

Raman Spectroscopy for the Direct Monitoring of Microbial Biotransformations

A thesis submitted to the University of Manchester for the degree of
Doctor of Philosophy in the Faculty of Science and Engineering

2017

Heidi Fisk

Supervised by Professor Jason Micklefield and Professor Royston Goodacre

School of Chemistry

Contents

ABSTRACT.....	25
DECLARATION	26
COPYRIGHT STATEMENT	27
PREFACE	28
ACKNOWLEDGEMENTS.....	29
1 INTRODUCTION	30
1.1 Electromagnetic Spectrum.....	31
1.2 Molecular Energy Levels	33
1.3 Vibrational Spectroscopy	34
1.4 Comparisons of IR and Raman	37
1.5 Fluorescence	39
1.6 Enhancement of the Raman Response	39
1.6.1 Surface-Enhanced Raman Scattering (SERS).....	40
1.6.1.1 Electromagnetic Enhancement.....	40
1.6.1.2 Chemical Enhancement (Charge-Transfer)	41
1.6.1.3 Parameter Optimisation in SERS.....	41
1.6.1.3.1 Choice of Metal.....	41
1.6.1.3.2 Type of Metal Substrate	42
1.6.1.3.3 Excitation Wavelength.....	43
1.6.1.3.4 Factors Influencing Analyte Interaction/Adsorption with the Surface.....	44
1.6.1.3.5 Aggregating Agent	44
1.6.1.3.6 Aggregation Time.....	44
1.6.1.4 Associated Advantages and Disadvantages of SERS	45
1.6.2 Resonance Raman Scattering (RRS)	46
1.7 Chemometrics.....	47
1.7.1 Data Pre-Processing	47

1.7.2	Principal Component Analysis (PCA)	48
1.7.3	Multivariate Curve Resolution-Alternating Least Squares	49
1.8	Applications of Raman Spectroscopy	49
1.9	Biocatalysts: Their Importance and Applications	50
1.10	High-Throughput Screening (HTS) Methods	53
1.11	Aim of the Thesis	56
1.12	References	57
2	Exploring Surface Enhanced Raman Scattering as a Method to Identify Regioselective Flavin-Dependent Halogenation	63
2.1	DECLARATION	63
2.2	ABSTRACT	64
2.3	INTRODUCTION	65
2.4	EXPERIMENTAL	69
2.4.1	Reagents and Materials	69
2.4.2	Halogenase Biotransformation Sample Preparation	69
2.4.3	SERS Sample Preparation	69
2.4.4	Raman Instrumentation	70
2.5	RESULTS AND DISCUSSION	71
2.5.1	Biotransformation Optimisation	71
2.5.2	SERS Optimisation	72
2.5.3	Whole-Cell Investigations	76
2.6	CONCLUSIONS	80
2.7	REFERENCES	81
2.8	SUPPORTING INFORMATION	83
3	Thiol-Functionalised Substrates for Sensitive and Rapid Detection of Regioselective Flavin-Dependent Halogenation using SERS	107
3.1	DECLARATION	107
3.2	ABSTRACT	108

3.3	INTRODUCTION.....	108
3.4	EXPERIMENTAL	112
3.4.1	Reagents and Materials.	112
3.4.2	Raman Instrumentation.....	112
3.4.3	SERS Sample Preparation of Standards.	112
3.4.4	FDH Activity Assays.	112
3.5	RESULTS AND DISCUSSION.....	113
3.5.1	Designing Thiol-Functionalised Substrates	113
3.5.2	Modelling Enzyme Activity.....	118
3.6	CONCLUSION.....	120
3.7	REFERENCES	120
3.8	SUPPORTING INFORMATION	124
4	Discrimination of Enzyme Enantioselectivity using <i>in situ</i> Surface Enhanced Raman Scattering Spectroscopy	137
4.1	DECLARATION	137
4.2	ABSTRACT.....	138
4.3	INTRODUCTION.....	138
4.4	EXPERIMENTAL	141
4.4.1	Reagents and Materials.	141
4.4.2	Raman Instrumentation.....	141
4.4.3	Biotransformation Sample Preparation	141
4.4.3.1	Small scale (1 mL) SERS biotransformation.	141
4.4.3.2	Large scale (20 mL) SERS biotransformation with comparative HPLC.....	141
4.4.4	Thioester Synthesis.	142
4.5	RESULTS AND DISCUSSION.....	142
4.5.1	Thioester Design and Optimisation.....	142
4.5.2	Highly Enantioselective Enzyme: α -Chymotrypsin.....	145
4.5.3	Low Enantioselectivity Enzyme: PLE	150

4.6	CONCLUSION.....	152
4.7	REFERENCES.....	154
4.8	SUPPORTING INFORMATION	156
5	Real-Time Monitoring of Enzyme-Catalysed Reactions using Deep UV Resonance Raman Spectroscopy	179
5.1	DECLARATION	179
5.2	ABSTRACT.....	180
5.3	INTRODUCTION.....	180
5.4	EXPERIMENTAL	183
5.4.1	Reagents and Materials.	183
5.4.2	Raman instrumentation.....	183
5.4.3	Reaction conditions for biotransformations 1 and 2.....	183
5.4.4	Reaction conditions for biotransformations 3 and 4.....	183
5.4.5	Reaction sample preparation and monitoring.....	183
5.5	RESULTS AND DISCUSSION.....	184
5.5.1	Optimisation of NHase and XO biotransformations	184
5.5.2	Instrument set up for real-time reaction monitoring.....	184
5.5.3	Monitoring NHase biotransformations 1 and 2.....	186
5.5.4	Monitoring XO biotransformations 3 and 4.....	188
5.5.5	Generalisation of the UVRR approach	190
5.6	CONCLUSION.....	191
5.7	REFERENCES.....	192
5.8	SUPPORTING INFORMATION	194
6	From Multi-Step Enzyme Monitoring to Whole-Cell Biotransformations: Development of Real-Time UVRR Spectroscopy	211
6.1	DECLARATION	211
6.2	ABSTRACT.....	212
6.3	INTRODUCTION.....	213

6.4	EXPERIMENTAL	215
6.4.1	Reagents and Materials.	215
6.4.2	Reaction sample preparation.....	215
6.4.3	UVRP Instrumentation.	215
6.4.4	Reaction set-up.	215
6.5	RESULTS AND DISCUSSION.....	216
6.5.1	Multi-step enzyme biocatalysis.....	216
6.5.2	Whole-cell biotransformations incorporating multi-step enzyme biocatalysis.....	220
6.6	CONCLUSIONS.....	223
6.7	REFERENCES	224
6.8	SUPPORTING INFORMATION	226
7	DISCUSSION	241
7.1	Discussion and Future Perspectives.....	242
7.2	REFERENCES	249
8	APPENDIX I: Published Work in Original Format	251
9	APPENDIX II: Additional Publications	258

Word Count: 62,809

LIST OF FIGURES

Figure 1.1. Diagram showing an electromagnetic wave. Oscillations of both magnetic (Z axis, horizontal) and electric fields (Y axis, vertical) occur in phase and perpendicular to one another with propagation through space (X axis, direction) (Sathyanarayana, 2015).....	31
Figure 1.2. A diagram representing the electromagnetic spectrum.	32
Figure 1.3. Diagram depicting the different ways in which light can interact with matter.....	33
Figure 1.4. A depiction of a molecular energy level diagram for a hypothetical molecule. Rotational energy levels (R) and vibrational energy levels (V) are depicted within the electronic ground state (S_0) and first excited state (S_1).	34
Figure 1.5. Jablonski diagram depicting the numerous mechanisms of electromagnetic radiation interacting with a molecule. Both infrared and fluorescence undergo absorption, whereas scattering processes can occur in either an elastic (Rayleigh, no change in energy) or inelastic manner (Raman scattering). The scattered photons may experience a reduction in energy of the incident light, known as Stokes Raman scattering. Alternatively, the molecule may lose energy, hence increasing the energy of the scattered photon, termed anti-Stokes Raman scattering. Fluorescence experiences non-radiative decay (black, dotted line), thus fluorescence emission is at a lower energy than the absorbed light.....	35
Figure 1.6. Normal vibrational modes of CO_2 A) symmetric stretch (1480 cm^{-1}), B) asymmetric stretch (2565 cm^{-1}), C) bending in-plane (526 cm^{-1}) D) bending out-of-plane (526 cm^{-1}).	38
Figure 1.7. Approximate wavelength ranges of copper (Cu), gold (Au) and silver (Ag).	42
Figure 1.8. An image depicting a colloidal SERS response. The electromagnetic field projecting from the nanoparticle (<i>i.e.</i> Au) is represented by the localised surface plasmon response (LSPR) edge. The analytes interact with the roughened nanoparticle and its electric field, experiencing enhancement of the Raman response. On the right, one can see mock responses for the analyte in regular Raman, SERS, SERS experiencing hot spots (further enhancement) and finally fluorescence, which experiences the greatest intensity response.	43
Figure 1.9. Illustration of an original data set represented by variables 1 and 2 (axis). The PCA model identifies suitable PCs (1 and 2) to uncover hidden relationships within the dataset.	48
Figure 1.10. Overview of directed evolution (DE), a mutation strategy for the development of biocatalysts. Beginning with the lead gene, it experiences diversification by mutations in the gene sequence, generating many different variants (gene library). The best variants are selected, through the application of screening techniques, which are then subjected to iterative rounds of subsequent mutagenesis, or the process is stopped as the desired traits are achieved.	52

Figure 2.1. Cycle of NADH-dependent reduction of FAD by a flavin reductase. Hypochlorous acid (HOCl) is generated within the active site.	65
Figure 2.2. The electrophilic addition of chlorine to tryptophan within the active site of FDHs. A lysine residue (K79) leads HOCl to the tryptophan substrate, whereby chlorine reacts through electrophilic addition to the π -system of the aromatic ring. A glutamate residue (E346) deprotonates the intermediate leading to the final chlorinated product.	66
Figure 2.3. Reactions of flavin-dependent tryptophan halogenases with L-tryptophan and their respective regiospecific products (X = Cl/Br).	67
Figure 2.4. An overview of SttH halogenation of tryptophan at C-6, involving glucose dehydrogenase (GDH) and flavin reductase (Fre) enzymes. Details of cloning, transformation, expression and purification are described in the SI, section 'Protein Expression and Purification'.	71
Figure 2.5. Average SERS spectra ($n = 5$) of tryptophan (blue) and 5-bromotryptophan (cyan) (2.5×10^{-5} M), characteristic peaks have been labelled. SERS spectra were acquired for 20 s with BRSC in HEPES buffer (pH 7.0).	73
Figure 2.6. Average SERS spectra ($n = 5$) of anthranilic acid (grey), 3-chloroanthranilic acid (purple), 4-chloroanthranilic acid (pink), 5-chloroanthranilic acid (orange), (2.5×10^{-5} M). Characteristic peaks have been labelled, SERS spectra were obtained for 20 s with BRSC in HEPES buffer (pH 7.0).	74
Figure 2.7. Average SERS spectra ($n = 6$) of (A) anthranilic acid, (B) FAD, (C) halogenase reaction mixture minus enzymes (analyte + cofactors + buffer) and (D) halogenase reaction mixture minus enzymes and FAD cofactor. Concentrations are representative of halogenase biotransformation conditions. SERS spectra were obtained for 20 s with BRSC in HEPES buffer (pH 7.0). SERS data has been baseline corrected and normalised (see SI, data processing for details).	75
Figure 2.8. Structures of chrysin (X = H, Y = H) and apigenin which were shown to have the greatest activity towards RadH whole-cell biotransformations. Structures of chlorochrysin (X = Cl, Y = H) and dichlorochrysin (X = Cl, Y = Cl), the products of RadH halogenation, are shown on the chrysin backbone structure.	77
Figure 2.9. Average SERS spectra ($n = 5$) of chrysin (black), chlorochrysin (red) and dichlorochrysin (blue) (2×10^{-5} M). Characteristic peaks have been labelled, SERS spectra were obtained for 10 s with CRSC in potassium phosphate buffer (pH 7.0) and KNO_3 (1×10^{-1} M).	78
Figure 2.10. SERS spectra ($n = 4$) of the <i>E. coli</i> supernatant control sample (orange), biotransformation sample at 0 min (purple) and the final biotransformation timepoint sample after 1 h incubation at 30 °C (dark cyan). Bands labelled in black were detectable, those in pale grey were unseen. SERS spectra were collected for 10 s with CRSC and KNO_3 . SERS data have been normalised, see SI, data processing for details.	79

Figure 3.1. The mechanism of flavin-dependent tryptophan halogenases. Glucose dehydrogenase (GDH) and flavin reductase (Fre) recycle essential cofactors, NAD^+ and FAD, to their reduced forms necessary for flavin-dependent halogenase (FDH) function. Position of halogenation (Cl/Br), denoted by X, is dependent on the FDH: $X_5 = \text{PyrH}$, $X_6 = \text{SttH}$, $X_7 = \text{PrnA/RebH}$110

Figure 3.2. Structures of SttH active substrates, X denotes the position of chlorination [3-indolepropionic acid produces two chlorinated products: $X_a = \text{Cl}$, $X_b = \text{H}$ (90 %) and $X_a = \text{H}$, $X_b = \text{Cl}$ (10 %)].113

Figure 3.3. Histogram representing the percentage conversion of L-tryptophan (5×10^{-4} M) to chlorinated product using different FDHs in the presence of various concentrations of thioacetate **1** or thiol **1**. Assay conditions: L-tryptophan (5×10^{-4} M) with either thioacetate **1** or thiol **1** (0, 1×10^{-4} , 2.5×10^{-4} , 5×10^{-4} , 7.5×10^{-4} M) in potassium phosphate buffer (1×10^{-3} M, pH 7.0), 1 h, 37 °C, 800 rpm (single measurements).115

Figure 3.4. SERS spectra ($n = 3$) of thiol **1** standard 2.5×10^{-5} M (black line). Thiol **1** was incubated for 1 h within assay conditions either without enzyme (red line) or with enzyme, PrnA, present (blue line). All spectra represent a final concentration of thiol **1** at 2.5×10^{-5} M, enzymes were heat denatured (95 °C, 5 min) and precipitated proteins removed using centrifugation ($21,000 \times g$ for 6 min) prior to SERS analysis. SERS conditions: 200 μL HRSC, 200 μL of assay/thiol **1** sample, 8 min aggregation time, 25 s acquisition using 633 nm excitation. Spectra were baseline corrected as detailed in the SI, data processing section.116

Figure 3.5. Structures of A) L-tryptophan (natural substrate of FDH), B) 3-(1-benzothiophen-3-yl)-L-alanine and C) thioacid **1**.....117

Figure 3.6. Principal component analysis (PCA) scores plot of known mixtures of thioacid **1** and 5-chlorothioacid (from dataset 1). Legend represents concentration of 5-chlorothioacid (i.e. $5 \mu\text{M} = 5 \times 10^{-6}$ M (10 % 5-chlorothioacid, 90 % thioacid **1**, $40 \mu\text{M} = 4 \times 10^{-5}$ M (90 % 5-chlorothioacid, 10 % thioacid **1**)). PCA was performed on full spectral range ($203 - 3400 \text{ cm}^{-1}$). Total concentration (5×10^{-5} M), 5x replicates, 633 nm laser excitation, 200 μL HRSC, 150 μL thioacid **1**/5-chlorothioacid mixture, 50 μL KNO_3 aggregating agent (5×10^{-1} M), 8 min aggregation, 20 s acquisition. Data were SNV normalised prior to chemometric analysis (as detailed in the SI, section data processing). TEV = total explained variance.....119

Figure 4.1. Basic overview of a pseudo-enantiomer approach towards esterase activity. If Y and X share identical functional groups between the two compounds, but are the two different enantiomeric forms (L- and D-/R- and S-), then they are enantiomers of one another as a chiral centre is present. The R-group differs between the two enantiomers (R_1 versus R_2), becoming pseudo-enantiomers. Upon enzyme-catalysed hydrolysis (i.e. esterase), the different R-groups will

indicate which enantiomer was selectively hydrolysed, thus uncovering enzyme enantioselectivity.140

Figure 4.2. SERS spectra (average of $n = 5$) of 1-propanethiol (red) and 2-propanethiol (blue). Structural isomers (straight versus branched chain) resulting in noticeably different SERS responses. 1×10^{-5} M concentration in potassium phosphate buffer (pH 6.2, 0.01 M) with HRSC, 785 nm irradiation using a 10 s acquisition time. Data have been baseline corrected and normalised (see SI, data processing for further details).144

Figure 4.3. Cumulative SERS spectra taken over a 30 min time course. The colour bar highlights the increasing intensity with time ($t = 0$ in blue, $t = 30$ in red). The SERS data have been normalised to the internal standard, t -BuOH at 747 cm^{-1} (see SI, data processing for further details). 1 mL reaction including an equimolar ratio of substrates L-1-A and D-2-A, with a total concentration of 2×10^{-5} M. The introduction of α -chymotrypsin (2×10^{-9} M) affords L-enantioselective hydrolysis, as demonstrated by the production of 1-propanethiol (L-1-A hydrolysis). Inset shows univariate analysis of the SERS response by plotting the peak area of 697 cm^{-1} against time.146

Figure 4.4. Data related to the 20 mL real-time SERS analysis paired with external HPLC measurements. A 20 mL biotransformation sample was prepared with L-1-A and D-2-A substrates at a concentration of 1×10^{-5} M each and $>60\%$ HRSC content. Introduction of 2×10^{-9} M α -chymotrypsin resulted in L-selective hydrolysis of L-1-A only (A), generating SERS responses correlating to 1-propanethiol. The cumulative SERS spectra are shown in (B). Normalisation to the internal standard, t -BuOH (10 % content) at 747 cm^{-1} , and the application of a colour bar enable clear observations of an increasing SERS response with respect to time. MCR-ALS was performed to the SERS data and off-line HPLC analysis enabling predictions of SERS thiol concentration, shown in (C). The SERS predicted concentrations and HPLC measured concentrations are shown in (D), giving a R^2 value of 0.9798, demonstrating excellent fit.148

Figure 4.5. (A) Cumulative SERS response of α -chymotrypsin (2 nM) in the presence of two different L-thioesters: L-1-A and L-2-A (1×10^{-5} M each). Spectra have been normalised to t -BuOH at 747 cm^{-1} (see SI, data processing for further details). (B) Plotted peak areas of 1-propanethiol (red, 697 cm^{-1}) and 2-propanethiol (blue, 590 cm^{-1}). NA 1-propanethiol gives a small peak at 624 cm^{-1} which appears as a shoulder on 2-propanethiols peak at 590 cm^{-1} , so this portion of the peak (624 cm^{-1}) was excluded from univariate analysis.150

Figure 4.6. Bar chart representing the average percentage conversions ($n = 3$) of L-1-B (orange, L-thioester) and D-1-B (green, D-thioester) with PLE by HPLC. Conditions: L-1-B (5×10^{-4} M), D-1-B (5×10^{-4} M), PLE (5 μL of 10 mg/mL solution), 10 % t -butanol and potassium phosphate buffer (1 x

10⁻² M, pH 6.2), using a 1.5 mL total volume. A 2:1 preference for the L-thioester was observed.151

Figure 5.1. Workflow of the UVRR approach for real-time reaction monitoring of multiple biotransformations.182

Figure 5.2. Average UVRR spectra (n = 5) of each analyte for both biotransformations: benzonitrile (blue), benzamide (red), p-tolunitrile (bright green) and p-toluamide (orange), hypoxanthine (green), xanthine (pink) and uric acid (purple). For NHase analytes spectra were obtained at 1.25 x 10⁻² mM, pH 7.2. For XO analytes spectra were obtained at 7.5 x 10⁻⁴ M, pH 7.6. All spectra are representative of starting reaction concentrations with characteristic peaks annotated. UVRR spectra were obtained for 20 s with baseline correction, normalisation and smoothing applied (see SI, 'data processing' for full details).185

Figure 5.3. An MCR-ALS model was applied to the UVRR data for the conversion of benzonitrile to benzamide (biotransformation 1). (a) Shows accumulative spectra taken over the 20 min time course. The colour bar highlights each time point monitored with the start (t = 0) in blue and the end point (t = 20) in red. (b) Shows the reaction dynamics from real-time UVRR measurements (denoted by outlined symbols) and off-line HPLC data (denoted by solid symbols) as a function of time. UVRR spectra were obtained for 20 s with baseline correction, normalisation and smoothing applied (see SI 'data processing' for full details). Data shown are from replicate 2.187

Figure 5.4. An MCR-ALS model was applied to the UVRR data for the conversion of xanthine to uric acid, biotransformation 3. (a) Accumulative spectra taken over the 18 min time course. The colour bar highlights each time point monitored with the start (t = 0) in blue and the end point (t = 18) in red. (b) Shows the reaction dynamics from real-time UVRR measurements (denoted by outlined symbols) and off-line HPLC data (denoted by solid symbols) as a function of time. UVRR spectra were obtained for 20 s with baseline correction, normalisation and smoothing applied (see SI 'data processing' for full details). Data shown are from replicate 1.189

Figure 5.5. An MCR-ALS model was applied to the UVRR data where it successfully deconvolved spectra into its pure components for biotransformation 4: (a) hypoxanthine (b) xanthine and (c) uric acid. (d) Shows the reaction dynamics from real-time UVRR measurements (denoted by outlined symbols) and off-line HPLC data (denoted by solid symbols) as a function of time for the conversion of hypoxanthine to xanthine to uric acid. UVRR spectra were obtained for 20 s with baseline correction, normalisation and smoothing applied (see SI 'data processing' for full details). Data shown are from replicate 1.190

Figure 6.1. Enzymatic conversion of benzonitrile to the corresponding carboxylic acid (benzoic acid) and/or carboxamide (benzamide).214

Figure 6.2. Average UVRR spectra ($n = 2$) of each analyte under investigation: benzonitrile (blue), benzamide (red) and benzoic acid (green). Spectra were obtained at 1.1×10^{-2} M in potassium phosphate buffer (2.5×10^{-2} M, pH 7.2), using conditions and concentrations that were representative of the initial reaction mixtures, with characteristic peaks identified (see Table S6.1 in SI for assignments). UVRR spectra were obtained for 20 s with baseline correction and normalisation applied (see the SI, “Data processing” for full details).	216
Figure 6.3. An MCR-ALS model was applied to the UVRR data where it successfully deconvolved spectra into its pure components for the biotransformation using pure enzymes. This figure shows the reaction dynamics from the real-time UVRR measurements (denoted by circular symbols) and off-line HPLC data (denoted by cross symbols) as a function of time for the conversion of benzonitrile (SM; blue) to benzamide (I; red) to benzoic acid (P; green), catalysed by NHase and amidase, correspondingly.	218
Figure 6.4. Multivariate curve resolution-alternating least squares-hard modelling (MCR-ALS-HM) results show good agreement with HPLC measured concentrations. Calculated kinetic rates of each enzyme were $k_1 = 0.3316$ and $k_2 = 0.0797$ for NHase and amidase, respectively.	219
Figure 6.5. MCR-ALS model applied to the whole-cell biotransformation of benzonitrile (SM, blue) to benzamide (P, red) using NHase-containing <i>E. coli</i> cells. Figure shows the reaction dynamics from the real-time UVRR measurements (denoted by circular symbols) and off-line HPLC data (denoted by cross symbols) as a function of time.	221
Figure 6.6. Modelling results for amidase-expressing bacterial whole-cells catalysing the conversion of benzamide (SM) to benzoic acid (P). (A) MCR-ALS soft modelling results and (B) MCR-ALS hard modelling results, kinetic rate was calculated as $k = 0.1986$	222
Figure 7.1. SERS spectra (averages of $n = 4$) of RadH biotransformation supernatant containing chrysin (2.5×10^{-5} M) and the equivalent of either 10 mL <i>E. coli</i> culture (top), 5 mL (middle) or 2 mL (bottom). Characteristic chrysin peaks have been highlighted in grey. Data have been baseline corrected and normalised.	245

LIST OF SUPPLEMENTARY FIGURES

Figure S2.1. Graph representing the variance in percentage conversion arising from changes in protein concentration (1×10^{-6} M). Results are shown for tryptophan (5×10^{-4} M) over a 1 h incubation period at 30 °C (single measurement shown).	87
Figure S2.2. Optimisation of SttH with GDH recycling system; optimal conditions were identified as SttH 1×10^{-5} M, Fre 1×10^{-6} M, GDH 6×10^{-6} M, MgCl_2 5×10^{-2} M, FAD 7.5×10^{-6} M, NADH 2×10^{-4} M, glucose 3×10^{-2} M (single measurement shown).	88
Figure S2.3. Structures of 29 substrates that were tested for activity towards SttH 6-halogenase. Active compounds are highlighted in boxes. Halogenated products confirmed by LRMS only are shown in blue. Halogenated products with full characterisation (^1H , ^{13}C NMR, HRMS, UV) are in green.	89
Figure S2.4. PCA scores plot depicting separation of anthranilic acid (AA, grey) 3-chloroanthranilic acid (3-Cl AA, purple), 4-chloroanthranilic acid (4-Cl AA, pink) and 5-chloroanthranilic acid (5-Cl AA, orange). Normalisation of data was performed prior to PCA. Eclipses are a guide to the eye and have no statistical significance. Total explained variance (TEV) is labelled on the axes as a percentage.	92
Figure S2.5. Loadings plot identifying key peaks enabling separation of regioisomers along the PC 1 and PC 2 axis of the PCA scores plot shown in Figure S2.4.	92
Figure S2.6. A flow diagram summarising the process undertaken to develop flavin-dependent halogenases for use in whole-cell biotransformations.	93
Figure S2.7. PCA scores plot representing separation of chrysin (circled in black), chlorochrysin (circled in red) and dichlorochrysin (circled in blue). Normalisation of data was performed prior to PCA. Eclipses are a guide to the eye and have no statistical significance. Total explained variance (TEV) is labelled on the axes as a percentage.	95
Figure S2.8. Loadings plot recognising key peaks enabling separation along the PC 1 and PC 2 axis of the PCA scores plot shown in Figure S2.7.	95
Figure S2.9. Overlaid SERS spectra of chrysin (black, 2×10^{-5} M), chlorochrysin (red, 2×10^{-5} M) and the control sample of <i>E. coli</i> supernatant (green). Characteristic peaks have been labelled. Data have been normalised (see data processing for details).	96
Figure S2.10. SEM images of borohydride-reduced silver colloid on silicon. Left, depicts images of 200x magnification and a scale bar of 100 nm. Right, 214x magnification and 100 nm scale bar.	97
Figure S2.11. SEM images of citrate-reduced silver colloid on silicon. Left, depicts images of 10x magnification and a scale bar of 1 μm . Right, 109x magnification and 200 nm scale bar.	97

Figure S2.12. SDS-PAGE of SttH protein expression in different <i>E. coli</i> competent cells.....	100
Figure S2.13. SDS-PAGE of Fre protein expression using Ni-NTA column.....	101
Figure S2.14. SDS-PAGE of GDH protein expression using Ni-NTA column.....	101
Figure S2.15. SDS-PAGE of SttH protein expression using Ni-NTA column.....	102
Figure S3.1. Figure showing comparative activity of four different flavin-dependent tryptophan halogenases towards tryptophol (blue) and indole-3-carbinol (orange). Assay conditions: [5×10^{-4} M] substrate, SttH/PrnA/PyrH/RebH (1×10^{-5} M), Fre (1×10^{-6} M), GDH (6×10^{-6} M), FAD (7.5×10^{-6} M), NADH (2×10^{-4} M), $MgCl_2$ (2×10^{-2} M), glucose (2×10^{-2} M) in potassium phosphate buffer (1×10^{-3} M, pH 7.0). 16 h incubation at 37 °C, 800 rpm (single measurements).....	127
Figure S3.2. Principal component analysis (PCA) scores plot of thiol 1 at 24 different concentrations (5 replicates shown and entire spectral range used). Colour bar represents decreasing concentrations (red to blue) with concentrations labelled. Initial concentration of 1.6×10^{-3} M, ensuing concentrations were reduced by 50 %. The LOQ has been highlighted as 9.77×10^{-8} M. Tight clustering between replicate data highlights the high reproducibility of the SERS responses. TEV = total explained variance in the first 2 PCs.....	128
Figure S3.3. SERS spectra (n = 5) of thioacid 1 (blue) and 5-chlorothioacid (red). A shorter spectral range ($400 - 2000 \text{ cm}^{-1}$) is shown as no discernible peaks occur outside of these perimeters. Total concentration (2.5×10^{-5} M), 5x replicates, 633 nm laser excitation, 200 μL HRSC, 150 μL thioacid 1 /5-chlorothioacid solution, 50 μL KNO_3 aggregating agent (5×10^{-1} M), 8 min aggregation, 20 s acquisition. Data were SNV normalised.....	129
Figure S3.4. Loadings plot of corresponding PCA for dataset 1 (Figure 3.5). Main peaks enabling PC separation along PC 1 are highlighted with asterisks (*) coloured in red, from left to right, 765, 1009, 1122, 1372 and 1616 cm^{-1} . All aforementioned peaks are characteristic of thioacid 1 and are absent in 5-chlorothioacid. Separation along PC 2 arises from subtler peaks, such as shifts in wavenumbers and changes in peak intensity.....	129
Figure S3.5. PCA scores plot of known mixtures of thioacid 1 and 5-chlorothioacid, dataset 2. PCA performed on the full spectral range ($203 - 3400 \text{ cm}^{-1}$). Total concentration (5×10^{-5} M), 5x replicates, 633 nm laser excitation, 200 μL HRSC, 200 μL thioacid 1 /5-chlorothioacid mixture, 50 μL KNO_3 aggregating agent (1×10^{-1} M), 20 s acquisition time. Data were SNV normalised prior to chemometric analysis.	130
Figure S3.6. Loadings plot of corresponding PCA for dataset 2 (Figure S3.5). Main peaks enabling PC separation along PC 1 are identical to dataset 1; from left to right, 765, 1009, 1122, 1372 and 1616 cm^{-1} (denoted by asterisks (*) in red). Separation along PC 2 arises from subtler peaks, such as shifts in wavenumbers and changes in peak intensity, again in sync with dataset 1.	130

Figure S3.7. SEM images of HRSC on silicon. Left, depicts images of 50x magnification and a scale bar of 200 nm. Right, 96x magnification and 200 nm scale bar.	131
Figure S3.8. PrnA active site with L-tryptophan bound. Key active site residues partaking in hydrogen bonding interactions are shown. Residues involved in π - π stacking (H101, F103 and W455) above and below the indole moiety are removed for clarity.	136
Figure S4.1. Graph of the ratio of peak areas of characteristic peaks for either 1-propanethiol versus 2-propanethiol (blue) or 1-pentanethiol versus 1-propanethiol (orange). Asterisk (*) identifies the concentration at which monolayer coverage is exceeded and we no longer observe a 1:1 ratio of thiols (competition between thiols due to monolayer coverage).	160
Figure S4.2. Figure represents 1-propanethiol at 2×10^{-5} M total concentration in a sealed vessel experiencing different pH conditions using various buffers. At various time points, samples were taken from the vessel for SERS analysis. As can be seen, the peak area is decreasing over time, independently of pH. After 40 min, a 50 % reduction in peak area is observed.	163
Figure S4.3. Cumulative SERS spectra taken over a 30 min time course. The colour bar highlights the increasing intensity with time (t = 0 in blue, t = 30 in red). The SERS data has been normalised to the internal standard, t-BuOH at 747 cm^{-1} . 1 mL reaction including substrates L-2-A and D-1-A, at a concentration of 1×10^{-5} M per substrate. The introduction of α -chymotrypsin gives rise to L-enantioselectivity, demonstrated by the production of 2-propanethiol (L-2-A hydrolysis). Inset shows univariate analysis of the SERS response by plotting the peak area of 590 cm^{-1} against time.	164
Figure S4.4. Image illustrating the process of monitoring the biotransformation with both SERS and HPLC analysis. The 20 mL reaction mixture contains: 62.5 % HRSC, 10 % t-BuOH, potassium phosphate buffer (5×10^{-3} M, pH 6.2 final), L-thioester [1×10^{-5} M], D-thioester [1×10^{-5} M] and α -chymotrypsin [2×10^{-9} M] added last. For SERS analysis, 0.4 mL of the reaction sample (un-modified) is added to a 1 mL glass vial for immediate analysis (10 s acquisition, 785 nm). For HPLC analysis, 1 mL of the reaction sample is instantaneously quenched using 1 % TFA. Subsequent protein removal (centrifugation at 21,000 xg for 6 min) and overnight lyophilisation of the supernatant to dryness, before the solid is re-dissolved in 70 μ L of acetonitrile and centrifuged (21,000 xg for 6 min) prior to HPLC analysis (24 min acquisition time, chiral column).	165
Figure S4.5. Flow diagram summarising the data pre-processing and MCR-ALS process used for predicting concentrations of product (<i>i.e.</i> 1-propanethiol) from the reaction.	166
Figure S4.6. MCR-ALS model was applied to the SERS data where it successfully deconvolved spectra into its pure components (A) 1-propanethiol (product) and (B) t-butanol (internal standard).	166

Figure S4.7. Peak areas of 1-propanethiol (red, 697 cm^{-1}) and 2-propanethiol (blue 590 cm^{-1}) over the reaction monitoring period for 3 replicates (**1 – 3**). Conditions: L-1-A and L-2-A, 1×10^{-5} M each, 2×10^{-9} M α -chymotrypsin addition. 2-propanethiol response has been adjusted by 1.6035 (average of 54 data points) to account for the increased SERS intensity relative to 1-propanethiol.167

Figure S4.8. Peak areas of 1-propanethiol (red, 697 cm^{-1}) and 2-propanethiol (blue 590 cm^{-1}) over the reaction monitoring period. **(A)** (L-1-B + D-2-B) and **(B)** (L-2-B and D-1-B) are combinations of L- and D-thioesters in the presence of PLE (50 μL of 10 mg/mL solution), whereas **(C)** (L-1-B + L-2-B) and **(D)** (D-1-B + D-2-B) are reactions with both L-/or D-thioesters. 2-propanethiol peaks areas have been adjusted to account for the greater intensity SERS response. In every reaction, 1-propanethiol always generates the greatest response, even when it is associated to the D-thioester SM. This identifies that PLE is demonstrating a preference towards the straight sidechain (1-propanethiol) over the branched isomer (2-propanethiol).168

Figure S4.9. Figures identify the importance of SERS optimisation. In the absence of buffer, no trend is observed with decreasing concentrations of 2-propanethiol (left). However, exploration of pH identified that pH 6.2 gave optimal responses, observing a sigmoidal-type response on decreasing concentration of 2-propanethiol (right). Sample 1 = 20 μM , consecutive samples each decrease by 2×10^{-6} M: 2 μM (i.e. sample 5 = 12 μM , sample 9 = 4 μM). Conditions: HRSC, 1 min. aggregation time, 6 s acquisition, 785 nm, potassium phosphate buffer (pH 6.2, 1×10^{-1} M) or no buffer.169

Figure S4.10. SEM images of HRSC deposited on silicon. Left, depicts images of 50x magnification and a scale bar of 200 nm. Right, 96x magnification and 200 nm scale bar.169

Figure S4.11. Marfey's reagent incubated with L- and D-thioester containing 2-propanethiol within its structure (L-2-A and D-2-A). The addition of Marfey's reagent and base (NaHCO_3) facilitates its reaction with primary amines to enable quick and easy separation of enantiomers on reverse-phase C18 column, by means of a second chiral centre (diastereomer).174

Figure S4.12. HPLC calibrations (chiral column) of substrate subset A and the corresponding L-product, N-Boc-L-phenylalanine. Absorbance measured at 220 nm, the averages of triplicate data shown are shown.176

Figure S4.13. HPLC calibrations (chiral column) of substrate subset B and the corresponding products, N-Cbz-L-/D-alanine. Absorbance measured at 254 nm, the averages of triplicate data shown are shown.177

Figure S5.1. Annotated instrument set-up to monitor biotransformation using UV resonance Raman spectroscopy. The x 40 UVR objective was carefully focused onto the reaction mixture (10 mL scale), with 100 % power on sample (~ 0.2 mW at sampling point). Throughout the course of the reaction, samples for HPLC analysis were removed as well as UVR data collected (20 s spectral

acquisitions). The solution was constantly stirred using the magnetic stirrer plate beneath the stage and the magnetic stirrer bar in the reaction vessel.....197

Figure S5.2. Average UVRR spectra ($n = 5$) of **a)** initial solid xanthine sample and **b)** degraded solid xanthine sample after 45 min (after each min, a 20 s spectral acquisition was acquired). A broadening of peaks around 1550 cm^{-1} indicates C-C presence suggesting photo-degradation...198

Figure S5.3. **(a)** Average UVRR spectra ($n = 5$) and **(b)** UV-Vis absorbance spectra of xanthine at various pH: 3.6, 5.0, 7.0, 7.6, 8.0, and 9.2. A bathochromic shift in the UV-Vis absorption spectra was observed on increasing the pH from pH 3.6 to 9.2, consequently, the UVRR spectra of xanthine changed due to it being in different ionisation states. **(c)** The intensity difference of the key peaks that change and **(b)** a plot of the centre of the peaks that shift on increasing the pH of the solution from pH 3.6 to 9.2.....201

Figure S5.4. **(a)** Average UVRR spectra ($n = 5$) and **(b)** UV-Vis absorbance spectra of hypoxanthine at various pH: 3.6, 5.0, 7.0, 7.6, 8.0, and 9.2. A bathochromic shift in the UV-Vis absorption spectra was observed on increasing the pH from pH 3.6 to 9.2, consequently, the UVRR spectra of hypoxanthine changed due to being in different ionisation states. **(c)** The intensity difference of the key peak that changes and **(d)** a plot of the centre of peak that shifts on increasing the pH of the solution from pH 3.6 to 9.2.....202

Figure S5.5. **(a)** Average UVRR spectra ($n = 5$) and **(b)** UV-Vis absorbance spectra of uric acid at various pH: 3.6, 5.0, 7.0, 7.6, 8.0, and 9.2. A bathochromic shift in the UV-Vis absorption spectra was observed on increasing the pH from pH 3.6 to 9.2, consequently, the UVRR spectra of uric acid changed due to being in different ionisation states. **(c)** The intensity difference of the key peaks that change and **(d)** a plot of the centre of peak that shifts on increasing the pH of the solution from pH 3.6 to 9.2.....203

Figure S5.6. Flow diagram summarising the data pre-processing and MCR-ALS process used for in predicting concentrations of each analyte from the reaction.....204

Figure S5.7. An MCR-ALS model was applied to the UVRR data where it successfully deconvolved spectra into its pure components **a)** benzonitrile (substrate) and **b)** benzamide (product) as shown for biotransformation 1.205

Figure S5.8. HPLC calibrations for **(a)** benzonitrile **(b)** benzamide **(c)** *p*-tolunitrile and **(d)** *p*-toluamide at 254 nm absorbance. Plots show the mean plus associated standard deviation (SD) error bars from triplicate data.....205

Figure S5.9. A MCR-ALS model was applied to the UVRR data where it successfully deconvolved spectra into its pure components **(a)** *p*-tolunitrile (green) and **(b)** *p*-toluamide (orange) as shown for biotransformation 2 **(c)** Shows the reaction dynamics from real-time UVRR measurements (denoted

by outlined symbols) and off-line HPLC data (denoted by solid symbols) as a function of time for the conversion of p-tolunitrile to p-toluamide. UVRR spectra were obtained for 20 s with baseline correction, normalisation and smoothing applied (as detailed in 'Materials and methods: data processing').....206

Figure S5.10. An MCR-ALS model was applied to the UVRR data where it successfully deconvolved spectra into its pure components (a) xanthine (substrate) and (b) uric acid (product) as shown for biotransformation 1.....207

Figure S5.11. HPLC calibrations for a) hypoxanthine b) xanthine and c) uric acid at 254 nm absorbance. Plots show the mean plus associated standard deviation (SD) error bars from triplicate data.207

Figure S5.12. (a) Overall schematic illustrating the two known pathways to catalyse the conversion of nitrile containing compounds into their corresponding carboxylic acid, either in a single step (nitrilase) or a multicomponent process (nitrile hydratase and amidase) Plots b-d) show average UVRR spectra (n = 5) of each analyte: (b) pure spectra of benzonitrile, benzamide and benzoic acid (1.25×10^{-2} M, pH 7.2) (c) pure spectra of 3-pyridinecarbonitrile, nicotinamide (vitamin B₃) and nicotinic acid (2.5×10^{-2} M, pH 7.2) and (d) pure spectra of pyrazinecarbonitrile and pyrazinamide (anti-tuberculosis drug) (2.5×10^{-2} M, pH 7.2). Characteristic peaks are annotated. UVRR spectra were obtained for 20 s with baseline correction, normalisation and smoothing applied (as detailed in 'Materials and methods: data processing').209

Figure S6.1. Annotated instrument set-up, including flow-cell apparatus, to monitor biotransformations using UV resonance Raman spectroscopy. The 40x UVRR objective was focussed into the quartz flow-cell containing the reaction mixture. 100 % power on the sample (~0.2 mW at sampling point) with UVRR data collection using a 20 s acquisition time throughout the time course. Mixing of the reaction sample was achieved using a stirrer bar within the reaction reservoir, as well as continuous flow of the mixture via the peristaltic pump (anti-clockwise direction of flow, plastic tubing connects the quartz flow-cell to the reaction reservoir). Samples for HPLC analysis were removed from the reaction reservoir. Note: set-up as shown above contains E. coli whole-cells, thus the reaction mixture is opaque in appearance.....229

Figure S6.2. Flow diagram summarising data pre-processing and MCR-ALS process used for predicting concentrations of each analyte from the reaction mixture.230

Figure S6.3. MCR-ALS deconvolved spectrum (dashed, red line) of individual analytes from the mixture compared with their known UVRR spectrum from standards (solid, blue line). The initial reaction mixture includes benzonitrile (SM) and is initiated upon the addition of NHase and amidase as purified enzymes.231

Figure S6.4. MCR-ALS prediction from UVRR of analyte concentrations versus HPLC measured concentrations (mM). With R^2 values ranging from 0.9076 to 0.9895 and indicates that results are in very good agreement with one another.....	232
Figure S6.5. Flow diagram summarising the MCR-ALS and HM processes. MCR-ALS model predicts the concentration profiles of each analyte, which are then taken forward to HM.....	233
Figure S6.6. Kinetic model used within the MCR-ALS-HM approach to calculate the concentrations of each analyte at any time point during the reaction where <i>A</i> is benzonitrile, <i>B</i> is Benzamide, <i>C</i> is Benzoic acid and <i>t</i> is time. We assume that the biotransformations followed a first order reaction, during which <i>A</i> had converted to <i>B</i> , and <i>B</i> had subsequently converted to <i>C</i> , at reaction rate constants k_1 and k_2 respectively.	233
Figure S6.7. MCR-ALS deconvolved spectrum (dashed, red line) of individual analytes from the mixture compared with the known UVRR spectrum (solid, blue line). The initial reaction mixture includes benzonitrile (SM) and is initiated upon the addition of NHase-containing <i>E. coli</i> cells. ...	235
Figure S6.8. MCR-ALS predicted UVRR concentration versus HPLC measured concentration (mM). NHase whole-cell biotransformation gave R^2 values of 0.9514 and 0.7286 for benzonitrile (SM) and benzamide (P), respectively.....	235
Figure S6.9. Kinetic model used by MCR-ALS-HM approach to calculate the concentrations of each reactant at any time during the one-step biotransformations where <i>A</i> is benzonitrile or benzamide (as the product), <i>B</i> is benzamide (as the starting material) or benzoic acid and <i>t</i> is time. We assume that reaction followed a first order reaction in which <i>A</i> had converted to <i>B</i> at a reaction rate constant k	236
Figure S6.10. MCR-ALS deconvolved spectrum (dashed, red line) of individual analytes from the mixture compared with their known UVRR spectrum (solid, blue line). Initial reaction mixture includes benzamide (SM) and is started by the addition of amidase-containing <i>E. coli</i> cells.	236
Figure S6.11. MCR-ALS predicted UVRR concentration versus HPLC measured concentration (mM). Amidase whole-cell biotransformation gave R^2 values of 0.9439 and 0.9862 for benzamide (SM) and benzoic acid (P), respectively.....	237
Figure S6.12. (A) UVRR spectra of in vivo biotransformation involving benzonitrile (SM) to benzamide (I) then benzoic acid (P) using NHase and amidase from within whole-cells. Colour bar represents the changing UVRR response with time. (B) UVRR response of standards: benzamide (red), benzoic acid (green) and <i>E. coli</i> BL21 (DE3) cells (grey) in potassium phosphate buffer (2.5×10^{-2} M, pH 7.2). Bacterial peaks at 1329 and 1481 cm^{-1} coincide with characteristic benzamide (1413 cm^{-1}) and benzoic acid (1389 cm^{-1}) peaks.....	237

Figure S6.13. SDS-PAGE (16 %) of NHase protein expression using Ni-NTA. Both the α and β subunits can be clearly visualised.....	239
Figure S6.14. SDS-PAGE of amidase protein expression using Ni-NTA.	240

LIST OF TABLES

Table 1.1. Common HTS techniques used within industry and academia, with a brief description about each technique and its associated advantages and disadvantages	54
Table 2.1. Conversion of substrate after 1 h with SttH.	72
Table 2.2. Discriminant bands from SERS (in cm^{-1}) for chrysin, chlorochrysin and dichlorochrysin.	77
Table 4.1. Table showing the structures of the 8 thioester compounds. Substrate set A are substrates towards α -chymotrypsin hydrolytic activity. Substrate set B are pig liver esterase (PLE) substrates. Each of these substrates have L- and D-enantiomeric forms, with two differing thioester groups: either 1-propanethiol or 2-propanethiol, generating 4 compounds per subset (A/B).	143
Table 5.1. A summary of the regression co-efficients (R^2) across all five replicates for the two separate biotransformations:	188

LIST OF SUPPLEMENTARY TABLES

Table S2.1. Raman and tentative SERS band assignments for tryptophan (Aliaga <i>et al.</i> , 2009, Leyton <i>et al.</i> , 2012). Raman of solid tryptophan collected at 633 nm, 60 s acquisition time at 50 % power and 785 nm, 60 s acquisition time at 25 % power.....	90
Table S2.2. Raman and tentative SERS band assignments for anthranilic acid (Govindarajan <i>et al.</i> , 2011). Raman of solid anthranilic acid collected at 633 nm, 5 s acquisition time at 1 % power and 785 nm, 30 s acquisition time and 50 % power.	91
Table S2.3. SERS and tentative band assignments of chrysin. Conditions used: CRSC, 10 s, acquisition, phosphate buffer (pH 7.0), chrysin (2×10^{-5} M) and 50 μL KNO_3 (1×10^{-1} M).....	94
Table S2.4. Concentration of imidazole within purification buffers.	99
Table S4.1. Table representing the structures of the thiols described in Figure S4.1 and the concentration of monolayer coverage.	161

Table S4.2. Tentative SERS band assignments of 1-propanethiolate and 2-propanethiolate (S ⁻) interacting with HRSC (785 nm irradiation).	161
Table S4.3. Table presenting the retention times of thioester compounds on a Phenomenex Lux-Cellulose 3 chiral column using isocratic gradients. Successful separation of all thioesters, identifying that they are enantiomerically pure.	175
Table S5.1. Tentative UVRR band assignments for nitrile hydratase catalysed biotransformations at pH 7.2: benzonitrile and benzamide (biotransformation 1) and <i>p</i> -tolunitrile and <i>p</i> -toluamide (biotransformation 2) *(Chatterjee <i>et al.</i> , 1978, Gao <i>et al.</i> , 1990, Mrozek <i>et al.</i> , 2001, Fleming <i>et al.</i> , 2008, Brittain, 2009).	199
Table S5.2. Tentative UVRR band assignments for xanthine oxidase catalysed biotransformations at pH 7.6: xanthine to uric acid (biotransformation 3) and hypoxanthine to xanthine to uric acid (biotransformation 4) *(Kodati <i>et al.</i> , 1990, Chowdhury <i>et al.</i> , 2000, Krishnakumar and Arivazhagan, 2004, Arivazhagan and Jeyavijayan, 2010, 2011, Goodall <i>et al.</i> , 2013).	199
Table S5.3. The R ² co-efficients obtained for three replicates for the conversion of <i>p</i> -tolunitrile to <i>p</i> -toluamide (biotransformation 2).	206
Table S5.4. A summary of the R ² co-efficients across all three replicates for the conversion of hypoxanthine to xanthine to uric acid (biotransformation 4).	208
Table S6.1. Tentative UVRR band assignments of benzonitrile, benzamide and benzoic acid at pH 7.2 *(Chatterjee <i>et al.</i> , 1978, Gao <i>et al.</i> , 1990, Mrozek <i>et al.</i> , 2001, Brittain, 2009, Sparrow <i>et al.</i> , 2001, Trout <i>et al.</i> , 2005)	229
Table S6.2. Table denoting the total concentration of analytes with the reaction mixture at various timepoints as calculated by HPLC (calibration adjusted concentrations).	234
Table S6.3. Table stating the OD ₆₀₀ values at various time points throughout the reaction monitoring process. Samples were diluted 1:5 prior to OD ₆₀₀ analysis.	234

LIST OF SCHEMES

Scheme 3.1 Synthesis overview and structures of all initial organosulfur substrates. (1) The conversion of tryptophol to thioacetate 1 , followed by hydrolysis to thiol 1 . (2) Indole-3-carbinol conversion to thioacetate 2 , ensued by reduction to thiol 2 . (3) Thiol 3 , 3-mercaptoindole, a commercially available compound.	114
Scheme 3.2. Synthesis of thioacid 1 (blue) via Lawesson's Reagent reaction using 3-indolepropionic acid. 5-chlorothioacid (red) synthesised using a two-step process: initial regioselective chlorination	

of 3-indolepropionic acid using PyrH (FDH) at C-5, subsequent conversion of carboxylic acid to thioacid using Lawesson's Reagent.....	118
Scheme 5.1. Biotransformations (1-4) selected for monitoring by UVRR: (a) bioconversions of nitriles to the corresponding amides by nitrile hydratase (NHase); (b) biooxidation of purines by xanthine oxidase (XO).	184

LIST OF SUPPLEMENTARY SCHEMES

Scheme S4.1. Common interactions of free thiols (-SH) or thiolates (-S ⁻). (1) Thiol oxidation resulting in disulfide bond (-S-S-) formation. (2) Thiolate exchanges its organic R-group (R ₃) with the thioester R-group (R ₂). (3) Similarly, a thiolate exchanges its R-group with an R-group associated to a disulfide bond.	162
--	-----

LIST OF EQUATIONS

Equation 1.1	$c = \lambda\nu$	31
Equation 1.2.	$E = h\nu$	33
Equation 1.3.	$h\nu + E_i = h\nu' + E_f$	36

LIST OF SUPPLEMENTARY EQUATIONS

Equation S3.1. LOD equation. SD = standard deviation of colloidal blank, c = y intercept, m = the gradient of a straight line. The LOD for thiol 1 was calculated as 24.2 nM with HRSC	128
Equation S3.2. LOQ equation. SD = standard deviation of colloidal blank, c = y intercept, m = the gradient of a straight line. 97.7 nM was identified as the LOQ for thiol 1 with HRSC.....	128

LIST OF ABBREVIATIONS

- 2x YT – 2x yeast extract tryptone
- ACN – Acetonitrile
- AE – Arctic express
- AI – Auto-induction
- BRSC – Borohydride-reduced silver colloid
- CARS – Coherent anti-Stokes Raman scattering
- CRSC – Citrate-reduced silver colloid
- DE – Directed evolution
- *E. coli* – *Escherichia coli*
- *ee* – Enantiomeric excess
- EM – Electromagnetic
- ep-PCR - Error-prone polymerase chain reaction
- FA – Formic acid
- FAD – Flavin adenine dinucleotide
- FDH – Flavin-dependent halogenase
- FTIR – Fourier transform infrared
- FWHM – Full width at half maximum
- GC – Gas chromatography
- GC-MS – Gas chromatography-mass spectrometry
- HEPES – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
- HPLC – High performance liquid chromatography
- HRSC – Hydroxylamine-reduced silver colloid
- HRMS – High-resolution mass spectrometry
- HTS – High-throughput screening
- IPTG - Isopropyl β -D-1-thiogalactopyranoside
- IR – Infrared
- LB – Luria-Bertani broth
- LC – Liquid chromatography
- LC-MS – Liquid chromatography-mass spectrometry
- LOD – Limit of detection
- LOQ – Limit of quantification

- LRMS – Low-resolution mass spectrometry
- LSPR - Localised surface plasmon resonances
- MCR-ALS – Multivariate curve resolution-alternating least squares
- MOPS – 3-(*N*-Morpholino)propanesulfonic acid
- MS – Mass Spectrometry
- MWCO – Molecular weight cut-off
- NAD(P)H - Nicotinamide adenine dinucleotide(phosphate)
- NHase – Nitrile hydratase
- NMR – Nuclear magnetic resonance
- Ni-NTA – Nickel nitrilotriacetic acid
- PAT – Process analytical technology
- PC – Principal component
- PCA – Principal component analysis
- PLE – Porcine liver esterase
- QbD – Quality by design
- SAMs – Self-assembled monolayers
- SEM – Scanning electron microscopy
- SE(R)RS – Surface enhanced Raman spectroscopy (resonance)
- SDS – Sodium dodecyl sulphate
- SDS-PAGE - Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
- SI – Supporting Information
- t-BuOH – tert-butanol
- TERS – Tip-enhanced Raman scattering
- TEV – Total explained variance
- TFA – Trifluoroacetic acid
- TRIS-HCl - Tris(hydroxymethyl)aminomethane hydrochloride
- UHPLC – Ultra-high performance liquid chromatography
- UHTS – Ultra-high-throughput screening
- UVRR – Ultraviolet resonance Raman
- UV – Ultraviolet
- UV-Vis – Ultraviolet-visible
- Vis – Visible
- XO – Xanthine oxidase

ABSTRACT

The University of Manchester

Heidi Fisk

Doctor of Philosophy

Raman Spectroscopy for the Direct Monitoring of Microbial Biotransformations

25th September 2017

Interest in biocatalysts is increasing as they catalyse reactions under mild conditions and often provide enhanced control surrounding regio-, stereo- and enantio-selectivity. Yet, their use within the pharmaceutical industry is limited compared to regular chemocatalysts. Advances in protein engineering strategies are increasing the suitability of biocatalysts within industrial processes. However, these strategies produce huge variant libraries ($10^6 - 10^8$ is typical) and the availability of the high-throughput screening methods to rapidly assess the mutant's fitness is currently lacking, ultimately slowing their development. Raman spectroscopy is a technique offering molecular specific information and is highly suited towards aqueous conditions, thus presenting itself as attractive alternative screening method. Raman scattering is inherently weak, thus enhancement techniques, such as surface-enhanced Raman scattering (SERS) and ultraviolet resonance Raman (UVR) spectroscopy, can enhance the weak response. Discussed within, we successfully implement SERS and UVR to monitor a wide range of biotransformations.

Firstly, the regioselective activity of flavin-dependent halogenases (FDHs) were investigated using SERS. These studies identified that SERS could distinguish between regioisomeric products. However, the complexity of the biotransformation, involving essential cofactors, caused significant problems in the detection of substrate and product, rendering this approach unsuccessful. Next, we developed thiol-functionalised substrates to combat competition experienced by the cofactors. Exploiting sulfur's high affinity towards silver proved to be extremely successful, resulting in a negligible cofactor response. However, FDHs did not display activity towards the sulfur-containing substrates, thus, we were unable to monitor activity. Furthermore, we implemented this thiol-functionalisation approach to effectively distinguish between the enantioselectivity of a protease: α -chymotrypsin. Synthesis of novel pseudo-enantiomer substrates, incorporating different thiol-linkers, led to easy, real-time discrimination of enantioselectivity using SERS.

Lastly, an on-line UVR approach was developed which allowed accurate and reproducible measurements of nitrile-hydrolysis, mainly nitrile hydratase. Additional development of a flow-cell set-up and introduction of whole-cells further established its aptness towards industrial processes, such as process analytical technology (PAT). The research presented within this thesis highlights the suitability of Raman enhancement techniques to rapidly analyse biotransformations, in some cases real-time measurements have afforded accurate and reproducible quantitative analysis.

DECLARATION

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

COPYRIGHT STATEMENT

- i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and she has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.
- ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.
- iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.
- iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see <http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420>), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see <http://www.library.manchester.ac.uk/about/regulations/>) and in The University’s policy on Presentation of Theses.

PREFACE

Overall, this thesis comprises of seven chapters. Throughout, the aim is to highlight the suitability of Raman spectroscopy, particularly SERS and UVRR, to monitor enzyme-catalysed reactions as an alternative screening technique. The first of these chapters is a general introduction to Raman spectroscopy and its enhancement techniques, specifically SERS and UVRR, along with an introduction to biocatalysis and current high-throughput screening methods.

The second chapter introduces flavin-dependent halogenases, their optimisation, as well as the development of SERS to monitor the regioselective halogenation of substrates.

Chapter 3 further investigates SERS to monitor flavin-dependent halogenases, yet this time incorporating thiol-functionalised substrates to overcome cofactor interactions which were problematic in the previous chapter. Chapter 4 implements a similar thiol-functionalisation type-approach to monitor enantioselective hydrolysis. These studies involve pseudo-enantiomer substrates and *in situ* SERS analysis (real-time) to afford quantitative measurements of enantioselectivity.

The next two chapters (Chapters 5 and 6) investigate UVRR to monitor the progression of biotransformations in real-time. Chapter 5 initially studies nitrile hydratase and xanthine oxidase catalysed reactions, producing reliable and reproducible quantitative measurements, in sync with off-line HPLC. Chapter 6 further explores UVRR to monitor nitrile hydratase activity combined with an amidase, to detect three analytes simultaneously. The development of a flow-cell apparatus and the introduction of whole-cells, as opposed to purified proteins, proved to be efficacious and further demonstrates the capabilities and versatility of Raman techniques to monitor biotransformations.

Finally, Chapter 7 provides a discussion summarising the previous chapters, along with an outlook on the future perspectives of the work.

ACKNOWLEDGEMENTS

Firstly, I would like to thank both of my supervisors, Jason Micklefield and Roy Goodacre. As my project has been multidisciplinary, combining chemical biology and analytical chemistry, both supervisors have been fundamental to my development and progression within these fields. I am extremely grateful for their advice and guidance throughout, as well as encouraging my tenacity and desire to investigate challenging and novel subjects. During my research, I have often been pessimistic, having supervisors that were idealists/realists has helped balance my often-cynical nature! I am also very grateful for their support and encouragement towards regularly presenting my work, attending international conferences and pursuing a 3-month placement at GlaxoSmithKline (Stevenage, UK). My confidence has grown immensely since the start of my studies, particularly towards presenting.

My upmost thanks go to Peter Sutton and Joe Adams for welcoming me into their lab at GSK. I am extremely thankful for this experience as I gained so much from my time in industry and met some fantastic people.

I would like to thank some important colleagues, which I am now able to call friends. Chloe Westley and Kat Hollywood, who have always provided help, advice and support throughout, as well as affording successful collaborations. Jonathan Latham for insights towards many aspects of biochemistry, particularly NHase investigations. Yun Xu was fundamental towards my understanding of chemometrics, as well as performing complex data analysis methodologies within this thesis. In addition, I appreciate all group members within the Micklefield and Goodacre groups, along with the sLola GSK project. Those group members have made my four years extremely enjoyable, but also listened to my queries and concerns when things may not have progressed as expected.

Special thanks are given to my boyfriend, James. He has always supported my desires to further my education and encouraged me to take risks. He has made my time in Manchester truly memorable and kept me sane during challenging times! '*Carpe Diem*' has been an important philosophy for us, especially throughout these four years. Finally, I must thank all of my crazy family, particularly my parents, Steph and Richard, and step-mum, Kirsty. They have always believed in me and given me strength at difficult times. Your unconditional support has meant the world to me.

1 INTRODUCTION

1.1 Electromagnetic Spectrum

Throughout this thesis different spectroscopic and spectrophotometric techniques are discussed, with emphasis placed upon vibrational spectroscopy. Therefore, it is important to introduce the properties of light and its fundamental role within these techniques. Light is a form of electromagnetic radiation, which refers to electromagnetic waves that consist of oscillating electric and magnetic fields. These electromagnetic waves are perpendicular to one another and display properties of a sine wave (Figure 1.1) (Atkins and De Paula, 2006, Gupta, 2007).

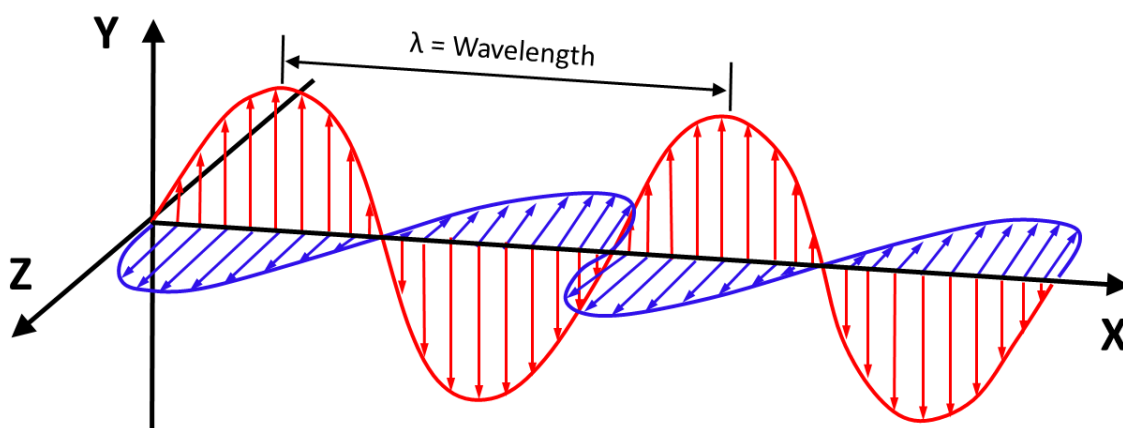


Figure 1.1. Diagram showing an electromagnetic wave. Oscillations of both magnetic (Z axis, horizontal) and electric fields (Y axis, vertical) occur in phase and perpendicular to one another with propagation through space (X axis, direction) (Sathyanarayana, 2015).

The electromagnetic spectrum organises electromagnetic radiation based on their varying frequencies (ν) or wavelengths (λ) of electromagnetic waves. The relationship between frequency and wavelength is described by Equation 1.1, where c = speed of light ($299792458 \text{ ms}^{-1}$), λ = wavelength (nm) and ν = frequency (Hz).

$$c = \lambda\nu$$

Equation 1.1

The electromagnetic spectrum ranges from very high energy electromagnetic radiation, *i.e.* gamma rays (γ -rays), which have short wavelengths and subsequent high frequencies, to low energy radiation, *i.e.* radio waves, with long wavelengths and low frequency. Figure 1.2 shows the various types of radiation and their corresponding properties in relation to frequency, wavelength and energy. Throughout this thesis, focus is placed upon vibrational excitation comprising of infrared (IR), visible (vis) and ultraviolet (UV) radiation.

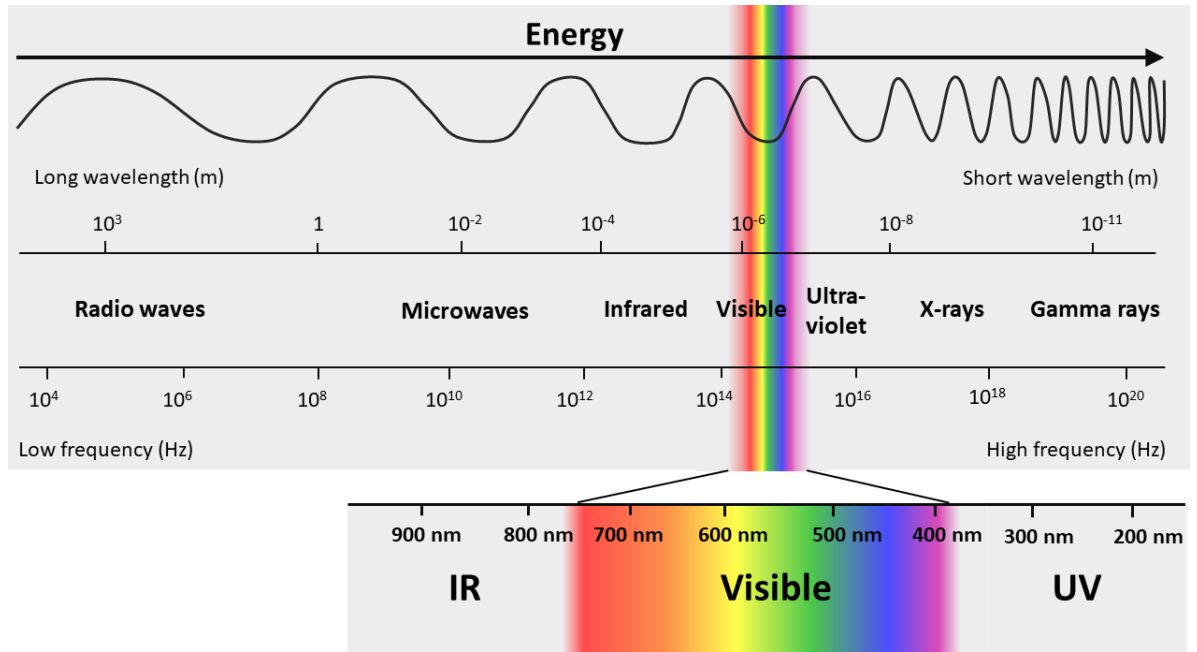


Figure 1.2. A diagram representing the electromagnetic spectrum.

In the late 19th century, the original theory that light behaved as a wave of infinite energy did not account for a series of experiments performed at that time, including the photoelectric effect as observed by Albert Einstein in 1905 (Einstein, 1905). In brief, light was shone onto a metal surface with subsequent detectable electrically charged particles, now known as electrons, emitted from the surface. If these observations were consistent with the proposed theory that light behaved solely as a wave possessing any amount of energy, then one would expect that increasing the intensity of the incident radiation (hence the amplitude) would eventually provide enough energy for the metal to emit those surface electrons. However, this was not the case. Alternatively, Einstein discovered that only when the frequency of the incident radiation was increased, it was then able to breach the threshold limit and emit electrons, independent of the intensity of the radiation. Thus, it was hypothesised that rather than a continuous wave of radiation, it was in fact arriving as discrete bundles of energy, known as quanta.

At a similar time, Maxwell Planck also experienced a similar effect when studying black-body radiation (Planck, 1914). He developed a hypothesis that particles present within a metal would only 'oscillate' at certain frequencies, again with energy differences between frequencies *i.e.* quanta. Ultimately, these observations could be explained using Planck's equation relating the energy (E) of particles, later named as photons, to the frequency (ν) of the electromagnetic radiation, by implementing Planck's constant ($h = 6.626 \times 10^{-34}$ J.s), Equation 1.2 (Lin-Vien *et al.*, 1991).

$$E = h\nu$$

Equation 1.2.

Planck's constant helped to explain Einstein's observation of the photoelectric effect, concluding that electromagnetic radiation comprises discrete quanta of light (photons), and that the energy of a photon was determined by its frequency (Fayer, 2010, Broglie, 1924). Wave-particle duality is the term given to describe these observations that electromagnetic radiation behaves as both a wave and particle. For simplicity, light will be referred to as behaving as a particle (photon), rather than a wave, when discussing the interactions of light with matter (Einstein and Infeld, 1966).

1.2 Molecular Energy Levels

Light can be categorised with regards to the type of interactions it has with a sample, these are seen in Figure 1.3. We are predominantly concerned with the scattering of light, with regards to Raman spectroscopy, and absorption, for IR spectroscopy. The analysis of light which is scattered, absorbed or emitted can generate detailed chemical information, as well as insights into the molecular composition of the compound(s) under investigation (molecular spectroscopy).

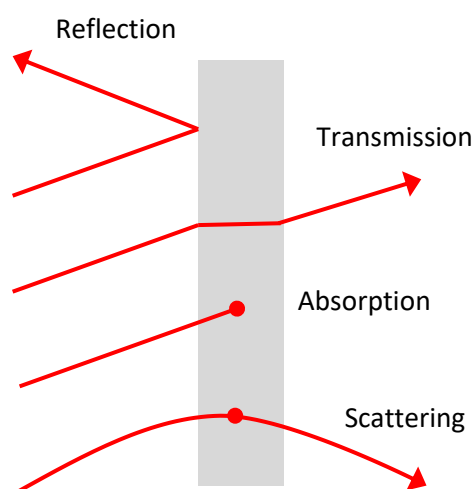


Figure 1.3. Diagram depicting the different ways in which light can interact with matter.

A free, quantum particle, *i.e.* an unbound electron, can have any energy. Whereas, a quantum mechanical system, *i.e.* a particle that is bound within part of an atom, ion or molecule, can only take on discrete energy values, known as energy levels. These energy levels consist of either rotational, vibrational or electronic states (Fayer, 2010). The interaction of light with matter may enable the molecule to transition between these distinct energy levels, when the appropriate amount of energy is given. There is a hierarchy between energy states, as shown in Figure 1.4; for

example, the energy difference between rotational levels (R) is less than that required for a transition between vibrational levels (V). Similarly, a much greater energy input is required transition between electronic energy levels *i.e.* to reach an excited electronic state (S_1) from the ground state (S_0) (Smith and Dent, 2005, Atkins and Friedman, 2011).

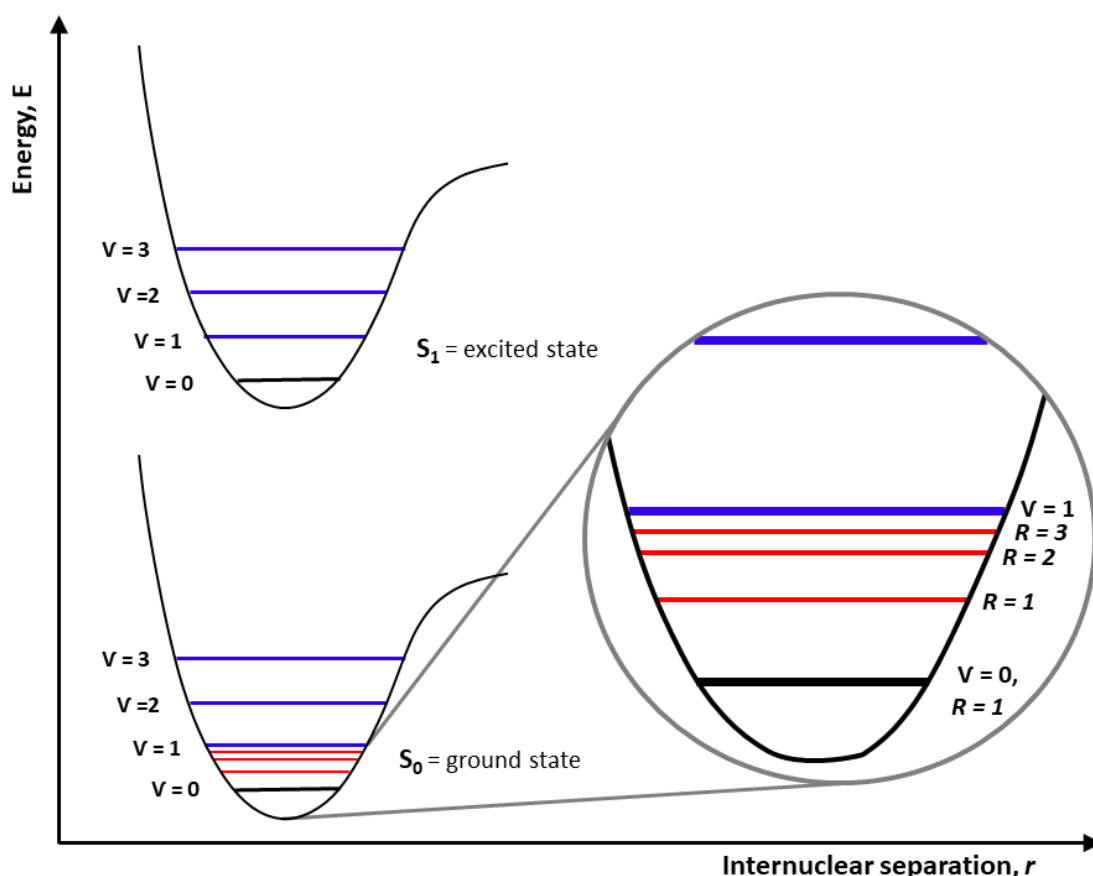


Figure 1.4. A depiction of a molecular energy level diagram for a hypothetical molecule. Rotational energy levels (R) and vibrational energy levels (V) are depicted within the electronic ground state (S_0) and first excited state (S_1).

1.3 Vibrational Spectroscopy

Vibrational spectroscopy is a collective term to describe infrared (IR) and Raman spectroscopy, which measure vibrational energy levels associated to chemical bonds within a molecule (Sathyanarayana, 2015). Both techniques are non-destructive and non-invasive, providing detailed information about the molecular structure and its interactions. IR spectroscopy measures the specific frequencies of polychromatic IR light which are absorbed by a molecule, resulting in vibrational transitions. For absorption to occur, the energy of an incident photon must match the energy difference between vibrational ground state ($V = 0$) and the vibrational energy levels ($V = 1, 2, \text{etc.}$), then the photon may be absorbed promoting it to the higher vibrational energy state. The

specific frequency of the IR light that is absorbed is characteristic of the functional groups that are present within the molecule, enabling functional group identification. IR frequencies are commonly divided into three spectral regions, these are near-, mid- and far-IR. Near-IR (NIR) ($14,000 - 4000 \text{ cm}^{-1}$) is the highest energy region which can excite overtone or harmonic vibrations, often used to study C-H, N-H or O-H stretching. Mid-IR ($4000-400 \text{ cm}^{-1}$), the most utilised region, enables insights into fundamental vibrations and associated rotational-vibrational structure. Lastly, far-IR ($400-10 \text{ cm}^{-1}$) is at the lowest energy and is used for rotational spectroscopy (Larkin, 2011).

By contrast, Raman spectroscopy uses monochromatic light to irradiate the sample, promoting the molecule to a virtual excited energy state. As the virtual excited energy state does not correspond to any formal molecular energy level, it is not necessary for the incident electromagnetic radiation to match the energy difference of the electronic ground and excited states. The scattered photons, upon relaxation, are detected in Raman spectroscopy. If there is a change of at least one vibrational unit of energy between the incident light and scattered photons, it is detected and relays information about the molecule (Smith and Dent, 2005). Figure 1.5 illustrates the different interactions of electromagnetic radiation within a molecule.

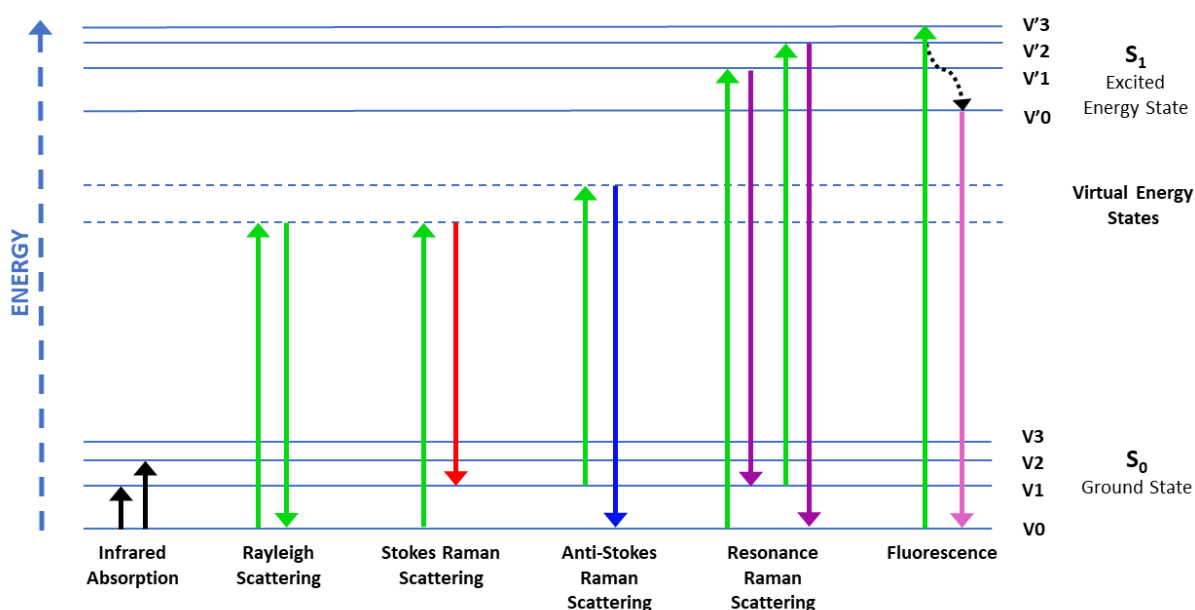


Figure 1.5. Jablonski diagram depicting the numerous mechanisms of electromagnetic radiation interacting with a molecule. Both infrared and fluorescence undergo absorption, whereas scattering processes can occur in either an elastic (Rayleigh, no change in energy) or inelastic manner (Raman scattering). The scattered photons may experience a reduction in energy of the incident light, known as Stokes Raman scattering. Alternatively, the molecule may lose energy, hence increasing the energy of the scattered photon, termed anti-Stokes Raman scattering. Fluorescence experiences non-radiative decay (black, dotted line), thus fluorescence emission is at a lower energy than the absorbed light.

There are two different types of scattering that can take place, these are elastic and inelastic scattering. Elastic scattering is the process whereby there is no change in the photon frequency, therefore the energy of the scattered light is identical to the incident beam ($\nu' = \nu$). This is the dominant process known as Rayleigh scattering, yet this phenomenon does not provide any information regarding the vibrational state of the molecule. Lord Rayleigh was the first to discover that light could be elastically scattered in 1871, naming the phenomenon after himself. A much smaller proportion of light experiences inelastic scattering (typically only 1 in 10 million photons which scatter are inelastically scattered) (Smith and Dent, 2005). In 1928, an Indian physicist, C.V. Raman, and collaborator, K. S. Krishnan, first discovered the phenomenon of inelastic scattering of light, known as the *Raman Effect* (Raman and Krishnan, 1928). The associated net change in the photon energy can either be due to the excitation or deactivation of molecular vibrations, this is called Raman scattering. The basic equation for Raman scattering is shown in Equation 1.3.

$$h\nu + E_i = h\nu' + E_f \quad \text{Equation 1.3.}$$

Where h is Planck's constant = 6.626×10^{-34} J.s, ν is the frequency of the incident photon, ν' the frequency of the scattered photon, E_i the initial energy state of the sample and E_f being the final energy state of the sample.

When the light interacts with a molecule, the electron cloud around the nucleus becomes polarised generating a short-lived state called the virtual state. This virtual state is very unstable, so the photon is instantly re-emitted (which is measured in Raman spectroscopy). If the interaction of the incident light leads to an increase in vibrational energy of the molecule from the photon, then the frequency of the scattered light from the incident beam will be lower in energy ($\nu' < \nu$), this is known as Stokes scattering. However, if the photon gains vibrational energy from the interaction with the molecule, then the scattered light will be higher in energy than the incident light ($\nu' > \nu$), known as anti-Stokes scattering (Figure 1.5). The intensity of Stokes scattering is greater than the anti-Stokes scattering as the populations of molecules in the vibrational ground state is always greater than in the vibrationally excited state, in accordance with the Boltzmann distribution law. Therefore, the Stokes shift is typically measured when performing Raman experiments (Wen, 2007).

Raman scattering may also undergo an additional phenomenon known as resonance, Figure 1.5, whereby a molecule experiences electronic absorption of UV/Vis excitation arising from chromophore bands within its structure, *i.e.* nucleic acids or aromatic amino acids. Excitation exceeds

that of the virtual state and is promoted to an excited energy state, S_1 , Figure 1.5. Thus, the scattered photon is at a greater energy than that of 'normal' Raman scattering, and the response is enhanced. Similarly, fluorescence also experiences excitation to the S_1 excited state in the visible region, however this is an absorbance – emission based system. Both resonance Raman scattering (RRS) and fluorescence can/will occur, with fluorescence often dominating the response and complicating, or preventing, the detection of the Raman response. This is a consequence of fluorescence occurring over a longer period of time due to its mode of emission experiencing non-radiative decay, whereas RRS is instantaneous and much shorter lived.

1.4 Comparisons of IR and Raman

IR and Raman spectroscopy may differ in the measurement of absorbed/scattered energy, yet they are considered complementary techniques. A vibrational transition that is active in Raman may not necessarily be IR active and *vice versa*, therefore these techniques can be used concurrently for the generation of a molecular 'fingerprint' (Ellis and Goodacre, 2006, Ellis *et al.*, 2007). To understand why a molecule may be IR or Raman active, there are basic selection rules. For a vibration to be Raman active there must be a change in polarisability (induced dipole moment) of the electron cloud around the molecule (Long, 2002). Symmetric vibrations cause the greatest change in polarisability and therefore are strong Raman scatterers. Electronegative functional groups, such as C-X (X = F, Cl, Br or I), C-NO₂ or C-S, exhibit large polarisability changes, consequently giving strong Raman signals (Dijkstra *et al.*, 2005). Alternatively, for IR active vibrations a change in dipole moment is required, hence asymmetric vibrations are the most intense (Colthup, 2012, Schrader, 2008). Selection rules are further reinforced by the mutual exclusion principle: if a molecule has a centre of symmetry then it will be Raman active but not IR active and *vice versa* (Smith and Dent, 2005). It is these selection rules that make IR and Raman spectroscopic techniques complementary as not all the vibrations present within a sample need to be/can be both IR and Raman active. Thus, the two techniques give different intensity patterns, but can be used simultaneously to provide more detailed information about the molecule(s) under investigation.

A molecule comprising of N number of atoms, will have $3N-6$ normal modes of vibration if it is non-linear (also called degrees of freedom), in contrast to a linear molecule where $3N-5$ normal modes are possible. For example, BF₃ a non-linear molecule has $(3 \times 4) - 6 = 6$ degrees of freedom, whereas linear CO₂ has $(3 \times 3) - 5 = 4$ degrees of freedom (Smith and Dent, 2005, Long, 2002). The vibrational modes of CO₂ are shown in Figure 1.6, the symmetric stretch (A) is IR inactive (no change in dipole

moment) but Raman active (distortion of the electron cloud). The opposite is true for the asymmetric stretch and degenerate bending modes, which are IR active yet Raman inactive.

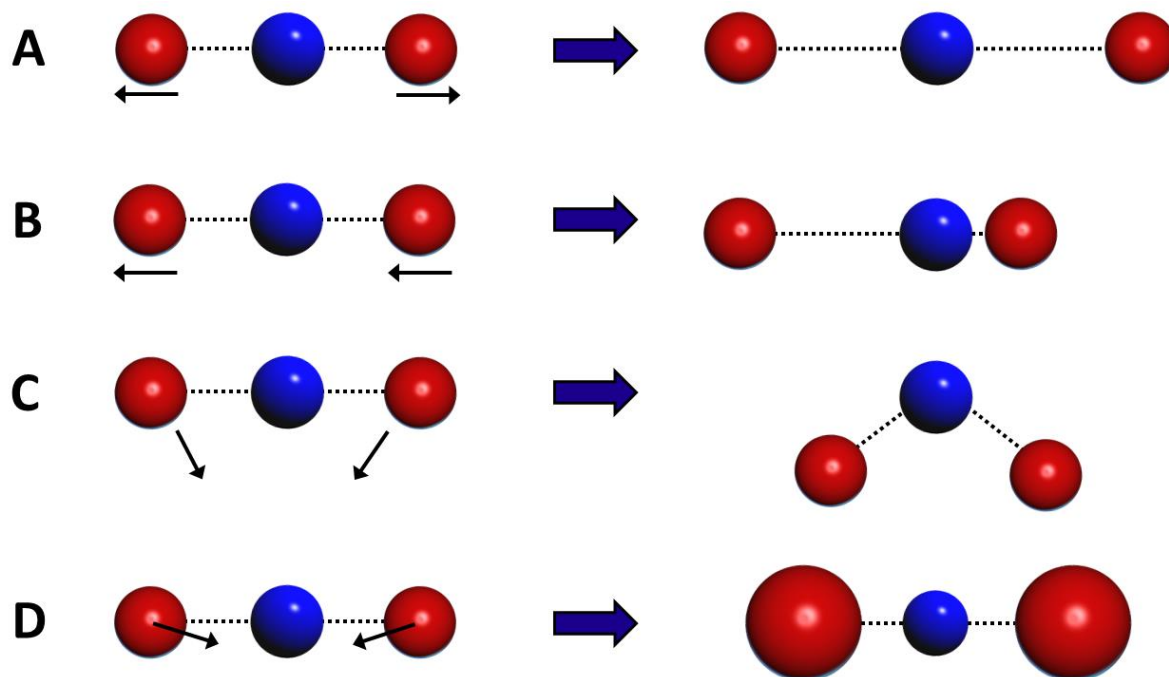


Figure 1.6. Normal vibrational modes of CO₂ **A)** symmetric stretch (1480 cm⁻¹), **B)** asymmetric stretch (2565 cm⁻¹), **C)** bending in-plane (526 cm⁻¹) **D)** bending out-of-plane (526 cm⁻¹).

Raman spectroscopy is less widely used in comparison to IR, predominantly a result of weak scattering signals, fluorescence interference and sample degradation. As mentioned before, Raman scattering is an inherently weak phenomenon as only 1 in 10⁶ – 10⁸ photons are inelastically scattered, significantly impacting the sensitivity of this technique. Water is a weak Raman scatterer, so it is a suitable medium for Raman investigations (*i.e.* biological studies which commonly use aqueous conditions). In contrast, water is a strong absorber of IR frequencies and often dominates an IR response, significantly reducing its compatibility towards aqueous environments. Fluorescence interference is common when using a radiation source in the visible region, as well as biological samples notoriously fluorescing. Adversely, fluorescence signals are much greater than that of Raman scattering, which may overwhelm the Raman signal (Figure 1.5). Infrared frequencies do not suffer fluorescence interference, thus this is not a consideration for IR spectroscopy.

1.5 Fluorescence

Fluorescence is a type of luminescence which occurs when the energy of the incident light is absorbed, resulting in photons of the singlet ground state to be promoted to a real excited energy state. Once the electronic excited state is reached, it is followed by partial dissipation yielding a relaxed singlet excited state, known as non-radiative decay. Finally, the photon will emit its energy, returning back to the ground state (Figure 1.5). The dissipation of energy causes the emitted photon to be at a lower energy and longer wavelength than the excitation photon (Stokes shift). Because of the non-radiative decay phase, fluorescence occurs over a longer time period ($\sim 10^{-9}$ s) than the Raman scattering process ($\sim 10^{-12}$ s) (Nair, 2006). It is the combination of these observations that leads to fluorescence interference: the emitted photons are at greater energy than those undergoing Raman scattering (as they achieve an excited state over a virtual energy transition), the photons experience non-radiative decay and consequently a change in energy, thus permitting their migration through Rayleigh filters into the detector, along with fluorescence occurring over a longer time frame, often dominating the Raman response. Furthermore, fluorescence is increasingly problematic when investigating biological or organic samples.

Consequently, investigations surrounding methods to quench fluorescence have been widely studied. Raman signal intensity is directly proportional to the fourth power of the inverse of the incident wavelength. Therefore, it is desirable to use a shorter wavelength to maximise the Raman scattering efficiency and increasing the number of detectable photons. However, the severity of fluorescence tends to increase with decreasing wavelengths, until deep-UV wavelengths are reached and fluorescence is less likely to occur (Asher and Johnson, 1984). Longer wavelengths, such as NIR, can help to avoid fluorescence interference, but this comes as a trade-off as the Raman intensity is significantly reduced. Alternatively, a Raman enhancement method, known as surface-enhanced Raman scattering (SERS) spectroscopy, actively quenches fluorescence arising from interactions with a nanoscale, roughened metal surface.

1.6 Enhancement of the Raman Response

The probability of Raman scattering events taking place is very low, thus emphasis is placed on increasing the likelihood of these events and increasing the Raman response. As we have just discussed, a logical approach may be to increase the laser power by using a shorter wavelength, however fluorescence interference subsequently increases, potentially affording no real improvement. Alternatively, enhancement techniques of 'normal' Raman spectroscopy can be used to increase the scattering efficiency. Several different enhancement methods have been

developed, including tip-enhanced Raman scattering (TERS) and coherent anti-Stokes Raman scattering (CARS). For the purpose of this thesis, two enhancement techniques are discussed in detail, these are surface-enhanced Raman scattering (SERS) and resonance Raman scattering (RRS).

1.6.1 Surface-Enhanced Raman Scattering (SERS)

Raman spectroscopy can often lack sensitivity for the analysis of some sample types, particularly when studying solutions. Surface-enhanced Raman scattering (SERS) spectroscopy is a technique in which the Raman scattering efficiency is enhanced by a factor of $\sim 10^4 - 10^6$ (where even single molecule detection has been reported (Kneipp *et al.*, 1997, Nie and Emory, 1997, Xu *et al.*, 1999)), as a result of molecules being in close contact with metal surfaces. The phenomenon was first observed in 1974 by Fleischmann, Hendra and McQuillan, who reported a notable enhancement in the Raman scattering response of pyridine when absorbed on the surface of a roughened silver electrode (Fleischmann *et al.*, 1974). At that present time, they hypothesised that the increased Raman response was a consequence of increased surface area arising from the roughened metal surface. Fast forward three years to 1977, two independent groups published successive findings of this enhancement effect, both proposing differing theories for the observed intensity, but neither attributing the effect to an increase in surface area. The first of these publications by Jeanmarie and Van Duyne (Jeanmaire and Van Duyne, 1977) proposed an electromagnetic effect, whereas Albrecht and Creighton (Albrecht and Creighton, 1977) hypothesised a charge-transfer effect. Despite these theories differing substantially by their mechanism of action, distinguishing between them experimentally has proven difficult over the decades, and consequently both are still considered today. These enhancement theories are described below.

1.6.1.1 Electromagnetic Enhancement

The first theory, electromagnetic enhancement, proposes that an interaction between the analyte and the plasmons of the metal surface results in an electromagnetic enhancement. The analyte absorbs onto or is in close proximity to the metal surface, whereby electrons covering the metal surface (plasmons) begin to oscillate when excited by the incident light. Surface plasmons have a resonance frequency at which they absorb and scatter light most efficiently; this is dependent on the metal and the nature of the surface. On application of the electromagnetic radiation, the plasmons are excited (referred to as localised surface plasmon resonances (LSPRs)) and result in the redistribution of the local field, generating enhancement of the electromagnetic field around

the nanoparticle (Figure 1.8) (Jeanmaire and Van Duyne, 1977, Moskovits, 1985). If Raman scattering is to occur, there must be an oscillation perpendicular to the surface plane, this is achieved by using a roughened metal surface or nanoparticles as they provide an area in which localised oscillations can occur (Jeanmaire and Van Duyne, 1977, Smith and Dent, 2005). 'Hot spots' is a term describing localised areas of high intensity local field enhancements. They are formed within the interstitial crevices of the roughened metallic nanoparticles, generating dramatic enhancements of the Raman response (enhancements of up to 10^{15} have been documented) (Kneipp *et al.*, 1997, Nie and Emory, 1997, Xu *et al.*, 1999). Generally, the degree of enhancement arising from 'hot spots' largely depends on the distance between nanoparticles, as well as their size and shape (Camden *et al.*, 2008, Kruszewski and Cyrankiewicz, 2012).

1.6.1.2 Chemical Enhancement (Charge-Transfer)

The second theory, chemical enhancement, proposes that the absorbate chemically bonds to the metal surface. Excitation is said to occur *via* the transfer of electrons from the metal surface to the chemisorbed substrate and back again, forming a charge-transfer complex. This theory is both site specific and dependent on the analyte interaction with metal (functional group specificity) (Albrecht and Creighton, 1977). Consequently, and after a decade long debate, it is generally agreed that the electromagnetic enhancement theory is the more dominant contributor towards most SERS processes, as it is less molecular dependent than the charge-transfer mechanism (Stiles *et al.*, 2008).

1.6.1.3 Parameter Optimisation in SERS

Many different factors can contribute towards SERS responses and signal enhancement. Therefore, numerous parameters need to be considered during the experimental design process, these variables are detailed below.

1.6.1.3.1 Choice of Metal

Silver, gold and copper contain plasmons (LSPRs) that oscillate within most of the visible and NIR wavelength range, thus they are extremely good substrates for use within SERS investigations (Figure 1.7) (Tian *et al.*, 2002, Smith and Dent, 2005). Silver substrates display a plasmon absorption maximum at ~ 400 nm, thus it has higher enhancement factors with visible frequencies. Whilst gold

has a maximum absorption of ~520 nm, which suffers a slightly lower enhancement factor in the visible region, yet it is most effective towards NIR frequencies (Kneipp *et al.*, 2006). Other metals, such as copper, are less widely used as the roughened surface can often form layers of oxide, altering the nature of the surface and hindering the absorption/interaction of the analyte under investigation (Smith and Dent, 2005). For a metal surface to be SERS active, there must be effective adsorption of the analyte onto the metal surface, which at times may be difficult to attain, however the mechanism actively quenches fluorescence (McNay *et al.*, 2011).

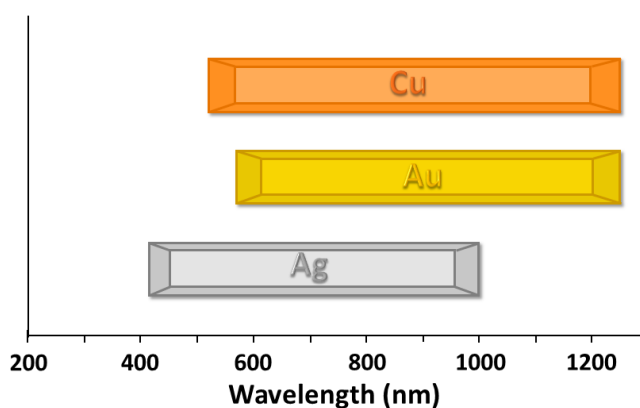


Figure 1.7. Approximate wavelength ranges of copper (Cu), gold (Au) and silver (Ag).

1.6.1.3.2 Type of Metal Substrate

There are two main classes of metal substrates used in SERS which exhibit nanoscale roughness, either a solid surface or a colloidal suspension of nanoparticles. Both types of metal substrates are required to incorporate nanoscale roughness (generally within the range of 5 – 100 nm) to enable SERS responses (Bordo and Rubahn, 2008). Solid state SERS combines a flat surface with a perpendicular roughened metal surface. These formulations are often expensive and difficult to produce in a reproducible manner, yet they have substantially less optimisation requirements.

On the other hand, colloidal suspensions are most commonly used due to their low cost, ease of preparation and disposal, stability, as well as providing strong enhancements for solution based assays (Figure 1.8) (Faulds *et al.*, 2004, Smith, 2008, Dougan *et al.*, 2011). There are several procedures that produce metal colloids, universally chemical reduction is used due to the simplicity of the method (Aroca *et al.*, 2005). Many different reducing agents can be used (such as sodium citrate and hydroxylamine hydrochloride) to reduce the metal ions to the metal and control nanoparticle size (Lee and Meisel, 1982, Leopold and Lendl, 2003). Previous studies have established that colloidal solutions permit the use of less expensive and portable instrumentation (evading microscope requirements) (Cowcher *et al.*, 2013, Mabbott *et al.*, 2013, Westley *et al.*,

2016). In addition, colloidal preparations present a truer indication of the analyte concentration. Suspended nanoparticles and the absorbed analyte(s) travel within the laser path under Brownian motion, therefore this generates an averaging effect rather than attempting to focus the laser on a fixed point manually, which would be required when using a roughened solid surface (Cowcher *et al.*, 2013). Furthermore, the implementation of colloidal substrates tends to reduce sample photodegradation, allowing higher laser powers to be used and increasing Raman scattering efficiency (Schlücker and Kiefer, 2011). Solid metallic surfaces have been shown to be advantageous as they can be ‘tuned’ to a particular wavelength by changing the thickness of the surface, nevertheless, the enhancement effect in comparison to colloidal substrates is usually far less.

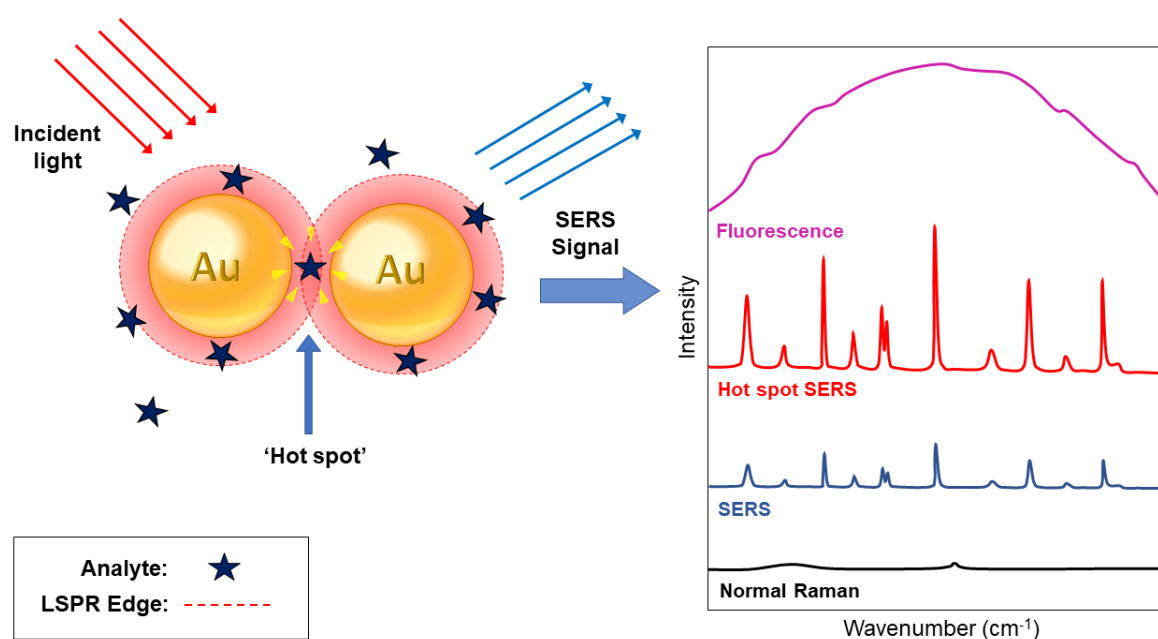


Figure 1.8. An image depicting a colloidal SERS response. The electromagnetic field projecting from the nanoparticle (*i.e.* Au) is represented by the localised surface plasmon response (LSPR) edge. The analytes interact with the roughened nanoparticle and its electric field, experiencing enhancement of the Raman response. On the right, one can see mock responses for the analyte in regular Raman, SERS, SERS experiencing hot spots (further enhancement) and finally fluorescence, which experiences the greatest intensity response.

1.6.1.3.3 Excitation Wavelength

The choice of laser wavelength can have a significant impact on the experimental capability for Raman spectroscopy, and the same is true for SERS. As previously discussed, wavelength selection is likely to depend on its capabilities of enhancing the Raman scattering yet minimising fluorescence interference. Metal nanoparticles have been shown to actively quench fluorescence through analyte-metal interactions (McNay *et al.*, 2011), thus increasing the flexibility in the excitation selection process. Visible wavelengths are typically used within SERS investigations (*i.e.* 488, 514.5,

532 and 633 nm) (Smith and Dent, 2005, Wen, 2007), yet examples in the literature include near UV (325 nm) (Ren *et al.*, 2003) and NIR frequencies (785, 830, 1064 and 1280 nm) (Premasiri *et al.*, 2012, Kneipp and Kneipp, 2006, Greeneltch *et al.*, 2012, Bedics *et al.*, 2015).

1.6.1.3.4 Factors Influencing Analyte Interaction/Adsorption with the Surface

A careful consideration when optimising SERS parameters is the analyte itself. Not all analytes will be SERS active arising from the functional groups present amongst its molecular structure, and their subsequent interaction with the metal surface. In general, polar, ionic and polarisable chemical species are favourable functionalities which are likely to give positive SERS responses, namely thiols (-SH), alcohols (-OH) and amines (-NH₂). Strategies to manipulate and modify the nanoparticle surface properties can be employed to promote favourable interactions. This may be achievable through the choice of the reducing agent, which influences the charge at the surface. For example, if the analyte and surface have the same charge, adsorption may be severely obstructed due to repulsion, which could cause the nanoparticles to aggregate and fall out of solution (colloidal preparation). To counteract charge repulsion, the analyte may be protonated or deprotonated by the addition of either acid or base to the SERS solution, promoting favourable interactions. Similarly, the addition of poly-L-lysine can be used as a method of charge neutralising the surface to promote favourable interactions and increase their stability (Marsich *et al.*, 2012).

1.6.1.3.5 Aggregating Agent

An aggregating agent may be used in conjunction with the colloidal suspension to cause different, irregular colloid particles to aggregate together. As a result, an increase in interactions between the colloid particles gives rise to larger surface plasmons generating surface enhancement (Yaffe and Blanch, 2008). Aggregation also increases the likelihood of 'hot spots' to form which can amplify the intensity of the Raman scattering, Figure 1.8 (Camden *et al.*, 2008, Kruszewski and Cyrankiewicz, 2012). On the other hand, if too much aggregating agent is used, then the colloidal nanoparticles begin precipitating out of solution resulting in weak/no SERS observations (Aroca *et al.*, 2005). Commonly used aggregating agents are sodium chloride, sodium nitrate and potassium chloride.

1.6.1.3.6 Aggregation Time

Finally, the time at which the SERS preparation is left to aggregate before being subjected to SERS analysis can have a significant impact on the reproducibility (Fisk *et al.*, 2016). This is usually the final parameter to optimise after all other parameters have been considered. Interpretation of

these results may identify a time frame in which replicate data show the greatest similarity in terms of Raman intensity, hence improving reproducibility between data sets.

1.6.1.4 Associated Advantages and Disadvantages of SERS

SERS spectroscopy offers a rapid method of identification that requires minimal sample preparation, accepts many sample types and has very high sensitivity and spatial resolution due to enhancements in scattering. Despite this, SERS does experience some shortcomings in comparison to normal Raman spectroscopy. Different colloid batches or solid surface preparations can often cause fluctuations on observed spectral features, as well as changes in the reproducibility of the data. Stability can also be a limiting factor; hydroxylamine-reduced silver colloid (HRSC) has been shown to be stable over a period of at least 4 months, yet other metal substrates have significantly reduced shelf-lives (Larmour *et al.*, 2012). Once the metal begins aggregating out of the colloidal solution, synthesis of a new colloid is required, again fluctuations are experienced within batch-to-batch synthesis. There are many factors that are influential on the size and distribution of the nanoparticles during their synthesis, for example the speed at which the solution is stirred, the rate of addition, pH, concentration and the order of addition (Kahraman *et al.*, 2007, Harrison *et al.*, 2008). Differing size and distribution of the nanoparticles can have inconsistent effects on the enhancement obtained with SERS, therefore the synthesis and experimental conditions must be carefully controlled to attain reproducible and consistent results.

Multiplex can be defined as “to enable a line to carry several signals simultaneously”. SERS spectroscopy is extremely valuable when studying a mixture of analytes as it allows the detection of multiple species, meaning that a mixture of analytes can be simultaneously detected without the need for separation (Nie and Emory, 1997, Alharbi *et al.*, 2015, Ngo *et al.*, 2014). Other spectroscopic methods can require isolation of individual analytes which is a time consuming and (often) difficult procedure, frequently resulting in loss of product. Sharp ‘fingerprint’ peaks are obtained in SERS due to their enhanced scattering efficiency, therefore, it is possible to assign multiple species within a single sample if the analytes have distinct spectra allowing spectral separation (Dougan and Faulds, 2012). This offers an improvement over the traditional Raman approach, particularly when studying biological samples as they are often in complex sample matrices.

In addition, SERS spectra are often difficult to assign; the analyte binding to/interacting with the metal surface can change any conventional symmetry which may have previously been seen in the

Raman spectra. As a consequence, an increase, as well a loss of some peaks is observed within the SERS spectral pattern (Smith and Dent, 2005).

1.6.2 Resonance Raman Scattering (RRS)

In 1972, Strekas and Spiro observed that chromophoric haems of haemoglobin and myoglobin resulted in an enhanced Raman spectrum, coining this observation resonance Raman (Spiro and Strekas, 1972). This method can enhance Raman scattering when the frequency of the incident laser coincides/lies close to the frequency of the electronic transition (*i.e.* the energy difference between the electronic ground state and electronic excited state), as shown in Figure 1.5. UV wavelengths are frequently used within resonance Raman studies as the UV frequency overlaps with an electronic transition of the resonant analyte under study - this technique is referred to as UV resonance Raman (UVR), whereby enhancements in scattering intensity are typically in the range of 10^2 - 10^6 (Asher, 1993). As the resonance Raman enhancement is several orders of magnitude greater than normal Raman scattering, much lower sample concentrations are required, frequently on the order of 1 mg/mL, in addition to shorter acquisition times (Couling *et al.*, 1998). Fluorescence is less likely to occur below 260 nm, therefore deep-UV wavelengths can be used and will not suffer from fluorescence, unlike all other commonly used Raman wavelengths mentioned earlier (Asher and Johnson, 1984).

A limitation of this method is the requirement of absorption of the laser light by a chromophore contained within the analyte under investigation, restricting its application towards structures that contain aromaticity or conjugation. On the other hand, RRS is very well-suited for studying complex biomolecules and biological materials. The resonant enhancement of the chromophore moieties permits the isolation of its corresponding Raman bands and structural units located in close proximity. This limits the amount of structural information that is generated and simplifies the analysis arising from enhancement of specific sites within the molecule, consequently, this may be considered advantageous when studying very large and complex biomaterials (Sathyanarayana, 2015). The requirement of a high-power UV laser may cause photochemical degradation of the analyte, so careful considerations must be taken as to avoid sample degradation, such as continuous stirring (liquid) or rotation (solid) of the sample. UVR instrumentation and lasers are generally very expensive to purchase and often require regular maintenance, which is likely to be a contributing reason to its limited use in the Raman community (Jarvis *et al.*, 2006).

Another resonant technique is surface-enhanced resonance Raman scattering (SERRS). SERRS is a technique further built on SERS: a metal surface is used to generate large scattering enhancements as beforementioned, however the analyte is chromophoric and lies close in energy to the laser

excitation wavelength (Stacy and Van Duyne, 1983). These resonant chromophores give a significant enhancement over SERS and have been shown to reach, or exceed, the sensitivity that is achievable with fluorescence (McNay *et al.*, 2011). The analyte must absorb onto the metal surface, as well as contain chromophores that are close to the laser frequency, otherwise SERRS is unattainable. Conversely, if an analyte does not possess these properties, SERRS active labels can be attached to enable SERRS responses. Labelling strategies present an advantage as a specific analyte within a biological matrix can be labelled, therefore the SERRS signal will exceed that of the non-specific SERS signals from other components. The capability of this approach within a multiplexed environment has been demonstrated within the literature, with six different SERRS labelled species identified in combination with chemometric analysis (Enright *et al.*, 2004, Laing *et al.*, 2016).

1.7 Chemometrics

Chemometrics is fundamentally the discipline of using mathematics to help improve the interpretation of data, which is aided by the application of computers. It can be defined as “*the application of mathematical and statistical methods to chemical measurements*” (Kowalski, 1980, Otto, 2016). The most straightforward approach to analysing Raman spectra is to use univariate methods, for example identifying characteristic peaks within a sample and then measuring the area under the peak. This information can be used as a qualitative and quantitative measure to make direct comparisons between samples. However, Raman spectra can often become difficult to interpret due to the multivariate nature of the data, as well as important changes in a sample often manifested in subtle modifications in their spectra (*i.e.* inconspicuous broadening of a peak) (Slutsky, 1998, Gouadec and Colombari, 2007). Therefore, simple univariate analysis may no longer be sufficient, consequently, the use of chemometric methods and algorithms have extensively assisted the interpretation, comparison and assignment of Raman spectra. Multivariate analysis methods can be used to show relationships between observed spectral changes and variables, such as concentration or binding studies (López-Díez and Goodacre, 2004, Ashton *et al.*, 2011).

1.7.1 Data Pre-Processing

When conducting Raman investigations, the spectral data set may experience physical distortions due to fluctuations associated with the instrumentation (*i.e.* spectral resolution or laser power) or the nature of the samples (*i.e.* fluorescence). These distortions cause nonlinearities to the Raman spectra model, so it may be appropriate to apply pre-processing to enable multivariate analysis to

perform effectively (Lasch, 2012). Data pre-processing is a method proven to resolve uncontrollable events that lead to errors in the data by transforming the raw data into an understandable format, removing these distortions so that true changes in the data set can be analysed. Common pre-processing methods include baseline correction (minimising fluorescence/background interference), normalisation (correcting for intensity differences between samples) or smoothing (to reduce noise) of the data set (Butler *et al.*, 2016).

A combination of pre-processing techniques may be applied to the data which can significantly affect the results of the multivariate analysis. Similarly, the order in which the pre-processing is applied can equally affect the overall outcome, therefore careful considerations must be taken as to avoid bias in the subsequent results (Brewster *et al.*, 2011).

1.7.2 Principal Component Analysis (PCA)

There are two main chemometric classification methods: supervised and unsupervised. Unsupervised classifications do not need *a priori* information about the samples being studied and results are generated based on the Raman data alone. Principal component analysis (PCA) is a popular unsupervised method designed to reduce complex data sets into lower dimensional data, often revealing hidden relationships (Wold *et al.*, 1987, Abdi and Williams, 2010).

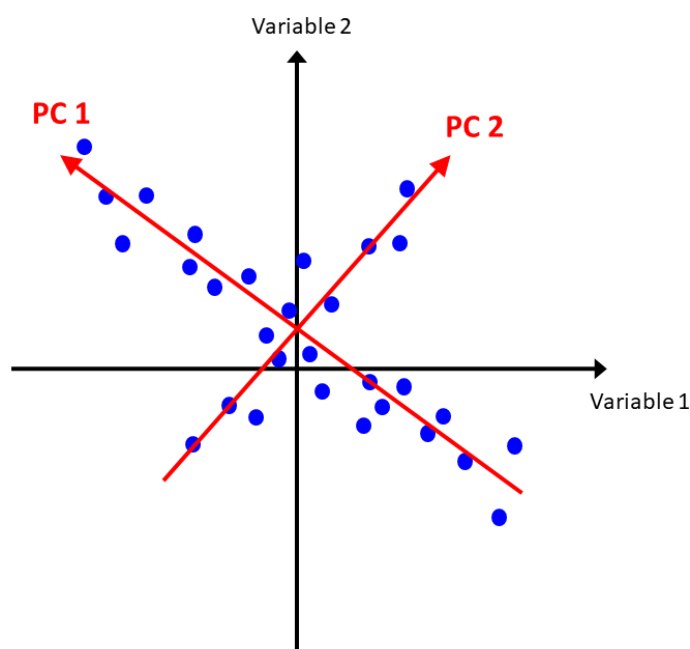


Figure 1.9. Illustration of an original data set represented by variables 1 and 2 (axis). The PCA model identifies suitable PCs (1 and 2) to uncover hidden relationships within the dataset.

This is achieved by the PCA algorithm extracting the important information from the data set and expressing this information as new orthogonal variables, referred to as principal components (PCs). The PCs identify directions along which the variation is greatest, whereby plots of these results enable visual interpretation of the variance within the data set, Figure 1.9. If one observes tight clustering between data points in the PCA, this indicates that the spectra are highly similar. Conversely, trends or outliers (an observation point that lies an abnormal distance from all other observations within the population) may be identifiable based on separation along the PC axes, for example the conversion of a starting material to its corresponding product may separate within the PCA due to spectral differences (Jolliffe, 1986).

1.7.3 Multivariate Curve Resolution-Alternating Least Squares

Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) is a soft-modelling method, and is used mathematically to deconvolve a spectroscopic response into the pure contributions of individual components present within a mixture (Garrido *et al.*, 2008, Azzouz and Tauler, 2008). The input of pure spectra (prior information) of each component of the mixture is an initial requirement. The iterative method works until the best solution is generated, acting to deconvolve the complex mixture and provide concentration profiles of each individual component. This multivariate model is particularly useful for the analysis of complex, biological samples such as an enzyme-catalysed reaction.

1.8 Applications of Raman Spectroscopy

Raman spectroscopy is an extremely useful vibrational technique allowing quantitative and qualitative analysis. Since its discovery in 1928 (Raman and Krishnan, 1928), it has enabled advances in many scientific fields including applications within polymer analysis, forensic science, geosciences, materials science and pharmaceuticals. Raman spectroscopy has been used to follow chemical and biochemical reactions, which has been made possible by the coupling of microscopy with Raman spectrometers. However, the Raman effect is inherently weak thus collection times can be slow. Methods which enhance signals, such as UVRR and SE(R)RS as introduced earlier, are continually being developed for high-throughput bioanalysis. Moore *et al.* successfully demonstrated that the activity of hydrolases could be monitored at ultra-low levels using SERRS reporter molecules (Moore *et al.*, 2004). Within this study, enzymatic substrates were designed with reporter groups that initially 'masked' SERRS signals, yet after biocatalytic hydrolysis, the

substrate released a surface targeting dye. This dye behaved as a reporter molecule, which bound strongly to the metal nanoparticles, generating SERRS signals and hence indicating enzyme activity. This method was shown to be applicable to a range of different hydrolases, including lipases, esterases and proteases. More recently, Westley *et al.* developed a SERS bioassay to monitor the enzyme-catalysed conversion of hypoxanthine to xanthine to uric acid using xanthine oxidase (Westley *et al.*, 2016). This research identified that SERS could directly quantify each component within the biotransformation, using a label-free and rapid method, achieving reproducible results that were in excellent agreement with off-line HPLC analysis. These investigations highlight the huge potential of Raman spectroscopic techniques as rapid methods for screening biochemical reactions.

1.9 Biocatalysts: Their Importance and Applications

Over recent years, interest and the development of biocatalysts has significantly grown due to their wide-ranging applications within industry, their practicality and low environmental impact (Schmid *et al.*, 2001, Leresche and Meyer, 2006, Turner and Truppo, 2013, Reetz, 2013, Patel, 2017). Biocatalysis can be defined as 'the use of natural substances, enzymes or microbes, as catalysts to carry out chemical modifications in synthetic chemistry'. Humans have been exploiting enzymes to perform chemical modifications for thousands of years, the fermentation of grapes into wine is an example dating from 4100 BC (Fernandes, 2010). A landmark discovery was made in 1897 by Eduard Buchner who reported the first incidence of using cell free yeast extracts for sugar fermentation. This discovery was irrefutable evidence that a biotransformation does not necessarily need a living cell for the reaction to occur (Buchner, 1897). More recently in 1950, pharmaceutical companies (Upjohn, Schering, Merck and Pfizer) unearthed another fundamental finding that microbes catalysed the oxidative hydroxylation of steroids in a regioselective and stereoselective manner, demonstrating their stringent control capabilities (Reetz, 2013). By the 1990s many enzymes were commercialised (*i.e.* biological detergents), and were increasingly incorporated within industrial processes such as the production of fine chemicals and chiral compounds.

At present, biocatalysts are mainly used within technical applications in industry, such as detergents and textiles. Other growing applications include the food and feed industry, production of fine chemicals within pharmaceutical and agrochemical industries, as well generating biofuels and bioplastics (Illanes *et al.*, 2013). Despite increasing interest and their rising achievements, biocatalysts are met with several long-standing limitations, as described on the following page:

- Enzymes have high substrate specificity, therefore a narrow substrate scope
- There are a finite number of enzymes that exist, limiting the types of chemical transformations
- Their availability and accessibility may be restricted
- Enzymatic products can often bind to the enzyme resulting in product inhibition
- Protein stability is often poor when compared to traditional chemocatalysts, and
- Proteins often require complex and costly co-substrates, such as co-factors (Bommarius and Riebel-Bommarius, 2004, Choi *et al.*, 2015).

Despite enzymes having narrow substrate specificity which is often considered as a disadvantage, their high stereo-, enantio- and regio-selectivity is extremely advantageous (Patel, 2011, Patel, 2017, Truppo, 2017). Some enzymes can successfully distinguish between the *R*- and *S*-enantiomers (L- and D-) and diastereomers, with many exhibiting >99 % enantiomeric excess (*ee*). The ability to distinguish between enantiomers is extremely important within pharmaceutical synthesis. Often, one enantiomer of a drug molecule can have the desired effect, whereas the other form may cause undesirable and harmful side effects. The most notable example being thalidomide, prescribed as a racemic mixture in 1957 to expecting mothers to alleviate morning sickness. The *R*-enantiomer was effective in nausea treatment, whereas the *S*-enantiomer is teratogenic, which resulted in birth defects.

Biocatalysts commonly operate at mild conditions with respect to temperature (frequently 20 – 40 °C), pH (neutral) and within aqueous media. In comparison, chemocatalysts regularly require harsh reaction conditions (high temperature and pressure), are less selective and experience poor atom economy by producing unwanted side products, thus often need protection/deprotection steps during synthesis. Furthermore, enzymes are biodegradable and are sourced from sustainable resources. This enables the use of greener chemistry and sustainable processes which have reduced environmental impact. In addition, proteins have high catalytic efficiency, therefore very small quantities of biocatalyst are required, further reducing costs (Wohlgemuth, 2010).

In spite of these advantages, in many cases the naturally occurring enzymes are insufficient at performing the industrial application of interest, which is unsurprising as they have evolved in nature for a specific function. However, their properties may be further tailored through the application of recombinant DNA methodology and directed enzyme evolution. There are three main approaches to protein engineering: (1) directed evolution (DE) is a method which mimics natural selection by making random mutations to the gene of interest and then assessing the results of such modifications. An advantage of DE is that there is no requirement of previous structural

information (*i.e.* crystal structure) and that mutations may be introduced at sites distant from the active site which has proven to bring about desirable traits (Saen-Oon *et al.*, 2008, Jiménez-Osés *et al.*, 2014). The combination of DE with high-throughput screening (HTS) techniques enable the rapid screening of vast gene libraries and subsequent selection of functionally improved mutants. These improved variants may be subjected to further rounds of mutagenesis, or selected as the end gene if the desired traits are reached (Figure 1.10). (2) Rational design, in which site-specific changes are made in accordance with previous knowledge about the protein structure, function and mechanism. An advantage of rational design is that the probability of beneficial mutations is increased and a substantial reduction in the number of variants significantly decreases the time and effort which is placed on screening. Lastly, (3) is the bioprospecting approach, which is the search for novel enzymes with improved performance that are isolated from living organisms often from unusual habitats (Arnold, 1998, Zhao *et al.*, 2002, Sheryl and Huimin, 2006, Bornscheuer *et al.*, 2012).

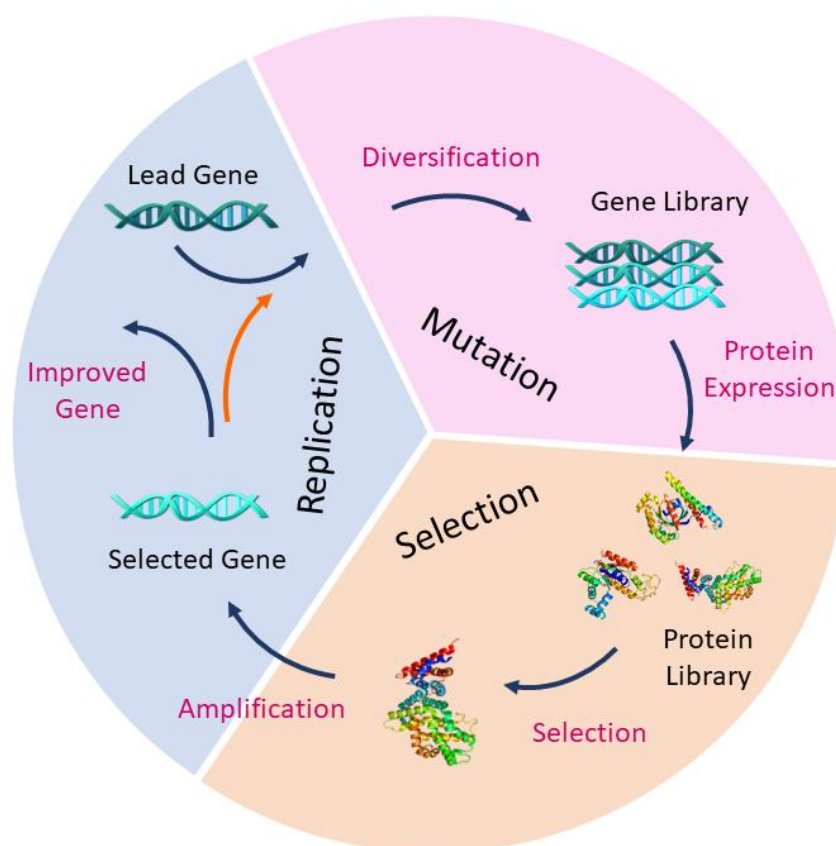


Figure 1.10. Overview of directed evolution (DE), a mutation strategy for the development of biocatalysts. Beginning with the lead gene, it experiences diversification by mutations in the gene sequence, generating many different variants (gene library). The best variants are selected, through the application of screening techniques, which are then subjected to iterative rounds of subsequent mutagenesis, or the process is stopped as the desired traits are achieved.

1.10 High-Throughput Screening (HTS) Methods

DE approaches are widely used throughout industry and academia, often favoured over rational design and bioprospecting methods. Major advancements in DE mutagenic strategies, for example error-prone polymerase chain reaction (ep-PCR) and gene shuffling, has led to significant reductions in the time needed to produce such large genetic libraries. However, the success of any DE process is reliant on the capabilities of the HTS method to rapidly screen and select the favourable variants. Despite there being a variety of well-established HTS methods capable of assessing biocatalytic processes, this step is the bottleneck of all DE programs. Their relative speed of analysis compared to the generation of the variants is hampering the overall effectiveness of the process, and ultimately the widespread application of biocatalysts within industrial synthesis (Leresche and Meyer, 2006, Whittall and Sutton, 2009, Reetz, 2013, Guazzaroni *et al.*, 2015).

HTS is an approach towards both biocatalyst and drug discovery that has gained increasing popularity over the past few decades and is now a standard technique for discovery investigations in the pharmaceutical industry. To summarise HTS, it is the process of screening and analysing a large number of experiments within a short period of time. Parallel tests may be performed on a small scale to achieve the desired speed to be considered as high-throughput (Szymański *et al.*, 2012). HTS methods can screen up to 10,000 compounds per day, with ultra-high-throughput screening (UHTS) techniques conducting in the region of 100,000 analyses per day (Martis *et al.*, 2011, Szymański *et al.*, 2012). HTS assays are used by both industrial scientists and academic researchers, with their widespread use facilitated by reductions in equipment costs. Several different types of libraries are screened, including many of the "omics" disciplines, such as genomics and proteomics, combinatorial chemistry, as well as protein and peptide libraries. HTS approaches often utilise robotics, detectors and sophisticated software that help to regulate and speed up the overall process. The main goal is to be able to identify active compounds ("hits") rapidly and accurately from vast volumes of chemical libraries.

Over the last three decades, industry has experienced a three-fold increase in the spending of research and development sectors for the ongoing search of 'blockbuster' drugs (Service, 2004, Liu *et al.*, 2014). Drug discovery is a time-consuming, expensive and inefficient process; therefore, the requirement of HTS methods that are fast, reliable and sensitive is essential. Commonly used HTS techniques include fluorometric and colorimetric assays, nuclear magnetic resonance (NMR) spectroscopy, liquid chromatography-mass spectrometry (LC-MS) and high performance liquid chromatography (HPLC) (Bornscheuer, 2013, Grunwald, 2014). A brief description of these methods, along with their associated advantages and disadvantages are in Table 1.1.

Table 1.1. Common HTS techniques used within industry and academia, with a brief description about each technique and its associated advantages and disadvantages (Kumar and Clark, 2006).

Assay	Principle of Method	Advantages (+) and Disadvantages (-)
Fluorometric Assays	Optical-readout technique which relies on a non-fluorogenic substrate being converted to a fluorescent product.	<ul style="list-style-type: none"> + Simple to use with rapid detection + Easily automated + Does not require purification - Limited application due to the requirement of a fluorogenic substrate
Colorimetric Assays	An assay which uses reagents that undergo a measurable colour change in response to an analyte.	<ul style="list-style-type: none"> + Simple to use with rapid detection + Easily automated - Sample recovery is restricted - Limited scope due to essential colorimetric reagent-analyte interaction
Nuclear magnetic resonance (NMR) Spectroscopy	Technique that exploits the interaction of electromagnetic radiation with the atomic nuclei. The intramolecular field around an atom in a compound provides detailed information about its structure, chemical environment and reaction state.	<ul style="list-style-type: none"> + Non-destructive + Flexible + Detailed and quantitative information - Low sensitivity - Purification often necessary - Expensive equipment
High-performance liquid chromatography (HPLC)	A method which physically separates a mixture into its individual components due to their polarity. Relies on pressurised solvents (mobile phase) containing the sample mixture to be passed through a column (stationary phase), with differences in components affecting their flow rates.	<ul style="list-style-type: none"> + Sensitive and selective + Good reproducibility + Automated - Little chemical information obtained - Considerable sample preparation - Expensive equipment - Commonly, large volumes of organic solvents required (excluding UPLC)
Mass spectrometry (MS)	A technique which analyses chemical species by initially ionizing them, then these ions are sorted by their mass-to-charge ratio which are subsequently detected. Fragmentation patterns enable further confirmation of the molecular structure.	<ul style="list-style-type: none"> + Highly sensitive and selective + High resolution + Good reproducibility + Automated - Considerable sample preparation - Expensive equipment - Internal standards necessary for quantification

The screening methods mentioned in Table 1.1 all have differing strengths and weaknesses, selecting an appropriate HTS method would vary depending on the type of library/reaction that was under study. With regards to screening biocatalytic activity, HPLC is often a 'go-to' method for quick identification of new product peaks. HPLC experiences high sensitivity, so only small sample volumes and concentrations are necessary, however this technique relays minimal structural information. Thus, it cannot provide enough evidence alone to identify if the desired product has been formed. Combining LC separation with MS (LC-MS) is advantageous as it can separate the mixture into its individual components, ensued by detection of their molecular weights and fragmentation patterns, enabling confirmation of the detected products molecular weight. Its shortcomings are experienced when similar masses are present within the same sample (*i.e.* regio- or stereo-isomers) or if the sample suffers from poor ionisation. Fluorometric and colorimetric based assays are popular HTS methods as they are robust, quick to detect 'hits' and are convenient to use. However, their application is very limited due to the requirement of a fluorescent/colour producing/activating product. Finally, NMR can generate a vast amount of detailed information about the compound under study and is quantitative, however the amount of sample required for analysis is much greater than other HTS methods, a consequence of its low sensitivity.

In summary, current screening methods used in academia and industry are not capable of rapidly screening large libraries in a universal manner. There are shortfalls experienced by the most popular methods (Table 1.1), consequently failing to meet the demands of experiments which generate large numbers of variants within a short period of time, such as in DE processes. Preferably, a method which experiences high sensitivity and selectivity, low sample volumes/concentrations and is robust and reproducible would greatly improve current screening capabilities. If it could generate molecular specific information (akin to NMR) rapidly and avoid the requirement of a specific reporter molecule (*i.e.* fluorophore or coloured reagent), it would greatly benefit discovery projects. In addition, if it were compatible with automation and sample miniaturization, this would further increase the speed of analysis and its overall utility.

1.11 Aim of the Thesis

Within this thesis, the overall aim is to introduce and discuss Raman spectroscopy as an alternative screening method within biocatalytic applications. Emphasis is placed on its enhancement techniques, mainly SERS and UVR, as described in the previous section. Their suitability is demonstrated throughout, showcasing fast acquisition times, information rich analysis and reproducible, quantitative measurements.

The main points I wish to address throughout this thesis are:

- The development of both SERS and UVR to monitor enzyme-catalysed conversions of substrate to corresponding product(s).
- The use of chemometric analysis, such as MCR-ALS, to enable quantitative measurements about the reaction components.
- Explore Raman enhancement methods to monitor enzyme reactions in real-time, as opposed to quenched, off-line analysis.
- Perform biotransformations using both *in vitro* (purified enzyme) and *in vivo* (bacterial whole-cells) preparations.
- Investigate the sensitivity and selectivity of these approaches, such as regioisomer identification and enantiomer discrimination.

1.12 References

- ABDI, H. & WILLIAMS, L. J. 2010. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2, 433-459.
- ALBRECHT, M. G. & CREIGHTON, J. A. 1977. Anomalously intense Raman spectra of pyridine at a silver electrode. *Journal of the American Chemical Society*, 99, 5215-5217.
- ALHARBI, O., XU, Y. & GOODACRE, R. 2015. Simultaneous multiplexed quantification of caffeine and its major metabolites theobromine and paraxanthine using surface-enhanced Raman scattering. *Analytical and Bioanalytical Chemistry*, 407, 8253-8261.
- ARNOLD, F. H. 1998. Design by Directed Evolution. *Accounts of Chemical Research*, 31, 125-131.
- AROCA, R. F., ALVAREZ-PUEBLA, R. A., PIECZONKA, N., SANCHEZ-CORTEZ, S. & GARCIA-RAMOS, J. V. 2005. Surface-enhanced Raman scattering on colloidal nanostructures. *Advances in Colloid and Interface Science*, 116, 45-61.
- ASHER, S. & JOHNSON, C. 1984. Raman spectroscopy of a coal liquid shows that fluorescence interference is minimized with ultraviolet excitation. *Science*, 225, 311-313.
- ASHER, S. A. 1993. UV resonance Raman spectroscopy for analytical, physical, and biophysical chemistry. Part 1. *Analytical Chemistry*, 65, 59A-66A.
- ASHTON, L., LAU, K., WINDER, C. L. & GOODACRE, R. 2011. Raman spectroscopy: lighting up the future of microbial identification. *Future Microbiology*, 6, 991-997.
- ATKINS, P. W. & DE PAULA, J. 2006. *Atkins' Physical Chemistry*, Oxford, Oxford University Press.
- ATKINS, P. W. & FRIEDMAN, R. S. 2011. *Molecular quantum mechanics*, Oxford university press.
- AZZOUZ, T. & TAULER, R. 2008. Application of multivariate curve resolution alternating least squares (MCR-ALS) to the quantitative analysis of pharmaceutical and agricultural samples. *Talanta*, 74, 1201-1210.
- BEDICS, M. A., KEARNS, H., COX, J. M., MABBOTT, S., ALI, F., SHAND, N. C., FAULDS, K., BENEDICT, J. B., GRAHAM, D. & DETTY, M. R. 2015. Extreme red shifted SERS nanotags. *Chemical Science*, 6, 2302-2306.
- BOMMARIUS, A. S. & RIEBEL-BOMMARIUS, B. R. 2004. *Biocatalysis: Fundamentals and Applications*, Wiley.
- BORDO, V. G. & RUBAHN, H.-G. 2008. *Optics and spectroscopy at surfaces and interfaces*, John Wiley & Sons.
- BORNSCHEUER, U. T. 2013. Protein Engineering as a Tool for the Development of Novel Bioproduction Systems. In: ZENG, A.-P. (ed.) *Fundamentals and Application of New Bioproduction Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- BORNSCHEUER, U. T., HUISMAN, G. W., KAZLAUSKAS, R. J., LUTZ, S., MOORE, J. C. & ROBINS, K. 2012. Engineering the third wave of biocatalysis. *Nature*, 485, 185-194.
- BREWSTER, V. L., ASHTON, L. & GOODACRE, R. 2011. Monitoring the Glycosylation Status of Proteins Using Raman Spectroscopy. *Analytical Chemistry*, 83, 6074-6081.
- BROGLIE, L. D. 1924. XXXV. A tentative theory of light quanta. *Philosophical Magazine*, 47, 446-458.
- BUCHNER, E. 1897. Alkoholische Gärung ohne Hefezellen. *Berichte der deutschen chemischen Gesellschaft*, 30, 1110-1113.
- BUTLER, H. J., ASHTON, L., BIRD, B., CINQUE, G., CURTIS, K., DORNEY, J., ESMONDE-WHITE, K., FULLWOOD, N. J., GARDNER, B., MARTIN-HIRSCH, P. L., WALSH, M. J., MCAINSH, M. R., STONE, N. & MARTIN, F. L. 2016. Using Raman spectroscopy to characterize biological materials. *Nature Protocols*, 11, 664-687.

- CAMDEN, J. P., DIERINGER, J. A., WANG, Y., MASIELLO, D. J., MARKS, L. D., SCHATZ, G. C. & VAN DUYN, R. P. 2008. Probing the Structure of Single-Molecule Surface-Enhanced Raman Scattering Hot Spots. *Journal of the American Chemical Society*, 130, 12616-12617.
- CHOI, J.-M., HAN, S.-S. & KIM, H.-S. 2015. Industrial applications of enzyme biocatalysis: Current status and future aspects. *Biotechnology Advances*, 33, 1443-1454.
- COLTHUP, N. 2012. *Introduction to Infrared and Raman Spectroscopy*, Elsevier Science.
- COULING, V. W., FISCHER, P., KLENERMAN, D. & HUBER, W. 1998. Ultraviolet Resonance Raman Study of Drug Binding in Dihydrofolate Reductase, Gyrase, and Catechol O-Methyltransferase. *Biophysical Journal*, 75, 1097-1106.
- COWCHER, D. P., XU, Y. & GOODACRE, R. 2013. Portable, Quantitative Detection of Bacillus Bacterial Spores Using Surface-Enhanced Raman Scattering. *Analytical Chemistry*, 85, 3297-3302.
- DIJKSTRA, R. J., ARIESE, F., GOOIJER, C. & BRINKMAN, U. A. T. 2005. Raman spectroscopy as a detection method for liquid-separation techniques. *Trends in Analytical Chemistry*, 24, 304-323.
- DOUGAN, J. A. & FAULDS, K. 2012. Surface enhanced Raman scattering for multiplexed detection. *Analyst*, 137, 545-554.
- DOUGAN, J. A., MACRAE, D., GRAHAM, D. & FAULDS, K. 2011. DNA detection using enzymatic signal production and SERS. *Chemical Communications*, 47, 4649-4651.
- EINSTEIN, A. 1905. Zur Elektrodynamik bewegter Körper. *Annalen der Physik*, 322, 891-921.
- EINSTEIN, A. & INFELD, L. 1966. *Evolution of Physics*, Touchstone.
- ELLIS, D. I., DUNN, W. D., GRIFFIN, J. L., ALLWOOD, W. J. & GOODACRE, R. 2007. Metabolic fingerprinting as a diagnostic tool. *Pharmacogenomics*, 8, 1243-1266.
- ELLIS, D. I. & GOODACRE, R. 2006. Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. *Analyst*, 131, 875-885.
- ENRIGHT, A., FRUK, L., GRONDIN, A., MCHUGH, C. J., SMITH, W. E. & GRAHAM, D. 2004. SERRS dyes Part 3. Synthesis of reactive benzotriazole azo dyes for surface enhanced resonance Raman scattering. *Analyst*, 129, 975-978.
- FAULDS, K., LITTLEFORD, R. E., GRAHAM, D., DENT, G. & SMITH, W. E. 2004. Comparison of Surface-Enhanced Resonance Raman Scattering from Unaggregated and Aggregated Nanoparticles. *Analytical Chemistry*, 76, 592-598.
- FAYER, M. D. 2010. *Absolutely Small: How Quantum Theory Explains Our Everyday World*, AMACOM.
- FERNANDES, P. 2010. Enzymes in Food Processing: A Condensed Overview on Strategies for Better Biocatalysts. *Enzyme Research*, 2010, 19.
- FISK, H., WESTLEY, C., TURNER, N. J. & GOODACRE, R. 2016. Achieving optimal SERS through enhanced experimental design. *Journal of Raman Spectroscopy*, 47, 59-66.
- FLEISCHMANN, M., HENDRA, P. J. & MCQUILLAN, A. J. 1974. Raman spectra of pyridine adsorbed at a silver electrode. *Chemical Physics Letters*, 26, 163-166.
- GARRIDO, M., RIUS, F. X. & LARRECHI, M. S. 2008. Multivariate curve resolution-alternating least squares (MCR-ALS) applied to spectroscopic data from monitoring chemical reactions processes. *Analytical and Bioanalytical Chemistry*, 390, 2059-2066.
- GOUADEC, G. & COLOMBAN, P. 2007. Raman Spectroscopy of nanomaterials: How spectra relate to disorder, particle size and mechanical properties. *Progress in Crystal Growth and Characterization of Materials*, 53, 1-56.
- GREENELTCH, N. G., DAVIS, A. S., VALLEY, N. A., CASADIO, F., SCHATZ, G. C., VAN DUYN, R. P. & SHAH, N. C. 2012. Near-Infrared Surface-Enhanced Raman Spectroscopy (NIR-SERS) for the Identification of Eosin Y: Theoretical Calculations and Evaluation of Two Different Nanoplasmonic Substrates. *The Journal of Physical Chemistry A*, 116, 11863-11869.

- GRUNWALD, P. 2014. *Industrial Biocatalysis*, Pan Stanford Publishing.
- GUAZZARONI, M.-E., SILVA-ROCHA, R. & WARD, R. J. 2015. Synthetic biology approaches to improve biocatalyst identification in metagenomic library screening. *Microbial Biotechnology*, 8, 52-64.
- GUPTA, M. C. 2007. *Atomic And Molecular Spectroscopy*, New Age International (P).
- HARRISON, R. G., WASHBURN, A. L., PICKETT, A. T. & CALL, D. M. 2008. Assembly of CdSe nanoparticles into microspheres by a liquid droplet emulsion process. *Journal of Materials Chemistry*, 18, 3718-3722.
- ILLANES, A., WILSON, L. & VERA, C. 2013. *Problem Solving in Enzyme Biocatalysis*, Wiley.
- JARVIS, R. M., BROOKER, A. & GOODACRE, R. 2006. Surface-enhanced Raman scattering for the rapid discrimination of bacteria. *Faraday discussions*, 132, 281-292.
- JEANMAIRE, D. L. & VAN DUYN, R. P. 1977. Surface raman spectroelectrochemistry. *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry*, 84, 1-20.
- JIMÉNEZ-OSÉS, G., OSUNA, S., GAO, X., SAWAYA, M. R., GILSON, L., COLLIER, S. J., HUISMAN, G. W., YEATES, T. O., TANG, Y. & HOUK, K. N. 2014. The Role of Distant Mutations and Allosteric Regulation on LovD Active Site Dynamics. *Nature Chemical Biology*, 10, 431-436.
- JOLLIFFE, I. T. 1986. Principal Component Analysis and Factor Analysis. *Principal Component Analysis*. Springer.
- KAHRAMAN, M., YAZICI, M. M., SLAHIN, F., BAYRAK, Ö. F. & ÇULHA, M. 2007. Reproducible Surface-Enhanced Raman Scattering Spectra of Bacteria on Aggregated Silver Nanoparticles. *Applied Spectroscopy*, 61, 479-485.
- KNEIPP, K. & KNEIPP, H. 2006. Surface-Enhanced Raman Scattering on Silver Nanoparticles in Different Aggregation Stages. *Israel Journal of Chemistry*, 46, 299-305.
- KNEIPP, K., MOSKOVITS, M. & KNEIPP, H. 2006. *Surface-Enhanced Raman Scattering: Physics and Applications*, Springer Berlin Heidelberg.
- KNEIPP, K., WANG, Y., KNEIPP, H., PERELMAN, L. T., ITZKAN, I., DASARI, R. R. & FELD, M. S. 1997. Single Molecule Detection Using Surface-Enhanced Raman Scattering (SERS). *Physical Review Letters*, 78, 1667-1670.
- KOWALSKI, B. R. 1980. Chemometrics. *Analytical Chemistry*, 52, 112-122.
- KRUSZEWSKI, S. & CYRANKIEWICZ, M. 2012. Aggregated silver sols as SERS substrates. *Acta Physica Polonica-Series A General Physics*, 121, A68.
- KUMAR, R. A. & CLARK, D. S. 2006. High-throughput screening of biocatalytic activity: applications in drug discovery. *Current Opinion in Chemical Biology*, 10, 162-168.
- LAING, S., GRACIE, K. & FAULDS, K. 2016. Multiplex in vitro detection using SERS. *Chemical Society Reviews*, 45, 1901-1918.
- LARKIN, P. 2011. *Infrared and Raman Spectroscopy: Principles and Spectral Interpretation*, Elsevier Science.
- LARMOUR, I. A., FAULDS, K. & GRAHAM, D. 2012. SERS activity and stability of the most frequently used silver colloids. *Journal of Raman Spectroscopy*, 43, 202-206.
- LASCH, P. 2012. Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging. *Chemometrics and Intelligent Laboratory Systems*, 117, 100-114.
- LEE, P. C. & MEISEL, D. 1982. Adsorption and surface-enhanced Raman of dyes on silver and gold sols. *The Journal of Physical Chemistry*, 86, 3391-3395.
- LEOPOLD, N. & LENDL, B. 2003. A New Method for Fast Preparation of Highly Surface-Enhanced Raman Scattering (SERS) Active Silver Colloids at Room Temperature by Reduction of Silver Nitrate with Hydroxylamine Hydrochloride. *The Journal of Physical Chemistry B*, 107, 5723-5727.

- LERESCHE, J. E. & MEYER, H.-P. 2006. Chemocatalysis and Biocatalysis (Biotransformation): Some Thoughts of a Chemist and of a Biotechnologist. *Organic Process Research & Development*, 10, 572-580.
- LIN-VIEN, D., COLTHUP, N. B., FATELEY, W. G. & GRASSELLI, J. G. 1991. *The Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules*, Elsevier Science.
- LIU, C., CONSTANTINIDES, P. P. & LI, Y. 2014. Research and development in drug innovation: reflections from the 2013 bioeconomy conference in China, lessons learned and future perspectives. *Acta Pharmaceutica Sinica. B*, 4, 112-119.
- LONG, D. A. 2002. *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules*, Wiley.
- LÓPEZ-DÍEZ, E. C. & GOODACRE, R. 2004. Characterization of Microorganisms Using UV Resonance Raman Spectroscopy and Chemometrics. *Analytical Chemistry*, 76, 585-591.
- MABBOTT, S., CORREA, E., COWCHER, D. P., ALLWOOD, J. W. & GOODACRE, R. 2013. Optimization of Parameters for the Quantitative Surface-Enhanced Raman Scattering Detection of Mephedrone Using a Fractional Factorial Design and a Portable Raman Spectrometer. *Analytical Chemistry*, 85, 923-931.
- MARSICH, L., BONIFACIO, A., MANDAL, S., KROL, S., BELEITES, C. & SERGO, V. 2012. Poly-L-lysine-Coated Silver Nanoparticles as Positively Charged Substrates for Surface-Enhanced Raman Scattering. *Langmuir*, 28, 13166-13171.
- MARTIS, E., R, R. & R.R, B. 2011. High-Throughput Screening: The Hits and Leads of Drug Discovery- An Overview. *Journal of Applied Pharmaceutical Science*, 01, 02-10.
- MCNAY, G., EUSTACE, D., SMITH, E. W., FAULDS, K. & GRAHAM, D. 2011. Surface-Enhanced Raman Scattering (SERS) and Surface-Enhanced Resonance Raman Scattering (SERRS): A Review of Applications. *Applied Spectroscopy*, 65, 825-837.
- MOORE, B. D., STEVENSON, L., WATT, A., FLITSCH, S., TURNER, N. J., CASSIDY, C. & GRAHAM, D. 2004. Rapid and ultra-sensitive determination of enzyme activities using surface-enhanced resonance Raman scattering. *Nature Biotechnology*, 22, 1133-1138.
- MOSKOVITS, M. 1985. Surface-enhanced spectroscopy. *Reviews of Modern Physics*, 57, 783-826.
- NAIR, K. P. R. 2006. *Atoms, Molecules and Lasers*, Alpha Science.
- NGO, H. T., WANG, H.-N., BURKE, T., GINSBURG, G. S. & VO-DINH, T. 2014. Multiplex detection of disease biomarkers using SERS molecular sentinel-on-chip. *Analytical and Bioanalytical Chemistry*, 406, 3335-3344.
- NIE, S. & EMORY, S. R. 1997. Probing Single Molecules and Single Nanoparticles by Surface-Enhanced Raman Scattering. *Science*, 275, 1102-1106.
- OTTO, M. 2016. What is Chemometrics? *Chemometrics*. Wiley-VCH Verlag GmbH & Co. KGaA.
- PATEL, R. N. 2011. Biocatalysis: Synthesis of Key Intermediates for Development of Pharmaceuticals. *ACS Catalysis*, 1, 1056-1074.
- PATEL, R. N. 2017. Biocatalysis for synthesis of pharmaceuticals. *Bioorganic & Medicinal Chemistry*.
- PLANCK, M. 1914. *The Theory of Heat Radiation*, P. Blakiston Son & Co.
- PREMASIRI, W. R., LEE, J. C. & ZIEGLER, L. D. 2012. Surface Enhanced Raman Scattering of Whole Human Blood, Blood Plasma and Red Blood Cells: Cellular Processes and Bioanalytical Sensing. *The Journal of Physical Chemistry. B*, 116, 9376-9386.
- RAMAN, C. V. & KRISHNAN, K. S. 1928. A New Type of Secondary Radiation. *Nature*, 121, 501-502.
- REETZ, M. T. 2013. Biocatalysis in Organic Chemistry and Biotechnology: Past, Present, and Future. *Journal of the American Chemical Society*, 135, 12480-12496.

- REN, B., LIN, X.-F., YANG, Z.-L., LIU, G.-K., AROCA, R. F., MAO, B.-W. & TIAN, Z.-Q. 2003. Surface-Enhanced Raman Scattering in the Ultraviolet Spectral Region: UV-SERS on Rhodium and Ruthenium Electrodes. *Journal of the American Chemical Society*, 125, 9598-9599.
- SAEN-OON, S., GHANEM, M., SCHRAMM, V. L. & SCHWARTZ, S. D. 2008. Remote Mutations and Active Site Dynamics Correlate with Catalytic Properties of Purine Nucleoside Phosphorylase. *Biophysical Journal*, 94, 4078-4088.
- SATHYANARAYANA, D. N. 2015. *Vibrational Spectroscopy: Theory and Applications*, New Age International (P) Limited.
- SCHLÜCKER, S. & KIEFER, W. 2011. *Surface Enhanced Raman Spectroscopy: Analytical, Biophysical and Life Science Applications*, Wiley.
- SCHMID, A., DORDICK, J. S., HAUER, B., KIENER, A., WUBBOLTS, M. & WITHOLT, B. 2001. Industrial biocatalysis today and tomorrow. *Nature*, 409.
- SCHRADER, B. 2008. *Infrared and Raman Spectroscopy: Methods and Applications*, Wiley.
- SERVICE, R. F. 2004. Surviving the Blockbuster Syndrome. *Science*, 303, 1796-1799.
- SHERYL, B. R.-P. & HUIMIN, Z. 2006. Recent Advances in Biocatalysis by Directed Enzyme Evolution. *Combinatorial Chemistry & High Throughput Screening*, 9, 247-257.
- SLUTSKY, B. 1998. Handbook of Chemometrics and Qualimetrics: Part A *Journal of Chemical Information and Computer Sciences*, 38, 1254-1254.
- SMITH, E. & DENT, G. 2005. *Modern Raman Spectroscopy: A Practical Approach*.
- SMITH, W. E. 2008. Practical understanding and use of surface enhanced Raman scattering/surface enhanced resonance Raman scattering in chemical and biological analysis. *Chemical Society Reviews*, 37, 955-964.
- SPIRO, T. G. & STREKAS, T. C. 1972. Resonance Raman Spectra of Hemoglobin and Cytochrome c: Inverse Polarization and Vibronic Scattering. *Proceedings of the National Academy of Sciences of the United States of America*, 69, 2622-2626.
- STACY, A. A. & VAN DUYN, R. P. 1983. Surface enhanced raman and resonance raman spectroscopy in a non-aqueous electrochemical environment: Tris(2,2'-bipyridine)ruthenium(II) adsorbed on silver from acetonitrile. *Chemical Physics Letters*, 102, 365-370.
- STILES, P. L., DIERINGER, J. A., SHAH, N. C. & VAN DUYN, R. P. 2008. Surface-Enhanced Raman Spectroscopy. *Annual Review of Analytical Chemistry*, 1, 601-626.
- SZYMAŃSKI, P., MARKOWICZ, M. & MIKICIUK-OLASIK, E. 2012. Adaptation of High-Throughput Screening in Drug Discovery—Toxicological Screening Tests. *International Journal of Molecular Sciences*, 13, 427-452.
- TIAN, Z.-Q., REN, B. & WU, D.-Y. 2002. Surface-Enhanced Raman Scattering: From Noble to Transition Metals and from Rough Surfaces to Ordered Nanostructures. *The Journal of Physical Chemistry B*, 106, 9463-9483.
- TRUPPO, M. D. 2017. Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Medicinal Chemistry Letters*, 8, 476-480.
- TURNER, N. J. & TRUPPO, M. D. 2013. Biocatalysis enters a new era. *Current Opinion in Chemical Biology*, 17, 212-214.
- WEN, Z.-Q. 2007. Raman spectroscopy of protein pharmaceuticals. *Journal of Pharmaceutical Sciences*, 96, 2861-2878.
- WESTLEY, C., XU, Y., CARNELL, A. J., TURNER, N. J. & GOODACRE, R. 2016. Label-Free Surface Enhanced Raman Scattering Approach for High-Throughput Screening of Biocatalysts. *Analytical Chemistry*, 88, 5898-5903.
- WHITTALL, J. & SUTTON, P. W. 2009. *Practical Methods for Biocatalysis and Biotransformations*, John Wiley & Sons.
- WOHLGEMUTH, R. 2010. Biocatalysis - key to sustainable industrial chemistry. *Current Opinion in Biotechnology*, 21, 713-724.

- WOLD, S., ESBENSEN, K. & GELADI, P. 1987. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2, 37-52.
- XU, H., BJERNELD, E. J., KÄLL, M. & BÖRJESSON, L. 1999. Spectroscopy of Single Hemoglobin Molecules by Surface Enhanced Raman Scattering. *Physical Review Letters*, 83, 4357-4360.
- YAFFE, N. R. & BLANCH, E. W. 2008. Effects and anomalies that can occur in SERS spectra of biological molecules when using a wide range of aggregating agents for hydroxylamine-reduced and citrate-reduced silver colloids. *Vibrational Spectroscopy*, 48, 196-201.
- ZHAO, H., CHOCKALINGAM, K. & CHEN, Z. 2002. Directed evolution of enzymes and pathways for industrial biocatalysis. *Current Opinion in Biotechnology*, 13, 104-110.

2 Exploring Surface Enhanced Raman Scattering as a Method to Identify Regioselective Flavin-Dependent Halogenation

Heidi Fisk, Jason Micklefield and Royston Goodacre

School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK

2.1 DECLARATION

This chapter consists of one piece of work written in the format of a journal article but not currently submitted for publication.

As primary author, I carried out all experimental work, data analysis and write-up. Prof. Jason Micklefield provided advice and support throughout these initial investigations, particularly towards the biotransformation. Prof. Royston Goodacre, as principal investigator, gave fundamental advice and guidance throughout this research, predominantly towards aspects surrounding SERS.

Investigations surrounding the optimisation of SttH activity and its substrate scope were included within a subsequent publication (third author) which can be found in Appendix II: Additional Publications:

SHEPHERD, S. A., MENON, B. R., FISK, H., STRUCK, A. W., LEVY, C., LEYS, D., MICKLEFIELD, J., 2016. A Structure-Guided Switch in the Regioselectivity of a Tryptophan Halogenase. *ChemBioChem*, 17, 821-824

2.2 ABSTRACT

There has been increasing interest in the application of halogenase enzymes to insert halogen atoms into small molecules in selective positions, using benign reagents and mild conditions. Their widespread use in industry and academia has been hampered by their relatively low activity and substrate scope. Evolutionary approaches (*i.e.* mutagenesis) to further the development and optimisation of flavin-dependent halogenases has been hindered by the lack of high-throughput screens enabling fast quantification and assignment of regioselective products. To attempt to develop a rapid technique that would meet these requirements, surface enhanced Raman scattering (SERS) spectroscopy was investigated.

First, the activity and substrate scope of a flavin-dependent tryptophan halogenase, SttH, was assessed. These results aided the optimisation process required for SERS method development, demonstrating that we can successfully distinguish between regioisomeric products. Initially, *in vitro* biotransformations were investigated using SERS, however their cofactor requirements and associated complexity led to unsuccessful detection of halogenase substrates and subsequent products. To simplify the enzymatic reaction, *in vivo* conditions were explored to exploit cofactors inherent to the bacterial cells; one could omit the addition of cofactors (FAD and NADH), reducing the complexity of the biotransformation and effectively aiding SERS detection. However, the low *in vivo* activity and consequent requirement of bacterial cells in high density led to the demise of this approach; specifically, discriminant peaks of substrate and product(s) were difficult to detect amongst the bacterial supernatant SERS response. Further optimisation to increase the inherently low *in vivo* activity of this class of enzyme would aid successful SERS detection and quantification of halogenated products.

2.3 INTRODUCTION

Organohalogen moieties are present in a vast number of pharmaceutical and agrochemical products (estimated at 20 % and 30 %, respectively) along with other valuable synthetic and chemical products (Herrera-Rodriguez *et al.*, 2011). The presence of carbon-halogen bonds has been shown to have a profound effect in altering the physicochemical properties and bioavailability of a multitude of drug compounds, examples include rebaccamycin (antitumour agent) and vancomycin (antibiotic), which display significantly reduced potency when dechlorinated (Harris *et al.*, 1985, Rodrigues Pereira *et al.*, 1996). Furthermore, the utilisation of halogen atoms within synthesis has become an indispensable tool for the generation of complex molecules; a result of their participation in transition metal cross-coupling reactions, such as the Pd-catalysed Suzuki-Miyaura cross-coupling (Miyaura *et al.*, 1979). Despite their widespread use throughout numerous industries, the synthetic method of installing halogens currently remains toxic and harmful to the environment, with very little regiocontrol, thus affording many unwanted by-products. Consequently, interest has shifted towards the implementation of halogenase enzymes that can insert halogen atoms within a molecule (both natural and synthetic scaffolds) at ambient temperature in aqueous conditions, whilst utilising benign reagents and demonstrating regioselective control.

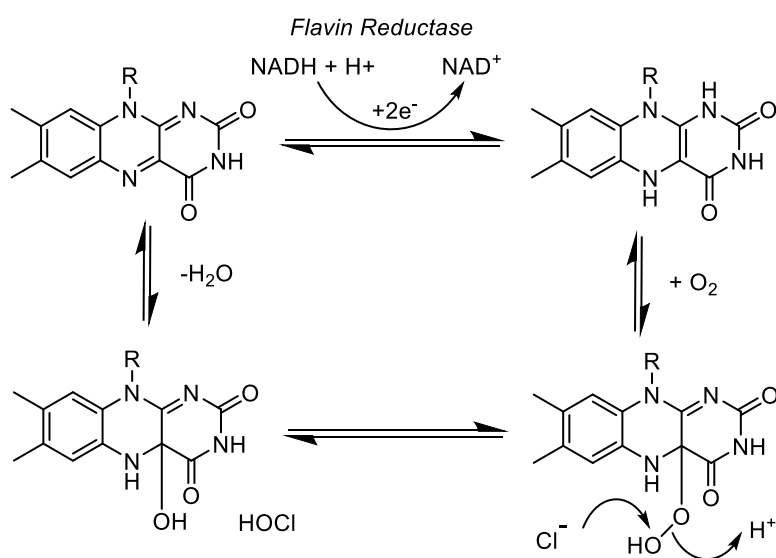


Figure 2.1. Cycle of NADH-dependent reduction of FAD by a flavin reductase. Hypochlorous acid (HOCl) is generated within the active site.

A variety of halogenating enzymes have been discovered, initially chloroperoxidase from the fungus, *Caldariomyces fumago* (Morris and Hager, 1966), followed by a range of vanadium-

and heme-dependent haloperoxidases (Vilter, 1984), yet these halogenases lack regioselectivity and substrate specificity (the generated chlorinating species are released from the active site to react freely with the substrate, without enzymatic control). More recently, Fe^{2+}/α -ketoglutarate (α KG)-dependent and flavin-dependent halogenases (FDHs) have been recognised as further examples of natural halogenating agents which show regioselectivity (Vaillancourt *et al.*, 2005). FDH enzymes make up part of the flavin-dependent monooxygenase superfamily (Dong *et al.*, 2005, Mascotti *et al.*, 2016). A main feature of this class of enzymes is their activation of molecular oxygen and reduced flavin (FADH_2), generating the highly reactive flavin hydrogen peroxide species (FAD-OOH). A chloride ion, located nearby, is oxidised to form free hypochlorous acid (HOCl), Figure 2.1. It is at this point that FDHs demonstrate regiocontrol as HOCl cannot diffuse and react uncontrollably, instead, it is directed through a tunnel towards the substrate binding site where halogenation ensues in a regiospecific manner, Figure 2.2 (Dong *et al.*, 2005, Yeh *et al.*, 2006). FDHs can be subdivided by their natural substrate: tryptophan, pyrrole or phenolic FDHs.

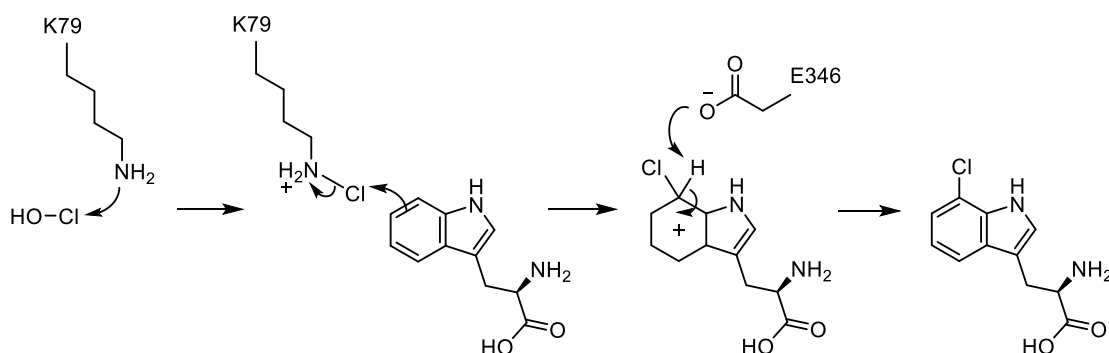


Figure 2.2. The electrophilic addition of chlorine to tryptophan within the active site of FDHs. A lysine residue (K79) leads HOCl to the tryptophan substrate, whereby chlorine reacts through electrophilic addition to the π -system of the aromatic ring. A glutamate residue (E346) deprotonates the intermediate leading to the final chlorinated product.

Flavin-dependent tryptophan halogenases are the most extensively studied and characterised class, displaying halogenation (chlorination and bromination) of the indole moiety of tryptophan at one of C-5, C-6 or C-7 (Figure 2.3). Those that have been characterised and their regiocontrol investigated include 7-halogenases, PrnA (Dong *et al.*, 2005, Shepherd *et al.*, 2015) and RebH (Bitto *et al.*, 2008), along with a 5-halogenase PyrH (Zhu *et al.*, 2009) (Figure 2.3). To date, little information has been elucidated about the 6-

halogenase SttH as no crystal structure and investigations into substrate scope have been explored (Zeng and Zhan, 2011).

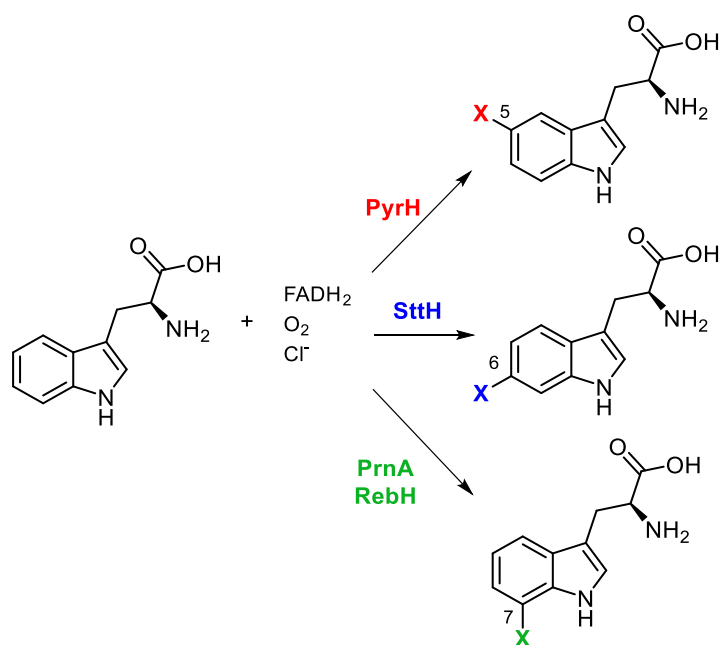


Figure 2.3. Reactions of flavin-dependent tryptophan halogenases with L-tryptophan and their respective regioselective products (X = Cl/Br).

Despite obvious advantages associated with regiocontrol demonstrated by flavin-dependent halogenases, their application has been hindered by their low efficiency and inherent instability. Advancements have been made to improve these pitfalls, such as cross-linked enzyme aggregates (CLEAs) (Frese and Sewald, 2015) and mutagenesis (Lang *et al.*, 2011, Poor *et al.*, 2014, Shepherd *et al.*, 2015), yet there is still a large obstacle to overcome before their widespread application. A contributing factor towards their limited development is the lack of a high-throughput screen that can rapidly screen for halogenase activity, as well as discriminating between regioselectivity. Commonly, low-throughput methods requiring UHPLC separation have been employed for initial observations of activity, however, full characterisation of regioselectivity needs more information rich techniques, such as NMR, requiring a much larger amount of sample and purification steps (Kumar and Clark, 2006). Progress into the development of faster halogenation screens has been made in the form of a quinone-amine coupling that can be used to colorimetrically monitor arylamine halogenation (Hosford *et al.*, 2014), as well as a fluorescence screen that can quantify halogenation after the generation of the fluorescent species *via* a cross-coupling Suzuki-Miyaura reaction (Schnepel *et al.*, 2016). Colorimetric and fluorescent techniques are frequently favoured as high-throughput methods arising from their ease of use and rapid

acquisition times (Bronson *et al.*, 2001, Bornscheuer, 2013, Grunwald, 2014). Nevertheless, these methods for monitoring halogenation are restricted as they have limited substrate scope, indirectly quantify product generation and require an extensive cross-coupling reaction, respectively.

Owing to the limited scope and extensive preparation of these approaches, we explored the development of a rapid and robust screen which would not only quantify the generation of product(s), but also differentiate between regioisomers. Raman spectroscopy, primarily its enhancement technique, surface enhanced Raman scattering (SERS), is an attractive technique arising from its compatibility towards aqueous conditions, its capacity for multiplexing (cofactor requirement of the halogenase biotransformation) and the speed of analysis (seconds). SERS demonstrates significant enhancements of Raman signals when analytes are adsorbed onto (or in close proximity to) a roughened nanoscale metallic surface. Typically, enhancements in the order of 10^4 – 10^6 are experienced, generating characteristic, molecular specific information about the molecule(s) of interest (Fleischmann *et al.*, 1974, Jeanmaire and Van Duyne, 1977, Albrecht and Creighton, 1977, Campion and Kambhampati, 1998). Recently, SERS has successfully been used to monitor biotransformations, further highlighting its suitability for this application (Westley *et al.*, 2016, Morelli *et al.*, 2017).

Herein, we investigate the suitability of SERS for quantifying flavin-dependent halogenation, as well as determining regioselective activity. Initially, the substrate scope, corresponding activity and regiospecificity of SttH (C-6, Figure 2.3) was determined using flavin-dependent tryptophan halogenases, enabling comparisons at all regiospecific positions. Next, SERS was explored to identify spectral differences arising from the position of the halogen around the aromatic structures. Clear spectral separation of regioisomers led us to investigate the application of SERS to monitor flavin-dependent halogenase biotransformations towards both *in vitro* and *in vivo* systems.

2.4 EXPERIMENTAL

2.4.1 Reagents and Materials.

All chemical reagents were of analytical grade and used with no additional purification unless otherwise stated. Full details of cloning, expression and protein purification is described in the SI section 'Protein Expression and Purification'.

2.4.2 Halogenase Biotransformation Sample Preparation.

***In vitro* (purified enzyme).** SttH (1×10^{-5} M) was incubated at 30 °C with agitation for 1 h with Fre (1×10^{-6} M), GDH (6×10^{-6} M), FAD (7.5×10^{-6} M), NADH (2×10^{-4} M), MgCl₂ (5×10^{-2} M), glucose (2×10^{-2} M) and substrate (5×10^{-4} M) in a total volume of 100 μL in potassium phosphate buffer (1×10^{-2} M, pH 7.0). Details surrounding biotransformation optimisation are described in the SI, section 'Optimisation of SttH Biotransformation'.

***In vivo* (whole-cell).** Arctic Express (DE3) *E. coli* cells enclosing two vectors (pACYC Duet-1 vector containing *Fre* and *GDH* genes, along with *RadH* pET 28b(+)) were transformed and expressed using protocols as stated in the SI section 'Protein Expression and Purification'. 10 equivalents of bacterial culture (v/v ratio) were resuspended in potassium phosphate buffer (1×10^{-2} M, pH 7.0), followed by the addition of substrate (5×10^{-4} M), MgCl₂ (1×10^{-2} M) and glucose (1×10^{-2} M), total volume of 1 mL. The reaction was incubated at 30 °C with shaking for up to 1 h.

2.4.3 SERS Sample Preparation.

Tryptophan halogenases. SERS analysis using borohydride-reduced silver colloid (BRSC): 350 μL of the silver colloid was pipetted into an 8 mm glass vial, followed by 50 μL of the sample (total sample concentration 2.5×10^{-5} M in HEPES buffer (1×10^{-2} M, pH 7.0)). The vial was capped and vortexed for 7 s and allowed to aggregate for 30 s before recording a 20 s SERS spectrum using 633 nm irradiation. A colour change from yellow to purple indicated successful aggregation. BRSC was prepared using a slightly modified procedure as reported by Lee and Meisel (Lee and Meisel, 1982), and is provided in the SI, Supplementary Methods.

Phenolic halogenases. SERS analysis using citrate-reduced silver colloid (CRSC): 200 μL of the silver colloid was pipetted into an 8 mm glass vial, followed by 200 μL of the sample (total sample concentration 2×10^{-5} M) and 50 μL of KNO₃ (1×10^{-1} M). The vial was capped and

vortexed for 7 s, left to aggregate for 30 s before collecting a 10 s SERS spectrum using 785 nm laser irradiation. CRSC was prepared according to the method of Lee and Meisel (Lee and Meisel, 1982).

2.4.4 Raman Instrumentation.

Tryptophan halogenase substrates. SERS spectra were recorded using a DeltaNu Advantage 200A portable spectrometer (DeltaNu, Laramie, WY, USA) equipped with a HeNe 633 nm laser with ~3 mW of power on the sample.

Phenolic halogenase substrates. SERS spectra were recorded using a DeltaNu Advantage 200A portable spectrometer (DeltaNu, Laramie, WY, USA) equipped with a HeNe 785 nm laser with ~60 mW of power on the sample.

2.5 RESULTS AND DISCUSSION

2.5.1 Biotransformation Optimisation

We began the development of a Raman-based assay by exploring the substrate scope of SttH (C-6) flavin-dependent tryptophan halogenase. Extensive investigations into substrate scope, optimisation of reaction conditions and the assignment of regioisomers have previously been conducted within the Micklefield group, along with others, for PyrH (C-5), PrnA and RebH (both C-7). These previous studies helped guide the choice of substrates and condition parameters towards SttH optimisation. SttH and FAD reductase (Fre, responsible for the reduction of FAD to FADH₂) were previously cloned into vectors by members of the Micklefield group, along with glucose dehydrogenase (GDH, involved in NADH recycling) which was cloned by the Scrutton group (The University of Manchester). Figure 2.4 illustrates the overall biotransformation process.

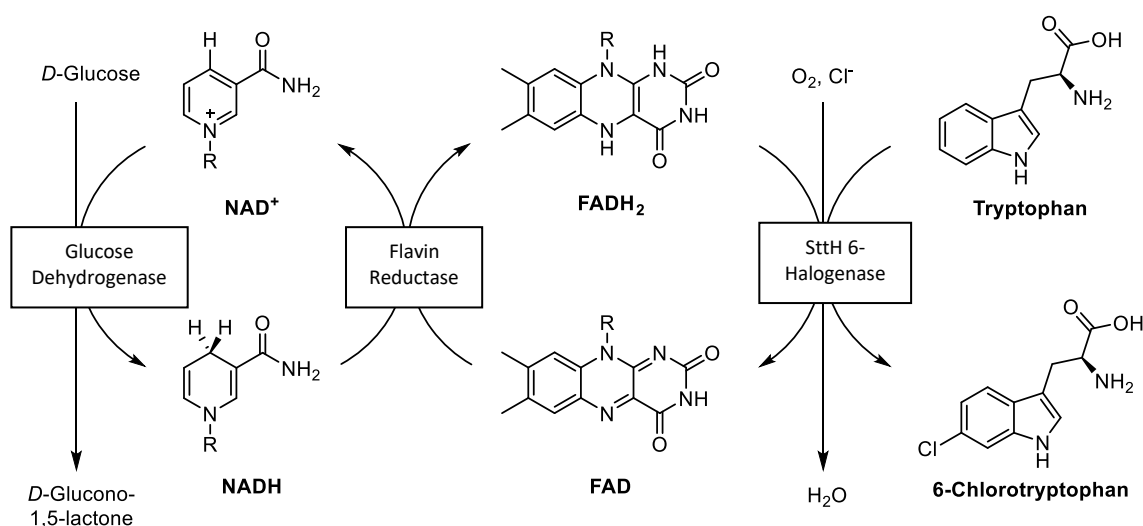
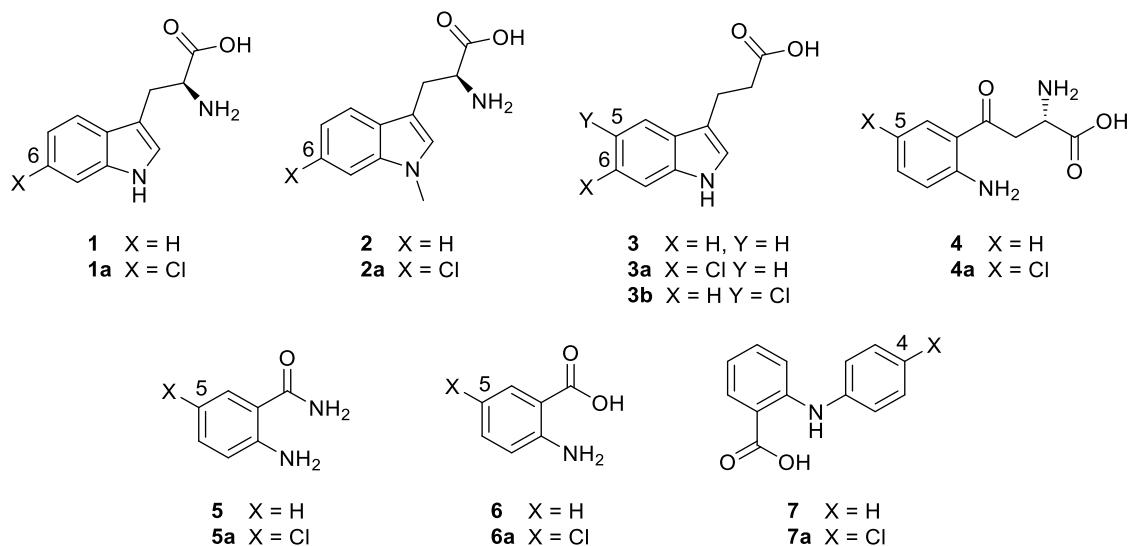


Figure 2.4. An overview of SttH halogenation of tryptophan at C-6, involving glucose dehydrogenase (GDH) and flavin reductase (Fre) enzymes. Details of cloning, transformation, expression and purification are described in the SI, section ‘Protein Expression and Purification’.

FDH biotransformations rapidly consume large quantities of expensive NADH cofactor as Fre reduces FAD into its active form, FADH₂. Consequently, it is common practice to install recycling systems, such as GDH, to reduce NAD⁺ using inexpensive and abundant substrates (*i.e.* glucose). Parameters, such as the ratio of enzymes and the concentration of cofactors were optimised to produce maximal halogenase activity, see SI section ‘Optimisation of SttH Biotransformation’, Figure S2.1 and S2.2. We then tested a large range of substrates for activity towards SttH to determine its substrate scope (Figure S2.3 in SI). Those that were

active substrates with good conversion were performed on a larger, preparative scale, enabling NMR analysis and assignment of regioselectivity (see Table 2.1 and SI section 'Preparative Halogenase Reactions and NMR Assignments' for full characterisation).

Table 2.1. Conversion of substrate after 1 h with SttH.



Substrate	Conversion [%]	Product
Tryptophan (1)	97 ± 4	1a
N-Methyltryptophan (2)	68 ± 3	2a
3-Indolepropionic acid (3)	57 ± 1.5	3a:3b (9:1)
Kynurenine (4)	79 ± 4	4a
Anthranilamide (5)	43 ± 2	5a
Anthranilic acid (6)	1.1 ± 0.1	6a
N-phenylanthranilic acid (7)	20 ± 2	7a

Red indicates conversion to expected product. Blue indicates conversion to chemically favoured products. Assay conditions: SttH (1×10^{-5} M) was incubated at 30 °C with agitation for 1 h with Fre (1×10^{-6} M), GDH (6×10^{-6} M), FAD (7.5×10^{-6} M), NADH (2×10^{-4} M), MgCl_2 (5×10^{-2} M), glucose (2×10^{-2} M) and substrate (5×10^{-4} M) in a total volume of 100 μL in potassium phosphate buffer (1×10^{-2} M, pH 7.0). Data was performed in triplicate, error shown.

2.5.2 SERS Optimisation

Following these investigations, all regiospecific flavin-dependent tryptophan halogenases (C-5, C-6, C-7) were available to develop the Raman-based assay. Tryptophan (Table 2.1, substrate 1) and its 5-brominated analog, along with anthranilic acid (Table 2.1, substrate 6) and its 3-, 4- and 5-chlorinated derivatives were selected as initial analytes (owing to the

commercial availability of their halogenated derivatives). Optimisation of each analyte for SERS responses was studied by varying multiple experimental parameters systematically (see SI, section 'SERS Optimisation'). Introduction of potassium phosphate buffer, which was previously used to buffer the halogenase biotransformation, resulted in a complete loss of SERS response, therefore several alternative biological buffers (pH 6.0 - 7.4) were screened as alternatives. From these tests, both HEPES and MOPS buffers (pH 7.0) retained SERS signals of substrate(s) and product(s) previously seen in water alone and were shown to have no effect on observed halogenase activity. Therefore, subsequent SERS optimisation and halogenase biotransformations were performed using HEPES buffer. Optimal SERS responses were obtained using 350 μ L of borohydride reduced silver colloid (BRSC), 2.5×10^{-5} M total analyte concentration in HEPES buffer (1×10^{-2} M, pH 7.0), 30 s aggregation time and 20 s SERS acquisition using 633 nm laser excitation.

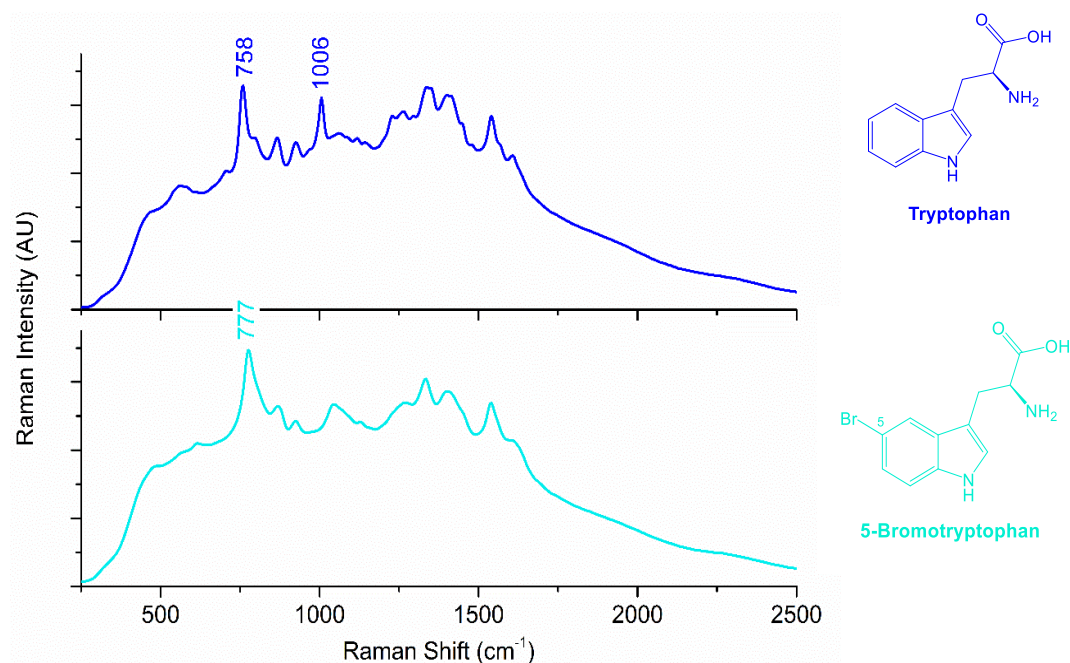


Figure 2.5. Average SERS spectra ($n = 5$) of tryptophan (blue) and 5-bromotryptophan (cyan) (2.5×10^{-5} M), characteristic peaks have been labelled. SERS spectra were acquired for 20 s with BRSC in HEPES buffer (pH 7.0).

If one compares the SERS responses for tryptophan and 5-bromotryptophan (Figure 2.5), characteristic peaks are easily identifiable, enabling one to distinguish between the two analytes (see Table S2.1 in SI for tentative band assignments); mainly, the vibration at 758 cm^{-1} which is red shifted to 777 cm^{-1} upon bromination, along with the loss of vibration at 1006 cm^{-1} (attributable to NH vibration). These results suggest that quantification of halogenase activity should be achievable. However, if we focus on anthranilic acid and its

chlorinated analogs (3-, 4- and 5-chloroanthranilic acid, Figure 2.6), it is apparent that discrimination between regioisomers is more complex.

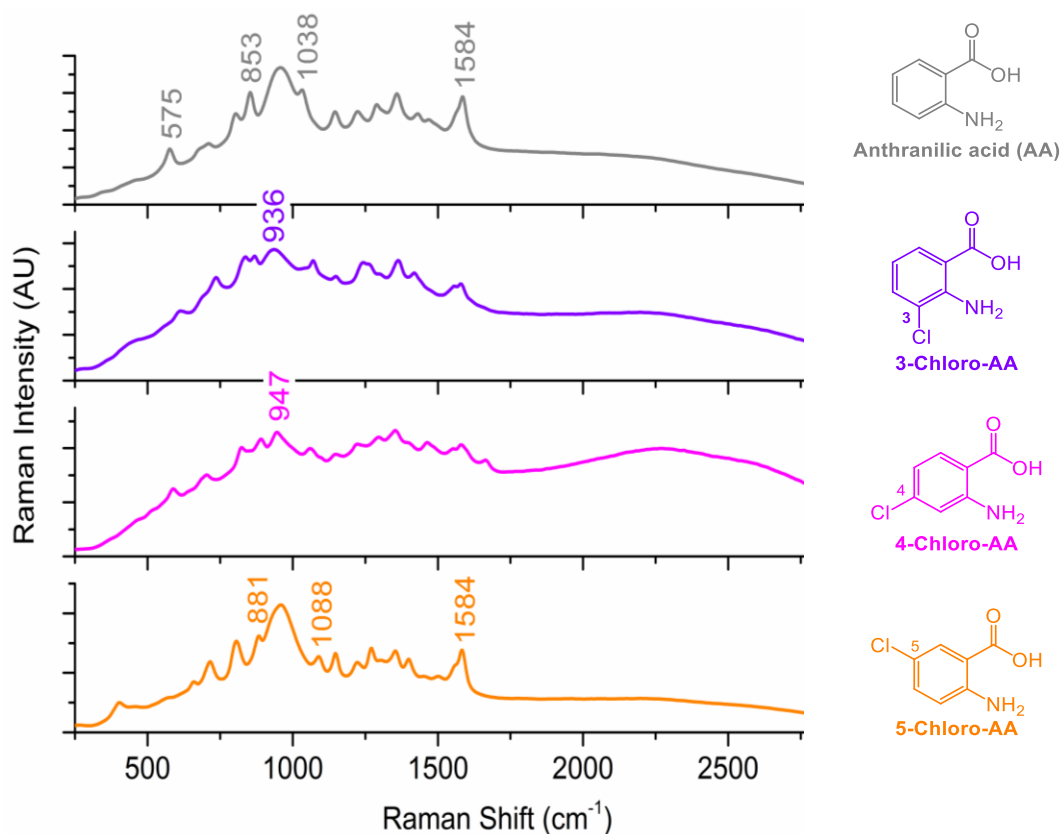


Figure 2.6. Average SERS spectra ($n = 5$) of anthranilic acid (grey), 3-chloroanthranilic acid (purple), 4-chloroanthranilic acid (pink), 5-chloroanthranilic acid (orange), (2.5×10^{-5} M). Characteristic peaks have been labelled, SERS spectra were obtained for 20 s with BRSC in HEPES buffer (pH 7.0).

Chemometrics is the use of mathematical and statistical methods to aid the analysis of chemical data. Principal component analysis (PCA) is a multivariate approach designed to reduce complex data sets into lower dimensional data, often revealing hidden relationships or patterns which can be seen in the PCA scores plots. The corresponding loadings plot identifies the relationship between the original variables and the subspace dimensions as indicated by the PCA scores plot. PCA was performed and the loadings plotted which are shown in Figure S2.4 and S2.5 in SI (see SI section 'Identification of Anthranilic acid Regioisomers'). We observe clear separation for all four analytes within the PCA scores plot arising from spectral differences, with subtle, discriminant peaks (as identified by the loadings plot) assigned in Figure 2.6 (see Table S2.2 in SI for tentative band assignments).

Following successful discrimination of halogenated regioisomers (Figure 2.6), a more complex approach was investigated involving a more realistic representation of the biotransformation by introduction of the necessary cofactors (Figure 2.4). All cofactors were screened for SERS responses using optimal SERS conditions as stated previously, with no observed SERS signal for NADH, MgCl_2 , or glucose. In contrast, FAD gave a strong SERS response and bands were consistent with those previously reported in the literature (Xu *et al.*, 1987) (Figure 2.7B). As the most intense peaks corresponding to FAD lie in the region of $1100 - 1700 \text{ cm}^{-1}$, it might be possible to differentiate between non-halogenated and halogenated regioisomers corresponding to characteristic peaks using the $500 - 1100 \text{ cm}^{-1}$ region (Figure 2.7A and Figure 2.7B).

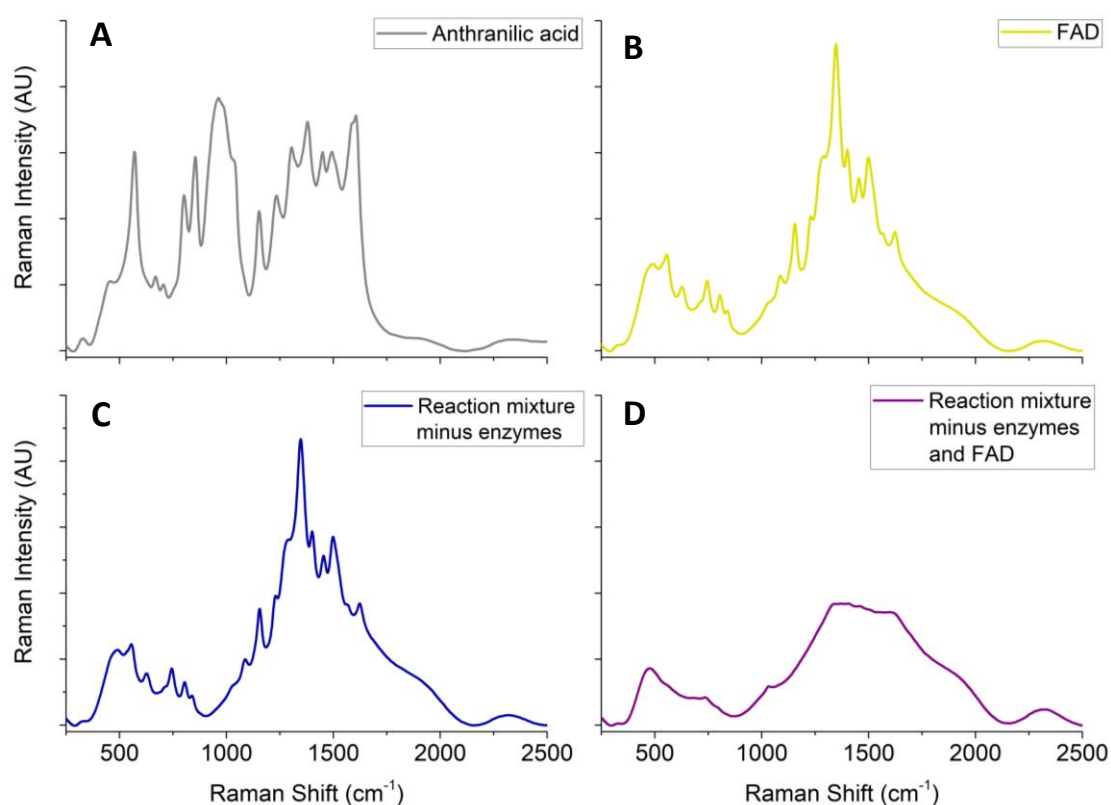


Figure 2.7. Average SERS spectra ($n = 6$) of (A) anthranilic acid, (B) FAD, (C) halogenase reaction mixture minus enzymes (analyte + cofactors + buffer) and (D) halogenase reaction mixture minus enzymes and FAD cofactor. Concentrations are representative of halogenase biotransformation conditions. SERS spectra were obtained for 20 s with BRSC in HEPES buffer (pH 7.0). SERS data has been baseline corrected and normalised (see SI, data processing for details).

Reaction mixtures were prepared representing conditions used within the halogenase reaction: tryptophan/anthranilic acid ($5 \times 10^{-4} \text{ M}$), FAD ($7.5 \times 10^{-6} \text{ M}$), NADH ($2 \times 10^{-4} \text{ M}$), MgCl_2 ($5 \times 10^{-2} \text{ M}$) and glucose ($2 \times 10^{-2} \text{ M}$) in HEPES ($1 \times 10^{-2} \text{ M}$, pH 7.0). Figure 2.7C shows the SERS reaction mixture response containing anthranilic acid. This SERS spectrum is identical

to that of FAD, with no peaks attributable to anthranilic acid. Following this observation, the reaction mixture was prepared, omitting FAD to restore the anthranilic acid signal. As shown in Figure 2.7D, the FAD response disappears as expected, however we were unable to retrieve the anthranilic acid SERS signal. Similar observations were made when tryptophan was introduced into the reaction mixtures. We assume that the loss of SERS response is arising from the complexity of the biotransformation, with the presence of numerous cofactors preventing favourable interactions of the analyte with the nanoparticle surface, thus hindering detection. For completeness, samples from an SttH biotransformation containing tryptophan/anthranilic acid were analysed with SERS, but as anticipated, no SERS responses were obtained (data not shown).

2.5.3 Whole-Cell Investigations

Considering our previous findings, we proposed that we could try to reduce the concentration of cofactors that are present within the reaction mixture by using whole-cells, thus exploiting cofactors that are innate to the *E. coli* cell (*i.e.* NADH, FAD), which may also be recycled *in situ*. To do this, halogenase biotransformations were explored using intact whole-cells, so that cofactor concentrations could be kept to a minimum and within the cell (which would be removed before analysis). We began by transforming a pACYCDuet-1 vector containing both *Fre* and *GDH* genes, along with a pET 28b(+) vector containing either *PyrH* (5-tryptophan halogenase) or *RadH* (phenolic halogenase) into competent *E. coli* cells (see SI section 'Halogenase Whole-Cell Optimisation'). SttH was omitted from these investigations as *PyrH* and *RadH* demonstrate greater activity and stability. Transformation into two different *E. coli* expression strains (BL21 (DE3) and Arctic Express (DE3)) were successful, obtaining colonies for all four strains (Figure S2.6 in SI). Subsequent tests for halogenase activity identified that only one cell line demonstrated activity; Arctic Express (DE3) containing *RadH*, *Fre* and *GDH* genes.

RadH, a phenolic FDH, accepts flavonoid-type analytes. These substrates differ greatly in structure from those previously explored towards tryptophan halogenases, so it was essential to reperform SERS optimisation. Optimal SERS conditions were found to be: citrate-reduced silver colloid (CRSC) (200 μ L), 2×10^{-5} M analyte concentration, KNO_3 aggregating agent (50 μ L, 1×10^{-1} M), 30 s aggregation time, 10 s acquisition time using 785 nm laser excitation. A panel of flavonoid-type substrates were screened for activity using whole-cell *RadH-Fre-GDH* in Arctic Express, amongst these substrates, chrysin and apigenin were found to be the most active (activity based on HPLC traces), Figure 2.8.

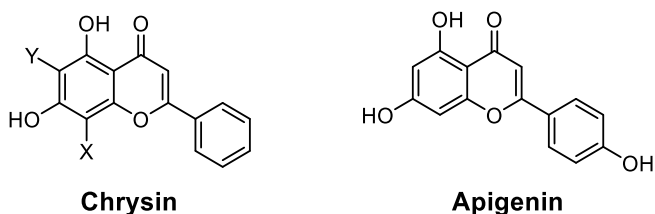


Figure 2.8. Structures of chrysin (X = H, Y = H) and apigenin which were shown to have the greatest activity towards RadH whole-cell biotransformations. Structures of chlorochrysin (X = Cl, Y = H) and dichlorochrysin (X = Cl, Y = Cl), the products of RadH halogenation, are shown on the chrysin backbone structure.

In contrast to the previously discussed tryptophan halogenases which regioselectively halogenate at a single position, these phenolic halogenases can undergo dihalogenation (substrate and time permitting). Chlorochrysin and dichlorochrysin had previously been isolated within the Micklefield group (using purified RadH) and the position(s) of halogenation assigned (Figure 2.8). Using optimal SERS conditions, one could identify spectral differences between chrysin, chlorochrysin and dichlorochrysin, along with similarities, Figure 2.9 (see SI section ‘Distinguishing between Chrysin, Chlorochrysin and Dichlorochrysin’, Figures S2.7 – S2.8). The Raman shifts in wavenumbers of discriminant bands are detailed in Table 2.2 (see Table S2.3 in SI for tentative SERS band assignments). Most spectral differences are observed in the region of 580 – 800 cm^{-1} which is in agreement with the literature, as these vibrations are of lower cm^{-1} due to the heavy halogens involved ($\nu(\text{C-Cl})$ 550 - 800 cm^{-1}) (Shakila *et al.*, 2011).

Table 2.2. Discriminant bands from SERS (in cm^{-1}) for chrysin, chlorochrysin and dichlorochrysin.

Chrysin	Chlorochrysin	Dichlorochrysin
		580 (w)
	606 (m)	
620 (s)		632 (s)
641 (s)		684 (m)
	684 (m)	
745 (s)		836 (m)

Abbreviations: s, strong; m, medium; w, weak.

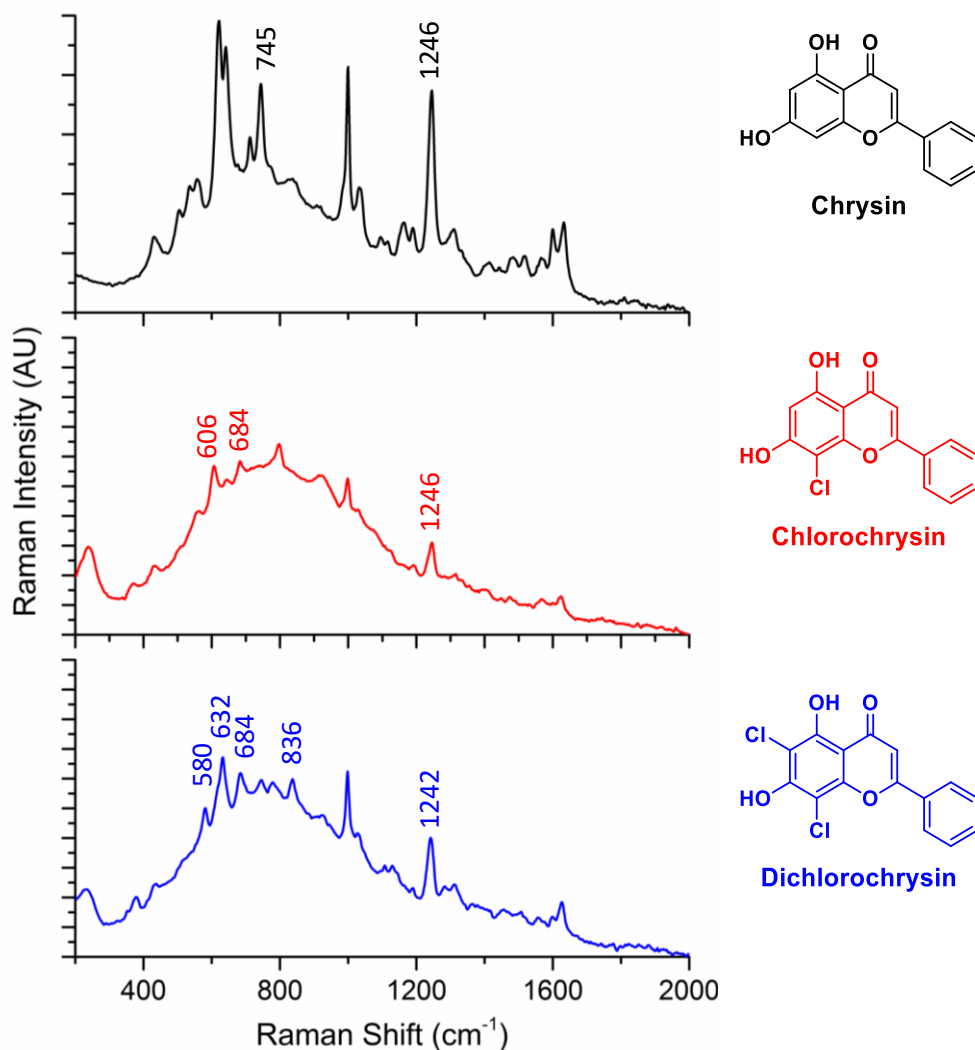


Figure 2.9. Average SERS spectra ($n = 5$) of chrysin (black), chlorochrysin (red) and dichlorochrysin (blue) (2×10^{-5} M). Characteristic peaks have been labelled, SERS spectra were obtained for 10 s with CRSC in potassium phosphate buffer (pH 7.0) and KNO_3 (1×10^{-1} M).

Next, optimised whole-cell biotransformations affording maximal halogenase activity was achieved (*n.b.* bacterial cells were in a resting state; after induction and overnight protein expression, cells were harvested and washed with buffer prior to use within the biotransformations). As we postulated, the addition of FAD and NADH is not needed for activity, however, a high density of bacterial cells was necessary for moderate activity (>30 % chlorination of chrysin (5×10^{-4} M) after 1 h incubation using 10 mL of bacterial culture resuspended in 1 mL reaction mixture). Multiple methods of quenching the biotransformation were investigated, including those that did not cause the bacteria to rupture, yet quenching using an organic solvent (EtOH/MeOH) was deemed the most appropriate method (see SI section ‘Whole-Cell SERS Observations’ for further information).

Despite a significant number of overlapping bands arising from *E. coli*, particularly within the region of interest ($580 - 800 \text{ cm}^{-1}$), SERS spectra of chrysin, chlorochrysin and the control sample of *E. coli* supernatant (post EtOH quenching) were successfully generated and assessed. Noticeably, subtle peaks were identified that should enable identification of chlorinated *versus* non-chlorinated chrysin (SI section 'Whole-Cell SERS Observations' Figure S2.9). Consequently, a whole-cell biotransformation was performed, taking various timepoints throughout, with the final sample at 1 h. At this final time point, HPLC confirmed 35 % conversion to chlorochrysin (incubation >1 h caused dichlorination and no further increase in chlorochrysin concentration). Comparing the control sample of the EtOH *E. coli* supernatant with the biotransformation reaction samples at 0 min and 1 h (Figure 2.10), it is obvious that the SERS response inherent to the *E. coli* supernatant dominates the spectra. Signature peaks at 1000 and 1246 cm^{-1} can be seen for chrysin/chlorochrysin, yet they are weak and do not distinguish the starting material from its chlorinated product. The discriminant bands for chlorochrysin at 684 and 797 cm^{-1} were not detectable within the complex supernatant, and as a result, the identification of halogenase activity using SERS with whole-cells was unsuccessful.

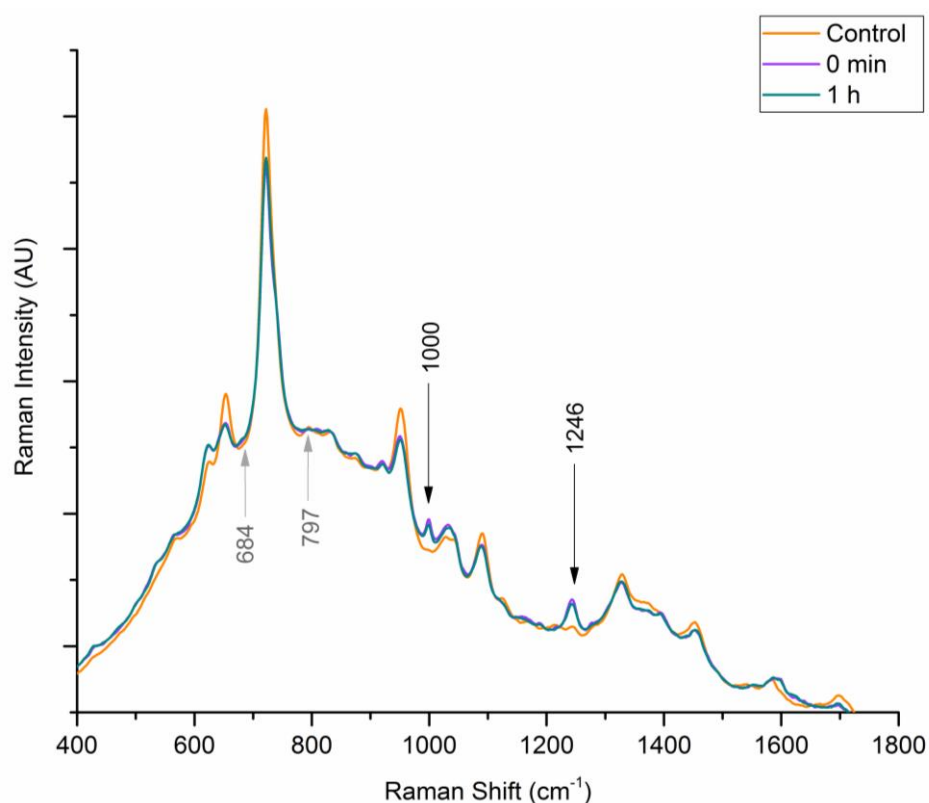


Figure 2.10. SERS spectra ($n = 4$) of the *E. coli* supernatant control sample (orange), biotransformation sample at 0 min (purple) and the final biotransformation timepoint sample after 1 h incubation at 30°C (dark cyan). Bands labelled in black were detectable, those in pale grey were unseen. SERS spectra were collected for 10 s with CRSC and KNO_3 . SERS data have been normalised, see SI, data processing for details.

2.6 CONCLUSIONS

We first attempted to discriminate between regioisomeric products of flavin-dependent tryptophan halogenase using SERS. Whilst this proved to be successful amongst standards, implementing these findings within real biotransformation samples became challenging. The leading factor resulting in the unsuccessful application of SERS to monitor these reactions was the complexity of the biotransformation. Complex cofactors, such as FAD and NADH, prevented favourable interactions of the analyte(s) with the nanoparticle surface. As a method to combat cofactor complexity, we developed a strategy of monitoring the biotransformation using whole-cells. The application of *in vivo* reactions meant that we could exploit cofactors inherent to the bacterial cells, only requiring additional MgCl₂ (chloride source) and glucose (NADH recycling), which we had previously established did not interfere with the SERS response of substrate(s) and product(s).

Tryptophan-halogenases (PryH) did not display activity within whole-cell reactions, yet RadH, a phenolic-halogenase, demonstrated good activity *in vivo*. As the substrate scope of RadH differed from those previously studied, SERS optimisation was re-performed, with citrate-reduced silver colloid being more favourable over borohydride-reduced nanoparticles, which was desirable due to its increased stability and reproducibility. Using these optimal SERS conditions, one could identify discriminant bands between chrysin, chlorochrysin and dichlorochrysin. Quenching the whole-cell biotransformation with ethanol was found to be the most suitable method, however this added complications to the SERS analyses due to the lysis and release of cell contents. The supernatant gave a noticeable SERS response which unfortunately coincided with all discriminant bands, meaning that identification of halogenation was not possible.

In conclusion, we have investigated SERS to monitor the generation of halogenated products using both *in vitro* and *in vivo* FDHs. However, the innate complexity of this enzyme-catalysed reaction and their relatively low activity prevented successful detection and subsequent quantification of halogenated products.

2.7 REFERENCES

- ALBRECHT, M. G. & CREIGHTON, J. A. 1977. Anomalously intense Raman spectra of pyridine at a silver electrode. *Journal of the American Chemical Society*, 99, 5215-5217.
- BITTO, E., HUANG, Y., BINGMAN, C. A., SINGH, S., THORSON, J. S. & PHILLIPS, G. N. 2008. The structure of flavin-dependent tryptophan 7-halogenase RebH. *Proteins: Structure, Function, and Bioinformatics*, 70, 289-293.
- BORNSCHEUER, U. T. 2013. Protein Engineering as a Tool for the Development of Novel Bioproduction Systems. In: ZENG, A.-P. (ed.) *Fundamentals and Application of New Bioproduction Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- BRONSON, D., HENTZ, N., JANZEN, W., LISTER, M., MENKE, K., WEGRZYN, J. & SITTA SITTAMPALAM, G. 2001. Basic Considerations in Designing High-Throughput Screening Assays. *Handbook of Drug Screening*. CRC Press.
- CAMPION, A. & KAMBHAMPATI, P. 1998. Surface-enhanced Raman scattering. *Chemical Society Reviews*, 27, 241-250.
- DONG, C., FLECKS, S., UNVERSUCHT, S., HAUPT, C., VAN PÉE, K.-H. & NAISMITH, J. H. 2005. Tryptophan 7-Halogenase (PrnA) Structure Suggests a Mechanism for Regioselective Chlorination. *Science*, 309, 2216-2219.
- FLEISCHMANN, M., HENDRA, P. J. & MCQUILLAN, A. J. 1974. Raman spectra of pyridine adsorbed at a silver electrode. *Chemical Physics Letters*, 26, 163-166.
- FRESE, M. & SEWALD, N. 2015. Enzymatic Halogenation of Tryptophan on a Gram Scale. *Angewandte Chemie International Edition*, 54, 298-301.
- GRUNWALD, P. 2014. *Industrial Biocatalysis*, Pan Stanford Publishing.
- HARRIS, C. M., KANNAN, R., KOPECKA, H. & HARRIS, T. M. 1985. The role of the chlorine substituents in the antibiotic vancomycin: preparation and characterization of mono- and didechlorovancomycin. *Journal of the American Chemical Society*, 107, 6652-6658.
- HERRERA-RODRIGUEZ, L. N., KHAN, F., ROBINS, K. T. & MEYER, H.-P. 2011. Perspectives on biotechnological halogenation. Part I: Halogenated products and enzymatic halogenation. *Chimica Oggi - Chemistry Today*, 29.
- HOSFORD, J., SHEPHERD, S. A., MICKLEFIELD, J. & WONG, L. S. 2014. A High-Throughput Assay for Arylamine Halogenation Based on a Peroxidase-Mediated Quinone-Amine Coupling with Applications in the Screening of Enzymatic Halogenations. *Chemistry - A European Journal*, 20, 16759-16763.
- JEANMAIRE, D. L. & VAN DUYN, R. P. 1977. Surface raman spectroelectrochemistry. *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry*, 84, 1-20.
- KUMAR, R. A. & CLARK, D. S. 2006. High-throughput screening of biocatalytic activity: applications in drug discovery. *Current Opinion in Chemical Biology*, 10, 162-168.
- LANG, A., POLNICK, S., NICKE, T., WILLIAM, P., PATALLO, E. P., NAISMITH, J. H. & VAN PÉE, K.-H. 2011. Changing the Regioselectivity of the Tryptophan 7-Halogenase PrnA by Site-Directed Mutagenesis. *Angewandte Chemie International Edition*, 50, 2951-2953.
- LEE, P. C. & MEISEL, D. 1982. Adsorption and surface-enhanced Raman of dyes on silver and gold sols. *The Journal of Physical Chemistry*, 86, 3391-3395.
- MASCOTTI, M. L., JURI AYUB, M., FURNHAM, N., THORNTON, J. M. & LASKOWSKI, R. A. 2016. Chopping and Changing: the Evolution of the Flavin-dependent Monooxygenases. *Journal of Molecular Biology*, 428, 3131-3146.
- MIYAURA, N., YAMADA, K. & SUZUKI, A. 1979. A new stereospecific cross-coupling by the palladium-catalyzed reaction of 1-alkenylboranes with 1-alkenyl or 1-alkynyl halides. *Tetrahedron Letters*, 20, 3437-3440.

- MORELLI, L., ZÓR, K., JENDRESEN, C. B., RINDZEVICIUS, T., SCHMIDT, M. S., NIELSEN, A. T. & BOISEN, A. 2017. Surface Enhanced Raman Scattering for Quantification of p-Coumaric Acid Produced by Escherichia coli. *Analytical Chemistry*, 89, 3981-3987.
- MORRIS, D. R. & HAGER, L. P. 1966. Chloroperoxidase. I. Isolation and properties of the crystalline glycoprotein. *The Journal of Biological Chemistry*, 241, 1763-8.
- POOR, C. B., ANDORFER, M. C. & LEWIS, J. C. 2014. Improving the Stability and Catalyst Lifetime of the Halogenase RebH Using Directed Evolution. *Chembiochem : a European Journal of Chemical Biology*, 15, 1286-1289.
- RODRIGUES PEREIRA, E., BELIN, L., SANCELME, M., PRUDHOMME, M., OLLIER, M., RAPP, M., SEVÈRE, D., RIOU, J.-F., FABBRO, D. & MEYER, T. 1996. Structure-Activity Relationships in a Series of Substituted Indolocarbazoles: Topoisomerase I and Protein Kinase C Inhibition and Antitumoral and Antimicrobial Properties. *Journal of Medicinal Chemistry*, 39, 4471-4477.
- SCHNEPEL, C., MINGES, H., FRESE, M. & SEWALD, N. 2016. A High-Throughput Fluorescence Assay to Determine the Activity of Tryptophan Halogenases. *Angewandte Chemie International Edition*, 55, 14159-14163.
- SHAKILA, G., PERIANDY, S. & RAMALINGAM, S. 2011. Molecular Structure and Vibrational Analysis of 1-Bromo-2-Chlorobenzene Using ab initio HF and Density Functional Theory (B3LYP) Calculations. *Journal of Atomic, Molecular, and Optical Physics*, 2011, 10.
- SHEPHERD, S. A., KARTHIKEYAN, C., LATHAM, J., STRUCK, A.-W., THOMPSON, M. L., MENON, B. R. K., STYLES, M. Q., LEVY, C., LEYS, D. & MICKLEFIELD, J. 2015. Extending the biocatalytic scope of regiocomplementary flavin-dependent halogenase enzymes. *Chemical Science*, 6, 3454-3460.
- VAILLANCOURT, F. H., YIN, J. & WALSH, C. T. 2005. SyrB2 in syringomycin E biosynthesis is a nonheme FeII α -ketoglutarate- and O₂-dependent halogenase. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 10111-10116.
- VILTER, H. 1984. Peroxidases from phaeophyceae: A vanadium(V)-dependent peroxidase from Ascophyllum nodosum. *Phytochemistry*, 23, 1387-1390.
- WESTLEY, C., XU, Y., CARNELL, A. J., TURNER, N. J. & GOODACRE, R. 2016. Label-Free Surface Enhanced Raman Scattering Approach for High-Throughput Screening of Biocatalysts. *Analytical Chemistry*, 88, 5898-5903.
- XU, J., BIRKE, R. L. & LOMBARDI, J. R. 1987. Surface-enhanced Raman spectroscopy from flavins adsorbed on a silver electrode: observation of the unstable semiquinone intermediate. *Journal of the American Chemical Society*, 109, 5645-5649.
- YEH, E., COLE, L. J., BARR, E. W., BOLLINGER, J. M., JR., BALLOU, D. P. & WALSH, C. T. 2006. Flavin redox chemistry precedes substrate chlorination during the reaction of the flavin-dependent halogenase RebH. *Biochemistry*, 45, 7904-12.
- ZENG, J. & ZHAN, J. 2011. Characterization of a tryptophan 6-halogenase from Streptomyces toxytricini. *Biotechnology Letters*, 33, 1607-1613.
- ZHU, X., DE LAURENTIS, W., LEANG, K., HERRMANN, J., IHLEFELD, K., VAN PEE, K. H. & NAISMITH, J. H. 2009. Structural insights into regioselectivity in the enzymatic chlorination of tryptophan. *Journal of Molecular Biology*, 391, 74-85.

2.8 SUPPORTING INFORMATION

SUPPLEMENTARY METHODS

Reagents and Materials.

All chemical reagents were of analytical grade and used with no additional purification. Chemicals and solvents were purchased from Acros Organic (New Jersey, USA), Alfa Aesar (Heysham, UK), Fischer Scientific (Loughborough, UK), Formedium (Hunstanton, UK) or Sigma Aldrich Ltd (Dorset, UK).

Synthesis of Silver Nanoparticles.

All glassware used within the synthesis of metal nanoparticles was cleaned using aqua regia; HNO₃: HCl (1:3) v/v, to ensure the removal of any residual metals. Glassware was then thoroughly rinsed with deionised water.

A modified Lee and Meisel (Lee and Meisel, 1982) procedure was used to produce borohydride-reduced silver colloid (**BRSC**). AgNO₃ (1 x 10⁻³ M, 25 mL) was added dropwise to an ice-cold solution of vigorously stirred NaBH₄ (2 x 10⁻³ M, 75 mL) to form a yellow colloid of Ag nanoparticles. Vigorous stirring was continued for a further 1 h to allow the colloidal suspension to return to room temperature. After 10 min of stirring, the colloidal solution changed to a dark black-grey colour, then returning to a yellow solution after 1 h.

Citrate-reduced silver colloid (**CRSC**) was prepared according to the method of Lee and Meisel (Lee and Meisel, 1982). AgNO₃ (1.1 x 10⁻³ M, 500 mL) was heated to boiling and trisodium citrate (1 % weight/volume, 10 mL) added dropwise with vigorous stirring. This solution was left to boil for 1 h then left to return to room temperature. A cloudy green-grey suspension was formed.

SERS Sample Preparation.

Tryptophan halogenases: SERS analysis using BRSC: 350 µL of the silver colloid was pipetted into an 8 mm glass vial, followed by 50 µL of the sample (total sample concentration 2.5 x 10⁻⁵ M in HEPES buffer (1 x 10⁻² M, pH 7.0)). The vial was capped and vortexed for 7 s and allowed to aggregate for 30 s before recording a 20 s SERS spectrum (see 'SERS Optimisation' for further details). A colour change from yellow to purple indicated successful aggregation. The addition of halogenase substrates (*i.e.* tryptophan or anthranilic acid) was immediately

followed by a purple colour change signifying self-aggregation, therefore aggregating agents were omitted for use with this colloid.

Phenolic halogenases: SERS analysis using CRSC: 200 μL of the silver colloid was pipetted into an 8 mm glass vial, followed by 200 μL of the sample (total sample concentration 2×10^{-5} M in potassium phosphate buffer (1×10^{-2} M, pH 7.0)), and then 50 μL of KNO_3 (1×10^{-1} M). The vial was capped and vortexed for 7 s and allowed to aggregate for 30 s before collecting a 10 s SERS spectrum (see 'SERS Optimisation' for further details). Upon addition of the aggregating agent, a darker grey colour change was observed.

INSTRUMENTATION AND DATA PROCESSING

Raman Spectrometers.

SERS spectra were recorded using a DeltaNu Advantage 200A portable spectrometer (DeltaNu, Laramie, WY, USA) equipped with either a HeNe 633 nm laser providing ~ 3 mW of power on the sample, or a HeNe 785 nm laser with ~ 60 mW of power at the sample. Toluene was pipetted into 8 mm glass vials in order to calibrate the systems, allowing the ideal distance between the laser and the point of sampling to be established. 633 nm SERS spectra were acquired for 20 s over a range of $200 - 3400 \text{ cm}^{-1}$; the spectral resolution was 10 cm^{-1} . 785 nm SERS spectra were acquired for 10 s over a range of $200 - 2000 \text{ cm}^{-1}$. Samples for SERS analysis were placed in an 8 mm glass vial, vortexed for 7 s and left for 30 s to aggregate, before subjected to laser irradiation once placed in the cell holder.

A Renishaw 2000 Raman microscope (Renishaw, Wotton-under-Edge, Gloucestershire, UK) was used to collect the Raman spectra of solid samples. Two different excitation wavelengths, 633 nm and 785 nm, were used. The power at the sampling point was between $2 - 4$ mW and the spectral resolution was 6 cm^{-1} . The instrument collected a static spectrum centred at 521 cm^{-1} for 1 s which was calibrated with a silicon wafer focused beneath the 50x objective. GRAMS WIRE software package (Galactic Industries Corp., 395 Main St., Salem, NH) operating in Windows 98 was used for data collection. Four repeats were taken and averages are shown, extended scans between 100 and 2500 cm^{-1} were acquired.

Raman spectra processing was performed using Matlab software version R2008a (The MathWorks, Natick, MA, USA) using scripts written in-house and available via GitHub (<http://www.biospec.net/resources/>). SERS spectra were either raw, baseline corrected or normalised (using DataNorm script) (as stated in the text).

UV-Vis Spectrophotometry.

Borohydride-reduced and citrate-reduced silver nanoparticles were characterised using UV-Vis spectroscopy to determine the surface plasmon resonance band λ_{max} of the nanoparticles. Colloid preparations were diluted 1:6 v/v with deionised water, to ensure that the UV-Vis absorbance did not exceed 2. 1 mL of the diluted nanoparticle solution was pipetted into a polystyrene cuvette and placed into the sample holder of a Thermo Biomate 5 (Thermo Fisher Scientific Inc., Massachusetts, USA). Absorbance spectra were collected over a range of 200-800 nm.

SEM.

SEM images were collected using a Field Emission Scanning Electron Microscope with an ultimate beam size of 1 nm and down to 100 eV electron energy (Carl-Zeiss-Straße 56, 73447 Oberkochen, Germany). A silicon surface was used to image the nanoparticles as silicon helps disperse negative charges and generates a greater contrast in comparison to glass, as a result higher resolution images can be obtained. To image BRSC/CRSC nanoparticles, 10 μL was spotted onto a silicon slide and left to dry overnight (Figure S2.10 and S2.11), respectively).

HPLC.

Halogenase assays were analysed by HPLC on an Agilent Technologies 1260 Infinity HPLC with an Agilent Zorbax Eclipse Plus C18 Rapid Resolution 4.6 x 100 mm 3.5 μm column. For tryptophan, 3-indolepropionic acid (33 % isocratic gradient $\text{H}_2\text{O}/\text{ACN} + 0.1\%$ FA) and *N*-phenylanthranilic acid (gradient elution involved: 5 – 95 % $\text{H}_2\text{O}/\text{ACN} + 0.1\%$ FA), absorbance was measured at 280 nm over 12.5 min. All other substrates and products absorbance were measured at 254 nm, with a 5 min gradient 5 – 75 % $\text{H}_2\text{O}/\text{ACN} + 0.1\%$ FA over 12.5 min. Flow rates were kept constant at 1 mL min^{-1} . The peak areas of analytes of interest were integrated using ChemStation (Agilent 1100 series) and reported % conversions were adjusted in accordance with HPLC calibrations.

Purification of halogenated products was performed on a semi-preparative scale using on Varian Prostar 210. 500 μL of solution containing crude reaction mixture was injected onto a Phenomenex Gemini® C18 HPLC column (5 μ packing, 110 Å, 250 x 10 mm). A gradient of 5 - 75 % $\text{H}_2\text{O}/\text{ACN} + 0.1\%$ FA was used over 35 min, using a 5 mL min^{-1} flow rate. UV absorbance was detected at 280 and 254 nm throughout.

NMR spectroscopy.

Purified samples were concentrated to dryness using the GeneVac (EZ-2 Series) before the addition of deuterated solvent for NMR analysis. NMR analysis was performed on the Bruker DXP 400 MHz or 800 MHz spectrometer. Water suppression was achieved by a w5 WATERGATE pulse sequence or a zgcppr pulse sequence. Spectra processing was performed using MestReNova software.

HRMS.

Waters LCT Time-of-Flight Mass Spectrometer coupled to a Waters Alliance 2790 LC, performing flow injection of 20 μ L samples into a mobile phase of 50/50 ACN/H₂O 0.1 % FA.

Nanoparticle Characterisation

To characterise the morphology of the nanoparticles, UV-Vis spectroscopy and scanning electron microscopy (SEM) were used. The absorption maximum (λ_{\max}) of the plasmon resonance band measured in UV-Vis spectroscopy provides information on the average particle size and the width of the band (FWHM; full width half maximum) is an indicator of size distribution; a narrow band suggest a more uniform distribution (Haiss *et al.*, 2007). SEM is another characterisation technique used for determining size, shape and distribution of colloidal nanoparticles (Buhr *et al.*, 2009). BRSC was shown to have a λ_{\max} of 398 nm and CRSC 410 nm, both in agreement with the literature (Larmour *et al.*, 2012, Chuang and Chen, 2009). SEM images identified that the size and shape distribution was good amongst both colloidal nanoparticles, and images are shown Figure S2.10 and S2.11. Several magnifications were taken, BRSC nanoparticles are predominantly spherical in shape, roughly $\sim 20 \pm 4$ nm in size, whereas CRSC contained spherical nanoparticles along with rod-shaped nanoparticles and were larger in size $\sim 45 \pm 5$ nm.

Optimisation of STTH Biotransformation

Flavin-dependent halogenases require reduced-FAD (FADH₂) to function which is produced using a flavin reductase (Fre) and NADH. Fre consumes large quantities of NADH which is a relatively expensive cofactor, thus it may restrict the application of halogenases within industry. Consequently, biocatalytic recycling systems are commonly installed to recycle NAD⁺ by regenerating reduced-NAD (NADH), as a route of reducing costs. Alcohol dehydrogenase and glucose dehydrogenase (GDH) are the most popular choice of recycling system, arising from their abundant and low-cost cofactors (an alcohol and glucose, respectively). We chose GDH as the biocatalytic recycling system due to its availability. Optimisation of the biotransformation began by varying the ratios of the three enzymes (SttH, Fre and GDH). Previous optimisation of flavin-dependent halogenases within the Micklefield group helped direct SttH optimisation. A concentration ratio of 10:1:6 x 10⁻⁶ M (SttH:Fre:GDH) was found to be optimal for halogenase activity Figure S2.1. In addition, the concentration of cofactors (FAD, NADH and glucose) were varied to obtain optimal conditions (Figure S2.2). Conversion to product was unaffected by the choice and concentration of chloride source (MgCl₂ or NaCl), therefore the chosen concentration corresponded to previous halogenase optimisation (5 x 10⁻² M).

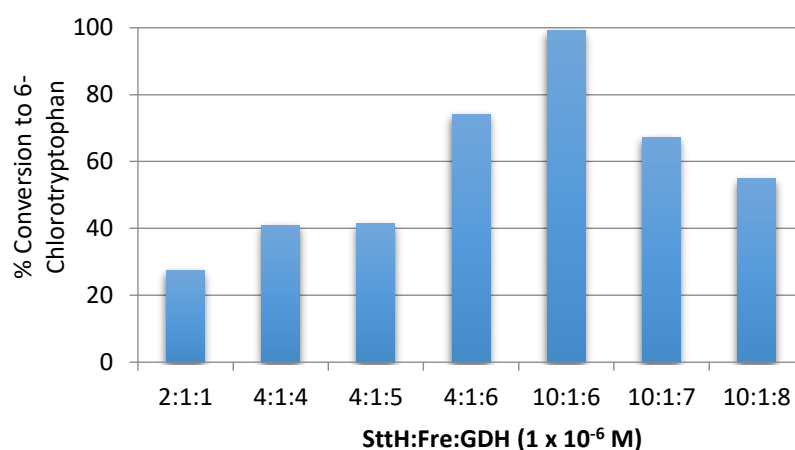


Figure S2.1. Graph representing the variance in percentage conversion arising from changes in protein concentration (1 x 10⁻⁶ M). Results are shown for tryptophan (5 x 10⁻⁴ M) over a 1 h incubation period at 30 °C (single measurement shown).

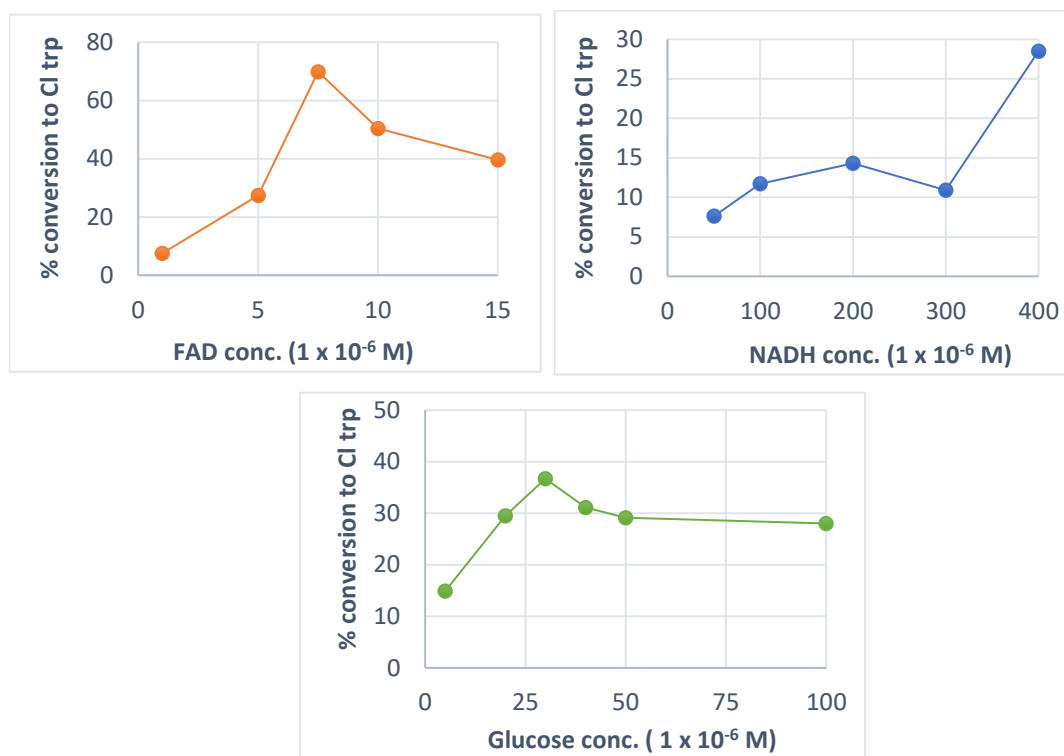


Figure S2.2. Optimisation of SttH with GDH recycling system; optimal conditions were identified as SttH 1×10^{-5} M, Fre 1×10^{-6} M, GDH 6×10^{-6} M, MgCl_2 5×10^{-2} M, FAD 7.5×10^{-6} M, NADH 2×10^{-4} M, glucose 3×10^{-2} M (single measurement shown).

SttH Substrate Scope

To determine regioselectivity and % conversions of SttH catalysed reactions, the following conditions were used: SttH (1×10^{-5} M) was incubated at 30°C with shaking for 1 h with Fre (1×10^{-6} M), GDH (6×10^{-6} M), FAD (7.5×10^{-6} M), NADH (2×10^{-4} M), MgCl_2 (5×10^{-2} M), glucose (2×10^{-2} M) and substrate (5×10^{-4} M) in a total volume of $100 \mu\text{L}$ in potassium phosphate buffer (1×10^{-2} M, pH 7.0). Reactions were stopped by incubating at 95°C for 5 min and the precipitated proteins were removed by centrifugation ($21,000 \times g$ for 6 min) before HPLC analysis. Structures of all substrates tested for SttH activity are shown in Figure S2.3.

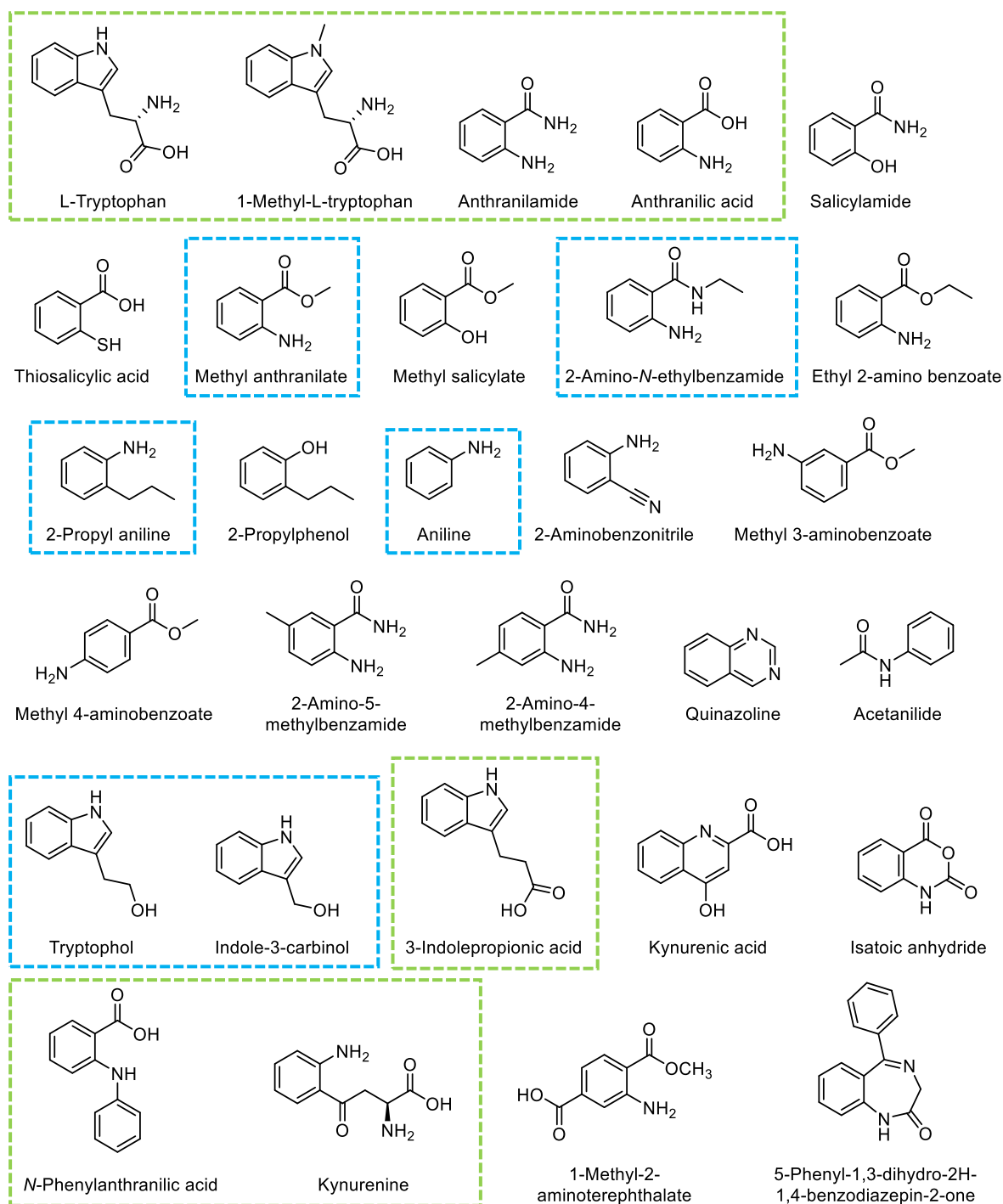


Figure S2.3. Structures of 29 substrates that were tested for activity towards SttH 6-halogenase. Active compounds are highlighted in boxes. Halogenated products confirmed by LRMS only are shown in blue. Halogenated products with full characterisation (^1H , ^{13}C NMR, HRMS, UV) are in green.

SERS Optimisation

Prior to experiments, several parameters were systematically varied to improve SERS profiles of analytes in terms of reproducibility and their enhancement. A significant amount of attention was placed upon the choice and optimisation of SERS active substrates. Numerous colloidal nanoparticles, as well solid-state surfaces, were investigated for their ability to generate SERS responses for the analytes under investigation. Successful SERS responses for tryptophan, anthranilic acid and their halogenated analogs were attainable using borohydride-reduced silver colloid (BRSC). Conversely, substrates used within *in vivo* investigations, chrysin, chlorochrysin and dichlorochrysin, gave a better response with citrate-reduced silver colloid (CRSC). Further optimisation involved studying the effect of colloid volume, analyte concentration, pH, aggregating agent and aggregation time.

Table S2.1. Raman and tentative SERS band assignments for tryptophan (Aliaga *et al.*, 2009, Leyton *et al.*, 2012). Raman of solid tryptophan collected at 633 nm, 60 s acquisition time at 50 % power and 785 nm, 60 s acquisition time at 25 % power.

Observed Raman		Previously reported Raman		Assignment	
633 nm	785 nm	634 nm [Aliaga 2009]	514 nm [Leyton 2012]	[Aliaga 2009]	[Leyton 2012]
1577 (m)	1541 (s)	1556 (m)	1555 (s)		NH ₃ ⁺
1475 (w)	1439 (w)	1450 (w)		δ(NH ₃)	
1444 (m)	1407 (s)	1427 (m)	1426 (s)	δ(COO ⁻)	ν(COO ⁻)
1377 (s)	1339 (vs)	1356 (vs)	1358 (s)	ν(COO ⁻)	ν(COO ⁻)
1358 (m)	1321 (m)	1340 (m)	1340		CH ₂
	1295 (vw)	1312 (w)	1312		CH ₂
1269 (vw)	1232 (w)	1255 (vw)		ν(CH), ν(C-N)	
1252 (m)	1213 (m)	1229 (w)	1232	δ(CH)	δ(CH)
1226 (vw)	1187 (vw)	1208 (w)	1208	δ(CH)	δ(CH)
1176 (w)	1141 (vw)	1165 (w)	1164	δ(CH)	δ(CH)
1135 (m)	1098 (w)	1118 (m)	1119	δ(NH)	δ(NH)
1026 (vs)	998 (vs)	1010 (s)	1007 (s)		δ(NH)
980 (vw)	943 (vww)	961 (w)		ν(C-COO ⁻)	
944 (w)	904 (w)	921 (w)		ν(C-COO ⁻)	
891 (s)	852 (m)	865 (m)	866	δ(NH)	δ(NH)
773 (vs)	733 (vs)	743 (m)	744		
612 (w)	570 (w)	594 (m)	595	δ(NH)	δ(NH)

vs – very strong; s – strong; m – medium; w – weak; vw – very weak; vww – very, very, weak.
ν – stretching; δ – twisting;

Table S2.2. Raman and tentative SERS band assignments for anthranilic acid (Govindarajan *et al.*, 2011). Raman of solid anthranilic acid collected at 633 nm, 5 s acquisition time at 1 % power and 785 nm, 30 s acquisition time and 50 % power.

Observed Raman		Previously reported 1064 nm [Govindarajan 2011]	Assignment [Govindarajan 2011]
633 nm	785 nm		
1652 (m)	1610 (m)	1626 (s)	α (NH ₂)
		1605 (s)	ν (C=C)
1590 (m)	1548 (m)	1585 (w)	ν (C=C)
1375 (m)	1330 (s)	1347 (vs)	ν (C-C)
1268 (m)	1221 (s)	1240 (vs)	β (NH)
		1175 (m)	β (NH)
1192 (w)	1145 (m)	1170 (s)	ν (C-NH ₂)
1058 (vww)	1042 (vw)	1052 (w)	ν (C-COOH)
	1009 (w)	1015 (w)	β (C=O)
886 (vw)	835 (vw)	840 (w)	φ (CH)
		810 (vw)	φ (CH)
799 (vs)	747(vs)		
601 (w)	544 (m)	565 (s)	β (C-COOH)
465 (w)	406 (s)	430 (s)	φ (C-NH ₂)
		385 (w)	φ (C-COOH)
290 (vww)	261 (w)	275 (w)	ω (NH ₂)
	229 (w)	255 (w)	δ (NH ₂)

vs – very strong; s – strong; m – medium; w – weak; vw – very weak; vww – very, very, weak.
 ν – stretching; β – in-plane bending; φ – out-of-plane bending; α – scissoring; δ – twisting; ω – wagging;

Identification of Anthranilic acid Regioisomers

Chemometric analysis (PCA scores plots and loadings) was employed to determine if halogenated regioisomers of anthranilic acid are distinguishable alongside their dehalogenated counterpart. The PCA scores plot (Figure S2.4) shows clear separation between chrysin and its chlorinated regioisomers. The loadings plot (Figure S2.5) demonstrates separation along principal component (PC) 1 predominantly arising from the reduction in peak intensity seen at 965 and 1584 cm⁻¹ when chlorination is present at C-3 and C-4 of anthranilic acid. Chlorination at C-5 retains peak intensity at these band vibrations, thus requiring PC 2 for further separation. PC 2 separation is mostly attributable to peaks

corresponding to either anthranilic acid or 5-chloroanthranilic acid as more noticeable spectral changes are seen.

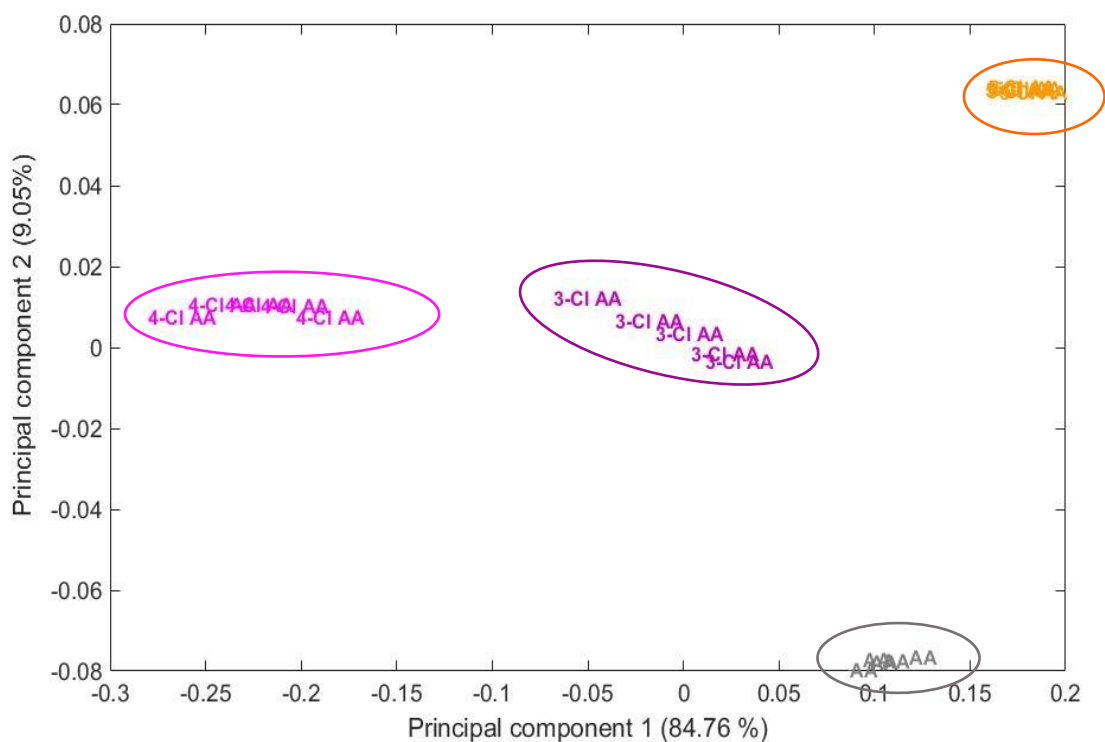


Figure S2.4. PCA scores plot depicting separation of anthranilic acid (AA, grey) 3-chloroanthranilic acid (3-Cl AA, purple), 4-chloroanthranilic acid (4-Cl AA, pink) and 5-chloroanthranilic acid (5-Cl AA, orange). Normalisation of data was performed prior to PCA. Eclipses are a guide to the eye and have no statistical significance. Total explained variance (TEV) is labelled on the axes as a percentage.

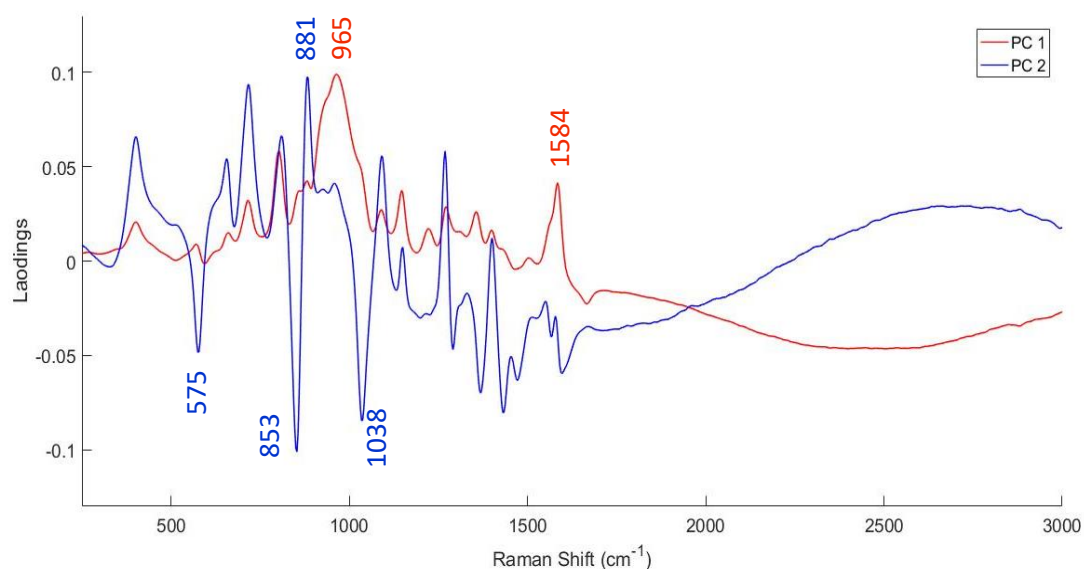


Figure S2.5. Loadings plot identifying key peaks enabling separation of regioisomers along the PC 1 and PC 2 axis of the PCA scores plot shown in Figure S2.4.

Halogenase Whole-Cell Optimisation

A pACYCDuet-1 vector containing both *Fre* and *GDH* genes was previously cloned by members of the Micklefield group (see section 'Protein Expression and Purification'). This plasmid, along with a second plasmid (pET 28b(+)) containing genes for either *PyrH* or *RadH* were transformed into two different cell lines, generating four different cell variants (see Figure S2.6). Having gained experience working with SttH (C-6), it was apparent that its activity and stability in comparison to other flavin-dependent halogenases (*PyrH* and *RadH*) was much lower. Therefore, SttH was omitted from further investigations. Halogenase activity *in vivo* was only observed with the *RadH-Fre-GDH* construct in Arctic Express (DE3), so this was taken on for further tests.

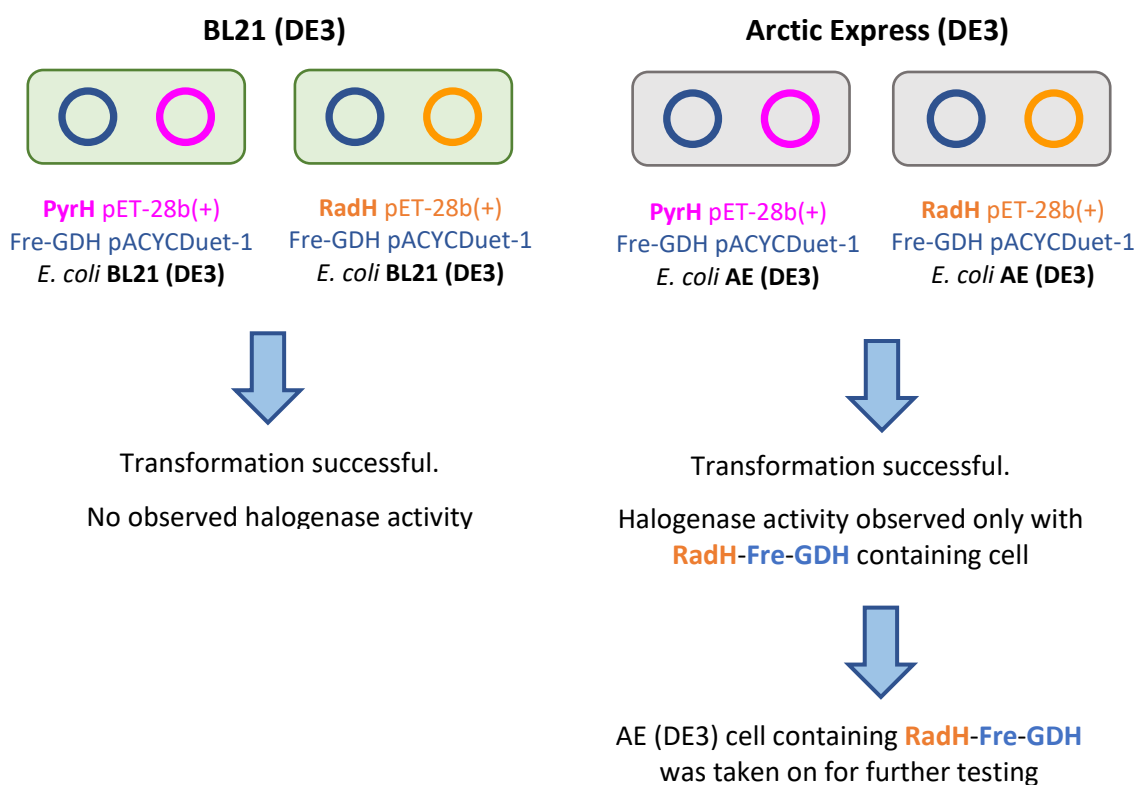


Figure S2.6. A flow diagram summarising the process undertaken to develop flavin-dependent halogenases for use in whole-cell biotransformations.

Distinguishing between Chrysin, Chlorochrysin and Dichlorochrysin

Optimisation of SERS conditions were re-performed for *RadH* substrates (flavonoids). These conditions were found to be entirely different to those that were used to investigate tryptophan-halogenase substrates tryptophan and anthranilic acid. We found that 785 nm laser irradiation gave an enhanced response (over 633 nm), using citrate-reduced silver

colloid (CRSC, a more reproducible and stable colloid) along with the introduction of an aggregating agent (KNO₃) which improved SERS responses. In Figure 2.9 some similarities in the SERS responses of chrysin and its singly and doubly chlorinated analogs are observed: mainly strong vibrations at 1000 cm⁻¹ (corresponding to ring stretching) and 1246 cm⁻¹ (C3H in-plane bend). Tentative SERS band assignments are shown in Table S2.3 (Corredor *et al.*, 2009). Many bands which induce spectral changes upon chlorination, therefore enabling discrimination, occur within the region 580-836 cm⁻¹. These observations are in agreement with the literature (Shakila *et al.*, 2011). PCA scores and loadings plots (Figure S2.7 and S2.8, respectively) clearly depict these spectral differences.

Table S2.3. SERS and tentative band assignments of chrysin (Corredor *et al.*, 2009). Conditions used: CRSC, 10 s, acquisition, phosphate buffer (pH 7.0), chrysin (2 x 10⁻⁵ M) and 50 μL KNO₃ (1 x 10⁻¹ M).

Observed SERS 785 nm	Previously reported SERS 785 nm [Corredor 2009]	Assignment [Corredor 2009]
429 (w)	429 (w)	Ring C-C
505 (vw)	504 (w)	Ring C-C
534 (vw)	534	
557 (vw)	556 (w)	Ring C-C
620 (s)	617 (m)	Ring C-C
641 (s)	642 (w)	Ring C-C
712 (w)	713 (w)	Ring C-C
745 (m)	743 (w)	
837 (br)	844 (w)	Ring CH
1000 (s)	1001 (m)	Ring trigonal str
1035 (w)	1041 (w)	CH bend
1094 (vw)	1099 (w)	COC; CH bend
1115 (vw)	1118 (w)	CH bend
1162 (w)	1165 (w)	Ring CH bend
1190 (w)	1193 (w)	5OH, 7OH bend
1246 (s)	1248 (s)	C3H bend
1310 (w)	1316 (w)	CH bend
1336 (sh)	1338	5OH, 7OH bend; Ring CC
1481 (vw)	1480 (m)	Ring CH bend
1560 (vw)	1569 (m)	Ring quinoid str
1599 (w)	1603 (s)	C=O str; Ring quinoid str
1632 (w)	1633 (s)	C=O str

vs – very strong; s – strong; m – medium; w – weak; vw – very weak; sh – shoulder; br - broad.
str – stretch.

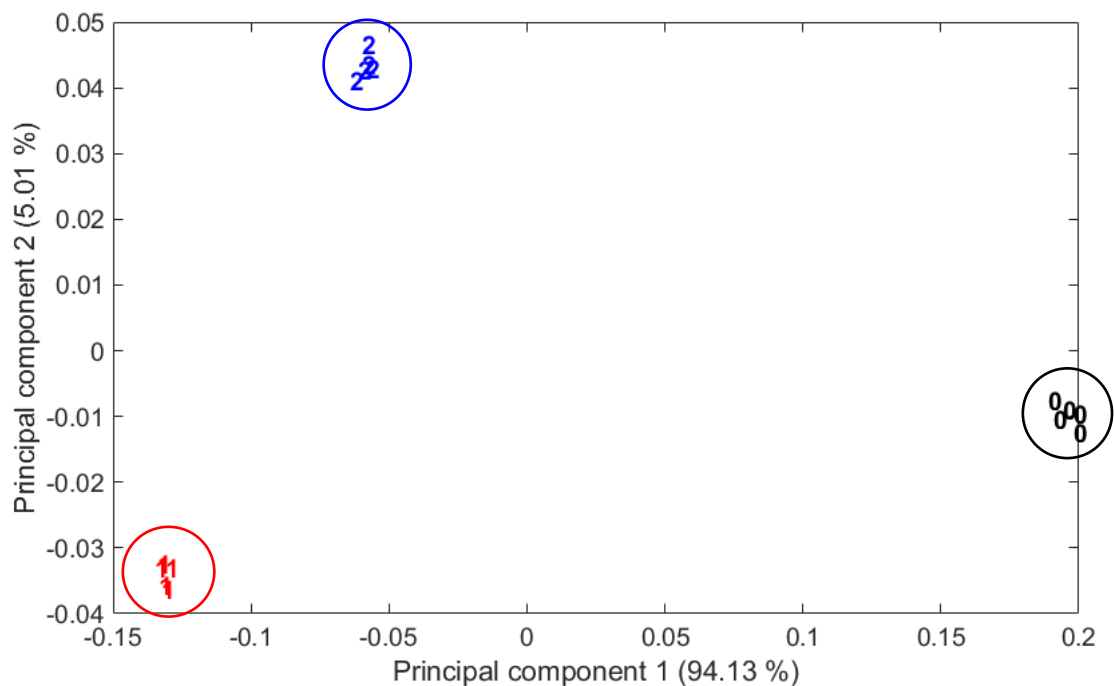


Figure S2.7. PCA scores plot representing separation of chrysin (circled in black), chlorochrysin (circled in red) and dichlorochrysin (circled in blue). Normalisation of data was performed prior to PCA. Eclipses are a guide to the eye and have no statistical significance. Total explained variance (TEV) is labelled on the axes as a percentage.

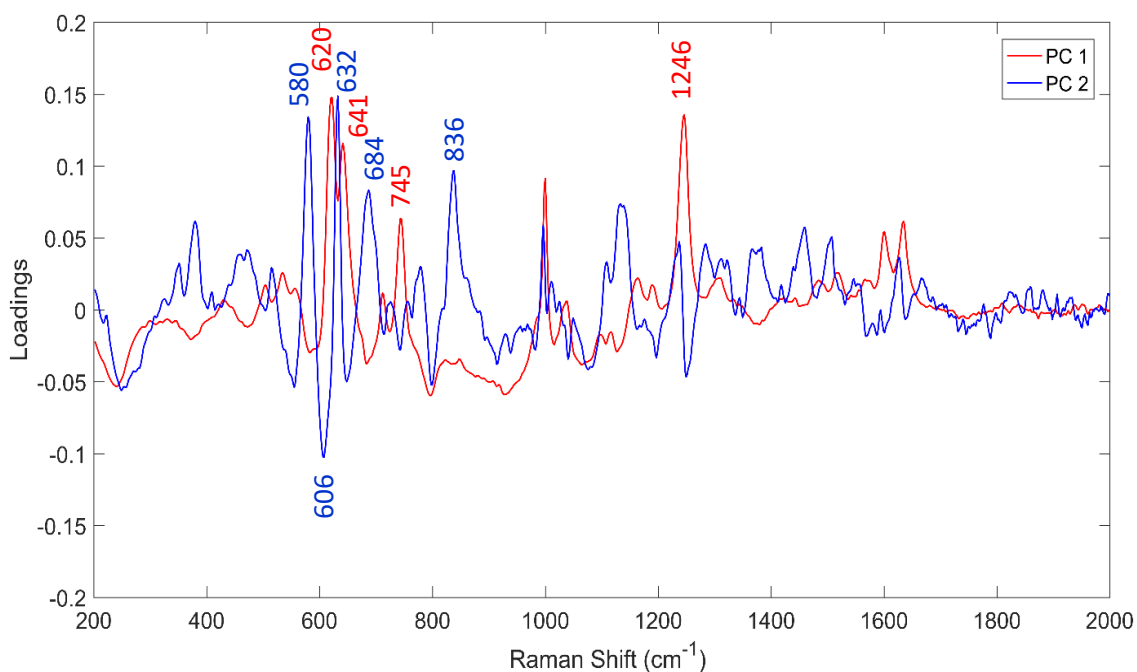


Figure S2.8. Loadings plot recognising key peaks enabling separation along the PC 1 and PC 2 axis of the PCA scores plot shown in Figure S2.7.

Whole-Cell SERS Observations

Several methods of quenching the whole-cell biotransformation were investigated, including those that did not disrupt the cell membrane leading to lysis (*i.e.* centrifugation and high molecular weight cut-off (MWCO) filters to remove bacterial content). Of these quenching methods, the application of an organic solvent (EtOH/MeOH) proved most successful, arising from the poor solubility of chrysin and its halogenated products, and that a significant amount of starting material and product(s) were retained within the cells (reducing total concentration). Consequently, it was important to determine if the supernatant of the lysed *E. coli* cells would produce bands which coincided with the bands of interest. In Figure S2.9, the SERS response of the quenched bacterial cells is overlaid with chrysin and chlorochrysin. It is evident that many bands from the cell supernatant coincide with those of interest, especially those between 645 – 741 cm^{-1} . Bands at 1000 and 1246 cm^{-1} did not have any overlapping bands, but as these are present in both chrysin and chlorochrysin, they cannot be used to determine halogenase activity. Much weaker bands, characteristic of chrysin: 1035, 1162, 1190 and 1560 cm^{-1} , and chlorochrysin, 684 and 797 cm^{-1} , did not overlap with the *E. coli* supernatant control, indicating potential characteristic peaks for discrimination and quantification.

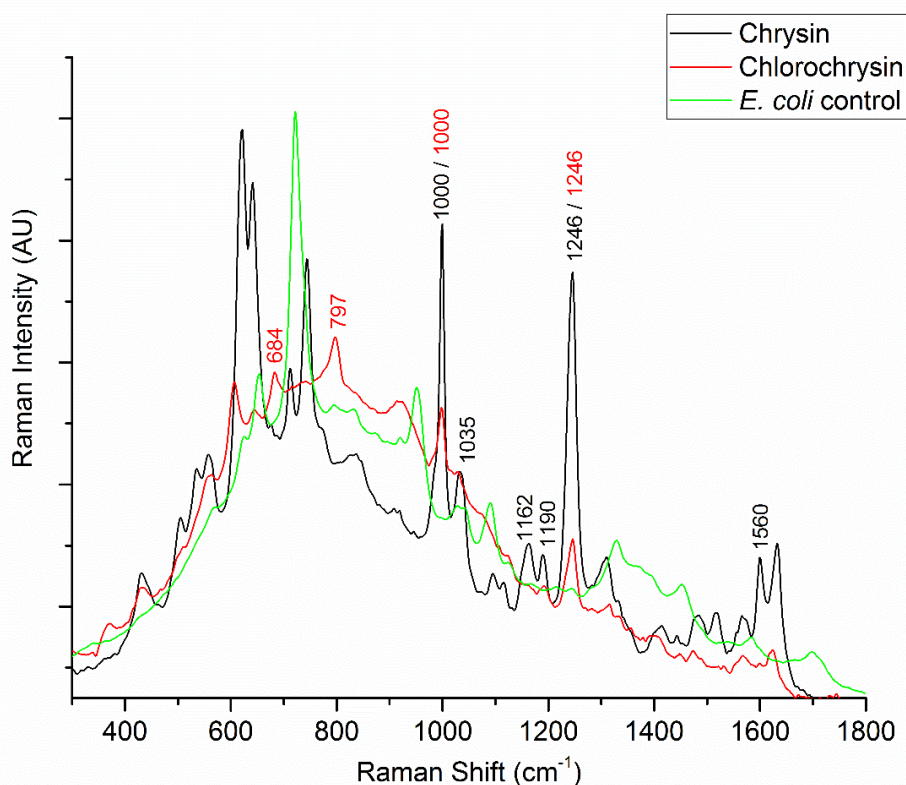


Figure S2.9. Overlaid SERS spectra of chrysin (black, 2×10^{-5} M), chlorochrysin (red, 2×10^{-5} M) and the control sample of *E. coli* supernatant (green). Characteristic peaks have been labelled. Data have been normalised (see data processing for details).

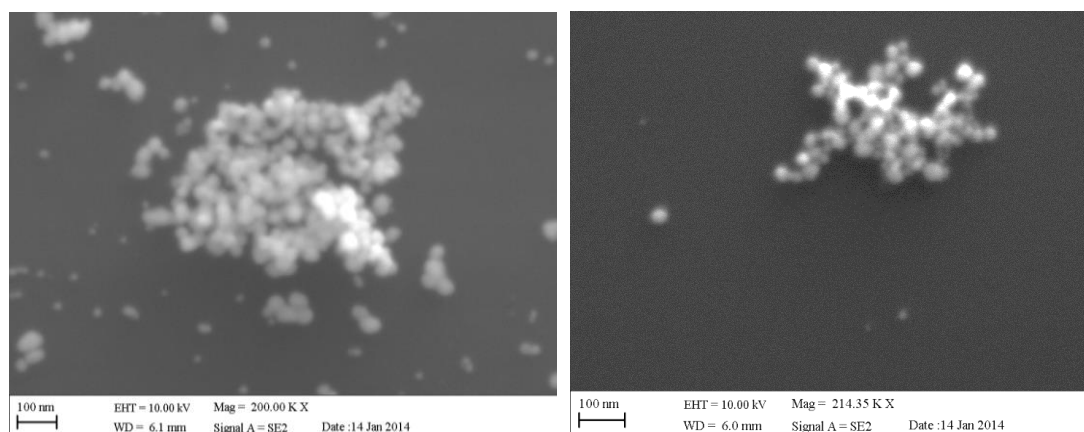


Figure S2.10. SEM images of borohydride-reduced silver colloid on silicon. Left, depicts images of 200x magnification and a scale bar of 100 nm. Right, 214x magnification and 100 nm scale bar.

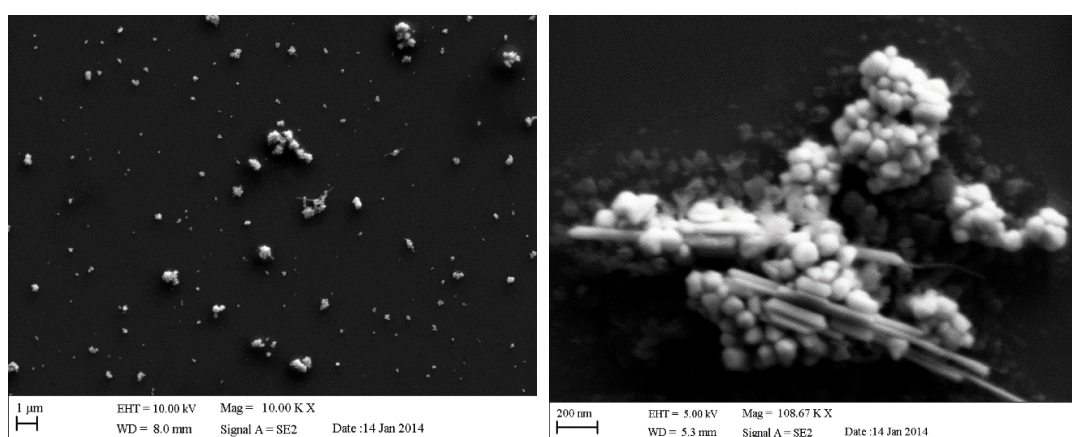


Figure S2.11. SEM images of citrate-reduced silver colloid on silicon. Left, depicts images of 10x magnification and a scale bar of 1 μ m. Right, 109x magnification and 200 nm scale bar.

PROTEIN EXPRESSION AND PURIFICATION

Cloning

The cloning of *SttH* and *PyrH* into pET 28a(+), *RadH* into pET 28b(+), and *Fre* into pET 45b(+) has previously been reported by our group (Shepherd *et al.*, 2015). A construct containing *GDH* in pET 21b(+) was generously provided by the Scrutton Group (University of Manchester). A pACYC Duet-1 vector containing both *Fre* and *GDH* genes was cloned by group member, Dr Binuraj Menon. *SttH* and *PyrH* plasmids were transformed into competent *E. coli* Arctic Express (DE3) for protein expression using kanamycin (50 μ g/mL) for selection. *Fre* and *GDH* plasmids were transformed into competent *E. coli* BL21 (DE3) using ampicillin (50 μ g/mL). pACYC Duet-1 vector containing *Fre* and *GDH*, along with *PyrH* pET 28a(+) was

transformed into both *E. coli* BL21 (DE3) and Arctic Express (DE3) using chloramphenicol (25 µg/mL) and kanamycin (50 µg/mL). A second preparation included pACYC Duet-1 vector containing *Fre* and *GDH*, along with *RadH* pET 28b(+) transformed into both competent *E. coli* BL21 (DE3) and Arctic Express (DE3) using chloramphenicol (25 µg/mL) and kanamycin (50 µg/mL).

Transformation

Plasmids were transformed into *E. coli* competent cells for protein expression using a standard heat shock protocol. First, competent cells were incubated on ice for 30 min with the plasmid. The cells were heat shocked at 42 °C for 45 s before incubating on ice for a further 5 min. LB (300 µL, RT) was then added to the cells, which were incubated at 37 °C for 1 h before plating aseptically onto LB agar with the suitable selection antibiotic(s) and overnight incubation at 37 °C.

Ligation mixtures were transformed into *E. coli* DH5α competent cells for DNA amplification using the heat shock method stated above. After overnight incubation at 37 °C, single colonies were selected aseptically to verify the presence of the plasmid. The colony was added to LB (3 mL) with the appropriate antibiotic(s) and incubated overnight at 37 °C. Following the *QIAprep Spin Miniprep* Kit Protocol, the plasmid DNA was isolated and purified before being sent for DNA sequencing. If the correct plasmid with the ligated insert was present, the plasmid was then transformed into expression strains.

Protein Expression

Fre and GDH. *Fre* BL21 (DE3)/*GDH* BL21 (DE3) single colonies of transformant were picked from LB agar plates and inoculated in LB medium (10 mL) containing ampicillin (50 µg/mL) and grown overnight at 37 °C with shaking. The resultant culture was diluted 100-fold in LB with ampicillin (50 µg/mL) and grown at 37 °C with shaking until reaching an $OD_{600} = 0.4 - 0.6$. IPTG (5×10^{-4} M) was then added to the culture for induction and grown at 30 °C with shaking for 4 h. After this time, cells were centrifuged at 4000 xg , 4 °C for 10 min to pellet the cells. Pelleted cells were stored at 4 °C until purification.

SttH and PryH. *SttH* pET 28a(+) was transformed into numerous competent *E. coli* expression strains to identify the most suitable for producing soluble protein, Arctic Express (DE3) was identified as the most suitable (see Figure S2.12). Single colonies of *SttH* AE (DE3)/*PryH* AE (DE3) transformants were selected from the incubated LB agar plates. The colony was inoculated in LB medium (10 mL) containing kanamycin (50 µg/mL) and grown overnight at 37°C with shaking. The resultant culture was diluted 100-fold in LB, excluding additional

antibiotic addition, then incubated at 30 °C with shaking until reaching an OD₆₀₀ = 0.4 – 0.6. The culture was cold shocked at 4 °C for 20 min, followed by induction with IPTG (1 x 10⁻⁴ M). The culture was grown overnight at 15 °C with shaking, after this time, cells were pelleted using centrifugation at 4000 xg, 4 °C for 10 min prior to purification.

Fre-GDH-PyrH. Single colonies of Fre-GDH-PyrH BL21 (DE3) transformants were selected from the incubated LB agar plates. The colony was inoculated in LB medium (10 mL) containing chloramphenicol (25 µg/mL), kanamycin (50 µg/mL) and grown overnight at 37 °C with shaking. The resultant culture was diluted 100-fold in Auto Induction (AI) media and grown for 3 days at 30 °C with shaking. After this time, cells were pelleted using centrifugation at 4000 xg, 4 °C for 10 min, then resuspended in potassium phosphate buffer (1 x 10⁻² M, pH 7.0) prior to use.

Fre-GDH-RadH. Single colonies of Fre-GDH-RadH BL21 (DE3) transformants were selected from the incubated LB agar plates. The colony was inoculated in LB medium (10 mL) containing chloramphenicol (25 µg/mL) and kanamycin (50 µg/mL) and grown overnight at 37 °C with shaking. The resultant culture was diluted 100-fold in AI medium and grown for 3 days at 30 °C with shaking. After this time, cells were pelleted using centrifugation at 4000 xg, 4 °C for 10 min, then resuspended in potassium phosphate buffer (1 x 10⁻² M, pH 7.0) prior to use.

Protein Purification Buffers

Buffers containing potassium phosphate buffer (pH 7.0, 5 x 10⁻² M), NaCl (5 x 10⁻¹ M) and imidazole (see Table S2.4) in distilled water were prepared and stored at 4 °C before protein purification.

Table S2.4. Concentration of imidazole within purification buffers.

Protein	Imidazole Wash Concentrations	Imidazole Elution Concentration
Fre/GDH	1 x 10 ⁻² M and 6 x 10 ⁻² M	2.5 x 10 ⁻¹ M
SttH/PyrH	1 x 10 ⁻² M and 6 x 10 ⁻² M	5 x 10 ⁻¹ M

Protein Purification

Cell pellets derived from the *E. coli* protein expression were resuspended in 25 mL of 1 x 10⁻² M imidazole buffer along with protease inhibitor tablets (1 tablet per 50 mL). The cell resuspension was lysed using sonication (10 min, 50 % pulse, 70 % power, 700 W, 4 °C) and

the lysate was clarified by centrifugation (4 °C, 40 min, 10,000 xg). Ni-NTA (Qiagen) was equilibrated with 1 x 10⁻² M imidazole buffer prior to loading with clarified lysate under gravity flow. The resin was then washed with imidazole buffer (3 column volumes, concentration shown in Table 2.6) prior to elution with imidazole elution buffer (5 column volumes, concentration shown in Table 2.6). The eluted protein fraction was subjected to buffer exchange with 1 x 10⁻¹ M potassium phosphate buffer (pH = 7.2) using either dialysis or spin concentration (Vivaspin 20 centricon, 10,000 MWCO). The final protein concentration was determined using Thermo Scientific NanoDrop 2000 spectrophotometer and the protein stored at -20 °C with 10 % glycerol. SDS-PAGE was used to separate proteins according to their size, enabling protein identification and gave an idea of purity, SDS-PAGE gels for each protein can be seen in Figure S2.13 – S2.15.

SDS-PAGE Images of Protein Expression and Purification Protocols

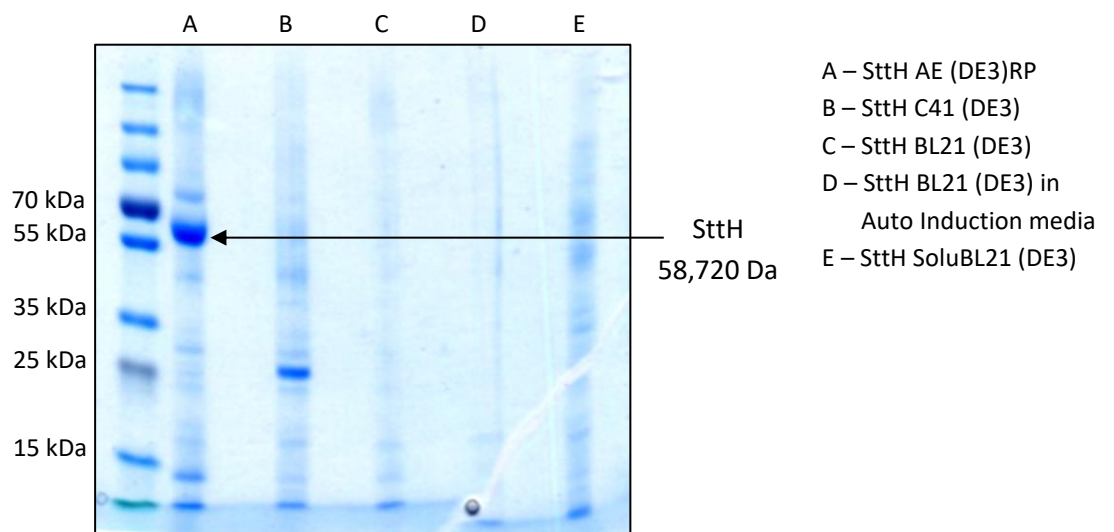


Figure S2.12. SDS-PAGE of SttH protein expression in different *E. coli* competent cells.

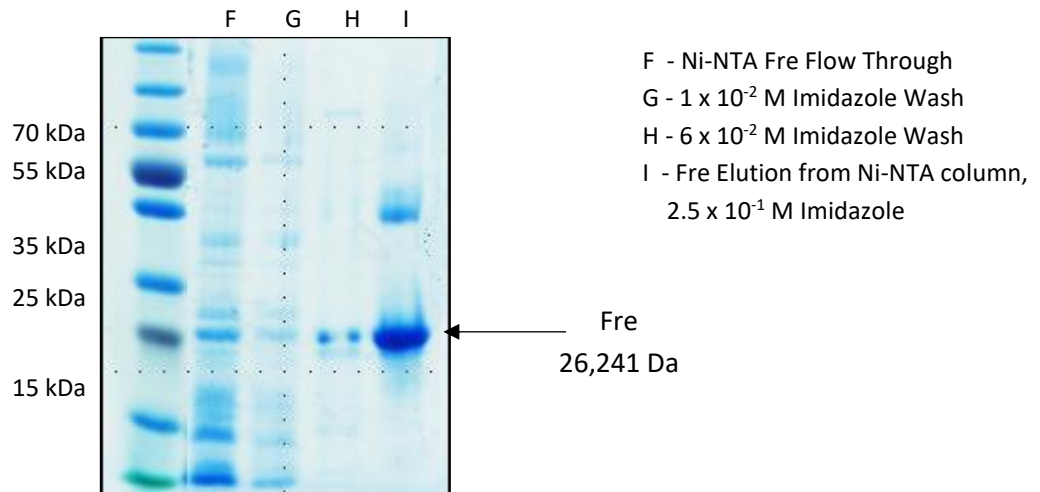


Figure S2.13. SDS-PAGE of Fre protein expression using Ni-NTA column.

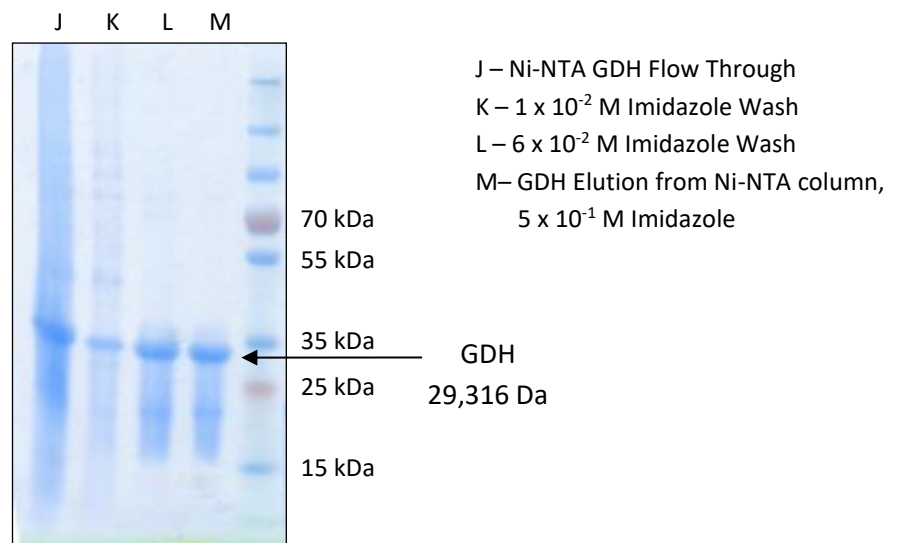


Figure S2.14. SDS-PAGE of GDH protein expression using Ni-NTA column.

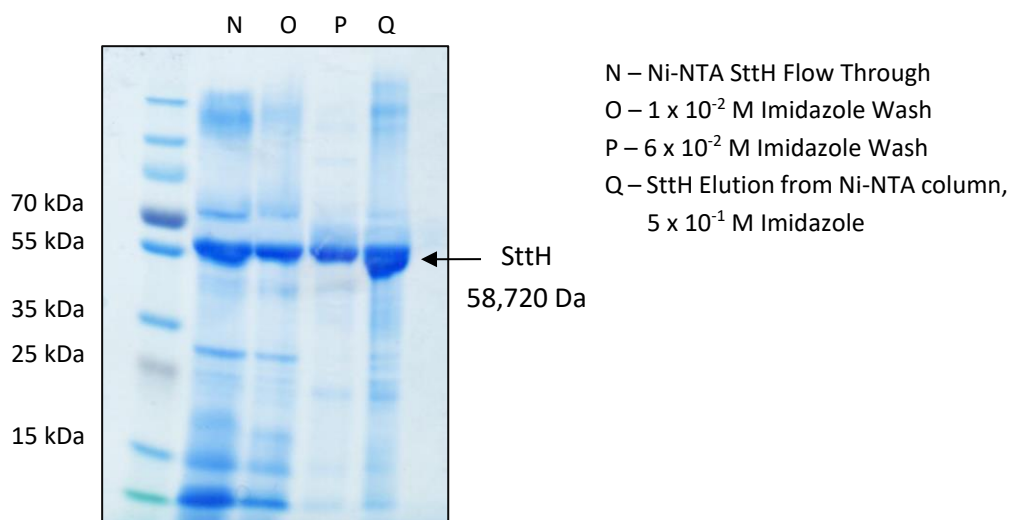
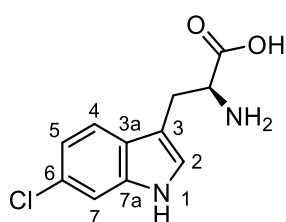


Figure S2.15. SDS-PAGE of SttH protein expression using Ni-NTA column.

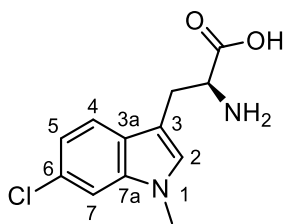
Preparative Halogenase Reactions and NMR Assignments

Larger scale assays were performed to obtain chlorinated products for characterisation using halogenase (1×10^{-5} M), Fre (1×10^{-6} M), GDH (6×10^{-6} M), substrate (2×10^{-3} M, excluding tryptophan), MgCl_2 (5×10^{-2} M), FAD (7.5×10^{-6} M), NADH (2×10^{-4} M), glucose (2×10^{-2} M) in 10 mL potassium phosphate buffer (1×10^{-2} M, pH 7.0). Assays were run at 30 °C with shaking, quenching was achieved by incubation at 95 °C for 5 min and precipitated protein was removed by centrifugation (4 °C, 10 min, 12000 xg) before analysis by HPLC. Low yielding reactions were performed multiple times and the crude mixture combined prior to purification by semi-preparative HPLC.



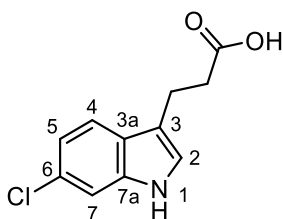
6-Cl tryptophan [1a]

6-Chlorotryptophan [1a] was prepared according to the general procedure described above with tryptophan (5×10^{-4} M) and SttH to give 6-chlorotryptophan (91 % yield). $^1\text{H NMR}$ (400 MHz, Deuterium oxide) 7.51 (1H, d, J 8.6 Hz, Ar-4H), 7.42 (1H, s, Ar-7H), 7.18 (1H, s, Ar-2H), 7.04 (1H, d, J 8.8 Hz, Ar-5H), 4.06 (1H, dd, J 7.8, 4.8 Hz, CH), 3.35 (1H, dd, J 15.8, 5.0 Hz, CH_2'), 3.23 (1H, dd, J 15.6, 7.8 Hz, CH_2''). **LRMS-ESI** $[\text{M}+\text{H}]^+$ expected m/z ^{35}Cl 239.0587 and ^{37}Cl 241.0558, observed m/z 239.0568 and 241.0539 (± 7.95 and 7.88 ppm).



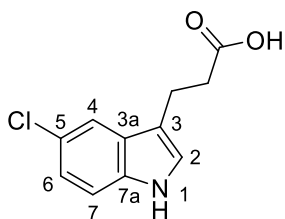
6-Cl-*N*-methyltryptophan [2a]

6-Chloro-*N*-methyltryptophan [2a] was prepared according to the general procedure described above with *N*-methyltryptophan and SttH to give 6-chloro-*N*-methyltryptophan (40 % yield). $^1\text{H NMR}$ (800 MHz, Methanol- d_4) 7.57 (1H, d, J 8.4 Hz, Ar-4H), 7.43 (1H, d, J 1.8 Hz Ar-7H), 7.14 (1H, s, Ar-2H), 7.07 (1H, dd, J 8.4, 1.8 Hz Ar-5H), 4.19 (1H, dd, J 8.0, 5.2 Hz, CH), 3.77 (3H, s, CH_3), 3.34 (1H, dd, J 15.4, 5.2 Hz, CH_2'), 3.30 (1H, dd, J 15.4, 7.8 Hz, CH_2''). $^{13}\text{C NMR}$ (201 MHz, Methanol- d_4) 171.8 (COOH), 139.3 (Ar-7aC) 130.8 (Ar-2C), 129.1 (Ar-3aC), 127.5 (Ar-6C), 120.9 (Ar-5C), 120.6 (Ar-4C), 110.7 (Ar-7C), 107.9 (Ar-3C), 54.6 (CH), 33.0 (CH_3), 27.3 (CH_2). **LRMS-ESI** $[\text{M}+\text{H}]^+$ expected m/z ^{35}Cl 253.0744 and ^{37}Cl 255.0714, observed m/z 253.0727 and 255.0695 (\pm 6.72 and 7.45 ppm).



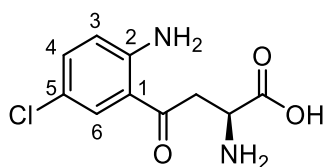
6-Cl-3-indolepropionic acid [3b]

Chlorination of 3-indolepropionic acid was carried out according to the general procedure described above with SttH to give a mixture of 5- and 6-chlorinated products. **6-chloro-3-indolepropionic acid [3b]** product (11 % yield) was isolated by HPLC. $^1\text{H NMR}$ (800 MHz, Methanol- d_4) 7.49 (1H, d, J 8.6 Hz, Ar-4H), 7.31 (1H, d, J 1.8 Hz Ar-7H), 7.07 (1H, s, Ar-2H), 6.97 (1H, dd, J 8.8, 1.8 Hz Ar-5H), 3.03 (2H, dd, J 7.8, 7.6 Hz, CH_2), 2.65 (2H, dd, J 7.8, 7.5 Hz, CH_2). $^{13}\text{C NMR}$ (201 MHz, Methanol- d_4) 176.3 (COOH), 137.0 (Ar-6C) 126.8 (Ar-7aC), 125.8 (Ar-3aC), 122.5 (Ar-2C), 118.9 (Ar-4C), 118.6 (Ar-5C), 114.1 (Ar-3C), 110.5 (Ar-7C), 34.8 (CH_2), 20.3 (CH_2). **LRMS-ESI** $[\text{M}-\text{H}]^-$ expected m/z ^{35}Cl 222.0327 and ^{37}Cl 224.0297, observed m/z 222.0317 and 224.0286 (\pm 4.50 and 4.91 ppm).



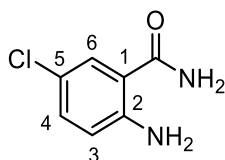
5-Cl-3-indolepropionic acid [3c]

5-Chloro-3-indolepropionic acid [3c] (65% yield) was acquired using PyrH (10 μ M) and was carried out in accordance with the general procedure stated above. $^1\text{H NMR}$ (800 MHz, Methanol- d_4) 7.51 (1H, d, J 2.0 Hz, Ar-4H), 7.28 (1H, d, J 8.6 Hz Ar-7H), 7.10 (1H, s, Ar-2H), 7.04 (1H, dd, J 8.6, 2.0 Hz Ar-6H), 3.01 (2H, dd, J 7.8, 7.6 Hz, CH_2), 2.65 (2H, dd, J 7.8, 7.6 Hz, CH_2). $^{13}\text{C NMR}$ (201 MHz, Methanol- d_4) 175.9 (COOH), 135.1 (Ar-5C) 128.2 (Ar-7aC), 123.9 (Ar-3aC), 123.3 (Ar-2C), 121.0 (Ar-6C), 117.2 (Ar-4C), 113.6 (Ar-3C), 111.9 (Ar-7C), 34.6 (CH_2), 20.2 (CH_2). **LRMS-ESI** $[\text{M}-\text{H}]^-$ expected m/z ^{35}Cl 222.0327 and ^{37}Cl 224.0297, observed m/z 222.0317 and 224.0285 (\pm 4.50 and 5.36 ppm).



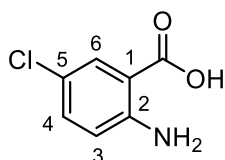
5-Cl-kynurenine [4a]

5-Chlorokynurenine [4a] was prepared according to the general procedure described above with kynurenine and SttH to give 5-chlorokynurenine (46 % yield) (Frese and Sewald, 2015). $^1\text{H NMR}$ (800 MHz, Methanol- d_4) 7.72 (1H, d, J 2.6 Hz, Ar-6H), 7.25 (1H, dd, J 9.0, 2.6 Hz, Ar-4H), 6.78 (1H, d, J 9.0 Hz, Ar-3H), 4.37 (1H, dd, J 7.2, 3.6 Hz, CH), 3.68 (1H, dd, J 18.6, 3.6 Hz, CH_2'), 3.64 (1H, dd, J 18.6, 7.2 Hz, CH_2''). $^{13}\text{C NMR}$ (201 MHz, Methanol- d_4) 198.1 (C=O), 171.9 (COOH), 151.9 (Ar-5C) 135.9 (Ar-4C), 130.5 (Ar-6C), 120.0 (Ar-3C) 118.8 (Ar-1C) 117.5 (Ar-2C) 50.3 (CH) 39.6 (CH_2). **LRMS-ESI** $[\text{M}+\text{H}]^+$ expected m/z ^{35}Cl 243.0536 and ^{37}Cl 245.0507, observed m/z 243.0517 and 245.0487 (\pm 7.82 and 8.16 ppm).



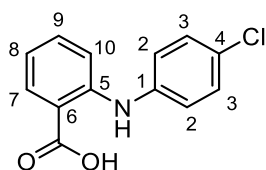
5-Cl-anthranilamide [5a]

5-Chloroanthranilamide [5a] was prepared according to the general procedure described above with anthranilamide and SttH to give 5-chloroanthranilamide (30 % yield). (Frese and Sewald, 2015) $^1\text{H NMR}$ (400 MHz, Methanol- d_4) 7.53 (1H, d, J 2.4 Hz, Ar-6H), 7.15 (1H, dd, J 8.6, 2.4 Hz, Ar-4H), 6.72 (1H, d, J 8.6 Hz, Ar-3H). **LRMS-ESI** $[\text{M}+\text{H}]^+$ expected m/z ^{35}Cl 171.0325 and ^{37}Cl 173.0296, observed m/z 171.0309 and 173.0278 (\pm 9.35 and 10.40 ppm).



5-Cl-anthranilic acid [6a]

5-Chloroanthranilic acid [6a] was prepared according to the general procedure described above with anthranilic acid and SttH to give 5-chloroanthranilic acid (12 % yield) (Frese and Sewald, 2015). $^1\text{H NMR}$ (800 MHz, Methanol- d_4) 7.73 (1H, d, J 2.6 Hz, Ar-6H), 7.17 (1H, dd, J 2.6 Hz, 9.0 Hz, Ar-4H), 6.72 (1H, d, J 9.0 Hz, Ar-3H). **LRMS-ESI** $[\text{M}+\text{H}]^+$ expected m/z ^{35}Cl 172.0165 and ^{37}Cl 174.0136, observed m/z 172.0141 and 174.0110 (\pm 13.95 and 14.94 ppm).



***N*-(4-Cl-phenyl)-anthranilic acid [7a]**

***N*-(4-chlorophenyl)anthranilic acid [7a]** was prepared according to the general procedure described above with *N*-phenylanthranilic acid and SttH to give *N*-(4-chlorophenyl)anthranilic acid (9 % yield) (Frese and Sewald, 2015). $^1\text{H NMR}$ (800 MHz, Methanol- d_4) 7.99 (1H, d, J 8.0 Hz, Ar-H), 7.38 – 7.30 (1H, m, Ar-H) 7.32 (2H, d, J 8.8 Hz, Ar-H), 7.26 – 7.20 (1H, m, Ar-H), 7.22 (2H, d, J 8.8 Hz, Ar-H), 6.76 (1H, t, J 8.0 Hz, Ar-H). **LRMS-ESI** $[\text{M}+\text{H}]^+$ expected m/z ^{35}Cl 248.0478 and ^{37}Cl 250.0449, observed m/z 248.0459 and 250.0428 (\pm 7.66 and 8.40 ppm).

REFERENCES

- ALIAGA, A. E., OSORIO-ROMÁN, I., LEYTON, P., GARRIDO, C., CÁRCAMO, J., CANIULEF, C., CÉLIS, F., DÍAZ F, G., CLAVIJO, E., GÓMEZ-JERIA, J. S. & CAMPOS-VALLETTE, M. M. 2009. Surface-enhanced Raman scattering study of L-tryptophan. *Journal of Raman Spectroscopy*, 40, 164-169.
- BUHR, E., SENFTLEBEN, N., KLEIN, T., BERGMANN, D., GNIESER, D., FRASE, C. G. & BOSSE, H. 2009. Characterization of nanoparticles by scanning electron microscopy in transmission mode. *Measurement Science and Technology*, 20, 084025.
- CHUANG, C.-H. & CHEN, Y.-T. 2009. Raman scattering of L-tryptophan enhanced by surface plasmon of silver nanoparticles: vibrational assignment and structural determination. *Journal of Raman Spectroscopy*, 40, 150-156.
- CORREDOR, C., TESLOVA, T., CAÑAMARES, M. V., CHEN, Z., ZHANG, J., LOMBARDI, J. R. & LEONA, M. 2009. Raman and surface-enhanced Raman spectra of chrysin, apigenin and luteolin. *Vibrational Spectroscopy*, 49, 190-195.
- FRESE, M. & SEWALD, N. 2015. Enzymatic Halogenation of Tryptophan on a Gram Scale. *Angewandte Chemie International Edition*, 54, 298-301.
- GOVINDARAJAN, M., GANASAN, K., PERIANDY, S., MOHAN, S. & TEDLAMELEKOT, F. 2011. Vibrational spectroscopic analysis of 2-bromobenzoic and anthranilic acids: A combined experimental and theoretical study. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 79, 2003-2011.
- HAISS, W., THANH, N. T. K., AVEYARD, J. & FERNIG, D. G. 2007. Determination of Size and Concentration of Gold Nanoparticles from UV-Vis Spectra. *Analytical Chemistry*, 79, 4215-4221.
- LARMOUR, I. A., FAULDS, K. & GRAHAM, D. 2012. SERS activity and stability of the most frequently used silver colloids. *Journal of Raman Spectroscopy*, 43, 202-206.
- LEE, P. C. & MEISEL, D. 1982. Adsorption and surface-enhanced Raman of dyes on silver and gold sols. *The Journal of Physical Chemistry*, 86, 3391-3395.
- LEYTON, P., BRUNET, J., SILVA, V., PAIPA, C., CASTILLO, M. V. & BRANDÁN, S. A. 2012. An experimental and theoretical study of l-tryptophan in an aqueous solution, combining two-layered ONIOM and SCRF calculations. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 88, 162-170.
- SHAKILA, G., PERIANDY, S. & RAMALINGAM, S. 2011. Molecular Structure and Vibrational Analysis of 1-Bromo-2-Chlorobenzene Using ab initio HF and Density Functional Theory (B3LYP) Calculations. *Journal of Atomic, Molecular, and Optical Physics*, 2011, 10.
- SHEPHERD, S. A., KARTHIKEYAN, C., LATHAM, J., STRUCK, A.-W., THOMPSON, M. L., MENON, B. R. K., STYLES, M. Q., LEVY, C., LEYS, D. & MICKLEFIELD, J. 2015. Extending the biocatalytic scope of regiocomplementary flavin-dependent halogenase enzymes. *Chemical Science*, 6, 3454-3460.

3 Thiol-Functionalised Substrates for Sensitive and Rapid Detection of Regioselective Flavin-Dependent Halogenation using SERS

Heidi Fisk, Yun Xu, Royston Goodacre and Jason Micklefield

School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK

3.1 DECLARATION

This chapter consists of one piece of work written in the format of a journal article but not currently submitted for publication.

As primary author, I carried out all experimental work, data analysis and write-up. Dr. Yun Xu provided assistance with the chemometrics and calculating the LOD and LOQ. Prof. Jason Micklefield provided advice and support throughout these initial investigations, particularly towards the biotransformation. Prof. Royston Goodacre, as principal investigator, gave fundamental advice and guidance throughout this research, predominantly towards aspects surrounding SERS.

3.2 ABSTRACT

The incorporation of biocatalysts within pharmaceutical synthesis is continually expanding and gaining pace. Directed evolution strategies have fuelled their progression, yet a lack of efficient high-throughput screening methods to analyse such large mutant libraries has hindered their development and further optimisation. Flavin-dependent halogenases are attractive biocatalysts owing to their regioselective halogenase activity, generating key pharmaceutical intermediates using mild conditions. Until now, their progress has faltered due to the lack of adequate screening techniques which are amenable to their complex cofactor requirements.

Herein, we discuss the development of SERS spectroscopy to monitor this complex reaction in a rapid and sensitive manner. Exploiting sulfur functionalisation of substrates enabled us to detect halogenase activity at low concentrations (2.42×10^{-8} M) without modification of the reaction conditions.

3.3 INTRODUCTION

The application of biocatalysts within pharmaceuticals, fine chemicals and food industries is continually gaining momentum, becoming a key consideration when developing cleaner and greener processes (Leresche and Meyer, 2006, Turner and Truppo, 2013, Reetz, 2013, Patel, 2017). Biocatalysts offer numerous advantages over regular synthetic routes, including high regio-, stereo- and enantio-selectivity, mild conditions, capability of recovering and reusing the catalyst, as well as avoiding toxic reagents and solvents (Patel, 2011, Truppo, 2017, Patel, 2017). However, their application in industry is often restricted resulting from their low-level activity, limited substrate scope and comparatively low stability to traditional chemocatalysts (Lutz and Bornscheuer, 2012, Bornscheuer, 2013). Directed evolution (DE) has proven to be an effective approach to combat these innate pitfalls by developing more suitable and robust biocatalysts. DE is the laboratory method of mimicking natural selection, generating biological entities with improved traits. In brief, genes are subjected to iterative rounds of mutagenesis (*via* various mutagenic strategies), creating genetic diversification within a population. Mutations displaying desirable traits are selected and subjected to further mutagenesis, until a defined end-goal is reached (*i.e.* 50x increase in activity) (Arnold, 1998, Cobb *et al.*, 2013, Bornscheuer, 2013, Packer and Liu, 2015, Grunwald, 2014). The bottleneck

of all mutagenesis strategies is the speed of the screening method which evaluates the effectiveness of the genetic modifications (Leresche and Meyer, 2006, Whittall and Sutton, 2009, Reetz, 2013, Guazzaroni *et al.*, 2015). Major advancements in DE methods have been made, meaning that huge mutant libraries can be generated in a short period of time, yet, the capabilities of high-throughput screening (HTS) methods to assess such large populations has not been achieved. The lack of suitable HTS methods has impacted the success of DE programs and consequently, impeded biocatalyst integration into synthetic routes.

Frequently used HTS methods include bespoke fluorometric or colorimetric assays, LC-MS, HPLC or NMR, however, these methods suffer drawbacks, such as long acquisition times, sample preparation/purification prior to analysis, requirement of a reporter molecule, large solvent volumes and costly equipment (Kumar and Clark, 2006, Bornscheuer, 2013, Grunwald, 2014). As a result, emphasis has been placed on developing new screening methods which can improve upon or completely avoid these weaknesses, one of which is surface enhanced Raman scattering (SERS). SERS is an enhancement technique of Raman spectroscopy, incorporating nanoscale roughened metal surfaces (usually Ag or Au) which interact with the analyte(s) under investigation (Fleischmann *et al.*, 1974). These interactions result in the enhancement of the Raman response in the order of $10^4 - 10^6$ (Jeanmaire and Van Duyne, 1977, Albrecht and Creighton, 1977, Campion and Kambhampati, 1998), frequently enabling detection of analytes at nano- or pico-molar concentration (single molecule detection has been reported (Kneipp *et al.*, 1997, Nie and Emory, 1997, Le Ru *et al.*, 2006, Kleinman *et al.*, 2011)). Molecule specific information is generated using rapid acquisition times (seconds), and this technique is well suited for aqueous conditions as water does not give an appreciable Raman response. Hence, it is evident that SERS is a method that would be well suited for the rapid screening of biotransformations, aiding mutagenesis programs.

Flavin-dependent halogenases (FDHs) are a class of enzyme that are attractive biocatalysts owing to their regioselective halogenase activity (Figure 3.1) (Dong *et al.*, 2005, Yeh *et al.*, 2006). Traditional methods of chemical halogenation require the use of hazardous and toxic reagents, experience poor atom economy and give very little regioselective control over the position at which the halogen is introduced (Rowlands *et al.*, 1994, Alonso *et al.*, 2002). Thus, there is increasing interest towards the development of FDHs to replace traditional methods, allowing for benign conditions and giving desirable regiocontrol. Despite previous achievements to increase the inherent low activity and limited substrate scope of FDHs (Lang *et al.*, 2011, Poor *et al.*, 2014, Frese and Sewald, 2015, Shepherd *et al.*, 2015, Payne *et al.*,

2015, Shepherd *et al.*, 2016, Andorfer *et al.*, 2016), their development has been hampered by the lack of HTS methods that can assess activity, as well as assign regioselectivity (determining the location of the halogen-carbon bond).

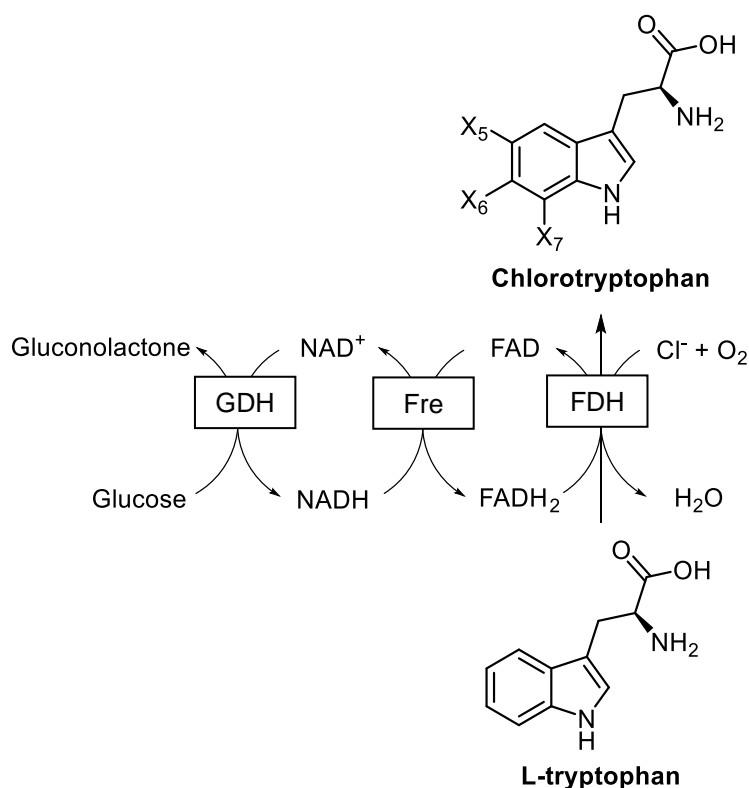


Figure 3.1. The mechanism of flavin-dependent tryptophan halogenases. Glucose dehydrogenase (GDH) and flavin reductase (Fre) recycle essential cofactors, NAD^+ and FAD, to their reduced forms necessary for flavin-dependent halogenase (FDH) function. Position of halogenation (Cl/Br), denoted by X, is dependent on the FDH: $X_5 = \text{PyrH}$, $X_6 = \text{SttH}$, $X_7 = \text{PrnA/RebH}$.

We previously investigated SERS as a method to monitor halogenase activity (Chapter 2: ‘Exploring Surface Enhanced Raman Scattering as a Method to Identify Regioselective Flavin-Dependent Halogenation’), these results identified that not only could we detect halogenation, but that the regiospecific position of the halogen atom could also be established, arising from subtle differences in the SERS response. Nevertheless, using this approach to analyse unmodified biotransformation reaction samples (both *in vitro* and *in vivo* preparations), containing essential cofactors was problematic. Once the required cofactors ($\text{FADH}_2/\text{NADH}$, Figure 3.1) were introduced into the system, detection of the starting material (*i.e.* L-tryptophan) and associated product (*i.e.* 5-chloro-L-tryptophan) was not possible due to the complex cofactors interfering with their interactions at the metal surface.

In this study, we explore the functionalisation of FDH substrates with sulfur-containing groups (*i.e.* thiols, -SH) to promote favourable interactions with the nanoparticle surface and exceed the binding affinity demonstrated by the cofactors. When exposed to transition metals, both thiols and dithiols form spontaneous transition metal-thiolate complexes, often referred to as self-assembled monolayers (SAMs) (Nuzzo and Allara, 1983). Sulfur is a 'soft' and relatively polarisable atom, having a strong affinity for metals, thus it is commonly used in SAMs. Gold-thiolate complexes are the most frequently used SAM due to its reasonably inert surface, gold-thiol affinity being so strong that it readily displaces adventitious structures from the surface, along with the complexes being stable over long periods of time (Nuzzo and Allara, 1983, Love *et al.*, 2005). Silver-thiolate SAMs are the second most studied complexes, however, their use is restricted arising from silver's increased susceptibility to oxidation and their toxicity to cells (Love *et al.*, 2005, Poon and Burd, 2004, AshaRani *et al.*, 2009). SAMs have wide ranging applications and are regularly used within the field of nanotechnology, examples include inks (lithography) (Liu *et al.*, 2000, Critchley *et al.*, 2006, Mu *et al.*, 2015), biosensors (Wink *et al.*, 1997, Chaki and Vijayamohanan, 2002), drug delivery (Paciotti *et al.*, 2004, Ghosh *et al.*, 2008) and as ultrathin corrosion protection layers (coatings) (Ramachandran *et al.*, 1996, Jennings and Laibinis, 1996).

In recent years, interest in monitoring metal-thiolate complexes using SERS has increased. One of the first metal-thiolate complexes studied using SERS was benzenethiol by Joo *et al.*, they postulated that benzenethiol interacts with the Ag surface through dissociative chemisorption of the S-H bond (Joo *et al.*, 1987). Since this initial discovery, many thiol-containing analytes have been studied using SERS, including nanosensors enabling quantitative measurements of the redox potential within eukaryotic cells (Jiang *et al.*, 2014, Thomson *et al.*, 2015), as well as emerging examples using thiols as linkers to facilitate the adsorption and subsequent detection of specific target analytes that might otherwise go undetected (Stewart and Bell, 2011, Graham *et al.*, 2011, Gühlke *et al.*, 2016).

It is this concept that we wish to employ within our investigations, whereby we design and synthesise organosulfur substrates (modelled on active substrates) to undergo halogenation by FDHs. The organosulfur substrate/product exhibit an enhanced binding affinity towards the metal nanoparticle surface, suppressing signals from interfering cofactor species (FADH₂/NADH) which do not relay any information regarding halogenase activity. The sulfur-functionalised substrates enable specific detection allowing measurements of halogenase activity to be determined.

3.4 EXPERIMENTAL

3.4.1 Reagents and Materials.

All chemical reagents were of analytical grade and used with no additional purification. Flavin-dependent halogenases, and their associated enzymes, were expressed and purified as previously reported (Chapter 2).

3.4.2 Raman Instrumentation.

A DeltaNu Advantage 200A portable Raman spectrometer (DeltaNu, Laramie, WY, USA) was used for spectra collection using a HeNe 633 nm laser with ~3 mW on the sample.

3.4.3 SERS Sample Preparation of Standards.

2×10^{-3} M stock concentrations of either thiol **1** or thiol **2** were prepared in MeOH (details of substrate synthesis are described in the SI, section 'Substrate Synthesis and Characterisation'). Prior to individual SERS analysis, stock samples were diluted using potassium phosphate buffer (1×10^{-1} M, pH 7.0). For SERS analysis of thiol **1** or thiol **2** standards: 200 μ L HRSC, followed by 200 μ L of analyte and buffer solution (at a desired concentration) in a glass vial was then vortexed for 7 s. The sample was left to aggregate for 8 min prior to SERS spectral collection (20 s acquisition).

For SERS responses of thioacid **1**/5-chlorothioacid, stock concentrations of 4×10^{-4} M were prepared in acetone (details of substrate synthesis are described in the SI, section 'Substrate Synthesis and Characterisation'). Stock samples were further diluted using potassium phosphate buffer (1×10^{-1} M, pH 7.0). SERS analysis: 200 μ L HRSC was added to a glass vial, next 150 μ L of analyte and buffer solution (at a chosen concentration), followed by 50 μ L of KNO_3 (5×10^{-1} M). The sample was vortexed for 7 s and left to aggregated for 8 min before SERS collection (25 s acquisition).

3.4.4 FDH Activity Assays.

Small scale activity assays were prepared using FDH (1×10^{-5} M), Fre (1×10^{-6} M), GDH (6×10^{-6} M), FAD (7.5×10^{-6} M), NADH (2×10^{-4} M), MgCl_2 (2×10^{-2} M), glucose (2×10^{-2} M) and substrate (5×10^{-4} M) in a total volume of 100 μ L in potassium phosphate buffer (1×10^{-2} M, pH 7.0). Reaction samples were incubated at 37 °C with agitation for 1 - 16 h, before being quenched (95 °C, 5 min). Precipitated proteins were removed by centrifugation (21,000 $\times g$ for 6 min) before analysis *via* HPLC or SERS.

3.5 RESULTS AND DISCUSSION

3.5.1 Designing Thiol-Functionalised Substrates

Previous investigations into the substrate scope of SttH, a C-6 flavin-dependent tryptophan halogenase, aided the design of organosulfur substrates (Chapter 2). Figure 3.2 identifies substrates that are active towards SttH halogenase, their respective chlorinated products have been characterised and the position of the C-Cl bond assigned. Knowing the structures of active substrates assisted the design process of developing a panel of organosulfur substrates.

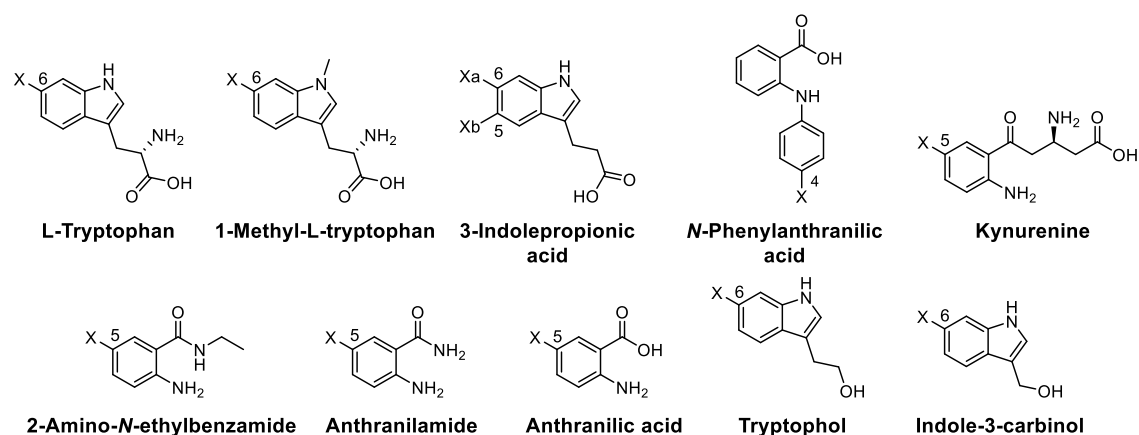
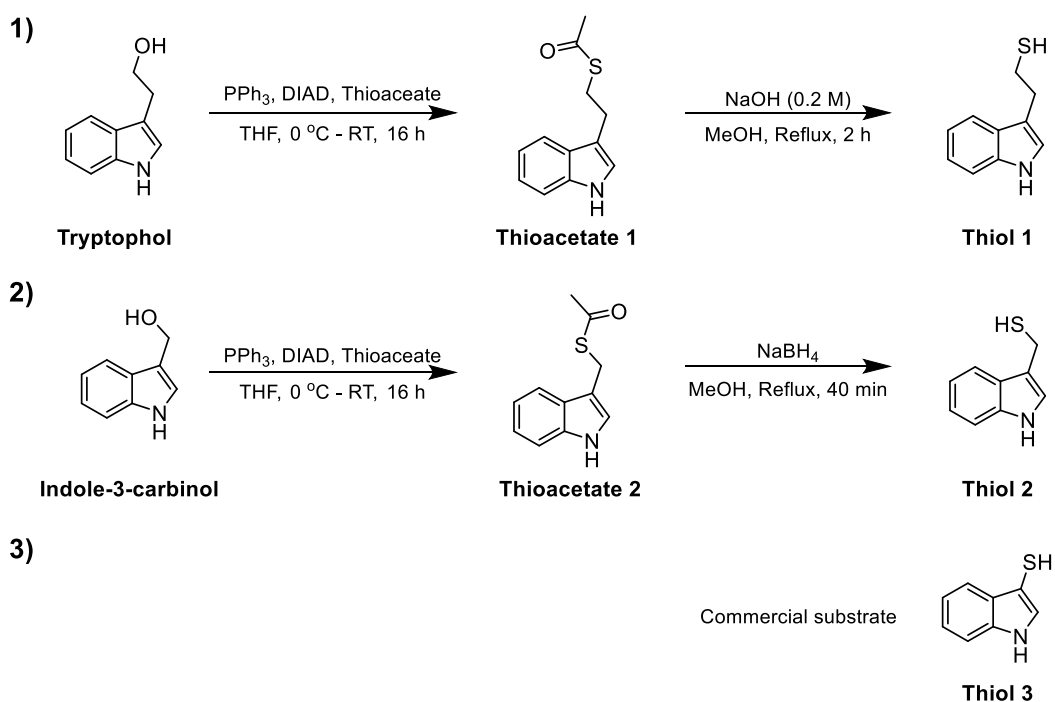


Figure 3.2. Structures of SttH active substrates, X denotes the position of chlorination [3-indolepropionic acid produces two chlorinated products: Xa = Cl, Xb = H (90 %) and Xa = H, Xb = Cl (10 %)].

The initial design process led us to investigate tryptophol and indole-3-carbinol (Scheme 3.1) as primary substrates for sulfur-functionalisation (their relative activity with a variety of flavin-dependent tryptophan halogenases is shown in the SI, Figure S3.1). The Mitsunobu reaction is an effective and versatile method which converts hydroxyl groups into potent leaving groups, which are readily displaced by a variety of nucleophiles (Mitsunobu and Yamada, 1967). This reaction was successfully performed on these primary alcohol substrates to afford thioacetates (thioacetates **1** and **2**, Scheme 3.1). Subsequent hydrolysis (Scheme 3.1.1) or reduction (Scheme 3.1.2) afforded thiols **1** and **2**, respectively (full synthesis details and characterisation found in SI, ‘Substrate Synthesis and Characterisation’). A third commercially available thiol, thiol **3** (Scheme 3.1.3) was included within our investigations, enabling comparisons into alkanethiol chain length.



Scheme 3.1 Synthesis overview and structures of all initial organosulfur substrates. (1) The conversion of tryptophol to thioacetate **1**, followed by hydrolysis to thiol **1**. (2) Indole-3-carbinol conversion to thioacetate **2**, ensued by reduction to thiol **2**. (3) Thiol **3**, 3-mercaptoindole, a commercially available compound.

Initial optimisation of thiol-functionalised analytes for SERS responses was investigated, and during this process numerous experimental parameters were varied. Optimal responses were attained using 200 μL of Ag-hydroxylamine reduced colloid (HRSC), 6.25×10^{-6} M total analyte concentration, 8 min aggregation time and 25 s acquisition time, using 633 nm laser excitation. The biotransformation contains potassium phosphate buffer 1×10^{-3} M at pH 7.0, preliminary optimisation revealed that this does not interfere with SERS responses. A SERS concentration profile of thiol **1** was performed to determine the limit of detection (LOD) and limit of quantification (LOQ). Principal component analysis (PCA) is a multivariate statistical approach often employed to explain the natural variance within a dataset, uncovering patterns that might otherwise go undetected. The PCA plot of the concentration profile, Figure S3.2 in SI, identifies a clear trend arising from increasing/decreasing concentration of thiol **1**. The LOQ identifiable at 9.77×10^{-8} M and the LOD calculated as 2.42×10^{-8} M (LOD and LOQ calculations, Equation S3.1 and S3.2 in SI). Nanomolar detection demonstrates the high sensitivity and low-level detection achievable with these thiol-functionalised analytes. The PCA plot also shows tight clustering of SERS data from replicate samples within each of the thiol concentrations, identifying that these optimal conditions give very good reproducibility.

Following SERS optimisation, all thioacetate intermediates and thiol substrates were tested for activity towards four different FDHs (SttH C-6, PrnA C-7, PyrH C-5 and RebH C-7). Despite demonstrable enzyme activity towards the alcohol starting materials (Figure S3.1 in SI), no observable activity was seen with any sulfur-containing substrate, despite numerous attempts and increasing enzyme concentration (confirmed using HPLC analysis).

Thiols are prone to oxidation, forming disulfide bonds (-S-S-). To ensure that oxidation was avoided and did not obstruct the acceptance of the substrate within the active site, dithioereitol (DTT) was introduced into the enzymatic assays. DTT is a small molecule that is widely used to reduce disulfide bonds, as well as stabilising proteins containing free thiol groups. DTT addition had no effect on activity and no halogenated product(s) was observed.

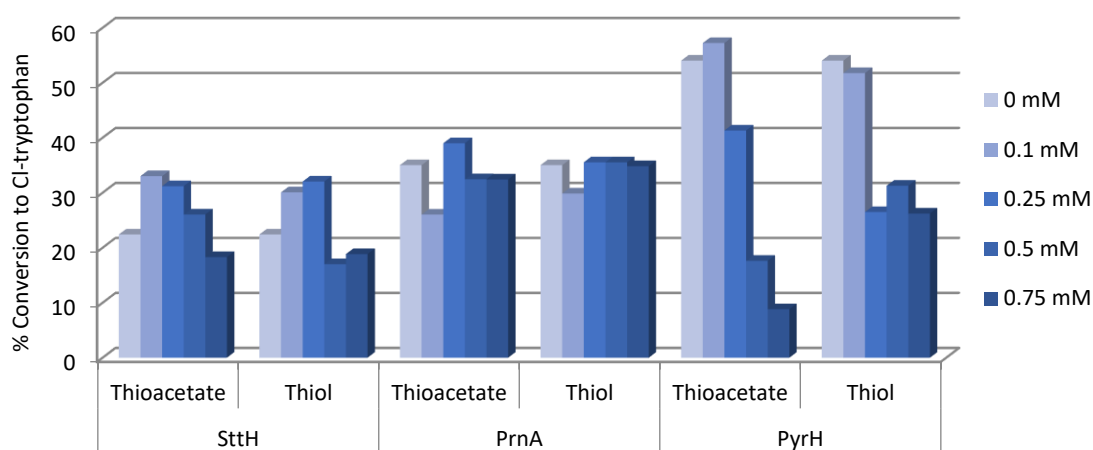


Figure 3.3. Histogram representing the percentage conversion of L-tryptophan (5×10^{-4} M) to chlorinated product using different FDHs in the presence of various concentrations of thioacetate **1** or thiol **1**. Assay conditions: L-tryptophan (5×10^{-4} M) with either thioacetate **1** or thiol **1** (0, 1×10^{-4} , 2.5×10^{-4} , 5×10^{-4} , 7.5×10^{-4} M) in potassium phosphate buffer (1×10^{-3} M, pH 7.0), 1 h, 37 °C, 800 rpm (single measurements).

Next, studies were performed to identify if the introduction of the sulfur atom was causing these substrates to behave as inhibitors. Increasing concentrations of either thioacetate **1** or thiol **1** were introduced into the assay, along with the natural substrate, L-tryptophan. If one compares the conversion of L-tryptophan to chlorinated product using three different FDHs, in the absence and presence of thioacetate **1**/thiol **1** (Figure 3.3), it is evident that there are no obvious deviations in activity. PyrH demonstrates a linear correlation between increasing concentrations of thioacetate **1** and decreasing conversion to chlorinated product. Nevertheless, in general the activity is unaffected by the organosulfur substrate, so it does not seem that these substrates are behaving as inhibitors.

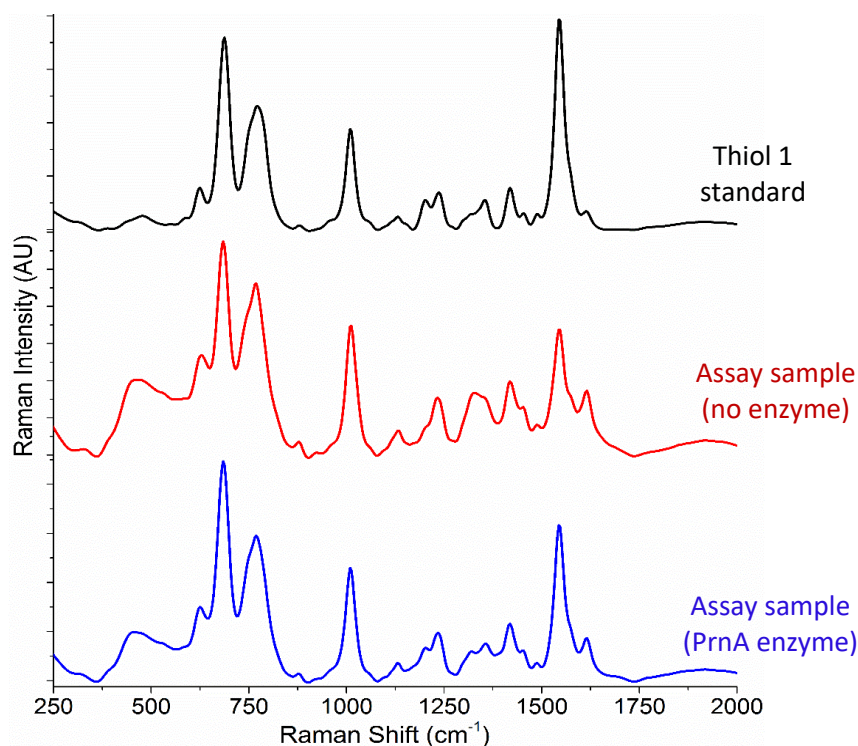


Figure 3.4. SERS spectra ($n = 3$) of thiol **1** standard 2.5×10^{-5} M (black line). Thiol **1** was incubated for 1 h within assay conditions either without enzyme (red line) or with enzyme, PrnA, present (blue line). All spectra represent a final concentration of thiol **1** at 2.5×10^{-5} M, enzymes were heat denatured (95°C , 5 min) and precipitated proteins removed using centrifugation ($21,000 \times g$ for 6 min) prior to SERS analysis. SERS conditions: 200 μL HRSC, 200 μL of assay/thiol **1** sample, 8 min aggregation time, 25 s acquisition using 633 nm excitation. Spectra were baseline corrected as detailed in the SI, data processing section.

Despite no enzyme activity towards thioacetates **1** and **2**, along with thiols **1** and **2**, real biotransformation samples were used to identify that organosulfur analytes exceed the binding affinity of other cofactors, leading to successful detection of the organosulfur analyte within the complex mixture. Two assay samples were prepared including the necessary cofactors and conditions needed for the FDH reaction (Figure 3.1), using thiol **1** as the substrate. One of those samples did not contain a FDH (Figure 3.4, red line), the other had PrnA, C-7 FDH, within the assay sample (Figure 3.4, blue line). The assay samples were incubated for 1 h at 37°C before protein precipitation and removal prior to SERS analysis. Figure 3.4 (black line) shows the standard spectrum of thiol **1**, the strong band at 684 cm^{-1} is characteristic of C-S vibration (C-S stretch occurs between $710\text{--}570\text{ cm}^{-1}$, (Joo *et al.*, 1987, Szafranski *et al.*, 1998, Bloxham *et al.*, 2002)). Observing the SERS standard of thiol **1** in the absence of cofactors and enzymes (Figure 3.4, black line), one can see that there is little alteration of that response when compared with the biotransformation samples (without/with enzyme, Figure 3.4 red/blue lines, respectively). Consequently, incorporation

of the thiol group has promoted favourable interactions of the analyte with the Ag nanoparticles, maintaining sensitive detection within the complex mixture when other molecules are competing for the nanoparticle surface/preventing successful interactions.

Next, the substrate scope of organosulfur analytes was increased by a further two substrates, in an attempt to identify active sulfur-functionalised analytes. These substrates included the commercially available 3-(1-benzothiophen-3-yl)-L-alanine, which is a structural analogue of L-tryptophan (Figure 3.5B). The indole moiety is replaced by a benzothiophene, retaining amino and carboxylic acid groups on the hydrocarbon side chain. For the second substrate, we converted FDH active 3-indolepropionic acid (Figure 3.2) to its corresponding thioacid (thioacid **1**) using Lawesson's Reagent (Figure 3.5C). This one-step protocol quickly (10 min) converts carboxylic acids to their respective thioacids using an inexpensive and commercially available Lawesson's Reagent (Scheme 3.2, full synthesis details and characterisation found in SI, 'Substrate Synthesis and Characterisation') (Rao *et al.*, 2009). Similarly, thioacid **1** shares structural similarities to L-tryptophan, differing in the absence of the amino group and substituting the carboxylic acid for thioacid functionality.

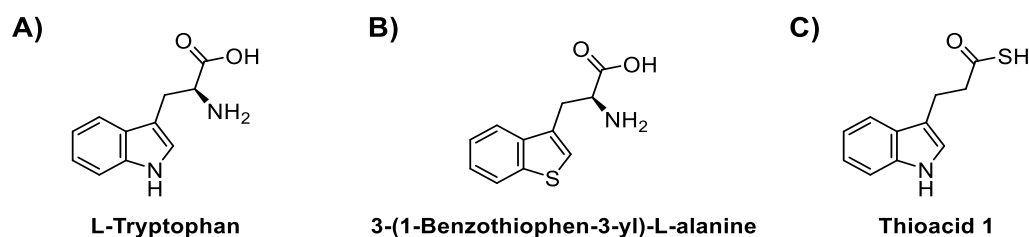


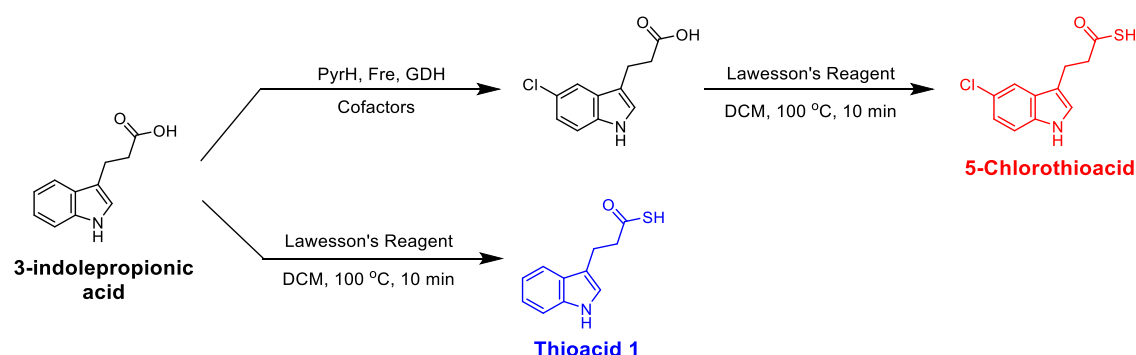
Figure 3.5. Structures of A) L-tryptophan (natural substrate of FDH), B) 3-(1-benzothiophen-3-yl)-L-alanine and C) thioacid **1**.

Due to the structural resemblances of these organosulfur compounds with the most active FDH substrate, L-tryptophan, halogenase activity was conceivable. Despite their heightened suitability, neither substrate demonstrated halogenase activity towards a panel of FDHs. Sulfur is a larger element than oxygen or nitrogen, its increased size lessens its electronegativity difference between sulfur and hydrogen (thiol) in comparison to oxygen and hydrogen (alcohol) or nitrogen and hydrogen (amine). Alcohols and amines interact with themselves and other groups through hydrogen bonding, dipole-dipole interactions, as well as Van der Waals dispersion forces. Whereas thiols display weak associations through hydrogen bonding (due to similarities in electronegativity) and interactions mainly occur through Van der Waals interactions. We hypothesise that these differences in binding

interactions are having adverse effects on the acceptance of the organosulfur substrates within the halogenase active site. It is also plausible that the increased size of the sulfur atom compared with oxygen/nitrogen could also be detrimental towards FDH activity, restricting their acceptance into the constrained active site. Key residues and interactions within the substrate binding site of FDHs are discussed in further detail within the SI, section 'FDH Active Site Interactions'.

3.5.2 Modelling Enzyme Activity

To complete our comprehensive study, we wanted to confirm with complete confidence that if FDHs had shown activity towards organosulfur substrates, that they would be distinguishable from their chlorinated product(s) and quantitative measurements could be made. To achieve this, 3-indolepropionic acid was halogenated on large scale with PyrH (C-5 FDH) to produce 5-chloro-3-indolepropionic acid (further details within SI 'Substrate synthesis and characterisation'). Conversion to a thioacid using Lawesson's Reagent was performed on the chlorinated product, synthesising 5-chlorothioacid, which we could compare with thioacid **1** (Scheme 3.2).



Scheme 3.2. Synthesis of thioacid **1** (blue) via Lawesson's Reagent reaction using 3-indolepropionic acid. 5-chlorothioacid (red) synthesised using a two-step process: initial regioselective chlorination of 3-indolepropionic acid using PyrH (FDH) at C-5, subsequent conversion of carboxylic acid to thioacid using Lawesson's Reagent.

To mimic enzymatic halogenation of thioacid **1** (had it been an active substrate), mixtures of thioacid **1** and 5-chlorothioacid were prepared and the SERS responses recorded. The total concentration was set to 5×10^{-5} M, model mixtures of 0 – 100 % chlorination were prepared at 5 % intervals (for example, 50 % chlorination = 2.5×10^{-5} M thioacid **1** and 2.5×10^{-5} M 5-chlorothioacid). The standard SERS spectrum of the non-chlorinated and chlorinated thioacids are shown in Figure S3.3 in SI, clear discriminant bands are easily identified. The

PCA plot of the known mixtures is shown in Figure 3.6 for the first dataset. As depicted by the colour gradient, one can see that on increasing concentration of 5-chlorothioacid (product) and corresponding decrease in thioacid **1** (starting material), the PCA can clearly recognise this trend. Most of the variance is captured by PC 1 (TEV 98.5 %), however further separation of the different mixtures is achieved using PC 2 (TEV 0.86 %). Product concentrations below 15 % ($<7.5 \times 10^{-6}$ M) are more difficult to distinguish between, thus product formation would need to exceed this level for accurate detection and quantification. The loadings plot in Figure S3.4 in SI identifies key peaks which enable PCA separation, predominantly arising from peaks associated to thioacid **1** which are absent in the chlorinated product (model mixtures were prepared and analysed twice, analysis of the second dataset is shown in Figure S3.5 and S3.6 in SI). These results suggest that if the FDH enzymes had shown activity towards the thiolated substrates, then SERS would have been a useful and rapid method to assess enzyme activity.

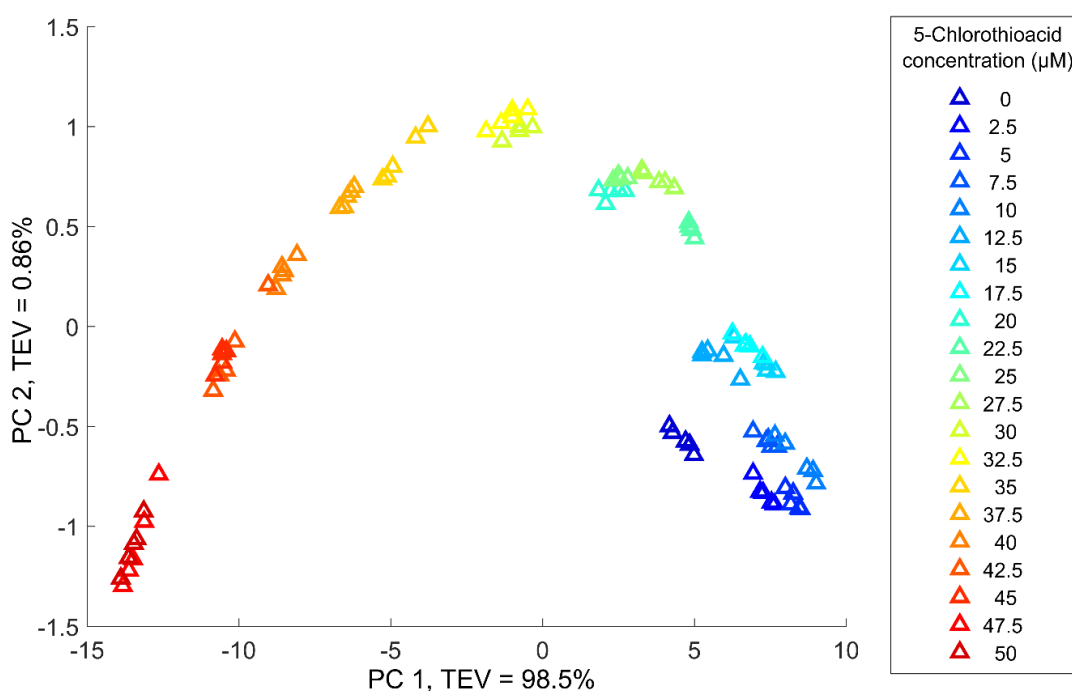


Figure 3.6. Principal component analysis (PCA) scores plot of known mixtures of thioacid **1** and 5-chlorothioacid (from dataset 1). Legend represents concentration of 5-chlorothioacid (i.e. 5 μ M = 5×10^{-6} M (10 % 5-chlorothioacid, 90 % thioacid **1**), 40 μ M = 4×10^{-5} M (90 % 5-chlorothioacid, 10 % thioacid **1**)). PCA was performed on full spectral range (203 – 3400 cm^{-1}). Total concentration (5×10^{-5} M), 5x replicates, 633 nm laser excitation, 200 μ L HRSC, 150 μ L thioacid **1**/5-chlorothioacid mixture, 50 μ L KNO_3 aggregating agent (5×10^{-1} M), 8 min aggregation, 20 s acquisition. Data were SNV normalised prior to chemometric analysis (as detailed in the SI, section data processing). TEV = total explained variance.

3.6 CONCLUSION

Previous investigations into SERS as a technique to monitor complex biotransformations, such as FDH activity, was unsuccessful arising from the sheer complexity of the reaction. Essential cofactors cause competition/interference at the colloidal nanoparticle surface, ultimately suppressing SERS responses of the molecules of interest (starting material and halogenated product). Thus, we investigated the utilisation of sulfur functionalisation to promote the binding affinity of interesting molecules, outcompeting cofactors which relay little information.

In total, five organosulfur substrates were investigated and proved to be successful at promoting these favourable interactions with the metal surface, exceeding binding interactions of the other assay substituents and rendering their SERS responses insignificant. Low level detection was achieved: thiol **1** permitted a LOQ of 9.77×10^{-8} M and the LOD as 2.42×10^{-8} M, demonstrating the high sensitivity of this SERS approach. Despite these sulfur-containing substrates showing no activity towards FDHs, we mimicked the enzymatic progression of this reaction, preparing model mixtures which represented the conversion of thioacid **1** to its chlorinated product, 5-chlorothioacid. Results established that quantification of halogenase activity using SERS is conceivable and that this approach could be applied towards other complex biotransformations which use enzymes that may be more accepting of sulfur-containing substrates.

3.7 REFERENCES

- ALBRECHT, M. G. & CREIGHTON, J. A. 1977. Anomalously intense Raman spectra of pyridine at a silver electrode. *Journal of the American Chemical Society*, 99, 5215-5217.
- ALONSO, F., BELETSKAYA, I. P. & YUS, M. 2002. Metal-Mediated Reductive Hydrodehalogenation of Organic Halides. *Chemical Reviews*, 102, 4009-4092.
- ANDORFER, M. C., PARK, H. J., VERGARA-COLL, J. & LEWIS, J. C. 2016. Directed evolution of RebH for catalyst-controlled halogenation of indole C-H bonds. *Chemical Science*, 7, 3720-3729.
- ARNOLD, F. H. 1998. Design by Directed Evolution. *Accounts of Chemical Research*, 31, 125-131.
- ASHARANI, P. V., LOW KAH MUN, G., HANDE, M. P. & VALIYAVEETIL, S. 2009. Cytotoxicity and Genotoxicity of Silver Nanoparticles in Human Cells. *ACS Nano*, 3, 279-290.
- BLOXHAM, S., EICHER-LORKA, O., JAKUBĚNAS, R. & NIAURA, G. 2002. Surface-enhanced Raman spectroscopy of ethanethiol adsorbed at copper electrode. *Chemija*, 13, 185-189.

- BORNSCHEUER, U. T. 2013. Protein engineering as a tool for the development of novel bioproduction systems. *Advances in Biochemical Engineering/Biotechnology*, 137, 25-40.
- CAMPION, A. & KAMBHAMPATI, P. 1998. Surface-enhanced Raman scattering. *Chemical Society Reviews*, 27, 241-250.
- CHAKI, N. K. & VIJAYAMOHANAN, K. 2002. Self-assembled monolayers as a tunable platform for biosensor applications. *Biosensors and Bioelectronics*, 17, 1-12.
- COBB, R. E., CHAO, R. & ZHAO, H. 2013. Directed Evolution: Past, Present and Future. *AIChE journal. American Institute of Chemical Engineers*, 59, 1432-1440.
- CRITCHLEY, K., ZHANG, L., FUKUSHIMA, H., ISHIDA, M., SHIMODA, T., BUSHBY, R. J. & EVANS, S. D. 2006. Soft-UV Photolithography using Self-Assembled Monolayers. *The Journal of Physical Chemistry B*, 110, 17167-17174.
- DONG, C., FLECKS, S., UNVERSUCHT, S., HAUPT, C., VAN PÉE, K.-H. & NAISMITH, J. H. 2005. Tryptophan 7-Halogenase (PrnA) Structure Suggests a Mechanism for Regioselective Chlorination. *Science*, 309, 2216-2219.
- FLEISCHMANN, M., HENDRA, P. J. & MCQUILLAN, A. J. 1974. Raman spectra of pyridine adsorbed at a silver electrode. *Chemical Physics Letters*, 26, 163-166.
- FRESE, M. & SEWALD, N. 2015. Enzymatic Halogenation of Tryptophan on a Gram Scale. *Angewandte Chemie International Edition*, 54, 298-301.
- GHOSH, P., HAN, G., DE, M., KIM, C. K. & ROTELLO, V. M. 2008. Gold nanoparticles in delivery applications. *Advanced Drug Delivery Reviews*, 60, 1307-1315.
- GRAHAM, D., STEVENSON, R., THOMPSON, D. G., BARRETT, L., DALTON, C. & FAULDS, K. 2011. Combining functionalised nanoparticles and SERS for the detection of DNA relating to disease. *Faraday Discussions*, 149, 291-299.
- GRUNWALD, P. 2014. *Industrial Biocatalysis*, Florida, Pan Stanford Publishing.
- GUAZZARONI, M.-E., SILVA-ROCHA, R. & WARD, R. J. 2015. Synthetic biology approaches to improve biocatalyst identification in metagenomic library screening. *Microbial Biotechnology*, 8, 52-64.
- GÜHLKE, M., HEINER, Z. & KNEIPP, J. 2016. Surface-Enhanced Raman and Surface-Enhanced Hyper-Raman Scattering of Thiol-Functionalized Carotene. *The Journal of Physical Chemistry C*, 120, 20702-20709.
- JEANMAIRE, D. L. & VAN DUYN, R. P. 1977. Surface raman spectroelectrochemistry. *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry*, 84, 1-20.
- JENNINGS, G. K. & LAIBINIS, P. E. 1996. Self-assembled monolayers of alkanethiols on copper provide corrosion resistance in aqueous environments. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 116, 105-114.
- JIANG, J., AUCHINVOLE, C., FISHER, K. & CAMPBELL, C. J. 2014. Quantitative measurement of redox potential in hypoxic cells using SERS nanosensors. *Nanoscale*, 6, 12104-12110.
- JOO, T. H., KIM, M. S. & KIM, K. 1987. Surface-enhanced Raman scattering of benzenethiol in silver sol. *Journal of Raman Spectroscopy*, 18, 57-60.
- KLEINMAN, S. L., RINGE, E., VALLEY, N., WUSTHOLZ, K. L., PHILLIPS, E., SCHEIDT, K. A., SCHATZ, G. C. & VAN DUYN, R. P. 2011. Single-Molecule Surface-Enhanced Raman Spectroscopy of Crystal Violet Isotopologues: Theory and Experiment. *Journal of the American Chemical Society*, 133, 4115-4122.
- KNEIPP, K., WANG, Y., KNEIPP, H., PERELMAN, L. T., ITZKAN, I., DASARI, R. R. & FELD, M. S. 1997. Single Molecule Detection Using Surface-Enhanced Raman Scattering (SERS). *Physical Review Letters*, 78, 1667-1670.
- KUMAR, R. A. & CLARK, D. S. 2006. High-throughput screening of biocatalytic activity: applications in drug discovery. *Current Opinion in Chemical Biology*, 10, 162-168.

- LANG, A., POLNICK, S., NICKE, T., WILLIAM, P., PATALLO, E. P., NAISMITH, J. H. & VAN PÉE, K.-H. 2011. Changing the Regioselectivity of the Tryptophan 7-Halogenase PrnA by Site-Directed Mutagenesis. *Angewandte Chemie International Edition*, 50, 2951-2953.
- LE RU, E. C., MEYER, M. & ETCHEGOIN, P. G. 2006. Proof of Single-Molecule Sensitivity in Surface Enhanced Raman Scattering (SERS) by Means of a Two-Analyte Technique. *The Journal of Physical Chemistry B*, 110, 1944-1948.
- LERESCHE, J. E. & MEYER, H.-P. 2006. Chemocatalysis and Biocatalysis (Biotransformation): Some Thoughts of a Chemist and of a Biotechnologist. *Organic Process Research & Development*, 10, 572-580.
- LIU, G.-Y., XU, S. & QIAN, Y. 2000. Nanofabrication of Self-Assembled Monolayers Using Scanning Probe Lithography. *Accounts of Chemical Research*, 33, 457-466.
- LOVE, J. C., ESTROFF, L. A., KRIEBEL, J. K., NUZZO, R. G. & WHITESIDES, G. M. 2005. Self-Assembled Monolayers of Thiolates on Metals as a Form of Nanotechnology. *Chemical Reviews*, 105, 1103-1170.
- LUTZ, S. & BORNSCHEUER, U. T. 2012. *Protein Engineering Handbook*, Wiley.
- MITSUNOBU, O. & YAMADA, M. 1967. Preparation of Esters of Carboxylic and Phosphoric Acid via Quaternary Phosphonium Salts. *Bulletin of the Chemical Society of Japan*, 40, 2380-2382.
- MU, X., GAO, A., WANG, D. & YANG, P. 2015. Self-Assembled Monolayer-Assisted Negative Lithography. *Langmuir*, 31, 2922-2930.
- NIE, S. & EMORY, S. R. 1997. Probing Single Molecules and Single Nanoparticles by Surface-Enhanced Raman Scattering. *Science*, 275, 1102-1106.
- NUZZO, R. G. & ALLARA, D. L. 1983. Adsorption of bifunctional organic disulfides on gold surfaces. *Journal of the American Chemical Society*, 105, 4481-4483.
- PACIOTTI, G. F., MYER, L., WEINREICH, D., GOIA, D., PAVEL, N., MCLAUGHLIN, R. E. & TAMARKIN, L. 2004. Colloidal Gold: A Novel Nanoparticle Vector for Tumor Directed Drug Delivery. *Drug Delivery*, 11, 169-183.
- PACKER, M. S. & LIU, D. R. 2015. Methods for the directed evolution of proteins. *Nature Reviews Genetics*, 16, 379-394.
- PATEL, R. N. 2011. Biocatalysis: Synthesis of Key Intermediates for Development of Pharmaceuticals. *ACS Catalysis*, 1, 1056-1074.
- PATEL, R. N. 2017. Biocatalysis for synthesis of pharmaceuticals. *Bioorganic & Medicinal Chemistry*, In press.
- PAYNE, J. T., POOR, C. B. & LEWIS, J. C. 2015. Directed Evolution of RebH for Site-Selective Halogenation of Large Biologically Active Molecules. *Angewandte Chemie International Edition*, 54, 4226-4230.
- POON, V. K. M. & BURD, A. 2004. In vitro cytotoxicity of silver: implication for clinical wound care. *Burns*, 30, 140-147.
- POOR, C. B., ANDORFER, M. C. & LEWIS, J. C. 2014. Improving the Stability and Catalyst Lifetime of the Halogenase RebH By Directed Evolution. *ChemBioChem*, 15, 1286-1289.
- RAMACHANDRAN, S., TSAI, B.-L., BLANCO, M., CHEN, H., TANG, Y. & GODDARD, W. A. 1996. Self-Assembled Monolayer Mechanism for Corrosion Inhibition of Iron by Imidazolines. *Langmuir*, 12, 6419-6428.
- RAO, Y., LI, X., NAGORNY, P., HAYASHIDA, J. & DANISHEFSKY, S. J. 2009. A Simple Method for the Conversion of Carboxylic Acids into Thioacids with Lawesson's Reagent. *Tetrahedron Letters*, 50, 6684-6686.
- REETZ, M. T. 2013. Biocatalysis in Organic Chemistry and Biotechnology: Past, Present, and Future. *Journal of the American Chemical Society*, 135, 12480-12496.
- ROWLANDS, S. A., HALL, A. K., MCCORMICK, P. G., STREET, R., HART, R. J., EBELL, G. F. & DONECKER, P. 1994. Destruction of toxic materials. *Nature*, 367, 223-223.

- SHEPHERD, S. A., KARTHIKEYAN, C., LATHAM, J., STRUCK, A.-W., THOMPSON, M. L., MENON, B. R. K., STYLES, M. Q., LEVY, C., LEYS, D. & MICKLEFIELD, J. 2015. Extending the biocatalytic scope of regiocomplementary flavin-dependent halogenase enzymes. *Chemical Science*, 6, 3454-3460.
- SHEPHERD, S. A., MENON, B. R. K., FISK, H., STRUCK, A.-W., LEVY, C., LEYS, D. & MICKLEFIELD, J. 2016. A Structure-Guided Switch in the Regioselectivity of a Tryptophan Halogenase. *ChemBioChem*, 17, 821-824.
- STEWART, A. & BELL, S. E. J. 2011. Modification of Ag nanoparticles with mixed thiols for improved SERS detection of poorly adsorbing target molecules: detection of MDMA. *Chemical Communications*, 47, 4523-4525.
- SZAFRANSKI, C. A., TANNER, W., LAIBINIS, P. E. & GARRELL, R. L. 1998. Surface-Enhanced Raman Spectroscopy of Aromatic Thiols and Disulfides on Gold Electrodes. *Langmuir*, 14, 3570-3579.
- THOMSON, P. I. T., CAMUS, V. L., HU, Y. & CAMPBELL, C. J. 2015. Series of Quinone-Containing Nanosensors for Biologically Relevant Redox Potential Determination by Surface-Enhanced Raman Spectroscopy. *Analytical Chemistry*, 87, 4719-4725.
- TRUPPO, M. D. 2017. Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Medicinal Chemistry Letters*, 8, 476-480.
- TURNER, N. J. & TRUPPO, M. D. 2013. Biocatalysis enters a new era. *Current Opinion in Chemical Biology*, 17, 212-214.
- WHITTALL, J. & SUTTON, P. W. 2009. *Practical Methods for Biocatalysis and Biotransformations*, West Sussex, UK, Wiley.
- WINK, T., J. VAN ZUILEN, S., BULT, A. & P. VAN BENNEKOM, W. 1997. Self-assembled Monolayers for Biosensors. *Analyst*, 122, 43R-50R.
- YEH, E., COLE, L. J., BARR, E. W., BOLLINGER, J. M., BALLOU, D. P. & WALSH, C. T. 2006. Flavin Redox Chemistry Precedes Substrate Chlorination during the Reaction of the Flavin-Dependent Halogenase RebH. *Biochemistry*, 45, 7904-7912.

3.8 SUPPORTING INFORMATION

SUPPLEMENTARY METHODS

Reagents and Materials.

All chemical reagents were of analytical grade and used with no additional purification. Chemicals and solvents were purchased from Acros Organic (New Jersey, USA), Alfa Aesar (Heysham, UK), Fischer Scientific (Loughborough, UK), Formedium (Hunstanton, UK) or Sigma Aldrich Ltd (Dorset, UK). Details of cloning, expression and protein purification can be found in Chapter 2.

Synthesis of Silver Nanoparticles.

All glassware used within the synthesis of metal nanoparticles was cleaned using aqua regia (HNO_3 : HCl (1:3) v/v) to ensure the removal of any residual metals. Glassware was then thoroughly rinsed with deionised water.

Hydroxylamine-reduced silver colloid (HRSC) was prepared as reported by Leopold and Lendl (Leopold and Lendl, 2003). Hydroxylamine hydrochloride (1.88×10^{-3} M, 180 mL) and NaOH (3.33×10^{-3} M) were prepared together. To this rapidly stirring solution, AgNO_3 (1×10^{-2} M, 20 mL) was added drop-wise and stirred for a further 15 min, forming an orange-yellow colloid.

SERS Sample Preparation of Standards.

Thiols 1 and 2. 2×10^{-3} M stock concentrations of either thiol **1** or thiol **2** were prepared in MeOH. Prior to individual SERS analysis, stock samples were diluted using potassium phosphate buffer (1×10^{-1} M, pH 7.0). For SERS analysis of standards, the following preparation was performed: 200 μL HRSC was added to a glass vial, followed by 200 μL of analyte and buffer solution (at the desired concentration) and then vortexed for 7 s. MeOH content did not exceed 2 % of total SERS sample volume and did not affect the SERS spectrum.

Thioacid 1 and 5-chlorothioacid. 4×10^{-4} M stock concentrations of thioacid **1** and 5-chlorothioacid were prepared using acetone. Stock samples were further diluted using potassium phosphate buffer (1×10^{-1} M, pH 7.0). SERS analysis preparation: 200 μL HRSC was added to a glass vial, next 150 μL of analyte and buffer solution (at desired concentration), followed by 50 μL of KNO_3 (5×10^{-1} M) and vortexed for 7 s. The acetone content did not exceed 2.5 % of the total SERS sample volume and did not affect the SERS spectrum.

INSTRUMENTATION AND DATA PROCESSING

Raman Spectrometer.

SERS spectra were recorded using a DeltaNu Advantage 200A portable spectrometer (DeltaNu, Laramie, WY, USA) equipped with a HeNe 633 nm laser providing ~3 mW of power on the sample. Toluene was pipetted into 8 mm glass vials to calibrate the system, allowing the ideal distance between the laser and the point of sampling to be established. 633 nm SERS spectra were acquired for 20 s (thiol 1/2) or 25 s (thioacid 1/5-chlorothioacid) over a range of 200 – 3400 cm^{-1} ; the spectral resolution was 10 cm^{-1} . Samples for SERS analysis were placed in an 8 mm glass vial, vortexed for 5 s and left for 8 min to aggregate (optimal aggregation time), before subjected to laser irradiation once placed in the cell holder.

Raman spectra processing was performed using Matlab software version R2008a (The MathWorks, Natick, MA, USA) using scripts written in-house and available via GitHub (<http://www.biospec.net/resources/>). SERS spectra were either raw, baseline corrected or SNV normalised (as stated in the text).

UV-Vis Spectrophotometry.

Hydroxylamine-reduced silver colloids (HRSC) were characterised using UV-Vis spectroscopy to determine the surface plasmon resonance band λ_{max} of the nanoparticles. HRSC was diluted 1:3 v/v with deionised water, to ensure that the UV-Vis absorbance did not exceed 2. For analysis 1 mL of the diluted nanoparticle solution was pipetted into a polystyrene cuvette and placed into the sample holder of a Thermo Biomate 5 (Thermo Fisher Scientific Inc., Massachusetts, USA). Absorbance spectra were collected over a range of 200-800 nm.

SEM.

SEM images were collected using a Field Emission Scanning Electron Microscope with an ultimate beam size of 1 nm at 100 eV electron energy (Carl-Zeiss-Straße 56, 73447 Oberkochen, Germany). A silicon surface was used to image the nanoparticles as silicon helps disperse negative charges and generates a greater contrast in comparison to glass, as a result higher resolution images can be obtained. To image HRSC nanoparticles, 10 μL was spotted onto a silicon slide and left to dry overnight (Figure S3.7).

HPLC.

Halogenase assays were analysed by HPLC on an Agilent Technologies 1260 Infinity HPLC with an Agilent Zorbax Eclipse Plus C18 Rapid Resolution 4.6 x 100 mm 3.5 μm column. Absorbance was measured at 254 nm, with a 3 min gradual gradient 20:80 % $\text{H}_2\text{O}/\text{ACN}$ + 0.05 % trifluoroacetic acid (TFA), followed by a gradual gradient of 5:95 $\text{H}_2\text{O}/\text{ACN}$ + 0.05 % TFA over 2.5 min. Flow rates were kept constant at 1 mL min^{-1} . The peak areas of analytes of interest were integrated using ChemStation (Agilent 1100 series) and reported % conversions were adjusted in accordance with HPLC calibrations.

Purification of PyrH biotransformation samples was performed on a semi-preparative scale using on Varian Prostar 210. 500 μL of solution containing concentrated crude reaction mixture was injected onto a Phenomenex Gemini[®] C18 HPLC column (5 μ packing, 110 \AA , 250 x 10 mm). A gradual gradient of 5 – 75 % $\text{H}_2\text{O}/\text{ACN}$ + 0.1 % FA was used over 35 min, using a 5 mL min^{-1} flow rate. UV absorbance was detected at 280 nm throughout.

NMR spectroscopy.

Purified samples were concentrated to dryness using the GeneVac (EZ-2 Series) before the addition of deuterated solvent for NMR analysis. NMR analysis was performed on the Bruker DXP 400 MHz or 800 MHz spectrometer. Water suppression was achieved by a w5 WATERGATE pulse sequence or a zgpcpr pulse sequence. Spectra processing was performed using MestReNova software.

HRMS.

Agilent 6510 QTOF coupled to an Agilent 1200 series LC. 5 μl of sample is flow injected at 0.3 ml/min 50 % ACN 0.1 % FA, using ESI in either positive or negative mode (as required, stated in the text).

Nanoparticle Characterisation

To characterise nanoparticle size distribution of HRSC, UV-Vis spectrometry was employed to enable comparisons between different batches. HRSC were shown to have an absorption maximum (λ_{max}) of the plasmon resonance band of 411 nm, which was in agreement with the literature (Leopold and Lendl, 2003). A sharp UV-Vis peak also identified that there was a narrow size distribution amongst the colloidal nanoparticles.

SERS optimisation

Prior to experiments, many different parameters were varied to attain optimal SERS profiles of each analyte. A large focus was placed on the choice of colloidal nanoparticles, along with investigations into laser excitation wavelength, analyte concentration, pH, aggregating agent, aggregation time, acquisition time, *etc.*

Biotransformation Sample Preparation

Optimisation of FDH reaction conditions, along with details of protein expression and purification, are previously reported in Chapter 2: 'Exploring Surface Enhanced Raman Scattering as a Method to Identify Regioselective Flavin-Dependent Halogenation'. Optimal conditions to test organosulfur substrate activity were: FDH (1×10^{-5} M), Fre (1×10^{-6} M), GDH (6×10^{-6} M), FAD (7.5×10^{-6} M), NADH (2×10^{-4} M), MgCl_2 (2×10^{-2} M), glucose (2×10^{-2} M) and organosulfur substrate (5×10^{-4} M) in a total volume of 100 μL in potassium phosphate buffer (1×10^{-2} M, pH 7.0). Reaction samples were incubated at 37 °C with agitation for 1 - 16 h, before being quenched (95 °C, 5 min). Precipitated proteins were removed by centrifugation (21,000 $\times g$ for 6 min) before analysis *via* HPLC or SERS.

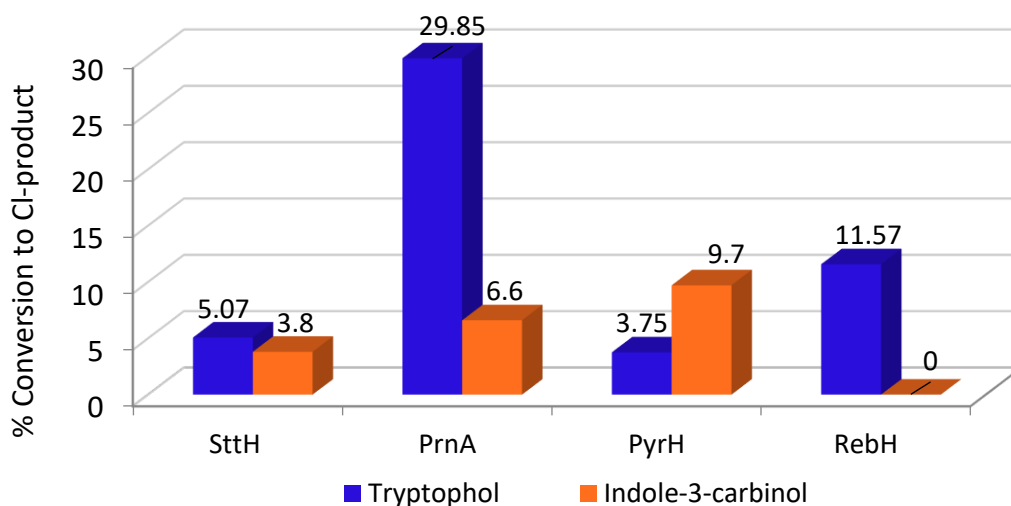


Figure S3.1. Figure showing comparative activity of four different flavin-dependent tryptophan halogenases towards tryptophol (blue) and indole-3-carbinol (orange). Assay conditions: [5×10^{-4} M] substrate, SttH/PrnA/PyrH/RebH (1×10^{-5} M), Fre (1×10^{-6} M), GDH (6×10^{-6} M), FAD (7.5×10^{-6} M), NADH (2×10^{-4} M), MgCl_2 (2×10^{-2} M), glucose (2×10^{-2} M) in potassium phosphate buffer (1×10^{-3} M, pH 7.0). 16 h incubation at 37 °C, 800 rpm (single measurements).

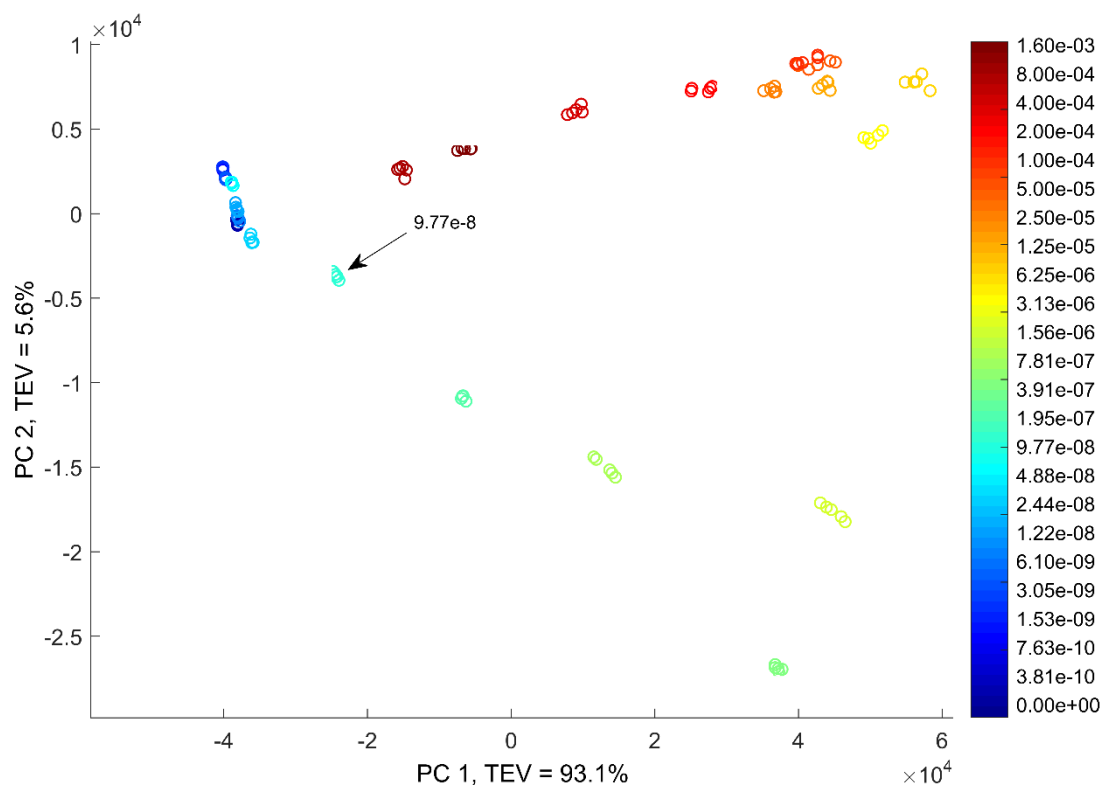


Figure S3.2. Principal component analysis (PCA) scores plot of thiol **1** at 24 different concentrations (5 replicates shown and entire spectral range used). Colour bar represents decreasing concentrations (red to blue) with concentrations labelled. Initial concentration of 1.6×10^{-3} M, ensuing concentrations were reduced by 50 %. The LOQ has been highlighted as 9.77×10^{-8} M. Tight clustering between replicate data highlights the high reproducibility of the SERS responses. TEV = total explained variance in the first 2 PCs.

$$\text{LOD} = \frac{((3 \times \text{SD of blank}) \pm c)}{m} = 2.42 \times 10^{-8} \text{ M} \quad \text{Equation S3.1.}$$

Limit of detection (LOD) equation SD = standard deviation of colloidal blank, c = y intercept, m = the gradient of a straight line. The LOD for thiol **1** was calculated as 24.2 nM with HRSC

$$\text{LOQ} = \frac{((10 \times \text{SD of blank}) \pm c)}{m} = 9.77 \times 10^{-8} \text{ M} \quad \text{Equation S3.2.}$$

LOQ equation SD = standard deviation of colloidal blank, c = y intercept, m = the gradient of a straight line. 97.7 nM was identified as the LOQ for thiol **1** with HRSC.

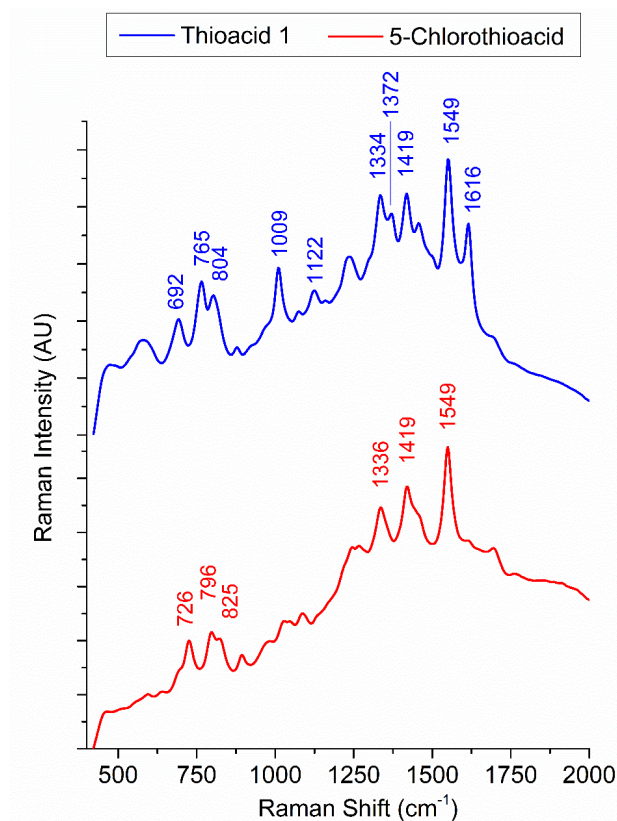


Figure S3.3. SERS spectra ($n = 5$) of thioacid **1** (blue) and 5-chlorothioacid (red). A shorter spectral range ($400 - 2000 \text{ cm}^{-1}$) is shown as no discernible peaks occur outside of these perimeters. Total concentration ($2.5 \times 10^{-5} \text{ M}$), 5x replicates, 633 nm laser excitation, 200 μL HRSC, 150 μL thioacid **1**/5-chlorothioacid solution, 50 μL KNO_3 aggregating agent ($5 \times 10^{-1} \text{ M}$), 8 min aggregation, 20 s acquisition. Data were SNV normalised.

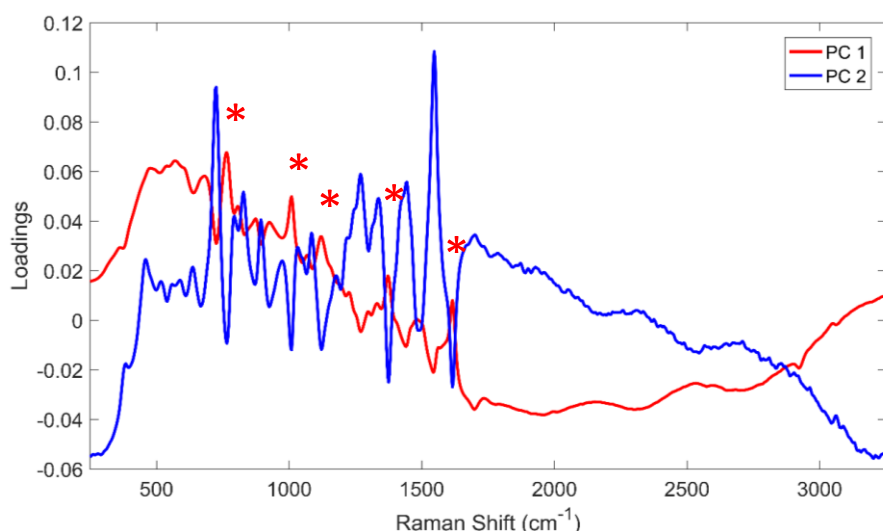


Figure S3.4. Loadings plot of corresponding PCA for dataset 1 (Figure 3.5). Main peaks enabling PC separation along PC 1 are highlighted with asterisks (*) coloured in red, from left to right, 765, 1009, 1122, 1372 and 1616 cm^{-1} . All aforementioned peaks are characteristic of thioacid **1** and are absent in 5-chlorothioacid. Separation along PC 2 arises from subtler peaks, such as shifts in wavenumbers and changes in peak intensity.

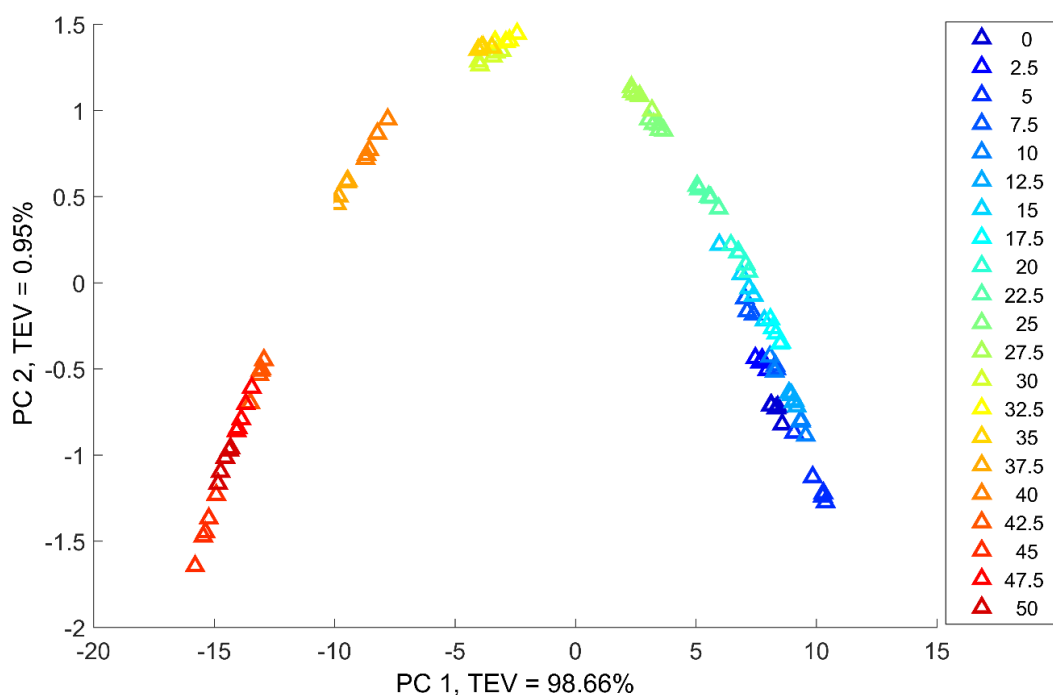


Figure S3.5. PCA scores plot of known mixtures of thioacid **1** and 5-chlorothioacid, dataset 2. PCA performed on the full spectral range (203 – 3400 cm^{-1}). Total concentration (5×10^{-5} M), 5x replicates, 633 nm laser excitation, 200 μL HRSC, 200 μL thioacid **1**/5-chlorothioacid mixture, 50 μL KNO_3 aggregating agent (1×10^{-1} M), 20 s acquisition time. Data were SNV normalised prior to chemometric analysis.

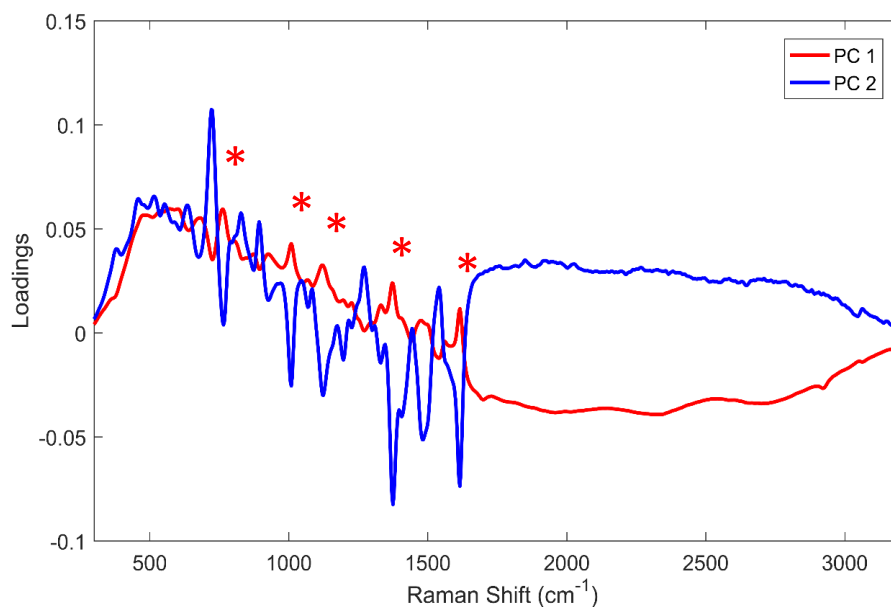


Figure S3.6. Loadings plot of corresponding PCA for dataset 2 (Figure S3.5). Main peaks enabling PC separation along PC 1 are identical to dataset 1; from left to right, 765, 1009, 1122, 1372 and 1616 cm^{-1} (denoted by asterisks (*) in red). Separation along PC 2 arises from subtler peaks, such as shifts in wavenumbers and changes in peak intensity, again in sync with dataset 1.

SEM images identified that the size and shape distribution of the nanoparticles was good, images are shown Figure S3.1. The average nanoparticle size was $\sim 30 \pm 5$ nm, once again agreeing with the literature.

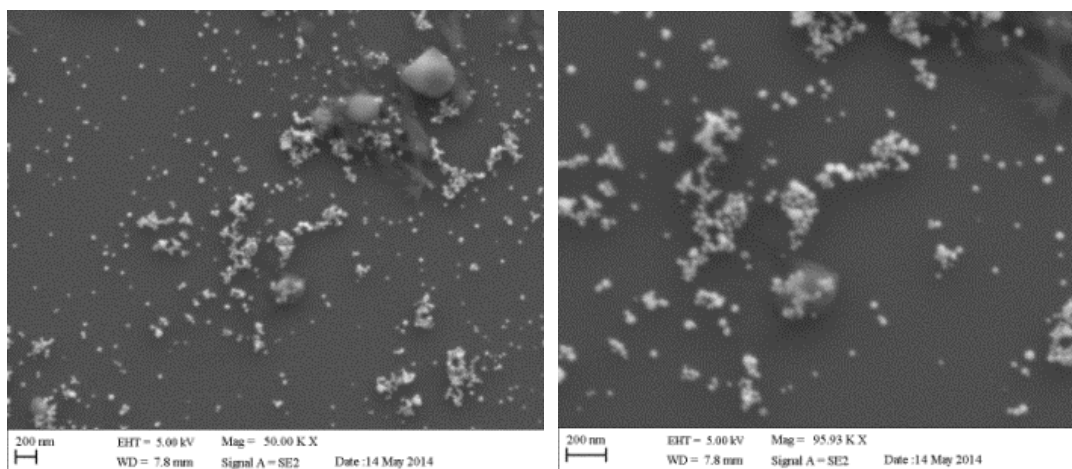
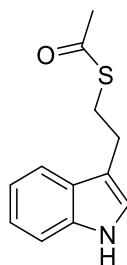


Figure S3.7. SEM images of HRSC on silicon. Left, depicts images of 50x magnification and a scale bar of 200 nm. Right, 96x magnification and 200 nm scale bar.

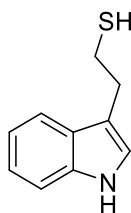
Substrate Synthesis and Characterisation



Thioacetate 1

Triphenylphosphine (5.25 g, 20.0 mmol) was dissolved in anhydrous THF (75 mL) and cooled to 0°C. DIAD (3.94 mL, 4.05 g) was added drop-wise under thorough stirring for 0.5 h. A solution of thioacetic acid (1.53 g, 20.0 mmol) and tryptophol (0.81 g, 5.0 mmol) in anhydrous THF (25 mL) was added drop-wise to the reaction mixture. Stirring at 0°C was continued for 1 h, then left to return to RT overnight with stirring. After 16 h, THF was removed *in vacuo*, the orange solid washed with cold Et₂O (2 x 25 mL) and dissolved in the minimum volume of EtOAc (~2 mL). The residue was purified by flash-column chromatography twice, an incremental gradient was used to remove PPh₃=O, eluting with hexane - ethyl acetate (7:3) to give the respective **thioacetate 1** (863.6 mg, 78 %) as a pale orange solid.

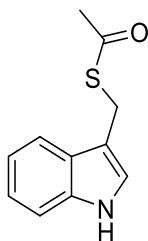
¹H NMR (400 MHz, Acetonitrile-*d*₃) 9.11 (1 H, s, NH), 7.68 – 7.62 (1 H, m, Ar-H), 7.42 (1 H, dd, *J* 8.0, 2.0 Hz, Ar-H), 7.21 – 7.04 (3 H, m, Ar-H), 3.24 – 3.15 (2 H, m, CH₂), 3.06 – 2.96 (2 H, m, CH₂), 2.33 (3 H, s, CH₃). **¹³C NMR** (101 MHz, Acetonitrile-*d*₃) 195.4 (C=O), 136.1 (Ar-C), 126.9 (Ar-C), 122.2 (Ar-C), 121.2 (Ar-C), 118.5 (Ar-C), 118.1 (Ar-C), 113.3 (Ar-C), 111.0 (Ar-C), 29.6 (CH₃), 29.3(CH₂), 25.0 (CH₂). **HRMS-ESI** [M+H]⁺ expected *m/z* 220.0791 observed *m/z* 220.0788 (± 1.36 ppm).



Thiol 1

Thioacetate **1** (226.9 mg, 1.03 mmol) was dissolved in degassed MeOH (5 mL) under inert atmosphere. NaOH (82.4 mg, 2.10 mmol) dissolved in degassed H₂O (1.25 mL) was added drop wise to the reaction mixture, which was refluxed under inert atmosphere for 2 h and then left to return to RT. Degassed 2 M HCl (3 mL) was added to neutralise the reaction mixture, forming a pale orange solid. Degassed H₂O (5 mL) was added and the organic layer which was extracted with degassed Et₂O (20 mL), washed with H₂O (3 x 10 mL), dried over MgSO₄ and concentrated under reduced pressure. The residue was rapidly purified by flash-column chromatography as thiols are susceptible to oxidation. The product was eluted with hexane - ethyl acetate (1:1) yielding **thiol 1** (129.84 mg, 71 %) as a pale-yellow solid.

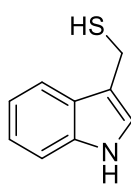
¹H NMR (400 MHz, Acetonitrile-*d*₃) 9.08 (1 H, s, NH), 7.57 (1 H, dq, *J* 8.0, 1.0 Hz, Ar-H), 7.40 (1 H, dt, *J* 8.2, 1.0 Hz, Ar-H), 7.17 – 7.01 (3 H, m, Ar-H), 3.03 (2 H, t, *J* 7.4 Hz, CH₂), 2.83 (2 H, q, *J* 7.4 Hz, CH₂), 1.63 (1 H, t, *J* 7.8 Hz, SH). **¹³C NMR** (101 MHz, Acetonitrile-*d*₃) 136.2 (Ar-C), 126.8 (Ar-C), 122.4 (Ar-C), 121.1 (Ar-C), 118.5 (Ar-C), 118.1 (Ar-C), 113.1 (Ar-C), 111.0 (Ar-C), 29.4 (CH₂), 24.6 (CH₂). **HRMS-ESI** [M+H]⁺ expected *m/z* 178.0685 observed *m/z* 178.0681 (± 2.25 ppm).



Thioacetate 2

Triphenylphosphine (5.25 g, 20.0 mmol) was dissolved in anhydrous THF (75 mL) and cooled to 0°C. DIAD (3.94 mL, 4.05 g) was added drop-wise under rapid stirring for 0.5 h. Thioacetic acid (1.53 g, 20.0 mmol) and indole-3-carbinol (0.736 g, 5.0 mmol) in anhydrous THF (25 mL) was added drop-wise to the reaction mixture. Stirring at 0°C was continued for a further 1 h, then left to return to RT overnight with stirring. After 16 h, THF was removed *in vacuo*, the yellow liquid washed with cold Et₂O (2 x 25 mL) and diluted with EtOAc (~1 mL). The mixture was purified by flash-column chromatography twice, an incremental gradient was used, eluting with hexane - ethyl acetate (7:3) affording **thioacetate 2** (0.87 mg, 85 %) as a yellow liquid.

¹H NMR (400 MHz, Acetonitrile-*d*₃) 9.16 (1 H, s, NH), 7.55 (1 H, dq, *J* 8.0, 1.0 Hz, Ar-H), 7.41 (1 H, dt, *J* 8.2, 1.0 Hz, Ar-H), 7.25 – 7.11 (2 H, m, Ar-H), 7.08 (1 H, ddd, *J* 8.0, 7.0, 1.2 Hz, Ar-H), 4.34 (2 H, s, CH₂), 2.31 (3 H, s, CH₃). ¹³C NMR (101 MHz, Acetonitrile-*d*₃) 195.3 (C=O), 136.1 (Ar-C), 126.2 (Ar-C), 123.7 (Ar-C), 121.5 (Ar-C), 118.9 (Ar-C), 118.1 (Ar-C), 111.2 (Ar-C), 110.4 (Ar-C), 29.4 (CH₃), 23.8 (CH₂). HRMS-ESI [M+H]⁺ expected *m/z* 206.0634 observed *m/z* 206.0630 (± 1.94 ppm).

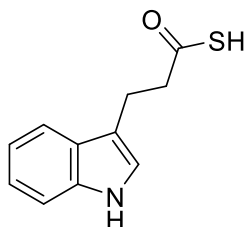


Thiol 2

Thioacetate **2** (220.0 mg, 1.07 mmol) was dissolved in degassed MeOH (2 mL), to this a solution of NaBH₄ (60.71 mg, 1.61 mmol) in degassed MeOH (2 mL) was added drop-wise and the mixture was refluxed for 1.5 h under inert atmosphere. Degassed 2 M HCl was added to quench the reaction until pH 6 was reached, degassed H₂O (3 mL) was added, along with degassed EtOAc (6 mL). The product was extracted using EtOAc, washed with H₂O, dried over MgSO₄ and concentrated *in vacuo* giving **thiol 2** (101.30 mg, 58 %) a pale-yellow solid.

¹H NMR (400 MHz, Acetonitrile-*d*₃) 9.13 (1 H, s, NH), 7.58 (1 H, dq, *J* 8.0, 1.0 Hz, Ar-H), 7.41 (1 H, dt, *J* 8.2, 1.0 Hz, Ar-H), 7.23 – 6.97 (3 H, m, Ar-H), 3.88 (2 H, d, *J* 0.8 Hz, CH₂), 2.17 (1 H,

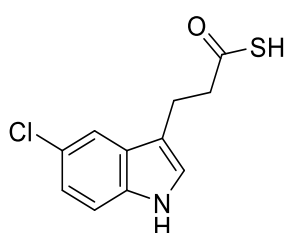
s, SH). ^{13}C NMR (101 MHz, Acetonitrile- d_3) 136.4 (Ar-C), 126.5(Ar-C), 123.3(Ar-C), 121.4(Ar-C), 118.6(Ar-C), 118.5(Ar-C), 111.3(Ar-C), 111.0(Ar-C), 26.0 (CH_2). HRMS-ESI $[\text{M}-\text{H}]^-$ expected m/z 162.0383 and 163.0416 observed m/z 162.0380 and 163.0406 (± 1.85 and 6.13 ppm).



Thioacid 1

A solution of 3-indolepropionic acid (94.61 mg, 0.5 mmol) and Lawesson's Reagent (113.25 mg, 0.28 mmol) in DCM (2 mL) was subjected to MW irradiation (100 °C, 10 min) in a sealed tube. The bright orange solution was purified using reverse phase C18 silica, product elution with 10 % MeOH in H_2O gave a dark yellow liquid, **thioacid 1** (53.36 mg, 52 %).

^1H NMR (400 MHz, Chloroform- d) 7.99 (1 H, s, NH), 7.59 (1 H, d, J 7.8 Hz, Ar-H), 7.37 (1 H, d, J 8.0 Hz, Ar-H), 7.22 (1 H, td, J 8.2, 7.6, 1.2 Hz, Ar-H), 7.14 (1 H, ddd, J 8.2, 7.0, 1.2 Hz, Ar-H), 7.02 (1 H, d, J 2.4 Hz, Ar-H), 4.47 (1 H, s, SH), 3.13 (2 H, ddt, J 9.6, 8.2, 2.2 Hz, CH_2), 3.03 (2 H, ddd, J 8.2, 7.2, 2.2 Hz, Ar-H). ^{13}C NMR (101 MHz, Chloroform- d) 197.4 (C=O), 136.3 (Ar-C), 127.0 (Ar-C), 122.2 (Ar-C), 121.6 (Ar-C), 119.5 (Ar-C), 118.6 (Ar-C), 114.1 (Ar-C), 111.2 (Ar-C), 46.3 (CH_2), 20.9 (CH_2). HRMS-ESI $[\text{M}-\text{H}]^-$ expected m/z 204.0489 observed m/z 204.0495 (± 2.94 ppm).



5-Chlorothioacid

The conversion of 3-indolepropionic acid to 5-chloro-3-indolepropionic acid was performed enzymatically using PyrH (C-5 FDH). Biotransformation conditions: PyrH (1×10^{-5} M), Fre (1×10^{-6} M), GDH (6×10^{-6} M), 3-indolepropionic acid (2.2×10^{-3} M), MgCl_2 (2×10^{-2} M), FAD (7.5×10^{-6} M), NADH (2×10^{-4} M), glucose (2×10^{-2} M) in 15 mL potassium phosphate buffer (1×10^{-2} M, pH 7.0). Assays were incubated for 16 h at 30 °C with shaking, after this time had passed, the reaction was quenched with MeOH (1:1 v/v). The biotransformation was performed many times (>10) and the concentrated crude mixture (*via* genevac concentrator) was

combined prior to purification by semi-preparative HPLC, producing 5-chloro-3-indolepropionic acid (21.24 mg).

A solution of purified 5-chloro-3-indolepropionic acid (21.24 mg, 0.095 mmol) and Lawesson's Reagent (21.13 mg, 0.052 mmol) in DCM (0.4 mL) was subjected to MW irradiation (100 °C, 11 min) in a sealed tube. The bright orange solution was purified using reverse phase C18 silica, product elution with 10 % MeOH in H₂O gave a pale-yellow liquid, **5-chlorothioacid** (12.19 mg, 54 %).

¹H NMR (400 MHz, Chloroform-*d*) 8.02 (1 H, s, NH), 7.57 – 7.51 (1 H, m, Ar-H), 7.29 – 7.24 (1 H, m, Ar-H), 7.15 (1 H, dd, *J* 8.6, 2.0 Hz, Ar-H), 7.06 – 6.98 (1 H, m, Ar-H), 4.49 (1 H, s, SH), 3.14 – 3.04 (2 H, m, CH₂), 3.04 – 2.95 (2 H, m, CH₂). ¹³C NMR (101 MHz, Chloroform-*d*) 195.5 (C=O), 133.0 (Ar-C), 126.5 (Ar-C), 123.7 (Ar-C), 121.5 (Ar-C), 120.9 (Ar-C), 116.5 (Ar-C), 112.3 (Ar-C), 110.6 (Ar-C), 44.5 (CH₂), 19.0 (CH₂). HRMS-ESI [M-H]⁻ expected *m/z* ³⁵Cl 238.0099 and ³⁷Cl 240.0069, observed *m/z* 238.0103 and 240.0074 (± 1.68 and 2.08 ppm).

FDH Active Site Interactions

The crystal structure of PrnA (C-7 FDH) was published by Dong *et al.*, (Dong *et al.*, 2005) these structural insights into the active site residues and binding interactions can help our understanding of why organosulfur substrates did not demonstrate halogenase activity. Figure S3.8 identifies the main interactions within the active site of PrnA, resulting in regioselective halogenation at position C-7 on the indole ring. The carboxylic acid and amino group of the hydrocarbon chain are involved in a series of hydrogen bonds to tyrosine (Y443 and Y444), phenylalanine (F454) and glutamate (E450). The indole moiety also participates in a hydrogen bond with glutamate (E346), as well as π-π stacking interactions (H101, F103 and W455).

If we consider the hydrogen bonding interactions in relation to the organosulfur substrates mentioned throughout, at least one of these H-bonding interactions are disrupted (Figure S3.8). Thiols **1** and **2**, along with their precursors thioacetate **1** and **2** (Scheme 3.1), retain interactions at E346 residue only, along with π-π stacking interactions. Unlike nitrogen and oxygen, sulfur exhibits very weak H-bonding, thus these substrates would experience weakened interactions at Y444, together with a lack of amino group preventing interactions at Y443, E450 and F454. Evidently, the loss of these interactions results in a loss of activity.

3-(1-benzothiophen-3-yl)-L-alanine (Figure 3.5B) is a structural analogue of L-tryptophan, which retains all interactions along the hydrocarbon sidechain. H-bonding at E346 is no longer permitted as the sulfur within the benzothiophene ring will not interact at this residue (no free H-bond). Absence of this interaction may prevent activity, or, as previously discussed, it may be due to the increase in atom size of the sulfur compared to nitrogen. Finally, thioacid **1** (Figure 3.5C) experiences a similar lack of H-bonding interactions surrounding the hydrocarbon sidechain, alike to thiol **1/2** and thioacetates **1/2**.

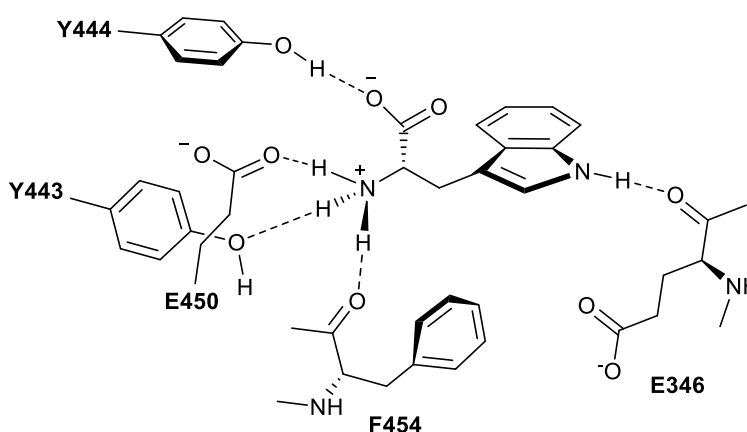


Figure S3.8. PrnA active site with L-tryptophan bound. Key active site residues partaking in hydrogen bonding interactions are shown. Residues involved in π - π stacking (H101, F103 and W455) above and below the indole moiety are removed for clarity.

In summary, it is reasonable to believe that these organosulfur substrates are inactive due to missing H-bonding interactions which are important for FDH activity. Conversely, it could be a consequence of the increased atomic radius and subsequent bond lengths associated with sulfur, thus the constrained active site may not be able to facilitate the acceptance of sulfur-containing substrates.

REFERENCES

- DONG, C., FLECKS, S., UNVERSUCHT, S., HAUPT, C., VAN PÉE, K.-H. & NAISMITH, J. H. 2005. Tryptophan 7-Halogenase (PrnA) Structure Suggests a Mechanism for Regioselective Chlorination. *Science*, 309, 2216-2219.
- LEOPOLD, N. & LENDL, B. 2003. A New Method for Fast Preparation of Highly Surface-Enhanced Raman Scattering (SERS) Active Silver Colloids at Room Temperature by Reduction of Silver Nitrate with Hydroxylamine Hydrochloride. *The Journal of Physical Chemistry B*, 107, 5723-5727.

4 Discrimination of Enzyme Enantioselectivity using *in situ* Surface Enhanced Raman Scattering Spectroscopy

Heidi Fisk[†], Yun Xu[†], Peter Sutton[‡], Jason Micklefield[†] and Royston Goodacre[†]

[†]School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK

[‡]Department of Chemical, Biological and Environmental Engineering, Autonomous University of Barcelona, Cerdanvola del Vallès, Barcelona, Spain

4.1 DECLARATION

This chapter consists of one piece of work written in the format of a journal article but not currently submitted for publication.

As primary author, I carried out all experimental work, data analysis and write-up. Dr. Yun Xu provided assistance with the chemometrics, including MCR-ALS modelling. Dr. Peter Sutton gave advice and insights surrounding the synthesis of substrates and enzyme classes. Prof. Jason Micklefield provided advice and support throughout these investigations, particularly towards the biotransformation. Prof. Royston Goodacre, as principal investigator, gave fundamental advice and guidance throughout this research, predominantly towards aspects surrounding SERS.

4.2 ABSTRACT

More than half of today's drug compounds contain a chiral centre, as a consequence, being able to distinguish between enantiomers at these centres is an increasing concern. Racemic compounds are not only wasteful, but can also have significant, adverse side effects. Biocatalysts are gaining momentum within pharmaceutical development as they enable mild conditions and sustainable processes, but their most attractive trait is their high regio-, stereo- and enantio-selectivity. Thus, enzymes can be incorporated within synthetic routes as to avoid the formation of the undesirable enantiomeric form(s) and reduce waste. There are few analytical techniques which are capable of enantiomer discrimination and the most common method, HPLC, suffers from lengthy optimisation, extended acquisition times and low sensitivity. Within this work, we successfully demonstrate SERS as a rapid technique to monitor enzyme enantioselectivity in real-time. By employing a pseudo-enantiomer approach, thioester substrates were selectively hydrolysed and their corresponding thiol products were easily detected *in situ* (that is to say, directly within the reaction mixture) by instantaneous capture using silver nanoparticles. Two enzymes were studied, the first having high selectivity towards L-enantiomers (α -chymotrypsin). The other, experiencing low enantioselectivity and hence hydrolysing both L- and D-enantiomers simultaneously (pig liver esterase). This SERS method offers high sensitivity (LOD of 1.56 μ M), rapid acquisition times (10 s) and successful discrimination between enantioselective activity in real-time.

4.3 INTRODUCTION

Most molecules that are of great importance to living organisms are chiral, these include amino acids, hormones, carbohydrates, nucleic acids, proteins and many natural products. In general, only one of the two enantiomeric forms of these biomolecules are produced in nature, this is a result of only a single enantiomer producing the desirable outcome (Crossley, 1995, Mohan *et al.*, 2009). This is also true within the development of chiral pharmaceuticals. The production of racemic drugs (both enantiomeric forms) can be wasteful, but more importantly, it may introduce undesirable side effects or adverse reactions, including teratogenicity. It is estimated that more than 50 % of marketed drugs are chiral. Evidently it is of great importance to be able to characterise and separate enantiomers to avoid toxic drug preparations (Walther and Netscher, 1996, Rentsch, 2002, Katzung *et al.*, 2016). Consequently, the development of techniques which can discriminate between enantiomeric

forms or methods to produce single enantiomers, is a growing field within academia and industry.

Commonly used analytical methods for chiral analysis include, high performance liquid chromatography (HPLC), gas chromatography (GC), supercritical fluid chromatography (SFC) and capillary electrophoresis (CE) (Nguyen *et al.*, 2006). The most extensively used method being HPLC, which can be performed by means of direct or indirect measurements. Direct analysis usually involves specialist chiral columns which require a significant amount of optimisation and frequently involve long acquisition times. Indirect approaches use chiral derivatization reagents. These reagents will react with the analyte under investigation (*i.e.* chiral drug) provided that specific functional groups are present. The derivatization reagent also has a chiral centre, leading to a diastereomer product which enables separation using regular HPLC columns (*i.e.* C18) (Toyo'oka, 2002). Despite a lot of research around the development of faster, universal and more information rich chiral techniques, little progress has been made.

Another approach within pharmaceutical development is the synthesis of single enantiomers: asymmetric synthesis. For this, the application of biocatalysts has been fundamental as they exhibit very high enantiomeric excess (*ee*), substrate specificity, along with mild operating conditions (Gröger, 2010). Yet their high substrate specificity can also limit their widespread application, therefore, protein engineering has been an important methodology to modify their properties (Kazlauskas and Bornscheuer, 2009). Directed evolution in particular, has rapidly emerged as a technique for improving an enzymes characteristics, examples include increasing their activity, substrate scope or switching their intrinsic enantioselectivity (Reetz, 2004, May *et al.*, 2000, Hibbert and Dalby, 2005, Chica *et al.*, 2005). These developments further support the application of biocatalysts within industry for the synthesis of enantiomerically pure pharmaceuticals.

Within this research, we will explore surface-enhanced Raman scattering (SERS) spectroscopy as a sensitive and rapid technique, to characterise enantioselectivity expressed by an enzyme (*i.e.* L- and/or D-selectivity). This vibrational technique has very recently proven to be extremely successful at quantifying biotransformations, achieving results that are in sync with more common analytical techniques, such as HPLC and UV/vis (Westley *et al.*, 2016, Hollywood *et al.*, 2010). The Goodacre group has previously investigated the sensitivity of SERS to distinguish between enantiomers. These studies involved the preparation of chiral colloids using L- and D-enantiomeric forms of reducing agents (literature known examples include tyrosine-reduced nanoparticles (Bhargava *et al.*, 2005)). However, SERS proved

unsuccessful at distinguishing between enantiomers, even when chiral nanoparticles were used. By contrast, our approach within this work is to utilise pseudo-enantiomers to generate molecule specific changes that are discriminant for identification in SERS. Pseudo-enantiomers retain the same substructure centred around the chiral moiety, however the enzymatically released counterpart differs slightly, depending on whether it is the L- or D-enantiomer, see Figure 4.1. This approach was shown to be extremely successful by Fernández-Álvarez, *et al.* (Fernández-Álvarez *et al.*, 2011). They demonstrated that pseudo-enantiomers modelled on an esterase substrate could enable selective recognition of *R*- or *S*-enantioselectivity. The chiral carboxylic acid was coupled to either a carbon source (*R*-enantiomer), leading to cell proliferation, or a brominated compound (*S*-enantiomer) causing cell death.

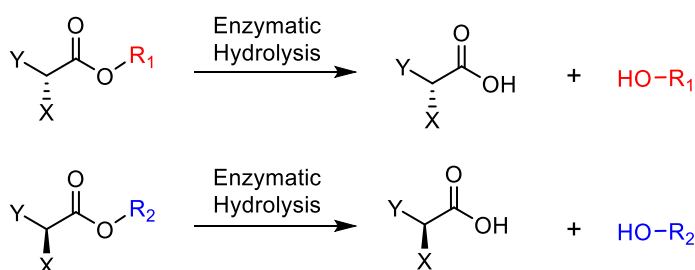


Figure 4.1. Basic overview of a pseudo-enantiomer approach towards esterase activity. If Y and X share identical functional groups between the two compounds, but are the two different enantiomeric forms (L- and D-/R- and S-), then they are enantiomers of one another as a chiral centre is present. The R-group differs between the two enantiomers (R₁ versus R₂), becoming pseudo-enantiomers. Upon enzyme-catalysed hydrolysis (i.e. esterase), the different R-groups will indicate which enantiomer was selectively hydrolysed, thus uncovering enzyme enantioselectivity.

As one can observe in Figure 4.1, if an enzyme demonstrates low enantioselectivity, *i.e.* hydrolyses both L- and D-compounds, then 6 different components may be present within the reaction mixture. Our previous research has explored complex biotransformations, namely flavin-dependent halogenases, which led to difficulties within the SERS analysis of the substrate and halogenated product amongst essential complex cofactors. Following these observations, selective thiol-functionalisation of the substrate and its corresponding product proved extremely successful, facilitating selective detection of these reaction components over the complex cofactors (see Chapter 3: ‘Thiol-Functionalised Substrates for Sensitive and Rapid Detection of Regioselective Flavin-Dependent Halogenation using SERS’). This was the result of the thiol group having a very high affinity towards the silver nanoparticles, exceeding the interactions of other un-modified molecules.

Herein, we aim to combine the pseudo-enantiomer approach with thiol-functionalised substrates. Thioester substrates will be designed so that hydrolysis of the thioester bond generates a free thiol group for sensitive SERS detection. The R-group attached to the thiol will be different between the two enantiomer forms, so that the cleaved R-group will be characteristic of the enantioselective behaviour expressed by the enzyme.

4.4 EXPERIMENTAL

4.4.1 Reagents and Materials.

All chemical reagents were of analytical grade and used with no additional purification. Commercial enzymes, α -chymotrypsin from bovine pancreas (EC 3.4.21.1) and pig liver esterase (EC 3.1.1.1) were purchased from Sigma Aldrich Ltd. (Dorset, UK).

4.4.2 Raman Instrumentation.

A DeltaNu Advantage 200A portable Raman spectrometer (DeltaNu, Laramie, WY, USA) was used for spectra collection using a HeNe 785 nm laser. The typical power on the sample was ~60 mW.

4.4.3 Biotransformation Sample Preparation

4.4.3.1 Small scale (1 mL) SERS biotransformation. Real-time SERS samples contained 625 μ L hydroxylamine-reduced silver colloid (HRSC), 100 μ L t-butanol (internal standard), potassium phosphate buffer (5×10^{-3} M final concentration, pH 6.2) and L- and D-thioesters at a total concentration of 2×10^{-5} M (1×10^{-5} M each). The introduction of α -chymotrypsin (2×10^{-9} M) initiates the reaction. SERS spectra were continually acquired, using a 10 s acquisition time.

4.4.3.2 Large scale (20 mL) SERS biotransformation with comparative HPLC. Real-time SERS samples contained 12.5 mL HRSC, 2 mL t-butanol (internal standard), potassium phosphate buffer (5×10^{-3} M final concentration, pH 6.2) and L- and D-thioesters at a total concentration of 2×10^{-5} M. The introduction of α -chymotrypsin (2×10^{-9} M) or PLE (50 μ L of 10 mg/mL ammonium sulfate solution) initiates the reaction. For SERS analysis, 0.4 mL was removed from the reaction mixture and analysed instantaneously using a 10 s acquisition time (mainly every minute over the monitoring period). For comparative HPLC analysis (largely every other

min), 1 mL of the reaction mixture was quenched using 1 % trifluoroacetic acid (TFA), followed by centrifugation (21,000 xg for 6 min) and overnight (16 h) lyophilisation of the supernatant affording a solid. This solid was re-dissolved in acetonitrile (70 μL) and centrifuged (1,000 xg for 6 min) prior to HPLC analysis (24 min acquisition time, chiral column). See SI, supplementary methods for further information.

4.4.4 Thioester Synthesis.

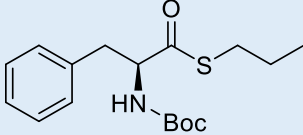
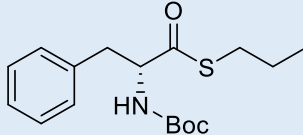
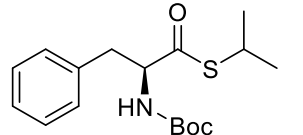
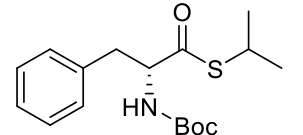
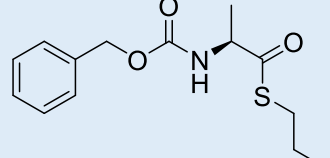
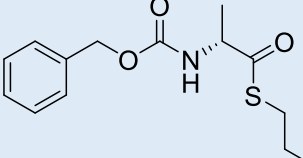
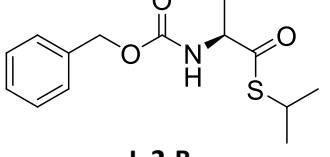
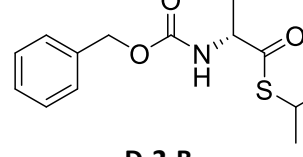
Details surrounding the synthesis, characterisation and confirmation of thioester enantiopurity can be found in the SI, section 'Thioester Synthesis and Characterisation'.

4.5 RESULTS AND DISCUSSION

4.5.1 Thioester Design and Optimisation

To enable successful enzyme-catalysed thioester hydrolysis, a major consideration was the design of suitable thioester substrates and choosing appropriate enzymes that would be active towards these compounds. Two different classes of commercial enzyme were studied, the first being α -chymotrypsin from bovine pancreas, which selectively cleaves peptide bonds with aromatic or large hydrophobic side chains (tyrosine, tryptophan, phenylalanine) on the carboxyl end of the bond. Secondly, an esterase from pig liver (PLE) which catalyses the hydrolysis of a broad range of esters to their corresponding carboxylic acids. Referring to Table 4.1, the first subset of thioester substrates (A) were designed around *N*-Boc-protected phenylalanine, knowing that L-phenylalanine esters are selectively hydrolysed by α -chymotrypsin. Secondly, a publication in 1994 established that a range of different lipases gave a low *ee* towards Cbz-protected alanine esters (Duthaler, 1994), *i.e.* active towards both L- and D-enantiomers. As we wanted to demonstrate the potential of this SERS approach towards screening within mutagenic strategies, such as those used in directed evolution, a low selectivity enzyme would identify if this method could correctly identify both thiol-hydrolysis products, with the potential to quantify their relative activity. Consequently, thioester substrates towards were designed around a Cbz-protected alanine substructure, substrate subset B (Table 4.1). Screening multiple lipases and esterases identified PLE as the most active and least enantioselective enzyme (low *ee*).

Table 4.1. Table showing the structures of the 8 thioester compounds. Substrate set **A** are substrates towards α -chymotrypsin hydrolytic activity. Substrate set **B** are pig liver esterase (PLE) substrates. Each of these substrates have L- and D-enantiomeric forms, with two differing thioester groups: either 1-propanethiol or 2-propanethiol, generating 4 compounds per subset (A/B).

A/B	Thiol	L-enantiomers	D-enantiomers
SUBSTRATE A	1-propanethiol	 L-1-A	 D-1-A
	2-propanethiol	 L-2-A	 D-2-A
SUBSTRATE B	1-propanethiol	 L-1-B	 D-1-B
	2-propanethiol	 L-2-B	 D-2-B

The thiol linkers on the thioester substrates were chosen to be 1-propanethiol and 2-propanethiol. This choice was predominantly a result of their SERS responses giving rise to peaks that were characteristically different (Figure 4.2), a preference over a small group (low MW) and that they were structural isomers of one another (for further discussion, see the SI section ‘Thiol Optimisation and Monolayer Coverage’, Figure S4.1 and Table S4.1). Information surrounding SERS parameter optimisation is provided in the SI, section ‘SERS Optimisation’, along with tentative band assignments of 1-propanethiol and 2-propanethiol in Table S4.2. Initially we performed at-line analysis of the biotransformation samples, involving the precipitation and removal of protein prior to SERS analysis. Early investigations identified that the thioester starting materials and their corresponding hydrolysed carboxylic

acid products do not give SERS responses (Table 4.1), it is purely the thiol products that give rise to SERS. Consequently, t-butanol (t-BuOH) was introduced into the reaction mixture to act as an internal standard for subsequent analysis of the SERS data (t-BuOH gives a single characteristic peak at 747 cm^{-1}).

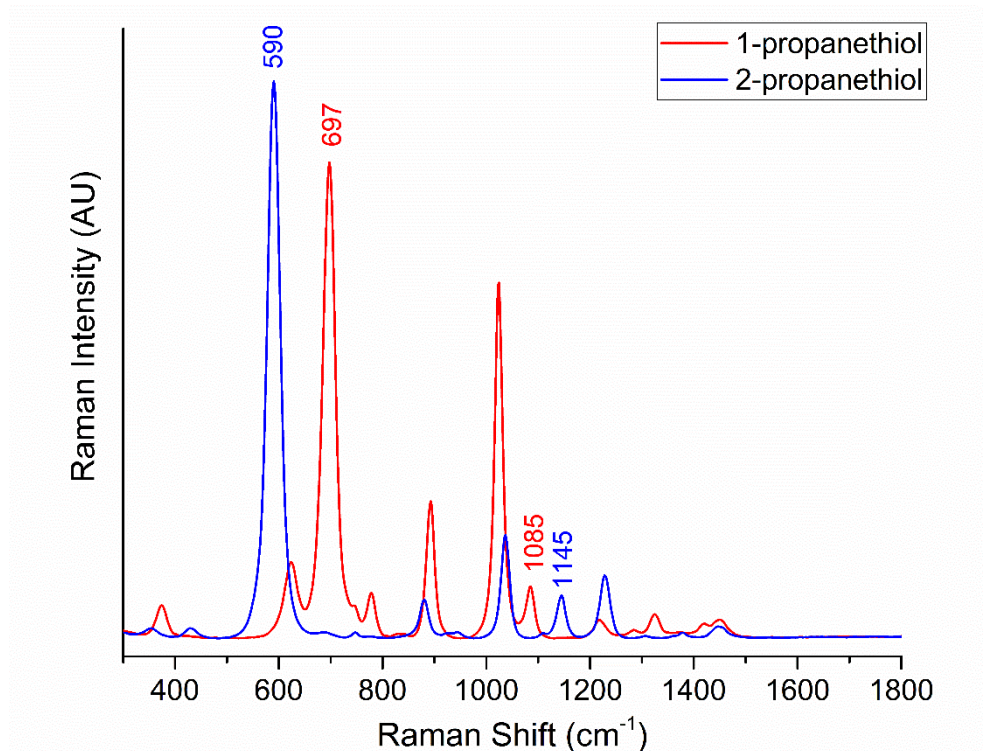


Figure 4.2. SERS spectra (average of $n = 5$) of 1-propanethiol (red) and 2-propanethiol (blue). Structural isomers (straight versus branched chain) resulting in noticeably different SERS responses. 1×10^{-5} M concentration in potassium phosphate buffer (pH 6.2, 0.01 M) with HRSC, 785 nm irradiation using a 10 s acquisition time. Data have been baseline corrected and normalised (see SI, data processing for further details).

During initial experiments, it became evident that the thiol products were oxidising and forming adverse disulfide bonds. We experienced a decreasing SERS response over time with a known thiol concentration, and a series of experiments proved that this was caused by dissolved oxygen within the reaction mixture (see SI, section ‘Oxidation: Disulfide Bond Formation’, Scheme S4.1 and Figure S4.2 for further discussion). If a proportion of the thiol was oxidising, this would significantly affect our ability to correctly identify and quantify the extent of thioester hydrolysis and hence, its enantioselectivity towards either L- and/or D-substrates. Similarly, the free thiols may also undergo transthioesterification with the unhydrolysed thioesters, which would also lead to false results (see Scheme S4.1 in SI). Therefore, we proposed that it might be possible to perform the biotransformation directly in the presence of the silver nanoparticles, facilitating instantaneous capture of the thiols as

they are generated *in situ*. This approach would avoid disulfide bond formation and enable real-time, on-line measurements of hydrolytic activity.

4.5.2 Highly Enantioselective Enzyme: α -Chymotrypsin

Initially the L-selective hydrolysis of L-1-A in the presence of D-2-A was performed, generating 1-propanethiol and *N*-Boc-L-phenylalanine products in equimolar concentrations. This 1 mL scale reaction, containing >60 % hydroxylamine-reduced silver nanoparticles, was constantly monitored using a 10 s acquisition time with a portable 785 nm Raman instrument. The corresponding SERS responses can be seen in Figure 4.3, the colour bar represents the time, enabling an increasing SERS response to be clearly visualised. Simple univariate analysis can be applied to the dataset by plotting the peak area of the major, discriminant peak at 697 cm^{-1} , shown inset. As one would expect, we can see an increasing peak area with respect to time, demonstrating that the thiols are interacting with nanoparticles upon their generation by enantioselective hydrolysis. A second real-time, 1 mL reaction was performed however using an equimolar concentrations of substrates L-2-A and D-1-A. Subsequent enzymatic hydrolysis towards the L-enantiomer produced a SERS response corresponding to 2-propanethiol, this response is shown in Figure S4.3 in SI. Controls were performed, confirming that this enantioselective hydrolysis was caused by the enzyme and not due to the metal nanoparticles behaving as a catalyst (nanozyme) (Manea *et al.*, 2004, Mancin *et al.*, 2016, Sloan-Dennison *et al.*, 2017).

As the thiols readily oxidise, a SERS calibration cannot be performed to quantify their response as it would incur many errors. Thus, it was necessary to incorporate another technique which would provide an external calibration of the reaction components, enabling predictions of the thiol concentrations *via* SERS analysis. For this, HPLC was employed to provide quantitative measurements of the reaction components, enabling an analytical calibration and training/validation for the multivariate curve resolution-alternating least squares (MCR-ALS) models. As the thioester starting materials (L- and D-forms) and the hydrolysed *N*-Boc-L-phenylalanine product do not give SERS signals, they do not interact with the metal nanoparticles. Thus, these compounds can be recovered from the reaction mixture and analysed by HPLC, unlike the thiol products which form irreversible bonds with the nanoparticles and cannot be easily recovered (nor do they contain a chromophore so they are not detectable in UV-vis or HPLC). *N*-Boc-L-phenylalanine and its associated thiol are hydrolysed in an equimolar ratio (1:1) upon enzyme catalysis, therefore, it is plausible to

hypothesise that the *N*-Boc-L-phenylalanine product and the two thioester starting materials could be used to quantify product generation, and hence the corresponding concentration of thiol being detected in SERS.

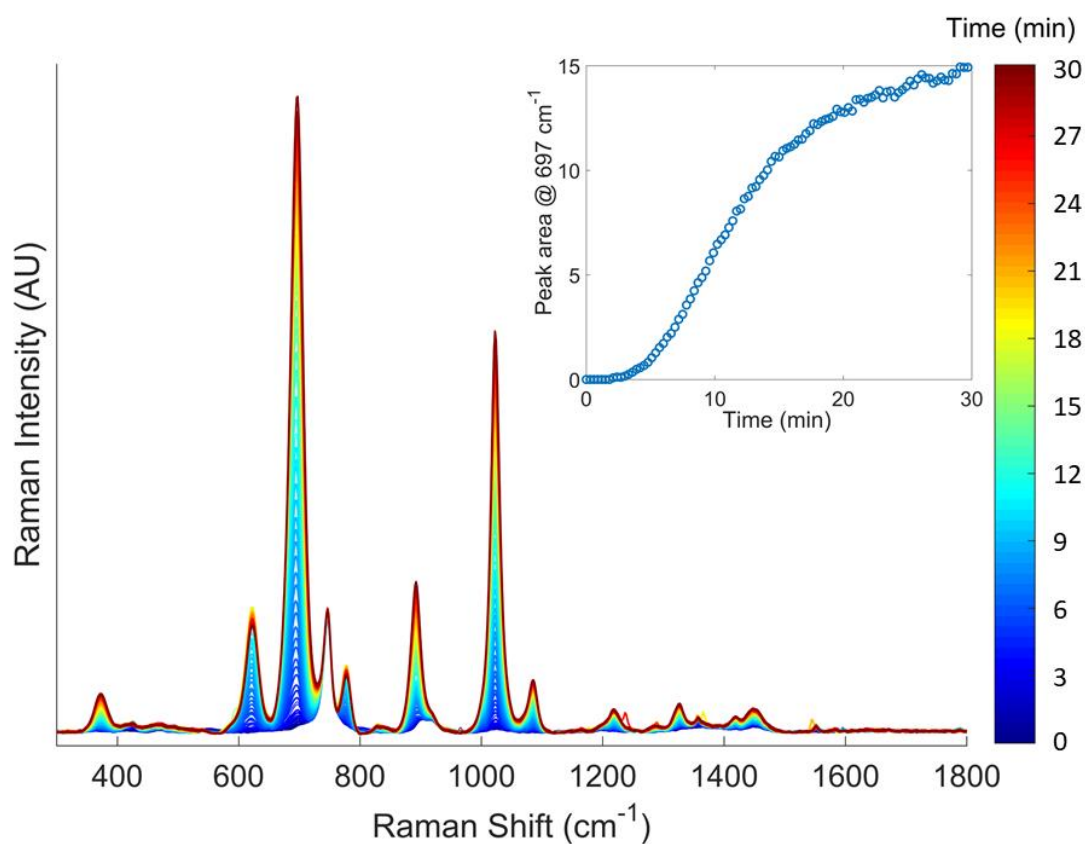


Figure 4.3. Cumulative SERS spectra taken over a 30 min time course. The colour bar highlights the increasing intensity with time ($t = 0$ in blue, $t = 30$ in red). The SERS data have been normalised to the internal standard, *t*-BuOH at 747 cm^{-1} (see SI, data processing for further details). 1 mL reaction including an equimolar ratio of substrates L-1-A and D-2-A, with a total concentration of $2 \times 10^{-5}\text{ M}$. The introduction of α -chymotrypsin ($2 \times 10^{-9}\text{ M}$) affords L-enantioselective hydrolysis, as demonstrated by the production of 1-propanethiol (L-1-A hydrolysis). Inset shows univariate analysis of the SERS response by plotting the peak area of 697 cm^{-1} against time.

The sensitivity of HPLC compared with SERS is very low. Regularly, concentrations in the range of 0.01 – 5 millimolar (10^{-3} M) are detectable using HPLC, whereas SERS commonly detects micromolar (10^{-6} M) or nanomolar (10^{-9} M) concentrations. It is essential that the SERS concentration remains below the point at which monolayer coverage is exceeded, otherwise competition at the metal surface will occur (Stewart *et al.*, 2012). Equimolar mixtures of the thiols, 1-propanethiol and 2-propanethiol, were prepared at various concentrations to determine the limit of detection (LOD) and the maximum concentration before monolayer coverage was exceeded. See the SI section ‘Thiol Optimisation and

Monolayer Coverage' and Figure S4.1 for further details. In brief, $1.56 \mu\text{M} \pm 0.156$ (1.56×10^{-6} M) was identified at the LOD combined for the two thiols and concentrations exceeding $37.5 \mu\text{M} \pm 3.75$ (3.75×10^{-5} M) led to competition at the surface (preference towards 1-propanethiol). It is worth noting that these concentrations will have some degree of error due to inherent oxidation, therefore a $\pm 10\%$ has been included to account for this. As a result, the SERS concentrations are far too low for HPLC detection, consequently, the reaction samples will require concentrating prior to HPLC analysis to enable quantification of the reaction components.

To monitor the biotransformation with both SERS and HPLC simultaneously, as well as permitting enough sample to be concentrated for HPLC analysis, the monitoring process needed to be scaled up to a 20 mL volume. Figure S4.4 in SI illustrates the overall monitoring process and the steps required for successful HPLC analysis, which include protein precipitation and removal, overnight lyophilisation of the supernatant to dryness and re-dissolving the solid in a small volume of acetonitrile prior to HPLC analysis. The SERS analysis is much simpler as the reaction sample requires no modification prior to analysis and takes only 10 s to acquire. SERS analysis must be performed instantaneously to represent the desired time point. Figure 4.4.A is the reaction schematic: within this example thioester substrates L-1-A and D-2-A were studied, ultimately generating *N*-Boc-L-phenylalanine and 1-propanethiol as products of L-selective hydrolysis. The cumulative SERS responses over a 20 min period can be seen in Figure 4.4.B, as one would expect, the intensity of the 1-propanethiol response increases with time. SERS spectra were acquired every minute, whereas HPLC analysis was performed every two mins. To be able to use the HPLC measured concentrations of *N*-Boc-L-phenylalanine to predict the SERS concentrations of 1-propanethiol, MCR-ALS was applied to the data.

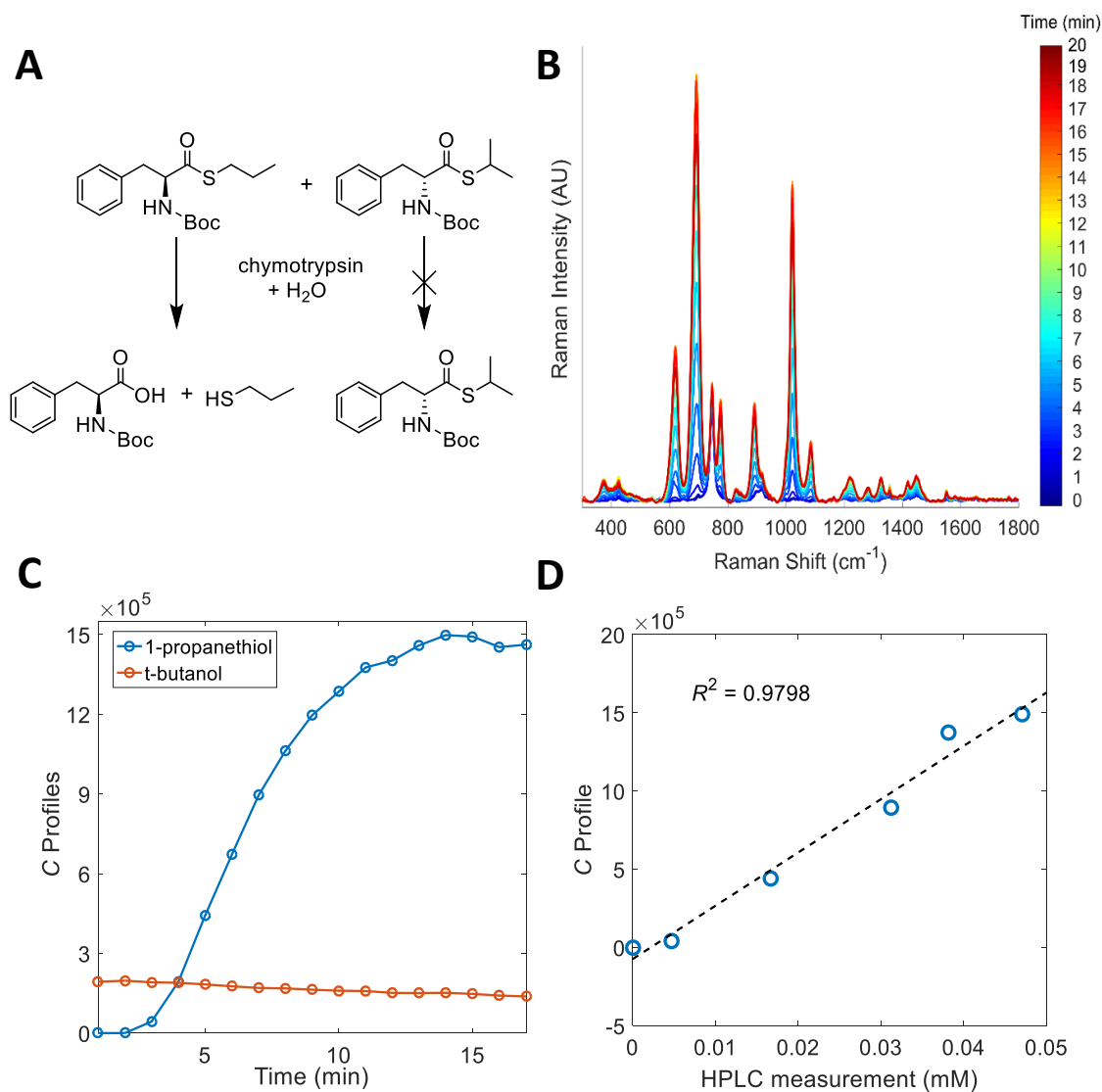


Figure 4.4. Data related to the 20 mL real-time SERS analysis paired with external HPLC measurements. A 20 mL biotransformation sample was prepared with L-1-A and D-2-A substrates at a concentration of 1×10^{-5} M each and >60 % HRSC content. Introduction of 2×10^{-9} M α -chymotrypsin resulted in L-selective hydrolysis of L-1-A only (A), generating SERS responses correlating to 1-propanethiol. The cumulative SERS spectra are shown in (B). Normalisation to the internal standard, t-BuOH (10 % content) at 747 cm^{-1} , and the application of a colour bar enable clear observations of an increasing SERS response with respect to time. MCR-ALS was performed to the SERS data and off-line HPLC analysis enabling predictions of SERS thiol concentration, shown in (C). The SERS predicted concentrations and HPLC measured concentrations are shown in (D), giving a R^2 value of 0.9798, demonstrating excellent fit.

MCR-ALS is statistical, multivariate method used as an extraction tool for mixture analysis, extracting the necessary information (pure component spectra and corresponding concentrations) to predict absolute levels of the analytes within a mixture (Tauler, 1995, Jaumot *et al.*, 2005). The MCR-ALS model was applied to the SERS responses of the thiol (in this case, 1-propanethiol) and internal standard (t-BuOH), along with the HPLC measured

concentrations of *N*-Boc-L-phenylalanine to try and correlate the two techniques, enabling MCR-ALS predictions of the thiol concentration (see Figure S4.5 in SI for a flow diagram summarising the MCR-ALS approach). These comparisons are shown in Figure 4.4.C and Figure 4.4.D. It is evident that the SERS and HPLC results correlate very well and that SERS is capable of monitoring the progression of the reaction, as depicted by the trajectories of 1-propanethiol and *t*-BuOH C profiles in Figure 4.4.C. As one would expect, the internal standard (*t*-BuOH) remains constant as the concentration of 1-propanethiol increases with time (resolved *S* profiles and pure spectra, as given to the MCR-ALS model, are shown in Figure S4.6 in SI). Regression-coefficient values (R^2) tells about the change in variability within a data set, as accounted for by the statistical model (MCR-ALS). R^2 values close to 1 demonstrate excellent fit and correlation between the SERS predicted concentrations and HPLC measured concentrations. Comparing the SERS predicted concentration of 1-propanethiol *versus* the HPLC measured concentration of *N*-Boc-L-phenylalanine gave a R^2 of 0.9798, thus the two techniques are in excellent agreement with one another (Figure 4.4.D). There is bound to be a degree of error incurred when the reaction samples are prepared for HPLC analysis (*i.e.* prolonged sample preparation including concentration to dryness and re-dissolving them in 1/14th ACN volume), so it is inevitable that the predicted SERS concentration will contain inaccuracies due to the measured HPLC concentrations not being perfect primary reference materials (gold standards). Nevertheless, R^2 of 0.9798 demonstrates that these inaccuracies are minimal.

It was important to determine if the choice of thiol side chain would affect the selectivity of the enzyme, α -chymotrypsin. For example, it may have a preference towards the straight chain, 1-propanethiol, over the branched isomer, 2-propanethiol, or *vice versa*. This would lead to false representations of the selectivity expressed by α -chymotrypsin. To investigate this, a 20 mL reaction (enabling both SERS and HPLC measurements) containing both L-thioesters (L-1-A and L-2-A) was monitored and performed in triplicate. The accumulative SERS response for replicate 1 is shown in Figure 4.5.A, peaks at 590 cm^{-1} (2-propanethiol) and 697 cm^{-1} (1-propanethiol) are easily identifiable. By simple univariate analysis, the peak areas of these major discriminant peaks were plotted to compare their relative rates and SERS intensities, shown in Figure 4.5.B. The hydrolysis of L-2-A appears to be at a greater rate than L-1-A, arising from the greater peak area of 2-propanethiol. However, previous attempts at a SERS calibration identified that 2-propanethiol gave a greater intensity SERS response relative to 1-propanethiol, so this is likely to be caused by differences in intensity rather than enzymatic rate. Comparative HPLC analysis established that the rate of hydrolysis was

identical between the L-1-A and L-2-A, confirming that no preference was given towards the thiol sidechain. This enabled us to adjust the peak area of the 2-propanethiol based on its greater intensity SERS response. Over 54 peak area data points (3 replicates) the average intensity difference between 1-propanethiol and 2-propanethiol was calculated to be 1.6035 (standard deviation of 0.0947, <6 % standard error). When 2-propanethiol's response is adjusted for by this average value, one can observe in Figure S4.7 in SI that the relative rates of hydrolysis are extremely similar. Therefore, we see no bias of α -chymotrypsin towards the thiol sidechain. These investigations also demonstrate that this approach would be extremely successful towards a low selectivity enzyme that would display activity towards both L- and D-thioesters, thus producing both thiols *in situ* with rates relative to enzyme enantioselectivity.

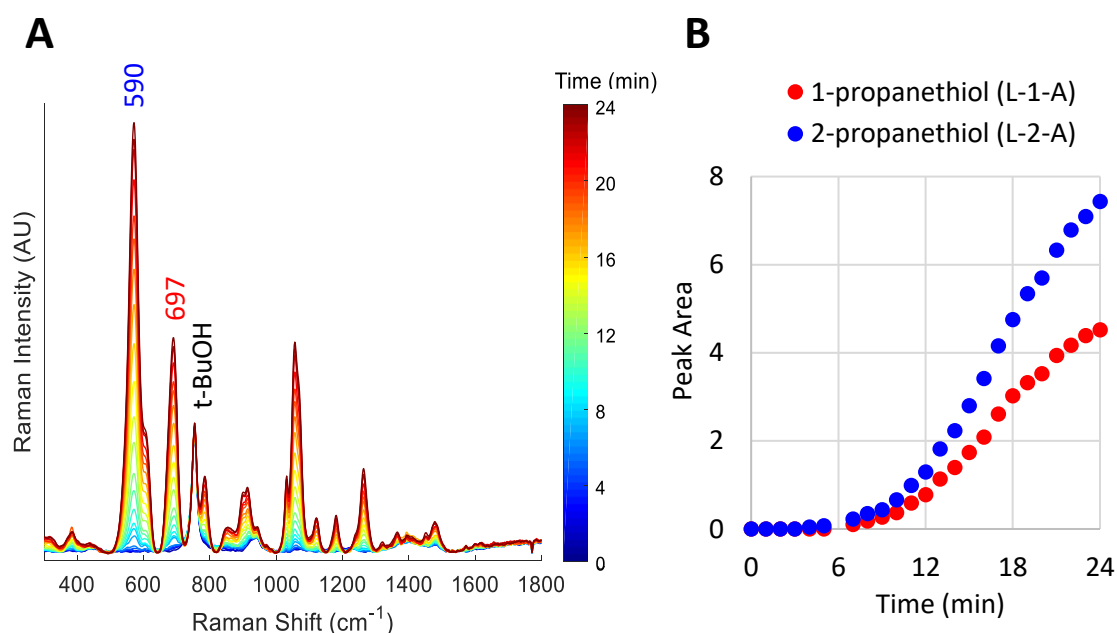


Figure 4.5. (A) Cumulative SERS response of α -chymotrypsin (2 nM) in the presence of two different L-thioesters: L-1-A and L-2-A (1×10^{-5} M each). Spectra have been normalised to t-BuOH at 747 cm^{-1} (see SI, data processing for further details). (B) Plotted peak areas of 1-propanethiol (red, 697 cm^{-1}) and 2-propanethiol (blue, 590 cm^{-1}). NA 1-propanethiol gives a small peak at 624 cm^{-1} which appears as a shoulder on 2-propanethiols peak at 590 cm^{-1} , so this portion of the peak (624 cm^{-1}) was excluded from univariate analysis.

4.5.3 Low Enantioselectivity Enzyme: PLE

Considering these observations, pig liver esterase (PLE) was investigated as an enzyme demonstrating low enantioselectivity towards N-Cbz-protected alanine thioesters: substrate subset B (Table 4.1). Ultimately, our aim is to demonstrate the suitability of this SERS pseudo-

enantiomer approach towards mutagenic strategies, such as directed evolution, where you might want to switch an enzyme's innate enantioselectivity. To compare PLE's enantioselectivity, substrates L-1-B and D-1-B (both including the 1-propanethiol sidechain) were incubated with PLE over a 30 min period. SERS analysis was omitted as the same thiol product is generated, so the HPLC percentages of the reaction components are shown in Figure 4.6. Chiral column separation of Cbz-L/D-alanine (product) was not possible, so the percentage of product is a combination of L- and D-product. Looking at Figure 4.6, it is evident that PLE has a preference towards the L-thioester (L-1-B) as it is consumed more rapidly than the D-thioester (2:1 ratio L:D preference). Nevertheless, we can still be confident that over the reaction monitoring period both thioesters will be hydrolysed, subsequently detecting the two different thiols and enabling rate comparisons.

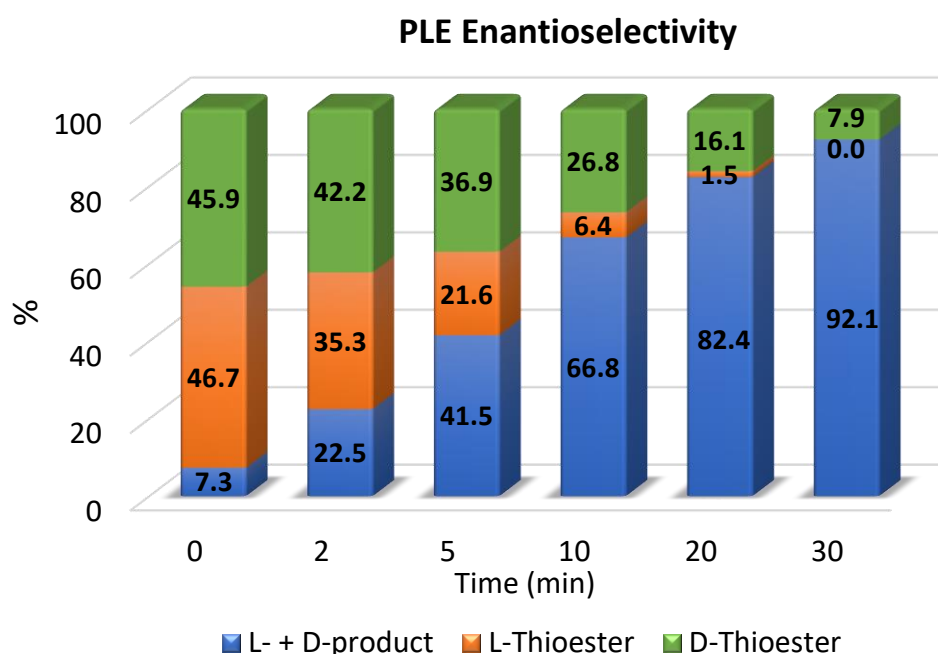


Figure 4.6. Bar chart representing the average percentage conversions ($n = 3$) of L-1-B (orange, L-thioester) and D-1-B (green, D-thioester) with PLE by HPLC. Conditions: L-1-B (5×10^{-4} M), D-1-B (5×10^{-4} M), PLE (5 μ L of 10 mg/mL solution), 10 % t-butanol and potassium phosphate buffer (1×10^{-2} M, pH 6.2), using a 1.5 mL total volume. A 2:1 preference for the L-thioester was observed.

Numerous large scale (20 mL) biotransformations were performed containing both L- and D-thioesters, see Figure S4.8 in SI. The first using L-1-B and D-2-B as substrates, one would expect to observe a 2:1 ratio of 1-propanethiol (L) to 2-propanethiol (D) as identified with the previous HPLC experiments (Figure 4.6). As can be seen in Figure S4.8A in SI, it experiences a peak area ratio of 10:1 1-propanethiol:2-propanethiol (results have been

adjusted for the increased SERS intensity of 2-propanethiol). Similarly, when the alternative thioester substrates are used, L-2-B and D-1-B, the D-thioester is the favoured as the relative peak area of 1-propanethiol (D) is greatest (Figure S4.8B in SI). These results inadvertently identified that PLE is demonstrating a preference towards the straight chain thiol (1-propanethiol) over the branched isomer (2-propanethiol). Consequently, it is misrepresenting the enantioselectivity innate to the enzyme. Further tests with PLE in the presence of both L-thioesters (L-1-B and L-2-B, Figure S4.8C in SI) or D-thioesters (D-1-B and D-2-B, Figure S4.8D in SI) confirmed these observations as the thiols were not produced in a 1:1 ratio, 1-propanethiol was always superior.

Experiments with PLE have further established that we can use this pseudo-enantiomer SERS approach to identify both thiol products when produced simultaneously, thus it would be a suitable screening technique for directed evolutionary projects aimed at modifying enantioselectivity. However, this example has also demonstrated that an enzyme may show selectivity towards the thiol side chain, ultimately causing falsifications of enantioselectivity. Therefore, it is extremely important to perform such tests to ensure that suitable thioester structures have been selected. Finally, experiments were performed to identify if the silver nanoparticles effected the rate of enzyme activity. Control experiments, both with and without nanoparticles, were performed and analysed by HPLC. α -Chymotrypsin was shown to be moderately slower in the presence of the nanoparticles, conversely PLE was unaffected by the presence of the silver colloid (data not shown).

4.6 CONCLUSION

In conclusion, we have investigated SERS as a technique to differentiate between enantioselective behaviour expressed by an enzyme. Previous investigations identified that SERS was not capable of distinguishing between enantiomeric forms, therefore we incorporated a pseudo-enantiomer approach. This method facilitates enantioselective behaviour and subsequent detection of the selectively hydrolysed products as they differ slightly in structure (avoiding modifications around the chiral centre). Within this work, substrates were designed to include a thioester bond and its corresponding hydrolysed thiol side chains were structural isomers of one another (1-propanethiol *and* 2-propanethiol). As a method of overcoming analyte competition at the nanoparticle surface, along with ensuring a strong affinity of the deterministic products to the nanoparticle surface, thiol functionality was incorporated.

Initial investigations quickly identified that the thiols were prone to oxidation, forming disulfide bonds. This led to a reduction in the SERS response over time and a misrepresentation of enzyme activity. Consequently, the nanoparticles were introduced into the enzymatic reactions to capture the hydrolysed thiols immediately, demonstrating a novel, real-time SERS approach.

Firstly, we explored this approach towards an L-selective enzyme: α -chymotrypsin. Thioester substrates modelled around *N*-Boc-protected phenylalanine were successful at demonstrating L-selective thioester hydrolysis. Both small scale (1 mL) and large scale (20 mL) reactions successfully demonstrated L-selective hydrolysis and capture of the corresponding thiol, with the latter enabling comparative HPLC analysis. The application of a MCR-ALS model to the SERS and HPLC data gave R^2 value of 0.9798, confirming that the SERS could successfully monitor the reaction progression and the results were in excellent agreement with HPLC. We further demonstrated that α -chymotrypsin did not show a preference towards the thiol sidechain (*i.e.* straight versus branched), hence this pseudo-enantiomer approach demonstrates the enzymes true selectivity towards the chiral centre.

A second, less enantioselective enzyme was studied to demonstrate this SERS approach within mutagenic screening applications. Thioester substrates were designed around a *N*-Cbz-protected alanine substructure and PLE was selected as the most suitable hydrolysing enzyme, demonstrating activity towards both L- and D-thioesters (2:1 ratio, respectively). SERS investigations soon identified that the enzyme had a preference towards the straight chain (1-propanethiol) linker over the branched isomer (2-propanethiol). This caused misrepresentations of the enzyme's enantioselectivity, so, this would not be suitable to use within enantioselectivity screening applications for this enzyme class. This highlights that careful considerations should be made when designing suitable pseudo-enantiomer substrates.

We hope that this novel, pseudo-enantiomer approach demonstrates that SERS can be implemented as a rapid (10 s) technique to discriminate between enantioselectivity in real-time. This proof-of-principle requires further optimisation and careful considerations should be taken during the design process. Nevertheless, it can correctly identify enantiomerically produced products, which is increasingly difficult to do in such a rapid and information rich manner.

4.7 REFERENCES

- BHARGAVA, S. K., BOOTH, J. M., AGRAWAL, S., COLOE, P. & KAR, G. 2005. Gold Nanoparticle Formation during Bromoaurate Reduction by Amino Acids. *Langmuir*, 21, 5949-5956.
- CHICA, R. A., DOUCET, N. & PELLETIER, J. N. 2005. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Current Opinion in Biotechnology*, 16, 378-384.
- CROSSLEY, R. J. 1995. *Chirality and Biological Activity of Drugs*, Boca Raton, CRC Press.
- DUTHALER, R. O. 1994. Recent developments in the stereoselective synthesis of α -aminoacids. *Tetrahedron*, 50, 1539-1650.
- FERNÁNDEZ-ÁLVARO, E., SNAJDROVA, R., JOCHENS, H., DAVIDS, T., BÖTTCHER, D. & BORNSCHEUER, U. T. 2011. A Combination of In Vivo Selection and Cell Sorting for the Identification of Enantioselective Biocatalysts. *Angewandte Chemie International Edition*, 50, 8584-8587.
- GRÖGER, H. 2010. Enzyme-Catalyzed Asymmetric Synthesis. *Catalytic Asymmetric Synthesis*. John Wiley & Sons, Inc.
- HIBBERT, E. G. & DALBY, P. A. 2005. Directed evolution strategies for improved enzymatic performance. *Microbial Cell Factories*, 4, 29.
- HOLLYWOOD, K. A., SHADI, I. T. & GOODACRE, R. 2010. Monitoring the Succinate Dehydrogenase Activity Isolated from Mitochondria by Surface Enhanced Raman Scattering. *The Journal of Physical Chemistry C*, 114, 7308-7313.
- JAUMOT, J., GARGALLO, R., DE JUAN, A. & TAULER, R. 2005. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemometrics and Intelligent Laboratory Systems*, 76, 101-110.
- KATZUNG, B. G., MASTERS, S. B. & TREVOR, A. J. 2016. *Basic & Clinical Pharmacology*, McGraw-Hill Medical.
- KAZLAUSKAS, R. J. & BORNSCHEUER, U. T. 2009. Finding better protein engineering strategies. *Nature Chemical Biology*, 5, 526-529.
- MANCIN, F., PRINS, L., PENGO, P., PASQUATO, L., TECILLA, P. & SCRIMIN, P. 2016. Hydrolytic Metallo-Nanozymes: From Micelles and Vesicles to Gold Nanoparticles. *Molecules*, 21, 1014.
- MANEA, F., HOUILLON, F. B., PASQUATO, L. & SCRIMIN, P. 2004. Nanozymes: Gold-Nanoparticle-Based Transphosphorylation Catalysts. *Angewandte Chemie International Edition*, 43, 6165-6169.
- MAY, O., NGUYEN, P. T. & ARNOLD, F. H. 2000. Inverting enantioselectivity by directed evolution of hydantoinase for improved production of L-methionine. *Nature Biotechnology*, 18, 317.
- MOHAN, S. J., MOHAN, E. C. & YAMSANI, M. R. 2009. Chirality and its importance in pharmaceutical field-an overview. *International Journal of Pharmaceutical Sciences and Nanotechnology*, 1, 309-316.
- NGUYEN, L. A., HE, H. & PHAM-HUY, C. 2006. Chiral Drugs: An Overview. *International Journal of Biomedical Science : IJBS*, 2, 85-100.
- REETZ, M. T. 2004. Changing the Enantioselectivity of Enzymes by Directed Evolution. *Methods in Enzymology*, 388, 238-256.
- RENTSCH, K. M. 2002. The importance of stereoselective determination of drugs in the clinical laboratory. *Journal of Biochemical and Biophysical Methods*, 54, 1-9.
- SLOAN-DENNISON, S., LAING, S., SHAND, N. C., GRAHAM, D. & FAULDS, K. 2017. A novel nanozyme assay utilising the catalytic activity of silver nanoparticles and SERRS. *Analyst*, 142, 2484-2490.

- STEWART, A., ZHENG, S., MCCOURT, M. R. & BELL, S. E. J. 2012. Controlling Assembly of Mixed Thiol Monolayers on Silver Nanoparticles to Tune Their Surface Properties. *ACS Nano*, 6, 3718-3726.
- TAULER, R. 1995. Multivariate curve resolution applied to second order data. *Chemometrics and Intelligent Laboratory Systems*, 30, 133-146.
- TOYO'OKA, T. 2002. Resolution of chiral drugs by liquid chromatography based upon diastereomer formation with chiral derivatization reagents. *Journal of Biochemical and Biophysical Methods*, 54, 25-56.
- WALTHER, W. & NETSCHER, T. 1996. Design and development of chiral reagents for the chromatographic ee determination of chiral alcohols. *Chirality*, 8, 397-401.
- WESTLEY, C., XU, Y., CARNELL, A. J., TURNER, N. J. & GOODACRE, R. 2016. Label-Free Surface Enhanced Raman Scattering Approach for High-Throughput Screening of Biocatalysts. *Analytical Chemistry*, 88, 5898-5903.

4.8 SUPPORTING INFORMATION

SUPPLEMENTARY METHODS

Reagents and Materials.

All chemical reagents were of analytical grade and used with no additional purification. Chemicals and solvents were purchased from Acros Organic (New Jersey, USA), Alfa Aesar (Heysham, UK), Fischer Scientific (Loughborough, UK), Fluorochem (Hadfield, UK) or Sigma Aldrich Ltd (Dorset, UK). Commercial enzymes, α -chymotrypsin from bovine pancreas (EC 3.4.21.1, lyophilised) and pig liver esterase (EC 3.1.1.1, ammonium sulfate solution) were purchased from Sigma Aldrich Ltd. (Dorset, UK).

Synthesis and Characterisation of Silver Nanoparticles.

All glassware used for the synthesis of metal nanoparticles was cleaned using aqua regia (HNO_3 : HCl (1:3) v/v) to ensure the removal of any residual metals. Glassware was then thoroughly rinsed with deionised water. Hydroxylamine-reduced silver colloid (**HRSC**) was prepared as reported by Leopold and Lendl (Leopold and Lendl, 2003). Hydroxylamine hydrochloride (1.88×10^{-3} M, 180 mL) and NaOH (3.33×10^{-3} M) were prepared together. To this rapidly stirring solution, AgNO_3 (1×10^{-2} M, 20 mL) was added drop-wise and stirred for a further 15 min, forming an orange-yellow colloid.

To characterise nanoparticle size distribution of HRSC, UV-Vis spectrometry was employed to enable comparisons between different batches. HRSC were shown to have an absorption maximum (λ_{max}) of the plasmon resonance band of 411 nm, in agreement with the literature (Leopold and Lendl, 2003). A sharp UV-Vis peak also identified that there was a narrow size distribution amongst the colloidal nanoparticles. Scanning electron microscopy images identified that the size and shape distribution of the nanoparticles was good. Images are shown Figure S4.9. The average nanoparticle size was $\sim 30 \pm 5$ nm, once again agreeing with the literature.

Small scale (1 mL) SERS biotransformation.

Real-time SERS samples contained 625 μL hydroxylamine-reduced silver colloid (HRSC), 100 μL t-butanol (internal standard), potassium phosphate buffer (5×10^{-3} M final concentration, pH 6.2) and L- and D-thioesters at a total concentration of 2×10^{-5} M (1×10^{-5} M each). The reaction mixture was prepared within an 8 mm glass vial which was inserted into the cell holder for continuous SERS analysis. The introduction of α -chymotrypsin (2×10^{-9} M) initiates the reaction, vigorous pipetting ensured a good distribution. SERS spectra were constantly acquired, using a 10 s acquisition time.

Large scale (20 mL) SERS biotransformation with comparative HPLC.

Real-time SERS samples contained 12.5 mL HRSC, 2 mL t-butanol (internal standard), potassium phosphate buffer (5×10^{-3} M final concentration, pH 6.2) and L- and D-thioesters at a total concentration of 20 μ M. The introduction of α -chymotrypsin (2×10^{-9} M in Tris HCl buffer 0.1 M, pH 7.2) or PLE (50 μ L of 10 mg/mL ammonium sulfate solution, 3.2 M, pH 8) initiates the reaction. Control experiments ensured that the protein buffers (Tris HCl/ammonium sulfate) did not affect the SERS response, induce non-enzymatic hydrolysis or alter the pH of the SERS sample.

For real-time SERS analysis, 0.4 mL was removed from the reaction mixture and analysed instantaneously using a 10 s acquisition time (mainly every minute over the monitoring period). For comparative, off-line HPLC analysis (mainly every other minute), 1 mL of the reaction mixture was quenched using 1 % trifluoroacetic acid (TFA). This was followed by centrifugation (21,000 xg for 6 min) and overnight (16 h) lyophilisation of the supernatant affording a solid. This solid was re-dissolved in acetonitrile (70 μ L) and centrifuged (1,000 xg for 6 min) prior to HPLC analysis (24 min acquisition time, chiral column). See Figure S4.5 for an image depicting the analysis process.

INSTRUMENTATION AND DATA PROCESSING

SERS Analysis.

SERS spectra were recorded using a DeltaNu Advantage 200A portable spectrometer (DeltaNu, Laramie, WY, USA) equipped with a HeNe 785 nm laser which provided ~ 60 mW power on the sample. Toluene was pipetted into 8 mm glass vials in order to calibrate the systems, allowing the ideal distance between the laser and the point of sampling to be established. 785 nm SERS spectra were acquired for 10 s over a range of 200 – 2000 cm^{-1} . Samples for SERS analysis were placed in an 8 mm glass vial before subjected to laser irradiation once placed in the cell holder.

All SERS data were exported from the respective instrument operating software and analysed using Matlab R2015a (The Mathworks, Natick, MA, USA) using scripts written in-house and available via GitHub (<http://www.biospec.net/resources/>). SERS spectra were either raw, baseline corrected or normalised to the internal standard (as stated in the text).

HPLC Analysis.

HPLC separation was conducted using an Agilent Zorbax Eclipse Plus HPLC system set-up for reverse phase separation consisting of a diode array detector. For all biotransformations, a chiral Phenomenex Lux 5 μm Cellulose-3 (250 x 4.6 mm) column was used and the solvents were $\text{H}_2\text{O} + 0.1\% \text{ FA}$ and $\text{ACN} + 0.1\% \text{ FA}$.

Phenylalanine-containing substrates (subset A) HPLC method: substrates were eluted using a gradual gradient of 60:40 $\text{H}_2\text{O}:\text{ACN} + 0.1\% \text{ FA}$ changing to 42:58 $\text{H}_2\text{O}:\text{ACN} + 0.1\% \text{ FA}$ over 16 min. A total run time of 24 min was used.

Alanine-containing substrates (subset B) HPLC method: substrates were eluted using an isocratic gradient of 50:50 $\text{H}_2\text{O}:\text{ACN} + 0.1\% \text{ FA}$ over 16 min. A total run time of 24 min was used.

The peaks of the target analytes were integrated with the results of the HPLC data and served as an external validation data set to independently verify the accuracies of the MCR-ALS models in prediction.

UV-Vis Spectrophotometry.

HRSCs were characterised using UV-Vis spectroscopy to determine the surface plasmon resonance band λ_{max} of the nanoparticles. HRSC was diluted 1:3 v/v with deionised water, to ensure that the UV-Vis absorbance did not exceed 2. For analysis 1 mL of the diluted nanoparticle solution was pipetted into a polystyrene cuvette and placed into the sample holder of a Thermo Biomate 5 (Thermo Fisher Scientific Inc., Massachusetts, USA). Absorbance spectra were collected over a range of 200-800 nm.

SEM.

SEM images were collected using a Field Emission Scanning Electron Microscope with an ultimate beam size of 1 nm at 100 eV electron energy (Carl-Zeiss-Straße 56, 73447 Oberkochen, Germany). A silicon surface was used to image the nanoparticles as silicon helps disperse negative charges and generates a greater contrast in comparison to glass, as a result higher resolution images can be obtained. To image HRSC nanoparticles, 10 μL was spotted onto a silicon slide and left to dry overnight (Figure S4.10).

NMR Spectroscopy.

Purified samples were concentrated to dryness using the GeneVac (EZ-2 Series) before the addition of deuterated solvent for NMR analysis. NMR analysis was performed on the Bruker

DXP 500 MHz or 800 MHz spectrometer. Spectra processing was performed using MestReNova software.

HRMS.

Agilent 6510 QTOF coupled to an Agilent 1200 series LC. 5 μ l of sample is flow injected at 0.3 ml/min 50 % ACN 0.1 % FA, using ESI in either positive or negative mode (as required, stated in the text).

SERS Data Analysis.

Multivariate curve resolution - alternating least squares (MCR-ALS) was employed to the SERS data. The only data pre-processing involved in MCR-ALS was baseline correction.

The spectra of all the monitored time points of a single reaction were then augmented to form a $t \times n$ data matrix X where t is the number of time points monitored and n is the number of wavenumbers recorded. Multivariate curve resolution using alternating least squares (MCR-ALS) was performed to deconvolve X into a product of two sub matrices C and S where C contains the “profiles” of the change in the concentrations the reactants during the reaction while S is the matrix storing the resolved spectra the reactants. Non-negativity constraint was applied to both concentration profile C and spectral profile S and each resolved pure spectrum had a unit norm (*i.e.* the sum of squares of each spectrum equals 1). For the time points when HPLC measures were also taken, a linear regression model was built between the concentration profile of each reactant in C and the corresponding concentration measured by HPLC. The regression model was then applied to the whole concentration profile to get SERS calibrated concentrations of the reactant over the whole monitored period of the reaction (see Figure S4.5 for flow diagram summarising MCR-ALS process).

Thiol Optimisation and Monolayer Coverage

As part of the optimisation process, we needed to find suitable thiols that would enable characteristic peak identification from one another (thus permitting identification of L-/D-thioester hydrolysis). Several low molecular weight thiols were investigated, those that were characteristically different from one another, but also similar in structure were the combinations of 1-propanethiol *versus* 2-propanethiol (structural isomers) and 1-pentanethiol *versus* 1-propanethiol (both straight chain thiols), see Table S4.1 for structures. 1:1 ratios of the comparative thiols were prepared at various concentrations (see Figure S4.1 for the total concentrations as shown on the graph axis). The ratio of the peak

areas (thiol 1 peak area/thiol 2 peak area) are plotted in Figure S4.1. We can use this information to calculate the point at which monolayer coverage is reached, by identifying the concentration when the ratio increases, indicating competition between the two thiols. 1-propanethiol and its structural isomer 2-propanethiol enable much lower limits of detection when compared to the combination of 1-pentanethiol and 1-propanethiol. The asterisk identifies the concentration when a deviation in the relative ratio is experienced (an increase), indicating competition because monolayer coverage has been reached. These are 5×10^{-5} M for 1-propanethiol + 2-propanethiol and 2×10^{-4} M for 1-pentanethiol + 1-propanethiol. Straight chains can pack more efficiently on the surface of the nanoparticle as opposed to branched chains, therefore they can permit a higher number of molecules around the nanoparticle surface, hence a higher concentration before monolayer coverage. We can conclude that if we do not exceed the point of monolayer saturation, then the thiols will have equal affinity for the nanoparticle surface, avoiding competition (Stewart *et al.*, 2012). Thus, the corresponding thiol SERS responses will be a true representation of enzymatic hydrolysis if both L- and D-thioesters are hydrolysed. The upper limits of the corresponding thiol mixtures are shown in Table S4.1. Due to the lower limits of detection and structural similarity (isomers), we chose 1-propanethiol and 2-propanethiol as the thiols to incorporate within the thioester structures, and subsequently detect post enzyme activity.

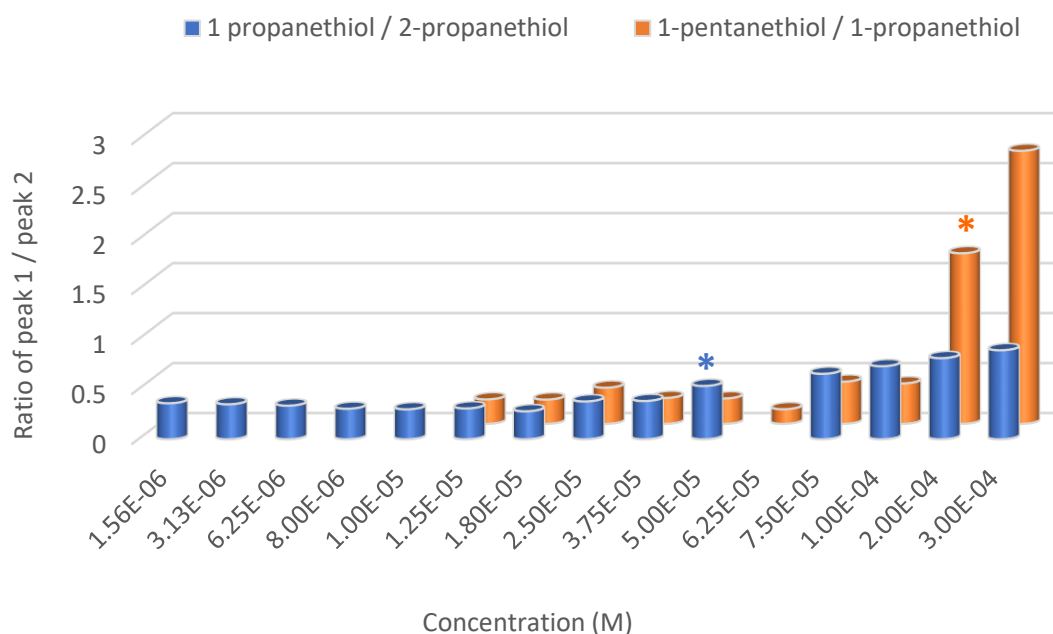

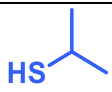




Figure S4.1. Graph of the ratio of peak areas of characteristic peaks for either 1-propanethiol versus 2-propanethiol (blue) or 1-pentanethiol versus 1-propanethiol (orange). Asterisk (*) identifies the concentration at which monolayer coverage is exceeded and we no longer observe a 1:1 ratio of thiols (competition between thiols due to monolayer coverage).

Table S4.1. Table representing the structures of the thiols described in Figure S4.1 and the concentration of monolayer coverage.

THIOL 1	THIOL 2	POINT OF MONOLAYER COVERAGE (M)
 1-propanethiol	 2-propanethiol	$>3.75 \times 10^{-5} (\pm 3.75 \times 10^{-6})$
 1-pentanethiol	 1-propanethiol	$>1 \times 10^{-4} \text{ M } (\pm 1.00 \times 10^{-5})$

Exceeding the concentrations stated in the table results in competition between the thiols, indicating that monolayer coverage has been exceeded. $\pm 10\%$ has been included to account for inherent oxidation that is unavoidable.

Table S4.2. Tentative SERS band assignments of 1-propanethiolate and 2-propanethiolate (S^-) interacting with HRSC (785 nm irradiation) (Joo *et al.*, 1986, Pang *et al.*, 1998, Strekal' *et al.*, 2000).

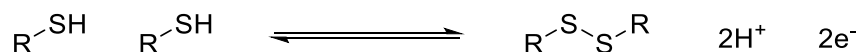
1-propanethiolate	2-propanethiolate	Assignment
624 (w)		C-S
697 (vs)	590 (vs)	C-S stretch
778 (vw)		CH_2
893 (m)	880 (w)	CH_2
1024 (s)	1036 (m)	CC
1085 (w)	1145 (w)	CC

vs – very strong; s – strong; m – medium; w – weak; vw – very weak.

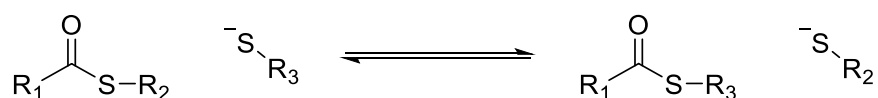
Oxidation: Disulfide Bond Formation

When performing at-line SERS analysis of quenched reaction samples (1 % TFA quench followed by centrifugation and removal of precipitated proteins), we observed fluctuations in the SERS response over time. To determine if this was due to undesirable side reactions, such as disulfide bond formation, we performed various experiments (see Scheme S4.1 for common thiol interactions). Firstly, to conclude if the low boiling point thiols were simply evaporating over time, two identical samples of 1-propanethiol (2×10^{-5} M) in conditions mimicking the biotransformation were prepared, omitting enzyme. Both samples were shaken over a 60 min period, one was left sealed throughout, the other was opened at various time points and SERS analysis performed. We observed that the SERS intensity decreased over time at every sampling time point (intermittently opened vessel). However, the sample which remained sealed until the final timepoint ($t = 60$ min) had a significant reduction in the relative SERS response of 1-propanethiol, which was comparative to the response at $t = 60$ min of the intermittently opened sample. Confirming that the reduction in SERS response was unlikely to be a consequence of thiol volatility.

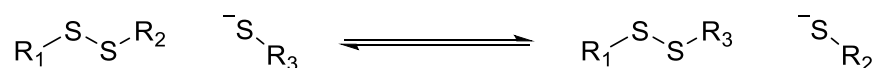
1) Thiol Oxidation



2) Transthioesterification



3) Thiol-disulphide exchange



Scheme S4.1. Common interactions of free thiols (-SH) or thiolates (-S⁻). (1) Thiol oxidation resulting in disulfide bond (-S-S-) formation. (2) Thiolate exchanges its organic R-group (R₃) with the thioester R-group (R₂). (3) Similarly, a thiolate exchanges its R-group with an R-group associated to a disulfide bond.

Secondly, we performed a similar study whereby we had the same total concentration of 1-propanethiol (2×10^{-5} M), however the pH was varied from acidic (pH 2) to basic (pH 10.8).

The literature suggests that the rate of disulfide bond formation will be slowest at acidic pH's as the sulfur is protonated (*i.e.* SH), whereas at basic pH they are deprotonated (*i.e.* S⁻) and disulfide bond formation is more likely (Singh and Whitesides, 2010). The pKa of 1-propanethiol is 10.2 (ChemAxon), therefore at pH 10.8 it should predominantly be deprotonated. Numerous pH values were investigated using differing buffers, as seen in Figure S4.2. The peak area of 1-propanethiol's major peak (697 cm⁻¹) has been plotted *versus* time. One can easily observe the significant reduction in peak area with time at all pH conditions. For example, >50 % reduction in the SERS peak area is experienced after a 40 min period by all of the samples. We hypothesise that the oxidation of the thiol to its corresponding disulfide is happening regardless of pH and is caused by dissolved oxygen. We can also assume that the silver nanoparticles are unable to cleave the disulfide bond once formed, otherwise the peak area would remain constant.

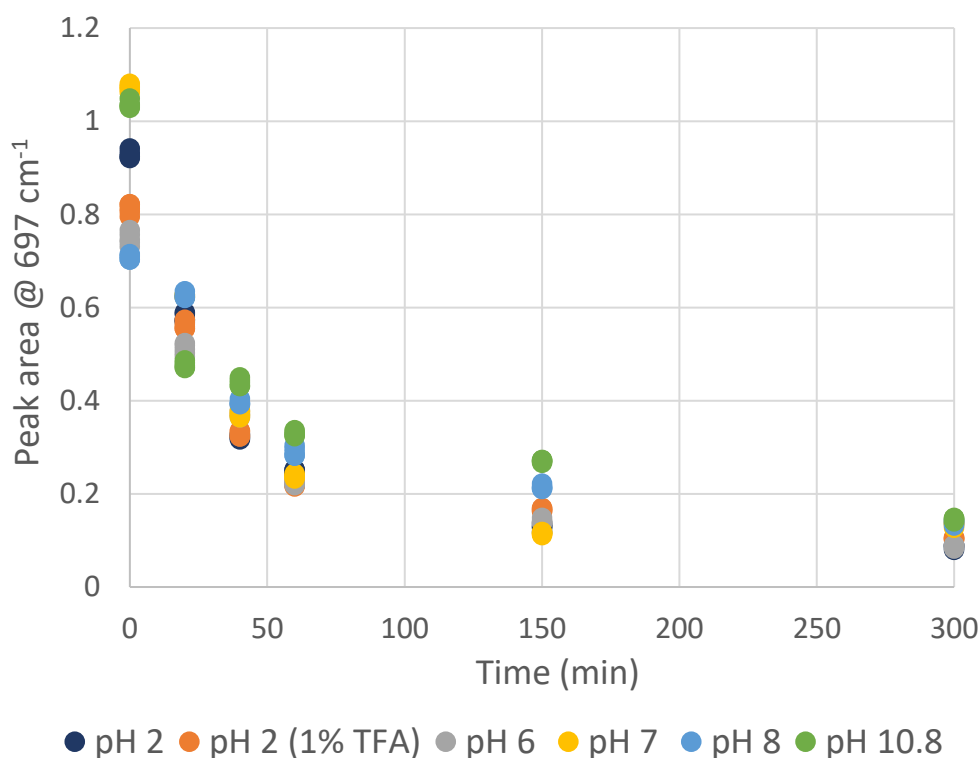


Figure S4.2. Figure represents 1-propanethiol at 2×10^{-5} M total concentration in a sealed vessel experiencing different pH conditions using various buffers. At various time points, samples were taken from the vessel for SERS analysis. As can be seen, the peak area is decreasing over time, independently of pH. After 40 min, a 50 % reduction in peak area is observed.

Attempts to deoxygenate the biotransformation samples were performed using both nitrogen degassing and freeze-thaw methods. Freeze-thaw proved most successful at reducing the rate of disulfide formation, however the extent of oxidation still had a significant

effect on the results, meaning that quantitative measurements were not feasible. Therefore, a new approach was developed to capture the thiol products *in situ* and instantaneously, avoiding disulfide formation completely.

Further experiments were performed to determine the stability of the thioester substrates, observing if they were prone to non-enzymatic hydrolysis or transthoesterification (Scheme S4.1). For the former hydrolysis investigations, all thioesters were incubated at 37 °C overnight in conditions mimicking the SERS biotransformation. After 24 h, no detectable hydrolysis was observed using HPLC analysis (*i.e.* no *N*-Boc-phenylalanine or *N*-Cbz-alanine product). Transthoesterification investigations were performed in a similar manner by incubating a thioester (*i.e.* L-1-A) with the opposite, free thiol (*i.e.* 2-propanethiol). If transthoesterification had occurred, we would expect a shift in the retention time resulting the formation of a new thioester (see Table S4.3 for thioester HPLC retention times). We did not observe any changes in retention times over the 24 h period (at multiple timepoints), confirming that transthoesterification did not occur.

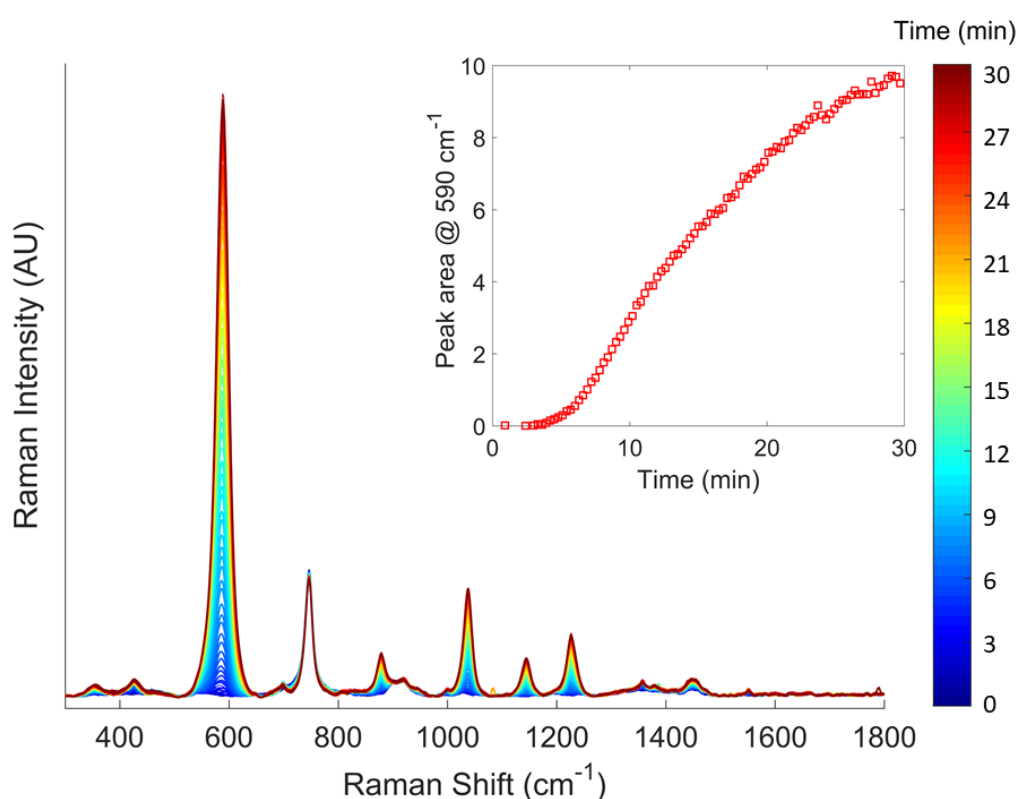


Figure S4.3. Cumulative SERS spectra taken over a 30 min time course. The colour bar highlights the increasing intensity with time ($t = 0$ in blue, $t = 30$ in red). The SERS data has been normalised to the internal standard, *t*-BuOH at 747 cm⁻¹. 1 mL reaction including substrates L-2-A and D-1-A, at a concentration of 1×10^{-5} M per substrate. The introduction of α -chymotrypsin gives rise to L-enantioselectivity, demonstrated by the production of 2-propanethiol (L-2-A hydrolysis). Inset shows univariate analysis of the SERS response by plotting the peak area of 590 cm⁻¹ against time.

20 mL Set-up of SERS Biotransformations with Comparative HPLC

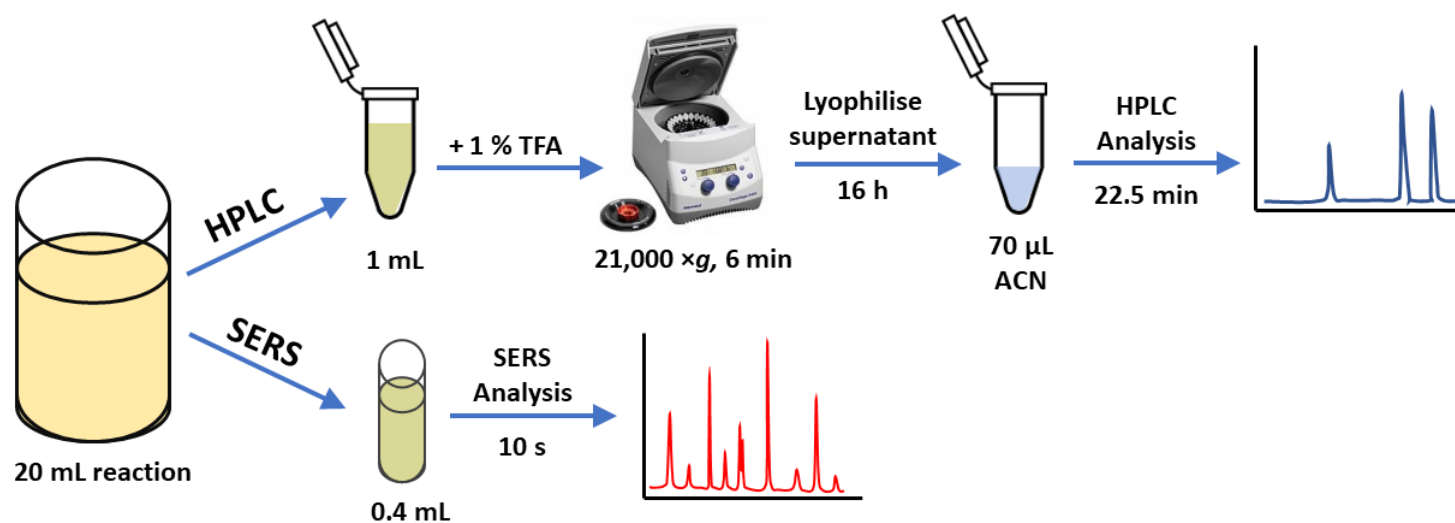


Figure S4.4. Image illustrating the process of monitoring the biotransformation with both SERS and HPLC analysis. The 20 mL reaction mixture contains: 62.5 % HRSC, 10 % t-BuOH, potassium phosphate buffer (5×10^{-3} M, pH 6.2 final), L-thioester [1×10^{-5} M], D-thioester [1×10^{-5} M] and α -chymotrypsin [2×10^{-9} M] added last. For SERS analysis, 0.4 mL of the reaction sample (un-modified) is added to a 1 mL glass vial for immediate analysis (10 s acquisition, 785 nm). For HPLC analysis, 1 mL of the reaction sample is instantaneously quenched using 1 % TFA. Subsequent protein removal (centrifugation at 21,000 xg for 6 min) and overnight lyophilisation of the supernatant to dryness, before the solid is re-dissolved in 70 µL of acetonitrile and centrifuged (21,000 xg for 6 min) prior to HPLC analysis (24 min acquisition time, chiral column).

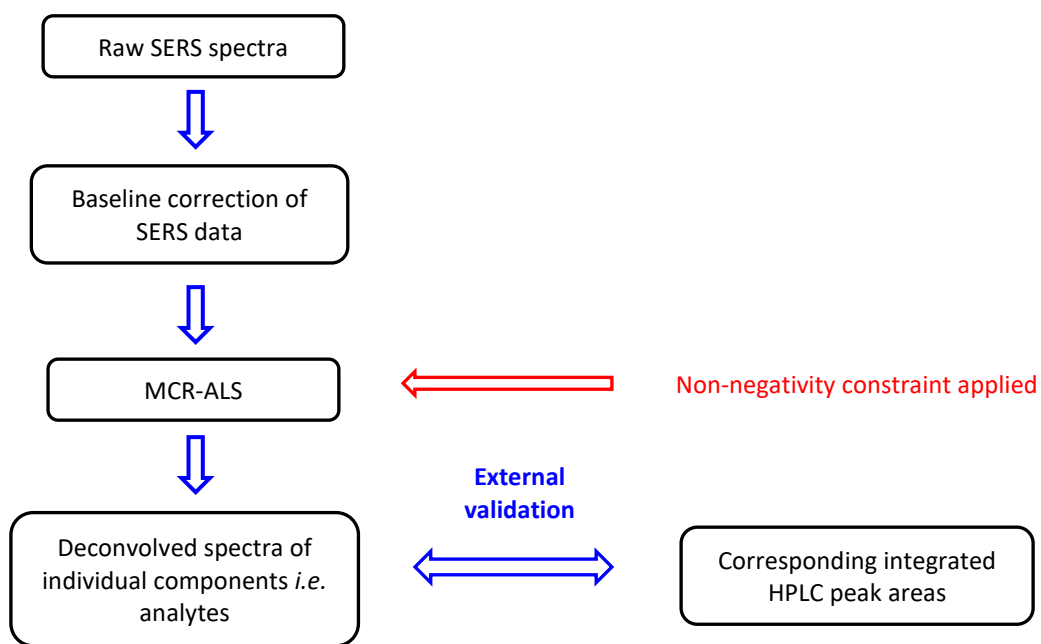


Figure S4.5. Flow diagram summarising the data pre-processing and MCR-ALS process used for predicting concentrations of product (*i.e.* 1-propanethiol) from the reaction.

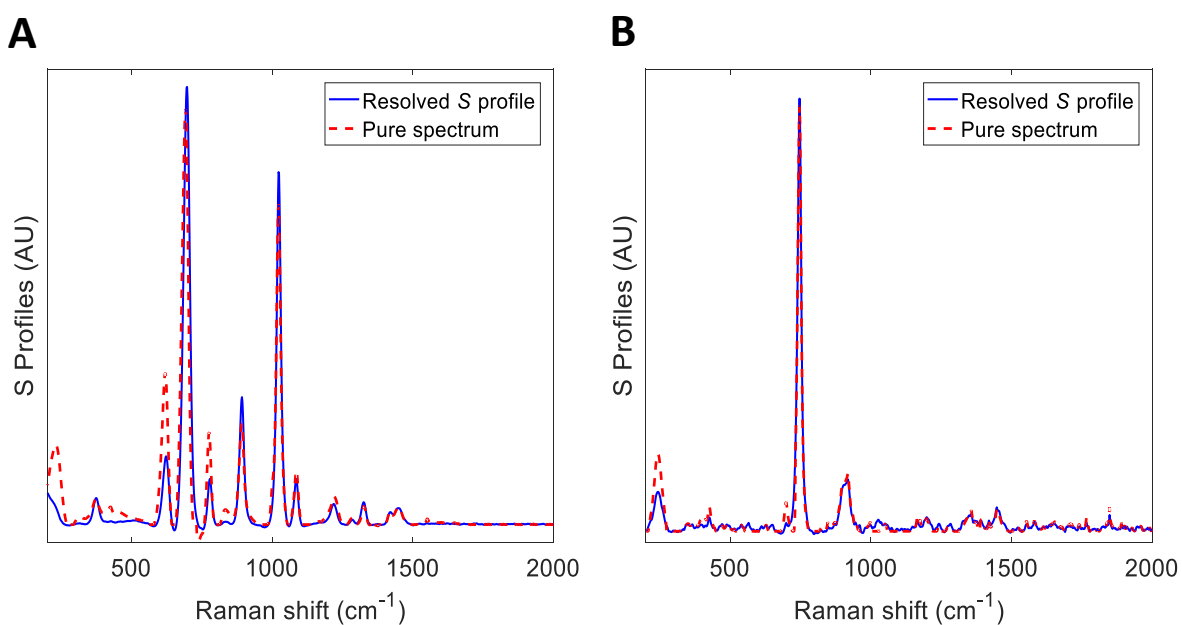


Figure S4.6. MCR-ALS model was applied to the SERS data where it successfully deconvolved spectra into its pure components (A) 1-propanethiol (product) and (B) t-butanol (internal standard).

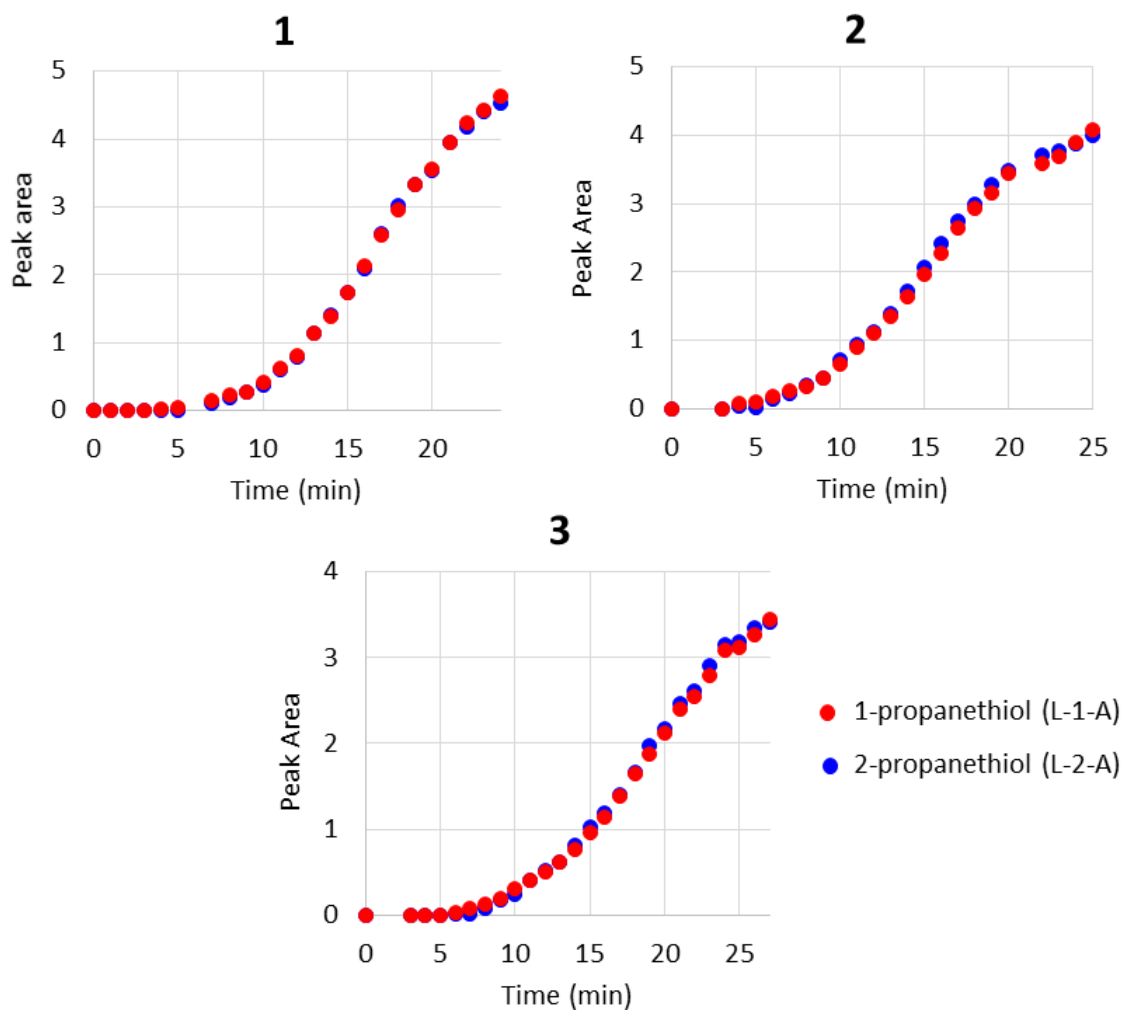


Figure S4.7. Peak areas of 1-propanethiol (red, 697 cm^{-1}) and 2-propanethiol (blue 590 cm^{-1}) over the reaction monitoring period for 3 replicates (**1 – 3**). Conditions: L-1-A and L-2-A, $1 \times 10^{-5}\text{ M}$ each, $2 \times 10^{-9}\text{ M}$ α -chymotrypsin addition. 2-propanethiol response has been adjusted by 1.6035 (average of 54 data points) to account for the increased SERS intensity relative to 1-propanethiol.

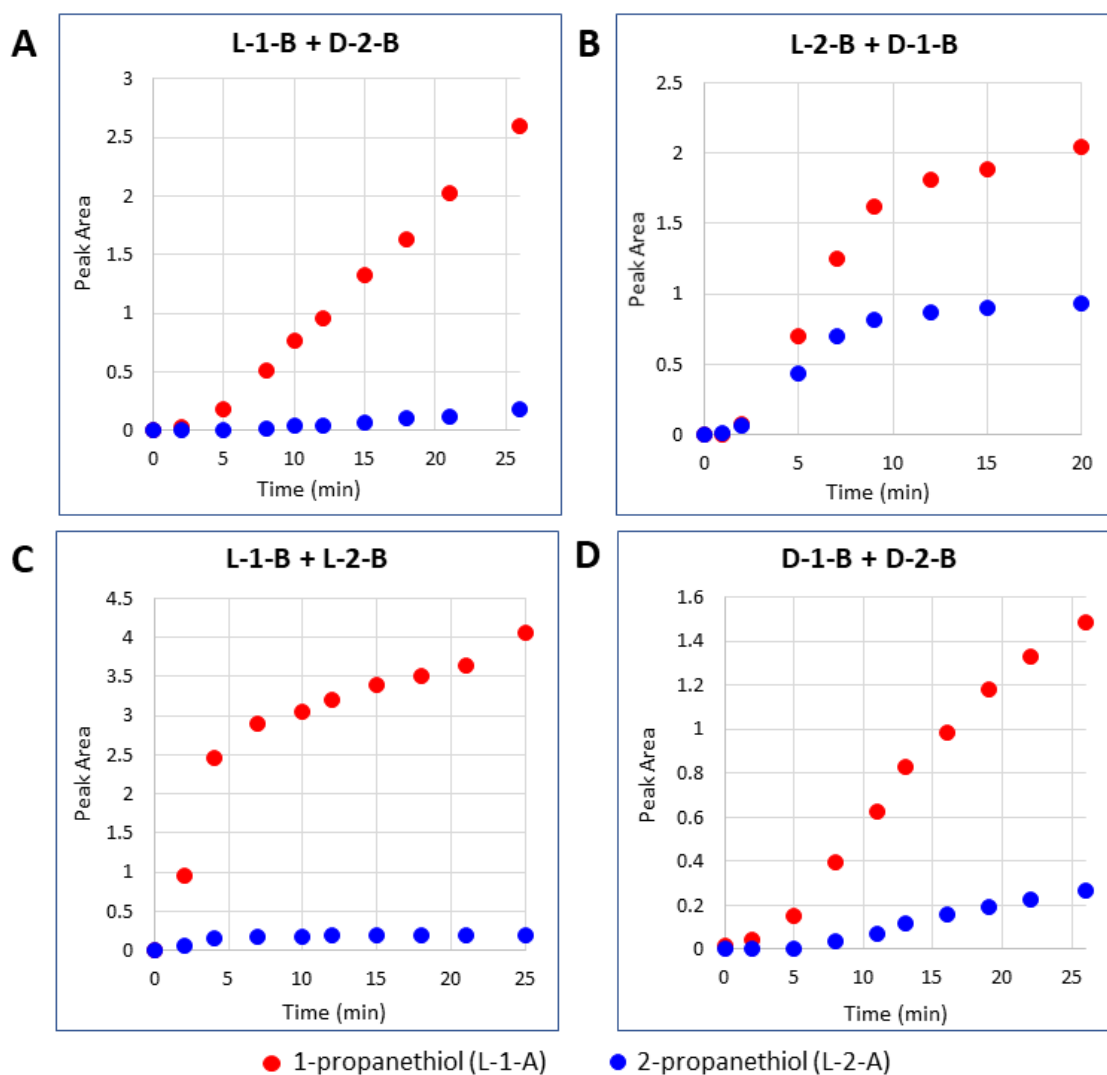


Figure S4.8. Peak areas of 1-propanethiol (red, 697 cm^{-1}) and 2-propanethiol (blue 590 cm^{-1}) over the reaction monitoring period. (A) (L-1-B + D-2-B) and (B) (L-2-B and D-1-B) are combinations of L- and D-thioesters in the presence of PLE (50 μL of 10 mg/mL solution), whereas (C) (L-1-B + L-2-B) and (D) (D-1-B + D-2-B) are reactions with both L-/or D-thioesters. 2-propanethiol peaks areas have been adjusted to account for the greater intensity SERS response. In every reaction, 1-propanethiol always generates the greatest response, even when it is associated to the D-thioester SM. This identifies that PLE is demonstrating a preference towards the straight sidechain (1-propanethiol) over the branched isomer (2-propanethiol).

SERS Optimisation

Prior to experiments, many different parameters were varied to attain optimal SERS profiles of each analyte. A large focus was placed on the choice of colloidal nanoparticles, along with investigations into laser excitation wavelength, analyte concentration, pH, aggregating agent and acquisition time. Figure S4.9 demonstrates the erratic concentration profile of 1-propanethiol in the absence of buffer: once potassium phosphate buffer was introduced (pH 6.2, $1 \times 10^{-1}\text{ M}$) a trend following a

sigmoidal curve response could be observed. These outcomes reconfirm the importance of parameter optimisation in SERS.

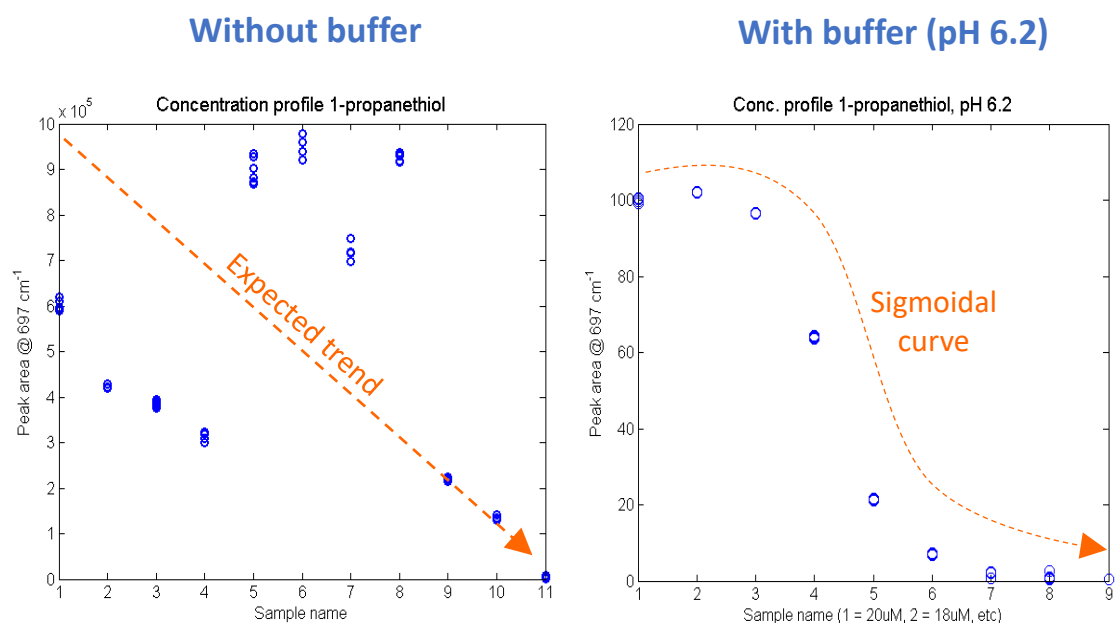


Figure S4.9. Figures identify the importance of SERS optimisation. In the absence of buffer, no trend is observed with decreasing concentrations of 2-propanethiol (left). However, exploration of pH identified that pH 6.2 gave optimal responses, observing a sigmoidal-type response on decreasing concentration of 2-propanethiol (right). Sample 1 = 20 μM , consecutive samples each decrease by 2×10^{-6} M: 2 μM (i.e. sample 5 = 12 μM , sample 9 = 4 μM). Conditions: HRSC, 1 min. aggregation time, 6 s acquisition, 785 nm, potassium phosphate buffer (pH 6.2, 1×10^{-1} M) or no buffer.

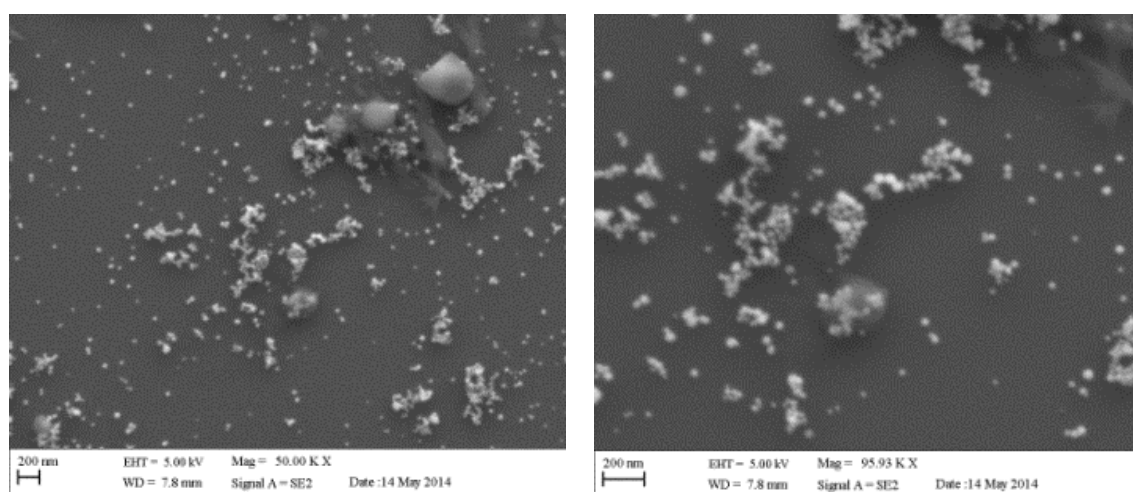
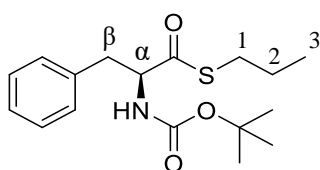


Figure S4.10. SEM images of HRSC deposited on silicon. Left, depicts images of 50x magnification and a scale bar of 200 nm. Right, 96x magnification and 200 nm scale bar.

THIOESTER SYNTHESIS AND CHARACTERISATION

General Method for Substrate Subset A Synthesis:

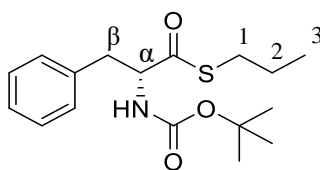
N-Boc-phenylalanine (L- or D-, 300 mg, 1.13 mmol) and 4-dimethylaminopyridine (12.3 mg, 0.113 mmol) were dissolved in dry DCM (3 mL), followed by the thiol (1-propanethiol or 2-propanethiol, 120.7 μ L, 1.47 mmol). The solution was cooled to 0 °C in an ice bath with stirring, then *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (260.7 mg, 1.36 mmol) was added. The solution was stirred at room temperature overnight before extraction with citric acid (0.5 M, 3 x 3 mL) and brine (2 x 3 mL). The organic was dried using MgSO₄ and concentrated *in vacuo*, giving a white solid.



L-1-A

L-1-A (*N*-Boc-L-phenylalanine + 1-propanethiol) yield = 272.93 mg (74.6 %)

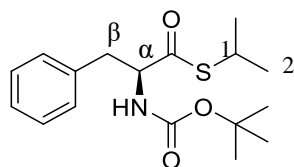
¹H NMR (500 MHz, Methanol-*d*₄) 7.38 – 7.05 (5 H, m, Ar-H), 4.34 (1 H, dd, *J* 31.4, 10.0 Hz, α H), 3.18 (1 H, d, *J* 14.0 Hz, β H''), 2.93 – 2.68 (3 H, m, β H' + C1H), 1.69 – 1.49 (2 H, m, C2H), 1.36 (9 H, s, Boc-H), 0.97 (3 H, t, *J* 6.2 Hz, C3H). ¹³C NMR (126 MHz, Methanol-*d*₄) 201.8 (C=O), 156.4 (C=O), 137.2 (Ar-C), 128.8 (Ar-C), 128.0 (Ar-C), 126.3 (Ar-C), 79.4 (C), 62.3 (α C), 37.3 (β C), 30.1 (C1), 27.3 (Boc-CH₃), 22.5 (C2), 12.2 (C3). HRMS-ESI [M+Na]⁺ expected *m/z* 346.1447 and 347.1481 observed *m/z* 346.1504 and 347.1499 (\pm 16.47 and 5.18 ppm).



D-1-A

D-1-A (*N*-Boc-D-phenylalanine + 1-propanethiol) yield = 275.63 mg (75.4 %)

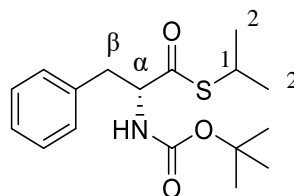
¹H NMR (500 MHz, Methanol-*d*₄) 7.34 – 7.13 (5 H, m, Ar-H), 4.34 (1 H, dd, *J* 31.4, 10.0 Hz, α H), 3.17 (1 H, d, *J* 14.2 Hz, β H''), 2.90 – 2.72 (3 H, m, β H' + C1H), 1.65 – 1.52 (2 H, m, C2H), 1.37 (9 H, s, Boc-H), 0.97 (3 H, t, *J* 6.8 Hz, C3H). ¹³C NMR (126 MHz, Methanol-*d*₄) 201.8 (C=O), 156.4 (C=O), 137.2 (Ar-C), 128.8 (Ar-C), 128.0 (Ar-C), 126.3 (Ar-C), 79.4 (C), 62.3 (α C), 37.3 (β C), 30.1 (C1), 27.3 (Boc-CH₃), 22.6 (C2), 12.2 (C3). HRMS-ESI [M+Na]⁺ expected *m/z* 346.1447 and 347.1481 observed *m/z* 346.1494 and 347.1489 (\pm 13.58 and 2.30 ppm).



L-2-A

L-2-A (*N*-Boc-L-phenylalanine + 2-propanethiol) yield = 244.96 mg (67.0 %)

¹H NMR (500 MHz, Methanol-*d*₄) 7.34 – 7.13 (5 H, m, Ar-H), 4.34 (1 H, dd, *J* 10.4, 4.6 Hz, αH), 3.55 (1 H, hept, *J* 6.8 Hz, C1H), 3.17 (1 H, dd, *J* 14.0, 4.6 Hz, βH''), 2.78 (1 H, dd, *J* 14.0, 10.4 Hz, βH'), 1.53 – 1.18 (15 H, m, C2H + Boc-H). **¹³C NMR** (126 MHz, Methanol-*d*₄) 201.7 (C=O), 156.3 (C=O), 137.2 (Ar-C), 128.8 (Ar-C), 128.0 (Ar-C), 126.3 (Ar-C), 79.4 (C), 62.2 (αC), 37.3 (βC), 34.1 (C1), 27.3 (Boc-CH₃), 21.8 (C2). **HRMS-ESI** [M+Na]⁺ expected *m/z* 346.1447 and 347.1481 observed *m/z* 346.1455 and 347.1482 (± 2.31 and 0.29 ppm).



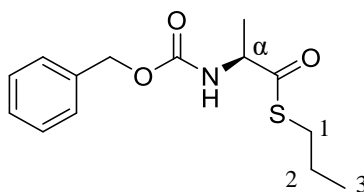
D-2-A

D-2-A (*N*-Boc-D-phenylalanine + 2-propanethiol) yield = 253.19 mg (69.2 %)

¹H NMR (500 MHz, Methanol-*d*₄) 7.32 – 7.17 (5 H, m, Ar-H), 4.34 (1 H, dd, *J* 10.4, 4.6 Hz, αH), 3.55 (1 H, hept, *J* 6.8 Hz, C1H), 3.17 (1 H, dd, *J* 14.0, 4.6 Hz, βH''), 2.78 (1 H, dd, *J* 14.0, 10.4 Hz, βH'), 1.52 – 1.19 (15 H, m, C2H + Boc-H). **¹³C NMR** (126 MHz, Methanol-*d*₄) 201.7 (C=O), 156.3 (C=O), 137.2 (Ar-C), 128.8 (Ar-C), 128.0 (Ar-C), 126.3 (Ar-C), 79.4 (C), 62.2 (αC), 37.3 (βC), 34.1 (C1), 27.3 (Boc-CH₃), 21.8 (C2). **HRMS-ESI** [M+Na]⁺ expected *m/z* 346.1447 and 347.1481 observed *m/z* 346.1458 and 347.1485 (± 3.18 and 1.15 ppm).

General Method for Substrate Subset B Synthesis:

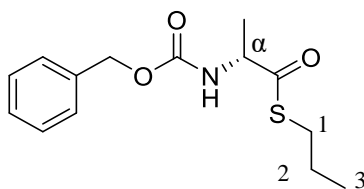
N-Cbz-alanine (L- or D-, 100 mg, 0.45 mmol) and 4-dimethylaminopyridine (5.43 mg, 0.045 mmol) were dissolved in dry DCM (2 mL), followed by the thiol (1-propanethiol or 2-propanethiol, 54 μL, 0.58 mmol). The solution was cooled to 0 °C in an ice bath with stirring, then *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (102.3 mg, 0.54 mmol) was added. The solution was stirred at room temperature overnight before extraction with citric acid (0.5 M, 3 x 3 mL) and brine (2 x 3 mL). The organic was dried using MgSO₄ and concentrated *in vacuo*, giving a clear liquid.



L-1-B

L-1-B (*N*-Cbz-L-alanine + 1-propanethiol) yield = 101.75 mg (72.4 %)

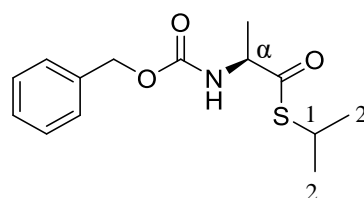
¹H NMR (800 MHz, Chloroform-*d*) 7.38 – 7.29 (5 H, m, Ar-H), 5.30 (1 H, d, *J* 8.4 Hz, NH), 5.18 – 5.10 (2 H, m, CH₂), 4.47 (1 H, p, *J* 7.4 Hz, αH), 2.86 (2 H, t, *J* 7.4 Hz, C1H), 1.60 (2 H, h, *J* 7.2 Hz, C2H), 1.40 (3 H, d, *J* 7.2 Hz, CH₃), 0.97 (3 H, t, *J* 7.4 Hz, C3H). **¹³C NMR** (201 MHz, Chloroform-*d*) 201.3 (C=O), 155.5 (C=O), 136.2 (Ar-C), 128.5 (Ar-C), 128.2 (Ar-C), 67.1 (CH₂), 56.7 (αC), 30.7 (C1), 22.7 (C2), 19.1 (CH₃), 13.3 (C3). **HRMS-ESI** [M+Na]⁺ expected *m/z* 304.0978 and 305.1011 observed *m/z* 304.0944 and 305.0966 (± 11.18 and 14.75 ppm).



D-1-B

D-1-B (*N*-Cbz-D-alanine + 1-propanethiol) yield = 127.94 (91 %)

¹H NMR (800 MHz, Chloroform-*d*) 7.38 – 7.28 (5 H, m, Ar-H), 5.30 (1 H, d, *J* 8.4 Hz, NH), 5.18 – 5.09 (2 H, m, CH₂), 4.47 (1 H, p, *J* 7.4 Hz, αH), 2.86 (2 H, t, *J* 7.4 Hz, C1H), 1.60 (2 H, h, *J* 7.2 Hz, C2H), 1.40 (3 H, d, *J* 7.2 Hz, CH₃), 0.96 (3 H, t, *J* 7.4 Hz, C3H). **¹³C NMR** (201 MHz, Chloroform-*d*) 201.3 (C=O), 155.5 (C=O), 136.2 (Ar-C), 128.5 (Ar-C), 128.2 (Ar-C), 67.1 (CH₂), 56.7 (αC), 30.7 (C1), 22.7 (C2), 19.1 (CH₃), 13.3 (C3). **HRMS-ESI** [M+Na]⁺ expected *m/z* 304.0978 and 305.1011 observed *m/z* 304.0938 and 305.0962 (± 13.15 and 16.06 ppm).

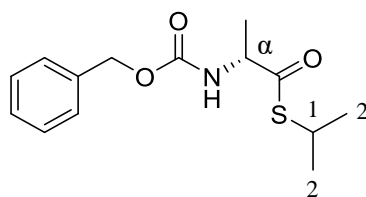


L-2-B

L-2-B (*N*-Cbz-L-alanine + 2-propanethiol) yield = 103.84 mg (73.8 %)

¹H NMR (500 MHz, Chloroform-*d*) 7.40 – 7.27 (5 H, m, Ar-H), 5.30 (1 H, d, *J* 8.4 Hz, NH), 5.20 – 5.06 (2 H, m, CH₂), 4.43 (1 H, p, *J* 7.4 Hz, αH), 3.61 (1 H, dq, *J* 14.8, 8.2, 7.4 Hz, C1H), 1.38 (3 H, d, *J* 7.4 Hz, CH₃), 1.30 (6 H, d, *J* 7.4 Hz, C2H). **¹³C NMR** (126 MHz, Chloroform-*d*) 201.2 (C=O), 155.5 (C=O), 136.2 (Ar-C), 128.5 (Ar-C), 128.2 (Ar-C), 128.2 (Ar-C), 67.1 (CH₂), 56.7 (αC), 34.8 (C1), 22.9 (C2), 19.1 (CH₃).

HRMS-ESI $[M+Na]^+$ expected m/z 304.0978 and 305.1011 observed m/z 304.0921 and 305.0946 (\pm 18.74 and 21.30 ppm).



D-2-B (*N*-Cbz-D-alanine + 2-propanethiol) yield = 87.50 mg (62.2 %)

$^1\text{H NMR}$ (500 MHz, Chloroform-*d*) 7.40 – 7.27 (5 H, m, Ar-H), 5.30 (1 H, d, J 8.4 Hz, NH), 5.20 – 5.06 (2 H, m, CH₂), 4.43 (1 H, p, J 7.4 Hz, α H), 3.62 (1 H, p, J 7.0 Hz, C1H), 1.38 (3 H, d, J 7.4 Hz, CH₃), 1.30 (6 H, d, J 7.4 Hz, C2H). **$^{13}\text{C NMR}$** (126 MHz, Chloroform-*d*) 201.2 (C=O), 155.5 (C=O), 136.2 (Ar-C), 128.5 (Ar-C), 128.2 (Ar-C), 128.2 (Ar-C), 67.1 (CH₂), 56.7 (α C), 34.8 (C1), 22.9(C2), 19.1 (CH₃). **HRMS-ESI** $[M+Na]^+$ expected m/z 304.0978 and 305.1011 observed m/z 304.0921 and 305.0943 (\pm 18.74 and 22.29 ppm).

Confirming Enantiopurity of Thioester Substrates

Marfey's reagent (1-fluoro-2-4-dinitrophenyl-5-L-alanine, FDAA) is a chiral derivatization reagent. It reacts with primary amines (through substitution of its fluoro-group), enabling quick and easy separation of enantiomers using regular HPLC conditions, arising from the generation of a diastereomer. Firstly, the thioester compounds were incubated in 1 HCl solution (1 M, 10 mg/mL) to remove the Boc-protecting group (thioester substrate subset A). Then, subsequent incubation of the thioester (50 μL) with Marfey's reagent (1 % w/v in acetone, 100 μL) and NaHCO₃ solution (1 M, 70 μL) was mixed thoroughly and incubated at 37 °C for 1 hour. The reaction mixture was diluted (1:50) with water, followed by centrifugation (16,000 xg , 5 min) before HPLC analysis.

Samples were analysed by HPLC on an Agilent Technologies 1260 Infinity HPLC with an Agilent Zorbax Eclipse Plus C18 Rapid Resolution 4.6 x 100 mm 3.5 μm column. Compounds were eluted using a 5 min gradient 5 – 75 % H₂O/ACN + 0.1 % FA over 12.5 min. Flow rates were kept constant at 1 mL min⁻¹ and elution was monitored by an absorbance at 340 nm.

Figure S4.11 shows the HPLC chromatogram of L-2-A and D-2-A thioesters incubated (separately) with Marfey's reagent. We can see that both thioesters give rise to two peaks, the first corresponding to either L- or D-phenylalanine, the second corresponding to 2-propanethiol (which

coordinates with Marfey's reagent and is now visible by HPLC). Thioester substrates were shown to be very stable in acidic conditions, however basic conditions (NaHCO_3) hydrolyse the thioester bond, returning it to its counterparts: phenylalanine and thiol. Nevertheless, these results determine that the thioesters (subset **A**) are synthesised without racemisation and are enantiomerically pure. Marfey's reagents could not be used to determine the enantiomeric purity of subset **B** thioesters as the CBz-protecting group proved difficult to remove (hydrogenation).

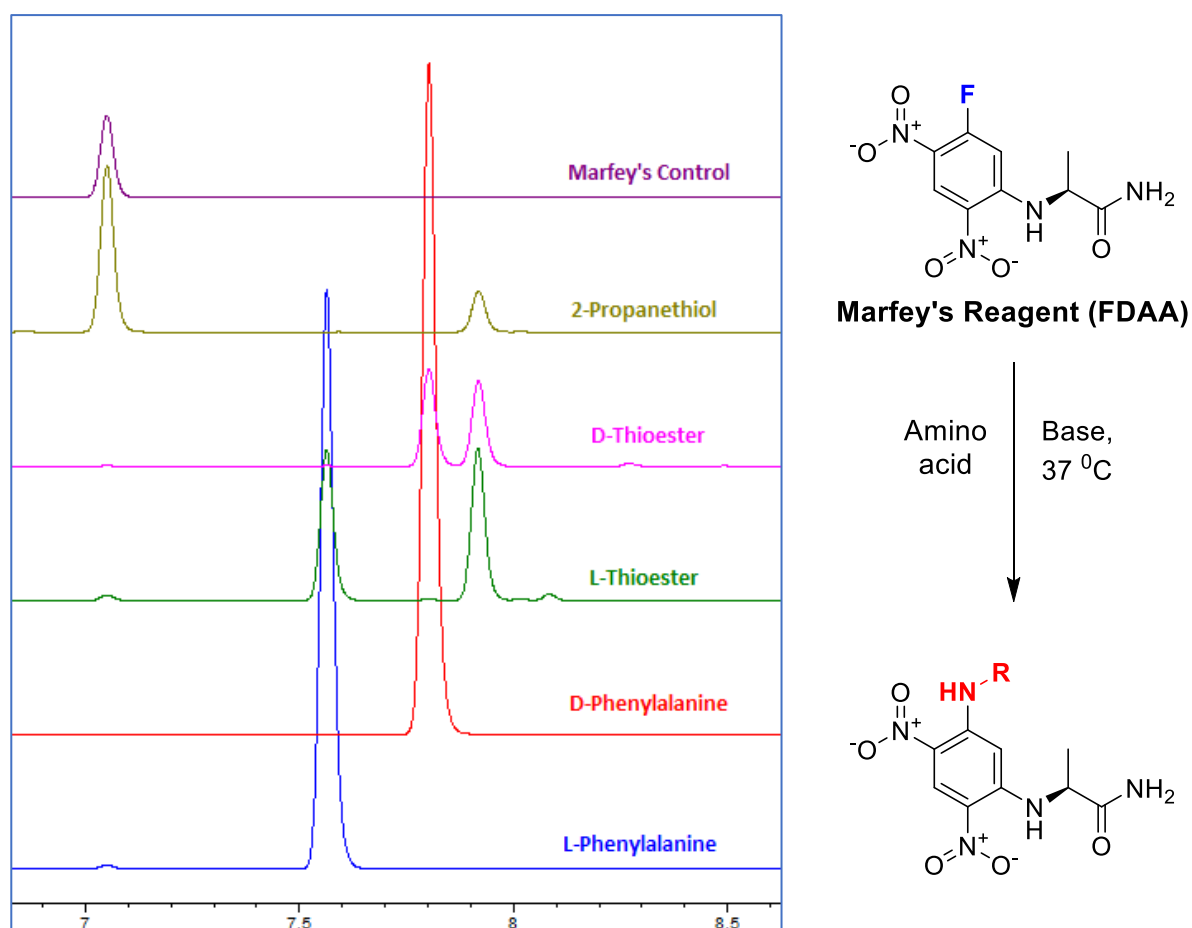


Figure S4.11. Marfey's reagent incubated with L- and D-thioester containing 2-propanethiol within its structure (L-2-A and D-2-A). The addition of Marfey's reagent and base (NaHCO_3) facilitates its reaction with primary amines to enable quick and easy separation of enantiomers on reverse-phase C18 column, by means of a second chiral centre (diastereomer).

In addition, for comparative HPLC analysis of the biotransformation samples, suitable HPLC conditions needed to be found so that substrates (thioesters) and product (*N*-Boc-L-phenylalanine or *N*-Cbz-alanine) could be quantified. Separation of all four thioester compounds (per subset **A/B**) proved successful using a Phenomenex Lux-Cellulose 3 chiral column and either a gradual gradient (subset **A**) or an isocratic gradient (subset **B**) with a 24 min acquisition time (further details of the

HPLC method are described in the instrumentation section, HPLC). Table S4.3 presents the retention times of all thioester compounds (using optimised, isocratic gradients), identifying that all compounds can be distinguished from one another and they are enantiomerically pure. Products (carboxylic acids) are much more polar, so they have shorter retention times on the column (7 – 8 min).

Table S4.3. Table presenting the retention times of thioester compounds on a Phenomenex Lux-Cellulose 3 chiral column using isocratic gradients. Successful separation of all thioesters, identifying that they are enantiomerically pure.

	Subset A				Subset B			
Thioester	L-1-A	D-1-A	L-2-A	D-2-A	L-1-B	D-1-B	L-2-B	D-2-B
Time (min)	18.63	18.81	18.11	17.76	13.97	14.43	12.47	13.54

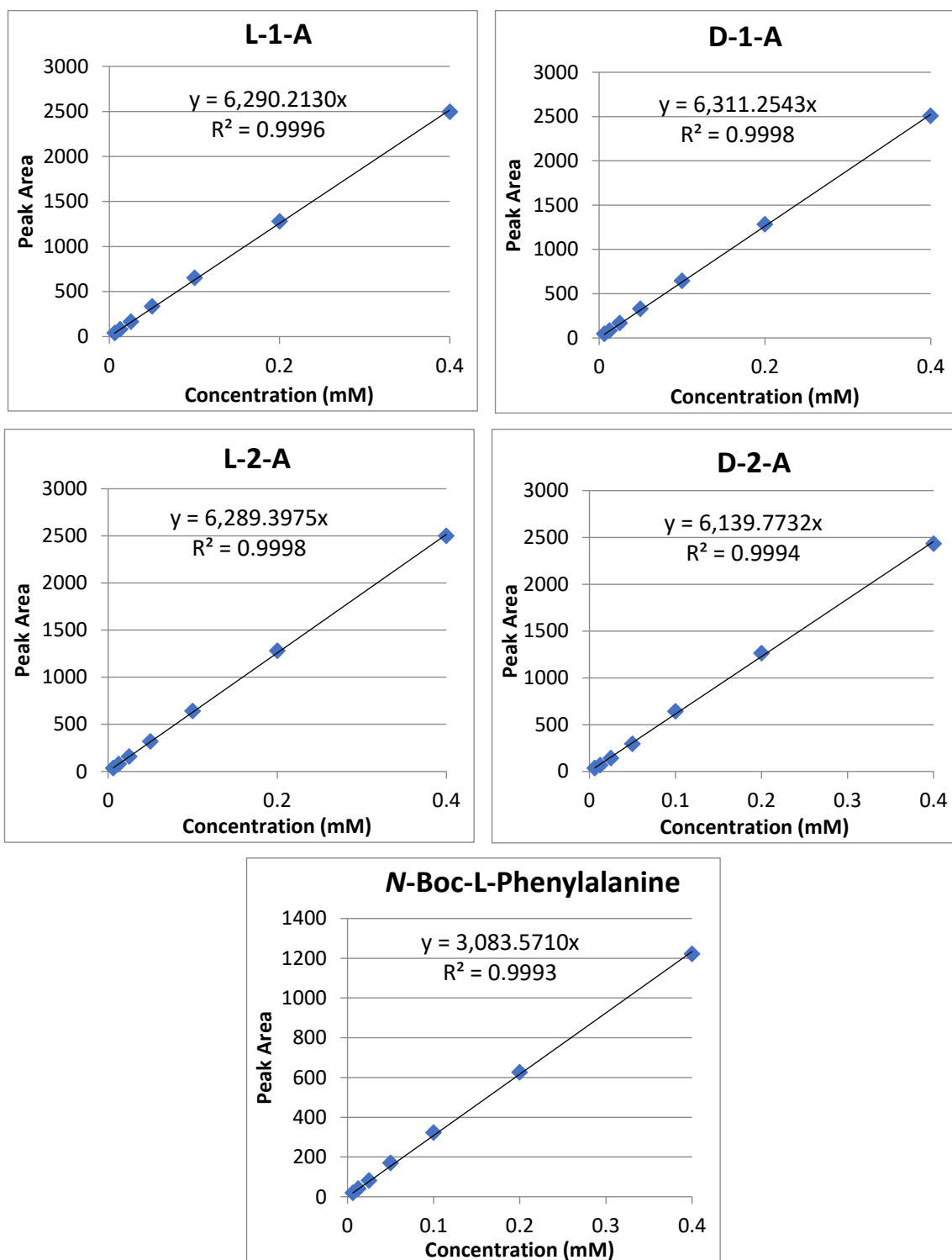


Figure S4.12. HPLC calibrations (chiral column) of substrate subset A and the corresponding L-product, N-Boc-L-phenylalanine. Absorbance measured at 220 nm, the averages of triplicate data shown are shown.

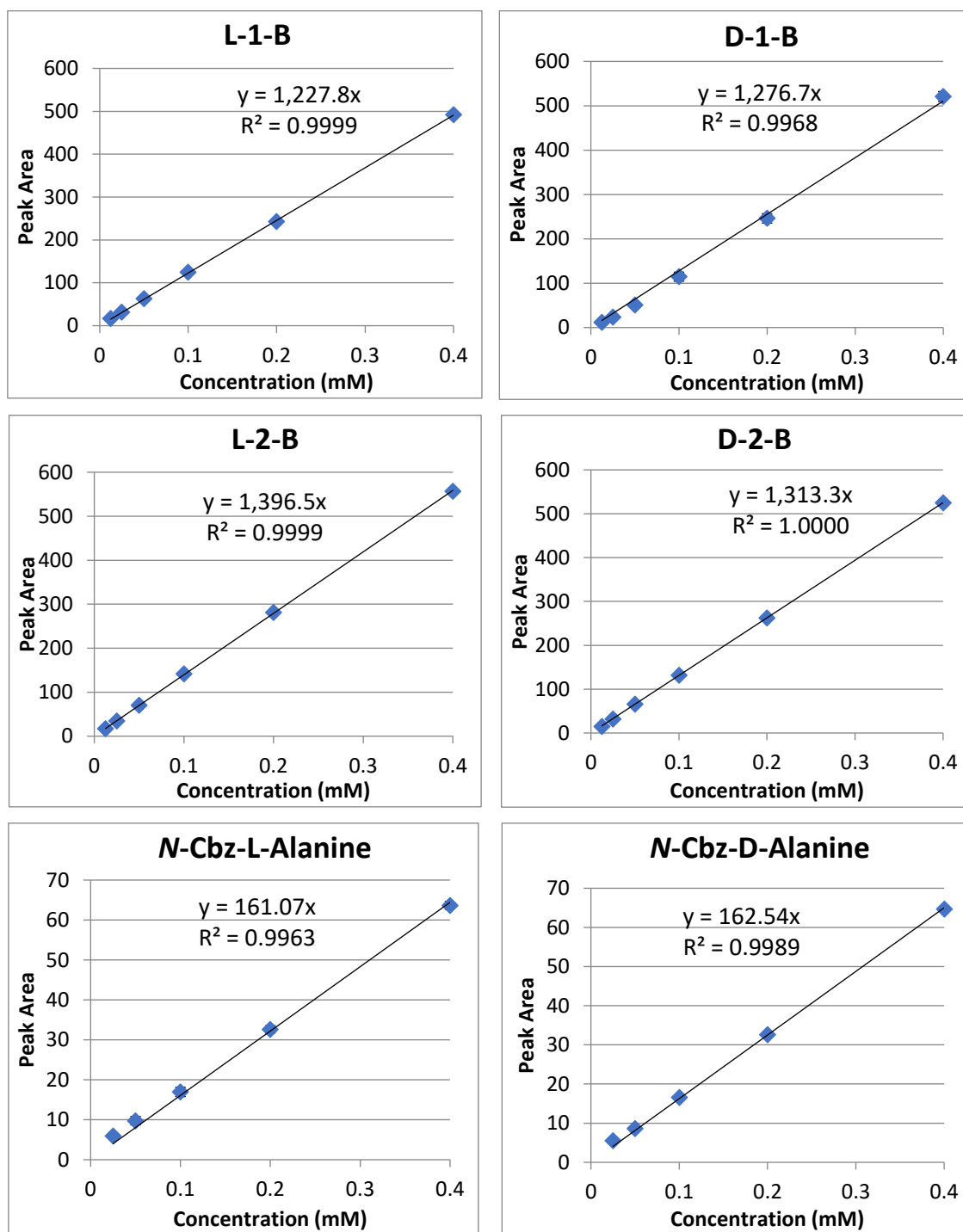


Figure S4.13. HPLC calibrations (chiral column) of substrate subset B and the corresponding products, N-Cbz-L-/D-alanine. Absorbance measured at 254 nm, the averages of triplicate data shown are shown.

REFERENCES

- JOO, T. H., KIM, K. & KIM, M. S. 1986. Surface-Enhanced Raman Scattering (SERS) of 1-propanethiol in silver sol. *The Journal of Physical Chemistry*, 90, 5816-5819.
- LEOPOLD, N. & LENDL, B. 2003. A New Method for Fast Preparation of Highly Surface-Enhanced Raman Scattering (SERS) Active Silver Colloids at Room Temperature by Reduction of Silver Nitrate with Hydroxylamine Hydrochloride. *The Journal of Physical Chemistry B*, 107, 5723-5727.
- PANG, Y. S., HWANG, H. J. & KIM, M. S. 1998. Reversible Temperature Dependence in Surface-Enhanced Raman Scattering of 1-Propanethiol Adsorbed on a Silver Island Film. *The Journal of Physical Chemistry B*, 102, 7203-7209.
- SINGH, R. & WHITESIDES, G. M. 2010. Thiol—disulfide interchange. *Sulphur-Containing Functional Groups (1993)*. John Wiley & Sons, Inc.
- STEWART, A., ZHENG, S., MCCOURT, M. R. & BELL, S. E. J. 2012. Controlling Assembly of Mixed Thiol Monolayers on Silver Nanoparticles to Tune Their Surface Properties. *ACS Nano*, 6, 3718-3726.
- STREKAL', N. D., GERMAN, A. E., GACHKO, G. A. & MASKEVICH, S. A. 2000. Spectra of surface-enhanced Raman scattering of 1-propanethiol and 3-mercaptopropionic acid chemisorbed on thin silver films. *Optics and Spectroscopy*, 89, 834-840.

5 Real-Time Monitoring of Enzyme-Catalysed Reactions using Deep UV Resonance Raman Spectroscopy

Chloe Westley^{††}, Heidi Fisk^{††}, Yun Xu[†], K. A. Hollywood[†], Andrew J. Carnell[‡], Jason Micklefield[†], Nicholas J. Turner[†], Royston Goodacre[†]

[†] School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK.

[‡] Department of Chemistry, University of Liverpool, Liverpool, L69 7ZD.

5.1 DECLARATION

This chapter consists of one recently published journal article:

WESTLEY, C., FISK, H., XU, Y., HOLLYWOOD, K. A., CARNELL, A. J., MICKLEFIELD, J., TURNER, N. J. & GOODACRE, R. 2017. Real-Time Monitoring of Enzyme-Catalysed Reactions using Deep UV Resonance Raman Spectroscopy. *Chemistry – A European Journal*, 23, 6983-6987.

This article has been reproduced in an unchanged format except for minor adjustments to incorporate it into this thesis.

As joint, primary author on this publication, I carried out the optimisation of NHase-related biotransformations. Dr. Chloe Westley optimised XO-related biotransformations. Together, we optimised the real-time UVRR set-up, performed all experiments and wrote the journal article. Dr. Yun Xu performed the chemometrics comprising of MCR-ALS. Dr. Katherine Hollywood was fundamental in the maintenance and troubleshooting of the UVRR instrument. Prof. Nick Turner and Prof. Andrew Carnell gave help and advice surrounding the biotransformations. Co-supervisor, Prof. Jason Micklefield, gave important insights into the biotransformations studied, as well as helping to shape the direction of the research. Prof. Royston Goodacre, as Principal Investigator, provided insights and advice on the project brief, giving direction during the research.

5.2 ABSTRACT

For enzyme-catalysed biotransformations, continuous *in situ* detection methods minimise the need for sample manipulation, ultimately leading to more accurate real-time kinetic determinations of substrate(s) and product(s). We have established for the first time an on-line, real-time quantitative approach to monitor simultaneously multiple biotransformations based on UV resonance Raman (UVR) spectroscopy. To exemplify the generality and versatility of this approach, multiple substrates and enzyme systems were used involving nitrile hydratase (NHase) and xanthine oxidase (XO), both of which are of industrial and biological significance, and incorporate multistep enzymatic conversions. Multivariate data analysis of the UVR spectra, involving multivariate curve resolution-alternating least squares (MCR-ALS), was employed to effect absolute quantification of substrate(s) and product(s); repeated benchmarking of UVR combined with MCR-ALS by HPLC confirmed excellent reproducibility

5.3 INTRODUCTION

Reaction monitoring based on analytical spectroscopy is broadly used to observe chemical changes in a variety of applications, including energy and fuel industries, bio-based technologies and processes, pharmaceuticals, as well as for biocatalyst discovery and optimization (Hinz, 2006, Workman *et al.*, 2007, 2009, Reetz, 2013). Reaction monitoring provides essential information in terms of molecular speciation, and affords key insights into reaction mechanisms, kinetics and the biochemical process of the system investigated. Furthermore, real-time (in contrast to off-line) reaction monitoring greatly improves the efficiency and accuracy of the overall process, with label-free spectroscopic-based methodologies being employed (Gardner *et al.*, 2013, Yan *et al.*, 2017). Laborious sample preparation methods and purification steps are no longer required prior to analysis, thus minimising the need for transfers and sample handling, ultimately reducing errors. Advancements in engineering, such as the incorporation of robotics and sophisticated computational programs, lead to overall improvements and as a consequence, there is a significant reduction in the time taken for analysis (Kourti, 2006, Workman *et al.*, 2009, Reetz, 2013).

However, for biocatalytic applications, real-time reaction monitoring provides specific challenges: the sensitivity required to monitor conversions is often an issue as low substrate concentrations are commonly used (Dadd *et al.*, 2000). As a result, monitoring conversions involving detection/presence of intermediates in multi-step biotransformations can be problematic. The most common method of measuring the rate of substrate turnover is the use of spectrophotometric

assays (Kumar and Clark, 2006, Acker and Auld, 2014, Packer and Liu, 2015). Although these assays are easy to use and interpret, a major limiting factor is the requirement for a fluoro-/chromo-genic reporter. However, in most cases, this means that the activity of the enzyme is detected indirectly or that improved enzyme activities may be selected based on the use of an idealised substrate, which may not translate to the real one. Whilst, other spectroscopic and spectrometric physicochemical techniques are commonly employed (*viz.* NMR, HPLC and LC-MS), these methods too have notable drawbacks such as extensive sample preparation, high equipment costs, large solvent volumes, long acquisition times, and in some instances provide limited structural information (Fox *et al.*, 2004, von Ahsen and Bömer, 2005, Kumar and Clark, 2006, Reymond and Babiak, 2007, Acker and Auld, 2014). Therefore, there is a need for rapid, robust and reagent free on-line high-throughput screening methods to overcome these significant drawbacks.

Raman spectroscopy presents itself as an ideal analytical technique to use for screening applications, as it is rapid, non-destructive and non-invasive. Moreover, it can be performed *in situ* in aqueous environments and provides molecular specific information. We have previously shown that the conversion of glucose to ethanol by yeast can be monitored by Raman spectroscopy with an NIR excitation wavelength (Shaw *et al.*, 1999). However, Raman scattering is a relatively weak physical phenomenon and is often further exacerbated by fluorescence interference when excitation involves lasers in the visible EM (Vankeirsbilck *et al.*, 2002). As a consequence, enhancement techniques are regularly employed to increase scattering efficiency. Surface enhanced Raman scattering (SERS), a surface-sensitive Raman enhancement technique, has previously been used to monitor enzymatic biotransformations indirectly (Moore *et al.*, 2004). Very recently, we successfully demonstrated a >30-fold reduction in acquisition times for multiple enzymatic steps measuring analytes directly. This delivered high levels of accuracy and reproducibility, highlighting its suitability as an alternative screening technique (Westley *et al.*, 2016). However, SERS requires a roughened metal surface that cannot be readily used for on-line assessment of enzymatic reactions, so at best is only suitable for at-line analysis.

Ultraviolet resonance Raman (UVRR) spectroscopy is a variant of 'normal' Raman and involves the enhancement of Raman scattering by UV (in this case, at 244 nm). When the frequency of the laser coincides/matches the frequency of the molecule's electronic transition, enhancements of $10^3 - 10^5$ can be observed (Asher, 1993). UVRR is an attractive technique for use in screening applications as the biotransformation(s) can be performed in real-time; with little interference from background fluorescence (Asher and Johnson, 1984). Moreover, the ability to measure analytes of interest directly without the need to quench the system, or have additional reagents as needed for SERS, is advantageous. Although this technique requires the absorption of laser light by chromophores in

the UV region (most notably from aromatics and fused ring systems), many complex biological systems fulfil this requirement, with nucleic acids and amino acids being particularly amenable to UVRR (Strekas and Spiro, 1972, Carey, 2012, Ashton *et al.*, 2015).

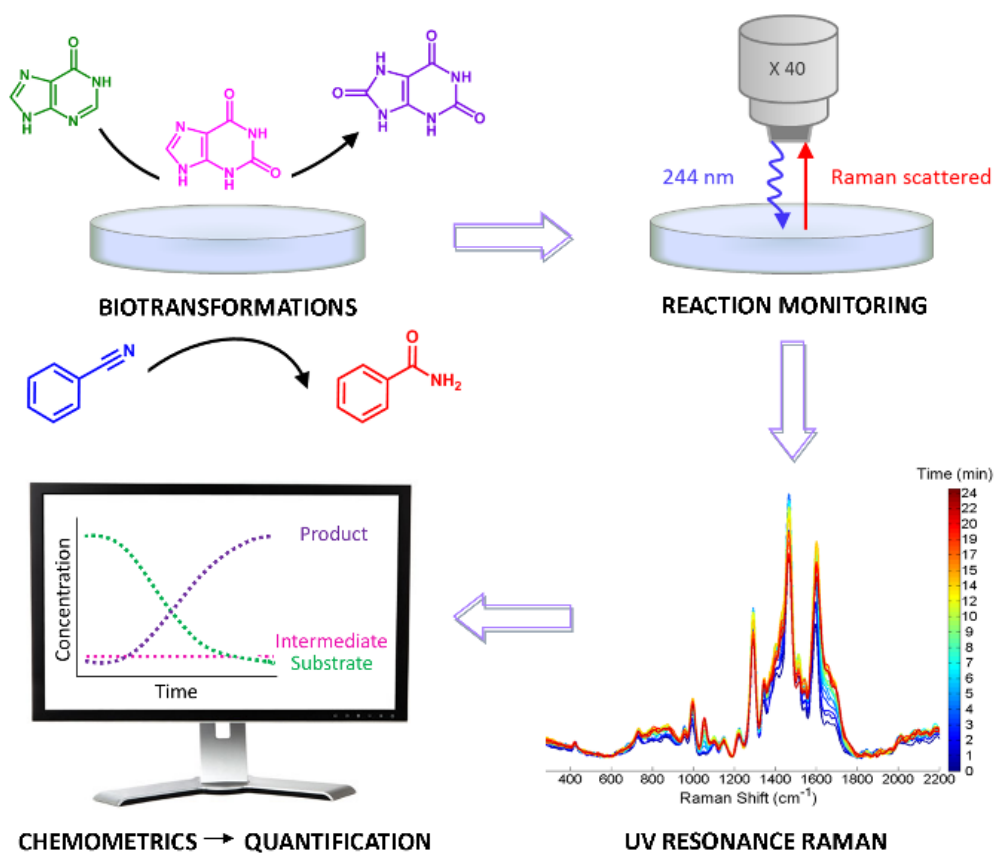


Figure 5.1. Workflow of the UVRR approach for real-time reaction monitoring of multiple biotransformations.

In the present study we demonstrate how UVRR can be used for real-time reaction monitoring using two different biocatalytic reactions (see Figure 5.1). First, we focus on the conversion of nitriles to their corresponding amides using nitrile hydratase (NHase) (Scheme 5.1a), a class of enzyme extensively used in chemical synthesis within various industries - with acrylamide, nicotinamide (Vitamin B₃) and pyrazinamide (anti-tuberculosis agent) being notable examples (Asano *et al.*, 1982, Kobayashi *et al.*, 1992, Mauger *et al.*, 1988, Nagasawa *et al.*, 1988, Ghisalba *et al.*, 2009). Second, to illustrate multiple reaction steps we have applied the method to xanthine oxidase (XO) catalysed biotransformations (Scheme 5.1b). XO catalyses the oxidation of a wide range of substrates including purines and xenobiotic compounds, with xanthine and hypoxanthine, its natural substrates, being the focus in this investigation (Hille, 1996, 2005, Harrison, 2002).

5.4 EXPERIMENTAL

5.4.1 Reagents and Materials.

All chemical reagents were of analytical grade and used with no additional purification unless otherwise stated. Nitrile hydratase (EC 4.2.1.84) was purchased from Prozomix (Northumberland, UK). Xanthine oxidase microbial lyophilized powder (EC 1.17.3.2) was purchased from Sigma Aldrich Ltd. (Dorset, UK).

5.4.2 Raman instrumentation.

UVRR analysis was performed using a Renishaw Raman 1000 system (Renishaw, Wotton-under-Edge, Gloucestershire, UK). Approximately 0.2 mW of power was delivered to the sampling point using a Lexel Model 95 ion laser emitting at 244 nm. The solution sample for analysis was continuously stirred beneath the laser to allow thorough mixing of the enzyme and substrate(s) and to avoid photo-degradation; no photo damage was observed (see SI, section 'photo-degradation of sample', Figure S5.1 and S5.2). Spectra were collected with an acquisition time of 20 s.

5.4.3 Reaction conditions for biotransformations 1 and 2.

A stock solution of NHase was prepared. The starting reaction mixture contained either benzonitrile (biotransformation 1) or *p*-tolunitrile (biotransformation 2) dissolved in potassium phosphate buffer at pH 7.2 (final concentration 1.25×10^{-2} M with 2.5 % MeOH) and NHase solution (40 μ L for biotransformation 1; 200 μ L for biotransformation 2).

5.4.4 Reaction conditions for biotransformations 3 and 4.

A stock solution of XO was prepared. Starting reaction mixture contained either xanthine (biotransformation 3) or hypoxanthine (biotransformation 4) dissolved in potassium phosphate buffer at pH 7.6 (final concentration 7.5×10^{-4} M) and XO (120 μ L for biotransformation 3; 150 μ L for biotransformation 4).

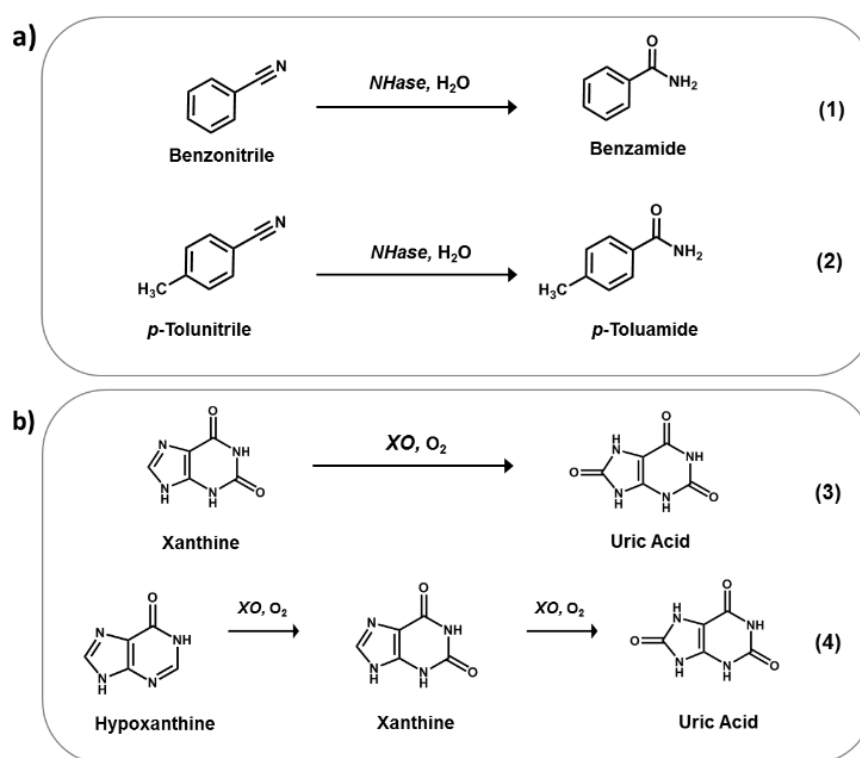
5.4.5 Reaction sample preparation and monitoring.

Initially, the substrate was focussed under the microscope objective. At various time points after enzyme addition: (i) on-line UVRR spectra were collected throughout the biotransformation; and (ii) samples collected, quenched and analysed by off-line HPLC.

5.5 RESULTS AND DISCUSSION

5.5.1 Optimisation of NHase and XO biotransformations

The biotransformations of interest (1-4) are shown in Scheme 5.1. Reaction conditions for all biotransformations were optimised accordingly for UVRR monitoring (see methods in SI for full details). For optimum UVRR spectra, a 20 second acquisition time and approximately 0.2 mW laser power at sample was required using an excitation wavelength in the deep UV at 244 nm. Characteristic UVRR spectra for each analyte, with unique peaks identified are summarised in Figure 5.2 (see Table S5.1 and S5.2 in SI for tentative band assignments).



Scheme 5.1. Biotransformations (1-4) selected for monitoring by UVRR: **(a)** bioconversions of nitriles to the corresponding amides by nitrile hydratase (NHase); **(b)** bioxidation of purines by xanthine oxidase (XO).

5.5.2 Instrument set up for real-time reaction monitoring

To monitor the enzyme-catalysed biotransformations, the instrument had to be modified and optimised (see Figure S5.1 in SI); briefly, a magnetic stirrer plate was inserted below the turntable, with the reaction vessel (containing a magnetic stirrer bar) on top, focused under the microscope

objective. The reaction was initiated upon the introduction of enzyme. Continuous stirring permitted maximal enzyme-substrate interaction throughout the reaction and provides a true representation of the conversion of substrate(s) to product(s). This set-up also allows the energy from the laser source to be evenly distributed over a much larger volume.

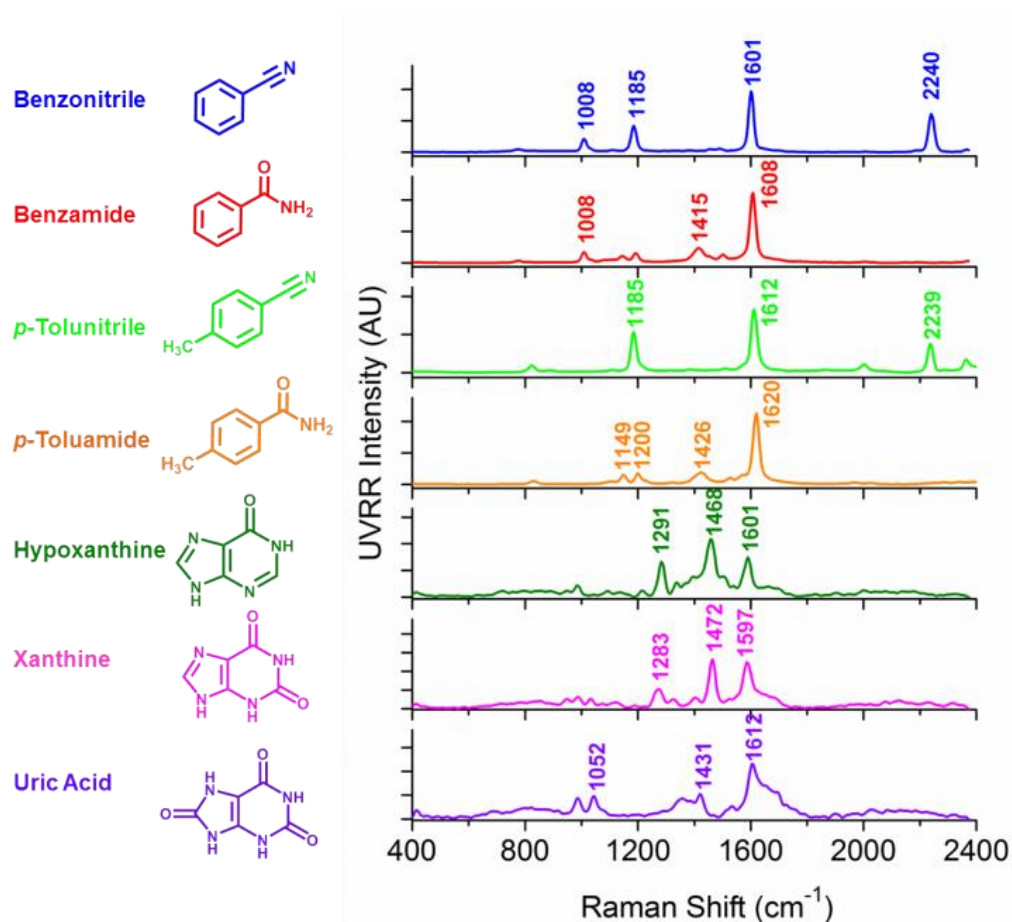


Figure 5.2. Average UVRR spectra ($n = 5$) of each analyte for both biotransformations: benzonitrile (blue), benzamide (red), p-tolunitrile (bright green) and p-toluamide (orange), hypoxanthine (green), xanthine (pink) and uric acid (purple). For NHase analytes spectra were obtained at 1.25×10^{-2} mM, pH 7.2. For XO analytes spectra were obtained at 7.5×10^{-4} M, pH 7.6. All spectra are representative of starting reaction concentrations with characteristic peaks annotated. UVRR spectra were obtained for 20 s with baseline correction, normalisation and smoothing applied (see SI, 'data processing' for full details).

Furthermore, to minimise the risk of reduced focus on the sample through solvent evaporation and removal of volume for HPLC analysis, the reaction was performed on a 10 mL scale. An initial concern was the integrity of the sample when subjected to a highly powered laser, however, no spectral changes (and hence no photo-degradation) was observed throughout the reaction time course (see the SI, section 'Photo-degradation of sample' and Figure S5.2). Interestingly, from these

investigations, we observed bathochromic shifts (as a function of pH) for XO analytes (see the SI, section ‘Bathochromic shifts of XO analytes’ and Figures S5.3 – 5.5).

5.5.3 Monitoring NHase biotransformations 1 and 2

Although there are characteristic peaks for each analyte, thus distinguishing starting material from product, the UVRR spectra were highly similar with many overlapping peaks (especially for XO analytes). Therefore, for all biotransformations, multivariate curve resolution - alternating least squares (MCR-ALS) was employed. MCR-ALS is a popular feature extraction tool for mixture analysis and was used to extract the necessary information (pure component spectra and corresponding concentrations) to predict absolute levels of the analytes within a mixture (see Figure S5.6 in SI for a flow diagram summarising this MCR-ALS approach) (Tauler, 1995, Jaumot *et al.*, 2005).

We initially looked at the conversion of benzonitrile to benzamide (biotransformation 1) with <50 % conversion achieved over a 20 min time period. The deconvolved spectra for each analyte were highly similar to the UVRR spectra from the pure substrate and product (Figure S5.7 in SI). For brevity purposes, we only represent the graphical results of one replicate. Figure 5.3a shows the UVRR spectra over the reaction time course, illustrating (by use of a colour bar) the increase and decrease of characteristic peaks with respect to time. Time points with both HPLC and UVRR data were used as the training set for the MCR-ALS model (*i.e.* HPLC was used as external validation – see Figure S5.8 in SI for HPLC calibration). Time points with UVRR data (but without HPLC data) were used as the test set. As one can easily observe from Figure 5.3b, the UVRR predictions are in excellent agreement with the HPLC results. This is further supported by high R^2 values across all replicates, with an average of 0.964 and 0.983 for substrate and product respectively (see Table 5.1). The coefficient of determination, R^2 , is the proportion of variability in a data set that is accounted for by a statistical model (in this case MCR-ALS) with R^2 values closer to one indicating an excellent fit. Notably, this experiment was conducted on five separate occasions, over a four-week period thus accounting for day-to-day instrument variance, ultimately demonstrating its robustness for on-line reaction monitoring.

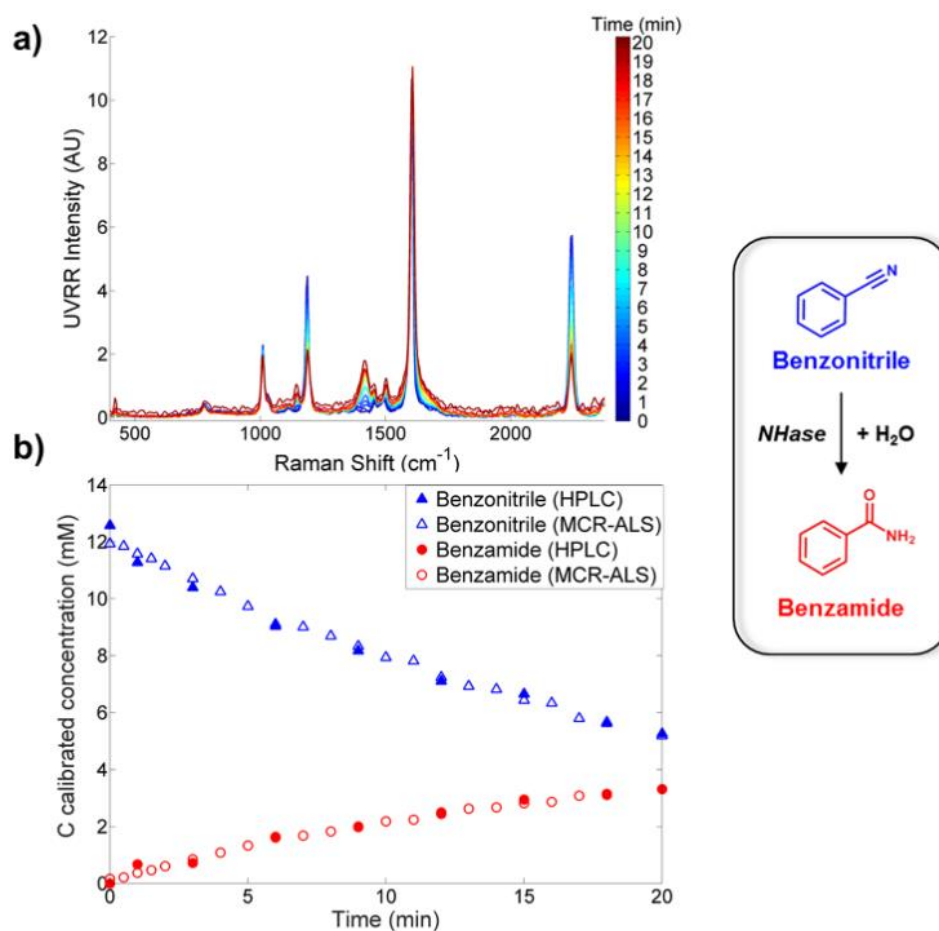


Figure 5.3. An MCR-ALS model was applied to the UVRR data for the conversion of benzonitrile to benzamide (biotransformation 1). (a) Shows accumulative spectra taken over the 20 min time course. The colour bar highlights each time point monitored with the start ($t = 0$) in blue and the end point ($t = 20$) in red. (b) Shows the reaction dynamics from real-time UVRR measurements (denoted by outlined symbols) and off-line HPLC data (denoted by solid symbols) as a function of time. UVRR spectra were obtained for 20 s with baseline correction, normalisation and smoothing applied (see SI ‘data processing’ for full details). Data shown are from replicate 2.

Moreover, to extend this approach, we next investigated a similar NHase substrate, *p*-tolunitrile – only differing by a CH₃ group, yet possessing unique peaks when compared to benzonitrile. Once again, with this biotransformation (biotransformation 2), we were able to monitor the reaction successfully. The UVRR and HPLC results were in very good agreement with one another, with typical R² values of 0.898 and 0.914 for *p*-tolunitrile and *p*-toluamide respectively (see Figure S5.9 and Table S5.3 in SI).

Table 5.1. A summary of the regression co-efficients (R^2) across all five replicates for the two separate biotransformations:

Replicate	NHase (Biotransformation 1)		XO (Biotransformation 3)	
	Benzonitrile R^2	Benzamide R^2	Xanthine R^2	Uric Acid R^2
1	0.959	0.993	0.990	0.955
2	0.987	0.987	0.954	0.977
3	0.969	0.982	0.916	0.965
4	0.962	0.980	0.973	0.987
5	0.942	0.973	0.957	0.983

Biotransformation 1 (benzonitrile to benzamide) using NHase, with overall mean R^2 values of 0.964 and 0.983, respectively. Biotransformation 3 (xanthine to uric acid) using XO, with overall mean R^2 values of 0.958 and 0.973, respectively.

5.5.4 Monitoring XO biotransformations 3 and 4

To demonstrate versatility of UVRR combined with MCR-ALS, we then performed analysis on a different, second enzyme system: XO (biotransformations 3 and 4). We have previously shown that we can monitor these conversions using SERS, and as already discussed, this involves the use of additional reagents preventing real-time monitoring (Westley *et al.*, 2016). Therefore, this UVRR approach should overcome this main drawback. Furthermore, this enzyme system in itself provided a challenge with the analytes being highly similar in structure, only differing by additional carbonyl groups (Scheme 5.1b). First, the two-analyte conversion (biotransformation 3) of xanthine to uric acid was investigated, with >50 % conversion achieved in 18 min. Adopting the same process, the MCR-ALS model was applied to the reaction data with results being in excellent agreement with the HPLC analysis (see Figure 5.4 and Figure S5.10 in SI for deconvolved spectra of each analyte and Figure S5.11 in SI for HPLC calibration). Average R^2 values of 0.958 and 0.973 were obtained for xanthine and uric acid, respectively (see Table 5.1). We then extended this to a third analyte to include the precursor hypoxanthine (biotransformation 4), ultimately demonstrating the flexibility of this real-time, on-line reaction monitoring screen for a more complex, multicomponent reaction system.

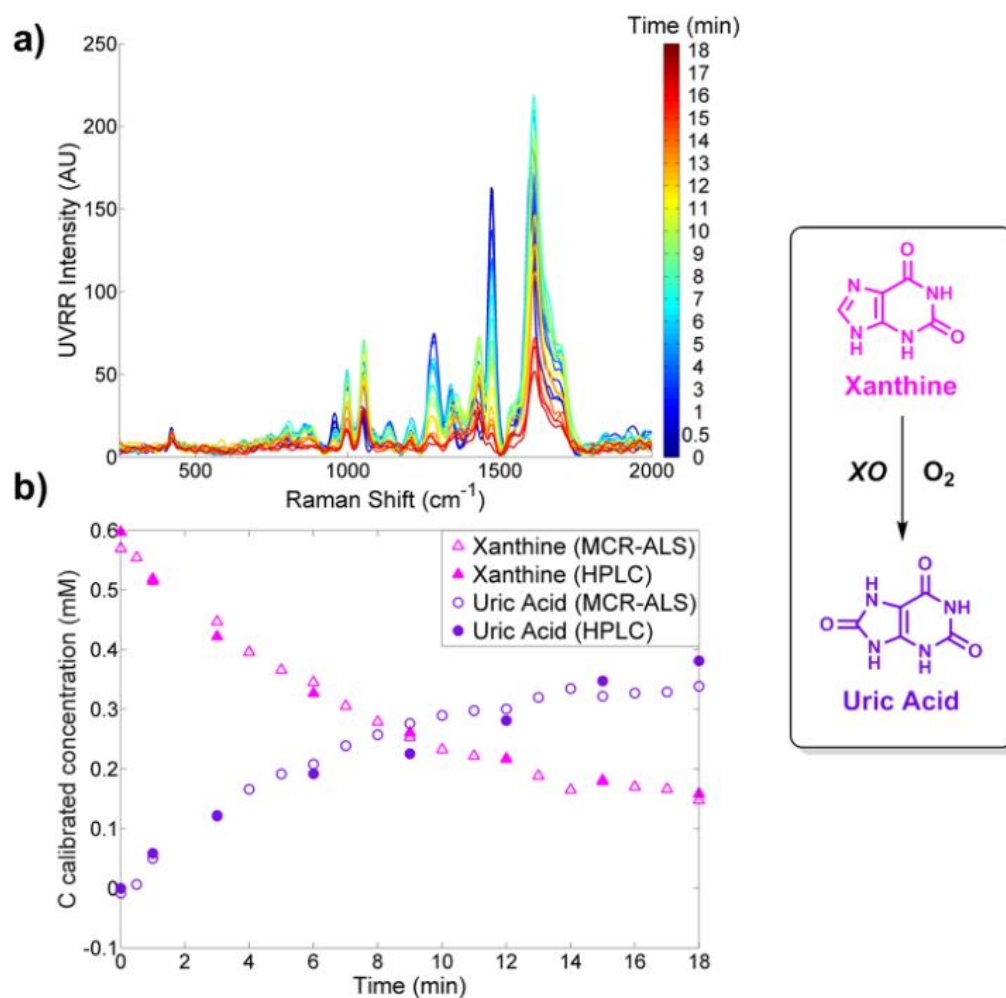


Figure 5.4. An MCR-ALS model was applied to the UVRR data for the conversion of xanthine to uric acid, biotransformation 3. **(a)** Accumulative spectra taken over the 18 min time course. The colour bar highlights each time point monitored with the start ($t = 0$) in blue and the end point ($t = 18$) in red. **(b)** Shows the reaction dynamics from real-time UVRR measurements (denoted by outlined symbols) and off-line HPLC data (denoted by solid symbols) as a function of time. UVRR spectra were obtained for 20 s with baseline correction, normalisation and smoothing applied (see SI ‘data processing’ for full details). Data shown are from replicate 1.

The reaction conditions were modified slightly, with <40 % conversion reached after 35 min. Again MCR-ALS analysis was employed with the deconvolved UVRR spectra being highly consistent with the pure spectra for each analyte (Figure 5.5a-c). The UVRR predictions were in very good agreement with the HPLC results (see Figure 5.5d). The R^2 values were slightly lower than biotransformation 3 (see Table S5.4 in SI), which was to be expected due to the increased complexity of the system, as well as the highly similar spectra between the three analytes. Noticeably, the R^2 value of xanthine was lower (biotransformation 4) than previous—this is due to the low overall concentration of xanthine (<8 %) throughout the reaction. This was further supported by the proposed mechanism of XO (based on xanthine dehydrogenase, XDH, from *Rhodobacter*

capsulatus), in which hypoxanthine binds to the active site and is converted to xanthine by oxidation at the C-2 position. Xanthine is then released, before binding in a different orientation to present the C-8 for oxidation to give uric acid (Dietzel *et al.*, 2009). This means that the concentration of the intermediate remains low throughout.

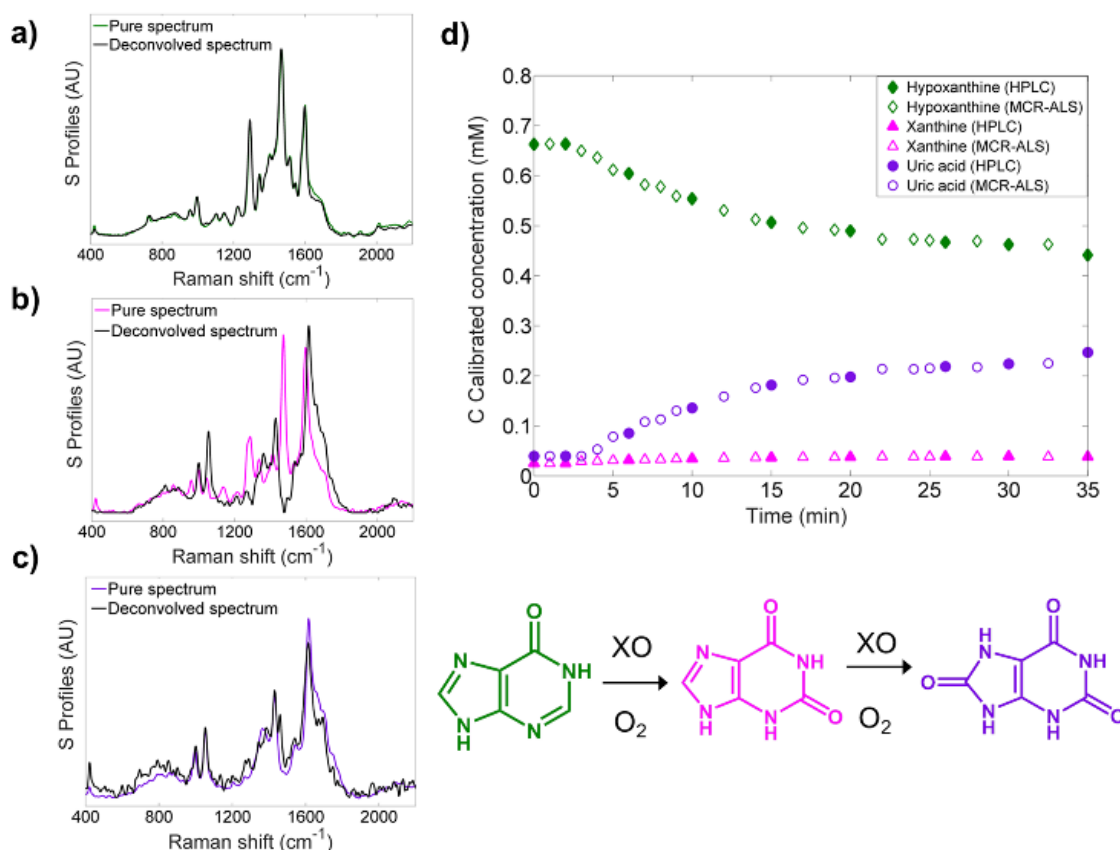


Figure 5.5. An MCR-ALS model was applied to the UVRR data where it successfully deconvolved spectra into its pure components for biotransformation 4: (a) hypoxanthine (b) xanthine and (c) uric acid. (d) Shows the reaction dynamics from real-time UVRR measurements (denoted by outlined symbols) and off-line HPLC data (denoted by solid symbols) as a function of time for the conversion of hypoxanthine to xanthine to uric acid. UVRR spectra were obtained for 20 s with baseline correction, normalisation and smoothing applied (see SI 'data processing' for full details). Data shown are from replicate 1.

5.5.5 Generalisation of the UVRR approach

Where this work could be further explored includes investigating the two separate pathways known to catalyse the conversion of nitrile containing compounds into their corresponding carboxylic acid: either in a single step (nitrilase) or a two-step process (nitrile hydratase and amidase; see the SI, Figure S5.12a). Fluorometric and colorimetric assays have previously been reported, including successful differentiation between the two pathways, however, only semi-quantitative analysis has

been possible (Reisinger *et al.*, 2006, Lin *et al.*, 2011, Angelini *et al.*, 2015). Extending on biotransformation 1, we have shown that we can potentially use this UVRR approach to monitor such cascades as the corresponding carboxylic acid involved in this pathway has unique, characteristic peaks (Figure S5.12b in SI). Furthermore, other nitrile containing substrates, *e.g.* 3-pyridinecarbonitrile and pyrazinecarbonitrile, which are precursors for important pharmaceutical products, can similarly be monitored (Figure S5.12c and S5.12d in SI). These results further demonstrate the general utility of the UVRR approach for enzyme reaction monitoring.

5.6 CONCLUSION

In this study, we have developed a label-free, rapid, on-line screening method to monitor biological and industrially relevant biotransformations based on UVRR spectroscopy. To demonstrate the general utility of this approach, multiple substrates and enzyme systems were investigated which included single, multiple and cascade enzyme systems. UVRR spectra acquisitions were rapid (20 s per measurement) and when combined with MCR-ALS produced substrate(s) and product(s) concentrations that were completely in agreement with off-line HPLC measurements. Additional bench marking involved repeat biotransformations conducted over several weeks and this established the excellent reproducibility and robustness of this novel analytical approach.

In conclusion, we believe that additional optimisation and configuration of the UVRR instrument set up will make this approach amenable to miniaturization and *in situ* point-and-shoot analyses (Ellis *et al.*, 2015), thus enhancing the potential for wider application. The method could also be developed as a high-throughput screening technique for enzyme activity, including the monitoring of cascade biotransformations, as well as for investigating enzyme inhibitors.

5.7 REFERENCES

- ACKER, M. G. & AULD, D. S. 2014. Considerations for the design and reporting of enzyme assays in high-throughput screening applications. *Perspectives in Science*, 1, 56-73.
- ANGELINI, L. M. L., DA SILVA, A. R. M., ROCCO, L. D. F. C. & MILAGRE, C. D. D. F. 2015. A high-throughput screening assay for distinguishing nitrile hydratases from nitrilases. *Brazilian Journal of Microbiology*, 46, 113-116.
- ASANO, Y., YASUDA, T., TANI, Y. & YAMADA, H. 1982. A New Enzymatic Method of Acrylamide Production. *Agricultural and Biological Chemistry*, 46, 1183-1189.
- ASHER, S. A. 1993. UV resonance Raman spectroscopy for analytical, physical, and biophysical chemistry. Part 1. *Analytical Chemistry*, 65, 59A-66A.
- ASHER, S. A. & JOHNSON, C. R. 1984. Raman spectroscopy of a coal liquid shows that fluorescence interference is minimized with ultraviolet excitation. *Science*, 225, 311-3.
- ASHTON, L., HOGWOOD, C. E. M., TAIT, A. S., KULIGOWSKI, J., SMALES, C. M., BRACEWELL, D. G., DICKSON, A. J. & GOODACRE, R. 2015. UV resonance Raman spectroscopy: a process analytical tool for host cell DNA and RNA dynamics in mammalian cell lines. *Journal of Chemical Technology & Biotechnology*, 90, 237-243.
- CAREY, P. 2012. *Biochemical applications of Raman and resonance Raman spectroscopes*, London, Academic Press Inc.
- DADD, M. R., SHARP, D. C. A., PETTMAN, A. J. & KNOWLES, C. J. 2000. Real-time monitoring of nitrile biotransformations by mid-infrared spectroscopy. *Journal of Microbiological Methods*, 41, 69-75.
- DIETZEL, U., KUPER, J., DOEBBLER, J. A., SCHULTE, A., TRUGLIO, J. J., LEIMKÜHLER, S. & KISKER, C. 2009. Mechanism of Substrate and Inhibitor Binding of Rhodobacter capsulatus Xanthine Dehydrogenase. *The Journal of Biological Chemistry*, 284, 8768-8776.
- ELLIS, D. I., MUHAMADALI, H., HAUGHEY, S. A., ELLIOTT, C. T. & GOODACRE, R. 2015. Point-and-shoot: rapid quantitative detection methods for on-site food fraud analysis—moving out of the laboratory and into the food supply chain. *Analytical Methods*, 7, 9401-9414.
- FOX, S., FARR-JONES, S., SOPCHAK, L., BOGGS, A. & COMLEY, J. 2004. High-Throughput Screening: Searching for Higher Productivity. *Journal of Biomolecular Screening*, 9, 354-358.
- GARDNER, P., ARNOLD, S. A., BROWN, F., CARR, R., NORDON, A., HARVEY, L. M. & MCNEIL, B. 2013. Investigating the feasibility of mid infrared spectroscopy for monitoring an industrial de-racemization biotransformation process. *Analytica Chimica Acta*, 779, 50-55.
- GHISALBA, O., MEYER, H.-P., WOHLGEMUTH, R. & FLICKINGER, M. C. 2009. Industrial Biotransformation. *Encyclopedia of Industrial Biotechnology*. John Wiley & Sons, Inc.
- HARRISON, R. 2002. Structure and function of xanthine oxidoreductase: where are we now? *Free Radicals in Biology and Medicine*, 33, 774-797.
- HILLE, R. 1996. The Mononuclear Molybdenum Enzymes. *Chemical Reviews*, 96, 2757-2816.
- HILLE, R. 2005. Molybdenum-containing hydroxylases. *Archives of Biochemistry and Biophysics*, 433, 107-116.
- HINZ, D. C. 2006. Process analytical technologies in the pharmaceutical industry: the FDA's PAT initiative. *Analytical and Bioanalytical Chemistry*, 384, 1036-1042.
- JAUMOT, J., GARGALLO, R., DE JUAN, A. & TAULER, R. 2005. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemometrics and Intelligent Laboratory Systems*, 76, 101-110.
- KOBAYASHI, M., NAGASAWA, T. & YAMADA, H. 1992. Enzymatic synthesis of acrylamide: a success story not yet over. *Trends in Biotechnology*, 10, 402-408.
- KOURTI, T. 2006. The Process Analytical Technology initiative and multivariate process analysis, monitoring and control. *Analytical and Bioanalytical Chemistry*, 384, 1043-1048.
- KUMAR, R. A. & CLARK, D. S. 2006. High-throughput screening of biocatalytic activity: applications in drug discovery. *Current Opinion in Chemical Biology*, 10, 162-168.

- LIN, Z. J., ZHENG, R.-C., LEI, L.-H., ZHENG, Y.-G. & SHEN, Y.-C. 2011. Ferrous and ferric ions-based high-throughput screening strategy for nitrile hydratase and amidase. *Journal of Microbiological Methods*, 85, 214-220.
- MAUGER, J., NAGASAWA, T. & YAMADA, H. 1988. Nitrile hydratase-catalyzed production of isonicotinamide, picolinamide and pyrazinamide from 4-cyanopyridine, 2-cyanopyridine and cyanopyrazine in *Rhodococcus rhodochrous* J1. *Journal of Biotechnology*, 8, 87-95.
- MOORE, B. D., STEVENSON, L., WATT, A., FLITSCH, S., TURNER, N. J., CASSIDY, C. & GRAHAM, D. 2004. Rapid and ultra-sensitive determination of enzyme activities using surface-enhanced resonance Raman scattering. *Nature Biotechnology*, 22, 1133-1138.
- NAGASAWA, T., MATHEW, C. D., MAUGER, J. & YAMADA, H. 1988. Nitrile Hydratase-Catalyzed Production of Nicotinamide from 3-Cyanopyridine in *Rhodococcus rhodochrous* J1. *Applied and Environmental Microbiology*, 54, 1766-1769.
- PACKER, M. S. & LIU, D. R. 2015. Methods for the directed evolution of proteins. *Nature Reviews Genetics*, 16, 379-394.
- REETZ, M. T. 2013. Biocatalysis in Organic Chemistry and Biotechnology: Past, Present, and Future. *Journal of the American Chemical Society*, 135, 12480-12496.
- REISINGER, C., VAN ASSEMA, F., SCHÜRMAN, M., HUSSAIN, Z., REMLER, P. & SCHWAB, H. 2006. A versatile colony assay based on NADH fluorescence. *Journal of Molecular Catalysis B: Enzymatic*, 39, 149-155.
- REYMOND, J. L. & BABIAK, P. 2007. Screening systems. *Advances in Biochemical Engineering/Biotechnology*, 105, 31-58.
- SHAW, A. D., KADERBHAI, N., JONES, A., WOODWARD, A. M., GOODACRE, R., ROWLAND, J. J. & KELL, D. B. 1999. Noninvasive, On-Line Monitoring of the Biotransformation by Yeast of Glucose to Ethanol Using Dispersive Raman Spectroscopy and Chemometrics. *Applied Spectroscopy*, 53, 1419-1428.
- STREKAS, T. C. & SPIRO, T. G. 1972. Cytochrome c: Resonance Raman spectra. *Biochimica et Biophysica Acta, Protein Structure*, 278, 188-192.
- TAULER, R. 1995. Multivariate curve resolution applied to second order data. *Chemometrics and Intelligent Laboratory Systems*, 30, 133-146.
- VANKEIRSBILCK, T., VERCAUTEREN, A., BAEYENS, W., VAN DER WEKEN, G., VERPOORT, F., VERGOTE, G. & REMON, J. P. 2002. Applications of Raman spectroscopy in pharmaceutical analysis. *Trends in Analytical Chemistry*, 21, 869-877.
- VON AHSEN, O. & BÖMER, U. 2005. High-Throughput Screening for Kinase Inhibitors. *ChemBioChem*, 6, 481-490.
- WESTLEY, C., XU, Y., CARNELL, A. J., TURNER, N. J. & GOODACRE, R. 2016. Label-Free Surface Enhanced Raman Scattering Approach for High-Throughput Screening of Biocatalysts. *Analytical Chemistry*, 88, 5898-5903.
- WORKMAN, J., KOCH, M., LAVINE, B. & CHRISMAN, R. 2009. Process Analytical Chemistry. *Analytical Chemistry*, 81, 4623-4643.
- WORKMAN, J., KOCH, M. & VELTKAMP, D. 2007. Process Analytical Chemistry. *Analytical Chemistry*, 79, 4345-4364.
- YAN, C., PARMEGGIANI, F., JONES, E. A., CLAUDE, E., HUSSAIN, S. A., TURNER, N. J., FLITSCH, S. L. & BARRAN, P. E. 2017. Real-Time Screening of Biocatalysts in Live Bacterial Colonies. *Journal of the American Chemical Society*, 139, 1408-1411.

5.8 SUPPORTING INFORMATION

MATERIAL AND METHODS

Reagents and Materials.

All chemical reagents were of analytical grade and used with no additional purification unless otherwise stated. Acetic acid (HPLC), acetonitrile (HPLC), benzonitrile ($\geq 99\%$), benzamide ($\geq 99\%$), benzoic acid ($\geq 99\%$), butanoic acid ($\geq 99\%$), hypoxanthine ($\geq 99\%$), methanol (HPLC), nicotinamide ($\geq 98.5\%$), nicotinic acid ($\geq 99.5\%$), 3-pyridinecarbonitrile ($\geq 98\%$), pyrazinecarbonitrile ($\geq 99\%$), *p*-tolunitrile ($\geq 98\%$), *p*-toluamide ($\geq 99\%$), sodium acetate trihydrate ($\geq 99\%$), uric acid ($\geq 99\%$), water (HPLC), xanthine ($\geq 99\%$), xanthine oxidase microbial lyophilized powder (EC 1.17.3.2) were purchased from Sigma Aldrich Ltd. (Dorset, UK). Potassium dihydrogen phosphate, dipotassium phosphate, pyrazinamide ($\geq 99\%$), sodium carbonate, and sodium hydroxide were obtained from Fischer Scientific (Loughborough, UK). Nitrile hydratase was purchased from Prozomix (PRO-NHASE (018), LOT 2013-1, EC 4.2.1.84) (Northumberland, UK).

Reaction conditions for biotransformations 1 and 2.

A stock solution of nitrile hydratase (as sold by Prozomix) was prepared as follows: 250 μL nitrile hydratase solution was pelleted by centrifugation (1 min, 21,000 *g*) and the supernatant removed. The pellet was then re-suspended in 2 mL of TRIS buffer (4.0×10^{-2} M) and butanoic acid (4.0×10^{-2} M) at pH 7.2 before use (stock concentration of enzyme not known). Starting reaction mixture contained either benzonitrile (biotransformation 1) or *p*-tolunitrile (biotransformation 4) dissolved in potassium phosphate buffer at pH 7.2 (final concentration 1.25×10^{-2} M with 2.5 % MeOH) and nitrile hydratase solution (40 μL for biotransformation 1, 200 μL for biotransformation 2).

Reaction conditions for biotransformations 3 and 4.

A stock solution of xanthine oxidase was prepared by dissolving 7 mg of xanthine oxidase microbial lyophilized powder in 1.1 mL potassium phosphate buffer at pH 7.6 (stock concentration 2.25×10^{-5} M). Starting reaction mixture contained either xanthine (biotransformation 3) or hypoxanthine (biotransformation 4) dissolved in water (final concentration 7.5×10^{-4} M), potassium phosphate buffer at pH 7.6 (final concentration 0.3 M) and xanthine oxidase (120 μL for biotransformation 2, 150 μL for biotransformation 3).

Reaction sample preparation and monitoring.

For all biotransformations, the reaction mixture was focused under the microscope objective in a small, glass Petri dish (see Figure S5.1). At various time points after enzyme addition, UVRR and HPLC spectra were collected throughout the biotransformation. To minimise the risk of reduced focus on the sample through solvent evaporation and removal of volume (for HPLC analysis), the reaction was performed on a 10 mL scale.

For biotransformation 1 and 2, at specific time points, 20 μL of sample was removed from the reaction mixture and immediately quenched and diluted with 180 μL with MeOH. The sample was then centrifuged at 21,000 $\times g$ for 6 min. 100 μL of the sample was then transferred to a HPLC vial and subjected to HPLC analysis.

For biotransformations 3 and 4, at specific time points, 20 μL of sample was removed from the reaction mixture and immediately diluted to 80 μL with water. The sample was then heated for 5 min at 80 $^{\circ}\text{C}$ and centrifuged at 14000 $\times g$ for 4 min. 60 μL of the sample was then transferred to a HPLC vial and subjected to HPLC analysis.

INSTRUMENTATION

UV-Vis Spectrophotometry.

UV-Vis absorption analysis was carried out using Thermo Biomate 5 (Thermo Fisher Scientific Inc., Massachusetts, USA). 1 mL of the sample was pipetted into a quartz cuvette and inserted into a sample holder. Data were acquired over a wavelength range of 210 – 350 nm.

HPLC.

Biotransformations 1 and 2. HPLC separation was conducted using an Agilent Zorbax Eclipse Plus HPLC system set up for reverse phase separation consisting of a diode array detector. For both biotransformations, the column was a 100 \times 4.6 mm, Phenomenex Eclipse Plus[®] C18 with a 3.5 μm particle size. For each injection, the run time was 12.0 min pumped at a flow rate of 1mL/min and at 30 $^{\circ}\text{C}$ column temperature. The mobile phase consisted of a linear gradient, starting conditions of 5% MeCN/H₂O (plus 0.05 % TFA) held for 2 min before increasing to 75 % MeCN/H₂O over 6 min. Prior to washing at 95 % MeCN/H₂O for 1.5 min and re-equilibration to initial conditions over 2.5 min (total run time 12 min). 5 μL of each sample was introduced using an auto-injector. UV absorbance was detected at 254 nm throughout.

Biotransformations 3 and 4. HPLC separation was conducted using an Agilent 1100 series HPLC system set up for reverse phase separation consisting of a diode array detector. For the biotransformation 3, the column was 150 \times 4.6 mm, Phenomenex Hyperclone C18 with a 5 μm

particle size. For each injection, the run time was 10.0 min. The mobile phase was 2×10^{-2} M aqueous potassium phosphate buffer at pH 6.5, pumped at a flow rate of 1.0 mL min^{-1} . For biotransformation 4, the column was a 250×4.6 mm, ACE 5 C18-AR (Reading, Berkshire) with a $5 \mu\text{m}$ particle size. For each injection, the run time was 12.0 min. The mobile phase was 0.047 M aqueous acetic acid buffer at pH 4.65, pumped at a flow rate of 1.0 mL min^{-1} . For both biotransformations, $50 \mu\text{L}$ of each sample was introduced using an auto-injector. UV absorbance detection was measured at 254 nm.

UVRR.

UVRR analysis for all biotransformations. UVRR was performed using a Renishaw Raman 1000 system (Renishaw, Wotton-under-edge, Gloucestershire, UK). Approximately 0.2 mW of power was delivered to the sampling point using a Lexel Model 95 ion laser emitting at 244 nm. The solution was continuously stirred under the laser to avoid photodegradation using a magnetic stirrer plate and magnetic bar. Spectra were collected with an acquisition time of 20 s. Only spectra with no demonstrable photodegradation and signal from the reaction vessel were used for analysis.

DATA PROCESSING

All data were exported from the respective instrument operating software and analysed using Matlab R2015a (The Mathworks, Natick, MA, USA).

HPLC data analysis

The peaks of the target analytes were integrated with the results of the HPLC data and served as an external validation data set to independently verify the accuracies of the MCR-ALS models in prediction.

UVRR data analysis

Multivariate curve resolution - alternating least squares (MCR-ALS) was employed due to the fact that the UVRR spectra of all three analytes are highly similar and do not possess characteristic peaks (Tauler, 1995, Jaumot *et al.*, 2005). In MCR-ALS, the UVRR spectra were first baseline corrected, smoothed using wavelet smoothing and then normalised so that the sum of squares of each spectrum equals 1. This is to account for the decreasing signal strength and relatively increasing noise (attributed to the removal of aliquots for HPLC analysis (And specifically for biotransformation 1 and 4 the inherent volatility of the solution)).

The spectra of all the monitored time points of a single reaction were then augmented to form a $t \times n$ data matrix \mathbf{X} where t is the number of time points monitored and n is the number of wavenumbers recorded. Multivariate curve resolution using alternating least squares (MCR-ALS) was performed to deconvolve \mathbf{X} into a product of two sub matrices \mathbf{C} and \mathbf{S} where \mathbf{C} contains the “profiles” of the change in the concentrations the reactants during the reaction while \mathbf{S} is the matrix storing the resolved spectra the reactants. Non-negativity constraint was applied to both concentration profile \mathbf{C} and spectral profile \mathbf{S} and each deconvolved pure spectrum had a unit norm (*i.e.* the sum of squares of each spectrum equals 1).

For the time points when HPLC measures were also taken, a linear regression model was built between the concentration profile of each reactant in \mathbf{C} and the corresponding concentration measured by HPLC. The regression model was then applied to the whole concentration profile to get UVRR calibrated concentrations of the reactant over the whole monitored period of the reaction.

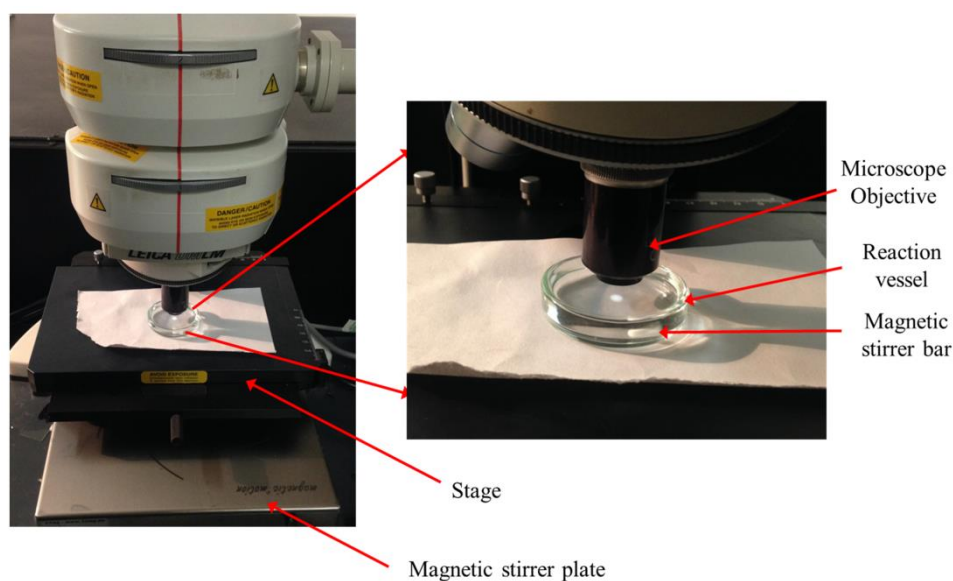


Figure S5.1. Annotated instrument set-up to monitor biotransformation using UV resonance Raman spectroscopy. The x 40 UVRR objective was carefully focused onto the reaction mixture (10 mL scale), with 100 % power on sample (~ 0.2 mW at sampling point). Throughout the course of the reaction, samples for HPLC analysis were removed as well as UVRR data collected (20 s spectral acquisitions). The solution was constantly stirred using the magnetic stirrer plate beneath the stage and the magnetic stirrer bar in the reaction vessel.

Photo-Degradation of Sample

We initially tried to photodegrade our sample so we were aware of the spectral changes to expect if the sample started to degrade or 'burn'. Using biotransformation 3 as an example, we tried to photo-degrade the starting material, xanthine (in solution). After 30 min of constant interrogation of the laser on the sample, there were no changes in the spectra and thus no photodamage. Our only observation was the evaporation of solvent meaning as time increased, the sample point became out of focus and hence the spectra became noisier. We next looked at photo-degrading the corresponding solid sample, and after 45 min (vastly exceeding the total reaction time) we noticed broadening of peaks around 1550 cm^{-1} region (from C-C), similar to spectra of graphitic carbon-type species and charcoal, thus indicating 'burning' of the sample. These observations are in agreement with the literature (Schwan *et al.*, 1996, Tallant *et al.*, 1997, Ishimaru *et al.*, 2007, Wang *et al.*, 2011).

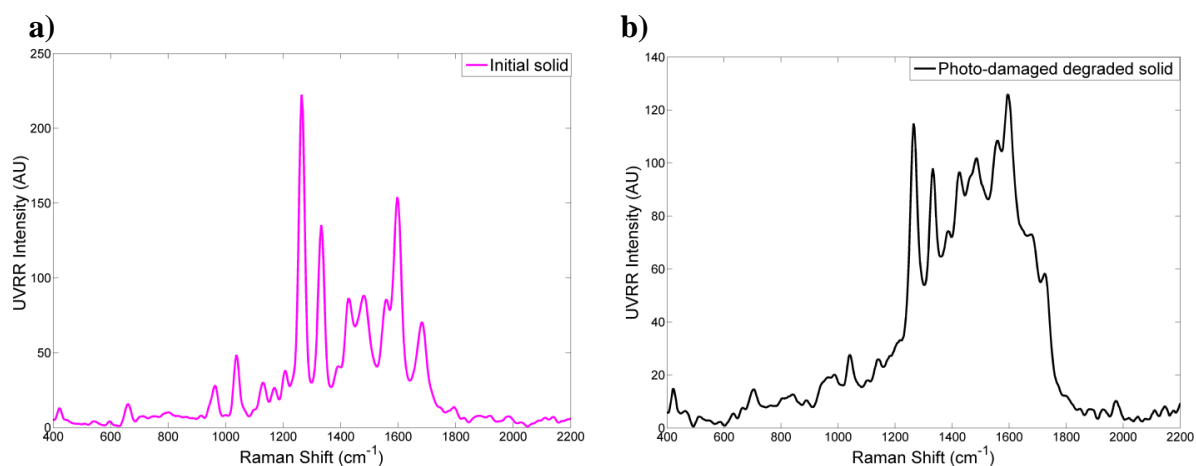


Figure S5.2. Average UVRR spectra ($n = 5$) of **a)** initial solid xanthine sample and **b)** degraded solid xanthine sample after 45 min (after each min, a 20 s spectral acquisition was acquired). A broadening of peaks around 1550 cm^{-1} indicates C-C presence suggesting photo-degradation.

Table S5.1. Tentative UVRR band assignments for nitrile hydratase catalysed biotransformations at pH 7.2: benzonitrile and benzamide (biotransformation 1) and *p*-tolunitrile and *p*-toluamide (biotransformation 2) *(Chatterjee *et al.*, 1978, Gao *et al.*, 1990, Mrozek *et al.*, 2001, Fleming *et al.*, 2008, Brittain, 2009).

Raman Shift (cm ⁻¹)				Tentative band assignment*
Benzonitrile	Benzamide	<i>p</i> -Tolunitrile	<i>p</i> -Toluamide	
		825 (w)	831 (w)	Ring breathing
1008 (m)	1008 (w)			C-C-C trigonal breathing
	1145 (vw)		1149 (m)	NH ₂ rocking mode
1185 (m)	1192 (w)	1185 (s)	1200 (m)	C-H in-plane bend
	1415 (m)		1426 (m)	C-N stretch
1601 (vs)	1608 (vs)	1612 (vs)	1620 (vs)	C-C in-plane stretch
2240 (s)		2239 (s)		C≡N stretch

vs-very strong, s-strong, m-medium, w-weak, vw-very weak, br-broad, sh-shoulder

Table S5.2. Tentative UVRR band assignments for xanthine oxidase catalysed biotransformations at pH 7.6: xanthine to uric acid (biotransformation 3) and hypoxanthine to xanthine to uric acid (biotransformation 4) *(Kodati *et al.*, 1990, Chowdhury *et al.*, 2000, Krishnakumar and Arivazhagan, 2004, Arivazhagan and Jeyavijayan, 2010, 2011, Goodall *et al.*, 2013).

Raman Shift (cm ⁻¹)			Tentative band assignment*
Hypoxanthine	Xanthine	Uric Acid	
1601 (s)	1597 (vs)	1612 (s)	C-N
		1540 (wsh)	-
1468 (vs)	1472 (vs)		C-N or C-C
1347 (m)	1410 (w)	1431 (m)	C-O
	1336 (s)	1359 (sh)	-
1291 (s)	1283 (s)		C-N, C-H bend
1223 (w)			Mixed ring vibrations, C-N
	1041 (w)	1052 (s)	Ring vibrations
995 (m)	995 (w)	995 (m)	N-H bend, ring trigonal deformation

vs-very strong, s-strong, m-medium, w-weak, vw-very weak, br-broad, sh-shoulder

Bathochromic shifts of XO analytes

Photo-degrading the sample led to an interesting observation regarding the UVRR spectra of xanthine. Notably, there were significant changes in spectral band positions and intensity between the solid spectra from xanthine and that in solution. Changes in environmental conditions (*e.g.* temperature, solvent, pH *etc.*) can lead to changes in some vibrational frequencies especially as for some molecules these functional groups will interact differently with the solvent (through H-bonding, as well as acidic/basic ions). This led to a pH investigation looking at the associated UV-Vis absorption of xanthine and UVRR spectra, with the results indicating the observation of a bathochromic shift on increasing pH. In acidic medium (*i.e.* low pH), the nitrogen atom is free to lose its lone pair of electrons thus decreasing the delocalisation in the ring, leading to a decrease in conjugation, as observed in Figure S5.3, consequently, the compound becomes less energetic. As the energy needed for excitation is higher, there is a shift to absorbing at shorter wavelengths. Conversely, in alkaline medium (*i.e.* high pH), the opposite phenomenon occurs, and the oxygens lone pairs increase delocalisation and conjugation to the ring, (as observed in Figure S5.3) meaning the compound becomes more energetic, so less energy is required for excitation thus shifting to absorbing at longer wavelengths. This is supported by the increase in wavelength absorbance from λ_{\max} 267 to 272 nm when going from pH 3.6 to 9.2. Figure S5.3c and S5.3d) highlight key peaks that are affected by changing the pH of the solution. A similar, less pronounced effect was observed for hypoxanthine and uric acid (see Figure S5.4 and S5.5) (Bergmann and Dikstein, 1955, Ascher *et al.*, 2014).

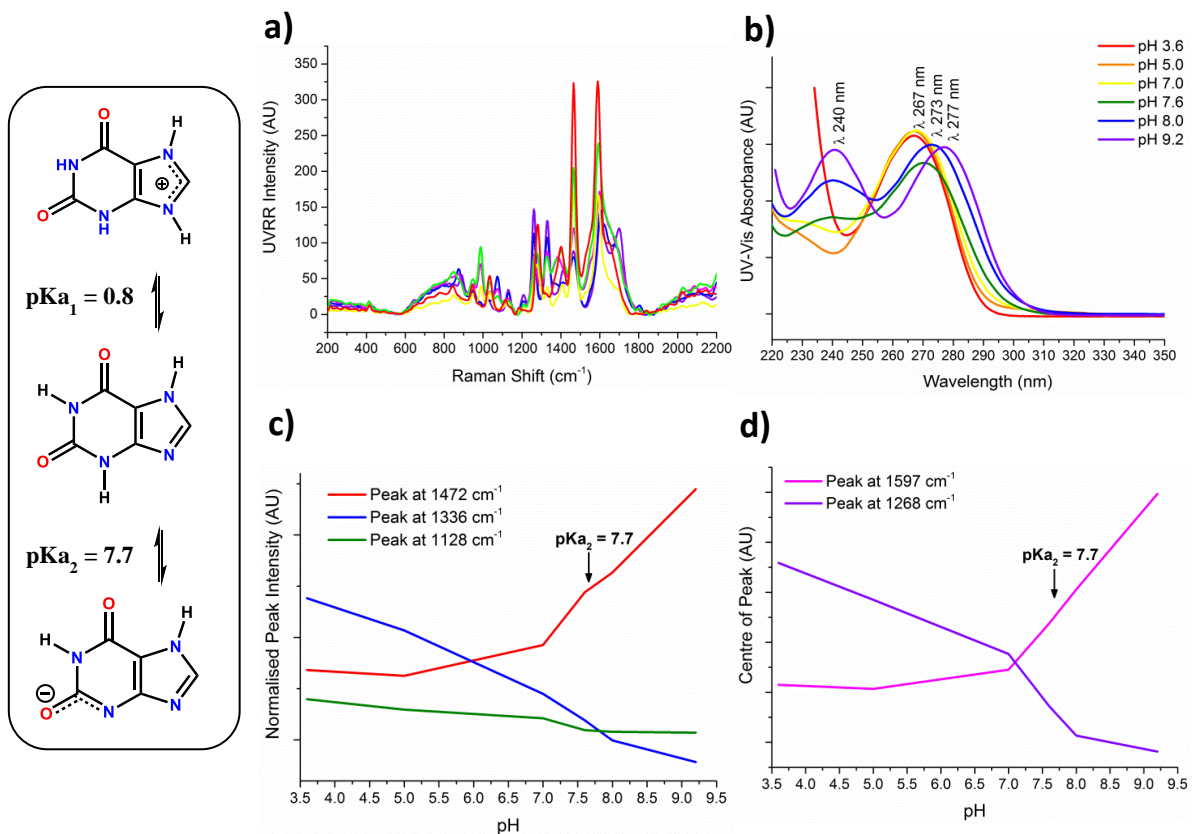


Figure S5.3. (a) Average UVRR spectra ($n = 5$) and (b) UV-Vis absorbance spectra of xanthine at various pH: 3.6, 5.0, 7.0, 7.6, 8.0, and 9.2. A bathochromic shift in the UV-Vis absorption spectra was observed on increasing the pH from pH 3.6 to 9.2, consequently, the UVRR spectra of xanthine changed due to it being in different ionisation states. (c) The intensity difference of the key peaks that change and (b) a plot of the centre of the peaks that shift on increasing the pH of the solution from pH 3.6 to 9.2.

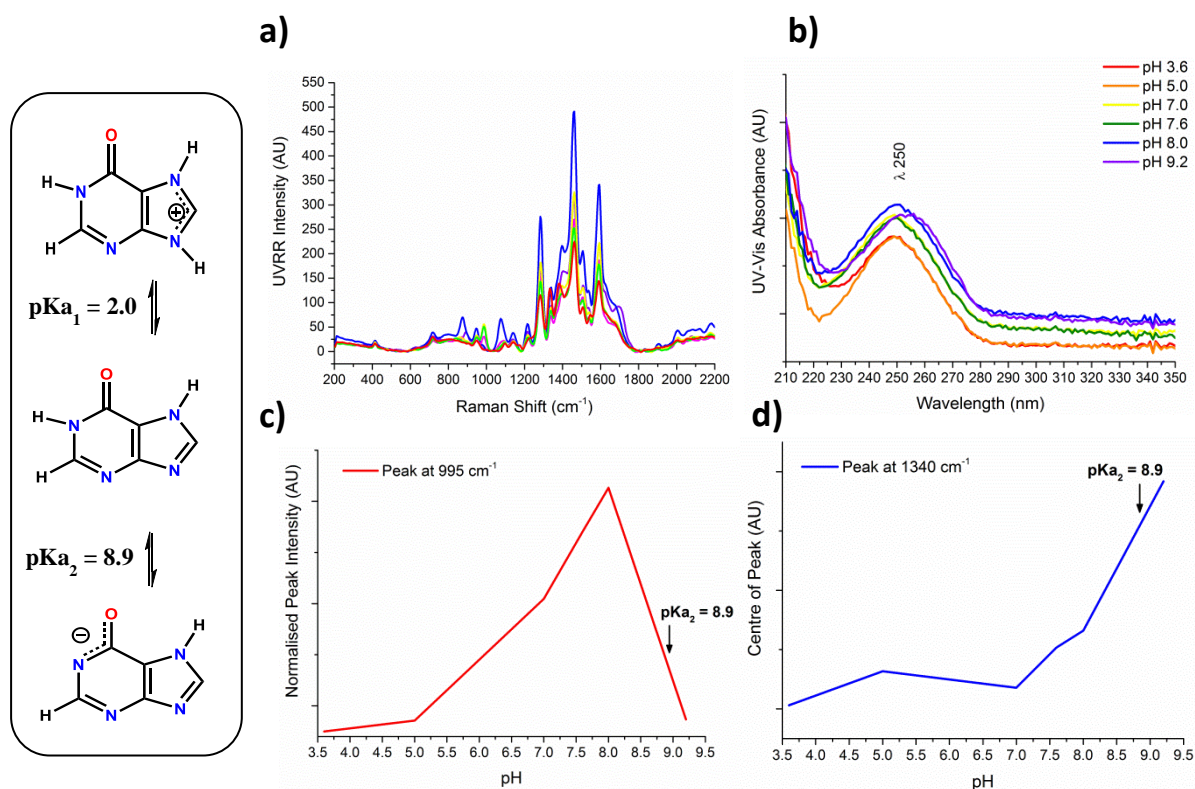


Figure S5.4. (a) Average UVRR spectra ($n = 5$) and (b) UV-Vis absorbance spectra of hypoxanthine at various pH: 3.6, 5.0, 7.0, 7.6, 8.0, and 9.2. A bathochromic shift in the UV-Vis absorption spectra was observed on increasing the pH from pH 3.6 to 9.2, consequently, the UVRR spectra of hypoxanthine changed due to being in different ionisation states. (c) The intensity difference of the key peak that changes and (d) a plot of the centre of peak that shifts on increasing the pH of the solution from pH 3.6 to 9.2.

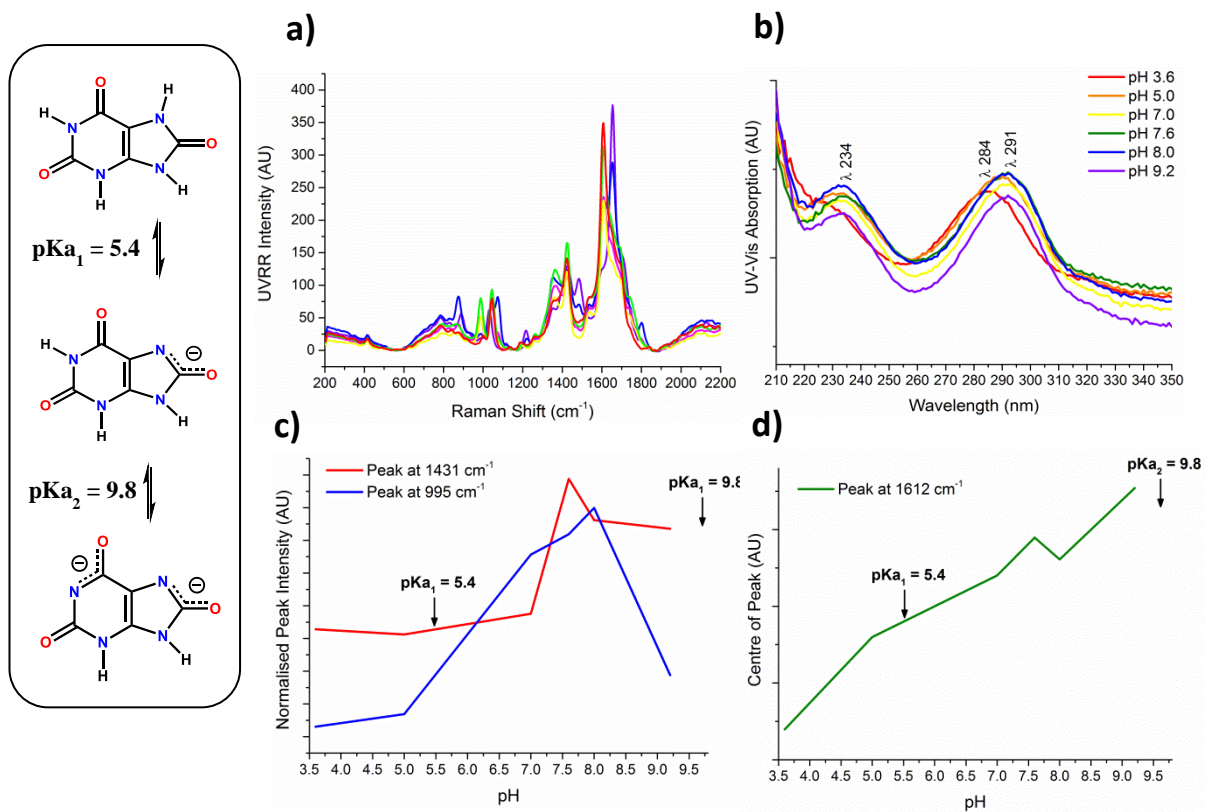


Figure S5.5. (a) Average UVRR spectra ($n = 5$) and (b) UV-Vis absorbance spectra of uric acid at various pH: 3.6, 5.0, 7.0, 7.6, 8.0, and 9.2. A bathochromic shift in the UV-Vis absorption spectra was observed on increasing the pH from pH 3.6 to 9.2, consequently, the UVRR spectra of uric acid changed due to being in different ionisation states. (c) The intensity difference of the key peaks that change and (d) a plot of the centre of peak that shifts on increasing the pH of the solution from pH 3.6 to 9.2.

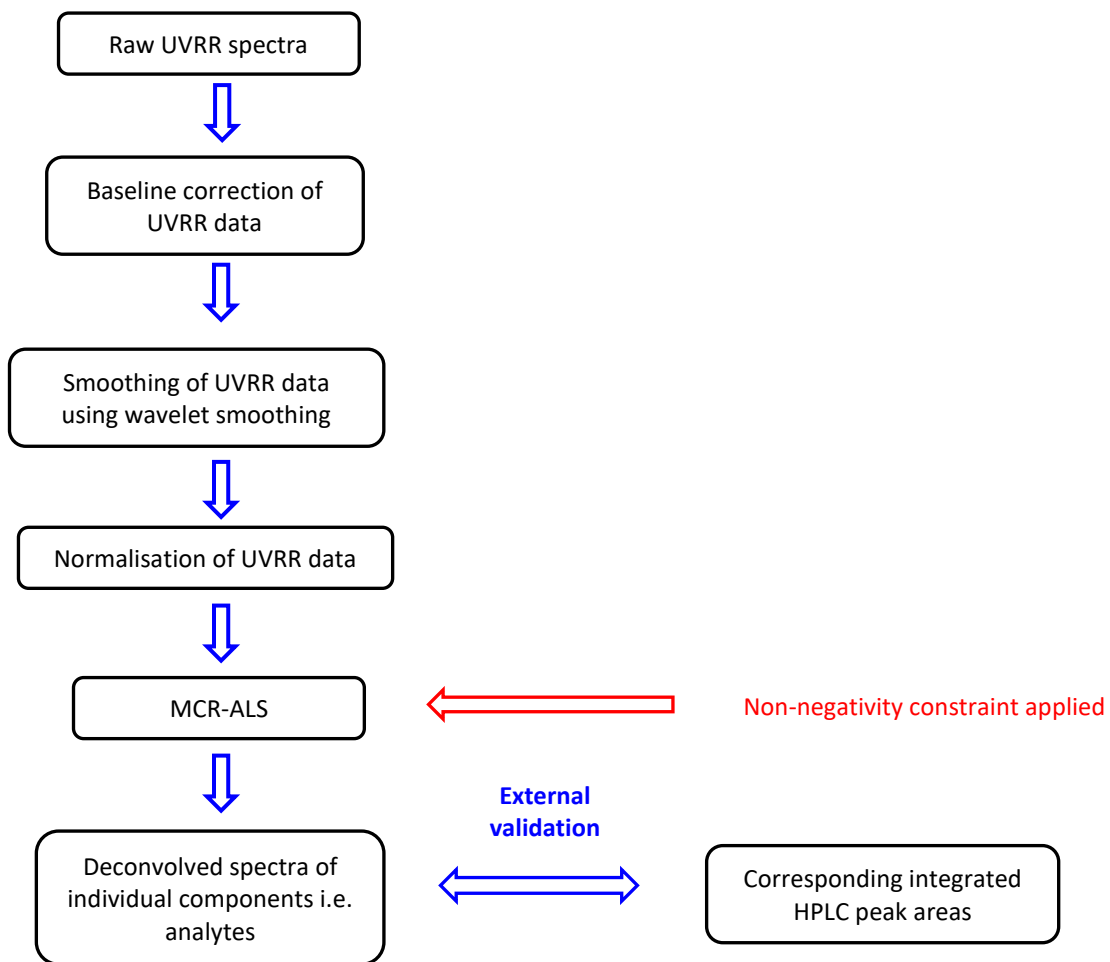


Figure S5.6. Flow diagram summarising the data pre-processing and MCR-ALS process used for in predicting concentrations of each analyte from the reaction

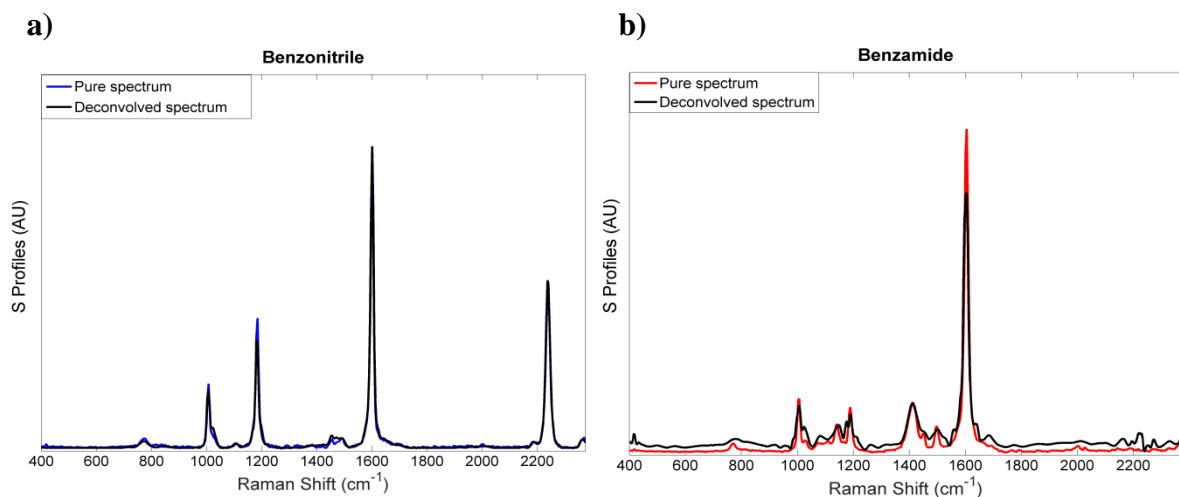


Figure S5.7. An MCR-ALS model was applied to the UVR data where it successfully deconvolved spectra into its pure components **a)** benzonitrile (substrate) and **b)** benzamide (product) as shown for biotransformation 1.

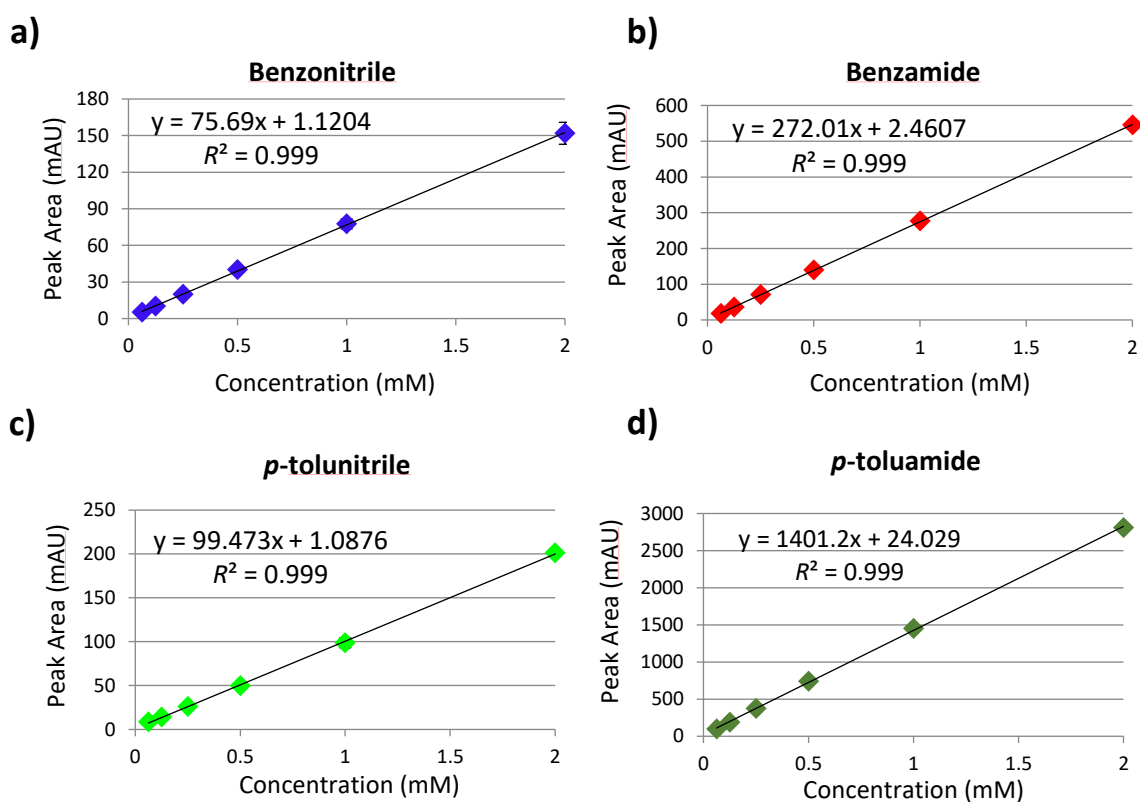


Figure S5.8. HPLC calibrations for **(a)** benzonitrile **(b)** benzamide **(c)** *p*-tolunitrile and **(d)** *p*-toluamide at 254 nm absorbance. Plots show the mean plus associated standard deviation (SD) error bars from triplicate data.

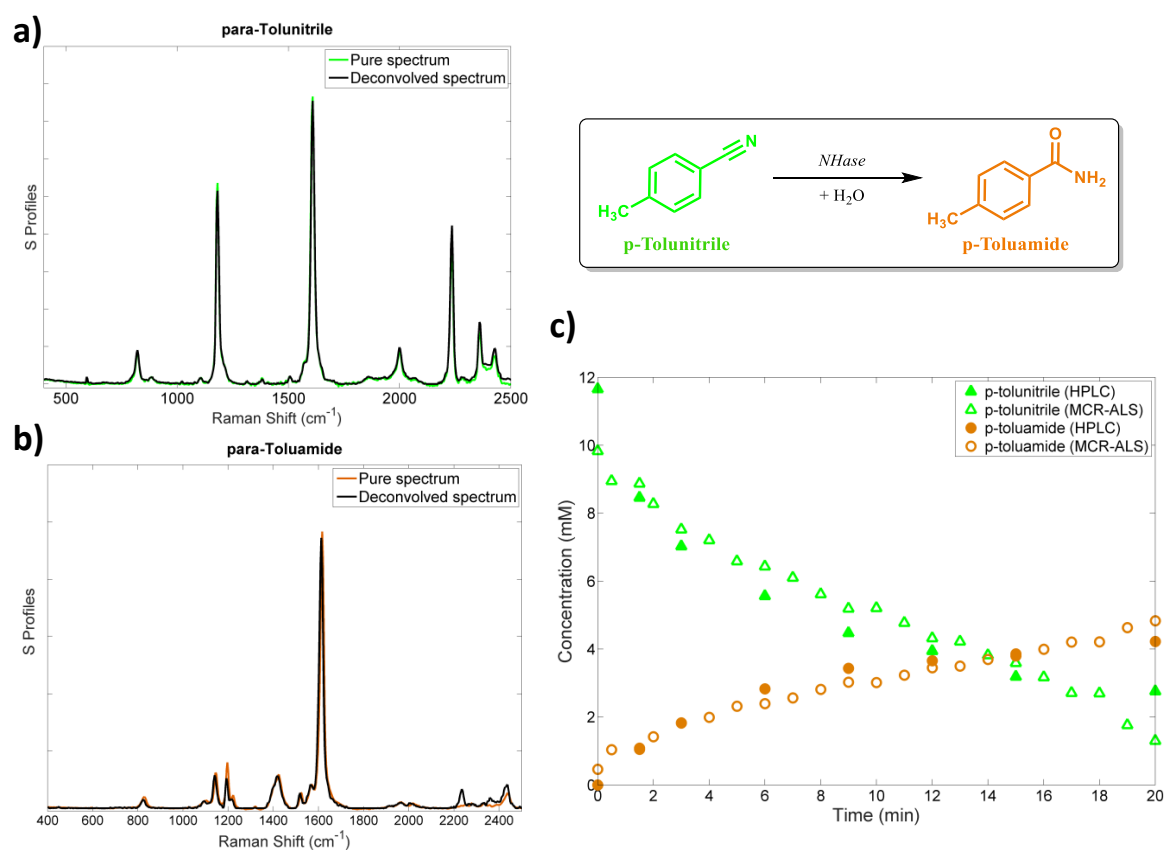


Figure S5.9. A MCR-ALS model was applied to the UVRR data where it successfully deconvolved spectra into its pure components **(a)** *p*-tolunitrile (green) and **(b)** *p*-toluamide (orange) as shown for biotransformation 2 **(c)** Shows the reaction dynamics from real-time UVRR measurements (denoted by outlined symbols) and off-line HPLC data (denoted by solid symbols) as a function of time for the conversion of *p*-tolunitrile to *p*-toluamide. UVRR spectra were obtained for 20 s with baseline correction, normalisation and smoothing applied (as detailed in ‘Materials and methods: data processing’).

Table S5.3. The R^2 co-efficients obtained for three replicates for the conversion of *p*-tolunitrile to *p*-toluamide (biotransformation 2).

Replicate	<i>p</i> -Tolunitrile R^2	<i>p</i> -Toluamide R^2
1	0.892	0.860
2	0.919	0.943
3	0.884	0.938

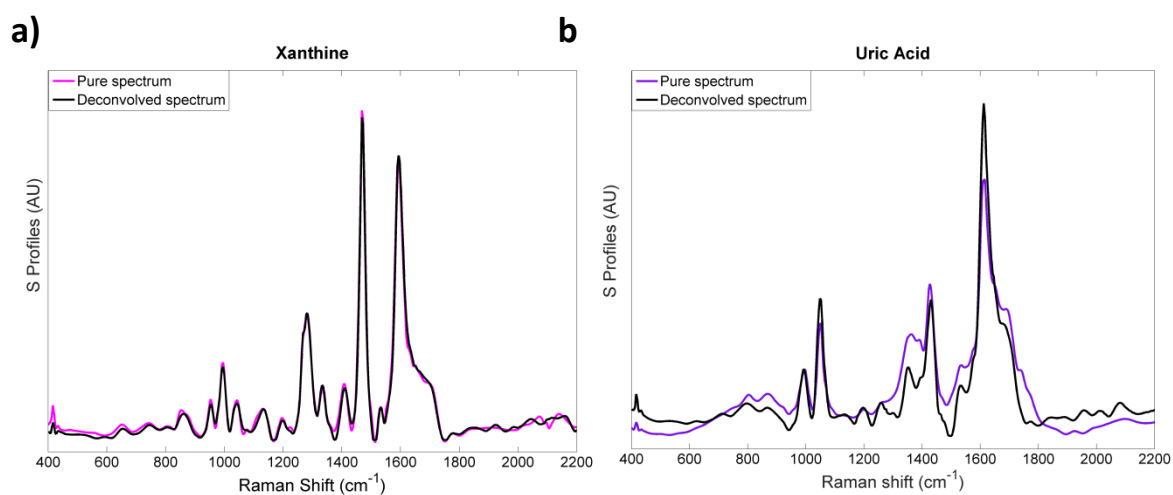


Figure S5.10. An MCR-ALS model was applied to the UVRR data where it successfully deconvolved spectra into its pure components (a) xanthine (substrate) and (b) uric acid (product) as shown for biotransformation 1.

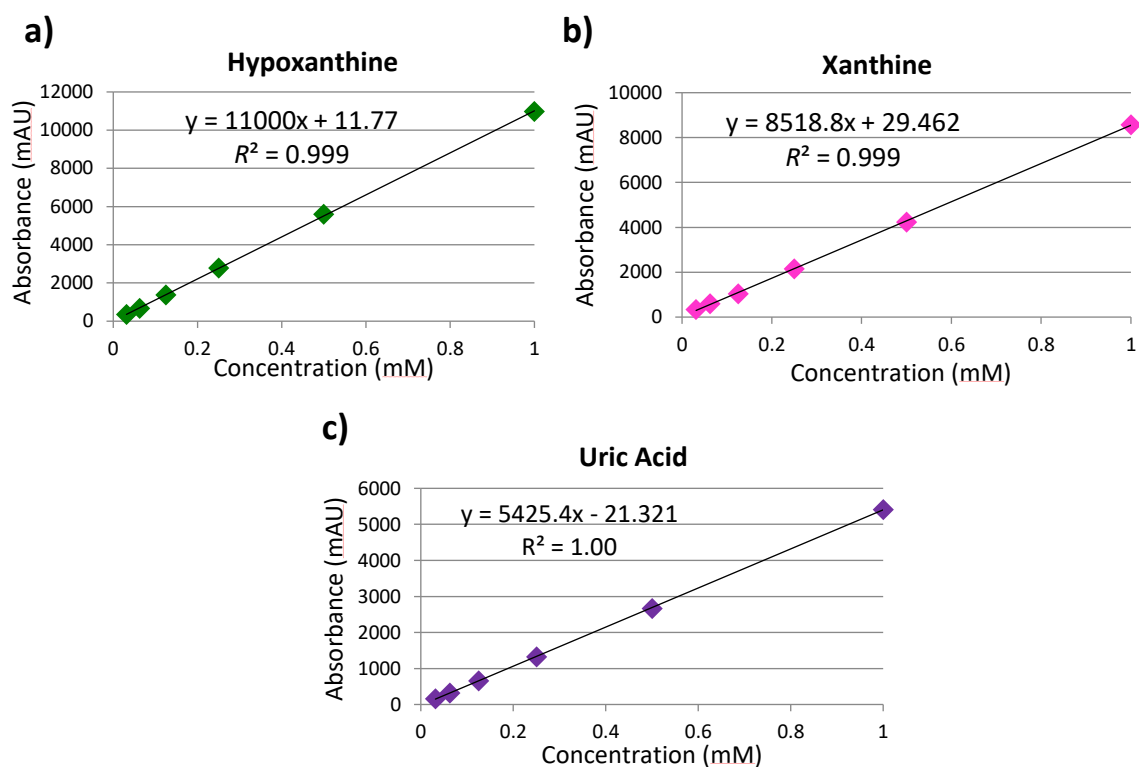


Figure S5.11. HPLC calibrations for a) hypoxanthine b) xanthine and c) uric acid at 254 nm absorbance. Plots show the mean plus associated standard deviation (SD) error bars from triplicate data.

Table S5.4. A summary of the R^2 co-efficients across all three replicates for the conversion of hypoxanthine to xanthine to uric acid (biotransformation 4).

Replicate	Hypoxanthine R^2	Xanthine R^2	Uric Acid R^2
1	0.904	0.171	0.935
2	0.736	0.001	0.928
3	0.628	0.245	0.765

XO biotransformation 4: supporting proposed mechanism

Noticeably, the R^2 value of xanthine was low for the three-analyte biotransformation (biotransformation 4) – this is due to the low overall concentration of xanthine (<8 %) throughout the reaction. This is further supported by the proposed mechanism of XO (based on xanthine dehydrogenase, *XDH*, from *Rhodobacter capsulatus*) whereby hypoxanthine binds to the active site and is converted to xanthine by oxidation at the C-2 position. Xanthine is then released, before binding in a different orientation to present the C-8 for oxidation to give uric acid (Dietzel *et al.*, 2009). This means that the concentration of the intermediate remains low throughout.

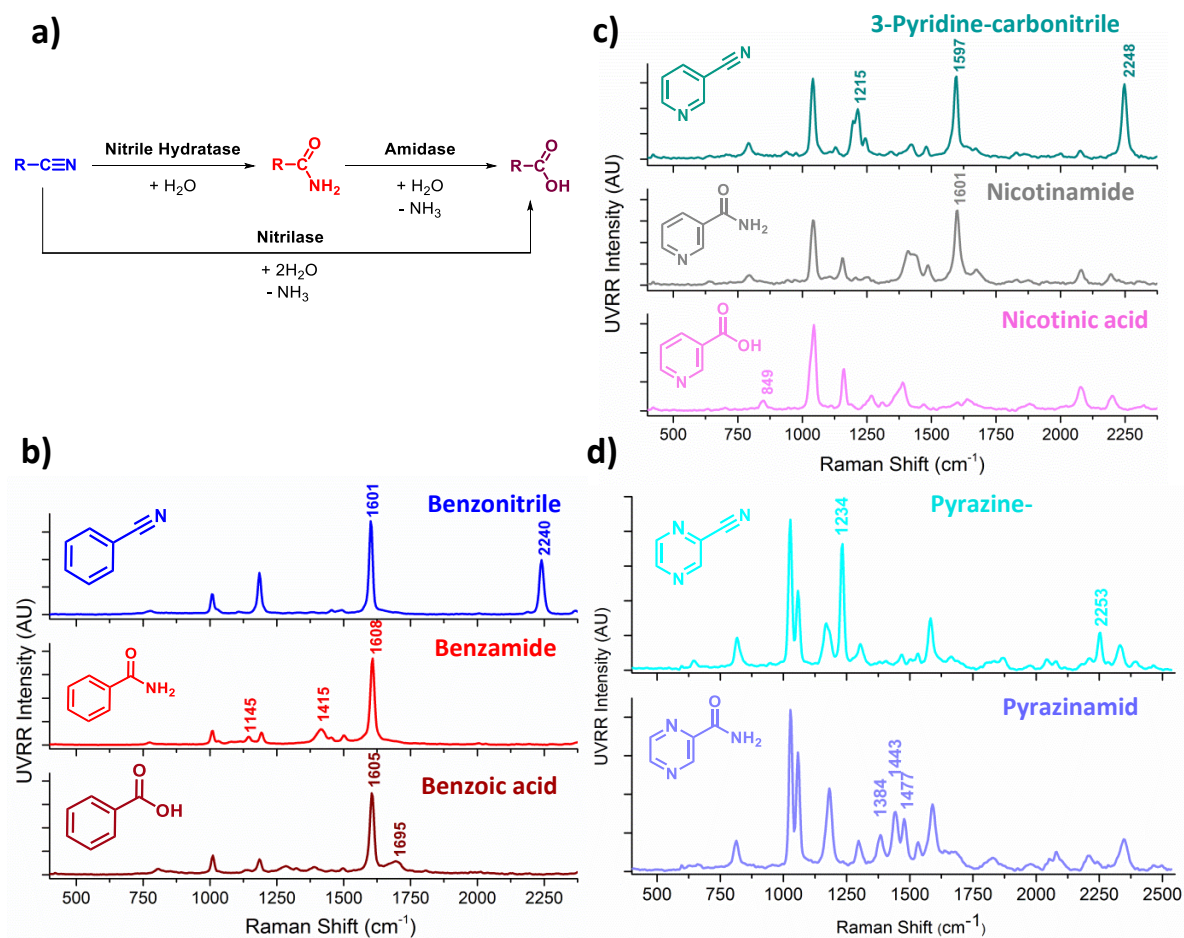


Figure S5.12. (a) Overall schematic illustrating the two known pathways to catalyse the conversion of nitrile containing compounds into their corresponding carboxylic acid, either in a single step (nitrilase) or a multicomponent process (nitrile hydratase and amidase) Plots b-d) show average UVRR spectra ($n = 5$) of each analyte: (b) pure spectra of benzonitrile, benzamide and benzoic acid (1.25×10^{-2} M, pH 7.2) (c) pure spectra of 3-pyridinecarbonitrile, nicotinamide (vitamin B₃) and nicotinic acid (2.5×10^{-2} M, pH 7.2) and (d) pure spectra of pyrazinecarbonitrile and pyrazinamide (anti-tuberculosis drug) (2.5×10^{-2} M, pH 7.2). Characteristic peaks are annotated. UVRR spectra were obtained for 20 s with baseline correction, normalisation and smoothing applied (as detailed in ‘Materials and methods: data processing’).

REFERENCES

- ARIVAZHAGAN, M. & JEYAVIJAYAN, S. 2010. Study of density functional theory and vibrational spectra of hypoxanthine. *Indian Journal of Pure & Applied Physics*, 48, 869-874.
- ARIVAZHAGAN, M. & JEYAVIJAYAN, S. 2011. FTIR and FT-Raman spectra, assignments, ab initio HF and DFT analysis of xanthine. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 79, 161-168.
- ASCHER, M. S., GOTTLIEB, A. A. & KIRKPATRICK, C. H. 2014. *Transfer factor: Basic properties and clinical applications*, London, Academic Press Inc.

- BERGMANN, F. & DIKSTEIN, S. 1955. The relationship between spectral shifts and structural changes in uric acids and related compounds. *Journal of the American Chemical Society*, 77, 691-696.
- BRITTAIN, H. G. 2009. Vibrational Spectroscopic Studies of Cocrystals and Salts. 1. The Benzamide–Benzoic Acid System. *Crystal Growth & Design*, 9, 2492-2499.
- CHATTERJEE, C. L., GARG, P. & JAISWAL, R. M. P. 1978. Vibrational spectrum and thermodynamic functions of para methylbenzonitrile molecule. *Spectrochimica Acta Part A: Molecular Spectroscopy*, 34, 943-947.
- CHOWDHURY, J., MUKHERJEE, K. M. & MISRA, T. 2000. A pH dependent surface-enhanced Raman scattering study of hypoxanthine. *Journal of Raman Spectroscopy*, 31, 427-431.
- DIETZEL, U., KUPER, J., DOEBBLER, J. A., SCHULTE, A., TRUGLIO, J. J., LEIMKÜHLER, S. & KISKER, C. 2009. Mechanism of Substrate and Inhibitor Binding of Rhodobacter capsulatus Xanthine Dehydrogenase. *Journal of Biological Chemistry*, 284, 8768-8776.
- FLEMING, G. D., GOLSIO, I., ARACENA, A., CELIS, F., VERA, L., KOCH, R. & CAMPOS-VALLETTE, M. 2008. Theoretical surface-enhanced Raman spectra study of substituted benzenes: II. Density functional theoretical SERS modelling of o-, m-, and p-methoxybenzonitrile. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 71, 1074-1079.
- GAO, X., DAVIES, J. P. & WEAVER, M. J. 1990. Test of surface selection rules for surface-enhanced Raman scattering: the orientation of adsorbed benzene and monosubstituted benzenes on gold. *The Journal of Physical Chemistry*, 94, 6858-6864.
- GOODALL, B. L., ROBINSON, A. M. & BROUSSEAU, C. L. 2013. Electrochemical-surface enhanced Raman spectroscopy (E-SERS) of uric acid: a potential rapid diagnostic method for early preeclampsia detection. *Physical Chemistry Chemical Physics*, 15, 1382-1388.
- ISHIMARU, K., HATA, T., BRONSVELD, P., NISHIZAWA, T. & IMAMURA, Y. 2007. Characterization of sp²- and sp³-bonded carbon in wood charcoal. *Journal of Wood Science*, 53, 442-448.
- JAUMOT, J., GARGALLO, R., DE JUAN, A. & TAULER, R. 2005. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemometrics and Intelligent Laboratory Systems*, 76, 101-110.
- KODATI, V. R., TU, A. T. & TURUMIN, J. L. 1990. Raman Spectroscopic Identification of Uric-Acid-Type Kidney Stone. *Applied Spectroscopy*, 44, 1134-1136.
- KRISHNAKUMAR, V. & ARIVAZHAGAN, M. 2004. Vibrational and normal coordinate analysis of xanthine and hypoxanthine. *Indian Journal of Pure and Applied Physics*, 42, 411-418.
- MROZEK, M. F., WASILESKI, S. A. & WEAVER, M. J. 2001. Periodic Trends in Electrode–Chemisorbate Bonding: Benzonitrile on Platinum-Group and Other Noble Metals As Probed by Surface-Enhanced Raman Spectroscopy Combined with Density Functional Theory. *Journal of the American Chemical Society*, 123, 12817-12825.
- SCHWAN, J., ULRICH, S., BATORI, V., EHRHARDT, H. & SILVA, S. R. P. 1996. Raman spectroscopy on amorphous carbon films. *Journal of Applied Physics*, 80, 440-447.
- TALLANT, D., FRIEDMANN, T., MISSERT, N., SIEGAL, M. & SULLIVAN, J. 1997. Raman spectroscopy of amorphous carbon. *MRS Online Proceedings Library*, 498, 37.
- TAULER, R. 1995. Multivariate curve resolution applied to second order data. *Chemometrics and Intelligent Laboratory Systems*, 30, 133-146.
- WANG, L., TUSCHEL, D. & ASHER, S. A. 2011. 229 nm UV photochemical degradation of energetic molecules. *SPIE Defense, Security, and Sensing, International Society for Optics and Photonics*, 808, 80181B-1 - 80181B-6.

6 From Multi-Step Enzyme Monitoring to Whole-Cell Biotransformations: Development of Real-Time UVRR Spectroscopy

Heidi Fisk, Yun Xu, Chloe Westley, Nick Turner, Jason Micklefield and Royston Goodacre

School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK

6.1 DECLARATION

This chapter consists of one journal article, currently pending submission at Analytical Chemistry:

FISK, H., XU, Y., WESTLEY, C., TURNER, N., MICKLEFIELD, J., & GOODACRE, R. 2017. From Multi-Step Enzyme Monitoring to Whole-Cell Biotransformations: Development of Real-Time UVRR Spectroscopy

This article has been reproduced in an unchanged format except for minor adjustments to incorporate it into this thesis.

As primary author on this publication, I carried out all protein optimisation (cloning, expression, purification), UVRR experiments and write-up. Dr Yun Xu performed the chemometrics comprising of MCR-ALS and MCR-ALS-HM. Dr. Chloe Westley was fundamental for the development of the flow-cell apparatus and general issues relating to the UVRR instrument. Prof. Nick Turner gave help and advice surrounding the biotransformations. Co-supervisor, Prof. Jason Micklefield, gave important insights into the biotransformations studied, as well as helping to shape the direction of the research. Prof. Royston Goodacre, as Principal Investigator, provided insights and advice on the project brief, giving direction during the research.

6.2 ABSTRACT

Process Analytical Technologies (PAT) are used within industry to give real-time measurements of critical quality parameters, ultimately improving the quality by design (QbD) of the final product and reducing manufacturing costs. Spectroscopic and spectrophotometric methods are readily employed within PAT due to their ease of use, compatibility towards a range of sample types, robustness and their multiplexing capabilities. We have developed a UV resonance Raman (UVRR) spectroscopy approach to quantify industrially-relevant biotransformations accurately, focusing on nitrile metabolising enzymes: nitrile hydratase (NHase) and amidase *versus* nitrilase activity. Sensitive detection of the amide intermediate by UVRR spectroscopy enabled discrimination between the two nitrile-hydrolysing pathways. Development of a flow-cell apparatus further exemplifies its suitability towards PAT measurements, incorporating *in situ* analysis within a closed system. Multivariate curve resolution-alternating least squares (MCR-ALS) was applied to the UVRR spectra, as well as off-line HPLC measurements, to enable absolute quantification of substrate, intermediate and product. Further application of hard modelling to MCR-ALS deconvolved concentration profiles enabled accurate kinetic determinations, thus removing the requirement for comparative off-line HPLC. Finally, successful quantitative measurements of *in vivo* activity using whole-cell biotransformations, where two *Escherichia coli* strains expressing either NHase (transforming benzonitrile to benzamide) or amidase (further conversion of benzamide to benzoic acid), illustrates the power, practicality and sensitivity of this novel approach for multi-step, and with further refinement we believe multiple micro-organism biotransformations.

6.3 INTRODUCTION

Nitrile hydrolysing enzymes, namely, nitrilases, nitrile hydratases (NHase) and amidases, are extensively used within chemical synthesis in industry owing to their high selectivity and activity as well as their broad substrate specificity (Zhou *et al.*, 2005). These nitrile metabolising enzymes work at ambient temperature(s) and pressure(s) and can offer enantioselective control, meaning they are attractive alternatives within pharmaceutical processes. At present, acrylamide (>30,000 tons per year), nicotinamide (Vitamin B₃, >6,000 tons per year), (*R*)-mandelic acid (a drug precursor), pyrazinamide (an antituberculosis agent) and 5-cyanovaleramide (a herbicide intermediate) are largely synthesised using nitrile hydrolysing biocatalytic routes (Asano *et al.*, 1982, Mauger *et al.*, 1988, Nagasawa *et al.*, 1988, Kobayashi *et al.*, 1992, Ghisalba *et al.*, 2009, Gong *et al.*, 2012). There has also been increasing interest in the detoxification of wastewater containing toxic nitriles of anthropogenic origins using nitrile hydrolysing enzymes (Wang *et al.*, 2004, Gong *et al.*, 2012).

Due to the appealing nature of these biocatalysts, different techniques to monitor the progression of these biotransformations have been investigated. On-line techniques (in contrast to off-line) are often more desirable as they directly monitor the reaction in real-time, providing rapid, continuous feedback. As such, sample manipulation, extraction or (partial) purification is no longer required prior to analysis, reducing the error by minimizing the need for sample transfers and handling. Furthermore, real-time reaction monitoring greatly improves the efficiency and accuracy of the overall process within PAT (process analytical technology), an essential component of QbD (quality by design) (Hinz, 2006, Undey *et al.*, 2011). Fluorometric and colorimetric assays are examples of on-line techniques that have been widely studied to monitor nitrile hydrolysis. Examples of colorimetric assays include those that rely on changes in pH (Banerjee *et al.*, 2003, Santoshkumar *et al.*, 2009, Angelini *et al.*, 2015, Black *et al.*, 2015), whereas spectrophotometric assays have involved monitoring ferrous and ferric ions (He *et al.*, 2011, Lin *et al.*, 2011), as well as NADH consumption when paired with an amidase and glutamate dehydrogenase (Reisinger *et al.*, 2006). Whilst these techniques are advantageous over other more time consuming and labour-intensive off-line techniques, such as HPLC, NMR or GC-MS or LC-MS, a major limiting factor is the requirement of a fluoro- or chromo-genic substrate, along with limited diagnostic and structural information (Kumar and Clark, 2006, Acker and Auld, 2014). Consequently, there is an increased effort towards the emergence of new and improved high-throughput screening (HTS) methods within biocatalysis and *in situ* PAT techniques. Raman spectroscopy offers considerable potential as a physicochemical technique for monitoring real-time enzymatic catalysed biotransformations. However, 'normal' Raman scattering is often too weak, limiting its application, yet enhancements

of $10^3 - 10^5$ are attainable when deep-UV (244 nm in this case) excitation is employed. In UV resonance Raman (UVRR) spectroscopy the laser frequency coincides with the electronic transitions of the molecules under investigation, specifically aromatic and conjugated systems, thus enhancing the Raman response (Asher and Johnson, 1984, Asher, 1993).

We recently demonstrated that UVRR could successfully monitor the real-time progression of biotransformations using either NHase or xanthine oxidase (conversion of hypoxanthine to xanthine, followed by xanthine to uric acid) (Westley *et al.*, 2017). UVRR achieved results that were in agreement with off-line HPLC analysis, but with a >30-fold reduction in acquisition time (20 s measurement time for UVRR *versus* >10 min for each HPLC run). Ultimately, rapid and sensitive detection of these analytes was possible by exploiting aromatic functionality which is inherent to their structure.

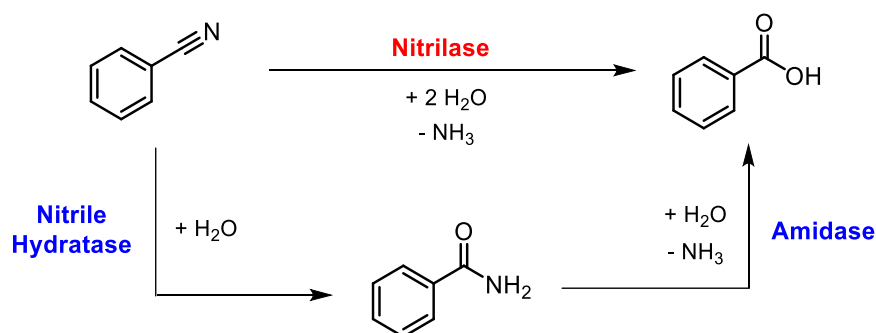


Figure 6.1. Enzymatic conversion of benzonitrile to the corresponding carboxylic acid (benzoic acid) and/or carboxamide (benzamide).

In this study, we further investigate the application of UVRR to discriminate quantitatively between nitrile metabolising enzymes. Nitrile catabolism comprises of two distinct pathways: (1) direct generation of carboxylic acids *via* nitrilase; and (2) a two-step process whereby nitrile hydratases catalyse the formation of an amide, which is subsequently hydrolysed by an amidase to give the corresponding carboxylic acid (Figure 6.1) (Zhou *et al.*, 2005). Consequently, discrimination between these pathways until now has been problematic as they generate the same final product, resulting in only semi-quantitative or qualitative analysis. We describe how UVRR spectroscopy can be used in real-time to easily distinguish between the two pathways, without the requirement of additional fluoro- or chromo-genic substrates, enabling full recovery of the reaction sample whilst affording quantitative analysis. Furthermore, a flow-cell apparatus has been developed to reduce signal-to-noise interference, further demonstrating its amenability for use in industrial processes (closed system enabling direct *in situ* measurements). Finally, investigations were extended to include whole-cell biotransformations (*in vivo*), omitting protein purification steps thus further increasing the overall speed and utility of this technique.

6.4 EXPERIMENTAL

6.4.1 Reagents and Materials.

All chemical reagents were of analytical grade and used with no additional purification unless otherwise stated. A nitrile hydratase (EC 4.2.1.84) construct was kindly provided by the Uwe Bornscheuer group (Greifswald University) (Rzeznicka *et al.*, 2010), and an amidase (EC 3.5.1.4) synthetic gene was purchased from Genewiz (New Jersey, USA). Full details of cloning, expression and protein purification is described within the SI.

6.4.2 Reaction sample preparation.

The starting reaction mixture contained either benzonitrile or benzamide (final concentration 1.25×10^{-2} M) in potassium phosphate buffer (2.5×10^{-2} M, pH 7.2). Reactions were performed using either purified enzyme (*in vitro*) or *Escherichia coli* whole-cells (resting state) which expressed either NHase or amidase protein (*in vivo*). For reactions involving purified enzymes, NHase + amidase (6.0×10^{-6} M and 3×10^{-6} M, respectively) were added to the reaction mixture to initiate the biotransformation. For whole-cell biotransformations, NHase or amidase (500 μ L or 250 μ L respectively, of concentrated whole-cells in buffer) were introduced into the reaction mixture or a combination of the two (375 μ L of NHase and 25 μ L of amidase of whole-cells in buffer). See SI, supplementary methods for further information.

6.4.3 UVRR Instrumentation.

UVRR was performed using a Renishaw Raman 1000 system (Renishaw, Wotton-under-edge, Gloucestershire, UK). Approximately ~ 0.2 mW of power was delivered to the sample using a Lexel Model 95 ion laser (frequency doubled) emitting at 244 nm.

6.4.4 Reaction set-up.

UVRR monitoring of biotransformations were performed using a flow-cell set-up. In brief, a quartz flow-cell was focussed beneath the UVRR microscope objective, which was connected to the reaction reservoir using tubing (Figure S6.1 in SI). Continuous stirring of the reaction mixture was achieved using a peristaltic pump, along with a stirrer bar agitating the reaction reservoir (the location at which enzyme is added and HPLC samples removed). A total reaction volume of 10 mL was used, UVRR analysis was performed using a 20 s acquisition time at various intervals (mainly every minute). In addition, every 2-3 min 20 μ L of the reaction mixture was removed and quenched with MeOH (180 μ L), before being subjected to comparative HPLC analysis.

6.5 RESULTS AND DISCUSSION

6.5.1 Multi-step enzyme biocatalysis

To explore if UVRR could distinguish between the two nitrile metabolising pathways, we investigated the more complex, two-step pathway combining both a nitrile hydratase and an amidase. A flow-cell apparatus was developed and optimized to reduce signal-to-noise interference by incorporating a quartz flow-cell at the site of UVRR analysis (images of the set-up are shown in Figure S6.1 in SI). Benzonitrile was investigated as the starting material (SM), which produces both benzamide as the intermediate (I) and benzoic acid as the final product (P) in the presence of NHase and amidase (Figure 6.1). Characteristic UVRR peaks for each analyte are highlighted in Figure 6.2 (see Table S6.1 in SI for tentative band assignments). Benzamide and benzoic acid produce very similar UVRR responses yet subtle differences are seen, most noticeably, one can observe the peak shift from 1413 to 1389 cm^{-1} (benzamide to benzoic acid) and the weak peak at 847 cm^{-1} which is absent in both benzonitrile and benzamide but present within the benzoic acid product.

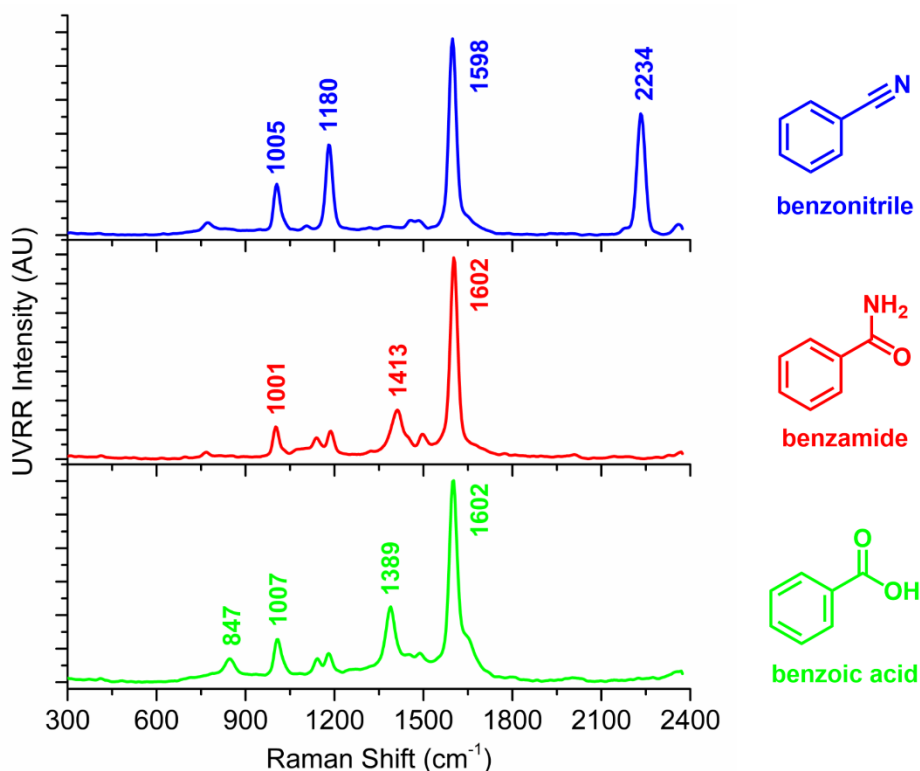


Figure 6.2. Average UVRR spectra ($n = 2$) of each analyte under investigation: benzonitrile (blue), benzamide (red) and benzoic acid (green). Spectra were obtained at 1.1×10^{-2} M in potassium phosphate buffer (2.5×10^{-2} M, pH 7.2), using conditions and concentrations that were representative of the initial reaction mixtures, with characteristic peaks identified (see **Error! Reference source not found.** in SI for assignments). UVRR spectra were obtained for 20 s with baseline correction and normalisation applied (see the SI, “Data processing” for full details).

Initially we looked at the conversion of benzonitrile to benzamide and subsequently to benzoic acid using NHase and amidase *in vitro* (purified enzymes), with comparative HPLC measurements acquired to confirm that the *in situ* UVRR measurements were accurate. Multivariate curve resolution-alternating least squares (MCR-ALS) is a soft-modelling method, used mathematically to deconvolve an instrumental response into the pure contributions of individual components present within a mixture. The input of pure spectra of each component of the mixture is an initial requirement, and the iterative method can then deconvolve the complex mixture and provide concentration profiles of each component (Tauler, 1995, Jaumot *et al.*, 2005). To benchmark this approach HPLC measurements of the same samples allows the comparison of the known analyte concentrations with the predicted UVRR concentrations *via* MCR-ALS modelling, thus serving as external validation (an overview of the spectral pre-processing and MCR-ALS process is shown in Figure S6.2 in SI). One can also compare how well the pure analyte spectra (as initially inputted into the model) and the resolved MCR-ALS spectra agree with one another and in this case the real and deconvolved were congruent (Figure S6.3 in SI).

It is important to note that UVRR spectra were acquired every minute throughout the monitoring process, whereas HPLC aliquots were taken less frequently (2-3 min intervals). Time points consisting of both HPLC and UVRR measurements were used as the training set for the MCR-ALS model, with only UVRR measurements used as the test set. Figure 6.3 shows the resolved concentration profiles, illustrating very good agreement between the HPLC and UVRR predictions. The MCR-ALS model correctly recognized all three components within the reaction mixture, despite their spectral similarities. Thus, enabling distinction between the two-step pathway (NHase + amidase) *versus* nitrilase catalysis (no intermediate). Regression-coefficient values (R^2) demonstrate the proportion of variability within a data set, as accounted for by the statistical model (MCR-ALS). Generally, R^2 values closer to 1 demonstrate excellent fit and correlation between the HPLC measured and UVRR predicted concentrations. Results are provided in Figure S6.4 in SI for all three analyte components – benzonitrile (SM), benzamide (I) and benzoic acid (P) – with their corresponding R^2 values: 0.9637, 0.9076 and 0.9895, respectively, indicating very good agreement between the two techniques.

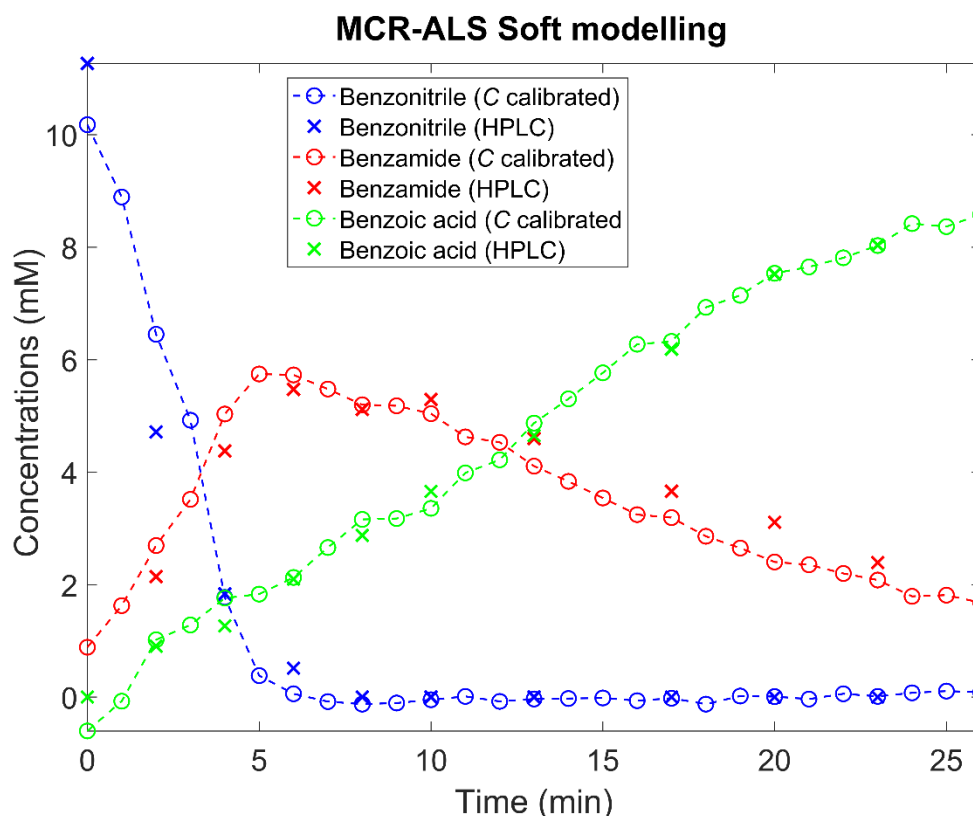


Figure 6.3. An MCR-ALS model was applied to the UVRR data where it successfully deconvolved spectra into its pure components for the biotransformation using pure enzymes. This figure shows the reaction dynamics from the real-time UVRR measurements (denoted by circular symbols) and offline HPLC data (denoted by cross symbols) as a function of time for the conversion of benzonitrile (SM; blue) to benzamide (I; red) to benzoic acid (P; green), catalysed by NHase and amidase, correspondingly.

To complement the incorporation of UVRR within on-line PAT processes further, we wanted to explore the application of hard modelling (HM) (Molloy *et al.*, 1999, de Carvalho *et al.*, 2006), as this would remove the requirement for external calibration (*i.e.* the additional HPLC measurements). The multivariate curve resolution-alternating least squares-hard modelling (MCR-ALS-HM) process is shown in Figure S6.5 in SI. In brief, assuming that the two steps of the biotransformations both followed a first-order reaction, the resolved concentration profiles from MCR-ALS were used to determine the reaction rate constants k_1 and k_2 using the kinetic models as shown in Figure S6.6 in SI. The concentrations of the reactants at any time points during the reaction were then derived using the estimated k_1 , k_2 and the known initial concentration of the starting material (*i.e.* benzonitrile). The MCR-ALS-HM results are shown in Figure 6.4 along with comparative HPLC results to highlight their consistency. One can see that the results are once again in good agreement, particularly for the benzonitrile (SM). However, benzamide and benzonitrile HM concentration profiles do not agree with the HPLC calibration concentrations quite as well. The MCR-ALS-HM approach assumes that 100 % of the SM (*i.e.* benzonitrile) is converted to product(s)

(i.e. benzamide + benzoic acid). Upon further inspection of the total concentration throughout the monitoring process, we experience significant fluctuations in total mass, consequently causing a disparity between the HPLC and MCR-ALS-HM results. The total analyte concentration at various timepoints (as calculated by HPLC) is shown in Table S6.2 in SI. After the initial timepoint, a substantial drop in the total concentration occurs and a third of the starting concentration is unaccounted for, which subsequently increases again with increasing time (final timepoint concentration = 96 % of starting concentration). We hypothesise that these variations in the total concentration is a consequence of reduced solubility of the intermediate (benzamide), resulting in partial insolubility before being converted to benzoic acid (product) where it is soluble once again due to the reduced concentration of benzamide. As the MCR-ALS and HPLC results (Figure 6.3) were in such good agreement with one another, we can conclude that both techniques (UVRR and HPLC) are detecting this reduction in concentration and it is a true effect. Unfortunately, the HM results cannot adjust for this mass imbalance, thus we see a discrepancy between the MCR-ALS-HM and HPLC results (Figure 6.4). Nevertheless, we can calculate reaction rate constants using the MCR-ALS-HM model and these constants were $k_1 = 0.3316$ and $k_2 = 0.0797$ for NHase and amidase, respectively (the equations for kinetic calculations are shown in Figure S6.6 in SI).

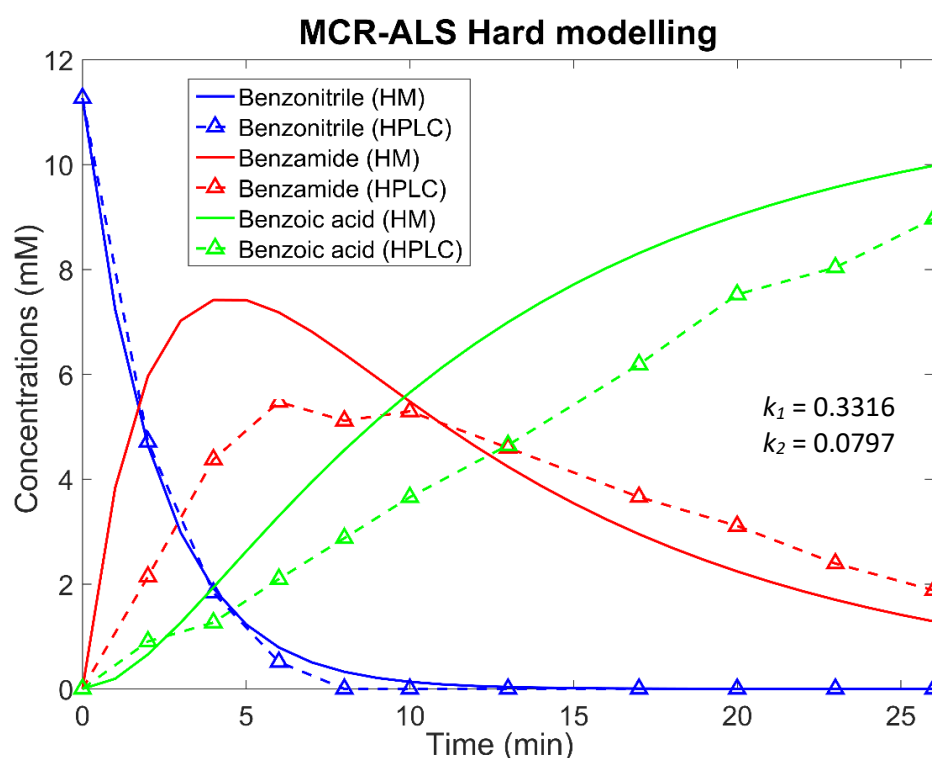


Figure 6.4. Multivariate curve resolution-alternating least squares-hard modelling (MCR-ALS-HM) results show good agreement with HPLC measured concentrations. Calculated kinetic rates of each enzyme were $k_1 = 0.3316$ and $k_2 = 0.0797$ for NHase and amidase, respectively.

6.5.2 Whole-cell biotransformations incorporating multi-step enzyme biocatalysis

In addition to the *in vitro* pure enzyme assessment above, we wanted to perform this biotransformation using *in vivo* conditions using whole-cells. Omitting protein purification steps would increase the overall efficiency and speed of this real-time monitoring approach, further proving its suitability within industrial processes. NHase and amidase were expressed within *E. coli* BL21 (DE3) cells using optimised expression conditions (see SI section 'Protein expression and purification'). Once cells were induced for protein expression and allowed to grow overnight, the cells were harvested by centrifugation and re-suspended in potassium phosphate buffer prior to use within the reaction. To ensure that bacterial cells were in the resting state and no further growth was experienced, optical density (OD) measurements at 600 nm were performed throughout the monitoring process. OD₆₀₀ measurements were relatively consistent throughout the 60 min period and thus we conclude that there was no significant further bacterial growth (see SI section 'OD₆₀₀ measurements' and Table S6.3). Initially we performed two separate whole-cell biotransformations: the first contained *E. coli* expressing NHase for conversion of benzonitrile (SM) to benzamide (P), and the second biotransformation with benzamide (SM) used *E. coli* expressing amidase, generating benzoic acid (P).

The first one-step, *in vivo* biotransformation studied was the conversion of benzonitrile to benzamide using *E. coli* cells expressing NHase activity. Consistent with our previous investigations, both UVRR and HPLC measurements were taken and MCR-ALS analysis performed. The MCR-ALS resolved profiles for starting material and product can be observed in Figure S6.7, along with the predicted concentration profiles and known HPLC concentrations comparisons of each analyte (Figure S6.8 in SI). The resolved MCR-ALS spectrum of benzamide (P) was less successful than benzonitrile (SM), reflected in their R^2 values of 0.7286 and 0.9514, correspondingly. This is likely due to the *E. coli* background response coinciding with the benzamide product peaks within the region of 1200 – 1800 cm^{-1} . Despite this the MCR-ALS modelled concentrations *versus* HPLC results (Figure 6.5) are evidently in very good agreement with one another and the UVRR approach can successfully quantify the whole-cell catalysed hydrolysis of benzonitrile to benzamide.

As with the pure enzyme reactions hard modelling was further applied to these results, however as previous, the reaction experienced a reduction in the total concentration (assuming benzamide insolubility) which had a substantial effect on MCR-ALS-HM. The final reaction timepoint ($t = 18$ min) had a 38 % mass loss when compared to the starting concentration, thus rendering the MCR-ALS-HM results inaccurate (MCR-ALS-HM kinetic equations are shown in Figure S6.9 in SI).

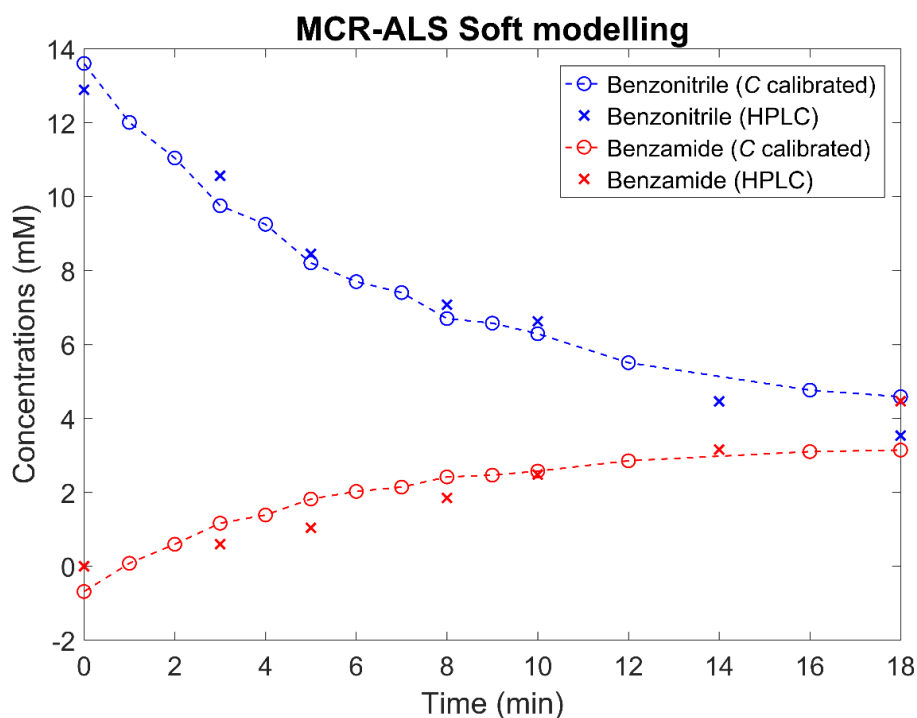


Figure 6.5. MCR-ALS model applied to the whole-cell biotransformation of benzonitrile (SM, blue) to benzamide (P, red) using NHase-containing *E. coli* cells. Figure shows the reaction dynamics from the real-time UVRR measurements (denoted by circular symbols) and off-line HPLC data (denoted by cross symbols) as a function of time.

Next, the second step of the biotransformation (benzamide to benzoic acid) was monitored in the presence of amidase-expressing *E. coli* cells. The increased expression of amidase *in vivo*, relative to NHase, enables a much lower quantity of cells to catalyse the reaction (which we estimated to be 1:15), hence the *E. coli* background does not overtly dominate the UVRR response and characteristic peaks are observed. Thus, the MCR-ALS soft modelling approach was applied to the UVRR and HPLC data and were shown to be in very good agreement with one another ($R^2 = 0.9439$ and 0.9862 for benzamide and benzoic acid, Figure S6.10 and S6.11 in SI). Figure 6.6A illustrates the similarity between the two methods. The total concentration remains stable throughout, so MCR-ALS hard modelling was applied to the data, removing the requirement of comparative HPLC (MCR-ALS-HM kinetic models shown in Figure S6.9 in SI). Results of the MCR-ALS-HM are shown in Figure 6.6B, where one can see that the hard modelling results (represented by the solid lines) are highly comparable to the off-line HPLC measured concentrations (dashed lines), enabling an accurate measurement of the kinetic rate, identified as $k = 0.1986$.

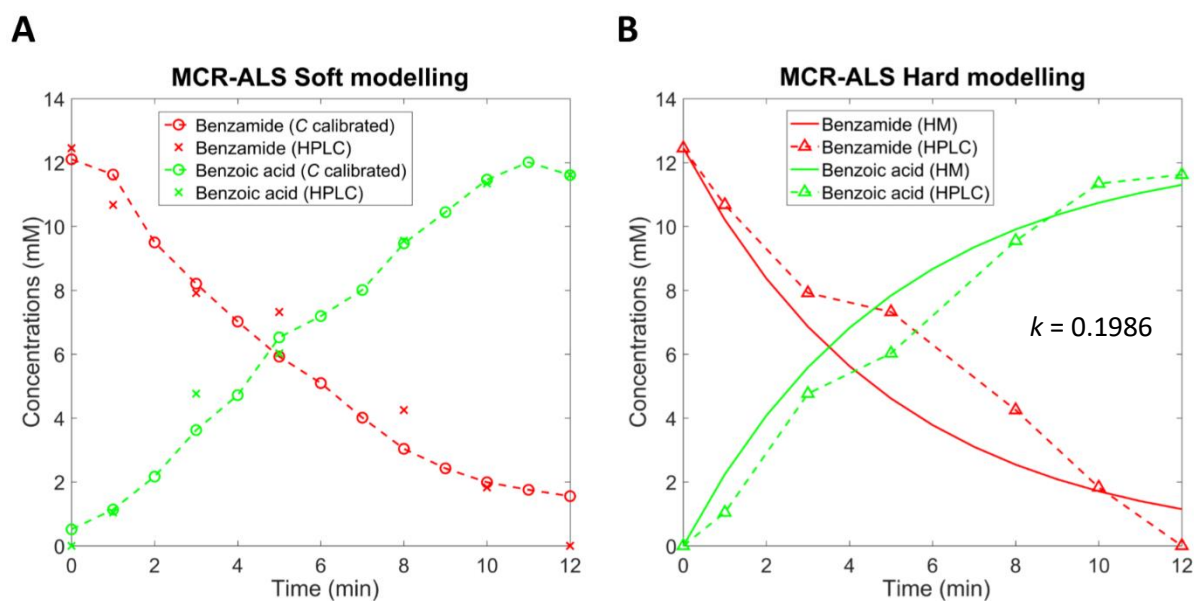


Figure 6.6. Modelling results for amidase-expressing bacterial whole-cells catalysing the conversion of benzamide (SM) to benzoic acid (P). **(A)** MCR-ALS soft modelling results and **(B)** MCR-ALS hard modelling results, kinetic rate was calculated as $k = 0.1986$.

Finally, we combined the two different *E. coli* together. As discussed above the relative expression levels of NHase were very low in comparison to that of amidase, and this required a much larger quantity of NHase-containing *E. coli* cells within the reaction mixture. We established using HPLC (data not shown) that the 15:1 ratio of NHase:amidase enabled detectable concentrations of benzamide before its subsequent hydrolysis to benzoic acid. Unfortunately, despite this optimisation to see the intermediate, *in vivo* UVRR reaction monitoring in the presence of NHase + amidase (15:1 ratio) was unsuccessful due to the high density of *E. coli* cells. Figure S6.12A in SI shows the corresponding UVRR spectra over the reaction time course, one can easily identify the decreasing nitrile peak at 2234 cm^{-1} , indicative of benzonitrile (SM) consumption. However, the bacterial cells produce a UVRR response which has peaks coinciding with those characteristic of benzamide and benzoic acid ($1200 - 1800\text{ cm}^{-1}$, see Figure S6.12B in SI) which meant that subsequent MCR-ALS modelling was unsuccessful. Despite trying to include the biomass peak profile within the MCR-ALS model (as an additional individual component) the deconvolution into four individual components (SM, I, P and biomass) was still unsuccessful. We believe that if the NHase expression had been higher that we would have been able to achieve an *in vivo* biotransformation using multiple organisms; however we are yet to find a clone that has sufficiently high expression at this time.

6.6 CONCLUSIONS

We have successfully demonstrated how UVRR can be used in a real-time, label-free and rapid (20 s) manner to discriminate between nitrile-metabolising pathways (NHase + amidase *versus* nitrilase) by observation of the amide intermediate. Combining on-line UVRR measurements with off-line HPLC analysis, MCR-ALS modelling could be implemented to quantify each of the components accurately within the reaction. High R^2 values demonstrate that the two analytical techniques are in very good agreement with one another. The similarity of the UVRR responses of the intermediate (benzamide) and product (benzoic acid) does not appear to impact the model's ability to distinguish them, further indicating the sensitivity and selectivity of this combined UVRR + MCR-ALS approach. In contrast to previous open pot reactions where products and substrates may evaporate (Westley *et al.*, 2017), the development of a flow-cell apparatus enables collection of UVRR measurements through a quartz flow-cell, highlighting its suitability within PAT processes and facilitating measurements within a closed system. To complement the application of a closed system further, hard modelling was combined with MCR-ALS (MCR-ALS-HM) as to remove the requirement of external HPLC measurements, with results shown to be in good agreement with off-line HPLC (validation). However, fluctuations in the total concentration (arising from reduced solubility of the intermediate) reduced the accuracy of the model as the mass of the system was no longer in equilibrium.

Finally, to extend our investigations and the applicability of this approach towards industrial development, we introduced *in vivo* conditions by using *E. coli* whole-cells which expressed either NHase or amidase activity. Omitting protein purification is attractive as it increases the overall speed, enabling faster turnover of the desired product. Low-level expression of NHase *in vivo* and the consequent requirement of a high density of bacterial cells led to unsuccessful detection and discrimination of the individual analytes when performed in the presence of both NHase and amidase bacterial cells. However, separating the two individual biotransformations proved to be very successful. MCR-ALS soft modelling and hard modelling approaches could be applied to the amidase-catalysed reaction, owing to its high-level expression (*in vivo*) and avoidance of solubility interference. We are confident that these results showcase the sensitivity and accuracy of UVRR spectroscopy for *in situ*, continuous measurements of enzyme activity, and its suitability for implementation within industry.

6.7 REFERENCES

- ACKER, M. G. & AULD, D. S. 2014. Considerations for the design and reporting of enzyme assays in high-throughput screening applications. *Perspectives in Science*, 1, 56-73.
- ANGELINI, L. M. L., DA SILVA, A. R. M., ROCCO, L. D. F. C. & MILAGRE, C. D. D. F. 2015. A high-throughput screening assay for distinguishing nitrile hydratases from nitrilases. *Brazilian Journal of Microbiology*, 46, 113-116.
- ASANO, Y., YASUDA, T., TANI, Y. & YAMADA, H. 1982. A New Enzymatic Method of Acrylamide Production. *Agricultural and Biological Chemistry*, 46, 1183-1189.
- ASHER, S. & JOHNSON, C. 1984. Raman spectroscopy of a coal liquid shows that fluorescence interference is minimized with ultraviolet excitation. *Science*, 225, 311-313.
- ASHER, S. A. 1993. UV resonance Raman spectroscopy for analytical, physical, and biophysical chemistry. Part 1. *Analytical Chemistry*, 65, 59A-66A.
- BANERJEE, A., KAUL, P., SHARMA, R. & BANERJEE, U. C. 2003. A High-Throughput Amenable Colorimetric Assay for Enantioselective Screening of Nitrilase-Producing Microorganisms Using pH Sensitive Indicators. *Journal of Biomolecular Screening*, 8, 559-565.
- BLACK, G. W., BROWN, N. L., PERRY, J. J. B., RANDALL, P. D., TURNBULL, G. & ZHANG, M. 2015. A high-throughput screening method for determining the substrate scope of nitrilases. *Chemical Communications*, 51, 2660-2662.
- DE CARVALHO, A. R., SÁNCHEZ, M. D. N., WATTOOM, J. & BRERETON, R. G. 2006. Comparison of PLS and kinetic models for a second-order reaction as monitored using ultraviolet visible and mid-infrared spectroscopy. *Talanta*, 68, 1190-1200.
- GHISALBA, O., MEYER, H.-P., WOHLGEMUTH, R. & FLICKINGER, M. C. 2009. Industrial Biotransformation. *Encyclopedia of Industrial Biotechnology*. John Wiley & Sons, Inc.
- GONG, J.-S., LU, Z.-M., LI, H., SHI, J.-S., ZHOU, Z.-M. & XU, Z.-H. 2012. Nitrilases in nitrile biocatalysis: recent progress and forthcoming research. *Microbial Cell Factories*, 11, 142.
- HE, Y.-C., MA, C.-L., XU, J.-H. & ZHOU, L. 2011. A high-throughput screening strategy for nitrile-hydrolyzing enzymes based on ferric hydroxamate spectrophotometry. *Applied Microbiology and Biotechnology*, 89, 817-823.
- HINZ, D. C. 2006. Process analytical technologies in the pharmaceutical industry: the FDA's PAT initiative. *Analytical and Bioanalytical Chemistry*, 384, 1036-1042.
- JAUMOT, J., GARGALLO, R., DE JUAN, A. & TAULER, R. 2005. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemometrics and Intelligent Laboratory Systems*, 76, 101-110.
- KOBAYASHI, M., NAGASAWA, T. & YAMADA, H. 1992. Enzymatic synthesis of acrylamide: a success story not yet over. *Trends in Biotechnology*, 10, 402-408.
- KUMAR, R. A. & CLARK, D. S. 2006. High-throughput screening of biocatalytic activity: applications in drug discovery. *Current Opinion in Chemical Biology*, 10, 162-168.
- LIN, Z.-J., ZHENG, R.-C., LEI, L.-H., ZHENG, Y.-G. & SHEN, Y.-C. 2011. Ferrous and ferric ions-based high-throughput screening strategy for nitrile hydratase and amidase. *Journal of Microbiological Methods*, 85, 214-220.
- MAUGER, J., NAGASAWA, T. & YAMADA, H. 1988. Nitrile hydratase-catalyzed production of isonicotinamide, picolinamide and pyrazinamide from 4-cyanopyridine, 2-cyanopyridine and cyanopyrazine in *Rhodococcus rhodochrous* J1. *Journal of Biotechnology*, 8, 87-95.
- MOLLOY, K. J., MAEDER, M. & SCHUMACHER, M. M. 1999. Hard modelling of spectroscopic measurements. Applications in non-ideal industrial reaction systems. *Chemometrics and Intelligent Laboratory Systems*, 46, 221-230.
- NAGASAWA, T., MATHEW, C. D., MAUGER, J. & YAMADA, H. 1988. Nitrile Hydratase-Catalyzed Production of Nicotinamide from 3-Cyanopyridine in *Rhodococcus rhodochrous* J1. *Applied and Environmental Microbiology*, 54, 1766-1769.

- REISINGER, C., VAN ASSEMA, F., SCHÜRMAN, M., HUSSAIN, Z., REMLER, P. & SCHWAB, H. 2006. A versatile colony assay based on NADH fluorescence. *Journal of Molecular Catalysis B: Enzymatic*, 39, 149-155.
- RZEZNICKA, K., SCHÄTZLE, S., BÖTTCHER, D., KLEIN, J. & BORNSCHEUER, U. T. 2010. Cloning and functional expression of a nitrile hydratase (NHase) from *Rhodococcus equi* TG328-2 in *Escherichia coli*, its purification and biochemical characterisation. *Applied Microbiology and Biotechnology*, 85, 1417-1425.
- SANTOSHKUMAR, M., NAYAK, A. S., ANJANEYA, O. & KAREGOUDAR, T. B. 2009. A plate method for screening of bacteria capable of degrading aliphatic nitriles. *Journal of Industrial Microbiology & Biotechnology*, 37, 111.
- TAULER, R. 1995. Multivariate curve resolution applied to second order data. *Chemometrics and Intelligent Laboratory Systems*, 30, 133-146.
- UNDEY, C., LOW, D., MENEZES, J. C. & KOCH, M. 2011. *PAT Applied in Biopharmaceutical Process Development And Manufacturing: An Enabling Tool for Quality-by-Design*, CRC Press.
- WANG, C.-C., LEE, C.-M. & CHEN, L.-J. 2004. Removal of Nitriles from Synthetic Wastewater by Acrylonitrile Utilizing Bacteria. *Journal of Environmental Science and Health, Part A*, 39, 1767-1779.
- WESTLEY, C., FISK, H., XU, Y., HOLLYWOOD, K. A., CARNELL, A. J., MICKLEFIELD, J., TURNER, N. J. & GOODACRE, R. 2017. Real-Time Monitoring of Enzyme-Catalysed Reactions using Deep UV Resonance Raman Spectroscopy. *Chemistry – A European Journal*, 23, 6983-6987.
- ZHOU, Z., HASHIMOTO, Y. & KOBAYASHI, M. 2005. Nitrile Degradation by *Rhodococcus*: Useful Microbial Metabolism for Industrial Productions. *Actinomycetologica* 19, 18-26.

6.8 SUPPORTING INFORMATION

SUPPLEMENTARY METHODS

Reagents and materials.

All chemical reagents were of analytical grade and used with no additional purification unless otherwise stated. Acetonitrile (HPLC grade), benzonitrile ($\geq 99\%$), benzamide ($\geq 99\%$), benzoic acid ($\geq 99\%$), butanoic acid ($\geq 99\%$), nitrilase (EC 3.5.5.1), TRIS HCl and water (HPLC grade) were purchased from Sigma Aldrich Ltd. (Dorset, UK). Potassium dihydrogen phosphate, dipotassium phosphate and sodium hydroxide were obtained from Fischer Scientific (Loughborough, UK).

Reaction conditions for *in vitro* biotransformations.

NHase + amidase. Nitrile hydratase and amidase were expressed and purified as described in the 'Protein Expression and Purification' section. The starting reaction mixture contained benzonitrile (final concentration 1.25×10^{-2} M with 2.5% MeOH) dissolved in potassium phosphate buffer (2.5×10^{-2} M, pH 7.2). The reaction was initiated upon the addition of NHase (6.0×10^{-6} M) and amidase (3×10^{-6} M).

Reaction conditions for *in vivo* biotransformations.

NHase. Nitrile hydratase was expressed as stated in 'Protein Expression and Purification' section. Bacterial culture was harvested by centrifugation ($4000 \times g$, 4°C for 10 min), followed by resuspension in potassium phosphate buffer (2.5×10^{-2} M, pH 7.2) in $1/40^{\text{th}}$ of the expression volume (for example, 400 mL of bacterial expression culture was resuspended in 10 mL buffer = 40x concentration). N.B. bacterial cells were in a resting state and did not display further growth (see Table S6.3 for OD_{600} measurements). The starting reaction mixture contained benzonitrile (final concentration 1.25×10^{-2} M with 2.5 % MeOH) dissolved in potassium phosphate buffer (2.5×10^{-2} M, pH 7.2). The reaction was initiated upon the addition of NHase *E. coli* suspension (500 μL).

Amidase. Amidase was expressed as stated in 'Protein Expression and Purification' section. Bacterial culture was harvested by centrifugation ($4000 \times g$, 4°C for 10 min), followed by resuspension in potassium phosphate buffer (2.5×10^{-2} M, pH 7.2) in $1/40^{\text{th}}$ of the expression volume. The starting reaction mixture contained benzamide (final concentration 1.25×10^{-2} M) dissolved in potassium phosphate buffer (2.5×10^{-2} M, pH 7.2). The reaction was initiated upon the addition of amidase *E. coli* suspension (250 μL).

NHase + amidase. Nitrile hydratase and amidase were expressed as stated in 'Protein Expression and Purification' section. Bacterial cultures were harvested by centrifugation ($4000 \times g$, 4°C for 10

min), followed by resuspension in potassium phosphate buffer (2.5×10^{-2} M, pH 7.2) in 1/40th of the expression volume. The starting reaction mixture contained benzonitrile (final concentration 1.25×10^{-2} M with 2.5 % MeOH) dissolved in potassium phosphate buffer (2.5×10^{-2} M, pH 7.2). The reaction was initiated by the addition of NHase *E. coli* suspension (375 μ L) and amidase *E. coli* suspension (25 μ L).

Reaction sample preparation and monitoring.

For all biotransformations (purified enzyme and *E. coli* whole-cell), the reaction mixture was focused under the microscope objective within a quartz cuvette attached to the flow-cell set-up (see Figure S6.1). At various time points after enzyme addition, UVRR and HPLC spectra were collected throughout the biotransformation monitoring period. The reaction was performed on a 10 mL scale, allowing for a large enough volume to fill the flow-cell set-up (~6 mL) and allow for ~4 mL of residual volume to enable HPLC sample collection (reaction reservoir, see Figure S6.1).

For all biotransformations, at specific time points, 20 μ L of sample was removed from the reaction mixture and immediately quenched and diluted with 180 μ L of MeOH. The sample was then centrifuged at $21,000 \times g$ for 6 min. 100 μ L of the diluted sample was then transferred to a HPLC vial and subjected to HPLC analysis.

INSTRUMENTATION AND DATA PROCESSING

HPLC Analysis.

HPLC separation was conducted using an Agilent Zorbax Eclipse Plus HPLC system set-up for reverse phase separation consisting of a diode array detector. For all biotransformations, the column was a 150 \times 4.6 mm, Phenomenex Eclipse Plus[®] C18 with a 3.5 μ m particle size. For each injection, the run time was 12.0 min pumped at a flow rate of 1 mL min⁻¹ and at 30 °C column temperature. The mobile phase consisted of a linear gradient, starting conditions of 5% MeCN/H₂O (plus 0.05 % TFA) held for 2 min before increasing to 75 % MeCN/H₂O over 6 min. Prior to washing at 95 % MeCN/H₂O for 1.5 min and re-equilibration to initial conditions over 2.5 min (total run time 12 min). 5 μ L of each sample was introduced using an auto-injector. UV absorbance was detected at 254 nm throughout.

UVRR analysis.

UVRR was performed using a Renishaw Raman 1000 system (Renishaw, Wotton-under-edge, Gloucestershire, UK). Approximately ~0.2 mW of power was delivered to the sampling point using a Lexel Model 95 ion laser emitting at 244 nm. The reaction mixture (10 mL) was continuously

stirred, using a magnetic stirrer plate and magnetic bar, as well as continuous circulation around the flow-cell set-up using a peristaltic pump to avoid photodegradation. Spectra were collected with an acquisition time of 20 s. Only spectra with no demonstrable photo-degradation and signal from the reaction vessel were used for analysis; the effects of photodegradation has been previously investigated (Westley *et al.*, 2017) and in the present study no observable photodegradation products were observed.

Data Processing.

All data were exported from the respective instrument operating software and analysed using Matlab R2015a (The Mathworks, Natick, MA, USA).

HPLC data analysis.

The peaks of the target analytes were integrated with the results of the HPLC data and served as an external validation data set to independently verify the accuracies of the MCR-ALS models in prediction.

UVRF data analysis.

The work-flow of the MCR-ALS soft modelling is provided in Figure S6.2. The data were exported as Galactic .spc files and then imported into Matlab. The spectra were baseline corrected and then normalized to unit norm (*i.e.* sum of squares of each spectrum equals 1). MCR-ALS was then applied to obtain the resolved concentration profile \mathbf{C} and pure spectra \mathbf{S} matrices. A linear regression model was built for each reactant using the HPLC measurements and the corresponding readings in \mathbf{C} at the same time points. This model was then applied to all the time points in \mathbf{C} to calculate the concentrations of the reactant on all the monitored time points.

For hard modelling, the concentration profiles were used to estimate the reaction rate constant k using the corresponding kinetic model as shown in Figure S6.6 and S6.7. The concentrations of each of the reactant of any time point during the reaction can then be determined by using the concentration of the starting material and the estimated k .

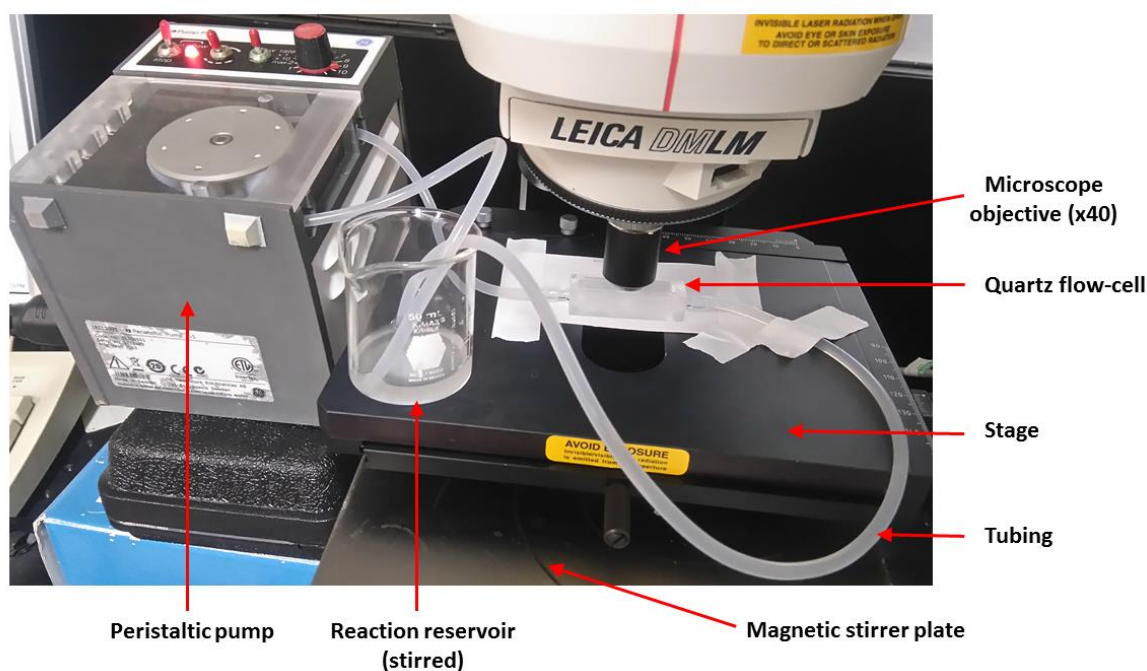


Figure S6.1. Annotated instrument set-up, including flow-cell apparatus, to monitor biotransformations using UV resonance Raman spectroscopy. The 40x UVRR objective was focussed into the quartz flow-cell containing the reaction mixture. 100 % power on the sample (~ 0.2 mW at sampling point) with UVRR data collection using a 20 s acquisition time throughout the time course. Mixing of the reaction sample was achieved using a stirrer bar within the reaction reservoir, as well as continuous flow of the mixture via the peristaltic pump (anti-clockwise direction of flow, plastic tubing connects the quartz flow-cell to the reaction reservoir). Samples for HPLC analysis were removed from the reaction reservoir. Note: set-up as shown above contains *E. coli* whole-cells, thus the reaction mixture is opaque in appearance.

Table S6.1. Tentative UVRR band assignments of benzonitrile, benzamide and benzoic acid at pH 7.2
* (Chatterjee *et al.*, 1978, Gao *et al.*, 1990, Mrozek *et al.*, 2001, Brittain, 2009, Sparrow *et al.*, 2001, Trout *et al.*, 2005)

Raman Shift (cm^{-1})			Tentative band assignment*
Benzonitrile	Benzamide	Benzoic acid	
		847 (w)	C-H ring breathing
1005 (m)	1001 (w)	1007 (m)	C-C-C trigonal breathing
	1138 (vw)	1141 (vw)	NH ₂ rocking mode
1180 (s)	1187 (w)	1180 (w)	C-H in-plane bend
	1413 (m)		C-N stretch
		1389 (m)	C-H ring stretch
1598 (vs)	1602 (vs)	1602 (vs)	C-C in-plane stretch
2234 (s)			C \equiv N stretch

vs-very strong, s-strong, m-medium, w-weak, vw-very weak.

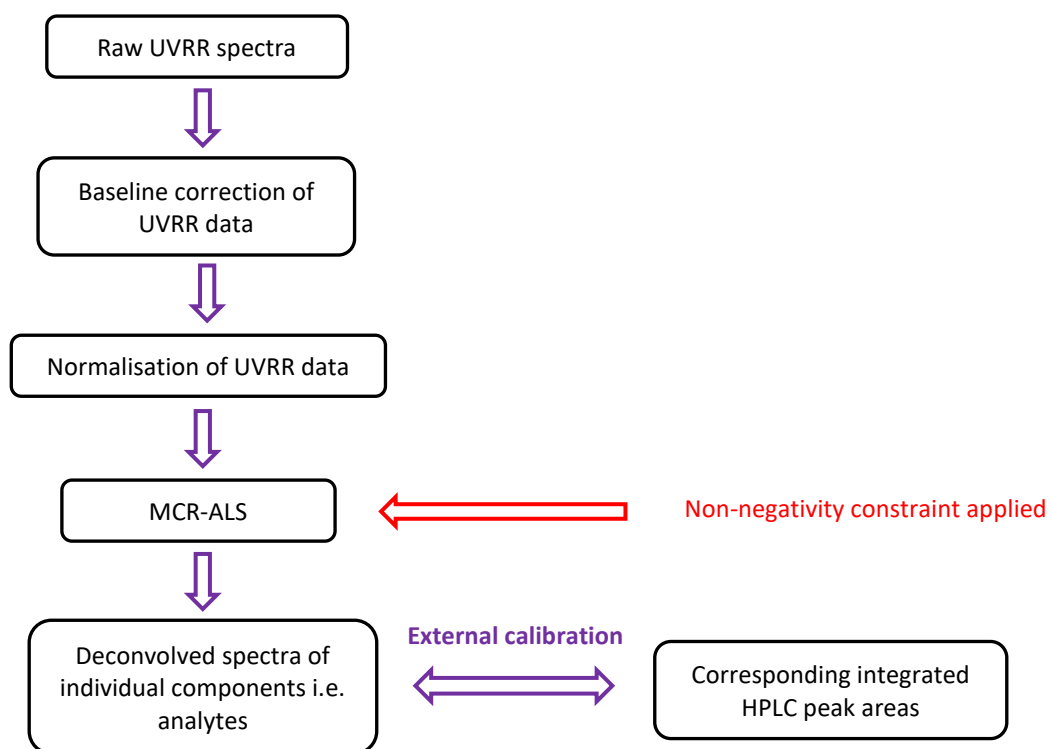


Figure S6.2. Flow diagram summarising data pre-processing and MCR-ALS process used for predicting concentrations of each analyte from the reaction mixture.

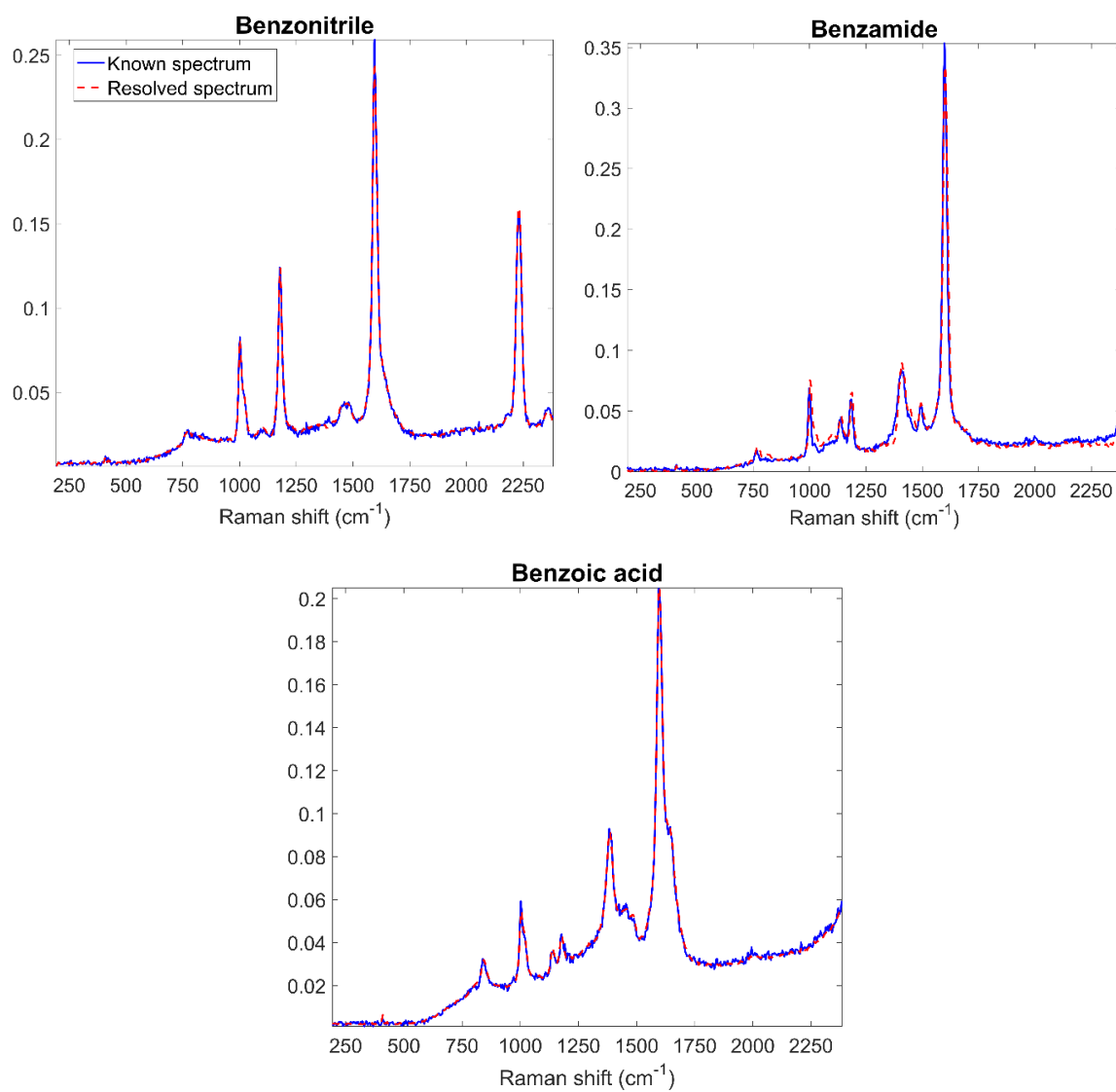


Figure S6.3. MCR-ALS deconvolved spectrum (dashed, red line) of individual analytes from the mixture compared with their known UVRR spectrum from standards (solid, blue line). The initial reaction mixture includes benzonitrile (SM) and is initiated upon the addition of NHase and amidase as purified enzymes.

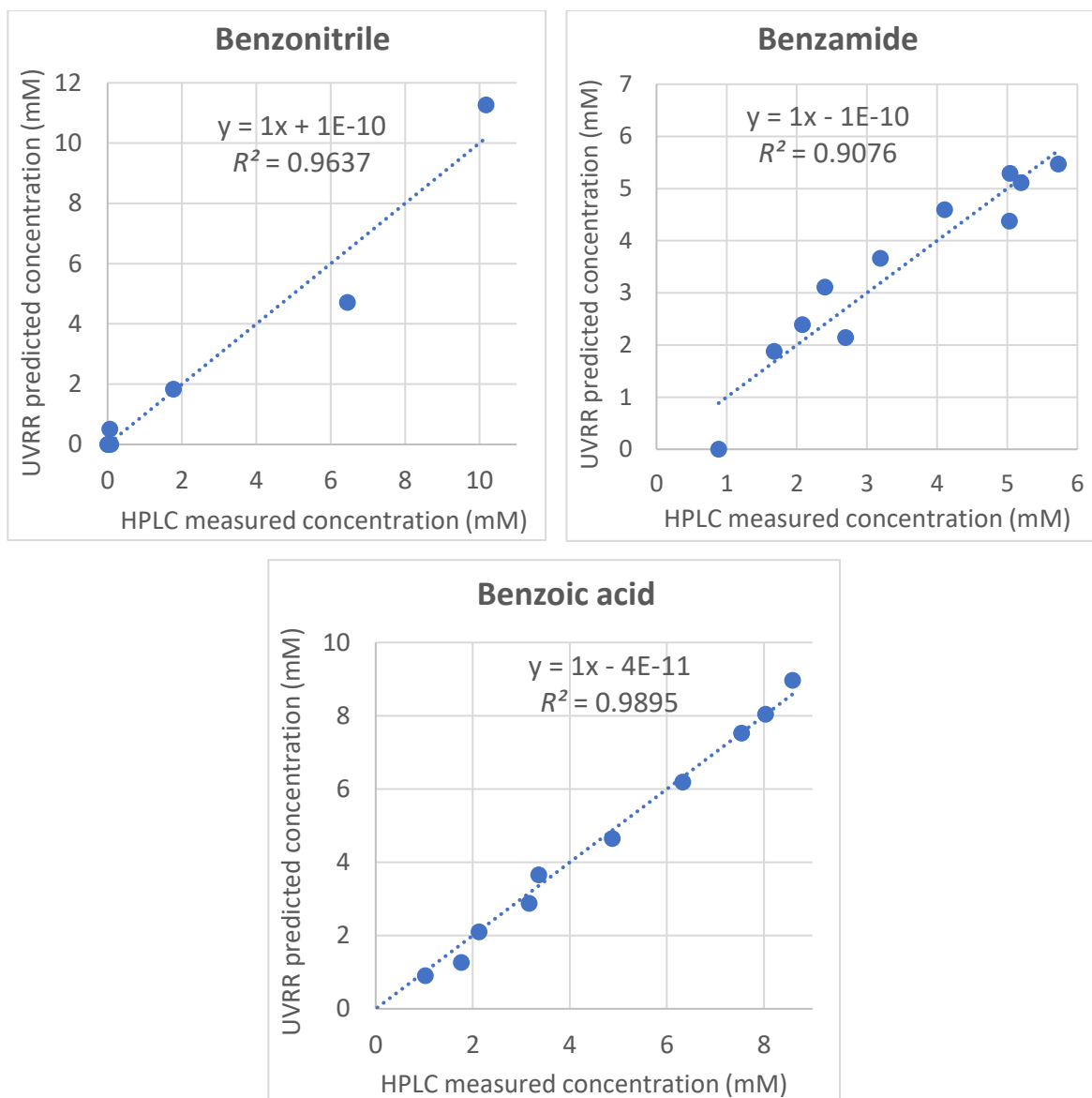


Figure S6.4. MCR-ALS prediction from UVRR of analyte concentrations versus HPLC measured concentrations (mM). With R^2 values ranging from 0.9076 to 0.9895 and indicates that results are in very good agreement with one another.

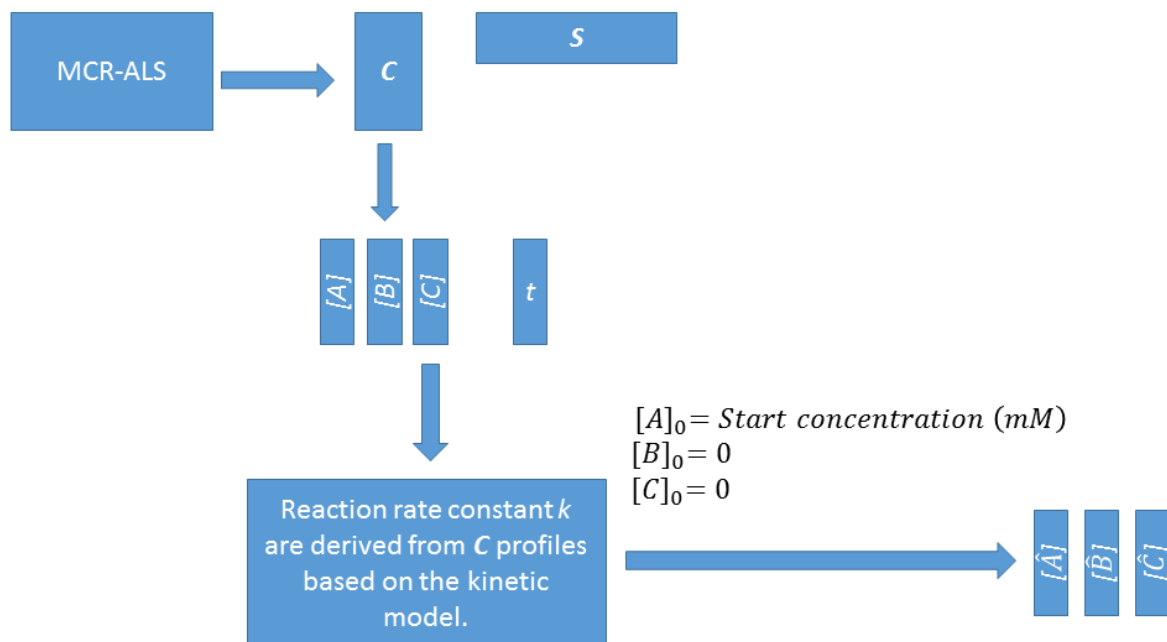


Figure S6.5. Flow diagram summarising the MCR-ALS and HM processes. MCR-ALS model predicts the concentration profiles of each analyte, which are then taken forward to HM.

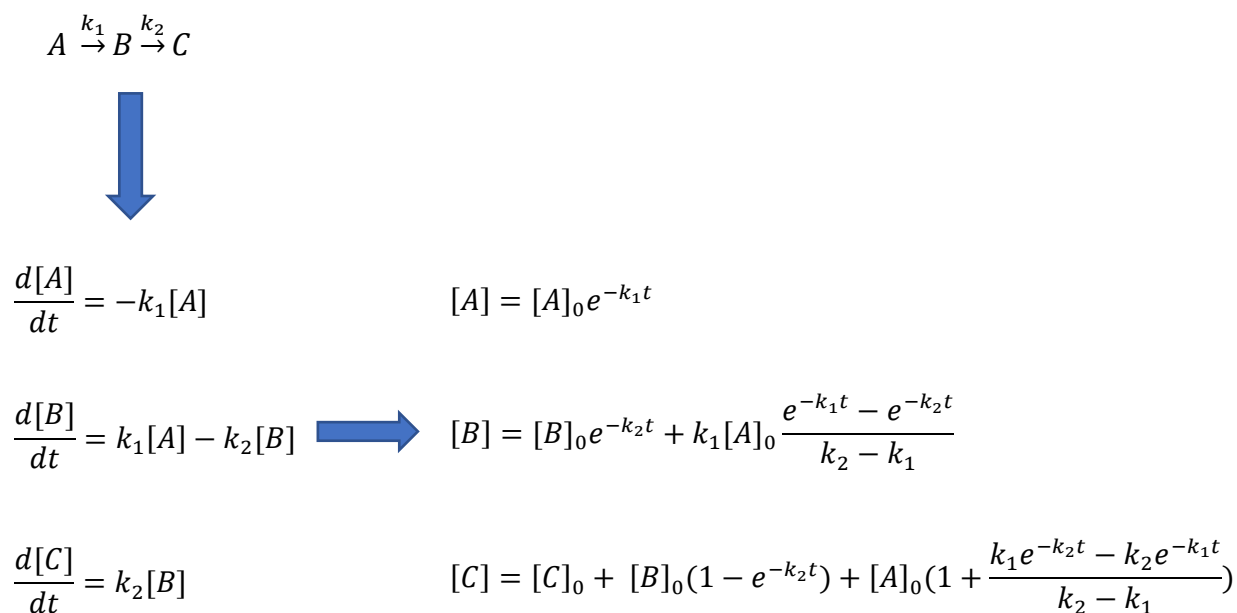


Figure S6.6. Kinetic model used within the MCR-ALS-HM approach to calculate the concentrations of each analyte at any time point during the reaction where A is benzonitrile, B is Benzamide, C is Benzoic acid and t is time. We assume that the biotransformations followed a first order reaction, during which A had converted to B , and B had subsequently converted to C , at reaction rate constants k_1 and k_2 respectively.

Table S6.2. Table denoting the total concentration of analytes with the reaction mixture at various timepoints as calculated by HPLC (calibration adjusted concentrations).

Time (min)	0	2	4	6	8	10	13	17	20	23	26
Total conc. (mM)	11.27	7.76	7.47	8.08	7.99	8.95	9.25	9.85	10.63	10.43	10.85

OD₆₀₀ measurements

To ensure that bacterial cells were in the resting state, and that the observed *in vivo* enzyme activity (substrate conversion to product(s)) was not a consequence of an increase in bacterial density, measurements at 600 nm (OD₆₀₀) were taken throughout the time course (Table S6.3). These controlled experiments (no substrate) enabled identification of the OD₆₀₀ values that remained relatively consistent throughout. Thus we could be confident that the *E. coli* cells were in a resting state and no further growth was occurring. OD₆₀₀ measurements were performed using the same reaction conditions used as the UVRR whole-cell monitoring experiments, but with the substrates omitted (*i.e.* benzonitrile/benzamide).

Table S6.3. Table stating the OD₆₀₀ values at various time points throughout the reaction monitoring process. Samples were diluted 1:5 prior to OD₆₀₀ analysis.

<i>E. coli</i> whole-cells	OD ₆₀₀ at time point				
	0 min	10 min	20 min	30 min	60 min
NHase	0.668	0.642	0.585	0.697	0.609
Amidase	0.360	0.360	0.341	0.377	0.335
NHase + Amidase	0.537	0.560	0.513	0.575	0.505

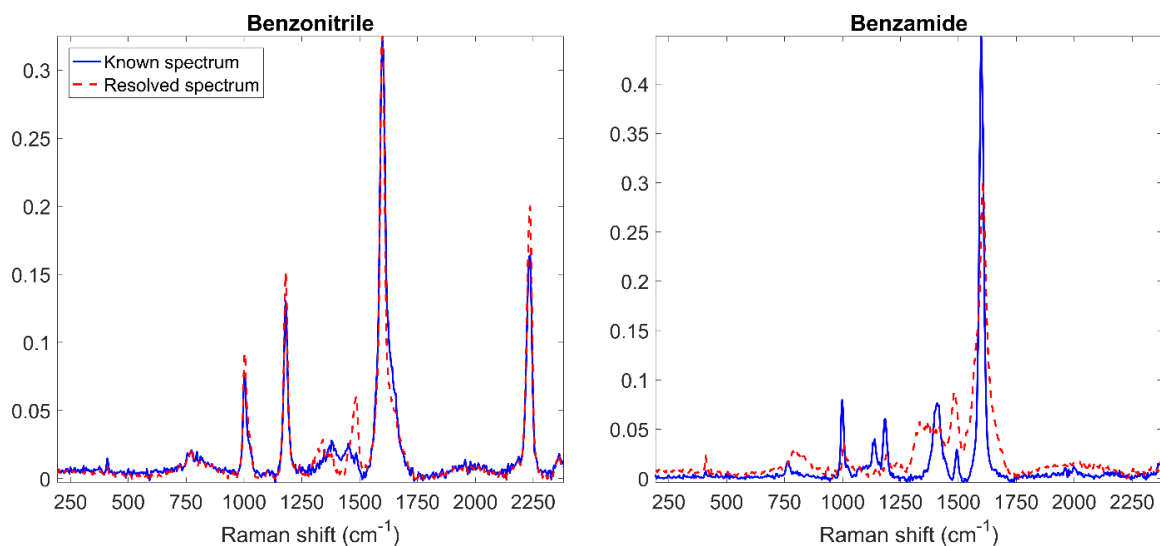


Figure S6.7. MCR-ALS deconvolved spectrum (dashed, red line) of individual analytes from the mixture compared with the known UVRR spectrum (solid, blue line). The initial reaction mixture includes benzonitrile (SM) and is initiated upon the addition of NHase-containing *E. coli* cells.

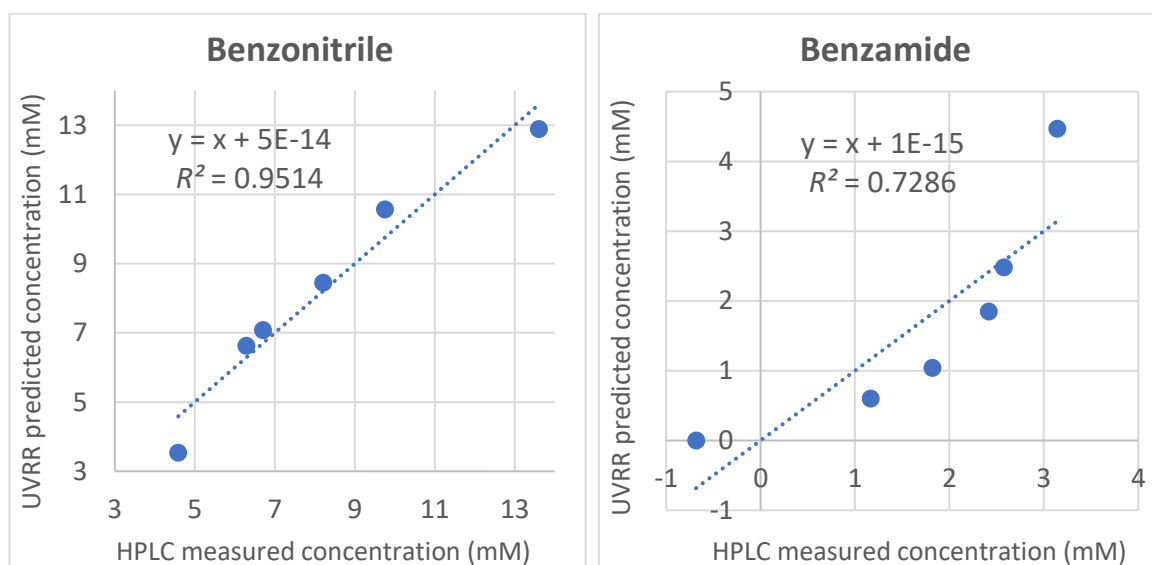
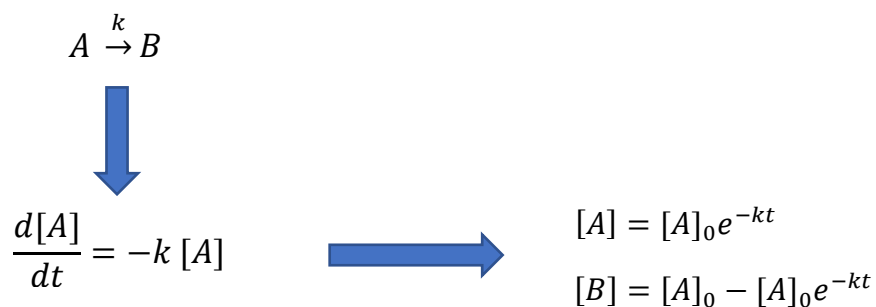


Figure S6.8. MCR-ALS predicted UVRR concentration versus HPLC measured concentration (mM). NHase whole-cell biotransformation gave R^2 values of 0.9514 and 0.7286 for benzonitrile (SM) and benzamide (P), respectively.



Note: Assuming that all A had converted to B!

Figure S6.9. Kinetic model used by MCR-ALS-HM approach to calculate the concentrations of each reactant at any time during the one-step biotransformations where A is benzonitrile or benzamide (as the product), B is benzamide (as the starting material) or benzoic acid and t is time. We assume that reaction followed a first order reaction in which A had converted to B at a reaction rate constant k.

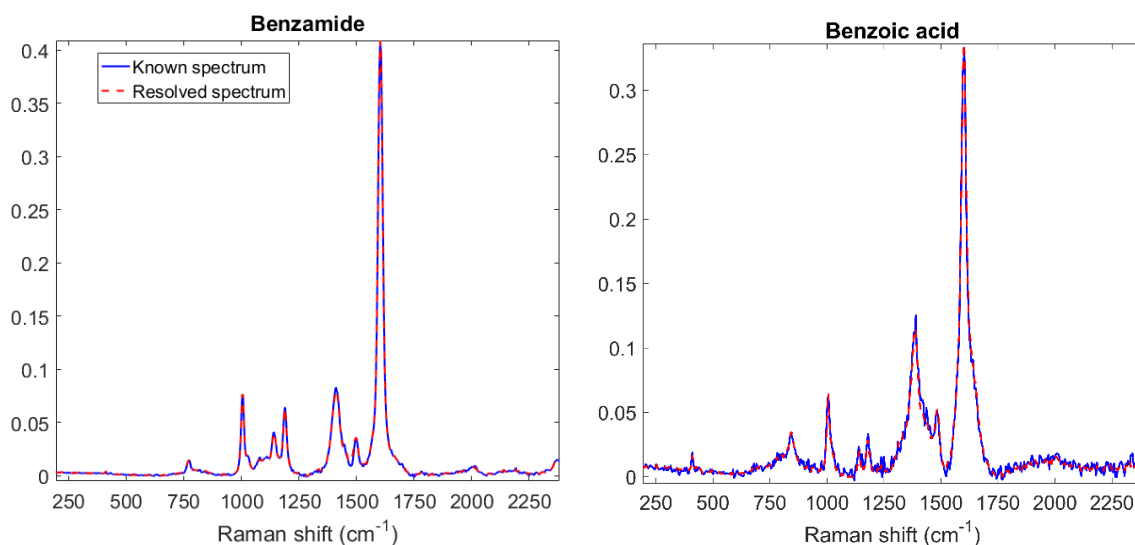


Figure S6.10. MCR-ALS deconvolved spectrum (dashed, red line) of individual analytes from the mixture compared with their known UVRR spectrum (solid, blue line). Initial reaction mixture includes benzamide (SM) and is started by the addition of amidase-containing *E. coli* cells.

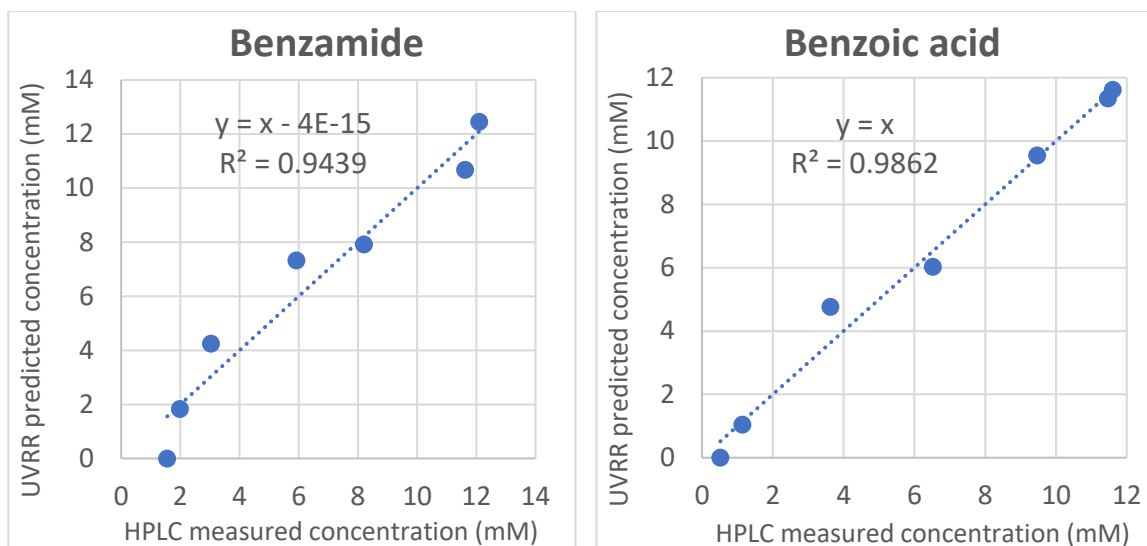


Figure S6.11. MCR-ALS predicted UVRR concentration versus HPLC measured concentration (mM). Amidase whole-cell biotransformation gave R^2 values of 0.9439 and 0.9862 for benzamide (SM) and benzoic acid (P), respectively.

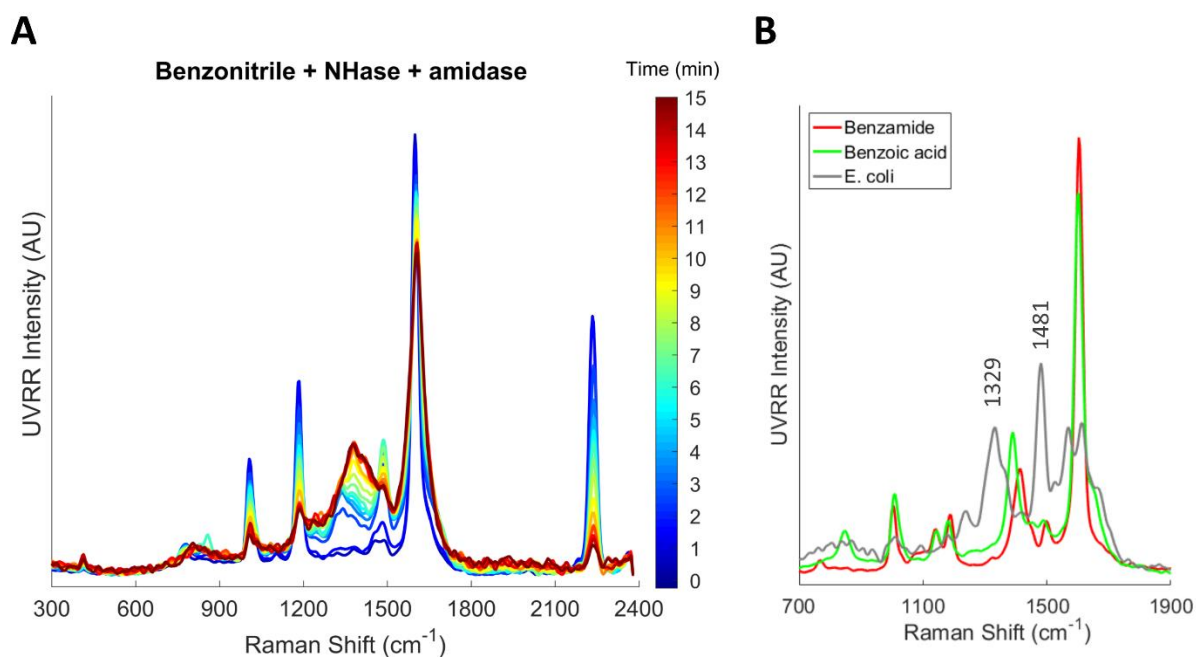


Figure S6.12. (A) UVRR spectra of in vivo biotransformation involving benzonitrile (SM) to benzamide (I) then benzoic acid (P) using NHase and amidase from within whole-cells. Colour bar represents the changing UVRR response with time. (B) UVRR response of standards: benzamide (red), benzoic acid (green) and *E. coli* BL21 (DE3) cells (grey) in potassium phosphate buffer (2.5×10^{-2} M, pH 7.2). Bacterial peaks at 1329 and 1481 cm^{-1} coincide with characteristic benzamide (1413 cm^{-1}) and benzoic acid (1389 cm^{-1}) peaks.

PROTEIN EXPRESSION AND PURIFICATION

Cloning and Expression.

Nitrile Hydratase. Constructs containing the genes encoding the α - and β - subunits of the nitrile hydratase (NHase) from *Rhodococcus equi* TG328-2 in pET21a and the associated NHase activator in pET28a were generously provided by the Uwe Bornscheuer group (Greifswald University) (Rzeznicka *et al.*, 2010). The pET28a construct containing the activator gene was transformed into *E. coli* BL21 (DE3) using kanamycin (50 $\mu\text{g}/\text{mL}$) selection. The resulting cells were then transformed with the pET21a construct containing both the α - and β -NHase subunits using kanamycin (50 $\mu\text{g}/\text{mL}$) and ampicillin (100 $\mu\text{g}/\text{mL}$) for selection. *E. coli* BL21 (DE3) cells transformed with recombinant NHase (α - and β -subunits) + activator plasmids were initially grown overnight at 37 °C in LB medium containing kanamycin (50 $\mu\text{g}/\text{mL}$) and ampicillin (100 $\mu\text{g}/\text{mL}$), before 100-fold dilution in 2x YT medium containing kanamycin (50 $\mu\text{g}/\text{mL}$) and ampicillin (100 $\mu\text{g}/\text{mL}$). Cultures were subsequently incubated at 37 °C with shaking until an $\text{OD}_{600} = 0.8$. Incubation was continued at 18 °C with shaking until an $\text{OD}_{600} = 1.0$, prior to induction with IPTG (1×10^{-4} M) and incubation at 18 °C with shaking for a further 20 h. After this time, cells were harvested by centrifugation at 4000 xg , 4 °C for 10 min. Pelleted cells were stored at 4 °C until purification or resuspension in potassium phosphate buffer (2.5×10^{-2} M, pH 7.2) for *in vivo* reactions.

Amidase. A synthetic *E. coli* codon optimised gene for an enantioselective amidase was purchased from Genewiz (US) using the nucleotide sequence as previously published by Trott (Trott *et al.*, 2002) from *Rhodococcus erythropolis* strain MP50. The amidase gene was sub-cloned into pET28a(+) vector containing N-terminal His-tag using restriction sites, *Hind*III and *Nde*I. The amidase encoding gene was amplified using *E. coli* DH5 α competent cells, following the *QIAprep Spin Miniprep* Kit Protocol. For protein expression, the pET28a(+) containing amidase plasmid was transformed into BL21 (DE3) competent cells, using kanamycin (50 $\mu\text{g}/\text{mL}$) for selection. *E. coli* BL21 (DE3) cells transformed with recombinant amidase plasmid were initially grown overnight at 37 °C in LB medium containing kanamycin (50 $\mu\text{g}/\text{mL}$). The resultant culture was diluted 100-fold in LB with kanamycin (50 $\mu\text{g}/\text{mL}$) and grown at 37 °C with shaking until reaching an $\text{OD}_{600} = 0.6$. IPTG (1×10^{-4} M) was then added to the culture for induction and grown at 18 °C with shaking for 20 h. After this time, cells were harvested by centrifugation at 4000 xg , 4 °C for 10 min. Pelleted cells were stored at 4 °C until purification or resuspension in potassium phosphate buffer (2.5×10^{-2} M, pH 7.2) for *in vivo* reactions.

Protein Purification.

Cell pellets harvested from the *E. coli* protein expression were resuspended in imidazole-containing loading buffer (NHase buffer containing TRIS HCl buffer (5×10^{-2} M), NaCl (1×10^{-1} M), butanoic acid (4×10^{-2} M) and imidazole (1×10^{-2} M), pH 7.5, amidase buffer containing phosphate (5×10^{-2} M), NaCl (5×10^{-1} M) and imidazole (1×10^{-2} M), pH 7.2). Cells were disrupted by sonication and the lysate was clarified by centrifugation (4°C , 40 min, $10,000 \times g$). The soluble extract was loaded onto Ni-NTA column (Qiagen) and the column washed with purification buffer containing 3×10^{-2} M imidazole (NHase) or 1×10^{-2} M and 6×10^{-2} M imidazole (amidase). Purified NHase and amidase were eluted using 3×10^{-1} M imidazole concentration with their respective buffers. Protein samples were subjected to buffer exchange with either TRIS HCl buffer (1×10^{-1} M) containing butanoic acid (4×10^{-2} M) pH 7.5 for NHase, or potassium phosphate buffer (1×10^{-1} M, pH 7.2) for amidase, using spin concentration (Vivaspin 20 centricon, 10,000 MWCO) before subsequent storage at -20°C . The final protein concentration was determined using Thermo Scientific NanoDrop 2000 spectrophotometer. SDS-PAGE was used to separate proteins according to their size, enabling protein identification and gave an idea of purity. SDS-PAGE gels for each protein can be seen in Figure S6.13 and S6.14.

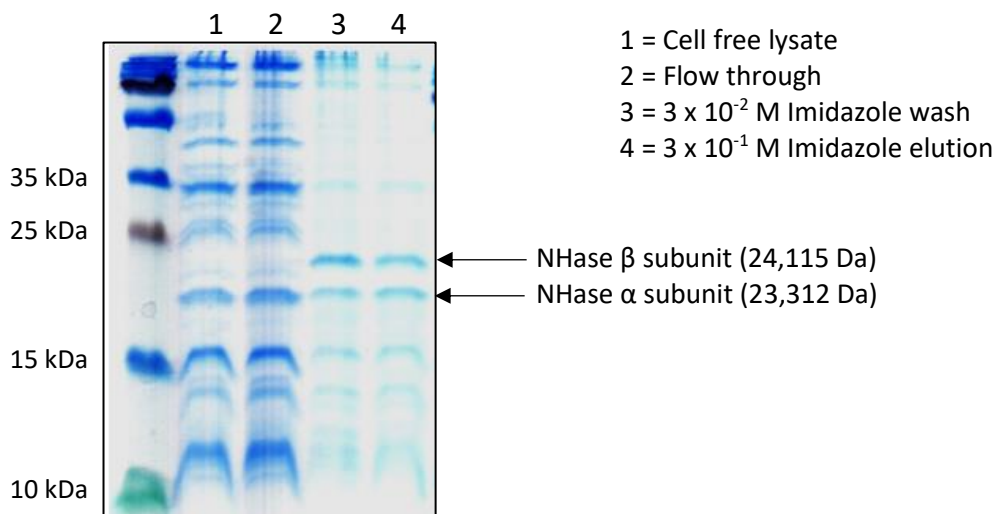


Figure S6.13. SDS-PAGE (16 %) of NHase protein expression using Ni-NTA. Both the α and β subunits can be clearly visualised.

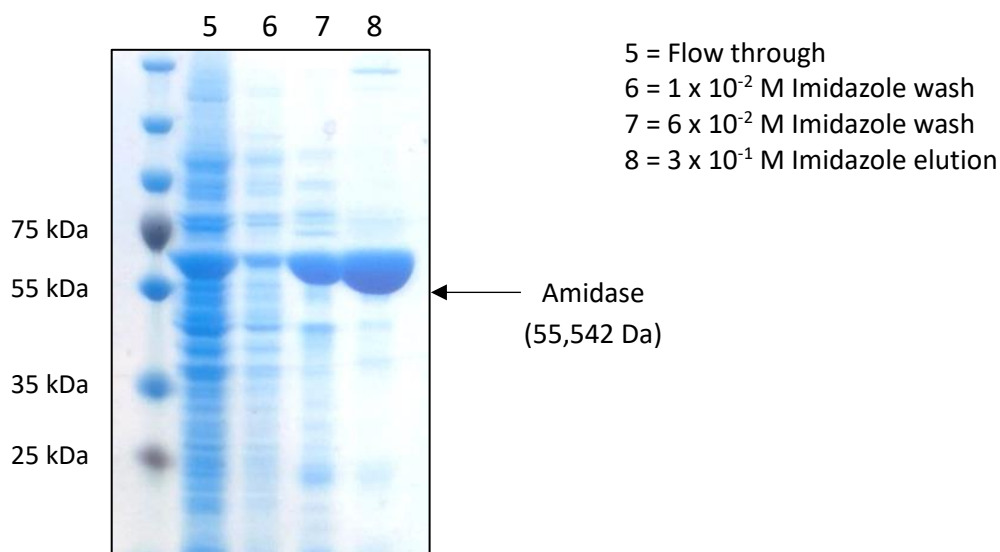


Figure S6.14. SDS-PAGE of amidase protein expression using Ni-NTA.

REFERENCES

- BRITTAI, H. G. 2009. Vibrational Spectroscopic Studies of Cocrystals and Salts. 1. The Benzamide–Benzoic Acid System. *Crystal Growth & Design*, 9, 2492-2499.
- CHATTERJEE, C. L., GARG, P. & JAISWAL, R. M. P. 1978. Vibrational spectrum and thermodynamic functions of para methylbenzonitrile molecule. *Spectrochimica Acta Part A: Molecular Spectroscopy*, 34, 943-947.
- GAO, X., DAVIES, J. P. & WEAVER, M. J. 1990. Test of surface selection rules for surface-enhanced Raman scattering: the orientation of adsorbed benzene and monosubstituted benzenes on gold. *The Journal of Physical Chemistry*, 94, 6858-6864.
- MROZEK, M. F., WASILESKI, S. A. & WEAVER, M. J. 2001. Periodic Trends in Electrode–Chemisorbate Bonding: Benzonitrile on Platinum-Group and Other Noble Metals As Probed by Surface-Enhanced Raman Spectroscopy Combined with Density Functional Theory. *Journal of the American Chemical Society*, 123, 12817-12825.
- RZEZNICKA, K., SCHÄTZLE, S., BÖTTCHER, D., KLEIN, J. & BORNSCHEUER, U. T. 2010. Cloning and functional expression of a nitrile hydratase (NHase) from *Rhodococcus equi* TG328-2 in *Escherichia coli*, its purification and biochemical characterisation. *Applied Microbiology and Biotechnology*, 85, 1417-1425.
- SPARROW, M. C., JACKOVITZ, J. F., MUNRO, C. H., HUG, W. F. & ASHER, S. A. 2001. New 224 nm Hollow Cathode Laser-UV Raman Spectrometer. *Applied Spectroscopy*, 55, 66-70.
- TROTT, S., BÜRGER, S., CALAMINUS, C. & STOLZ, A. 2002. Cloning and Heterologous Expression of an Enantioselective Amidase from *Rhodococcus erythropolis* Strain MP50. *Applied and Environmental Microbiology*, 68, 3279-3286.
- TROUT, C. C., TAMBACH, T. J. & KUBICKI, J. D. 2005. Correlation of observed and model vibrational frequencies for aqueous organic acids: UV resonance Raman spectra and molecular orbital calculations of benzoic, salicylic, and phthalic acids. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 61, 2622-2633.
- WESTLEY, C., FISK, H., XU, Y., HOLLYWOOD, K. A., CARNELL, A. J., MICKLEFIELD, J., TURNER, N. J. & GOODACRE, R. 2017. Real-Time Monitoring of Enzyme-Catalysed Reactions using Deep UV Resonance Raman Spectroscopy. *Chemistry – A European Journal*, 23, 6983-6987.

7 DISCUSSION

7.1 Discussion and Future Perspectives

The research presented throughout this thesis has demonstrated that Raman spectroscopy has successfully enabled rapid and sensitive measurements of enzyme activity, and in some cases, permitted reliable quantitative determinations – including on-line assessments. Enhancement techniques, such as SERS and UVR, have been the focus in this work, as normal Raman responses were too weak to generate reasonable signals and this is because of the concentrations used in these biotransformations. Several different classes of enzyme were studied, and there are some important conclusions that can be drawn from the work as a whole.

This research has demonstrated the capabilities of Raman spectroscopic techniques to detect analytes of interest directly; *i.e.* this method is label-free and does not require reporter molecules. Numerous biotransformations have been studied throughout, including flavin-dependent halogenases (Chapter 2 and 3), protease (α -chymotrypsin, Chapter 4), esterase (PLE, Chapter 4), oxidase (xanthine oxidase, Chapter 5), nitrile hydratase (Chapter 5 and 6) and an amidase (Chapter 6). Each of these biotransformations involve different aromatic substrates with varying functional group moieties, including nitriles, carboxylic acids, amines, alcohols, etc. Through the application of either SERS or UVR, and optimisation of the associated condition parameters, we have successfully characterised their corresponding substrates and/or product(s) directly.

As Raman spectroscopy offers characteristic chemical information that is molecule specific, it was important to address its sensitivity at distinguishing between structurally very similar compounds (including regioisomers and enantiomers). It is of great importance to be able to distinguish between structural isomers/stereoisomers as the different forms can have a significant impact on the bioavailability and reactivity of a drug molecule, as well as inducing adverse effects. SERS investigations could successfully detect subtle differences arising from differences in the regiospecific position of halogen atoms around small aromatic structures, as demonstrated in Chapter 2. In addition, the development of a novel, pseudo-enantiomer type-approach (Chapter 4) facilitated observations into an enzymes enantioselectivity using SERS, easily distinguishing between the two enantiomeric forms. Similarly, UVR proved to be very sensitive as it easily discriminated between the different oxidised states of purine structures (hypoxanthine, xanthine and uric acid, Chapter 5).

Moreover, both enhancement techniques utilised in this research could be performed as on-line (*i.e.* real-time) or at the very least, at-line (*i.e.* quick analysis of external samples) methods. This enabled rapid analysis (10 – 20 s) of the biotransformation samples, which is an essential parameter for use as an alternative HTS assay. On-line measurements completely avoid the need to modify the system prior to analysis, *i.e.* quenching and removing protein content. Not only does this

increase the efficiency by reducing the overall analysis time, but it also improves accuracy by avoiding errors incurred through sample handling and transfers (Workman *et al.*, 2007).

Finally, throughout these investigations, the pairing of chemometric analysis with Raman has aided the interpretation and understanding of the data. Simple univariate methods, such as plotting the peak area of characteristic bands, as well as more complex, multivariate methods have been used to uncover trends. MCR-ALS (Chapter 4, 5 and 6) models have been used to quantify analytes successfully, generating results that are in excellent agreement with robust, off-line HPLC analysis. Furthermore, the application of hard modelling (MCR-ALS-HM, Chapter 6) revealed that comparative, low-throughput HPLC measurements were not a necessity for quantification. Further demonstrating the utility of Raman and its potential within industrial monitoring processes, such as PAT.

As presented in the thesis introduction (Chapter 1), it is extremely important to develop new and improved techniques that are amenable for high-throughput screening, particularly when huge libraries of enzyme variants are produced during mutagenic strategies. Directed evolution has played an important role in the development of enzymes and their increasing incorporation within synthetic routes, such as pharmaceutical synthesis. Current HTS methods often suffer from long acquisition times and require high concentrations (HPLC/NMR). Conversely, much faster methods with lower limits of detection are used (fluorometry/colorimetry/MS), however these techniques give very limited structural information surrounding the analyte(s) under investigation and are limited due to their requirement of a fluorescent/colorimetric reagent. Ultimately, advancements in HTS would lead to huge improvements within this field and further support the incorporation of biocatalysts within synthetic methodologies (Kumar and Clark, 2006, Packer and Liu, 2015). SERS and UVRR investigations have been very successful at demonstrating their potential as alternative HTS methods, with notable advantages: both techniques offer rapid analysis, using either a 10 or 20 s acquisition time, which is a significant improvement over 10 – 24 min HPLC analysis (in the studies presented). Moreover, throughout all SERS investigations portable instrumentation was used, showcasing its versatility as an inexpensive instrument capable of excellent reproducibility and sensitivity.

Chapter 2 explored SERS to discriminate between regioisomers of small, halogenated aromatic structures, primarily chloroanthranilic acid. This proved to be very effective at identifying subtle spectral differences arising from the regiospecific position of a halogen atom around an aromatic ring. Combining chemometric analysis, such as PCA and its associated loadings plot, helped to identify these discreet spectral differences. These findings reiterate that SERS is a very sensitive and information rich technique, providing structural information that HPLC is not capable of. However,

the innate mode of action to attain SERS responses, *i.e.* analyte(s) experience favourable interactions with the metal nanoparticles, caused numerous complications when analysing complex halogenase reaction samples. Favourable interactions are analyte specific, and unless they have been previously documented within the literature, it is very much a trial and error approach to attain SERS responses, if at all. Ultimately, the analysis of halogenase reaction samples proved unsuccessful due to strong competition/favoured interactions of the cofactors (predominantly FAD) over the starting material and product(s).

This highlights a notable drawback of SERS: unless the reaction/analyte(s) have previously been investigated, then the discovery and optimisation of the SERS process can be extensive and time consuming, as there are many variable parameters (*i.e.* aggregating agent, metal, reducing agent, pH, etc). Alternatively, algorithms or fractional factorial designs have been effectively incorporated within experimental design to significantly reduce the quantity of experiments to achieve optimal responses (Fisk *et al.*, 2016). Both methods help to identify key parameters to test during the optimisation process, substantially reducing the number to experiments. Mabbott and colleagues present a prime example whereby the optimisation of mephedrone detection using SERS was reduced to 288 experiments from the original 1,722, by adopting a fractional factorial design (Mabbott *et al.*, 2013). Furthermore, the complexity (*i.e.* more than one analyte present) can complicate SERS analysis as it can lead to competition for the metal surface. If competition is experienced, positive interactions of the desired analyte(s) may be inhibited and hence obstruct/prevent SERS responses (as experienced in Chapter 2). It is these factors that may limit the widespread adoption of SERS within industrial screening processes.

To combat competition at the nanoparticle surface, as experienced by flavin-dependent halogenases in Chapter 2, whole-cell biotransformations were developed. Whole-cells are advantageous as they reduce analysis time by avoiding protein purification steps, along with benefiting from *in situ* recycling of essential cofactors (*i.e.* FADH₂ or NADH), ultimately leading to reductions in cost and complexity. Chapter 2 highlights that the application of whole-cells significantly reduced complexity by utilising cofactors that were innate to the cell. However, RadH (flavin-dependent halogenase amenable to whole-cell activity) suffered from low-level expression, hence a substantially high density of bacterial cells was required to achieve moderate activity (equivalent of 10 mL of bacterial culture resuspended in 1 mL reaction volume, halogenating >30 % chrysin at 5 x 10⁻⁴ M concentration). Ultimately, bands associated to the bacterial cells dominated the SERS response and led to difficulties identifying the substrate and product. Figure 7.1 demonstrates that if the quantity of bacterial culture is reduced (*i.e.* the number of cells), then chrysin (RadH substrate) becomes increasingly detectable as more characteristic SERS bands are

observed amongst the bacterial response. Therefore, with further optimisation of whole-cell conditions to increase protein expression, it would seem reasonable to suggest that SERS measurements would be successful towards halogenase biotransformations.

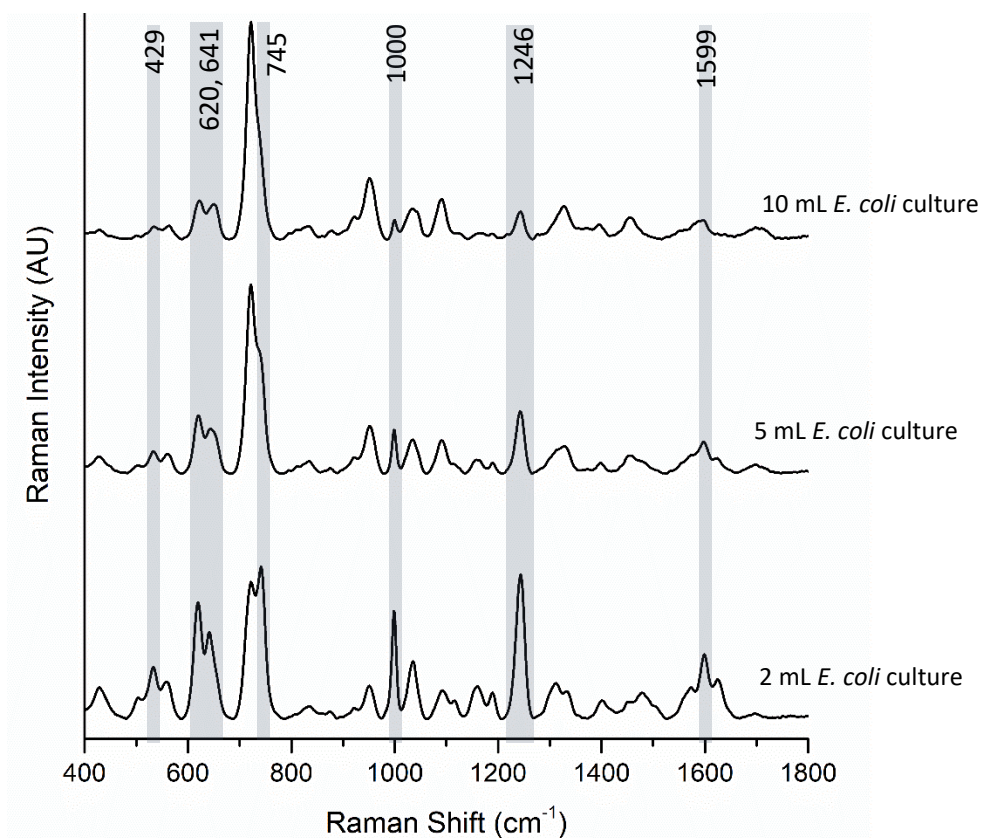


Figure 7.1. SERS spectra (averages of $n = 4$) of RadH biotransformation supernatant containing chrysin (2.5×10^{-5} M) and the equivalent of either 10 mL *E. coli* culture (top), 5 mL (middle) or 2 mL (bottom). Characteristic chrysin peaks have been highlighted in grey. Data have been baseline corrected and normalised.

Yan *et al.* recently showcased the capabilities of desorption electrospray ionization coupled with ion mobility mass spectrometry imaging (DiBT-IMMS) to image whole-cell bacterial colonies and semi-quantitatively measure product generation in real-time. Their results were noticeably improved when a halogen atom was incorporated within the analytes structure, as this permitted the natural isotope ratios to be detected, increasing the analytes signal within the complex biological matrix (Yan *et al.*, 2017). Perhaps this MS imaging approach would be suited towards the analysis of halogenase activity as the sensitive halogen atom is ‘installed’ by the enzyme. In addition, we have already demonstrated that RadH halogenase is compatible towards *in vivo* (whole-cell) activity, further supporting its fitness.

To actively combat competition at the nanoparticle surface within SERS analysis of halogenase samples, sulfur functionalisation was investigated (Chapter 3). Previous examples within the

literature have demonstrated that interactions experienced by a molecule and the metal nanoparticle surface can be substantially increased using thiol-modifications (Stewart and Bell, 2011, Gühlke *et al.*, 2016). Hence, halogenase active-substrates were designed to incorporate thiol-functionality to exploit its strong affinity towards silver. This proved efficacious, resulting in specific targeting of the thiol-functionalised substrate and its corresponding product. It was found that complex cofactors became a negligible contribution in the SERS spectra, as desired. However, this was accompanied by a complete loss of enzyme activity towards these substrates. Despite no-demonstrable enzyme activity in Chapter 3, these observations identified that thiols were extremely good functional groups to promote favourable and strong interactions in SERS, which supported our investigations in Chapter 4.

To examine the sensitivity of SERS further, having proved its effectiveness towards regioisomers, it was used to distinguish between enantiomers (structural mirror-images around a chiral centre). Previous attempts within the Goodacre group determined that a direct approach of enantiomer discrimination using SERS was not attainable (C. Westley's PhD thesis, 2017 'Raman Spectroscopy and its Enhancement Techniques for Direct Monitoring of Biotransformations'), therefore we developed a pseudo-enantiomer type-approach (Chapter 4). Chiral thioester substrates were designed so that subsequent enantioselective hydrolysis of the thioester would afford a thiol-containing product, employing advantageous interactions with the nanoparticles as determined in Chapter 3. The thiol-linker associated to the thioester bond was varied dependent on the enantiomeric form (L- versus D- and 1-propanethiol versus 2-propanethiol). Within this work, the sulfur-containing substrates (thioesters) were shown to be active towards two different classes of enzyme, α -chymotrypsin and PLE. Not only could we quantitatively monitor enzyme activity in real-time using SERS (MCR-ALS modelling with additional HPLC analysis), but enantioselectivity could be determined relative to the hydrolysed, detected thiol.

Nevertheless, this thiol-functionalisation approach experienced several complications. Firstly, thiols are prone to oxidation, forming disulfide bonds. We demonstrated (in Chapter 4) that instantaneous capture of thiols was achievable by introducing the nanoparticles into the biotransformation samples, thus avoiding oxidation. This also enabled real-time SERS measurements which is a novel concept for monitoring biotransformations in an on-line manner. However, this method is only suitable towards mixtures that contain un-modified reaction components that do not interact with the nanoparticles (*i.e.* no SERS response is generated). If other analytes interact (*i.e.* the starting material(s)), the nanoparticles may experience undesirable competition at the surface upon generation of the thiol product. In addition, the nanoparticles might also begin to aggregate out of solution before the product (thiol) can interact, resulting in

non-quantitative responses and problematic analysis. Secondly, our investigations revealed that PLE showed a preference towards the thiol linker of the thioester bond (*i.e.* the straight chain, 1-propanethiol, was favoured over the branched chain isomer, 2-propanethiol). This preference gave falsifications of the enzymes innate enantioselectivity. If a similar thioester approach was employed within the screening of mutagenic strategies (*i.e.* directed evolution to modify enantioselectivity), and a similar preference towards the pseudo-part of the thioesters was experienced, this would cause misrepresentations of activity and ultimately render the screen invalid. Thirdly, cysteine residues present on a protein's structure can participate in thiol-disulfide redox exchange reactions (Trivedi *et al.*, 2009). Commonly, colour-producing reagents, such as Ellman's reagent (5,5'-dithiobis-(2-nitrobenzoic acid) or DTNB), are employed to quantitatively determine the presence of a free thiol group in the cysteine residues (Ellman, 1959). More recently, sophisticated isotopic labelling paired with intact MS measurements have also been used to determine the redox-state of thiols in a protein (Thurlow *et al.*, 2016). Noticing that these reagents (*i.e.* Ellman's reagent) often contain free thiol groups, hence enabling the formation of a disulfide bond, it is likely that the thiol products produced upon enzymatic thioester hydrolysis are also likely to interact with cysteine residues (if there are cysteines within the proteins structure). This would likely render the thioester pseudo-enantiomer screen qualitative rather than quantitative if a portion of the thiol is interacting with the protein, rather than being detected at the nanoparticle surface.

Therefore, Chapter 4 has highlighted that pseudo-enantiomer substrates must be carefully designed to enable enzymatic activity but avoid linker preference, as well as identifying that many control experiments must be performed before it could be confirmed that the substrates are suitable. Again, this lengthy optimisation process may hinder its acceptability and use throughout industry and academia; however, little progress has been made towards methods that are capable of enantiomer discrimination. Hence, this proof-of-principle research illustrates that Raman could be an attractive alternative.

Once again, Chapters 5 and 6 explored real-time measurements of enzyme activity, yet this time utilising UVRR spectroscopy. UVRR measurements avoid the complications that are often associated with SERS as nanoparticles are not involved, thus the (tedious) optimisation steps are avoided. However, this technique requires analytes to contain a chromophore, and have electronic transitions that coincide with the laser frequency (244 nm in this case). Therefore, this limits its universal application as many molecules may not possess these qualities. On the other hand, most biological molecules possess such functionality, therefore supporting its suitability within biocatalysis and the pharmaceutical industry. Firstly, in Chapter 5, nitrile hydratase and xanthine oxidase biotransformations were studied using an open pot reaction set-up. MCR-ALS modelling,

combined with HPLC analysis confirmed that UVRR was extremely successful at monitoring the biotransformations, enabling reliable and reproducible quantitative measurements (Westley *et al.*, 2017). However, the monitoring process experienced an increase of noise over time, arising from the openness of the set-up (attributable to the removal of aliquots for HPLC analysis and evaporation of reaction components). As a result, we extended this approach in Chapter 6, to incorporate a flow-cell apparatus, which increased the signal-to-noise ratio, reducing the level of pre-processing when compared to the previous open vessel set-up. Development of this closed, flow system also established its suitability to be incorporated within PAT. Conventional Raman methods are frequently used within PAT for at-line or on-line analysis, which further supports the likelihood of UVRR being introduced into similar processes. UVRR lasers, and their associated optics, are comparatively expensive and are not as robust as normal Raman, so this could affect its adoption within industry and processes control. However, having demonstrated its aptness towards monitoring biological applications, plus this technique avoids additional reagents/solvents (*i.e.* HPLC) and full recovery of the reaction sample, the initial financial investment might not be so unappealing in the long-term.

Furthermore, in Chapter 6, we used additional hard modelling analysis combined with MCR-ALS (MCR-ALS-HM) to yield quantitative measurements, removing the necessity of external HPLC validation (provided that 100 % solubility of all reaction components is evidenced). *E. coli* whole-cells expressing amidase protein were highly successful at performing the biotransformation, with the substrate and product easily detectable within the complex biological mixture, despite their UVRR responses being highly similar. This example illustrates the sensitivity of UVRR, combined with chemometric analysis, to uncover subtle spectral differences. Therefore, it would seem reasonable to suggest that this approach could be suitable towards a wide range of analytes and biotransformations, both *in vitro* and *in vivo* sample types, provided a chromophore is present. Moreover, as we successfully demonstrated monitoring three analytes *in situ*, it would be interesting to explore this further by increasing the number of analytes under study (*i.e.* $n = 5, 6$, etc). Provided that characteristic bands for each analyte were present, this method could rival SERS' current multiplexing capabilities (Laing *et al.*, 2016).

Overall, the outcomes of the studies presented in this thesis have been extremely positive. Despite no publishable work on the first two research chapters encompassing halogenase activity (Chapters 2 and 3), this work helped develop the ensuing research. Initially tackling a complex and problematic biotransformation highlighted several features that needed to be overcome: simplification of the biotransformations (*i.e.* avoiding cofactors) and sulfur functionalisation to promote strong and favourable interactions (regarding SERS). These opening investigations also

demonstrated the power of SERS to discriminate between regioisomers, which subsequently inspired efficacious enantiomer investigations (Chapter 4). Experiments utilising UVRR enhancement were extremely effective and demonstrated that a simpler method (in comparison to SERS) could produce highly reliable and reproducible results in real-time.

Clearly, we have shown Raman's ability for monitoring enzymatic biotransformations and its use as an alternative HTS approach. As there is an unmet demand for rapid, universal HTS assays, especially when considering analytes involving regioisomers and enantiomers, the results within this thesis support the suitability of SERS and UVRR to meet these requirements. With further optimisation (*i.e.* improvements in whole-cell protein expression), increasing the number of biotransformations studied and reductions in associated Raman instrument costs, it is reasonable to believe that Raman will gain increasing interest and development within monitoring processes. To conclude, several novel Raman approaches have been developed, including real-time SERS analysis, regioisomer and enantiomer discrimination, along with on-line UVRR measurements of both *in vitro* (purified enzyme) and *in vivo* (whole-cells) biotransformations, ultimately affording (in some cases) quantitative determinations of substrate and/or product concentrations.

7.2 REFERENCES

- ELLMAN, G. L. 1959. Tissue sulfhydryl groups. *Archives of Biochemistry and Biophysics*, 82, 70-77.
- FISK, H., WESTLEY, C., TURNER, N. J. & GOODACRE, R. 2016. Achieving optimal SERS through enhanced experimental design. *Journal of Raman Spectroscopy*, 47, 59-66.
- GÜHLKE, M., HEINER, Z. & KNEIPP, J. 2016. Surface-Enhanced Raman and Surface-Enhanced Hyper-Raman Scattering of Thiol-Functionalized Carotene. *The Journal of Physical Chemistry C*, 120, 20702-20709.
- KUMAR, R. A. & CLARK, D. S. 2006. High-throughput screening of biocatalytic activity: applications in drug discovery. *Current Opinion in Chemical Biology*, 10, 162-168.
- LAING, S., GRACIE, K. & FAULDS, K. 2016. Multiplex *in vitro* detection using SERS. *Chemical Society Reviews*, 45, 1901-1918.
- MABBOTT, S., CORREA, E., COWCHER, D. P., ALLWOOD, J. W. & GOODACRE, R. 2013. Optimization of Parameters for the Quantitative Surface-Enhanced Raman Scattering Detection of Mephedrone Using a Fractional Factorial Design and a Portable Raman Spectrometer. *Analytical Chemistry*, 85, 923-931.
- PACKER, M. S. & LIU, D. R. 2015. Methods for the directed evolution of proteins. *Nature Reviews Genetics*, 16, 379-394.
- STEWART, A. & BELL, S. E. J. 2011. Modification of Ag nanoparticles with mixed thiols for improved SERS detection of poorly adsorbing target molecules: detection of MDMA. *Chemical Communications*, 47, 4523-4525.
- THURLOW, S. E., KILGOUR, D. P., CAMPOPIANO, D. J., MACKAY, C. L., LANGRIDGE-SMITH, P. R. R., CLARKE, D. J. & CAMPBELL, C. J. 2016. Determination of Protein Thiol Reduction Potential by Isotope Labeling and Intact Mass Measurement. *Analytical Chemistry*, 88, 2727-2733.

- TRIVEDI, M. V., LAURENCE, J. S. & SIAHAAN, T. J. 2009. The role of thiols and disulfides in protein chemical and physical stability. *Current Protein & Peptide Science*, 10, 614-625.
- WESTLEY, C., FISK, H., XU, Y., HOLLYWOOD, K. A., CARNELL, A. J., MICKLEFIELD, J., TURNER, N. J. & GOODACRE, R. 2017. Real-Time Monitoring of Enzyme-Catalysed Reactions using Deep UV Resonance Raman Spectroscopy. *Chemistry – A European Journal*, 23, 6983-6987.
- WORKMAN, J., KOCH, M. & VELTKAMP, D. 2007. Process analytical chemistry. *Analytical chemistry*, 79, 4345-4364.
- YAN, C., PARMEGGIANI, F., JONES, E. A., CLAUDE, E., HUSSAIN, S. A., TURNER, N. J., FLITSCH, S. L. & BARRAN, P. E. 2017. Real-Time Screening of Biocatalysts in Live Bacterial Colonies. *Journal of the American Chemical Society*, 139, 1408-1411.

8 APPENDIX I: Published Work in Original Format

WESTLEY, C.[†], FISK, H.[†], XU, Y., HOLLYWOOD, K. A., CARNELL, A. J., MICKLEFIELD, J., TURNER, N. J. & GOODACRE, R. 2017. Real-Time Monitoring of Enzyme-Catalysed Reactions using Deep UV Resonance Raman Spectroscopy. *Chemistry – A European Journal*, 23, 6983-6987.

[[†]] These authors contributed equally to this work.

Reaction Monitoring

Real-Time Monitoring of Enzyme-Catalysed Reactions using Deep UV Resonance Raman Spectroscopy

Chloe Westley^{+, [a]} Heidi Fisk^{+, [a]} Yun Xu,^[a] Katherine A. Hollywood,^[a] Andrew J. Carnell,^[b] Jason Mickfield,^[a] Nicholas J. Turner,^[a] and Royston Goodacre^{*, [a]}

Abstract: For enzyme-catalysed biotransformations, continuous *in situ* detection methods minimise the need for sample manipulation, ultimately leading to more accurate real-time kinetic determinations of substrate(s) and product(s). We have established for the first time an on-line, real-time quantitative approach to monitor simultaneously multiple biotransformations based on UV resonance Raman (UVR) spectroscopy. To exemplify the generality and versatility of this approach, multiple substrates and enzyme systems were used involving nitrile hydratase (NHase) and xanthine oxidase (XO), both of which are of industrial and biological significance, and incorporate multistep enzymatic conversions. Multivariate data analysis of the UVR spectra, involving multivariate curve resolution-alternating least squares (MCR-ALS), was employed to effect absolute quantification of substrate(s) and product(s); repeated benchmarking of UVR combined with MCR-ALS by HPLC confirmed excellent reproducibility.

Reaction monitoring based on analytical spectroscopy is broadly used to observe chemical changes in a variety of applications, including energy and fuel industries, bio-based technologies and processes, pharmaceuticals, as well as for biocatalyst discovery and optimization.^[1] Reaction monitoring provides essential information in terms of molecular speciation, and affords key insights into reaction mechanisms, kinetics and the biochemical process of the system investigated. Furthermore, real-time (in contrast to off-line) reaction monitoring greatly improves the efficiency and accuracy of the overall pro-

cess, with label-free spectroscopic-based methodologies being employed.^[2] Laborious sample preparation methods and purification steps are no longer required prior to analysis, thus minimising the need for transfers and sample handling, ultimately reducing errors. Advancements in engineering, such as the incorporation of robotics and sophisticated computational programs, lead to overall improvements and as a consequence, there is a significant reduction in the time taken for analysis.^[1, c, d, 3]

However, for biocatalytic applications, real-time reaction monitoring provides specific challenges: the sensitivity required to monitor conversions is often an issue, because low substrate concentrations are commonly used.^[4] As a result, monitoring conversions involving detection/presence of intermediates in multi-step biotransformations can be problematic. The most common method of measuring the rate of substrate turnover is the use of spectrophotometric assays.^[5] Although these assays are easy to use and interpret, a major limiting factor is the requirement for a fluoro-/chromo-genic reporter. However, in most cases, this means that the activity of the enzyme is detected indirectly or that improved enzyme activities may be selected based on the use of an idealised substrate, which may not translate to the real one. Whilst other spectroscopic and spectrometric physicochemical techniques are commonly employed (*viz.*, NMR, HPLC and LC-MS), these

[a] C. Westley,⁺ H. Fisk,⁺ Dr. Y. Xu, Dr. K. A. Hollywood, Prof. J. Mickfield, Prof. N. J. Turner, Prof. R. Goodacre
School of Chemistry, Manchester Institute of Biotechnology
University of Manchester, 131 Princess street, Manchester, M1 7DN (UK)
E-mail: roy.goodacre@manchester.ac.uk

[b] Dr. A. J. Carnell
Department of Chemistry, University of Liverpool
Liverpool, L69 7ZD (UK)

[*] These authors contributed equally to this work.

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/chem.201701388>.

© 2017 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

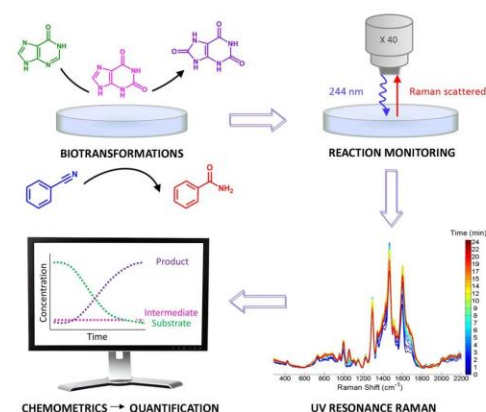


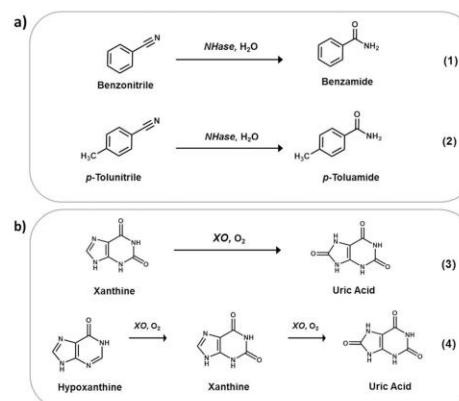
Figure 1. Workflow of the UVR approach for real-time reaction monitoring of multiple biotransformations.

methods too have notable drawbacks, such as extensive sample preparation, high equipment costs, large solvent volumes, long acquisition times, and in some instances provide limited structural information.^[5a,b,6] Therefore, there is a need for rapid, robust and reagent-free on-line high-throughput screening methods to overcome these significant drawbacks.

Raman spectroscopy presents itself as an ideal analytical technique to use for screening applications, because it is rapid, non-destructive and non-invasive. Moreover, it can be performed *in situ* in aqueous environments and provides molecular specific information. We have previously shown that the conversion of glucose to ethanol by yeast can be monitored by Raman spectroscopy with a NIR excitation wavelength.^[7] However, Raman scattering is a relatively weak physical phenomenon and is often further exacerbated by fluorescence interference when excitation involves lasers in the visible EM.^[8] As a consequence, enhancement techniques are regularly employed to increase scattering efficiency. Surface-enhanced Raman scattering (SERS), a surface-sensitive Raman enhancement technique, has previously been used to monitor enzymatic biotransformations indirectly.^[9] Very recently, we successfully demonstrated a >30-fold reduction in acquisition times for multiple enzymatic steps measuring analytes directly. This delivered high levels of accuracy and reproducibility, highlighting its suitability as an alternative screening technique.^[10] However, SERS requires a roughened metal surface that cannot be readily used for on-line assessment of enzymatic reactions, so at best is only suitable for at-line analysis.

Ultraviolet resonance Raman (UVRR) spectroscopy is a variant of "normal" Raman and involves the enhancement of Raman scattering by UV (in this case, at 244 nm). When the frequency of the laser coincides/matches the frequency of the molecule's electronic transition, enhancements of 10^3 – 10^5 can be observed.^[11] UVRR is an attractive technique for use in screening applications as the biotransformation(s) can be performed in real time; with no interference from background fluorescence (there is no fluorescence below 260 nm excitation).^[12] Moreover, the ability to measure analytes of interest directly without the need to quench the system, or have additional reagents as needed for SERS, is advantageous. Although this technique requires the absorption of laser light by chromophores in the UV region (most notably, from aromatics and fused ring systems), many complex biological systems fulfil this requirement, with nucleic acids and amino acids being particularly amenable to UVRR.^[13]

Herein, we demonstrate how UVRR can be used for real-time reaction monitoring using two different biocatalytic reactions (Figure 1). First, we focus on the conversion of nitriles to their corresponding amides using nitrile hydratase (NHase) (Scheme 1a), a class of enzyme extensively used in chemical synthesis within various industries—with acrylamide, nicotinamide (vitamin B₃) and pyrazinamide (anti-tuberculosis agent) being notable examples.^[14] Second, to illustrate multiple reaction steps, we have applied the method to xanthine oxidase (XO) catalysed biotransformations (Scheme 1b). XO catalyses the oxidation of a wide range of substrates including purines



Scheme 1. Biotransformations (1–4) selected for monitoring by UVRR: (a) bioconversions of nitriles to the corresponding amides by nitrile hydratase (NHase); (b) oxidation of purines by xanthine oxidase (XO).

and xenobiotic compounds, with xanthine and hypoxanthine, its natural substrates, being the focus in this investigation.^[15]

The biotransformations of interest (1–4) are shown in Scheme 1. Reaction conditions for all biotransformations were optimised accordingly for UVRR monitoring (see methods in the Supporting Information for full details). For optimum UVRR spectra, a 20 second acquisition time and approximately 0.2 mW laser power at sample was required using an excitation wavelength in the deep UV at 244 nm. Characteristic UVRR

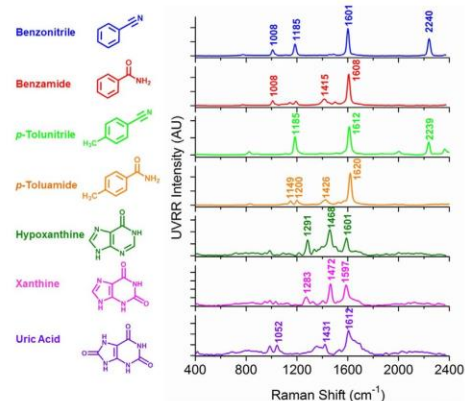


Figure 2. Average UVRR spectra ($n=5$) of each analyte for both biotransformations: benzotrile (blue), benzamide (red), *p*-tolunitrile (bright green) and *p*-toluamide (orange), hypoxanthine (green), xanthine (pink) and uric acid (purple). For NHase analytes, spectra were obtained at 12.5 mM, pH 7.2. For XO analytes, spectra were obtained at 0.75 mM, pH 7.6. All spectra are representative of starting reaction concentrations with characteristic peaks annotated. UVRR spectra were obtained for 20 s with baseline correction, normalisation and smoothing applied (see the Supporting Information, "Data processing" for full details).

spectra for each analyte, with unique peaks identified are summarised in Figure 2 (see Tables S1 and S2 in the Supporting Information for tentative band assignments). To monitor the enzyme-catalysed biotransformations, the instrument had to be modified and optimised (see Figure S1 in the Supporting Information); briefly, a magnetic stirrer plate was inserted below the turntable, with the reaction vessel (containing a magnetic stirrer bar) on top, focused under the microscope objective. The reaction was initiated upon the introduction of enzyme. Continuous stirring permitted maximal enzyme-substrate interaction throughout the reaction and provides a true representation of the conversion of substrate(s) to product(s). This set-up also allows the energy from the laser source to be evenly distributed over a much larger volume. Furthermore, to minimise the risk of reduced focus on the sample through solvent evaporation and removal of volume for HPLC analysis, the reaction was performed on a 10 mL scale. An initial concern was the integrity of the sample when subjected to a highly powered laser; however, no spectral changes (and hence, no photo-degradation) was observed throughout the reaction time course (see the Supporting Information, "Photo-degradation of sample" section and Figure S2). Interestingly, from these investigations, we observed bathochromic shifts (as a function of pH) for XO analytes (see the Supporting Information, section "Bathochromic shifts of XO analytes" and Figures S3–S5). Although there are characteristic peaks for each analyte, thus distinguishing starting material from product, the UVRR spectra were highly similar with many overlapping peaks (especially for XO analytes). Therefore, for all biotransformations, multivariate curve resolution-alternating least squares (MCR-ALS) was employed. MCR-ALS is a popular feature extraction tool for mixture analysis and was used to extract the necessary information (pure component spectra and corresponding concentrations) to predict absolute levels of the analytes within a mixture (see the Supporting Information, Figure S6 for a flow diagram summarising this MCR-ALS approach).¹⁷

We initially looked at the conversion of benzonitrile to benzamide (biotransformation 1) with <50% conversion achieved over a 20 minute time period. The deconvolved spectra for each analyte were highly similar to the UVRR spectra from the pure substrate and product (see the Supporting Information, Figure S7). For brevity purposes, we only represent the graphical results of one replicate. Figure 3a shows the UVRR spectra over the reaction time course, illustrating (by use of a colour bar) the increase and decrease of characteristic peaks with respect to time. Time points with both HPLC and UVRR data were used as the training set for the MCR-ALS model (i.e., HPLC was used as external validation—see SI Figure S8 for HPLC calibration). Time points with UVRR data (but without HPLC data) were used as the test set. As can be observed from Figure 3b, the UVRR predictions are in excellent agreement with the HPLC results, which is reflected by high R^2 values across all replicates, with an average of 0.964 and 0.983 for substrate and product, respectively (see Table 1). The coefficient of determination, R^2 , is the proportion of variability in a data set that is accounted for by a statistical model (in this

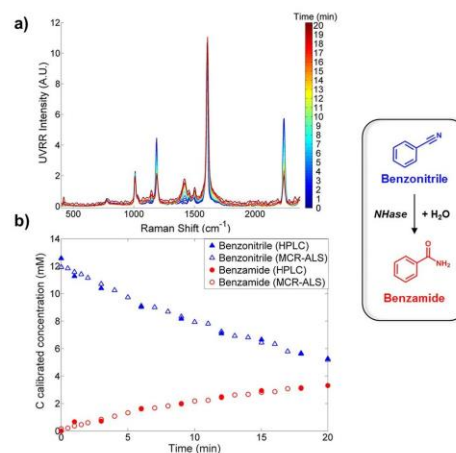


Figure 3. MCR-ALS model was applied to the UVRR data for the conversion of benzonitrile to benzamide (biotransformation 1). a) Accumulative spectra taken over the 20 min time course. The colour bar highlights each time point monitored with the start ($t=0$) in blue and the end point ($t=20$) in red. b) Reaction dynamics from real-time UVRR measurements (denoted by outlined symbols) and off-line HPLC data (denoted by solid symbols) as a function of time. UVRR spectra were obtained for 20 s with baseline correction, normalisation and smoothing applied (see the Supporting Information "Data processing" for full details). Data shown from replicate 2.

Table 1. Summary of the regression coefficients (R^2) across all five replicates for the two separate biotransformations.^[a]

Replicate	NHase (biotransformation 1)		XO (biotransformation 3)	
	Benzonitrile R^2	Benzamide R^2	Xanthine R^2	Uric acid R^2
1	0.959	0.993	0.990	0.955
2	0.987	0.987	0.954	0.977
3	0.969	0.982	0.916	0.965
4	0.962	0.980	0.973	0.987
5	0.942	0.973	0.957	0.983

[a] Biotransformation 1 (benzonitrile to benzamide) using NHase, with overall mean R^2 values of 0.964 and 0.983, respectively. Biotransformation 3 (xanthine to uric acid) using XO, with overall mean R^2 values of 0.958 and 0.973, respectively. These high R^2 (that are close to 1) indicate excellent fit.

case MCR-ALS) with R^2 values closer to one indicating an excellent fit. Notably, this experiment was conducted on five separate occasions, over a four-week period thus accounting for day-to-day instrument variance, ultimately demonstrating its robustness for on-line reaction monitoring.

Moreover, to extend this approach, we next investigated a similar NHase substrate, *p*-tolunitrile—only differing by a CH_3 group, yet possessing unique peaks compared to benzonitrile. Once again, with this biotransformation (biotransformation 2), we were able to monitor the reaction successfully. The UVRR and HPLC results were in very good agreement with one another, with typical R^2 values of 0.898 and 0.914 for *p*-tolunitrile

and *p*-toluamide, respectively (see the Supporting Information, Figure S9 and Table S3).

To demonstrate versatility of UVRR combined with MCR-ALS, we then performed analysis on a different, second enzyme system: XO (biotransformations 3 and 4). We have previously shown that we can monitor these conversions using SERS, and as has been already discussed, this involves the use of additional reagents preventing real-time monitoring.^[10] Therefore, this UVRR approach should overcome this main drawback. Furthermore, this enzyme system in itself provided a challenge with the analytes being highly similar in structure, only differing by additional carbonyl groups (Scheme 1b). First, the two-analyte conversion (biotransformation 3) of xanthine to uric acid was investigated, with >50% conversion achieved in 18 minutes. Adopting the same process, the MCR-ALS model was applied to the reaction data with results being in excellent agreement with the HPLC analysis (see Figure 4 and the Supporting Information, Figure S10 for deconvolved spectra of each analyte and Figure S11 for HPLC calibration). Average R^2 values of 0.958 and 0.973 were obtained for xanthine and uric acid, respectively (Table 1). We then extended this to a third analyte to include the precursor hypoxanthine (biotransformation 4), ultimately demonstrating the flexibility of this real time, on-line reaction monitoring screen for a more complex, multicomponent reaction system. The reaction conditions were modified slightly, with <40% conversion reached after 35 minutes. Again, MCR-ALS analysis was employed with the deconvolved UVRR spectra being highly consistent with the

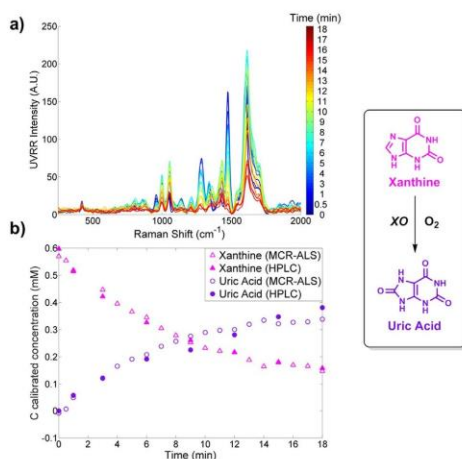


Figure 4. MCR-ALS model was applied to the UVRR data for the conversion of xanthine to uric acid, biotransformation 3. a) Accumulative spectra taken over the 18 min time course. The colour bar highlights each time point monitored with the start ($t=0$) in blue and the end point ($t=18$) in red. b) Reaction dynamics from real-time UVRR measurements (denoted by outlined symbols) and off-line HPLC data (denoted by solid symbols) as a function of time. UVRR spectra were obtained for 20 s with baseline correction, normalisation and smoothing applied (see the Supporting Information "Data processing" for full details). Data shown from replicate 1.

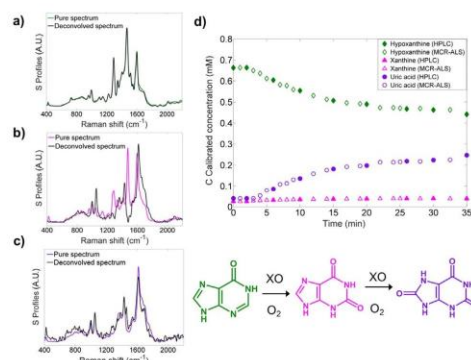


Figure 5. MCR-ALS model was applied to the UVRR data, in which it successfully deconvolved spectra into its pure components for biotransformation 4: a) hypoxanthine; b) xanthine; and c) uric acid. d) Reaction dynamics from real-time UVRR measurements (denoted by outlined symbols) and off-line HPLC data (denoted by solid symbols) as a function of time for the conversion of hypoxanthine to xanthine to uric acid. UVRR spectra were obtained for 20 s with baseline correction, normalisation and smoothing applied (see the Supporting Information "Data processing" for full details). Data shown from replicate 1.

pure spectra for each analyte (Figure 5a–c). The UVRR predictions were in very good agreement with the HPLC results (see Figure 5d). The R^2 values were slightly lower than biotransformation 3 (see the Supporting Information, Table S4), which was to be expected due to the increased complexity of the system, as well as the highly similar spectra between the three analytes. Noticeably, the R^2 value of xanthine was lower (biotransformation 4) than previous—this is due to the low overall concentration of xanthine (<8%) throughout the reaction. This was further supported by the proposed mechanism of XO (based on xanthine dehydrogenase, XDH, from *Rhodobacter capsulatus*), in which hypoxanthine binds to the active site and is converted to xanthine by oxidation at the C-2 position. Xanthine is then released, before binding in a different orientation to present the C-8 for oxidation to give uric acid.^[17] This means that the concentration of the intermediate remains low throughout.

Where this work could be further explored includes investigating the two separate pathways known to catalyse the conversion of nitrile containing compounds into their corresponding carboxylic acid: either in a single step (nitrilase) or a two-step process (nitrile hydratase and amidase; see the Supporting Information, Figure S12a). Fluorometric and colorimetric assays have previously been reported, including successful differentiation between the two pathways; however, only semi-quantitative analysis has been possible.^[18] Extending on biotransformation 1, we have shown that we can potentially use this UVRR approach to monitor such cascades as the corresponding carboxylic acid involved in this pathway has unique, characteristic peaks (see the Supporting Information, Figure S12b). Furthermore, other nitrile containing substrates, for example, 3-pyridinecarbonitrile and pyrazinecarbonitrile,

which are precursors for important pharmaceutical products, can similarly be monitored (see the Supporting Information, Figure S12c and S12d). These results further demonstrate the general utility of the UVRR approach for enzyme reaction monitoring.

In this study, we have developed a label-free, rapid, on-line screening method to monitor biological and industrially relevant biotransformations based on UVRR spectroscopy. To demonstrate the general utility of this approach, multiple substrates and enzyme systems were investigated, which included single, multiple and cascade enzyme systems. UVRR spectra acquisitions were rapid (20 s per measurement) and when combined with MCR-ALS produced substrate(s) and product(s) concentrations that were completely in agreement with off-line HPLC measurements. Additional benchmarking involved repeat biotransformations conducted over several weeks and this established the excellent reproducibility and robustness of this new analytical approach. In conclusion, we believe that additional optimisation and configuration of the UVRR instrument set-up will make this approach amenable to miniaturization and *in situ* point-and-shoot analyses,^[19] thus enhancing the potential for wider application. The method could also be developed as a high-throughput screening technique for enzyme activity, including the monitoring of cascade biotransformations, as well as for investigating enzyme inhibitors.

Acknowledgements

C.W. and H.F. are grateful to BBSRC for their PhD studentships. H.F. thanks Jonathan Latham for NHase insights. Y.X. thanks the Cancer Research UK for funding (including an Experimental Cancer Medicine Centre award). A.J.C. would like to acknowledge BBSRC for funding (BB/M028631/1). N.J.T., J.M. and R.G. are also indebted to BBSRC and GSK for financial funding (grant BB/K00199X/1). N.J.T. thanks the Royal Society for a Wolfson Research Merit Award.

Conflict of interest

The authors declare no conflict of interest

Keywords: biotransformations · on-line monitoring · quantification · reaction monitoring · UV resonance Raman (UVRR) spectroscopy

- [1] a) D. C. Hinz, *Anal. Bioanal. Chem.* **2006**, *384*, 1036–1042; b) J. Workman, M. Koch, D. Veltkamp, *Anal. Chem.* **2007**, *79*, 4345–4364; c) J.

- Workman, M. Koch, B. Lavine, R. Chrisman, *Anal. Chem.* **2009**, *81*, 4623–4643; d) M. T. Reetz, *J. Am. Chem. Soc.* **2013**, *135*, 12480–12496.
 [2] a) P. Gardner, S. A. Arnold, F. Brown, R. Carr, A. Nordon, L. M. Harvey, B. McNeil, *Anal. Chim. Acta* **2013**, *779*, 50–55; b) C. Yan, F. Parmeggiani, E. A. Jones, E. Claude, S. A. Hussain, N. J. Turner, S. L. Flitsch, P. E. Barran, *J. Am. Chem. Soc.* **2017**, *139*, 1408–1411.
 [3] T. Kourti, *Anal. Bioanal. Chem.* **2006**, *384*, 1043–1048.
 [4] M. R. Dadd, D. C. A. Sharp, A. J. Pettman, C. J. Knowles, *J. Microbiol. Methods* **2000**, *41*, 69–75.
 [5] a) R. A. Kumar, D. S. Clark, *Curr. Opin. Chem. Biol.* **2006**, *10*, 162–168; b) M. G. Acker, D. S. Auld, *Perspect. Sci.* **2014**, *1*, 56–73; c) M. S. Packer, D. R. Liu, *Nat. Rev. Genet.* **2015**, *16*, 379–394.
 [6] a) S. Fox, S. Farr-Jones, L. Sopchak, A. Boggs, J. Comley, *J. Biomol. Screening* **2004**, *9*, 354–358; b) O. von Ahlsen, U. Bömer, *ChemBioChem* **2005**, *6*, 481–490; c) J. L. Reymond, P. Babiak, *Adv. Biochem. Eng./Biotechnol.* **2007**, *105*, 31–58.
 [7] A. D. Shaw, N. Kaderbhai, A. Jones, A. M. Woodward, R. Goodacre, J. J. Rowland, D. B. Kell, *Appl. Spectrosc.* **1999**, *53*, 1419–1428.
 [8] T. Vankeirsbilck, A. Vercauteren, W. Baeyens, G. Van der Weken, F. Verpoort, G. Vergote, J. P. Remon, *TrAC Trends Anal. Chem.* **2002**, *21*, 869–877.
 [9] B. D. Moore, L. Stevenson, A. Watt, S. Flitsch, N. J. Turner, C. Cassidy, D. Graham, *Nat. Biotechnol.* **2004**, *22*, 1133–1138.
 [10] C. Westley, Y. Xu, A. J. Carnell, N. J. Turner, R. Goodacre, *Anal. Chem.* **2016**, *88*, 5898–5903.
 [11] S. A. Asher, *Anal. Chem.* **1993**, *65*, 59A–66A.
 [12] S. A. Asher, C. R. Johnson, *Science* **1984**, *225*, 311–313.
 [13] a) T. C. Strekas, T. G. Spiro, *Biochim. Biophys. Acta Protein Struct.* **1972**, *278*, 188–192; b) P. Carey, *Biochemical Applications of Raman and Resonance Raman Spectroscopies*, Academic Press, London, **2012**; c) L. Ashton, C. E. M. Hogwood, A. S. Tait, J. Kuligowski, C. M. Smales, D. G. Bracewell, A. J. Dickson, R. Goodacre, *J. Chem. Technol. Biotechnol.* **2015**, *90*, 237–243.
 [14] a) Y. Asano, T. Yasuda, Y. Tani, H. Yamada, *Agric. Biol. Chem.* **1982**, *46*, 1183–1189; b) M. Kobayashi, T. Nagasawa, H. Yamada, *Trends Biotechnol.* **1992**, *10*, 402–408; c) J. Mauger, T. Nagasawa, H. Yamada, *J. Biotechnol.* **1988**, *8*, 87–95; d) T. Nagasawa, C. D. Mathew, J. Mauger, H. Yamada, *Appl. Environ. Microbiol.* **1988**, *54*, 1766–1769; e) O. Ghisalba, H.-P. Meyer, R. Wohlgemuth, M. C. Flickinger in *Encyclopedia of Industrial Biotechnology*, Wiley, Hoboken, **2009**.
 [15] a) R. Hille, *Chem. Rev.* **1996**, *96*, 2757–2816; b) R. Hille, *Arch. Biochem. Biophys.* **2005**, *433*, 107–116; c) R. Harrison, *J. Free Radicals Biol. Med.* **2002**, *33*, 774–797.
 [16] a) R. Tauler, *Chemom. Intell. Lab. Syst.* **1995**, *30*, 133–146; b) J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, *Chemom. Intell. Lab. Syst.* **2005**, *76*, 101–110.
 [17] U. Dietzel, J. Kuper, J. A. Doebbler, A. Schulte, J. J. Truglio, S. Leimkühler, C. Kisker, *J. Biol. Chem.* **2009**, *284*, 8768–8776.
 [18] a) C. Reisinger, F. van Assema, M. Schürmann, Z. Hussain, P. Remler, H. Schwab, *J. Mol. Catal. B* **2006**, *39*, 149–155; b) Z. J. Lin, R.-C. Zheng, L.-H. Lei, Y.-G. Zheng, Y.-C. Shen, *J. Microbiol. Methods* **2011**, *85*, 214–220; c) L. M. L. Angelini, A. R. M. da Silva, L. D. F. C. Rocco, C. D. D. F. Milagre, *Braz. J. Microbiol.* **2015**, *46*, 113–116.
 [19] D. I. Ellis, H. Muhamadali, S. A. Haughey, C. T. Elliott, R. Goodacre, *Anal. Methods* **2015**, *7*, 9401–9414.

Manuscript received: March 28, 2017

Accepted Article published: March 29, 2017

Final Article published: May 2, 2017

9 APPENDIX II: Additional Publications

SHEPHERD, S. A., MENON, B. R., FISK, H., STRUCK, A. W., LEVY, C., LEYS, D., MICKLEFIELD, J.,
2016. A Structure-Guided Switch in the Regioselectivity of a Tryptophan Halogenase.
ChemBioChem, 17, 821-824

A Structure-Guided Switch in the Regioselectivity of a Tryptophan Halogenase

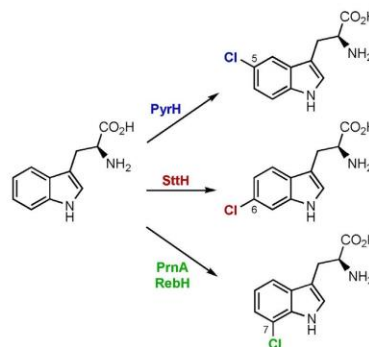
Sarah A. Shepherd, Binuraj R. K. Menon, Heidi Fisk, Anna-Winona Struck, Colin Levy, David Leys, and Jason Micklefield*^[a]

Flavin-dependent halogenases are potentially useful biocatalysts for the regioselective halogenation of aromatic compounds. Haloaromatic compounds can be utilised in the synthesis and biosynthesis of pharmaceuticals and other valuable products. Here we report the first X-ray crystal structure of a tryptophan 6-halogenase (SttH), which enabled key residues that contribute to the regioselectivity in tryptophan halogenases to be identified. Structure-guided mutagenesis resulted in a triple mutant (L460F/P461E/P462T) that exhibited a complete switch in regioselectivity; with the substrate 3-indolepropionate 75% 5-chlorination was observed with the mutant in comparison to 90% 6-chlorination for the wild-type SttH. This is the first clear example of how regiocomplementary halogenases can be created from a single parent enzyme. The biocatalytic repertoire of SttH was also expanded to include a range of indolic and non-indolic substrates.

Enzymes that can catalyse the regioselective halogenation of aromatic substrates could provide an attractive alternative to the traditional halogenation methods that are commonly used in synthesis. Halogenated aromatic compounds find extensive synthetic applications, particularly in transition-metal-catalysed crosscoupling reactions,^[1] and are important constituents of pharmaceuticals,^[2] agrochemicals^[3] and other valuable materials.^[4] Despite this, the traditional methods of producing haloaromatic compounds utilise harsh reaction conditions and often require harmful reagents, catalysts and solvents. These non-enzymatic methods also lack regiocontrol resulting in unwanted by-products that can be difficult to separate and are problematic to dispose of owing to their toxicity or persistence in the environment.^[5] Consequently, there has been major interest in harnessing nature's halogenases, which employ benign inorganic halides in aqueous media, to effect cleaner and more regioselective halogenation reactions.

The first halogenating enzyme to be identified was the chloroperoxidase from the fungus, *Caldariomyces fumago*, which lacked regiocontrol owing to the free hypochlorous acid (HOCl) produced as the halogenating agent. Further examples of both haem- and vanadium-dependent haloperoxidases were later identified, which also generally lacked substrate specificity and regioselectivity.^[6] More recently, Fe²⁺/α-ketoglutarate (αKG)-dependent and flavin-dependent halogenases, which effect the regioselective halogenation of precursors in the biosynthesis of a wide range of halogenated natural products, have been identified.^[7] Many of the αKG- and flavin-dependent halogenases utilise substrates that are tethered to the carrier proteins of biosynthetic assembly-line enzymes thus making their use as biocatalysts limited. However, there are a number of flavin-dependent tryptophan halogenases that can regioselectively halogenate free tryptophan, and therefore have more potential for synthetic purposes.^[8]

Previously, X-ray structures have been elucidated for the tryptophan 7-halogenases, PrnA^[9] and RebH,^[10] as well as a tryptophan 5-halogenase PyrH (Scheme 1).^[11] These structures provided insights into the mechanism and regiocontrol of flavin-dependent halogenases. It is suggested, that the halogenases utilise O₂ to oxidise FADH₂ giving C4a-hydroperoxyflavin, which then reacts with chloride to produce HOCl. It is then proposed that HOCl reacts with an active site lysine to generate a chloramine, which chlorinates the substrate.^[9,12] Although the position of the active-site lysine relative to the substrate is likely to be important in determining the regiochemical out-



Scheme 1. Reactions of flavin-dependent halogenases with tryptophan, and their respective products.

[a] Dr. S. A. Shepherd, Dr. B. R. K. Menon, H. Fisk, Dr. A.-W. Struck, Dr. C. Levy, Prof. D. Leys, Prof. J. Micklefield
School of Chemistry and Manchester Institute of Biotechnology
The University of Manchester
131 Princess Street, Manchester, M1 7DN (UK)
E-mail: Jason.micklefield@manchester.ac.uk

Supporting information for this article can be found under <http://dx.doi.org/10.1002/cbic.201600051>.

© 2016 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

come of the reactions, the factors that effect the regiocontrol of these enzymes are still not fully understood. For example, using the PrnA and PyrH X-ray crystal structures,^[9,11] Lang et al. attempted to switch the regioselectivity of PrnA to that of PyrH by targeted mutagenesis.^[13] However, of all the mutants tested only one mutation, F103A, had any effect on the regioselectivity of PrnA. The F103A mutant showed a modest change in regioselectivity with bromide giving a 2:1 mixture of 7- and 5-bromotryptophan, whereas the wild-type PrnA gives exclusively 7-bromotryptophan. This shift to produce 33% 5-bromotryptophan falls some way short of the change that would be required to create a new regiocomplementary enzyme. Here we describe the first structure of a tryptophan 6-halogenase, SttH, which provides further insights into the factors affecting the regioselectivity of these flavin-dependent halogenases. Moreover, these structural insights were used to expand the biocatalytic repertoire of this enzyme, and to guide mutagenesis leading to a complete switch in the regioselectivity from 90% chlorination at the 6-position to 75% in favour of chlorination at the 5-position of 3-indolepropionic acid.

Given that there is no structure for a tryptophan 6-halogenase, we explored expression of a number of candidate enzymes for crystallography trials. From this, we found that SttH from *Streptomyces toxytricini* gave good expression in *Escherichia coli* and catalysed the halogenation of tryptophan to give exclusively 6-chlorotryptophan, as reported previously.^[14] The X-ray structure of SttH was then determined at 2.7 Å (Figure 1)

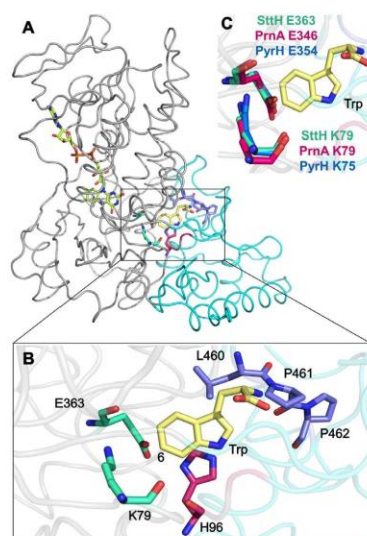


Figure 1. A) Crystal structure of SttH (PDB ID: 5HY5) showing typical box (grey) and triangular pyramid (cyan) of flavin-dependent tryptophan halogenases. B) Selected active-site residues of SttH with tryptophan modelled. C) Overlay of catalytic lysine and glutamate in PyrH, SttH and PrnA.

to reveal a dimer, with each monomer exhibiting a box and triangular pyramid structure, as previously observed with PrnA, RebH and PyrH.^[9–11] Within the box structure are the conserved flavin-dependent tryptophan halogenase sequences GxGxxG and WxWxIP. At the interface with the triangular pyramid, are the catalytic residues K79 and E363, which align with the active-site lysine and glutamate of PrnA and PyrH (Figure 1 C).^[14,15] In addition, SttH residues H96 and F98 are positioned for π -stacking with the indole moiety of the substrate (Figure S2), whilst the P97 carbonyl and Y463 hydroxy groups can potentially hydrogen bond with the indole NH (Figure S3).

From sequence alignments it is evident that SttH is more like PyrH than PrnA, with insertions present in PyrH and SttH between residues SttH 155 and 167 and a deletion between SttH 457 and 464 compared with PrnA (Figure S1). These subtle differences around the active site of the enzymes lead to alterations in the binding mode of tryptophan and effect the regiocontrol observed with these enzymes. By comparing the apo structures of PyrH and SttH, it can also be noted that many of the other active-site residues are very closely aligned, including sequences of residues such as QFPYAYHF (SttH residues 171–178) and PYYHGxxxYS (SttH residues 455–464). Other than residues between SttH G148 and G167, which lack electron density in the SttH structure, the only differences evident in the active-site region between the structures of PyrH and SttH are those of PyrH residues F451, E452 and T453 and SttH L460, P461 and P462. These residues are of particular interest because they are in close proximity to the active site in PyrH and SttH, and are positioned directly above the α -amino acid moiety of the substrate, tryptophan (**1**). Moreover, there is a loop insertion in PrnA in this region that is suggested to contribute to its regioselectivity.^[11] Each of these residues was mutated in SttH to the corresponding residue in PyrH, that is, L460F, P461E and P462T. Individually, each mutation reduced the relative activity of the enzyme with **1**, but did not have a significant effect on the observed regioselectivity, with 6-chlorotryptophan (**1 a**) remaining the major product (Figure 2). Interestingly however, the triple mutant SttH L460F/P461E/P462T exhibited similar activity to the wild-type SttH, with tryptophan as a substrate, but produced 32% 5-chlorotryptophan (**1 b**) and 68% 6-chlorotryptophan, whereas the wild-type SttH only produces the 6-chlorinated product.

A second substrate 3-indolepropionic acid (**2**), which lacks the amino group of tryptophan and therefore has more flexibility in the active site owing to the absence of a potential interaction with the backbone carbonyl of SttH G459 (Figure S4), causes a more significant shift in regioselectivity than that of the tryptophan (Figure 2). With wild-type SttH 10% 5-chloro-3-indolepropionic acid (**2 b**) is produced and 90% 6-chloro-3-indolepropionic acid (**2 a**), whereas the SttH triple mutant produces 75% **2 b** and 25% **2 a**. Notably, the relative activity of the SttH triple mutant was similar to that of the wild-type enzyme, with 3-indolepropionic acid (**2**) as substrate, thus indicating that the complete switch in regioselectivity can be achieved without impacting on catalytic efficiency. The wild-type PyrH enzyme, with either **1** or **2** as a substrate, produced exclusively 5-chlorinated products. In an effort to switch the regioselectivity

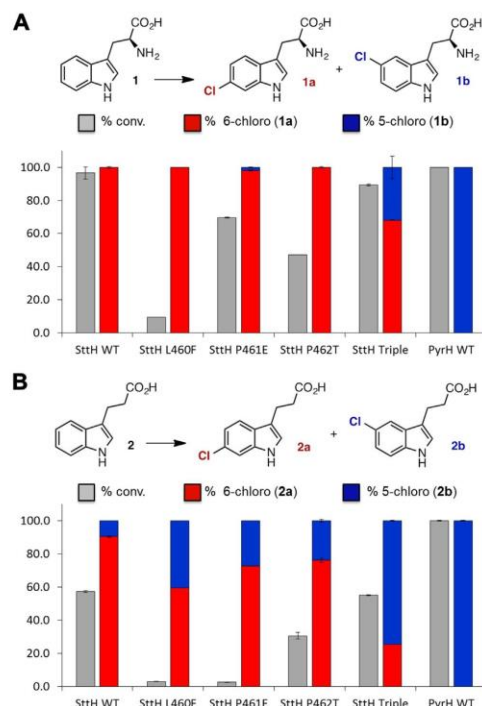


Figure 2. Percentage conversion of A) tryptophan or B) 3-indolepropionic acid with SttH wild type, single and triple mutants, as well as PyrH wild type. Purified halogenase (10 μM) was incubated with agitation at 30 °C for 1 h with Fre (1 μM), GDH2 (6 μM), FAD (7.5 μM), NADH (200 μM), MgCl_2 (50 mM), glucose (20 mM) and substrate (0.5 mM) in a total volume of 100 μL in potassium phosphate buffer (10 mM, pH 7.0).

ty of PyrH from 5- to 6-halogenation, the corresponding PyrH triple mutant (F451L/E452P/T453P) was generated with the corresponding residues observed in SttH. However, this triple mutant was found to be inactive with both substrates 1 and 2.

Previous studies have indicated that PrnA^[16] and RebH^[17] can halogenate indolic substrates, as well as tryptophan. Here we investigate the substrate specificity of SttH with *N*-methyltryptophan (3) in addition to non-indolic aromatic substrates such as kynurenine (4), anthranilamide (5) and other anilines (Table 1). As with the natural substrate 1, halogenation occurs solely at the 6-position of 3 resulting in 6-chloro product (3a). However, upon moving to the non-indolic substrate 4, chlorination did not occur at the 4-position, *meta* to the amino group, as might be expected. Instead kynurenine is chlorinated at the intrinsically more reactive 5-position by SttH. This suggests the greater flexibility of the kynurenine side chain, and perhaps also reduced π -stacking interactions with H96 and F98 (Figure S2) compared with tryptophan, enables the active-site lysine residue, K79, to deliver the electrophilic chloroamine to the most reactive 5-position, *para* to the amino group. Presum-

Table 1. Conversion of substrates after 1 hour with SttH wild type.

Substrate	Conversion [%]	Product
tryptophan (1)	97 ± 4	1a
3-indolepropionic acid (2)	57 ± 1.5	2a:2b (9:1)
<i>N</i> -methyltryptophan (3)	68 ± 3	3a
kynurenine (4)	79 ± 4	4a
anthranilamide (5)	43 ± 2	5a
anthranilic acid (6)	1.1 ± 0.1	6a
<i>N</i> -phenylanthranilic acid (7)	20 ± 2	7a

Red indicates conversion to expected product. Blue indicates conversion to chemically favoured products. Assay conditions shown in Figure 2.

ably, the subtle differences in the active-site architecture of SttH compared with PyrH are not sufficient to prevent the movement of kynurenine aryl group, so the more electronically favoured *para*-chlorination reaction predominates. The same regioselectivity is also evident with the smaller aromatic substrates 5 and anthranilic acid (6), which are both solely chlorinated at the 5-position. Finally, the biaryl compound *N*-phenylanthranilic acid (7) is also halogenated by SttH, thus showing the potential for the halogenation of larger aromatic compounds with this enzyme.

When comparing the activity of SttH with various substrates (Table 1), it is unsurprising that tryptophan was the best substrate. The addition of a methyl group to the nitrogen of the tryptophan indole ring (3) reduced activity. The activity was further decreased by the loss of the α -amino group (2) and this also led to reduced regiocontrol, about 90% 6-chlorination observed (Figure 2). With non-indolic substrates, the most electronically favoured products were produced with 4 (with its similar side chain to tryptophan) displaying close to 80% conversion. Anilines 5 and 7 also displayed good conversion; however, upon switching the amide of 5 to the acid of 6, activity is severely reduced.

From the kinetics of the selected substrates 1, 4 and 5 (Table 2), it is evident that substrate binding has a significant

Table 2. Kinetics of selected substrates with SttH wild type.

Substrate	K_m [μM]	k_{cat} [min^{-1}]	$k_{cat}/K_m \times 10^{-3}$ [$\text{min}^{-1} \mu\text{M}^{-1}$]
tryptophan (1)	0.8 ± 0.1	0.65 ± 0.02	825 ± 95
kynurenine (4)	241 ± 32	0.51 ± 0.02	2.1 ± 0.30
anthranilamide (5)	1075 ± 154	1.21 ± 0.04	1.1 ± 0.17

effect on the overall catalytic efficiency of the enzyme. Generally, the turnover varies between 0.6 and 1.2 min⁻¹; however, the K_m varied between 0.8 μ M for tryptophan and 1 mM for anthranilamide. Presumably kynurenine has lower affinity for the active site, compared with tryptophan, owing to greater side chain flexibility and reduced π -stacking interactions; this is consistent with the observed regioselectivity. In addition to this, anthranilamide also loses contacts with residues S54 and Q171, which are likely to bind the α -amino acid moiety of **1** and **4**; this leads to even lower binding affinity (Figure S4).

Previously Frese et al. demonstrated that crosslinked enzyme aggregates (CLEAs) incorporating the 7-chlorotryptophan halogenase RebH can be used to chlorinate the natural substrate tryptophan on a gram scale.^[18] By applying this method, CLEAs of SttH were produced and used to halogenate the unnatural substrate anthranilamide on a 100 mg scale (isolated yield 25%). This could be improved by optimisation and catalyst recycling.

In summary, the first crystal structure of a tryptophan 6-halogenase (SttH) has been determined. By comparing the structure of SttH with those of other halogenases, including PyrH, it is clear how subtle differences in the active site, π -stacking interactions and contacts to the α -amino acid moiety can alter the position of the aromatic moiety relative to the catalytic Lys residue thereby affecting the orientation of the subsequent electrophilic substitution reaction. The observed structural differences between the halogenases were exploited to create a SttH triple mutant, L460F/P461E/P462T, which showed the first complete switch in regioselectivity of this class of enzymes: with 3-indolepropionate as substrate, wild-type SttH gives 6-chloro-3-indolepropionate, whereas 5-chloro-3-indolepropionate was predominately produced by the triple mutant. The new regiocomplementary SttH variant displayed similar activity to the wild-type enzyme. Further assays revealed an additional five substrates that can be regioselectively halogenated by SttH, and with CLEAs, the halogenase can be stabilised for use on a preparative scale. Taken together, these results provide guidance for future efforts to engineer regiocomplementary halogenases for a wider range of aromatic substrates of synthetic utility.^[17–19]

Acknowledgements

We acknowledge the BBSRC (grants BB/K00199X/1 and BB/I008055/1), GlaxoSmithKline and CoEbio3 for support. We thank Diamond Light Source for access to beamline i24 (MX-7146), which contributed to the results presented here.

Keywords: aryl halides · biocatalysis · halogenation · mutagenesis · regioselectivity

- [1] a) S. Z. Tasker, E. A. Standley, T. F. Jamison, *Nature* **2014**, *509*, 299–309; b) J. F. Hartwig, *Nature* **2008**, *455*, 314–322; c) F.-S. Han, *Chem. Soc. Rev.* **2013**, *42*, 5270–5298; d) E. J. Cho, T. D. Senecal, T. Kinzel, Y. Zhang, D. A. Watson, S. L. Buchwald, *Science* **2010**, *328*, 1679–1681; e) K. C. Nicolaou, P. G. Bulger, D. Sarlah, *Angew. Chem. Int. Ed.* **2005**, *44*, 4442–4489; *Angew. Chem.* **2005**, *117*, 4516–4563.
- [2] Y. Lu, Y. Liu, Z. Xu, H. Li, H. Liu, W. Zhu, *Expert Opin. Drug Discovery* **2012**, *7*, 375–383.
- [3] P. Jeschke, *Pest Manage. Sci.* **2010**, *66*, 10–27.
- [4] O. Bolton, K. Lee, H.-J. Kim, K. Y. Lin, J. Kim, *Nat. Chem.* **2011**, *3*, 205–210.
- [5] a) F. Alonso, I. P. Beletskaya, M. Yus, *Chem. Rev.* **2002**, *102*, 4009–4091; b) S. A. Rowlands, A. K. Hall, P. G. McCormick, R. Street, R. J. Hart, G. F. Ebell, P. Donecker, *Nature* **1994**, *367*, 223; c) B. C. Kelly, M. G. Ikonomou, J. D. Blair, A. E. Morin, F. A. P. C. Gobas, *Science* **2007**, *317*, 236–239; d) B. Xu, H.-Z. Zhu, Y.-L. Lin, K.-Y. Shen, W.-H. Chu, C.-Y. Hu, K.-N. Tian, S. A. Mwakaganda, X.-Y. Bi, *Water Air Soil Pollut.* **2012**, *223*, 4429–4436; e) J. M. Sufflita, J. A. Robinson, J. M. Tiedje, *Appl. Environ. Microbiol.* **1983**, *45*, 1466–1473.
- [6] a) K.-H. van Pée, C. Dong, S. Flecks, J. Naismith, E. P. Patallo, T. Wage, *Adv. Appl. Microbiol.* **2006**, *59*, 127–157; b) P. Bernhardt, T. Okino, J. M. Winter, A. Miyanaga, B. S. Moore, *J. Am. Chem. Soc.* **2011**, *133*, 4268–4270.
- [7] a) A. Butler, M. Sandy, *Nature* **2009**, *460*, 848–854; b) C. S. Neumann, D. G. Fujimori, C. T. Walsh, *Chem. Biol.* **2008**, *15*, 99–109; c) P. Gao, Y. Huang, *Appl. Environ. Microbiol.* **2009**, *75*, 4813–4820; d) S. Cadel-Six, C. Dauga, A. M. Castets, R. Rippka, C. Bouchier, N. T. de Marsac, M. Welker, *Mol. Biol. Evol.* **2008**, *25*, 2031–2041.
- [8] a) D. Khare, B. Wang, L. Gu, J. Razelun, D. H. Sherman, W. H. Gerwick, K. Håkansson, J. L. Smith, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 14099–14104; b) P. C. Dorrestein, E. Yeh, S. Garneau-Tsodikova, N. L. Kelleher, C. T. Walsh, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13843–13848.
- [9] C. Dong, S. Flecks, S. Unversucht, C. Haupt, K.-H. van Pée, J. H. Naismith, *Science* **2005**, *309*, 2216–2219.
- [10] E. Bitto, Y. Huang, C. Bingman, *Proteins Struct. Funct. Bioinf.* **2008**, *289*–293.
- [11] X. Zhu, W. De Laurentis, K. Leang, J. Herrmann, K. Ihlefeld, K.-H. van Pée, J. H. Naismith, *J. Mol. Biol.* **2009**, *391*, 74–85.
- [12] a) E. Yeh, L. Cole, E. Barr, J. Bollinger, *Biochemistry* **2006**, *45*, 7904–7912; b) E. Yeh, L. C. Blasiak, A. Koglin, C. L. Drennan, C. T. Walsh, *Biochemistry* **2007**, *46*, 1284–1292.
- [13] A. Lang, S. Polnick, T. Nicke, P. William, E. P. Patallo, J. H. Naismith, K.-H. van Pée, *Angew. Chem. Int. Ed.* **2011**, *50*, 2951–2953; *Angew. Chem.* **2011**, *123*, 3007–3010.
- [14] J. Zeng, J. Zhan, *Biotechnol. Lett.* **2011**, *33*, 1607–1613.
- [15] a) X. Robert, P. Gouet, *Nucleic Acids Res.* **2014**, *42*, W320–W324; b) S. Zehner, A. Kotzsch, B. Bister, R. D. Süssmuth, C. Méndez, J. A. Salas, K.-H. van Pée, *Chem. Biol.* **2005**, *12*, 445–452.
- [16] M. Hölzer, W. Burd, H.-U. Reibig, K.-H. van Pée, *Adv. Synth. Catal.* **2001**, *343*, 591–595.
- [17] a) J. T. Payne, M. C. Andorfer, J. C. Lewis, *Angew. Chem. Int. Ed.* **2013**, *52*, 5271–5274; *Angew. Chem.* **2013**, *125*, 5379–5382; b) J. T. Payne, C. B. Poor, J. C. Lewis, *Angew. Chem. Int. Ed.* **2015**, *54*, 4226–4230; *Angew. Chem.* **2015**, *127*, 4300–4304.
- [18] M. Frese, N. Sewald, *Angew. Chem. Int. Ed.* **2015**, *54*, 298–301; *Angew. Chem.* **2015**, *127*, 302–305.
- [19] S. A. Shepherd, C. Karthikeyan, J. Latham, A.-W. Struck, M. L. Thompson, B. R. K. Menon, M. Q. Styles, C. Levy, D. Leys, J. Micklefield, *Chem. Sci.* **2015**, *6*, 3454–3460.

Manuscript received: January 27, 2016

Accepted article published: February 3, 2016

Final article published: March 30, 2016

FISK, H.[†], WESTLEY, C. [†], TURNER, N. J. & GOODACRE, R. 2016. Achieving optimal SERS through enhanced experimental design. *Journal of Raman Spectroscopy*, 47, 59-66.

[[†]] These authors contributed equally to this work.

Achieving optimal SERS through enhanced experimental design

Heidi Fisk,[†] Chloe Westley,[†] Nicholas J. Turner and Royston Goodacre*^{*}

One of the current limitations surrounding surface-enhanced Raman scattering (SERS) is the perceived lack of reproducibility. SERS is indeed challenging, and for analyte detection, it is vital that the analyte interacts with the metal surface. However, as this is analyte dependent, there is not a single set of SERS conditions that are universal. This means that experimental optimisation for optimum SERS response is vital. Most researchers optimise one factor at a time, where a single parameter is altered first before going onto optimise the next. This is a very inefficient way of searching the experimental landscape. In this review, we explore the use of more powerful multivariate approaches to SERS experimental optimisation based on design of experiments and evolutionary computational methods. We particularly focus on colloidal-based SERS rather than thin film preparations as a result of their popularity. © 2015 The Authors. *Journal of Raman Spectroscopy* published by John Wiley & Sons, Ltd.

Keywords: SERS; chemometrics; optimisation; design of experiment; genetic algorithm

Introduction

Surface-enhanced Raman scattering (SERS) is a vibrational spectroscopic technique that employs the use of roughened metal substrates, giving rise to large enhancements of the Raman signal and overcoming the inherent weakness of the traditional technique. Typically, enhancements in the range of 10^4 – 10^6 can be observed as well as reports of single molecule detection.^[1–3] Additional enhancement may be observed when the SERS technique is coupled with resonance Raman, a technique referred to as surface-enhanced resonance Raman scattering. The SERS phenomenon was first observed by Fleischmann *et al.* in 1974, and these authors observed interactions of pyridine at the surface of a roughened silver electrode, which led to a substantial increase in Raman intensity, which at that time could not be explained.^[4]

Despite SERS being discovered over 40 years ago, the mechanism of enhancement is still under debate within the SERS community. It is thought that there are two principal mechanisms that give rise to the dramatic enhancement: the electromagnetic (EM) and chemical transfer mechanisms. The first theory, thought to be the more dominant,^[5] occurs because of an interaction between the analyte and the plasmon excitation on the roughened metal surface. The incident laser light excites coherent wave oscillations of the surface electrons in the metal, called a localised surface plasmon. These oscillations at the nanostructured surface result in amplification of EM fields, which can reach out to the analyte located in close proximity.^[6,7] The second theory proposes that the analyte forms a chemical bond to the metal surface and excitation occurs via the transfer of electrons from the metal to the analyte and *vice versa*, generating a charge-transfer complex thus increasing the molecular polarisability.^[8]

The SERS technique has become increasingly popular in a wide variety of research fields because of its rapid, non-destructive and label-free nature, whilst generating highly specific structural information. Examples of these broad applications include biosensors,^[9] detection of illicit drugs,^[10] identification of DNA bases^[11] and the discrimination of bacteria.^[12]

Despite the increasing number of SERS applications, a common limitation, and real bottleneck within the community, is the difficulty in generating reproducible spectra and consistent enhancements within and between different experiments.^[13–16] Moreover, the literature can be confusing because of the many different substrates and conditions used: although this can be advantageous (due to the myriad of different SERS-active substrates and their varying properties), there tends to be an over-reliance on specific conditions meaning suboptimal enhancements are frequent. Consequently, the need to optimise the system to obtain optimum SERS responses (for each specific analyte) is an essential step in SERS experiments.

Parameters to consider in optimisation

Central to SERS is the production of a surface that is nanoscale in roughness, produced in a reproducible way, and thus, it is important that several batches of the same substrate are produced during the optimisation process. There are several parameters that need to be carefully considered, as well as various characterisation techniques that can be used to assess the suitability of these conditions (Fig. 1). To note, although the optimisation parameters we discuss in this review are more applicable for colloidal solutions, the same theory applies for thin films (with a different parameter set employed).

The selection of laser excitation wavelength has a significant impact on experimental capabilities, and whilst typically visible excitation is

* Correspondence to: Royston Goodacre, School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK. E-mail: roy.goodacre@manchester.ac.uk

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

[†] These authors contributed equally to this review.

School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester M1 7DN, UK

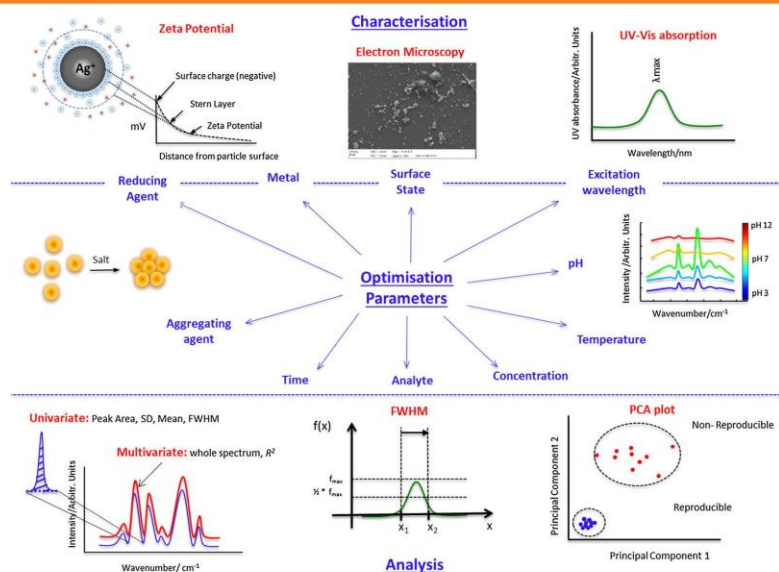


Figure 1. Schematic depicting which variables should be considered when optimising a system for achieving the best surface-enhanced Raman scattering response in terms of the signal enhancement being strong, robust and reproducible. This schematic highlights the individual parameters, various characterisation techniques as well as the data analysis approach that need to be considered during the design of experiments phase. FWHM, full width half maximum; PCA, principal component analysis; SD, standard deviation.

used (e.g. 488, 514.5, 532 and 633 nm),^[17,18] SERS has been reported from the near UV (325 nm),^[19] the near IR (785 and 830 nm) as well as recently at even longer wavelengths such as 1064, 1280 and 1550 nm.^[20–23] This selection process is often determined by a need to compromise between minimising sample fluorescence and maximising scattering efficiencies (especially if resonance is also used). Most biological samples are fluorescent in nature; therefore, choosing high frequency lasers (high power) would be deemed appropriate, although there is evidence that the metal particles can reduce the level of fluorescence.^[24,25]

A requirement for SERS is a metal surface with nanoscale roughness (normally in the range 5–100 nm), and typically, SERS substrates are in two forms: either solid state or in a colloidal suspension.^[26] Solid state SERS involves a flat surface with a roughened metal layer on top. They are usually quite expensive and more difficult to produce, as well as being site specific, and consequently, there is less control in terms of optimisation. On the other hand, colloids involve metal nanoparticles suspended in solution and are generally the more favoured substrate due to their low cost and ease of preparation. Various reducing agents are available (such as sodium citrate and hydrochloride hydroxylamine), which reduce the metal and control the nanoparticle size.^[27,28]

Most SERS substrates consist of the coinage metals, predominantly gold and silver, as their surface plasmons lie in the visible region of the EM spectrum and thus coincide with common Raman excitation wavelengths.^[18] Generally, analytes that contain thiol groups will exhibit much stronger binding to gold surfaces, whereas analytes that contain amine groups tend to have a higher affinity for silver.^[29] Manipulation of the analyte/surface binding can be achieved through varying the pH either of the analyte itself or the surface charge (through choice of reducing agent as well as charge neutralising chemicals such as poly-Lysine or spermine).

Evidently, small changes in the metal particles surface charge can have serious implications in terms of stability, sensitivity to the environment as well as electro-kinetic properties. The nanoparticles surface charge is of the utmost importance as interaction between analyte molecules and colloidal particles is a primary requisite for obtaining strong surface enhancement. When the analyte and surface have the same charge, the adsorption process can be strongly hindered, and if the colloidal particles fail to exceed a minimum repulsion with one another, they will aggregate and precipitate out of solution.^[30] In order to make the system more acidic (i.e. protonation of analyte) HCl or citric acid is typically added. In contrast, to deprotonate an analyte and make a more basic environment, NaOH is commonly used. Michota *et al.* (2003) and Alharbi *et al.* (2014) have demonstrated how the different binding modes are dependent on pH in relation to 4-mercaptobenzoic acid and nicotine, respectively.^[31,32]

Employing an aggregating agent (common examples include NaCl and KNO₃) can also have a major influence on the SERS response and so careful consideration and screening is often required. Aggregating agents are commonly employed to aggregate the different, irregular nanoparticles together. This leads to increased interactions between the colloid particles, affording larger surface plasmon resonances and greater surface enhancement. However, if too much aggregating agent is added, the colloid particles will quickly precipitate out of solution and the SERS signal will be lost.^[16,33]

Lastly, perhaps key to improving the reproducibility and often overlooked in experiments is the time allowed for optimal aggregation to occur; i.e. when do you get the most stable SERS responses and this is often analyte/system dependent. Therefore, time studies should be performed for each SERS experiment.

Characterisation techniques for nanoparticle synthesis

In order to establish the optimum conditions, especially when initial decisions regarding substrate and metal are key, various techniques are typically employed by researchers.

After synthesising the various colloids available, the first initial step is to measure its UV–Vis absorption (Fig. 1). The λ_{max} of silver colloids is around 400 and 520 nm for gold colloids (with specific surface plasmon bands available for differently reduced colloids), and comparisons with the literature are often made, rather than further characterisation (e.g. electron microscopy; EM).^[27,28] The λ_{max} full width half maximum (FWHM) of the prepared colloid allows for assessment of the nanoparticle size distribution: a narrower peak is indicative of a more monodisperse and reproducible colloid. A combination of both values should be used to establish which batch of colloid is selected for further experimental studies, although we note that this alone does not guarantee SERS for the analyte of interest.

In addition, further characterisation techniques can be employed such as EM and zeta potential analysis. EM is used to determine the size, morphology and distribution of the nanoparticles – and commonly, formation of agglomerated nanoparticles occurs more readily with silver substrates than gold. The zeta potential provides useful information regarding the charge carried by the nanoparticles and therefore the stability and ability to interact with analyte molecules. Larmour *et al.* (2012) explains that a colloid is considered stable if the zeta potential value is less than -30 mV or greater than $+30$ mV – with citrate-reduced, hydroxylamine-reduced and ethylenediamine-tetraacetic acid-reduced metal ions considered the most stable. Borohydride-reduced silver colloids are notoriously unstable with a very high zeta potential of around -7.8 mV.^[34]

Data analysis methods and characterisation techniques

There are various data analysis methods that can be employed in order to establish the suitability and reproducibility of a set of conditions. The most simple techniques involve univariate analysis – whereby the peak area of a characteristic vibration is plotted against a certain parameter, e.g. concentration and time point. Normally, other statistical assessments are calculated such as the mean and standard deviation, with the latter describing the associated errors. The FWHM can also be used to evaluate how suitable conditions are, with sharper peaks (reflected by a reduced FWHM) more favourable. In addition, it is believed that the stronger, characteristic and narrow peaks readily allow multiplexing whereby several analytes are detected simultaneously.^[32,35,36]

Although currently less common, multivariate analysis can be employed whereby the whole SERS spectrum is considered in the assessment process. One approach is to consider how well correlated repeat experiments of the same conditions are. Here, correlation coefficients are used and R^2 values closer to 1 indicate a good fit. As well as this, various chemometric approaches can be applied to simplify complex multivariate SERS data for interpretation and analysis, such as the use of principal component analysis (PCA). PCA is a widely used, unsupervised method, i.e. it does not require *a priori* information, representing the natural variance within a data set. Tight clustering of replicate measurements in PCA scores space (Fig. 1) indicates that the spectra are more similar and thus the

conditions are more reproducible. Feng, Webster and Mabbott present good examples of how PCA has been used on Raman data.^[37–39]

Combining all these characterisation techniques illustrated in Fig. 1, with the various data analysis methods available, should help determine which colloid is most suitable to use, especially in terms of batch-to-batch variation. Moreover, we advocate that multiple batches of the same metal substrate are always produced so that adequate statistics can be generated; far too often, a single experiment is reported, which may be atypical.

Design of experiment

Most researchers require suitably characterised and optimised protocols and instruments for their research purposes with a particular focus on increased sensitivity, resolution, reproducibility and lower limits of detection and quantification. Optimisation usually involves conducting appropriate experiments that will provide data on one or more performance criteria under a variety of conditions. Commonly, researchers employ the one factor at a time principle, where a single parameter is optimised first before going onto optimise the next, and so on and so forth.^[40] However, whilst relying on the knowledge and skill of the researcher, a major flaw in this approach is that this assumes a lack of statistical interaction of the parameters. For many cases, parameters within an experiment are interdependent thus contributing to a joint effect, meaning this approach rarely provides a definitive solution.

A similar problem, and mentioned earlier, is the dependence on certain conditions when conducting different experiments. It is common for researchers to focus on using a set of conditions that are familiar or those that they have had success with, rather than exploring other alternatives. A common starting point that seems to be adopted is to replicate 'optimal' conditions presented by others and switch to a one factor at a time/trial and error' approach if and when problems arise. By contrast, we believe that a more elegant approach, perhaps most effective of all, and often overlooked, is to perform a systematic design of experiment (DoE) prior to research.

Design of experiment is a well-established proven statistical method first pioneered in the 1920s by R.A. Fisher.^[41] This statistical design of experiments uses replication, blocking, randomisation and orthogonality to recognise the statistical interaction of variables and employs statistics as an objective means of drawing conclusions. These general principles of experimental design are described in detail by Morgan, Underwood, Quinn and Keough and indeed any standard statistical textbook.^[42–44]

There are different levels of design that can be applied depending on the nature of the experiment. In its simplest form, a fractional factorial design can be employed – whereby experiments are performed in order to identify which factors are the most critical. This approach is often used at the beginning of an optimisation project where many factors are likely to have little or no effect on the response (SERS enhancement). Conversely, a full factorial design combines all possible combinations of factors, affording the identification of significant interactions between them, but can be more time consuming.^[45] For example, if one wanted to optimise a modest eight parameters (e.g. type of metal, reducing agent, aggregating agent, volumes, concentrations and time of acquisition), and each of these could take one of just ten values, then the number of possible experiments is 8^{10} or just over 10^9 ! Clearly, an exhaustive search of all possible experiments is not plausible.

The common phases identified in any DoE approach include^[46]

- First, the identification of factors that may affect the outcome of the experiment and a response that can give an objective measure of this outcome. For example, this could be any optimisation parameter discussed in Fig. 1 and the effect this has on the intensity of the SERS signal (response).
- Next, is the choice of an appropriate experimental design – for example, either a full factorial design or a fractional factorial design – this will result in the generation of the design matrix, which aims to carefully select which experimental conditions need to be conducted. This small fraction of selected experiments is expected to be sufficient to reveal the most important features of the problem studied.
- These conditions are then performed in the lab and assessed by various data analysis methods, i.e. peak area(s), standard deviation, mean and PCA, as described earlier. Ultimately, whichever data analysis method is selected, the plots should describe the same trends in the results, enabling one to draw conclusions and plan the next step(s) to be taken.

A generic example of this process is illustrated in Fig. 2. A fractional factorial design has been generated to find the best solution(s) in the optimisation of a specific analyte for optimal SERS enhancement. The design is for a set of experiments in which three factors are thought to be important (these could for example be pH, concentration and aggregating agent) and is modelled on a cube to represent the experimental region being explored. The blue circles represent the initial experiments to be conducted and analysed using SERS, and these have been chosen to span the experimental search space adequately. After SERS assessment (and

data analysis), certain sets of conditions are identified as the best solutions (denoted by the red spheres) and so more experiments based on these conditions are tested (denoted by the green spheres) and analysed to find the most optimum solution(s). This iterative process should determine which sets of factors are important with the number of experiments conducted significantly reduced.

It should be noted that when optimising parameters within a SERS experiment, it is not necessarily true that the conditions optimal to signal enhancement are also optimal to reproducibility and so a trade-off between the two objectives must be established. This behaviour is an example of the Pareto principal. If one set of conditions leads to greater enhancement but lower reproducibility (or *vice versa*) than another set, one cannot assume that either set is superior. The sets of solutions within the entire search space (i.e. all combinations of parameters) that are not dominated by other solutions are termed the Pareto optimal front (denoted by the dashed arc in Fig. 2).^[47]

There are few SERS examples that have utilised the idea of DoE, with Mabbott and colleagues (2013) providing a key example in the optimisation of mephedrone detection. In this optimisation, the number of statistically significant experiments was greatly reduced from 1722 to 288 by adopting the fractional factorial approach and allowing excellent quantification of this illicit drug.^[39]

Whilst there are currently few examples in the SERS area, DoE is extensively utilised in other disciplines including: engineering, agriculture, social sciences as well as other analytical techniques. For example, in mass spectrometry based investigations for quantitative proteomics, Morris *et al.* (2010) used DoE for monitoring patterns of protein abundance in biological samples under various conditions

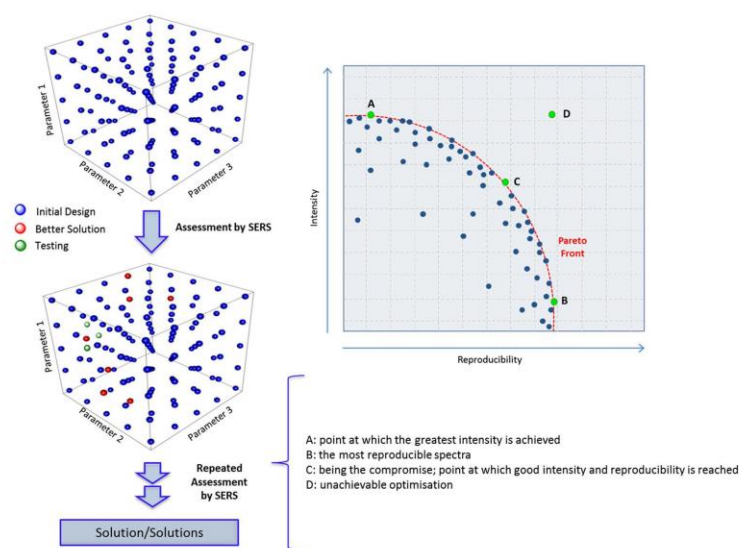


Figure 2. A 3D representation of a typical design of experiment that incorporates different parameters that needs to be optimised in order to achieve the most optimal surface-enhanced Raman scattering (SERS) response. Sequential rounds of assessment are performed until a realistic solution(s) is obtained. The concept of Pareto optimality is demonstrated: when optimising parameters within a SERS experiment, it is not necessarily true that the conditions optimal to signal enhancement are also optimal to reproducibility and so a trade-off between the two objectives must be established. As an example, the parameters to be optimised concurrently include parameter 1 = pH, parameter 2 = concentration of colloid and parameter 3 = aggregating agent.

and states. This enabled further understanding of the functioning of living organisms in search for early detection, diagnosis and prognosis of disease.^[48] Ultimately, the employment of DoE afforded the generation of reproducible and accurate results.

One of the more current areas where DoE is heavily exploited is in drug discovery of pharmaceuticals. Time and money are the major limitations holding back big drug discovery breakthroughs. DoE has widely been used in the optimisation and screening of experimental parameters. This allows key decisions to be made in the development of robust and reliable protocols in chemical synthesis, leading to optimal reaction conditions being identified in shorter periods of time.^[49] Moreover, in many lead discovery operations, assay development has been a major obstacle, consuming vast amounts of time and does not necessarily meet the desired assay quality parameters/signal window. DoE, in combination with high-throughput technologies for drug efficacy, has emerged as a leading approach in overcoming this problem by effectively reducing the time taken without compromising on quality.^[50,51]

Evolutionary computational approaches

One of the challenges in DoE is to navigate the experimental search space sufficiently, which is problematic as one cannot perform all possible experiments. Indeed, as the number of parameters that one wants to optimise increases linearly, the number of possible solutions increases exponentially. This is a so-called NP hard (non-deterministic polynomial-time hard) problem and alternative search algorithms are needed.

Genetic algorithms (GAs) are heuristic search algorithms inspired by the Darwinian principle of evolution through natural selection.^[52] First proposed by Holland (1992), these computing techniques exploit a highly abstract version of evolutionary processes in order to solve problems efficiently for which there may be more than one potential solution; that is to say the central theme is that a good solution is appropriate rather than trying to find the best overall experimental conditions, as this is impossible (there simply is not enough time) to establish without conducting every single possible solution.^[53]

One can consider a mountain range as an analogy for the experimental search space, where the height of the mountain (z-axis) represents the assessment of the analytical result, and the x-axis

and y-axis represent the parameters to be considered for optimisation. Here the higher the peak the better the experiment and the idea is to negotiate this landscape to reach the summit of the mountain range.

For relatively well-behaved (some may say easy) optimisation search spaces, there may be a simple, obvious route for a given problem. This is illustrated by Mount Fuji in Fig. 3A, where the path to the summit is clear, and such a search may be possible using simple univariate statistics. By contrast, Fig. 3B depicts a more complex mountainous landscape, where the route taken to reach the summit is unclear whereby several paths could be taken. In such instances, the search is highly dimensional in nature and multivariate approaches are necessary. Implementation of a GA to this 'Himalayan' landscape could simplify and uncover the optimal route for reaching Everest.

Following on from the evolutionary analogy, GAs can be considered as a biological representation of DNA that is changed over a series of evolutions: each GA operates on a population of chromosomes (i.e. solutions to be tested), consisting of a number of genes (i.e. variables; one per parameter to be optimised). Each gene is binary encoded (0, 1), which can be thought of as an allele, where 0 = do not use this trait (e.g. a particular aggregating agent) and 1 = use this trait in the experiment. We note that there are richer encodings but these will not be discussed here.^[53] Implementation of this binary encoding allows a series of traits to be selected. In Fig. 4B, one can translate 'Child 2' (1, 0, 1, 0, 1, 0, 0) to represent a GA for use in SERS optimisation whereby the variables represent an experiment that uses Au citrate-reduced colloid with NaCl as the reducing agent.

The overall GA process is depicted in Fig. 4A. The initial population is randomly generated and does not use any prior information that the analyst may have in terms of which parts of the search space may be better; this often requires some persuasion of the non-cognoscenti. Next, these experiments are conducted in the laboratory. The SERS enhancement for each set of conditions is assessed and this allows one to rank these experiments according to their effectiveness at solving the problem (i.e. their fitness is calculated, with higher fitness being assigned to better experimental outcomes). A subset of this population is generated by means of 'survival of the fittest'.

New experimental conditions now need to be generated (*vide infra*) to replenish the population and this process is analogous to

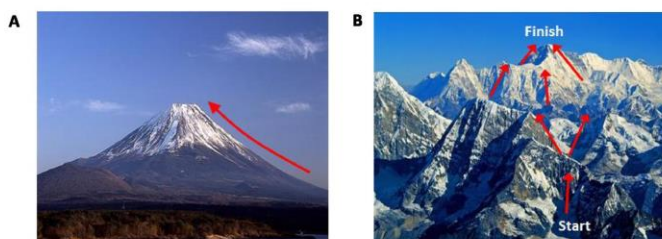


Figure 3. An illustration of the genetic algorithm (GA) approach to surface-enhanced Raman scattering (SERS) optimisation using a mountain analogy. A GA reduces the number of steps taken in order to reach the highest fitness value through the evolution of solutions to find the highest peak. (A) Denotes a simple fitness landscape (e.g. Mount Fuji), where this may mean that only a single variable needs optimising. In this case, a GA approach is not necessarily required, and simple hill climbing algorithms would suffice. Whereas (B) highlights a more complex solution where multiple routes can be taken to reach the summit of the fitness landscape, i.e. multiple variables need to be optimised simultaneously, until the highest point is reached (here depicted by the Himalayas), and the application of GA may simplify the number of solutions to reach the optimum fitness value. The figures are available from the Creative Commons license agreement. For more details, see Mount Fuji image on Flickr <https://www.flickr.com/photos/9177053@N05/4469232631/in/photostream/> and Himalayas image on Deviant art <http://citizenfresh.deviantart.com/art/Himalaya-Mountains-1-Nepal-72353246>

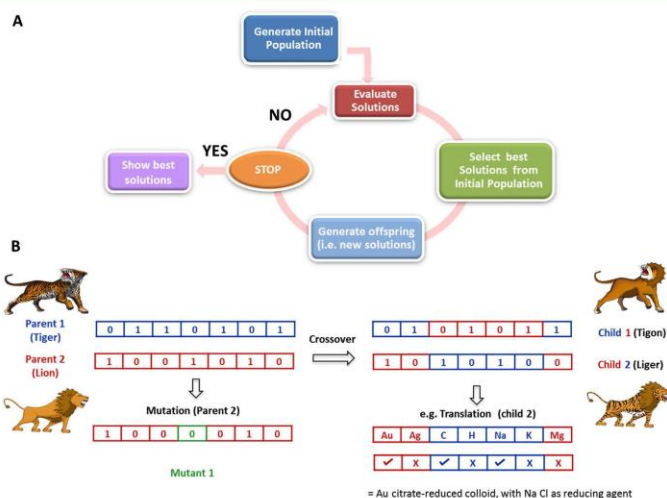


Figure 4. A schematic of the overall evolutionary approach that is used in experimental design. (A) Outlines a workflow of the genetic algorithm (GA) approach. The procedure continues to evolve until the population converges or when a maximum number of iterations is reached. (B) Denotes the methods used to generate a new population. These include offspring (in this example tigers and ligers) from the parents (a lion and a tiger). Mutations can also be generated and introduced from parents; e.g. mutant 1 is generated from a single-point mutation to parent 2. This GA process needs translation in terms of surface-enhanced Raman scattering optimisation: in the example provided, child 2 is translated as a solution where a Au citrate-reduced colloid with NaCl as the reducing agent is used for surface-enhanced Raman scattering.

biological reproduction (Fig. 4B). Chromosomes with the highest fitness values are selected as parents to recombine through crossover (mate) and produce the next population of children; alternatively, mutation may be used to modify the chromosome. In general the total population size remains constant throughout the evolutionary process.

The fitness of this new population is assessed. If the optimum solution is reached, the GA will stop, otherwise the evolutionary process will continue. The stop criterion is specific to the GA: common examples include a fixed number of generations, an observed convergence to a predetermined target (that is to say a solution has been generated that has a fitness greater than a particular value), or that after numerous generations the individual with the highest fitness value remains constant for several iterations.

Central to the GA is the way in which the population evolves over time. As well as asexual reproduction in which the children are merely clones of the parents, resulting in no diversification of the successive population, there are two main methods of producing new individuals from the previous population:

- Sexual reproduction, or crossover, occurs when two parents with high fitness values mate and swap alleles, generating two children chromosomes. Figure 4B illustrates this process where a 'Tiger' and 'Lion' are the parent chromosomes. A double crossover is shown, which generates two genetically different children – a 'Tigon' and a 'Liger' – in the hope of producing individuals with higher fitness values to that of its parents.
- Mutation can also occur whereby a single allele is randomly altered on a parent with a high fitness value, allowing for random divergence of the population. Figure 4B also shows a single mutation for Parent 2 (Lion), altering a random allele to the opposite gene (1 to 0).

There are many different factors to be considered when designing a GA. Examples include the following:

- Type of fitness function: This could be univariate *versus* multivariate assessment; SERS enhancement *versus* reproducibility or both combined in Pareto optimality.
- Population size.
- Rates of crossover and mutation operators – in general, mutation rates have a low probability of occurring, but their main purpose is to retain variation so that premature convergence does not occur.
- Evolutionary scheme that will be applied – some GAs use multiple populations that evolve independently that then cross fertilise.
- Finally the stop criterion (as discussed in the preceding section).

In summary, GAs are evolutionary computational-based algorithms that are considered as powerful explanatory techniques. Whilst the primary aim here is to search the experimental landscape efficiently to generate a good SERS substrate and set of protocols for analyte detection/quantification, in other areas, GAs have been combined with chemometric approaches to effect variable selection and the interested reader is directed to.^[54,55]

Application of evolutionary computational approaches to experimental optimisation

Genetic algorithms have proven to be highly efficient search models within a wide range of computer based science fields. He and Mort and Watanabe have successfully demonstrated GAs employment within communication network design and routing,

ultimately minimising the path and increasing the reliability between routers.^[56,57] Yao *et al.* optimised setup parameters for the simulation of a F1 race car, leading to enhanced performance and faster lap times.^[58] In addition, it would be remiss not to mention the numerous examples of the use of GAs in logistics and 'the travelling salesman problem'. This scenario relates to the optimisation process involved in identifying the shortest route required to pass through each node (city) of a tour only once. An approach that has applications in the scheduling of shipping and routing of ships, as described by Al-Hamad.^[59]

More relevant examples have been described within the analytical chemistry field, demonstrating GAs effectiveness at deconvolving complex spectral datasets into simpler solutions. Tapp successfully established that different olive oils could be distinguished as a result of geographical origin by means of GA-LDA (linear discriminant analysis) approach using Fourier transform infrared spectroscopy data.^[60] Moreover, metabolomics has employed GAs because of the high dimensional multivariate data acquired. The field focuses on characterising small molecular metabolites involved in biological processes, using combinatorial techniques such as liquid chromatography–mass spectrometry or gas chromatography–mass spectrometry.^[61] Correa and Goodacre demonstrated that *Bacillus* species could be correctly identified and classified as a result of biomarker features that were selected by the GA from complex mass spectrometry data.^[62]

To date, there are limited examples in the literature that exploit evolutionary computational approaches for the analysis of Raman and SERS data. Lavine and co-workers successfully implemented a Raman spectroscopy-GA approach in order to perform pattern recognition on several wood types, ultimately leading to their classification based on specific features, such as intensity at characteristic wavelengths.^[63]

With respect to experimental optimisation, Jarvis *et al.* identified a key example in which SERS conditions for the detection of L-cysteine were optimised by comparing the application of a multi-objective evolutionary algorithm (MOEA) versus a full factorial design. The overall aim was to increase enhancement, as well as the reproducibility of SERS spectra. Two hundred and sixteen initial conditions were assessed using the later approach [(3 × colloidal substrate) × (6 × aggregating agent) × (3 × v/v colloid ratio) × (4 × aggregating agent concentration)], whereas the MOEA consisted of four generations, 20 experiments in each, composing of the variables previously stated. The evolutionary algorithm approach was shown to be superior, showcasing a 32% improvement in reproducibility and enhancement, using far less evaluations and thus being more cost effective.^[47] Finally, Levene and colleagues were able to decrease the limits of detection of propranolol (a β-adrenergic blocker drug) 25-fold lower than what was previously published. This substantial increase in sensitivity was achieved by developing a SERS-MOEA, based on Pareto optimality. The various experimental variables (metal type, aggregating agent, laser wavelength, etc.) investigated would have resulted in a full search consisting of 7785 experiments; however, enhanced experimental conditions were determined using only 4% of all the possible combinations.^[64]

Conclusion

From the literature, it is clear that there are many different approaches towards the optimisation of SERS experiments. However, as the SERS process is analyte dependent, with different chemical species having differing hydrophobicities, charges, sizes, as well

as functional groups, there is no uniform optimisation protocol for one to follow. Notably, there is a compromise between intensity of SERS signals and reproducibility, and this is often determined by the researchers' experimental aim. For example, multiplexing experiments are not necessarily concerned with high levels of reproducibility, if the aim is just to detect more than one analyte at a time correctly. This optimisation approach would require intensity to be the most important factor, as similar levels of response for each analyte is required. However, if the aim was to quantify each of the analytes within the mixture, then reproducibility would become an increased concern. Evidently, identifying the ultimate goal prior to optimisation is key, as unavoidably there will be a trade-off between the two objectives.

Throughout this review, we have highlighted examples of the different approaches that can be applied to enhance experimental results, such as DoE and GA. It is apparent that these models have been extensively used in active research fields yet seem to be currently limited within the optimisation of SERS systems. Perhaps a contributing factor in the opposition of DoE and GAs is the fear of statistics, with many researchers considering them to be a complication. In order to implement them successfully, there is a need for the researcher to have some understanding and appreciation of the statistical tests that underpin DoE and GAs before these approaches can become standard practice. We hope that the reader has found this review useful in this regard.

Acknowledgements

H. F. and C. W. are both very grateful to BBSRC for funding their PhD studentships. R. G. and N. J. T. are also indebted to BBSRC and GSK for financial funding (grant BB/K00199X/1).

References

- [1] K. Kneipp, Y. Wang, H. Kneipp, L. T. Perelman, I. Itzkan, R. R. Dasari, M. S. Feld, *Phys. Rev. Lett.* **1997**, *78*, 1667.
- [2] S. Nie, S. R. Emory, *Science* **1997**, *275*, 5303.
- [3] H. Xu, E. J. Bjerneld, M. Käll, L. Börjesson, *Phys. Rev. Lett.* **1999**, *83*, 4357.
- [4] M. Fleischmann, P. J. Hendra, A. J. McQuillan, *Chem. Phys. Lett.* **1974**, *26*, 2.
- [5] P. Kambhampati, C. M. Child, M. C. Foster, A. Campion, *J. Chem. Phys.* **1998**, *108*, 12.
- [6] D. L. Jeanmarie, R. P. V. Duyne, *J. Electroanal. Chem.* **1977**, *84*, 1.
- [7] M. Moskovits, *Rev. Mod. Phys.* **1985**, *57*, 783.
- [8] M. G. Albrecht, J. A. Creighton, *J. Am. Chem. Soc.* **1977**, *99*, 15.
- [9] T. Vo-Dinh, *Sensors. Actuators, B.* **1995**, *29*, 1–3.
- [10] K. Faulds, W. E. Smith, D. Graham, R. J. Lacey, *Analyst* **2002**, *127*, 2.
- [11] K. Kneipp, H. Kneipp, V. B. Kartha, R. Manoharan, G. Deinum, I. Itzkan, R. R. Dasari, M. S. Feld, *Phys. Rev. E.* **1998**, *57*, 6.
- [12] R. M. Jarvis, R. Goodacre, *Anal. Chem.* **2004**, *76*, 17.
- [13] C. S. S. R. Kumar, *Raman Spectroscopy for Nanomaterials Characterization*, Springer, Berlin, **2012**.
- [14] R. Tantra, R. J. C. Brown, M. J. T. Milton, *J. Raman Spectrosc.* **2007**, *38*, 11.
- [15] N. R. Yaffe, A. Ingram, D. Graham, E. W. Blanch, *J. Raman Spectrosc.* **2010**, *41*, 6.
- [16] N. R. Yaffe, E. W. Blanch, *Vib. Spectrosc.* **2008**, *48*, 2.
- [17] Z. Q. Wen, *J. Pharm. Sci.* **2007**, *96*, 11.
- [18] E. Smith, G. Dent, *Modern Raman Spectroscopy: A Practical Approach*, Wiley, New Jersey, **2005**.
- [19] B. Ren, X. F. Lin, Z. L. Yang, G. K. Liu, R. F. Aroca, B. W. Mao, Z. Q. Tian, *J. Am. Chem. Soc.* **2003**, *125*, 32.
- [20] H. N. Xie, L. A. Larmour, Y. C. Chen, A. W. Wark, V. Tileli, D. W. McComb, K. Faulds, D. Graham, *Nanoscale* **2013**, *5*, 2.
- [21] N. G. Greeneltch, A. S. Davis, N. A. Valley, F. Casadio, G. C. Schatz, R. P. Van Duyne, N. C. Shah, *J. Phys. Chem. A* **2012**, *116*, 48.
- [22] M. A. Bedics, H. Kearns, J. M. Cox, S. Mabbott, F. Ali, N. C. Shand, K. Faulds, J. B. Benedict, D. Graham, M. R. Detty, *Chem. Sci.* **2015**, *6*, 4.

- [23] H. Huang, C. Shende, A. Sengupta, F. Inscore, C. Brouillette, W. Smith, S. Farquharson, *J. Raman Spectrosc.* **2012**, *43*, 6.
- [24] G. McNay, D. Eustace, W. E. Smith, K. Faulds, D. Graham, *Appl. Spectrosc.* **2011**, *65*, 8.
- [25] W. E. Smith, *Chem. Soc. Rev.* **2008**, *37*, 5.
- [26] V. G. Bordo, H.-G. Rubahn, *Optics and Spectroscopy at Surfaces and Interfaces*, Wiley, Weinheim, **2008**.
- [27] P. C. Lee, D. Meisel, *J. Phys. Chem.* **1982**, *86*, 17.
- [28] N. Leopold, B. Lendl, *J. Phys. Chem. B* **2003**, *107*, 24.
- [29] J. R. Lombardi, R. L. Birke, T. Lu, J. Xu, *J. Chem. Phys.* **1986**, *84*, 8.
- [30] R. A. Alvarez-Puebla, E. Arceo, P. J. G. Goulet, J. J. Garrido, R. F. Aroca, *J. Phys. Chem. B* **2005**, *109*, 9.
- [31] A. Michota, J. Bukowska, *J. Raman Spectrosc.* **2003**, *34*, 1.
- [32] O. Alharbi, Y. Xu, R. Goodacre, *Analyst* **2014**, *139*, 19.
- [33] S. Sánchez-Cortés, J. V. García-Ramos, G. Morcillo, *J. Colloid Interface Sci.* **1994**, *167*, 2.
- [34] I. A. Larmour, K. Faulds, D. Graham, *J. Raman Spectrosc.* **2012**, *43*, 2.
- [35] C. J. McHugh, F. T. Docherty, D. Graham, W. E. Smith, *Analyst* **2004**, *129*, 1.
- [36] K. Gracie, E. Correa, S. Mabbott, J. A. Dougan, D. Graham, R. Goodacre, K. Faulds, *Chem. Sci.* **2014**, *5*, 3.
- [37] S. Feng, R. Chen, J. Lin, J. Pan, G. Chen, Y. Li, M. Cheng, Z. Huang, J. Chen, H. Zeng, *Biosens. Bioelectron.* **2010**, *25*, 11.
- [38] G. T. Webster, J. Dusting, S. Balabani, E. W. Blanch, *J. Phys. Chem. B* **2011**, *115*, 11.
- [39] S. Mabbott, E. Correa, D. P. Cowcher, J. W. Allwood, R. Goodacre, *Anal. Chem.* **2013**, *85*, 2.
- [40] C. Daniel, *J. Am. Stat. Assoc.* **1973**, *68*, 342.
- [41] R. A. Fisher, *Statistical Methods for Research Workers*, Oliver & Boyd, London, **1934**.
- [42] E. Morgan, *Journal of Chemometrics*, Wiley, Chichester, **1995**.
- [43] A. J. Underwood, *Experiments in Ecology: Their Logical Design and Interpretation Using Analysis of Variance*, Cambridge University Press, Cambridge, **1997**.
- [44] G. P. Quinn, M. J. Keough, *Experimental Design and Data Analysis for Biologists*, Cambridge University Press, Cambridge, **2002**.
- [45] C. F. J. Wu, M. Hamada, *Experiments: Planning, Analysis, and Parameter Design Optimization*, Wiley India Pvt. Limited, New Delhi, **2009**.
- [46] R. A. Bailey, *Design of Comparative Experiments*, Cambridge University Press, Cambridge, **2008**.
- [47] R. M. Jarvis, W. Rowe, N. R. Yaffe, R. O'Connor, J. D. Knowles, E. W. Blanch, R. Goodacre, *Anal. Bioanal. Chem.* **2010**, *397*, 5.
- [48] J. S. Morris, K. A. Baggerly, H. B. Gutstein, K. R. Coombes, *Methods in Molecular Biology*, Humana Press, New York, **2010**.
- [49] M. R. Owen, C. Luscombe, L. W. Lai, S. Godbert, D. L. Crookes, D. Emiabata-Smith, *Org. Process Res. Dev.* **2001**, *5*, 3.
- [50] M. W. Lutz, J. A. Menius, T. D. Choi, R. G. Laskody, P. L. Domanico, A. S. Goetz, D. L. Saussy, *Drug Discov. Today* **1996**, *1*, 7.
- [51] H. Tye, *Drug Discov. Today* **2004**, *9*, 11.
- [52] C. Darwin, *On the Origin of Species*, Murray, London, **1859**.
- [53] J. H. Holland, *Sci. Am.* **1992**, *267*, 1.
- [54] R. O. Duda, P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, **1973**.
- [55] T. M. Mitchell, *Machine Learning*, McGraw-Hill, New York, **1997**.
- [56] L. He, N. Mort, *Brit. Tele. Technol.* **2000**, *18*, 4.
- [57] S. Watanabe, T. Hiroyasu, M. Miki, Proceedings of the EUROGEN 2001 Conference, Athens, **2001**.
- [58] X. Yao, E. Burke, J. Lozano, J. Smith, J. Merelo-Guervós, J. Bullinaria, J. Rowe, P. Tiño, A. Kabán, H. P. Schwefel, K. Wloch, P. Bentley, *Parallel Problem Solving from Nature – PPSN VIII*, Springer, Berlin, **2004**.
- [59] K. Al-Hamad, M. Al-Ibrahim, E. Al-Enezy, *Ann. Oper. Res.* **2012**, *2*, 3.
- [60] H. S. Tapp, M. Defernez, E. K. Kemsley, *J. Agric. Food Chem.* **2003**, *51*, 21.
- [61] S. G. Oliver, M. K. Winson, D. B. Kell, F. Baganz, *Trends Biotechnol.* **1998**, *16*, 9.
- [62] E. Correa, R. Goodacre, *BMC Bioinf.* **2011**, *12*, 33.
- [63] B. K. Lavine, C. E. Davidson, A. J. Moores, P. R. Griffiths, *Appl. Spectrosc.* **2001**, *55*, 8.
- [64] C. Levene, E. Correa, E. W. Blanch, R. Goodacre, *Anal. Chem.* **2012**, *84*, 18.