Enzyme activity by design: an artificial rhodium hydroformylase for linear aldehydes

Amanda G. Jarvis,*[a] Lorenz Obrecht,[a] Peter J. Deuss,[a, b] Wouter Laan,[a, g] Emma K. Gibson,[c,d] Peter P. Wells,[d,e,f] and Paul C. J. Kamer.*[a][h]

Abstract: Artificial metalloenzymes (ArMs) are hybrid catalysts that offer a unique opportunity to combine the superior performance of natural protein structures with the unnatural reactivity of transition metal catalytic centers. Therefore, they provide the prospect of highly selective and active catalytic chemical conversions for which natural enzymes are unavailable. Here we show that by rationally combining robust site-specific phosphine bioconjugation methodology and a lipid-binding protein (SCP-2L), an artificial rhodium hydroformylase can be developed that displays remarkable activities and selectivities for the biphasic production of long chain linear aldehydes under benign aqueous conditions. Overall, this study demonstrates that judiciously chosen protein binding scaffolds can be adapted to obtain metalloenzymes that provide the reactivity of the introduced metal center combined with specifically intended product selectivity.

The development of substrate and product specific catalytic processes that operate efficiently at mild reaction temperatures is a major challenge for the synthetic chemical community.[1] Enzymes are nature's main catalysts which catalyse numerous chemical transformations, typically at benign conditions. However, many desired chemical reactions are not performed by nature and therefore suitable natural enzymes are lacking. ArMs provide a way to bridge that gap between nature and chemical production, providing enzymes for unnatural reactions.[2] Despite these successes, most ArMs do not meet the rates and performances achieved by natural enzymes.[3] In addition, the molecular recognition and shape selectivity of proteins has typically not been exploited. The most successful approach to create ArMs has been the use of non-covalent anchoring strategies utilizing protein scaffolds with strong supramolecular recognition motifs e.g. avidin.[4] In these ArMs the binding site is used to carry the active metal centre or bind metal-containing cofactors, restricting the possible applications of a protein's binding properties. An alternative approach utilizes site-selective protein modification methodology[5] to incorporate transition metals into a wide range of protein scaffolds whilst leaving the protein's innate binding capabilities largely intact. Any protein scaffold can be used, allowing the exploitation of the almost unlimited range of highly specific substrate binding capabilities of proteins. Therefore, virtually any organometallic non-natural catalytic reaction can be merged with the sophisticated biological performance of enzymes and this approach offers significant opportunities for the design of ArMs aiming at high selectivity via shape selective product formation. Here we demonstrate the potential of such ArMs by the development of artificial rhodium enzymes derived from a protein scaffold that was selected for its apolar substrate binding properties. These ArMs enable selective aldehyde formation in the biphasic hydroformylation of long chain linear alkenes, a reaction for which no natural enzymes are known and which is

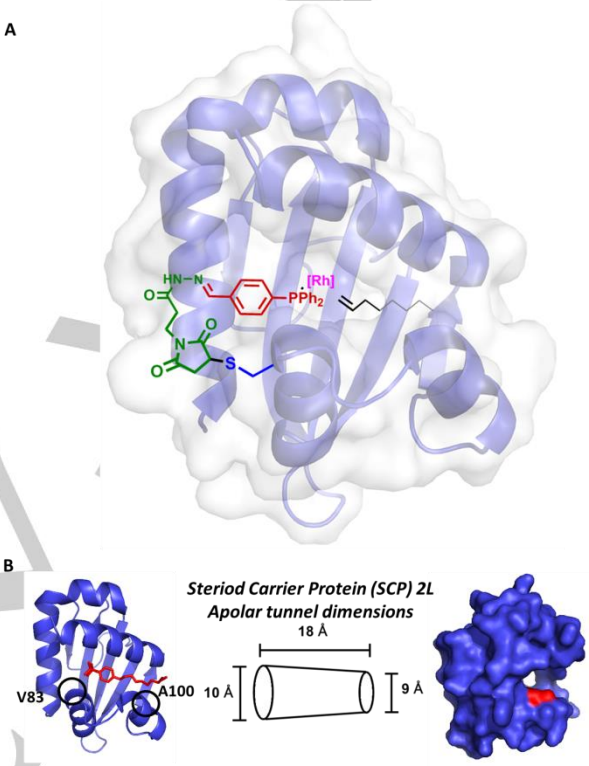


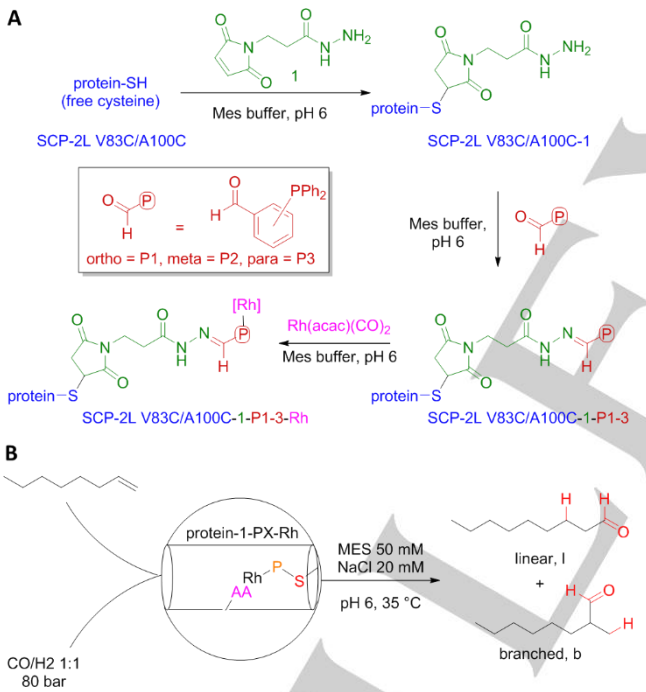
Figure 1. A) Conceptual illustration of the use of the apolar tunnel to introduce regioselectivity into hydroformylation. **B**) Illustrations of the apolar tunnel in SCP-2L, showing the position of Triton X-100 in the tunnel in the original crystal structure pdb: 1IKT[13], the tunnel dimensions and positions of A100 and V83.

[a] Dr. A. Jarvis*, Prof. P. C. J. Kamer*, Dr. L. Obrecht, Dr. P. J. Deuss, Dr. W. Laan.
School of Chemistry, University of St Andrews, North Haugh, St Andrews, Fife, KY16 9ST.
E-mail: agj2@st-andrews.ac.uk; Paul.Kamer@catalysis.de
[b] Department of Chemical Engineering (ENTEG), University of Groningen, Nijenborgh 4, 9747 AG Groningen, the Netherlands.
[c] Dr. E. K. Gibson
Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, United Kingdom.
[d] Dr. E. K. Gibson, Dr. P. P. Wells
UK Catalysis Hub, Research Complex at Harwell, Rutherford Appleton Laboratory, Harwell Science & Innovation Campus, Didcot, Oxfordshire, OX11 0FA.
[e] Dr. P. P. Wells
School of Chemistry, University of Southampton, Southampton, SO17 1BJ.
[f] Dr P. P. Wells
Diamond Light Source, Harwell Science & Innovation Campus, Didcot, Oxfordshire, OX11 0DE.
[g] Current Address: Phoreon, Bioincubator I, Gaston Geenslaan 1, B-3001 Leuven, Belgium.
[h] Bioinspired Homo- & Heterogeneous Catalysis, Leibniz Institute for Catalysis, Albert-Einstein-Straße 29 a, 18059 Rostock, Germany.

Supporting information for this article is given via a link at the end of the document.

challenged in current industrial applications by low solubility of the substrates in the aqueous phase.[6]

We have previously reported a methodology that enables such an approach for the synthesis of ArMs containing metal-binding phosphine ligands (Scheme 1a).[5a,7] Rhodium-phosphine complexes are known to be highly active and robust hydroformylation catalysts and thus our strategy provides the prospect of enzymatic hydroformylation reactions. Rhodium-protein hybrids tested to date in hydroformylation have utilized dative protein-rhodium interactions with limited success.[8] Although, this has led to unprecedented linear selectivities for the hydroformylation of styrene[9] the exact nature of the active species, and thus origin of the selectivity is still unclear.[10] Rhodium-catalyzed hydroformylation is used on a 800,000 tonne scale to produce butyraldehyde from propene under biphasic conditions,[11] allowing recovery of the expensive rhodium-3,3',3''-phosphanetriyltris(benzenesulfonic acid) trisodium salt catalyst (Rh-TPPTS). Long chain aldehydes are desired by industry, as they are important precursors for the production of detergent and plasticizers.[12] This process is not feasible for long chain alkenes (>5 carbon atoms) due to their low solubility in water[6].



Scheme 1. (A) Scheme showing the synthesis of the artificial metalloproteins. (B) Reaction scheme for the hydroformylation of 1-octene, AA = amino acid.

Fatty acid transporter proteins contain apolar tunnels and clefts to bind their hydrophobic cargo. The steroid carrier protein type 2 like (SCP-2L) domain of the human multifunctional enzyme 2 (MFE-2) was identified as a suitable linear alkene binding protein scaffold as it is known to bind a variety of linear aliphatic substrates[13] and can be obtained in high yields.[5a] Our hypothesis was that the apolar tunnel present in SCP-2L (Figure 1b) would be able to facilitate the transport of alkenes to the aqueous environment and orient the starting alkene along the tunnel towards the rhodium enabling shape selective production of the desired linear product (Figure 1a). To introduce the catalytic rhodium-phosphines, two mutants containing unique cysteines at either end of the tunnel were prepared (SCP-2L V83C and SCP-2L A100C, Figure 1b).[5a,14] These two mutants, obtained in excellent yields, showed little structural permutations and similar aliphatic substrate binding capabilities as the WT protein (see SI: Table S3). Both SCP-2L mutants were successfully modified with aldehyde phosphines **P1**-**P3** through a cysteine modification strategy (Scheme 1a, characterisation data for SCP-2L V83C-**1**-**P1-3** see ref 5a and SI for SCP-2L A100C-**1**-**P1-3**).

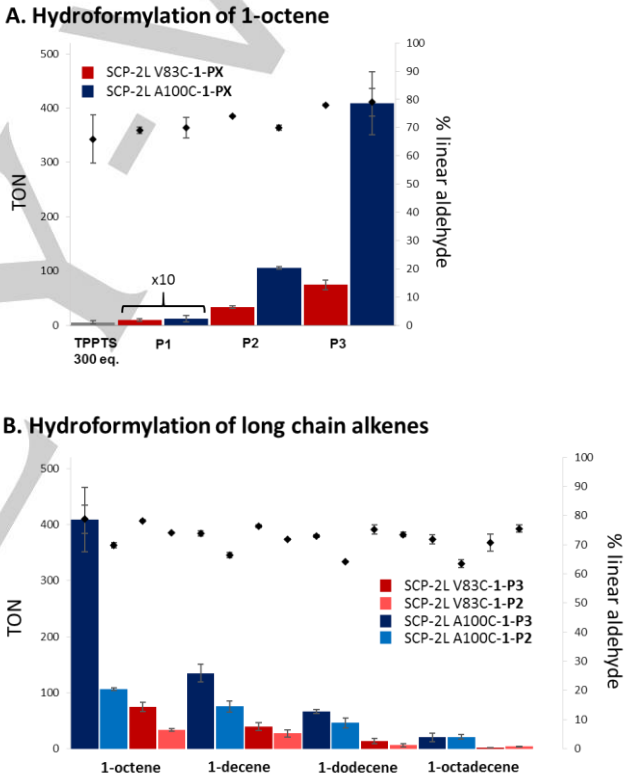


Figure 2. Activity (coloured bars) and selectivity (black squares) of the catalytic hydroformylation: (A) using different artificial metalloenzymes in the hydroformylation of 1-octene (the values for **P1** have been magnified), (B) across different length alkenes. Standard conditions: 80 bar $CO:H_2$ (1:1), 35 °C, 625 rpm, 0.5 mL of catalyst solution and 0.5 mL of alkene containing 9% (v/v) n-heptane and 1% (v/v) diphenyl ether as internal standards. Rh concentrations were obtained by ICP-MS and used to calculate TON, recorded values between 20 and 100 nmol Rh. Conversions and linear selectivities were obtained by GC using a minimum of 3 runs. Error bars show standard deviation. Further details available in the SI section 4.

Table 1: Control reactions for the aqueous hydroformylation of 1-octene.

Entry	Catalyst	TON	% linear aldehyde
1	ArM: SCP-2L A100C-**1**-**P3**-Rh[a]	408.7 (57.79)	78.8 (4.86)
2	$Rh(acac)(CO)_2$[b]	529.7 (53.30)	55.3 (0.67)
3	Protein scaffold A100C treated with Rh[c,d]	123.5 (38.19)	57.8 (0.07)
4	Rh-TPPTS 1:2[e]	2245 (674)	58.9 (0.41)
5	Rh-TPPTS 1:20[e]	700 (190)	56.5 (0.08)
6	Rh-TPPTS 1:300[e]	5.4 (3.25)	65.9 (8.55)
7	Rh-TPPTS 1:300 + SCP-2L A100C[f]	9.4 (1.01)	60.7 (1.48)
8	Rh-TPPTS 1:300 + WT SCP-2L[f]	5.1 (2.48)	67.0 (0.64)

Standard conditions: 80 bar syn gas (1:1), 35 °C, stirring 625 rpm, 0.5 mL of catalyst solution and 0.5 mL of 1-octene containing 9% (v/v) n-heptane and 1% (v/v) diphenyl ether. Rh concentration was obtained by ICP-MS for the ArM entry. Conversions and linear selectivities were obtained by GC using a minimum of 3 runs. Standard deviation in brackets. a) P:Rh 1.5, 23 nmol Rh, b) 41.25 h, total volumnn 0.4 ml 1-octene, no water 150 nmol Rh, c) Treated with Rh and washed in the same manner as the metalloproteins. d) 10.0 nmol Rh, e) 30 nmol Rh, f) 2 eq. protein to Rh.

The rhodium proteins (SCP-2L V83C/A100C-**1**-**P1**-**3**-Rh) were obtained by the addition of $Rh(acac)(CO)_2$. Other Rh precursors did not selectively bind to the phosphine (see Table S4 for MS and metal loading analysis). Their catalytic activity investigated in the hydroformylation of 1-octene at 35 °C and 80 bar synthesis gas (Figure 2, and see SI: Table S5). To minimise rhodium leaching and therefore false results from 'free Rh' (leading to low selectivity (~55% linearity) and high TON's (>500) Table 1, entry 2), a slight excess of protein (2 eq.) was used. Even though these reactions were performed at a relatively low temperature (typically industrial conditions are 125 °C), significant hydroformylation activity was detected over 48 h for several of the rhodium-phosphine ArMs (Figure 2a). Reactions over time showed that the enzyme was active across the whole 48 h (see SI: Table S6). Control reactions with the ArM phosphine selenide or ArM phosphine gold complex showed that the rhodium was required for the hydroformylation reaction to occur (see SI: Table S5).

The structure of the phosphine cofactor was found to have a large effect on the turnover number, with the activity increasing as the phosphine moved from the *ortho* (**P1**) to *meta* position (**P2**), and the *ortho* (**P1**) to *para* position (**P3**) by 30 fold and 70 fold, respectively for SCP-2L V83C. The protein scaffold also influenced the reactivity, with a marked improvement in turnover found when using SCP-2L A100C-**1**-**P3**, achieving TONs of >400, versus 75 for SCP-2L V83C-**1**-**P3**. The selectivity of the reaction was also found to vary depending on the combination of protein scaffold and phosphine cofactor applied. It ranged between 69 and 79% for the linear product (nonanal), matching typical selectivities for double phosphine ligated rhodium catalyzed hydroformylation.[15] Selectivities of 80% are rarely seen with monoligated P-Rh systems. Control reactions with **P1-3** or $Rh(acac)(CO)(PPh_3)$ in neat alkene show that with monophosphine's in the organic phase of our reaction selectivities up to 74% can be achieved (refer to SI Table S8-9). Catalyst degradation and leaching of the phosphine and rhodium to the organic phase was not responsible for the observed selectivities as only minimal rhodium leaching and degradation were observed (see SI: Table S11 and Figs S12 and S13). It should be noted that our ArM system has a low P:Rh ratio, which, when using a benchmark biphasic Rh-TPPTS catalytic system only gives activity and selectivities that correspond to rhodium leaching into the organic phase (Table 1, entries 2, 4 and 5). At the same concentration as our ArM reactions, over 300 eq of TPPTS ligand is needed to prevent metal leaching (Table 1, entries 4-6) and obtain high selectivities. The same effect of rhodium leaching was observed when the protein was simply mixed with $Rh(acac)(CO)_2$ (Table 1, entry 3). This, alongside the differences between the two protein mutants, shows that it is the hybrid catalysts that are responsible for the hydroformylation results. SCP-2L A100C-**1**-**P3** gave the best performance overall (79% nonanal, 409 TON; Table 1, entry 1). These conversions and selectivities are remarkable as a benchmark catalyst (Rh-TPPTS) gives negligible conversion approaching the detection limit (TON ≈ 1) when the TPPTS:Rh ratio is optimised to give similar selectivities to the metalloenzyme (TPPTS:Rh 30:1 at ~10 times the Rh conc. of the ArM reaction gives 72% linearity; see SI: Table S7).

Following the successful hydroformylation of 1-octene, these artificial metalloenzymes were tested in the hydroformylation of 1-decene, 1-dodecene and 1-octadecene (Figure 2b).[6] When TPPTS is used a 10-fold rate decrease is observed on the addition of two carbons to the chain length due to the reduced water solubility of the alkene.[6] Using the ArM a < 4-fold decrease in activity was observed on going from 1-octene to 1-decene, and only 10-fold when going to octadecene. Under selectivity optimised conditions (high ligand concentrations to give adequate linear selectivity but prevention of Rh-leaching), Rh-TPPTS displayed no significant activity (Figure 2a, Table 1, entry 6, and SI: Table S7). Control experiments using the Rh-TPPTS system in the presence of the protein scaffold showed no significant difference in turnover for 1-octene (Table 1, entry 6 vs. 7 and 8), providing evidence that the increase in activity for the ArMs cannot solely be explained by the protein acting as a phase transfer reagent. We therefore attribute the higher than expected activity to the presence of the lipid-binding tunnel in the protein scaffold in direct proximity of the Rh center.

Overall, the selectivities for the linear hydroformylation products were remarkably high for monophosphine-ligated rhodium in water indicating that the protein scaffold counter balances the lack of phosphine ligands. In addition, both the phosphine-cofactor and protein mutant affect the activity of the reaction. To better understand the observed selectivity of our hydroformylase we investigated the local environment of the Rh in the protein scaffold. Both the X-ray absorption near edge

structure (XANES) and extended X-ray absorption fine structure (EXAFS) of SCP-2L A100C-**1**-**P3**-Rh at the Rh K edge were assessed. Comparing the XANES of SCP-2L A100C-**1**-**P3**-Rh to model Rh complexes (see SI: Figure S10) suggested the loss of carbonyl functionalities from the Rh(acac)$(CO)_2$ precursor. This was further supported by the lack of CO stretches in the IR and fitting the EXAFS data of SCP-2L A100C-**1**-**P3**-Rh (see SI: Table S1).

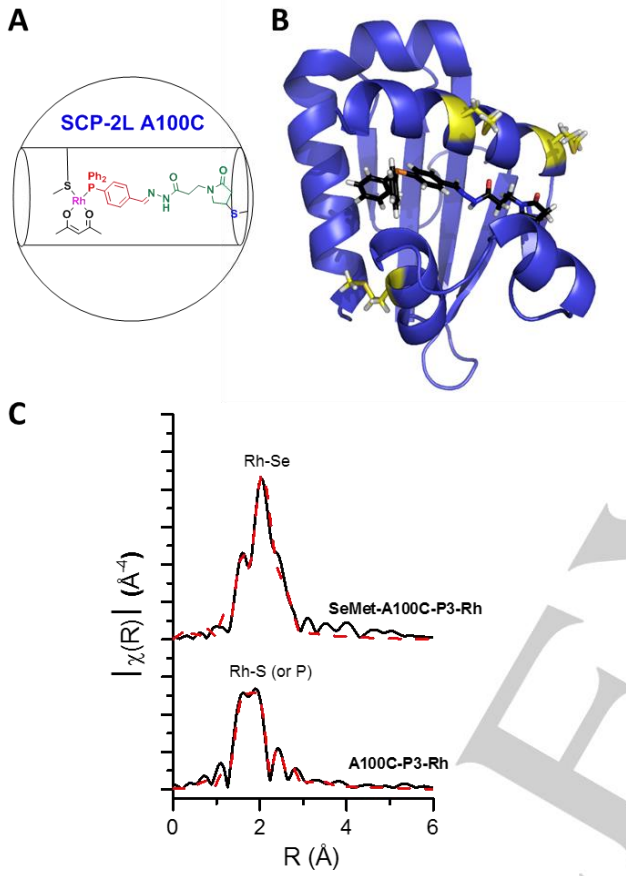


Figure 3. (A) Cartoon of the hypothesised metal environment. (B) Model of A100C (Swiss model) with **1**-**P3** docked using Gold to illustrate the reach of this added cofactor. Methionines within range are highlighted in yellow. (C) k3 weighted Rh K edge EXAFS data (black lines) and associated fit (dashed red line) for A100C-**1**-**P3**-Rh and SeMet-A100C-**1**-**P3**-Rh.

The fitting model applied used characteristic scattering paths from both acac and PPh_2Ar ligands, with a derived Rh-P coordination number of two. EXAFS is unable to distinguish between scattering neighbours of like atomic number, especially where Z = ± 1. The prospect of two phosphorous atoms coordinated to the rhodium appeared unlikely due to the resulting steric congestion of placing two protein scaffolds around the metal. Thus, we postulated the possibility of one of the observed Rh-P neighbours arising from a Rh-S interaction (Figure 3b); the protein scaffold contains functionalised sulfur in the forms of methionine, at the N-terminus and on the flexible alpha helices, and the introduced cysteine (Figure 3a).

To probe the possibility of a near-by methionine to the Rh center, the selenomethionine derivative of the protein scaffold were expressed and purified. Rhodium complexation of the phosphine modified SeMet protein using Rh(acac)$(CO)_2$ gave similar results as for SCP-2L A100C-**1**-**P3**, exhibiting the RhCO adducts in the mass spectrometry, and also showed a small shift in the ^{77}Se NMR upon Rh addition (see SI: Figure S11). There was a profound change in the Rh K edge EXAFS data of SeMet-A100C-**1**-**P3**-Rh compared to SCP-2L A100C-**1**-**P3**-Rh (Figure 3c). The differences observed can be rationalised by the coordination to a neighbour of higher atomic number with a larger backscattering amplitude; between the analogous systems Rh-S interactions have been replaced by Rh-Se. The EXAFS analysis supports our hypothesis that there is monophosphine coordination, coupled with further interaction with S from a methionine residue. Moreover, through observation of this Rh-S/Se interaction we have direct evidence of the coordination of the protein scaffold with the Rh centre.

We were intrigued if the methionine coordination would have an effect on the catalytic performance of the ArM, or if the methionine would just decoordinate under the reaction conditions to give the same active catalyst, and thus same catalytic results in all cases. To probe this further 4 mutants of SCP-2L A100C were prepared in which each Met residue was replaced by alanine, and the proteins then modified as above. The initial catalytic results are ambiguous, the activity of SCP-2L A100C M105A-**1**-**P3** was on average higher than A100C or the other mutants (TON = 112,(±33) and TON 60-86 for the others, see SI: Table S10) which could indicate that M105 is indeed involved in rhodium coordination. However, this can also be due to other factors like structural changes or decreased protein stability. More detailed studies are required before firm conclusions can be drawn on the role of the methionines in the hydroformylation by these ArMs.

In summary, we show that rhodium phosphine modified SCP-2Ls are linear-selective catalysts in the hydroformylation of long chain alkenes. SCP-2L A100C-**1**-**P3**-Rh showed a rate-enhancement of at least 10^3 compared to the traditional Rh/TPPTS system in the biphasic hydroformylation of 1-octene and 1-decene. This demonstrates that a protein chosen for its specific binding properties can be converted into an enzyme in which these properties are used to transmit product selectivity. Combining this technology with the recent advances in chemical biology will allow us to rapidly engineer highly selective catalysts that operate under benign conditions. Moving forward we believe this approach has the potential to be used for a whole range of reactions, which traditionally use phosphines as ligands and convert these into biocatalytic processes. In the long term, as chemogenetic optimisation is used to improve activity, this could open the door to a new era of biocatalytic chemical production.

Acknowledgements

This work was supported by the EU through NEST Adventure STREP Project Artizymes (contract no. FP6-2003-NEST-B3 15471); Marie Curie Excellence Grants, Artizyme Catalysis, (contract no. MEXT-2004-014320), and through a Marie Curie

Individual Fellowship project ArtOxiZymes to AGJ (contract no. H2020-MSCA-IF-2014-657755). The authors thank the EPSRC for funding through EPSRC Critical mass grant 'Clean catalysis for sustainable development' (EP/J018139/1) and Sasol (CASE studentship to P.J.D.). The UK Catalysis Hub is kindly thanked for resources and support provided via our membership of the UK Catalysis Hub Consortium and funded by EPSRC (grants EP/K014706/2, EP/K014668/1, EP/K014854/1, EP/K014714/1 and EP/M013219/1). We acknowledge Diamond Light Source for beamtime through award SP10306, and thank the beamline scientists Dr Diego Gianolio and Dr Giannantonio Cibin. We thank Dr Sally Shirran and Dr Catherine Botting for advice on mass spectrometry, and Dr Thomas Lebl for help with ^{77}Se NMR. We also thank Dr T. Glumoff (University of Oulu, Finland) for providing plasmids. A.G.J thanks Dr Irshad Baig for assistance with WT SCP-2L protein production. P.J.D thanks Yiteng Zhang for assistance.

Keywords: Artificial metalloenzymes • Catalyst design • Hydroformylation • Phosphines • Rhodium

[1] *Contemporary Catalysis: Science, Technology and Applications*. (Eds.: P. C. J. Kamer, D. Vogt, J. Thybaut), RSC, Cambridge, **2017**.

[2] a) J. C. Lewis, *ACS Catal.* **2013**, *3*, 2954-2975; b) F. Rosati, Roelfes, G. *ChemCatChem* **2010**, *2*, 916-927; c) A. Chatterjee, H. Mallin, J. Klehr, J. Vallapurackal, A. D. Finke, L. Vera, M. Marsh, T. R. Ward, *Chem. Sci.* **2016**, *7*, 673-677; d) P. Srivastava, H. Yang, K. Ellis-Guardiola, J. C. Lewis, *Nat. Commun.* **2015**, *6*, 7789, doi:10.1038/ncomms8789; e) Y. Lu, N. Yeung, N. Sieracki, N. M. Marshall, *Nature* **2009**, *460*, 855-862; f) O. Pàmies, M. Diéguez, J-E. Bäckvall, *Adv. Synth. Catal.* **2015**, *357*, 1567-1586.

[3] a) T. Heinisch, T. R. Ward, *Acc. Chem. Res.* **2016**, *49*, 1711-1721; b) F. Nastri, M. Chino, O. Maglio, A. Bhagi-Damodaran, Y. Lu, A. Lombardi, *Chem. Soc. Rev.* **2016**, *45*, 5020-5054.

[4] a) T. R. Ward, *Acc. Chem. Res.* **2011**, *44*, 47-57; b) H. M.Key, P. Dydio, D. S. Clark, J. F. Hartwig, *Nature*, **2016**, *534*, 534-537.

[5] a) P. J. Deuss, G. Popa, C. H. Botting, W. Laan, P. C. J. Kamer, *Angew. Chem. Int. Ed.* **2010**, *49*, 5315-5317; b) *Bioconjugate Techniques*, 2nd ed. (Ed. G. T. Hermanson), Acadamic Press, San Diego, **2008**.

[6] H. Bahrmann, S. Bogdanovic, P. W. N. M. van Leeuwen in *Aqueous Phase Organometallic Catalysis*, 2nd ed. (Eds.: B. Cornils, W. Herrmann), Wiley VCH, Weinheim, **2004**, pp. 391-409.

[7] a) R. den, Heeten, B. K. Muñoz, G. Popa, W. Laan, P. C. J. Kamer, *Dalton Trans.* **2010**, *39*, 8477-8483; b) W. Laan, B. K. Muñoz, R. den Heeten, P. C. J. Kamer, *ChemBioChem* **2010**, *11*, 1236-1239.

[8] a) M. Marchetti, G. Mangano, S. Pagnelli, C. Botteghi, *Tetrahedron Lett.* **2000**, *41*, 3717-3720; b) C. Bertucci, C. Botteghi, D. Giunta, M. Marchetti, S. Paganelli, *Adv. Synth. Catal.* **2002**, *344*, 556-562.

[9] Q. Jing, R. J. Kazlauskas, *ChemCatChem* **2010**, *2*, 953-957.

[10] H. M. Key, D. S. Clark, J. F. Hartwig, *J. Am. Chem. Soc.* **2015**, *137*, 8261-8268.

[11] E. Wiebus, B. Cornils, in *Catalyst Separation, Recovery and Recycling. Chemistry and Process Design* (Eds. D. J. Cole-Hamilton, R. P. Tooze) Springer, The Netherlands, **2006**, pp. 105.

[12] C. D. Frohning, C. W. Kohlpaintner, H.-W. Bohene in *Applied Homogeneous Catalysis with Organometallic Compounds*, Vol 1. 2nd ed. (Eds. B. Cornils, W. A. Herrmann) Wiley VCH, Weinheim, **2002** pp. 29-103.

[13] A. M. Haapalainen *et al. J. Mol. Biol.* **2001**, *313*, 1127-1138.

[14] P. J. Deuss, G. Popa, A. M. Z. Slawin, W. Lann, P. C. J. Kamer, *ChemCatChem* **2013**, *5*, 1184-1191.

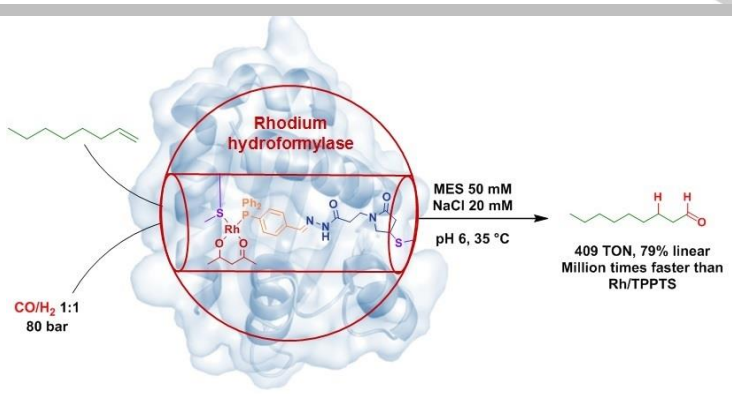
[15] *Rhodium Catalysed Hydroformylation* (Eds. Van Leeuwen, P. W. N. M., & Claver, C.) Kluwer Academic Publishers, New York, **2002**.

Entry for the Table of Contents (Please choose one layout)

Layout 1:

COMMUNICATION

Artificial metalloenzymes for the linear hydroformylation of long chain alkenes have been prepared which show enzyme like rate enhancements.



A. G. Jarvis, L. Obrecht, P. J. Deuss, W. Laan, E. K. Gibson, P. P. Wells, and P. C. J. Kamer.**

XX-XX

Enzyme activity by design: an artificial rhodium hydroformylase for linear aldehydes

Layout 2:

COMMUNICATION

*Author(s), Corresponding Author(s)**

Page No. – Page No.

Title

Text for Table of Contents