
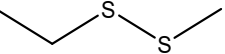
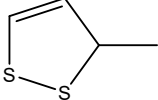
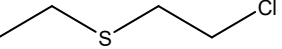
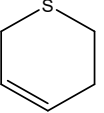
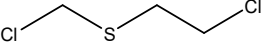
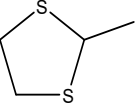
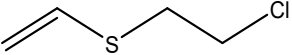
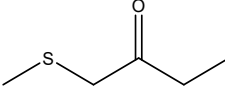

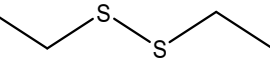
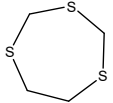
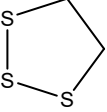
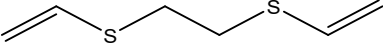
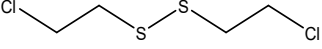
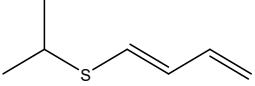
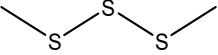
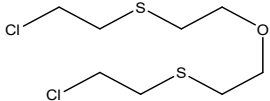
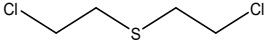
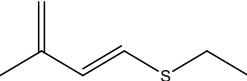
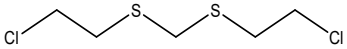
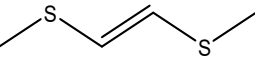
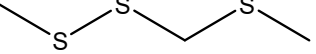
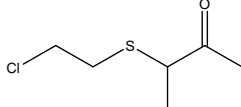
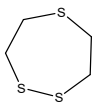


No.	Structure	CAS number	No.	Structure	CAS number	No.	Structure	CAS number
1		15980-15-1	10		20333-39-5	19		118023-96-4
2		693-07-2	11		40697-99-2	20		2625-76-5
3		5616-51-3	12		81142-02-1	21		13678-58-5
4		505-29-3	13		110-81-6	22		81451-19-6
5		6669-39-2	14		4413-12-1	23		1002-41-1
6		88017-34-9	15		3658-80-8	24		63918-89-8
7		505-60-2	16		49563-09-9	25		63869-13-6
8		19698-38-5	17		42474-44-2	26		117239-10-8

**Table S1** All CAS ultimately screened as valuable for attribution

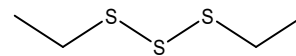
9



6576-93-8 18



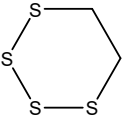
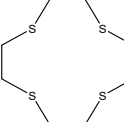
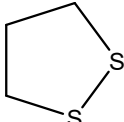
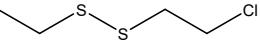
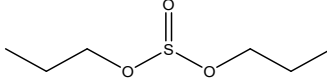
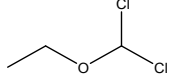
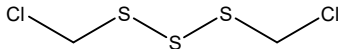
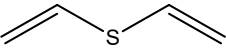
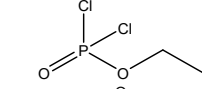
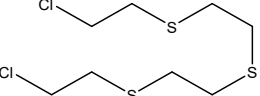
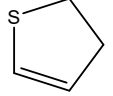
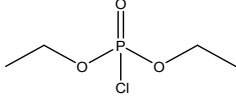
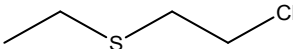
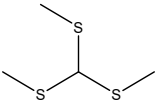
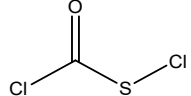
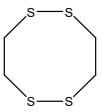
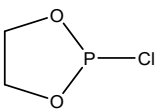
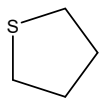
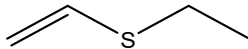
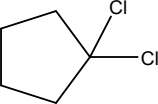
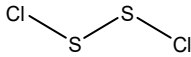
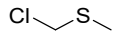
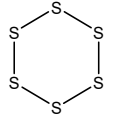
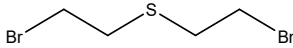
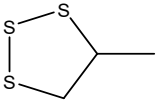
420-12-2 27



3600-24-6

---

**Table S1 (Continued)** All CAS ultimately screened as valuable for attribution

No.	Structure	CAS number	No.	Structure	CAS number	No.	Structure	CAS number
28		290-81-3	37		25423-56-7	46		557-22-2
29		7579-29-7	38		623-98-3	47		1498-42-6
30		19149-77-0	39		627-51-0	48		1498-51-7
31		51472-73-2	40		1120-59-8	49		814-49-3
32		693-07-2	41		5418-86-0	50		2757-23-5
33		1940-01-8	42		822-39-9	51		110-01-0
34		627-50-9	43		31038-06-9	52		10025-67-9
35		2373-51-5	44		13798-23-7			
36		7617-64-3	45		116664-29-0			

**Table S2- S9** Delineate the predictive results of the various classification models for the test set. Each model assesses the matching degree between a test sample and individual synthetic routes and assigns a corresponding similarity score; the route with the highest score is defined as the attributed route for the sample, corresponding to the highest inferred similarity. Green denotes correct classifications and red denotes incorrect ones.

**Table S2** M1 Predictions

Test sample \ Predict	Predict		
	SOCl <sub>2</sub>	PCl <sub>3</sub>	No intermediates
A	-0.009817	0.979193	0.0306233
B	1.01414	0.019638	-0.0337799
C	-0.28476	1.31222	-0.027463
D	1.02844	0.094258	-0.122698
E	-0.053148	0.92644	0.126708
F	0.616178	0.378016	0.00580586
G	0.218256	0.78025	0.00149384
H	1.43198	-0.38365	-0.0483359
I	-3.07257	4.53973	-0.467153
J	1.07471	-0.15085	0.0761442
K	-0.105073	1.12263	-0.0175594
L	1.04472	-0.05508	0.0103634
M	0.0846811	0.904707	0.010612
N	0.927266	0.015132	0.0576023
O	0.277807	0.753088	-0.0308955
P	0.548637	0.478067	-0.0267036
Q	0.397719	0.614909	-0.0126279
R	0.521532	0.447458	0.0310101
S	0.409603	0.585707	0.00469029
T	2.18828	-1.32636	0.138079
U	-0.35923	0.125645	1.23359
V	0.0804285	-0.07169	0.991259

**Table S3** M2 Predictions

Test sample \ Predict	Predict	
	Stepwise	One-pot
B	1.24089	-0.24089
D	-0.38822	1.38822
J	1.16539	-0.16539
L	0.184094	0.815906
F	0.69928	0.30072
H	0.157212	0.842788
N	1.23488	-0.23488
P	0.106581	0.893419
R	0.10186	0.89814
T	-0.62002	1.62002

**Table S4** M3 Predictions

Test sample \ Predict	Predict	
	Stepwise	One-pot
A	1.06421	-0.06421
C	-0.11245	1.11245
I	1.51952	-0.51952
K	-0.04071	1.04071
E	0.970514	0.029486
G	0.306599	0.693401
M	0.80604	0.19396
O	0.014203	0.985797
Q	0.009398	0.990602
S	0.186363	0.813637

**Table S5** M4 Predictions

Test sample \ Predict	Predict					
	(D)	(H)	(L)	(P)	(R)	(T)
D	1.7081	-0.04313	-0.06723	-0.16224	-0.30561	-0.12988
H	-0.34454	2.31296	-0.25035	-0.07142	-0.40252	-0.24413
L	-0.00075	0.008024	0.969172	-0.01261	0.018483	0.017685
P	-0.07978	-0.06221	-0.06127	1.39283	-0.10349	-0.08608
R	0.198374	0.049757	0.065522	0.207498	0.400812	0.078036
T	-0.33251	-0.16316	-0.16153	-0.28815	-1.24281	3.18815

**Table S6** Extended Data Table 6 M5 Predictions

Test sample \ Predict	Predict			
	(B)	(F)	(N)	(J)
B	1.17001	-0.08998	-0.03589	-0.04414
F	0.024678	0.855713	0.045842	0.073767
N	-0.03109	-0.04452	0.963371	0.112236
J	0.00447	-0.17726	-0.09397	1.26676

**Table S7** M6 Predictions

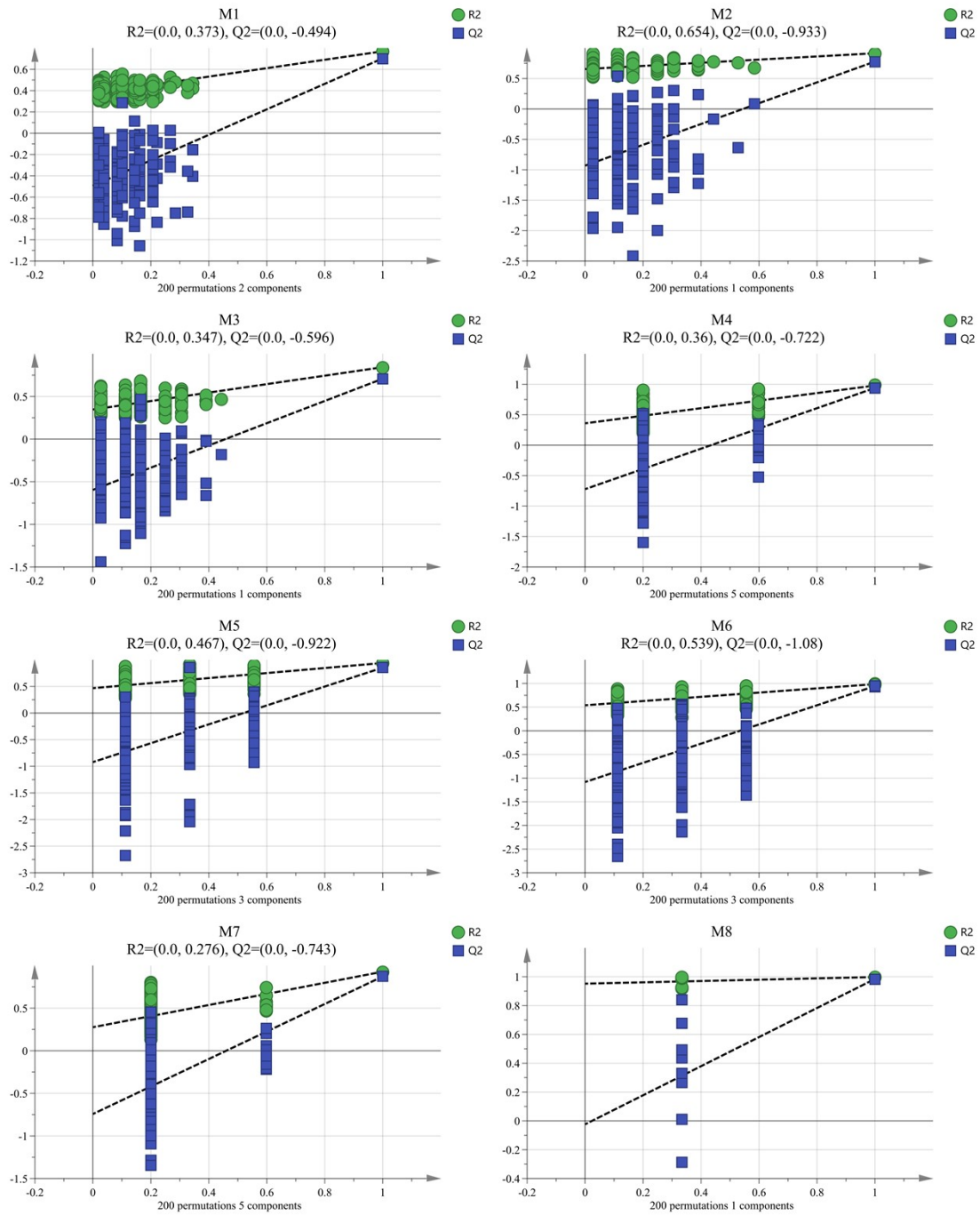
Test sample \ Predict	Predict			
	(A)	(E)	(I)	(M)
A	1.06587	-0.03536	-0.03372	0.003209
E	0.069486	1.21211	-0.35404	0.072448
I	-0.85238	0.626813	1.27868	-0.05311
M	-0.0064	0.046581	0.237111	0.72271

**Table S8** M7 Predictions

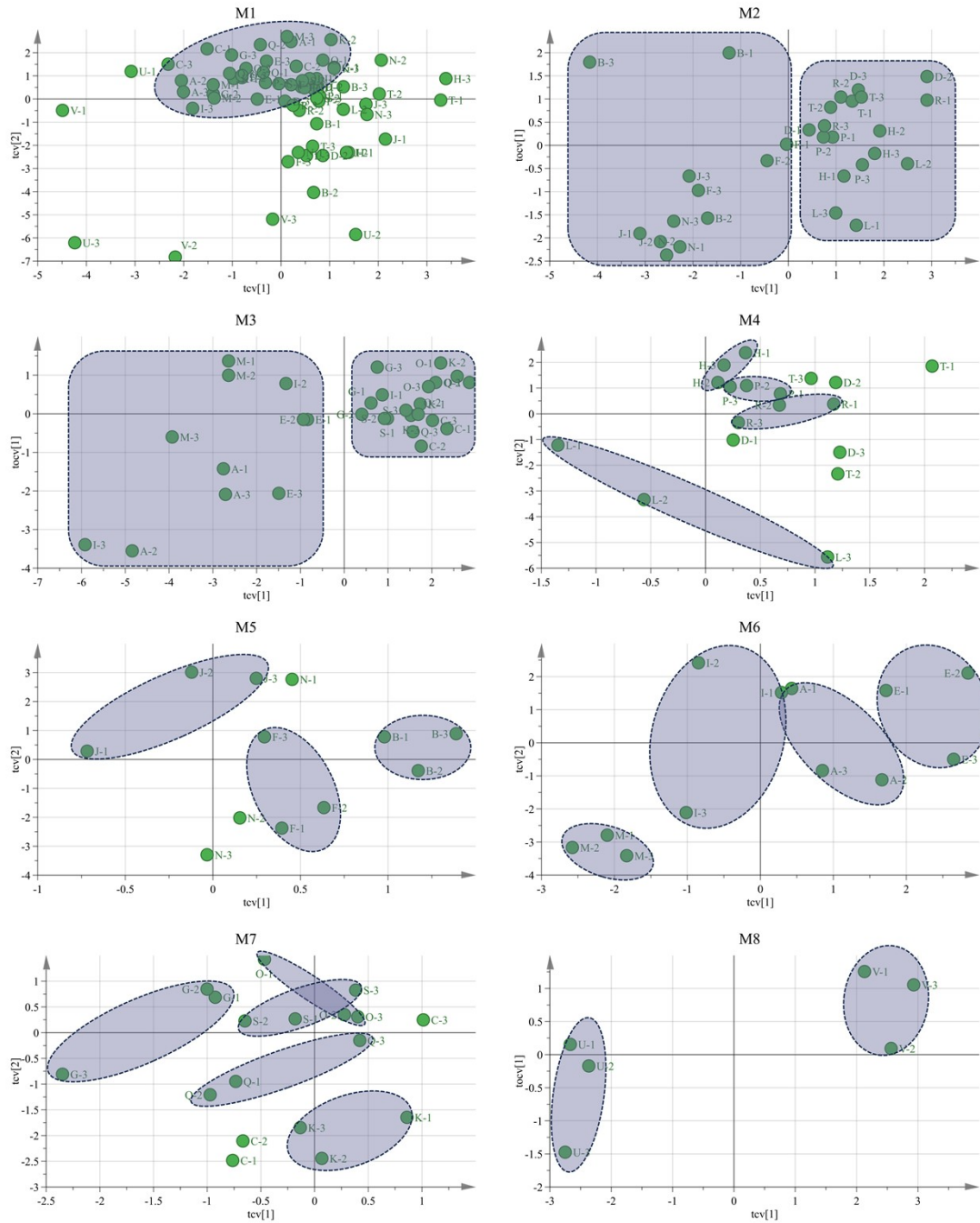
<b>Predict</b> <b>Test sample</b>	<b>(C)</b>	<b>(G)</b>	<b>(K)</b>	<b>(O)</b>	<b>(Q)</b>	<b>(S)</b>
<b>C</b>	1.53906	0.101548	-0.10691	-0.09792	0.008813	-0.4446
<b>G</b>	-0.08806	1.05638	-0.04565	-0.1793	0.010976	0.245654
<b>K</b>	0.087572	-0.09427	1.04185	-0.07444	0.012629	0.026658
<b>O</b>	0.012739	-0.00132	0.049822	0.8692	-0.00474	0.074299
<b>Q</b>	-0.19639	-0.15338	-0.1195	-0.12805	0.57802	1.0193
<b>S</b>	0.122357	0.145059	0.0884	0.042453	0.081466	0.520265

**Table S9** M8 Predictions

<b>Predict</b> <b>Test sample</b>	<b>(U)</b>	<b>(V)</b>
<b>U</b>	1.22045	-0.22045
<b>V</b>	-0.78072	1.78072



**Fig. S1** Permutation test validation (n=200) for the OPLS-DA hierarchical classification framework. The plot shows the regression of  $R^2$  (green) and  $Q^2$  (blue) against the permutation ratio.



**Fig. S2** Cross validated plots