Engineering Thermostability in Artificial Metalloenzymes to Increase Catalytic Activity

Engineering Thermostability in Artificial Metalloenzymes to Increase Catalytic Activity

Megan V. Doble[a], Lorenz Obrecht,[a] Henk-Jan Joosten[b], Misun Lee[c], Henriette J. Rozeboom[c], Emma Branigan,[a] James. H. Naismith[a, d], Dick B. Janssen[c], Amanda G. Jarvis[a,e], Paul C. J. Kamer[a,f#]*

AUTHOR ADDRESS [a] School of Chemistry, University of St Andrews, KY16 9ST, United Kingdom. [b] Bio-Prodict, Nieuwe Marktstraat 54E, 6511 AA Nijmegen, The Netherlands. [c] Biotransformation and Biocatalysis, Groningen Biomolecular Science and Biotechnology Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands. [d] Rosalind Franklin Institute, Harwell Campus, Didcot, OX11 0FA United Kingdom. [e] School of Chemistry, Joseph Black Building, David Brewster Rd, Kings Buildings, University of Edinburgh, Edinburgh, EH9 3FJ.[f] Bioinspired Homo- & Heterogeneous Catalysis, Leibniz Institute for Catalysis, Albert-Einstein-Straße 29 a, 18059 Rostock, Germany.

[#]Professor Paul Kamer passed away November 19, 2020.

Corresponding author: *amanda.jarvis@ed.ac.uk

Protein engineering has shown widespread use in improving the industrial application of enzymes and broadening the conditions they are able to operate under by increasing their thermostability and solvent tolerance. Here we show that protein engineering can be used to increase the thermostability of an artificial metalloenzyme. Thermostable variants of the human steroid carrier protein 2L, modified to bind a metal catalyst, were created by rational design using structural data and a 3DM database. These variants were tested to identify mutations that enhanced the stability of the protein scaffold and a significant increase in melting temperature was observed with a number of the modified metalloenzymes. The ability to withstand higher reaction temperatures resulted in increased activity in the hydroformylation of 1-octene, with >5-fold improvements in turnover number (TON), while the selectivity for linear aldehyde remained high up to 80%.

KEYWORDS Artificial Metalloenzyme, Hydroformylation, Bioengineering, Protein Thermostability, Biocatalysis, Bioinformatics

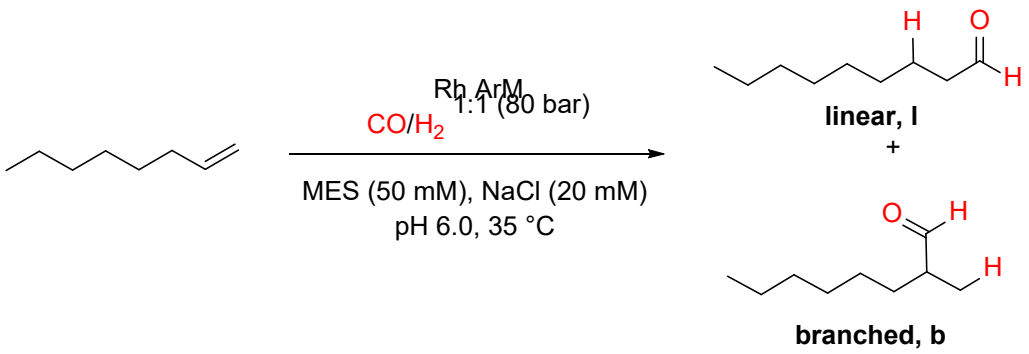
Introduction

Over the last decades, the power of biocatalysis in chemical synthesis, particularly within an industrial setting, has become undeniable. Enzymes such as transaminases and ketoreductases are increasingly used in the synthesis of chiral intermediates in pharmaceutical processes.[1-4] The use of biocatalytic technology to replace chemical steps as part of existing drug syntheses is attractive due to its simple nature and beneficial sustainability metrics.[5,6] The recent development of artificial metalloenzymes (ArMs) has added to the potential of biocatalysis by providing an elegant approach to combining enzymatic and chemical methods, thereby expanding the catalytic toolbox.[7-11] ArMs combine the molecular recognition properties of proteins, which are responsible for the high selectivity observed in many enzymatic processes, with the industrially relevant reactivity of homogeneous transition metal catalysts. The introduction of synthetic transition metal cofactors provides catalytic versatility by allowing unnatural reactions to be introduced into the biocatalysis arena. To date, a whole range of ArMs has been developed for unnatural reactions including transfer hydrogenation[12], hydroformylation,[13, 14] metathesis,[15] cross coupling reactions[16] and other carbon-carbon bond forming reactions such as the Diels-Alder reaction.[17-19] Nonetheless, the combination of proteins with organometallic catalysis remains challenging due to the different and contrasting conditions often required for efficient catalysis.[20] Proteins have evolved over time to work in a biological environment; therefore, harsh conditions, such as high temperatures and the presence of organic solvents favored in chemocatalysis, often lead to protein denaturation and deactivation of the catalyst.[21] Within the realm of biocatalysis, enzymes are nowadays routinely engineered to increase their stability to organic solvents and temperature, leading to more active enzymes.[22, 23] One example showed that by using a computationally engineered variant with strongly improved thermostability, a peptide amidase could be applied in an organic solvent to achieve versatile peptide C-terminal functionalization.[24] Based on such documented examples we hypothesized that by enhancing the thermostability of the protein scaffold it would be possible to increase the activity of ArMs and their tolerance to organic solvents.[25] This would expand the repertoire of possible reactions that ArMs could conduct and increase the experimental design space they can operate in.

A successful strategy used to engineer thermostability into a protein is the rational design of mutations based on structural information. Such mutations are focused on flexible regions in the protein with the aim to reduce local unfolding.[26] Reduced flexibility can be achieved by a number of strategies including the introduction of disulphide bridges to stabilize early unfolding regions or destabilize the unfolded state by reducing its entropy.[27] Another strategy to enhance stability is the introduction of surface-located charged groups that can form salt bridges.[28] Such electrostatic interactions contribute to local stability, so it can be beneficial to introduce these into flexible regions within the protein scaffold.[29] Several studies show that modifying other electrostatic interactions, such as the removal of isolated surface charges, also contributes to the stability of a protein.[30, 31] Another approach to engineering thermostability employs phylogenetic or structure-based multiple sequence alignments to identify mutations and sequences that match the consensus

sequence of a protein family. This approach is based on the hypothesis that the consensus amino acids (i.e. the most abundant amino acids at corresponding positions within a large set of aligned sequences or structures) generally add more to the stability of the protein than non-consensus amino acids.[32]

The Kamer group has previously reported covalent approaches for synthesizing ArMs containing metal binding ligands via a maleimide linker to unique cysteine residues in the protein scaffold human steroid carrier protein SL (SCP-2L).[33] The SCP-2L protein is a 120 amino acid domain from the multifunctional enzyme type 2 (MFE-2, involved in β-oxidation) and was specifically selected for its ability to bind linear substrates in its hydrophobic tunnel.[34] Cysteine residues for protein modification were introduced at various locations within the tunnel including V83C and A100C to take advantage of substrate binding in the tunnel.[13, 35] Hybrid catalysts based on these SCP-2L mutants in combination with phosphine-liganded Rh cofactors showed high activities and selectivities for the production of long chain linear aldehydes in the rhodium-catalyzed hydroformylation of alkenes, under benign aqueous conditions (Scheme 1).[13] This reaction is of industrial interest due to the high added-value of the aldehyde products in comparison to the alkene substrates. The catalyst performance is highly sensitive to the structure of the phosphine ligands, coordination number and coordination mode.[36] Typically, highest selectivities for the desired linear aldehyde are obtained with two phosphine ligands coordinated in the equatorial plane of the trigonal bipyramidal $RhH(CO)_2L_2$ catalyst resting state.[37] Monoligated complexes show generally much lower selectivities, and for ligand-free rhodium carbonyls the linear to branched ratios go down to 1 :1. In industry, rhodium-catalyzed hydroformylation is used on a large scale under biphasic conditions, enabling the recovery of the expensive rhodium catalyst. However, this reaction is challenging for longer alkenes (>5 carbon atoms) due to their low solubility under aqueous conditions.[38] ArMs provide a catalyst for longer alkene substrates that can work at low temperatures (<35 °C) compared to industrial conditions (typically 110-120 °C) and also offer high selectivity and turnover numbers, overall providing a much greener reaction. We set out to examine the use of sequence-based / rational protein engineering techniques to increase the stability of our artificial hydroformylase and thus increase the reaction temperature and the rate of hydroformylation.



Scheme 1. Rhodium catalyzed hydroformylation of alkenes to aldehyde products.

Results and Discussion

The design of mutations that could increase stability of SCP-2L was carried out both using rational design and sequence-based mutant selection with an SCP-like family 3DM database.[39] The 3DM databases consist of structure-based alignments of homologous sequences and assigns each residue a position number that allows comparison of positions among proteins that are structurally related within the protein family.[40] Both approaches require a 3D structure; therefore we decided to solve the crystal structures of two variants that carry an introduced cysteine for attachment of the artificial metal cofactor. The variants chosen were SCP-2L V83C and SCP-2L A100C which had been successfully used previously by the group in the development of hybrid catalysts for hydroformylation.[13] These mutants contained unique cysteines at either end of the hydrophobic tunnel for the introduction of the catalytic rhodium–phosphine complexes (Figure 1).

SCP-2L V83C and SCP-2L A100C were produced as previously described.[35] Structures of the two variants were determined by X-ray crystallography (see Supporting Information, pdb entries: 6Z1X and 6Z1W). The general fold of the wild-type SCP-2L with an α/β-fold consisting of five β-strands and five α-helices (PDB 1IKT)[34] is preserved, thus the mutations lead to no structural change. The structural alignment between the variants and wild-type SCP-2L have a root-mean square deviation (RMSD) of 0.3 - 0.5 Å over 115 aligned residues with 99.1% sequence identity.

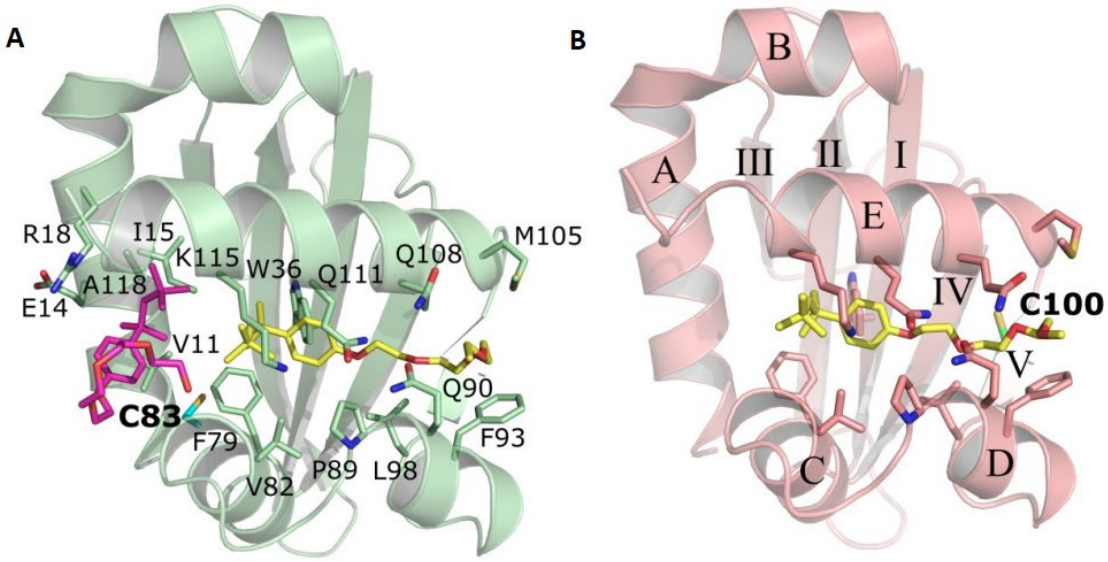


Figure 1. Crystal structures of the V83C (panel A, pdb: 6Z1X) and A100C (panel B, pdb: 6Z1W) mutants of SCP-2L The ligand bound to the protein is Triton X-100 (2-[4-(2,4,4-trimethylpentan-2-yl)phenoxy]ethanol), shown in yellow or magenta sticks. The β-strands and α-helices are indicated by roman numerals (I-V) and capitals (A-E), respectively. The mutated residues are shown in cyan (Structure A) and green (Structure B).

Both mutants contain Triton X-100, an analog of a lipid molecule observed in the wild-type structure with similar interactions. Mutant V83C comprises two Triton X-100 molecules with the (1,1,3,3-tetramethyl)butyl parts at ~4.0 Å from each other (Figure 1A). The second Triton X-100 has hydrophobic interactions with the sidechains of V11, I15, C83 and A118, and with the hydrophobic part of the sidechains of E14 and R18. The ethoxy repeats of the Triton X-100 fold back into the SCP-2L molecule. The last ethoxy repeat visible in electron density has hydrophobic contacts with the phenyl group. The introduced C83 is situated on the C-terminal part of helix C and the new C100 is in the middle of strand V (Figure 1B). The sulfur atom of C100 has hydrophobic contacts with one of the methyl groups of the butyl chain of the retained Triton X-100.

To stabilize the protein, we evaluated a number of mutation strategies, including stabilizing flexible loop regions by salt bridge formation, strengthening α helix dipoles, and mutations to introduce more conserved amino acids identified from 3DM database alignments of the SCP-2L protein family. Stabilisation by introduction of disulfide bridges was avoided because of the requirement of a unique cysteine for modification of the protein to introduce the metal cofactor and their incompatibility with phosphines.

Using the 3DM database and structural analysis of the protein, positions were identified where mutations towards consensus could be introduced without any expected loss of structural integrity (Figure 2). We excluded the four helices which surround the hydrophobic tunnel from mutagenesis

because these are involved in binding of ligands such as long-chain fatty acyl-CoAs[41] and cholesterol derivatives.[42] Mutations were especially sought in three loops that appear more flexible as concluded from crystallographic B factors, which reflect fluctuation of atoms from the average position due to protein dynamics.

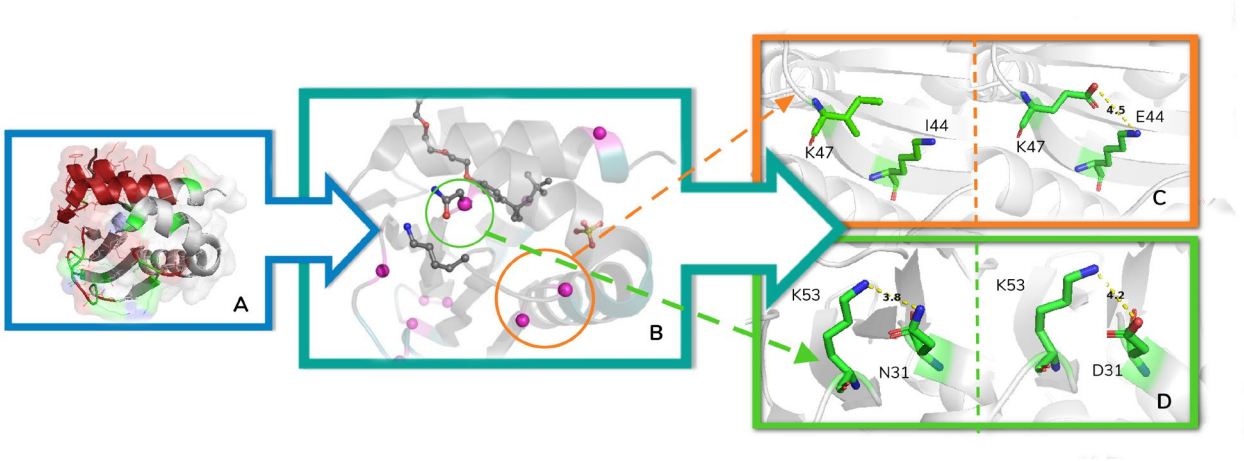


Figure 2. Position of mutations in SCP-2L (pdb 1IKT). A) Positions and regions to consider for mutagenesis are highlighted in green, positions and regions to avoid are indicated in red. B) Positions of confirmed stabilizing mutations highlighted in pink spheres. C) Introduction of a possible E44-K47 salt bridge by mutation I44E. D) Introduction of a possible salt bridge K53-D31and strengthened dipole by mutation N31D. Image created in PyMOL.

Using the 3DM database we found mutations that enable the introduction of more conserved residues within SCP-2L. Residue V26 is positioned within an α helix and is 15.5% conserved in the SCP-like 3DM database (Figure S1). However, F26 is the more abundant residue in the whole protein family (27.5% conserved). The V26F mutation also introduces an aromatic group that could create π-stacking interactions. The following residues were also identified for mutation to more conserved residues: K40, G41, S54, and D116 (Table 1). Whilst residue A68 is the most conserved at 70%, proline is found in the same position for most of the remaining proteins (21%) and was thus also chosen as a conservative mutation.

We also explored the protein structure to identify residues that would lead to the formation of salt bridges that can stabilize flexible loops and regions on the proteins surface. Residue I44 is positioned within a loop between β-sheets in a flexible region and creating a salt bridge here was anticipated to increase stability. It is also in close proximity to K47 and by introducing a E or D an electrostatic interaction can be obtained (Figure 2C). Residues S56, K58, K65 and E81 were also identified as sites for salt bridge formation, which can be achieved by reversing their charge to create new ionic interactions.

The alignment of the dipoles of peptide bonds parallel to the axis of an α-helix causes a net dipole moment with its positive pole at the amino terminus and negative pole at the carboxy terminus. The total dipole moment of the entire helix is due to individual dipoles of the C=O groups involved in hydrogen bonding.[43] Phosphate moieties bind frequently at the positive N-termini of helices in proteins. Introducing a charged residue at the C-terminal positions can stabilize a helix by hydrogen bonding to a main-chain CO group.[44] Residue N31 is present on the C-terminus of α-helix 1 of SCP-2L with the side chain close to K53. To strengthen the interaction with the positively charged lysine we introduced mutation N31D (Figure 2D). The stronger dipole and the formation of a new D31-K53 salt bridge were expected to increase protein stability.

Site directed mutagenesis[45] was performed both for SCP-2L A100C and V83C to separately obtain plasmids carrying the mutations G41D, S56D, K58E, K58D, I44E, I44D, E81K, S54N, S56E, D116K, I44D, K40N, A68P, K65E, N31D, or V26F (Table 1). All mutants were confirmed by DNA sequencing. Using circular dichroism (CD) spectroscopy apparent melting temperatures ($T_{m,app}$) values were measured [46] and compared with the parent proteins SCP-2L A100C and V83C (Fig. S4). The CD method used a fast scan approach by measuring at 6 different temperatures. Reversibility was not established and therefore the $T_{m,app}$ values are approximate since they can be influenced by incubation conditions; the values serve the purpose of indicating how the thermostability is affected by the mutations. The results showed that several mutants had a large improvement in thermostability over the original SCP-2L A100C and V83C proteins. The $T_{m,app}$ values increased by up to 12 and 10 °C for mutants constructed in an A100C and V83C background, respectively. In general, the A100C protein template was more responsive to the designed mutations. The introduction of N31D gave an increase of 12 °C for A100C, but only a 3 °C increase for V83C (Table 1, entry 1). The introduction of residues likely to form salt bridges raised the $T_{m,app}$ of both mutants, with both E81K and S56D giving a large increase in $T_{m,app}$. The $T_{m,app}$ of mutant A100C increased by 9 and 11 °C for E81K and S56D, respectively, and with V83C increases of 6 and 10 °C were observed (Table 1, entries 10, 4). Mutation S56D improved the $T_{m,app}$ with both templates whereas S56E led to a decrease in $T_{m,app}$ for both mutants. Introducing more conserved amino acids based on the results found using the 3DM database led to more modest increases in $T_{m,app}$. An exception was the G41D mutation in A100C which increased the $T_{m,app}$ by 7 °C. In comparison, this mutation in V83C led to a decrease in $T_{m,app}$ of 6 °C (Table 1, entry 13).

Table 1. Effect of designed mutations on thermostability of SCP-2L V83C and A100C.

Entry	Design method	Mutation	Template	
			A100C	V83C
			$\Delta T_{m,app}$ (°C)	$\Delta T_{m,app}$ (°C)
1	Stabilizing α-helix dipole	N31D	12	3
2	Introducing a salt bridge	I44D	2	-1
3		I44E	4	1
4		S56D	11	10
5		S56E	-2	-7
6		K58D	5	n.p.
7		K58E	u	1
8		K65D	n.p.	n.p.
9		K65E	7	1
10		E81K	9	6
11	Towards consensus mutations	V26F	u	3
12		K40N	4	3
13		G41D	7	-6
14		S54N	n.p.	n.p.
15		A68P	3	4
16		D116K	2	0

n.p, no data due to lack of protein; u, unreliable $T_{m,app}$ obtained.

There is ample evidence in the literature that combining mutations that increase thermostability can give an additive effect.[47] Therefore, the most favourable mutations from the first round of rational design (S56D, E81K, N31D) were combined in different ways and six double mutants were produced according to the same protocol as for the single mutants. Using CD spectroscopy, the melting temperature ($T_{m,app}$) of V83C and A100C and the thermostable double mutants were measured using data points taken with 1°C intervals at 222 nm (Table 2, Figure 3). This CD method provides a more accurate value than the fast scan approach used in the previous screen as a larger number of data points are utilised.

Table 2. Apparent melting temperatures ($T_{m,app}$) of SCP-2L mutants determined using CD, measuring the absorbance at 222 nm at 1°C steps

Variant	$T_{m,app}$ (°C)	ΔT (°C)
V83C template	48	
A100C template	42	
A100C+N31D+S56D	41	-1
A100C+N31D+E81K	55	+13
A100C+E81K+S56D	58	+16
V83C+N31D+S56D	60	+12
V83C+N31D+E81K	62	+14
V83C+E81K+S56D	64	+16

The designed double mutants showed impressive improvements in thermostability, increasing the $T_{m,app}$ by 12-16 °C in all but one case (Table 2). Variant A100C+N31D+S56D revealed a negative effect of combining the N31D and S56D mutations and the $T_{m,app}$ decreased by 1 °C. This protein was obtained in very low yield (2 mg/L) so it was not used further in this work.

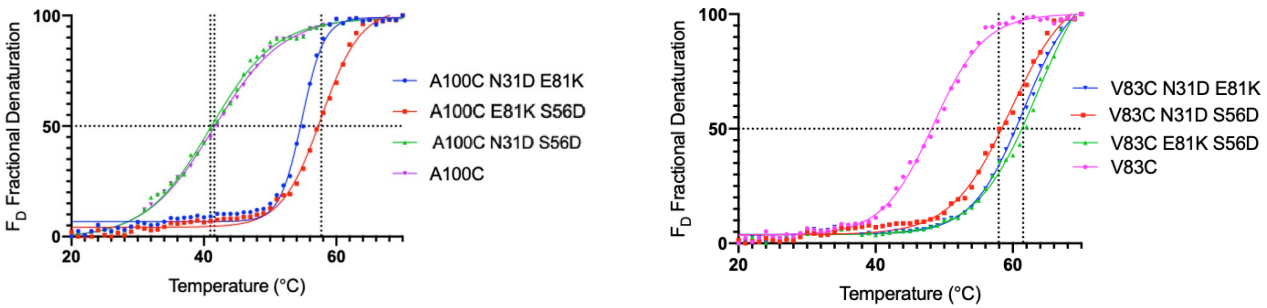
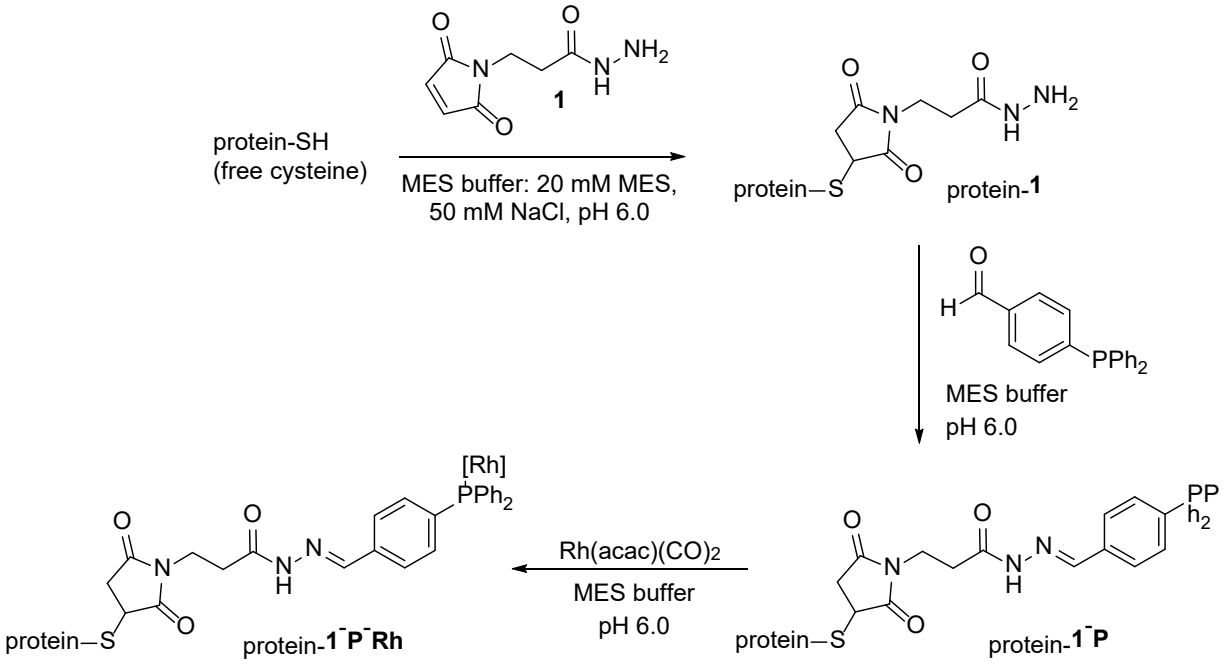


Figure 3. Thermal denaturation curves of SCP-2L double mutants

These results show that rational design was successfully used to increase the stability of the SCP-2L-derived ArM protein hosts. Next, we examined if these stabilized scaffolds were suitable for creating more robust catalytic ArMs. Hydroformylation was chosen as a benchmark reaction because it is the most studied reaction using SCP-2L-derived ArMs. Previous results reported by the Kamer group showed that SCP-2L A100C and V83C rhodium proteins showed hydroformylation activity reaching up to 400 and 100 turnover numbers (TONs), respectively,

over a 48 h reaction period.[13] These activities were surprisingly high considering that the reactions were carried out at 35 °C. With a more stable protein scaffold, the activity might increase further as less catalyst degradation will occur over time allowing the use of higher temperatures.

The proteins were treated as previously described to give the phosphine-modified proteins, SCP2L X-**1**-**P**. Next, the rhodium metalloproteins for all 5 triple mutants (SCP-2L X variants with mutations X = A100C+N31D+E81K, A100C+E81K+S56D, V83C+N31D+E81K, V83C+E81K+S56D or V83C+N31D+S56D) were obtained by the addition of $Rh(acac)(CO)_2$. The resulting proteins were analysed by LC-MS to confirm modification was effective (Scheme 2, see SI for details).[33]



Scheme 2. Synthesis of the artificial SCP-2L-derived metalloproteins.

The hydroformylation activities of each of the five double mutants were investigated using 1-octene as the substrate and compared to SCP-2L A100C-**1-P-Rh** and V83C-**1-P-Rh** (Scheme 1). The reactions were carried out under the same conditions as the previous work except for the reaction time, which was shortened to 16 h (Table 3). Similar linear selectivities were observed as previously reported and TONs were as expected for shorter reaction times (entries 1 and 9).[13] As seen previously, the use of the ArM gave higher linear selectivity (~80%) than just $Rh(acac)(CO)_2$ which exhibits only a slight preference for the linear aldehyde (Table 3, entry 15). The higher TON

recorded for $Rh(acac)(CO)_2$ is consistent with Rh leaching into the organic layer in the absence of the phosphine-modified protein.

All three of the new thermostable SCP-2L V83C mutants (V83C+E81K+S56D, V83C+N31D+E81K, V83C+N31D+S56D) showed an improvement in TONs compared to SCP-2L V83C at 35 °C. SCP-2L V83C E81K S56D gave a TON of 92 (TOF 6.5 h^{-1}), more than double

Table 3. Hydroformylation of 1-octene catalyzed by SCP-2L derived rhodium-hydroformylases[a]

Entry	Mutant	Temp °C	16 h		48 h	
			Linear selectivity[b] (SD)	TON[c] (SD)	Linear selectivity[b] (SD)	TON[c] (SD)
1	V83C	35	79 (0.2)	43 (0.3)	78 (1.2)	75.7 (1.3)
2		45	78 (0.1)	12 (1.3)	76 (1.0)	27.3 (1.0)
3	V83C+E81K+S56D	35	78 (0.1)	92 (1.0)	80 (0.7)	221 (1.9)
4		45	80 (0.1)	190 (1.1)	79 (0.9)	385 (3.9)
5	V83C+N31D+E81K	35	80 (0.1)	104 (0.9)	80 (1.1)	199 (2.0)
6		45	78 (0.3)	201 (0.9)	78 (0.7)	404 (4.2)
7	V83C+N31D+S56D	35	75 (0.1)	99 (1.0)	78 (1.9)	204 (2.8)
8		45	77 (0.1)	180 (1.2)	80 (0.8)	402 (4.3)
9	A100C	35	80 (0.1)	211 (0.7)	79 (0.8)	401 (3.9)
10		45	75 (0.2)	99 (2.9)	70 (1.6)	171 (2.7)
11	A100C+E81K+S56D	35	79 (0.7)	202 (1.3)	77 (1.2)	406 (5.2)
12		45	78 (1.1)	322 (3.1)	73 (0.9)	415 (3.9)
13	A100C+N31D+E81K	35	72 (0.2)	213 (0.4)	72 (1.3)	403 (3.5)
14		45	73 (0.5)	346 (2.1)	73 (1.3)	326 (2.8)
15	$Rh(acac)CO_2$ (no protein) **Error! Bookmark not defined.**	35			55 (0.7)	530 (53)

[a]Conditions: 80 bar syngas (1:1), stirring 625 rpm, 16 h in degassed MES buffer, 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. Results are the average of three repeated reactions, with standard deviation (SD) between brackets. Yields determined by GC. [b] Linear selectivity refers to the amount of linear aldehyde (%). [c] TON based on the Rh concentration of the catalyst solution, measured by ICP-MS.

that seen with SCP-2L V83C under the same conditions (TON 43, TOF 2.9 h^{-1}, Table 3, entry 3 vs. 1). This result, alongside the visual identification of precipitated protein after the reaction for SCP-2L V83C only, suggests that significant denaturation occurred when using SCP-2L V83C as the protein scaffold and that increasing the thermostability of the scaffold protein increased the performance of the catalyst under these reaction conditions. Additional evidence was provided by comparing the CD spectra before and after the reaction. For V83C an indistinct spectrum was obtained after the reaction, whilst for V83C N31D E81K the far UV CD spectra still showed the expected features (see SI Fig. S10 and S11). To test this further, the reaction temperature was increased to 45 °C, the hypothesis being that rates would increase at higher temperatures if the protein remains stable, and therefore a larger TON would be observed. For SCP-2L V83C (Table 3, entry 2) a significantly reduced TON was observed, most likely due to rapid denaturation of the protein. However, the new thermostable rhodium-complexed proteins remained active at this higher temperature and produced high TONs of up to 200 (Table 3, entry 6) thus showing the expected increase in rate on increasing temperature. The stabilized variants showed an approximate doubling of activity upon a 10 °C increase in reaction temperature, i.e. for V83C+E81K+S56D TOFs of 5.75 *vs* ~11.9 h^{-1} were observed (Table 3 entry 3 vs 4, see Table S1 for data as TOF). For all V83C mutants the selectivity towards the linear aldehyde remained high (75- 80%), and thus provides evidence that the designed mutations did not interfere with the active site of the catalysts.

The two new thermostable A100C-derived mutants (A100C N31D+E81K, A100C E81K+S56D) showed no improvement but consistent results with the parent A100C. However, when the reaction temperature was increased to 45 °C these rhodium proteins were still stable, unlike the parent A100C, and showed increased TONs from 202 and 213 to 322 and 346 for A100C E81K+S56D and A100C N31D+E81K, respectively (Table 3, entries 11 vs 12 and 13 vs 14). This increase in TON is somewhat lower than what would be expected by increasing the reaction temperature by 10 °C, indicating that the newly designed A100C mutants are at the limit of their stability which is also indicated by the relatively low increase in TON after longer reaction times (Table 3, entry 12 and 14). The linear selectivity of the catalysts was not significantly affected by the mutations, though a small decrease in selectivity was observed for A100C N31D+E81K.

To test the long-term stability of the artificial metalloenzymes, the reaction time was further increased to 48 h to see if the TON could be improved (Table 3). As was seen after 16 h, the new thermostable V83C mutants showed an increase of TON from the original 76 of V83C up to 221 for V83C+E81K+S56D at 35 °C (Table 3 entries 1 and 3, respectively) representing an almost tripling of rate (TOF 1.6 vs 4.6 h^{-1}). When the temperature was raised to 45 °C the TONs doubled for both V83C+N31D+E81K and V83C+N31D+S56D giving rise to an expected increase in TOF from ~4 h^{-1} to 8 h^{-1} (Table 3 entry 8, for TOFs see Table S1). This suggests that the SCP-2L V83C-derived thermostable mutants are capable of retaining their structure at this higher temperature.

However, comparing the average reaction rate over 16 h and 48 h for all the V83C proteins, a drop of ~30% over the second timeframe is observed in all cases, regardless of temperature. Thus, whilst thermostability has increased, some loss of activity over during the reaction still occurs, potentially due to exposure of the protein to mechanical stress through stirring and pressure.

In comparison to the C83 derivatives, the A100C-derived thermostable mutants showed less improvement in activity over the longer 48 h reaction period. This matched observations from the 16 h reactions which already implied that after 16 h the protein scaffold begins to denature, as shown by a large build-up of white precipitate. This indicates that inactivation is accompanied by aggregate formation and thus is irreversible. The loss of catalyst activity over the 48 h reaction period at 45 °C resulted in a drop of the average TOF for the A100C+N31D+E81K enzyme of almost 70%, i.e. from 21.6 h^{-1} over a 16 h reaction to 6.8 h^{-1} for a 48 h reaction (Table 3, entry 14). This relatively large drop in activity does not appear to be related to thermostability since the mutations introduced in the C100 and C83 scaffolds gave similar $\Delta_{Tm,app}$ values. Instead, loss of activity may be related to conversion since the most active variants (Table 3, entry 12 and 14) show the largest reduction of activity.

Overall, the enhanced activity was correlated to the increase in stability of the SCP-2L-derived ArMs as reflected in $T_{m,app}$ values. All of the results show high linear selectivity in hydroformylation reactions indicating the catalysis occurs due to the artificial catalytic center introduced into the protein tunnel and not from 'free Rh'. Furthermore, the linearity of up to 80% remains very high for a monoligated Rh-catalyst, corroborating the importance of the second coordination sphere of the protein scaffold.[13]

Conclusion

In conclusion, we have shown that the stability of the protein scaffold within an ArM is a crucial factor with respect to the catalytic performance. Space time yields are of crucial importance for industrial applications, especially when costly catalysts such as ArMs based on noble metals are concerned. The advantage of mild and green operating conditions of biocatalysts can be counterbalanced by low stability. By improving the thermostability of the SCP-2L V83C scaffold by up to 16 °C, we were able to increase the reactivity of our ArM at higher temperatures. With the stabilized hybrid enzymes, the TON improved over 5-fold reaching values of up to over 400. Whilst this study did not explore the increased solvent tolerance of the ArMs explicitly, the increased TON under biphasic conditions strongly indicates an increased stability in the presence of organic solvents. The increased organic solvent tolerance opens the door to study reactions of substrates that need organic cosolvents for solubility. Thus, rational mutant design for improving ArM stability yielded a biocatalyst that enables a wider range of reaction conditions and showed higher activities. In addition to other powerful tools in biocatalysis and protein engineering such

as directed evolution, this may offer an attractive approach to tailor biocatalysts to fit real industrial applications.

Funding Sources

M.V.D thanks the BBSRC for support through an EastBio studentship BB/J01446X/1. A.G.J. thanks the University of Edinburgh for support through a Christina Miller Fellowship and the EU for support through a Marie Curie Individual Fellowship, project ArtOxiZymes, (H2020-MSCA-IF-2014-657755). The authors thank the EPSRC for funding through the EPSRC critical mass grant "Clean catalysis for sustainable development" (EP/J018139/1). The UK Catalysis Hub is kindly thanked for resources and support provided via our membership of the UK Catalysis Hub Consortium, which is funded by the EPSRC (EP/K014706/2, EP/K014668/1, EP/K014854/1, EP/K014714/1, and EP/M013219/1).

Supporting information

This information is available free of charge on the ACS Publications website.
Experimental procedures, instrumental information, and product characterization. (PDF)

ABBREVIATIONS

ArM, Artificial Metalloenzyme; SCP-2L, Sterol Carrier Protein type 2- like domain.

ACKNOWLEDGMENT

We thank Dr. T. Glumoff (University of Oulu, Finland) and Prof. Dr. K. J. Hellingwerf (University of Amsterdam) for providing plasmids. In addition we thank the St Andrews Mass Spectrometry and Proteomics facility for help with the mass spectrometry instruments.

REFERENCES

1. Fuchs, M.; Farnberger, J. E.; Kroutil, W. The Industrial Age of Biocatalytic Transamination, *Eur, J. Org. Chem*. **2015**, *32*, 6965-6982. 10.1002/ejoc.201500852
2. Patel, R. N. Biocatalytic Synthesis of Chiral Alcohols and Amino Acids for Development of Pharmaceuticals, *Biomolecules* **2013**, 3, 741-777. 10.3390/biom3040741
3. Slabu, I.; Galman, J. L.; Lloyd, R. C.; Turner, N. J. Discovery, Engineering, and Synthetic Application of Transaminase Biocatalysts, *ACS Catal.* **2017**, *7*, 8263-8284. 10.1021/acscatal.7b02686
4. Moore, J. C.; Pollard, D. J.; Kosjek, B.; Devine, P. N. Advances in the Enzymatic Reduction of Ketones, *Acc. Chem. Res.* **2007**, *40*, 1412-1419. 10.1021/ar700167a
5. Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J.

Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture, *Science* **2010**, *329*, 305-309. 10.1126/science.1188934
6. Huffman, M. A.; Fryszkowska1, A.; Alvizo, O.; Borra-Garske, M.; Campos, K. R.; Canada, K. A.; Devine, P. N.; Duan, D.; Forstater, J. H.; Grosser, S. T.; Halsey, H. M.; Hughes, G. J.; Jo, J.; Joyce, L. A.; Kolev, J. N.; Liang, J.; Maloney, K. M.; Mann, B. F.; Marshall, N. M.; McLaughlin, M.; Moore, J. C.; Murphy, G. S.; Nawrat, C. C.; Nazor, J.; Novick, S.; Patel, N. R.; Rodriguez-Granillo, A.; Robaire, S. A.; Sherer, E. C.; Truppo, M. D.; Whittaker, A. M.; Verma, D.; Xiao, L.; Xu, Y.; Yang H. Design of an in vitro biocatalytic cascade for the manufacture of islatravir, *Science* **2019**, *366*, 1255-1259. DOI: 10.1126/science.aay8484
7. Schwizer, F.; Okamoto, Y.; Heinisch, T.; Gu, Y.; Pellizzoni, M. M.; Lebrun, V.; Reuter, R.; Köhler, V.; Lewis, J. C.; Ward, T. R. Artificial Metalloenzymes: Reaction Scope and Optimization Strategies, *Chem. Rev*. **2018**, *118*, 142-231. 10.1021/acs.chemrev.7b00014
8. Doble, M. V.; Ward, A. C.; Deuss, P. J.; Jarvis, A. G.; Kamer, P. C. J. Catalyst design in oxidation chemistry; from KMnO4 to artificial metalloenzymes, *Bioorg. Med. Chem*. **2014**, *22*, 5657-5677. doi: 10.1016/j.bmc.2014.07.002
9. Rosati, F.; Roelfes, G. Artificial Metalloenzymes, *ChemCatChem* **2010**, 2, 916-927. 10.1002/cctc.201000011
10. Deuss, P. J.; den Heeten, R.; Laan, W.; Kamer, P. C. J. Bioinspired Catalyst Design and Artificial Metalloenzymes, *Chem. Eur. J*, **2011**, *17*, 4680-4698. DOI: 10.1002/*chem*.201003646
11. Heinisch, T.; Ward, T. R. Design strategies for the creation of artificial metalloenzymes, *Curr. Opin. Chem. Biol.* **2010**, *14*, 184-199. doi: 10.1016/j.cbpa.2009.11.026
12. Durrenberger, M.; Heinisch, T.; Wilson, Y. M.; Rossel, T.; Nogueira, E.; Knorr, L.; Mutschler, A.; Kersten, K.; Zimbron, M. J.; Pierron, J.; Schirmer, T.; Ward, T. R. Artificial Transfer Hydrogenases for the Enantioselective Reduction of Cyclic Imines, *Angew. Chem., Int. Ed*. **2011**, *50*, 3026-3030. doi: 10.1002/anie.201007820
13. Jarvis, A. G.; Obrecht, L.; Deuss, P. J.; Laan, W.; Gibson, E. K; Wells, P. P.; Kamer, P. C. J. Enzyme Activity by Design: An Artificial Rhodium Hydroformylase for Linear Aldehydes, *Angew. Chem., Int. Ed*. **2017**, *56*, 13784-13788. DOI:10.1002/anie.201705753
14. Imam, H. T.; Jarvis, A. G.; Celorrio, V.; Baig, I.; Allen, C. C. R.; Marr, A. C.; Kamer, P. C. J. Catalytic and biophysical investigation of rhodium hydroformylase *Catal. Sci. Technol.*, **2019**, *9*, 6428-6437. DOI: 10.1039/c9cy01679a
15. Jeschek, M.; Reuter, R.; Heinisch, T.; Trindler, C.; Klehr, J.; Panke, S.; Ward, T. R. Directed evolution of artificial metalloenzymes for in vivo metathesis, *Nature* **2016**, *537*, 661-665. doi: 10.1038/nature19114
16. Chatterjee, A.; Mallin, H.; Klehr, J.; Vallapurackal, J.; Finke, A. D.; Vera, L.; Marsh, M.; Ward, T. R. An enantioselective artificial Suzukiase based on the biotin–streptavidin technology, *Chem. Sci.* **2016**, *7*, 673-677. DOI: 10.1039/c5sc03116h
17. Deuss, P. J.; Doble, M. V.; Jarvis, A. G.; Kamer, P. C. J. in Artificial Metalloenzymes and MetalloDNAzymes in Catalysis: From Design to Applications, Eds. Diéguez, M.; Bäckvall, J.-E.; Pàmies, **2018**, O. Wiley. p. 285-319. 10.1002/9783527804085.ch10
18. Reetz, M. T. Artificial Metalloenzymes as Catalysts in Stereoselective Diels–Alder Reactions, *Chem. Rec.* **2012**, *12*, 391-406. 10.1002/tcr.201100043
19. Deuss, P. J.; Popa, G.; Slawin, A. M. Z.; Laan, W.; Kamer, P. C. J. Artificial Copper Enzymes for Asymmetric Diels–Alder Reactions, *ChemCatChem*, **2013**, *5*, 1184-1191. 10.1002/cctc.201200671

20. Rudroff, F.; Mihovilovic, M. D.; Gröger, H.; Snajdrova, R.; Iding, H.; Bornscheuer, U. T. Opportunities and challenges for combining chemo- and biocatalysis, *Nat. Catal.* **2018**, *1*, 12-22. 10.1038/s41929-017-0010-4
21. Wijma, H. J.; Floor, R. J.; Janssen. D. B. Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability, *Curr. Opin. Struct. Biol*. **2013**, *23*, 588-594. doi: 10.1016/j.sbi.2013.04.008.
22. Arabnejad, H.; Dal Lago, M.; Jekel, P. A.; Floor, R. J.; Thunnissen, A. W. H.; Terwisscha van Scheltinga, A. C.; Wijma, H. J.; Janssen, D. B. A robust cosolvent-compatible halohydrin dehalogenase by computational library design, *Protein Eng. Des. Sel*. **2017**, *30*, 173-187. doi: 10.1093/protein/gzw068
23. Cerdobbel, A.; De Winter, K.; Aerts, D.; Kuipers, R.; Joosten, H.-J.; Soetaert, W.; Desmet, T. Increasing the thermostability of sucrose phosphorylase by a combination of sequence- and structure-based mutagenesis, *Protein Eng. Des. Sel.*, **2011**, *24*, 829-834. 10.1093/protein/gzr042
24. Wu, B.; Wijma, H. J.; Song, L.; Rozeboom, H. J.; Poloni, C.; Tian, Y.; Arif, M. I.; Nuijens, T.; Quaedflieg, P. J. L. M.; Szymanski, W.; Feringa, B. L.; Janssen‘ D. B. Versatile Peptide C-Terminal Functionalization via a Computationally Engineered Peptide Amidase, *ACS Catal*. **2016**, *6*, 5405-5414. 10.1021/acscatal.6b01062
25. Arnold, F. H. Engineering enzymes for non-aqueous solvents, *Trends in Biotech*. **1990**, *8*, 244-249. 10.1016/0167-7799(90)90186-2
26. Yu, H.; Huang, H. Engineering proteins for thermostability through rigidifying flexible sites, *Biotechnology Advances*, **2014**, *32*, 308-315. 10.1016/j.biotechadv.2013.10.012
27. Eijsink, V. G.; Bjørk, A.; Gåseidnes, S.; Sirevåg, R.; Synstad, B.; van den Burg, B.; Vriend, G. Rational engineering of enzyme stability, *J. Biotechnol*, **2004**, *113*, 105-120. DOI: 10.1016/j.jbiotec.2004.03.026
28. Perutz, M. F. Electrostatic effects in proteins, *Science* **1978**, *201*, 1187-1191. 10.1126/science.694508
29. Chen, J.; Yu, H.; Liu, C.; Liu, J.; Shen, Z. Improving stability of nitrile hydratase by bridging the salt-bridges in specific thermal-sensitive regions, *J. Biotechnol*. **2012**, *164*, 354-362. 10.1016/j.jbiotec.2013.01.021
30. Zarrine-Afsar, A.; Zhang, Z.; Schweiker, K. L.; Makhatadze, G. I.; Davidson, A. R.; Chan, H. S. Kinetic consequences of native stateoptimization of surface-exposed electrostaticinteractions in the Fyn SH3 domain, *Proteins*, **2012**, *80*, 858
31. Lee, C. F.; Makhatadze, G. I.; Wong, K. B. Effects of charge-to-alanine substitutions on the stability of ribosomal protein L30e from Thermococcus celer, *Biochemistry*. **2005**, *44*, 16817. 10.1021/bi0519654
32. Steiner, K.; Schwab, H. Recent advances in rational approaches for enzyme engineerings in rational approaches for enzyme engineering, *Comput. Struct. Biotechnol. J*. **2012**, 2, 1. 10.5936/csbj.201209010
33. Deuss, P. J.; Popa, G.; Botting, C. H.; Laan, W.; Kamer, P. C. J. Highly efficient and site-selective phosphane modification of proteins through hydrazone linkage: development of artificial metalloenzymes, *Angew. Chem. Int. Ed*. **2010**, *49*, 5315-5317. 10.1002/anie.201002174
34. Haapalinen, A. M.; van Aalten, D. M.; Meriläinen, G.; Jalonen, J. E.; Pirilä, P.; Wierenga, R. K.; Hiltunen, J. K.; Glumoff, T. Crystal structure of the liganded SCP-2-like domain of human peroxisomal multifunctional enzyme type 2 at 1.75 Å resolution, *J. Mol. Biol*, **2001**, *313*, 1127-1138. 10.1006/jmbi.2001.5084

35. Deuss, P. J.; Popa, G.; Slawin, A. M. Z.; Laan, W.; Kamer, P.C. J. Artificial Copper Enzymes for Asymmetric Diels-Alder Reactions, *ChemCatChem*, **2013**, 5, 1184-1191. 10.1002/cctc.201200671
36. van Leeuwen, P. W. N. M.; Casey, C. P.; Whiteker, G. T. Phosphines as ligands in *Rhodium Catalysed Hydroformylation* Eds.: P. W. N. M. Van Leeuwen, C. Claver, **2002**, Kluwer Academic Publishers, New York, p. 63-105.
37. van Leeuwen, P. W. N. M. Rhodium Catalysed Hydroformylation in *Homogeneous catalysis, understanding the art*, **2004**, Kluwer Academic Publishers, Dordrecht, p. 139-174.
38. Bahrmann, H.; Bogdanovic, S.; van Leeuwen, P. W. N. M. Higher Alkenes in *Aqueous-Phase Organometallic Catalysis* Eds. Cornils, B.; Herrmann. W. A. **2004**. Wiley. Weinheim, p. 391-409.
39. Joosten, H. J. **2008**, Method of Generating a Protein database. WO/2008/035970
40. Kourist, R.; Jochens, H.; Bartsch, S.; Kuipers, R.; Kumar Padhi, S.; Gall, M.; Böttcher, D.; Joosten, H.-J.; Bornscheuer, U. T. The α/β-Hydrolase Fold 3DM Database (ABHDB) as a Tool for Protein Engineering, *ChemBioChem* **2010**, *11*¸1635-1643. 10.1002/cbic.201000213
41. Lensink, M. F.; Haapalinen, A. M.; Hiltunen, J. K.; Glumoff, T.; Juffer, A. H. Response of SCP-2L Domain of Human MFE-2 to Ligand Removal: Binding Site Closure and Burial of Peroxisomal Targeting Signal, *J. Mol. Biol*, **2002**, *323*, 99-113. 10.1016/S0022-2836(02)00939-7
42. Filipp, F. V.; Sattler, M. Conformational Plasticity of the Lipid Transfer Protein SCP2, *Biochemistry*, **2007**, *46*, 7980-7991. 10.1021/bi6025616
43. Berendsen, H. J. C. The α-helix dipole and the properties of proteins, *Nature*, **1978**, 273, 443-446. 10.1038/273443a0
44. Armstrong, K. M.; Baldwin, R. L. Charged histidine affects alpha-helix stability at all positions in the helix by interacting with the backbone charges, *Proc Natl Acad Sci Usa*, **1993**, 90, 11337. 10.1073/pnas.90.23.11337
45. Carter, P. Site-directed mutagenesis, *Biochem J*, **1986**, 237, 1-7. 10.1042/bj2370001
46. Greenfield, N. J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions, *Nat Protoc*, **2006**, *1*, 2527-2535. 10.1038/nprot.2006.204
47. Wijma, H. J.; Floor, R. J.; Jekel, P. A.; Baker, D.; Marrink, S. J.; Janssen, D. B. Computationally designed libraries for rapid enzyme stabilization, *Protein Eng. Des. Sel*. **2013**, *27*, 49-58. 10.1093/protein/gzt061