Electronic Supplementary Material (ESI) for Reaction Chemistry & Engineering.

Supporting Information

Exploring the chemical space of phenyl sulfide oxidation by automated optimization

Pia Mueller, Aikaterini Vriza, Adam D. Clayton, Oliver S. May, Stuart Notman, Steven V. Ley, Thomas W. Chamberlain and Richard A. Bourne

Contents

1 Experimental Details

1.1 Chemical Overview

Table 1. Used chemicals with given purities and suppliers. HEPSoxide was purified as described in Section S1.5. Water was distilled. Chemicals were sourced from Sigma-Aldrich, Alfa Aesar, Chem Cruz, Fluorochem and Fischer Science and used as provided without further purification .

Reaction	Chemical	CAS#	Abbreviation	Purity	Supplier
General	Trifluorotoluene (internal standard)	98-08-8	TFT	99%	Alfa Aesar
	1-Butanol	71-39-3	/	99%	Alfa Aesar
	Water	7732-18-5	H_2O	distilled	/
	Acetonitrile	75-05-8	MeCN	>99.9%	Sigma-Aldrich
	Phosphotungstic acid hydrate	12501-23-4	Cat	/	Acros Organics
	Hyrdogen peroxide	7722-84-1	H_2O_2	>30% w/v	Fischer Science
TA oxidation	m-ph-sulfide	100-68-5	TA	N/A	Fluorochem
	m-ph-sulfoxide	1193-82-4	TAoxide	>97%	Sigma-Aldrich
	m-ph-sulfone	3112-85-4	TAone	>98%	Alfa Aesar
HEPS oxidation	2-hydroxyethyl ph-sulfide	699-12-7	HEPS	N/A	Fluorochem
	2-hydroxyethyl-ph-sulfoxide	22063-21-4	HEPSoxide	N/A	Not available see S1.5
	2-(phenylsulfonyl) ethanol	20611-21-6	HEPSone	N/A	Fluorochem
DPS oxidation	diphenylsulfide	139-66-2	DPS	99%	Acros Organics
	Diphenylsulfoxide	945-51-7	DPSoxide	97%	Acros Organics
	diphenylsulfone	127-63-9	DPSone	N/A	Acros Organics
CEPS oxidation	2-Cl-ethyl-ph-sulfide	5535-49-9	CEPS	N/A	fluorochem
	[(2-chloroethyl)sulfinyl]benzene		CEPSoxide	N/A	ChemCruz
	2-Cl-ethyl-ph-sulfoxide	938-09-0	CEPSone	98%	Alfa Aesar

1.2 Experimental set-up

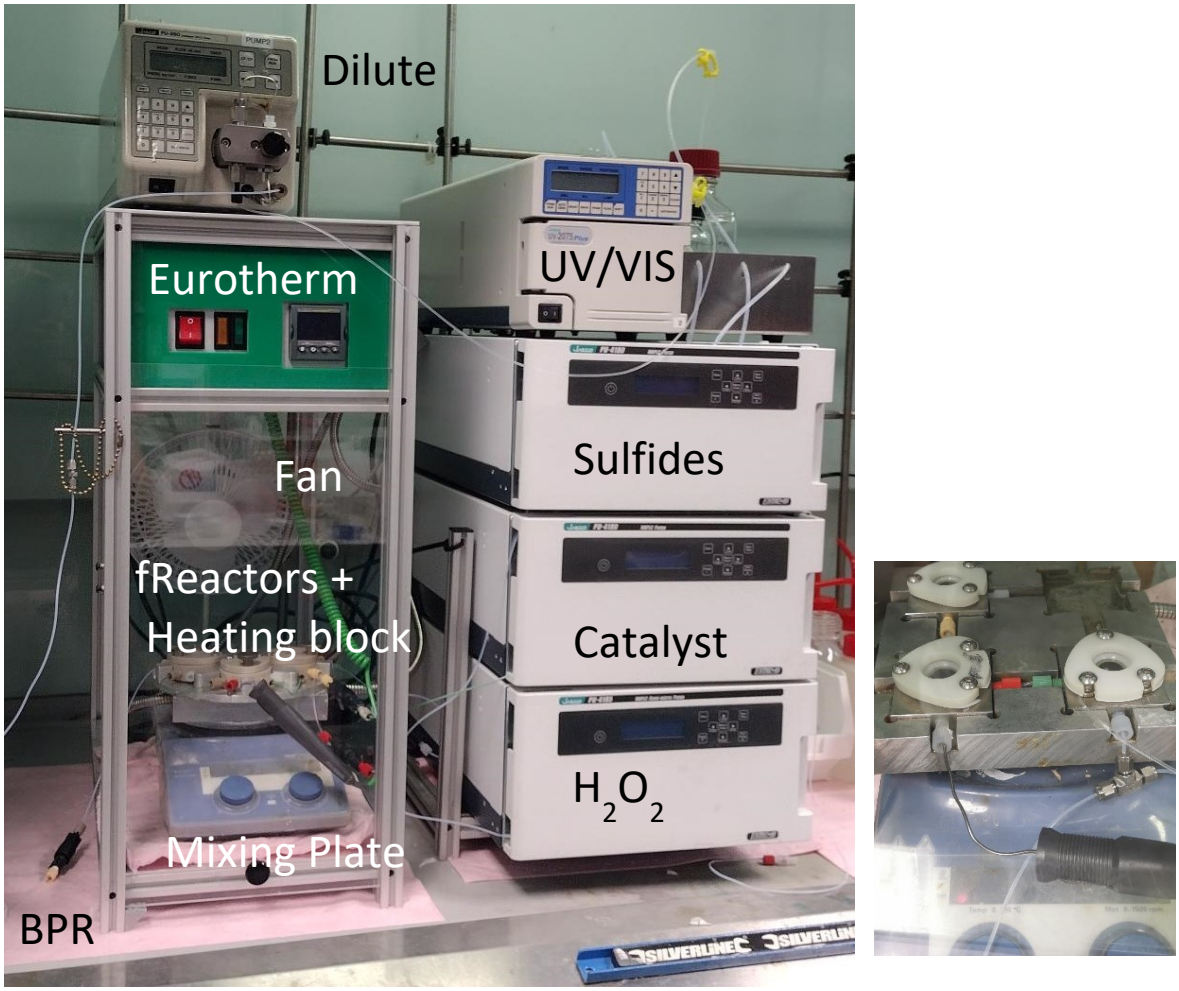


Figure S1. Photo of automated CSTR cascade reaction system. Thioanisol being methyl phenyl sulfide. For TA, DPS, HEPS metal fReactor cells of identically dimensions and connection were used for better heat transfer. For CEPS PEEK fReactor cells where chosen due to corrosion caused by hydrochloric acid formed by the elimination side reaction in water.

Experimental details:

Hydroxyethyl phenyl sulfoxide was purified from the crude reaction solution with a purity of >98% by NMR. Reagent solutions were prepared using volumetric glassware; hydrogen peroxide and the catalyst were diluted in water while the sulfides and α,α,α-trifluorotoluene (TFT) as internal standard were diluted in n-butanol. Pure samples of reagents and products including the internal standard were used to calibrate the GCMS offline. Quantitative analysis was performed on a Shimadzu GCMS-QP2010 SE instrument fitted with a SH-Rxi-5Sil MS column (30 m length, 2.5 mm ID and 2.5 µm film thickness). Methods were tailored for each screened system separation; method times varied between 6-8 minutes.

Reagents were pumped using JASCO PU4080 and PU4085 dual piston HPLC pumps. Mini-CSTRs from Asynt (fReactors) were used to mix the biphasic system.[1] For all reagents except chloroethylphenylsulfide (CEPS) metal CSTRs were used for improved heat transfer. For the chlorinated compound the system was replaced by PEEK CSTRs due to the formation of corrosive hydrochloric acid. Each CSTRs had a volume of 1.5 mL. It was found that using 3 CSTRs in a row enabled steady state to be achieved in 3 reactor volumes. The CSTRs were placed into a fitted aluminium plate and heated by a Eurotherm 980 controlled

by a thermocouple inside the CSTR and mixed by a mixing plate from IKA at 1000 rpm. Sampling was achieved by diluting both phases in acetonitrile in a 10:1 excess to achieve one homogenous stream that was sampled by a Shimadzu Flow cell, from which the AOC 6000 auto sampler extracts a volume of 0.5 µL. The reactor was maintained under the desired fixed back pressure using an Upchurch Scientific psi back pressure regulator set at 1.7 bar. Polyflon PTFE tubing (1/16" OD, 1/32" ID) was used throughout the reactor.

The system consisted out of 4 pump, 3 of which supplied the reagents and catalyst to the mini CSTR cascade. Temperature was controlled over an eurotherm measuring the temperature inside the middle of the 3 fReactor cells.

The system was connected to a flow cell of the GCMS which automatically sampled into the GC column, shown in our previous publication. [2]

Between optimisation runs the system was set to reagents saving mode till the next conditions were predicted and a temperature set point reached to minimise material consumption (SI, S2.2

Reservoir solutions were prepared by dissolving the desired reagents in solvent under stirring at ambient conditions. Reagent 1 pump: sulfides (0.4 mol L^{-1}) and trifluorotoluene (3.65 g, 25 mmol) in n-butanol in a measuring cylinder of 500 ml; Reagent 2 pump H_2O_2: hydrogen peroxide (0.6 mol/L-1) in water in a measuring cylinder of 500 ml; Catalyst pump Phosphotungstic acid hydrate (0.004 mol/L-1) in Water in a measuring cylinder of 500 ml. Solvent pump for dilution: acetonitrile. The automated reactor was set up according to the schematic shown in Figure 2 in the manuscript, where the total reactor volume = 4.8 mL and the fixed back pressure = 50 psi, the total experimental volume including the flow cell was 5 ml, used for steady state calculations only.

Bayesian details:

During the Bayesian optimisation a surrogate model is created using Gaussian process (GP) regression. An acquisition function is then optimised to determine the next point to sample. In this case, Adaptive Expected Improvement was used, which dynamically controls the exploration/exploitation trade-off. 39 The GP model is updated sequentially, and the process repeated iteratively to converge on the optimum. This leads to constrains of the optimisation to only differential equations and therefore falls in the category of supervised learning.

1.3 Residence time study

RTD for 1-3 fReactors

RTD studies were conducted for 1, 2 and 3 fReactors in series, the experimental conditions are presented in Table S2.

Table S2. Parameters employed for each RTD run using the Freactor experimental set up with the times dye flow was started.

Number of fReactor in series	Reactor Volume (mL)	Flow rate of dye (mL/min)	Time Flow Started (s)
1	1.6	4	0
2	3.2	4	0
3	4.8	4	0

The UV absorption of methylene blue was used as step tracer. Translated to concentrations and Exit-age Distribution plotted in Figure S2.

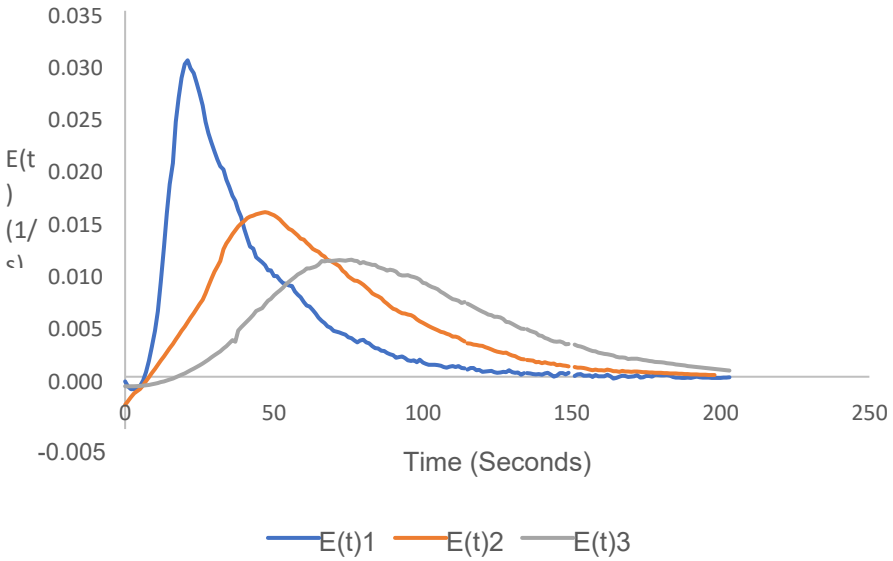


Figure S2. Exit-age Distribution curves for experiments 1,2 and 3 plotted using excel and smoothed using OriginPro.

Figure S2 shows that the graphs each has a defined peak and that the height of this peak reduces and the tail increases as the number of fReactors in series increases.

Table S3. Characterisation data based on the collected experimental data for the fReactors in series.

Number of fReactor in series	Residence Time (s)	Variance (s^2)	Dimensionless Variance	Number of ideal tanks in series
1	39.6	567	0.362	2.77
2	67.9	970	0.210	4.75
3	91.6	1450	0.173	5.77

From the data in Table S3, it can be concluded that as the number of fReactors in series increases the number of ideal tanks in series increases accordingly, however the calculated number of ideal tanks is greater than the number of actual fReactors. This difference is due to the connection tubing between the cells. Based on the 3 fReactor in series it can be calculated that steady state is over 95% achieved after 3 full Space times, which was used as wait time in the automation.

1.4 GC methods

For each of the 4 reactions a GC method was developed to separate products in the shortest amount of time to reduce experimental downtime.

TA

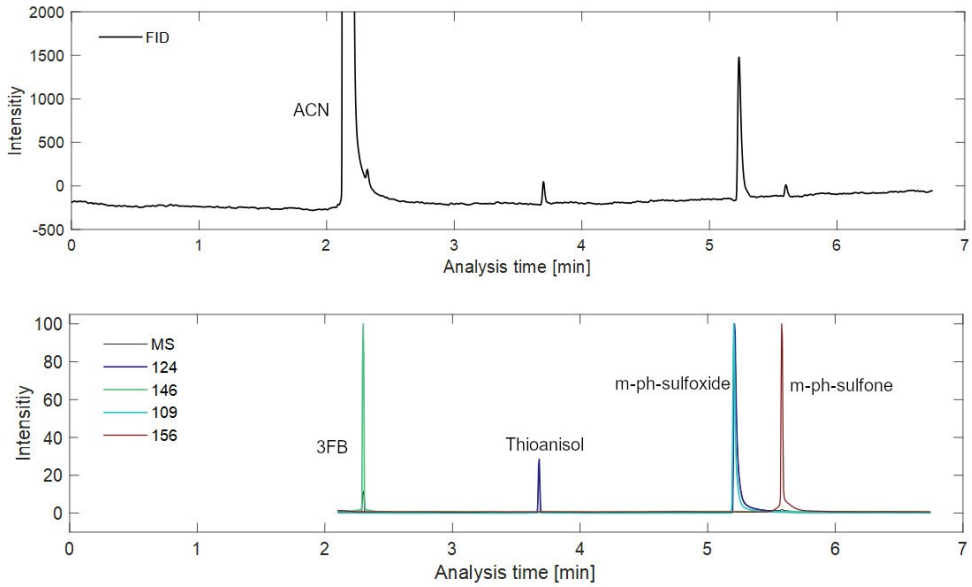


Figure S3. Exemplary GC output for m-ph-sulfide oxidation. Analytical results adopted from Shimadzu software showing the FID and MS signals in parallel for on experiment (122 °C, 0.1 ml/min TA, 0.4 ml/min H_2O_2), Thioanisol = methyl phenyl sulfide.

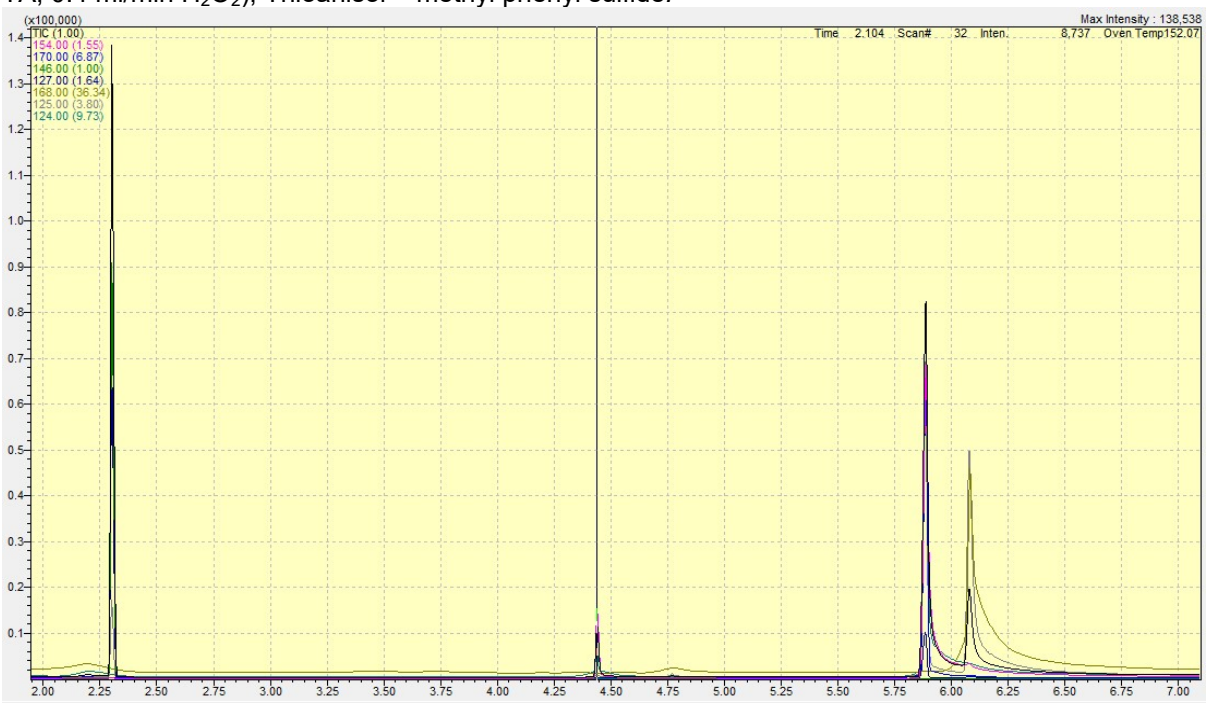


Figure S4. GCMS spectrum of HEPS oxidation. Analytical results adopted from Shimadzu software showing the FID and MS signals in parallel for on experiment (122 °C, 0.1 ml/min TA, 0.4 ml/min H_2O_2).

The MS follows each mass during the program, allowing for deconvolution of overlapping signals. Samples were calibrated for methyl phenyl sulfide and reactants with a minimum value of 99.8 obtained for the calibration factor.

Table S4. GC method descriptive parameter used for each reaction system oxidation system. Hydrogen peroxide could not be measured, but was tested before reaction to still contain 30% w/v.

TA				
Oven Ramp			ion source temperature	200 °C
Rate °C/min	Temp. °C	hold time [min]	interface temperature	250 °C
-	80	0	injection temperature	250 °C
20	215	0	oven temperature	80 °C
MS			injection volume	0.5 µl
Mode selective	140	146	split	100
124	156	109	total program time	**6.75 min**

HEPS				
Oven Ramp			ion source temperature	200 °C
Rate °C/min	Temp. °C	hold time [min]	interface temperature	250 °C
-	70	0	injection temperature	250 °C
40	150	0	oven temperature	70 °C
20	200	1		
30	250	0		
MS			injection volume	0.5 µl
Mode selective	154	146	split	100
124	125	186	total program time	
170	168			**7.17 min**

DPS				
Oven Ramp			ion source temperature	200 °C
Rate °C/min	Temp. °C	hold time [min]	interface temperature	275 °C
-	90	0	injection temperature	275 °C
40	200	0	oven temperature	90 °C
20	275	0		
MS			injection volume	0.5 µl
Mode selective	127	146	split	100
202	185	186	total program time	
218	125	168		**6.38 min**

CEPS

Oven Ramp			ion source temperature	200 °C
Rate °C/min	Temp. °C	hold time [min]	interface temperature	250 °C
-	80	0	injection temperature	250 °C
30	190	0	oven temperature	80 °C
10	205	1.3		
MS			injection volume	0.5 µl
Mode selective	91	109	split	100
172	123	135	total program time	
146	125	168		**6.47 min**
136	127	96		
170				

1.5 Purification procedure for HEPSoxide

For HEPSoxide (2-(Phenylsulfinyl)ethanol) no available commercial pure compound was found therefore HEPS was continuously oxidized in the platform without catalyst at 95 °C, 0.22 mL/min HEPS (0.4 M without internal standard), and 0.32 H_2O_2 (0.6 M) for 3 hours after reaching steady state, collecting a total of 97.2 mL reaction mixture.0.22

The crude reaction mixture was extracted with DCM (3 x 20 mL). The organic layers were combined and concentrated *in vacuo*. The residue was purified by column chromatography (hexane:EtOAc, 70:30) to yield the HEPSone (0.148 g, 6%).

H-NMR

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.04 – 7.92 (m, 2H), 7.76 – 7.67 (m, 1H), 7.63 (tt, *J* = 6.8, 1.4 Hz, 2H), 4.03 (s, 2H), 3.45 – 3.32 (m, 2H), 2.76 (s, 1H).

The spectra were in agreement with found literature[3] as well as the mass spectra in the GCMS being identical with the product peak from previous runs, proofed the purity of the compound

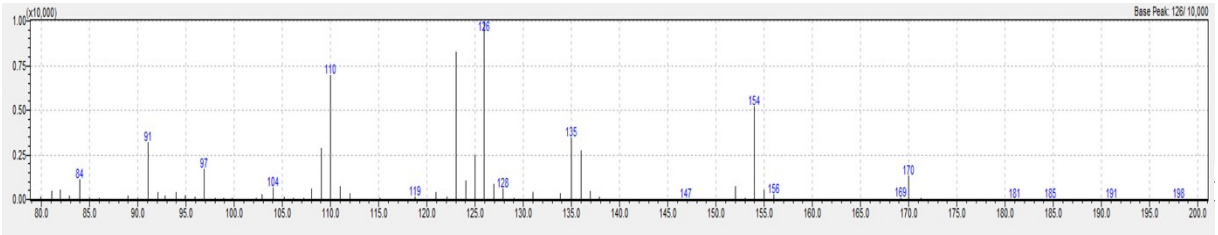


Figure S5. Mass spectrum by quadruple core MS showing the highest mass weight at 170 (total mass of HEPSoxide) and fragmented masses of 154, 135,126.

1.6 Solvent study in batch

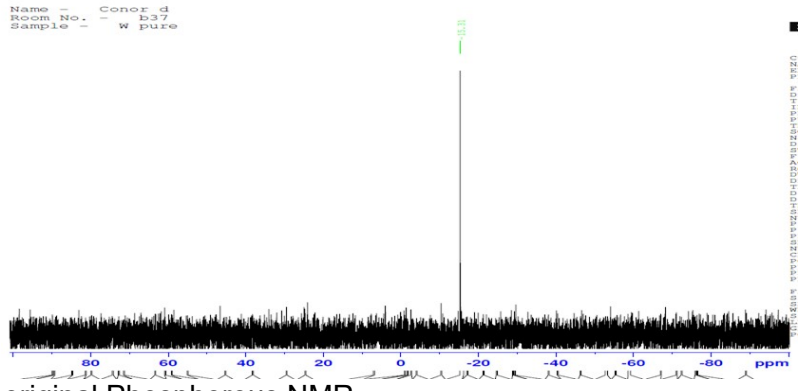
Prior to the flow synthesis the effect of solvents on the oxidation of TA with the Cat was tested. Alcohols and Acetonitrile showed good conversion and selectivity's but the Cat was only limited soluble in them. Water was itself already very active and led to quick over

oxidation. Combining aqueous and organic phase allows for high yields and solubility of catalyst.

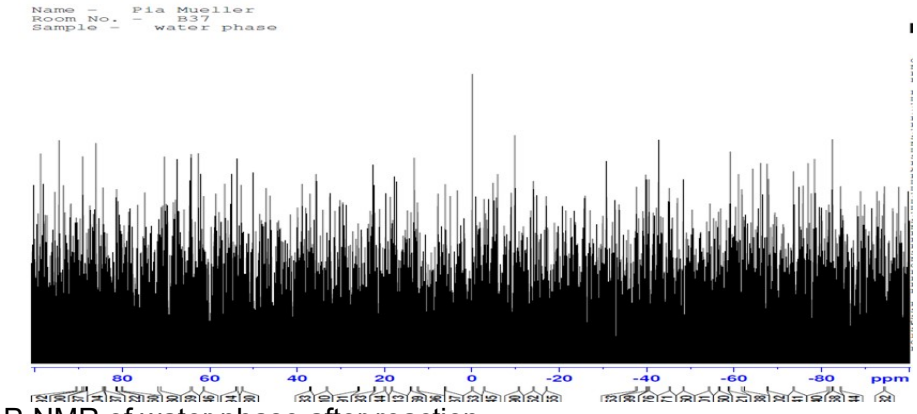
Table S5. Catalytic study of TA conversion in different solvents in batch with 1:1 ratio of 0.4 M TA and 0.6 M H_2O_2. DMSO MTF, DCM led to the formation of side products in some cases only residue of standard products could be found.

	Iso-propanol	**Aceto-nitrile**	**Aceto-nitrile, 0.9 /0.1 water**	**water**	**DMSO**	**Meta-tetrahydrofuran**	**DCM**
1.5 h no cat	Conversion 0.10	0.05	0.10	0.55	0.00	0.00	0.00
	Selectivity 0.96	0.94	0.96	0.92	0.00	0.00	0.00
+2 h with cat 0.005 % mol of TA	Conversion 0.79	0.87	0.89	1.00	0.77	0.88	0.85
	Selectivity 0.99	0.98	0.94	0.00	0.60	0.87	0.50
+10 hours	Conversion 1.00	0.99	0.97	1.00	0.82	0.93	0.99
	Selectivity 0.97	0.96	0.95	0.00	0.66	0.92	0.33

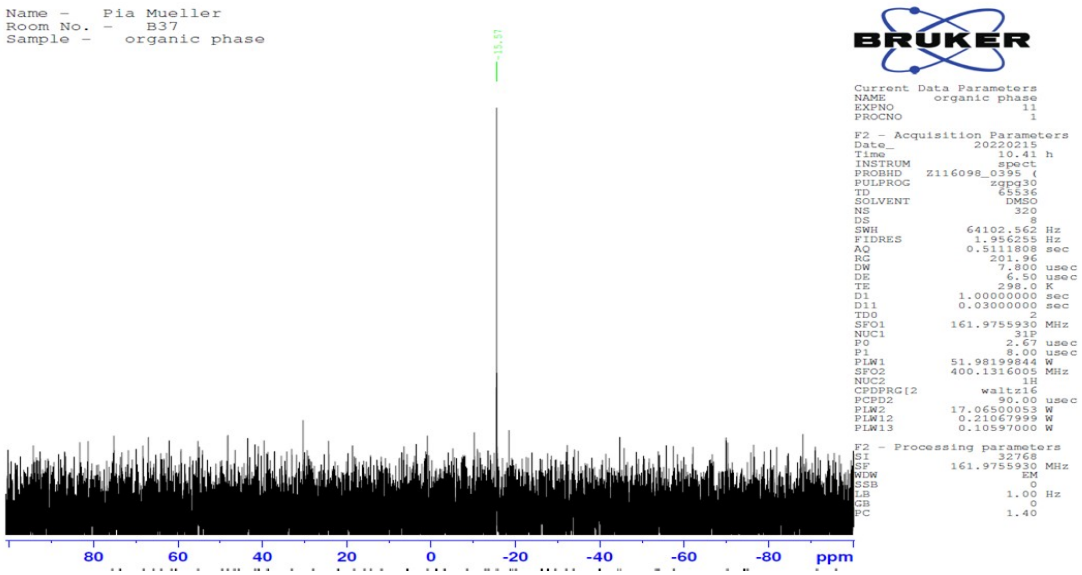
1.7 Phase study



original Phosphorous NMR



P-NMR of water phase after reaction



P-NMR of organic phase after reaction

Figure S6. P-NMR spectra of catalyst solution, and the organic and aqueous phases after reaction.

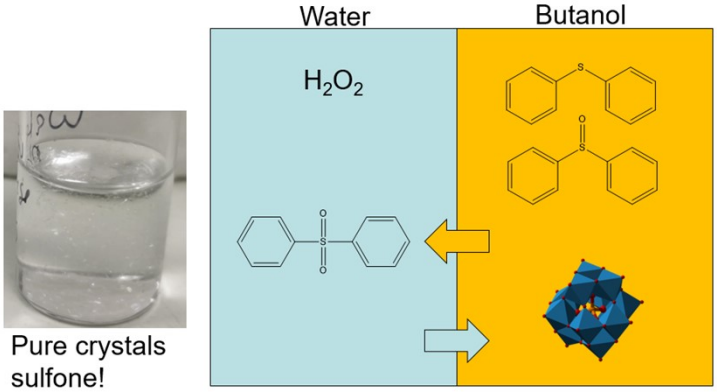
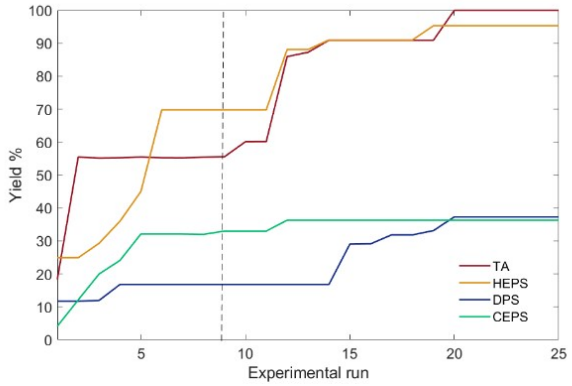


Figure S7. Schematic transfer of reagents and catalyst in DPS oxidation.

For DPS the DPSone crystalized in the aqueous phase. Due to the lower melting points this was not observed for the other tested sulfones. However the phase equilibrium between sulfones and sulfides could be further manipulated with the addition of sodium chloride. Based on the P-NMR it is assume the catalyst was preferring the organic phase, which is in agreement with Can et al.[4]. Based on this observation a recycling of the catalyst and separation of the product from catalyst would be not as trivial but could offer through the separation of the sulfone a route for purer sulfoxide.

2 Results



2.1 Optimisation Result

Figure S8. . For all four substrates the BOAEI algorithm was able to find global optima in less than 25 experiments, shown in the cumulative yield over the experimental numbers (b).

Table S6a. Optimisation experimental bounds for all experiments and sulfides. Parameters were automatically translated to

	Temperature /°C	Residence Time/min	Equiv. H_2O_2 to Sulfide	Equiv. Cat. to Sulfide
Lower bound	35	2	0.5	0.002
Upper bound	95	10	3	0.02
Step size	5	0.5	0.05	0.0002

Table S6b. Optimisation experimental overview, sorted by sulfide reagents

TA				
Temperature / °C	**Residence time / min**	**ratio H_2O_2 / TA**	**ratio Cat/TA / %**	**Yield / %**
36.55	7.52	1.40	1.8	18.3
43.54	9.69	2.32	1.5	55.3
49.34	6.37	1.92	0.5	22.3
60.44	4.78	2.85	1	43.5
64.92	3.25	1.31	1.2	25.6
72.41	8.64	0.87	1.2	51.1
78.69	4.64	0.56	1.9	18.1
87.07	2.38	1.65	0.3	24.0
92.20	7.17	2.71	0.8	51.2
74.59	9.55	2.56	0.6	60.1
94.98	10.00	1.26	1	43.4
91.67	10.00	1.59	0.4	85.9
94.76	9.99	1.34	0.5	87.2
95.00	10.00	1.89	1.4	90.9

94.93	9.33	1.76	1.3	55.1
78.42	10.00	2.19	1.2	64.2
94.81	9.99	2.99	1.1	55.5
94.99	9.85	1.68	0.5	69.7
95.00	10.00	1.10	0.3	59.8
94.98	10.00	1.67	1.2	100.0
61.54	9.40	2.12	1.3	57.3
35.54	9.22	2.63	1.2	19.2
52.45	9.20	2.93	0.4	46.7
79.48	9.07	2.43	1	100.0
79.77	10.00	2.32	0.3	100.0
77.13	9.98	2.96	1.5	56.4
75.56	10.00	1.93	1.6	78.4
41.72	10.00	1.94	0.7	38.1
40.64	10.00	1.93	0.3	37.9

HEPS				
Temperature / °C	**Residence time / min**	**ratio H_2O_2 / HEPS**	**ratio Cat/HEPS / %**	**Yield / %**
35.47	4.81	2.72	1.6	24.9
46.37	6.99	0.65	0.6	19.6
52.21	2.08	1.49	1.2	29.3
61.21	9.42	3.00	0.9	36
62.92	7.78	1.89	1.8	45
70.58	8.37	1.75	0.2	69.8
81.45	3.82	2.42	1.6	44
87.67	6.28	0.96	1.4	53.1
91.52	3.56	1.14	0.5	30.5
70.50	8.56	2.20	0.2	57.8
58.57	9.98	2.30	0.4	40.3
73.97	8.03	1.24	0.2	88.1
76.13	9.42	0.74	0.3	62.3
76.20	8.65	1.51	1	91
75.59	7.55	1.52	0.9	85.4
75.51	8.51	1.52	0.2	83.9
71.53	9.83	1.42	2	48.3
73.22	8.59	1.42	0.6	57.1
77.98	7.79	1.70	1.6	95.3
79.06	7.85	2.11	1.5	78.5
77.59	7.70	1.35	1.5	57.7
75.99	7.85	1.43	0.8	78.5
76.43	8.02	1.48	0.8	74.5
78.74	8.60	0.92	1.7	54
74.37	7.33	2.88	0.2	54.8

DPS				
Temperature	**Residence time / min**	**ratio H_2O_2 / DPS**	**ratio Cat/DPS**	**Yield**

/ °C			/ %	/ %
35.47	4.81	2.72	1.6	11.7
46.37	6.99	0.65	0.6	5.2
52.21	2.08	1.49	1.2	12.0
61.21	9.42	3.00	0.9	16.8
62.92	7.78	1.89	1.8	10.0
70.58	8.37	1.75	0.2	9.4
81.45	3.82	2.42	1.6	5.0
87.67	6.28	0.96	1.4	5.4
91.52	3.56	1.14	0.5	3.8
48.10	10.00	3.00	0.8	12.9
60.41	6.01	1.11	0.8	9.5
62.14	10.00	1.57	1.2	8.6
58.73	2.28	2.97	2.0	4.0
60.89	9.30	3.00	0.7	11.2
76.04	4.68	2.69	1.4	29.1
80.50	2.66	1.83	0.5	4.2
71.92	3.20	2.52	1.4	31.7
71.90	2.09	2.81	1.4	12.6
63.06	3.99	2.55	1.4	33.1
94.82	7.84	2.44	1.4	37.3
90.38	5.30	2.69	1.4	30.7
70.42	2.00	2.68	1.4	4.5
75.48	4.21	2.56	1.4	15.8
52.38	4.23	2.52	1.4	23.4
79.84	4.18	2.57	1.4	6.3
74.70	5.96	2.58	1.4	16.4
85.29	5.95	2.56	1.4	15.8
94.10	6.12	2.55	1.4	10.4
64.91	6.81	2.48	1.4	4.7
76.81	6.48	2.52	1.4	29.4

CEPS				
Temperature / °C	**Residence time / min**	**ratio H_2O_2 / CEPS**	**ratio Cat/CEPS / %**	**Yield / %**
36.02	6.23	2.42	1.6	4
42.76	9.83	1.28	1.3	12
53.95	3.63	2.89	0.8	20
57.80	7.34	1.72	0.4	24
63.78	3.88	2.70	1.9	32
74.17	2.10	0.76	0.9	3
79.69	8.28	0.97	0.2	31
82.57	6.92	1.42	1.7	32
91.69	5.00	2.14	1.1	33
88.79	7.60	2.29	1.2	28
48.61	9.07	0.70	1.3	6

83.37	7.89	2.88	1.9	36
83.77	2.61	0.76	2	5
58.10	3.66	1.25	1.9	4
95.00	7.36	1.49	0.4	22
69.67	6.64	1.24	0.3	6
87.06	8.43	1.59	0.2	30
80.40	2.79	0.80	1.3	5
35.42	9.00	3.00	0.9	4
59.50	2.27	0.91	0.7	3
85.97	7.82	1.60	0.7	26
83.15	6.92	1.28	0.9	27
84.09	6.88	1.70	2	34
81.03	6.84	2.10	1.7	25
83.09	6.56	0.50	1.3	34

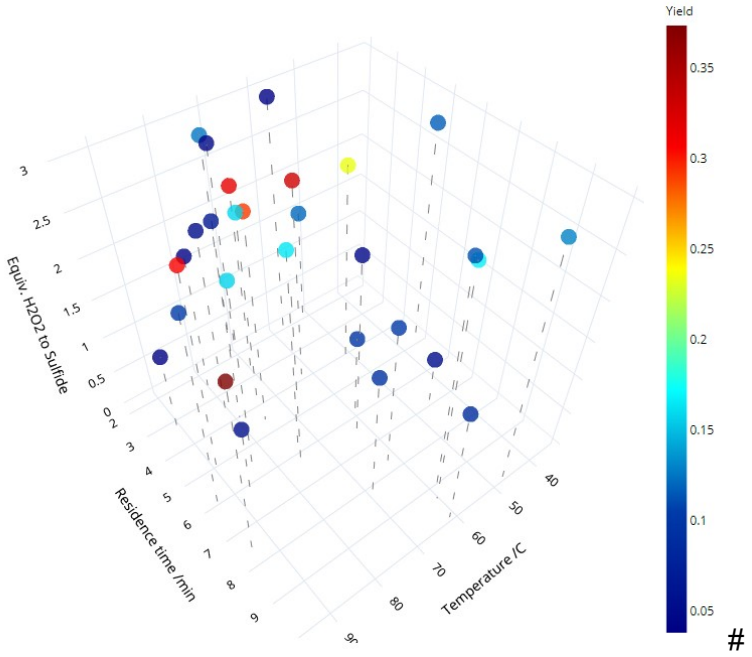


Figure S9. DPS optimisation, interactive graph under: https://chart-studio.plotly.com/~pmueller2209/20/#/

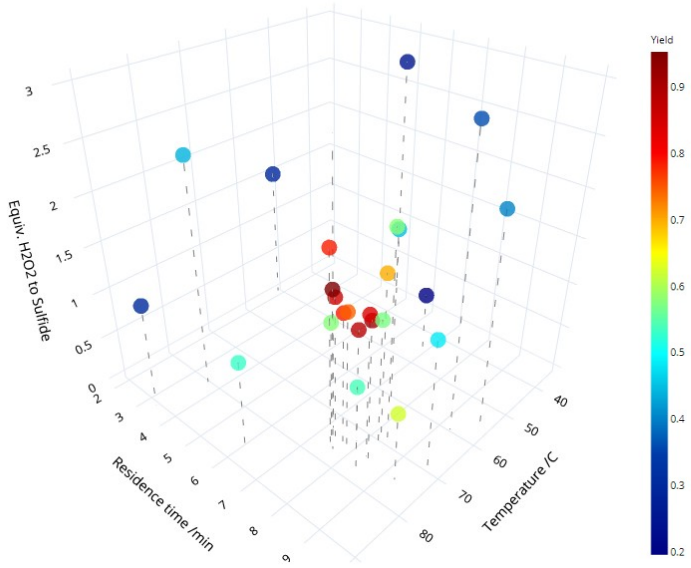


Figure S10. HEPS optimisation, interactive graph under: https://chart-studio.plotly.com/~pmueller2209/22/#/

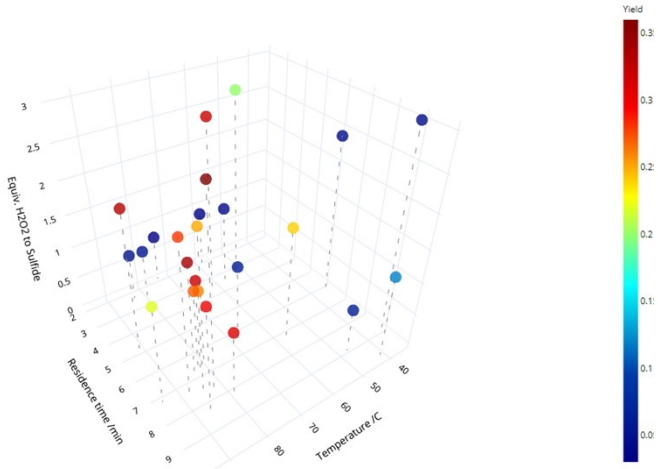


Figure S11. CEPS optimisation, interactive graph under:https://chart-studio.plotly.com/~pmueller2209/25/#/

2.2 Meta data of experiments

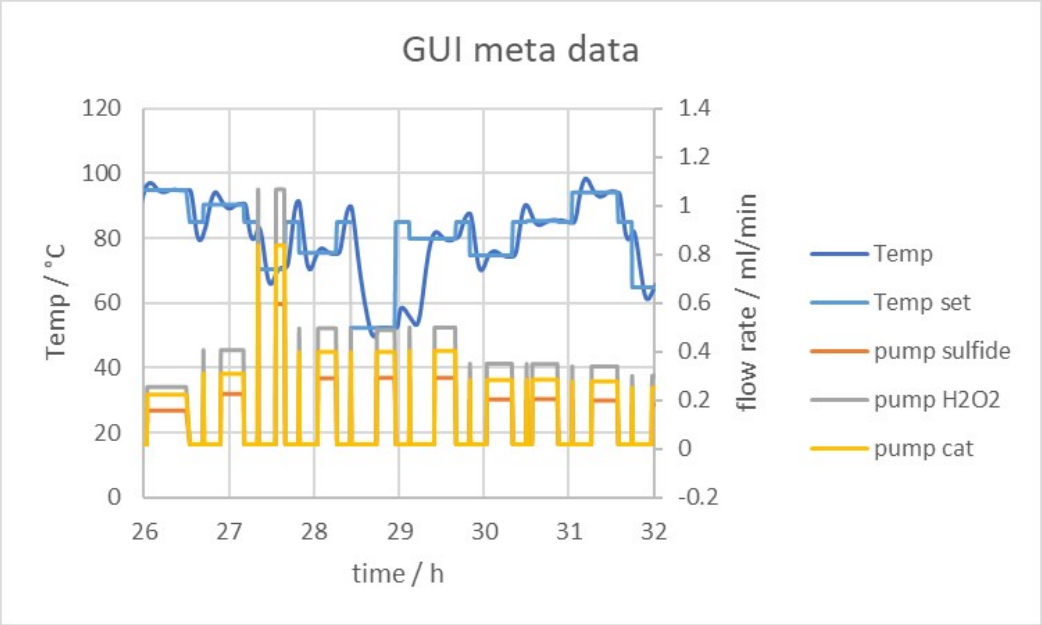


Figure S12. Fraction of continuous experiments. 10 successive experiments of the DPS optimisation are shown. Temperature was set in between runs to an average value to decrease heating time between runs.

2.3 Hyperparameters, GP Surrogate Models and Simulation

Hyperparameters

The hyperparameters can be extracted from the surrogate models built during the optimisation to reveal important process information. For the hyperparameters of the input variables (Θ_i) a lower value indicates a greater contribution to the output. The σ_n^2 hyperparameter corresponds to the noise of the system, which is medium low for the objectives in both algorithm studies. This indicates high quality and consistent data.

Table S7. Overview Hyperparameters for the 4 optimisations.

Variable	GP (TA)	GCP (HEPS)	GP (DPS)	GP (CEPS)
$\Theta_{temperature}$	40.50	11.02	7.20	15.36
$\Theta_{residence\ time}$	4.06	3.22	0.46	1.31
$\Theta_{H2O2\ ratio.}$	1.56	1.32	0.66	1.99
$\Theta_{Cat\ ratio.}$	18.27	0.05	0.01	0.03
σ_n^2	0.79	0.73	0.41	1.05

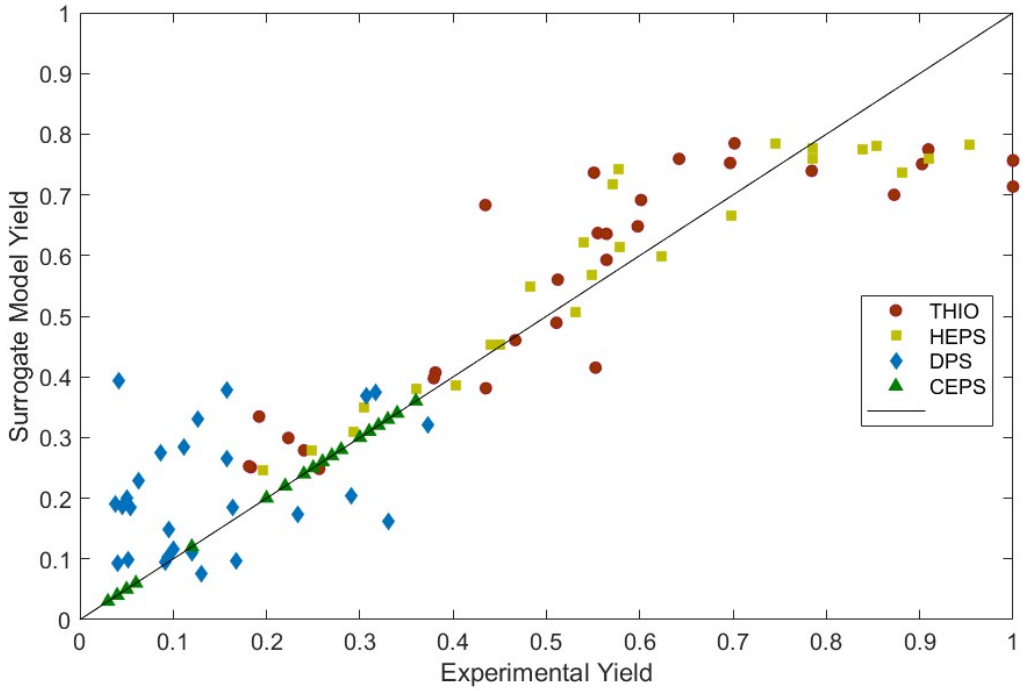


Figure 13. Pareto front between experimental yield and predicted yield by fully trained Bayesian algorithms. (THIO = Thioanisol)

3 Chemical Space Screening using machine learning

General observations: TA and HEPS seem to have similar properties in terms of polarity and solubility which might give an explanation why they perform similarly in terms of the oxidation reaction. Although HEPS and CEPS are structurally very similar, their molecular properties are quite different and thus they perform differently in the oxidation reaction.

3.1 Screening the molecular space of commercial phenyl sulfides

ZINC20 database was screened for finding all the in-stock molecules which are readily available for purchase. The extracted dataset was further filtered to find all the phenyl sulfides with relatively low complexity, *i.e.*, length of SMILES stings less than 20. This process gave a final dataset of 84 structurally diverse molecules.

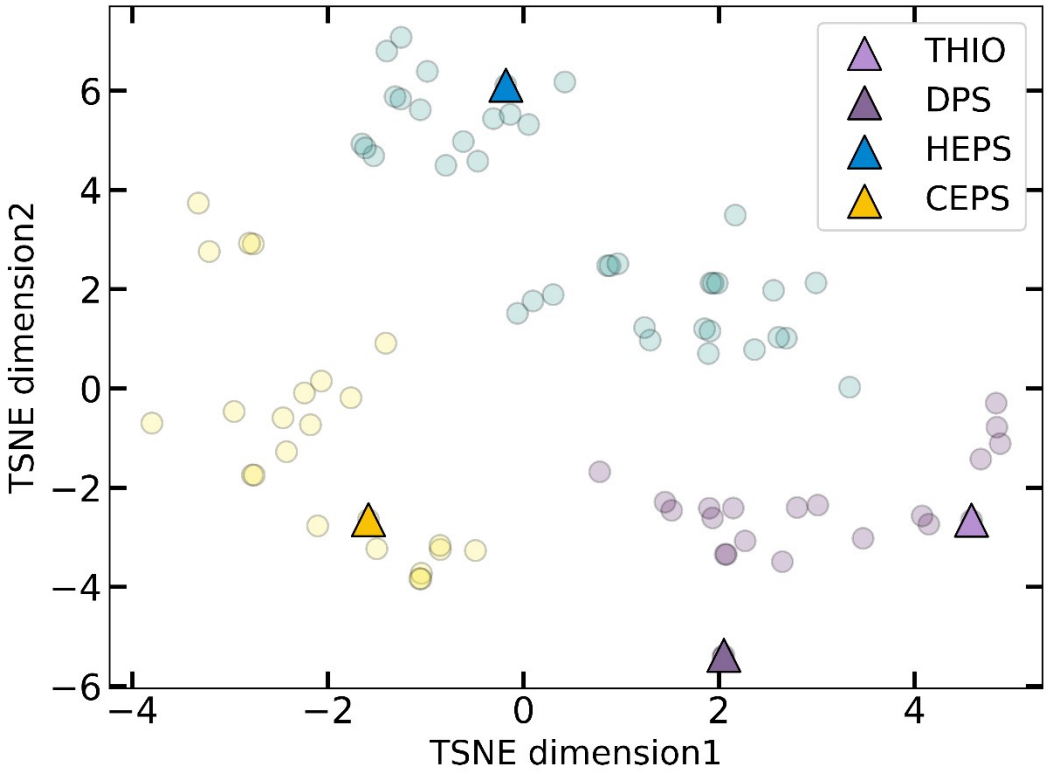


Figure 14. Chemical space of the purchasable phenyl sulfides extracted from ZINC20 database. [5] Each molecule is represented as a vector of molecular descriptors extracted from SwissADME library[6]. The clustering was performed with a Gaussian Mixture model from sklearn and the projection to the 2D space using TSNE algorithm. It can be seen that the four selected molecules cover substantially the chemical space of interest. (THIO = Thioanisol)

3.2 Tanimoto similarity between the four selected phenyl sulfides

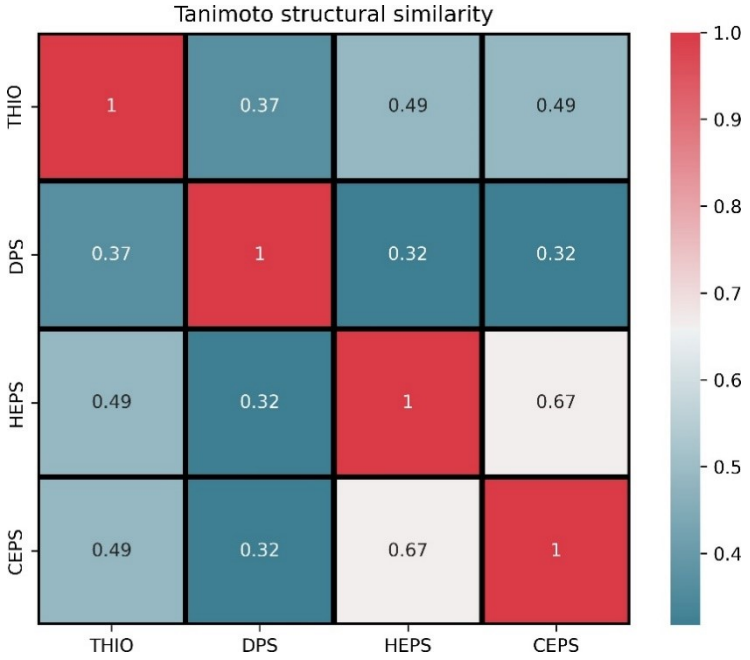


Figure 15. Correlation matrix showing the structural similarity (Tanimoto similarity) between the 4 selected sulfides. (THIO = Thioanisol)

3.3 Molecular properties comparison of the selected molecules

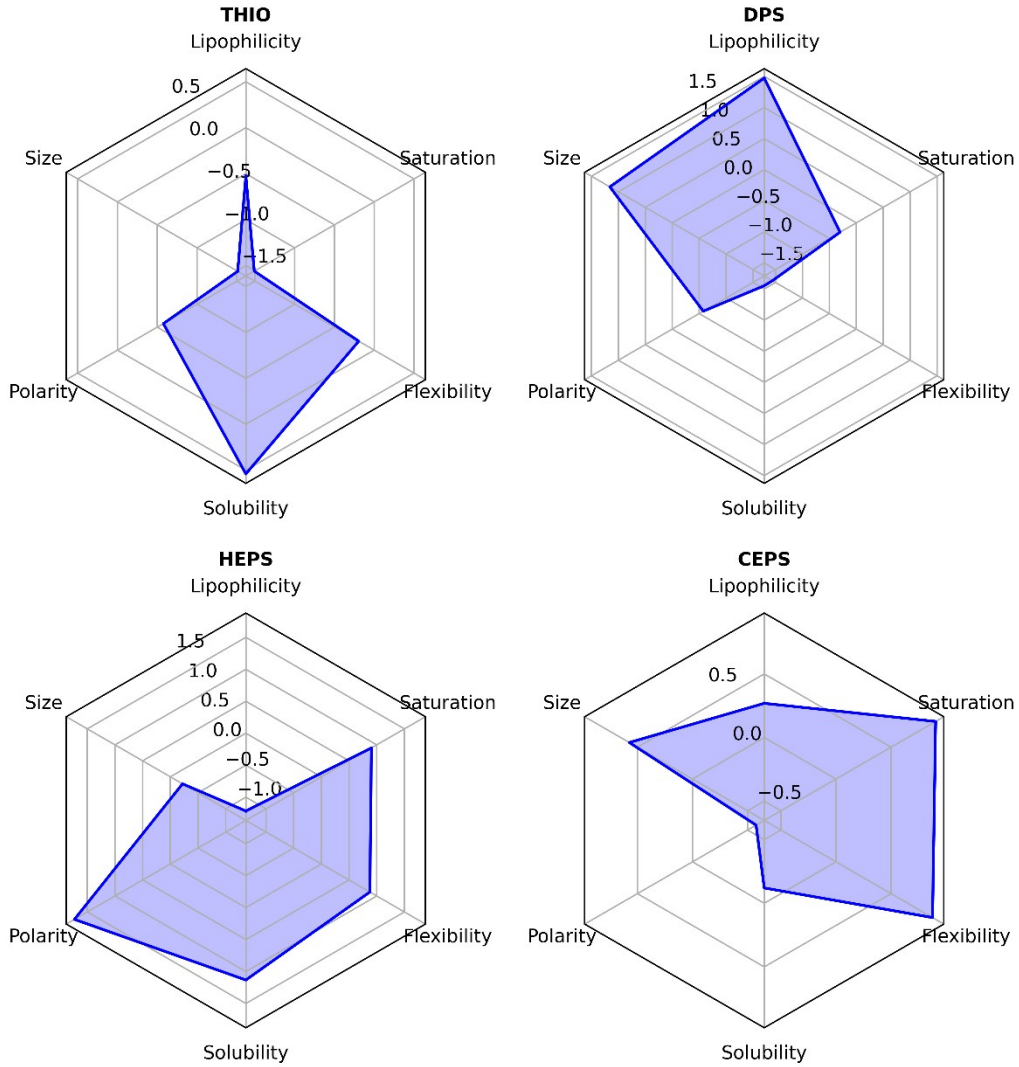


Figure 16. Spider plots showing the distribution of some important molecular features for each of the four sulfides. The molecular descriptors were generated using the SwissADME library.

4 References

1 Asynt | Laboratory Equipment Manufacturer | Supplier, https://www.asynt.com/, (accessed 10 June 2022).

2 P. Müller, A. D. Clayton, J. Manson, S. Riley, O. S. May, N. Govan, S. Notman, S. V. Ley, T. W. Chamberlain and R. A. Bourne, *React. Chem. Eng.*, 2022, **7**, 987–993.

3 A. Ghorbani-Choghamarani and P. Zamani, *J. Iran. Chem. Soc.*, 2011, **8**, 142–148.

4 F. Can, X. Courtois and D. Duprez, *Catalysts*, 2021, **11**, 703.

5 J. J. Irwin, K. G. Tang, J. Young, C. Dandarchuluun, B. R. Wong, M. Khurelbaatar, Y. S. Moroz, J. Mayfield and R. A. Sayle, *J. Chem. Inf. Model.*, 2020, **60**, 6065–6073.

6 A. Daina, O. Michielin and V. Zoete, *Sci. Rep.*, 2017, **7**, 42717.