Supplementary Information for

一种新多肽表征方法及支持向量机用于肽 HPLC 定量结构-保留建模预测

梁桂兆 [1, 2]　李志良 [1, 2*]　周　原 [1, 2]　何　留 [1]　周　鹏 [1]

([1] 重庆大学化学化工学院；[2] 重庆大学生物力学与组织工程教育部重点实验室，重庆 400030)

A New Peptide Sequences Representation Technique and Support Vector Machine for Quantitative Structure-Retention Modeling of Peptides in HPLC

LIANG, Gui-Zhao[1, 2]　LI, Zhi-Liang[1, 2*]　ZHOU, Yuan[1, 2]　He, Liu[1]　ZHOU, Peng[1]

([1]College of Chemistry and Chemical Engineering, Chongqing University, Chongqing 400030, P. R. China;

[2]Key Laboratory of Biomechanics and Tissue Engineering, MOE, Chongqing University, Chongqing 400030, P, R. China)

[*]Correspondent, E-mail: zlli2662@163.com; Tel: 023-65106677

附表1　71种肽的序列及实验与预测值

Table s1.　Sequences of 71 peptides with observed and predicted activities

No	Sequences	lg*k*		
		Observed	PLS	RSVM
1	AS	−0.6990	−0.7156	−0.499
2	GV	−0.6021	−0.7071	−0.4021
3	LG	−0.5229	−0.3652	−0.3229
4	APK	−0.5229	−0.7148	−0.3229
5	PGK	−0.5229	−0.3136	−0.3229
6	PSK	−0.4559	−0.0213	−0.2559
7	QSNNK	−0.4559	0.3814	−0.2559
8	AV	−0.4437	−0.1739	−0.2459
9	PF	0.2253	0.2944	0.4253
10	VDIK	0.4654	0.3308	0.6654
11	TPGSP	0.5132	0.1953	0.3132
12	ADSSPVK	0.5922	0.6218	0.7922
13	AGVETTK	0.6180	0.6531	0.7656
14	KNSISPE	0.6335	0.8138	0.7024
15	AGVETTTPSK	0.6415	0.5428	0.4415
16	VTALSQPK	0.6794	0.7139	0.8794
17	DGDGTITTK	0.6875	0.8251	0.8875
18	GF	0.7292	0.8182	0.5292
19	VFDK	0.7679	0.7039	0.5679
20	YINEHK	0.8669	1.3992	1.0669
21	ADYEK	0.8704	0.9449	0.6704
22	LTVLGQPK	0.8733	0.5740	0.6733
23	EAFR	0.8887	0.8618	0.8770
24	RVY	0.9196	1.1398	0.8613
25	LY	0.9243	0.9941	1.1243
26	DGHAHSHLIQQHIEK	0.9400	0.8792	0.7415
27	VDNALQSGNSQESVTEQDSK	0.9479	0.8822	0.7479
28	HGLDNYR	0.9881	0.7129	0.7881
29	ISRGQHKYEPE	0.9983	0.9036	0.7983
30	HVLFGGGTK	0.9987	0.9709	0.9518
31	KLSGHIYE	1.0174	0.8943	0.9416
32	VQWK	1.0504	1.2962	1.2504
33	AVRYINE	1.0542	0.9075	0.8604
34	DSTYSLSSTLTLSK	1.0704	0.6802	0.8704
35	DTDSEEEIR	1.0730	1.3191	0.8730
36	ANPTVTLFPPSSEELQANK	1.0917	1.0263	0.8917
37	ANPSVTLFPPSSEELQANK	1.0924	1.0011	0.8933
38	VHVIFNYK	1.1021	0.9666	0.9888
39	GW	1.1106	1.0251	0.9106

Continued

No	Sequences	lg*k*		
		Observed	PLS	RSVM
40	IHPF	1.1238	1.1257	1.0030
41	WKPRQIDNPE	1.1483	1.0962	1.3381
42	DPTVYFK	1.1670	1.142	0.9670
43	NTDGSTDYGILQINSR	1.1709	0.8294	0.9709
44	DRVYIHPFHL	1.1775	1.3499	1.1421
45	HHQEHPYTAGE	1.1844	1.2033	0.9844
46	VKIDNSQVE	1.1853	1.1107	1.1608
47	LLISDNYNRPSGVPARFSGSK	1.2122	0.7835	1.0122
48	GTDVQAWIR	1.2460	0.975	1.0460
49	DRVYIHPF	1.2943	1.3619	1.0943
50	VFDKDGDGYISAAELR	1.3610	1.2370	1.5610
51	LI	−0.6990	−0.5784	−0.6009
52	HK	−0.5229	1.1209	−0.6073
53	MK	0.4048	0.4458	0.4381
54	HAVE	0.5490	0.8994	0.8334
55	ADGSPVK	0.6415	0.6403	1.0377
56	HASLEKPKDE	0.7059	0.3692	0.7291
57	TFKRD	0.8129	0.8638	0.6811
58	LTVLRQPK	0.8274	0.6891	0.6629
59	VFGR	0.8876	0.7576	0.8521
60	YVLNKHNE	0.9106	0.7497	0.8023
61	ELGTVMR	0.9415	0.8687	0.7346
62	VPVVFVKKE	0.9567	0.6741	0.7070
63	VKDGHAHSHLIQQHIE	0.9590	0.7984	0.8633
64	TWGVTKAAELQ	1.0183	0.9080	0.9933
65	GQTLVVQFTVK	1.0770	0.9736	1.2640
66	TWGVTKAAE	1.0927	0.7809	1.3491
67	AKNWADD	1.1193	0.5813	1.4224
68	EAFSLFDKDGDGTITTK	1.1709	0.8201	1.0193
69	NTDGSTDYGILQIN	1.1878	0.7998	1.5460
70	RTVAAPSVFIFPPSDEQLK	1.2737	0.8789	1.3951
71	RVYIHPE	1.3705	1.3529	1.3759

附表 2　正交实验表挑选 SVM 参数结果

Table s2　Features and results of orthogonal design for SVM

No	Factor1(C)	Factor2(ε)	Factor3(γ)	Q^2_{ext}
1	1(100)	1(0.1)	1(0.001)	0.830
2	1(100)	2(0.2)	2(0.002)	0.832
3	1(100)	3(0.3)	3(0.003)	0.799
4	1(100)	4(0.4)	4(0.004)	0.794
5	2(200)	1(0.1)	2(0.002)	0.802
6	2(200)	2(0.2)	1(0.001)	0.815
7	2(200)	3(0.3)	4(0.004)	0.768
8	2(200)	4(0.4)	3(0.003)	0.839
9	3(300)	1(0.1)	3(0.003)	0.798
10	3(300)	2(0.2)	4(0.004)	0.857
11	3(300)	3(0.3)	1(0.001)	0.838
12	3(300)	4(0.4)	2(0.002)	0.850
13	4(400)	1(0.1)	4(0.004)	0.776
14	4(400)	2(0.2)	3(0.003)	0.796
15	4(400)	3(0.3)	2(0.002)	0.747
16	4(400)	4(0.4)	1(0.001)	0.803