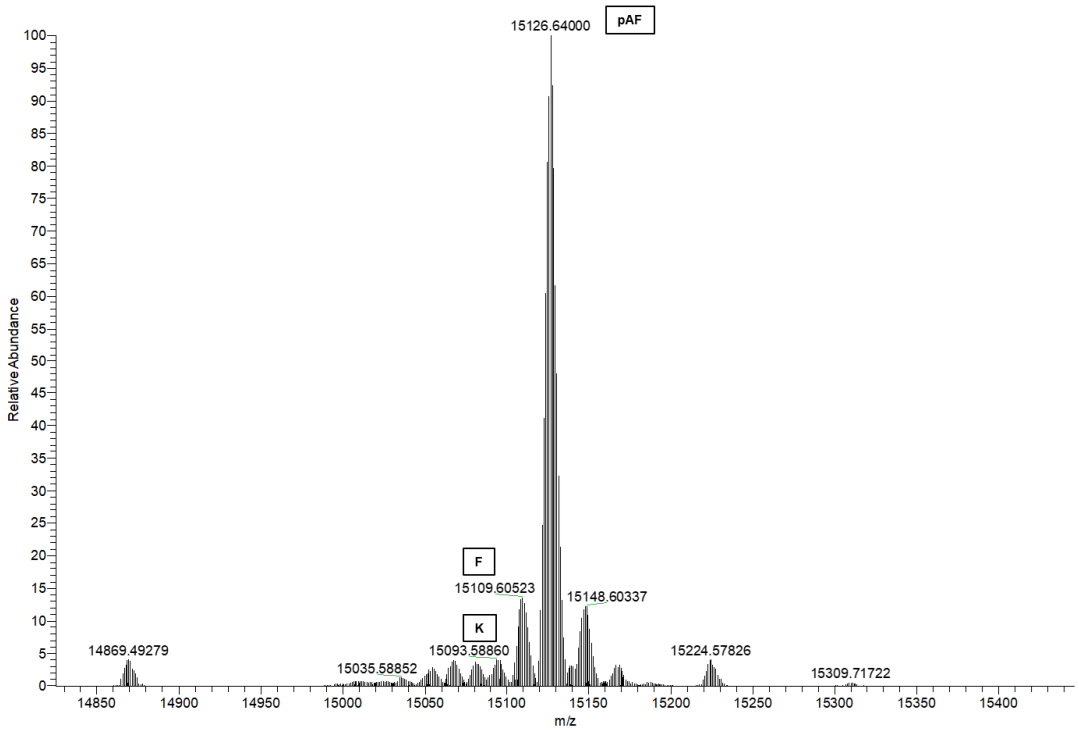



**Figure S6.** Mass of LmrR-V15pAF_RMH expressed in MMV. After deconvolution, an observed mass of 15126.64, which corresponds to the expected mass of LmrR-V15pAF_RMH, was detected. Although expression in this media gives some misincorporation of phenylalanine (F) and Lysine (K), around 80 % of incorporated amino acid was pAF (unnatural amino acid).

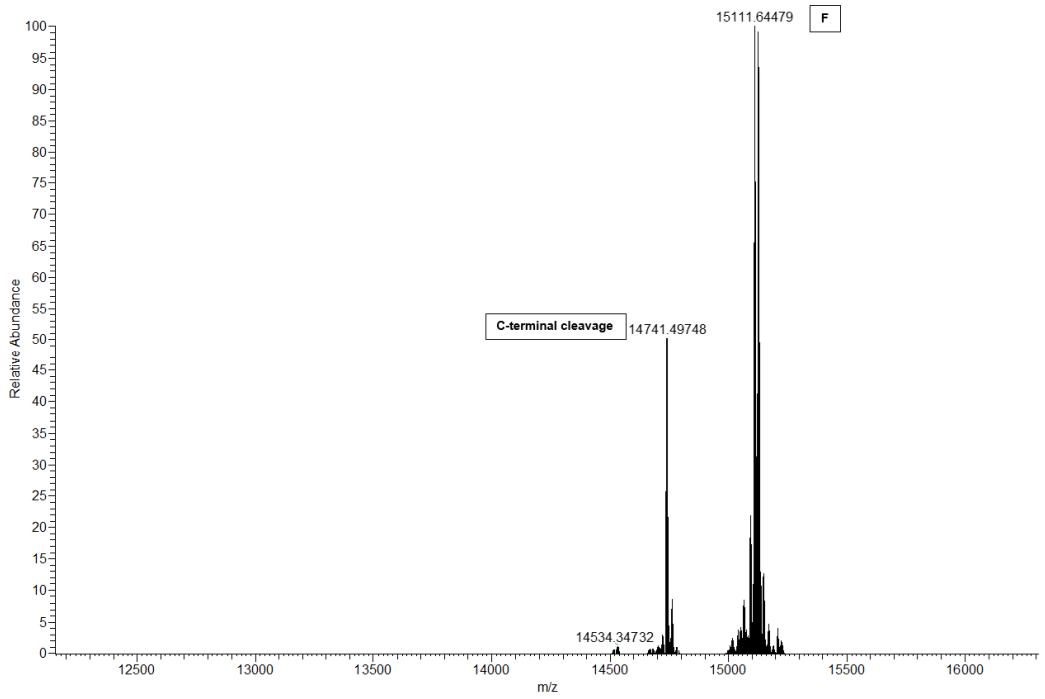



**Figure S7.** Mass of LmrR-V15X_RMH expressed in M9 media. After deconvolution, an observed mass of 15111.64 which corresponds to the expected mass of LmrR-V15X_RMH when phenylalanine (F) is incorporation instead of pAF. A mass corresponding to the cleavage of 4 amino acids from the C-terminal was also detected.

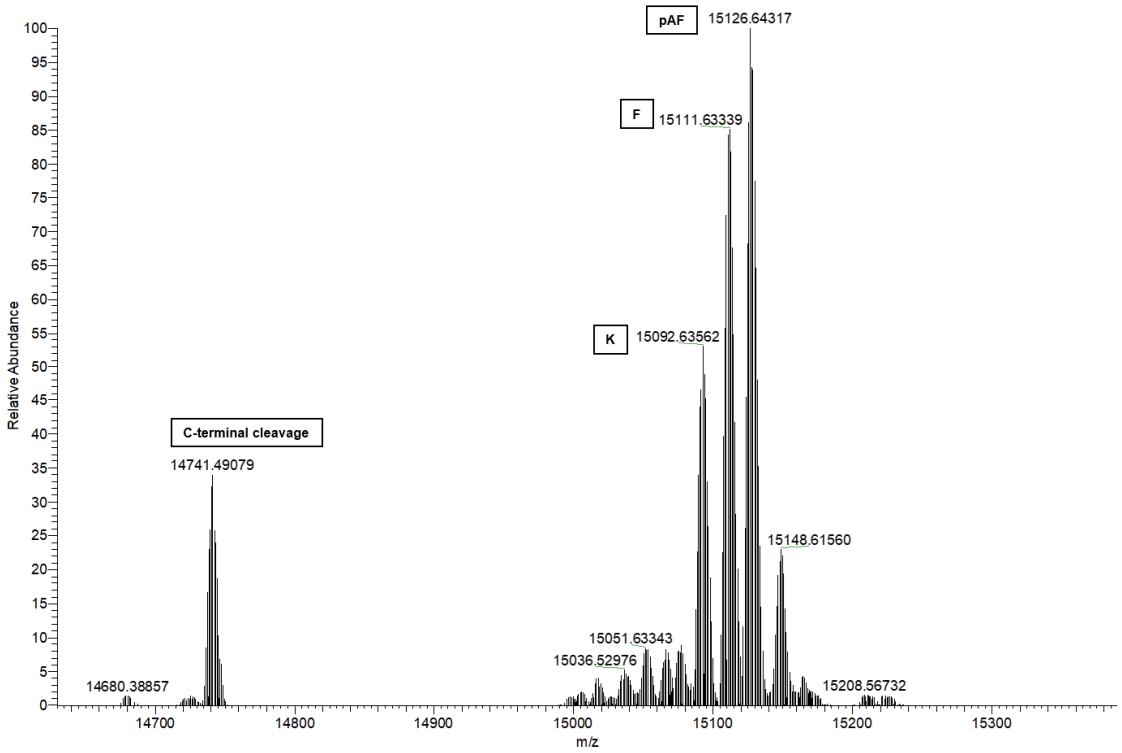



**Figure S8.** Mass of LmrR-V15X_RMH expressed in LB media. Although after deconvolution an observed mass of 15126.64 which corresponds to the expected mass of LmrR-V15pAF_RMH was detected, there was also a lot of misincorporation of phenylalanine (F) and Lysine (K). The presence of C-terminal cleavage was also detected as it was with M9.

## Uv-vis spectroscopy

After hydrazone formation reaction with para hydroxy benzaldehyde (**1a**) and NBDH (**2**), the absorbance of 1 ml of product **3a** solution was measured.

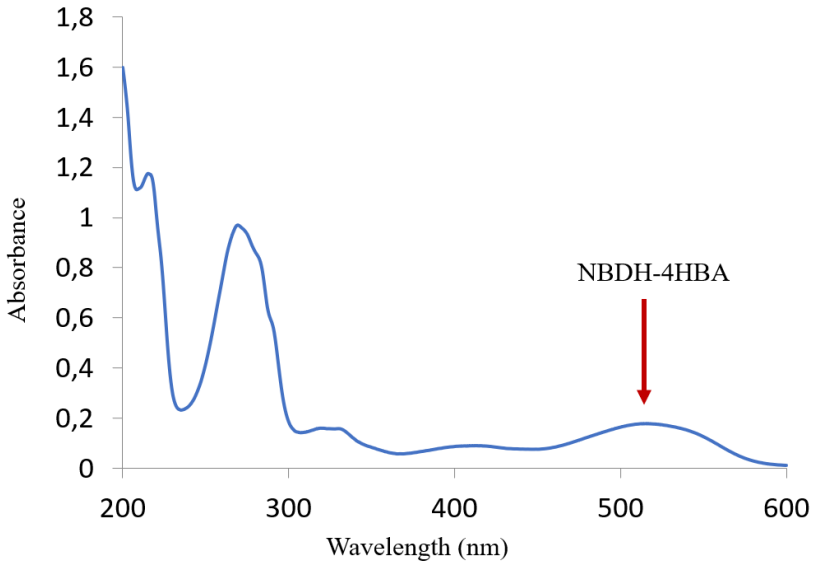



**Figure S9.** Uv-Vis spectra analysis of product **3a**. The characteristic absorption band of the NBDH-4HBA product (**3a**) was detected at a wavelength of 520 nm.

## UPLC-MS Analysis

Reactions samples were pelleted by centrifugation, and aqueous supernatant was collected and filtered with 2 µm filter for UPLC-MS analysis using ACQUITY UPLC® HSS T3 1.8µm. The solvents used were (A) acetonitrile + 0.1% formic acid (TFA); (B) water +0.1% formic acid Run for 20 mins.

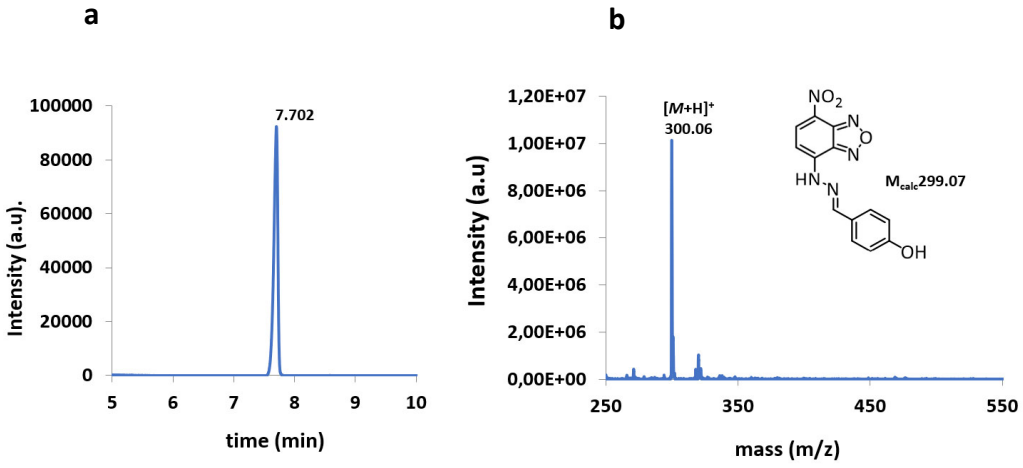



**Figure S10.** UPLC-MS analysis to confirm the mass of hydrazone product**. (a)** single peak at time 7.7 min corresponded to the retention time of hydrazone on the UPLC. (**b**)The mass of the calculated hydrazone was confirmed.

## Hydrazone formation reaction *in vivo* / *in vitro*

**Hydrazone formation reaction in vivo**: Experimental culture cells were grown in 50 ml MMV supplemented with antibiotics at 37°C, 135 rpm until $OD_{600}$ of 0.4 – 0.7. Gene expression was induced with 1 mM isopropyl-β-d-1-thiogalactopyranoside (IPTG) and 0.02 % L- arabinose. During induction 1 mM pAF was added in solid form, incubated at 24 °C, and agitated at 135 rpm for 48 hrs. Cultures were pelleted and resuspended in newly prepared MMV, to wash away non-incorporated UAA, or in reaction buffer. In the 2.5 mL cell culture with newly prepared MMV medium was added 25 μL NBDH (**2**) (5 mM in DMF) solution (final concentration 50 μM) and 62.5 μL carbonic acid or aldehyde (200 mM in DMF) solution (final concentration 5 mM). Then the mixture was put into the 24 °C shaker at 135 rpm for 1 to 24 hours. The formation of the product was determined by HPLC. Take 500 μL reaction culture into 2 ml Eppendorf, added 50 μL IS (1 mM Fluorescein in DMF), then extracted with 1ml ethyl acetate, after the evaporation, the residue was resuspended in 1200 μL $CH_3CN$. 200 μL of this solution was placed into HPLC vial and the sample was analyzed by HPLC.

**Hydrazone formation reaction in vitro was set-up as follows**: To 280 μL of 5 μM HMFO and 4 μM LmrR_V15pAF_RMH in freshly prepared phosphate buffer pH 6.5 was added 10 μL NBDH (1.5 mM in DMF) solution (final concentration 50 μM) and 10 μL carbonic acid (150 mM in DMF) solution (final concentration 5 mM). Then the mixture was put into the 24 °C shaker at 135 rpm for 1 to 3 hours. The yield of the products was determined by HPLC. To the reaction mixture was added 50 μL internal standard (1 mM Fluorescein in DMF), then it was extracted with 200 μL ethyl acetate. After the evaporation, the residue was resuspended in 240 μL $CH_3CN$. 200 μL of this solution was placed into HPLC vial and the sample was analyzed by HPLC.

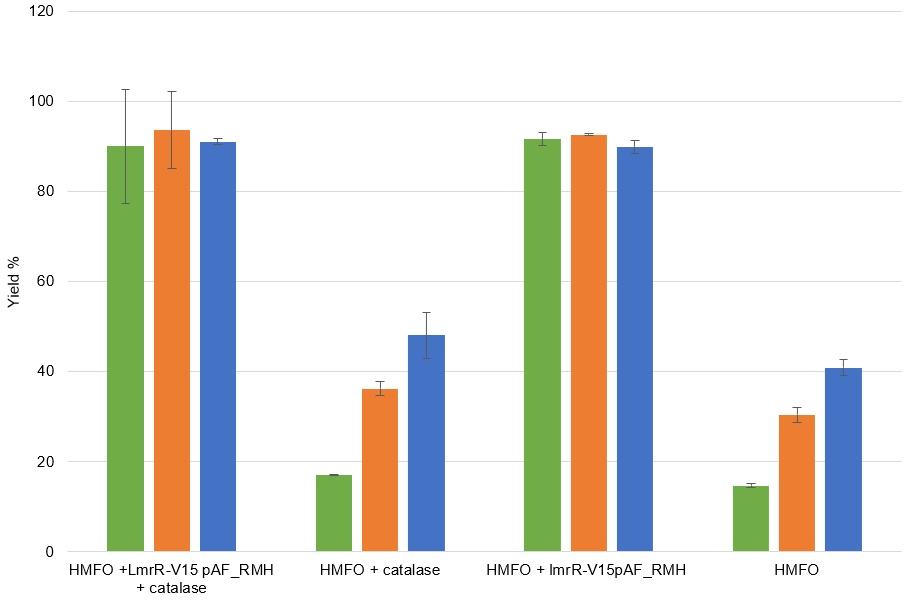



**Figure S11.** Hydrazone formation by the combination of HMFO and LmrR_V15pAF_RMH *in vitro*.
The in vitro reaction using the combination of HMFO and LmrR-V15pAF_RMH gave 85% yield after 1 h (green). Monitoring at 2 h (orange) and 3 h (blue) didn't give rise to further increased product formation. The control experiments with catalase showed that $H_2O_2$ has no effect on the cascade reactions.

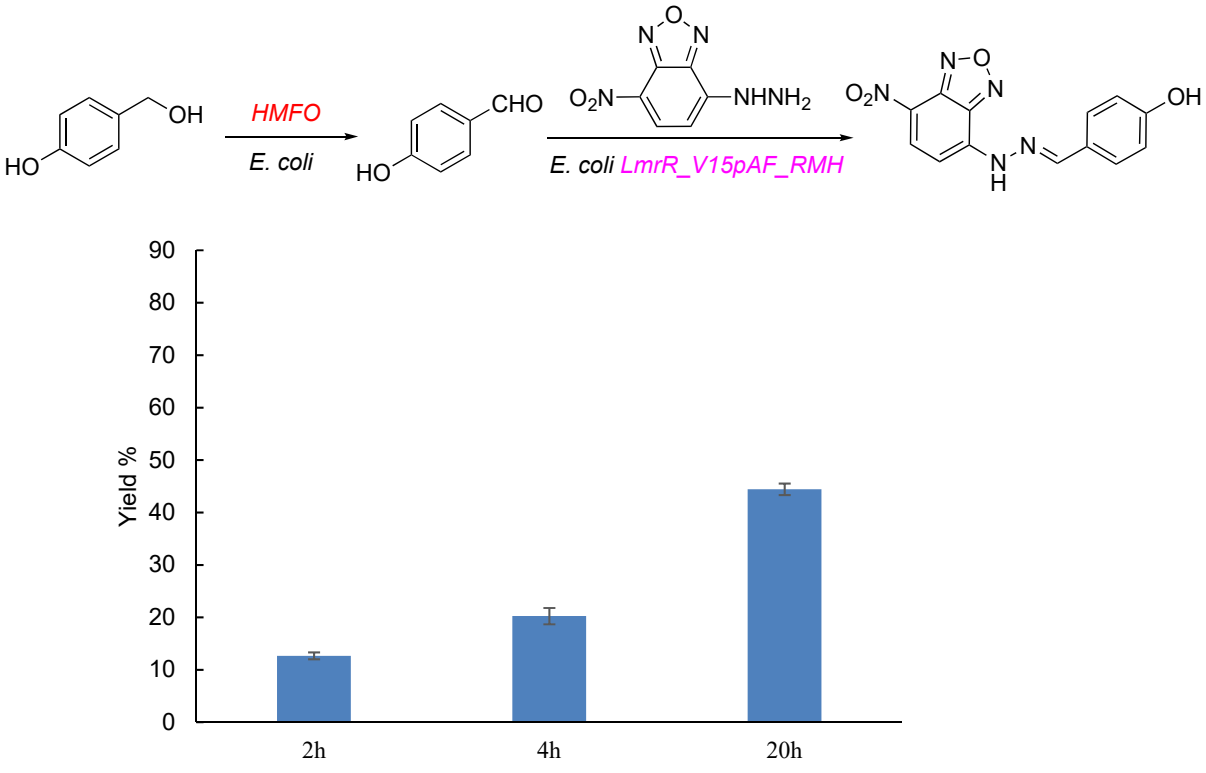



**Figure S12.** Hydrazone formation reaction of **1a** and **2** by integration of HMFO and LmrR_V15pAF_RMH *in vivo***.** The reaction was followed during 20 h with a final yield of 50 % of product **3a** formed.

## HPLC Chromatogram for hydrazones

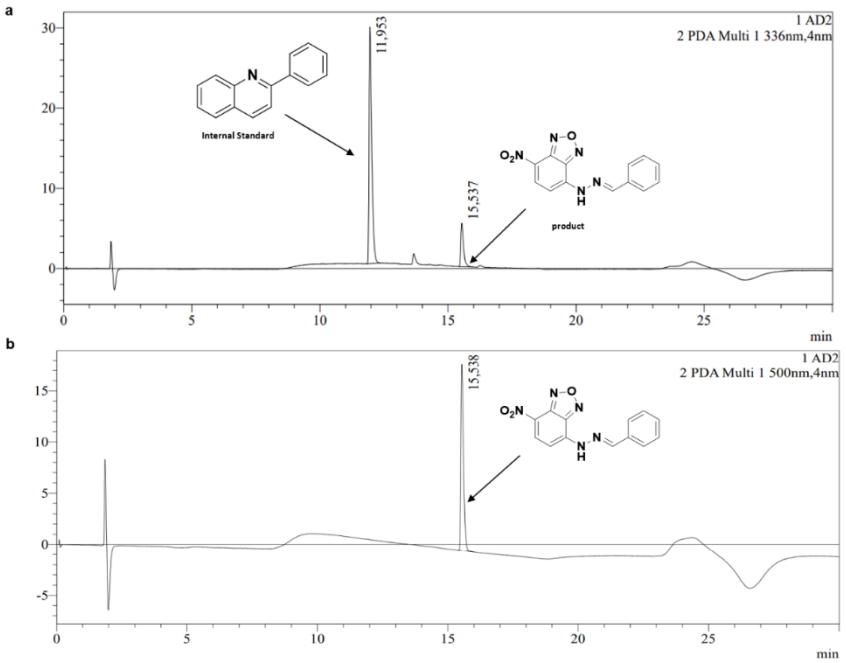



**Figure S13.** HPLC chromatogram of Hydrazone product **3b** formed by integration of HMFO and LmrR_V15pAF_RMH *in vivo*. (**a**) At wavelength 336 nm. (**b**) At wavelength 500 nm.

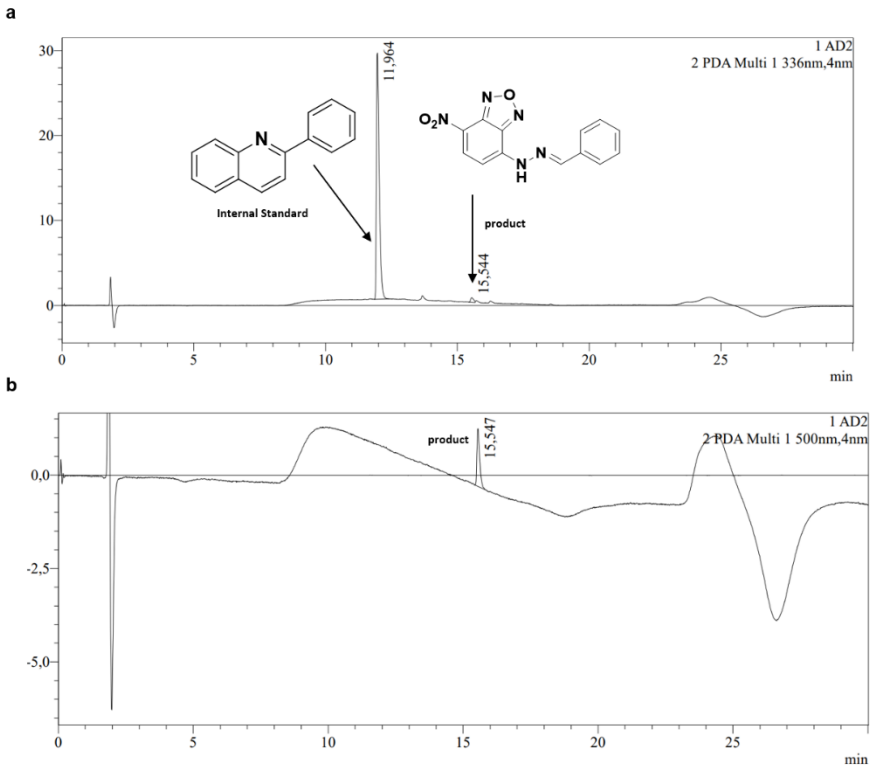



**Figure S14.** HPLC chromatogram of Hydrazone product **3b** in the presence only HMFO *in vivo* (i.e. in absence of LmrR_V15pAF_RMH). (**a**) At wavelength 336 nm. (**b**) At wavelength is 500 nm.

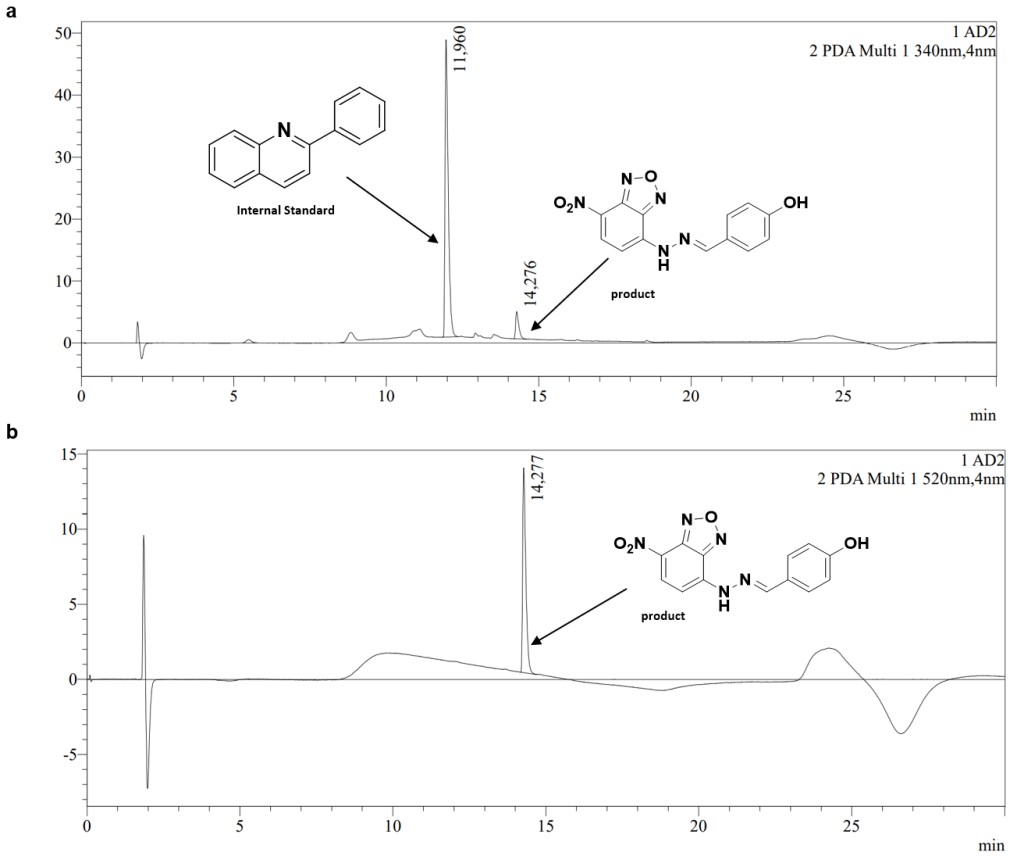
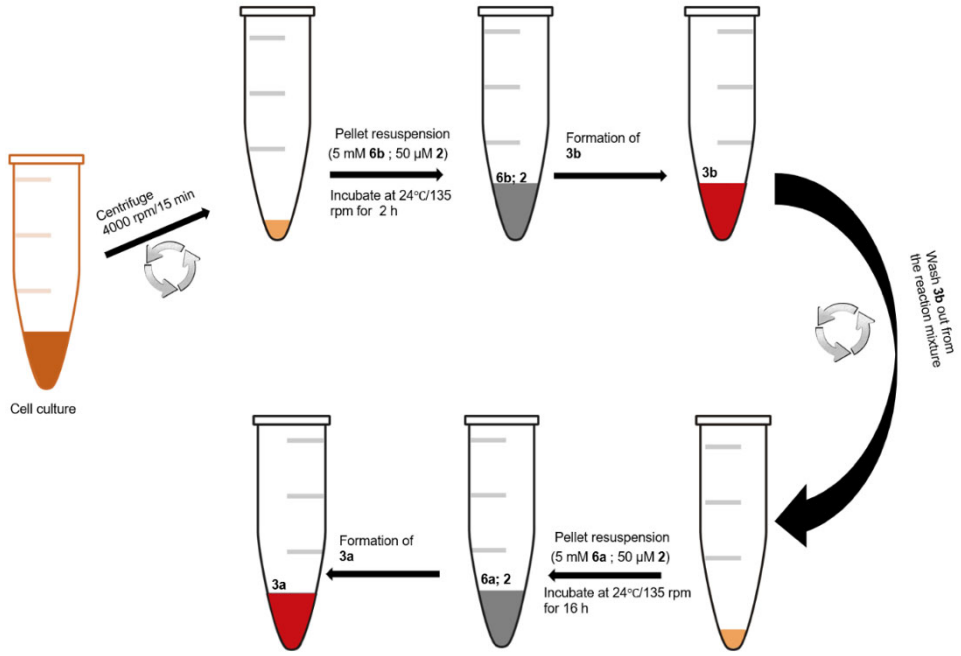



**Figure S15.** HPLC chromatogram of Hydrazone product **3a** formed by integration of HMFO and LmrR_V15pAF_RMH *in vivo*. (**a**) At wavelength 340 nm. (**b**) At wavelength is 520 nm.

## Recycling Experiment

**a**




**b**

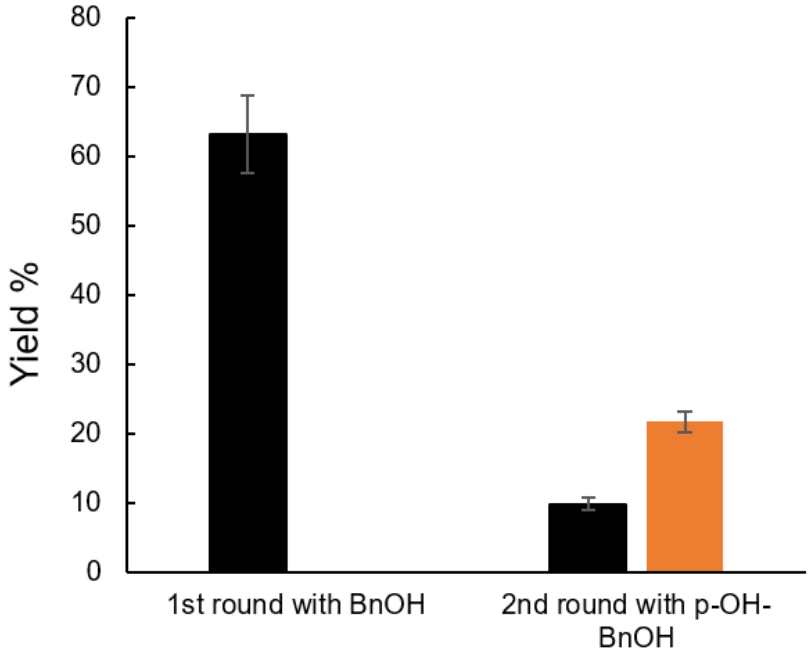



**Figure S16.** Recycling experiment to establish that catalysis takes place in the cell and the cell remains intact during and after catalysis (**a**) Schematic workflow of an *in vivo* recycling experiment of hydrazone formation reaction using HMFO/LmrR_V15pAF _RMH. Products (**3a,b**) were analyzed with the HPLC after extraction. (**b**) Product yields detected after each catalysis experiment, first round of reaction resulted in 63 % of **3b** (black) and the second round yielded 22 % of **3a** (orange) with 10 % residual **3a**.

**Cell viability test after catalysis**

After the *in vivo* catalysis, reaction mixtures were spun down at 8000 rpm for 30 mins and resuspended in LB medium. Serial dilutions (10x - $10^{-8}$) of selected samples were plated on LB plates containing 100 µg/mL ampicillin, 10 µg/mL spectinomycin and 50 µg/mL Kanamycin for 3 plasmid system and plates containing 100 µg/mL ampicillin and 10 µg/mL spectinomycin for 2 plasmid system. Plates were then incubated at 37 °C overnight. The following day, colonies were counted, plates with 10x diluted were confluent and therefore not countable. However plates with dilutions from $10^{-5}$ were countable.

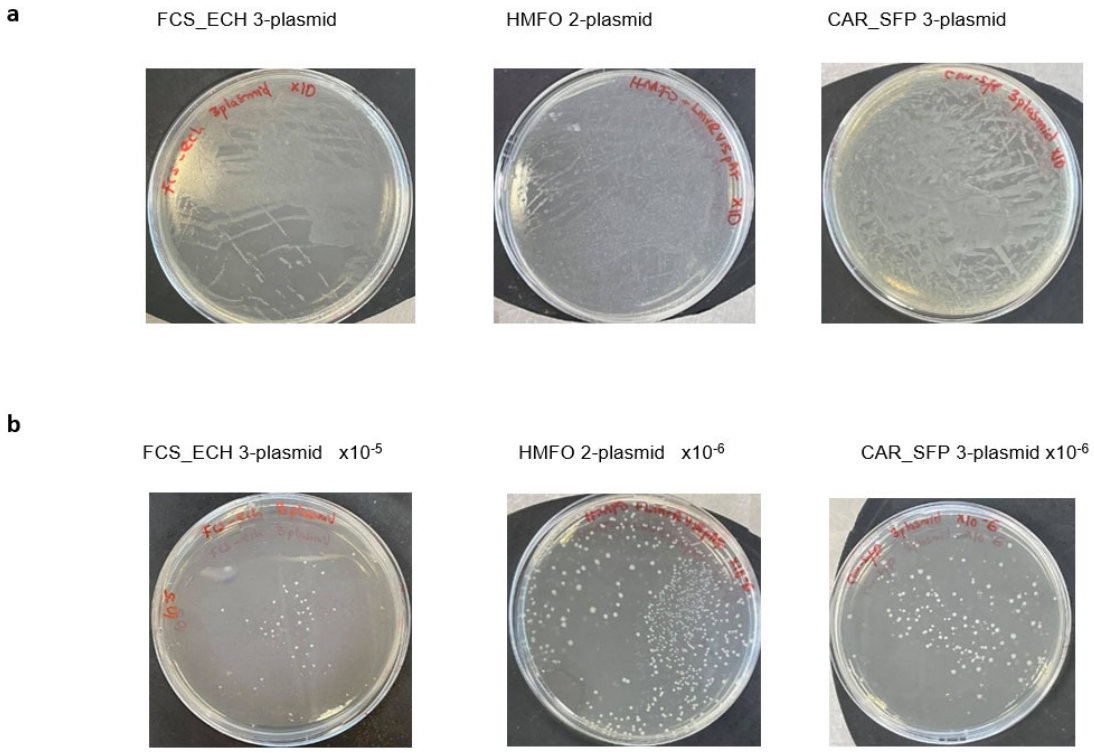



**Figure S17.** Cell viability test after hydrazone formation reaction. (**a**) Image of the confluency of colonies after 10 times dilution. (**b**) Plate with FCS_ECH 3-plasmid had 45 colonies, HMFO 2-plasmid had 336 colonies and CAR_SFP 3-plasmid had 141 colonies after several serial dilutions.

After the viability test, LmrR_V15pAF_RMH/HMFO system was used to repeat the cell viability test and this time with 2 different substrates. After protein expression at 24℃/135

rpm for 48 hours, cultures were collected and spun down at 4000 rpm for 10 mins. The pellet was resuspended in 50 mM potassium phosphate buffer at pH 6.5, reaction was performed by adding 5 mM of benzylalcohol and 50 µM of NBD-H to the cell and incubated at 24℃/135 rpm for 2 hours. The reaction mixtures were then spun down and the cells resuspended in LB media, serial dilutions were done down to $10^{-6}$ and 200 µl was plated on LB agar each dilution and incubated at 37℃ overnight. A single colony from the $10^{-6}$ dilution plate was picked and grown in LB media while expressing LmrR_V15pAF_RMH/HMFO. Catalysis was performed as previously described with the new cells, now using 5 mM p-hydroxybenzyl alcohol and 50 µM NBDH as substrates in 50 mM potassium phosphate buffer at pH 6.5. After overnight reaction at 24℃/135 rpm, the samples were again spun down, resuspended in LB and serial dilutions down to $10^{-9}$ were plated on agar. Colonies on the plates with $10^{-8}$ and $10^{-9}$ dilutions yielded 419 and 60 colonies respectively.

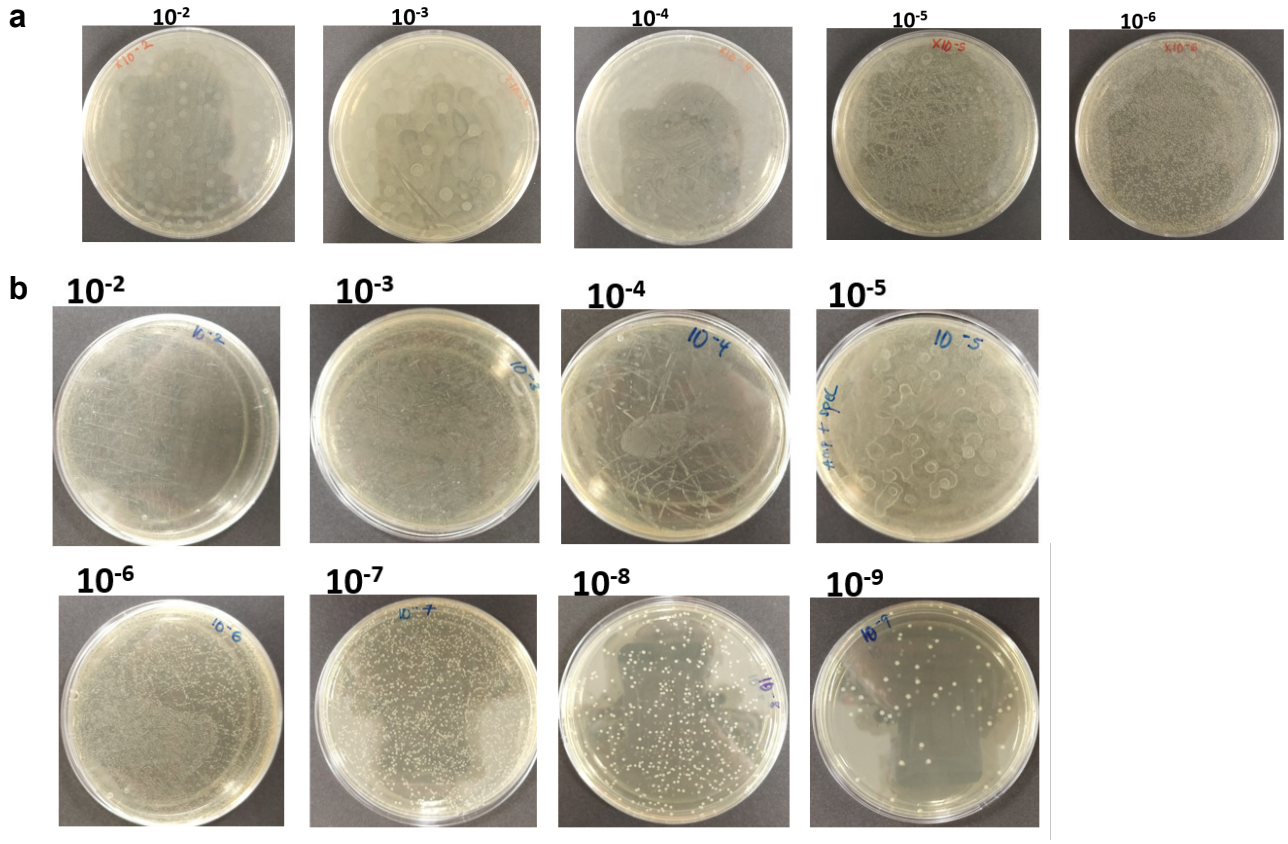



**Figure S18.** Agar plates with colonies after viability test with LmrR_V15pAF_RMH/HMFO system (**a**) images of colony plates after the first round of catalysis with 5 mM of benzylalcohol and 50 µM of NBD-H for 2 hours. (**b**)Colony plates after second round of catalysis 5 mM p-hydroxybenzyl alcohol and 50 uM NBDH for overnight. Colonies counted on the $10^{-8}$ and $10^{-9}$ plates were estimated to be $6\times10^{7}$ cells/ml for plate $10^{-8}$ and $2\times10^{7}$ cells/ml for plate $10^{-9}$ viable cells after catalysis, compared to $1.92\times10^{10}$ cells/ml at the onset.

## Gene sequence used in the work

***LmrR-V15TAG RMH***

ATGGGTGCCGAAATCCCGAAAGAAATGCTGCGTGCTCAAACCAATTAGATCCTGCTGATGGTCCTGAAACAA
GGCGATAACTATGTGTATGGCATTATCAAACAGGTGAAAGAAGCGAGCAACGGTGAAATGGAACTGAATGAA
GCCACCCTGTATACGATTTTTGATCGTCTGGAACAGGACGGCATTATCAGCTCTTACTGGGGTGATGAAAGT
CAAGGCGGTCGTCGCAAATATTACCGTCTGACCGAAATCGGCCATGAAAACATGCGCCTGCGGCATGAATC
CTGGAGTCGTGTGGACAAAATCATTGAAAATCTGGAAGCAAACAAAAAATCTGAAGCGATCAAATCTAGAGG
TGGCAGCGGTGGCTGGAGCCACCCGCAGTTCGAAAAATAA

***HMFO***

ATGACTGATACGATTTTTGACTACGTGATTGTTGGCGGTGGCACGGCGGGTAGCGTTCTGGCCAACCGTCT
GTCCGCCCGTCCGGAGAATCGCGTGTTGCTGATTGAGGCCGGTATTGATACCCCGGAAAACAATATTCCGC
CGGAGATCCACGATGGCCTGCGCCCGTGGCTGCCGCGTCTGAGCGGTGATAAGTTCTTTTGGCCGAATCT
GACCGTCTACCGTGCCGCGGAACACCCGGGTATCACGCGCGAGCCGCAGTTCTATGAACAAGGCCGTCTG
CTGGGCGGTGGTAGCAGCGTGAACATGGTCGTTTCTAACCGTGGTCTGCCTCGCGACTATGACGAATGGCA
GGCACTGGGCGCAGATGGTTGGGATTGGCAGGGTGTTCTGCCGTACTTCATCAAGACCGAGCGTGACGCG
GACTACGGTGACGACCCGTTGCATGGCAATGCGGGTCCGATTCCGATCGGTCGCGTCGATTCGCGTCACT
GGAGCGACTTCACGGTGGCGGCAACCCAAGCTCTGGAAGCGGCTGGCCTGCCGAACATTCACGACCAAAA
CGCACGTTTTGATGACGGTTACTTCCCACCGGCATTTACGTTGAAAGGTGAAGAGCGCTTCAGCGCCGCAC
GCGGTTATCTGGATGCGAGCGTCCGTGTGCGTCCGAACCTGAGCCTGTGGACTGAGAGCCGTGTCCTGAA
GCTGCTGACCACTGGCAATGCAATCACCGGTGTGAGCGTGCTGCGTGGTCGCGAAACCCTGCAAGTTCAAG
CGCGCGAGGTCATCCTGACCGCCGGTGCGTTGCAAAGCCCAGCGATTCTGTTGCGCACCGGCATCGGCCC
TGCGGCGGATCTGCACGCACTGGGTATTCCTGTTCTGGCAGACCGTCCGGGTGTTGGTCGCAATCTGTGG
GAGCACAGCTCTATCGGTGTGGTTGCCCCGCTGACCGAGCAGGCACGTGCAGACGCCAGCACGGGTAAAG
CCGGCTCTCGCCATCAACTGGGTATCCGTGCGTCGTCCGGCGTAGATCCGGCGACGCCTAGCGACCTGTT
TCTGCATATCCATGCTGATCCAGTCAGCGGTCTGGCAAGCGCTCTGTTCTGGGTGAACAAGCCAAGCTCCA
CCGGCTGGCTGAAGCTGAAGGACGCGGACCCGTTTAGCTACCCGGACGTAGACTTCAATCTGCTGAGCGAT
CCGCGCGACTTGGGTCGTCTGAAAGCGGGCCTGCGTCTGATCAAACATTACTTCGCATATCCGTCCCTGGC
GAAATATGGTTTGGCGCTGGCATTGAGCCGTTTTGAGGCACCGCAGCCGGGTGGTCCGCTGCTGAACGAC
CTGTTGCAGGACGAAGCCGCCCTGGAACGCTATTTGCGTACGAACGTCGGCGGTGTTTGGCATGCGAGCG
GCACGGCGCGTATCGGCCGTGCGGATGATTCCCAGGCTGTTGTCGATAAAGCGGGTCGTGTGTACGGCGT
CACCGGCCTGCGTGTTGCGGACGCAAGCATTATGCCGACCGTTCCGACCGCCAATACCAATCTGCCGACG
CTGATGCTGGCTGAGAAAATTGCGGATGCGATTCTGACCCAGGCTTAA

***CAR-SFP***

CCATGGGCAGCAGCCATCACCATCATCACCACAGCCAGGATCCGAATTCGATGGCTGTGGACTCGCCGGAT
GAACGCCTGCAACGCCGTATCGCCCAACTGTTTGCCGAAGATGAACAAGTGAAAGCTGCCCGCCCGCTGGA
AGCAGTTAGCGCGGCCGTCTCTGCACCGGGTATGCGTCTGGCTCAGATCGCAGCTACGGTGATGGCTGGT
TATGCGGATCGTCCGGCGGCGGGCCAGCGTGCTTTCGAACTGAATACCGATGACGCAACCGGCCGTACCA
GCCTGCGTCTGCTGCCGCGTTTTGAAACCATTACGTACCGCGAACTGTGGCAGCGTGTCGGCGAAGTGGCA
GCTGCGTGGCATCACGACCCGGAAAACCCGCTGCGTGCGGGTGATTTTGTGGCCCTGCTGGGCTTCACCA
GCATTGATTATGCAACGCTGGATCTGGCTGACATCCATCTGGGTGCGGTTACCGTGCCGCTGCAAGCGAGC
GCGGCGGTGTCCCAACTGATTGCAATCCTGACCGAAACGAGTCCGCGCCTGCTGGCGTCCACCCCGGAAC
ATCTGGATGCTGCGGTGGAATGCCTGCTGGCAGGCACCACGCCGGAACGTCTGGTGGTTTTCGATTATCAC
CCGGAAGATGACGATCAGCGCGCCGCATTTGAAAGTGCGCGTCGCCGTCTGGCAGATGCAGGTTCCCTGG
TGATCGTTGAAACCCTGGACGCGGTGCGTGCGCGTGGCCGTGATCTGCCGGCTGCGCCGCTGTTTGTCCC
GGATACCGACGATGACCCGCTGGCGCTGCTGATTTATACGTCAGGTTCGACCGGCACGCCGAAAGGTGCC
ATGTACACCAATCGTCTGGCCGCAACGATGTGGCAGGGCAACTCAATGCTGCAAGGCAACAGCCAACGCGT
TGGCATTAACCTGAATTATATGCCGATGAGTCATATTGCGGGTCGTATCTCCCTGTTCGGCGTGCTGGCGCG
TGGCGGCACCGCATACTTTGCTGCGAAATCAGACATGAGCACCCTGTTTGAAGATATTGGCCTGGTTCGCC

CGACCGAAATCTTTTTCGTTCCGCGTGTCTGTGACATGGTGTTTCAGCGCTATCAAAGCGAACTGGATCGCC
GTTCTGTCGCTGGTGCGGATCTGGACACCCTGGACCGCGAAGTGAAAGCGGATCTGCGTCAGAATTACCTG
GGCGGTCGCTTCCTGGTTGCAGTCGTGGGCTCGGCTCCGCTGGCCGCAGAAATGAAAACGTTTATGGAAA
GCGTGCTGGACCTGCCGCTGCATGATGGTTATGGCAGTACCGAAGCCGGCGCATCCGTTCTGCTGGATAAC
CAGATCCAACGTCCGCCGGTCCTGGACTATAAACTGGTCGATGTGCCGGAACTGGGTTACTTTCGCACGGA
TCGTCCGCACCCGCGTGGCGAACTGCTGCTGAAAGCAGAAACCACGATTCCGGGTTATTACAAACGCCCGG
AAGTTACGGCGGAAATCTTTGATGAAGACGGCTTCTATAAAACCGGCGATATTGTGGCCGAACTGGAACATG
ACCGCCTGGTTTACGTGGATCGTCGTAACAATGTTCTGAAACTGTCCCAGGGCGAATTTGTGACCGTTGCGC
ACCTGGAAGCTGTGTTCGCGAGCAGCCCGCTGATCCGTCAAATTTTTATCTATGGTAGTTCCGAACGCAGTT
ACCTGCTGGCCGTCATTGTGCCGACCGATGACGCACTGCGTGGCCGCGATACCGCTACGCTGAAAAGCGC
TCTGGCGGAATCTATTCAGCGTATCGCCAAAGACGCAAATCTGCAACCGTATGAAATTCCGCGCGATTTTCT
GATCGAAACCGAACCGTTCACGATTGCCAATGGCCTGCTGAGCGGTATCGCAAAACTGCTGCGCCCGAACC
TGAAAGAACGTTATGGTGCGCAGCTGGAACAAATGTACACCGACCTGGCTACGGGCCAGGCAGATGAACTG
CTGGCCCTGCGCCGTGAAGCTGCGGATCTGCCGGTGCTGGAAACCGTTAGCCGTGCCGCAAAAGCGATGC
TGGGTGTGGCAAGCGCGGATATGCGTCCGGACGCACATTTTACCGATCTGGGCGGTGACAGCCTGTCTGC
ACTGAGTTTTTCCAACCTGCTGCACGAAATCTTCGGTGTTGAAGTCCCGGTGGGTGTTGTCGTGTCTCCGGC
AAACGAACTGCGTGATCTGGCGAATTATATTGAAGCCGAACGCAACAGTGGCGCAAAACGTCCGACCTTCA
CGTCAGTGCATGGCGGTGGCTCGGAAATTCGTGCTGCGGATCTGACCCTGGACAAATTTATCGATGCACGC
ACGCTGGCCGCAGCTGATTCTATTCCGCACGCCCCGGTGCCGGCACAGACCGTTCTGCTGACGGGTGCGA
ATGGCTATCTGGGTCGTTTCCTGTGCCTGGAATGGCTGGAACGCCTGGATAAAACCGGCGGCACCCTGATT
TGTGTTGTCCGTGGTAGCGACGCGGCGGCGGCACGTAAACGTCTGGATTCAGCCTTTGATAGCGGCGATC
CGGGCCTGCTGGAACATTATCAGCAACTGGCAGCACGTACCCTGGAAGTGCTGGCAGGCGATATTGGTGAC
CCGAACCTGGGCCTGGATGACGCGACCTGGCAGCGTCTGGCAGAAACGGTCGATCTGATTGTGCATCCGG
CAGCTCTGGTGAATCACGTTCTGCCGTACACCCAGCTGTTTGGCCCGAACGTGGTTGGCACCGCGGAAATT
GTGCGCCTGGCTATCACCGCGCGTCGTAAACCAGTGACCTATCTGTCTACGGTTGGCGTCGCAGATCAGGT
TGACCCGGCTGAATACCAAGAAGATAGCGATGTGCGTGAAATGTCTGCGGTGCGTGTCGTGCGCGAAAGCT
ATGCCAACGGTTACGGCAATTCTAAATGGGCTGGTGAAGTGCTGCTGCGCGAAGCGCATGATCTGTGCGGT
CTGCCGGTGGCAGTTTTTCGTTCAGATATGATTCTGGCACACTCGCGCTATGCTGGTCAGCTGAATGTCCAA
GATGTGTTCACCCGTCTGATTCTGTCACTGGTTGCTACGGGCATCGCGCCGTATTCGTTTTACCGCACCGAT
GCAGACGGTAACCGTCAGCGCGCCCATTACGATGGTCTGCCGGCAGATTTCACCGCGGCGGCGATTACGG
CGCTGGGTATCCAGGCCACCGAAGGCTTTCGCACGTATGATGTGCTGAATCCGTATGATGACGGTATTAGT
CTGGACGAATTTGTTGATTGGCTGGTCGAATCCGGCCATCCGATTCAGCGTATCACGGATTATTCAGACTGG
TTTCACCGCTTCGAAACCGCCATCCGTGCACTGCCGGAAAAACAGCGTCAAGCCAGCGTGCTGCCGCTGCT
GGATGCATACCGTAACCCGTGTCCGGCCGTTCGCGGTGCAATTCTGCCGGCTAAAGAATTTCAGGCTGCGG
TCCAAACCGCGAAAATTGGCCCGGAACAGGATATTCCGCACCTGAGTGCCCCGCTGATTGATAAATACGTG
TCTGACCTGGAACTGCTGCAACTGCTGGGTAGTGGCTCTGGACTGGTGGGTGCCCTGATGCACGTGATGCA
GAAGCGCAGCCGCGCCATCCACTCCTCCGACGAAGGGGAGGACCAGGCTGGCGATGAAGATGAAGATTGA
GCGGCCGC

***FSC***

GGATCCGGTGCGTAACCAGGGTCTGGGTTCTTGGCCGGTTCGTCGTGCTCGTATGTCTCCGCACGCTACCG
CTGTTCGTCACGGTGGTACCGCTCTGACCTACGCTGAACTGTCTCGTCGTGTTGCTCGTCTGGCTCACGGT
CTGCGTGAAGCTGGTGTTCGTCCGGGTGACCGTGTTGCTTACCTGGGTCCGAACCACCCGGCTTACCTGGA
AACCCTGTTCGCTTGCGGTCAGGCTGGTGCTGTTTTCGTTCCGCTGAACTTCCGTCTGGGTGTTCCGGAACT
GGACCACGCTCTGGCTGACTCTGGTGCTTCTGTTCTGATCCACACCCCGGAACACGCTGAAACCGTTGCTG

CTCTGGCTGGTGACCGTCTGCTGCGTGTTCCGGCTGGTGAACTGGAAGCTGCTGACGACGAACCGCTGGA
CCTGCCGGTTGGTCTGGACGACGTTTGCCTGCTGATGTACACCTCTGGTTCTACCGGTCGTCCGAAAGGTG
CTATGCTGACCCACGGTAACCTGACCTGGAACTGCGTTAACGTTCTGGTTGAAACCGACCTGGCTTCTGAC
GAACGTGCTCTGGTTGCTGCTCCGCTGTTCCACGCTGCTGCTCTGGGTATGGTTTGCCTGCCGACCCTGCT
GAAAGGTGGTACCGTTATCCTGCACTCTGCTTTCGACCCGGGTGCTGTTCTGTCTGCTGTTGAACAGGAAC
GTGTTACCCTGGTTTTCGGTGTTCCGACCATGTACCAGGCTATCGCTGCTCACCCGCGTTGGCGTTCTACC
GACCTGTCTTCTCTGCGTACCCTGCTGTGCGGTGGTGCTCCGGTTCCGGCTGACCTGGCTGGTCGTTACCT
GGACCGTGGCCTGGCGTTCGTTCAGGGTTACGGTATGACCGAAGCTGCTCCGGGTGTTCTGGTTCTGGAC
CGTGCTCACGTTGCTGAAAAAATCGGTTCTGCTGGTGTTCCGTCTTTCTTCACCGACGTTCGTGTTGCTGGT
CCGTCTGGTGAACCGGTTCCGCCGGGTGAAAAAGGTGAAATCGTTGTTTCTGGTCCGAACGTTATGAAAGG
TTACTGGGGTCGTCCGGAAGCTACCGCTGAAGTTCTGCGTGACGGTTGGTTCCGTTCTGGTGACGTTGCTA
CCGTTGACGGTGACGGTTACTTCCACGTTGTTGACCGTCTGAAAGACATGATCATCTCTGGTGGTGAAAACA
TCTACCCGGCTGAAGTTGAAAACGAACTGTACGGTTACCCGGGTGTTGAAGCTTGCGCTGTTATCGGTGTTC
CGGACCCGCGTTGGGGTGAAGTTGGTAAAGCTGTTGTTGTTCCGGCTGCTGGTTCTCGTATCGACGGTGCT
GAACTGCTGGCTTGGCTGCGTACCCGTCTGGCTGGTTACAAAGTTCCGAAATCTGTTGAATTCACCGACCGT
CTGCCGACCACCGGTTCTGGTAAAATCCTGAAAGGTGAAGTTCGTCGTCGTTTCGGTTAAGCGGCCGC

***ECH***

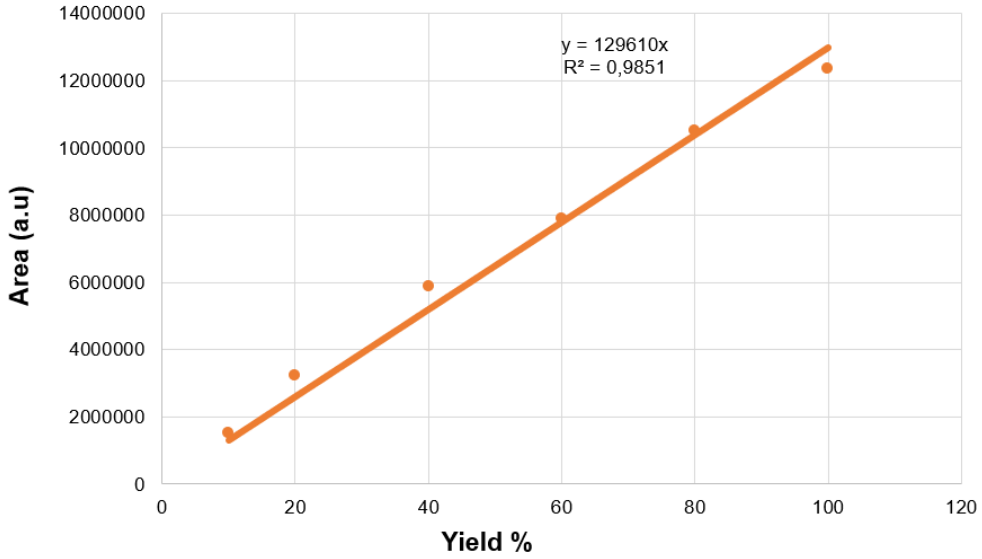
CATATGTCTACCGCTGTTGGTAACGGTCGTGTTCGTACCGAACCGTGGGGTGAAACCGTTCTGGTTGAATTC
GACGAAGGTATCGCTTGGGTTACCCTGAACCGTCCGGACAAACGTAACGCTATGAACCCGACCCTGAACGA
CGAAATGGTTCGTGTTCTGGACCACCTGGAAGGTGACGACCGTTGCCGTGTTCTGGTTCTGACCGGTGCTG
GTGAATCTTTCTCTGCTGGTATGGACCTGAAAGAATACTTCCGTGAAGTTGACGCTACCGGTTCTACCGCTG
TTCAGATCAAAGTTCGTCGTGCTTCTGCTGAATGGCAGTGGAAACGTCTGGCTAACTGGTCTAAACCGACCA
TCGCTATGGTTAACGGTTGGTGCTTCGGTGGTGCTTTCACCCCGCTGGTTGCTTGTGACCTGGCGTTCGCT
GACGAAGACGCTCAGTTCGGTCTGTCTGAAGTTAACTGGGGTATCCCACCGGGTGGCGTTGTAAGCCGTGC
TCTGGCTGCTACCGTTCCGCAGCGTGACGCTCTGTACTACATCATGACCGGTGAACCGTTCGACGGTCGTC
GTGCTGCTGAAATGCGTCTGGTTAACGAAGCTCTGCCGGCTGACCGTCTGCGTGAACGTACCCGTGAAGTT
GCTCTGAAACTGGCTTCTATGAACCAGGTTGTTCTGCACGCTGCTAAAACCGGTTACAAAATCGCTCAGGAA
ATGCCGTGGGAACAGGCTGAAGACTACCTGTACGCTAAACTGGACCAGTCTCAGTTCGCTGACAAAGCTGG
TGCTCGTGCTAAAGGTCTGACCCAGTTCCTGGACCAGAAATCTTACCGTCCGGGTCTGTCTGCTTTCGACCC
GGAAAAATAACTCGAG

## Note

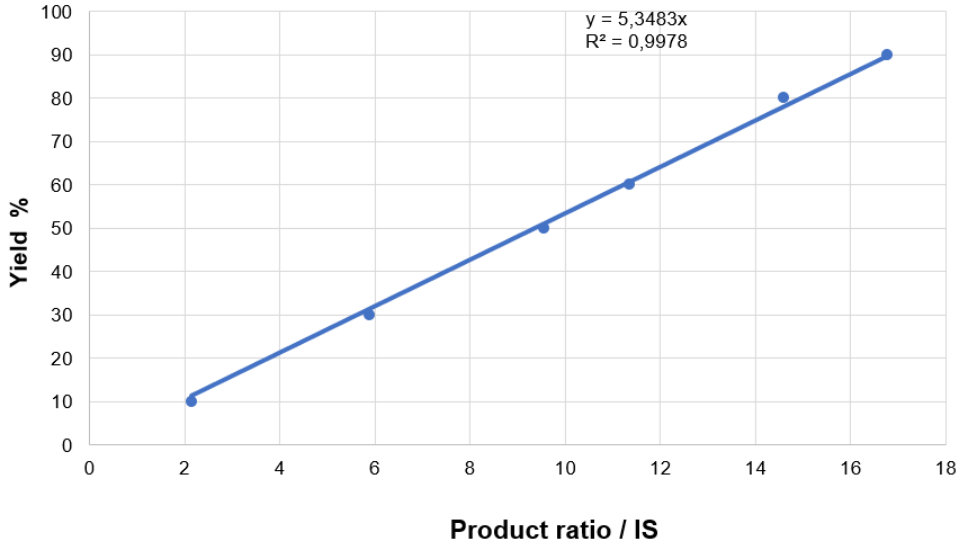
All underlined sequence are the restriction sites used for cloning

**Calibration Curves for Aldehydes**

**Benzaldehyde (1b)** IS: 2-methylanisole

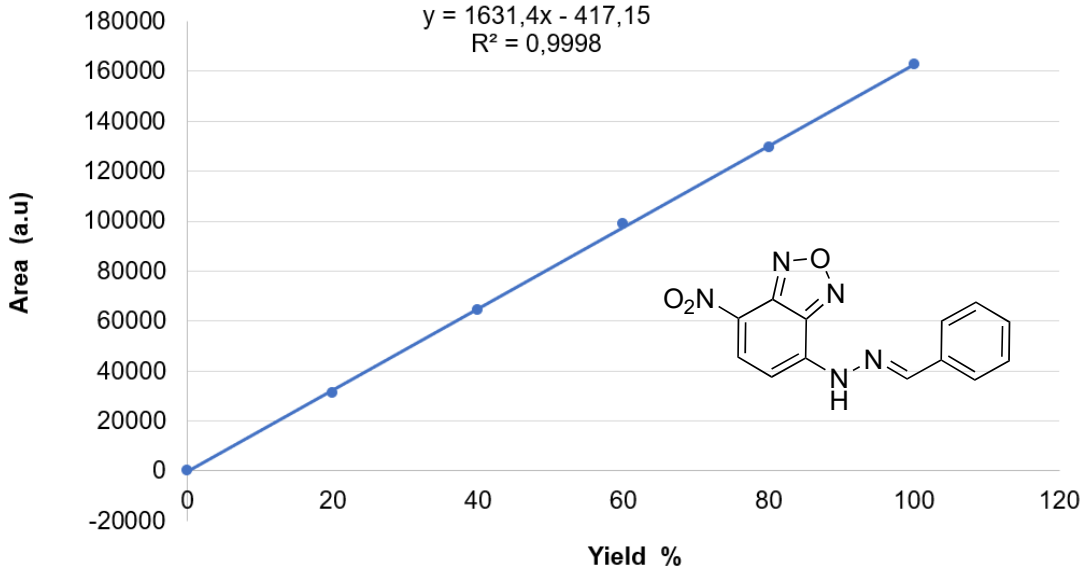


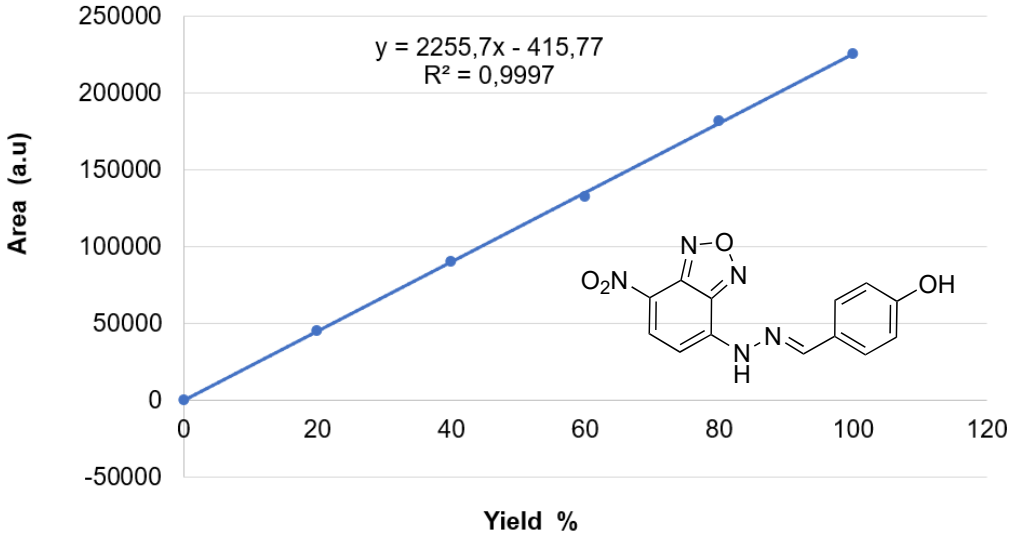

***p*-OH Benzaldehyde (1a)** IS: (2-Phenylquinoline)



**Chemical synthesis and calibration curves of hydrazone products 3a and 3b.**

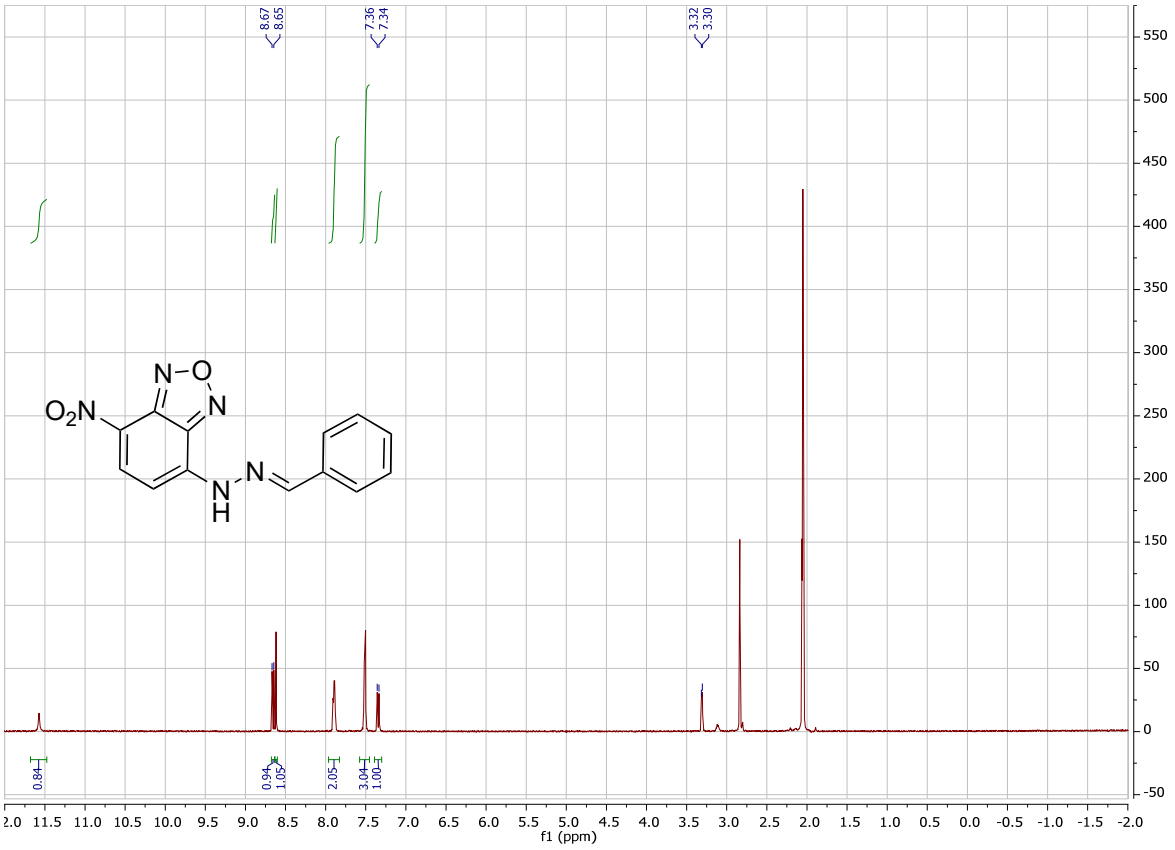
The hydrazone product between NBD-H and 4-hydroxybenzaldehyde (4-HBA) was prepared according to the same procedure as we previously reported.[3] The hydrazone product of the reaction between NBD-H and benzaldehyde was prepared as previously described in the literature.[4] In brief, NBDH (25 mg, 0.13 mmol, 1 equiv) was dissolved in methanol (10 mL). Benzaldehyde (0.13 mL, 1.3 mmol, 10 equiv) was added, turning the reaction mixture a red color. The product gradually formed as a dark red-black precipitate and was isolated by vacuum filtration. $^1$H NMR (400 MHz, $(CD_3)_2CO$): 8.62 (d, 1H, 3J = 8.8 Hz), 8.59 (s, 1H), 7.92–7.79 (m, 2H), 7.56–7.39 (m, 3H), 7.31 (d, 1H, 3J = 8.8 Hz)



## NMR spectra

1H-NMR product **3b** in acetone-d6




## References


[1] R. A. Mehl, J. C. Anderson, S. W. Santoro, L. Wang, A. B. Martin, D. S. King, D. M. Horn, P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 935–939.

[2] C. Mayer, C. Dulson, E. Reddem, A. M. W. H. Thunnissen, G. Roelfes, *Angew. Chemie - Int. Ed.* **2019**, *58*, 2083–2087.

[3] I. Drienovská, C. Mayer, C. Dulson, G. Roelfes, *Nat. Chem.* **2018**, *10*, 946–952.

[4] J. A. Key, C. Li, C. W. Cairo, *Bioconjugate Chem.* **2012**, *23*, 363–371.