

基于逐步非线性回归的血管紧张素转化酶抑制肽 QSAR 建模

周 恒 巴庆芳 袁哲明 代志军

附表 1 141 个 ACE 抑制二肽的序列与观察活性值

Table S1 Sequences and activities of 141 ACE-inhibitory dipeptides

No.	Pep.	Obsd.									
1	AA	2.25	37	GM	3.15	73	MY	2.29	109	WM	1.98
2	AF	1.83	38	GP	2.63	74	NF	1.67	110	WV	2.70
3	AG	3.40	39	GQ	3.74	75	NG	4.08	111	WY	1.58
4	AP	2.40	40	GR	3.51	76	NK	2.91	112	YA	2.66
5	AR	1.98	41	GS	3.58	77	NP	3.36	113	YE	2.80
6	AV	2.90	42	GT	3.76	78	NY	1.51	114	YG	3.18
7	AW	1.08	43	GV	3.28	79	PG	4.23	115	YH	0.71
8	AY	1.61	44	GW	1.48	80	PR	0.61	116	YK	2.79
9	DA	3.58	45	GY	2.26	81	QG	3.94	117	YL	1.91
10	DF	2.56	46	HG	3.80	82	QK	2.95	118	YN	1.71
11	DG	2.62	47	HL	3.51	83	RA	2.66	119	YP	2.91
12	DL	3.30	48	HY	1.42	84	RF	2.17	120	YV	2.76
13	DM	2.78	49	IA	2.18	85	RG	3.08	121	YW	0.97
14	DW	1.11	50	IF	2.97	86	RL	3.39	122	CF	0.30
15	DY	2.00	51	IG	3.08	87	RP	1.91	123	AI	0.53
16	EA	4.00	52	IL	1.74	88	RR	2.43	124	FE	0.16
17	EG	3.94	53	IP	2.11	89	RW	1.27	125	WW	1.91
18	EP	3.08	54	IR	2.88	90	RY	1.37	126	AH	2.72
19	ER	2.82	55	IW	0.67	91	SF	2.11	127	DP	0.33
20	EY	0.43	56	IY	0.48	92	SG	3.93	128	FK	2.42
21	FG	3.57	57	KA	2.04	93	SY	1.82	129	QN	1.95
22	FL	1.20	58	KF	1.76	94	TF	1.26	130	ST	0.61
23	FP	2.50	59	KG	3.51	95	TG	4.00	131	VL	1.11
24	FQ	1.71	60	KL	1.70	96	TK	3.21	132	WP	2.34
25	FR	2.96	61	KP	1.50	97	TP	2.89	133	WT	2.80
26	FW	0.77	62	KR	2.58	98	VF	1.64	134	YS	2.61
27	FY	0.81	63	KW	0.89	99	VG	3.04	135	DK	2.71
28	GA	3.30	64	KY	1.00	100	VK	1.11	136	MA	2.18
29	GD	3.96	65	LA	2.49	101	VP	2.69	137	MP	2.13
30	GE	3.79	66	LF	2.47	102	VQ	3.11	138	PL	2.53
31	GF	2.75	67	LG	3.94	103	VR	2.40	139	QP	1.82
32	GG	3.90	68	LW	1.24	104	VW	0.49	140	RS	1.79
33	GH	3.49	69	LY	1.26	105	VY	1.34	141	WH	0.71
34	GI	3.10	70	MF	1.65	106	WA	2.44			
35	GK	3.73	71	MG	3.68	107	WG	3.77			
36	GL	3.40	72	MW	1.23	108	WL	1.65			

附表 2 150 个 ACE 抑制三肽的序列与观察活性值

Table S2 Sequences and activities of 150 ACE-inhibitory tripeptides

No.	Pep.	Obsd.									
1	AAP	4.52	39	GPM	4.77	77	LIY	6.09	115	RPG	2.91
2	ADA	3.83	40	GPP	4.92	78	LKA	5.07	116	RPP	4.22
3	AEL	4.24	41	GQP	5.49	79	LKP	6.02	117	RRR	4.23
4	AFL	4.20	42	GRP	4.70	80	LKY	6.11	118	SVY	5.09
5	AGP	3.25	43	GSH	4.49	81	LLF	4.10	119	TNP	3.68
6	ALP	3.62	44	GVV	4.18	82	LLL	4.65	120	VAA	4.89
7	AQL	4.24	45	GYG	3.67	83	LLP	4.80	121	VAF	4.45
8	AVP	3.47	46	GYY	4.93	84	LPF	4.40	122	VGP	4.58
9	DLP	5.32	47	HIR	3.02	85	LPP	5.02	123	VIY	5.12
10	FAL	4.58	48	HLL	4.24	86	LRP	6.21	124	VLP	4.09
11	FCF	4.96	49	HQG	3.13	87	LQP	5.83	125	VLY	4.51
12	FDK	3.41	50	IAE	4.46	88	LQW	5.42	126	VPP	5.05
13	FEP	4.92	51	IAP	5.57	89	LSA	5.11	127	VQV	5.06
14	FFF	4.80	52	IAQ	4.46	90	LSP	5.77	128	VRP	5.66
15	FFG	3.29	53	IFL	4.35	91	LTf	5.56	129	VSP	5.00
16	FFL	4.43	54	IKP	5.68	92	LVL	5.19	130	VSW	4.63
17	FFP	4.92	55	IKY	6.68	93	LVQ	4.85	131	VTR	3.87
18	FGF	4.71	56	ILP	4.49	94	LVR	4.85	132	VVF	4.45
19	FGG	3.21	57	IMY	5.74	95	LVY	5.74	133	VVV	4.37
20	FGK	3.80	58	IPA	3.85	96	LWA	4.90	134	VWY	5.03
21	FIV	3.96	59	IPP	5.30	97	LWY	5.30	135	VYP	3.83
22	FNF	5.16	60	IRA	5.01	98	LYP	5.18	136	YPF	4.40
23	FPF	4.68	61	IRP	5.74	99	MNP	4.18	137	YPR	4.78
24	FPK	3.55	62	ITF	4.31	100	PFP	4.26	138	YYY	4.46
25	FPP	4.50	63	IVQ	4.02	101	PGI	3.77	139	AMY	5.26
26	FQP	4.92	64	IVY	5.84	102	PGG	2.86	140	FAP	5.42
27	FWN	4.74	65	IWH	5.46	103	PGL	4.86	141	GGY	5.89
28	FYN	4.74	66	IYP	4.21	104	PGP	4.18	142	HHL	5.27
29	GEG	3.72	67	KPF	4.49	105	PGR	3.33	143	IKW	6.68
30	GFF	4.98	68	LAA	4.89	106	PIP	4.31	144	LRY	6.82
31	GFG	3.47	69	LAP	5.73	107	PLW	4.44	145	MKY	5.14
32	GGF	4.89	70	LAY	5.41	108	PPG	2.82	146	PRY	5.60
33	GGG	3.39	71	LDP	4.37	109	PPP	4.14	147	RIY	4.55
34	GGP	4.72	72	LEE	4.00	110	PSY	4.80	148	TVY	4.82
35	GKV	5.41	73	LEL	4.81	111	PWP	3.66	149	VAP	5.70
36	GLG	3.55	74	LEP	5.24	112	PYP	3.66	150	YEY	5.40
37	GLY	5.05	75	LGI	4.54	113	RFH	3.48			
38	GPL	5.59	76	LGL	4.48	114	RGP	4.27			

附表 3

	V1	V2	V3	V4	V5	V6	V7	V8
Ala(A)	0.870	1.520	0.400	1.800	19.200	0.058	5.080	0.007
Arg(R)	0.850	1.520	0.300	12.500	74.770	0.085	4.750	0.044
Asn(N)	0.090	1.520	0.900	-5.600	42.180	0.091	5.750	0.005
Asp(D)	0.660	1.520	0.800	5.050	40.470	0.081	5.960	-0.024
Cys(C)	1.520	1.520	0.500	-16.500	40.470	0.128	2.950	-0.037
Gln(Q)	0.000	1.520	0.700	6.300	52.130	0.098	4.240	0.049
Glu(E)	0.670	1.520	1.300	12.000	51.790	0.064	6.040	0.007
Gly(G)	0.100	1.000	0.000	0.000	0.000	0.152	8.200	0.179
His(H)	0.870	1.520	1.000	-38.500	55.560	0.054	2.100	-0.011
Ile(I)	3.150	1.900	0.400	12.400	56.590	0.056	4.950	0.022
Leu(L)	2.170	1.520	0.600	-11.000	57.620	0.070	8.030	0.052
Lys(K)	1.640	1.520	0.400	14.600	65.850	0.095	4.930	0.018
Met(M)	1.670	1.520	0.300	-10.000	64.140	0.055	2.610	0.003
Phe(F)	2.870	1.520	0.700	-34.500	79.910	0.065	4.360	0.038
Pro(P)	2.770	1.520	0.900	-86.200	43.560	0.068	4.840	0.240
Ser(S)	0.070	1.520	0.400	-7.500	25.030	0.106	6.410	0.005
Thr(T)	0.070	1.730	0.400	-28.000	38.410	0.079	5.870	0.003
Trp(W)	3.770	1.520	0.600	-33.700	100.150	0.167	2.310	0.038
Tyr(Y)	2.670	1.520	1.200	-10.000	88.830	0.125	4.550	0.024
Val(V)	1.870	1.900	0.400	5.630	43.900	0.053	6.070	0.057

The G-scale descriptor includes 8 properties of amino acid, i.e., hydrophobicity (V1), STERIMOL minimum width of the side chain (V2), loss of side chain hydrophathy by helix formation (V3), optical rotation (V4), side chain molecular volume (V5), frequency of the 4th residue in turn (V6), AA composition of EXT of multi-spanning proteins (V7), and net charge index (V8), which are selected by stepwise multiple regression (SMR) on the original 457 physicochemical properties classified into three sorts of parameters including hydrophobic, steric, and electric properties.

附表 4 ACE 抑制二肽的预测与拟合精度

Table S4 Prediction and fitting accuracies on ACE-inhibitory dipeptides

FS Method	Descriptor	SVR				PLSR				MLR						
		RMSE _{pred}	R ² _{pred}	Q ² _{pred}	RMSE _{fit}	R ² _{fit}	RMSE _{pred}	R ² _{pred}	Q ² _{pred}	RMSE _{fit}	R ² _{fit}	RMSE _{pred}	R ² _{pred}	Q ² _{pred}	RMSE _{fit}	R ² _{fit}
SSNR	DPPS	0.5740	0.6547	0.6548	0.5387	0.7030	0.5477	0.6856	0.6857	0.6791	0.5281	0.5518	0.6810	0.6810	0.6770	0.5310
	G-scale	0.5946	0.6295	0.6295	0.5502	0.6837	0.6224	0.5940	0.5940	0.5788	0.6500	0.6046	0.6169	0.6169	0.5639	0.6677
	ST-scale	0.6288	0.5856	0.5857	0.5796	0.6740	0.6147	0.6041	0.6051	0.7107	0.5098	0.6361	0.5760	0.5770	0.6954	0.5307
	VHSE	0.5505	0.6824	0.6827	0.6100	0.6265	0.5680	0.6619	0.6622	0.6552	0.5691	0.6389	0.5722	0.5722	0.6156	0.6196
	VSW	0.6897	0.5014	0.5026	0.6499	0.5876	0.6718	0.5270	0.5282	0.6585	0.5765	0.6774	0.5191	0.5202	0.6459	0.5926
	ICM	0.5571	0.6747	0.6748	—	—	0.5585	0.6732	0.6732	—	—	0.5564	0.6756	0.6757	—	—
GA-PLS	DPPS	0.9273	0.0988	0.0990	0.5391	0.7026	0.7111	0.4700	0.4702	0.7024	0.4950	0.7833	0.3570	0.3571	0.6886	0.5148
	G-scale	0.7623	0.3910	0.3910	0.6573	0.5486	0.6509	0.5560	0.5560	0.5893	0.6372	0.6206	0.5964	0.5964	0.5679	0.6630
	ST-scale	0.6222	0.5943	0.5954	0.6731	0.5602	0.6850	0.5082	0.5095	0.7318	0.4802	0.6850	0.5082	0.5095	0.7318	0.4802
	VHSE	0.6028	0.6192	0.6196	0.6808	0.5348	0.6375	0.5741	0.5745	0.6974	0.5119	0.6100	0.6101	0.6104	0.6927	0.5184
	VSW	0.7552	0.4022	0.4037	0.7300	0.4797	0.7878	0.3497	0.3512	0.7728	0.4168	0.7820	0.3591	0.3607	0.7666	0.4262
	ICM	0.7037	0.4810	0.4811	—	—	0.6450	0.5640	0.5641	—	—	0.6424	0.5675	0.5676	—	—
SLR	DPPS	0.5966	0.6269	0.6270	0.5504	0.6899	0.7086	0.4737	0.4737	0.6049	0.6255	0.7142	0.4654	0.4654	0.6048	0.6256
	G-scale	0.6477	0.5603	0.5603	0.5397	0.6956	0.6456	0.5632	0.5632	0.6006	0.6231	0.6566	0.5482	0.5482	0.5996	0.6244
	ST-scale	0.6276	0.5872	0.5883	0.6742	0.5589	0.6422	0.5677	0.5688	0.7003	0.5240	0.6422	0.5677	0.5688	0.7003	0.5240
	VHSE	0.6129	0.6063	0.6067	0.6307	0.6008	0.6384	0.5729	0.5733	0.6234	0.6100	0.6384	0.5729	0.5733	0.6234	0.6100
	VSW	0.6684	0.5317	0.5317	0.6614	0.5728	0.7703	0.3781	0.3796	0.6719	0.5592	0.7655	0.3859	0.3874	0.6706	0.5609
	ICM	0.5934	0.6310	0.6310	—	—	0.6471	0.5612	0.5613	—	—	0.6493	0.5582	0.5582	—	—

附表 5 ACE 抑制三肽的预测与拟合精度

Table S5 Prediction and fitting accuracies on ACE-inhibitory tripeptides

FS Method	Descriptor	SVR				PLSR				MLR						
		RMSE _{pred}	R ² _{pred}	Q ² _{pred}	RMSE _{fit}	R ² _{fit}	RMSE _{pred}	R ² _{pred}	Q ² _{pred}	RMSE _{fit}	R ² _{fit}	RMSE _{pred}	R ² _{pred}	Q ² _{pred}	RMSE _{fit}	R ² _{fit}
SSNR	DPPS	0.5889	0.4745	0.4745	0.2126	0.9291	0.4553	0.6858	0.6858	0.4649	0.6610	0.4548	0.6866	0.6866	0.4649	0.6610
	G-scale	0.5845	0.4822	0.4322	0.4367	0.7010	0.5834	0.4841	0.4841	0.5382	0.5457	0.6112	0.4340	0.4340	0.5294	0.5604
	ST-scale	0.5734	0.5017	0.5017	0.3285	0.8308	0.5442	0.5512	0.5512	0.5157	0.5829	0.5545	0.5341	0.5341	0.5121	0.5887
	VHSE	0.5041	0.6149	0.6149	0.3978	0.7518	0.4347	0.7136	0.7137	0.4372	0.7002	0.4254	0.7258	0.7258	0.4303	0.7096
	VSW	0.5266	0.5798	0.5798	0.3728	0.7821	0.5715	0.5051	0.5051	0.5024	0.6042	0.5598	0.5251	0.5251	0.4932	0.6185
	ICM	0.4819	0.6480	0.6480	—	—	0.4672	0.6692	0.6692	—	—	0.4652	0.6720	0.6720	—	—
GA-PLS	DPPS	0.5654	0.5155	0.5155	0.5014	0.6057	0.7011	0.2550	0.2550	0.5609	0.5065	0.6985	0.2607	0.2607	0.5608	0.5067
	G-scale	0.7544	0.1375	0.1375	0.6188	0.3995	0.7554	0.1352	0.1352	0.6186	0.3999	0.7541	0.1382	0.1382	0.6170	0.4029
	ST-scale	0.5285	0.5767	0.5767	0.4525	0.6789	0.5328	0.5698	0.5698	0.5372	0.5474	0.5328	0.5698	0.5698	0.5372	0.5474
	VHSE	0.4930	0.6317	0.6317	0.4650	0.6610	0.5115	0.6036	0.6036	0.4915	0.6212	0.5122	0.6025	0.6025	0.4915	0.6212
	VSW	0.5036	0.6157	0.6157	0.4131	0.7324	0.6096	0.4369	0.4369	0.4789	0.6404	0.6074	0.4409	0.4409	0.4784	0.6410
	ICM	0.4779	0.6538	0.6538	—	—	0.5305	0.5734	0.5734	—	—	0.5313	0.5722	0.5722	—	—
SLR	DPPS	0.6082	0.4395	0.4395	0.3283	0.8310	0.4923	0.6327	0.6327	0.4539	0.6768	0.4917	0.6336	0.6336	0.4636	0.6773
	G-scale	0.6035	0.4480	0.4480	0.4790	0.6401	0.5492	0.5429	0.5429	0.4840	0.6326	0.5491	0.5431	0.5431	0.4840	0.6326
	ST-scale	0.5479	0.5451	0.5451	0.4592	0.6693	0.5725	0.5033	0.5033	0.5318	0.5566	0.5725	0.5033	0.5033	0.5318	0.5566
	VHSE	0.4768	0.6555	0.6555	0.4284	0.7122	0.4714	0.6632	0.6632	0.4429	0.6923	0.4714	0.6632	0.6632	0.4429	0.6923
	VSW	0.5079	0.6090	0.6090	0.4674	0.6575	0.5506	0.5406	0.5406	0.4984	0.6105	0.5352	0.5660	0.5660	0.4972	0.6124
	ICM	0.4708	0.6641	0.6641	—	—	0.4591	0.6806	0.6806	—	—	0.4585	0.6814	0.6814	—	—