GDCh

Communications

Angewandte
Chemie
International Edition
www.angewandte.org

Gene Fusion and Directed Evolution to Break Structural Symmetry and Boost Catalysis by an Oligomeric C–C Bond-Forming Enzyme

*Guangcai Xu+, Andreas Kunzendorf+, Michele Crotti, Henriëtte J. Rozeboom,
Andy-Mark W. H. Thunnissen, and Gerrit J. Poelarends**

Abstract: *Gene duplication and fusion are among the primary natural processes that generate new proteins from simpler ancestors. Here we adopted this strategy to evolve a promiscuous homohexameric 4-oxalocrotonate tautomerase (4-OT) into an efficient biocatalyst for enantioselective Michael reactions. We first designed a tandem-fused 4-OT to allow independent sequence diversification of adjacent subunits by directed evolution. This fused 4-OT was then subjected to eleven rounds of directed evolution to give variant 4-OT(F11), which showed an up to 320-fold enhanced activity for the Michael addition of nitromethane to cinnamaldehydes. Crystallographic analysis revealed that 4-OT(F11) has an unusual asymmetric trimeric architecture in which one of the monomers is flipped 180° relative to the others. This gene duplication and fusion strategy to break structural symmetry is likely to become an indispensable asset of the enzyme engineering toolbox, finding wide use in engineering oligomeric proteins.*

Gene duplication and fusion are among the primary natural processes that generate new proteins from simpler ancestors.[1–5] This is nicely illustrated, for example, by the collection of more than 11 000 proteins that make up the five main subgroups within the tautomerase superfamily (TSF).[6,7] The TSF members share a characteristic β-α-β building block and often possess an amino-terminal proline (Pro-1) that plays a key role as a catalytic residue. The 4-oxalocrotonate tautomerase (4-OT) subgroup is the largest of the five subgroups and its members are composed of a single β-α-β unit (58–84 amino acids) to form homo- or heterohexamers, whereas the members of the other four subgroups are composed of two successively joined β-α-β units (110–150 amino acids) to form homotrimers. The prevalence of this structural arrangement suggests that gene duplication and fusion events took place in the evolutionary history of the TSF, resulting in the diversification of structure and function that is seen nowadays.[6,8]

Research in our laboratory largely focuses on homo-hexameric 4-OT from *Pseudomonas putida* mt-2, which has unique properties that make it an attractive enzyme for study. The small monomer size (only 62 amino acids), the absence of cofactors, and its stability make 4-OT easy to manipulate and amenable to study by a set of diverse techniques. Furthermore, we have demonstrated that 4-OT is capable of utilizing its Pro-1 residue to promiscuously catalyze several synthetically useful transformations via enamine or iminium ion intermediates, including asymmetric Michael addition, aldol addition, epoxidation and cyclo-propanation reactions.[9–15] Although we have been successful in engineering 4-OT variants with improved activity and enantioselectivity, the further optimization of 4-OT to develop efficient biocatalysts with potential for application in chemical synthesis remains challenging.

The crystal structure of homohexameric 4-OT shows that the enzyme consists of a trimeric arrangement of strongly interacting pairs of monomers, with two active sites located in each pair of adjacent subunits (Figure 1). Importantly, this symmetry relationship within homohexa-meric 4-OT, with any point mutation reflected in all six subunits, imposes a significant limitation for its genetic optimization. Inspired by the gene duplication and fusion strategy observed in nature to create new enzymes within the TSF, we report herein our efforts toward the design of a tandem fused 4-OT, enabling independent sequence diversi-fication of neighboring subunits. This tandem fused 4-OT was then subjected to multiple rounds of directed evolution to generate an efficient enzyme for the enantioselective Michael addition of nitromethane to cinnamaldehydes. Our results show that gene fusion to reduce symmetry at the level of the quaternary structure is a powerful strategy for efficient engineering of oligomeric enzymes.

To enlarge the genetic optimization potential of 4-OT by enabling the independent mutagenesis of two adjacent subunits, we set out to construct a tandem fused 4-OT in which the C-terminus of the first monomer is fused to the N-terminus of the second monomer (Figure 1). Inspection of the crystal structure[16] of mature 4-OT revealed that the distance between the C-terminus of the first monomer and

[*] G. Xu,+ A. Kunzendorf,+ M. Crotti, G. J. Poelarends
Department of Chemical and Pharmaceutical Biology
Groningen Research Institute of Pharmacy
University of Groningen, Antonius Deusinglaan 1, 9713 AV
Groningen (The Netherlands)
E-mail: g.j.poelarends@rug.nl

H. J. Rozeboom, A.-M. W. H. Thunnissen
Molecular Enzymology Group, Groningen Institute of Biomolecular
Sciences and Biotechnology, University of Groningen, Nijenborgh
4, 9747 AG Groningen (The Netherlands)

[+] These authors contributed equally to this work.
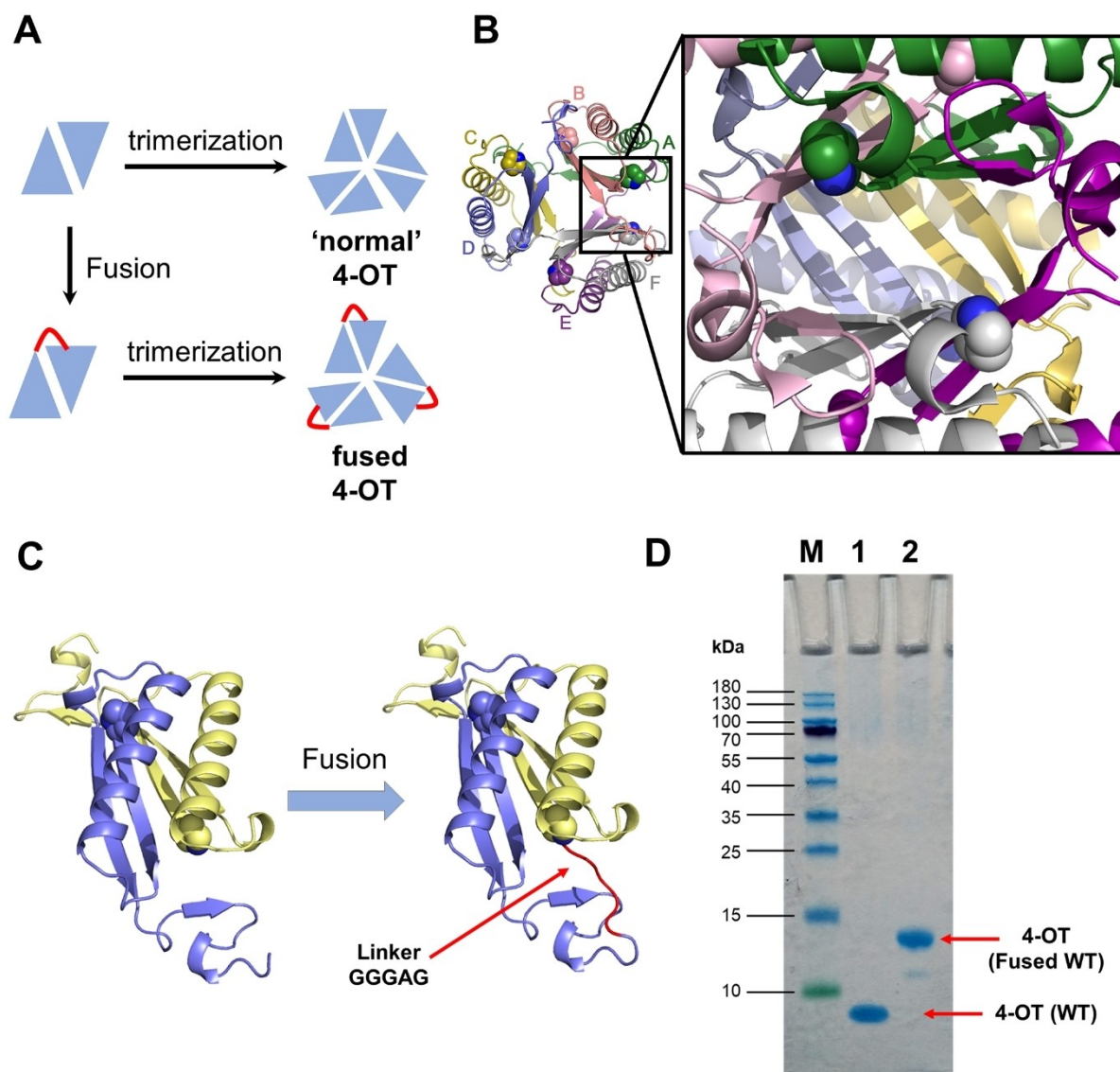
Figure 1. Design of the tandem fused wild-type 4-OT (Fu-WT). A) Cartoon scheme showing the design of the tandem fused 4-OT; one blue triangle represents one β-α-β unit (62 amino acids) and the fusion linkers are shown in red. B) Normal homohexameric 4-OT with a close-up view of the two active sites at the interface of subunits A and F; the six key catalytic N-terminal proline side chains are shown as spheres. C) Structure representation of the tandem fusion of 4-OT. The two β-α-β units are shown in blue and yellow; the side chain of the Pro-1 residue in the first half of the monomer and that of the Pro-68 residue in the second half of the monomer are shown as spheres, and the fusion linker region is shown in red. D) SDS-PAGE gel comparing the purified normal wild-type 4-OT (6.8 kDa) and fused wild-type 4-OT (13.9 kDa). M = protein marker; lane 1 = 4-OT(WT); lane 2 = 4-OT(Fu-WT).

the N-terminus of the second neighboring monomer is around 15 Å. Therefore, a short flexible linker consisting of 5 amino acids (GGGAG) was introduced to connect the two 4-OT monomers, minimizing potential conformational strain on the native head-to-tail termini topology. To evaluate whether this designed tandem fused 4-OT (Fu-WT hereafter) possesses enzymatic activity, we selected as a model reaction the Michael addition of nitromethane (**1**) to cinnamaldehyde (**2a**) yielding γ-nitroaldehyde **3a**, which is a precursor to the well-known pharmaceutical Phenibut (Figure 2). Gratifyingly, Fu-WT catalyzes this synthetically useful Michael addition reaction at a similar rate compared to 'unfused' wild-type 4-OT (Figure S3).

Having designed a functional tandem fused 4-OT, we started our engineering campaign by introducing two mutations into Fu-WT that were previously identified to significantly enhance the Michael addition activity of "unfused" wild-type 4-OT.[12,17] Thus the beneficial mutations M45I and F50A found for "unfused" 4-OT were introduced at the corresponding positions in the second half (M112I and F117A) of Fu-WT. The resulting double mutant 4-OT(Fu-IA) possesses a comparable activity to that of the corresponding "unfused" 4-OT(M45I/F50A) variant, showing a 22-fold enhancement in activity compared to the wild-type enzyme (Figures 2 and S3). These results indicate that the designed tandem fused 4-OT has great optimization

A

B

C
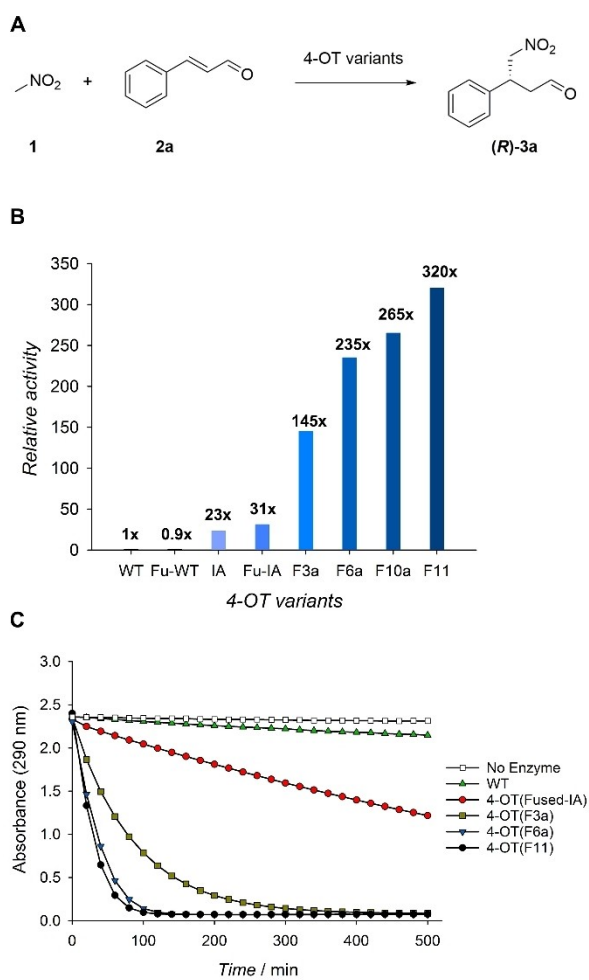
Figure 2. Michael addition activity of 4-OT variants. A) Reaction scheme of the 4-OT-catalyzed Michael addition of **1** to **2a** to yield **3a**. B) Comparison of the specific activity of 4-OT variants. The specific activities are as follows: WT (1.3 mU mg⁻¹), Fu-WT (1.2 mU mg⁻¹), IA (29.9 mU mg⁻¹), Fu-IA (40.3 mU mg⁻¹), F3a (188.5 mU mg⁻¹), F6a (305.5 mU mg⁻¹), F10a (344.5 mU mg⁻¹), F11 (415.5 mU mg⁻¹). C) Reaction progress curves using different 4-OT variants. Reaction conditions: 1 mM cinnamaldehyde, 25 mM nitromethane, 5 µM 4-OT variant, 25 mM HEPES pH 6.5 with 5% (v/v) EtOH.

potential, making it an attractive template to develop efficient "Michaelases" by directed evolution.

To further improve the Michael addition activity of 4-OT(Fu-IA), we performed repeated rounds of directed evolution, combining focused iterative saturation mutagenesis (ISM) and global random mutagenesis (error-prone PCR) (Figure S11, see Supporting Information for a schematic overview of the directed evolution campaign).[18] In the first round, we targeted multiple active-site residue positions individually using ISM, followed in round 2 by shuffling (using overlap extension PCR) of variants possessing beneficial mutations. The directed evolution trajectory was then continued with repeated rounds of error-prone PCR (rounds 3–7) with the intention of randomly sampling the sequence space of fused 4-OT for beneficial mutations. The genes of the identified variants from rounds 6 and 7

were shuffled (using staggered extension PCR) in round 8 to recombine beneficial diversity. After an additional round of error-prone PCR (round 9), the 'hotspot' positions identified by random mutagenesis were revisited in the final two rounds using ISM. To reduce the screening effort for large libraries created by error-prone PCR and DNA shuffling, we used a tailor-made iminium-activated colorimetric "turn-on" probe in agar plate-based pre-screening assays to quickly distinguish active from inactive mutants.[17] Improved enzyme variants were identified by monitoring the depletion of substrate **2a** in a spectrophotometric kinetic assay in multiwell plates (800–1200 transformants were assayed in each round).

After 11 rounds of laboratory evolution, the most active variant, 4-OT(F11), showed a remarkable 320-fold enhancement in catalytic activity for the Michael addition of **1** to **2a** compared to the wild-type enzyme (Figure 2). 4-OT(F11) catalyzes the Michael addition of **1** to **2a** with a catalytic efficiency k_{cat}/K_M of 2.8×10^3 M⁻¹s⁻¹ (Figure S4), making this evolved designer enzyme an effective "Michaelase" with potential for biocatalytic application. 4-OT(F11) also has enhanced enantioselectivity, allowing the production of (R)-**3a** with 98% ee (Figures S13 and S14). Notably, 4-OT(F11) contains 17 mutations on top of Fu-WT, with 7 mutations in the first half and 10 in the second half of the tandem fused protein. Interestingly, the originally designed linker peptide was unaltered in the final variant, although a few mutations emerged in the linker during the evolution campaign, suggesting other feasible linker designs (Figure S11).

To investigate the contribution of the tandem fusion to the catalytic performance of 4-OT(F11), we performed gene fission by removing the linker peptide and separating the enzyme into two "unfused" monomers, named 4-OT(F11α) and 4-OT(F11β) (Figure S5). The genes for both monomers were individually cloned into the same expression vector as originally used for 4-OT(F11). Although the "unfused" monomers 4-OT(F11α) and 4-OT(F11β) are apparently less stable with at least half of the amount of protein produced being present in the insoluble cell fraction (Figure S5), we were able to purify sufficient quantities of both proteins to perform activity assays. Whereas 4-OT(F11α) (first monomer) completely lost its enzymatic activity, 4-OT(F11β) (second monomer) still retained activity albeit 14-fold decreased (Figure S5). These results underscore the importance of the tandem fusion for optimization of the enzyme's catalytic performance, allowing independent sequence diversification of the neighboring monomers during the evolution process.

The biocatalytic potential of 4-OT(F11) was further demonstrated by performing the gram-scale synthesis of the important chiral synthon (R)-**3a**. Even though substrate **2a** (40 mM) was not completely soluble in the reaction buffer containing 5% (v/v) EtOH as cosolvent (Figure S6), nearly full conversion was achieved yielding product (R)-**3a** in good isolated yield (2.5 g, 65%) and with excellent enantiopurity (98% ee). Interestingly, the preparative-scale reaction with 40 mM **2a** can even be performed under otherwise identical conditions but without adding organic cosolvent,

maintaining very good conversion (>95 %). This is probably attributed to the low K_M (43±2 μM) of the evolved designer enzyme for substrate **2a** (Figure S4), allowing efficient catalysis even at low substrate concentrations in the aqueous phase. Substrate scope analysis showed that 4-OT(F11) can accept various cinnamaldehydes (**2b**–**2g**) with *ortho*-, *meta*- and *para*-substitutions on the aromatic ring, providing enzymatic access to the corresponding (R)-γ-nitroaldehydes (**3b**–**3g**) with good to excellent enantiopurity (78–99 % *ee*; Figure S7, S15–S20).

Finally, we determined the crystal structure of 4-OT-(F11) at 2.35 Å resolution (Figure 3, Table S1). Consistent with the existence of the trimeric species in solution (Figure S21), 4-OT(F11) adopts a homotrimeric architecture in which each monomer shows an overall fold that is highly similar to that of a corresponding 'unfused' 4-OT homo-dimer, except for the presence of the engineered linker. Cα-backbone superimpositions of 4-OT(F11) fused monomers to "unfused" homodimers of 4-OT (PDB entries 4X19, 5CLN, 6FPS) result in root-mean-square-deviations of 0.4–0.6 Å, emphasizing that the engineered linker and the
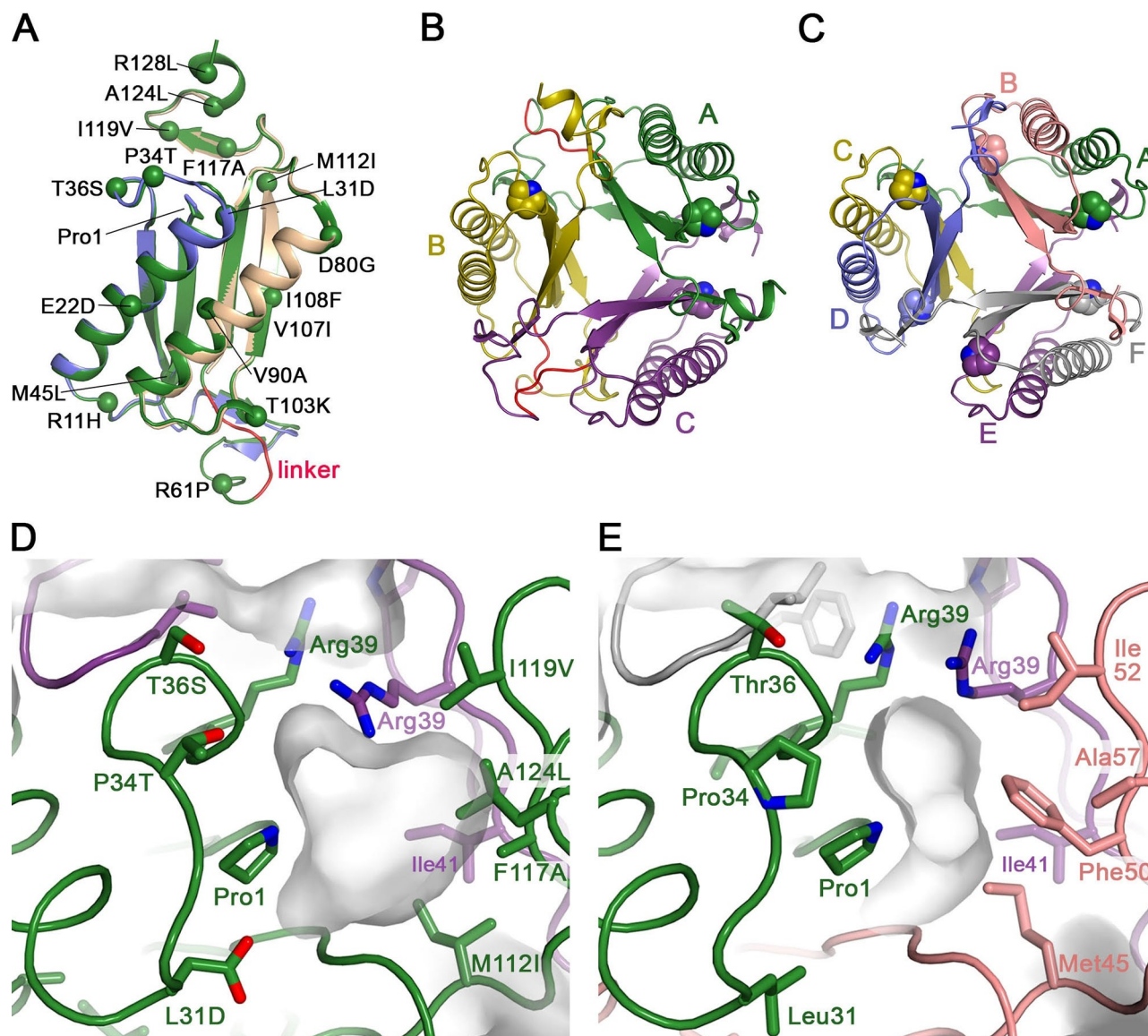
Figure 3. Crystal structure of 4-OT(F11) and comparison with "unfused" wild-type 4-OT. A) Backbone overlay of a fused 4-OT(F11) monomer (in green) to an "unfused"' homodimer of wild-type 4-OT (PDB entry 4X19, in blue and wheat). Spheres indicate the Cα atoms of the mutated residues in the fused 4-OT(F11) monomer. B) Cartoon representation of the asymmetric 4-OT(F11) trimer (PDB entry 7PUO), with each monomer indicated by a label and a different color (green, chain A; gold, chain B; magenta, chain C). The Pro-1 side chains are shown as spheres. The linker regions are shown in red. C) Similar cartoon representation of a wild-type 4-OT hexamer (PDB entry 4X19). D) Close-up view of the 4-OT(F11) active site in subunit A (green), at the interface with subunit C (magenta), showing the side chains of the mutated residues that surround the hydrophobic active site pocket near Pro-1. The cut-away surface in grey depicts the boundaries of the hydrophobic pocket. E) Equivalent view of an active site in homohexameric wild-type 4-OT, showing that the active site pocket in the natural wild-type enzyme is substantially smaller.

seventeen mutations do not significantly affect the 4-OT backbone fold. Interestingly, though, the 4-OT(F11) trimer lacks the 3-fold rotational symmetry present in the homotrimeric structures of most naturally fused members of the tautomerase superfamily (Figure S8). Instead, it forms an asymmetric trimer in which one of the monomers (chain C) is flipped 180° relative to the other two monomers (chains A and B). Consequently, the Pro-1 residues and active sites in subunits A and C are positioned adjacent to each other across one of the trimeric interfaces, while at another interface the Pro-1 residue of chain B is positioned adjacent to Pro-68 of chain A (Figure 3B). Intriguingly, a similar uncommon asymmetric trimeric architecture has recently been observed for certain naturally fused homotrimeric 4-OT homologs.[19,20] Analysis of the subunit interfaces indeed suggests that for 4-OT(F11) the assembly of an asymmetric trimer is energetically more favorable than that of a symmetric trimer (Table S2).[20] Hence, both natural and laboratory evolution of tandem fused 4-OT enzymes may thus result in unusual asymmetric trimers.

A comparison of the 4-OT(F11) and wild-type 4-OT structures further reveals clear structural effects for mutations M112I, F117A, I119V and A124L in opening up and reshaping a hydrophobic pocket next to Pro-1 (Figure 3D,E). Molecular modeling of the putative reaction intermediate (**Int3**, Figure S9) formed between Pro-1 and (*R*)-**3a** (Figure S10) suggests that this pocket is suitable for binding cinnamaldehyde and its conversion to γ-nitroaldehyde **3a** via Pro-1 assisted iminium catalysis (Figure S9, S10). Located at the back of the active site of subunit B, mutation I108F from subunit A additionally contributes to reshaping the hydrophobic pocket. Instead, in the other two active sites, residue Ile-41 is located at this position, which is the non-mutated equivalent residue of Phe-108 in the first half of the 4-OT(F11) monomers. Additional mutations near the active site pocket are L31D, P34T, and T36S. It is unlikely that these mutated residues play a direct role in catalysis or substrate binding, but they could potentially be required for providing a suitable polar environment for binding nitromethane, for stabilising the reaction intermediates, or for activating a nucleophilic water in the final step of the reaction mechanism. The additional nine mutations in the 4-OT(F11) variant are located away from the active site pockets (Figure 3A), at the protein surface (R11H, E22D, R61P, D80G, T103K, R128L), the subunit interface (M45L) or subunit interior (V90A, V107I). The 4-OT(F11) crystal structure does not provide a clear understanding how these mutations contribute to the catalytic activity of the enzyme, but possibly they improve the stability of the fused 4-OT trimer.

In summary, we have engineered an efficient biocatalyst for the Michael addition of nitromethane **1** to cinnamaldehyde **2a** by directed evolution of a rationally designed tandem fused 4-OT enzyme. After eleven rounds of directed evolution, the evolved variant 4-OT(F11) displayed a 320-fold enhancement in catalytic activity compared to the wild-type enzyme. Gratifyingly, 4-OT(F11) outperforms all previously engineered 'unfused' 4-OT variants in terms of activity (40-fold enhance activity compared to the previously

engineered 4-OT F50A variant),[12] accepts various cinnamaldehydes as Michael acceptor in the enantioselective addition of nitromethane, and enables the gram-scale synthesis of the (*R*)-γ-nitroaldehyde precursor to phenibut with excellent enantiopurity (98 % *ee*). Compared to an established organocatalytic methodology for the synthesis of the same compound, 4-OT(F11) requires lower catalyst loading (0.08 mol % vs 10 mol %) and enables catalysis in aqueous buffer instead of organic solvent.[21] The tandem fused 4-OT has a reduced symmetry compared to "unfused" wild-type 4-OT, expanding its genetic optimization potential by allowing independent sequence diversification of adjacent subunits, and thus enlarging the protein sequence space that can be sampled by directed evolution. The applied engineering strategy recapitulates the path nature took to create new enzymes within the TSF and will undoubtedly prove to be useful for the optimization of other promiscuous activities of 4-OT, supporting the construction of a new family of 4-OT derived biocatalysts with synthetically useful activities.

It is important to stress that different evolved enzyme variants may possess different selectivity profiles.[22] Current work in our group is therefore focused on testing the versatile set of novel fused 4-OT variants against a panel of substrates (e.g., other α,β-unsaturated aldehydes, hydroperoxides, diethyl (halo)malonates, and thiols). The initial results show that improved fused 4-OT variants, indentified in round three of our directed evolution program, have also significantly enhanced activity for the epoxidation of cinnamaldehydes (using H_2O_2 instead of nitromethane) compared to a previously reported "unfused" 4-OT mutant.[14] These preliminary results will be reported in due course.

Although widely used by nature to create new enzymes, this powerful approach of gene fusion to break protein symmetry at the level of the quaternary structure has rarely been explored in the laboratory evolution of oligomeric enzymes. Notable exceptions are recent elegant works by both the Hilvert and Ward groups, who have adopted a similar gene fusion strategy to engineer efficient artificial metalloenzymes.[23,24] Thus, gene fusion and most likely also gene fission, which are among the primary natural processes that generate new genes (and their corresponding protein products), appear to be indispensable assets of the enzyme engineering toolbox.

Conflict of Interest

The authors declare no conflict of interest.

[1] M. L. Romero Romero, A. Rabin, D. S. Tawfik, *Angew. Chem. Int. Ed.* **2016**, *55*, 15966–15971; *Angew. Chem.* **2016**, *128*, 16198–16203.

[2] R. V. Eck, M. O. Dayhoff, *Science* **1966**, *152*, 363–366.

[3] S. Balaji, *Curr. Opin. Struct. Biol.* **2015**, *32*, 156–166.

[4] D. Lang, R. Thoma, M. Henn-Sax, R. Sterner, M. Wilmanns, *Science* **2000**, *289*, 1546–1550.

[5] J. Lee, M. Blaber, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 126–130.

[6] G. J. Poelarends, V. P. Veetil, C. P. Whitman, *Cell. Mol. Life Sci.* **2008**, *65*, 3606–3618.

[7] C. P. Whitman, *Arch. Biochem. Biophys.* **2002**, *402*, 1–13.

[8] R. Davidson, B. J. Baas, E. Akiva, G. L. Holliday, B. J. Polacco, J. A. LeVieux, C. R. Pullara, Y. J. Zhang, C. P. Whitman, P. C. Babbitt, *J. Biol. Chem.* **2018**, *293*, 2342–2357.

[9] E. Zandvoort, E. M. Geertsema, B. J. Baas, W. J. Quax, G. J. Poelarends, *Angew. Chem. Int. Ed.* **2012**, *51*, 1240–1243; *Angew. Chem.* **2012**, *124*, 1266–1269.

[10] E. M. Geertsema, Y. Miao, P. G. Tepper, P. Dehaan, E. Zandvoort, G. J. Poelarends, *Chem. Eur. J.* **2013**, *19*, 14407–14410.

[11] J. Y. Van Der Meer, H. Poddar, B. J. Baas, Y. Miao, M. Rahimi, A. Kunzendorf, R. Van Merkerk, P. G. Tepper, E. M. Geertsema, A. M. W. H. Thunnissen, W. J. Quax, G. J. Poelarends, *Nat. Commun.* **2016**, *7*, 10911.

[12] C. Guo, M. Saifuddin, T. Saravanan, M. Sharifi, G. J. Poelarends, *ACS Catal.* **2019**, *9*, 4369–4373.

[13] M. Saifuddin, C. Guo, L. Biewenga, T. Saravanan, S. J. Charnock, G. J. Poelarends, *ACS Catal.* **2020**, *10*, 2522–2527.

[14] G. Xu, M. Crotti, T. Saravanan, K. M. Kataja, G. J. Poelarends, *Angew. Chem. Int. Ed.* **2020**, *59*, 10374–10378; *Angew. Chem.* **2020**, *132*, 10460–10464.

[15] A. Kunzendorf, G. Xu, M. Saifuddin, T. Saravanan, G. J. Poelarends, *Angew. Chem. Int. Ed.* **2021**, *60*, 24059–24063; *Angew. Chem.* **2021**, *133*, 24261–24265.

[16] L. Biewenga, T. Saravanan, A. Kunzendorf, J. Y. Van Der Meer, T. Pijning, P. G. Tepper, R. Van Merkerk, S. J. Charnock, A. M. W. H. Thunnissen, G. J. Poelarends, *ACS Catal.* **2019**, *9*, 1503–1513.

[17] L. Biewenga, M. Crotti, M. Saifuddin, G. J. Poelarends, *ACS Omega* **2020**, *5*, 2397–2405.

[18] G. Qu, A. Li, C. G. Acevedo-Rocha, Z. Sun, M. T. Reetz, *Angew. Chem. Int. Ed.* **2020**, *59*, 13204–13231; *Angew. Chem.* **2020**, *132*, 13304–13333.

[19] B. J. Baas, B. P. Medellin, J. A. Levieux, M. De Ruijter, Y. J. Zhang, S. D. Brown, E. Akiva, P. C. Babbitt, C. P. Whitman, *Biochemistry* **2019**, *58*, 2617–2627.

[20] B. P. Medellin, E. B. Lancaster, S. D. Brown, S. Rakhade, P. C. Babbitt, C. P. Whitman, Y. J. Zhang, *Biochemistry* **2020**, *59*, 1592–1603.

[21] H. Gotoh, H. Ishikawa, Y. Hayashi, *Org. Lett.* **2007**, *9*, 5307–5309.

[22] C. G. Acevedo-Rocha, C. G. Gamble, R. Lonsdale, A. Li, N. Nett, S. Hoebenreich, J. B. Lingnau, C. Wirtz, C. Fares, H. Hinrichs, A. Deege, A. J. Mulholland, Y. Nov, D. Leys, K. J. McLean, A. W. Munro, M. T. Reetz, *ACS Catal.* **2018**, *8*, 3395–3410.

[23] S. Studer, D. A. Hansen, Z. L. Pianowski, P. R. E. Mittl, A. Debon, S. L. Guffy, B. S. Der, B. Kuhlman, D. Hilvert, *Science* **2018**, *362*, 1285–1288.

[24] S. Wu, Y. Zhou, J. G. Rebelein, M. Kuhn, H. Mallin, J. Zhao, N. V. Igareta, T. R. Ward, *J. Am. Chem. Soc.* **2019**, *141*, 15869–15878.