Supplementary Information

Bayesian Optimization of Comprehensive Two-dimensional Liquid Chromatography Separations

Jim Boelrijk[a,d], Bob Pirok[a,b], Bernd Ensing[a,c,*], Patrick Forré[a,d]

[a] *AI4Science Lab*
[b] *Analytical Chemistry Group, Van 't Hoff Institute for Molecular Sciences*
[c] *Computational Chemistry Group, Van 't Hoff Institute for Molecular Sciences*
[d] *AMLab, Informatics Institute*
University of Amsterdam, The Netherlands
[*] *Corresponding author. Email: b.ensing@uva.nl*

S.1. Influence of c_1, c_2, c_3 and c_4 in sampling strategy B and D

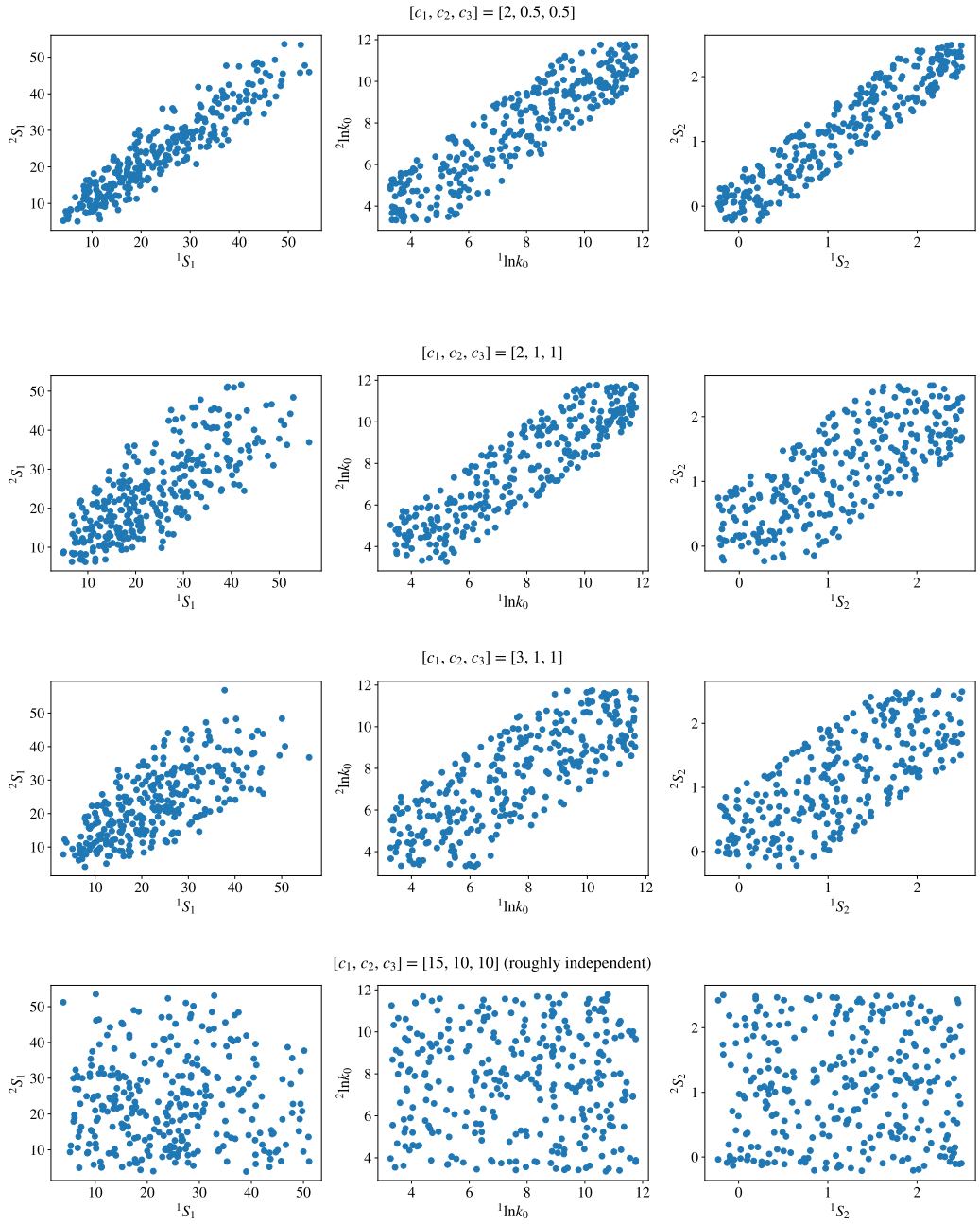


Figure S-1: Influence of parameters c_1, c_2 and c_3 on dependence of first- and second dimension retention parameters in sampling strategy B. Notice that the parameters of each dimension become more independent as c_1, c_2 and c_3 increase.

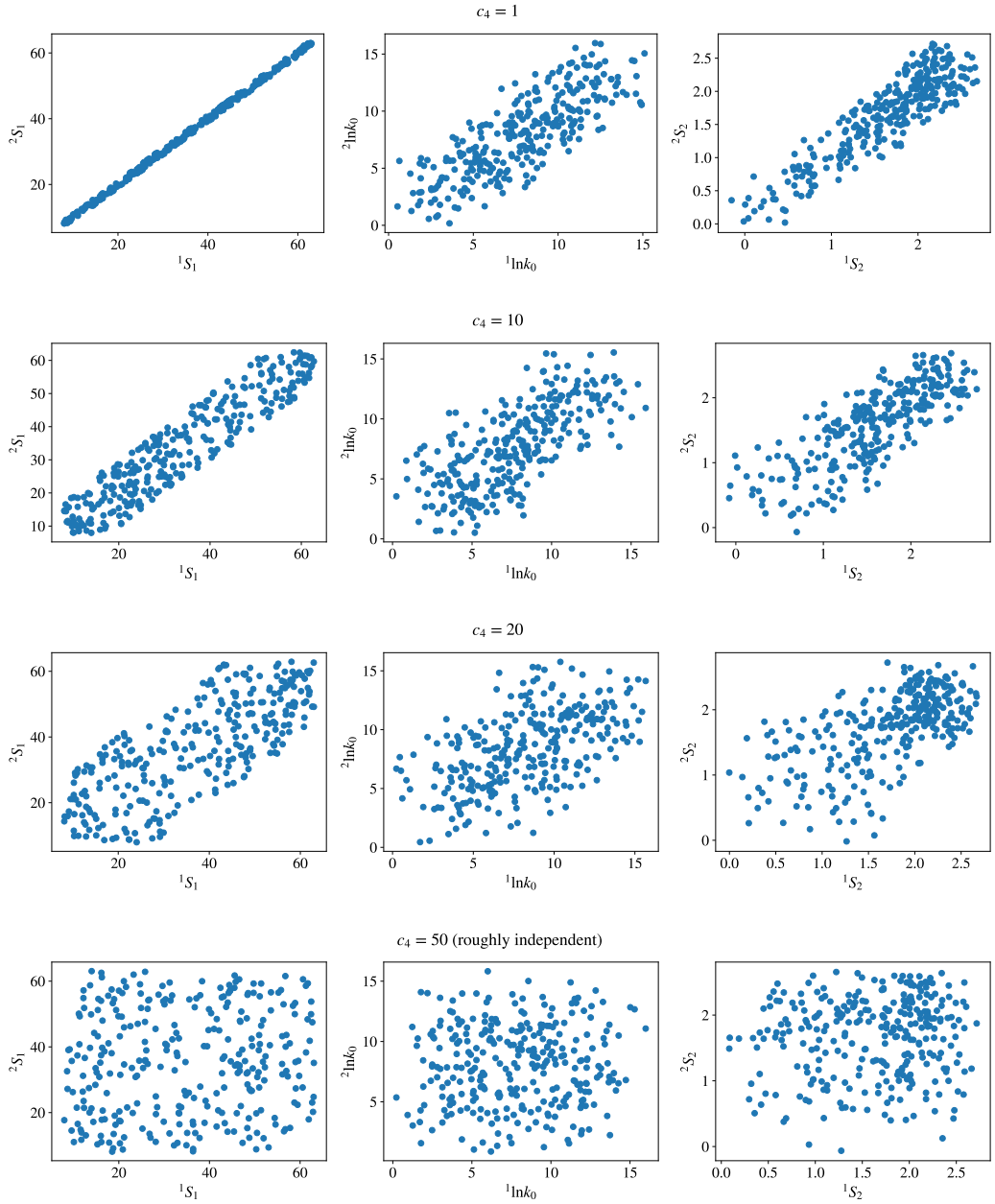


Figure S-2: Influence of parameters c_4, c_5 and c_6 on dependence of first- and second dimension retention parameters in sampling strategy D. Notice that the parameters of each dimension become more independent as c_4, c_5 and c_6 increase.

S.2. Quantitative description of an undirected graph, connected components and the objective function

Definition 1 (Undirected graph). *An* (undirected) graph *is a tuple* $G = (V, E)$*, where* V *is the set of vertices/nodes and* $E \subseteq V \times V$ *is the set of (undirected) edges, such that* $(v_1, v_2) \in E$ *then also* $(v_2, v_1) \in E$*.*
Instead of the tuple (v_1, v_2) *we will also write* $v_1 - v_2$*, mirroring the graphical representation of edges.*

Definition 2 (Walks in graphs). *Let* $G = (V, E)$ *be an undirected graph. A* walk π *in* G *is a sequence of vertices*

$$\pi = (w_0 - w_2 - \cdots - w_{n-1} - w_n),$$

such that $n \geq 0$ *and* $w_i \in V$ *and* $w_i - w_{i-1} \in E$ *for all* $i = 1, \ldots, n$*.*

Definition 3 (Connected vertices in an undirected graph). *Let* $G = (V, E)$ *be an undirected graph and* $v_1, v_2 \in V$ *vertices. We say that* v_1 *and* v_2 *are* connected *in* G *if there exists a walk* $\pi = (w_0 - \cdots - w_n)$ *inside* G *such that* $w_0 = v_1$ *and* $w_n = v_2$*. If this is the case we write:*

$$v_1 \sim v_2.$$

Lemma 1. *Let* $G = (V, E)$ *be an undirected graph. Then the relation* $\sim$ *is an equivalence relation on the set of vertices* V*.*

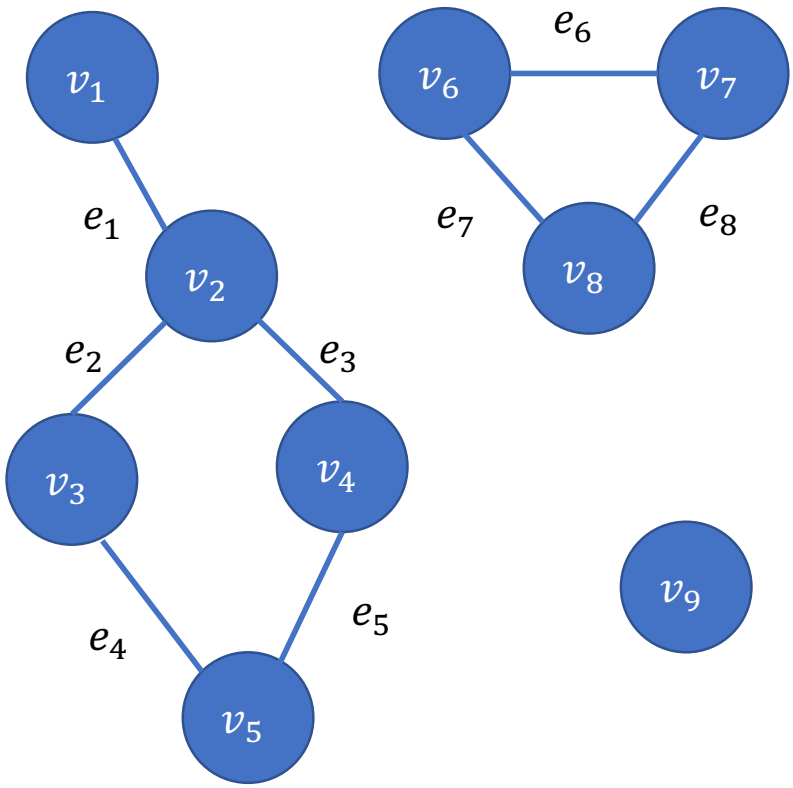
Proof: Reflexivity: For every $v \in V$ we have the trivial walk $\pi = (v)$ without edges (with $n = 0$). So we get $v \sim v$.
Symmetry: If $v_1 \sim v_2$ then there exists a walk $\pi = (w_0 - \cdots - w_n)$ such that $w_0 = v_1$ and $w_n = v_2$. Then the inverse walk $\pi' = (w'_0 - \cdots - w'_n)$ with $w'_m := w_{n-m}$ is again a walk in G and we have $w'_0 = w_n = v_2$ and $w'_n = w_0 = v_1$. Note that if $w_{i-1} - w_i \in E$ then also $w_i - w_{i-1} \in E$. This shows that $v_2 \sim v_1$.
Transitivity: If $v_1 \sim v_2$ and $v_2 \sim v_3$ then there exists walks $\pi_1 = (w_0 - \cdots - w_n)$ and $\pi_2 = (w'_0 - \cdots - w'_m)$ with $w_0 = v_1$, $w_n = v_2 = w'_0$ and $w'_m = v_3$. Then the composite walk:

$$\pi := (w_0 - \cdots - w_n = w'_0 - \cdots - w'_m)$$

is a walk in G and satisfies $w_0 = v_1$ and $w'_m = v_3$. This shows $v_1 \sim v_3$.

Definition 4 (Connected components of a graph). *Let $G = (V, E)$ be an undirected graph. Then the equivalence classes of the equivalence relation $\sim$ are called the* connected components *of G.*

An illustration of a graph of nine nodes $(v_1, ..., v_9)$ and eight edges $(e_1, ..., e_8)$ is shown in Figure S-3. The graph has three connected components: $(v_1, v_2, v_3, v_4, v_5)$, (v_6, v_7, v_8) and (v_9). In the objective function proposed in this work, every analyte peak is defined as a node in the graph. An edge is defined between two nodes if the resolution (as defined in Equation 28 in the main text) between the two nodes is lower than 1.



S-3 A graph with three connected components: $(v_1, v_2, v_3, v_4, v_5)$, (v_6, v_7, v_8) and (v_9).

S.3. Iterations needed to find 29 connected components for sample A

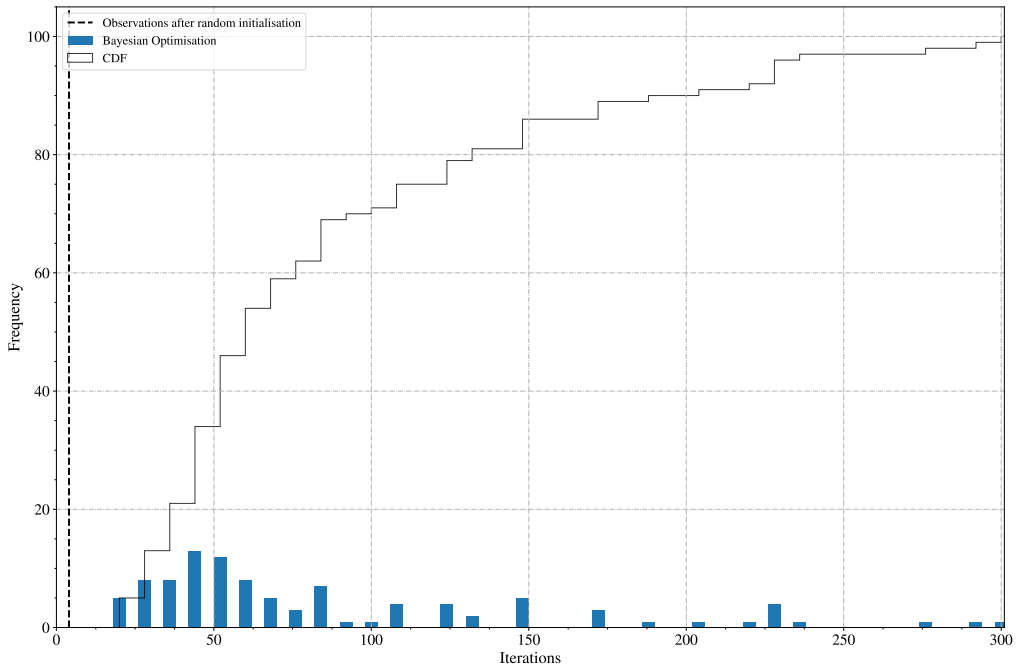


Figure S-4: Number of iterations needed for the BO algorithm to reach a score of 29 connected components for sample A, for 100 trials with different random seeds. The grey line denotes the cumulative distribution function (CDF). The black vertical line denotes the number of initial random observations.

S.4. Retention parameters of synthetic samples

Table S-1: Retention parameters for 50 compounds sampled using strategy A

${}^1\ln k_0$	${}^1S_1$	${}^1S_2$	${}^2\ln k_0$	${}^2S_1$	${}^2S_2$
4.321	19.582	1.594	8.371	35.271	1.804
5.579	28.656	2.176	7.920	27.070	1.376
7.422	25.177	1.454	9.775	36.675	1.701
6.449	17.499	0.960	11.693	44.872	1.859
5.948	26.501	1.626	3.773	7.213	0.249
10.099	41.046	1.998	8.765	20.780	0.771
3.340	24.778	2.195	8.678	14.449	0.308
7.832	10.685	0.066	10.324	33.433	1.517

7.321	31.528	1.907	3.906	24.802	2.342
9.111	37.497	2.019	9.153	14.249	0.205
8.400	7.665	-0.157	11.628	27.092	0.818
4.971	10.734	0.358	5.652	18.758	1.177
4.397	23.875	1.794	5.952	25.241	1.560
3.556	8.557	0.321	10.356	15.092	0.216
5.114	14.779	0.949	9.225	29.234	1.270
5.865	14.189	0.641	5.345	9.103	0.202
11.145	29.771	1.129	7.788	15.491	0.482
6.535	36.824	2.472	3.545	10.847	0.716
7.928	19.089	0.734	6.947	27.095	1.520
5.981	22.870	1.303	5.435	12.024	0.528
9.365	40.053	1.928	11.113	41.009	1.801
10.963	19.120	0.456	3.554	26.565	2.479
7.575	36.889	2.065	10.458	19.457	0.567
8.587	38.841	2.213	11.461	24.658	0.689
3.888	17.120	1.377	8.198	21.732	0.909
9.353	8.864	-0.190	4.561	27.908	2.176
4.080	5.676	-0.033	5.151	10.921	0.509
11.143	34.326	1.421	6.156	18.438	0.942
5.635	19.804	1.266	5.677	9.129	0.156
11.269	23.508	0.676	11.264	24.358	0.764
8.553	15.823	0.475	10.947	45.510	1.983
5.087	7.648	0.129	8.097	27.008	1.321
10.888	48.614	2.154	9.362	27.166	1.231
4.505	14.353	0.895	3.759	23.405	2.080
4.395	9.891	0.446	8.316	10.006	0.016
7.092	16.722	0.644	3.570	21.108	1.756
5.392	29.861	2.195	4.393	5.545	-0.142
7.819	23.612	1.149	5.987	34.363	2.219
7.062	33.828	2.098	3.403	7.924	0.324
11.222	34.302	1.323	7.381	24.507	1.398
4.072	13.810	0.974	3.722	6.047	0.069
4.003	9.968	0.442	5.771	21.279	1.269
6.250	16.443	0.776	10.349	16.782	0.299
9.632	27.790	1.111	5.181	25.019	1.977

11.634	37.855	1.586	6.635	15.507	0.605
4.826	20.415	1.427	8.005	17.093	0.545
4.347	4.802	-0.220	8.559	20.614	0.806
3.483	5.937	0.036	11.211	18.958	0.374
3.320	11.541	0.870	7.107	24.829	1.397
5.800	14.666	0.784	11.728	50.852	2.169

Table S-2: Retention parameters for 50 compounds sampled using strategy B

${}^1\ln k_0$	${}^1S_1$	${}^1S_2$	${}^2\ln k_0$	${}^2S_1$	${}^2S_2$
9.542	16.754	0.343	10.334	12.324	0.025
4.711	13.206	0.743	3.453	11.126	0.669
9.219	7.850	-0.228	9.071	9.219	-0.157
5.448	30.869	2.246	6.771	33.799	2.167
7.034	6.447	-0.171	5.740	6.321	-0.092
4.459	28.035	2.299	4.830	31.311	2.445
5.368	20.837	1.507	7.324	26.117	1.430
8.999	19.981	0.672	8.671	25.702	1.194
8.555	19.565	0.763	7.070	10.524	0.111
8.582	28.053	1.420	7.384	33.523	2.022
9.758	39.250	1.961	10.668	49.191	2.301
11.208	35.554	1.386	9.774	36.329	1.767
11.038	29.563	1.008	11.494	45.673	1.873
9.145	13.233	0.163	9.404	15.053	0.360
3.895	21.061	1.646	4.600	26.882	2.279
4.598	26.847	2.079	3.657	18.684	1.635
4.647	7.280	0.156	5.380	15.658	0.776
4.567	11.343	0.511	5.900	20.493	1.312
6.184	8.620	0.050	6.138	16.400	0.855
6.267	16.308	0.860	5.549	9.308	0.295
10.636	31.778	1.286	11.030	24.421	0.807
8.957	45.211	2.309	10.716	33.686	1.348
3.811	4.243	-0.179	4.117	7.528	0.290
10.146	39.479	1.766	10.911	45.715	2.167
4.952	22.347	1.717	5.434	24.306	1.800
5.378	27.470	1.803	5.709	27.752	1.825

4.252	26.078	2.045	5.783	30.761	2.076
6.775	17.333	0.809	8.709	18.848	0.725
8.824	29.697	1.458	10.100	47.864	2.258
10.991	42.920	1.928	10.407	26.777	0.986
10.653	25.487	0.918	10.745	33.148	1.305
9.440	13.255	0.208	8.478	17.379	0.600
8.092	11.348	0.111	7.812	7.605	-0.155
7.932	20.596	0.966	6.436	17.898	0.895
9.474	38.586	1.914	7.513	25.788	1.427
6.064	15.322	0.663	5.025	17.429	1.050
8.598	36.011	2.028	9.962	42.121	2.069
4.102	24.863	2.049	3.427	26.926	2.499
4.793	14.961	0.898	3.405	14.656	1.184
10.381	37.768	1.664	10.895	49.378	2.312
3.634	12.678	0.864	4.703	10.962	0.440
8.490	34.783	1.889	7.795	22.096	0.937
8.983	31.576	1.465	10.201	42.808	2.141
10.560	13.392	0.052	9.246	10.641	-0.030
10.233	35.301	1.504	8.784	41.355	2.235
7.777	22.727	1.173	6.092	27.513	1.935
5.219	22.724	1.721	6.482	21.869	1.208
8.229	36.020	1.996	7.518	37.750	2.351
8.778	24.202	0.981	8.139	29.052	1.469
10.258	19.529	0.485	10.612	20.571	0.546

Table S-3: Retention parameters for 50 compounds sampled using strategy C

${}^1\ln k_0$	${}^1S_1$	${}^1S_2$	${}^2\ln k_0$	${}^2S_1$	${}^2S_2$
10.130	43.197	1.939	6.141	35.912	1.744
4.484	8.072	0.001	11.761	57.255	2.223
10.011	45.336	1.902	2.812	9.082	0.188
2.128	16.434	0.754	12.558	61.933	2.158
8.350	28.435	1.546	4.141	19.866	1.242
10.803	58.441	2.393	7.588	35.653	2.033
9.454	57.072	1.994	7.942	30.867	1.530
10.730	56.890	2.435	4.327	28.359	1.307

8.863	30.792	1.352	8.862	41.448	1.831
4.909	29.524	1.317	3.781	18.820	1.351
4.399	10.772	0.444	12.293	45.174	2.250
13.354	59.989	2.703	13.621	50.850	2.376
8.756	39.217	2.102	10.120	40.970	1.915
6.553	30.128	1.517	3.623	14.722	0.843
8.477	29.524	1.547	6.894	32.415	1.602
5.157	22.630	1.252	5.154	8.330	-0.062
14.554	57.133	2.055	8.179	33.672	1.898
7.003	38.270	1.662	4.150	10.981	0.656
9.286	36.340	2.059	9.549	45.613	1.930
3.271	19.375	1.267	13.655	58.927	2.081
1.179	11.987	0.296	7.507	42.195	2.297
14.973	59.299	2.481	2.397	17.384	0.732
7.511	35.854	1.548	8.566	26.360	1.160
14.495	59.782	2.564	7.721	44.687	1.771
10.308	53.240	2.343	11.173	50.624	2.323
3.006	9.555	0.171	9.390	47.812	2.392
6.498	21.111	1.262	6.041	19.755	0.886
7.638	27.791	1.392	6.994	38.930	1.987
9.034	32.144	1.508	7.930	41.505	2.216
7.585	24.167	1.562	4.610	14.979	0.581
6.415	14.522	0.765	10.008	59.589	2.524
3.306	24.816	1.107	1.786	10.491	0.294
11.232	56.159	2.148	3.139	9.393	0.510
6.252	36.474	1.771	1.789	16.635	1.090
7.088	17.823	1.188	7.970	21.140	1.444
6.471	20.924	1.486	11.534	56.713	2.299
6.775	30.193	1.690	14.826	57.252	2.234
8.742	36.089	1.708	8.263	36.337	2.105
12.615	50.604	2.480	9.431	37.890	1.738
5.629	17.208	1.144	12.682	50.448	1.938
4.656	24.711	1.501	11.780	46.883	1.891
10.408	46.806	2.363	5.384	32.337	1.776
8.585	30.167	1.579	3.805	16.348	0.879
6.651	40.083	1.906	5.479	16.932	1.304

6.533	27.135	1.724	1.304	10.471	0.431
7.511	36.258	1.827	12.264	44.806	2.055
12.089	59.578	2.574	9.307	43.136	2.076
3.945	19.474	1.021	12.183	58.006	2.517
14.742	59.666	2.692	0.308	7.962	0.450
10.605	52.350	2.126	8.480	27.782	1.324

Table S-4: Retention parameters for 50 compounds sampled using strategy D

${}^1\ln k_0$	${}^1S_1$	${}^1S_2$	${}^2\ln k_0$	${}^2S_1$	${}^2S_2$
4.682	9.651	0.566	7.361	26.769	1.182
8.454	43.403	1.994	14.002	61.706	2.101
5.127	12.166	0.703	4.533	11.402	0.611
7.459	28.303	1.486	2.985	10.697	0.495
5.920	21.926	1.345	10.377	36.506	1.602
5.288	11.350	0.534	7.638	21.438	1.413
6.659	21.863	1.513	10.887	40.725	1.668
3.525	20.830	1.548	8.046	40.569	1.972
4.742	8.074	0.310	6.451	22.510	1.057
9.827	47.300	2.176	9.404	54.794	2.013
4.869	16.016	1.231	8.349	23.438	1.272
12.086	50.506	2.388	9.209	49.261	2.094
10.968	49.436	1.868	11.776	46.238	2.368
3.815	18.668	1.190	4.769	8.408	0.513
3.511	12.144	0.664	3.726	26.776	1.569
7.957	23.146	1.132	6.973	18.558	1.264
9.321	40.853	2.242	4.765	22.973	1.393
4.459	11.045	0.294	6.264	30.008	1.554
7.008	19.961	1.158	2.821	16.084	1.031
4.603	11.091	0.823	5.066	17.301	0.918
8.016	44.598	2.072	11.448	49.636	2.361
14.015	53.180	1.999	11.314	47.224	2.276
14.741	59.557	2.243	9.822	52.461	2.075
10.399	58.852	2.304	11.317	48.707	2.333
5.711	12.330	0.741	7.043	27.311	1.451
5.104	23.264	1.431	8.535	35.019	1.819

10.323	50.755	2.394	11.464	39.764	2.038
12.257	52.256	2.259	9.388	50.738	2.296
6.338	36.747	2.050	9.299	50.851	1.856
8.508	48.059	2.242	10.661	52.652	2.426
7.842	34.847	1.438	6.846	18.535	1.358
4.446	14.410	0.856	2.852	11.106	0.489
6.780	17.870	1.050	4.810	11.779	0.743
6.453	40.832	1.918	9.453	38.155	1.698
6.857	18.603	1.368	7.941	37.119	1.759
13.800	54.733	2.457	13.000	50.699	2.153
6.232	30.371	1.748	9.728	30.058	1.643
8.319	38.896	1.671	11.851	45.904	2.134
9.379	49.256	2.456	15.256	62.422	2.129
2.889	13.971	0.602	7.453	18.746	0.997
9.711	39.012	1.688	6.921	19.483	1.088
2.997	22.597	1.338	9.022	36.982	1.733
7.319	34.127	1.458	3.774	16.842	1.076
6.024	30.971	1.359	5.211	18.082	1.237
11.211	61.072	2.209	9.190	47.130	2.327
8.380	48.230	2.309	10.722	48.563	2.187
3.005	10.097	0.291	4.619	20.632	1.079
6.268	13.442	0.613	5.093	11.197	0.730
9.675	45.173	2.302	15.382	61.841	2.524
5.606	17.521	1.041	1.927	17.339	0.858