Electronic Supplementary Material (ESI) for Catalysis Science & Technology.

Supporting Information

Catalytic and biophysical investigation of rhodium hydroformylase

Hasan T. Imam, Amanda G. Jarvis* Veronica Celorrio, Irshad Baig, Christopher C. R. Allen, Andrew C. Marr* and Paul C. J. Kamer*

S1 Experimental Procedures

S1.1 Protein expression and purification procedures

S1.1.1 SCP-2L V83C, SCP-2L A100C and SCP-2L A100C MXA

The mutagenesis, expression and purification of the proteins was carried out as previously described.[1]

S1.1.2 SCP-2L Q111C and W36C

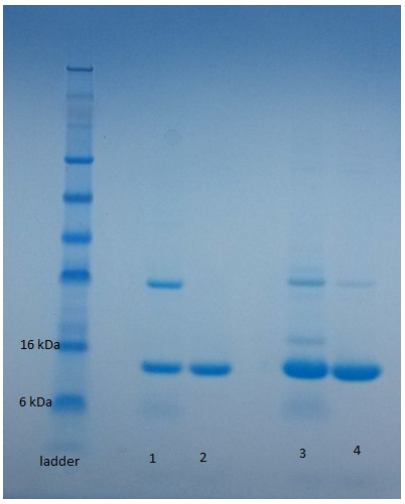
Primer	Mutation acquired	Sequence
Primer 1 fw	W36C TGA/TGC	TATTTGAG**TGC**CATATAACCAAAGGCGGAAATATTGGGGCTAA
Primer 1 rv		GGTTATATG**GCA**CTCAAATACAGCATTTACTTTCTTCACCACC
Primer 2 fw	Q111C CAG/TGC	AGAAACTT**TGC**ATGATTCTTAAA GAC TAT GCC AAG CTC TGA
Primer 2 rv		AAGAATCAT**GCA**AAGTTTCTGGCTCAGCATGATGTTCCCT

The reaction was set up as follows: Primer fw: 100 nM (10 μM, 1μL), primer rv: 100 nM (10 μM, 1μL), SCP-2L template DNA (2μL), dNTPs: 200 μM (2 mM, 2.5 μL), $MgSO_4$ (25 mM, 1 μL), KOD buffer 10X (5 μL) KOD Hot Start DNA polymerase (1 μL) and made up to a total of 25 μL using Millipore H_2O. PCR reaction times:

1. 94 °C, 5 mins
2. 94 °C, 1 min
3. 56 °C, 1 min
4. 70 °C, 3 min
5. Repeat steps 2-4, 11 times.
6. 95 °C, 1 min
7. 49 °C, 1 min
8. 70 °C, 3 min
9. Repeat steps 6-8, 2 times.
10. 70 °C, 7 min

Digestion of the wild type template strand was then performed as follows: PCR product (8.5 μL), 10x Fast digest buffer (1 μL) and Dpn1 (0.5 μL) were incubated at 37 °C for 10 min, followed by inactivation at 80 °C for 5 min. The digested products were transformed into commercial DH5α cells (Selectline silver) and grown on kanamycin (50 μg/mL) agar plates at 37 °C overnight. One to two colonies were observed for each mutant. The DNA was obtained via a miniprep kit and sent for sequencing at GATC Biotech Co., UK.

The expression and purification of Q111C and W36C was carried out following the same procedures as those described for A100C and V83C.



SDS page gel of purified SCP_2L-Q111C and SCP_2L-W36C

Ladder: SeeBlue prestained ladder

1: Q111C after TEV, before purification

2: Pure Q111C

3: W36C after TEV, before purification

4: Pure W36C

Yields: SCP-2L W36C = 37.5 mgL^{-1} media, SCP-2L Q111C = 33 mgL^{-1} media.

S1.1.3 SCP-2L V83C MXA

The alanine mutants of SCP-2L V83C MXA were introduced using the Haunting's method of site directed mutagenesis.

Primer	Mutation acquired	Sequence
Primer 1 fw	M1A ATG/GCA	*AGGGCGCc**GCA**gaGG*GAGGGAAGCTTCAGAGtACC
Primer 1 rv		*CCtc**TGC**gGCGCCCT*GAAAAtACAGGTTTTCGGTcgt
Primer 2 fw	M80A ATG/GCC	*TCAGATGAAG*ATTTC**GCC**GAGGTGTGCCTG
Primer 2 rv		*CTTCATCTGA*AAGTATGATTGTTGTATCAGCAGCACC
Primer 3 fw	M105A ATG/GCA	*GAACATCGCACTGAGCC*AGAAACTTCAGATGATTCTTAAAG
Primer 3 rv		*GGCTCAGTGCGATGTTC*CCTCTGGCCTTCAgCCTGC
Primer 4 fw	M112A ATG/GCA	*TCAGGCAATTCTTAAAGAC*TATGCCAAGCTCTGAGGATCC
Primer 4 rv		*GTCTTTAAGAATTGCCTGA*AGTTTCTGGCTCAGCATGATG

The reaction was set up as follows: Primer fw (10 µM, 1µL), primer rv (10 µM, 1µL), SCP-2L template DNA (1µL), dNTPs (2 mM, 2.5 µL), $MgSO_4$ (25 mM, 1.75 µL), KOD buffer 10X (2.5 µL) KOD Hot Start DNA polymerase (1 µL) and made up to a total of 25 µL using Millipore H_2O. PCR reaction times:

1. 95 °C, 5 mins
2. 95 °C, 1 min
3. 63 °C (M80A), 58 °C (M1A, M105A, M112A), 1 min
4. 70 °C, 3 min
5. Repeat steps 2-4, 11 times.
6. 95 °C, 1 min
7. 49 °C, 1 min
8. 70 °C, 3 min
9. Repeat steps 6-8, 2 times.

10. 70 °C, 7 min

Digestion was then performed as follows: PCR product (8.5 µL), 10x Fast digest buffer (1 µL) and Dpn1 (0.5 µL) were incubated at 37 °C for 10 min, followed by inactivation at 80 °C for 5 min. The digested products were transformed into commercial DH5α cells (Selectline silver) and grown on kanamycin (50 µg/mL) agar plates at 37 °C overnight. One to two colonies were observed for each mutant. The DNA was obtained via a miniprep kit and sent for sequencing at GATC Biotech.

Expression of unlabeled protein was performed as previously described.[1] Plasmids contain the genes of pEHISTEV::dΔhΔSCP-2L V83C MXA, were transformed into Rosetta (DE3) competent cells. The cells were grown (2-3 h, 37°C, 200 rpm) until an optical density (OD600) of 0.6-0.7 was reached. Protein expression was than initiated by addition of isopropyl-1-thio-β-D-galactopyranoside (IPTG, 0.2 mM from a 0.5 M stock solution). The culture was grown overnight (16-18h, 16°C, 200 rpm).

S1.1.4 Expression and purification of labeled proteins: ^{13}C-methyl-L-methionine protein and SeMet SCP-2L A100C and SeMet SCP-2L A100C_M105A

A single transformed colony was transferred into 100-250 mL LB medium (Lysogeny broth medium contains: 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, pH 7.0) and allowed to grow overnight (37 °C, 200 rpm). The resulting culture was pelleted and washed with base media (NH_4Cl 1 g/L, KH_2PO_4 3 g/L and $Na_2HPO_4.7H_2O$ 6 g/L). Bacterial cells were resuspended in base media complimented with nutrient mix (Molecular Dimensions, 50 ml/L, dissolving 3.5-5 mg of nutrient mix in 50 ml of ddH_2O, filter sterilized). The cells were incubated at 37 °C, 200 rpm for 20 min. ^{13}C-methyl-L-methionine was added to the cell culture to a final concentration of 60 µg/mL. Incubation at 37 °C continued until the OD600 reached 0.6-0.7 (about 2-3 h), and the protein expression was induced by addition of 0.2 mM IPTG (final concentration). The cells continued to grow overnight (16-20 h, 16 °C, 200 rpm). Selenomethionine proteins were expressed following the similar procedures, however, instead ^{13}C-Methyl-methionine, Selenomethionine (60 µg/mL) was added. Proteins were harvested and stored as previously described.[1]

The protein was purified using a Ni column (5 ml, HisTrap FF, GE Healthcare) connected to a peristaltic pump (P-1, GE Healthcare). The syringe filtered (0.45µm) cell supernatant was passed through pre-equilibrated washed with buffer, (5 CV, 30 mM Tris-HCl, 20 mM Imidazole, 150 mM NaCl, pH 8) Ni column (5 ml, HisTrap FF, GE Healthcare). The protein elution was achieved with elution buffer (30 mM Tris-HCl, 330 mM Imidazole, 150 mM NaCl, pH 8). The protein solution was dialyzed (Snake skin dialysis tubing, 3.5K MWCO, Thermo Scientific) against dialysis buffer (30 mM Tris-HCl, 10 mM Imidazole, 150 mM NaCl, pH 8) overnight with constant slow stirring at 4°C. The dialyzed protein solution was treated with DTT (1 mM, 100 mM stock)), EDTA (0.5 mM, 50mM stock)) and TEV protease (0.014 equivalents to protein, 1.47 mg/ml stock) overnight with slow rotation at room temperature to cleave the histag. The tag free protein was obtained by passing the TEV protease treated protein solution through a pre- equilibrated column (30 mM Tris-HCl, 20 mM Imidazole, 150 mM NaCl, pH 8). Unlabeled

SCP-2LV83C and methionine mutants' SCP-2L-V83C-M1A, SCP-2L-M105A and SCP-2L-M112A proteins yield 30-45mg/L culture, SCP-2L-V83C-M80A for which the yield was 5-7 mg/L. For selenomethionine & ^{13}C-Methyl methionine labelled protein yield was 8-12 mg/L.

S1.2 Generation of Artificial metalloenzymes: Protein modification

Chemicals and reagents
The chemicals, reagents and solvents were purchased commercially high quality and pure, and used without further purification, unless otherwise mentioned in the text. 3-Maleimidopropionic acid hydrazide hydrochloride (1), purchased from Speed chemical. 4-(Diphenylphosphino)benzaldehyde (P), was synthesized following literature procedures[2] Dicarbonyl-acetylacetonato-rhodium(I) (Rh) was purchased from Sigma Aldrich.

All reactions involving phosphines were performed under an argon atmosphere using degassed solvents and standard Schlenk techniques. Buffer solutions (20mM MES, 50 mM NaCl, pH 6) were degassed by bubbling argon through the solution. Protein solutions were degassed by washing and concentrating three times with degassed buffer).

The protein (various concentration) was treated with 3-maleimidopropionic acid hydrazide hydrochloride (**1,** 10-20 moles equivalents) and kept in a water bath overnight (20°C, 60 rpm, 16-18h). After centrifugral buffer exchange to remove the excess maleimide, the maleimide-modified proteins were allowed to react with 4-(diphenylphosphino)benzaldehyde (10 equivalents) under an argon atmosphere in a water bath (20°C, 60 rpm, 16-18h). Excess phosphine was removed by centrifugal buffer exchange and [Rh(acac)$(CO)_2$] (0.5 equivalents with respect to the protein concentration , 25mM stock in DMF) was added to the phosphine modified protein and left overnight. The Rh added protein was buffer wash and concentrated to obtain the artificial metalloenzyme.

S1.3 Characterisation

S1.3.1 MS of proteins

LC-MS(ES+) was used for the analysis of protein and performed on a Waters Alliance HT 2795 equipped with a Micromass LCT-TOF mass spectrometer, using positive electrospray ionisation and applying a Waters MASSPREP® On-line Desalting 2.1x10 mm cartridge using a gradient of 1% formic acid in H2O to 1% formic acid in acetonitrile. ESI-MS results were analysed by MassLynx V. 4.0 and its MaxEnt algorithm.

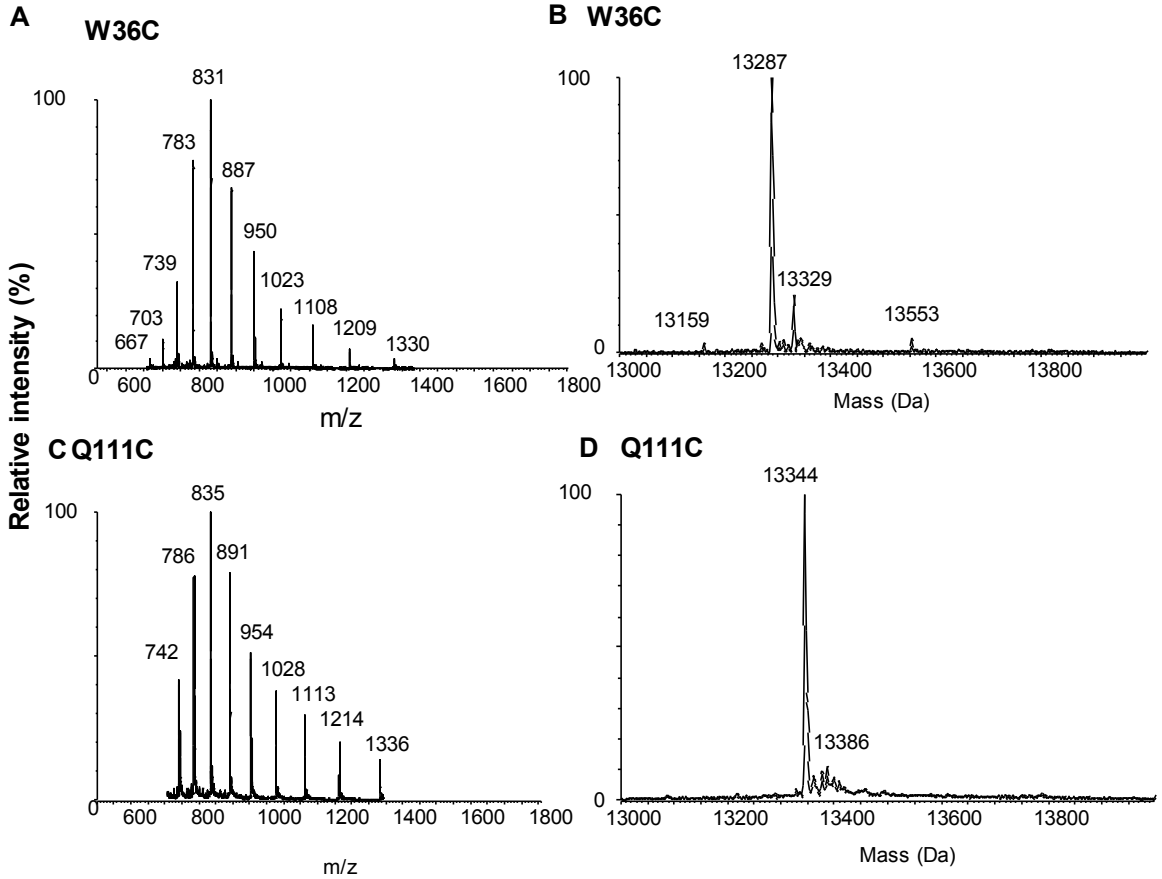


Fig S1 Mass spectra of proteins. SCP-2L-W36C (A) Raw (B) Deconvoluted, SCP-2L-Q111C (C) Raw (D) Deconvoluted. The data were recorded over 20-30 min and the results were analysed by MassLynx V. 4.0 and its MaxEnt algorithm.

S1.3.2 MS of modified proteins

The samples for phosphine-modified proteins were prepared immediately before injection to reduce the oxidation of the sample.

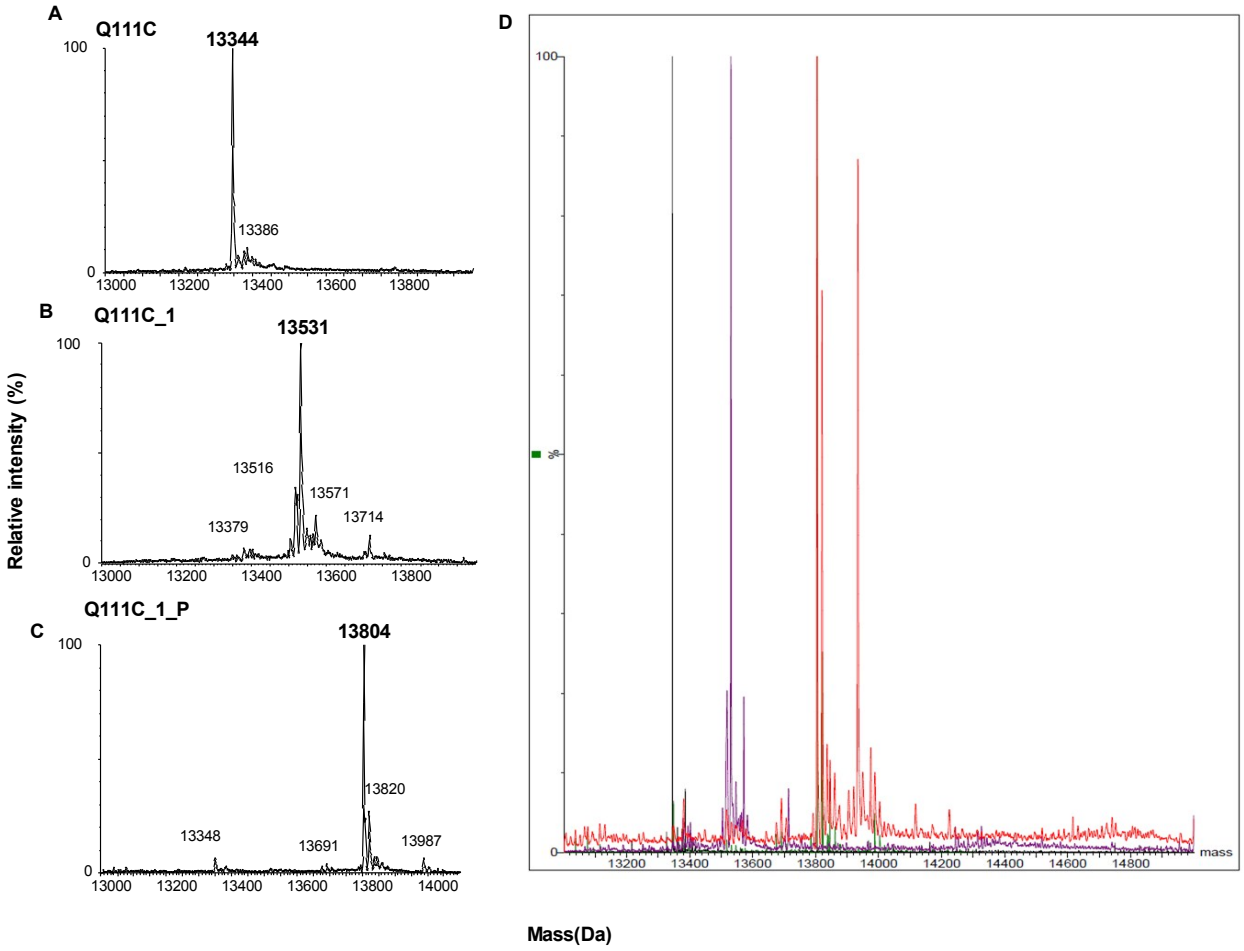


Fig S2 Deconvoluted mass spectra of SCP-2L-Q111C modification. (A) SCP-2L-Q111C (B) Maleimide modified (1) SCP-2L-Q111C (C) Phosphine (P) modified SCP-2L-Q111C-1. (D) Overlay of the Deconvoluted mass spectrum of Q111C (black), Q111C-1 (purple), Q111C-1-P (green) and Q111C-1-P-RhCO (Red). The data were recorded over 20-30 min and the results were analysed by MassLynx V. 4.0 and its MaxEnt algorithm.

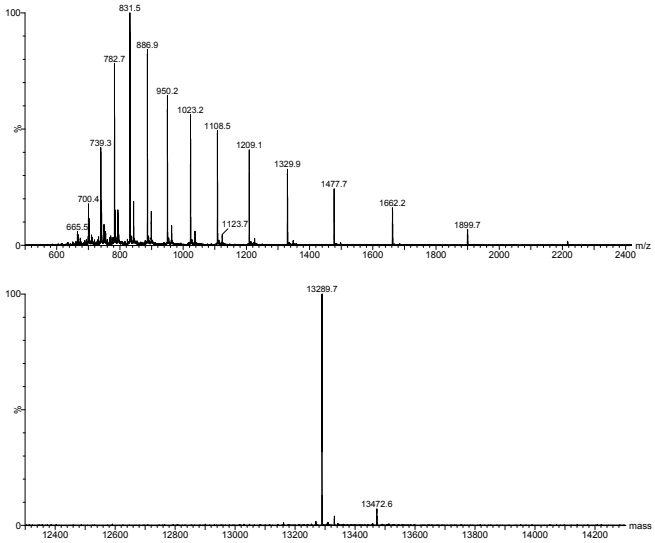


Fig S3 SCP_2L-W36C-1, partial modification only. Raw mass spectrum (top) and deconvoluted mass spectrum (bottom, MaxEnt)

S1.3.3 Circular Dichroism (CD) spectroscopy

The protein folding of the mutants and the modified mutants were determined by using Circular Dichroism (Biologic MOS-500 spectrometer) mounted with a Xe lamp. The near UV spectra were recorded using the following parameters- wavelength 190 – 260 nm, Quartz cuvette, path length 10 mm, step 0.25 nm, acquisition period 0.5 s, repeats 3, ±MD 30 and slit size 2 nm.

The CD spectra were recorded on samples at concentrations of between 0.5-1.5 mg/mL and were corrected for concentration according to the following equation to give the molar ellipticity [θ] values:

$[\theta]\ (degcm^2dmol^{-1}) = (mD\ sample - mD\ blank)/\ 10\ *conc.\ (M)*pathlength\ (cm))$

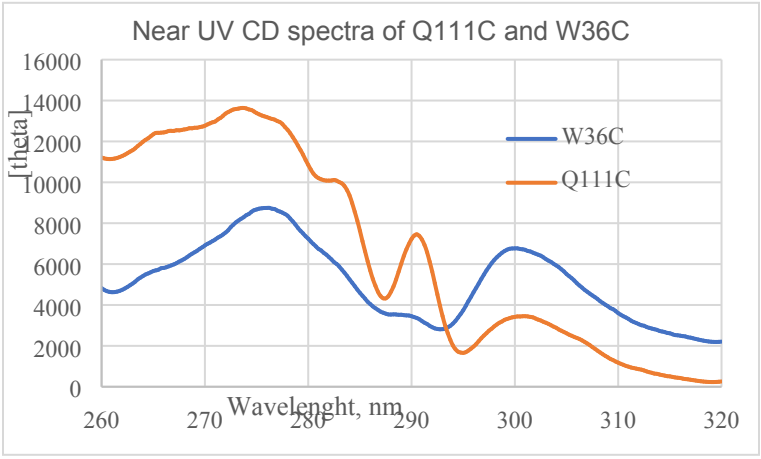


Fig S4 Near UV CD spectra of SCP_2L-Q111C and SCP_2L-W36C

S1.3.4 Nuclear magnetic resonance (NMR) spectroscopy

Protein/catalyst samples (various concentrations) were prepared in standard buffer (30 mM Tris-D11, 10% D_2O, 50 mM NaCl, pH 7.43). The Protein/catalyst NMR spectra were recorded in a Bruker 700 MHz spectrometer mounted with TCI Prodigy inverse triple resonance cryoprobe at temperature 303K. 1D ^{1}H NMR spectra were recorded at an operation frequency of 700.13 MHz and ^{13}C Spectra were recorded at 176.05 MHz. 1D 1H-NMR spectra were recorded with 32k complex data points, 32 scans and a spectral width of 16 ppm. 2D [^{1}H ^{13}C] heteronuclear single-quantum coherence (HSQC) spectra were recorded with 2k data points in F2 (^{1}H) and 512 increments in F1 (^{13}C). Spectra were acquired with a spectral width of 16 ppm in F2 and 70 ppm in F1 over 4 scans, with a $^{1}J(^{1}H\text{-}^{13}C)$ coupling constant of 145 Hz. The raw data were apodized and Fourier transformed into 4k × 1k data points in the F2 × F1 dimensions, respectively. ^{1}H chemical shifts were referenced to the residual water and ^{13}C is referenced externally to DSS (2,2′-dimethyl-2-silapentane-5-sulfonate. Recorded spectra were processed using Bruker Topspin v. 2.1 software. For 2D, phase- and baseline-corrected Topspin files were transferred to Sparky v. 3.114 [75] for analysis. [^{1}H ^{13}C] HSQC ^{13}C-methyl-Met-V83C and its methionine mutant's NMR spectra were recorded in a Bruker 800 MHz spectrometer ^{1}H at operation frequency 799.57 MHz and ^{13}C Spectra were recorded at 201.05 MHz with a spectral width of 16 ppm in F2 and 35 ppm in F1 in the same experimental conditions as mentioned above.

S1.3.5 EXAFS and XANES

XAFS spectra were collected on beamline B18 at Diamond Light Source.[3] The monochromator comprises Si(311) crystals operating in Quick EXAFS mode. Calibration of the monochromator was carried using a Rh foil prior to the measurements. The protein samples were measured in fluorescence mode at the Rh K-edge (23219 eV) using a 36 element solid state Ge detector.

Sample preparation

The Selenomethionine labelled proteins SCP-2L_A100C and SCP-2L_A100C_M105A were expressed and purified as described in section S1.1.4 The catalysts SeMet_A100C-1-P-Rh and SeMet_A100C_M105A-1-P-Rh were prepared following the protocol as mentioned in section 1.2. Catalysts SeMet_A100C-1-P-Rh and SeMet_A100C_M105A-1-P-Rh (0.12 ml, 3.5-5mM, 20mM MES, 50 mM NaCl, pH 6) were transferred to gelatine capsules (size 4, 0.21 ml, 5.0 mm, Agar Scientific Limited) and flash frozen immediately in liquid nitrogen. The frozen samples were kept and transported within a dry shipper (CX100, Taylor-Wharton) until the experiment, where they were transferred to a cryostat (78K) mounted on the beamline.

In order to avoid beam damaged, samples were placed inside a liquid nitrogen cooled cryostat. A total of 25 spectra were averaged to improve the signal to noise. The data were analysed using the Athena and Arthemins programs,[4] which implement the FEFF6 and IFEFFIT codes.[5]

The XANES for SCP_2L-A100C-M105A-1-P-Rh plus the Rh foil is shown in Fig S5. A slight shift of the edge to higher energy is observed for the Rh hydroformylase compared to the foil, which is normally indicative of a higher oxidation state, which would be consistent with Rh^{+}.

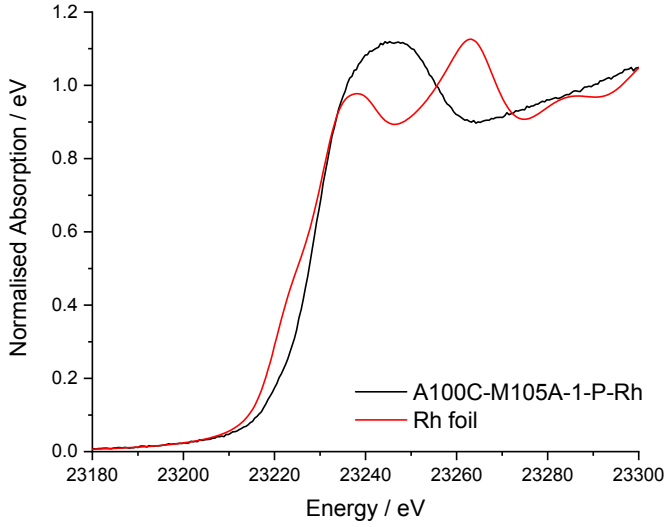


Fig S5 Normalized Rh K-edge XANES spectra of Rh hydroformylase and Rh foil

The features that occur around the main edge are representative of electronic transitions, and by inference the available orbitals the different bonding types allow. It is therefore a suitable means to understand ligand type. Both reference structures containing CO ligands, show a double maximum pattern between 23234-23260 eV. This suggests that features in these positions are indicative of bound carbonyls to the Rh centre. It can be seen that this pattern is absent from the XANES spectrum of the SCP_2L-A100C-M105A-1-P-Rh, confirming a significant change in ligand environment, and suggesting the loss of carbonyl (Fig S6).

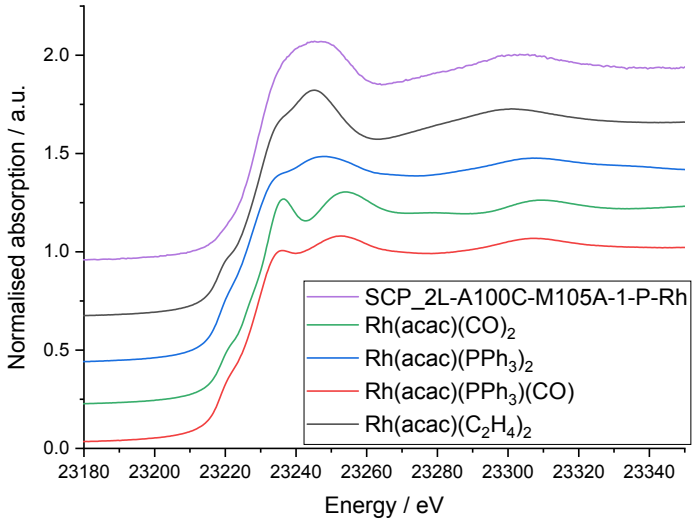


Fig S6 Normalized Rh K-edge XANES spectra of Rh hydroformylase and Rh foil

Even compared with previous published samples, it is clear that the coordination is different:

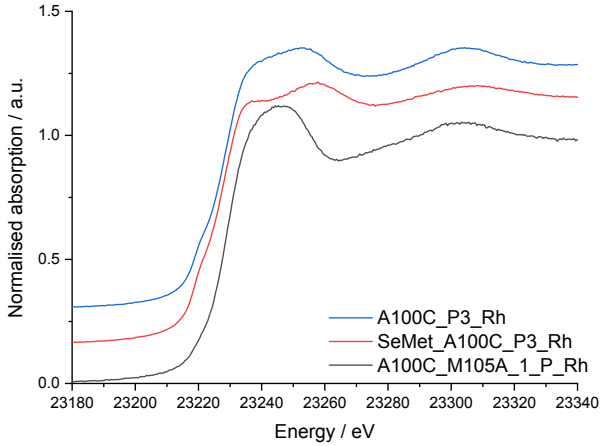


Fig S7 XANES of the previously published samples versus SCP_2L-A100C-M105A-1-P-Rh (labelled Rh hydroformylase)

The fit for A100C:

In order to get an insight into the molecular structure, three models of the Extended X-ray Absorption Fine Structure were studied. Fig S8 shows the comparison between the Fourier transform (FT) of the experimental spectra (black line) and the best-fit simulations (red line), as well as the signals in *k*-space. The bond lengths, Debye-Waller factors (σ^2) and the energy shift parameter (ΔE_0) were refined. The best-fit parameters are summarized in Table S1.

The first model (model A) assumes that no Se atom is in the near vicinity of Rh. One hypothesis was that M105 could be close to the active site and may be the residue responsible for the sulfur/selenium coordination shown in previous EXAFS experiments.[1] If this assumption is correct, mutating the methionine to an alanine should result in the removal of the sulfur coordination. We prepared the SeMet variant of A100C-M105A-1-P-Rh to test this hypothesis and see if we did indeed lose the Se coordination we had previously observed. When model A was compared with both model B (same model from previous EXAFS) and C, it was observed that it had a much higher R_f value and thus model A was disregarded.

Model B and C are quite similar, and both have reasonable fits, importantly both suggest a Se atom still in close proximity to the Rh center. It is therefore unlikely that M105 was the methionine responsible for the observed Rh binding in A100C-1-P-Rh. Model C achieved the most satisfactory degree of fit and uses four scattering paths: Rh-O, Rh-P, Rh-Se and Rh-C. The shortest path is that of the Rh-O, where a coordination number of 3 was required to achieve a good level of fit. From current data it is unclear what the extra oxygen atom corresponds to.

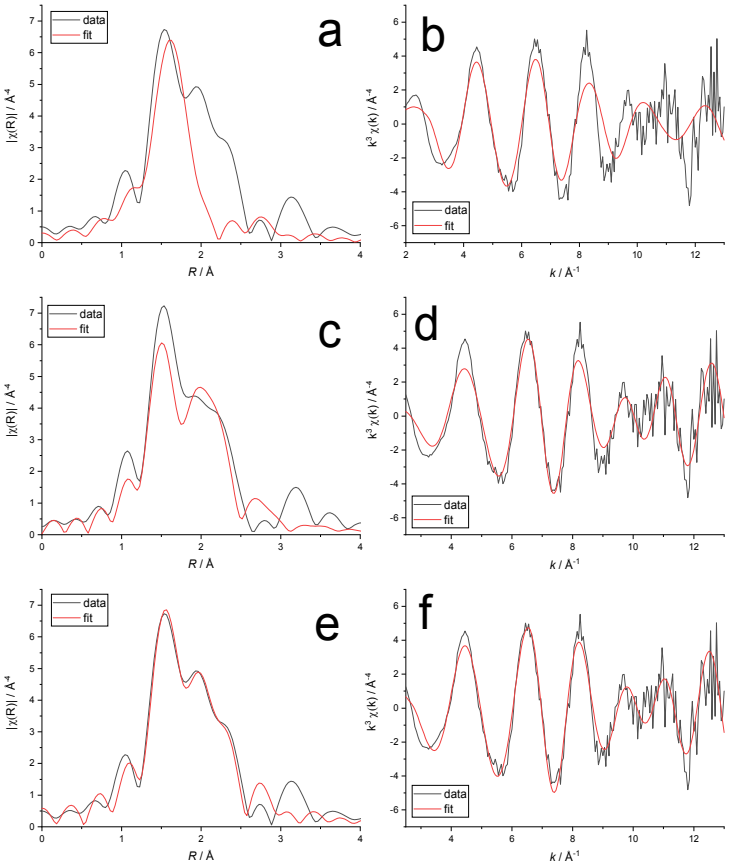


Fig S8: *k^3* weighted χ data (right) and *k^3* weighted Fourier transform EXAFS data (left) for the Rh hydroformylase catalysts recorded at Rh K edge using model A (a, b), model B (c, d) and model C (e, f). For each sample the red line represents the simulated fit.

Table S1 Energy shift and the best fit results from the structural analysis of the different models used at the Rh K-edge. N is the coordination number, R is the interatomic distance and σ^2 is the Debye-Waller factor. R_f is the R-factor, which represents the relative error of the fit and data.: S_0^2 = 0.81; Fit range: 3.5 < k < 13, 1 < R <3.

		N (coordination number)	R / Å	10^3 σ^2 / Å2	ΔE_0 / eV	R_{factor}
Model A	Rh-O	3 (fixed)	2.05 ± 0.05	2.3 ± 2.1	1.4 ± 4.9	0.13
	Rh-P	1 (fixed)	2.34 ± 0.08	3.0 (fixed)		
	Rh-C	4 (fixed)	3.44 ± 0.36	34.3 ± 11.9		
Model B	Rh-O	2 (fixed)	2.01 ± 0.04	1.0 (fixed)	-4.7 ± 5.9	0.04
	Rh-P	1 (fixed)	2.29 ± 0.11	3.0 (fixed)		
	Rh-Se	1 (fixed)	2.44 ± 0.03	3.1 ± 2.4		
	Rh-C	4 (fixed)	3.31 ± 0.13	9.9 ± 4.5		
Model C	Rh-O	3 (fixed)	2.03 ± 0.02	1.7 ± 0.7	-0.5 ± 2.4	0.01
	Rh-P	1 (fixed)	2.35 ± 0.06	3.0 (fixed)		
	Rh-Se	1 (fixed)	2.44 ± 0.02	3.5 ± 1.1		
	Rh-C	4 (fixed)	3.37 ± 0.09	9.3 ± 9.1		

S1.4 Catalytic procedures

S1.4.1 General hydroformylation procedure

Hydroformylation was carried out using the procedure previously described.[1] Glass reaction vials (4 ml volume) with a septum caps containing a magnetic stirring bar were placed in a stainless steel autoclave (accommodating 7 vials). A needle was pierced through the septum cap of the vial. The autoclave containing the vials was flushed with 20 bar of Ar gas three times to create an inert atmosphere. Immediately, the catalyst (0.5 ml) and substrate (0.5ml 1-octene containing 9% (v/v) heptane and 1% (v/v) diphenyl ether as internal standards) were added to the vials (3-4 vials) under the Ar blanket. The autoclave was flushed with syngas (20 bar, CO:H_2, 1:1) three times and finally pressurized with syngas (80 bar) and placed in a preheated (35°C) oil bath with stirring (600 rpm). Prior to every reaction the 1-octene was passed over activated alumina (dried under vacuum for 4-5 hours at 150°C) to remove peroxides. The reaction was carried out for the desired reaction time and stopped by putting the autoclave in ice bath for 20-30 minutes. The pressure was released slowly and the organic phase analysed by GC. All experiments were run in triplicate, unless otherwise stated. A protein-P : Rh ratio of 2 : 1 was aimed for and an aliquot of this solution was saved for ICP-MS analysis to obtain the actual Rh concentration.

S1.4.2 Gas chromatography

Hydroformylation products were analyzed on an Agilent 7820A GC system installed with an Agilent HP-5 column 30m x 0.32 mm x 0.25 μm and an autosampler. Conditions: injector temperature 250 °C; FID detector temperature 300 °C; 6.5 mL/min constant flow; 75:1 split ratio; 1 μL injection; oven method: 25 °C isotherm for 6 min, 10 °C/min to 60 °C, 20 °C/min to 300 °C. Retention times: heptane 2.05 min, 1-octene 4.45 min, 2-methyloctanal 11.34 min, n-nonanal 11.83 min, diphenyl ether 14.19 min. Small amounts of isomerization products could not be detected due to the large excess of 1-octene resulting in a broad tailing alkene peak.

S1.4.3 Collection of Time Course data

The time course experiments were run as batch experiments (4 vials per time point) for each time point (1, 2, 3, 4, 5, 7 and 12 hours). All reactions were set up within 24 h of each other, using the same batch of SCP-2L A100C-**1-P3-Rh**. Two different autoclaves were used for the experiments.

S1.4.4 Catalyst testing at different catalyst concentrations

An aliquot of SCP-2L A100C-**1-P3-Rh** (528 μM Rh), was diluted to give 5 different overall concentrations. The reaction was run in duplicate and stopped after 5 h.

S1.4.5 ICP MS

Samples preparation: Trace metals grade concentrated nitric acid (0.5 mL) added slowly to the catalyst's solution (0.5ml containing 20mM MES, 50 mM NaCl, pH 6). The resulting solution was heated at 80°C for 4-5 h. The solution was diluted with water (11 ml, MilliQ) to make a total volume of 12ml. Control sample was made in the similar way, with buffer (0.5 ml, 20mM MES, 50 mM NaCl, pH 6) was added, instead of catalyst.ICP-MS was carried out using a Perkin Elmer Elan 6100 DRC Quad and a Thermo-Finnegan Element 2 at Edinburgh University ICP-MS facilities.

S2 Supplementary Figures & Tables:

Table S2 Hydroformylation of 1-Octene using SCP-2L-Q111C-1-P-Rh

Entry	Catalyst	TON	% linear aldehyde
1	SCP-2L Q111C[a]	252.02 (29.95)	71.7 (0.50)

Standard conditions: 80 bar syn gas (1:1), 35 °C, stirring 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. Rh concentration was obtained by ICP-MS and used to calculate TON. Conversions and linear selectivities were obtained by GC using a minimum of 3 runs. Standard deviation in brackets.a) 45 h, Rh 33 nmole.

Table S3 Time dependent hydroformylation of 1-Octene using SCP-2L-A100C-1-P-Rh

Entry	Time (mins)	TON	% linear aldehyde
1	60	4.47 (0.62)	81.47 (1.06)
2	120	7.50 (0.40)	82.91 (0.41)
3	180	12.10 (1.77)	83.11 (0.34)
4	240	20.32 (1.07)	82.79 (0.38)
5	300	24.75 (0.54)	83.16 (0.25)
6	420	38.96 (2.17)	81.71 (0.31)
7	720	66.88 (12.86)	79.58 (0.76)
8[a]	960	112 (4)	83 (1.4)
9[a]	2880	409 (58)	79 (5)

Standard conditions: 80 bar syn gas (1:1), 35 °C, stirring 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. Rh 42 nmole. Conversions and linear selectivities were obtained by GC using 4 runs. A) From a different batch of catalyst.

Table S4 Hydroformylation of 1-Octene at different [Rh] using SCP-2L-A100C-1-P-Rh

Entry	Rh concentratiuon (μM)	Moles of aldehyde (μmol)	TON	% linear aldehyde
1	528	3.54	16.35	84.7
2	264	2.11	21.77	83.8
3	106	0.95	24.51	81.2
4	53	0.60	31.07	79.5
5	26	0.31	32.55	74.2

Standard conditions: 80 bar syn gas (1:1), 35 °C, stirring 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. 5 h. Conversions and linear selectivities were obtained by GC using 2 runs. Average given.

Table S5 Hydroformylation of 1-Octene using SCP_2L_100C-1-P-Rh & SCP_2L_V83C-1-P-Rh, and their MXA mutants

Entry	Catalyst	[Rh], M	TON		% Liner Aldehyde	
1	SCP-2L-V83C-1-P-Rh	3.87×10^{-08}	6.2	(0.45)	72.3	(1.27)
2	SCP-2L-V83C-M1A-1-P-Rh	3.70×10^{-08}	57.7	(22.8)	75.6	(0.7)
3	SCP-2L-V83C-M105A-1-P-Rh	2.04×10^{-08}	40.2	(17.8)	69.2	(1.6)
4	SCP-2L-V83C-M112A-1-P-Rh	2.45×10^{-08}	41.6	(10.6)	75.1	(0.3)
5	SCP-2L-100C-1-P-Rh	2.00×10^{-08}	60.9	(3.2)	77.4	(3.2)
6	SCP-2L-100C-M1A-1-P-Rh	1.62×10^{-08}	86.0	(9.1)	80.8	(1.0)
7	SCP-2L-100C-M80A-1-P-Rh	1.85×10^{-08}	67.4	(6.7)	76.1	(0.4)
8	SCP-2L-100C-M105A-1-P-Rh	1.82×10^{-08}	112.1	(33.1)	74.4	(2.3)
9	SCP-2L-100C-M112A-1-P-Rh	1.40×10^{-08}	66.7	(5.0)	75.5	(0.5)

Reaction Conditions: 80 bar syngas (1:1), 35°C, 630 rpm, 5 h. Samples: Catalyst: 500 µl, substrate: 450 µl (1-Octene), Internal standard: 50 µl (9% v/v n-heptane, diphenyl ether, 1%, from 10% stock).

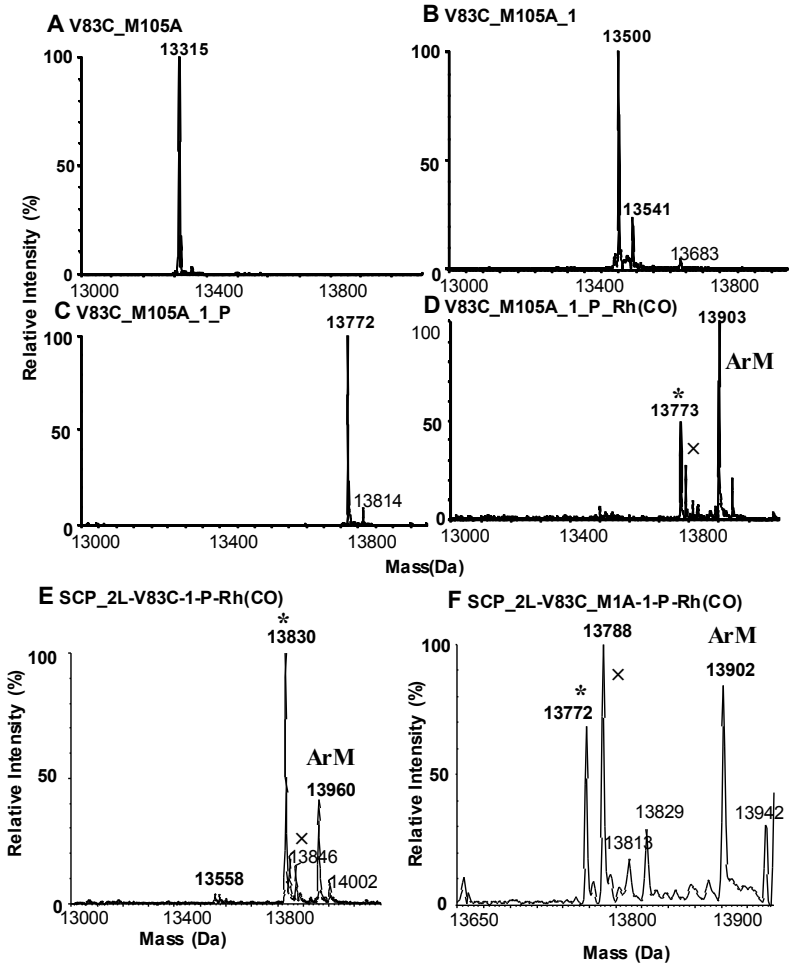


Fig S9 Deconvoluted mass spectra of hydroformylation catalyst of V83C and its methionine mutants. Demonstration of catalyst design on V83C_M105A (A) V83C_M105A (B) Maleimide (1) modification (C) Phosphine (P) modification and (D) Addition of [Rh(acac)$(CO)_2$] results in V83C_M105A-1-P-Rh(CO) catalyst. (E) V83C-1-P-Rh(CO) and (F) V83C-M1A-1-P-Rh(CO). *Free phosphine × Oxidized phosphine and ArM is Rh(CO) incorporated species. Protein samples for mass spectrometry were 10-25 µM, in 20 mM MES and 50 Mm NaCl buffer, Ph 6.0, acquiring using a gradient of 1% formic acid in H_2O to 1% formic acid in acetonitrile. The data were recorded over 20-30 min and the results were analysed by MassLynx V. 4.0 and its MaxEnt algorithm.

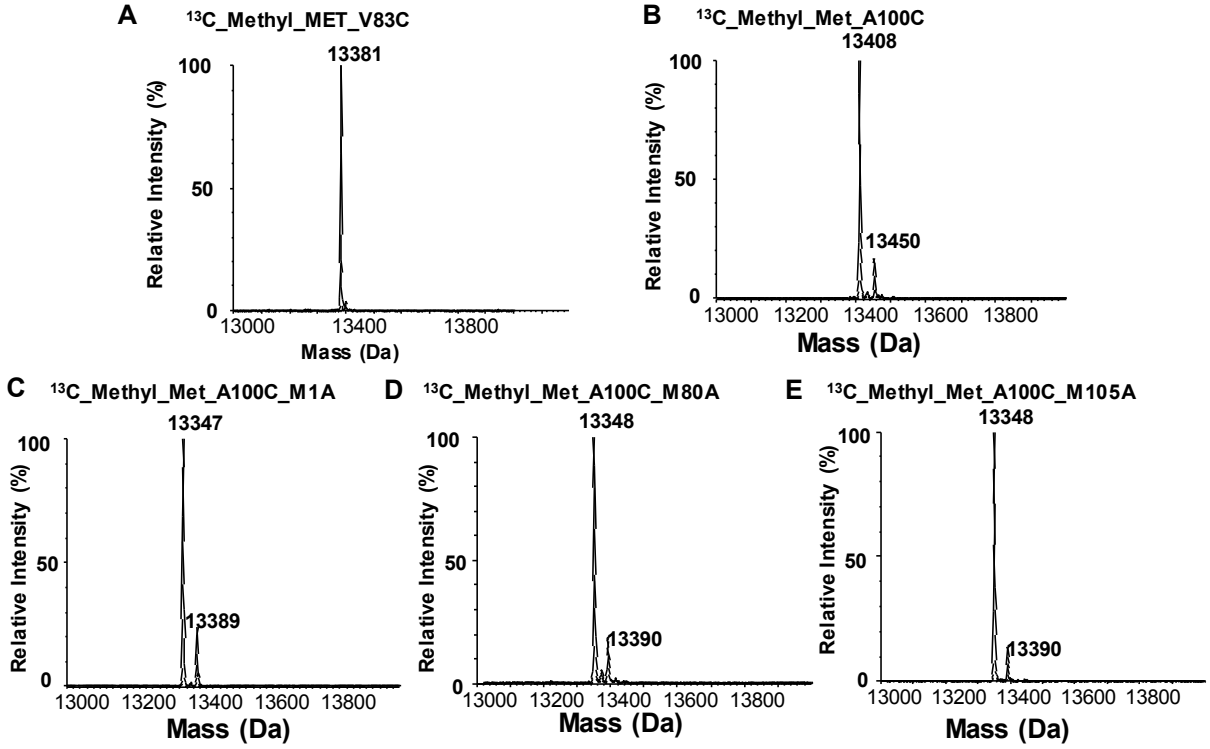


Fig S10 Deconvoluted mass spectra of homogenously ^{13}C-Methyl-methionine label protein of SCP-2L (A) V83C (B) A00C (C) A100C_M1A, (D)A100C_M80Aand (E) A100C_M105A. Protein samples for mass spectrometry were 10-25 μM, in 20 mM MES and 50 Mm NaCl buffer, Ph 6.0, acquiring using a gradient of 1% formic acid in H_2O to 1% formic acid in acetonitrile. The data were recorded over 20-30 min and the results were analysed by MassLynx V. 4.0 and its MaxEnt algorithm.

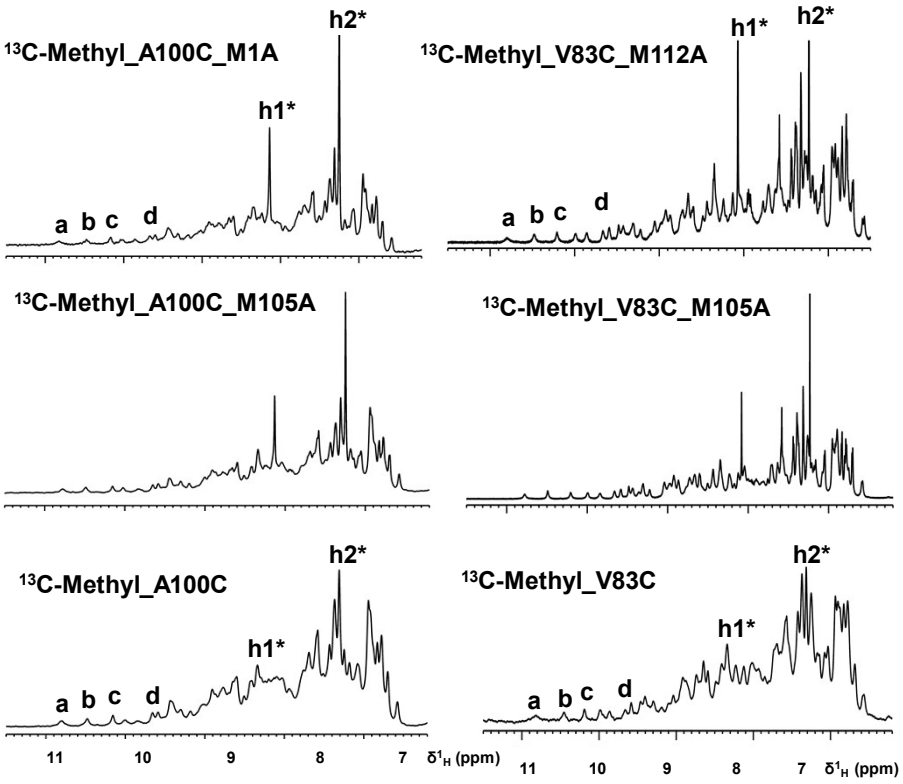


Fig S11 Characterization and monitoring the protein folding by 1D 1H NMR of SCP-2L_^{13}C-Methyl_Met_V83C (B) (A) SCP-2L_^{13}C-Methyl_Met_A00C and their methionine mutants. Well dispersioned protons peaks in the NH and aromatic (6-11ppm) region suggesting folded proteins. A homogenously labelled (^{13}C-Methyl-methionine label) protein sample (~65-100µM, 30 mM Tris-D_{11}, 10 % D_2O, 50 mM NaCl, pH 7.43) at an operating frequency of 700.13 MHz for ^{1}H using a cryoprobe at 303 K were used for these experiments.

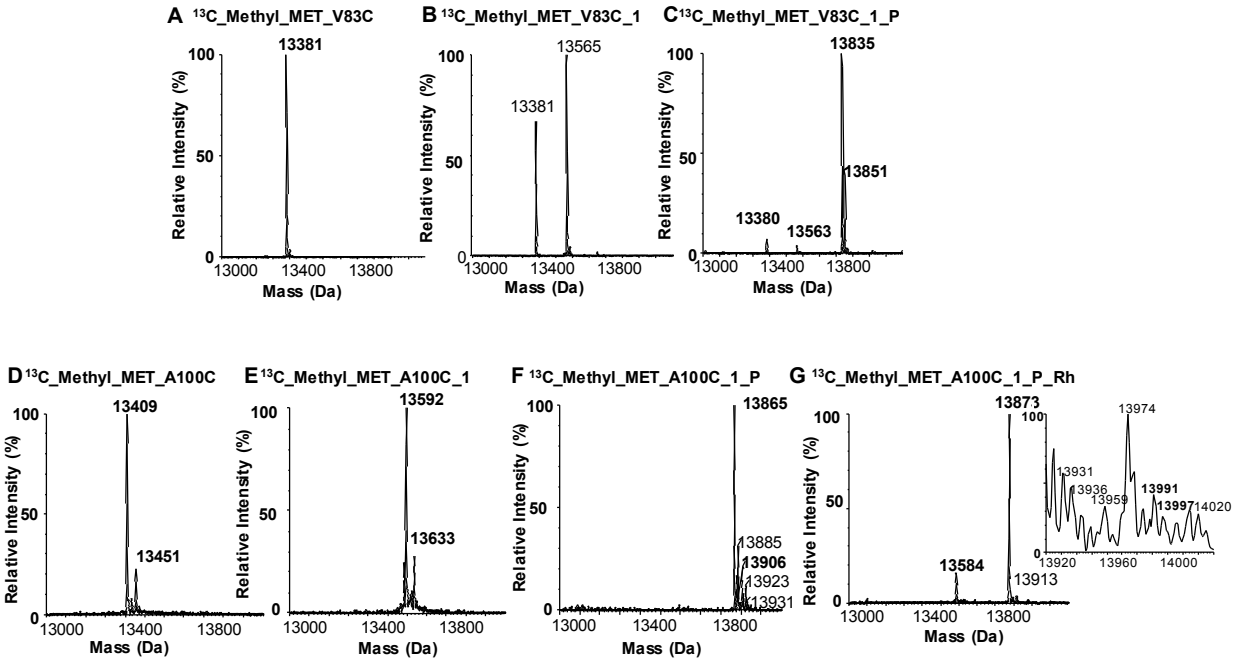


Fig S12 Deconvoluted mass spectra of ^{13}C-Methyl_Met_ V83C and of ^{13}C-Methyl_Met_A100C as we monitor the catalyst design by 2D [^{1}H ^{13}C] HSQC. (A) ^{13}C-Methyl_Met_ V83C (B) Maleimide (1) modified ^{13}C-Methyl_Met_ V83C (C) Phosphine (P) modified ^{13}C-Methyl_Met_ V83C. (D) ^{13}C-Methyl_Met_A100C (E) Maleimide (1) modified ^{13}C-Methyl_Met_A100C (F) Phosphine (P) modified ^{13}C-Methyl_Met_A100C , (G) Addition of 1 equivalent of [Rh(acac)$(CO)_2$] to the phosphine modified protein, (inset) only a small fraction of Rh(CO) (mass 13997, less than 5% as indicated from mass spectrum peak intensity) bound. Protein samples for mass spectrometry were 10-25 μM, in 30 mM Tris-D_{11}, 10 % D_2O, 50 mM NaCl, pH 7.43 acquired using a gradient of 1% formic acid in H_2O to 1% formic acid in acetonitrile. The data were recorded over 20-30 min and the results were analysed by MassLynx V. 4.0 and its MaxEnt algorithm.

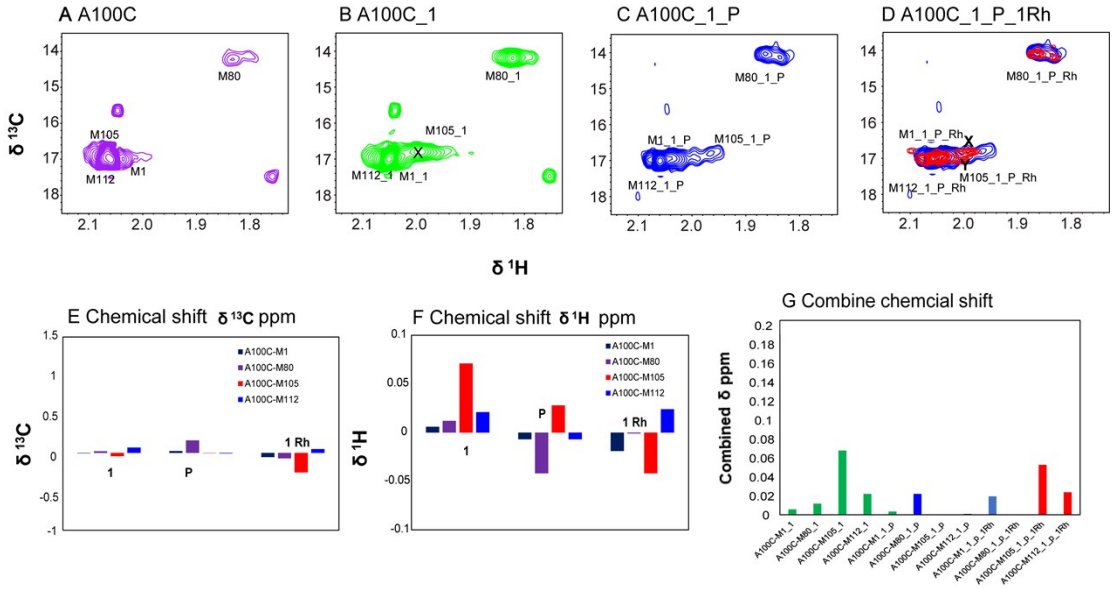


Fig S13 Monitoring the catalyst design by [^{1}H ^{13}C] HSQC NMR spectroscopy. (A) SCP-2L_^{13}C-Methyl_MetA100C (B) Maleimide (1) modified SCP-2L_^{13}C-Methyl_Met_A100C (C) Phosphine (P) conjugated maleimide (1) modified SCP-2L_^{13}C-Methyl_Met_A100C. (D) Addition of 1 equivalent of [Rh(acac)$(CO)_2$] (Rh) to phosphine-maleimide- SCP-2L_^{13}C-Methyl_Met_A100C. Changes in chemical shifts (ppm) of methionine methyl group due to covalent modifications (E) Chemical shift change in δ^{13}C dimension (F) Chemical shift change in δ ^{1}H dimension (G) combined chemical shifts δ(combine). A homogenously labelled (^{13}C-Methyl-methionine label) protein sample (~65-100µM, 30 mM Tris-D_{11}, 10 % D_2O, 50 mM NaCl, pH 7.43) at an operating frequency of 700.13 MHz for ^{1}H using a cryoprobe at 303 K were used for these experiments.

Table S6 NMR chemical shift of ^{13}C and ^{1}H of ^{13}C-methyl-methionine of SCP-2L-A100C due chemical modification during catalyst design

Entry	M1	M80	M105	M112
	^{13}C/^{1}H ppm	^{13}C/^{1}H ppm	^{13}C/^{1}H ppm	^{13}C/^{1}H ppm
SCP-2L-A100C	16.99/2.03	14.21/1.83	16.75/2.06	17.07/2.07
SCP-2L-A100C-1	16.99/2.03	14.19/1.82	16.79/1.99	17.00/2.07
SCP-2L-A100C-1-P	16.97/2.03	14.03/1.86	16.79/1.96	17.01/2.05
SCP-2L-A100C-1-P-Rh	17.02/2.05	14.10/1.86	17.05/2.0	16.96/2.05

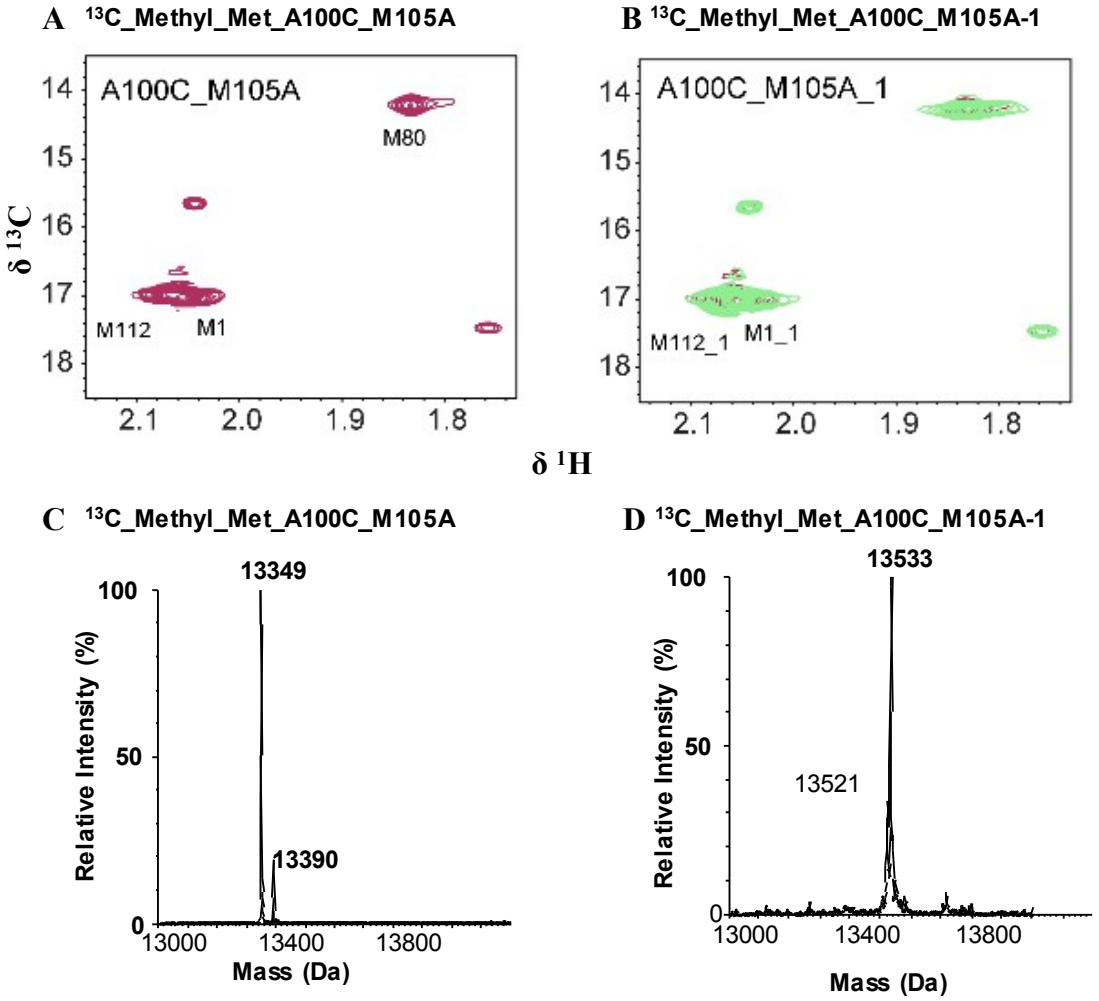


Fig S14 2D [^{1}H ^{13}C] HSQC (A,B) and Deconvoluted mass spectra (C,D) ^{13}C-Methyl_Met_SCP_2L_A100C-M105A and ^{13}C-Methyl_Met_ SCP_2L_A100C-M105A-1Protein samples for mass spectrometry were 10-25 µM, in 20 mM MES and 50 Mm NaCl buffer, Ph 6.0, acquiring using a gradient of 1% formic acid in H_2O to 1% formic acid in acetonitrile. The data were recorded over 20-30 min and the results were analysed by MassLynx V. 4.0 and its MaxEnt algorithm. For NMR, a homogenously labelled (^{13}C-Methyl-methionine label) protein sample (~60-100µM, 30 mM Tris-D_{11}, 10 % D_2O, 50 mM NaCl, pH 7.43) at an operating frequency of 700.13 MHz for ^{1}H using a cryoprobe at 303 K were used for these experiments.

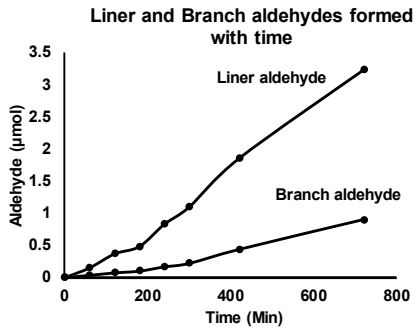


Fig S15 Hydroformylation of 1-octene. Production of linear and branched aldehydes with time.

S3 References

1 A. G. Jarvis, L. Obrecht, P. J. Deuss, W. Laan, E. K. Gibson, P. P. Wells and P. C. J. Kamer, *Angew. Chem. Int. Ed.*, 2017, **56**, 13596–13600.

2 Y.-S. Hon, C.-F. Lee, R.-J. Chen and P.-H. Szu, *Tetrahedron*, 2001, **57**, 5991–6001.

3 A. J. Dent, G. Cibin, S. Ramos, A. D. Smith, S. M. Scott, L. Varandas, M. R. Pearson, N. A. Krumpa, C. P. Jones and P. E. Robbins, *J. Phys. Conf. Ser.*, , DOI:10.1088/1742-6596/190/1/012039.

4 B. Ravel and M. Newville, *J. Synchrotron Radiat.*, 2005, **12**, 537–541.

5 M. Newville, *J. Synchrotron Radiat.*, 2001, **8**, 96–100.