Supporting Information

New Insights into the Biphasic CO-free Pauson-Khand Cyclisation Reaction through Combined In Situ Spectroscopy and Multiple Linear Regression Modelling

Robert Geitner[a,†,*], Tianbai Huang[b], Stephan Kupfer[b], Stefanie Gräfe[b], Florian Meirer[a] and Bert M. Weckhuysen[a,*]

[a] *Inorganic Chemistry and Catalysis Group, Debye Institute for Nanomaterials Science, Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands*

[b] *Institute for Physical Chemistry and Abbe Center of Photonics, Friedrich Schiller University Jena, Helmholtzweg 4, 07743 Jena, Germany*

†Present address: Institute for Organic and Macromolecular Chemistry, Friedrich Schiller University Jena, Philosophenweg 7, 07743, Jena, Germany.

*robert.geitner@uni-jena.de and b.m.weckhuysen@uu.nl

Table of Contents:

1. Experimental Section

Table S1. Experimental conditions for individual in situ Raman spectroscopy measurements.

Exp.	**T / °C**	**$[Rh(cod)Cl]_2$ / mg**	**dppp / mg**	**TPPTS / mg**	**SDS / mg**	**PFA / mg**	**Enyne / mL**
1	100	15.15	33.42	36.14	361.72	189.35	0.20
2	100	30.78	67.34	72.88	374.58	95.13	0.20
3	100	7.74	17.25	18.11	359.21	95.76	0.20
4	100	15.86	34.53	35.53	354.92	93.71	0.40
5	100	15.35	33.72	35.57	362.73	91.19	0.10
6	100	15.25	33.31	35.42	722.30	93.71	0.20
7	100	15.79	34.24	36.43	181.15	92.95	0.20
8	100	15.47	4.95	67.89	363.56	95.95	0.20
9	100	15.24	41.84	14.16	363.95	94.07	0.20
10	100	15.76	34.16	35.00	365.74	93.31	0.20
11	110	15.83	34.41	36.08	365.33	93.05	0.20
12	90	15.57	33.86	35.39	360.28	93.58	0.20
13	100	15.61	33.95	36.73	357.49	47.68	0.20
14	100	15.73	34.01	35.35	361.04	187.96	0.20
15	102	24.77	35.60	45.91	392.61	108.19	0.10
16	101	17.90	21.82	46.90	512.34	59.51	0.11
17	93	18.22	25.46	67.52	347.41	133.08	0.18
18	96	15.40	37.27	67.55	381.10	143.37	0.22

pH Measurements

The pH of the reaction suspension prior to the reaction was measured using a PHM210 pH meter from Radiometer Analytical (Lyon, France). The pH was 8.26.

Data Processing

The Raman spectra were preprocessed using R (4.0.2).[1] First the Raman spectra were despiked using a moving average approach (window size = 23, standard deviation = 3).[2] Subsequently the wavenumber axis for each individual spectrum was interpolated from 560 to 3400 cm^{-1} (1 cm^{-1} step size) using a cubic spline fit to have all intensity values correspond to the same x-axis values. This step is necessary to enable the subsequent fitting procedure. Afterwards the Raman spectra were background corrected using a SNIP algorithm[3,4] (iterations = 40 or 80 for in situ and reference spectra, respectively, order = 2, smoothing window = 3). The in situ Raman spectra were slightly smoothed by a regression function using local polynomials and a kernel bandwidth of 1.[5,6] Finally, the in situ Raman spectra were normalized to the band between 1380 and 1470 cm^{-1}. This band can be assigned to δ(C-H)[7] and thus can serve as a reference signal as many C-H groups especially from SDS, **Enyne** and **BCP** are not changing over the course of the reaction. The reference spectra were vector normalized by dividing each spectrum vector by its magnitude so that the magnitude becomes 1 and all 32 spectra were averaged (see Figure S1 - Figure S14).

For the kinetic analysis each individual in situ spectrum S was fitted using a linear least-square fit and the reference Raman spectra of the pure components s_x. The fit coefficients c_x are adjusted to best represent S by s_x.

$$S = \sum c_x \cdot s_x$$

For this fitting procedure the wavenumber regions between 1838 and 2136 as well between 2276 and 2674 cm^{-1} were cut from all in situ and reference Raman spectra as they only contain noise and thus disturb the fit. None of the reference components shows Raman signals in these wavenumber areas (compare Figure S1 - Figure S14). At first 8 reference Raman spectra (**Enyne**, **BCP**, H_2O in microwave tube, dppp, TPPTS, paraformaldehyde, SDS and $[Rh(cod)Cl]_2$) were used to estimate the contribution of each individual component to a given in situ spectrum. After inspecting the time dependent fit coefficients c_x of all components it became clear that most components had only a minor effect on the quality of the fitting procedure evaluated by the R^2 value (see Figure S16 and Figure S18). Thus, the number of components was reduced until a major drop in the R^2 value was observed. This was the case when the number of components changed from 3 to 2 (see Figure S18). In this iterative procedure the component with the smallest contribution to the fit was dropped from the fitting procedure. As expected the 3 components with the strongest contribution to the fitting procedure are **Enyne**, **BCP** and H_2O as they have the largest concentration in the reaction mixtures (see Figure S17). To validate this statement all 56 combinations to use 3 components from a total pool of 8 available components were tested for all in situ Raman measurements. On average the previously identified combination of **Enyne**, **BCP** and H_2O performs best in this fitting procedure when comparing all available in situ measurements.

To validate this approach principal component analysis (PCA) was used.[8,9] The exemplary PCA reveals that 3 principal components (PCs) explain a large portion of the variance (46.3 % of cumulative variance) found in the example data set (see Figure S20). This is indicated by the scree (elbow criterion) in the cumulative variance plot.

To prove that the coefficient of an individual component in the least square fit of the in situ Raman spectra is proportional to the concentration of this component in the reaction mixture the concentration dependent Raman signal of **Enyne** in H_2O was studied. The Raman peak area of the ν(C≡C) vibration was integrated from 2220 to 2280 cm^{-1} and plotted against the concentration of **Enyne** in H_2O (see Figure S21). This calibration plot shows a linear dependency so that the Raman peak area of the ν(C≡C) vibration can serve as an indicator for the **Enyne** concentration. Subsequently, the same peak area was integrated in an exemplary in situ measurement (see Figure S22). The resulting time dependent peak area curve looks very similar to the time dependent fit coefficient of **Enyne** c_{Enyne} (Figure S17). This finding is supported when also the peak area of **BCP** – integrated from 1620 to 1690 cm^{-1} – is compared to the corresponding fit coefficient c_{BCP}. Thus, it can be concluded that the coefficients c_x describing the contribution of each individual component to a given in situ Raman spectrum is a good descriptor of this components concentration in the reaction mixture.

As a final validation of the linear least-square fitting approach the Multivariant Curve Resolution – Alternating Least Square (MCR-ALS) algorithm was used.[10,11] MCR-ALS is an unsupervised data analysis approach and thus inertly avoids biases which may stem from the choice of the individual fitting components. The MCR-ALS algorithm was initialized with 3 components and the previously integrated time-dependent Raman peak areas as well as a constant to value the presence of H_2O. The resulting spectra extracted by the MCR-ALS algorithm are shown in Figure S23 and Figure S24 while the concentration profiles are shown in Figure S25. A comparison between the MCR-ALS spectra and the pure Raman spectra of **Enyne** and **BCP** reveals that the unsupervised MCR-ALS approach identified **Enyne** and **BCP** as the main compounds in the in situ Raman spectra without a-priori knowledge of their respective spectral signature. As the supervised linear least-square fit and the unsupervised

MCR-ALS approach arrived at the same result the linear least-square fit was chosen as the go-to model as it is more robust and carries more direct chemical information.

After the individual fit coefficients were determined a first-order kinetic fit was used to extract the kinetic rate constants for each in situ measurement. To achieve this the time dependent fit coefficients of **Enyne** c_{Enyne} were fitted with an exponential function according to

$$c_{Enyne} = c_{0,Enyne} \cdot e^{-k \cdot t} + \ c_{f,Enyne}$$

where t is the time, k the kinetic rate constant, $c_{0,\mathrm{Enyne}}$ the starting fit coefficient and $c_{\mathrm{f,Enyne}}$ the fit coefficient at the end of the reaction. The fit coefficients c_{Enyne} and the corresponding time t were weighted with the R^2 values of the previous linear spectra fitting procedure. Following from this equation k was extract for each in situ Raman measurements. k describes the consumption of **Enyne** while the decreasing **Enyne** concentration is an indicator for the progress of the on-going Pauson-Khand reaction.

To prove that the underlying kinetic is indeed of first order the integrated peak areas (see Figure S22) were fitted not only with a first order kinetic function but also with a zeroth order function according to

$$c_{Enyne} = -k_{zero} \cdot t + \ c_{0,Enyne}$$

and a second order kinetic fit according to

$$c_{Enyne} = \frac{1}{2 \cdot k_{sec} \cdot t + \frac{1}{c_{0,Enyne}}} + \ c_{f,Enyne}$$

The zeroth and second order fit give a slightly worse description of the data set as revealed by the associated standard residual errors (see Figure S22).

Finally, the first-order kinetic rate constants k were combined with the experimental starting conditions shown in Table S1. To evaluate the influence of all 7 experimental reaction conditions x a multiple linear regression model was built. The model was defined as follows:

$$k = \sum a_x \cdot x = a_T \cdot T + a_{Rh} \cdot [Rh]_0 + a_{dppp} \cdot [dppp]_0 + a_{TPPTS} \cdot [TPPTS]_0 + a_{SDS} \cdot [SDS]_0 + a_{PFA} \cdot [PFA]_0 + a_{Enyne} \cdot [Enyne]_0$$

Where k is the kinetic rate constant of a given in situ measurement, a_x is the influence parameter with a specific reaction condition while x is one of the 7 experimental reaction conditions, namely the reaction temperature T and the starting concentrations $[x]_0$. The influence parameters a_x were subsequently optimized by multiple linear regression to best describe the influence of each reaction condition on the kinetic rate constant k. This is achieved by minimizing the squared residuals which are calculated by squaring the difference between the experimental kinetic rate constants and the predicted kinetic rate constants from the multiple linear regression model. The R^2 value of the optimized model was 0.91.

2. Additional Raman Spectra

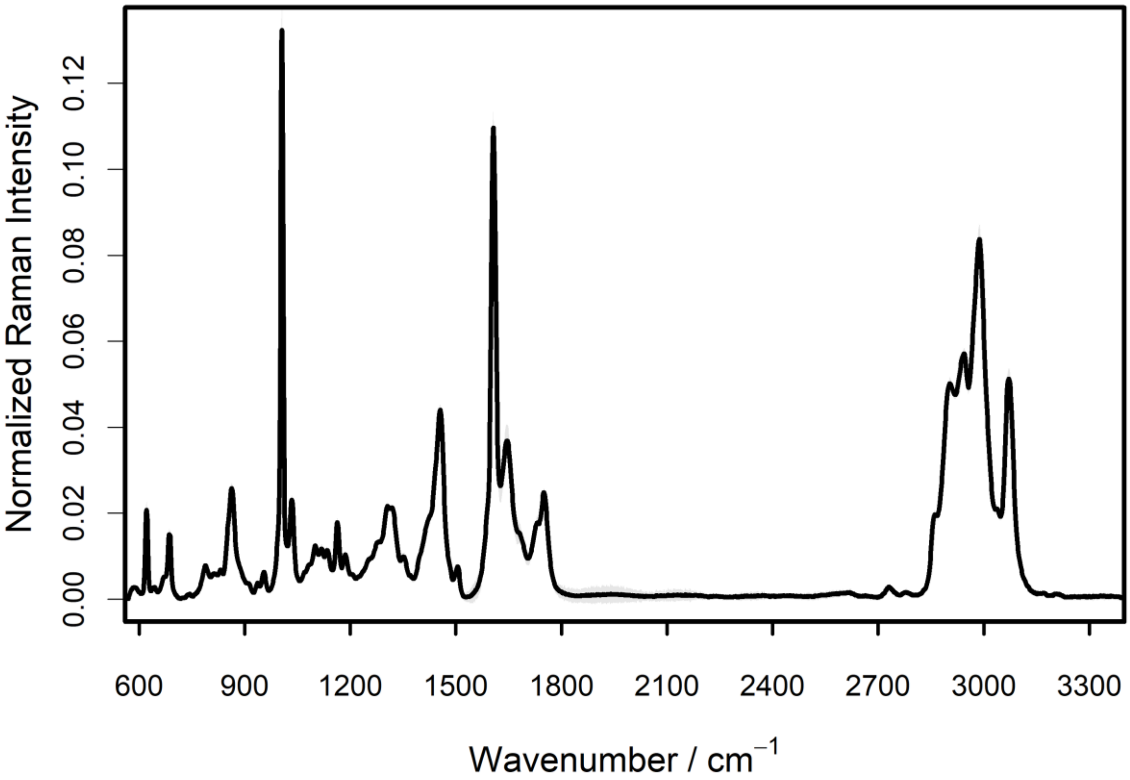

Figure S1. Background corrected and normalized Raman spectrum of **BCP**. The grey area indicates the 95 % confidence interval. In this case it is very narrow and therefore hard to see. Band assignment[7]: 3073 (ν(=C-H)), 2988 (ν(C-H)), 1752 (ν(C=O)), 1732 (ν(C=O)), 1646 (ν(C=C)), 1608 (ν(C=C)), 1455 (δ(C-H)), 1005 (phenyl ring breathing) cm^{-1}.

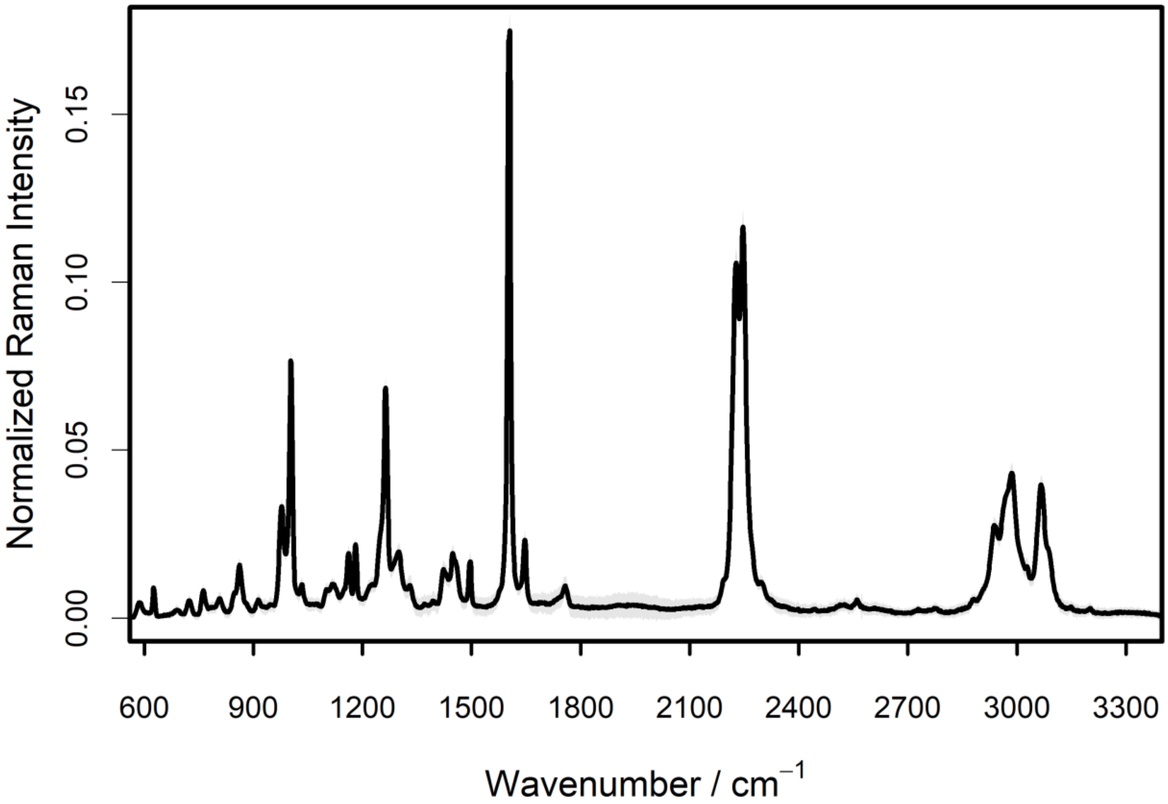

Figure S2. Background corrected and normalized Raman spectrum of **Enyne**. The grey area indicates the 95 % confidence interval. In this case it is very narrow and therefore hard to see. Band assignment[7]: 3067(ν(=C-H)), 2985(ν(C-H)), 2246 (ν(C≡C)), 2229 (ν(C≡C)), 1761 (ν(C=O)), 1649 (ν(C=C)), 1602 (ν(C=C)), 1264, 1002 (phenyl ring breathing) cm^{-1}.

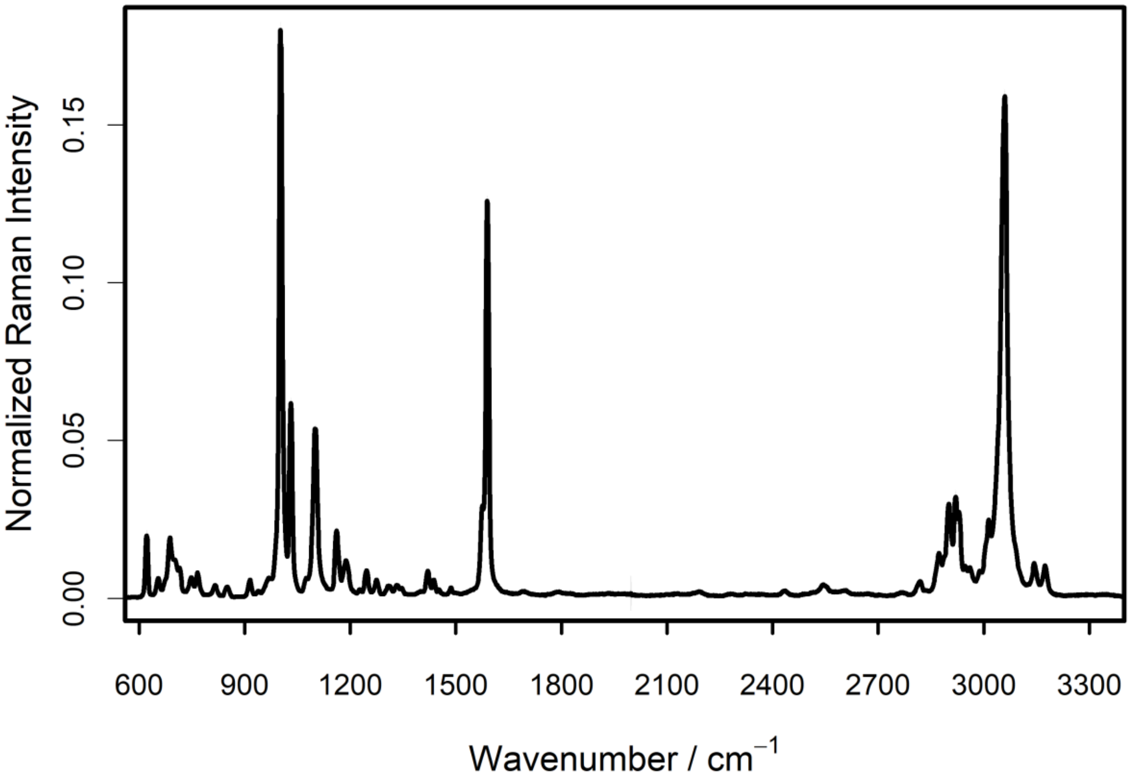

Figure S3. Background corrected and normalized Raman spectrum of dppp. The grey area indicates the 95 % confidence interval. In this case it is very narrow and therefore hard to see. Band assignment[7]: 3058 (ν(=C-H)), 1591 (ν(C=C)), 1576 (ν(C=C)), 1099, 1029, 1002 (phenyl ring breathing) cm^{-1}.

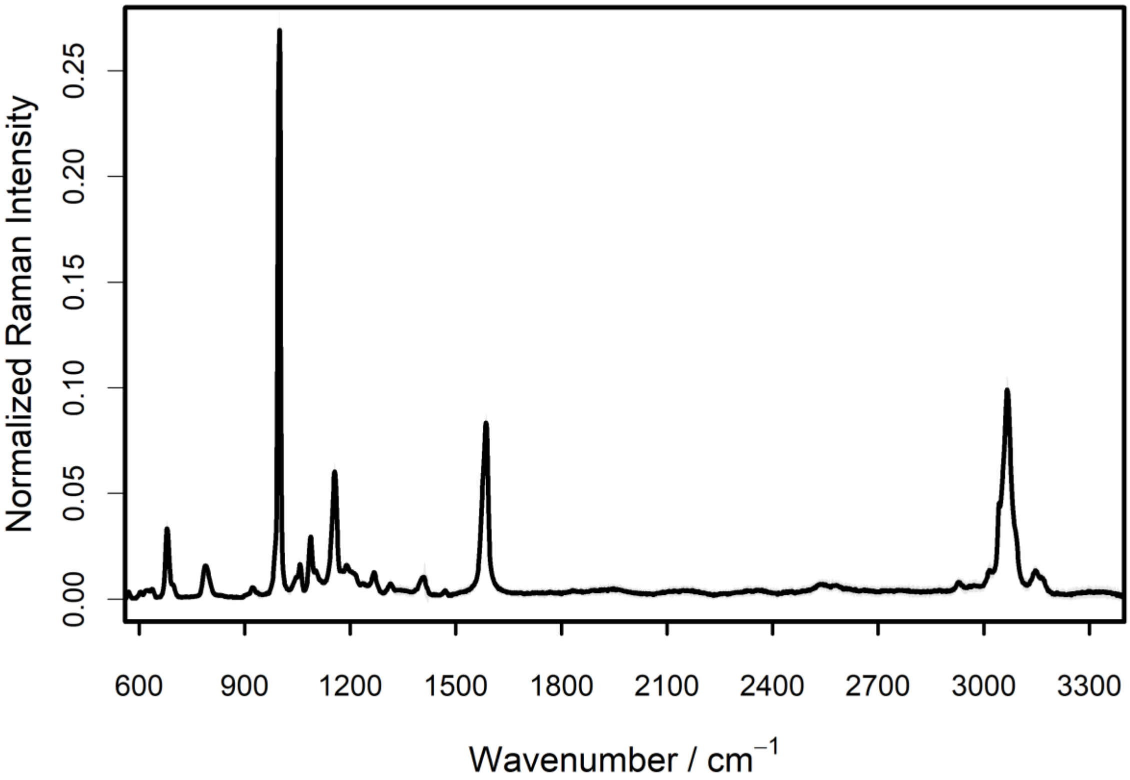

Figure S4. Background corrected and normalized Raman spectrum of TPPTS. The grey area indicates the 95 % confidence interval. In this case it is very narrow and therefore hard to see. Band assignment[7]: 3067 (ν(=C-H)), 1584 (ν(C=C)), 1154, 1000 (phenyl ring breathing) cm^{-1}.

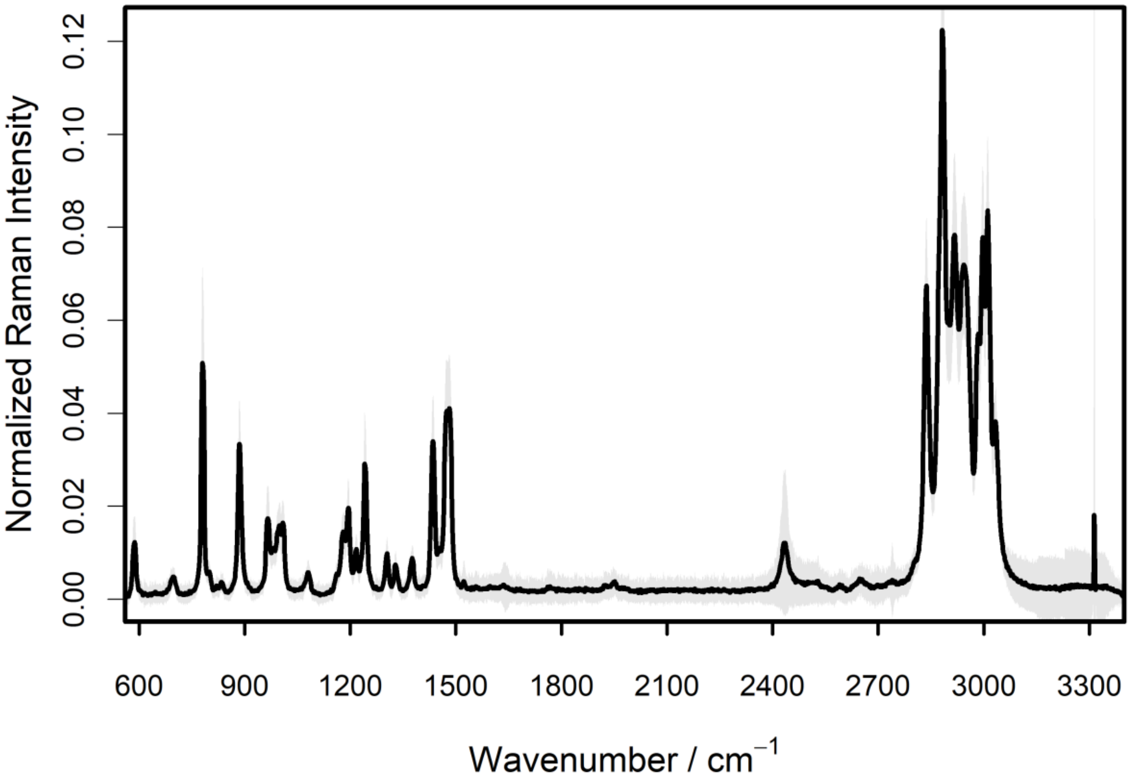

Figure S5. Background corrected and normalized Raman spectrum of [Rh(cod)Cl]$_2$. The grey area indicates the 95 % confidence interval. Band assignment[7]: 2885 (ν(C-H)), 2435, 1476 (δ(C-H)), 1435 (δ(C-H)), 1244, 888, 782 cm^{-1}.

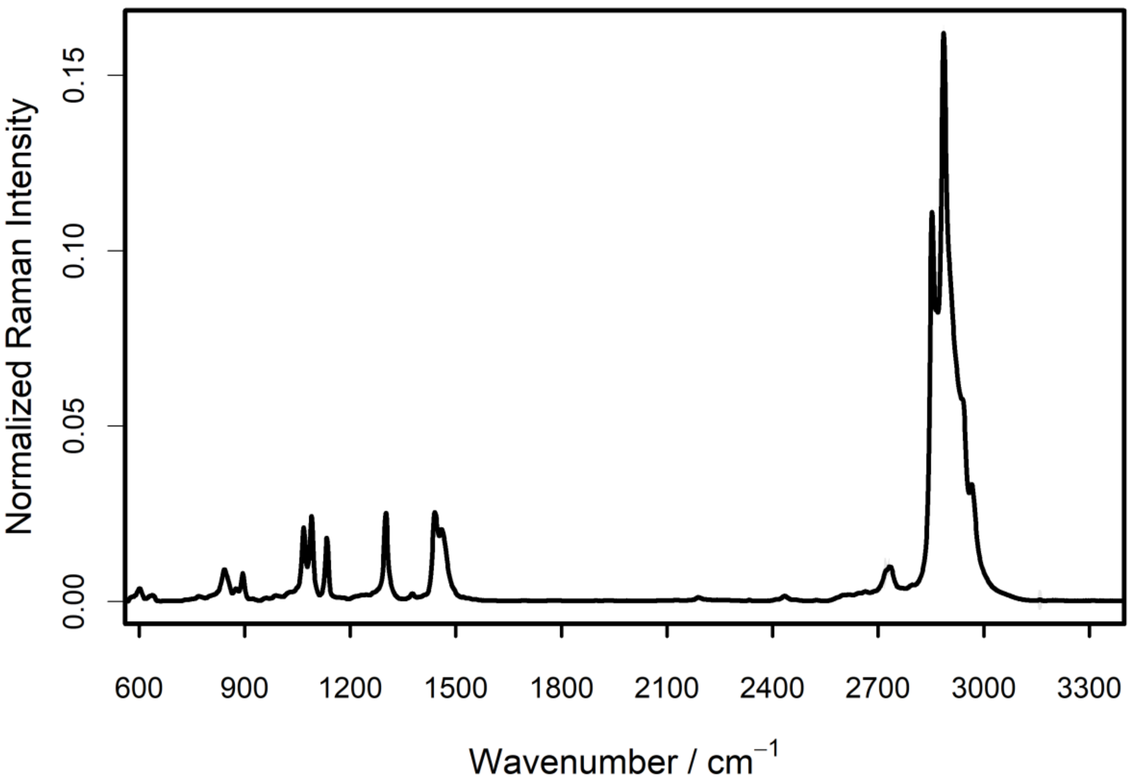

Figure S6. Background corrected and normalized Raman spectrum of SDS. The grey area indicates the 95 % confidence interval. In this case it is very narrow and therefore hard to see. Band assignment[7]: 2967 (ν(C-H)), 2888 (ν(C-H)), 2849 (ν(C-H)), 1461 (δ(C-H)), 1441 (δ(C-H)) cm^{-1}.

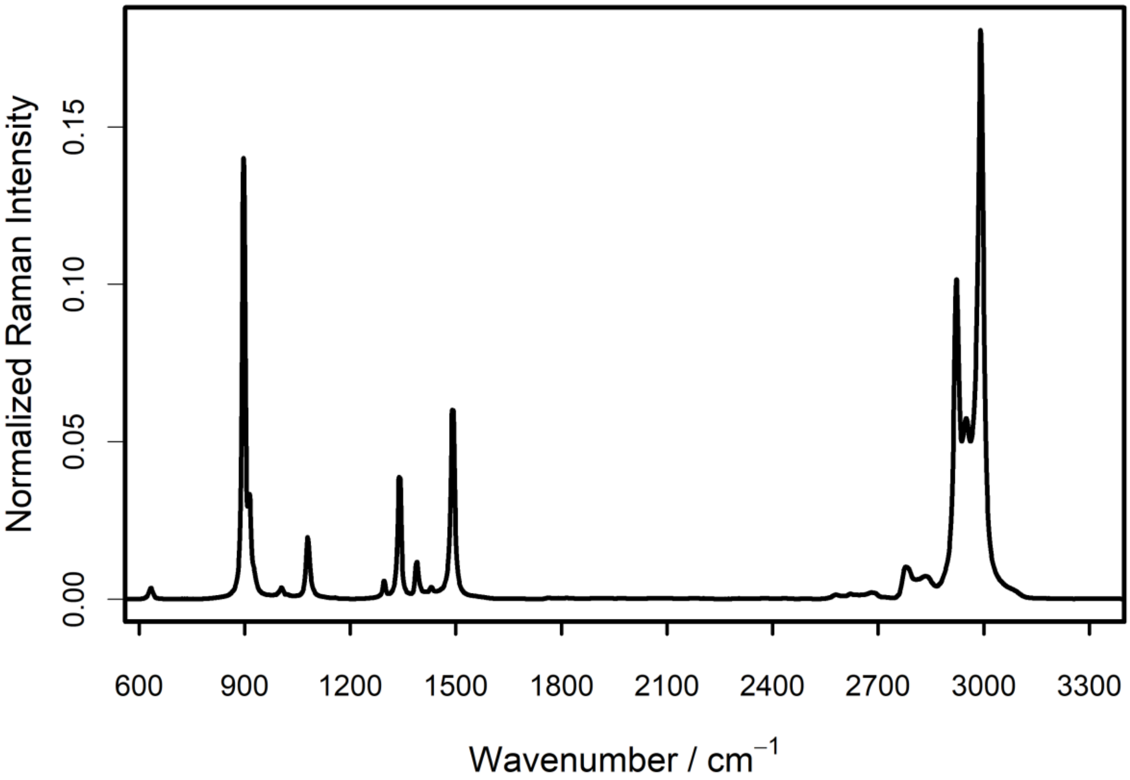

Figure S7. Background corrected and normalized Raman spectrum of ^{13}C-PFA. The grey area indicates the 95 % confidence interval. In this case it is very narrow and therefore hard to see. Band assignment[7]: 2993 (ν(C-H)), 2923 (ν(C-H)), 1494 (δ(C-H)), 1341, 1082, 897 (ν(O-C-O)) cm^{-1}.

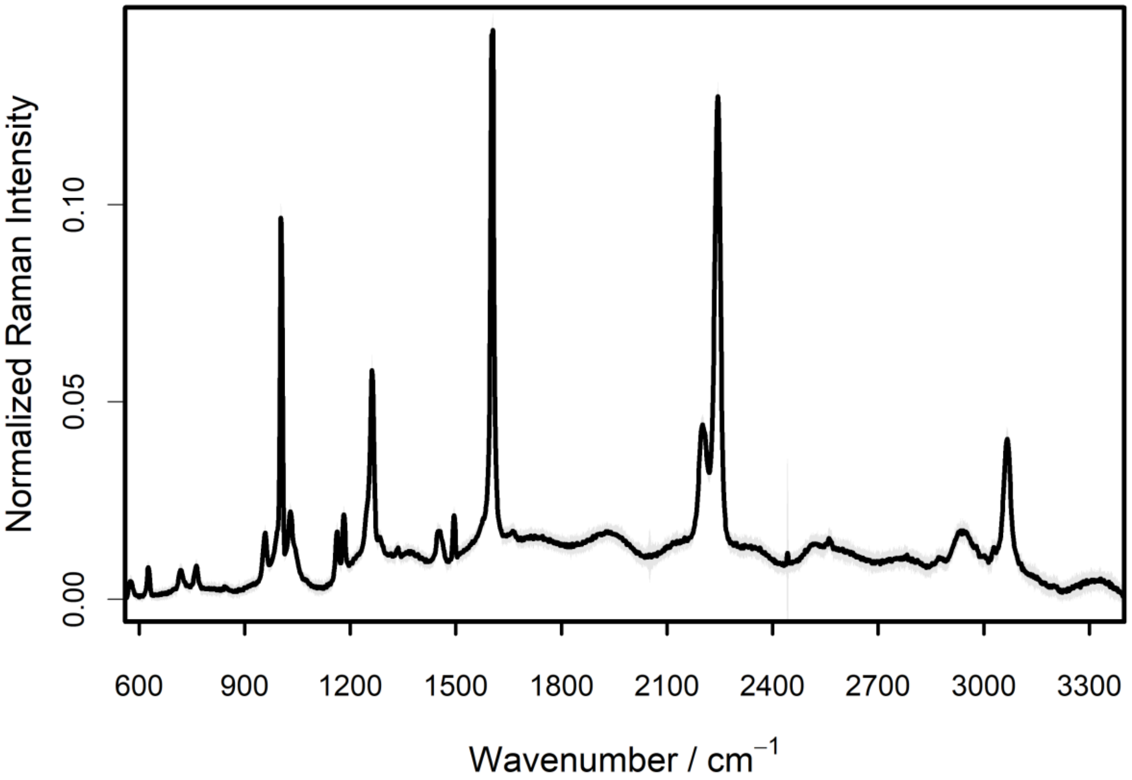

Figure S8. Background corrected and normalized Raman spectrum of 3-phenyl-2-propyne-1-ol. The grey area indicates the 95 % confidence interval. In this case it is very narrow and therefore hard to see. Band assignment[7]: 3067 (ν(=C-H)), 2245 (ν(C≡C)), 2201 (ν(C≡C)), 1602 (ν(C=C)), 1262, 1005 (phenyl ring breathing) cm^{-1}.

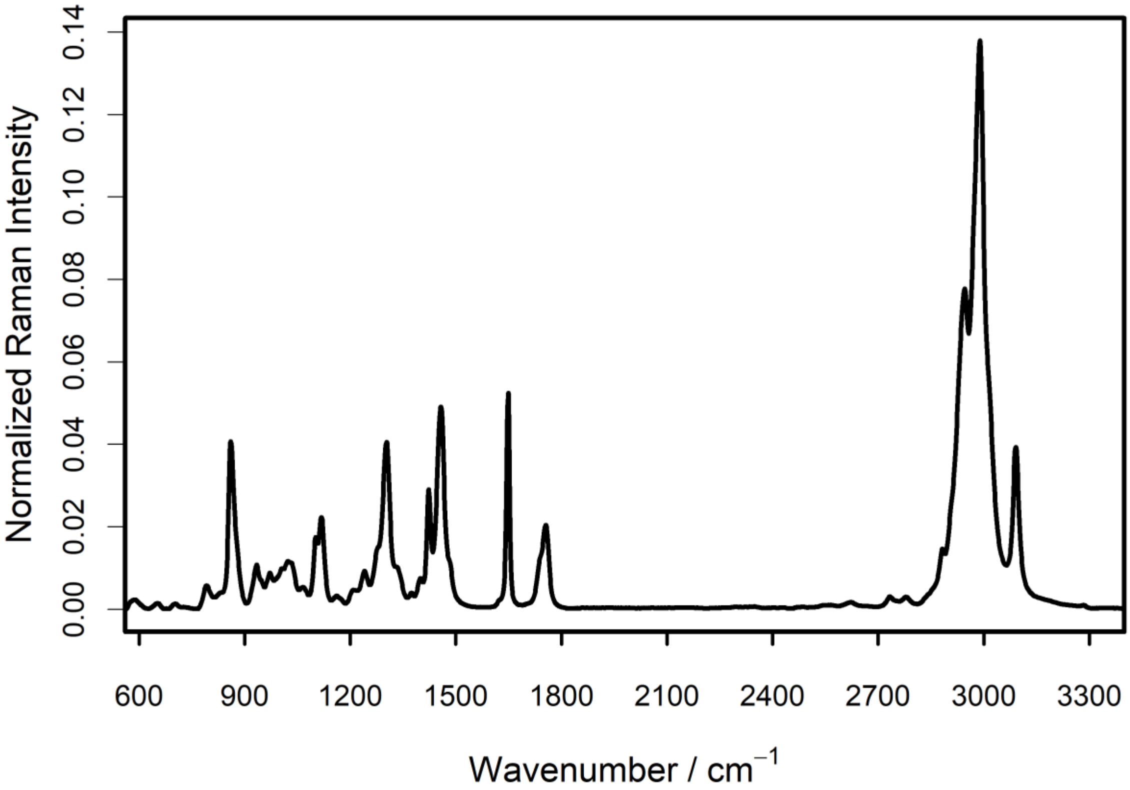

Figure S9. Background corrected and normalized Raman spectrum of diethylallylmalonate. The grey area indicates the 95 % confidence interval. In this case it is very narrow and therefore hard to see. Band assignment[7]: 3090 (ν(=C-H)), 2988 (ν(C-H)), 1757 (ν(C=O)), 1651 (ν(C=C)), 1458 (δ(C-H)) cm^{-1}.

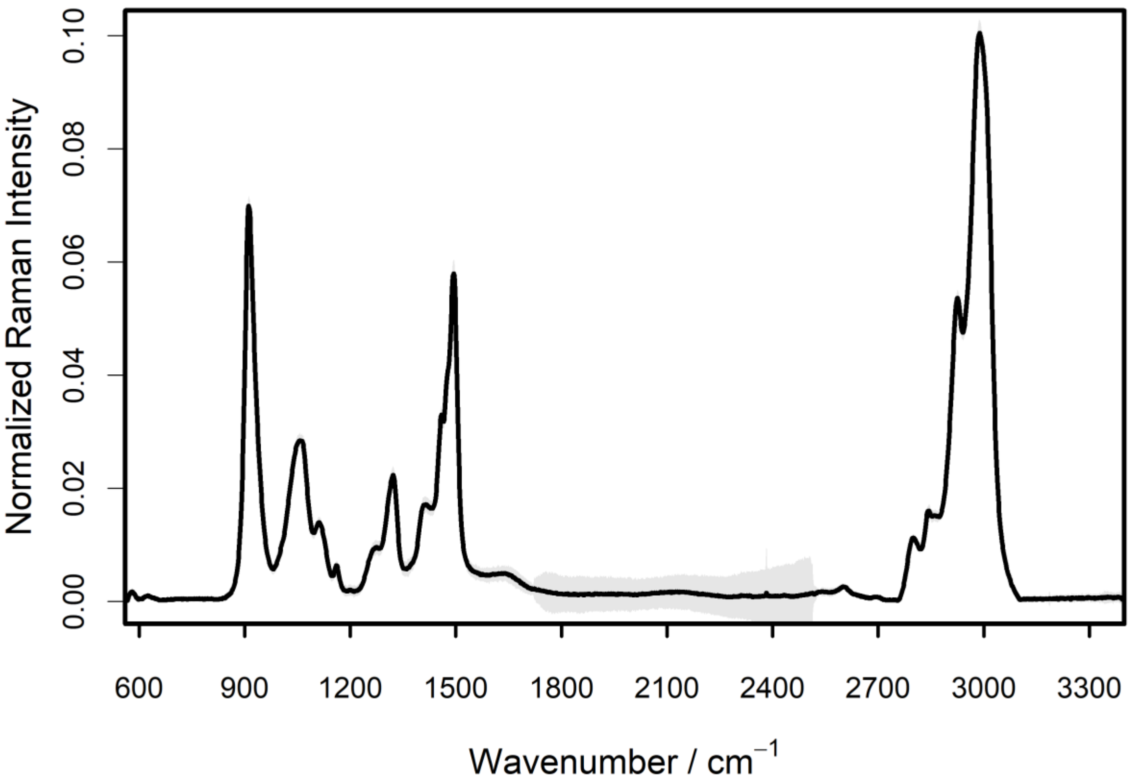

Figure S10. Background corrected and normalized Raman spectrum of 37 w% formaldehyde in H_2O. The grey area indicates the 95 % confidence interval. Band assignment[7]: 2991 (ν(C-H)), 1493 (δ(C-H)), 914 (ν(O-C-O)) cm^{-1}.

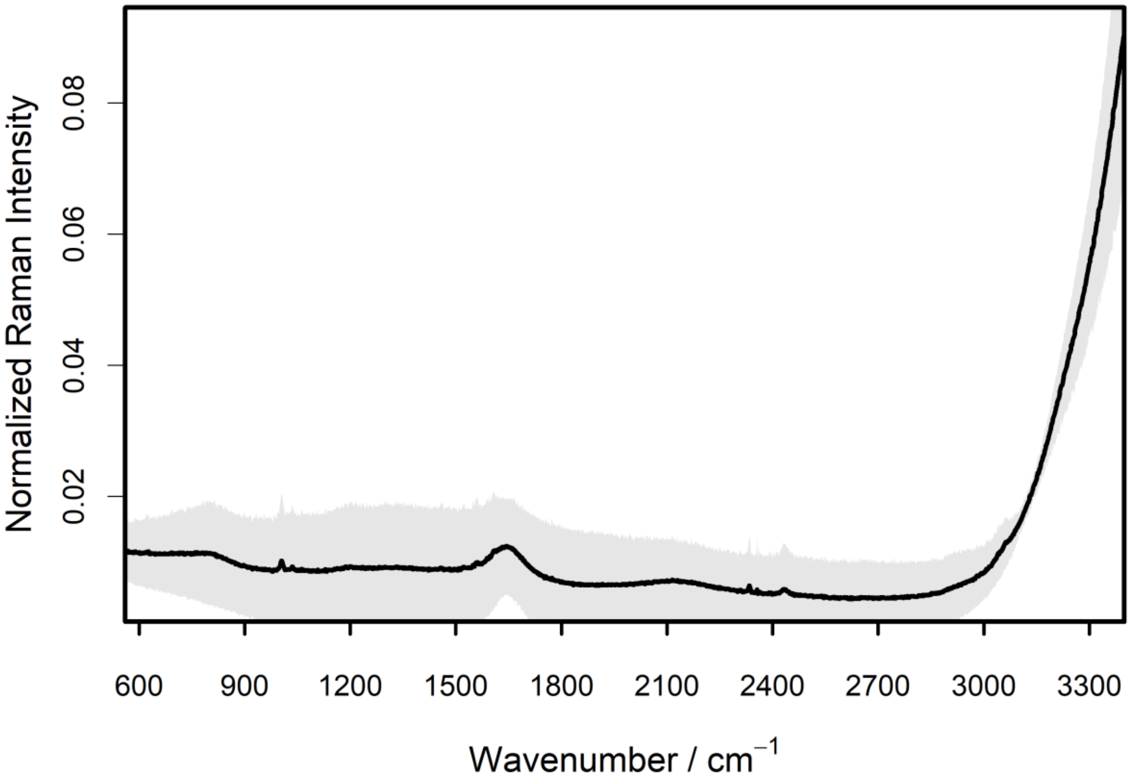

Figure S11. Normalized Raman spectrum of H_2O. The grey area indicates the 95 % confidence interval.

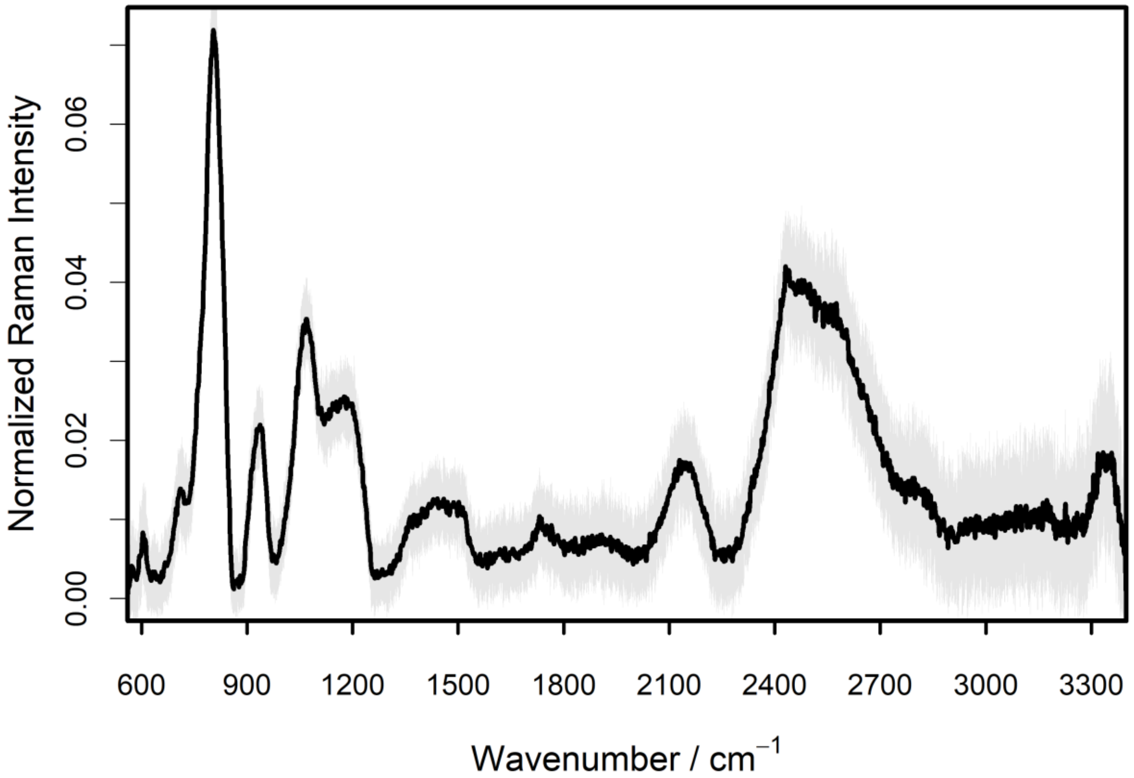

Figure S12. Background corrected and normalized Raman spectrum of H_2O in a microwave tube. The grey area indicates the 95 % confidence interval.

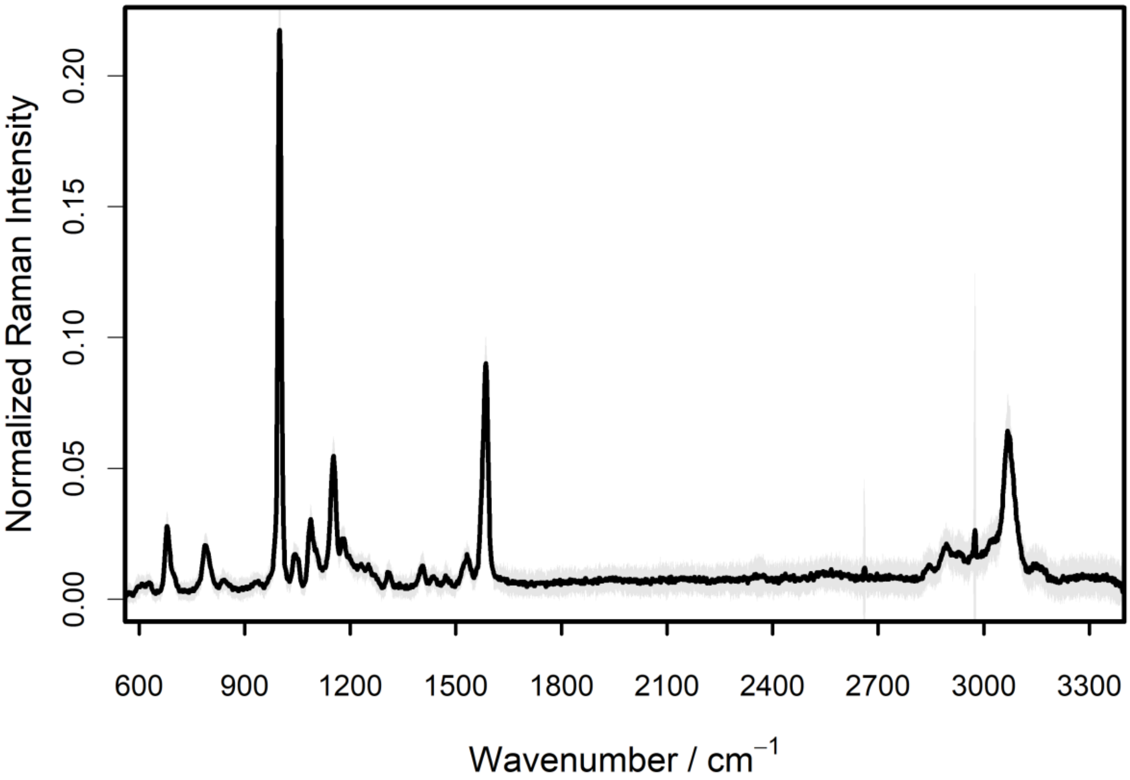

Figure S13. Background corrected and normalized Raman spectrum of [Rh(TPPTS)$_3$]Cl. The grey area indicates the 95 % confidence interval. Band assignment[7]: 3066 (ν(=C-H)), 1585 (ν(C=C)), 1158, 992 (phenyl ring breathing) cm^{-1}.

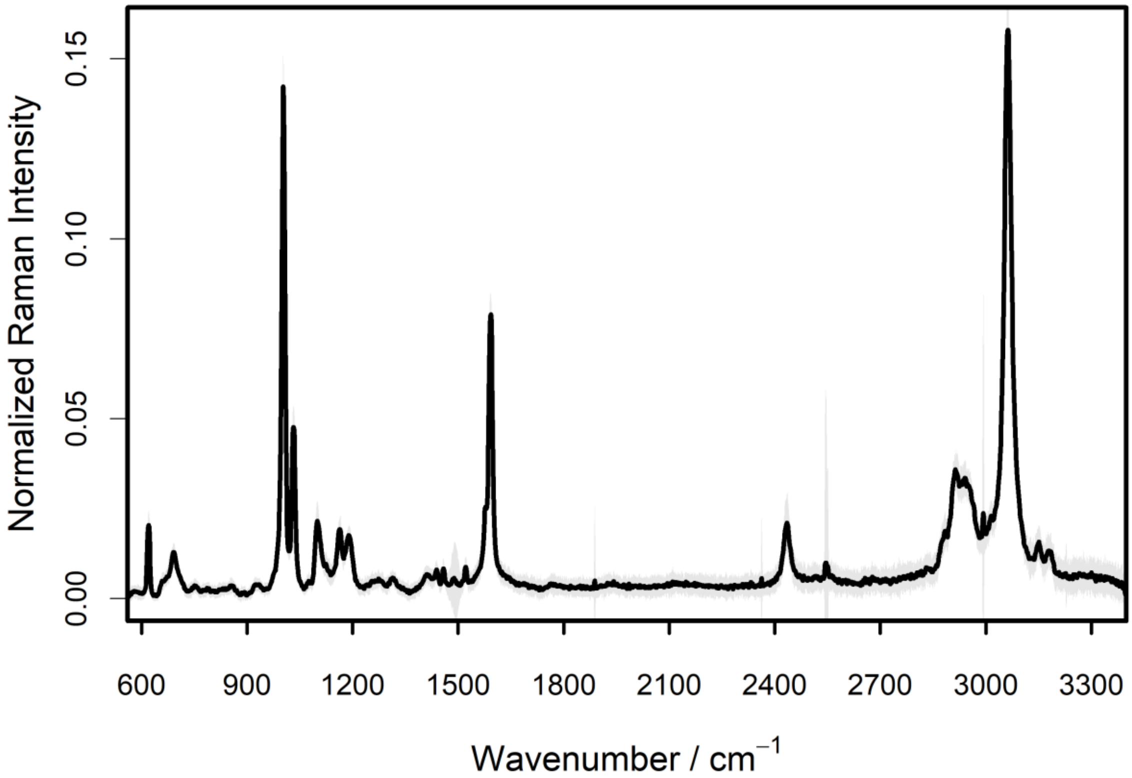

Figure S14. Background corrected and normalized Raman spectrum of $[Rh(dppp)_2]Cl$. The grey area indicates the 95 % confidence interval. Band assignment[7]: 3066 (ν(=C-H)), 2433, 1596 (ν(C=C)), 1095, 1032, 1004 (phenyl ring breathing) cm^{-1}.

3. Additional Details about Advanced Data Analysis

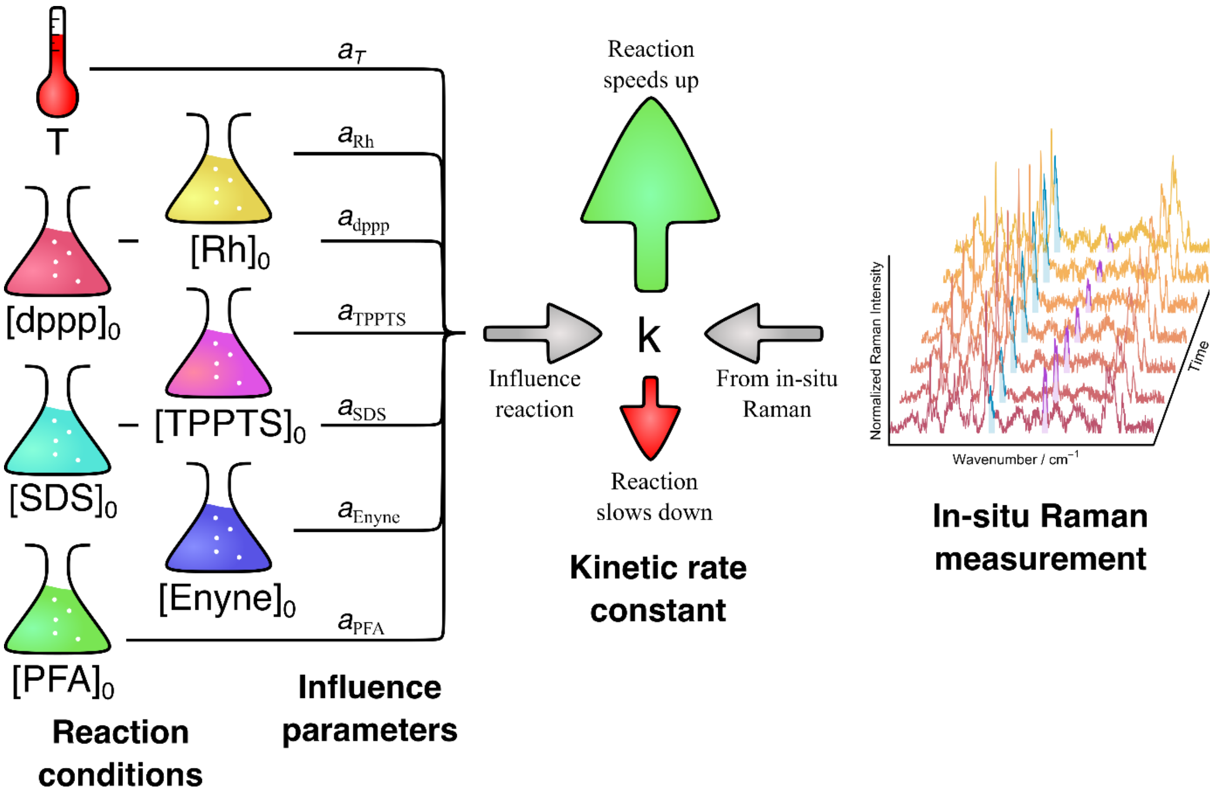

Figure S15. Multiple linear regression model and its parameters. The left side illustrates the 7 considered reaction conditions and their influence parameters a_x. The right side shows some in situ Raman spectra (also shown in Figure 1a). Highlighted in purple and blue are the C≡C and C=O stretching of **Enyne** and **BCP**, respectively. From the in situ Raman spectra the kinetic rate constant k is extracted. The linear model then describes which changes in the reaction condition accelerates or inhibits the Pauson Khand cyclisation.

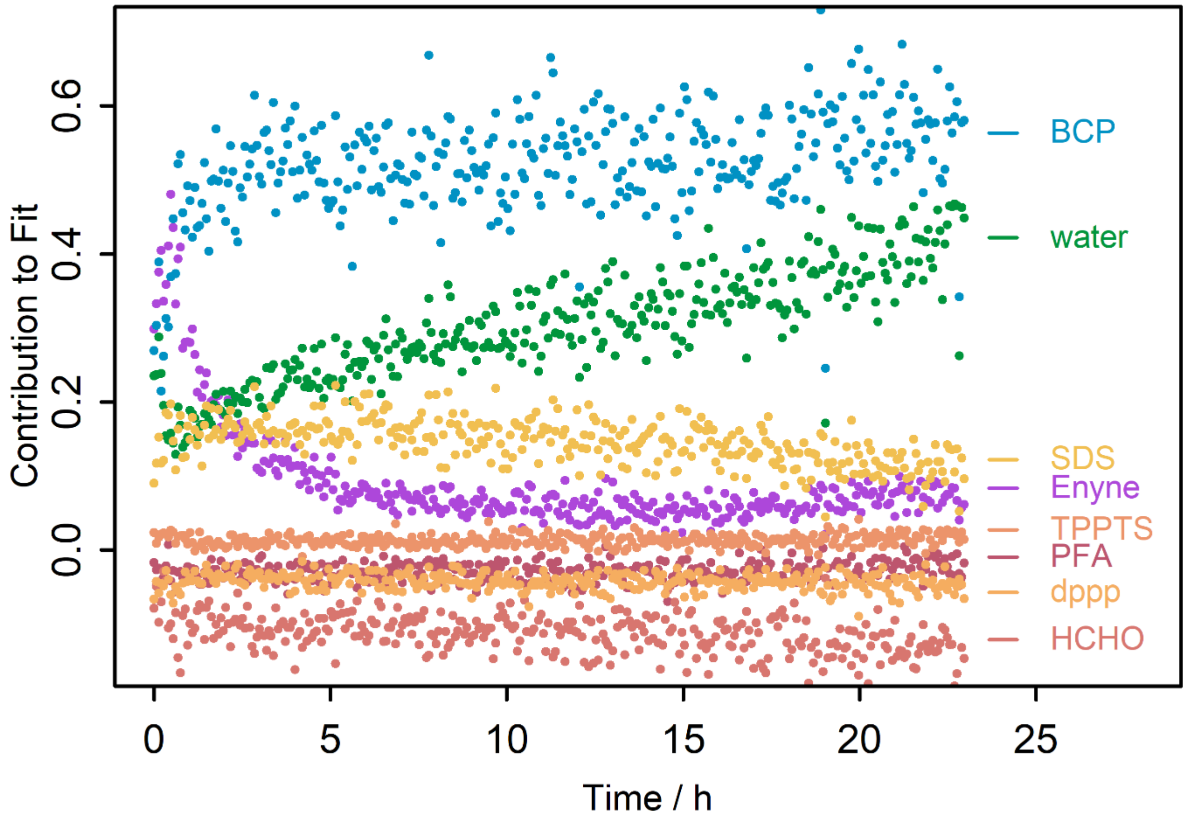

Figure S16. Time-dependent fit concentrations from the linear least-square fitting procedure of an in situ measurement when 8 compounds are used. Most compounds have contributions close to 0. For the 8-compound fit the Raman spectra of PFA, HCHO in H_2O, **Enyne**, **BCP**, TPPTS, dppp, SDS and H_2O were used.

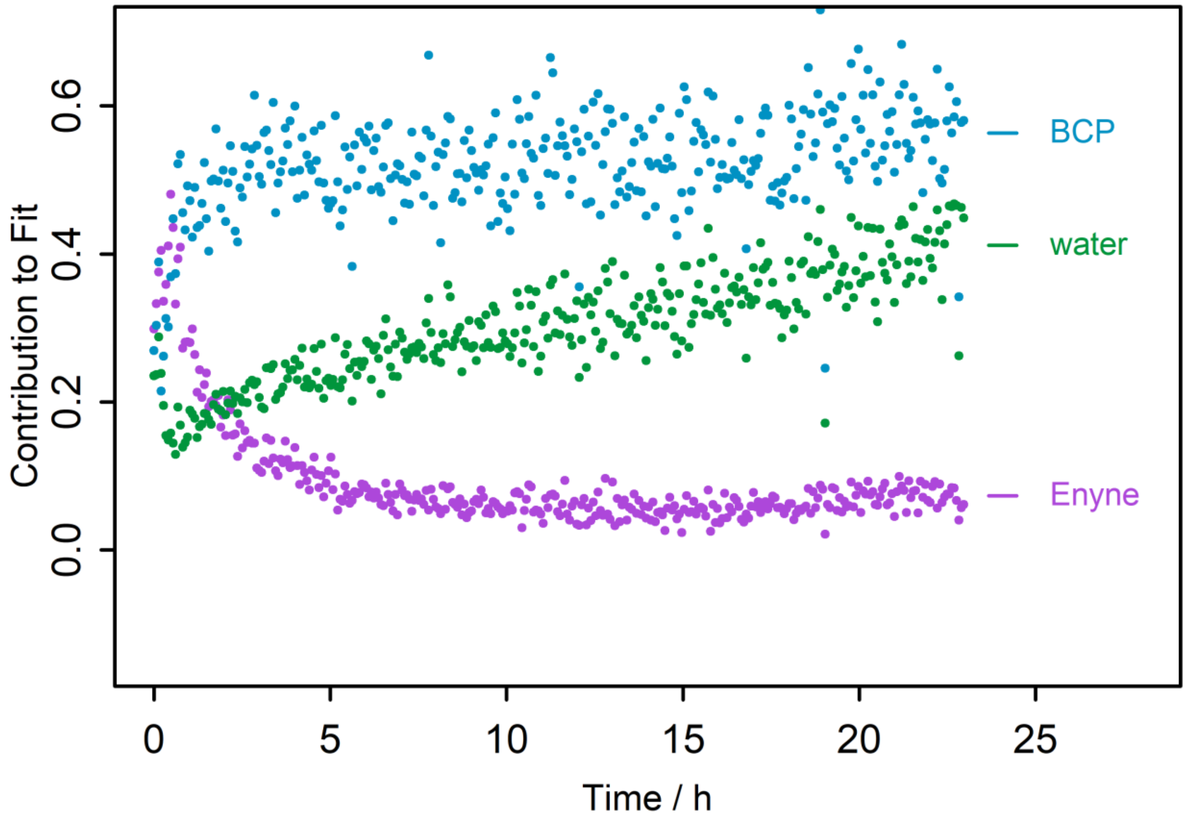

Figure S17. Time-dependent fit concentrations from the linear least-square fitting procedure of an in situ measurement when 3 compounds are used. For the 3-compound fit the Raman spectra of **Enyne**, **BCP** and H_2O were used.

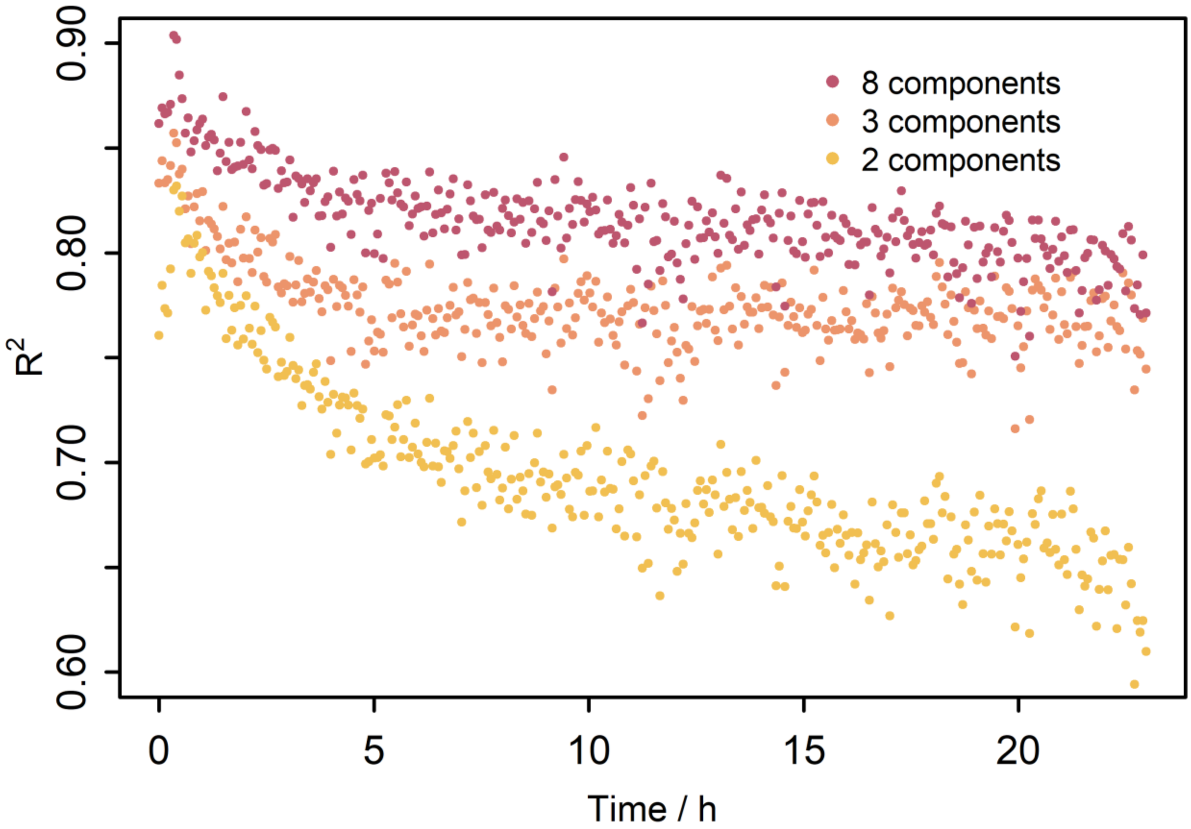

Figure S18. R^2 from the linear least-square fitting procedure of an in situ measurement when 8, 3 or 2 components are used, respectively. R^2 does not change significantly when the number of components is reduced to 3 but drops significantly when only 2 components are used. For the 8-component fit the Raman spectra of PFA, HCHO in H_2O, **Enyne**, **BCP**, TPPTS, dppp, SDS and H_2O were used. For the 3-component fit only the Raman spectra of **Enyne**, **BCP** and H_2O were used, while only the Raman spectra of **Enyne** and **BCP** were used for the 2-component fit.

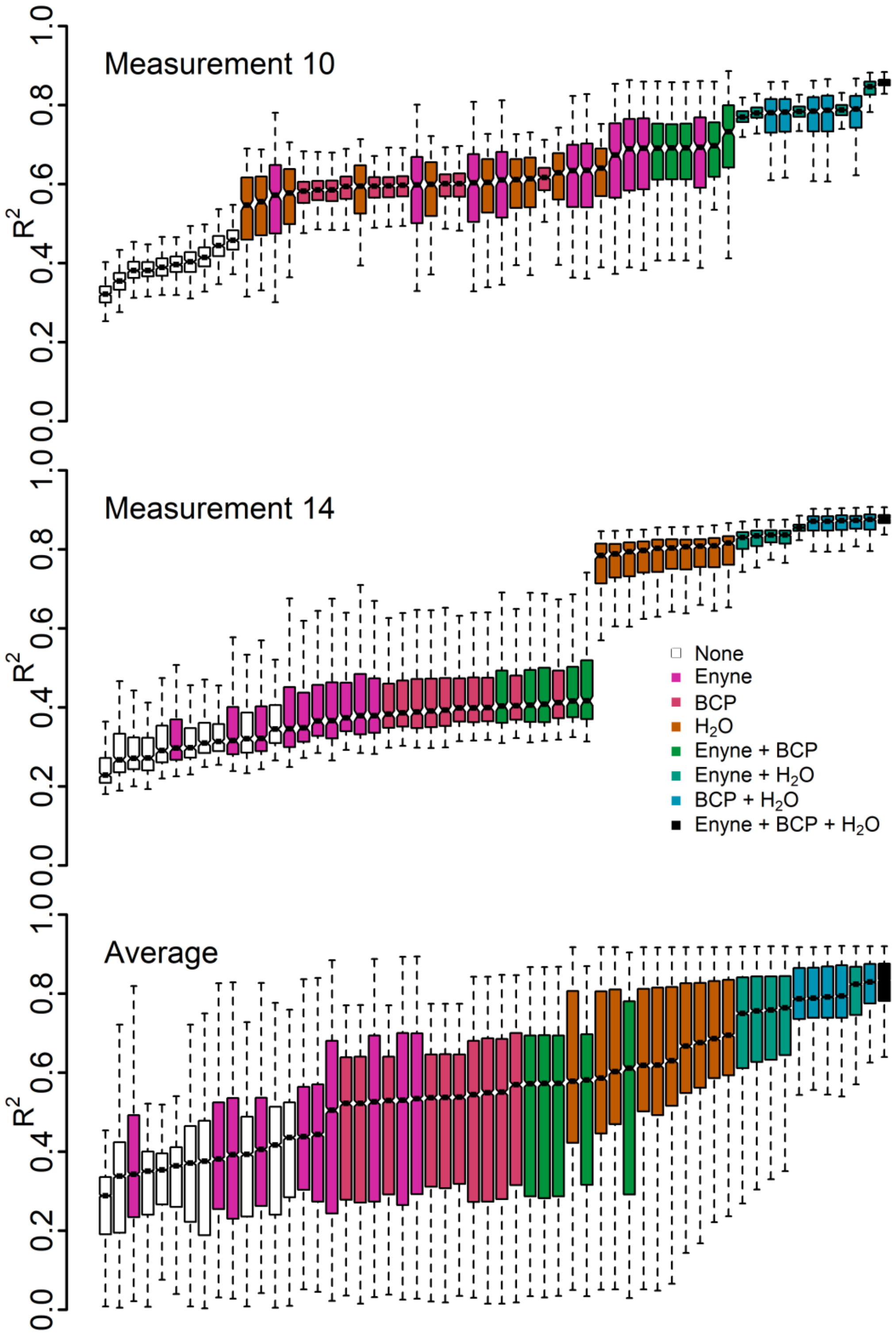

Figure S19. Boxplots (lower whisker, 1st quantile, median, 3rd quantile, upper whisker; outliers are not shown for clarity) representing the R^2 for each of 56 possible combinations to use 3 components from a total of 8 components to fit an in situ Raman measurement. The first 2 plots represent the performance of the fitting procedure on a given individual measurement, here measurements 10 and 14, respectively. The third plot shows the average over all in situ measurements in this work. The colors represent when **Enyne**, **BCP** and/or H_2O are part of the fitting procedure. The black box represents the approach when **Enyne**, **BCP** and H_2O are together used. As can be seen this approach gives the best fit when comparing the fits of all in situ Raman measurements.

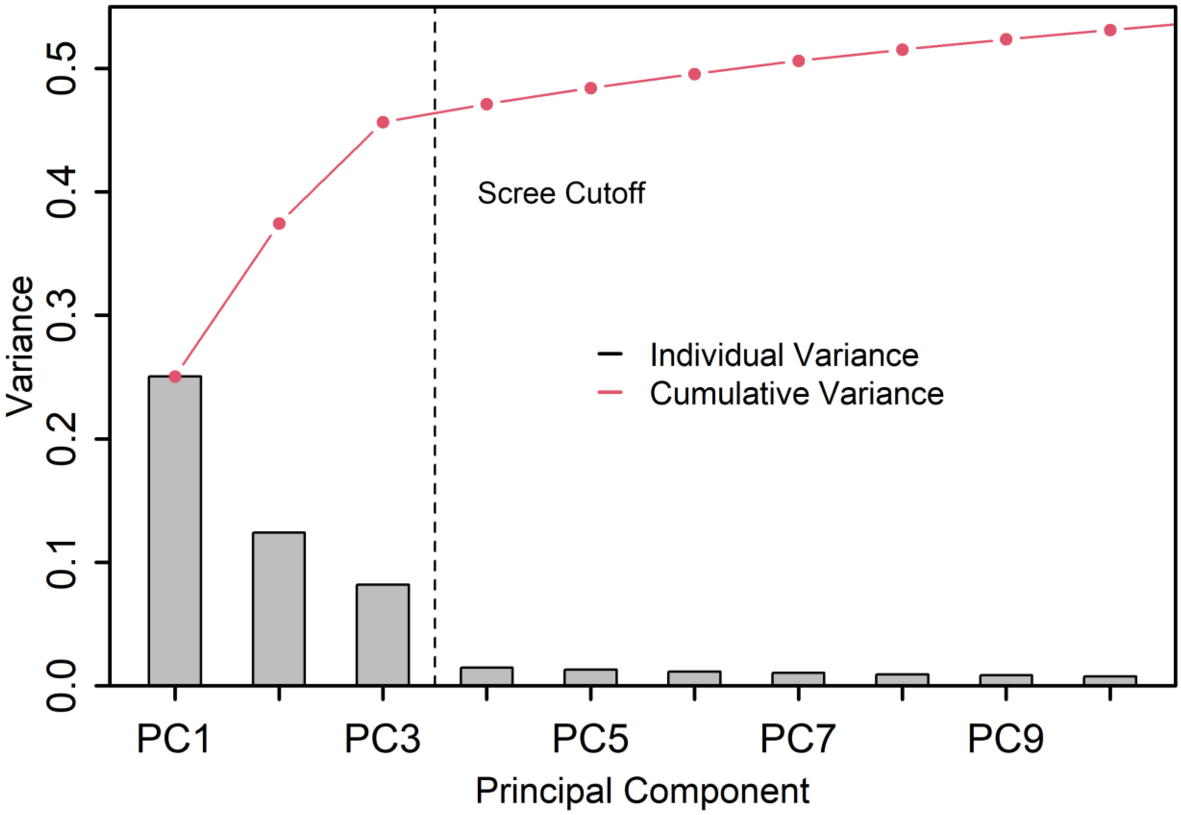

Figure S20. Scree plot from an exemplary PCA. The scree plot indicates that 3 components are mainly responsible for the observed variance according to the elbow criterion.

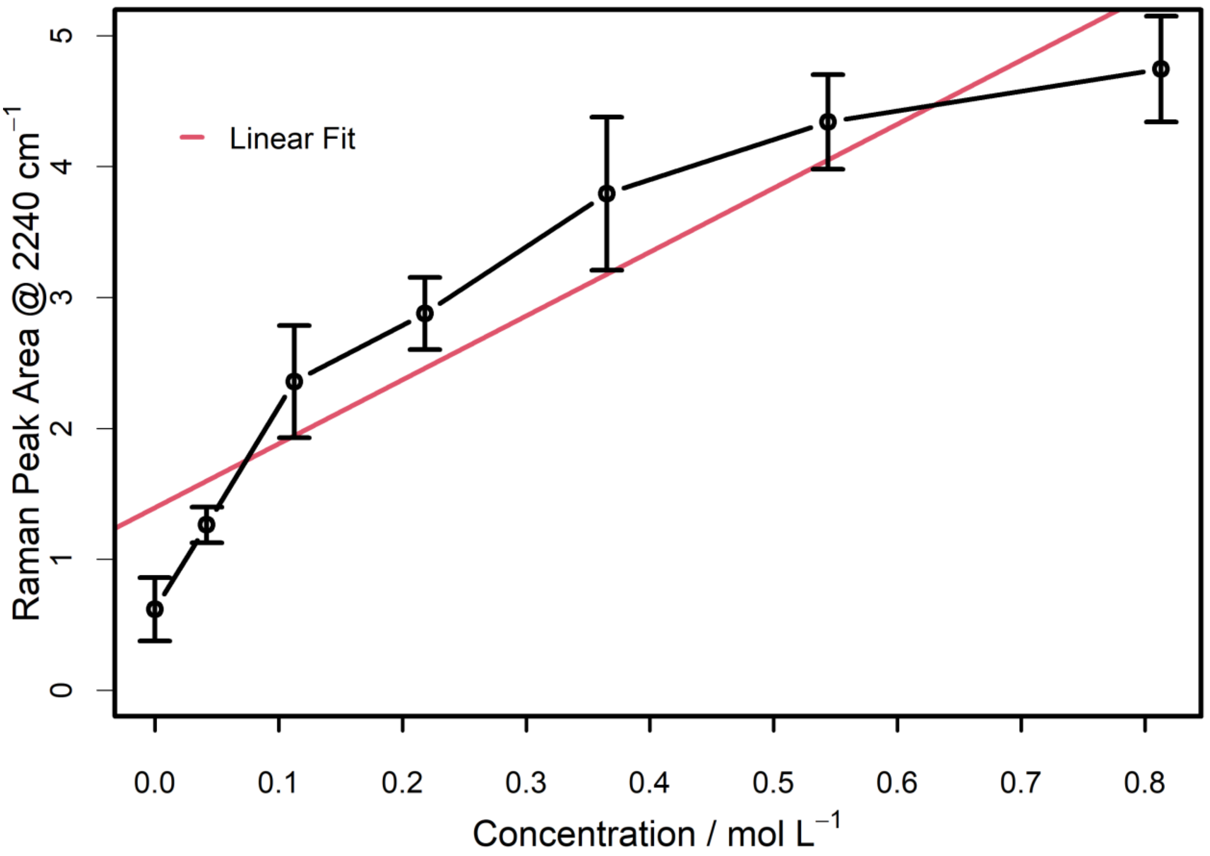

Figure S21. Linear calibration curve for **Enyne** concentration (in H_2O) against the Raman peak at 2238 cm^{-1}. The error bars show the single standard deviation calculated from 9 individual measurements.

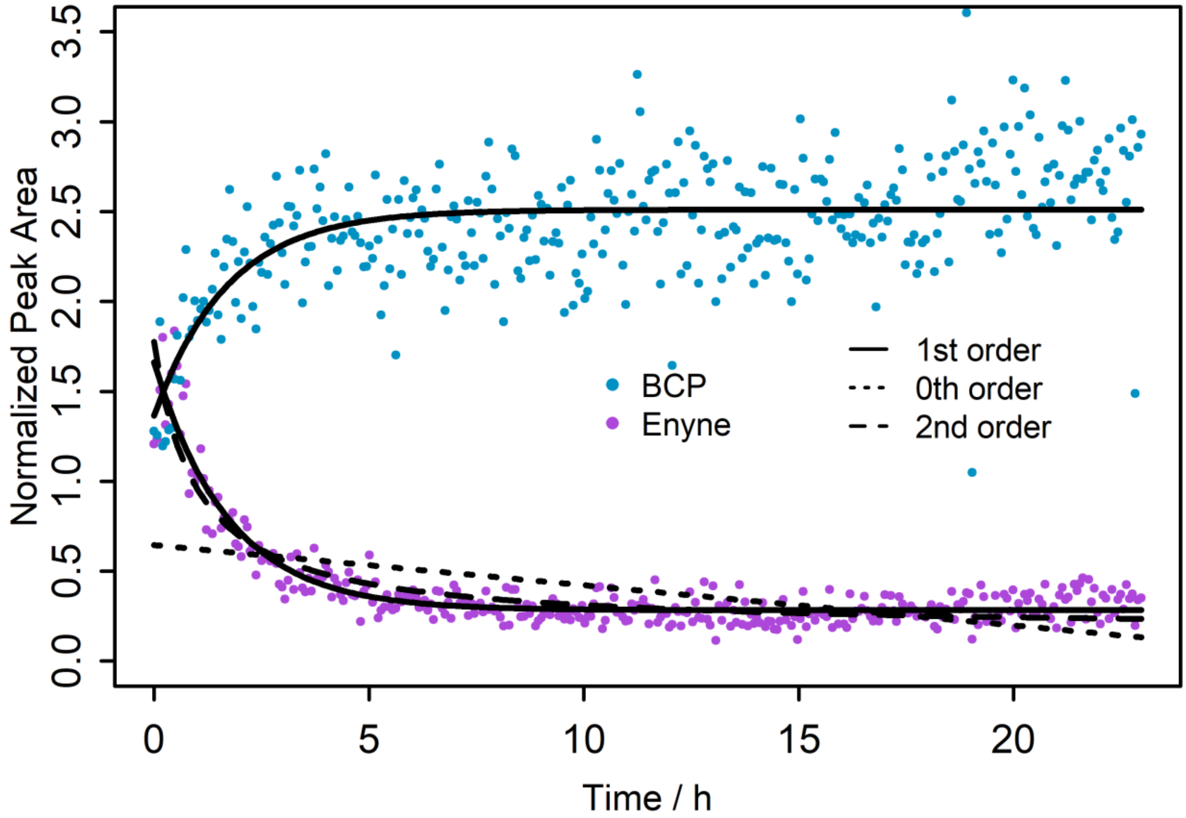

Figure S22. Exemplary peak integration areas for Enyne (2220-2260 cm^{-1}) and BCP (1620-1690 cm^{-1}) from an in situ Raman measurement. The resulting peak areas are used to initialize the subsequent MCR-ALS algorithm. The peak area-time curve for **Enyne** is fitted with a zeroth, first and second order kinetic fit. The first order fit describes the dataset best. Residual standard errors: 0.2232 (0th order), 0.0924 (1st order), 0.1084 (2nd order). 1st order rate constant for **Enyne** consumption: 1.61 * 10^{-4} s^{-1}, 1st order rate constant for **BCP** generation: 1.63 * 10^{-4} s^{-1}.

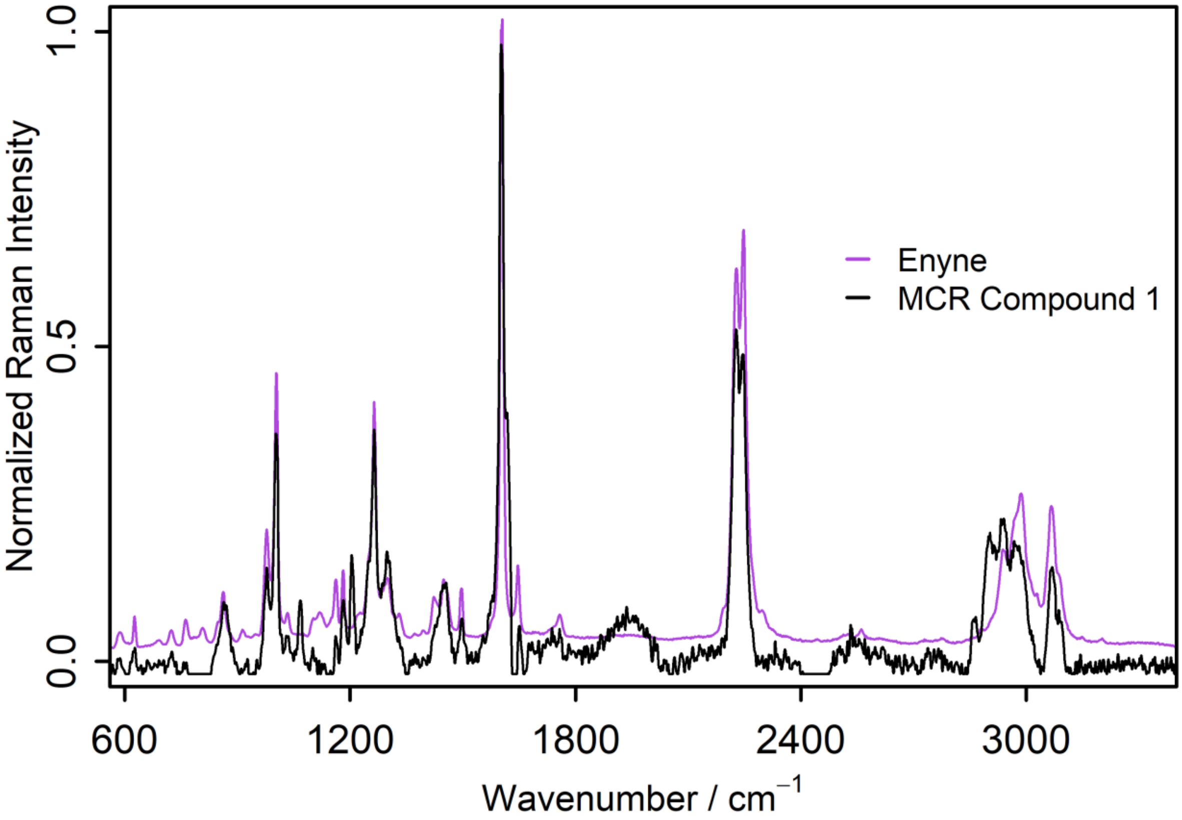

Figure S23. Raman spectrum of pure **Enyne** in comparison to the Raman spectrum extracted by MCR from the in situ measurements.

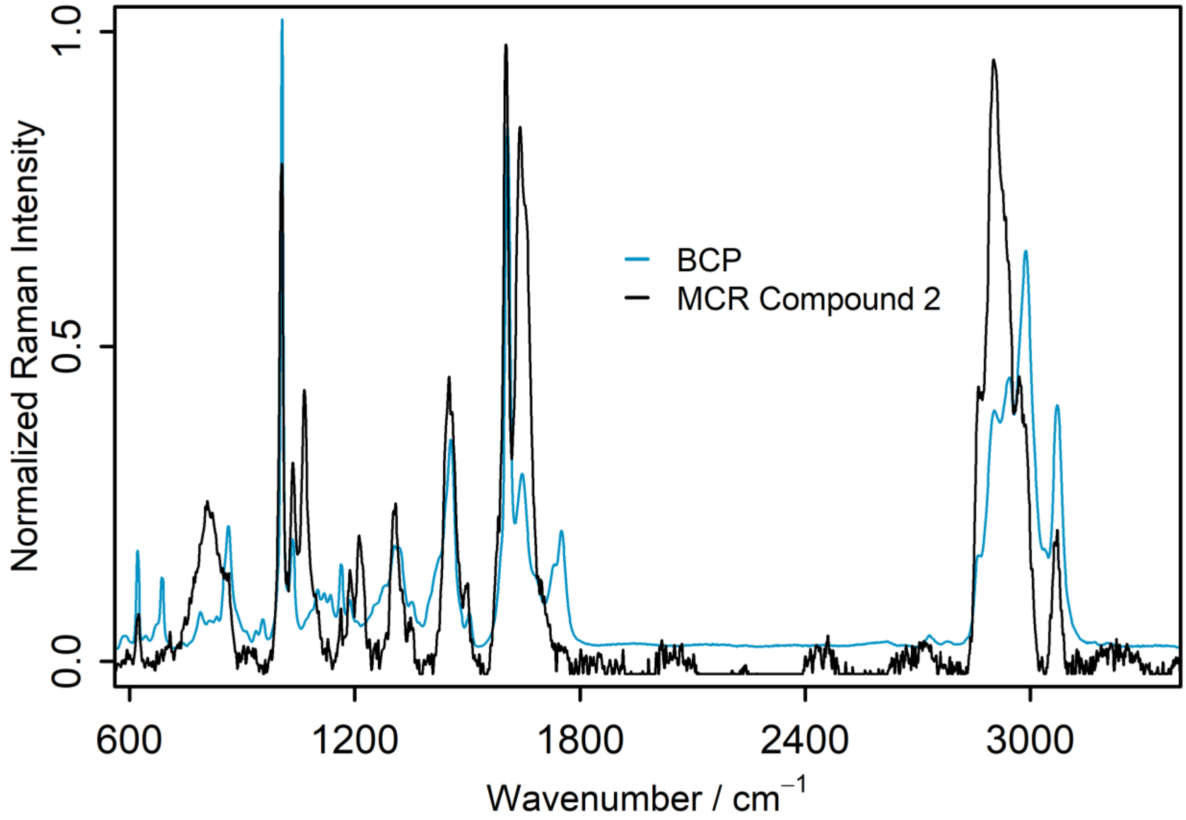

Figure S24. Raman spectrum of pure **BCP** in comparison to the Raman spectrum extracted by MCR from the in situ measurements.

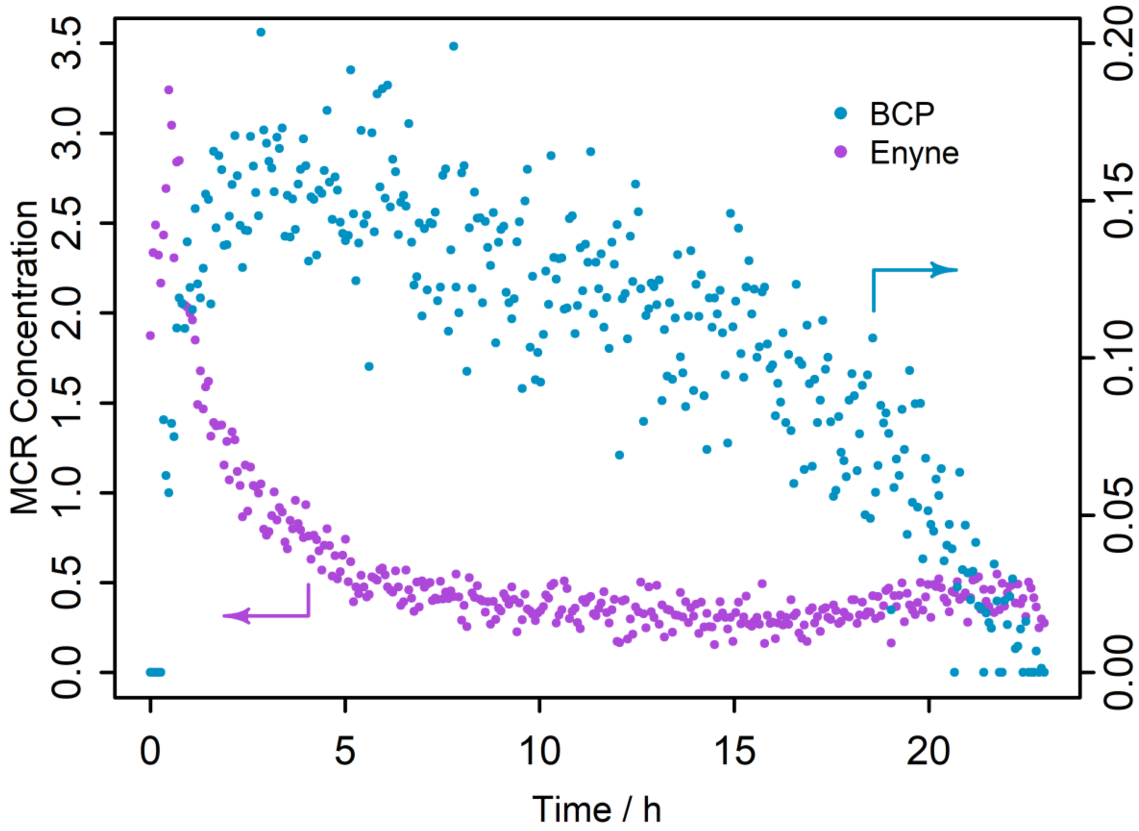

Figure S25. MCR concentration profiles for **Enyne** and **BCP**.

4. Additional NMR Spectra

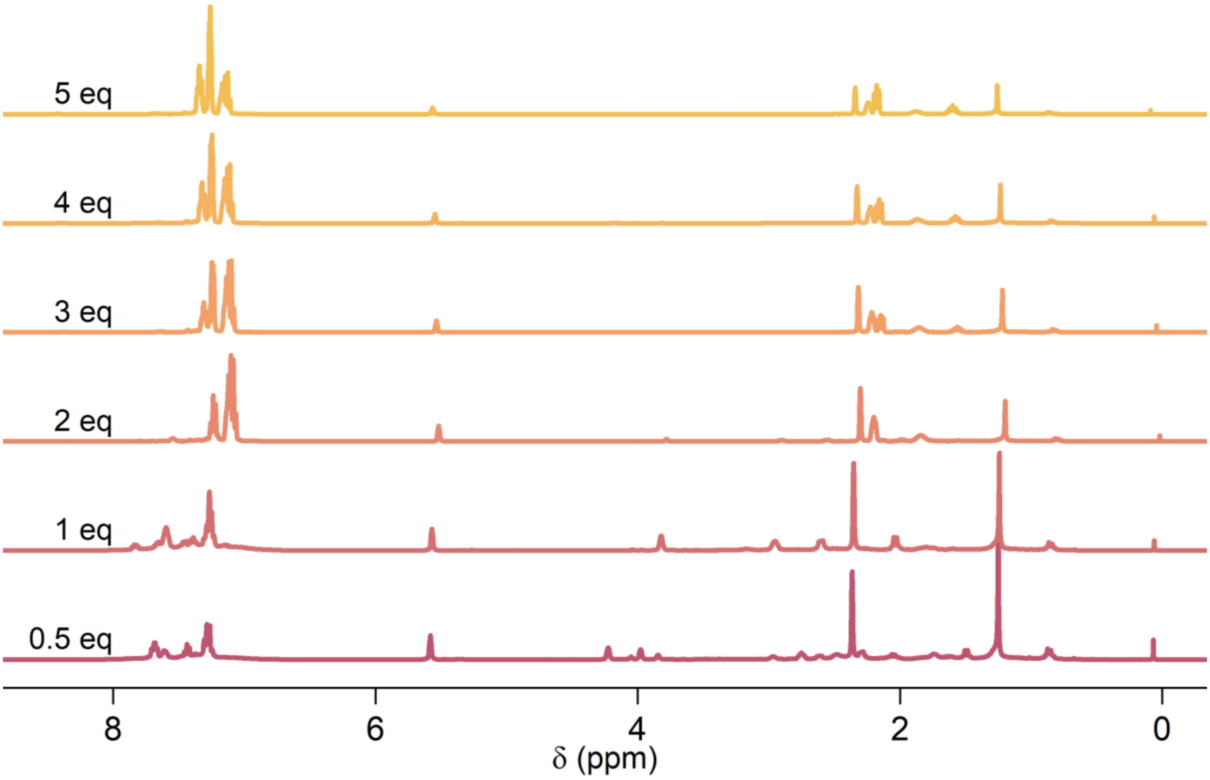

Figure S26. Background corrected ^{1}H-NMR spectra of $[RhCl(cod)]_2$ + different equivalents of dppp in $DCCl_3$. The spectra were referenced to the solvent signal at 7.26 ppm.

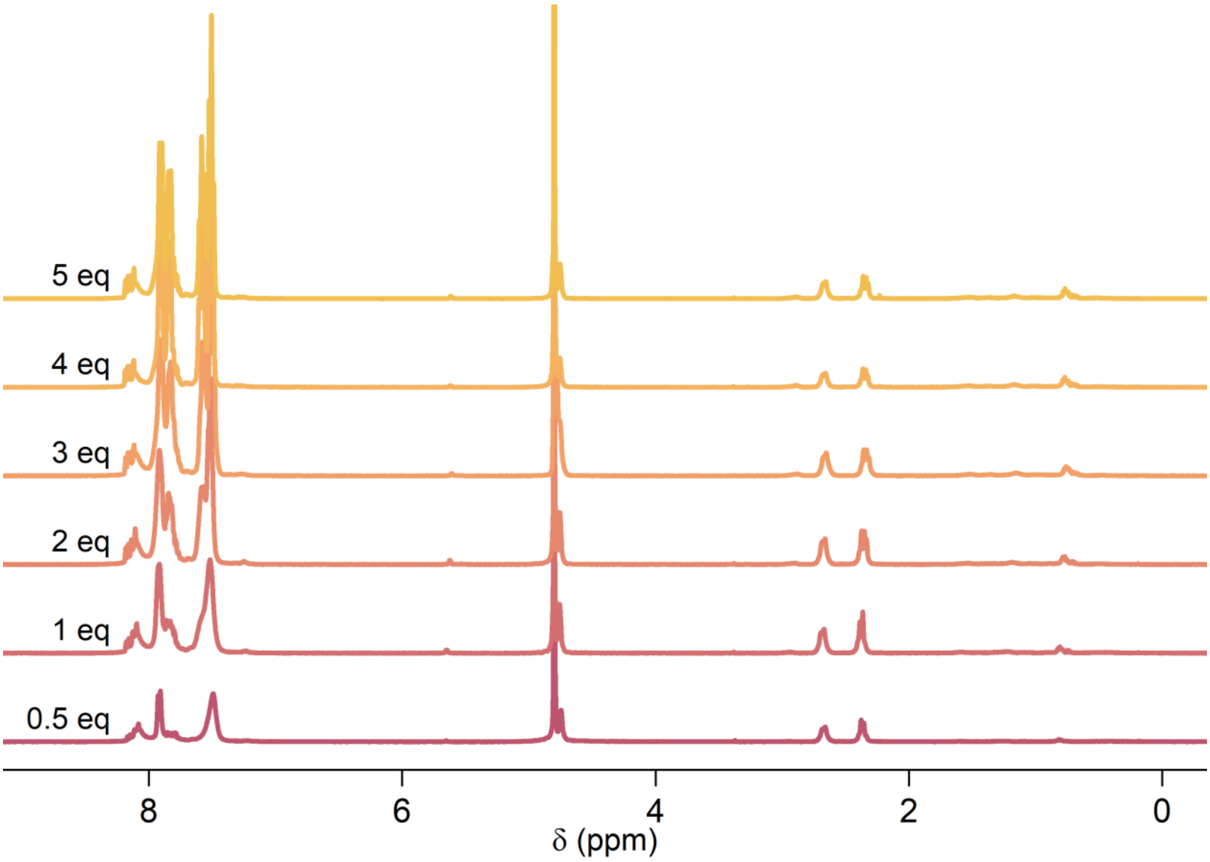

Figure S27. Background corrected ^{1}H-NMR spectra of $[RhCl(cod)]_2$ + different equivalents of TPPTS in H_2O. The spectra were referenced to the solvent signal at 4.86 ppm.

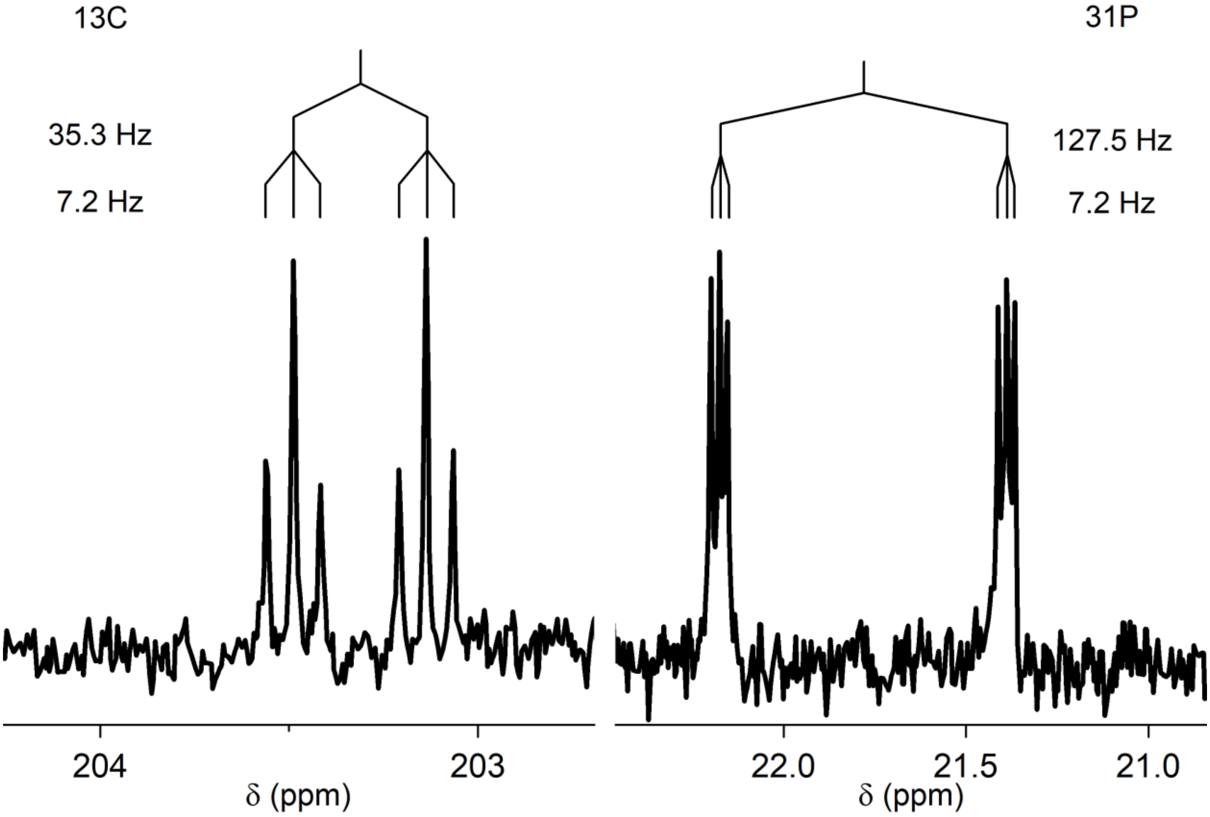

Figure S28. Background corrected ^{13}C- and ^{31}P-NMR spectra of a Rh-catalyzed Pauson-Khand type cyclization after 990 mins in $DCCl_3$ using ^{13}C-PFA as CO surrogate. The spectra prove the presents of $[Rh(dppp)(CO)_2]^+$. The ^{31}P signal at 21.8 ppm is the only visible signal in the ^{31}P spectrum.

5. Density Functional Theory Calculated Raman Spectra

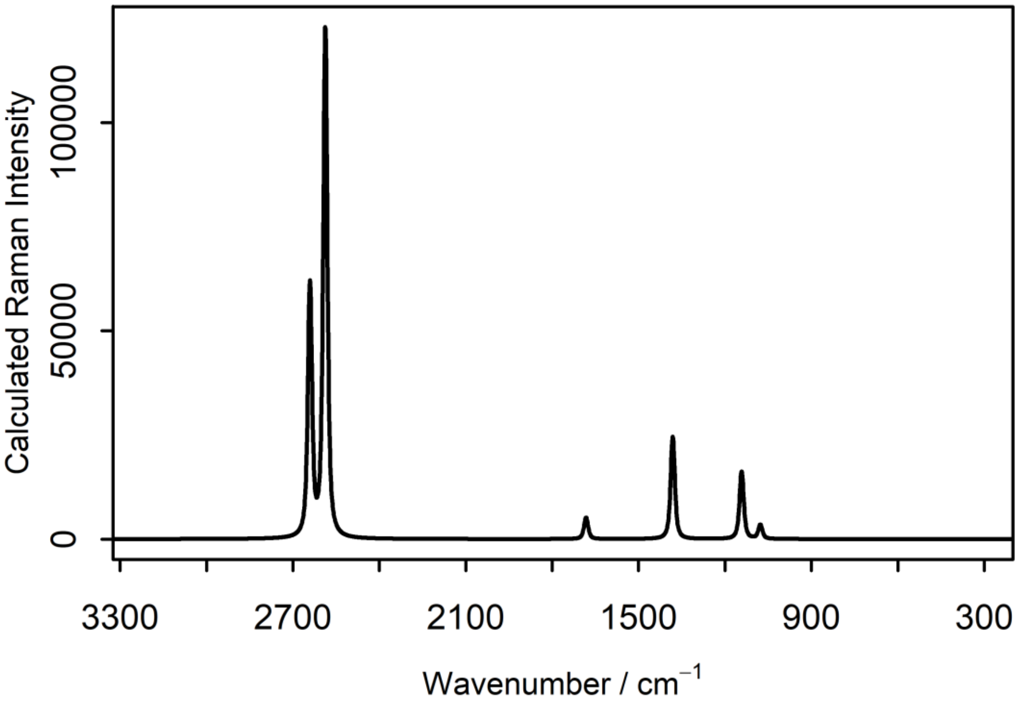

Figure S29. DFT calculated Raman spectrum of HCHO (FWHM: 18 cm^{-1}).

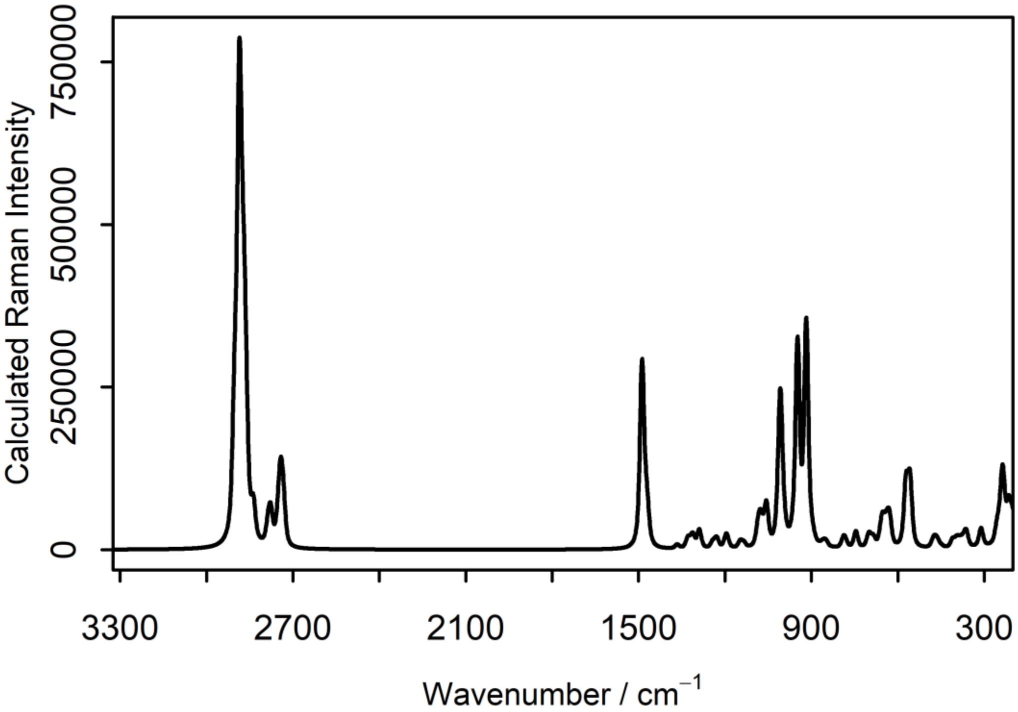

Figure S30. DFT calculated Raman spectrum of dppp (FWHM: 18 cm^{-1}).

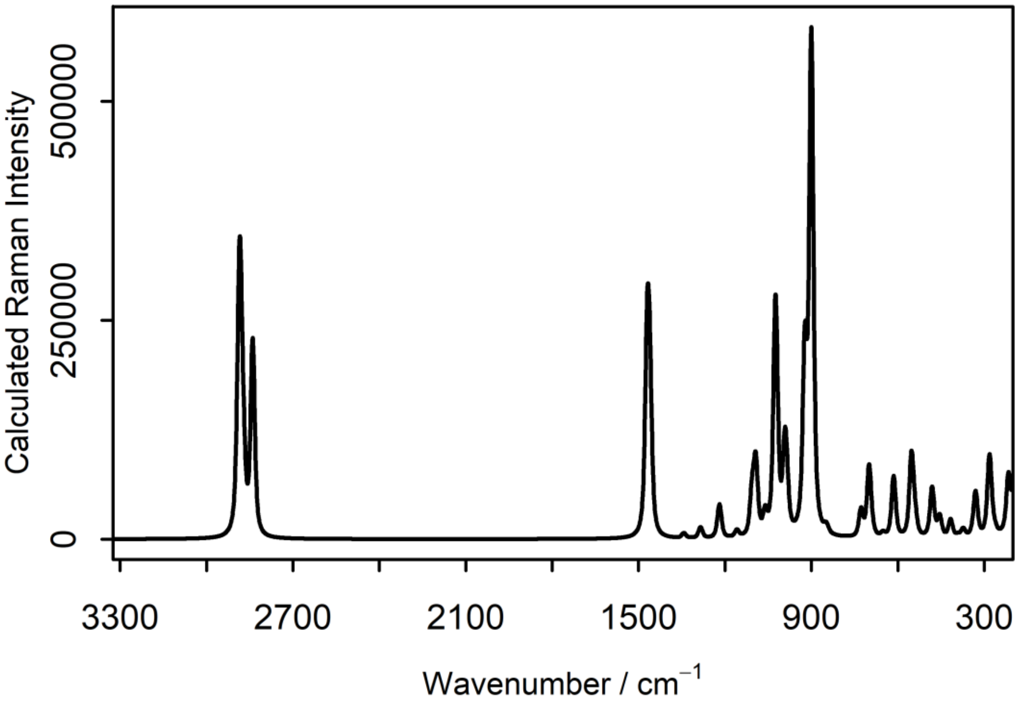

Figure S31. DFT calculated Raman spectrum of TPPTS (FWHM: 18 cm^{-1}).

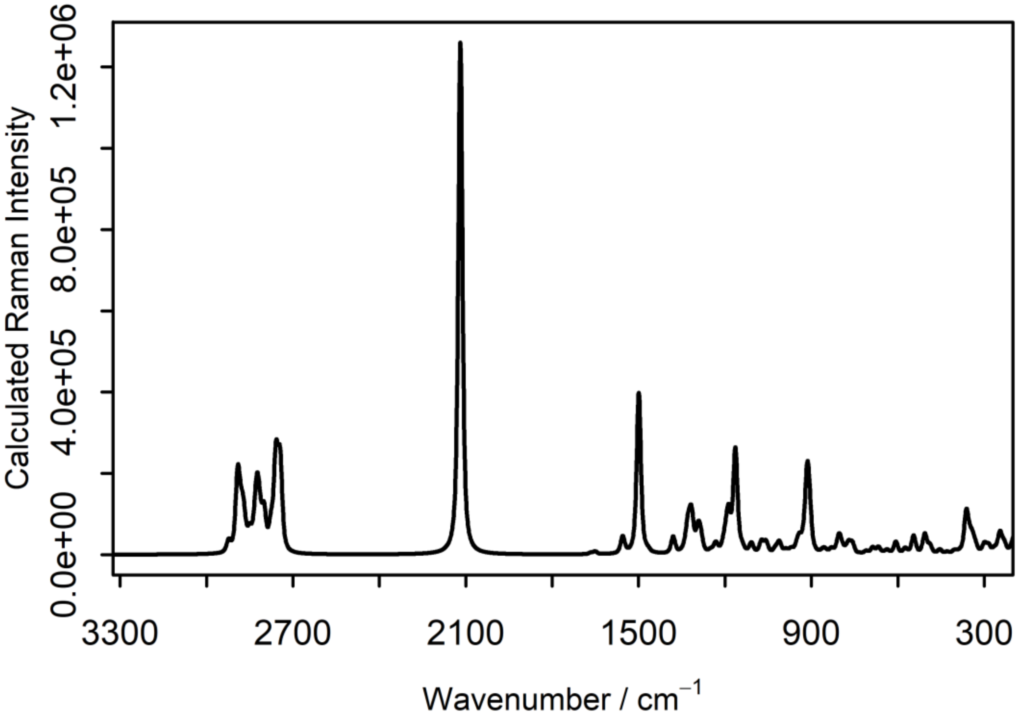

Figure S32. DFT calculated Raman spectrum of **Enyne** (FWHM: 18 cm^{-1}).

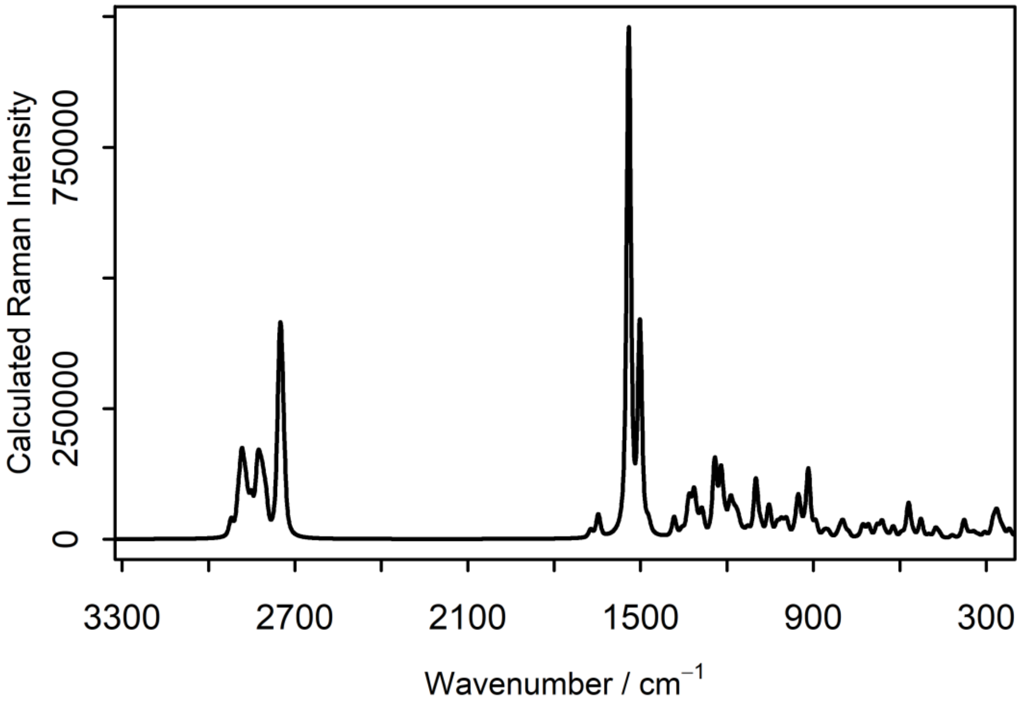

Figure S33. DFT calculated Raman spectrum of (*S*)-**BCP** (FWHM: 18 cm^{-1}).

6. References

[1] R Core team, *R: A Language and Environment for Statistical Computing;* R Foundation for Statistical Computing, Vienna, Austria, **2020**.

[2] T. Bocklitz, A. Walter, K. Hartmann, P. Rösch, J. Popp, *Anal. Chim. Acta* **2011**, *704*, 47–56.

[3] C. G. Ryan, E. Clayton, W.L. Griffin, S.H. Sie, D.R. Cousens, *Nucl. Instrum. Methods Phys. Res, Sect. B* **1988**, *34*, 396–402.

[4] M. Morhác, *Peaks: Peaks*, **2012**.

[5] M. P. Wand, M. C. Jones, *Kernel Smoothing;* Springer US, Boston, MA, **1995**.

[6] M. Wand, *KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones (1995)*, **2020**.

[7] G. Socrates, *Infrared and Raman Characteristic Group Frequencies*: Tables and Charts; John Wiley & Sons, West Sussex, **2007**.

[8] K. Pearson, *London Edinburgh Dublin Phil. Mag. J. Sci.* **1901**, *2*, 559–572.

[9] I. T. Jolliffe, J. Cadima, *Philos. Trans. A Math Phys. Eng. Sci.* **2016**, *374*, 20150202.

[10] M. Garrido, F. X. Rius, M. S. Larrechi, *Anal. Bioanal. Chem.* **2008**, *390*, 2059–2066.

[11] K. M. Mullen, *ALS: Multivariate Curve Resolution Alternating Least Squares*, **2015**.