SUPPORTING INFORMATION

Enzyme Classification Using Complex Dynamic Hemithioacetal Systems

Yan Zhang,[a] H. Surangi N. Jayawardena,[b] Mingdi Yan,*[a],[b] and Olof Ramström*[a]

[a] KTH - Royal Institute of Technology, Department of Chemistry, Teknikringen 30, S-10044 Stockholm, Sweden
[b] Department of Chemistry, University of Massachusetts Lowell, 1 University Ave., Lowell, MA 01854, United States

---

Contents

---

1. General

Reagents were obtained from commercial suppliers and used as received. Lipase preparation PSIm was purchased from Amano Enzyme Inc. The other lipases were purchased from Sigma-Aldrich. ^{1}H-NMR and ^{13}C-NMR data were recorded on a Bruker Avance 400 (100) MHz and/or a Bruker Avance 500 (125) MHz, respectively. Chemical shifts are reported as δ values (ppm) with $CDCl_3$ (^{1}H-NMR δ 7.26, ^{13}C-NMR δ 77.16) as an internal standard. J values are given in Hertz (Hz). High resolution mass spectroscopy was performed by the Instrument Station of National Center for Genetic Engineering and Biotechnology (BIOTEC), Thailand. Thin layer chromatography (TLC) was performed on precoated Polygram® SIL G/UV 254 silica plates (0.20 mm, Macherey-Nagel), visualized with UV-detection. Flash column chromatography was performed on silica gel 60, 0.040-0.063 mm (SDS) or neutral aluminum oxide.

Lipase abbreviations: *Aspergillus niger* lipase (ANL), *Candida antarctica* lipase A, (CALA), *Candida antarctica* lipase B, (CALB), *Candida rugosa* lipase (CRL), *Mucor javanicus* lipase (MJL), *Mucor miehei* lipase (MML), *Pseudomonas fluorescens* lipase (PFL), porcine pancreas lipase (PPL), *Burkholderia cepacia* lipase (PSCI and PSIm), *Rhizopus arrhizus* lipase (RAL), and *Rhizopus niveus* lipase (RNL).

2. Generation of complex dynamic system for lipase screening

Pyridine-2-carbaldehyde, 3-phenylpropanal, butanal, octanal, 2-methylpropanal, 3-methylbutanal, ethyl 2-oxoacetate solution, cyclohexanecarbaldehyde, 2-ethylbutanal (0.1 mmol each), triethylamine (0.1 mmol), methyl 2-sulfanylacetate (0.2 mmol), butane-1-thiol (0.2 mmol), phenyl acetate (0.3 mmol), together with dry toluene (0.6 mL) were mixed in a vial, then transferred into a sealed-cap vial (1.5 mL) containing lipase (5 mg for CALB, 30 mg for CALA, 10 mg for MML, 50 mg for each other lipase) and ground 4 Å molecular sieves (20 mg). The sealed vials were then kept standing at r.t. NMR was used to follow the reaction progress.

3. Reaction system with 3-phenylpropanal

3-Phenylpropanal (0.1 mmol), triethylamine (0.1 mmol), methyl 2-sulfanylacetate (0.1 mmol), butane-1-thiol (0.1 mmol), phenyl acetate (0.3 mmol) and dry toluene (0.6 mL) were mixed in a vial, then transferred into a sealed-cap vial (1.5 mL) containing lipase (5 mg for CALB, 50 mg for each other lipase) and ground 4 Å molecular sieves (20 mg). The sealed vials were then kept standing at r.t. ^{1}H-NMR was used to follow the reaction progress. The reactions were quenched with dilute HCl (1 M), extracted with CH_2Cl_2 (2 mL × 3), and the combined organic layer was dried over $MgSO_4$ and the solvent evaporated. The crude mixture was purified by column chromatography.

4. Synthesis of compounds 1C and 1D

Methyl 2-((1-acetoxy-3-phenylpropyl)thio)acetate (1C). Colorless oil. ^{1}H-NMR (500 MHz, $CDCl_3$, 25 °C) δ 2.08 (s, 3H, CH3), 2.11-2.25 (m, 2H, CH2), 2.74 (t, J = 8.3 Hz, 2H, CH2), 3.37 (d, J = 16.0 Hz, 1H, CH2), 3.55 (d, J = 16.0 Hz, 1H, CH2), 3.72 (s, 3H, CH3), 6.02 (t, J = 6.8 Hz, 1H, CH), 7.17 (d, J = 7.6 Hz, 2H, CH), 7.21 (t, J = 7.1 Hz, 1H, CH), 7.29 (t, J = 7.6 Hz, 2H, CH); ^{13}C-NMR (125 MHz, $CDCl_3$, 25 °C) δ 21.2, 32.1, 32.9, 36.6, 52.7, 78.9, 126,4, 128.5, 128.7, 140.5, 170.6, 170.7. HRMS: found 305.0818, calc. for $C_{14}H_{18}NaO_4S$ [M + Na^+] 305.0818.

2-Phenethyl-1,3-oxathiolan-5-one (1D). Colorless oil. ^{1}H-NMR (500 MHz, $CDCl_3$, 25 °C) δ 2.10-2.19 (m, 1H, CH2), 2.29-2.39 (m, 1H, CH2), 2.75-2.88 (m, 2H, CH2), 3.65 (d, J = 16.2 Hz, 1H, CH2), 3.71 (d, J = 16.2 Hz ,1H, CH2), 5.46 (t, J = 6.2 Hz, 1H, CH), 7.18-7.25 (m, 3H, CH), 7.36-7.85 (t, J = 7.4 Hz, 2H, CH); ^{13}C-NMR (125 MHz, $CDCl_3$, 25 °C) δ 31.5, 31.8, 38.5, 81.5, 126.6, 128.6, 128.8, 139.9, 172.8. HRMS: found 231.0450, calc. for $C_{11}H_{12}NaO_2S$ [M + Na^+] 231.0450.

5. NMR-spectra of compounds 1C and 1D

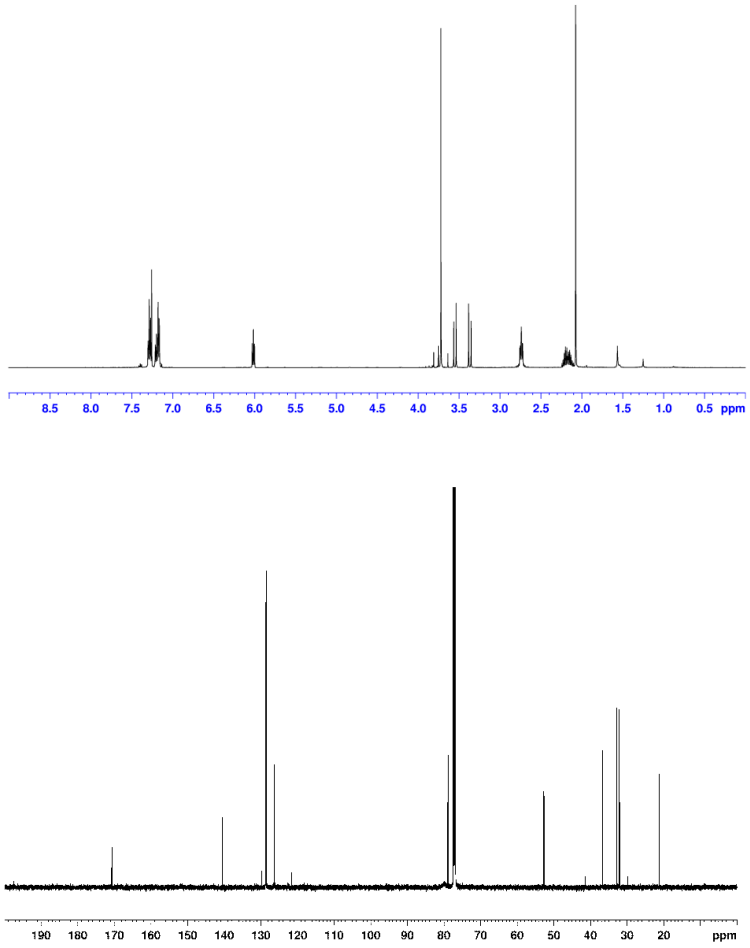

Figure S1. ^{1}H-NMR and ^{13}C-NMR spectra of compound **1C**

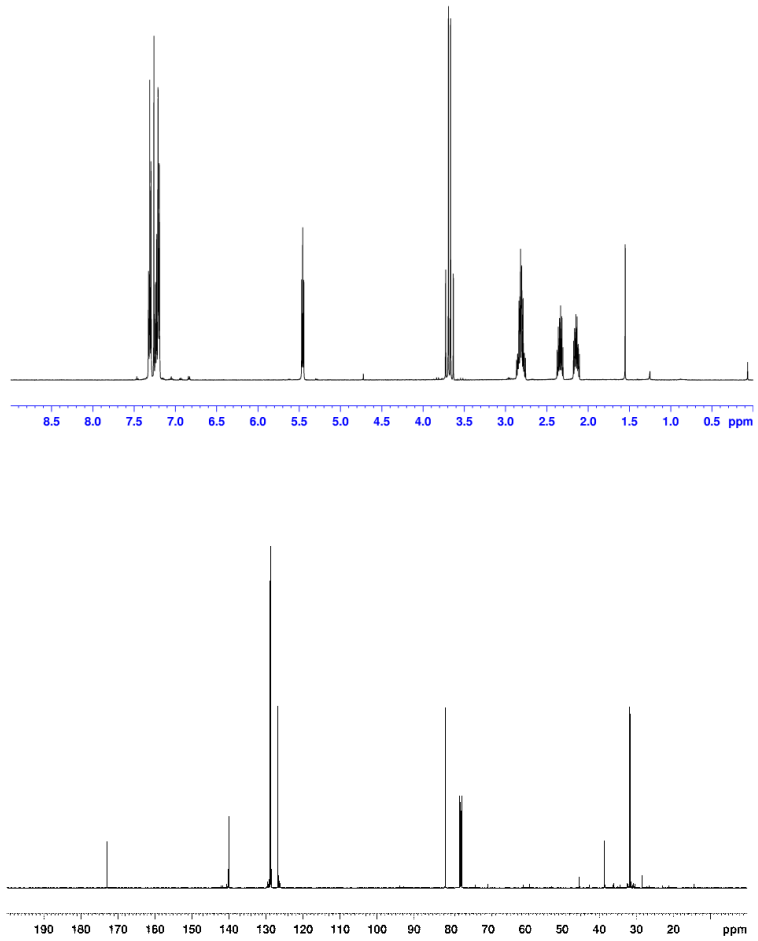

Figure S2. ^{1}H-NMR and ^{13}C-NMR spectra of compound **1D**

6. NMR response patterns

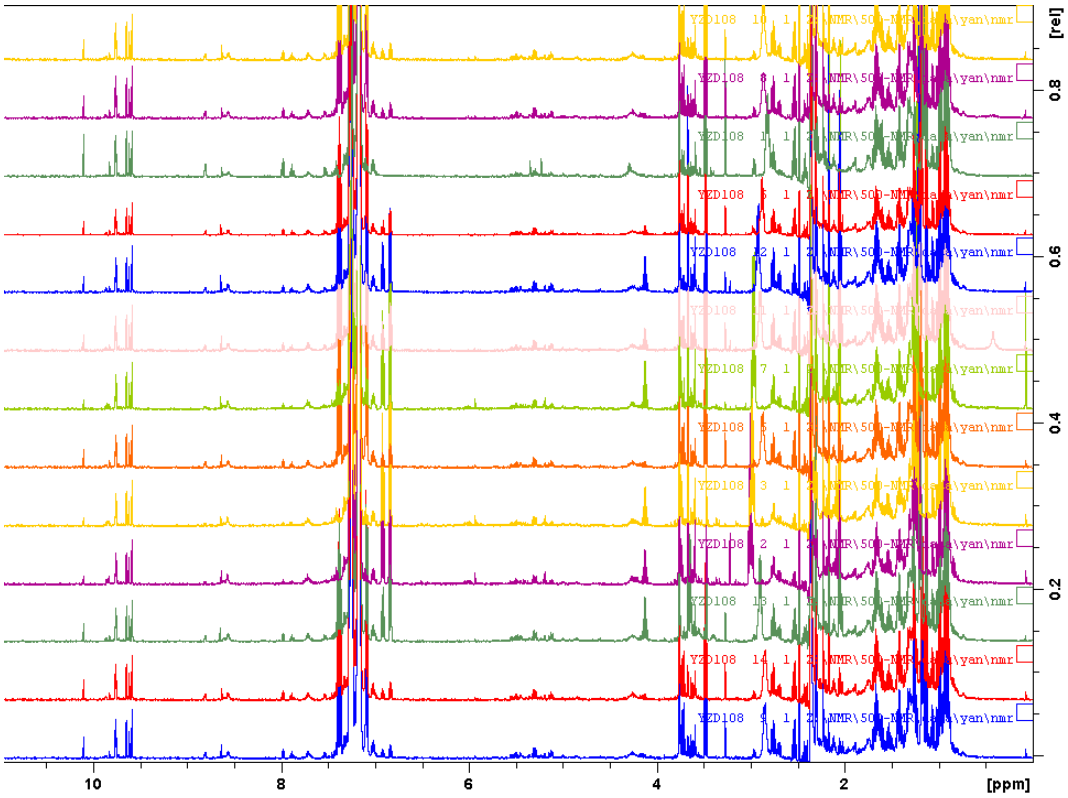

Figure S3. ^{1}H-NMR spectra for monitoring the dynamic hemithioacetal systems; full spectra with ANL, CALA, CALB, CRL, MJL, MML, PFL, PPL, PSCI, PSIm, RAL, RNL, probe (from top to bottom)

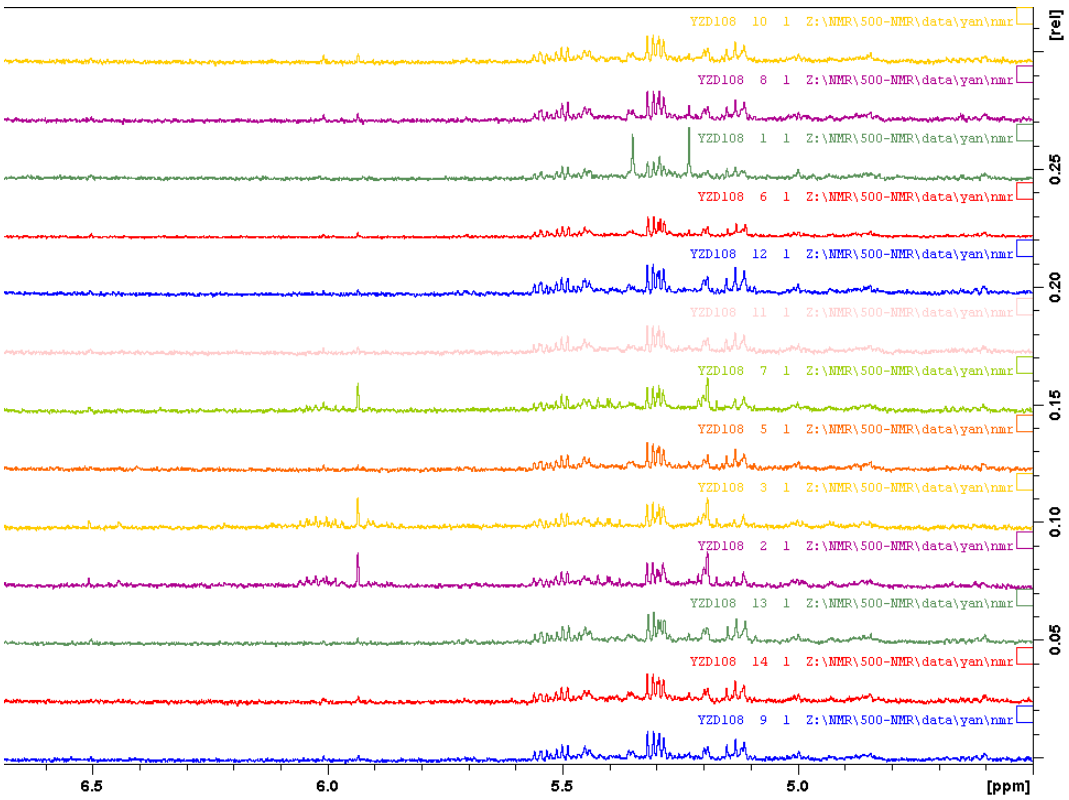

Figure S4. ^{1}H-NMR spectra for monitoring the dynamic hemithioacetal systems; top) enlarged area with ANL, CALA, CALB, CRL, MJL, MML, PFL, PPL, PSCI, PSIm, RAL, RNL, probe (from top to bottom)

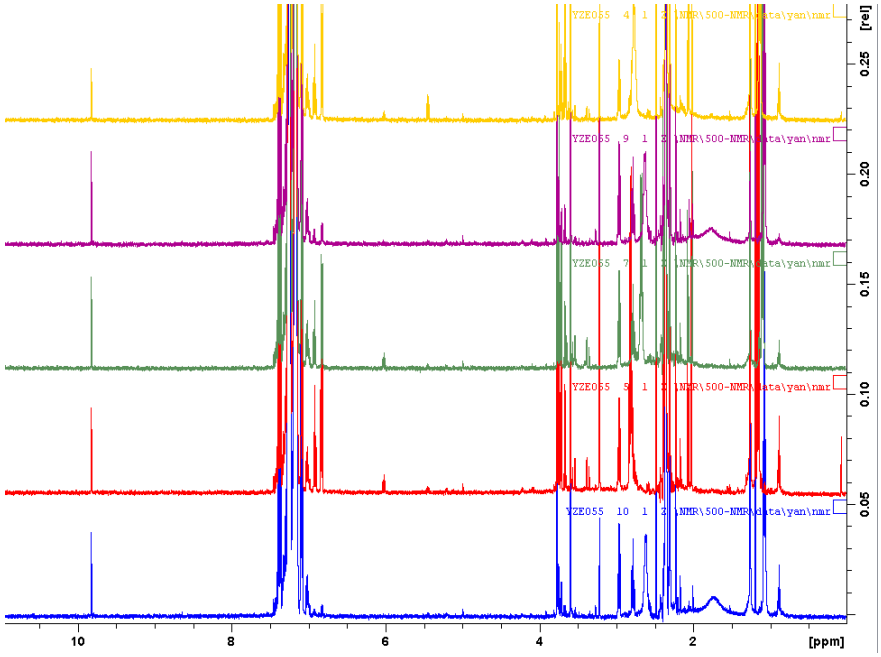

Figure S5. ^{1}H-NMR spectra for monitoring the reaction system with 3-phenylpropanal using different lipases; full spectra with CALB, CRL, PFL, PSIm, probe (from top to bottom)

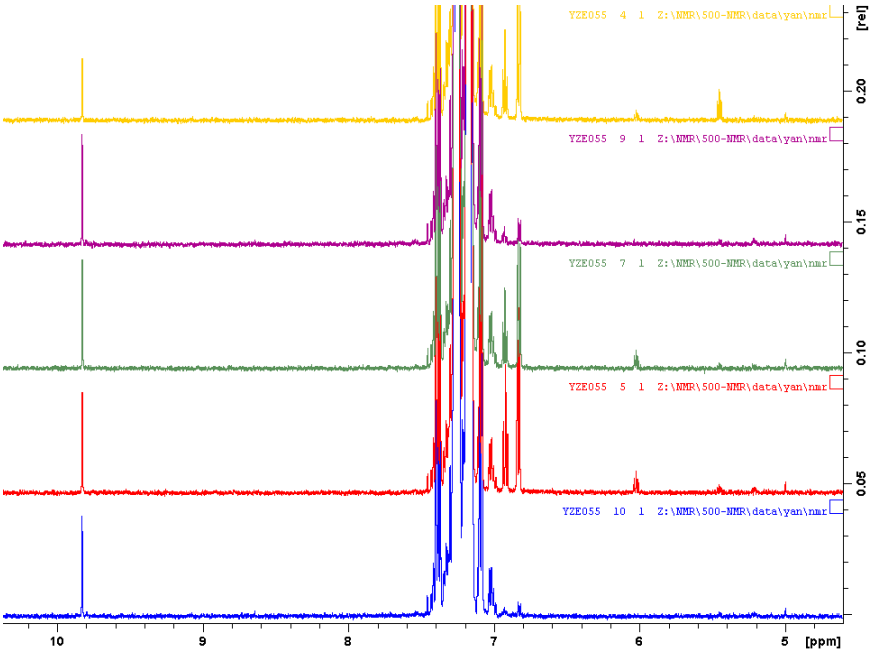

Figure S6. ^{1}H-NMR spectra for monitoring the reaction system with 3-phenylpropanal using different lipases; enlarged area with CALB, CRL, PFL, PSIm, probe (from top to bottom)

7. Linear discriminant analysis (LDA)

LDA analyses were performed using R (R Core Team, R Foundation For Statistical Computing, Vienna, Austria, 2015), and XLSTAT-Pro statistical analysis software version 2013.1 (Addinsoft, New York, NY) using forward stepwise discrimination. The dependent variables in this case, 12 classes of lipases, were grouped according to the discriminant functions calculated from the explanatory (or predictor) variables. The region from 6.7 - 4.6 ppm of the ^{1}H-NMR spectra of the 12 classes of lipases were divided into 21 regions at 0.1 ppm interval, and the integrated NMR intensities at each 0.1 ppm was recorded as explanatory variables. The training matrix (Table S1) was built with the 12 lipases using 21 integrated intensities (I01-I21), where each experiment was done in three replicates. The training matrix was then subjected to LDA analysis, and the factor scores for each of the 12 lipases were calculated (Table S3). Results from manova tests and cross-validation (jackknife) analysis are displayed in Table S5, Table S6, and Table S7. The ^{1}H-NMR spectrum of the probe lipase PFL was recorded and processed in the same manner (Table S2). The factor scores calculated by posterior probability (Table S4) were used to predict the classification of the probe lipase. Cross-validation was also performed on the dataset where lipases ANL, CALA, CRL, MJL, MML, PPL, RNL, and RAL were grouped as lipase class LIP. The calculated contigency table and test results are displayed in Table S8 and Table S9, respectively.

Table S1. Training matrix of NMR-response patterns from 6.7 – 4.6 ppm

Lipase	I01	I02	I03	I04	I05	I06	I07	I08	I09	I10	I11	I12	I13	I14	I15	I16	I17	I18	I19	I20	I21
ANL	0.008	0.003	−0.019	−0.056	−0.072	−0.042	−0.038	−0.046	−0.079	−0.035	−0.018	0.204	0.382	0.434	0.478	0.505	0.119	0.141	0.210	0.003	0.132
CALA	−0.027	−0.044	−0.081	−0.078	−0.067	−0.077	−0.060	−0.069	−0.079	−0.053	−0.017	0.199	0.474	0.481	0.536	0.530	0.139	0.156	0.247	0.052	0.142
CALB	−0.048	−0.022	−0.040	−0.057	−0.031	−0.011	0.003	−0.013	−0.020	0.012	0.046	0.408	0.520	0.395	0.451	0.526	0.150	0.137	0.207	0.039	0.150
CRL	−0.008	−0.020	−0.036	−0.047	−0.059	−0.041	−0.040	−0.037	−0.053	−0.040	−0.005	0.177	0.407	0.445	0.481	0.497	0.155	0.137	0.232	0.036	0.158
MJL	0.053	0.024	−0.008	−0.027	−0.033	−0.036	−0.017	−0.025	−0.035	0.013	0.022	0.252	0.441	0.476	0.462	0.598	0.201	0.210	0.290	0.085	0.182
MML	−0.005	−0.029	−0.035	−0.054	−0.060	−0.043	−0.065	−0.038	−0.058	−0.030	−0.009	0.211	0.415	0.464	0.458	0.499	0.158	0.156	0.220	0.037	0.162
PFL	−0.027	−0.042	−0.026	−0.023	0.001	−0.002	0.107	0.161	0.017	0.029	0.088	0.266	0.519	0.527	0.530	0.515	0.184	0.178	0.233	0.021	0.122
PPL	0.026	0.013	−0.002	−0.028	−0.041	−0.073	−0.075	−0.040	−0.071	−0.022	−0.032	0.205	0.392	0.458	0.439	0.500	0.133	0.118	0.195	−0.013	0.101
PSCI	−0.035	0.010	0.019	−0.010	0.027	0.047	0.247	0.246	0.090	0.057	0.097	0.246	0.407	0.324	0.429	0.378	0.116	0.151	0.146	−0.025	0.060
PSIm	0.034	0.036	0.053	0.044	0.045	0.042	0.184	0.182	0.031	0.015	0.042	0.186	0.331	0.260	0.387	0.346	0.066	0.081	0.135	−0.044	0.042
RAL	−0.002	−0.025	−0.026	−0.040	−0.041	−0.031	−0.018	−0.005	−0.014	0.021	0.054	0.284	0.512	0.534	0.556	0.602	0.214	0.202	0.301	0.092	0.219
RNL	0.013	−0.015	−0.016	−0.044	−0.038	−0.053	−0.039	−0.042	−0.079	−0.008	−0.010	0.186	0.388	0.494	0.476	0.466	0.124	0.123	0.225	0.012	0.117
ANL	0.044	0.029	0.026	−0.003	−0.031	−0.023	−0.034	−0.003	−0.055	−0.036	−0.001	0.198	0.433	0.486	0.476	0.495	0.142	0.130	0.195	0.016	0.128
CALA	−0.056	−0.034	−0.039	−0.082	−0.088	−0.077	−0.094	−0.077	−0.092	−0.045	−0.056	0.210	0.377	0.513	0.469	0.438	0.126	0.131	0.240	0.018	0.098
CALB	0.067	0.061	0.031	0.014	0.020	0.004	0.014	−0.000	−0.014	0.016	0.038	0.373	0.496	0.352	0.481	0.549	0.127	0.135	0.163	0.000	0.150
CRL	−0.021	−0.070	−0.047	−0.042	−0.050	−0.050	−0.063	−0.034	−0.062	−0.016	0.018	0.205	0.461	0.487	0.487	0.556	0.166	0.177	0.288	0.071	0.167
MJL	−0.022	−0.036	−0.048	−0.066	−0.077	−0.044	−0.059	−0.041	−0.056	−0.008	0.022	0.251	0.426	0.494	0.456	0.547	0.170	0.171	0.257	0.046	0.162
MML	0.001	−0.019	−0.012	−0.036	−0.043	−0.036	−0.048	−0.041	−0.075	−0.012	−0.033	0.206	0.404	0.489	0.455	0.486	0.172	0.154	0.246	0.034	0.137
PFL	0.010	0.041	0.016	0.033	0.034	0.013	0.128	0.158	0.018	0.029	0.033	0.219	0.494	0.400	0.473	0.484	0.115	0.140	0.193	−0.008	0.083
PPL	−0.020	−0.030	−0.041	−0.045	−0.058	−0.070	−0.054	−0.039	−0.061	−0.032	−0.019	0.218	0.354	0.444	0.444	0.465	0.130	0.147	0.244	0.022	0.132
PSCI	0.007	−0.001	0.005	−0.004	0.017	−0.013	0.175	0.180	0.018	−0.026	−0.001	0.147	0.284	0.246	0.334	0.305	0.069	0.074	0.095	−0.049	0.025
PSIm	0.007	0.018	0.027	0.020	0.046	0.034	0.203	0.197	0.018	−0.001	0.021	0.167	0.381	0.315	0.358	0.354	0.074	0.071	0.127	−0.043	0.042
RAL	−0.041	−0.008	−0.011	−0.053	−0.061	−0.059	−0.038	−0.012	−0.052	0.009	0.029	0.274	0.448	0.526	0.476	0.572	0.179	0.163	0.251	0.072	0.172
RNL	0.042	0.016	−0.008	−0.028	−0.027	−0.034	−0.041	−0.009	−0.048	−0.010	0.003	0.197	0.440	0.466	0.464	0.500	0.158	0.155	0.246	0.049	0.144
ANL	0.007	−0.020	−0.036	−0.041	−0.054	−0.056	−0.046	−0.033	−0.047	−0.033	0.016	0.220	0.405	0.470	0.493	0.532	0.166	0.162	0.257	0.033	0.170
CALA	−0.036	−0.025	−0.050	−0.048	−0.077	−0.077	−0.035	−0.042	−0.068	−0.029	−0.020	0.214	0.453	0.505	0.517	0.487	0.140	0.135	0.234	0.035	0.136
CALB	0.009	0.004	−0.007	−0.015	−0.000	−0.022	0.012	−0.010	−0.016	0.010	0.042	0.384	0.518	0.393	0.476	0.516	0.126	0.138	0.171	0.026	0.164
CRL	−0.051	−0.053	−0.062	−0.077	−0.088	−0.071	−0.060	−0.052	−0.075	−0.043	−0.009	0.245	0.418	0.509	0.474	0.522	0.166	0.149	0.224	0.058	0.167
MJL	0.060	0.045	0.027	0.009	−0.020	−0.016	−0.015	−0.030	−0.038	0.004	0.028	0.207	0.395	0.351	0.424	0.542	0.182	0.172	0.251	0.056	0.138
MML	−0.036	−0.050	−0.068	−0.068	−0.064	−0.064	−0.057	−0.029	−0.052	−0.013	−0.010	0.211	0.454	0.476	0.487	0.571	0.204	0.204	0.323	0.085	0.208
PFL	0.019	0.028	0.032	0.020	0.046	0.025	0.137	0.156	0.018	0.041	0.061	0.199	0.465	0.414	0.471	0.467	0.128	0.130	0.202	−0.015	0.069
PPL	0.018	0.016	−0.011	−0.042	−0.067	−0.060	−0.056	−0.034	−0.079	−0.025	−0.043	0.218	0.421	0.456	0.460	0.514	0.142	0.148	0.213	0.034	0.132
PSCI	0.057	0.033	0.065	0.045	0.033	0.053	0.193	0.200	0.039	−0.014	0.028	0.172	0.334	0.254	0.399	0.372	0.085	0.080	0.115	−0.038	0.062
PSIm	0.034	0.056	0.038	0.012	0.038	0.020	0.204	0.236	0.012	−0.004	0.021	0.200	0.369	0.296	0.434	0.372	0.081	0.090	0.117	−0.050	0.034
RAL	−0.021	−0.025	−0.024	−0.037	−0.035	−0.028	−0.036	−0.018	−0.029	0.009	0.038	0.260	0.459	0.482	0.510	0.556	0.175	0.192	0.242	0.052	0.168
RNL	−0.087	−0.048	−0.094	−0.100	−0.096	−0.100	−0.086	−0.064	−0.084	−0.047	−0.026	0.197	0.430	0.474	0.466	0.510	0.135	0.104	0.223	0.012	0.134

I01: 6.7-6.6 ppm, I02: 6.6-6.5 ppm, I03: 6.5-6.4 ppm, I04: 6.4-6.3 ppm, I05: 6.3-6.2 ppm, I06: 6.2-6.1 ppm, I07: 6.1-6.0 ppm, I08: 6.0-5.9 ppm, I09: 5.9-5.8 ppm, I10: 5.8-5.7 ppm, I11: 5.7-5.6 ppm, I12: 5.6-5.5 ppm, I13: 5.5-5.4 ppm, I14: 5.4-5.3 ppm, I15: 5.3-5.2 ppm, I16: 5.2-5.1 ppm, I17: 5.1-5.0 ppm, I18: 5.0-4.9 ppm, I19: 4.9-4.8 ppm, I20: 4.8-4.7 ppm, I21: 4.7-4.6 ppm

Table S2. NMR response pattern for probe lipase from 6.7 – 4.6 ppm

Lipase	I01	I02	I03	I04	I05	I06	I07	I08	I09	I10	I11	I12	I13	I14	I15	I16	I17	I18	I19	I20	I21
Probe	0.007	0.020	−0.014	−0.040	−0.044	−0.048	−0.031	−0.019	−0.054	−0.008	0.002	0.209	0.434	0.498	0.492	0.528	0.147	0.171	0.229	0.035	0.159
Probe	−0.028	−0.021	−0.027	−0.046	−0.035	−0.049	−0.042	−0.037	−0.062	−0.035	−0.008	0.213	0.471	0.499	0.499	0.521	0.166	0.162	0.232	0.033	0.150
Probe	−0.012	−0.033	−0.054	−0.078	−0.063	−0.065	−0.082	−0.048	−0.061	−0.025	0.000	0.187	0.421	0.472	0.484	0.485	0.157	0.144	0.261	0.023	0.136

Factor scores

Table S3. Factor scores

Lipase	LD1	LD2	LD3	LD4	LD5	LD6	LD7	LD8	LD9	LD10	LD11
ANL	−16.84	9.69	7.72	4.19	−0.69	1.43	2.61	1.52	0.70	0.05	−0.51
CALA	−22.65	16.86	8.27	−5.45	3.07	−1.02	−0.22	1.07	−2.32	0.69	0.41
CALB	−13.75	−47.01	4.74	−0.81	0.75	−0.55	−1.51	−0.11	−2.41	0.25	0.20
CRL	−12.05	8.40	−4.49	−0.49	0.82	1.43	0.95	1.99	−0.77	1.29	−0.33
MJL	−16.15	7.52	1.96	6.56	−1.41	−2.89	−2.30	−0.45	−1.01	−0.81	−1.20
MML	−12.75	−1.46	−7.46	1.90	0.85	2.85	0.66	−0.11	0.33	1.28	0.52
PFL	22.53	5.18	0.22	−1.86	−4.05	−2.84	1.71	−2.33	−0.13	1.60	0.87
PPL	−12.10	5.26	3.36	1.33	0.74	0.34	−2.32	−2.86	3.36	0.57	0.85
PSCI	66.58	3.78	0.59	1.65	4.09	−1.42	−0.70	−0.43	0.35	2.00	−2.45
PSIm	47.99	−4.66	0.19	−0.80	−1.76	0.51	−1.26	1.78	−0.09	−1.14	0.47
RAL	−11.22	−2.62	−7.88	−1.98	3.17	−2.38	−0.12	0.55	0.43	0.72	−1.27
RNL	−19.20	3.39	−3.53	−3.65	−2.76	1.88	−2.25	2.05	0.62	−0.07	−0.35
ANL	−16.86	10.73	5.99	0.92	−1.18	−0.34	3.39	1.25	1.33	0.67	0.07
CALA	−21.51	17.15	8.43	−4.14	4.62	1.12	0.38	−1.04	0.16	−1.30	−1.38
CALB	−16.17	−46.99	4.36	1.01	−0.45	−0.71	0.42	−0.48	1.34	0.02	−0.16
CRL	−11.67	7.12	−3.48	−0.64	0.57	−0.90	0.68	−0.39	−0.76	−0.12	2.66
MJL	−17.75	4.64	0.45	5.15	−3.31	−0.50	−1.42	0.24	−0.19	0.76	1.05
MML	−15.68	−1.71	−8.39	0.31	−0.45	2.81	1.45	−2.98	−1.73	−0.68	−0.76
PFL	22.89	3.33	1.23	−3.91	−4.52	−2.44	0.23	−3.10	0.12	−0.81	0.35
PPL	−12.57	6.29	4.12	0.97	−0.01	0.96	0.08	−0.51	0.83	−1.84	0.49
PSCI	66.62	4.67	0.38	1.97	4.29	2.58	−2.65	−0.36	−0.05	−0.00	1.08
PSIm	49.27	−5.89	1.15	0.17	−3.97	1.87	1.74	1.07	−1.36	−0.12	0.15
RAL	−11.46	−1.30	−9.47	−0.62	1.28	−3.87	−0.28	2.61	1.90	−2.18	−0.33
RNL	−20.30	1.82	−2.96	−1.36	−3.62	2.33	−2.90	1.16	0.16	0.87	−0.74
ANL	−17.28	8.54	5.32	3.30	−1.11	−0.15	2.11	0.98	1.09	0.56	−0.07
CALA	−19.58	17.37	8.15	−4.35	2.00	−0.00	−0.52	−0.75	−0.97	−0.11	−0.19
CALB	−15.17	−48.01	4.83	−1.33	3.04	0.53	1.06	0.26	0.62	0.29	0.16
CRL	−11.54	4.66	−4.78	0.20	2.61	−1.25	0.61	1.22	−1.21	−0.86	1.33
MJL	−17.02	5.09	2.15	7.18	−1.75	−1.79	−1.19	0.24	−2.28	−0.33	−0.53
MML	−13.40	−1.55	−10.91	1.32	0.43	3.13	1.94	−2.18	−0.54	−0.42	−1.00
PFL	24.19	2.90	−0.73	−3.80	−4.79	−2.16	−0.09	−0.44	0.17	0.77	−0.88
PPL	−13.50	6.58	1.36	1.44	1.72	1.43	−1.32	−1.94	0.64	−0.92	0.13
PSCI	67.59	3.64	0.34	2.34	4.75	−0.23	0.92	0.80	0.18	0.03	1.13
PSIm	50.14	−5.86	1.72	−2.35	−3.87	1.62	0.92	0.98	0.38	−1.98	−0.81
RAL	−9.08	−1.32	−10.63	−0.96	2.72	−2.97	1.07	−0.44	0.24	0.06	0.28
RNL	−20.54	3.74	−2.33	−3.41	−1.81	1.59	−1.89	1.14	0.87	1.21	0.75

Table S4. Predicted lipase and scores for probe lipase

Lipase	Prediction	LD1	LD2	LD3	LD4	LD5	LD6	LD7	LD8	LD9	LD10	LD11
Probe	RNL	−21.36	7.45	−2.70	0.43	−1.90	0.36	1.83	1.73	1.91	−0.22	−1.70
Probe	RNL	−21.39	0.57	0.80	−1.08	2.25	0.64	7.39	−0.80	−0.44	0.64	−1.49
Probe	RNL	−24.47	5.34	1.85	−1.69	−1.83	4.27	−2.46	0.70	1.17	3.05	−1.12

Manova tests and cross-validation

Table S5. Manova tests

Test	Df	Residuals	Measure	Approx. F	Df (num)	Df (den)	Prob>F
Wilks	11	24	1.64e-12	4.32	231	66.6	9e-11***
Pillai	11	24	8.06	1.83	231	154	3.4e-05***
Hotelling-Lawley	11	24	1639	15.5	231	24	7.8e-11***
Roy	11	24	1212	808	21	14	<2e-16***

Table S6. Cross-validation (jackknife) contingency table

Lipase	ANL	CALA	CALB	CRL	MJL	MML	PFL	PPL	PSCI	PSIm	RAL	RNL
ANL	2	1	0	0	0	0	0	0	0	0	0	0
CALA	1	0	0	0	0	0	0	1	0	0	0	1
CALB	0	0	3	0	0	0	0	0	0	0	0	0
CRL	0	0	0	0	0	1	0	1	0	0	0	1
MJL	1	0	0	0	0	0	0	1	0	0	1	0
MML	0	0	0	0	0	0	0	0	0	0	2	1
PFL	0	0	0	0	0	0	3	0	0	0	0	0
PPL	1	0	0	1	0	0	0	1	0	0	0	0
PSCI	0	0	0	0	0	0	0	0	2	1	0	0
PSIm	0	0	0	0	0	0	0	0	1	2	0	0
RAL	0	0	0	0	0	2	0	0	0	0	0	1
RNL	0	0	0	0	1	0	0	1	0	0	0	1

Table S7. Cross-validation results (total = 0.3889)

ANL	CALA	CALB	CRL	MJL	MML	PFL	PPL	PSCI	PSIm	RAL	RNL
0.6667	0.0000	1.0000	0.0000	0.0000	0.0000	1.0000	0.3333	0.6667	0.6667	0.0000	0.3333

Cross-validation on grouped dataset

Table S8. Cross-validation contingency table, grouped dataset

Lipase	CALB	LIP	PFL	PSCI	PSIm
CALB	3	0	0	0	0
LIP	0	24	0	0	0
PFL	0	0	3	0	0
PSCI	0	0	0	2	1
PSIm	0	0	0	0	3

Table S9. Cross-validation results, grouped dataset (total = 0.9722)

CALB	LIP	PFL	PSCI	PSIm
1.0000	1.0000	1.0000	0.6667	1.0000