

Automated LC-MS Analysis and Data Extraction for High-Throughput Chemistry

Supporting Information

Joseph Mason,^{*a} Harry Wilders,^{b,c} David J. Fallon,^b Ross P. Thomas,^{b,c} Jacob T. Bush,^b
Nicholas C. O. Tomkinson^c and Francesco Rianjongdee^a

^a Medicinal Chemistry, GSK Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, United Kingdom.

^b Chemical Biology, GSK Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, United Kingdom.

^c Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow, G1 1XL, United Kingdom.

Contents

Installation Information:	3
Pre-Requisite Libraries	3
Required Input Files	3
Running an Analysis	4
Additional Optional Parameters	4
Metrics for Analysis.....	6
Standard Options	6
Normalised Options	6
Compound Summary Table.....	6
Experimental Information.....	8
General Experimental	8
Parallel Synthesis of Photoaffinity Bit Library and Meta-Analysis by PyParse	13
Synthetic Procedure.....	13
Meta-Analysis by PyParse	13
C-H Activation Reaction Optimisation Plate	16
Plate Design	16
Procedure to Prepare the Reaction Optimisation Plate	18
Hit Validation Graph for the Desired Product.....	19
Heatmap for Regioisomer at 0.71 min (4.2)	19
Heatmap for Regioisomer at 0.73 min (4.1)	20
Synthesis of (2-(quinolin-3-yl)oxazol-4-yl)methanol (4.1)	22
Synthesis of (5-(quinolin-3-yl)oxazol-4-yl)methanol (4.2)	23
NMR Spectra	24
¹ H NMR for (2-(quinolin-3-yl)oxazol-4-yl)methanol (4.1)	24
¹³ C NMR for (2-(quinolin-3-yl)oxazol-4-yl)methanol (4.1)	24
ROESY NMR for (2-(quinolin-3-yl)oxazol-4-yl)methanol (4.1)	25
¹ H NMR for (5-(quinolin-3-yl)oxazol-4-yl)methanol (4.2)	26
¹³ C NMR for (5-(quinolin-3-yl)oxazol-4-yl)methanol (4.2)	26
ROESY NMR for (5-(quinolin-3-yl)oxazol-4-yl)methanol (4.2)	27
References	28

Installation Information:

Full documentation for installing and running PyParse can be found at the GitHub Pages site: <https://thatchemistryguy.github.io/PyParse/index.html>

The above documentation also contains instructions and an additional example dataset to help users get started with using PyParse.

Pre-Requisite Libraries

PyParse was developed and run using Python 3.7 and the following libraries. Libraries which are not part of the standard Python 3.7 (or equivalent version) library should be installed as per the installation guide for each library.

A conda.txt file to generate a suitable conda environment available from the PyParse GitHub repository: <https://github.com/thatchemistryguy/PyParse>

- os (part of standard Python library)
- sys (part of standard Python library)
- time (part of standard Python library)
- argparse (part of standard Python library)
- re (part of standard Python library)
- math (part of standard Python library)
- statistics (part of standard Python library)
- zipfile (part of standard Python library)
- shutil (part of standard Python library)
- logging (part of standard Python 3.7 library)
- matplotlib (version 3.3.2)
- seaborn (version 0.11.0)
- pandas (version 1.1.3)
- rdkit (version 2020.09.1)

PyParse.py should be saved locally with read, write and execute privileges.

Required Input Files

Two input files are required for a PyParse analysis:

- The OpenLynx™ Browser Report file generated by a Waters™ Corporation LC-MS machine
- A platemap, saved as a .csv file

The platemap contains information on the compound expected in each well of the analysis plate and should include the following headers as a minimum:

- Well: The well ID of the plate, in the format "A1", "A2", etc
- Desired product SMILES: the smiles string for the desired product in each well

Optionally, this platemap can also contain the following headers, which are case-insensitive:

- Limiting reactant SMILES: the SMILES string for the limiting reactant
- ByproductX SMILES: the smiles string for a by-product, where X corresponds to a positive integer
- [COMPOUND CLASS] RT: the retention time of a compound, where [COMPOUND CLASS] is replaced with either “desired product”, “limiting reactant” or “byproductX”.
- [COMPOUND CLASS] Name: the specified common of a compound, where [COMPOUND CLASS] is replaced with either “desired product”, “limiting reactant” or “byproductX”. If a name is not provided, a generic name will be generated, e.g. “Product1”

An example platemap is shown below:

	A	B	C	D	E
1 Well	Desired Product SMILES	Limiting Reactant SMILES	InternalSTD SMILES	Byproduct1 SMILES	
2 A1	<chem>OCC1=COC(C2=CC(C=CC=3)=C3N=C2)=N1</chem>	<chem>BrC1=CC2=CC=CC=C2N=C1</chem>	<chem>C1(N(C2=CC=CC=C2)CC3=CC=CC=C3)=CC=CC=C1</chem>	<chem>OCC1=C(C2=CN=C(C=CC=3)C3=C2)OC(C4=CC5=C(N=C4)C=CC=C5)=N1</chem>	
3 A2	<chem>OCC1=COC(C2=CC(C=CC=3)=C3N=C2)=N1</chem>	<chem>BrC1=CC2=CC=CC=C2N=C1</chem>	<chem>C1(N(C2=CC=CC=C2)CC3=CC=CC=C3)=CC=CC=C1</chem>	<chem>OCC1=C(C2=CN=C(C=CC=3)C3=C2)OC(C4=CC5=C(N=C4)C=CC=C5)=N1</chem>	
4 A3	<chem>OCC1=COC(C2=CC(C=CC=3)=C3N=C2)=N1</chem>	<chem>BrC1=CC2=CC=CC=C2N=C1</chem>	<chem>C1(N(C2=CC=CC=C2)CC3=CC=CC=C3)=CC=CC=C1</chem>	<chem>OCC1=C(C2=CN=C(C=CC=3)C3=C2)OC(C4=CC5=C(N=C4)C=CC=C5)=N1</chem>	
5 A4	<chem>OCC1=COC(C2=CC(C=CC=3)=C3N=C2)=N1</chem>	<chem>BrC1=CC2=CC=CC=C2N=C1</chem>	<chem>C1(N(C2=CC=CC=C2)CC3=CC=CC=C3)=CC=CC=C1</chem>	<chem>OCC1=C(C2=CN=C(C=CC=3)C3=C2)OC(C4=CC5=C(N=C4)C=CC=C5)=N1</chem>	
6 A5	<chem>OCC1=COC(C2=CC(C=CC=3)=C3N=C2)=N1</chem>	<chem>BrC1=CC2=CC=CC=C2N=C1</chem>	<chem>C1(N(C2=CC=CC=C2)CC3=CC=CC=C3)=CC=CC=C1</chem>	<chem>OCC1=C(C2=CN=C(C=CC=3)C3=C2)OC(C4=CC5=C(N=C4)C=CC=C5)=N1</chem>	
7 A6	<chem>OCC1=COC(C2=CC(C=CC=3)=C3N=C2)=N1</chem>	<chem>BrC1=CC2=CC=CC=C2N=C1</chem>	<chem>C1(N(C2=CC=CC=C2)CC3=CC=CC=C3)=CC=CC=C1</chem>	<chem>OCC1=C(C2=CN=C(C=CC=3)C3=C2)OC(C4=CC5=C(N=C4)C=CC=C5)=N1</chem>	
8 A7	<chem>OCC1=COC(C2=CC(C=CC=3)=C3N=C2)=N1</chem>	<chem>BrC1=CC2=CC=CC=C2N=C1</chem>	<chem>C1(N(C2=CC=CC=C2)CC3=CC=CC=C3)=CC=CC=C1</chem>	<chem>OCC1=C(C2=CN=C(C=CC=3)C3=C2)OC(C4=CC5=C(N=C4)C=CC=C5)=N1</chem>	
9 A8	<chem>OCC1=COC(C2=CC(C=CC=3)=C3N=C2)=N1</chem>	<chem>BrC1=CC2=CC=CC=C2N=C1</chem>	<chem>C1(N(C2=CC=CC=C2)CC3=CC=CC=C3)=CC=CC=C1</chem>	<chem>OCC1=C(C2=CN=C(C=CC=3)C3=C2)OC(C4=CC5=C(N=C4)C=CC=C5)=N1</chem>	
10 A9	<chem>OCC1=COC(C2=CC(C=CC=3)=C3N=C2)=N1</chem>	<chem>BrC1=CC2=CC=CC=C2N=C1</chem>	<chem>C1(N(C2=CC=CC=C2)CC3=CC=CC=C3)=CC=CC=C1</chem>	<chem>OCC1=C(C2=CN=C(C=CC=3)C3=C2)OC(C4=CC5=C(N=C4)C=CC=C5)=N1</chem>	
11 A10	<chem>OCC1=COC(C2=CC(C=CC=3)=C3N=C2)=N1</chem>	<chem>BrC1=CC2=CC=CC=C2N=C1</chem>	<chem>C1(N(C2=CC=CC=C2)CC3=CC=CC=C3)=CC=CC=C1</chem>	<chem>OCC1=C(C2=CN=C(C=CC=3)C3=C2)OC(C4=CC5=C(N=C4)C=CC=C5)=N1</chem>	
12 A11	<chem>OCC1=COC(C2=CC(C=CC=3)=C3N=C2)=N1</chem>	<chem>BrC1=CC2=CC=CC=C2N=C1</chem>	<chem>C1(N(C2=CC=CC=C2)CC3=CC=CC=C3)=CC=CC=C1</chem>	<chem>OCC1=C(C2=CN=C(C=CC=3)C3=C2)OC(C4=CC5=C(N=C4)C=CC=C5)=N1</chem>	
13 A12	<chem>OCC1=COC(C2=CC(C=CC=3)=C3N=C2)=N1</chem>	<chem>BrC1=CC2=CC=CC=C2N=C1</chem>	<chem>C1(N(C2=CC=CC=C2)CC3=CC=CC=C3)=CC=CC=C1</chem>	<chem>OCC1=C(C2=CN=C(C=CC=3)C3=C2)OC(C4=CC5=C(N=C4)C=CC=C5)=N1</chem>	
14 B1	<chem>OCC1=COC(C2=CC(C=CC=3)=C3N=C2)=N1</chem>	<chem>BrC1=CC2=CC=CC=C2N=C1</chem>	<chem>C1(N(C2=CC=CC=C2)CC3=CC=CC=C3)=CC=CC=C1</chem>	<chem>OCC1=C(C2=CN=C(C=CC=3)C3=C2)OC(C4=CC5=C(N=C4)C=CC=C5)=N1</chem>	
15 B2	<chem>OCC1=COC(C2=CC(C=CC=3)=C3N=C2)=N1</chem>	<chem>BrC1=CC2=CC=CC=C2N=C1</chem>	<chem>C1(N(C2=CC=CC=C2)CC3=CC=CC=C3)=CC=CC=C1</chem>	<chem>OCC1=C(C2=CN=C(C=CC=3)C3=C2)OC(C4=CC5=C(N=C4)C=CC=C5)=N1</chem>	
16 B3	<chem>OCC1=COC(C2=CC(C=CC=3)=C3N=C2)=N1</chem>	<chem>BrC1=CC2=CC=CC=C2N=C1</chem>	<chem>C1(N(C2=CC=CC=C2)CC3=CC=CC=C3)=CC=CC=C1</chem>	<chem>OCC1=C(C2=CN=C(C=CC=3)C3=C2)OC(C4=CC5=C(N=C4)C=CC=C5)=N1</chem>	
17 B4	<chem>OCC1=COC(C2=CC(C=CC=3)=C3N=C2)=N1</chem>	<chem>BrC1=CC2=CC=CC=C2N=C1</chem>	<chem>C1(N(C2=CC=CC=C2)CC3=CC=CC=C3)=CC=CC=C1</chem>	<chem>OCC1=C(C2=CN=C(C=CC=3)C3=C2)OC(C4=CC5=C(N=C4)C=CC=C5)=N1</chem>	

Note: Multiple products and limiting reagents can be specified in a plate (one per well), but only one internal standard may be used in a plate. Multiple by-products may be specified, but each new by-product should be placed in a new column with a new header (E.g. “Byproduct1 SMILES”, “Byproduct2 SMILES”, etc). There is no limit placed on the number of by-products which can be specified.

Running an Analysis

The analysis should be run specifying the processed LC-MS data file (.rpt file) and platemap (saved as .csv file) as positional arguments; values for optional parameters may also be specified as required. The following optional parameters are specific to the plate being analysed:

- **plate_col_no (-r): Integer value**
Number of columns in the analysis plate
- **plate_row_no (-c): Integer value**
Number of rows in the analysis plate
- **plot_type (-pt): String value - Choose from “Parea”, “P/STD”, “P/SM+P”, “corrParea”, “corrP/STD” or “corrP/SM+P”**
Metric to use for generating outputs, see section titled “Metrics for Analysis” below.
- **output (-o): String value**
Directory to save the PyParse output to

An example of a typical analysis, run from the command line, where the working directory is the same as that of the input files and PyParse.py, is shown here:

```
python pyparse.py LCMS_data.rpt my_platemap.csv -r 8 -c 12 -pt corrP/STD -o my_output_directory
```

Additional Optional Parameters

These parameters are unlikely to require changing from one plate analysis to the next, but may require adjustment for alternative setups/chromatographic methods at different institutions.

- **validate (-V): True/False**
Specify whether or not to run hit validation algorithm.
- **verbose (-v): True/False**
Specify whether logging should be run in debug mode or not.
- **mass_abs_tol (-mat): Float value**
How close an observed m/z should be to a calculate mass to be considered a match.
- **time_abs_tol (-tat): Float value**
How close two retention times should be to be considered part of the same cluster/relate to the same compound.
- **uv_abs_tol (-uat): Integer value**
How close should the wavelengths of the maxima in a UV absorption profile be for two eluting peaks to be considered a match.
- **min_peak_area (-mpa): Float value**
Minimum percentage UV peak area required for inclusion in analysis.
- **min_massconf_threshold (-mmt): Integer value**
The lowest mass confidence (percentage height of m/z peak divided by the sum of percentage heights of all observed m/z peaks) at which a hit will still be recorded.
- **min_uv_threshold (-mut): Integer value**
Minimum height of a peak in the UV absorption profile (AU vs wavelength) required to be saved for use in the hit validation algorithm.
- **uv_match_threshold (-umt): Float value**
Parameter to determine how many UV maxima a hit needs to match those found for the bulk values before it is considered a tentative hit and flagged to the user for further analysis.
- **uv_cluster_threshold (-uct): Float value**
Parameter to determine how many times a UV maximal wavelength should be observed for a compound (across all hits in a cluster) compared to the number of times the most commonly observed wavelength is seen, as a fraction, to be considered a “required value” for that compound.
- **massconf_threshold (-mt): Float value**
The value for which, when it’s multiplied by the mean mass confidence for a cluster, the mass confidence for a hit in that cluster is deemed too small and is flagged to the user for further analysis.
- **cluster_size_threshold (-cst): Float value**
The value for which, when it’s multiplied by the number of hits in the largest cluster, the number of hits in a cluster is deemed too small and is discarded from further analysis.
- **min_no_of_wells (-now): Integer value**
Minimum number of hit-containing wells required for a compound to allow the full validation process to run.
- **points_per_trace (-ppt): Integer value**
The maximum number of data points of a UV chromatogram to save. This value should be reduced if the program is found to be running exceptionally slowly.
- **mass_or_area (-moa): “mass_conf” or “area”**
Parameter to determine how a hit is chosen when more than one hit is found in a well for a compound.
- **calc_higherions (-chi): True/False**
Specifies to PyParse whether higher ions (such as [M+2H]²⁺ and [M+3H]³⁺) should be used to assign LC-MS peaks to a compound when they are observed.
- **generate_csv (-g): True/False**
Choose whether to generate and save a csv of the outputTable.

- **generate_zip (-z): True/False**
Choose whether to generate and save a zip file containing all visualisations and outputs.

Metrics for Analysis

These options determine the value by which the heatmap and trellised pie charts are generated, and which is the best well in the case of product compounds.

Standard Options

- **Parea:** Heatmap is coloured by the UV percentage peak area of the product. Suggested option for parallel synthesis and D2B experiments.
- **P/SM+P:** Heatmap is coloured by the UV area of the product, divided by the sum of peak areas for the product and reactant in that well. Requires a reactant to be specified for each well.
- **P/STD:** Heatmap is coloured by the UV area of the product divided by the UV area of the internal standard in that well. Requires an internal standard to be specified.

Normalised Options

Useful for plates with multiple products with different UV activities or conversions. Values are between 0 and 1.

- **corrParea:** The observed UV area is normalised to the maximum UV area observed for that product across the whole plate.
- **corrP/SM+P:** The value of P/SM+P for each well is normalised against the maximum value that was observed across the plate for that product.
- **corrP/STD:** The value of P/STD for each well is normalised against the maximum value that was observed across the plate for that product. Suggested option for reaction screening plate and library validation experiments.

Compound Summary Table

The compound summary table lists all compounds which were provided in the platemap, and provides a summary for each, including: retention time, observed m/z values and the best well. The “Best Well” is calculated using the selected metric for analysis in the case of products (e.g. “corrP/STD”), or the LC-MS UV percentage area for all other compound classes. A “Best Purity” column is also present, and this relates to the LC-MS UV percentage area for that compound in the “Best Well”.

The Overlap Detection column is designed to alert the user where PyParse detects peak overlap with the compound in question in the “Best Well” for that compound. This feature is designed to caution the user to scrutinise the well manually before proceeding further with analysis, i.e. submission to a biological assay. It is important to note that PyParse is designed to work in conjunction with analysis by the chemist, who should ensure the quality of the data obtained. This is particularly pertinent in cases where an internal standard is used, which should be selected to ensure it does not co-elute with another compound.

An example of the compound summary table is provided below, and was taken from the PyParse output of the reaction optimisation plate described below.

An example of the compound summary table is also provided on the GitHub repository for PyParse.

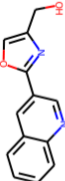
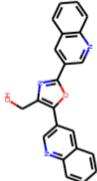
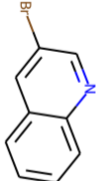
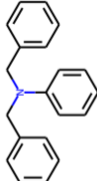
Name	SMILES	Mass 1	Mass 2	Mass 3	Retention Time	m/z-	m/z+	Best Well	Purity of Best Well	Overlap Detection	Potential Conflicts
Product1		226.07	0.0	0	0.71	225.1	227.24	D6	70.51	No peak overlap detected.	No potential conflicts found.
Byproduct1		353.12	0.0	0	1.01	352.11	354.32	C7	61.77	No peak overlap detected.	No potential conflicts found.
Reactant1		206.97	208.97	0	1.04	206.29	208.11	C10	61.89	No peak overlap detected.	No potential conflicts found.
InternalSTD		273.15	0.0	0	1.49	-	274.28	D10	23.31	No peak overlap detected.	No potential conflicts found.

Figure S1. The compound summary table generated by PyParse for the reaction optimisation plate described.

Experimental Information

General Experimental

All reagents were obtained from commercial suppliers and were used without further purification unless otherwise stated.

Anhydrous solvents were obtained from Sigma-Aldrich and transferred from the Sure Seal™ bottles by syringe, under an atmosphere of nitrogen.

All intermediate compounds are at least 90% pure. These purity assessments are made using the UV absorption spectra collected as part of LC-MS analysis or NMR. For known compounds, only LC-MS and ¹H NMR data is presented.

Reaction mixtures from both the D2B dataset and the reaction optimisation dataset, and isolated products, were analysed using a Waters Acquity UPLC equipped with a Waters Acquity PDA (UPLC LG 500 nm), configured for a sample rate of 40 points/sec, at a range of 210 nm to 350 nm with a resolution of 1.2 nm.

LCMS System A

Column: 50 mm x 2.1 mm ID, Acquity UPLC CSH C₁₈ column.

Flow Rate: 1 mL/min

Temp: 40 °C

UV detection range: 210 to 350 nm

Mass spectrum: Recorded on a mass spectrometer using alternate-scan positive and negative mode electrospray ionisation, with a mass range of 100 – 1000.

Solvents:

A: Water + 0.1% Formic Acid.

B: Acetonitrile + 0.1% Formic Acid

Gradient:	<i>Time (min.)</i>	<i>A%</i>	<i>B%</i>
	0	97	3
	1.5	0.1	99.9
	1.9	0.1	99.9
	2.0	97	3

LCMS System B

Column: 50 mm x 2.1 mm ID, Acquity UPLC CSH C₁₈ column.

Flow Rate: 1 mL/min

Temp: 40 °C

UV detection range: 210 to 350 nm

Mass spectrum: Recorded on a mass spectrometer using alternate-scan positive and negative mode electrospray ionisation, with a mass range of 100 – 1000.

Solvents:

A: 10 mM solution of ammonium bicarbonate in water adjusted to pH 10 with ammonia solution.

B: Acetonitrile

Gradient:	<i>Time (min.)</i>	<i>A%</i>	<i>B%</i>
	0	100	0
	1.5	3	97
	1.9	3	97
	2.0	100	0

Mass Directed Auto Purification (MDAP) was undertaken using the conditions given below. The UV detection was an averaged signal across wavelengths between 210 nm to 350 nm and mass spectra were recorded on a mass spectrometer using alternate-scan positive and negative mode electrospray ionisation.

MDAP Method A

MDAP Method A was conducted on an XBridge Prep Phenyl (typically 150 mm x 30 mm i.d. 5 µm packing diameter) at ambient temperature.

Temp: ambient

Injection volume: varied, typically 800 µL

UV detection range: 210 to 350 nm

Mass spectrum: Recorded on a mass spectrometer using alternate-scan positive and negative mode electrospray ionisation.

Solvents:

A = 0.1% Formic acid in water.

B = Methanol

The gradient was delivered by two pumps in order to use at-column dilution when injecting the sample. The gradient program is summarised below:

Chromatographic Pump:

Gradient:	Time (min.)	A%	B%	Flow Rate (ml/min)
	0	65	35	4
	4	65	35	4
	4.5	65	35	20
	16	35	65	20
	16.5	10	90	20
	16.6	0	100	20
	19	0	100	20

Regeneration Pump:

Gradient:	Time (min)	A%	B%	Flow Rate (ml/min)
	0	75	25	36
	4	75	25	36
	4.5	65	35	20
	16	35	65	20
	16.5	10	90	20
	16.6	0	100	20
	19	0	100	20

MDAP Method B

MDAP Method B was conducted on an XBridge Prep C18 (typically 150 mm x 30 mm i.d. 5 µm packing diameter) at ambient temperature.

Temp: ambient

Injection volume: varied, typically 1300 µL

UV detection range: 210 to 350 nm

Mass spectrum: Recorded on a mass spectrometer using alternate-scan positive and negative mode electrospray ionisation.

Solvents:

A = 0.1% Formic acid in water.

B = Methanol

The gradient was delivered by two pumps in order to use at-column dilution when injecting the sample. The gradient program is summarised below:

Chromatographic Pump:

Gradient:	<i>Time (min.)</i>	<i>A%</i>	<i>B%</i>	<i>Flow Rate (ml/min)</i>
	0	65	35	4
	4	65	35	4
	4.5	65	35	20
	16	45	55	20
	16.5	10	90	20
	16.6	0	100	20
	19	0	100	20

Regeneration Pump:

Gradient:	<i>Time (min)</i>	<i>A%</i>	<i>B%</i>	<i>Flow Rate (ml/min)</i>
	0	75	25	36
	4	75	25	36
	4.5	65	35	20
	16	35	55	20
	16.5	10	90	20
	16.6	0	100	20
	19	0	100	20

Thin layer chromatography was carried out using POLYGRAM[®] SIL G/UV₂₅₄ plates with 0.2 mm silica gel and fluorescent indicator UV₂₅₄ (available from Macherey-Nagel, Germany). TLC plates were analysed using an Ultra-Violet Products Inc. Chromato-Vue[®] CC20G and developed using potassium permanganate solution as appropriate.

Chromatographic purification was typically performed using pre-packed silica gel cartridges (2 g to 220 g).

Solvent removal using a stream of nitrogen was performed at 40-50 °C on a GreenHouse Blowdown system available from Radleys[™] Discovery Technologies, Saffron Walden, Essex, CB11 3AZ, United Kingdom.

¹H NMR, ¹³C{¹H} and ROESY spectra were recorded in either CDCl₃, CD₃OD, D₂O or DMSO-*d*₆ by the internal service at GSK using a Bruker AVII+ spectrometer (600 MHz, 151 MHz and 600 MHz

respectively). Chemical shifts are reported in ppm and coupling constants are reported in Hz and refer to $^3J_{H-H}$ interactions unless otherwise stated. The internal standard used was either tetramethylsilane or the residual protonated solvent at 7.26 ppm for $CDCl_3$, 3.31 ppm for CD_3OD , 4.79 ppm for D_2O or 2.50 ppm for $DMSO-d_6$. Chemical shifts (δ) are reported in parts per million (ppm) to the nearest 0.01 ppm for 1H NMR and to the nearest 0.1 ppm for ^{13}C NMR. Coupling constants (J) are quoted in Hz to the nearest 0.5 Hz and splitting patterns are recorded as singlets (s), doublets (d), triplets (t), quartets (q), combinations thereof or multiplets (m).

Infrared spectra were obtained using a Perkin Elmer Spectrum Two™ FT-IR Spectrometer fitted with a Perkin Elmer UATR diamond sampling accessory, either from solids or by allowing a thin film of compound as a solution in DCM to evaporate. Absorption maxima ($\tilde{\nu}_{max}$) are listed in wavenumbers (cm^{-1}). The intensity of the absorption maxima are characterised as either strong (s), medium (m) or weak (w).

HRMS were run on a Waters XEVO G2-XS QToF mass spectrometer using positive electrospray ionisation with a scan range of 100 to 1200 atomic mass units (AMU).

Parallel Synthesis of Photoaffinity Bit Library and Meta-Analysis by PyParse

Synthetic Procedure

Details for the synthesis of the photoaffinity bit library have been previously reported by Thomas *et al.*¹ The LC-MS data obtained in this work, which was subsequently used for the meta-analysis described here, used LC-MS System A.

R. P. Thomas, R. E. Heap, F. Zappacosta, E. K. Grant, P. Pogány, S. Besley, D. J. Fallon, M. M. Hann, D. House, N. C. O. Tomkinson, J. T. Bush. *Chem. Sci.* 2021, **12**, 12098-12106

Meta-Analysis by PyParse

For each 384-well plate, a .csv platemap was prepared using the SMILES representations for each of the amide-coupled products. The Browser Report LC-MS file (that was obtained by Thomas *et al.*) and platemap were specified as arguments to PyParse in a command line operation, where “Parea” was selected as the metric for analysis and “calc_higherions” was set to False.

e.g.

```
python PyParse.py LCMS_data.rpt platemap.csv -c 24 -r 16 -o output_directory -pt Parea -chi False
```

The PyParse analyses were run using a laptop equipped with a 11th Generation Intel® Core™ i7-11850H @ 2.50 GHz (16 CPUs) and 32 Gb RAM.

The compound summary table for each reactive fragment plate was then compared to the tabulated manual assignment performed and recorded by Thomas *et al.*¹ Individual wells that had not been analysed by Thomas *et al.* due to the incorrect mass:charge ratio being observed for the amine starting material were removed from further consideration, leaving 1026 reactions to conduct the meta-analysis.

The analysis conducted by Thomas *et al.* and the analysis from PyParse were compared on a well-by-well basis; the difference in the LC-MS UV percentage peak area reported by each was calculated, and was used to assign wells to one of three categories.

- Where this difference in peak area was less than 1 (to account for the rounding of peak areas to the nearest integer in the original analysis by Thomas *et al.*), the well was tagged as having been correctly assigned by PyParse.
- In cases where the peak area was equal to or greater than 1, the PyParse compound summary table “Overlap Detection” column was taken into account. If PyParse noted that the peak assigned to the product overlapped with another peak, the well was tagged as correct, on the basis that the program had correctly specified (as an exceptional case) to the user that this well required manual intervention.
- In cases where the peak area was equal to or greater than 1, and no comment was made in the “Overlap Detection” column from PyParse, the conclusions from each were compared. In cases where both PyParse and Thomas *et al.* had selected a UV peak where the desired mass:charge ratio was observed, the well was marked as having an Ambiguous Assignment.
- In all other cases, the well was marked as having been incorrectly assigned by PyParse.

On this basis, it was found that:

- 976 wells (95.1%) were correctly assigned by PyParse
- 37 wells (3.6%) were incorrectly assigned by PyParse
- 13 wells (1.3%) were marked as having an ambiguous assignment

It was found during the course of conducting the meta-analysis that the cause of incorrect assignment by PyParse was frequently the result of smaller shoulder peaks that were mistakenly assigned as the desired product for that well where, in fact, the more likely peak in these cases was that with the larger peak percentage area. To demonstrate this, an example has been selected from one of the wells (Figure S2).

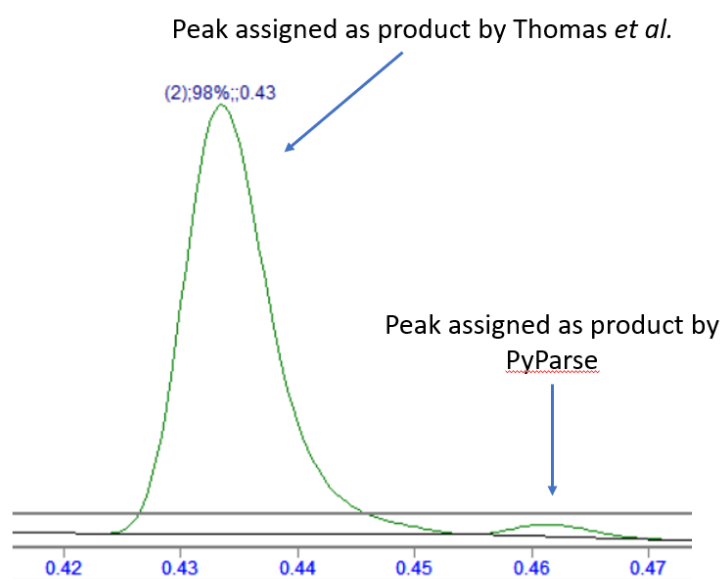


Figure S2: An example of mis-assignment by PyParse, when compared to the original assignment made by Thomas *et al.*

The time taken to conduct the analysis of each plate by PyParse was calculated by taking the summation of the times taken for each step of the analysis, which were reported by PyParse in the HTML output.

To exemplify the range of analysis difficulty for the D2B dataset, two illustrative examples have been selected: Figure S3, where a simpler profile was observed; and Figure S4, where a more complicated profile was observed.

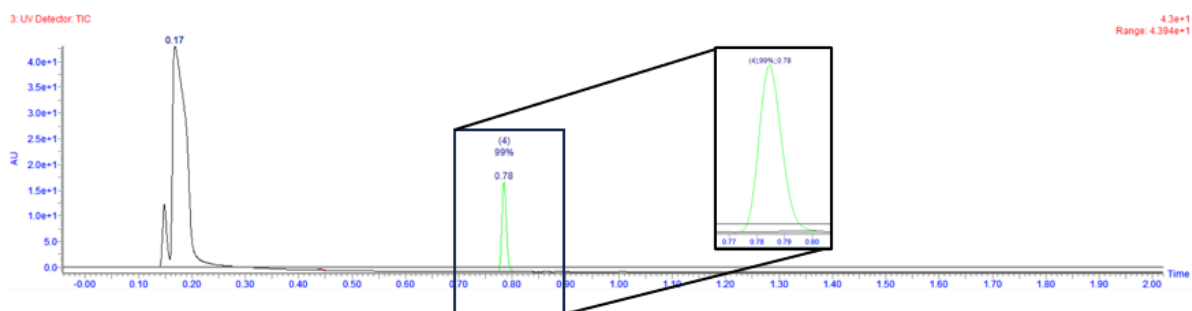


Figure S3. Illustrative example where a simpler reaction profile was observed.

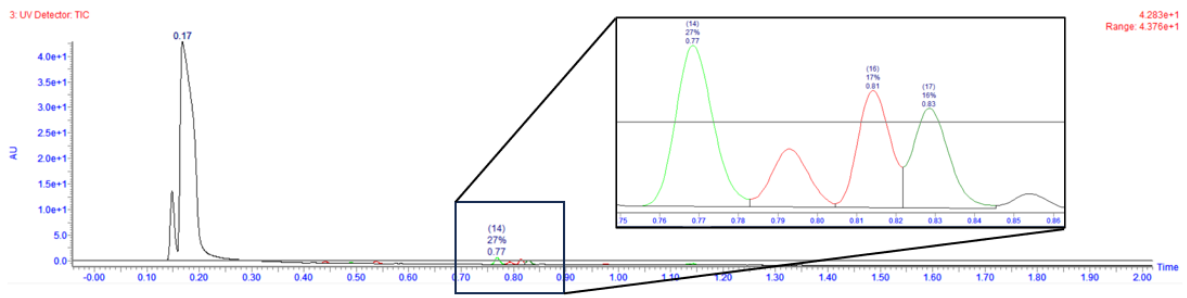


Figure S4. Illustrative example where a more complicated reaction profile was observed.

C-H Activation Reaction Optimisation Plate

Plate Design

The following palladium pre-catalysts were used to prepare the pre-dispensed plates (subsequently used in the reaction plates), for which the phosphine ligand is shown in each case and the CAS number for the precatalyst is given. The 48-well optimisation (split over 2 x 24-well metal heating blocks) was designed to include duplicate pre-catalysts to determine plate-to-plate variability.

Note: Duplicate conditions are found in wells A2/A8, A3/A9, A5/A11 and A6/A12, then again in wells C2/C8, C3/C9, C5/C11 and C6/C12. This duplication was done to ensure reproducibility between the two 24-well metal heating blocks.

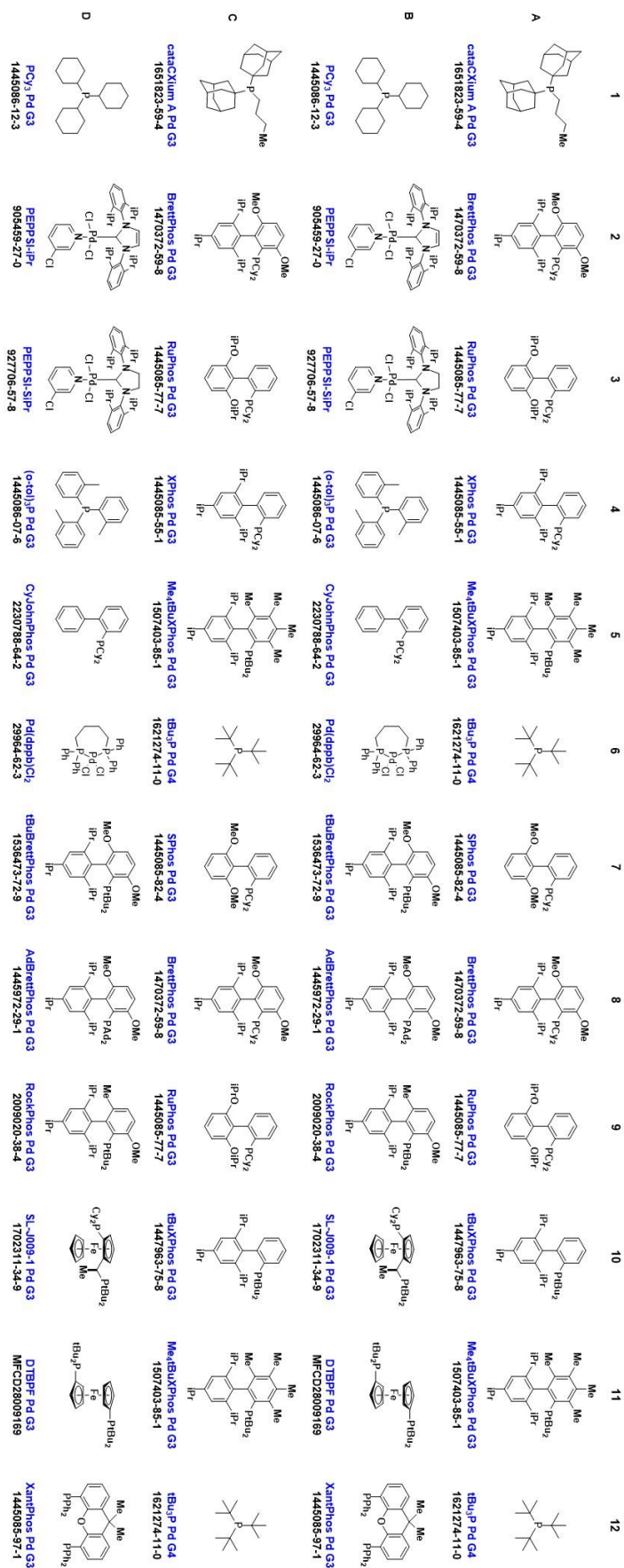
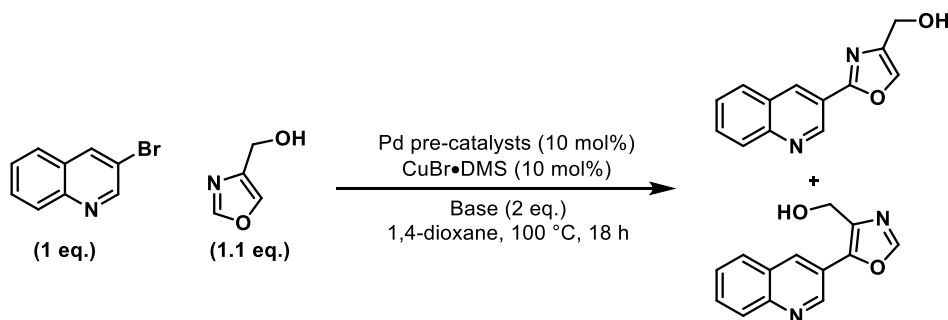


Figure S5. A schematic of the palladium pre-catalysts used in the reaction optimisation described, labelled by the row and column in which they were used in the plate.

Procedure to Prepare the Reaction Optimisation Plate



Inside a glovebox, a solution of each palladium precatalyst was prepared by dissolving 20 μmol of each in a sufficient quantity of 2-methyltetrahydrofuran to give 2000 μL of solution. 100 μL of each solution was then dosed to 8 x 24 well plates, and the solvent was removed *in vacuo* using the Genevac for 1 h. The plates were stored in the glovebox and used within 1 year of creation.

Inside a glovebox, a plate of pre-dispensed palladium precatalysts (1 μmol) was taken and tumble stirrer bars were added to each vial. Cesium carbonate (6.5 mg, 20 μmol , 2 eq.) was added by solid addition (aided by a metal transfer template) to rows A and B of the two pre-dispensed plates, and potassium pivalate (2.81 mg, 20 μmol , 2 eq.) was added to rows C and D. The following solutions were prepared and dosed to each vial:

- 3-bromoquinoline (150 mg, 720 μmol) was dissolved in 1,4-dioxane (1650 μL), and 25 μL of the resulting solution was added to each reaction vial such that each vial contained 10 μmol (1 eq.) of the starting material.
- Oxazol-4-ylmethanol (78 mg, 792 μmol) was dissolved in 1,4-dioxane (1722 μL), and 25 μL of the resulting solution was added to each reaction vial such that each vial contained 11 μmol (1.1 eq.) of the starting material.
- Copper(I) bromide dimethylsulfide complex (14.8 mg, 72 μmol) was dissolved in 1,4-dioxane (3585 μL), and 50 μL of the resulting solution was added to each reaction vial such that each vial contained 1 μmol (0.1 eq.) of the copper co-catalyst.

The metal heating blocks were then sealed, removed from the glovebox, and heated to 100 °C using a heated tumble stirrer for 18 h. The reactions were then allowed to cool to room temperature, then each reaction vial was diluted with 500 μL of a pre-prepared solution consisting of 1 μmol dibenzylaniline (labelled as an internal standard for the purposes of PyParse analysis) and 20 μmol dimethylsulfoxide in acetonitrile. The reactions were sealed and shaken to ensure a homogenous mixture; 50 μL of each reaction mixture was then aspirated and dispensed to the analysis plate. Each well of the analysis plate was diluted with 150 μL of dimethylsulfoxide, and the analysis plate was sealed with polypropylene film. An LC-MS (System B) was run using this analysis plate and analysed using PyParse. The following visualisations were generated using this LC-MS data via PyParse.

Hit Validation Graph for the Desired Product

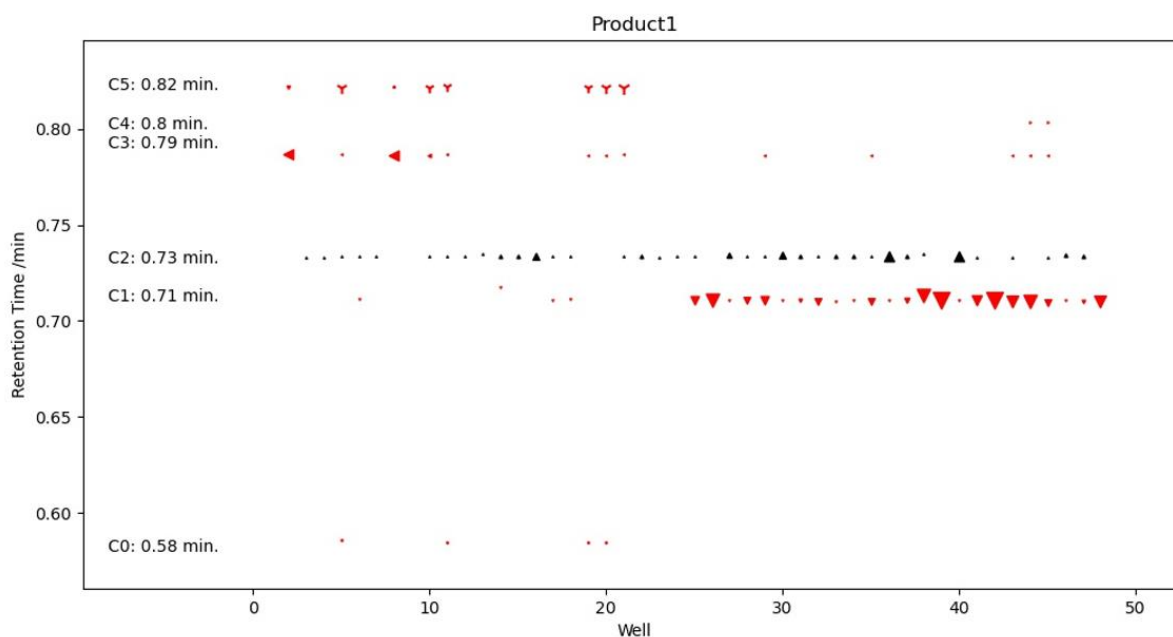


Figure S6. Hit validation graph generated by PyParse for the desired product, (2-(quinolin-3-yl)oxazol-4-yl)methanol (**4.1**), in the reaction optimisation plate.

Two distinct sets of hits were identified at 0.71 min and 0.73 mins from this hit validation graph, which suggested the synthesis of the two expected regioisomers, **4.1** and **4.2**. The platemap was adjusted to include the expected retention time of the product (0.71 min and 0.73 min), and the data were re-analysed by PyParse to generate two distinct outputs for each assumed regioisomer. In each case, the metric used for analysis was “corrP/STD”; this metric was selected as it shades each well in the heatmap visualisation according to the normalised ratio of product to internal standard.

Heatmap for Regioisomer at 0.71 min (**4.2**)

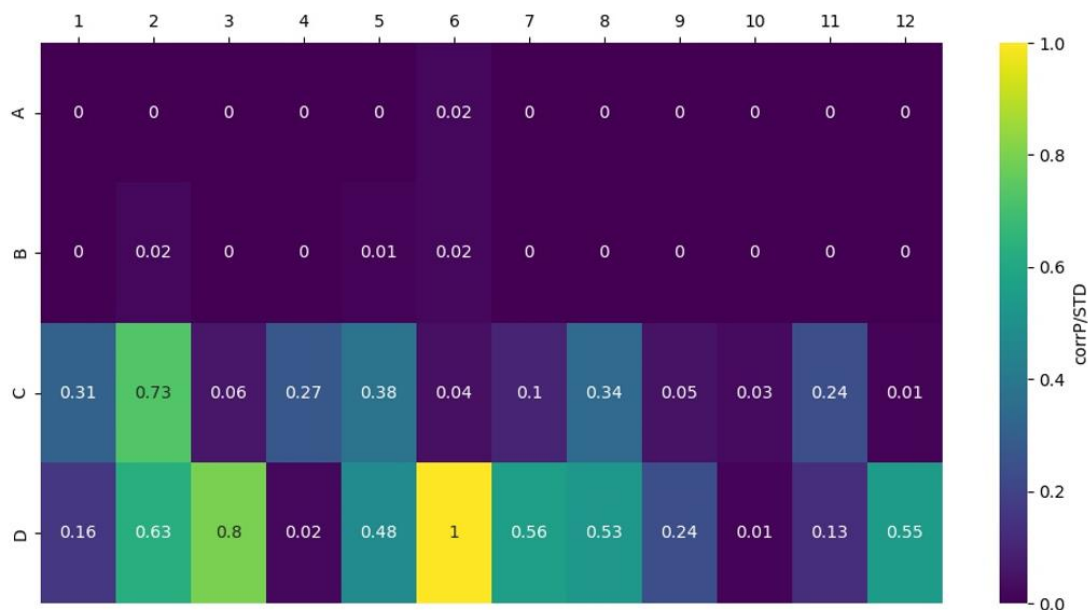


Figure S7. Heatmap generated by PyParse where a retention time of 0.71 minutes for the desired product was specified in the platemap.

From this heatmap, it was concluded that the conditions found in well D6 were most effective out of those tested for the synthesis of the regioisomer that elutes at 0.71 min.

- 3-bromoquinoline (1 eq.)
- Oxazol-4-ylmethanol (1.1 eq.)
- Pd(dppb)Cl₂ (10 mol%)
- CuBr.DMS (10 mol%)
- Potassium pivalate (2 eq.)
- 1,4-dioxane (0.1 M), 100 °C, 18 h

Heatmap for Regioisomer at 0.73 min (4.1)

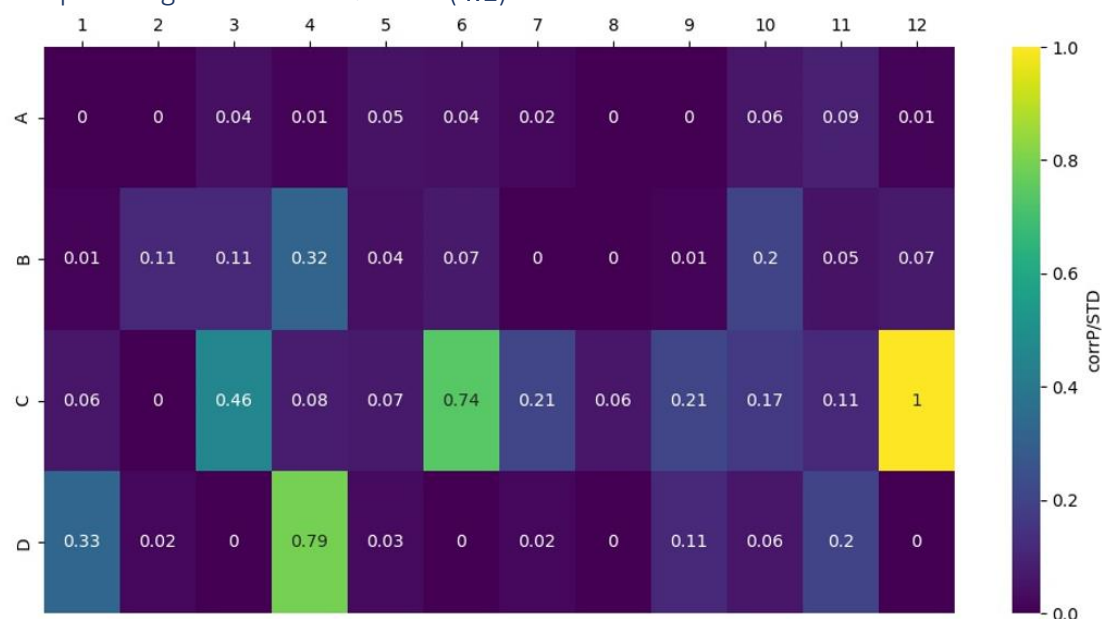


Figure S8. Heatmap generated by PyParse where a retention time of 0.73 minutes for the desired product was specified in the platemap.

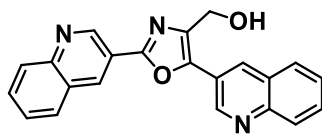
From this heatmap, it was concluded that the conditions found in well C12 were the most effective out of those tested for the synthesis of the regioisomer that elutes at 0.73 min.

- 3-bromoquinoline (1 eq.)
- Oxazol-4-ylmethanol (1.1 eq.)
- *t*-Bu₃P Pd G4 (10 mol%)
- CuBr.DMS (10 mol%)
- Potassium pivalate (2 eq.)
- 1,4-dioxane (0.1 M), 100 °C, 18 h

The best conditions for each regioisomer were repeated on a 0.96 mmol scale to facilitate isolation and characterisation; the retention times obtained for the purified samples were then used to identify which regioisomer corresponded to the LC-MS peaks observed in the reaction optimisation plate, after accounting for small differences in the retention time between crude and purified samples.

In the course of this analysis, the mass-to-charge ratio for (2,5-di(quinolin-3-yl)oxazol-4-yl)methanol **4.3** was observed for LC-MS UV peaks with a consistent retention time of 1.01 minutes. These peaks

were tentatively assigned to this by-product **4.3**, and were used to calculate the quantity (by UV percentage peak area) of this compound in each well.

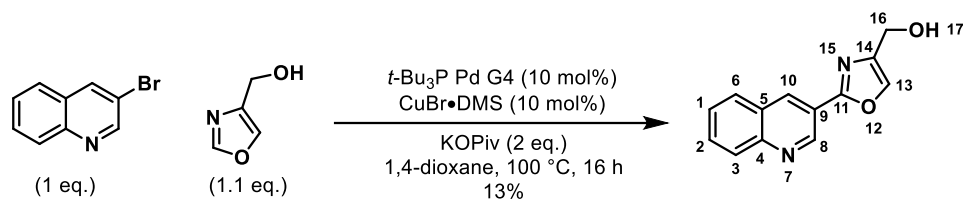


(2,5-di(quinolin-3-yl)oxazol-4-yl)methanol (**4.3**)

Exact Mass: 353.11643

tentative assignment by observed $m/z = 354.30 [M+H]^+$

Synthesis of (2-(quinolin-3-yl)oxazol-4-yl)methanol (**4.1**)



In the glovebox, a microwave vial was charged with methanesulfonato(tri-*t*-butylphosphino)(2'-methylamino-1,1'-biphenyl-2-yl)palladium(II) (56.4 mg, 0.1 eq., 96 μ mol), copper(I) bromide dimethylsulfide complex (19.8 mg, 0.1 eq., 96 μ mol), potassium trimethylacetate (269.6 mg, 2 eq., 1.90 mmol) and (1,3-oxazol-4-yl)methanol (104.8 mg, 1.1 eq., 1.06 mmol). 1,4-Dioxane (9 mL) was added followed by 3-bromoquinoline (200.0 mg, 1 eq., 0.96 mmol). The microwave vial was sealed, removed from the glovebox and the reaction was heated to 100 °C for 16 h. The reaction was quenched with water, and extracted with EtOAc. The combined organics were filtered through a hydrophobic frit and loaded onto celite; this sample was purified by normal phase chromatography (0-25% ethanol in EtOAc) using a 24 g silica cartridge. The desired fractions were combined and reduced in vacuo to give 60 mg of a mixture of the desired product **4.1** and an additional peak with an LC-MS (system A) retention time of 1.01 min. This additional peak was tentatively assigned as 2,5-di(quinolin-3-yl)oxazol-4-yl)methanol **4.3** on the basis of the observed m/z of 354.01, $[M+H]^+$.

This sample (60 mg) was further purified by dissolving it in 3.32 mL DMSO, and subjecting it to reverse phase MDAP (Method B). The desired fractions were combined and taken to pH 10 by the addition of aqueous ammonia solution, then reduced under a stream of nitrogen. This sample was redissolved in EtOAc and water; the organic layer was separated off and the aqueous layer extracted with EtOAc. The combined organics were reduced in vacuo to afford the *title compound* **4.1** (29.6 mg, 13%).

^1H NMR (600 MHz, DMSO- d_6): δ 9.43 (d, J = 2.0 Hz, 1 H, **8**), 8.93 (d, J = 2.0 Hz, 1 H, **10**), 8.18 - 8.15 (m, 2 H, **6**, **13**), 8.08 (d, J = 8.5 Hz, 1 H, **3**), 7.85 (ddd, J = 8.5, 7.0, 1.5 Hz, 1 H, **2**), 7.70 (m, 1 H, **1**), 5.32 (t, J = 5.5 Hz, 1 H, **17**), 4.50 (dd, J = 5.5, 0.5 Hz, 2 H, **16**).

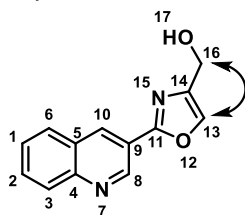
^{13}C NMR (151 MHz, DMSO- d_6): δ 158.7 (**11**), 147.8 (**4**), 147.3 (**8**), 143.3 (**14**), 136.8 (**13**), 133.1 (**10**), 131.0 (**2**), 129.0 (**6**), 128.8 (**3**), 127.6 (**1**), 126.9 (**5**), 120.2 (**9**), 55.7 (**16**).

LC-MS (System B): t_R = 0.74 min; m/z = 227.09 ($M+H$) $^+$, 225.17 ($M-H$) $^-$.

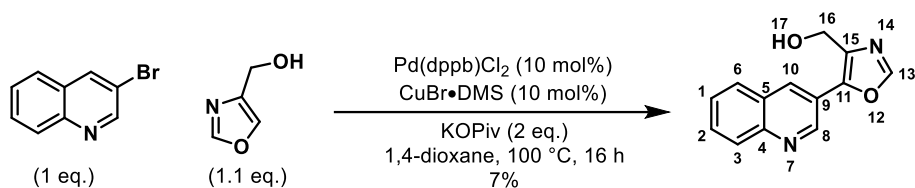
HRMS (ESI $^+$): $\text{C}_{13}\text{H}_{11}\text{N}_2\text{O}_2$ [$M+H$] $^+$ requires 227.0820 found 227.0834.

FT-IR $\tilde{\nu}_{\text{max}}$ (cm^{-1}): 3298 (broad, w, O-H), 1618 (m), 1018 (s, C-H).

Key ROESY correlations observed:



Synthesis of (5-(quinolin-3-yl)oxazol-4-yl)methanol (**4.2**)



In the glovebox, a microwave vial was charged with 1,4-bis(diphenylphosphino)butane-palladium(II)chloride (58.0 mg, 0.1 eq., 96 μ mol), copper(I) bromide dimethylsulfide complex (19.8 mg, 0.1 eq., 96 μ mol) and potassium trimethylacetate (269.6 mg, 2 eq., 1.90 mmol) and (1,3-oxazol-4-yl)methanol (104.8 mg, 1.1 eq., 1.06 mmol). 1,4-Dioxane (9 mL) was added followed by 3-bromoquinoline (200.00 mg, 1 eq., 0.96 mmol). The vial was sealed, removed from the glovebox and the reaction was heated to 100 °C for 16 h. The reaction was quenched with water, and extracted with EtOAc. The combined organics were filtered through a hydrophobic frit and loaded onto Celite™ to give the crude product. This was purified by normal phase chromatography (0-100% EtOAc in cyclohexane, then 0-20% ethanol in EtOAc) using a 24 g silica cartridge. The desired fractions were combined and reduced in vacuo to give 23 mg of a mixture of the two regioisomers **4.1** and **4.2**, eluting at 0.71 min and 0.73 min on LC-MS (system A). This sample was further purified by dissolving it in 4.14 mL DMSO, and subjecting it to reverse phase MDAP (Method A). The desired fractions were combined and taken to pH 10 by the addition of aqueous ammonia solution, then reduced under a stream of nitrogen. This sample was redissolved in EtOAc and water; the organic layer was separated off and the aqueous layer extracted with EtOAc. The combined organics were reduced in vacuo to afford the *title compound* **4.2** (15.1 mg, 7%).

¹H NMR (400 MHz, DMSO-*d*₆): δ 9.26 (d, *J* = 2.0 Hz, 1H, **8**), 8.66 (d, *J* = 2.0 Hz, 1H, **10**), 8.53 (s, 1H, **13**), 8.09 (d, *J* = 8.0 Hz, 1H, **6**), 8.06 (d, *J* = 8.5 Hz, 1H, **3**), 7.81 (ddd, *J* = 8.5, 7.0, 1.5 Hz, 1H, **2**), 7.68 (ddd, *J* = 8.0, 7.0, 1.0 Hz, 1H, **1**), 5.47 (t, *J* = 5.5 Hz, 1H, **17**), 4.66 (d, *J* = 5.5 Hz, 1H, **16**).

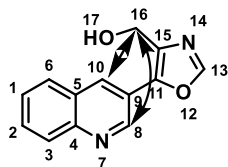
¹³C NMR (151 MHz, DMSO-*d*₆): δ 151.2 (**13**) 147.7 (**8**) 146.8 (**4**) 144.3 (**11**) 136.9 (**15**) 132.3 (**10**) 130.2 (**2**) 128.8 (**3**) 128.5 (**6**) 127.5 (**1**) 127.1 (**5**) 121.4 (**9**) 55.7 (**16**).

LC-MS (System B): *t*_R = 0.72 min; *m/z* = 227.06 (M+H)⁺, 225.30 (M-H)⁻.

HRMS (ESI+): C₁₃H₁₁N₂O₂ [(M+H)⁺] requires 227.0820 found 227.0832.

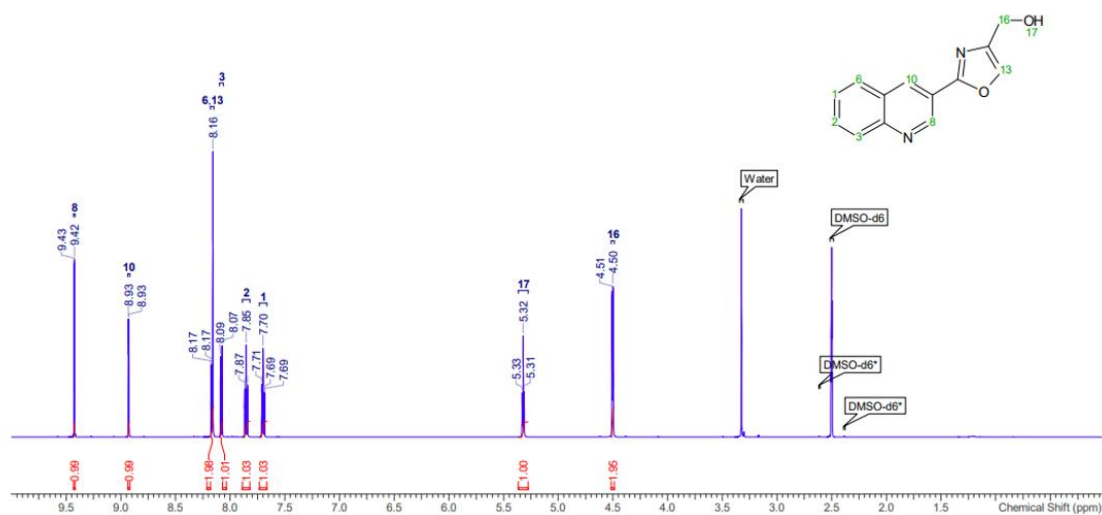
FT-IR $\tilde{\nu}_{\max}$ (cm⁻¹): 3187 (broad, m, O-H), 1510 (m), 1030 (s, C-H).

Key ROESY correlations observed:

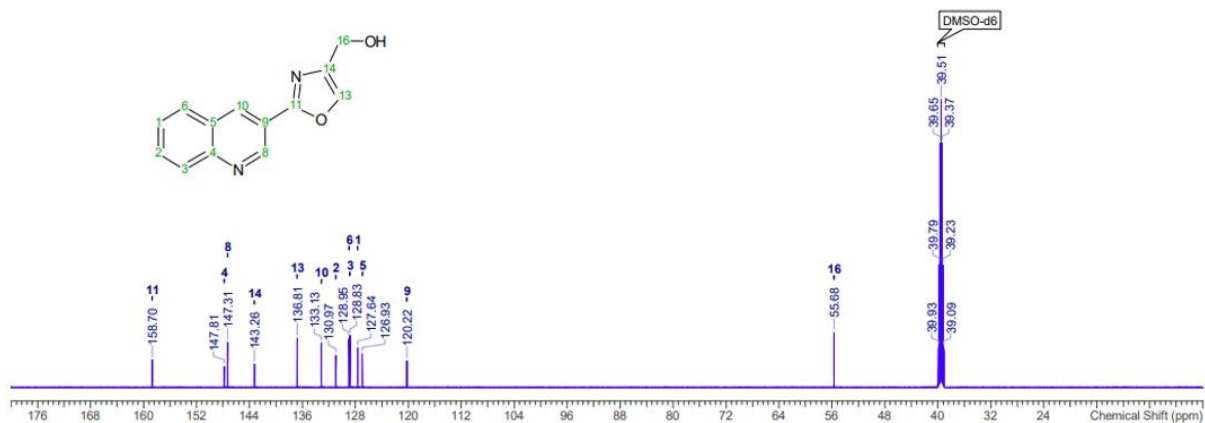


NMR Spectra

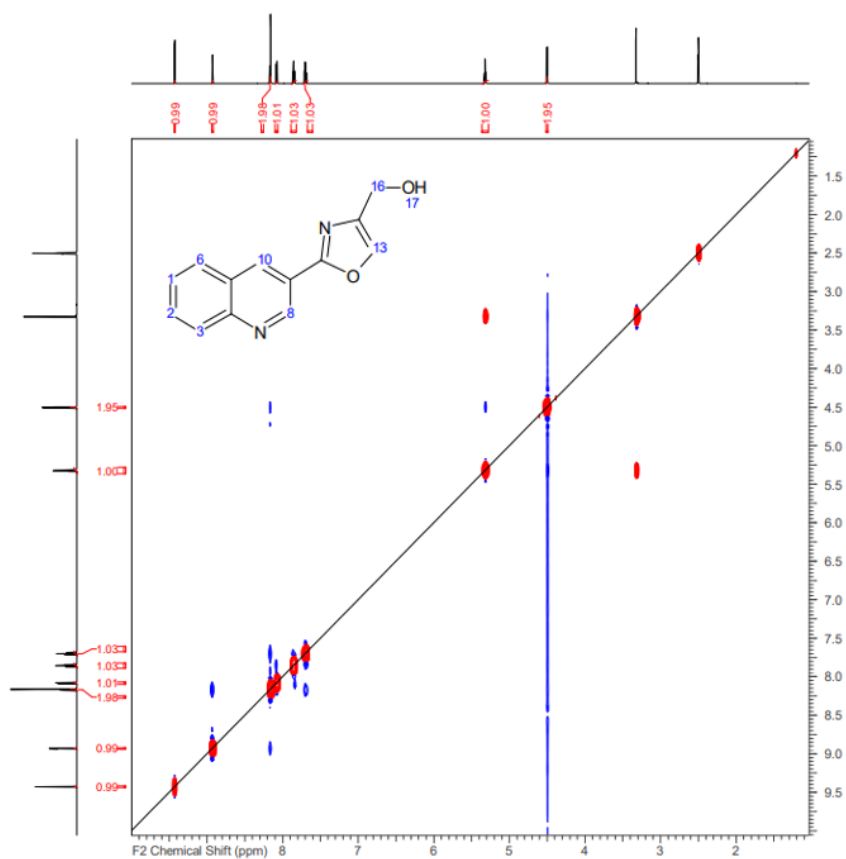
¹H NMR for (2-(quinolin-3-yl)oxazol-4-yl)methanol (4.1)



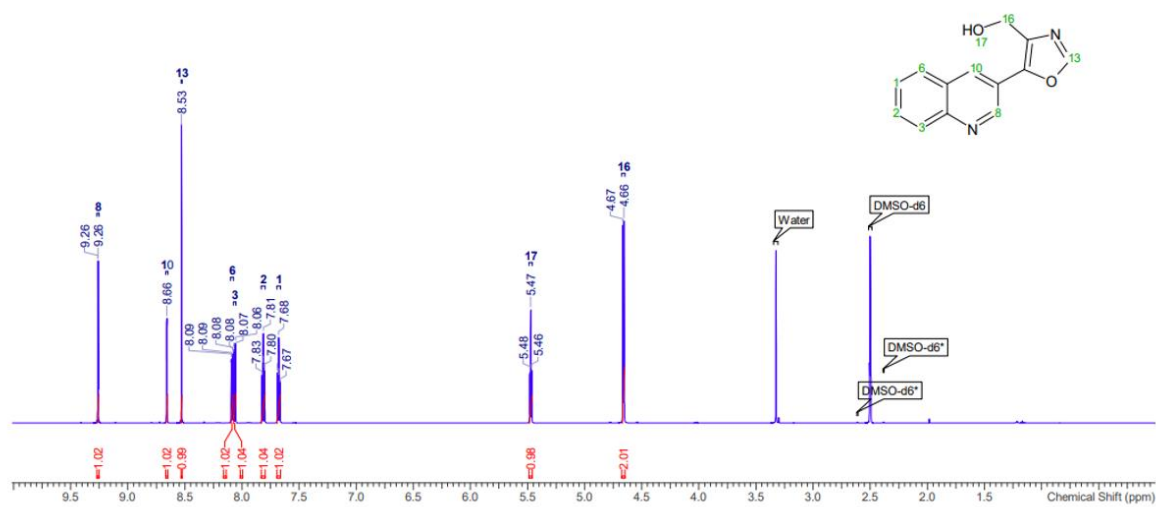
¹³C NMR for (2-(quinolin-3-yl)oxazol-4-yl)methanol (4.1)



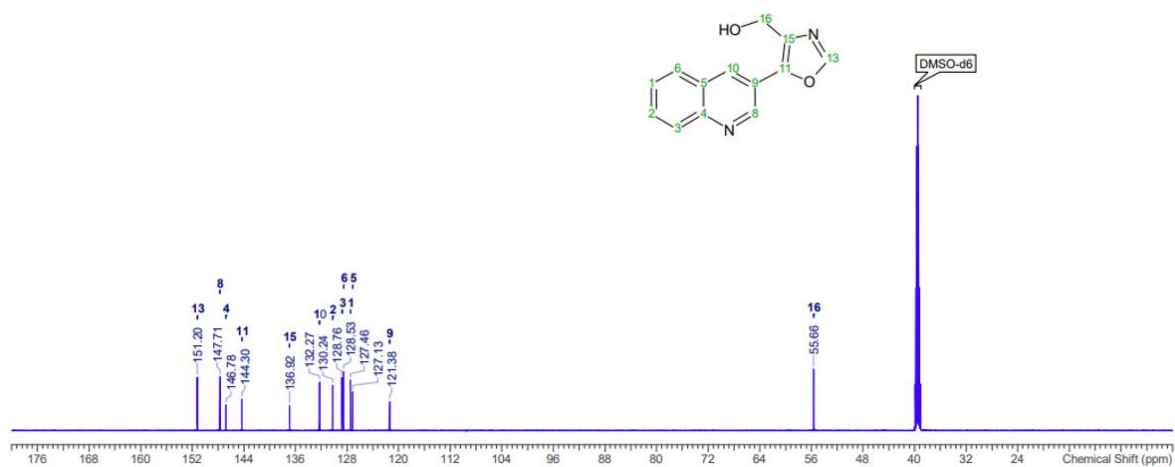
ROESY NMR for (2-(quinolin-3-yl)oxazol-4-yl)methanol (**4.1**)



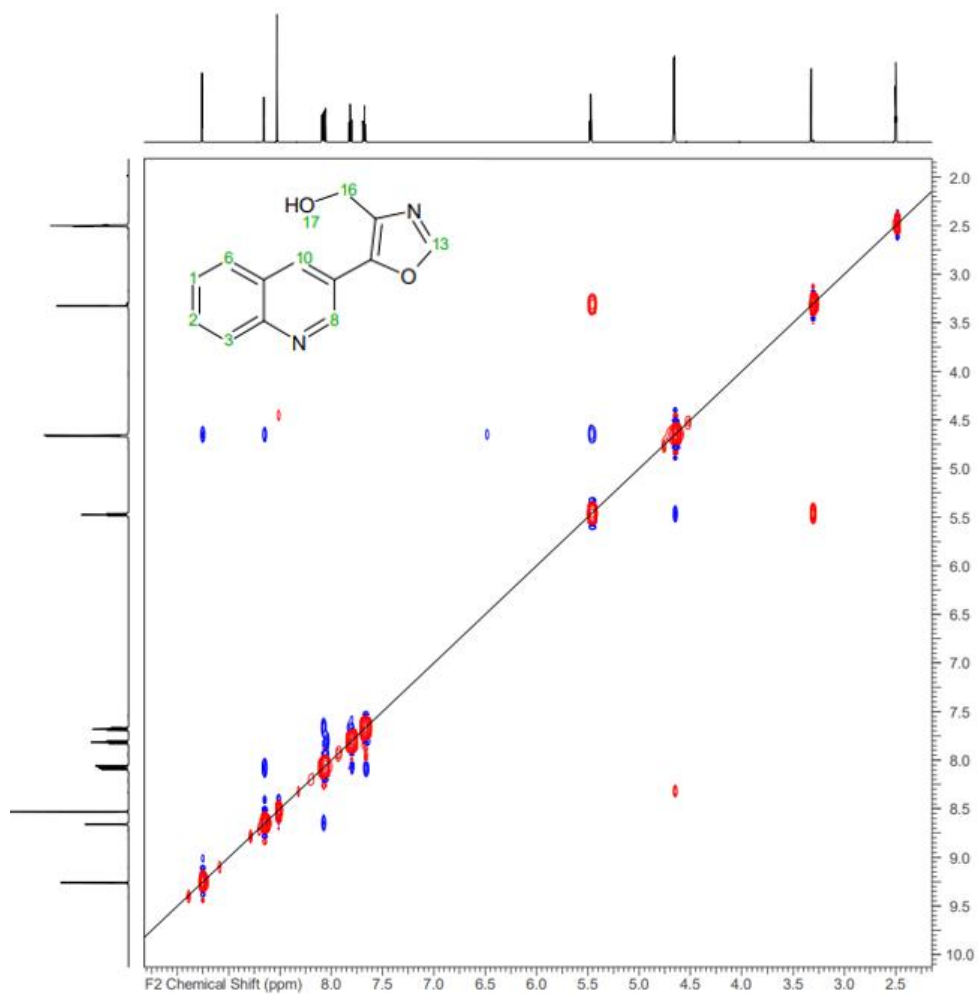
¹H NMR for (5-(quinolin-3-yl)oxazol-4-yl)methanol (4.2)



¹³C NMR for (5-(quinolin-3-yl)oxazol-4-yl)methanol (4.2)



ROESY NMR for (5-(quinolin-3-yl)oxazol-4-yl)methanol (4.2)



References

- (1) R. P. Thomas, R. E. Heap, F. Zappacosta, E. K. Grant, P. Pogány, S. Besley, D. J. Fallon, M. M. Hann, D. House, N. C. O. Tomkinson, J. T. Bush. *Chem. Sci.* 2021, **12**, 12098-12106